Particle–boson interactions and the weak coupling limit

E. B. Davies[a)]

St. John's College, Oxford, England
(Received 17 January 1978)

We consider the interaction of a finite number of nonrelativistic particles with a positive or zero mass quantum field. We show that in the weak coupling limit the quantum field gives rise to an effective interaction between the particles of a Yukawa or Coulomb type, as well as, in some cases, a mass renormalization. In a simple exactly soluble model we investigate the higher order radiation processes.

1. INTRODUCTION

We present a rigorous derivation of the nonrelativistic N-body Schrödinger Hamiltonian

$$H' = -\sum_{r=1}^{N} \frac{1}{2m_r} \Delta_r + \sum_{r<s} V(q_r - q_s)$$

for N quantum particles, starting from a quantum field theoretic Hamiltonian where the interaction between the particles depends upon their exchange of virtual mesons, or phonons. The Hamiltonian H' is obtained by taking the weak coupling limit between the particles and the meson field at the same time as the particle masses become infinite. These limits are closely related to the usual treatment, where one supposes the particle are static so that the dynamics of the quantum field becomes exactly soluble.

Although our model is similar to several others describing particles coupled to infinite reservoirs,[1–3] it has several significant variations. The first, that the asymptotic evolution of the particles is of Hamiltonian type, with no dissipative terms, is a consequence of the infinite mass of the particles and the zero temperature of the quantum field. Because of the absence of dissipative terms we can describe the time evolution at the Hilbert space level instead of on the space of trace class operators. Moreover, because we do not need an infinite order perturbation expansion it is possible to remove the volume cutoff on the interaction Hamiltonian.

In an earlier version of the present paper we consider only the case where the mesons have positive mass. The extension to zero mass field particles (photons or phonons) leads to mathematical problems caused by the infrared singularity. However, since we study the time evolution for finite times this singularity does not have effects as serious as in scattering theory,[4,5] we are now able to solve the case where the field particles are relativistic spinless bosons of positive or zero mass.

In the second part of the paper we study the emission of bosons by the particles, a process which depends on the zero mass of the bosons. By means of a detailed analysis of the Friedrichs's model we show that this process involves terms which are of higher than second order in the coupling constant λ. Non integer powers of the coupling constant are shown to arise because of the form of the infrared singularity.

For maximum generality we specify our model in an abstract form; some examples are described in Sec. 3. The particles are represented by a Hilbert space $\mathcal{H}$ with Hamiltonian H. The quantum field is described by a Hilbert space $\mathcal{J}$ with a vacuum state Ω and a Hamiltonian F such that $F\Omega = 0$. The interaction between the particles and the quantum field is represented by the operator A on $\mathcal{H} \otimes \mathcal{J}$. The total Hamiltonian of the system is then

$$H_\lambda = \lambda^{-2}(\lambda^2 H + \lambda A + F) \tag{1.1}$$

which we study in the weak coupling limit $\lambda \to 0$. The factor λ^2 on H may be interpreted as a mass scaling for the particles while the factor λ^{-2} on the whole Hamiltonian may be interpreted as a time scaling.

Our first goal is to find conditions on the various operators under which

$$\lim_{\lambda \to 0} \exp(-iH_\lambda t)(\psi \otimes \Omega) = (\exp[-i(H+K)t]\psi) \otimes \Omega \tag{1.2}$$

for all $\psi \in \mathcal{H}$, $t \in \mathbb{R}$, and some effective interaction operator K on $\mathcal{H}$. The conditions are given in terms of the operators appearing in the interaction Hamiltonian

$$A = \sum_{r=1}^{N} \int_{\mathbb{R}^3} \delta_{rx} \otimes \varphi_{rx} d^3x, \tag{1.3}$$

where δ_{rx} are operators on $\mathcal{H}$ localized at the point $x \in \mathbb{R}^3$ and φ_{rx} are field operators on $\mathcal{J}$ localized at the point $x \in \mathbb{R}^3$. Our technical assumptions on these operators require that ultraviolet cutoffs be incorporated into their definition.

2. THE LIMITING HAMILTONIAN

We suppose that $\mathcal{H}$ contains dense subspaces

$$\mathcal{H}_0 \subseteq \mathcal{H}_1 \subseteq \mathcal{H}_2 \subseteq \mathcal{H}_3 = \mathcal{H}$$

which are again Banach spaces with respect to norms $\|\cdot\|_i$ such that each injection in the above sequence is a contraction. We also suppose that $\mathcal{J}$ contains subspaces

$$\mathbb{C}\Omega = \mathcal{J}_0 \subseteq \mathcal{J}_1 \subseteq \mathcal{J}_2 \subseteq \mathcal{J}_3 = \mathcal{J},$$

which are again Banach spaces with respect to norms $\|\cdot\|_i$ such that each injection in the above sequence is a contraction. We identify $\mathcal{H}$ with the subspace $\mathcal{H} \otimes \Omega$ of $\mathcal{H} \bar{\otimes} \mathcal{J}$, define P_0 to be the orthogonal projection L of $\mathcal{H} \bar{\otimes} \mathcal{J}$ onto this subspace, and put $P_1 = 1 - P_0$.

The above spaces and norms are used to formulate our conditions on the various operators.

(A1) H is a bounded operator from $\mathcal{H}_2$ into $\mathcal{H}$ which is symmetric as an operator on $\mathcal{H}$.

(A2) F is a nonnegative self-adjoint operator on $\mathcal{J}$ such that $F\Omega = 0$. Moreover e^{-tF} are bounded operators on $\mathcal{J}_i$ depending strongly continuously on $t \geq 0$ for all i.

a)Permanent address. On leave in Mathematics Department, Princeton University. Work partly supported by NSF Grant No. MPS 75-11864A04.

 0022-2488/79/030345-07$01.00

(A3) δ_{rx} are bounded operators from H_i to H_{i+1} depending strongly continuously on $x \in \mathbb{R}^3$. Moreover

$$\int_{\mathbb{R}^3} \| \delta_{rx}\psi \|_{i+1} d^3x \leq c \| \psi \|_i$$

for $i = 0, 1, 2,$ and $1 \leq r \leq n$.

(A4) φ_{rx} are bounded operators from $\mathcal{F}_i$ to $\mathcal{F}_{i+1}$ depending strongly continuously on $x \in \mathbb{R}^3$. Moreover

$$\| \varphi_{rx}\psi \|_{i+1} \leq c \| \psi \|_i$$

for all $x \in \mathbb{R}^3$ and $i = 0, 1, 2$ and $1 \leq r \leq n$.

(A5) $\langle \varphi_{rx} \rangle_{\text{vac}} = 0$ for all $x \in \mathbb{R}^3$ and $1 \leq r \leq n$. For our next two conditions we suppose that β is a constant such that

$$0 < \beta < \frac{3 - \sqrt{5}}{2} \doteq 0.382. \tag{2.1}$$

(A6) $\| e^{-tF} \varphi_{rx} \Omega \|_1 \leq ct^{\beta-1}$ for all $t > 0$ and $x \in \mathbb{R}^3$ and $1 \leq r \leq n$.

(A7) $\| e^{-tF} \varphi_{rx} e^{-uF} \varphi_{sy} \Omega \|_2 \leq ct^{\beta-1}u^{\beta-1}$ for all $t, u > 0$ and $x, y \in \mathbb{R}^3$ and $1 \leq r, s \leq n$.

In order to make sense of the formal expression

$$A = \sum_{r=1}^{n} \int_{\mathbb{R}^3} \delta_{rx} \otimes \varphi_{rx} d^3x, \tag{2.2}$$

we introduce some subspaces of $H \bar{\otimes} \mathcal{F}$. Defining D_i to be the Banach space projective tensor product $H_i \bar{\otimes} \mathcal{F}_i$ we note that

$$D_0 \subseteq D_1 \subseteq D_2 \subseteq D_3 \subseteq H \bar{\otimes} \mathcal{F},$$

where each injection is a contraction. Moreover under the identification of H with $H \otimes \Omega$, D_0 is a dense subspace of H.

Lemma 2.1: Under conditions (A1)–(A7) the formula (2.2) defines a symmetric operator A on D_2 satisfying conditions (B1)–(B7) below.

Proof: If $\psi_1 \in H_i$ and $\psi_2 \in \mathcal{F}_i$, then

$$\left\| \sum_{r=1}^{n} \int_{\mathbb{R}^3} \| \delta_{rx} \otimes \varphi_{rx} d^3x \quad (\psi_1 \otimes \psi_2) \right\|_{i+1}$$

$$\leq \sum_{r=1}^{n} \int_{\mathbb{R}^3} \| \delta_{rx}\psi_1 \|_{i+1} \| \varphi_{rx}\psi_2 \|_{i+1} d^3x$$

$$\leq nc^2 \| \psi_1 \|_i \| \psi_2 \|_i$$

by (A3) and (A4). The definition of the projective tensor product norm leads to

$$\| A\psi \|_{i+1} \leq nc^2 \| \psi \|_i$$

for all $\psi \in D_i$, which proves (B3).

To prove (B4) we note that by Eq. (2.1)

$$0 > (\beta - 1) > -1 > 2(\beta - 1).$$

Since

$$| \langle \varphi_{sy} e^{-tF} \varphi_{rx} \rangle_{\text{vac}} | \leq \| e^{-tF} \varphi_{rx} \Omega \| \, \| \varphi_{sy} \Omega \| \leq c^2 t^{\beta-1}$$

and

$$| \langle \varphi_{sy} e^{-tF} \varphi_{rx} \rangle_{\text{vac}} | \leq e^{-tF/2} \phi_{sy} \Omega \| \, \| e^{-tF/2} \varphi_{rx} \Omega \| \leq c' t^{2(\beta-1)},$$

it follows that

$$\int_0^\infty | \langle \varphi_{sy} e^{-tF} \varphi_{rx} \rangle_{\text{vac}} | dt \leq c'' < \infty \tag{2.3}$$

for some c'' independent of $x, y \in \mathbb{R}^3$ and of r, s. (B4) now follows from the expression

$$K_\mu \psi = - \sum_{r,s=1}^{n} \int_{\mathbb{R}^6} \int_0^\infty e^{-t\mu} \delta_{sy} \delta_{rx} \langle \varphi_{sy} e^{-tF} \varphi_{rx} \rangle_{\text{vac}} \times dt \, d^3x \, d^3y \tag{2.4}$$

whose convergence for all $\psi \in D_0$ and as $\mu \downarrow 0$ is a consequence of (A3).

(B5) is an obvious consequence of (A5). Since (B6) and (B7) are proved in a similar fashion from (A6) and (A7), respectively, we only treat the first one. Noting that

$$\psi_1 = \sum_{r=1}^{n} \int_{\mathbb{R}^3} \int_0^\infty e^{-t\mu} e^{-tF} (\delta_{rx} \otimes \varphi_{sx})(\psi \otimes \Omega) dt \, d^3x,$$

we obtain

$$\| \psi_1 \|_1 \leq \sum_{r=1}^{n} \int_{\mathbb{R}^3} \int_0^\infty e^{-t\mu} \| e^{-tF} \varphi_{rx} \Omega \|_1 \| \delta_{sx} \psi \|_1 dt \, d^3x$$

$$\leq nc^2 \| \psi \|_0 \int_0^\infty e^{-t\mu} t^{\beta-1} dt$$

by (A3) and (A6), whence

$$\| \psi_1 \|_1 \leq nc^2 \| \psi \|_0 \Gamma(\beta) \mu^{-\beta},$$

as required.

The conditions (B1)–(B7) are:

(B1) H is a bounded operator from H_2 into H which is symmetric as an operator on H.

(B2) F is a nonnegative self-adjoint operator on $\mathcal{F}$ such that $F\Omega = 0$.

(B3) The operator A on D_2 is bounded from D_i to D_{i+1} for $i = 0, 1, 2$.

(B4) $K_\mu = -P_0 A (F + \mu)^{-1} A P_0$ are bounded operators from D_0 into H for all $\mu > 0$. Moreover

$$\lim_{\mu \downarrow 0} K_\mu \psi = K\psi$$

exists for all $\psi \in D_0$.

(B5) $\langle A\varphi, \psi \rangle = 0$ for all $\varphi, \psi \in D_0$.

(B6) If $\psi \in D_0$ and $\mu > 0$, then

$$\psi_1 = (F + \mu)^{-1} A \psi$$

lies in D_1 and satisfies

$$\| \psi_1 \|_1 \leq c\mu^{-\beta} \| \psi \|_0.$$

(B7) If $\psi \in D_0$ and $\mu, \nu > 0$, then

$$\psi_2 = (F + \mu)^{-1} P_1 A (F + \nu)^{-1} A \psi$$

lies in D_2 and satisfies

$$\| \psi_2 \|_2 \leq c\mu^{-\beta}\nu^{-\beta} \| \psi \|_0.$$

Our last two conditions, which do not follow from (A1)–(A7), are

(B8) The operator $(H + K)$ on D_0 is essentially self-adjoint as an operator on H.

(B9) There are self-adjoint operators H_λ on $H \bar{\otimes} \mathcal{F}$ such that

$$H_\lambda \psi = H\psi + \lambda^{-1} A\psi + \lambda^{-2} F\psi \tag{2.5}$$

for all ψ in $D_2 \cap \text{dom} F$.

Before proving our main theorem we make some comments on the above conditions.

(i) Concerning (B9) we point out that it is not necessary for H_λ to be essentially self-adjoint on $\mathcal{J}_2 \cap \mathrm{dom} F$. Moreover it is not relevant for our purposes to enquire as to which of the self-adjoint extensions is the physically "correct" one.[5] The existence of a self-adjoint extension is necessary, but this is a fairly weak condition as von Neumann's condition (Ref. 6, p. 143) shows.

(ii) Condition (B8) is of course satisfied if K is norm bounded and H is essentially self-adjoint on $\mathcal{D}_0$. However the condition in its present form allows us to deal with much more singular perturbations K. In order to verify (B8) one may use the expression

$$K = -\sum_{r,s=1}^{n} \int_{\mathbb{R}^6} \delta_{sy}\delta_{rx} \langle \varphi_{sy} F^{-1} \varphi_{rx} \rangle_{\mathrm{vac}} d^3x\, d^3y \tag{2.6}$$

deduced from Eq. (2.4).

(iii) Ignoring technicalities our only important conditions are (A5)—(A7), which are essentially estimates on the truncated Euclidean n-point functions for $n \leq 4$. If the field particles have positive mass m the quantities $t^{\beta-1}$ may be replaced by e^{-mt}.

(iv) The peculiar restriction (2.1) on the range of values of β is surely an artifact of our proof. We conjecture that any β with $0 < \beta < \frac{1}{2}$ will do. These bounds are necessary for our proof of Eq. (2.3) which in turn is needed to prove the existence of the operator K in (B4).

(v) (A5) can actually be deduced from (A6) by taking the limit $t \to +\infty$ and using (A2).

Theorem 2.2: Under conditions (B1)—(B9) we have

$$\lim_{\lambda \to 0} \exp(-iH_\lambda t)\psi = \exp(-i(H+K)t)\psi \tag{2.7}$$

for all $\psi \in \mathcal{H}$ and $t \in \mathbb{R}$.

Proof: This is yet another variation on calculations in Refs. 7 and 8. It is sufficient to show that for all $\psi_0 \in \mathcal{D}_0$ there exist $\psi_\lambda \in \mathcal{D}_2 \cap \mathrm{dom} F$ such that

$$\lim_{\lambda \to 0} \psi_\lambda = \psi_0, \quad \lim_{\lambda \to 0} H_\lambda \psi_\lambda = (H+K)\psi_0. \tag{2.8}$$

We take

$$\psi_\lambda = \psi_0 + \psi_1 + \psi_2$$

where

$$\psi_1 \in \mathcal{D}_1 \cap \mathrm{dom} F, \quad \psi_2 = \mathcal{D}_2 \cap \mathrm{dom} F$$

are specified precisely below.

We have to estimate

$$\begin{aligned}\xi_\lambda &= (H + \lambda^{-1}A + \lambda^{-2}F)(\psi_0 + \psi_1 + \psi_2) - (H+K)\psi_0 \\ &= \{\lambda^{-1}A\psi_0 + \lambda^{-2}F\psi_1 + \lambda^{-2\gamma}\psi_1\} \\ &\quad + \{\lambda^{-1}P_0 A\psi_1 - K\psi_0\} \\ &\quad + \{\lambda^{-1}P_1 A\psi_1 + \lambda^{-2}F\psi_2 + \lambda^{-2+\delta}\psi_2\} \\ &\quad + H\psi_1 - \lambda^{-2\gamma}\psi_1 + H\psi_2 - \lambda^{-2+\delta}\psi_2 + \lambda^{-1}A\psi_2\end{aligned}$$

where γ, δ are to be specified.

We eliminate the first term by putting

$$\psi_1 = -\lambda(F + \lambda^\gamma)^{-1} A\psi_0$$

so that $\psi_1 \in \mathcal{D}_1 \cap \mathrm{dom} F$ and

$$\|\psi_1\|_1 \leq c\,\lambda^{1-\beta\gamma}$$

by (B6). Hence

$$\lim_{\lambda \to 0} \{\|\psi_1\| + \|H\psi_1\| + \|\lambda^{-2+\gamma}\psi_1\|\} = 0$$

provided

$$1 - \beta\gamma > 0 \tag{2.9}$$

and

$$-1 - \beta\gamma + \gamma > 0. \tag{2.10}$$

According to (B4) the second term converges to zero as $\gamma \to 0$ provided

$$\gamma > 0. \tag{2.11}$$

We eliminate the third term by putting

$$\begin{aligned}\psi_2 &= -\lambda(F + \lambda^\delta)^{-1} P_1 A\psi_1 \\ &= \lambda^2 (F + \lambda^\delta)^{-1} P_1 A (F + \lambda^\delta)^{-1} A\psi_0\end{aligned}$$

so that $\psi_2 \in \mathcal{D}_2 \cap \mathrm{dom}\, F$ and

$$\|\psi_2\|_2 \leq c\lambda^{2-\beta\gamma-\beta\delta}$$

by (B7). Hence

$$\lim_{\lambda \to 0} \{\|\psi_2\| + \|H\psi_2\| + \|\lambda^{-2+\delta}\psi_2\| + \|\lambda^{-1}A\psi_2\|\} = 0$$

provided

$$2 - \beta\gamma - \beta\delta > 0 \tag{2.12}$$

and

$$\delta - \beta\gamma - \beta\delta > 0 \tag{2.13}$$

and

$$1 - \beta\gamma - \beta\delta > 0. \tag{2.14}$$

To prove that Eq. (2.8) is satisfied we therefore need only find γ, δ satisfying Eqs. (2.9)—(2.14).

The last three equations may be rewritten as

$$\frac{\beta\gamma}{1-\beta} < \delta < \frac{1-\beta\gamma}{\beta}$$

assuming $0 < \beta < 1$. This is soluble provided

$$\beta^2\gamma < (1-\beta)(1-\beta\gamma) \tag{2.15}$$

or

$$1 - \beta - \beta\gamma > 0$$

or

$$\gamma < \frac{1-\beta}{\beta}.$$

The final conditions on γ are therefore

$$\frac{1}{1-\beta} < \gamma < \frac{1-\beta}{\beta}$$

which are soluble provided

$$\beta < (1-\beta)^2. \tag{2.16}$$

This is equivalent to Eq. (2.1).

3. SOME EXAMPLES

A. Example 1

We show how to use the above theorems to derive the Yukawa or Coulomb interaction between N particles exchanging virtual bosons. The particles are taken to have Hamiltonians $H^r = -(1/2m_r)\Delta_r$ on the Hilbert spaces $\mathcal{H}^r = L^2(\mathbb{R}^3)$ and we put

$$H = \sum_{r=1}^{N} H^r$$

on the Hilbert space tensor product $\mathcal{H}$ of the $\mathcal{H}^r$. If μ is a C^∞ function of compact support on $\mathbb{R}^3$ we let M denote the bounded operator

$$(M_x\Psi)(y) = \mu(y-x)\Psi(y)$$

on $L^2(\mathbb{R}^3)$, this being considered as a regularization of the δ-function at x. We then define the bounded operators δ_x on $\mathcal{H}$ by

$$\delta_x = \sum_{r=1}^{N} M_x^r$$

If Q_i, P_j are the canonical position and momentum operators on $L^2(\mathbb{R}^3)$ with domains equal to the Schwartz space $\mathcal{S}$, then for any 6-tuple $\alpha = (\alpha_1, \ldots, \alpha_6)$ of non-negative integers we put

$$X_\alpha = Q_1^{\alpha_1} Q_2^{\alpha_2} Q_3^{\alpha_3} P_1^{\alpha_4} P_2^{\alpha_5} P_3^{\alpha_6}$$

and then define $\mathcal{H}_i^r \subseteq \mathcal{H}^r$ for $0 \leq i \leq 3$ to be the completion of $\mathcal{S}$ for the norm

$$\|\Psi\|_i = \sum_{|\alpha| \leq 15-5i} \|X_\alpha \Psi\|,$$

where $|\alpha| = \sum_{j=1}^{6} \alpha_j$. We define $\mathcal{H}_i \subseteq \mathcal{H}$ to be the Banach space projective tensor product

$$\mathcal{H}_i = \mathcal{H}_i^1 \hat{\otimes} \mathcal{H}_i^2 \hat{\otimes} \cdots \hat{\otimes} \mathcal{H}_i^N.$$

The truth of (A1) is a consequence of the trivial inequality

$$\|(P_1^2 + P_2^2 + P_3^2)\Psi\| \leq \sum_{|\alpha| \leq 5} \|X_\alpha \Psi\|.$$

Using the commutation relations

$$[P_r, M_x] = M_x', \quad [Q_s, M_x] = 0$$

to commute the operator M_x in $X_\alpha M_x \psi$ to the left-hand side, and then applying the estimate

$$\|M_x(1 + Q_1^2 + Q_2^2 + Q_3^2)^{-2}\| \leq c(1+x^2)^{-2}$$

one obtains

$$\|X_\alpha \delta_{rx}\psi\| \leq c(1+x^2)^{-2} \sum_{|\beta| \leq |\alpha|+5} \|X_\beta \psi\|.$$

This implies

$$\|\delta_{rx}\psi\|_{i+1} \leq c(1+x^2)^{-2}\|\psi\|_i$$

and hence (A3).

We define $\mathcal{J}$ to be the boson Fock space with single particle space $L^2(\mathbb{R}^3)$. F is defined to be the free Hamiltonian on $\mathcal{J}$ which acts on the single particle space according to

$$(F\psi)(k) = (m^2 + k^2)^{1/2}\psi(k),$$

where $m \geq 0$.

If $i = 0, 1, 2$, we define $\mathcal{J}_i$ to be the sum of the m-particle subspaces of $\mathcal{J}$ such that $0 \leq m \leq i$, with the usual Hilbert space norm, and observe that (A2) is obviously satisfied.

The field operators φ_x on $\mathcal{J}$ are defined by

$$\varphi_x = (2\pi)^{-3/2}[a^*(f_x) + a(f_x)],$$

where f_x is defined as a function on momentum space by

$$f_x(k) = f(k)e^{-ix\cdot k}$$

and the infrared and ultraviolet behaviors of f are controlled by the assumptions that $f \in L^2(\mathbb{R}^3)$ and

$$|f(k)| \leq c|k|^{-1/2} \tag{3.1}$$

for all $k \in \mathbb{R}^3$. Conditions (A4) and (A5) are immediate consequences of the action of the creation and annihilation operators on Fock space. Because the Hamiltonian F is free, (A6) and (A7) are immediate consequences of the following lemma; the presence of P_1 in (A7) is essential to ensure that only the creation parts of φ_x and φ_y have any effect.

Lemma 3.1: If $f \in L^2(\mathbb{R}^3)$ satisfies Eq. (3.1) and $f_{xt} \in L^2(\mathbb{R}^3)$ is defined for $t > 0$ and $x \in \mathbb{R}^3$ by

$$f_{xt}(k) = f(k)\exp[-ix\cdot k - t(m^2 + k^2)^{1/2}],$$

then for any β such that $0 \leq \beta \leq 1$ we have

$$\|f_{xt}\| \leq c_\beta t^{-\beta}. \tag{3.2}$$

Proof: The case $\beta = 0$ follows from

$$\|f_{xt}\| \leq \|f\|,$$

so we need only prove the other extreme case $\beta = 1$ and interpolate,

$$\begin{aligned}\|f_{xt}\|^2 &\leq \int_{\mathbb{R}^3} |f(k)|^2 e^{-2t|k|} d^3k \\ &\leq \int_{\mathbb{R}^3} c^2|k|^{-1} e^{-2t|k|} d^3k \\ &= 4\pi c^2 \int_0^\infty r e^{-2tr} dr \\ &= c' t^{-2}.\end{aligned}$$

We next compute the form of the operator K. If we define

$$\begin{aligned}W(x-y) &\equiv \langle \varphi_x F^{-1} \varphi_y \rangle_{\text{vac}} \\ &= \int_{\mathbb{R}^3} (m^2 + k^2)^{-1/2} |f(k)|^2 \exp[-i(x-y)\cdot k] d^3k,\end{aligned}$$

then our assumptions on f ensure that W is a bounded continuous function of $(x-y)$. If we remove the ultraviolet cutoff by putting

$$f(k) = (m^2 + k^2)^{-1/4},$$

then W equals the Yukawa potential, or the Coulomb potential if $m = 0$. In terms of W the operator K is given by Eq. (2.6), namely

$$K = -\sum_{r,s=1}^{N} \int_{\mathbb{R}^6} M_x^r M_y^s W(x-y) d^3x\, d^3y. \tag{3.3}$$

The terms with $r \neq s$ describe two-body potentials while the terms with $r = s$ are finite self-energy constants. The validity of condition (B8) is an immediate consequence of the boundedness of the operator K.

We do not verify (B9) but refer to Comment (i) before Theorem (2.2).

B. Example 2

In our second example we consider a more singular interaction where the particle self-energy may be described as a mass renormalization. In order to simplify the treatment as far as possible we consider a single particle with a first order derivative coupling to a three component phonon field. We use the notation of the last example as far as possible, simplified to the case of one particle. The operators δ_{rx} are defined for $1 \leq r \leq 3$ by

$$\delta_{rx} = P_r M_x + M_x P_r$$

and (A1) and (A3) are verified as before.

The boson field is represented by the Fock space $\mathcal{F}$ with single particle space $L^2(\mathbb{R}^3) \otimes \mathbb{C}^3$, with F being the free Hamiltonian which acts on each single-particle component according to

$$(F\psi)(k) = |k|\,\psi(k).$$

The field operators φ_{rx} are defined for $x \in \mathbb{R}^3$ and $1 \leq r \leq 3$ by

$$\varphi_{rx} = (2\pi)^{-3/2}\{a^*(f_{rx}) + a(f_{rx})\},$$

where f_{rx} lies in the rth component of $L^2(\mathbb{R}^3) \otimes \mathbb{C}^3$ but is otherwise independent of r and satisfies the conditions of Lemma 3.1. The interaction Hamiltonian is defined by Eq. (2.2) with $n = 3$. The verification of the axioms is exactly as for the previous example, except for (B8).

The verification of (B8) depends upon finding an explicit expression for the operator K. Eq. (2.6) may be rewritten as

$$\begin{aligned} K &= -\sum_{r=1}^{3} \int_{\mathbb{R}^6} \delta_{rx}\delta_{rz} \langle \varphi_{rx} F^{-1} \varphi_{rz} \rangle_{\text{vac}}\, d^3x\, d^3z \\ &= -\sum_{r=1}^{3} \int_{\mathbb{R}^6} (P_r M_x + M_x P_r)(P_r M_z + M_z P_r)\,\rho(x-z) \\ &\quad \times d^3x\, d^3z, \end{aligned}$$

where

$$\rho(u) = \int_{\mathbb{R}^3} |k|^{-1} |f(k)|^2 \exp(-iu\cdot k)\, d^3k$$

is a continuous bounded function on $\mathbb{R}^3$ which, for simplicity, we assume to be spherically symmetric. Putting

$$(M_x^r \psi)(y) = (-i\partial_r \mu)(y - x)\psi(y)$$

we obtain

$$K = -\sum_{r=1}^{3} \int_{\mathbb{R}^6} (2P_r M_x - M_x^r)(2M_z P_r + M_z^r)\,\rho(x-z)\, d^3x\, d^3z.$$

Straightforward calculations show that for any two C^∞ functions μ^1 and μ^2 of compact support on $\mathbb{R}^3$, one has the operator identity

$$\begin{aligned} &\int_{\mathbb{R}^6} M_x^1 M_z^2 \rho(x-z)\, d^3x\, d^3z \\ &= \int_{\mathbb{R}^6} \mu^1(x)\mu^2(z)\rho(z-x) d^3x\, d^3z. \quad 1 \\ &= \int_{\mathbb{R}^3} \hat{\mu}^1(k)\, \hat{\mu}^2(-k)\rho(k)\, d^3k \,.\quad 1, \end{aligned}$$

where $\hat{}$ denotes the Fourier transform. Using the spherical symmetry of μ and ρ we obtain

$$\begin{aligned} K = &-4(P_1^2 + P_2^2 + P_3^2) \int_{\mathbb{R}^3} \{\hat{\mu}(k)\}^2 \hat{\rho}(k)\, d^3k \\ &- 1 \int_{\mathbb{R}^3} k^2 \{\hat{\mu}(k)\}^2 \hat{\rho}(k)\, d^3k. \end{aligned}$$

The operator $(H+K)$ is therefore of the form

$$H + K = -c - \frac{1}{2m^*}\Delta$$

with a renormalized mass m^*. The validity of (B8) is immediate from the fact that H_0 contains $\mathcal{S}$.

C. Variations of the models

We mention some variations of the above models. Since there is no necessity for a phonon field in the limit of small lattice spacing to be rotationally invariant we should not have supposed this above. The modifications are straightforward and lead to a tensor renormalized mass.

There is no difficulty in dealing with higher order derivative couplings along the lines described above. The renormalized particle Hamiltonian then involves powers of the Laplacian in the spherically symmetric case.

The condition (3.1) on the infrared behavior of f is sufficient to deal with a variety of particle-field interactions (Ref. 9, p. 133; Ref. 10, p. 43). Because of the unclear status of the restrictions on β in Eq. (2.1) we are not sure how far (3.1) may be weakened. However the assumption

$$|f(k)| \leq c\,|k|^{-1}$$

would correspond to the value $\beta = \frac{1}{2}$ in Eqs. (2.1) and (3.2), and is not sufficient for the convergence of the integral in Eq. (3.3).

We mention that by developing the techniques above it is possible to examine electron-atom interactions in the low density limit, for a fixed coupling constant.[11] In that case the effective Hamiltonian of the atoms involves multibody potentials of all orders.

For the examples of this section it is also possible to derive the effective interaction by the use of dressing transformations $e^{T(\Omega)}$ as in Ref. 12. Indeed in the proof of Theorem 2.2, Ψ_λ is essentially the sum of the terms of order ≤ 2 in the expansion of $e^{T(\Omega_\lambda)}\Psi_0$ in powers of λ.

4. BOSON EMISSION

In the above analysis nothing corresponding to boson emission appeared. This is rather puzzling in view of the many rigorous calculations obtaining such an effect from asymptotic calculations to the second order in the coupling constant.[2,3,13] In this section we clarify the situation by yet another study of the Wigner-Weisskopf atom. We assume the reader is familiar with this model, described in Refs. 1, 14, and 15 for example.

One starts with a Hilbert space $H = \mathbb{C} \oplus H_0$ where $\mathbb{C}$ represents the particle and H_0 the quantum field. The Hamiltonian H_λ is defined for $\lambda > 0$ by the matrix

$$H_\lambda = \begin{bmatrix} \lambda^2\omega & \lambda f \\ \lambda f & H \end{bmatrix}, \tag{4.1}$$

where $\omega \in \mathbb{R}$, $f \in H_0$, and $H \geq 0$ is a self-adjoint operator on H_0. We put $\Omega = \binom{1}{0}$ and study the asymptotic form as $\lambda \to 0$ of $\langle \exp(-iH_\lambda t)\Omega, \Omega\rangle$. If $\lambda^2\omega$ is replaced by ω in H_λ, then it is shown in Refs. 1,2, and 13 that under

suitable conditions

$$\lim_{\lambda\to 0} \exp(i\omega\lambda^{-2}t)\,\langle \exp(-iH_\lambda\lambda^{-2}t)\Omega,\Omega\rangle = \exp[-(a+ib)t] \tag{4.2}$$

for all $t \geq 0$, where $b \in \mathbb{R}$ and $a > 0$. However in order to relate the Wigner–Weisskopf atom to the calculations of our previous sections we must take H_λ as in Eq. (4.1). It then turns out that Eq. (4.2) is still valid but with $a = 0$, so that for times $0 \leq t \leq O(\lambda^{-2})$ there is no decay. We shall show that there is still exponential decay in this case but that it is a higher order phenomenon. For related results concerning the higher order behavior of the Wigner–Weisskopf atom we refer the reader to Ref. 12.

It turns out that the critical quantity for our analysis is the form near $z = 0$ of the function

$$\mu(z) = \langle (z+H)^{-1}f, f\rangle \tag{4.3}$$

defined and analytic for $|\arg z| < \pi$. This behavior determines the form of the infrared singularity of the quantum field. We assume that for some constants $\omega_0 > 0$, $c > 0$ and $0 < \beta < 1$

$$\mu(z) = \omega_0 - cz^\beta + o(|z|^\beta) \tag{4.4}$$

as $|z| \to 0$ with $|\arg z| < \pi$. We make some comments to motivate this condition.

(i) If $\rho(dk)$ is the measure on $[0,\infty]$ such that

$$\langle (H+z)^{-1}f, f\rangle = \int_0^\infty (k+z)^{-1}\rho(dk),$$

then the existence of a condition such as (4.4) depends only on the form of ρ near the origin. For we can always write

$$\mu(z) = \int_0^\epsilon (k+z)^{-1}\rho(dk) + \int_\epsilon^\infty (k+z)^{-1}\rho(dk)$$

and the second integral is analytic in the region $\{|z| < \epsilon\}$.

(ii) If $\mathcal{H}_0 = L^2(\mathbb{R}^3)$, $(H\psi)(k) = |k|\psi(k)$ for all $\psi \in \mathcal{H}_0$ and

$$f(k) = |k|^{\beta/2-1}(a+|k|)^{-1}$$

for some $a > 0$ and $0 < \beta < 1$, then by Ref. 10, p. 215,

$$\mu(z) = 4\pi \int_0^\infty k^\beta (a+k)^{-2}(z+k)^{-1}dk$$

$$= \frac{1}{a-z}\left\{\frac{\pi(a^\beta - z^\beta)}{(a-z)\sin\pi\beta} - \frac{\beta\pi a^{\beta-1}}{\sin\pi\beta}\right\}$$

which satisfies (4.4).

(iii) There are analogous conditions for $\beta \geq 1$ for which more complicated versions of the following analysis may be carried out.

It is easy to show that the Hamiltonian H_λ has a negative energy eigenvalue if and only if $\omega < \omega_0$. In mathematical terms $\omega = \omega_0$ is a critical threshold for the bound state to dissolve into the continuum; physically it is the threshold for the emission of bosons in an analogous fashion to Cerenkov radiation (see Ref. 9, p. 136 and Ref. 10, p. 46). From now on we assume that $\omega > \omega_0$.

Lemma 4.1: If $\mathrm{Im}\, z > 0$, then

$$\lim_{\lambda\to 0} \lambda^\alpha \langle (H_\lambda - E_\lambda + \lambda^\alpha z)^{-1}\Omega, \Omega\rangle = (z + a + ib)^{-1}, \tag{4.5}$$

where

$$\alpha = 2 + 2\beta,\quad E_\lambda = \lambda^2(\omega - \omega_0)$$

$$a + ib = c\,|\omega - \omega_0|^\beta e^{\pi i\beta}. \tag{4.6}$$

Proof: Applying Eq. (4.4) to the standard formula

$$\{\langle (H_\lambda + z)^{-1}\Omega, \Omega\rangle\}^{-1} = \lambda^2\omega + z - \lambda^2\mu(z), \tag{4.7}$$

we obtain

$$\{\lambda^\alpha \langle (H_\lambda - E_\lambda + \lambda^\alpha z)\Omega, \Omega\rangle\}^{-1}$$

$$= \lambda^{-\alpha}\{\lambda^{-2}\omega - E_\lambda + \lambda^\alpha z - \lambda^2[\omega_0 - c(-E_\lambda + \lambda^\alpha z)^\beta + o(|E_\lambda - \lambda^\alpha z|^\beta)]\}$$

$$= z + \lambda^{2-\alpha}c\{-\lambda^2(\omega - \omega_0) + \lambda^\alpha z\}^\beta + o(\lambda^{2-\alpha+2\beta})$$

$$= z + \{-(\omega - \omega_0) + 0i\}^\beta + o(1)$$

$$= z + c\,|\omega - \omega_0|^\beta e^{\pi i\beta} + o(1)$$

as required.

Theorem 4.2: If $\omega > \omega_0$ and Eq. (4.4) is satisfied, then

$$\lim_{\lambda\to 0} \exp(iE_\lambda\lambda^{-\alpha}t)\,\langle \exp(-iH_\lambda\lambda^{-\alpha}t)\,\Omega, \Omega\rangle = e^{-iat-bt} \tag{4.8}$$

for all $t \geq 0$, where E_λ, α, a, b are defined by Eq. (4.6).

Proof: Defining

$$K_\lambda = \lambda^{-\alpha}(H_\lambda - E_\lambda)$$

Eq. (4.5) may be rewritten in the form

$$\lim_{\lambda\to 0} \langle (K_\lambda + z)^{-1}\,\Omega, \Omega\rangle = (z + a + ib)^{-1}.$$

It is a consequence of Ref. 17 or of Fourier analysis that

$$\lim_{\lambda\to 0} \langle \exp(-iK_\lambda t)\,\Omega, \Omega\rangle = \exp(-iat - b|t|)$$

for all $t \in \mathbb{R}$, and this implies Eq. (4.8).

In less precise terms Eq. (4.8) states that

$$\langle \exp(-iH_\lambda t)\,\Omega, \Omega\rangle \sim \exp\{-i\lambda^2(\omega - \omega_0)t - ia\lambda^\alpha t - b\lambda^\alpha t\}$$

as $\lambda \to 0$ for $0 \leq t \leq O(\lambda^{-\alpha})$. Thus the bound state is unstable and decays at the rate $2b\lambda^\alpha$. To obtain the decay rate we needed to work to the order α in the coupling constant; we see that $\alpha > 2$ and α is not necessarily an integer. The occurrence of such noninteger powers of λ is a commonplace of higher order analysis and is frequently associated with nonexponential decay.[7]

A glance at Eq. (4.5) suggests that the exponential decay is associated with a pole on the unphysical sheet of the resolvent near the point

$$z_\lambda = \lambda^2(\omega - \omega_0) + \lambda^\alpha(a + ib).$$

As $\lambda \to 0$ this point converges to the origin, where the resolvent has an essential singularity so that an application of the methods of Refs. 17–19 would not be straightforward. In our approach, however, we needed estimates only on the matrix elements of the resolvent in the physical sheet. We did not assume the existence of boundary values on the real axis as in Ref. 20 let alone the existence of an analytic continuation to the unphysical sheet.

ACKNOWLEDGMENTS

I should like to thank J. Fröhlich for some very helpful discussions, and the members of the Princeton Mathematics Department for their great hospitality.

[1]E. B. Davies, "Dynamics of a multilevel Wigner—Weisskopf atom," J. Math. Phys. **15**, 2036—2041 (1974).
[2]E. B. Davis, "Markovian master equations," Commun. Math. Phys. **39**, 91—110 (1974).
[3]E. B. Davis, "A model of atomic radiation," Ann. Inst. H. Poincaré (to appear).
[4]P. Blanchard, "Discussion mathématique du modèle de Pauli et Fierz relatif à la catastrophe infrarouge," Commun. Math. Phys. **15**, 156—172 (1969).
[5]J. Fröhlich, "On the infrared problem in a model of scalar electrons and massless scalar bosons," Ann. Inst. H. Poincaré A **19**, 1—103 (1973).
[6]M. Reed and B. Simon, Methods of Modern Mathematical Physics (Academic, New York, 1975), Vol. 2.
[7]E. B. Davies, "Asymptotic analysis of some abstract evolution equations," J. Functional Anal. **25**, 81—101 (1977).
[8]T. G. Kurtz, "A limit theorem for perturbed operator semigroups with applications to random evolutions," J. Funct. Anal. **12**, 55—67 (1973).
[9]C. Kittel, *Quantum Theory of Solids* (Wiley, New York, 1963).
[10]J. M. Ziman, *Elements of Advanced Quantum Theory* (Cambridge U. P., Cambridge, 1969).
[11]E. B. Davies, "Electron-atom interactions in the low density limit" (in preparation).
[12]E. Nelson, "Interaction of nonrelativistic particles with a quantized scalar field," J. Math. Phys. **5**, 1190—7 (1964).
[13]E. B. Davies, "The harmonic oscillator in a heat bath," Commun. Math. Phys. **33**, 171—86 (1973).
[14]R. Davidson and J. J. Kozak, "Relaxation to quantum statistical equilibrium of the Wigner—Weisskopf atom in a one-dimensional radiation field. III. The quantum-mechanical solution." J. Math. Phys. **12**, 903—917 (1973).
[15]A. P. Grecos and I. Prigogine, "Kinetic and ergodic properties of quantum systems—the Friedrichs model," Physica **59**, 77—96 (1972).
[16]A. Erdelyi (editor), Bateman manuscript project. *Tables of Integral Transforms* (McGraw-Hill, New York, 1954), Vol. 2.
[17]J. S. Howland, "Spectral concentrations and virtual poles II.," Trans. Amer. Math. Soc. **162**, 141—56 (1971).
[18]J. S. Howland, "The Livsic matrix in perturbation theory," J. Math. Anal. Appl. **50**, 415—37 (1975).
[19]B. Simon, "Resonances in n-body quantum systems with dilatation analytic potentials and the foundations of time-dependent perturbation theory," Ann. Math. **97**, 247—74 (1973).
[20]E. B. Davies, "Resonances, spectral concentration and exponential decay," Lett. Math. Phys. **1**, 31—5 (1975).

Microscopic theory of superfluid Fermi systems. I. Binary expansion

L. E. Reichl

Center for Statistical Mechanics, The University of Texas, Austin, Texas 78712

This is the first of two papers in which the binary expansion is used to study the properties of hard core fermi systems with broken gauge symmetry. In this paper, which is primarily formal, symmetry-breaking source terms are introduced into the generating function for thermodynamic quantities, and an exact expression is obtained for the generating function in terms of single-particle and two-particle sources and the reaction matrices for the two-body problem. In a subsequent paper, the theory is applied to a dilute hard sphere fluid.

I. INTRODUCTION

At the present time, a discontinuity exists between the theory of classical fluids and the theory of degenerate quantum fluids. It is our aim in this and a subsequent paper to try to bridge the gap.

The theory of classical fluids is largely based on the use of binary expansions (expansions in terms of exact two body processes). For example, for equilibrium systems, the Ursell—Mayer expansion is used to calculate virial coefficients and provides an important tool for determining the form of the interparticle potential in real gases.[1] In nonequilibrium systems, binary expansions have been used to show that the current—correlation functions for moderately dense gases exhibit long time tails.[2] The importance of binary expansions is that they provide a theory of many-body systems in which the integrity of the basic two-body processes is preserved.

On the other hand, the theory of degenerate quantum fluids has been based largely on the use of perturbation expansions, i.e., quantum field theory. While quantum field theory has led to huge advances in our understanding of real degenerate quantum fluids, it can only be considered as a phenomenological theory for fluids in which the particles have hard cores or can form bound states. In quantum field theory, the integrity of the basic two-body processes is completely lost.

The binary expansion has been used to study degenerate quantum fluids but has had a varied history. It was first studied by Lee and Yang[3] for the case of normal Fermi fluids and for boson fluids with broken gauge symmetry. While Lee and Yang never introduced a systematic method for dealing with self-energy effects (the proper treatment of self-energy terms is of fundamental importance in a symmetry broken fluid because they determine the form of the gap equation), they did introduce the so-called x-ensemble as a means of summing terms in the binary expansion for boson fluids below the critical temperature. The original work of Lee and Yang was later extended by Mohling,[4] who was able to resum self-energy effects in a consistent manner by means of his Λ-transformation. The binary expansion has never yet been applied to a Fermi fluid with broken gauge symmetry.

In a recent series of papers,[5-7] (referred to here as RI, RII, and RIII) the author has reexamined the binary expansion for normal Fermi fluids and has shown that there is an alternative method (to the Λ-transformation) for treating self-energy effects in the binary expansion. The method introduced in RI—RIII agrees more in spirit with the methods used in quantum field theory. In the present series of papers, we shall use some of the results of RI—RIII to study the behavior of symmetry broken Fermi fluids using the binary expansion.

The breaking of gauge symmetry in a Fermi fluid appears in the form of off-diagonal long-range order (ODLRO)[8] in the reduced two-body density matrices. In quantum field theory the standard method of treating such systems is to introduce symmetry breaking source terms into the density operator[9,10] and to resum self-energy effects with the source terms present. At the end of the calculation the source terms are set equal to zero, and as a result one obtains a self-consistent equation for the gap function.

In the present paper, we shall show that such methods can also be used for the binary expansion and indeed provide a simpler and more transparent means of studying symmetry broken quantum fluids than the x-ensemble introduced by Lee and Yang. Since most properties of the binary expansion that we shall use have been discussed at great length in RI—RIII we shall continually refer to those papers and only discuss differences introduced by the symmetry-breaking source terms. The present paper will concentrate on formal properties of the binary expansion in the presence of symmetry-breaking sources. In a subsequent paper we shall apply our results to the case of a low-density hard-sphere Fermi fluid. This application will make the differences between the binary expansion and quantum field theory quite clear.

The ultimate purpose of this work, of course, is to obtain a deeper understanding of the processes leading to super fluidity in liquid He^3. Both the perturbation expansion (quantum field theory) and the binary expansion are valid solutions to the differential equation for the N-body density operator. It therefore is of interest to see what they both have to say on the subject.

We shall begin in Sec. II, by writing a generating function for various thermodynamic quantities in terms of a density operator containing one- and two-body source operators. The one-body source operators will enable us to define Green's functions and expectation values; the two-body source operators will allow us to

 0022-2488/79/030352-09$01.00

obtain an equation for the gap function below the critical temperature. In Sec. III, the density operator will be expanded in terms of binary operators and the source operators, and a partial resummation of the density operator will be performed. In Sec. IV, we explicitly write the generating function in terms of antisymmetrized matrix elements in the momentum representation and introduce an expansion for it in terms of 0 diagrams. In Sec. V, we resum terms in the expression for the generating function and write it in terms of a new generating function which depends only on connected 0 diagrams and therefore is proportional to the volume. We then go on to define all expectation values and the grand potential in terms of connected diagrams. Finally, in Sec. VI, we obtain an expression for the generating function in terms of source operators and reaction matrices for two-body scattering processes in the fluid.

II. GENERATING FUNCTION FOR THERMODYNAMIC QUANTITIES

Expressions for the grand potential, various expectation values, and the finite temperature Green's functions for an equilibrium hard core Fermi fluid in which gauge symmetry has been broken can be derived from a single generating function.

We shall consider a system of spin-$\frac{1}{2}$ fermions with magnetic moment μ, which interact with one another via a spherically symmetric potential $V(|\mathbf{r}_{ij}|)$ ($\mathbf{r}_{ij}$ is the relative position between particles i and j). We shall assume that $V(|\mathbf{r}_{ij}|)$ is short ranged and spin independent with a large repulsive core for small values of $|\mathbf{r}_{ij}|$ and a weak attractive region for larger values of $|\mathbf{r}_{ij}|$. We shall not include the magnetic dipole interaction, although there would be no difficulty in doing so.

In order to include the possibility of broken gauge symmetry in the Fermi fluid, we shall explicitly break the gauge symmetry by including one-particle and two-particle source terms in the density operator for the system. The one-particle source terms will enable us to obtain the finite temperature Green's functions and expectation values for the system. The two-particle source terms will enable us to obtain the anomalous Green's functions. At the end of the calculation we will set the strength of the source terms equal to zero. In a Fermi system, the broken symmetry appears as ODLRO (off-diagonal long-range order) in the two body reduced density. Thus, setting the single-particle source strength to zero has no effect. However, the two-particle source terms select out precisely those quantities which exhibit ODLRO. If the system has a spontaneously broken symmetry, the terms selected by the two-body sources will not vanish when the source strength goes to zero. We can use this fact to determine the temperature at which ODLRO sets in.

We can obtain the grand partition function, expectation values, and the temperature dependent Green's function from the following generating function:

$$L(\beta, g, \nu_{\mathrm{I}}, \nu_{\mathrm{II}}) = \sum_{N=0}^{\infty} \mathrm{Tr}_N e^{-\beta(H^N + \Delta H_{\mathrm{I}}^N + \Delta H_{\mathrm{II}}^N)}, \tag{II. 1}$$

where

$$H^N = T^N - gN + V^N, \tag{II. 2}$$

$$\Delta H_{\mathrm{I}}^N = S_{\mathrm{I}}^{N(+)} + S_{\mathrm{I}}^{N(-)}, \tag{II. 3}$$

and

$$\Delta H_{\mathrm{II}}^N = S_{\mathrm{II}}^{N(+)} + S_{\mathrm{II}}^{N(-)}. \tag{II. 4}$$

In Eq. (II. 1), Tr_N denotes a trace with respect to a complete set of N-body states, $\beta = (K_B T)^{-1}$, where K_B is Boltzmann's constant and T is the absolute temperature, g is the chemical potential, ν_{I} is the single-particle source strength, and ν_{II} is the two-particle source strength.

In Eq. (II. 2), T^N is the kinetic energy operator

$$T^N = \sum_{i=1}^{N} \frac{K_i^2}{2m} \tag{II. 5}$$

where Planck's $\hbar \simeq 1$, $\mathbf{K}_i$ is the wave vector of the ith particle, and m is the mass of a fermion; V^N is the interaction potential energy operator

$$V^N = \sum_{i<j=1}^{N(N-1)/2} V_{ij}, \tag{II. 6}$$

where $V_{ij} \equiv V(|\mathbf{r}_{ij}|)$.

In Eq. (II. 3), $S_{\mathrm{I}}^{N(+)}$ is an N-particle source which emits single particles and $S_{\mathrm{I}}^{(-)}$ is an N-particle sink which absorbs single particles. They may be written

$$S_{\mathrm{I}}^{N(\pm)} = \nu_{\mathrm{I}} \sum_{i=1}^{N} s_i^{(\pm)} \tag{II. 7}$$

where s_i^+ is a source for particle i and s_i^- is a sink for particle i. Similarly, in Eq. (II. 4) $S_{\mathrm{II}}^{N(\pm)}$ are N-particle sources and sinks for pairs of particles. They may be written

$$S_{\mathrm{II}}^{N(\pm)} = \nu_{\mathrm{II}} \sum_{i<j=1}^{N(N-1)/2} s_{ij}^{(\pm)} \tag{II. 8}$$

where s_{ij}^+ is a source for pair ij and $s_{ij}^{(-)}$ is a sink for pair ij. In Sec. III, we explain how these source operators are to be used.

Expectation values of ordinary one- and two-body operators

$$0_1^N = \sum_{i=1}^{N} 0_i \tag{II. 9}$$

and

$$0_2^N = \sum_{i<j=1}^{N(N-1)/2} 0_{ij} \tag{II. 10}$$

(these can include the source operators themselves) are defined

$$\langle 0_{1(2)} \rangle = \lim_{\nu_{\mathrm{I}}, \nu_{\mathrm{II}} \to 0} \left[\left(\sum_{N=0}^{\infty} \mathrm{Tr}_N e^{-\beta(H^N + \Delta H_{\mathrm{I}}^N + \Delta H_{\mathrm{II}}^N)} 0_{1(2)}^N \right) \times \left(\sum_{N=0}^{\infty} \mathrm{Tr}_N e^{-\beta(H^N + \Delta H_{\mathrm{I}}^N + \Delta H_{\mathrm{II}}^N)} \right)^{-1} \right]. \tag{II. 11}$$

III. LADDER DIAGRAMS

In order to evaluate the trace in Eqs. (II. 1) and (II. 11), it is convenient to study the quantity $e^{\beta H_0^N} e^{-\beta(H^N + \Delta H_{\mathrm{I}}^N + \Delta H_{\mathrm{II}}^N)}$ which may be written

$$e^{\beta H_0^N} e^{-\beta(H^N + \Delta H_{\mathrm{I}}^N + \Delta H_{\mathrm{II}}^N)} = W^N(\beta, 0) + \sum_{n=1}^{\infty} (-1)^n \int_0^{\beta} d\lambda_1 \int_0^{\lambda_1} d\lambda_2 \cdots \times \int_0^{\lambda_{n-1}} d\lambda_n W^N(\beta, \lambda_1) [\Delta H_{\mathrm{I}}^N(\lambda_1)$$

$$+\Delta H_{II}^N(\lambda_2)]\,W^N(\lambda_1,\lambda_2)[\Delta H_I^N(\lambda_2)$$
$$+\Delta H_{II}^N(\lambda_2)]\times\cdots\times[\Delta H_I^N(\lambda_n)$$
$$+\Delta H_{II}^N(\lambda_n)]W^N(\lambda_n,0), \qquad \text{(III. 1)}$$

where

$$H_0^N \equiv T^N - g^N, \qquad \text{(III. 2)}$$

$$\Delta H_{I(II)}^N(\lambda) = e^{\lambda H_0^N}\Delta H_{I(II)}^N e^{-\lambda H_0^N}, \qquad \text{(III. 3)}$$

and

$$W^N(\lambda,\lambda_0) = e^{\lambda H_0^N} e^{-(\lambda-\lambda_0)H^N} e^{-\lambda H_0^N}. \qquad \text{(III. 4)}$$

The operator $W^N(\lambda,\lambda_0)$ can be expanded either in a perturbation expansion or a binary expansion. If we expand it in a perturbation expansion, then the procedure which we shall introduce in subsequent sections would enable us to reproduce all the results of field theory. However, by expanding in a perturbation expansion, we destroy the integrity of the basic scattering processes in the fluid. If we expand $W^N(\lambda,\lambda_0)$ in a binary expansion, we can obtain expressions for thermodynamic quantities in terms of the exact solution of the two-body problem and, in the classical limit, we can recover the Ursell–Mayer expansion.

In this paper, we wish to write the thermodynamic properties of a symmetry broken Fermi fluid in terms of solutions to the two-body problem, and therefore we will expand $W^N(\lambda,\lambda_0)$, wherever it appears, in terms of the following binary expansion:

$$W^N(\lambda,\lambda_0) = 1 + \sum_{\mu=1}^{N(N-1)/2}\int_{\lambda_0}^{\lambda} d\lambda_1 R_\mu(\lambda,\lambda_1) M_\mu^N(\lambda_1,\lambda_0), \qquad \text{(III. 5)}$$

where

$$M_\mu^N(\lambda_1,\lambda_0) = 1 + \sum_{\nu\neq\mu=1}^{N(N-1)/2}\int_{\lambda_0}^{\lambda_1} d\lambda_2 R_\nu(\lambda_1,\lambda_2) M_\nu^N(\lambda_2,\lambda_0). \qquad \text{(III. 6)}$$

The binary operator $R_\mu(\lambda_i,\lambda_j)$ depends on the pair of particles μ and is defined

$$R_\mu(\lambda_i,\lambda_j) = -\frac{\partial}{\partial\lambda_j} W^\mu(\lambda_i,\lambda_j), \qquad \text{(III. 7)}$$

where W^μ is defined in Eq. (III. 4) and depends on the pair of particles μ. The binary operator obeys the identity

$$R_\mu(\lambda_3,\lambda_1) = [1+\int_{\lambda_2}^{\lambda_3} ds\, R_\mu(\lambda_3,s)]\,R_\mu(s,\lambda_1) \qquad \text{(III. 8)}$$

(cf. RI for more discussion).

It is useful to express Eq. (III. 1) geometrically:

$$e^{\beta H_0^N} e^{-\beta(H^N+\Delta H_I^N+\Delta H_{II}^N)} = \sum_{Q=0}^{\infty} \text{(all different } Q\text{th order } N\text{-particle ladder diagrams)}. \qquad \text{(III. 9)}$$

To construct a Qth order N-particle ladder diagram, first draw N vertical lines and label them from left to right from 1 to N. At the top draw a horizontal line labeled β and below it draw Q horizontal lines labeled from λ_1 to λ_Q from top to bottom. Ladder diagrams are completed by inserting X^2s into boxes and triangular one- and two-particle source vertices on horizontal lines according to the rules in Appendix A.

Some examples of four-particle ladder diagrams are given in Fig. 1. Algebraic expressions for the diagrams in Fig. 1 are

$$\nu_I\nu_{II}\int_0^\beta d\lambda_1\int_0^{\lambda_1} d\lambda_2\int_0^{\lambda_2} d\lambda_3\int_0^{\lambda_3} d\lambda_4\int_0^{\lambda_4} d\lambda_5\, R_{12}(\beta,\lambda_1)\, s_{34}^-(\lambda_2)$$
$$\times R_{23}(\lambda_2,\lambda_3)\, s_3^+(\lambda_4)\, R_{34}(\lambda_4,\lambda_5), \qquad \text{Fig. 1(a)}$$

$$-\nu_I^2\nu_{II}\int_0^\beta d\lambda_1\int_0^{\lambda_1} d\lambda_2\int_0^{\lambda_2} d\lambda_3\int_0^{\lambda_3} d\lambda_4\int_0^{\lambda_4} d\lambda_5\int_0^{\lambda_5} d\lambda_6\, s_{12}^+(\lambda_1)\, s_2^-(\lambda_2)$$
$$\times R_{23}(\lambda_2,\lambda_3)\, R_{12}(\lambda_3,\lambda_4)\, R_{23}(\lambda_4,\lambda_5)\, s_1^-(\lambda_6). \qquad \text{Fig. 1(b)}$$

At this point it is possible to write Eq. (III. 1) in terms of a cluster expansion (cf. RI, Sec. II) by summing large classes of unconnected ladder diagrams. We would then obtain an expression for Eq. (III. 1) in terms of a sum of products of connected ladder diagrams. If no source terms were present, this would be a useful way to proceed because matrix elements of connected ladder diagrams would then be proportional to the volume and we would have located all the volume dependence of our generating function. However, the source terms give rise to additional volume dependence in the connected ladder diagrams. Therefore, we shall work directly with the ladder diagrams themselves.

The ladder diagrams can be simplified if we make use of Eq. (III. 8) to sum large classes of terms which contain disjoint internal chains of crosses (cf. RI, Sec. II and RI, Appendix A). We then can write Eq. (III. 9) as follows:

$$e^{\beta H_0^N} e^{-\beta(H^N+\Delta H_I^N+\Delta H_{II}^N)} = \sum_{Q=0}^{\infty} \text{(all different } Q\text{th order } N\text{-particle contracted ladder diagrams)}. \qquad \text{(III. 10)}$$

Contracted ladder diagrams are drawn and evaluated in the same way as ladder diagrams except that Rule (A. 1) of the Appendix is changed to read

(III. i) One and only one cross can occur between any two horizontal lines. The upper end of each cross must either rest on the line $\lambda=\beta$ or must connect to a higher cross or triangle.

We can now use Eq. (III. 10) to evaluate the trace in Eq. (II. 1).

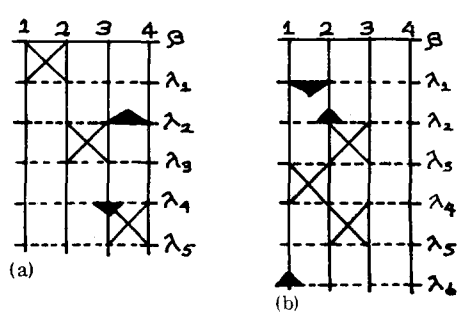


FIG. 1. Four-particle ladder diagrams.

IV. GENERATING FUNCTION IN THE MOMENTUM REPRESENTATION

Since we are dealing with a system of identical fermions, we will evaluate the trace in Eq. (II. 1) with respect to a complete set of N-particle antisymmetrized momentum and spin eigenstates of H_0^N which we denote

$$|K_1, \ldots, K_N\rangle^{(a)} = \sum_p (\epsilon)^p |K_1, \ldots, K_N\rangle, \tag{IV. 1}$$

where $\sum_p$ is the sum of all permutations of quantities K_j, and $\epsilon = -1$ (we write ϵ rather than -1 because it allows us to keep track of terms which result from exchange effects). As usual, each particle is assigned a definite position in the ket, the quantity K_j which occurs in the jth position gives the momentum and spin of the particle; i.e., $K_j = (\mathbf{K}_j, \gamma_j)$. The ket $|K_1, \ldots, K_N\rangle^{(a)}$ is not normalized but must be multiplied by a factor $(N!)^{-1/2}$. We can only use the antisymmetrized states to obtain matrix elements of operators which are symmetric under interchange of particles.

The generating function can now be written

$$L(\beta, g, \nu_{\mathrm{I}}, \nu_{\mathrm{II}}) = \sum_{N=0}^{\infty} \left(\frac{1}{N!}\right) \sum_{K_1, \ldots, K_N} \exp\left(-\beta \sum_{i=1}^{N} \omega_i\right) \langle K_1, \ldots, K_N | \times e^{\beta H_0^N} e^{-\beta(H^N + \Delta H_{\mathrm{I}}^N + \Delta H_{\mathrm{II}}^N)} | K_1, \ldots, K_N \rangle^{(a)}, \tag{IV. 2}$$

where

$$\omega_j = K_j^2/2m - g. \tag{IV. 3}$$

With the definition of $e^{\beta H_0^N} e^{-\beta(H^N + \Delta H_{\mathrm{I}}^N + \Delta H_{\mathrm{II}}^N)}$ given in Eq. (III. 10), the method of evaluation of the matrix element in Eq. (IV. 2) is straightforward and has been discussed elsewhere (cf. RI, for example). However, the source terms introduce some new features which we shall discuss in some detail.

The source terms can be thought of as "half-potentials." They give rise to vertices in which particles can only enter or only leave. One way to evaluate Eq. (IV. 2) would be to treat $s_j^{(\pm)}$ and $s_{ij}^{(\pm)}$ as normal potentials and at the end of the calculation, eliminate all contributions from lines which leave $s_j^{(-)}$ or $s_{ij}^{(-)}$ and all contributions from lines which enter $s_j^{(+)}$ or $s_{ij}^{(+)}$, since these would be contributions from nonexistent particles. The main difficulty in evaluating Eq. (IV. 2) is in counting terms correctly. We can treat the sources as "half-potentials" if we introduce the following devices. We define the sources in the momentum representation as

$$s_i^- \equiv \nu_{\mathrm{I}} \sum_{K_a} s_{K_a}^- (|0_a\rangle\langle K_a|)_i, \tag{IV. 4}$$

$$s_i^+ \equiv \nu_{\mathrm{I}} \sum_{K_a} s_{K_a}^+ (|K_a\rangle\langle 0_a|)_i, \tag{IV. 5}$$

$$s_{ij}^- = \nu_{\mathrm{II}} \sum_{K_a K_b} s_{K_a K_b}^- (|0_a 0_b\rangle\langle K_a K_b|)_{ij}, \tag{IV. 6}$$

and

$$s_{ij}^+ = \nu_{\mathrm{II}} \sum_{K_a K_b} s_{K_a K_b}^+ (|K_a K_b\rangle\langle 0_a 0_b|)_{ij}. \tag{IV. 7}$$

The states $|0_a 0_b\rangle$ are introduced for counting purposes and are defined to satisfy the relation

$$\langle K_1 K_2 | 0_a 0_b\rangle = \epsilon\langle K_2 K_1 | 0_a 0_b\rangle = \langle K_1 | 0_a\rangle\langle K_2 | 0_b\rangle = \epsilon\langle K_2 | 0_a\rangle\langle K_1 | 0_b\rangle. \tag{IV. 8}$$

With the above definitions of the source terms, the procedure for writing the generating function in terms of matrix elements of two-body and one-body operators is as follows: We first use Eq. (III. 10) to write $L(\beta, g, \nu_{\mathrm{I}}, \nu_{\mathrm{II}})$ in terms of matrix elements of the binary operators, $R_u(\lambda_i, \lambda_j)$, and the source operators. We then use Eq. (IV. 8) to combine terms. *At the very end of the calculation* we set $\langle K_i | 0_a\rangle = \langle 0_a | K_i\rangle = \delta_{K_i 0_a}$ and sum over K_i. Terms which have matrix elements of binary operators $\langle K_1 K_2 | R_u(\lambda_i, \lambda_j) | K_3 K_4\rangle$ or source terms $s_{K_1}^{\pm}$, $s_{K_1, K_2}^{\pm}$ with one or more momenta equal to zero are then set equal to zero since they represent processes involving nonexistent particles.

As a result of the above procedure, we are lead to the following expression for the generating function:

$$L(\beta, g, \nu_{\mathrm{I}}, \nu_{\mathrm{II}}) = \sum (\text{all different 0 diagrams}). \tag{IV. 9}$$

A 0 diagram is a collection of R-vertices, $s_{\mathrm{I}}^{(\pm)}$-vertices, $s_{\mathrm{II}}^{(\pm)}$-vertices, solid lines, and wavy lines. R-vertices have two lines entering and two lines leaving, s_{I}^--vertices have one line entering, s_{I}^+-vertices have one line leaving, s_{II}^--vertices have two lines entering, s_{II}^+-vertices have two lines leaving. Wavy lines can only be directed to the left but solid lines can be directed to the left or right. No wavy line double bonds can connect two R-vertices. No lines in the diagram can be broken. Two 0 diagrams differ if they have different topological structure or if they have the same topological structure but the lines are of different types or directions.

Algebraic expressions can be associated with the 0 diagrams according to the rules in Appendix B.

Some examples of 0 diagrams are given in Fig. 2. Algebraic expressions for the diagrams in Fig. 2 are

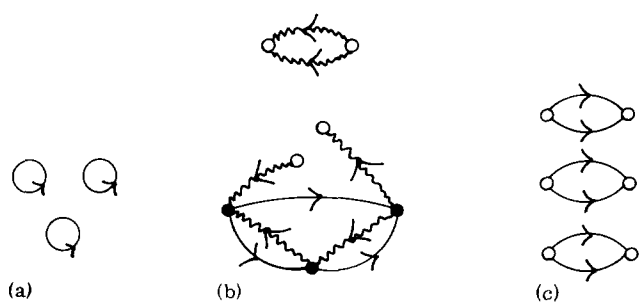
$$\frac{1}{3!}\left[\sum_{K_1} \epsilon \ln(1 - \epsilon e^{-\beta\omega_1})\right]^3 \qquad \text{Fig. 2(a)}$$

$$\frac{1}{2}\left[\sum_{K_1 K_2} \int_0^\beta \int_0^\beta d\lambda_1 d\lambda_2 \theta(\lambda_1 - \lambda_2)\, s_{\mathrm{II}}^- \left(\begin{matrix} 0 & 0 \\ K_1 & K_2 \end{matrix}\right)_{\lambda_1} s_{\mathrm{II}}^+ \left(\begin{matrix} K_1 & K_2 \\ 0 & 0 \end{matrix}\right)_{\lambda_2}\right]$$

$$\times \epsilon^4 \sum_{K_1, \ldots K_7} \int_0^\beta \int_0^\beta d\lambda_1 \ldots d\lambda_5 \theta(\lambda_1 - \lambda_3)\theta(\lambda_1 - \lambda_2)\theta(\lambda_4 - \lambda_5)$$

$$\times \theta(\lambda_3 - \lambda_5)(\epsilon\nu_3)(\epsilon\nu_4)(\epsilon\nu_6)\, R\left(\begin{matrix} 3 & 4 \\ 1 & 5 \end{matrix}\right)_{\lambda_1} R\left(\begin{matrix} 5 & 6 \\ 4 & 7 \end{matrix}\right)_{\lambda_3} R\left(\begin{matrix} 2 & 7 \\ 3 & 6 \end{matrix}\right)_{\lambda_5}$$

$$\times s_{\mathrm{I}}^+ \left(\begin{matrix} K_1 \\ 0 \end{matrix}\right)_{\lambda_2} s_{\mathrm{I}}^- \left(\begin{matrix} 0 \\ K_2 \end{matrix}\right)_{\lambda_4} \qquad \text{Fig. 2(b)}$$



(a) (b) (c)

FIG. 2. 0 diagrams.

and

$$\frac{1}{3!}\left[\sum_{K_1K_2}\int_0^\beta\int_0^\beta d\lambda_1 d\lambda_2\,\theta(\lambda_1-\lambda_2)(\epsilon\nu_1)(\epsilon\nu_2)\,s_{\rm II}^{+}\begin{pmatrix}K_1 & K_2\\ 0 & 0\end{pmatrix}_{\lambda_1} s_{\rm II}^{-}\begin{pmatrix}0 & 0\\ K_1 & K_2\end{pmatrix}_{\lambda_2}\right]^3$$

We now can sum the 0 diagrams. Fig. 2(c)

V. GENERATING FUNCTION IN TERMS OF CONNECTED DIAGRAMS

In Section IV, we wrote the generating function $L(\beta,g,\nu_{\rm I},\nu_{\rm II})$ in terms of 0 diagrams which depend on various powers of the volume. However, the 0 diagrams can be summed and the generating function can be written as an exponential of a function, $B(\beta,g,\nu_{\rm I},\nu_{\rm II})$ which depends only on connected 0 diagrams and is proportional to the volume:

$$L(\beta,g,\nu_{\rm I},\nu_{\rm II})=e^{B(\beta,g,\nu_{\rm I},\nu_{\rm II})}, \tag{V.1}$$

where

$$B(\beta,g,\nu_{\rm I},\nu_{\rm II})=\sum(\text{all different connected 0 diagrams}). \tag{V.2}$$

Connected 0 diagrams are drawn in the same way as 0 diagrams except that all vertices must be completely interconnected by solid and wavy lines. Rules (B.1)–(B.10) may be used to associate algebraic expressions with them.

In terms of the generating function, $B(\beta,g,\nu_{\rm I},\nu_{\rm II})$, the grand potential is defined

$$\Omega(\beta,g)=-\frac{1}{\beta}\lim_{\substack{\nu_{\rm I}\to 0\\ \nu_{\rm II}\to 0}} B(\beta,g,\nu_{\rm I},\nu_{\rm II}). \tag{V.3}$$

Expectation values of arbitrary one- and two-body operators $0_{1(2)}^N$ are defined

$$\langle 0_1\rangle=e^{\beta\Omega}\lim_{\substack{\nu_{\rm I}\to 0\\ \nu_{\rm II}\to 0}}\lim_{\substack{\lambda_b\to 0^+\\ \lambda_a\to 0}}\sum_{K_aK_b}\langle K_a|0_1|K_b\rangle\frac{\delta^2L(\beta,g,\nu_{\rm I},\nu_{\rm II})}{\delta s_{\rm I}^{-}(b)\delta s_{\rm I}^{+}(a)} \tag{V.4}$$

and

$$\langle 0_2\rangle=e^{\beta\Omega}\lim_{\substack{\nu_{\rm I}\to 0\\ \nu_{\rm II}\to 0}}\lim_{\substack{\lambda_a\to\lambda_b\to 0\\ \lambda_c\to\lambda_d\to 0^+}}\frac{1}{2}\sum_{K_aK_bK_cK_d}\langle K_aK_b|0_2|K_cK_d\rangle^{(a)}$$
$$\times\frac{\delta^4L(\beta,g,\nu_{\rm I},\nu_{\rm II})}{\delta s_{\rm I}^{+}(a)\delta s_{\rm I}^{+}(b)\delta s_{\rm I}^{-}(c)\delta s_{\rm I}^{-}(d)}. \tag{V.5}$$

Expectation values of one- and two-body source operators $S_{\rm I}^{N(\pm)}$ and $S_{\rm II}^{N(\pm)}$ are

$$\langle s_{\rm I}^{(\pm)}\rangle=e^{\beta\Omega}\lim_{\substack{\nu_{\rm I}\to 0\\ \nu_{\rm II}\to 0}}\lim_{\lambda_a\to 0}\sum_{K_a}s_{K_a}^{(\pm)}\frac{\delta L(\beta,g,\nu_{\rm I},\nu_{\rm II})}{\delta s_{\rm I}^{\pm}(a)} \tag{V.6}$$

and

$$\langle s_{\rm II}^{(\pm)}\rangle=e^{\beta\Omega}\lim_{\substack{\nu_{\rm I}\to 0\\ \nu_{\rm II}\to 0}}\lim_{\lambda_a\to 0}\sum_{K_aK_b}s_{K_aK_b}^{(\pm)}\frac{\delta L(\beta,g,\nu_{\rm I},\nu_{\rm II})}{\delta s_{\rm II}^{(\pm)}(a,b)} \tag{V.7}$$

In Eqs. (V.4)–(V.7) we have used the notation

$$s_{\rm I}^{+}(a)\equiv s_{\rm I}^{+}\begin{pmatrix}K_a\\ 0\end{pmatrix}_{\lambda_a},\quad s_{\rm I}^{-}(a)\equiv s_{\rm I}^{-}\begin{pmatrix}0\\ K_a\end{pmatrix}_{\lambda_a},\quad s_{\rm II}^{+}(a,b)\equiv s_{\rm II}^{+}\begin{pmatrix}K_a & K_b\\ 0 & 0\end{pmatrix}$$

and $S_{\rm II}^{-}(a,b)\equiv S_{\rm II}^{-}\begin{pmatrix}0 & 0\\ K_a & K_b\end{pmatrix}_{\lambda_a}$. The variational derivatives act on the source operators in the expression for $L(\beta g\nu_{\rm I}\nu_{\rm II})$. They are defined

$$\frac{\delta s_{\rm I}^{\pm}(j)}{\delta s_{\rm I}^{\pm}(a)}\equiv\frac{(-1)}{\nu_{\rm I}}\delta(\lambda_j-\lambda_a)\,\delta_{K_j,K_a}, \tag{V.8}$$

$$\frac{\delta s_{\rm I}^{\mp}(j)}{\delta s_{\rm I}^{\pm}(a)}\equiv 0, \tag{V.9}$$

$$\frac{\delta s_{\rm II}^{\pm}(i,j)}{\delta s_{\rm II}^{\pm}(a,b)}\equiv\frac{(-1)}{\nu_{\rm II}}\delta(\lambda_i-\lambda_a)\delta_{K_iK_a}\delta_{K_jK_b}, \tag{V.10}$$

and

$$\frac{\delta s_{\rm II}^{\mp}(i,j)}{\delta s_{\rm II}^{\pm}(a,b)}\equiv 0. \tag{V.11}$$

The effect of each variational derivative is to remove a source vertex from $L(\beta g\nu_{\rm I}\nu_{\rm II})$ and attach its line to a vertex representing $0_{1(2)}^N$ or $S_{\rm I(II)}^{N(\pm)}$ on the extreme right. In Eqs. (V.4) the temperature of the sink $S_{\rm I}^{-}(b)$ must be kept larger than the temperature of the source $S_{\rm I}^{+}(a)$ as they are both set equal to zero to ensure that the attached vertex appears to the right and not the left. The same is true in Eq. (V.5).

If we substitute Eq. (V.1) into Eqs. (V.4)–(V.7) and use Eq. (V.3), we can obtain an expression for the expectation values which is proportional to $B(\beta g,\nu_{\rm I},\nu_{\rm II})$. Since ODLRO does not occur for single-particle objects in a Fermi fluid, any terms containing single-particle sources after the variational derivatives are taken will be identically zero. That is,

$$\lim_{\nu_{\rm I}\to 0}\frac{\delta B}{\delta s_{\rm I}^{\pm}(a)}\equiv 0$$

since single-particle sources occur in pairs.

Thus, we obtain the following expressions for the various expectation values.

$$\langle 0_1\rangle=\lim_{\substack{\nu_{\rm I}\to 0\\ \nu_{\rm II}\to 0}}\lim_{\substack{\lambda_a\to 0^+\\ \lambda_b\to 0}}\sum_{K_aK_b}\langle K_b|0_1|K_a\rangle\frac{\delta^2B}{\delta s_{\rm I}^{+}(a)\delta s_{\rm I}^{-}(b)}, \tag{V.12}$$

$$\langle 0_2\rangle=\lim_{\substack{\nu_{\rm I}\to 0\\ \nu_{\rm II}\to 0}}\lim_{\substack{\lambda_a\to\lambda_b\to 0\\ \lambda_c\to\lambda_d\to 0^+}}\frac{1}{2}\sum_{K_aK_bK_cK_d}\langle K_aK_b|0_2|K_cK_d\rangle^{(a)}$$
$$\times\left[\frac{\delta^4B}{\delta s_{\rm I}^{+}(a)\delta s_{\rm I}^{+}(b)\delta s_{\rm I}^{-}(c)\delta s_{\rm I}^{-}(d)}+\frac{\delta^2B}{\delta s_{\rm I}^{+}(b)\delta s_{\rm I}^{-}(c)}\frac{\delta^2B}{\delta s_{\rm I}^{+}(a)\delta s_{\rm I}^{-}(d)}\right. \tag{V.13}$$
$$\left.+\frac{\delta^2B}{\delta s_{\rm I}^{+}(a)\delta s_{\rm I}^{-}(c)}\frac{\delta^2B}{\delta s_{\rm I}^{+}(b)\delta s_{\rm I}^{-}(d)}+\frac{\delta^2B}{\delta s_{\rm I}^{-}(c)\delta s_{\rm I}^{-}(d)}\frac{\delta^2B}{\delta s_{\rm I}^{+}(a)\delta s_{\rm I}^{+}(b)}\right],$$

$$\langle s_{\rm I}^{\pm}\rangle=0, \tag{V.14}$$

$$\langle s_{\rm II}^{(\pm)}\rangle=\lim_{\substack{\nu_{\rm I}\to 0\\ \nu_{\rm II}\to 0}}\lim_{\lambda_a\to 0}\sum_{K_aK_b}s_{K_aK_b}^{(\pm)}\frac{\delta B}{\delta s_{\rm II}^{\pm}(a,b)}. \tag{V.15}$$

VI. GENERATING FUNCTION IN TERMS OF REACTION MATRICES

In writing our expressions for thermodynamic properties of superfluid He^3, we wish to make as close a contact to the structure of quantum field theory as possible, since our present understanding of the properties of superfluid He^3 is based on field theoretic methods. In RI, we wrote expressions for the binary operators in terms of the exact solution to the two-body problem. We found there that the binary operator splits into two parts, one part which depends on a single temperature and another part which depends on two temperatures, as does the binary operator. Because of this splitting, we were able to isolate terms in the binary expansion to which we can apply techniques for resummation

introduced by Matsubara. These terms may be thought to give the most coherent contributions to the thermodynamic properties of superfluid He^3.

In RI, Eq. (VI. 13) we obtained the following expression for the binary operator in terms of reaction matrices

$$R\begin{pmatrix}1&2\\3&4\end{pmatrix}^{\lambda_2}_{\lambda_1}=-\Big[C^2(\mathbf{K}_{12})_0\langle K_1K_2|A|K_3K_4\rangle_0^{(a)}e^{-\lambda_1(\omega_1+\omega_2-\omega_3-\omega_4)}$$
$$+\frac{1}{2}\sum_{K_5K_6}C^2(\mathbf{K}_{56})e^{\lambda_1(\omega_1+\omega_2-\omega_5-\omega_6)}e^{-\lambda_2(\omega_3+\omega_4-\omega_5-\omega_6)}$$
$$\times P\left(\frac{1}{\omega_5+\omega_6-\omega_1-\omega_2}\right)_0\langle K_1K_2|A|K_5K_6\rangle_0^{(a)}$$
$$\times{}_0\langle K_5K_6|A|K_3K_4\rangle_0^{(a)}\Big]. \qquad \text{(VI. 1)}$$

In Eq. (VI. 1), $C(\mathbf{K}_{12})$ is a normalization constant, $\mathbf{K}_{12}$ is the relative momentum

$$\mathbf{K}_{12}=\tfrac{1}{2}(\mathbf{K}_1-\mathbf{K}_2),\quad {}_0\langle K_1K_2|A|K_3K_4\rangle_0^{(a)}$$

is the reaction matrix and P denotes principal part. We shall represent the first and second terms on the right hand side of Eq. (VI. 1) by separate vertices, as was done in RI. We then obtain the following expression for the generating function $F(\beta, g, \nu_I, \nu_{II})$:

$$B(\beta,g,\nu_I,\nu_{II})=\sum(\text{all different } A\text{-matrix 0 diagrams}) \qquad \text{(VI. 2)}$$

As in RI, A-matrix 0 diagrams are obtained from connected 0 diagrams by splitting the binary operators into two parts; and by expanding the temperature Heaviside functions in the connected 0 diagrams until we obtain all diagrams which cannot be deformed into one another without changing the direction of at least one solid line. Furthermore, we shall add together all diagrams which have the same topological structure but differ only in the number of left directed lines which are solid or wavy (cf. RIII). We can then define the A-matrix 0 diagrams as follows.

An A-matrix 0 diagram is a collection of A-vertices, D-vertices, and $s_{I(II)}^{(\pm)}$-vertices, ordered with respect to one another from left to right. All vertices are completely connected by directed solid and wavy lines. A-vertices and D-vertices each have two lines leaving and two lines entering. The s_I^+-vertex has one line leaving and the s_I^--vertex has one line entering. The s_{II}^+-vertex has two lines entering and the s_{II}^--vertex has two lines leaving. All lines are solid except for left directed internal lines of the D-vertices or left directed lines which begin and end on vertices with the same temperature labels. Left directed solid line double bonds are allowed (we explicitly subtract off a wavy line double bond each time one occurs between A-vertices, D-vertices, or A- and D-vertices). (D-vertices with two solid lines leaving must be placed so that both solid lines are directed to the right.) Two A-matrix 0 diagrams differ if they different topological structure, or if they have the same topological structure but the lines are of different types or directions, or the D-vertices have different temperature labeling.

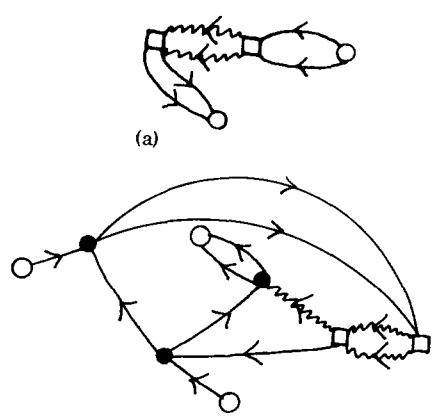
Algebraic expressions can be associated to the A-matrix 0 diagrams according to the rules in Appendix C.

Some examples of A-matrix 0 diagrams are given in Fig. 3. Algebraic expressions for the diagrams in Fig. 4 are

$$\left(\frac{1}{2}\right)\sum_{K_1,\ldots,K_6}\int_{-\infty}^{\infty}\cdots\int d\lambda_1\cdots d\lambda_3\theta(\beta-\lambda_1)\theta(\beta-\lambda_2)\theta(\lambda_1-\lambda_3)\theta(\lambda_2)\theta(\lambda_3)$$
$$\times(\epsilon\nu_1)(\epsilon\nu_2)(1+\epsilon\nu_5)(1+\epsilon\nu_6)D\begin{pmatrix}K_1&K_2\\K_3&K_4\\K_5&K_6\end{pmatrix}s_{II}^-\begin{pmatrix}0&0\\K_1&K_2\end{pmatrix}$$
$$\times s_{II}^+\begin{pmatrix}K_5&K_6\\0&0\end{pmatrix}e^{(\beta-\lambda_2)(\omega_1+\omega_2)}e^{(\beta-\lambda_1)(\omega_3+\omega_4)}e^{-(\lambda_1-\lambda_3)(\omega_5+\omega_6)},$$
Fig. 3(a)

$$\epsilon^6\sum_{K_1\cdots K_{12}}\left(\frac{1}{2}\right)^3\int_{-\infty}^{\infty}\cdots\int d\lambda_1\cdots d\lambda_7\theta(\beta-\lambda_1)\theta(\lambda_1-\lambda_2)\theta(\lambda_2-\lambda_3)$$
$$\times\theta(\lambda_3-\lambda_6)\theta(\lambda_2-\lambda_7)\theta(\lambda_3-\lambda_5)\theta(\beta-\lambda_4)\theta(\lambda_4-\lambda_5)\theta(\lambda_6)$$
$$\times\theta(\lambda_3-\lambda_6)\theta(\lambda_5-\lambda_7)\theta(\lambda_7)(\epsilon\nu_1)(\epsilon\nu_2)(\epsilon\nu_3)(\epsilon\nu_5)$$
$$\times(1+\epsilon\nu_4)(1+\epsilon\nu_8)(1+\epsilon\nu_9)(1+\epsilon\nu_7)(1+\epsilon\nu_6)$$
$$\times s_I^+\begin{pmatrix}K_1\\0\end{pmatrix}A\begin{pmatrix}K_2&K_3\\K_1&K_4\end{pmatrix}A\begin{pmatrix}K_4&K_5\\K_6&K_7\end{pmatrix}s_{II}^-\begin{pmatrix}0&0\\K_8&K_9\end{pmatrix}s_I^+\begin{pmatrix}K_6\\0\end{pmatrix}A\begin{pmatrix}K_8&K_9\\K_5&K_{10}\end{pmatrix}D\begin{pmatrix}K_7&K_{10}\\K_{11}&K_{12}\\K_2&K_3\end{pmatrix}$$
$$\times e^{(\lambda_1-\lambda_2)\omega_1}e^{(\lambda_2-\lambda_7)(\omega_2+\omega_3)}e^{-(\lambda_2-\lambda_3)\omega_4}e^{-(\lambda_4-\lambda_5)(\omega_8+\omega_9)}$$
$$\times e^{-(\lambda_3-\lambda_6)\omega_6}e^{(\lambda_3-\lambda_5)\omega_5}e^{-(\lambda_3-\lambda_5)\omega_7}e^{-(\lambda_6-\lambda_7)(\omega_{11}+\omega_{12})}. \quad \text{Fig. 3(b)}$$

The expression we have obtained for the generating function, $B(\beta, g, \nu_I, \nu_{II})$, in terms of A-matrix 0 diagrams is exact. We therefore have been able to express all thermodynamic quantities in terms of the reaction matrix for two-body scattering processes in the fluid. As a first step in obtaining a useful theory of symmetry broken Fermi fluids, we must resum self-energy effects. In RIII, we showed that it is possible to do so using techniques similar to those used in quantum field theory. However, the method used in RIII is different from the Matsubara method of temperature dependent field theory in that we introduce a Laplace transformation of temperature dependent variables rather than a Fourier expansion, as is done in field theory. The Matsubara method makes use of the fact that all propagators and vertices in temperature dependent field theory are periodic functions of temperature. In the binary expansion, this is no longer true. The appearance of D-vertices and exclusion of wavy line double bonds destroy the periodicity (particle–hole symmetry) of $B(\beta, g, \nu_I, \nu_{II})$.



(a)

(b)

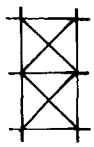
FIG. 3. A-matrix 0 diagrams.

However, there is one case in which the binary expansion and quantum field theory become almost identical in structure, and that is when we can approximate all reaction matrices by their values for forward scattering. Then the D-vertices give no contribution and the A-matrix 0 diagrams, except for a small subclass of terms, become periodic functions of temperature.

In a subsequent paper, we shall apply our theory to the case of a low-density hard-sphere Fermi fluid in the forward scattering approximation.

APPENDIX A: RULES FOR LADDER DIAGRAMS

(A. 1) One and only one cross can occur between any two horizontal lines. Crosses must connect neighboring horizontal lines.

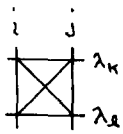
(A. 2) Two crosses cannot have two points in common, i. e., the structure  is forbidden.

(A. 3) Source and sink vertices are represented by inverted and upright triangles, respectively. Only one triangle can occur on any horizontal line.

(A. 4) The bottom points of a cross may not rest on the same horizontal line to which the base of a triangle is attached.

(A. 5) If N_t is the number of triangles and N_x is the number of crosses, then $N_t + N_x = Q$.

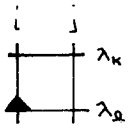
(A. 6) With each cross associate a factor



$$= R_{ij}(\lambda_K, \lambda_l),$$

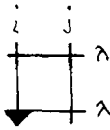
where i and j are particle labels and λ_K and λ_l are temperature labels.

(A. 7) With the one-particle source vertices, associate factors



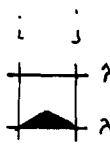
$$= -\nu_{\mathrm{I}} S_i^{-}(\lambda_l)$$

and



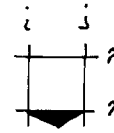
$$= -\nu_{\mathrm{I}} S_i^{+}(\lambda_l).$$

(A. 8) With the two-particle source vertices, associate factors



$$= -\nu_{\mathrm{II}} S_{ij}^{-}(\lambda_l)$$

and



$$= -\nu_{\mathrm{II}} S_{ij}^{+}(\lambda_l).$$

(A. 9) Order the algebraic expressions for the cross and triangle vertices from left to right in the same order that they appear when reading the ladder diagram from top to bottom.

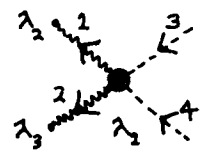
(A. 10) Integrate over each temperature from zero to the next higher temperature (the temperature of the horizontal line immediately above).

APPENDIX B: RULES FOR 0 DIAGRAMS

(B. 1) Label all lines from 1 to n, where n is the number of lines and assign to the jth line a momentum and spin $K_j = (\mathbf{K}_j, \gamma_j)$.

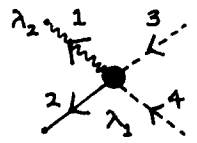
(B. 2) Label the R-vertices, and $S_{\mathrm{I}(\mathrm{II})}^{(\pm)}$-vertices from left to right from λ_1 to λ_Q, where Q is the number of R- and $S_{\mathrm{I}(\mathrm{II})}^{(\pm)}$-vertices.

(B. 3) To each R-vertex, associate factors



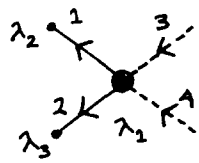
$$= R\begin{pmatrix} K_1 & K_2 \\ K_3 & K_4 \end{pmatrix}_{\lambda_1}^{\lambda_2} \theta(\lambda_3 - \lambda_2)\theta(\lambda_2 - \lambda_1)$$

if λ_3 appears to the left of λ_2



$$= R\begin{pmatrix} K_1 & K_2 \\ K_3 & K_4 \end{pmatrix}_{\lambda_1}^{\lambda_2} \theta(\beta - \lambda_2)\theta(\lambda_2 - \lambda_1)$$

and



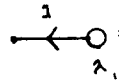
$$= R\begin{pmatrix} K_1 & K_2 \\ K_3 & K_4 \end{pmatrix}_{\lambda_1}^{\beta} \theta(\beta - \lambda_1)$$

In the above vertices, the dotted lines may be solid or wavy and

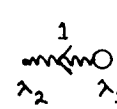
$$R\begin{pmatrix} K_1 & K_2 \\ K_3 & K_4 \end{pmatrix}_{\lambda_1}^{\lambda_2} \equiv \langle K_1 K_2 | R(\lambda_2, \lambda_1) | K_3 K_4 \rangle^{(a)}$$

The temperatures λ_2 and λ_3 are the temperatures of the vertices to which the wavy lines attach.

(B. 4) With the $s_1^{\pm}$-vertices, associate factors

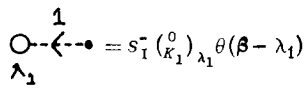


$$= s_1^{+}\begin{pmatrix} K_1 \\ 0 \end{pmatrix}_{\lambda_1} \theta(\beta - \lambda_1),$$



$$= s_1^{+}\begin{pmatrix} K_1 \\ 0 \end{pmatrix}_{\lambda_1} \theta(\lambda_2 - \lambda_1),$$

and



$$= S_I^-\begin{pmatrix}0\\K_1\end{pmatrix}_{\lambda_1}\theta(\beta-\lambda_1)$$

where

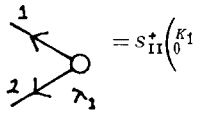
$$S_I^+\begin{pmatrix}K_1\\0\end{pmatrix}_{\lambda_1} = (-1)\nu_I S_{K_1}^+ e^{\lambda_1\omega_1}$$

and

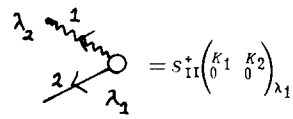
$$S_I^-\begin{pmatrix}0\\K_1\end{pmatrix} = (-1)\nu_I S_{K_1}^- e^{-\lambda_1\omega_1}.$$

The dotted line may be solid or wavy.

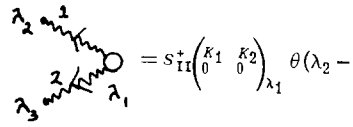
(B. 5) With the $S_{II}^{(\pm)}$-vertices, associate factors



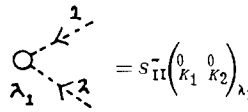
$$= S_{II}^+\begin{pmatrix}K_1 & K_2\\0 & 0\end{pmatrix}_{\lambda_1}\theta(\beta-\lambda_1),$$



$$= S_{II}^+\begin{pmatrix}K_1 & K_2\\0 & 0\end{pmatrix}_{\lambda_1}\theta(\lambda_2-\lambda_1),$$



$$= S_{II}^+\begin{pmatrix}K_1 & K_2\\0 & 0\end{pmatrix}_{\lambda_1}\theta(\lambda_2-\lambda_1)\,\theta(\lambda_3-\lambda_1),$$



$$= S_{II}^-\begin{pmatrix}0 & 0\\K_1 & K_2\end{pmatrix}_{\lambda_1}\theta(\beta-\lambda_1),$$

where

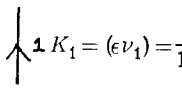
$$S_{II}^+\begin{pmatrix}K_1 & K_2\\0 & 0\end{pmatrix}_{\lambda_1} = (-1)\nu_{II}\, S_{K_1K_2}^+ e^{\lambda_1(\omega_1+\omega_2)},$$

$$S_{II}^-\begin{pmatrix}0 & 0\\K_1 & K_2\end{pmatrix}_{\lambda_1} = (-1)\nu_{II} S_{K_1K_2}^{-(a)} e^{-\lambda_2(\omega_1+\omega_2)}$$

and

$$S_{K_1K_2}^{-(a)} \equiv S_{K_1K_2}^- + \epsilon S_{K_2K_1}^-$$

(B. 6) With each solid line connecting two vertices associate a factor

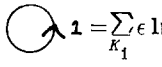


$$1\,K_1 = (\epsilon\nu_1) = \frac{\epsilon\exp(-\beta\omega_1)}{1-\epsilon\exp(-\beta\omega_1)}$$

where

$$\omega_1 = K_1^2/2m - g.$$

(B. 7) With each closed loop which does not pass through any vertices, associate a factor



$$1 = \sum_{K_1}\epsilon\ln(1-\epsilon e^{-\beta\omega_1})$$

(B. 8) Multiply by an overall factor $\epsilon^{P_B+N_I^-}(1/S)$, where N_I^- is the number of S_I^--vertices, and S is the symmetry number of the diagram (S is the number of permutations of labels on lines which leave the diagram topologically unchanged). P_B is defined as follows: Associate each momentum in the $S_{I(II)}^-$-vertices with a momentum in the $S_{I(II)}^+$-vertices by locating the momentum from an $S_{I(II)}^-$-vertex in an R-vertex matrix element and following a chain of momenta through the R-vertex matrix elements (reading from top to bottom) until the first momentum index associated to an $S_{I(II)}^+$-vertex is reached. These two momentum indices are "associated." Do this for each momentum in the $S_{I(II)}^-$-vertices until each is "associated" with a momentum from the $S_{I(II)}^-$-vertices. P_B is then obtained by counting the number of permutations of bottom line momentum indices with respect to top line indices in the R-vertex matrix elements (this includes the number of permutations needed to line up associated momenta).

(B. 9) Sum over all momenta and spins.

(B. 10) Integrate over the temperature of each vertex from 0 to β

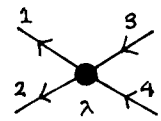
In associating temperature Heaviside functions with the vertices, as we have done in Rules (B. 3)—(B. 5), we have summed over a large number of contracted X diagrams.

APPENDIX C: RULES FOR A MATRIX 0 DIAGRAM

(C. 1) Label each line from 1 to n, where n is the number of lines, and associate to the jth line a momentum and spin $K_j = (\mathbf{K}_j\gamma_j)$.

(C. 2) Label the vertices from left to right from λ_1 to λ_Q. (D-vertices require two temperature labels. One is determined by its horizontal position; the other by the type of lines that leave it and the temperature of the vertices to which they attach.)

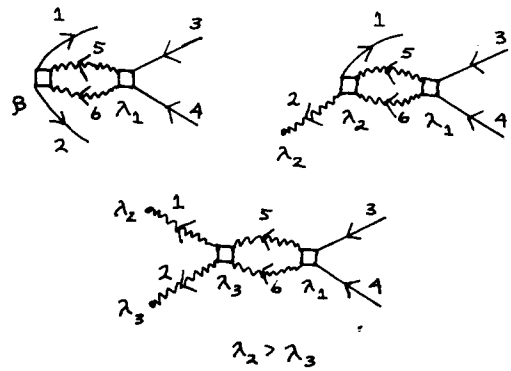
(C. 3) With each A-vertex associate a factor



$$= A\begin{pmatrix}K_1 & K_2\\K_3 & K_4\end{pmatrix} = -C^2(\mathbf{K}_{12})\,{}_0\langle K_1K_2|A|K_3K_4\rangle_0^{(a)}.$$

where the dotted lines may be wavy or solid.

(C. 4) To each D-vertex assign temperature labels according to the following conventions:



$\lambda_2 > \lambda_3$

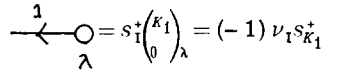
To each of the above D-vertices, assign a factor

$$D\begin{pmatrix}K_1 & K_2\\K_5 & K_6\\K_3 & K_4\end{pmatrix} = -C^2(\mathbf{K}_{56})\,{}_0\langle K_1K_2|A|K_5K_6\rangle_0^{(a)}\,{}_0\langle K_5K_6|A|K_3K_4\rangle_0^{(a)}$$

$$\times P\left(\frac{1}{\omega_5+\omega_6-\omega_1-\omega_2}\right).$$

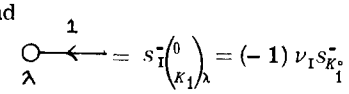
The factor $\frac{1}{2}$ in Eq. (VI. 1) is now contained in the symmetry number S^{-1} of Rule (B. 8).

(C. 5) To each $s_{\mathrm{I}}^{(\pm)}$-vertex assign factors



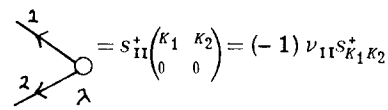
$$= s_{\mathrm{I}}^{+}\begin{pmatrix} K_1 \\ 0 \end{pmatrix}_{\lambda} = (-1)\,\nu_{\mathrm{I}} s_{K_1}^{+}$$

and



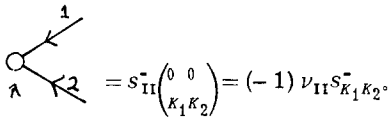
$$= s_{\mathrm{I}}^{-}\begin{pmatrix} 0 \\ K_1 \end{pmatrix}_{\lambda} = (-1)\,\nu_{\mathrm{I}} s_{K_1}^{-}.$$

(C. 6) To each $s_{\mathrm{II}}^{(\pm)}$-vertex assign factors



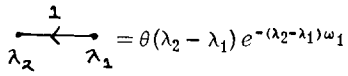
$$= s_{\mathrm{II}}^{+}\begin{pmatrix} K_1 & K_2 \\ 0 & 0 \end{pmatrix} = (-1)\,\nu_{\mathrm{II}} s_{K_1 K_2}^{+}$$

and



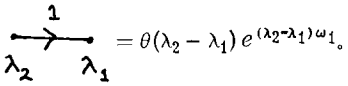
$$= s_{\mathrm{II}}^{-}\begin{pmatrix} 0 & 0 \\ K_1 & K_2 \end{pmatrix} = (-1)\,\nu_{\mathrm{II}} s_{K_1 K_2}^{-}.$$

(C. 7) With each left directed line, solid or wavy, associate a factor



$$= \theta(\lambda_2 - \lambda_1)\, e^{-(\lambda_2 - \lambda_1)\omega_1}$$

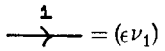
and with each right directed line associate a factor



$$= \theta(\lambda_2 - \lambda_1)\, e^{(\lambda_2 - \lambda_1)\omega_1}.$$

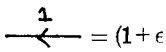
(C. 8) With each A- or $s_{\mathrm{I(II)}}^{(\pm)}$-vertex with no lines leaving or entering on the left assign a factor $\theta(\beta - \lambda_1)$, where λ_1 is the temperature of the vertex. To each A-, D-, or $s_{\mathrm{I(II)}}^{(\pm)}$-vertex with no lines leaving or entering on the right, assign a factor $\theta(\lambda_1)$, where λ_1 is the temperature of the vertex.

(C. 9) With each right directed solid line associate a factor



$$= (\epsilon\nu_1)$$

and to each left directed solid line associate a factor

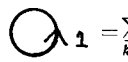


$$= (1 + \epsilon\nu_1)$$

(C. 10) For each left directed solid line double bond which connects two A-vertices, or an A-vertex to a D-vertex, subtract a factor

$$= \theta(\lambda_1 - \lambda_2)\, e^{-(\lambda_1 - \lambda_2)(\omega_1 + \omega_2)}.$$

(C. 11) With the closed loop which does not pass through any vertices, associate a factor



$$= \sum_{k_1} \epsilon \ln(1 - \epsilon e^{-\beta\omega_1}).$$

For each closed loop which passes through a single vertex assign a factor $(\epsilon\nu_1)$.

(C. 12) Multiply the entire expression by a factor $\epsilon^{P_B + N_{\mathrm{I}}}(1/S)$ as in Rule (B. 8). (Now the A-vertex and D-vertex matrix elements in obtaining P_B.)

(C. 13) Sum over all momenta and spins. Integrate over all temperatures from $-\infty$ to ∞.

[1] J. O. Hirchfelder, C. F. Curtus, and R. B. Bird, *Molecular Theory of Gases and Liquids* (Wiley, New York, 1954).
[2] M. H. Ernst, L. K. Haines, and J. R. Dorfman, Rev. Mod. Phys. **41**, 296 (1969).
[3] T. D. Lee and C. N. Yang, Phys. Rev. **113**, 1165 (1959); **116**, 25 (1959); **117**, 12, 22, 897 (1960).
[4] F. Mohling, Phys. Rev. **122**, 1043, 1062 (1961); **124**, 538 (1961); **128**, 1365 (1962); **135**, A831, A855, A876 (1964).
[5] L. E. Reichl, J. Math. Phys. **17**, 2007 (1976) (referred to as RI).
[6] L. E. Reichl, J. Math. Phys. **17**, 2023 (1976) (referred to as RII).
[7] L. E. Reichl, J. Math. Phys. **17**, 2034 (1976) (referred to as RIII).
[8] C. N. Yang, Rev. Mod. Phys. **34**, 694 (1962).
[9] C. de Dominicis and P. C. Martin, J. Math. Phys. **5**, 14, 31 (1964).
[10] B. Johansson, Physica **32**, 2164 (1966).

Microscopic theory of superfluid Fermi systems. II. Application to low density hard sphere system

L. E. Reichl

Center for Statistical Mechanics, The University of Texas, Austin, Texas 78712

This is the second of two papers in which the binary expansion is used to study properties of hard core Fermi systems with broken gauge symmetry. In this paper, we make contact with the phenomenology of superfluid Fermi systems commonly used in field theory. This requires, however, that each scattering event is approximated by its forward scattering value, an approximation which, at best is valid at very low densities. Within this approximation we show that a hard sphere Fermi fluid can exhibit superfluidity.

I. INTRODUCTION

In a previous paper,[1] we derived an exact expression for the generating function for thermodynamic properties of a symmetry broken Fermi fluid in terms of reaction matrices for the two-body problem. In the present paper, in order to illustrate use of the formalism, we shall apply that theory to the case of a low density hard sphere fermi fluid. Such a system, as has been pointed out by Lazer and Fay,[2] can exhibit superfluidity if exchange effects are taken into account because an attractive interaction can exist between pairs of particles with total spin 1. This is now thought to be the mechanism for superfluidity in liquid He^3.[3]

We shall assume that the dominant contribution to each scattering event in the fluid comes from forward scattering processes. Such an approximation is at best valid only for a very low density gas.

In the forward scattering approximation, the A-matrix 0 diagrams of Ref. 1 simplify because all D-vertices are then identically zero. The only troublesome terms which remain are those involving double bonds. As we have seen in Ref. 1 (cf. Ref. 1, Rule C.10), the binary expansion excludes the possibility of having wavy line double bonds. These correspond to processes in which a pair of particles collides and then recollides without one of the particles first having interacted with another particle in the medium or having been exchanged with another particle in the medium. Even though wavy line double bonds are excluded, we can still have double bonds which depend on one wavy line and one solid line or on two solid lines since these involve exchange effects (we are now using the terminology of Ref. 1, Appendix B). However, as we shall see, at low density these terms give no contribution to scattering processes involving pairs of particles with total spin 1, and therefore have no effect on the possible superfluidity of the hard sphere system. However, they can give a sizeable repulsive contribution to scattering processes involving particles with total spin 0. Since in the subsequent sections, we are primarily interested in superfluid properties, we shall exclude all double bond structures. When we do this, the remaining A-matrix 0 diagrams exhibit particle–hole symmetry (are periodic functions of temperature), and we can use Matsubara methods to resum them (after the single-particle sources have been set equal to zero, however).

We shall begin our discussion in Sec. II by writing the generating function $B(\beta, g, \nu_I, \nu_{II})$ in terms of the subclass of A matrix 0 diagrams which contain no D-vertices and no double bonds. We then can write expressions for the Green's functions and irreducible self-energies for the system. In Sec. III, we use a method due to Johansson[4] to obtain the gap equations for the fluid, and in Sec. IV we obtain an expression for the effective interaction of spin-1 particle pairs at low density at the transition temperature. In Sec. V, we make some concluding remarks.

II. GENERATING FUNCTION AND GREEN'S FUNCTIONS

If we restrict ourselves to forward scattering processes and if we are interested primarily in the superfluid properties of a hard core fermi fluid, we can write the generating function as

$$B'(\beta, g, \nu_I, \nu_{II}) = \sum (\text{all different connected } 0' \text{ diagrams}) \tag{II.1}$$

$0'$ diagrams a e A-matrix 0 diagrams (cf. Ref. 1) which contain no D-vertices and no double bonds. (They are similar to the type I diagrams of Ref. 5.) The vertices in $0'$ diagrams need have no particular left–right ordering.

The first step in obtaining a meaningful theory is to resum *all* self-energy structures, since they give rise to terms polynomial in β. It is convenient to introduce the following propagator:

$$\begin{aligned} G(K_a, K_b, \lambda_a, \lambda_b) &= \lim_{\nu_I \to 0} \frac{\delta^2 B'}{\delta s_I^+(a)\delta s_I^-(b)} \\ &= \delta_{\mathbf{k}_a, \mathbf{k}_b} G_{\gamma_a\gamma_b}(\mathbf{K}_a, \lambda_a - \lambda_b) \\ &= \lim_{\nu_I \to 0} \sum[\text{all different } (1,1) \text{ diagrams}] \end{aligned} \tag{II.2}$$

and two anomalous propagators

$$G^{(+)}(K_a, K_b; \lambda_a, \lambda_b) \equiv \lim_{\nu_I \to 0} \frac{\delta^2 B'}{\delta s_I^-(a)\delta s_I^-(b)} = \delta_{\mathbf{K}_a + \mathbf{K}_b, 0} G^{(+)}_{\gamma_a\gamma_b}(\mathbf{K}_a, \lambda_a + \lambda_b)$$

$$= \lim_{\nu_I \to 0} \sum[\text{all different } (2,0) \text{ diagrams}] \tag{II.3}$$

and

$$G^{(-)}(K_a, K_b; \lambda_a, \lambda_b) \equiv \lim_{\nu_I \to 0} \frac{\delta^2 B'}{\delta s_I^+(a)\delta s_I^+(b)} = \delta_{\mathbf{K}_a + \mathbf{K}_b, 0} G^{(-)}_{\gamma_a\gamma_b}(\mathbf{K}_a, \lambda_a + \lambda_b)$$

$$= \lim_{\nu_I \to 0} \sum[\text{all different } (0,2) \text{ diagrams}]. \tag{II.4}$$

 0022-2488/79/030361-06$01.00

In addition, for future reference, we shall introduce a generalized interaction

$$I(K_a\lambda_a;K_b\lambda_b;K_c\lambda_c;K_d\lambda_d)$$
$$=\lim_{\nu_I\to 0}\frac{\delta^4 B'}{\delta s_I^-(a)\delta s_I^-(b)\delta s_I^+(c)\delta s_I^+(d)}$$
$$=I_{\gamma_a\gamma_b;\gamma_c\gamma_d}(\mathbf{K}_a\lambda_a;\mathbf{K}_b\lambda_b;\mathbf{K}_c\lambda_c;\mathbf{K}_d\lambda_d)$$
$$=\lim_{\nu_I\to 0}\sum[\text{all different (2,2) diagrams}]. \qquad \text{(II. 5)}$$

An (n_1, n_2) diagram is a connected diagram with n_1 broken lines entering and n_2 broken lines leaving.

Once the propagators and generalized interactions are defined, we can take the limit $\nu_I \to 0$ and Fourier transform the 0′ diagrams and (n_1, n_2) diagrams so that they are expressed in terms of momenta and "frequencies" rather than momenta and temperatures (cf. Ref. 4). No diagram will then contain single-particle sources. Rules for evaluating these diagrams are given in the Appendix.

It is convenient also to introduce an irreducible self-energy

$$=\Sigma_{\gamma_1\gamma_2}(1)=\lim_{\nu_I\to 0}\sum[\text{all different irreducible (1,1) diagrams}] \qquad \text{(II. 6)}$$

and two anomalous irreducible self-energies

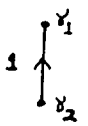
$$\equiv\Sigma^{(+)}_{\gamma_1\gamma_2}(1)=\lim_{\nu_I\to 0}\sum[\text{all different irreducible (2,0) diagrams}] \qquad \text{(II. 7)}$$

and

$$\equiv\Sigma^{(-)}_{\gamma_1\gamma_2}(1)=\lim_{\nu_I\to 0}\sum[\text{all different irreducible (0,2) diagrams}]. \qquad \text{(II. 8)}$$

We have now defined the self-energies in terms of frequencies rather than temperatures. In Eqs. (II. 6)–(II. 8), the dotted lines are included for labeling purposes only. No factors are to be associated with them. An irreducible diagram is one which cannot be cut into two parts by cutting one line. Irreducible (n_1, n_2) diagrams are evaluated by the rules in the Appendix. We have used the notation $\Sigma_{\gamma_1\gamma_2}(1)\equiv\Sigma_{\gamma_1\gamma_2}(\mathbf{K}_1, z_1)$.

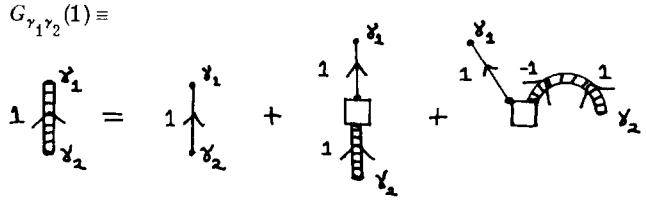
We now can obtain expressions for the propagators in Eqs. (II. 2)–(II. 4) in terms of the self-energies in Eqs. (II. 6)–(II. 8). Let us denote



$$=G^0_{\gamma_1\gamma_2}(1)=\delta_{\gamma_1,\gamma_2}/(iz_1-\omega_1). \qquad \text{(II. 9)}$$

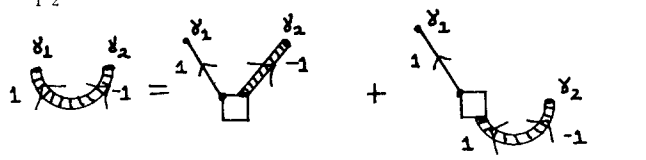
We can then write the following equations for the propagators in Eqs. (II. 2)–(II. 4):

$G_{\gamma_1\gamma_2}(1)\equiv$



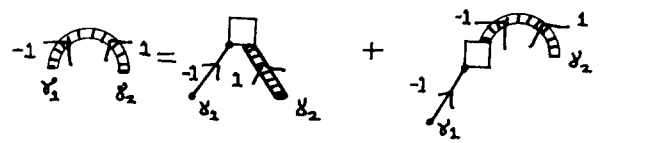
(II. 10)

$G^{(+)}_{\gamma_1\gamma_2}(1)\equiv$



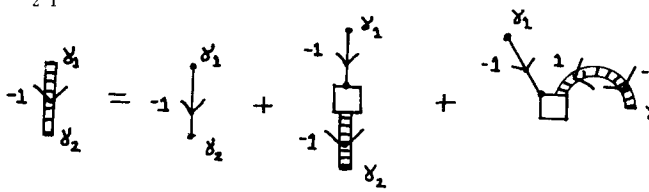
(II. 11)

$G^{(-)}_{\gamma_1\gamma_2}(1)\equiv$



(II. 12)

$G_{\gamma_2\gamma_1}(1)\equiv$



(II. 13)

where $G_{\gamma_1\gamma_2}(-1)\equiv G_{\gamma_1\gamma_2}(-\mathbf{K}_1, -z_1)$. Equation (II. 13) is identical to Eq. (II. 10) but is included for completeness. Note that by symmetry

$$G^{(+)}_{\gamma_1\gamma_2}(\mathbf{K}_1 z_1)=G^{(+)}_{\gamma_2\gamma_1}(-\mathbf{K}_1, -z_1) \qquad \text{(II. 14)}$$

and

$$G^{(-)}_{\gamma_1\gamma_2}(\mathbf{K}_1, z_1)=G^{(-)}_{\gamma_2\gamma_1}(-\mathbf{K}_1, -z_1). \qquad \text{(II. 15)}$$

The propagators and self-energies are 2×2 matrices in spin space. Equations (II. 10)–(II. 13) can be written in terms of 2×2 matrices if we introduce the notation

$$\hat{G}(1)\equiv\begin{pmatrix} G_{\uparrow\uparrow}(1) & G_{\uparrow\downarrow}(1) \\ G_{\downarrow\uparrow}(1) & G_{\downarrow\downarrow}(1) \end{pmatrix} \qquad \text{(II. 16)}$$

with similar notation for the other propagators and self-energies. Equations (II. 10)–(II. 13) respectively can now be written

$$\hat{G}(1)=\hat{G}^0(1)+\hat{G}^0(1)\,\hat{\Sigma}(1)\,\hat{G}(1)+\hat{G}^0(1)\,\hat{\Sigma}^{(+)}(1)\,\hat{G}^{(-)}(1), \qquad \text{(II. 17)}$$

$$\hat{G}^{(+)}(1)=\hat{G}^0(1)\,\hat{\Sigma}^{(+)}(1)\,\hat{G}(-1)^T+\hat{G}^0(1)\,\hat{\Sigma}(1)\,\hat{G}^{(+)}(1), \qquad \text{(II. 18)}$$

$$\hat{G}^{(-)}(1)=\hat{G}^0(-1)^T\,\hat{\Sigma}^{(-)}(1)\,\hat{G}(1)+\hat{G}^0(-1)^T\hat{\Sigma}\,(-1)^T\hat{G}^{(-)}(1), \qquad \text{(II. 19)}$$

and

$$\hat{G}(-1)^T - \hat{G}^0(-1)^T + \hat{G}^0(-1)^T\,\hat{\Sigma}\,(-1)^T\hat{G}(-1)^T + \hat{G}^0(-1)^T\,\hat{\Sigma}^{(-)}(1)G^{(+)}(1). \quad \text{(II. 20)}$$

In Eqs. (II. 18)—(II. 20) T denotes transpose of the spin matrix. Equations (II. 17)—(II. 20) yield the following solutions for the propagators in terms of the self-energies:

$$\hat{G}(1) = [1 - \hat{G}^0(1)\,\hat{\Sigma}\,(1) - \hat{G}^0(1)\,\hat{\Sigma}^+(1) \times [1 - G^0(-1)^T\,\hat{\Sigma}\,(-1)^T]^{-1}\hat{G}^0(-1)^T\textstyle\sum^{(-)}(1)]^{-1}\hat{G}^0(1), \quad \text{(II. 21)}$$

$$\hat{G}(-1)^T = [1 - \hat{G}^0(-1)^T\,\hat{\Sigma}\,(-1) - \hat{G}^0(-1)^T\,\hat{\Sigma}^{(-)}(1) \times [1 - \hat{G}^0(1)\,\hat{\Sigma}\,(1)]^{-1}\hat{G}^0(1)\,\hat{\Sigma}^{(+)}(1)]^{-1}G^0(-1)^T, \quad \text{(II. 22)}$$

$$G^{(+)}(1) = [1 - \hat{G}^0(1)\,\hat{\Sigma}\,(1)]^{-1}\hat{G}^0(1)\,\hat{\Sigma}^{(+)}(1)\hat{G}(-1)^T, \quad \text{(II. 23)}$$

and

$$G^{(-1)}(1) = [1 - \hat{G}^0(-1)^T\,\hat{\Sigma}\,(-1)^T]^{-1}G^0(-1)^T\,\hat{\Sigma}^{-}(1)\hat{G}(1). \quad \text{(II. 24)}$$

The four equations Eq. (II. 10)—(II. 13) can be written in terms of a single 4×4 matrix equation if we introduce the notation

$$\bar{\bar{G}}(1) = \begin{pmatrix} \hat{G}(1) & \hat{G}^{(+)}(1) \\ \hat{G}^{(-)}(1) & \hat{G}(-1)^T \end{pmatrix}, \quad \text{(II. 25)}$$

$$\bar{\bar{\Sigma}}(1) = \begin{pmatrix} \hat{\Sigma}(1) & \Sigma^+(1) \\ \hat{\Sigma}^{(-)}(1) & \hat{\Sigma}(-1)^T \end{pmatrix}, \quad \text{(II. 26)}$$

and

$$\bar{\bar{G}}_0(1) = \begin{pmatrix} \hat{G}_0(1) & 0 \\ 0 & \hat{G}_0(-1)^T \end{pmatrix}. \quad \text{(II. 27)}$$

Equations (II. 17)—(II. 20) now reduce to the simple form

$$\bar{\bar{G}}(1) = \bar{\bar{G}}_0(1) + \bar{\bar{G}}_0(1)\,\bar{\bar{\Sigma}}(1)\bar{\bar{G}}(1). \quad \text{(II. 28)}$$

The generalized propagator $\bar{\bar{G}}(1)$ has the solution

$$\bar{\bar{G}}(1) = [\bar{\bar{G}}_0(1)^{-1} - \bar{\bar{\Sigma}}(1)]^{-1}. \quad \text{(II. 29)}$$

We have expressed all the propagators in terms of irreducible self-energy structures. We can now use a method due to Luttinger and Ward[5] to write the generating function $B'(\beta, g, 0, \nu_{II})$ in terms of propagators and self-energies. The result is

$$B'(\beta, g, \nu_I, \nu_{II}) = \frac{1}{2}\sum_{K_1}\sum_{n_1}\mathrm{Tr}_4\ \bar{\bar{\Sigma}}(1)\bar{\bar{G}}(1) + \ln[\bar{\bar{\Sigma}}(1) - \bar{\bar{G}}_0(1)^{-1}] + B''\{\bar{\bar{G}}\}, \quad \text{(II. 30)}$$

$B''\{\bar{\bar{G}}\} = \sum$ [all different irreducible 0′ diagrams with solid lines replaced by propagators Eqs. (II. 10)—(II. 13)].

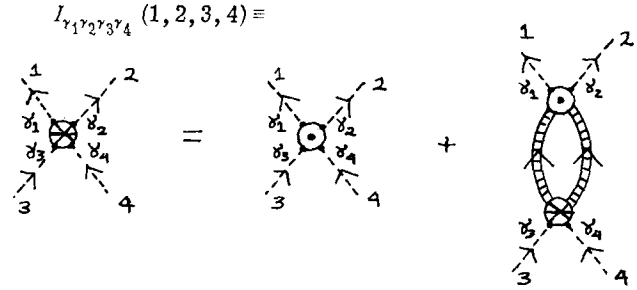
Irreducible 0′ diagrams contain no self-energy structures. (See also Ref. 7 where a similar expression is given for the grand potential of a superfluid Fermi system.)

In Eqs. (II. 29) and (II. 30 we have expressed the propagators and the generating function $B'(\beta, g, 0, \nu_{II})$ in a form in which all self-energy effects have been resumed. We can now take the limit $\nu_{II} \to 0$ and obtain an equation for the gap in terms of reaction matrices.

III. GAP EQUATION

We shall use a method due to Johansson[4] to find a self-consistent equation for the gap. Let us first note that the generalized interaction $I_{\gamma_1\gamma_2\gamma_3\gamma_4}(1,2,3,4)$ (now written in terms of frequencies) satisfies the equation

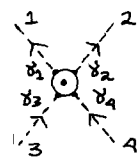
$I_{\gamma_1\gamma_2\gamma_3\gamma_4}(1,2,3,4) \equiv$



(III. 1)

where

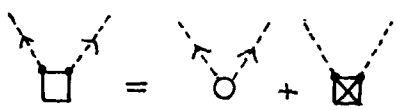
$J_{\gamma_1\gamma_2\gamma_3\gamma_4}(1,2,3,4) \equiv$



$= \sum$[call different ⟲-irreducible (2, 2) diagrams]. (III. 2)

A ⟲-irreducible (2, 2) diagram has two broken lines entering and two broken lines leaving and cannot be cut in two parts by breaking two identically directed lines. (2, 2) diagrams are evaluated in the same way as 0′ diagrams except that no factors are to be associated with the dotted lines. We have used the notation $I_{\gamma_1\gamma_2\gamma_3\gamma_4}(1,2,3,4) \equiv I_{\gamma_1\gamma_2\gamma_3\gamma_4}(\mathbf{K}_1 z_1; \mathbf{K}_2 z_2; \mathbf{K}_3 z_3; \mathbf{K}_4 z_4)$.

We can now write the following integral equations for the anomalous self-energies:

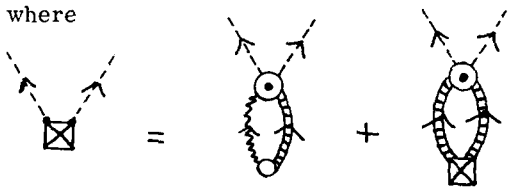


(III. 3)

and

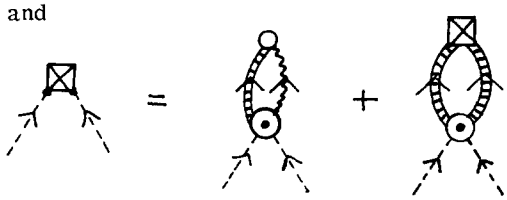
(III. 4)

where

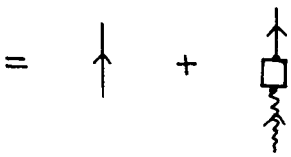


(III. 5)

and



(III. 6)



(III. 7)

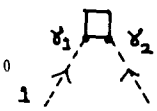
The gap functions are defined as

$$\Delta^{(+)}_{\gamma_1\gamma_2}(1) \equiv \lim_{\nu_{II}\to 0}$$

$$= \lim_{\nu_{II}\to 0} \sum\nolimits^{(+)}_{\gamma_1\gamma_2}(1) \qquad \text{(III. 8)}$$

and

$$\Delta^{(-)}_{\gamma_1\gamma_2}(1) \equiv \lim_{\nu_{II}\to 0}$$



$$= \lim_{\nu_{II}\to 0} \sum\nolimits^{(-)}_{\gamma_1\gamma_2}(1). \qquad \text{(III. 9)}$$

When we take the limit $\nu_{II} \to 0$, the first terms on the right-hand side of Eqs. (III. 3)–(III. 6) will go to zero. If the generalized interaction $I_{\gamma_1\gamma_2\gamma_3\gamma_4}(1,2,3,4)$ is well behaved, then the second terms will also be zero and the gap itself will go to zero in the limit $\nu_{II} \to 0$. In this case there will be no condensed phase in the Fermi fluid. However, if the generalized interaction is singular (infinite), the limit $\nu_{II} \to 0$ is not well defined and the second term on the right-hand side of Eqs. (III. 4)–(III. 8) need not be zero. In this case the gap can be finite and gauge symmetry of the Fermi fluid can be broken.

In the case when gauge symmetry is broken, Eqs. (III. 1)–(III. 9) yield the following equations for the gap:

$$\Delta^{(+)}_{\gamma_1\gamma_2}(1) = \sum_{\gamma_3\cdots\gamma_6} \sum_{2} J_{\gamma_1\gamma_2;\gamma_3\gamma_4}(1,-1,2,-2) \times G_{\gamma_3\gamma_4}(2) G_{\gamma_4\gamma_6}(-2) \Delta^{(+)}_{\gamma_5\gamma_6}(2) \qquad \text{(III. 10)}$$

and

$$\Delta^{(-)}_{\gamma_1\gamma_2}(1) = \sum_{\gamma_3\cdots\gamma_6} \sum_{2} \Delta^{(-)}_{\gamma_5\gamma_6}(2) G_{\gamma_5\gamma_3}(2) \times G_{\gamma_6\gamma_4}(-2) J_{\gamma_3\gamma_4\gamma_1\gamma_2}(2,-2;1,-1). \qquad \text{(III. 11)}$$

[Note that our equation for the gap is more general than that of Johansson because we have used the full propagators $G_{\gamma_1\gamma_2}(1)$.] In order to determine if gauge symmetry of the Fermi fluid is broken, we must study the effective interaction $J_{\gamma_1\gamma_2\gamma_3\gamma_4}(1,-1,2,2)$.

IV. GAP EQUATION FOR A LOW DENSITY HARD SPHERE GAS

We shall now obtain an expression for the effective interaction $J_{\gamma_1\gamma_2\gamma_3\gamma_4}(1,-1;2,-2)$ in the forward scattering approximation. From Refs. 1 and 8, we can write the basic vertex function

$$A\begin{pmatrix} \mathbf{K}_1\gamma; \mathbf{K}_2\gamma_2 \\ \mathbf{K}_3\gamma_3; \mathbf{K}_4\gamma_4 \end{pmatrix}$$

as

$$A\begin{pmatrix} \mathbf{K}_1\gamma_1 \mathbf{K}_2\gamma_2 \\ \mathbf{K}_3\gamma_3 \mathbf{K}_\gamma{}_4 \end{pmatrix} = C^2(\mathbf{K}_{12}) \langle \mathbf{K}_1\gamma_1; \mathbf{K}_2\gamma_2 | A | \mathbf{K}_3\gamma_3; \mathbf{K}_4\gamma_4 \rangle^{(a)}$$

$$= \delta(\mathbf{K}_1 + \mathbf{K}_2 - \mathbf{K}_3 - \mathbf{K}_4) \sum_{l,m} \frac{2l+1}{4\pi} P_l(\hat{K}_{12}\cdot\hat{K}_{34}) \times [\langle \gamma_1\gamma_2 | \gamma_3\gamma_4 \rangle - (-1)^l \langle \gamma_1\gamma_2 | \gamma_4\gamma_3 \rangle \left(\frac{2}{\pi} \frac{\cos^2\delta_l(K_{12})}{K_{12}} \langle K_{34} l | A | K_{12} l \rangle \right), \qquad \text{(IV. 1)}$$

where $P_l(\hat{K}_{12}\cdot\hat{K}_{34})$ is the lth order Legendre polynomial, $\mathbf{K} = \frac{1}{2}(\mathbf{K}_1 - \mathbf{K}_2)$, $\hat{K}_{12} = \mathbf{K}_{12}/|\mathbf{K}_{12}|$, and $\delta_l(k_{12})$ is the lth order phase shift. We will be interested primarily in the effect of long wavelength spin fluctuations on the particle–pair interaction. For this case, we can approximate the basic vertex function by its value for forward scattering

$$A\begin{pmatrix} \mathbf{K}_1\gamma_1 & \mathbf{K}_2\gamma_2 \\ \mathbf{K}_3\gamma_3 & \mathbf{K}_4\gamma_4 \end{pmatrix} \approx \delta(\mathbf{K}_1 + \mathbf{K}_2 - \mathbf{K}_3 - \mathbf{K}_4)(-1) \sum_{l,m} \left(\frac{2l+1}{4\pi} \right) \times [\langle \gamma_1\gamma_2 | \gamma_3\gamma_4 \rangle - (-1)^l \langle \gamma_1\gamma_2 | \gamma_4\gamma_3 \rangle] \left(\frac{2}{\pi} \frac{\cos^2\delta_l(K_{12})}{K_{12}} \tan\delta_l(K_{12}) \right). \qquad \text{(IV. 2)}$$

At very low densities, where the momentum of the interacting particles is low, we have

$$\tan\delta_l(K) \simeq -(Ka)^{2l+1}/(2l+1)[(2l-1)!!]^2 \qquad \text{(IV. 3)}$$

where K is the relative momentum of the scattering particles and a is the radius of the hard sphere. The phase shifts take the form

$$\delta_0(K) \approx -Ka + O[(Ka)^3] \qquad \text{(IV. 4)}$$

and

$$\delta_l(K) \approx O[(Ka)^{2l+1}]. \qquad \text{(IV. 5)}$$

Thus at low density, for forward scattering, we have

$$A\begin{pmatrix} \mathbf{K}_1\gamma_1 & \mathbf{K}_2\gamma_2 \\ \mathbf{K}_3\gamma_3 & \mathbf{K}_4\gamma_4 \end{pmatrix} \approx \delta(\mathbf{K}_1 + \mathbf{K}_2 - \mathbf{K}_3 - \mathbf{K}_4) \frac{4\pi a}{(2\pi)^3} [\langle \gamma_1\gamma_2 | \gamma_3\gamma_4 \rangle - \langle \gamma_1\gamma_2 | \gamma_4\gamma_3 \rangle \qquad \text{(IV. 6)}$$

Let us now expand the spin wavefunctions in terms of eigenstates of the total spin of the particles, $|s, s_z\rangle$. Thus $|\uparrow\uparrow\rangle = |1,1\rangle$, $|\downarrow\downarrow\rangle = |1,-1\rangle$, $|\uparrow\downarrow\rangle = (1/\sqrt{2})\{|1,0\rangle + |0.0\rangle\}$ and $|\downarrow\uparrow\rangle = (1/\sqrt{2})\{|1,0\rangle - |0,0\rangle\}$. The states $|s, s_z\rangle$ are orthonormal. It is easy to see that

$$A\begin{pmatrix} \mathbf{K}_1\uparrow & \mathbf{K}_2\uparrow \\ \mathbf{K}_3\uparrow & \mathbf{K}_4\uparrow \end{pmatrix} = 0, \qquad \text{(IV. 7)}$$

$$A\begin{pmatrix}\mathbf{K}_1\uparrow & \mathbf{K}_2\downarrow\\ \mathbf{K}_3\uparrow & \mathbf{K}_4\downarrow\end{pmatrix}=\delta(K_1+K_2-K_3-K_4)\frac{(4\pi a)}{(2\pi)^3}, \qquad \text{(IV. 8)}$$

and

$$A\begin{pmatrix}\mathbf{K}_1\uparrow & \mathbf{K}_2\downarrow\\ \mathbf{K}_3\downarrow & \mathbf{K}_4\uparrow\end{pmatrix}=-\delta(K_1+K_2-K_3-K_4)\frac{(4\pi a)}{(2\pi)^3}. \qquad \text{(IV. 9)}$$

Thus, parallel spin pairs ($S=1$, $S_z=1$) do not experience the hard core directly because of exchange effects. We further note that the effective interaction for particles with spin $S=1$ and $S_z=0$ is

$$A\begin{pmatrix}K_1 & K_2\\ K_3 & K_4\end{pmatrix}_{\substack{S=1\\ S_z=0}}=\left[A\begin{pmatrix}\mathbf{K}_1\uparrow & \mathbf{K}_2\downarrow\\ \mathbf{K}_3\uparrow & \mathbf{K}_4\downarrow\end{pmatrix}+A\begin{pmatrix}\mathbf{K}_1\uparrow & \mathbf{K}_2\downarrow\\ \mathbf{K}_3\downarrow & \mathbf{K}_4\uparrow\end{pmatrix}\right]=0, \qquad \text{(IV. 10)}$$

and for particles pairs with $S=0$ and $S_z=0$ is

$$A\begin{pmatrix}\mathbf{K}_1 & \mathbf{K}_2\\ \mathbf{K}_3 & \mathbf{K}_4\end{pmatrix}=\left[A\begin{pmatrix}\mathbf{K}_1\uparrow & \mathbf{K}_2\downarrow\\ \mathbf{K}_3\uparrow & \mathbf{K}_4\downarrow\end{pmatrix}-A\begin{pmatrix}\mathbf{K}_1\uparrow & \mathbf{K}_2\downarrow\\ \mathbf{K}_3\downarrow & \mathbf{K}_4\uparrow\end{pmatrix}\right]$$
$$=\delta(\mathbf{K}_1+\mathbf{K}_2-\mathbf{K}_3-\mathbf{K}_4)\frac{(4\pi a)}{(2\pi)^3}. \qquad \text{(IV. 11)}$$

Thus only pairs with total spin zero experience the hard core directly.

The entire effective interaction for particle pairs with spin $S=1$, $S_z=1$ mediated by spin fluctuations is given by the diagrams in Fig. 1. At low densities, the important contribution comes from those terms containing one loop since each loop is proportional to K_f. It is easiest to write the effective interaction at the critical temperature where the gap is zero since the gap will introduce anisotropies into the system.[3] Furthermore, for simplicity we shall neglect self-energy effects in the propagators. We then obtain

$$J(\mathbf{Q},\epsilon_l)_{S=1,\,S_z=1}=J(\mathbf{Q},\epsilon_l)_{\uparrow\uparrow;\uparrow\uparrow}$$
$$=\frac{2\pi a}{(2\pi)^3}\frac{1}{1+4\pi a\chi_0(\mathbf{Q}\epsilon_l)}-\frac{1}{1-4\pi a\chi_0(\mathbf{Q}\epsilon_l)}$$
$$\approx\frac{-(4\pi a)^2}{(2\pi)^3}\chi_0(\mathbf{Q}\epsilon_l) \qquad \text{(IV. 12)}$$

where $\mathbf{Q}=\mathbf{K}_1-\mathbf{K}_2$ and $\epsilon_l=z_1-z_2$. The response function $\chi_0(\mathbf{Q},\epsilon_l)$ is defined as usual as

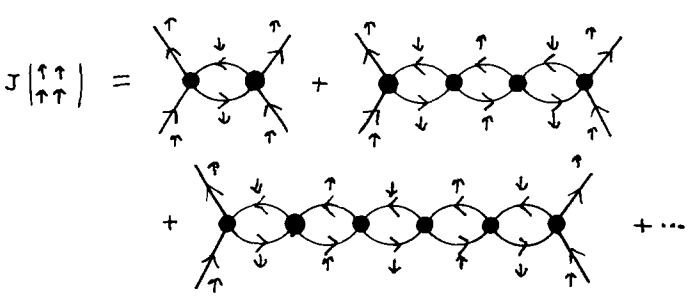


FIG. 1. Diagrams which contribute to the effective interaction of particle pairs with spin $S=1$, $S_z=1$.

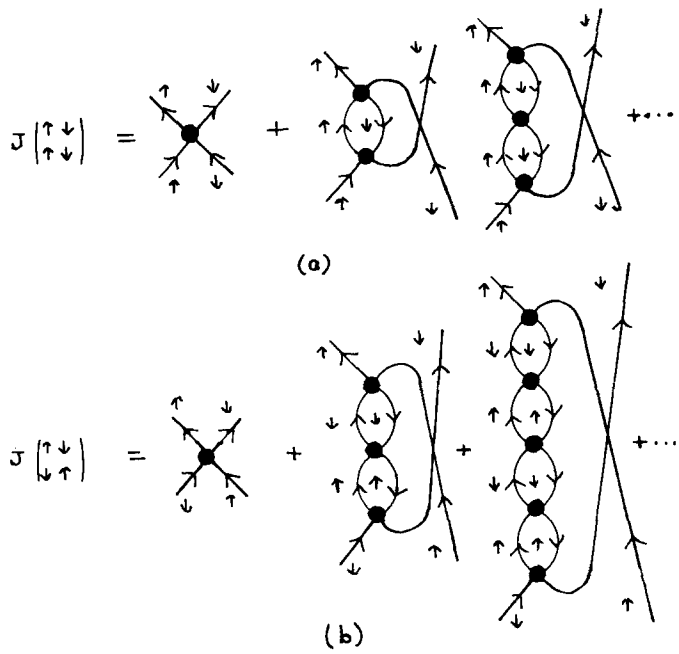


FIG. 2. Diagrams which contribute to the effective interaction of particle pairs with spin $S=1$, $S_Z=0$.

$$\chi_0(\mathbf{Q},\epsilon_l)=-\frac{1}{\beta}\sum_n\int\frac{d\mathbf{K}}{(2\pi)^3}\left(\frac{1}{i\epsilon_l+iz-\omega(\mathbf{K}+\mathbf{Q})}\right)\left(\frac{1}{iz-\omega(\mathbf{K})}\right) \qquad \text{(IV. 13)}$$

and is positive in the limit $\mathbf{Q}\to 0$, $\epsilon_l\to 0$. In considering long wavelength and low frequency spin fluctuations, we are limiting ourselves to small $\mathbf{Q}$ and ϵ_l and forward scattering, $|\mathbf{K}_1|\approx K_f$ and $|\mathbf{K}_2|\sim K_f$ and $\chi_0(\mathbf{Q},\epsilon_l)\approx\chi_0(00)\approx mK_f/2\pi^2$. The effective interaction between parallel spin particles is negative.

The effective interaction for particle pairs of spin $S=1$ and $S_z=0$ is given by the sum of the diagrams in Figs. 2a and 2b. We find that for particle pairs with $S=1$, $S_z=0$, the effective interaction is also

$$J(\mathbf{Q}\epsilon_l)_{s=1,\,s_z=0}=J(\mathbf{Q},\epsilon_l)_{\uparrow\downarrow;\uparrow\downarrow}+J(\mathbf{Q}\epsilon_l)_{\uparrow\downarrow;\downarrow\uparrow}$$
$$=\frac{2\pi a}{(2\pi)^3}\left[\frac{1}{1+4\pi a\chi_0(\mathbf{Q}\epsilon_l)}-\frac{1}{1-4\pi a\chi_0(\mathbf{Q}_1\epsilon_l)}\right]$$
$$\approx\frac{-(4\pi a)^2}{(2\pi)^3}\chi_0(\mathbf{Q}\epsilon_l). \qquad \text{(IV. 14)}$$

Thus, for low-density hard-core particles at the Fermi surface which are mediated by long-wavelength low-frequency spin fluctuations, the effective interaction between pairs of particles with total spin $S=1$ is attractive. At the critical temperature, the effective interaction for pairs with ($S=1$, $S_z=1$) and ($S=1$, $S_z=0$) are equal. This result is rigorous for particle pairs at the Fermi surface. We can now apply the usual arguments to Eqs. (IV. 11) and (IV. 12) to obtain an expression for the critical temperature (see, for example, Ref. 9, Chap. 7).

V. CONCLUSIONS

In this paper, we have considered a rather simple application of the binary expansion in order to illustrate

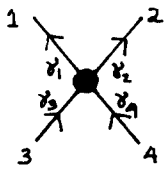
the meaning of various quantities in the theory, and because it enables us to compare various features of the binary expansion and quantum field theory. We see that the basic structure of the two theories is quite similar except that the binary expansion contains extra terms (the D-vertices) and the two theories give different weights to repeated scattering processes. The main difference in the two theories, however, is in their treatment of basic interactions in the fluid. The basic vertices in the binary expansion are given in terms of the exact solution to the two-body problem, whereas the basic vertices in quantum field theory, for particles with hard cores, are infinite and meaningless, since they involve the Fourier transform of the Lennard–Jones potential.

If we go to higher densities, use of the forward scattering approximation will certainly not be satisfactory, and we must consider the full scattering problem. In that case, the simple method of Matsubara for summing "secular" effects is no longer applicable and we must use the method of Laplace transform discussed in Ref. 10. While the structure of the propagators will be more complicated when using the Laplace transformation, the method for obtaining the gap will be the same as that considered here.

APPENDIX: RULES FOR EVALUATING 0′ DIAGRAMS AND (n_1, n_2) DIAGRAMS

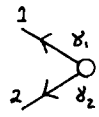
(A. i) Label the lines from 1 to n, where n is the number of lines, and associate to the jth line a momentum K_i and an energy $Z_j = \pi n_j/\beta$, where n_j is a positive or negative odd integer.

(A. ii) Associate with each A-vertex a factor

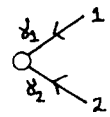


$$= -\beta\delta_{n_1+n_2,\,n_3+n_4} A\begin{pmatrix} \mathbf{K}_1\gamma_1;\mathbf{K}_2\gamma_2 \\ \mathbf{K}_3\gamma_3;\mathbf{K}_4\gamma_4 \end{pmatrix}.$$

(A. iii) Associate with each S_{II}^{+}-vertex a factor

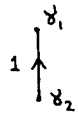


$$= -\beta\delta_{n_1+n_2,\,0} S^{(+)}\begin{pmatrix} \mathbf{K}_1\gamma_1 & \mathbf{K}_2\gamma_2 \\ 0 & 0 \end{pmatrix},$$



$$= -\beta\delta_{n_1+n_2,\,0} S^{(-)}\begin{pmatrix} 0 & 0 \\ \mathbf{K}_1\gamma_1 & \mathbf{K}_2\gamma_2 \end{pmatrix}.$$

(A. iv) With the jth line associate a factor



$$= \frac{\delta_{\gamma_1\gamma_2}}{(iz_1 - \omega_1)} \equiv G^0_{\gamma_1\gamma_2}(1),$$

where $z_1 = \pi n_1/\rho$ (n_1 odd integer) and $\omega_1 = k_1^2/2m - q$.

(A. vi) Multiply the entire expression by a factor $\epsilon^{P_B}(1/S)$, where P_B and S are defined in Ref. (Rule B8).

$$= \epsilon \sum_{k_1} \ln(1 - \epsilon e^{-\beta\omega_1}).$$

With each closed loop that attaches to a single vertex, associate a factor

$$= \frac{\delta_{\gamma_1\gamma_2} e^{z_1 0^+}}{(iz_1 - \omega_1)}.$$

(A. vi) Multiply the entire expression by a factor $\epsilon^{P_B}(1/S)$, where P_B and S are defined in Ref. (Rule B8).

(A. vii) Sum over momenta $\mathbf{K}_i$ spins, γ_i and odd integers n_i, and multiply by $1/\beta$ for each internal line.

[1] L. E. Reichl, J. Math. Phys. 20, 352 (1979).
[2] A. Layzer and D. Fay, Intern. J. Mag. 1, 135 (1971).
[3] P. W. Anderson and W. F. Brinkman, Phys. Rev. Lett. 30, 1108 (1973) and A. J. Leggett, Rev. Mod. Phys. 47, 331 (1975).
[4] B. Johansson, Physica 32, 2164 (1966).
[5] L. E. Reichl, J. Math. Phys. 13, 2023 (1976).
[6] J. M. Luttinger and J. C. Ward, Phys. Rev. 118, 1417 (1960).
[7] D. Rainer and J. W. Serene, Phys. Rev. B 13, 4745 (1976).
[8] L. E. Reichl, J. Math. Phys. 17, 2007 (1976).
[9] J. R. Schieffer, *Theory of Superconductivity* (Benjamin, New York, 1964).
[10] L. E. Reichl, J. Math. Phys. 17, 2034 (1976).

Classification and construction of finite dimensional irreducible representations of the graded algebras; application to the (Sp(2n); 2n) algebra

Miroslav Bednář and Vladimír F. Šachl

Institute of Physics, Czechoslovak Academy of Sciences, 180 40 Prague 8, Czechoslovakia
(Received 16 February 1978)

A method, which enables us to construct the finite-dimensional representations of the graded Lie algebras [explicitly the GLA (Sp(2N); 2N)] on the irreducible tensors is suggested. Those tensors are constructed with the aid of the specifically symmetrized products of vectors from the fundamental [i.e., $(2N+1)$-dimensional] representation space of the graded Lie algebra. The tensors, on which it is possible to represent the graded algebra (Sp(2N); 2N) irreducibly, represent a generalization of tensors which are known from the general representation theory of the symplectic Lie algebra Sp(2N). The knowledge of the irreducible tensors of the algebra Sp(2N) gives us then the possibility of solving the problems of classification as well as construction of the irreducible tensors of the graded algebra (Sp(2N); 2N). For illustration, by using the suggested method of tensors the irreducible representations of the simplest graded algebra, i.e., of the algebra (Sp(2); 2) are constructed.

1. INTRODUCTION

In the physical as well as in the mathematical literature much attention has been devoted to the study of graded Lie algebras during the several recent years. By using various methods a number of authors[1-4] solved the problem of classification of the graded Lie algebras and at present it seems that the problem of classification has been fully solved.[4] The problem which is left and which should necessarily be further solved, is the problem of classification and construction of the irreducible representations of the particular graded Lie algebras.

As far as the use of the graded algebras, for instance, in elementary particle physics for the purpose of classification of particles, it is naturally necessary to know the structure of the irreducible representations—of the multiplets of the graded algebras, since the vectors of the basis of these irreducible multiplets should correspond to the particular elementary particles. One of the first attempts to using the graded Lie algebra spl(2, 1),[5] [denoted also by (SU(2)⊗ SU(1); $2 \oplus \bar{2}$) according to Ref. 3] was for the purpose of classification of elementary particles (see Ref. 6). The particles as mesons and baryons with the same isospin and hypercharge, but different spins and baryon numbers were classified in the frame of so-called supersymmetric octets—the eight-dimensional irreducible multiplets of the graded algebra spl (2, 1). We can, in principle, also use more complicated graded algebras in a similar way, of course, only under the assumption that we know the structure of the irreducible multiplets, i. e., the structure of the representations of these graded algebras. The irreducible representations have been studied systematically and explicitly constructed only for the two simplest graded Lie algebras: (Sp(2); 2), see Refs. 2 and 5, and the already above-mentioned algebra spl(2, 1), see Ref. 5.

Some general statements, which regard representations of the graded algebras are given in the paper by Kac.[4] The most important of them is the theorem according to which the irreducible representations of the graded algebras are characterized by the highest weights, which is similar to the case of the representations of the classical Lie algebras. At present, however, the problem of the structure of the representations (the representation spaces) which are characterized by the highest weights (e. g., the dimensions of these representations) are not known and the matrix realizations of the generators of the graded Lie algebra in the representation spaces specified by the corresponding highest weights has not been solved yet.

One of the possibilities of how to approach the solution of these—up to now not yet, generally, solved—problems in the case of the graded algebra (Sp(2N); 2N) is the subject of this paper.

The content of the particular sections is as follows: The structure of the graded algebra (Sp(2N); 2N) is mentioned in Sec. 2. In Sec. 3 we study the matrix realization of generators of the graded algebra (Sp(2N); 2N) in the $(2N+1)$-dimensional space $\mathcal{R}$ (in which the lowest-dimensional, i. e., the fundamental, representation acts), and the matrix realization of the generators of the graded algebra in the spaces $\mathcal{R}^{\otimes}\mathcal{R}$, $\mathcal{R}\otimes\mathcal{R}\otimes\mathcal{R}$, $\mathcal{R}^{\otimes}\mathcal{R}\otimes\mathcal{R}\otimes\cdots$ of tensor products of the fundamental representations $\mathcal{R}$. In Sec. 4 the method of construction of the irreducible tensors of the graded algebra (Sp(2N); 2N) is presented. In the last section the irreducible tensor representations of the simplest graded algebra (Sp(2); 2) are explicitly given.

2. STRUCTURE OF THE GRADED ALGEBRA (Sp(2N); 2N)

The graded Lie algebra (GLA) (Sp(2N); 2N) is generated by the following operators:

(i) The Sp(2N) operators X_{ij} $i,j = -N,\ldots,-1,1,\ldots,N$ (see Ref. 7), from which, of course, only $N(2N+1)$ are independent generators.

The operators X_{ij} fulfill the relations

$$X_{ij} = -\epsilon_i\epsilon_j X_{-j-i}, \tag{2.1}$$

where $\epsilon_i = +1$ for $i > 0$, $\epsilon_i = -1$ for $i < 0$.

 0022-2488/79/030367-05$01.00

(ii) The vector [w. r. t. the Sp(2N)] operators V_k, $k=-N,\ldots,-1,1,\ldots,N$.

The graded algebra generated by the X_{ij} and V_k is then defined by the following commutation and anticommutation relations, see Ref. 8:

$$[X_{ij},X_{kl}]=\delta_{kj}X_{il}-\delta_{il}X_{kj}+\epsilon_i\epsilon_j\delta_{-ij}X_{k-i}+\epsilon_j\epsilon_k\delta_{-ik}X_{-jl}, \tag{2.2}$$

$$[X_{ij},\ V_k]=(Q_{ij})_{lk}V_l,$$

$$\{V_k,\ V_l\}=(gQ_{ij})_{kl}X_{ji},$$

where g is the matrix with the matrix elements $g_{ij}=\epsilon_i\delta_{i-j}$, $i,j=-N,\ldots,-1,1,\ldots,N$, and Q_{ij} are matrices with the matrix elements

$$(Q_{ij})_{kl}=\delta_{il}\delta_{kj}-\epsilon_i\epsilon_j\delta_{-ik}\delta_{-jl}. \tag{2.3}$$

The matrices Q_{ij} represent the generators X_{ij} in the lowest dimensional (i. e., 2N-dimensional) representation of the symplectic algebra Sp(2N).

3. REPRESENTATIONS OF GENERATORS OF THE GLA (Sp(2N); 2N)

A. The fundamental (2N + 1)-dimensional representation in the space R

It is possible to represent the algebra of operators X_{ij} and V_k by the matrices in the representation space R of dimension (2N+1). This representation is the fundamental (lowest dimensional) representation of the GLA (Sp(2N); 2N). If we decompose each of the (2N+1) $\times$(2N+1) matrices A in R into four blocks according to the scheme

$$A=\begin{pmatrix}(2N\times 2N) & (2N\times 1)\\ (1\times 2N) & 1\times 1\end{pmatrix},$$

in which the row and column matrix indices take up (2N+1) values $-N,\ldots,-1,1,\ldots,N,0$, (in the given order), then the generators X_{ij} in R are represented by the matrices

$$q_{ij}=\begin{pmatrix}(Q_{ij}) & (0)\\ (0) & (0)\end{pmatrix}, \tag{3.1}$$

with the nonzero "diagonal" blocks Q_{ij}, while the operators V_k are represented by the matrices

$$v_k=\begin{pmatrix}(0) & (\neq 0)\\ (\neq 0) & (0)\end{pmatrix}, \tag{3.2}$$

with the nonzero "nondiagonal" blocks, in which the nonzero matrix elements are given by

$$(v_k)_{m,0}=\sqrt{2}\,\delta_{km},\quad (v_k)_{0,m}=\sqrt{2}\,g_{km}. \tag{3.3}$$

This representation is the grade star representation in the sense of the definition given in Ref. 5. If we define for the operators $v=\left(\begin{smallmatrix}0 & A\\ B & 0\end{smallmatrix}\right)$ [see Eq. (3.2)], the grade star operation (‡) as $v^{\ddagger}=\left(\begin{smallmatrix}0 & -B^{\dagger}\\ A^{\dagger} & 0\end{smallmatrix}\right)$, then it is possible to easily verify, that for the matrices v_k, given by Eq. (3.3), the following relation

$$v_k^{\ddagger}=-g_{km}v_m$$

is valid.

B. The tensor products of the fundamental representations R in the spaces $R\otimes R$, $R\otimes R\otimes R$, $R\otimes R\otimes R\otimes R\otimes\ldots$

We mention the realization of the generators of the graded algebra (Sp(2N); 2N) in the space $R\otimes R$ of the direct product of two fundamental representations first. The generators of the algebra (Sp(2N); 2N) which act in $R\otimes R$ are denoted by $X_{ij}^{(2)}$ and $V_k^{(2)}$ for simplicity.

Both these generators are expressible with the aid of matrices q_{ij} and v_k, which act in the corresponding spaces R.

The Lie generators $X_{ij}^{(2)}$ are on $R\otimes R$ given by the prescription

$$X_{ij}^{(2)}=q_{ij}\otimes\mathrm{I}+\mathrm{I}\otimes q_{ij}, \tag{3.4}$$

where I is the (2N+1)-dimensional diagonal matrix

$$\mathrm{I}=\begin{pmatrix}(\mathrm{I}_{2N}) & 0\\ 0 & 1\end{pmatrix}. \tag{3.5}$$

The operators V_k are constructed in a somewhat complicated manner, namely

$$V_k^{(2)}=v_k\otimes\mathrm{I}+\mathrm{II}\otimes v_k, \tag{3.6}$$

where II is the (2N+1)-dimensional diagonal matrix

$$\mathrm{II}=\begin{pmatrix}(\mathrm{I}_{2N}) & (0)\\ (0) & (-1)\end{pmatrix}, \tag{3.7}$$

where I_{2N} is the 2N-dimensional unit matrix. The operators $X_{ij}^{(2)}$ and $V_k^{(2)}$, defined in $R\otimes R$ by Eqs. (3.4) and (3.6), fulfill relations (2.2). It is possible to immediately verify this statement by applying the identities

$$\{A\otimes B,\ C\otimes D\}=\tfrac{1}{2}[A,C]\otimes[B,D]+\tfrac{1}{2}[A,C]\otimes\{B,D\}, \tag{3.8}$$

$$[A\otimes B,\ C\otimes D]=\tfrac{1}{2}[A,C]\otimes\{B,D\}+\tfrac{1}{2}[A,C]\otimes[B,D], \tag{3.9}$$

and the relations

$$[v_k,\mathrm{I}]=0,\qquad \{v_k,\mathrm{II}\}=0, \tag{3.10}$$

which follow from the definitions (3.1), (3.2), (3.5), and (3.7), in which the generators of the graded algebra are expressed with the aid of matrices for the operators q_{ij}, v_k, I, and II.

We are going to consider the general case of the direct product of n-fundamental representations R, i. e.,

$$\underbrace{R\otimes R\otimes\cdots\otimes R}_{n\text{ times}}.$$

The generators of the graded algebra in

$$\underbrace{R\otimes R\otimes\cdots\otimes R}_{n\text{ times}}$$

will be denoted by $X_{ij}^{(n)}$ and $V_k^{(n)}$. These generators can be expressed in terms of the generators $X_{ij}^{(n-1)}$ and $V_k^{(n-1)}$ [which act in the space

$$\underbrace{R\otimes R\otimes\cdots\otimes R}_{(n-1)\text{ times}}$$

of the direct product of $(n-1)$ fundamental representations] in the following way:

$$X_{ij}^{(n)} = X_{ij}^{(n-1)} \otimes \mathrm{I} + \mathrm{I}^{(n-1)} \otimes q_{ij}, \qquad V_k^{(n)} = V_k^{(n-1)} \otimes \mathrm{I} + \mathrm{II}^{(n-1)} \otimes v_k. \tag{3.11}$$

Here, $\mathrm{I}^{(n-1)}$ is the unit matrix

$$\mathrm{I}^{(n-1)} = \underbrace{\mathrm{I} \otimes \mathrm{I} \otimes \cdots \otimes \mathrm{I}}_{(n-1) \text{ times}} \tag{3.12}$$

and $\mathrm{II}^{(n-1)}$ is the diagonal matrix

$$\mathrm{II}^{(n-1)} = \underbrace{\mathrm{II} \otimes \mathrm{II} \otimes \cdots \otimes \mathrm{II}}_{(n-1) \text{ times}}, \tag{3.13}$$

both defined in

$$\underbrace{R \otimes R \otimes R \otimes \cdots \otimes R}_{(n-1) \text{ times}}.$$

Using Eqs. (3.11) we can express the generators $X_{ij}^{(n-1)}$ and $V_k^{(n-1)}$ with the aid of $X_{ij}^{(n-2)}$ and $V_k^{(n-2)}$. By repeating this procedure we get, at the end, the explicit expressions for the operators $X_{ij}^{(n)}$ and $V_k^{(n)}$, which are expressed with the aid of the fundamental matrices q_{ij} and v_k (those represent the generators of the GLA in R).

4. THE METHOD OF CONSTRUCTION OF TENSORS, IRREDUCIBLE W.R.T. THE GRADED ALGEBRAS (Sp(2N); 2N)

We denote by $\mathbf{x}$ a vector from R (i.e. $\mathbf{x} \in R$); $\mathbf{x}$ is a $(2N+1)$-dimensional vector with the components x_α, $\alpha = -N, \ldots, -1, 1, \ldots, N, 0$. It is evident that the operators $X_{ij}^{(n)}$ and $V_k^{(n)}$ act in the space of $n(2N+1)$ tensor components $x_{\alpha_1}^1 x_{\alpha_2}^2 \ldots x_{\alpha_n}^n$ $(\alpha_1, \alpha_2, \alpha_3, \ldots, \alpha_n = -N, \ldots, -1, 1, \ldots, N, 0)$ of the tensor given by the direct product of n vectors, i.e.,

$$\mathbf{x}^1 \otimes \mathbf{x}^2 \otimes \cdots \otimes \mathbf{x}^n.$$

The GLA (Sp(2N); 2N) is obviously not represented irreducibly in the space of $n(2N+1)$ independent components $x_{\alpha_1}^1 x_{\alpha_2}^2 \cdots x_{\alpha_n}^n$. Naturally, the following questions arises: (i) How the irreducible tensors (w.r.t. the GLA) are classified, (ii) what is the structure of these irreducible tensors, and (iii) how they are constructed explicitly? Finally, to which irreducible components can the direct product $R \otimes R \otimes R \otimes \cdots \otimes R$ be decomposed?

To give an exhaustive answer to these questions is far from being easy. However, it is possible to easily find the answers to questions (i) and (iii).

It is well known, see Ref. 9, that any irreducible representation of the Lie algebra Sp(2N) is specified with the aid of the Young tableau $[\lambda_1, \lambda_2, \ldots, \lambda_N]$, which in a unique way determines the type of the symmetry of the corresponding tensor $T^{(r)}$ of the order $r = \lambda_1 + \lambda_2 + \cdots + \lambda_N$, on which the symplectic algebra Sp(2N) is represented irreducibly.

The components $T^{(r)}_{k_1, \ldots, k_r}$, of the tensor $T^{(r)}$ have r tensor indices $k_1, k_2, \ldots, k_r = -N, \ldots, -1, 1, \ldots, N$. These components are constructed with the aid of the specifically symmetrized products of r vectors $\mathbf{x}^1$, $\mathbf{x}^2, \ldots, \mathbf{x}^r$ (the type of the symmetry of the tensor is specified by the corresponding Young tableau), i.e., with the aid of the components

$$x_{k_1}^1 x_{k_2}^2 \cdots x_{k_r}^r, \quad k_1, k_2, \cdots, k_r = -N, \cdots, -1, 1, \cdots, N.$$

We shall therefore assume, that we know explicitly what the tensor components of the irreducible [w.r.t. Sp(2N)] tensor of the rth order are which correspond to the Young tableau $[\lambda_1, \lambda_2, \ldots, \lambda_N]$.

If we act on these irreducible components [with r indices of the type k $(= -N, \ldots, -1, 1, \ldots, N)$] by the operators $V_k^{(r)}$, we get tensor components of the new type: with $(r-1)$ indices of type k and one index of type 0. In case we act on these new components further by the operator $V_k^{(r)}$, we get the next new-type components, which have the $(r-2)$ indices of type k and two indices of type 0. In this way we can construct all tensor components of the new tensor—of the same order r—the tensor indices which already take up $(2N+1)$ values: $-N, \ldots, -1, 1, \ldots, N, 0$.

It is possible on the tensor components of the new tensor, the construction of which has been just described, to represent the graded algebra (Sp(2N); 2N) irreducibly. At the same time the representation of the GLA (Sp(2N); 2N), realized on tensors constructed in this way, is specified by the Young tableau $[\lambda_1, \ldots, \lambda_N]$, which describes the type of the symmetry of just those components [of the irreducible w.r.t. (Sp(2N);2N) tensor], which have all tensor indices of the type k $(= -N, \ldots, -1, 1, \ldots, N)$. In this way, for the GLA (Sp(2N); 2N) the problem of classification of representations can be solved. Further, the problem of the construction of the basis in the representation space and the matrix realizations of the generators of the GLA in these representations can be solved. In the next section we apply the method of tensors to the description of the irreducible representations of one of the simplest algebras, namely, the GLA (Sp(2); 2).

5. THE GRADED ALGEBRA (Sp(2); 2) REPRESENTED ON IRREDUCIBLE TENSORS

The Lie algebra Sp(2) is isomorphic to the Lie algebra SU(2) and consequently the graded algebra (Sp(2); 2) is isomorphic to the GLA (SU(2); 2) (see Ref. 2). For formal simplicity we are therefore going to consider the GLA (SU(2); 2) and to study its representations.

The GLA (SU(2); 2) is generated, Ref. 2, by the SU(2) generators Q_m, $m = 1, 2, 3$, and by the two components V_α, $\alpha = \pm \frac{1}{2}$, of the spinor [w.r.t. the SU(2)] operator V. The commutation and anticommutation relations of these operators may be written down in the form (Ref. 2):

$$[Q_m, Q_n] = i\epsilon_{mnp} Q_p, \qquad [Q_m, V_\alpha] = \tfrac{1}{2} (\tau_m)_{\beta\alpha} V_\beta, \qquad \{V_\alpha, V_\beta\} = \tfrac{1}{2} (C\tau_m)_{\alpha\beta} Q_m, \tag{5.1}$$

where $\tau_m (m = 1, 2, 3)$ are the Pauli matrices, represented by

$$\tau_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad \tau_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \qquad \tau_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

The charge conjugation matrix C is in this representa-

tion of Pauli matrices given by

$$C=\begin{pmatrix}0 & 1\\ -1 & 0\end{pmatrix}. \tag{5.2}$$

It is purposeful to introduce in the algebra SU(2) the spherical basis of the operators $Q_\pm$, Q_3, defined as

$$Q_\pm = Q_1 \pm iQ_2. \tag{5.3}$$

Definition equation (5.1) of the GLA (SU(2); 2) may be then rewritten to the following form (Ref. 5);

$$\begin{aligned}
&[Q_3, Q_\pm] = \pm Q_\pm, && [Q_+, Q_-] = 2Q_3,\\
&[Q_3, V_{\pm 1/2}] = \pm \tfrac{1}{2} V_{\pm 1/2}, && [Q_\pm, V_{\pm 1/2}] = 0, \quad [Q_\pm, V_{\mp 1/2}] = V_{\pm 1/2}.\\
&\{V_{\pm 1/2}, V_{\pm 1/2}\} = \pm \tfrac{1}{2} Q_\pm, && \{V_{\pm 1/2}, V_{\mp 1/2}\} = -\tfrac{1}{2} Q_3,
\end{aligned} \tag{5.4}$$

The space for the fundamental representation of the GLA (SU(2); 2) is three-dimensional and therefore the generators of GLA on it are represented by 3×3 matrices. It is possible to define the vectors of the basis in $\mathcal{R}$ in the following manner:

$$\chi_{1/2}=\begin{pmatrix}1\\0\\0\end{pmatrix}, \quad \chi_{-1/2}=\begin{pmatrix}0\\1\\0\end{pmatrix}, \quad \psi_0=\begin{pmatrix}0\\0\\1\end{pmatrix}. \tag{5.5}$$

The SU(2) generators Q_3 and $Q_\pm$ are represented in this basis by the matrices

$$Q_3=\tfrac{1}{2}\begin{pmatrix}1 & 0 & 0\\ 0 & -1 & 0\\ 0 & 0 & 0\end{pmatrix}, \quad Q_+=\begin{pmatrix}0 & 1 & 0\\ 0 & 0 & 0\\ 0 & 0 & 0\end{pmatrix}, \quad Q_-=\begin{pmatrix}0 & 0 & 0\\ 1 & 0 & 0\\ 0 & 0 & 0\end{pmatrix}. \tag{5.6}$$

In the first grade-star representation the operators $V_{\pm 1/2}$ are represented by the following matrices:

$$V_{+1/2}=\tfrac{1}{2}\begin{pmatrix}0 & 0 & 1\\ 0 & 0 & 0\\ 0 & 1 & 0\end{pmatrix}, \tag{5.7}$$

$$V_{-1/2}=\tfrac{1}{2}\begin{pmatrix}0 & 0 & 0\\ 0 & 0 & 1\\ -1 & 0 & 0\end{pmatrix} \tag{5.8}$$

It is easy to verify that the matrices (5.6), (5.7), and (5.8), correspondingly, represent the generators of the GLA (SU(2); 2), i.e., they fulfill relations (5.1).

We are next going to discuss the construction of further representations of the algebra (SU(2); 2) of higher dimensions.

It is well known that the algebra SU(2) is represented irreducibly on the multispinors. It is defined as (See Ref. 10)

$$\xi_s(\sigma)=\left(\frac{(2s)!}{(s+\sigma)!\,(s-\sigma)!}\right)^{1/2}\sum_P \xi_{\overbrace{1/2\cdots 1/2}^{(s+\sigma)}\ \overbrace{-1/2\cdots -1/2}^{(s-\sigma)}}^{2s} \tag{5.9}$$

(here the sum runs over all permutations P of the $2s$ indices which correspond to different orders of the indices $\frac{1}{2}, -\frac{1}{2}$), where

$$\xi_{\overbrace{1/2\cdots 1/2}^{(s+\sigma)}\ \overbrace{-1/2\cdots -1/2}^{(s-\sigma)}} = \underbrace{\chi_{1/2}\chi_{1/2}\cdots\chi_{1/2}}_{(s+\sigma)}\underbrace{\chi_{-1/2}\cdots\chi_{-1/2}}_{(s-\sigma)} \tag{5.10}$$

is the product of the $(s+\sigma)$ spinors $\chi_{1/2}$ and the $(s-\sigma)$ spinors $\chi_{-1/2}$.

The accessible values of the spin s are $\frac{1}{2}$, 1, $\frac{3}{2}$, $2,\cdots$ and that of the third spin-component σ equals $\sigma = -s, \ldots, s$.

On these multispinors the representation (characterized by the spin s) of the algebra SU(2) is realized as:

$$\begin{aligned}
Q_1\xi_s(\sigma) &= \tfrac{1}{2}\sqrt{(s+\sigma)(s-\sigma+1)}\,\xi_s(\sigma-1)\\
&\quad + \tfrac{1}{2}\sqrt{(s-\sigma)(s+\sigma+1)}\,\xi_s(\sigma+1),\\
Q_2\xi_s(\sigma) &= (i/2)\sqrt{(s+\sigma)(s-\sigma+1)}\,\xi_s(\sigma-1)\\
&\quad - (i/2)\sqrt{(s-\sigma)(s+\sigma+1)}\,\xi_s(\sigma+1), \quad Q_3\xi_s(\sigma)=\sigma\xi_s(\sigma).
\end{aligned} \tag{5.11}$$

Thus, we know the irreducible "tensors" (i.e., the multispinors) of the order $2s$, on which it is possible to represent the algebra SU(2) irreducibly. The "tensor" indices of components of this "tensor" are of the type $k(=+\frac{1}{2}, -\frac{1}{2})$, in the sense of the terminology of Sec. 4.

We shall aim further at the construction of the next tensor components of the tensor of the order $2s$, on which it would be [together with the known components denoted by $\xi_s(\sigma)$] possible to represent the whole graded algebra (SU(2); 2) irreducibly.

If we act on $\xi_s(\sigma)$ by the operator $V^{(2s)}_{-1/2}$ we get the new expressions, which contain $(2s-1)$ indices of type k $(=\pm\frac{1}{2})$ and one of type 0. These new expressions, normalized to the unit and denoted as $\xi_{s-1/2}(\sigma-\frac{1}{2})$, may be written in the form

$$\xi_{s-1/2}(\sigma-\tfrac{1}{2}) = \frac{1}{\sqrt{s+\sigma}}\left(\frac{(2s)!}{(s+\sigma)!\,(s-\sigma)!}\right)^{1/2} \sum_P \xi_{\overbrace{0\,1/2\cdots 1/2}^{(s+\sigma-1)}\ \overbrace{-1/2\cdots -1/2}^{(s-\sigma)}}^{2s}, \tag{5.12}$$

where

$$\xi_{\mu\gamma\rho\cdots\Lambda} = \eta_\mu\eta_\gamma\eta_\rho\cdots\eta_\Lambda, \tag{5.13}$$

in which $\eta_\mu = \chi_{1/2}$ for $\mu = \frac{1}{2}$, $\eta_\mu = \chi_{-1/2}$ for $\mu = -\frac{1}{2}$ and $\eta_\mu = \psi_0$ for $\mu = 0$.

We can easily verify, that the following relation is valid,

$$V^{(2s)}_{-1/2}\,\xi_s(\sigma) = -\tfrac{1}{2}\sqrt{s+\sigma}\;\xi_{s-1/2}(\sigma-\tfrac{1}{2}). \tag{5.14}$$

If we act further by the operator $V^{(2s)}_{-1/2}$ on the terms $\xi_{s-1/2}(\sigma-\frac{1}{2})$, we find simply that its action is

$$V^{(2s)}_{-1/2}\xi_{s-1/2}(\sigma-\tfrac{1}{2}) = -\tfrac{1}{4}Q_-\xi_s(\sigma). \tag{5.15}$$

It follows from Eq. (5.15) that the tensor components, which should contain two or more tensor indices of

type 0, are not included in the irreducible [w.r.t. the (SU(2); 2)] representation space, which contains $\xi_s(\sigma)$, $\xi_{s-1/2}(\sigma - \frac{1}{2})$.

Therefore, the GLA (SU(2); 2) is represented irreducibly on the space (of tensors of the order $2s$) in which the basis is formed by

$$\xi_s(\sigma), \qquad \sigma = -s, \ldots, s \quad (\text{spin } s),$$
$$\xi_{s-1/2}(\sigma), \quad \sigma = -s + \tfrac{1}{2}, \ldots, s - \tfrac{1}{2} \quad (\text{spin } s - \tfrac{1}{2}).$$

The action of the generators $V^{(2s)}_{\pm 1/2}$ of the GLA on the ξ_s and $\xi_{s-1/2}$ follows from the equations:

$$\begin{aligned} V^{(2s)}_{\pm 1/2}\xi_{s-1/2}(\sigma - \tfrac{1}{2}) &= \tfrac{1}{2}\sqrt{s - \sigma + 1}\,\xi_s(\sigma - 1), \\ V^{(2s)}_{-1/2}\xi_s(\sigma) &= -\tfrac{1}{2}\sqrt{s + \sigma}\,\xi_{s-1/2}(\sigma - \tfrac{1}{2}), \\ V_{+1/2}\xi_{s-1/2}(\sigma - \tfrac{1}{2}) &= \tfrac{1}{2}\sqrt{s + \sigma}\,\xi_s(\sigma), \\ V^{(2s)}_{+1/2}\xi_s(\sigma) &= \tfrac{1}{2}\sqrt{s - \sigma}\,\xi_{s-1/2}(\sigma + \tfrac{1}{2}). \end{aligned} \tag{5.16}$$

It follows explicitly from these equations, that the $(4s+1)$-dimensional subspace [with the basis formed by the functions $\xi_s(\sigma)$ and $\xi_{s-1/2}(\sigma)$ in the space

$$\underbrace{R \otimes R \otimes R \otimes \cdots \otimes R}_{2s \text{ times}}$$

actually forms the subspace for the irreducible representation (specified by s) of the graded algebra (SU(2); 2).

These representations were, of course, found only recently by a quite different method, and studied in Refs. 2 and 5.

The reply to the question: How to clasify the irreducible representations of the GLA (SU(2); 2) is contained in the following statement: To any integral and half-integral s, i.e., $s = \frac{1}{2}, 1, \frac{3}{2}, 2, \frac{5}{2}, 3, \cdots$ there exists the irreducible $(4s+1)$-dimensional representation of the graded Lie algebra (SU(2); 2), realized on the tensor components $\xi_s(\sigma)$ and $\xi_{s-1/2}(\sigma)$ [Eqs. (5.9) and (5.12)] of the tensor of the $2s$-order.

By the method suggested in Sec. 4, we have constructed the irreducible representations of the simplest of the graded algebras, i.e., (Sp(2); 2).

The explicit construction of the representations of the more complicated graded Lie algebras (Sp(2N); 2N), $N = 2, 3, \cdots$ and others will be given in further papers.

6. CONCLUSION

In this paper the problem of classification and the construction of the irreducible representations of graded Lie algebras (Sp(2N); 2N) is studied. A method is suggested, which enables us to construct the irreducible representations of these algebras, constructed in the framework of the tensor products of vectors from the space of the fundamental (i.e., the lowest-dimensional representation of the graded algebras (Sp(2N); 2N). The method, with the help of which the irreducible representations of the symplectic GLA's (Sp(2N); 2N) are discussed, has, of course, a general validity and can be applied to the study of representations of any classical graded algebras G ($= G_{\bar{0}} + G_{\bar{1}}$, where $G_{\bar{0}}$ is the Lie subalgebra) which is simple, or semisimple, and at which the representation $G_{\bar{0}}$ in $G_{\bar{1}}$ is fully reducible (Ref. 4).

In general, only the problem of classification of the representations of the graded Lie algebras with the aid of the highest weights (the theorem by Kac, Ref. 4) was discussed in the literature. Meanwhile, till the present day, no one had solved the problem of structure of the representations, which are specified by the highest weights. To solve this problem in the graded algebras is however, a substantially more complicated task than for the classical Lie algebras, mainly due to the fact that the representations of the graded algebras could be not fully decomposable, i.e., they can contain the invariant (w.r.t. the graded algebra) subspaces (see Ref. 4).

The constructive method presented in this paper enables us to solve the problem of the structure of representations of the GLA. It enables us to explicitly construct the basis in the representation space, which is specified by the highest weight. Furthermore, in such a basis in the representation space it enables us to represent all the generators of the GLA by the explicit matrices.

For this reason the method which uses tensors is a useful tool; with its help the up to now open problems in the theory of representations of the graded algebras can now be solved.

ACKNOWLEDGMENTS

The authors would like to thank Doctor J. Niederle and Doctor L. Trlifaj for useful comments.

[1] P.G.O. Freund and I. Kaplansky, J. Math. Phys. **17**, 228 (1976).
[2] A. Pais and V. Rittenberg, J. Math. Phys. **16**, 2062 (1975).
[3] W. Nahm, V. Rittenberg, and M. Scheunert, Phys. Lett. B **61**, 383 (1976); M. Scheunert *et al.*, J. Math. Phys. **17**, 1626, 1640 (1976).
[4] V.G. Kac, Commun. Math. Phys. **53**, 31 (1977).
[5] M. Scheunert *et al.*, J. Math. Phys. **18**, 146, 155 (1977).
[6] R.C. Hwa and C.S. Lam, Phys. Rev. D **12**, 3730 (1975).
[7] A.M. Perelomov and V.S. Popov, Yad. Fiz. **3**, 1127 (1965), C.O. Nwachuku and M.A. Rashid, J. Math. Phys. **17**, 1611 (1976).
[8] M. Bednář and V. Šachl, "Fourth Degree Casimir Operator of the Semi-Simple Graded Lie Algebra (Sp(2N); 2N)," J. Math. Phys. **19**, 1487 (1978).
[9] H. Weyl, *The Classical Groups* (Russian transl.) (Publishing House Inostrannoj Lit., Moscow, 1947), M. Hammermesh, *Group Theory* (Addison-Wesley, Reading, Mass., 1964).
[10] L.D. Landau and I.M. Lifsic, *Quantum Mechanics* (in Russian) (Publ. House GIFM Lit., Moscow, 1963).

Off-energy-shell results for scattering by a nonlocal potential. I[a)]

B. Talukdar, U. Das,[b)] and S. Mukhopadhyaya

Department of Physics, Visva-Bharati University, Santiniketan 731235, West Bengal, India
(Received 19 September 1977; revised manuscript received 2 March 1978)

The importance of analyzing the off-shell effects due to scattering by nonlocal potentials is emphasized. Analytical expressions for l-wave off-shell wavefunctions associated with Jost (irregular) and physical (outgoing wave) boundary conditions are derived for an N-term separable potential by using a differential equation approach. The half-off-shell and fully off-shell T matrices are expressed in terms of appropriate Jost functions, Fredholm determinants, and transforms of the form factors of the potential. The general results presented are then used to construct exact expressions for T matrices for Tabakin, Beregi, and Mongan potentials. In limiting cases, each of our results can be seen to yield the off-shell T matrix for the one-term Yamaguchi potential calculated by other techniques.

I. INTRODUCTION

An awkward analytical constraint associated with the short-range local potential is that the phase shift $\delta_l(k)$ is a continuous function of momentum k. There exist situations where relaxation of the constraint is necessary in order to accomodate experimental results. For example, in the vicinity of an isolated compound resonance, the phase shift for the resonant partial wave develops a jump of magnitude π. The change in phase becomes discontinuous as the width of the resonance approaches zero. Recently, it has been emphasized by Mulligan *et al.*[1] that this constraint can be relaxed in going from a local to a nonlocal potential. The nonlocal potential is thus effective in treating a much wider variety of phenomena than that encompassed with a short-range local potential.

One of the tasks in developing the description of physical processes characteristic of a nonlocal potential must be the analysis of off-shell effects due to such a potential. We attempt to do this by dealing with an N-term separable potential and solving for off-shell wavefunctions with physical (out-going wave) and Jost (irregular) boundary conditions for scattering by such potentials. These wavefunctions are useful in determining half-off-shell and off-shell T matrices. For example, the behavior of the irregular solution near the origin determines the Jost functions. Expressions for the on- and off-shell Jost functions can be employed to calculate the half-off-shell T matrix. The Fredholm determinant associated with the physical wavefunction is used to obtain the off-shell T matrix. A merit of the T matrix calculation for a separable potential based on the wavefunction approach is that it does not involve the evaluation of typical contour integrals associated with the Tabakin's[2] procedure. Further, the method can be easily extended to potentials of arbitrary rank.

In Sec. II we derive expressions for the l -wave off-shell wavefunctions by using the van Leeuwen–Reiner approach.[3,4] In Sec. III we specialize to the s-wave case, and employ our formal results to obtain T matrices for one-term separable potentials in terms of elementary transcendental functions. In particular, we consider potentials introduced by Tabakin[5] and by Beregi.[6] In Sec. IV we present similar results for two-term separable potentials. In this paper we present two such cases, the Mongan[7] and Tabakin[8] (two-term) potentials. Finally, we conclude by making some observations on our results.

II. OFF-SHELL WAVEFUNCTIONS AND FREDHOLM DETERMINANTS

We consider the scattering of a particle by a nonlocal potential $V(\mathbf{r},\mathbf{s})$. Let $\mathbf{k}$ denote the on-shell momentum related to the energy by $E=k^2+i\epsilon$, $0<\epsilon\ll 1$ and $\mathbf{q}$, an off-shell momentum. The radial part of the off-shell wavefunction $\psi_l^+(k,q,r)$ relating to the physical boundary condition satisfies the projected van Leeuwen–Reiner equation[9]

$$\left(\frac{d^2}{dr^2}+k^2-\frac{l(l+1)}{r^2}\right)\psi_l^+(k,q,r)-\int_0^\infty v_l(r,s)\psi_l^+(k,q,s)\,ds = (k^2-q^2)\hat{j}_l(qr). \tag{1}$$

Here

$$v_l(r,s)=2\pi rs\int_{-1}^{+1}d(\cos\theta)\,P_l(\cos\theta)\,v(\mathbf{r},\mathbf{s}), \tag{2}$$

and $\hat{j}_l(qr)$ stands for the Riccati-Bessel function. For an N-term separable potential $v_l(r,s)=\sum_{i=1}^N \lambda_i v_l^{(i)}(r)\,v_l^{(i)}(s)$, Eq. (1) reduces to

$$\left(\frac{d^2}{dr^2}+k^2-\frac{l(l+1)}{r^2}\right)\psi_l^+(k,q,r) = (k^2-q^2)\hat{j}_l(qr)+\sum_{i=1}^N\int_0^\infty \lambda_i v_l^{(i)}(r)\,v_l^{(i)}(s)\psi_l^+(k,q,s)\,ds. \tag{3}$$

Similarly, the off-shell wavefunction $f_l(k,q,r)$ irregular at the origin satisfies

$$\left(\frac{d^2}{dr^2}+k^2-\frac{l(l+1)}{r^2}\right)f_l(k,q,r) = (k^2-q^2)\,e^{il\pi/2}\hat{h}_l^+(qr)+\sum_{i=1}^N\lambda_i v_l^{(i)}(r)v_l^{(i)}(s)f_l(k,q,s)\,ds, \tag{4}$$

with $\hat{h}_l^+(qr)$, the Riccati–Hankel function of the first kind. The Bessel functions used here follow Messiah's[10] convention so that $\hat{h}_l^+(qr)=\hat{\eta}_l(qr)+i\hat{j}_l(qr)$. The function $\hat{\eta}_l(qr)$ stands for the Riccati–Neumann function. We work in units in which $\hbar^2/2m$ is unity.

a)Supported in part by the Department of Atomic Energy, Government of India.
b)Based in part on a thesis to be submitted by one of the authors (UD) to the Visva-Bharati University.

 0022-2488/79/030372-06$01.00

The free Green's functions[11] appropriate to physical and Jost boundary conditions are given below.

$$G_k^+(r,r') = G_k^R(r,r') - \frac{e^{-il\pi}}{k} j_l(kr)\,\hat{h}_l^+(kr') \tag{5}$$

with

$$G_k^R(r,r') = \begin{cases} (1/k)\{\hat{j}_l(kr)\hat{\eta}_l(kr') - \hat{j}_l(kr')\hat{\eta}_l(kr)\}, & r > r', \\ 0 & r < r'. \end{cases}$$

$G_k^+(r,r')$ represents the outgoing wave Green's function and

$$G_k^J(r,r') = \begin{cases} -(1/k)\{\hat{j}_l(kr)\,\hat{\eta}_l(kr') - \hat{j}_l(kr')\hat{\eta}_l(kr)\}, & r < r', \\ 0 & r > r', \end{cases} \tag{6}$$

the Green's function associated with the irregular boundary condition. In view of Eqs. (3)–(6) the integral equations for the physical, and irregular wavefunctions are given by

$$\begin{aligned}\psi_l^+(k,q,r) = {} & \hat{j}_l(qr) + \frac{1}{k}\sum_{i=1}^{N}\lambda_i\Big[\int_0^r [-\hat{j}_l(kr')\,\hat{\eta}_l(kr) \\ & + \hat{\eta}_l(kr')\,\hat{j}_l(kr)]\,v_i^{(i)}(r')\,dr' \\ & - e^{-il\pi}\hat{j}_l(kr)\int_0^\infty \hat{h}_l^+(kr')\,v_i^{(i)}(r')\,dr'\Big\} \\ & \times \int_0^\infty v_i^{(i)}(s)\psi_l^+(k,q,s)ds,\end{aligned} \tag{7}$$

and

$$\begin{aligned}f_l(k,q,r) = {} & e^{il\pi/2}\,\hat{h}_l^+(qr) \\ & - \frac{1}{k}\sum_{i=1}^{N}\lambda_i\int_r^\infty [\hat{j}_l(kr)\hat{\eta}_l(kr') - \hat{\eta}_l(kr)\hat{j}_l(kr')] \\ & \times v_i^{(i)}(r')\,dr'\int_0^\infty v_i^{(i)}(s)\,f_l(k,q,s)\,ds.\end{aligned} \tag{8}$$

One can also write an equation similar to Eq. (8) for $f_l(k,-q,r)$. The boundary conditions prescribed for $f_l(k,\pm q,r)$ have been stated by Fuda and Whiting.[4]

Equations (7) and (8) can be solved by multiplying them by $v_i^{(j)}(r)$ and integrating. This yields the following matrix equations:

$$_NA^P(k)\,X^P(k,q) = Y^P(q), \tag{9}$$

and

$$_NA^J(k)\,X^J(k,q) = Y^J(q). \tag{10}$$

In these equations the superscripts P and J relate to the physical and Jost solutions, respectively. The left-hand subscript N on A indicates that we are dealing with an N-term separable potential. The components of the column vectors X and Y are defined by

$$X_j^P(k,q) = \int_0^\infty v_i^{(j)}(s)\psi_l^+(k,q,s)\,ds, \tag{11}$$

$$X_j^J(k,q) = \int_0^\infty v_i^{(j)}(s)\,f_l(k,q,s)\,ds, \tag{12}$$

$$Y_i^P(q) = \int_0^\infty v_i^{(i)}(s)\,\hat{j}_l(qs)\,ds, \tag{13}$$

and

$$Y_i^J(q) = e^{il\pi/2}\int_0^\infty v_i^{(i)}(s)\hat{h}_l^+(qs)\,ds. \tag{14}$$

The elements of the Fredohlm determinant A are written in the form

$$\begin{aligned}{}_NA_{ij}^P(k) = {} & \delta_{ij} - \frac{\lambda_j}{k}\int_0^\infty v_i^{(i)}(r)\,dr \\ & \times\{\int_0^r[-\hat{j}_l(kr')\hat{\eta}_l(kr) + \hat{\eta}_l(kr')\hat{j}(kr)]\,v_i^{(j)}(r')\,dr' \\ & - e^{-il\pi}\int_0^\infty v_i^{(j)}(r')\hat{j}_l(kr)\hat{h}_l^+(kr')dr'],\end{aligned} \tag{15}$$

and

$$\begin{aligned}{}_NA_{ij}^J(k) = {} & \delta_{ij} + \frac{\lambda_j}{k}\int_0^\infty v_i^{(i)}(r)dr \\ & \times\int_r^\infty[\hat{j}_l(kr)\hat{\eta}_l(kr') - \hat{\eta}_l(kr)\hat{j}_l(kr')]v_i^{(j)}(r')dr'.\end{aligned} \tag{16}$$

In analogy with the on-shell case treated by Coz *et al.*[12] the algorithms presented in Eqs. (9)–(16) can be used to write the off-shell physical and Jost solutions. We have

$$\begin{aligned}\psi_l^+(k,q,r) = {} & j_l(qr) + \frac{1}{k\det_N A^P(k)}\sum_{i,j=1}^{N}\lambda_i a_{ij}^P(k)\,Y_j^P(q) \\ & \times\{\int_0^r[-\hat{j}_l(kr')\hat{\eta}_l(kr) + \hat{\eta}_l(kr')\hat{j}_l(kr)]v_i^{(i)}(r')dr' \\ & - e^{-il\pi}\hat{j}_l(kr)\int_0^\infty \hat{h}_l^+(kr')v_i^{(i)}(r')dr'\},\end{aligned} \tag{17}$$

and

$$\begin{aligned}& f_l(k,q,r) \\ & = e^{il\pi/2}\hat{h}_l^+(qr) - \frac{1}{k\det_N A^J(k)}\sum_{i,j=1}^{N}\lambda_i a_{ij}^J(k)\,Y_j^J(q) \\ & \times\int_r^\infty[\hat{j}_l(kr)\hat{\eta}_l(kr') - \hat{\eta}_l(kr)\hat{j}_l(kr')]v_i^{(i)}(r')dr'.\end{aligned} \tag{18}$$

Here a_{ij}^P and a_{ij}^J stand for the cofactors of $_NA_{ji}^P$ and $_NA_{ji}^J$, respectively.

The off-shell Jost function is defined by

$$f_l(k,q) = \frac{q^l e^{-il\pi/2}(2l+1)}{(2l+1)!!}\lim_{r\to 0} r^l f_l(k,q,r). \tag{19}$$

We have normalized the off-shell Jost function $f_l(k,q)$ so that when $q=k$, it becomes the ordinary Jost function[13] $f_l(k)$. In terms of on- and off-shell Jost functions the half-off-shell T matrix

$$T_l(k,q,k^2) = \left(\frac{k}{q}\right)^l \frac{f_l(k,q) - f_l(k,-q)}{i\pi q f_l(k)}. \tag{20}$$

For a rank N separable potential the off-shell T matrix can be written as

$$\begin{aligned}T_l(p,q,k^2) = {} & \frac{2}{\pi pq}\sum_{i=1}^{N}\lambda_i\int_0^\infty\int_0^\infty dr\,dr'\hat{j}_l(pr)v_i^{(i)}(r) \\ & \times v_i^{(i)}(r')\psi_l^+(k,q,r') \\ = {} & \frac{2}{\pi pq\det_N A^P(k)}\sum_{i,j=1}^{N}\lambda_i a_{ij}^P(k)\,Y_j^P(q)\,Y_i^P(p).\end{aligned} \tag{21}$$

Equations (9)–(21) represent the basic equations for computing T matrices for a separable potential of arbitrary rank by the wavefunction approach.

III. T MATRIX FOR ONE-TERM SEPARABLE POTENTIALS

In order to illustrate the usefulness of the general results presented in the previous section, we consider the problem of nucleon–nucleon scattering in the s state. We omit the subscript $l=0$ throughout. For reasons of physical interest we focus our attention on two such potentials—the Tabakin[5] and Beregi.[6] In configuration space both these potentials can be written in the form

$$V(r,r') = \lambda v(r)v(r'). \tag{22}$$

We shall see presently, the wavefunctions as well as T

matrices in both cases can be obtained in rather simple analytical forms.

A. Tabakin potential

Tabakin has introduced a separable nonlocal potential of the form (22). The function $v(r)$ is written as

$$v(r)=(A_1\cos\alpha_1 r+A_2\sin\alpha_1 r)e^{-\alpha_1 r}+A_3 e^{-\alpha_2 r}. \tag{23}$$

The parameters A_1, A_2, A_3, α_1, α_2, and λ for $^1s_0$ as well as for $^3s_1$ have been given in Ref. 5.

The Jost solution for this potential can be obtained in a straightforward way by specializing Eq. (18) to $l=0$ and $N=1$ and then using Eq. (23) for the potential. We thus have

$$f(k,q,r)=e^{iqr}-\frac{\lambda Y^J(q)}{\det_1 A^J(k)}\left[\frac{2A_1\alpha_1^2-A_2k^2}{4\alpha_1^4+k^4}e^{-\alpha_1 r}\sin\alpha_1 r-\frac{A_1k^2+2A_2\alpha_1^2}{4\alpha_1^4+k^4}e^{-\alpha_1 r}\cos\alpha_1 r-\frac{A_3}{\alpha_2^2+k^2}e^{-\alpha_2 r}\right], \tag{24}$$

with

$$Y^J(q)=\frac{1}{4\alpha_1^4+q^4}\left[A_1\{\alpha_1(2\alpha_1^2+q^2)+iq^3\}+A_2\{\alpha_1(2\alpha_1^2-q^2)+2i\alpha_1^2 q\}\right]+\frac{A_3(\alpha_2+iq)}{\alpha_2^2+q^2}, \tag{25}$$

and

$$\begin{aligned}\det_1 A^J(k)=&1+\frac{\lambda A_1^2(2\alpha_1^2-3k^2)}{8\alpha_1(4\alpha_1^4+k^4)}-\frac{\lambda A_2^2(2\alpha_1^2+k^2)}{8\alpha_1(4\alpha_1^4+k^4)}\\&-\frac{\lambda A_3^2}{2\alpha_2(\alpha_2^2+k^2)}-\frac{\lambda A_1A_2(2\alpha_1^2+k^2)}{4\alpha_1(4\alpha_1^4+k^4)}\\&-\frac{\lambda A_1A_3}{(2\alpha_1^2+2\alpha_1\alpha_2+\alpha_2^2)}\left[\frac{\alpha_1+\alpha_2}{\alpha_2^2+k^2}+\frac{\alpha_1k^2+\alpha_2k^2-2\alpha_1^3}{4\alpha_1^4+k^4}\right]\\&-\frac{\lambda A_2A_3}{(2\alpha_1^2+2\alpha_1\alpha_2+\alpha_2^2)}\left[\frac{\alpha_1}{\alpha_2^2+k^2}+\frac{\alpha_1k^2+2\alpha_1^2\alpha_2+2\alpha_1^3}{4\alpha_1^4+k^4}\right].\end{aligned} \tag{26}$$

The off-shell Jost functions are obtained as

$$f(k,\pm q)=1+\frac{\lambda Y^J(\pm q)}{\det_1 A^J(k)}\left(\frac{A_1k^2+2A_2\alpha_1^2}{4\alpha_1^4+k^4}+\frac{A_3}{\alpha_2^2+k^2}\right). \tag{27}$$

As noted earlier the ordinary Jost functions

$$f(\pm k)=\lim_{q\to\pm k}f(k,q). \tag{28}$$

The physical solution $\psi^+(k,q,r)$ for this potential is given by

$$\begin{aligned}\psi^+(k,q,r)=&\sin qr+\frac{\lambda Y^P(q)}{\det_1 A^P(k)}\left[\frac{A_3}{\alpha_2^2+k^2}e^{-\alpha_2 r}-\left(\frac{A_1k^2+2A_2\alpha_1^2}{4\alpha_1^4+k^4}+\frac{A_3}{\alpha_2^2+k^2}\right)e^{ikr}\right.\\&+\frac{A_1k^2+2A_2\alpha_1^2}{4\alpha_1^4+k^4}e^{-\alpha_1 r}\cos\alpha_1 r\\&\left.+\frac{A_2k^2-2A_1\alpha_1^2}{4\alpha_1^4+k^4}e^{-\alpha_1 r}\sin\alpha_1 r\right],\end{aligned} \tag{29}$$

where

$$\begin{aligned}\det_1 A^P(k)=&\det_1 A^J(k)+c(k)\left[\frac{A_1\alpha_1(2\alpha_1^2+k^2)}{4\alpha_1^4+k^4}+\frac{A_2\alpha_1(2\alpha_1^2-k^2)}{4\alpha_1^4+k^4}+\frac{A_3\alpha_2}{\alpha_2^2+k^2}\right]\\&+ikc(k)\left[\frac{A_1k^2+2A_2\alpha_1^2}{4\alpha_1^4+k^4}+\frac{A_3}{\alpha_2^2+k^2}\right]\end{aligned} \tag{30a}$$

with

$$c(k)=-\lambda\frac{A_1k^2+2A_2\alpha_1^2}{4\alpha_1^4+k^4}+\frac{A_3}{\alpha_2^2+k^2}, \tag{30b}$$

and

$$Y^P(q)=-qc(q)/\lambda. \tag{30c}$$

Combining Eqs. (20), (27), and (28), we obtain the half-off-shell T matrix

$$T(k,q,k^2)=\frac{2\lambda}{\pi kq}\frac{Y^P(k)Y^P(q)}{\det_1 A^P(k)}. \tag{31}$$

Combining Eqs. (23), (30), and the s-wave version Eq. (21) with $N=1$, we obtain the off-shell T matrix in the form

$$T(p,q,k^2)=\frac{2\lambda}{\pi pq}\frac{Y^P(p)Y^P(q)}{\det_1 A^P(k)}. \tag{32}$$

Combining Eqs. (31) and (32), we see that the expression for half-off-shell T matrix obtained from the s-wave version of Eq. (20) coincides with the limit of off-shell T matrix $T(p,q,k^2)$ as $p\to k$. This serves as a useful check on our expression for $T(k,q,k^2)$ in terms Jost functions. For other potentials which we consider in this paper we shall not include results for half-off-shell T matrices since the latter can be obtained as $T(k,q,k^2)=\lim_{p\to k}T(p,q,k^2)$. We shell, however, present results for Jost solutions and Jost functions because of their particular relevance in studies of bound state problems and half-shell reactions.

B. Beregi potential

Bergei has suggested a one-term separable potential for which the function $v(r)$ is given by

$$v(r)=e^{-\alpha_1 r}-ae^{-\alpha_2 r}. \tag{33}$$

With the values of the parameters α_1, α_2, and "a" given by Beregi it produces the characteristic features of the Tabakin $^3s_1$ potential. For the Beregi potential

$$f(k,q,r)=e^{iqr}-\frac{\lambda}{\det_1 A^J(k)}\left[\frac{1}{\alpha_1-iq}-\frac{a}{\alpha_2-iq}\right]\times\left[\frac{a}{\alpha_2^2+k^2}e^{-\alpha_2 r}-\frac{1}{\alpha_1^2+k^2}e^{-\alpha_1 r}\right]. \tag{34}$$

The corresponding off-shell Jost functions are

$$f(k,\pm q)=1-\frac{\lambda}{\det_1 A^J(k)}\left[\frac{1}{\alpha_1\mp iq}-\frac{a}{\alpha_2\mp iq}\right]\times\left[\frac{a}{\alpha_2^2+k^2}-\frac{1}{\alpha_1^2+k^2}\right]. \tag{35}$$

In Eq. (35)

$$\det_1 A^J(k)=1-\frac{\lambda}{2\alpha_1(\alpha_1^2+k^2)}-\frac{\lambda a^2}{2\alpha_2(\alpha_2^2+k^2)}$$

$$+\frac{\lambda a}{\alpha_1+\alpha_2}\left[\frac{1}{\alpha_2^2+k^2}+\frac{1}{\alpha_1^2+k^2}\right]. \tag{36}$$

The physical solution and the off-shell T matrix for the Beregi potential are as follows: the physical solution,

$$\psi^+(k,q,r)=\sin qr-\frac{\lambda q}{\det_1 A^P(k)}\left[\frac{1}{\alpha_1^2+q^2}-\frac{a}{\alpha_2^2+q^2}\right]\times\left[\left(\frac{1}{\alpha_1^2+k^2}-\frac{a}{\alpha_2^2+k^2}\right)e^{ikr}-\frac{1}{\alpha_1^2+k^2}e^{-\alpha_1 r}+\frac{a}{\alpha_2^2+k^2}e^{-\alpha_2 r}\right]; \tag{37}$$

off-shell T matrix,

$$T(p,q,k^2)=\frac{2\lambda}{\pi\det_1 A^P(k)}\left[\frac{1}{\alpha_1^2+p^2}-\frac{a}{\alpha_2^2+p^2}\right]\times\left[\frac{1}{\alpha_1^2+q^2}-\frac{a}{\alpha_2^2+q^2}\right]. \tag{38}$$

In Eqs. (37) and (38), $\det_1 A^P(k)$ is given by

$$\det_1 A^P(k)=\det_1 A^J(k)+\frac{\lambda\alpha_1}{(\alpha_1^2+k^2)^2}-\frac{\lambda a(\alpha_1+\alpha_2)}{(\alpha_1^2+k^2)(\alpha_2^2+k^2)}+\frac{\lambda a^2\alpha_2}{(\alpha_2^2+k^2)^2}+ik\left[\frac{\lambda}{(\alpha_1^2+k^2)^2}-\frac{2\lambda a}{(\alpha_1^2+k^2)(\alpha_2^2+k^2)}+\frac{\lambda a^2}{(\alpha_2^2+k^2)^2}\right]. \tag{39}$$

IV. T MATRIX FOR TWO-TERM SEPARABLE POTENTIALS

A one-term separable potential does not exhibit the characteristic short-range repulsion of the nucleon–nucleon interaction. In view of this attempts have been made to construct separable interactions of rank more than one. The simplest is, however, a rank two potential. Clearly, the purpose in introducing the second term is somehow to inject the repulsive part of the nucleon–nucleon interaction which dominates at high energies in as much the same way as a repulsive core in a local interaction. As noted earlier, in this paper we shall be concerned with two such potentials. In configuration space a two-term separable potential is written as

$$V(r,s)=-g(r)g(s)+h(r)h(s). \tag{40}$$

A. Mongan potential

Mongan has introduced a two-term separable potential in fitting the $^1s_0$ nucleon–nucleon phase shifts. For this potential the form factors are given by

$$g(r)=\sqrt{-\lambda_2}\,e^{-\alpha_2 r}, \tag{41a}$$

$$h(r)=\sqrt{\lambda_1}\,e^{-\alpha_1 r}. \tag{41b}$$

For the values of the parameters given by Mongan, his potential is a real symmetric one. Here the off-shell Jost solution obtained by combining Eqs. (18), (40), (41a), and (41b) is of the form

$$f(k,q,r)=e^{iqr}+\frac{1}{\det_2 A^J(k)}\times[I_1(k,q)e^{-\alpha_1 r}+I_2(k,q)e^{-\alpha_2 r}], \tag{42}$$

where

$$\det_2 A^J(k)=1-\frac{\lambda_1}{2\alpha_1(\alpha_1^2+k^2)}-\frac{\lambda_2}{2\alpha_2(\alpha_2^2+k^2)}+\frac{\lambda_1\lambda_2(\alpha_1-\alpha_2)^2}{4\alpha_1\alpha_2(\alpha_1+\alpha_2)^2(\alpha_1^2+k^2)(\alpha_2^2+k^2)}, \tag{43}$$

$$I_1(k,q)=\frac{\lambda_1(\alpha_1+iq)}{(\alpha_1^2+k^2)(\alpha_1^2+q^2)}-\frac{\lambda_1\lambda_2(\alpha_1+iq)}{2\alpha_2(\alpha_1^2+k^2)(\alpha_2^2+k^2)(\alpha_1^2+q^2)}+\frac{\lambda_1\lambda_2(\alpha_2+iq)}{(\alpha_1+\alpha_2)(\alpha_1^2+k^2)(\alpha_2^2+k^2)(\alpha_2^2+q^2)}, \tag{44}$$

and

$$I_2(k,q)=\frac{\lambda_2(\alpha_2+iq)}{(\alpha_2^2+k^2)(\alpha_2^2+q^2)}-\frac{\lambda_1\lambda_2(\alpha_2+iq)}{2\alpha_1(\alpha_1^2+k^2)(\alpha_2^2+k^2)(\alpha_2^2+q^2)}+\frac{\lambda_1\lambda_2(\alpha_1+iq)}{(\alpha_1+\alpha_2)(\alpha_1^2+k^2)(\alpha_2^2+k^2)(\alpha_1^2+q^2)}. \tag{45}$$

As in the case of one-term separable potentials, the off-shell Jost functions can be obtained by substituting $r=0$ in Eq. (42). The on-shell Jost functions are $f(\pm k)=\lim_{q\to\pm k}f(k,q)$. The physical wavefunction for this potential is given by

$$\psi^+(k,q,r)=\sin qr+\frac{1}{\det_2 A^P(k)}\times[Q_2(k,q)e^{-\alpha_1 r}+Q_3(k,q)e^{-\alpha_2 r}-Q_1(k,q)e^{ikr}] \tag{46}$$

where

$$Q_3(k,q)=\frac{\lambda_2\beta_2(k,q)}{\alpha_2^2+k^2}, \tag{47a}$$

$$Q_2(k,q)=\frac{\lambda_1\beta_1(k,q)}{\alpha_1^2+k^2} \tag{47b}$$

and

$$Q_1(k,q)=Q_2(k,q)+Q_3(k,q). \tag{47c}$$

In these equations $\beta_1(k,q)$ and $\beta_2(k,q)$ are

$$\beta_1(k,q)=G_2(k)Y_1^P(q)-H_{12}(k)Y_2^P(q) \tag{48a}$$

and

$$\beta_2(k,q)=G_1(k)\,Y_2^P(q)-H_{21}(k)\,Y_1^P(q), \tag{48b}$$

with

$$Y_i^P(q)=\frac{q}{\alpha_i^2+q^2}, \tag{48c}$$

$$G_i(k)=1+\frac{\lambda_i(\alpha_i^2-k^2)}{2\alpha_i(\alpha_i^2+k^2)^2}+\frac{i\lambda_i k}{(\alpha_i^2+k^2)^2}, \tag{48d}$$

and

$$H_{ij}(k)=-\frac{\lambda_j}{(\alpha_i^2+k^2)(\alpha_j^2+k^2)}\left[\frac{\alpha_i\alpha_j-k^2}{\alpha_i+\alpha_j}+ik\right],\quad i,j=1,2,\ i\neq j. \tag{48e}$$

The Fredholm determinant associated with the physical solution can be written in terms of the Fredholm determinant for the Jost solution:

$$\det_2 A^P(k)=\det_2 A^J(k)+R(k)+iI(k) \tag{49}$$

with

$$R(k)=\frac{\lambda_1\alpha_1}{(\alpha_1^2+k^2)^2}+\frac{\lambda_2\alpha_2}{(\alpha_2^2+k^2)^2}$$

$$-\frac{\lambda_1\lambda_2 k^2(\alpha_1-\alpha_2)^2}{2\alpha_1\alpha_2(\alpha_1^2+k^2)^2(\alpha_2^2+k^2)^2} \tag{50}$$

and

$$I(k)=\frac{\lambda_1 k}{(\alpha_1^2+k^2)^2}+\frac{\lambda_2 k}{(\alpha_2^2+k^2)^2}$$
$$+\frac{\lambda_1\lambda_2 k(\alpha_1\alpha_2-k^2)(\alpha_1+\alpha_2)^2}{2\alpha_1\alpha_2(\alpha_1+\alpha_2)(\alpha_1^2+k^2)^2(\alpha_2^2+k^2)^2}. \tag{51}$$

For the Mongan case IV potential the off-shell T matrix is given by

$$T(p,q,k^2)$$
$$=\frac{2}{\pi pq\det_2 A^P(k)}[\lambda_1 Y_1^P(p)\beta_1(k,q)+\lambda_2 Y_2^P(p)\beta_2(k,q)]. \tag{52}$$

B. Tabakin potential

Tabakin has parametrized a two-term separable interaction which fits the nucleon–nucleon data reasonably well. For this potential, the form factors to be used in Eq. (40) are

$$g(r)=\gamma e^{-\alpha r}, \tag{53a}$$

and

$$h(r)=\beta e^{-br}\left[\frac{d^2-b^2}{2db}\sin dr+\cos dr\right]. \tag{53b}$$

The parameters α,β,γ,b and d are defined in Ref. 8. The Jost solution for this potential is obtained in the form

$$f(k,q,r)=e^{iqr}-\frac{1}{\det_2 A^J(k)}[J_1(k,q)e^{-\alpha r}-J_2(k,q)e^{-br}$$
$$\times\cos dr+J_3(k,q)e^{-br}\sin dr] \tag{54}$$

with

$$J_1(k,q)=\frac{\gamma}{\alpha^2+k^2}\chi_1(k,q), \tag{55a}$$

$$J_2(k,q)=\frac{\beta k^2}{(b^2+k^2)^2+2d^2(b^2-k^2)+d^4}\chi_2(k,q), \tag{55b}$$

and

$$J_3(k,q)=\frac{\beta[(b^2+d^2)^2-k^2(d^2-b^2)]}{2db[(b^2+k^2)^2+2d^2(b^2-k^2)+d^4]}\chi_2(k,q). \tag{55c}$$

In Eq. (59) χ_i's are given

$$\chi_1(k,q)=\frac{\gamma(\alpha+iq)}{\alpha^2+q^2}\left[1+\frac{\beta^2\{(b^2+d^2)(b^2-k^2+d^2)-4b^2k^2\}}{16b^3\{(b^2+k^2)^2+2d^2(b^2-k^2)+d^4\}}\right]$$
$$-\frac{\gamma\beta^2\{(b^2+d^2)(b^2-k^2+d^2)-2b\alpha k^2\}}{2b\{(\alpha+b)^2+d^2\}\{(b^2+k^2)^2+2d^2(b^2-k^2)+d^4\}}$$
$$\times\left[\frac{(b^2+d^2)(b^2-q^2+d^2)+4b^2q^2}{2b\{(b^2+q^2)^2+2d^2(b^2-q^2)+d^4\}}\right.$$
$$\left.+\frac{iq^3}{(b^2+q^2)^2+2d^2(b^2-q^2)+d^4}\right] \tag{56a}$$

and

$$\chi_2(k,q)$$
$$=\beta\left[1+\frac{\gamma^2}{2\alpha(\alpha^2+k^2)}\right]\left[\frac{(b^2+d^2)(b^2-q^2+d^2)+4b^2q^2}{2b\{(b^2+q^2)^2+2d^2(b^2-q^2)+d^4\}}\right.$$
$$\left.+\frac{iq^3}{(b^2+q^2)^2+2d^2(b^2-q^2)+d^4}\right]$$
$$-\frac{\beta\gamma^2(\alpha+iq)(b^2+d^2+2b\alpha)}{2b(\alpha^2+k^2)(\alpha^2+q^2)((\alpha+b)^2+d^2)}. \tag{56b}$$

The Fredholm determinant associated with the Jost solution is

$$\det_2 A^J(k)$$
$$=\left[1+\frac{\gamma^2}{2\alpha(\alpha^2+k^2)}\right]$$
$$\left[1+\frac{\beta^2\{(b^2+d^2)(b^2-k^2+d^2)-4b^2k^2\}}{16b^3\{(b^2+k^2)^2+2d^2(b^2-k^2)+d^4\}}\right]$$
$$-\frac{\gamma^2\beta^2\{(b^2+d^2)(b^2-k^2+d^2)-2b\alpha k^2\}\{b^2+d^2+2b\alpha\}}{4b^2\{(\alpha+b)^2+d^2\}^2\{\alpha^2+k^2\}\{(b^2+k^2)^2+2d^2(b^2-k^2)+d^4\}}. \tag{57}$$

Jost functions are obtained by substituting $r=0$ in Eq. (54). The physical wavefunction for the Tabakin two-term potential is given by the expression

$$\psi^+(k,q,r)=\sin qr+\frac{1}{\det_2 A^P(k)}[s_3(k,q)e^{ikr}-s_1(k,q)e^{-\alpha r}$$
$$+s_2(k,q)e^{-br}\cos dr+s_4(k,q)e^{-br}\sin dr] \tag{58}$$

with s_i's written as

$$s_1(k,q)=\frac{\gamma\xi_1(k,q)}{\alpha^2+k^2}, \tag{59a}$$

$$s_2(k,q)=\frac{\beta k^2\xi_2(k,q)}{(b^2+k^2)^2+2d^2(b^2-k^2)+d^4}, \tag{59b}$$

$$s_3(k,q)=s_1(k,q)-s_2(k,q), \tag{59c}$$

and

$$s_4(k,q)=\frac{\beta\xi_2(k,q)\{k^2(d^2-b^2)-(b^2+d^2)^2\}}{2db\{(b^2+k^2)^2+2d^2(b^2-k^2)+d^4\}}. \tag{59d}$$

Here $\xi_1(k,q)$ and $\xi_2(k,q)$ are as follows.

$$\xi_1(k,q)=$$
$$\left[1+\frac{\beta^2\{(b^2+d^2)(b^2-k^2+d^2)-4b^2k^2\}}{16b^3\{(b^2+k^2)^2+2d^2(b^2-k^2)+d^4\}}\right.$$
$$\left.+\frac{\beta^2k^2\{(b^2+d^2)(b^2-k^2+d^2)+4b^2k^2+2ibk^3\}}{2b\{(b^2+k^2)^2+2d^2(b^2-k^2)+d^4\}^2}\right]Y_1^P(q)$$
$$-\frac{\gamma\beta}{2b(\alpha^2+k^2)}\left[\frac{\alpha^2}{(\alpha+b)^2+d^2}+\frac{k^2(b^2-k^2+d^2)+2ibk^3}{(b^2+k^2)^2+2d^2(b^2-k^2)+d^4}\right]Y_2^P(q), \tag{60a}$$

and

$$\xi_2(k,q)=\left[1+\frac{\gamma^2}{2\alpha(\alpha^2+k^2)}-\frac{\gamma^2\alpha}{(\alpha^2+k^2)^2}-\frac{i\gamma^2k}{(\alpha^2+k^2)}\right]Y_2^P(q)$$
$$+\frac{\gamma\beta}{(b^2+k^2)^2+2d^2(b^2-k^2)+d^4}\left[\frac{(b^2+d^2)(b^2-k^2+d^2)}{2b((\alpha+b)^2+d^2)}\right.$$
$$\left.-\frac{\alpha k^2}{(\alpha+b)^2+d^2}+\frac{\alpha k^2}{\alpha^2+k^2}+\frac{ik^3}{\alpha^2+k^2}\right]Y_1^P(q) \tag{60b}$$

with

$$Y_1^P(q)=\frac{\gamma q}{\alpha^2+q^2}, \tag{61a}$$

and

$$Y_2^P(q)=\frac{\beta q^3}{(b^2+q^2)^2+2d^2(b^2-q^2)+d^4}. \tag{61b}$$

The Fredholm determinant $\det_2 A^P(k)$ in Eq. (58) is given by

$$\begin{aligned}
&\det_2 A^P(k)\\
&\quad=\det_2 A^J(k)\\
&\quad+\frac{\beta^2\{(b^2+d^2)(b^2-k^2+d^2)+4b^2k^2+2ibk^3\}}{2b\{(b^2+k^2)^2+2d^2(b^2-k^2)+d^4\}^2}\\
&\quad\times\left[k^2+\frac{\gamma^2k^2}{2\alpha(\alpha^2+k^2)}+\frac{\gamma^2\{(b^2+d^2)(b^2-k^2+d^2)-2\alpha bk^2\}}{2b(\alpha^2+k^2)((\alpha+b)^2+d^2)}\right]\\
&\quad-\frac{\gamma^2(\alpha+ik)}{(\alpha^2+k^2)^2}\left[1+\frac{\beta^2}{16b^3\{(b^2+k^2)^2+2d^2(b^2-k^2)+d^4\}}\right.\\
&\quad\times\left\{(b^2+d^2)(b^2-k^2+d^2)-4b^2k^2\right.\\
&\quad\left.\left.+\frac{8b^2k^2(b^2+d^2+2\alpha b)}{(\alpha+b)^2+d^2}\right\}\right].
\end{aligned} \tag{62}$$

For this potential the off-shell T matrix is

$$T(p,q,k^2)=\frac{2}{\pi pq\det_2 A^P(k)}[Y_2^P(p)\xi_2(k,q)-Y_1^P(p)\xi_1(k,q)]. \tag{63}$$

V. SUMMARY AND CONCLUSION

In this paper we emphasize the importance of analyzing off-shell effects due to scattering by nonlocal potentials. We solve the van Leeuwen–Reiner equation for a separable potential of arbitrary rank with physical and Jost boundary conditions. Using these results we derive expressions for half-off-shell and off-shell T matrices for some realistic nucleon–nucleon interactions. All results are obtained in terms of elementary transcendental functions. The algebraic expressions are, however, fairly complicated. To ensure the correctness of the results presented one can perform a couple of checks.

(i) It can be seen that in the appropriate limit all results for T matrices of this paper go over to the corresponding results for the Yamaguchi form factors,[14] obtained by Bagchi and Mulligan[15] by using a relatively complicated procedure.

(ii) From Eq. (20) it is apparent that the phase of half-off-shell T matrix is the phase shift. The phase shifts obtained from our expressions for the half-off-shell T matrices can easily be seen to agree with those obtained by solving the relevant Schrödinger equation.

The results for the half-off-shell T matrices will be useful in describing the half-shell reactions. The off-shell amplitudes can be employed to investigate the physical properties of three-nucleon systems by means of Faddeev equations. In the accompanying paper we shall present results for the off-shell wavefunctions with regular and standing wave boundary conditions. We shall also obtain results for half-off-shell and off-shell **k** matrices.

[1] B. Mulligan, L.G. Arnold, B. Bagchi, and T.O. Krause, Phys. Rev. C **13**, 2131 (1976).
[2] F. Tabakin, Ann. Phys. (N.Y.) **30**, 51 (1964).
[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] F. Tabakin, Phys. Rev. **174**, 1208 (1968).
[6] P. Beregi, Nucl. Phys. A **206**, 217 (1973).
[7] T.R. Mongan, Phys. Rev. **175**, 1260 (1968); **178**, 1597 (1969).
[8] F. Tabakin, Phys. Rev. **137**, B75 (1965).
[9] S.B. Sarkar, B. Talukdar, and D. Chattarji, Nucl. Phys. A **250**, 73 (1975).
[10] A. Messiah, Quantum Mechanics (Wiley Interscience, New York, 1961, 1962), Vols 1 and 2.
[11] R.G. Newton, *Scattering Theory of Waves and Particles* (McGraw-Hill, New York, 1966).
[12] M. Coz, L.G. Arnold, and D. Mackellar, Ann. Phys. **59**, 219 (1970).
[13] R. Jost, Helv. Phys. Acta **20**, 256 (1947).
[14] Y. Yamaguchi, Phys. Rev. **95**, 1628 (1954).
[15] B. Bagchi and B. Mulligan, Phys. Rev. C **10**, 2197 (1974).

Off-energy-shell results for scattering by a nonlocal potential. II[a)]

B. Talukdar, U. Das,[b)] N. Mallick, and M. Sen Gupta

Department of Physics, Visva-Bharati University, Santiniketan 731235 West Bengal, India
(Received 2 March 1978)

The wavefunction approach to off-shell scattering on nonlocal potentials is used to obtain half-off-shell and off-shell K matrices in terms of Jost functions and the Fredholm determinant associated with the off-shell principal value wavefunction. The results are employed to construct exact expressions for K matrices for Tabakin, Beregi, and Mongan case IV potentials.

I. INTRODUCTION

In the preceeding paper[1] (hereafter referred to as paper I) we solved the van Leeuwen–Reiner equation for a separable potential of arbitrary rank to obtain off-shell wavefunctions associated with physical and Jost boundary conditions for scattering by such a potential. We used these wavefunctions to determine the half-off-shell and off-shell T matrices for a number of realistic nucleon–nucleon interactions. In this paper we derive the off-shell wavefunctions with regular and standing wave boundary conditions. The half-off-shell K matrix is written directly in terms of the Jost functions introduced in I. The off-shell K matrix is, however, expressed in terms of form factors of the potential and of the Fredholm determinant associated with the principal value wavefunction (standing wave boundary condition). The results of the present paper will be useful in the studies of nuclear scattering reactions.[2]

In Sec. II we obtain expressions for the l-wave off-shell wavefunctions. Specializing to the s-wave case we employ our formal results in Sec. III to construct exact expressions for K matrices for the one- and two-term separable potentials treated in paper I. We conclude by noting that our expressions for the K matrices satisfy the usual relations involving T and K matrices.

II. REGULAR AND PRINCIPAL VALUE WAVEFUNCTIONS

Recently, it has been shown[3,4] that for a short-range local potential the physical, regular as well as principal value wavefunctions satisfy the same van Leeuwen–Reiner equation. For the Jost solution, however, the inhomogeneous term differs. The regular and principal value wavefunctions for an N-term separable potential will satisfy Eq. (3) of Ref. 1. The appropriate Green's functions[5] are the following:

The Green's function associated with the regular boundary condition is

$$G_k^{\mathscr{R}}(r,r') = \left\{ \begin{array}{ll} \frac{1}{k}\{\hat{j}_l(kr)\hat{\eta}_l(kr') - \hat{j}_l(kr')\hat{\eta}_l(kr)\}, & r>r', \\ 0, & r<r', \end{array} \right\} \tag{1}$$

and that with the standing wave boundary condition is

$$G_k^{\mathscr{P}}(r,r') = G_k^{\mathscr{R}}(r,r') - \frac{e^{-i\pi l}}{k}\hat{j}_l(kr)\hat{\eta}_l(kr'). \tag{2}$$

Employing the technique of paper I we can construct the off-shell regular and principal value wavefunctions in terms of the matrix elements of the free particle Green's functions between the form factors of the potential. For the regular wavefunction we have

$$\begin{aligned}\phi_l(k,q,r) &= \hat{j}_l(qr) + \frac{1}{k\det_N A^{\mathscr{R}}(k)}\sum_{ij=1}^{N}\lambda_i a_{ij}^{\mathscr{R}}(k)Y_j^{\mathscr{R}}(q) \\ &\times\left[\int_0^r\{-\hat{j}_l(kr')\hat{\eta}_l(kr)+\hat{\eta}_l(kr')\hat{j}_l(kr)\}v_l^{(i)}(r')\,dr'\right].\end{aligned} \tag{3}$$

The principal value wavefunction can be written in the form

$$\begin{aligned}\psi_l^{\mathscr{P}}(k,q,r) &= \hat{j}_l(qr) + \frac{1}{k\det_N A^{\mathscr{P}}(k)}\sum_{ij=1}^{N}\lambda_i a_{ij}^{\mathscr{P}}(k)Y_j^{\mathscr{P}}(q) \\ &\times\left[\int_0^r\{-\hat{j}_l(kr')\hat{\eta}_l(kr)+\hat{\eta}_l(kr')\hat{j}_l(kr)\}v_l^{(i)}(r')\,dr'\right. \\ &\left. - e^{-il\pi}\hat{j}_l(kr)\int_0^\infty\hat{\eta}_l(kr')v_l^{(i)}(r')\,dr'\right].\end{aligned} \tag{4}$$

The elements of the Fredholm determinants $\det_N A^{\mathscr{R}}(k)$ and $\det_N A^{\mathscr{P}}(k)$ in Eqs. (3) and (4) are given by

$$\begin{aligned}{}_N A_{ij}^{\mathscr{R}}(k) &= \delta_{ij} - \frac{\lambda_j}{k}\int_0^\infty v_i^{(i)}(r)\,dr\left\{\int_0^r[-\hat{j}_l(kr')\hat{\eta}_l(kr) \right. \\ &\left. + \hat{\eta}_l(kr')\hat{j}_l(kr)]v_l^{(j)}(r')\,dr'\right\},\end{aligned} \tag{5}$$

and

$$\begin{aligned}{}_N A_{ij}^{\mathscr{P}}(k) &= \delta_{ij} - \frac{\lambda_j}{k}\int_0^\infty v_i^{(i)}(r)\,dr\left\{\int_0^r[-\hat{j}_l(kr')\hat{\eta}_l(kr) \right. \\ &\left. + \hat{\eta}_l(kr')\hat{j}_l(kr)]v_l^{(j)}(r')\,dr' - e^{-il\pi}\right.\end{aligned}$$

[a)]Supported in part by the Department of Atomic Energy, Government of India.

[b)]Based in part on a thesis to be submitted by one of the authors (U.D.) to the Visva-Bharati University.

 0022-2488/79/030378-04$01.00

$$\times \int_0^\infty v_l^{(\mathscr{D})}(r')\hat{j}_l(kr)\hat{\eta}_l(kr')\,dr'\Big\}. \tag{6}$$

Also $a_{ij}^{\mathscr{R}}$ and $a_{ij}^{\mathscr{D}}$ stand for the cofactors of ${}_N\mathcal{A}_{ji}^{\mathscr{R}}$ and ${}_N\mathcal{A}_{ji}^{\mathscr{D}}$ respectively. The components of the column vector Y are given by

$$Y_i^{\mathscr{R}}(q) = \int_0^\infty v_l^{(i)}(\delta)\hat{j}_l(qs)\,ds, \tag{7}$$

and

$$Y_i^{\mathscr{D}}(q) = Y_i^{\mathscr{R}}(q). \tag{8}$$

In terms of the wavefunction in Eq. (4) the off-shell K matrix is written as

$$K_l(p,q,k^2) = \frac{2}{\pi pq}\sum_{i=1}^{N}\lambda_i\int_0^\infty\int_0^\infty dr dr'\hat{j}_l(pr)v_l^{(i)}(r)v_l^{(i)}(r')\psi_l^{\mathscr{D}}(k,q,r')$$
$$= \frac{2}{\pi pq\det_N\mathcal{A}^{\mathscr{D}}(k)}\sum_{i,j=1}^{N}\lambda_i a_{ij}^{\mathscr{D}}(k)Y_j^{\mathscr{D}}(q)Y_i^{\mathscr{D}}(p). \tag{9}$$

The half-off-shell K matrix can, however, be calculated by using the Jost functions in paper I with the help of the relation

$$K_l(k,q,k^2) = \left(\frac{k}{q}\right)^l \frac{2}{\pi iq}\frac{f_l(k,q)-f_l(k,-q)}{f_l(k)+f_l(-k)}. \tag{10}$$

Equations (9) and (10) represent the basic equations for computing off-shell and half-off-shell K matrices for a non-local separable potential by the wavefunction method.

III. EXPRESSIONS FOR K MATRICES

A. One-term Tabakin potential

Using Eqs. (27) and (28) of paper I for the off-shell Jost functions in the s-wave version of Eq. (10), we obtain the half-off-shell K matrix for the Tabakin potential in the form

$$K(k,q,k^2) = \frac{2\lambda Y^{\mathscr{D}}(k)Y^{\mathscr{D}}(q)}{\pi kq[\det_1 A^J(k) + (\lambda/k)Y^{\mathscr{D}}(k)F(k)]}, \tag{11}$$

where the form factors $Y^{\mathscr{D}}$ and F are given by

$$Y^{\mathscr{D}}(x) = x\left[\frac{\mathscr{A}_1 x^2 + 2\mathscr{A}_2\alpha_1^2}{4\alpha_1^4 + x^4} + \frac{\mathscr{A}_3}{\alpha_2^2 + x^2}\right], \tag{12}$$

and

$$F(x) = \frac{2\alpha_1^3(\mathscr{A}_1+\mathscr{A}_2) + \alpha_1 x^2(\mathscr{A}_1 - \mathscr{A}_2)}{4\alpha_1^4 + x^4} + \frac{\mathscr{A}_3\alpha_2}{\alpha_2^2 + x^2}. \tag{13}$$

Here $\det_1 A^J(k)$ stands for the Fredholm determinant for the Jost solution. Note that the Jost Fredholm determinants for all potentials which we consider in this paper are given in paper I. The Fredholm determinant associated with the principal value wavefunction is

$\det_1 A^{\mathscr{D}}(k)$

$$= 1 - \lambda\left[\frac{\mathscr{A}_3}{\alpha_2^2+k^2}\left(\frac{\mathscr{A}_2\alpha_2 + \mathscr{A}_1(\alpha_1+\alpha_2)}{\alpha_1^2 + (\alpha_1+\alpha_2)^2}\right.\right.$$
$$\left. + \frac{\mathscr{A}_3}{2\alpha_2}\right) - \left(\frac{\mathscr{A}_1 k^2 + 2\mathscr{A}_2\alpha_1^2}{4\alpha_1^4 + k^4} + \frac{\mathscr{A}_3}{\alpha_2^2+k^2}\right)$$
$$\times\left(\frac{\mathscr{A}_1\{\alpha_1(2\alpha_1^2+k^2)\} + \mathscr{A}_2\{\alpha_1(2\alpha_1^2-k^2)\}}{4\alpha_1^4+k^4}\right.$$
$$\left. + \frac{\mathscr{A}_3\alpha_2}{\alpha_2^2+k^2}\right) + \left(\frac{\mathscr{A}_1 k^2 + 2\mathscr{A}_2\alpha_1^2}{4\alpha_1^4+k^4}\right)$$
$$\times\left(\frac{3\mathscr{A}_1 + \mathscr{A}_2}{8\alpha_1} + \frac{\mathscr{A}_3(\alpha_1+\alpha_2)}{\alpha_1^2 + (\alpha_1+\alpha_2)^2}\right)$$
$$+ \left(\frac{\mathscr{A}_2 k^2 - 2\mathscr{A}_1\alpha_1^2}{4\alpha_1^4+k^4}\right)$$
$$\left.\times\left(\frac{\mathscr{A}_1+\mathscr{A}_2}{8\alpha_1} + \frac{\mathscr{A}_3\alpha_1}{\alpha_1^2+(\alpha_1+\alpha_2)^2}\right)\right]. \tag{14}$$

In terms of the Fredholm determinant in Eq. (14), the off-shell K matrix for this potential comes out to be

$$K(p,q,k^2) = \frac{2Y^{\mathscr{D}}(p)Y^{\mathscr{D}}(q)}{\pi pq\,\det_1\mathscr{A}^{\mathscr{D}}(k)}. \tag{15}$$

After some algebraic manipulations it can be seen that the quantity inside the square brackets in the denominator of Eq. (11) coincides with the principal value Fredholm determinant in (14). Thus we get the relation

$$K(k,q,k^2) = \lim_{p\to k} K(p,q,k^2). \tag{16}$$

Equation (16) serves as a check on our expression for $K(k,q,k^2)$ in terms of Jost functions. For other potentials under consideration we shall not include results for half-off-shell K matrices since the latter can be obtained by the limiting procedure given in (16).

B. Beregi potential

The Fredholm determinant associated with the principal value wavefunction is given by

$$\det_1 A^{\mathscr{D}}(k) = 1 + \frac{\lambda}{\alpha_1^2+k^2}\left[\frac{\alpha_1}{\alpha_1^2+k^2} - \frac{1}{2\alpha_1}\right]$$
$$+ \frac{\lambda a^2}{\alpha_2^2+k^2}\left[\frac{\alpha_2}{\alpha_2^2+k^2} - \frac{1}{2\alpha_2}\right]$$
$$+ \frac{2\lambda a(\alpha_1\alpha_2 - k^2)}{(\alpha_1+\alpha_2)(\alpha_1^2+k^2)(\alpha_2^2+k^2)}. \tag{17}$$

The off-shell K matrix is

$$K(p,q,k^2) = \frac{2\lambda}{\pi pq} \frac{Y^{\mathscr{P}}(p)Y^{\mathscr{P}}(q)}{\det_1 A^{\mathscr{P}}(k)} \tag{18}$$

with

$$Y^{\mathscr{P}}(x) = x\left[\frac{1}{\alpha_1^2 + x^2} - \frac{a}{\alpha_2^2 + x^2}\right]. \tag{19}$$

C. Mongan case IV potential

For this potential the principal value Fredholm determinant and off-shell K matrix are given by the following expressions:

$$\det_2 A^{\mathscr{P}}(k) = 1 - \frac{\lambda_1}{2\alpha_1(\alpha_1^2 + k^2)} - \frac{\lambda_2}{2\alpha_2(\alpha_2^2 + k^2)} + \frac{\lambda_1\alpha_1}{(\alpha_1^2 + k^2)^2} + \frac{\lambda_2\alpha_2}{(\alpha_2^2 + k^2)^2} + \frac{\lambda_1\lambda_2(\alpha_1 - \alpha_2)^2}{4\alpha_1\alpha_2(\alpha_1 + \alpha_2)^2(\alpha_1^2 + k^2)(\alpha_2^2 + k^2)} - \frac{\lambda_1\lambda_2 k^2(\alpha_1 - \alpha_2)^2}{2\alpha_1\alpha_2(\alpha_1^2 + k^2)^2(\alpha_2^2 + k^2)^2}, \tag{20}$$

and

$$K(p,q,k^2) = \frac{2}{\pi pq \det_2 A^{\mathscr{P}}(k)} \sum_{i,j=1}^{2} \lambda_i Y_i^{\mathscr{P}}(p) \times [L_j(k)Y_i^{\mathscr{P}}(q) - M_{ij}(k)Y_j^{\mathscr{P}}(q)], \quad j \neq i. \tag{21}$$

Here

$$Y_i^{\mathscr{P}}(x) = \frac{x}{\alpha_i^2 + x^2}, \tag{22}$$

$$L_j(k) = 1 + \frac{\lambda_j(\alpha_j^2 - k^2)}{2\alpha_j(\alpha_j^2 + k^2)^2}, \tag{23}$$

and

$$M_{ij}(k) = \frac{\lambda_j(\alpha_i\alpha_j - k^2)}{(\alpha_i + \alpha_j)(\alpha_i^2 + k^2)(\alpha_j^2 + k^2)}, \tag{24}$$

D. Two-term Tabakin potential

For the two-term Tabakin potential the relevant Fredholm determinant and off-shell K matrix can be written as,

$$\det_2 A^{\mathscr{P}}(k) = \det_2 A^{J}(k) - \frac{\gamma^2\alpha}{(\alpha^2 + k^2)^2}\left[1 + \frac{\beta^2}{16b^3[(b^2 + k^2)^2 + 2d^2(b^2 - k^2) + d^4]} \times \left\{(b^2 + d^2)(b^2 - k^2 + d^2) - 4b^2k^2 + \frac{8b^2k^2(b^2 + d^2 + 2\alpha b)}{(\alpha + b)^2 + d^2}\right\}\right] + \frac{\beta^2[(b^2 + d^2)(b^2 - k^2 + d^2) + 4b^2k^2]}{2b[(b^2 + k^2)^2 + 2d^2(b^2 - k^2) + d^4]^2}\left[k^2 + \frac{\gamma^2k^2}{2\alpha(\alpha^2 + k^2)} + \frac{\gamma^2[(b^2 + d^2)(b^2 - k^2 + d^2) - 2\alpha bk^2]}{2b(\alpha^2 + k^2)((\alpha + b)^2 + d^2)}\right], \tag{25}$$

and

$$K(p,q,k^2) = \frac{2}{\pi pq \det_2 A^{\mathscr{P}}(k)}[Y_2^{\mathscr{P}}(p)\rho_2(k,q) - Y_1^{\mathscr{P}}(p)\rho_1(k,q)]. \tag{26}$$

In Eq. (26)

$$Y_1^{\mathscr{P}}(x) = \frac{\gamma x}{\alpha^2 + x^2}, \tag{27a}$$

$$Y_2^{\mathscr{P}}(x) = \frac{\beta x^3}{(b^2 + x^2)^2 + 2d^2(b^2 - x^2) + d^4}, \tag{27b}$$

$$\rho_1(k,q) = \left[1 + \frac{\beta^2[(b^2 + d^2)(b^2 - k^2 + d^2) - 4b^2k^2]}{16b^3[(b^2 + k^2)^2 + 2d^2(b^2 - k^2) + d^4]} + \frac{\beta^2k^2[(b^2 + d^2)(b^2 - k^2 + d^2) + 4b^2k^2]}{2b[(b^2 + k^2)^2 + 2d^2(b^2 - k^2) + d^4]^2}\right] Y_1^{\mathscr{P}}(q)$$

$$+\frac{\gamma\beta}{2b(\alpha^2+k^2)}\left[\frac{b^2+d^2+2\alpha b}{(\alpha+b)^2+d^2}-\frac{(b^2+d^2)(b^2-k^2+d^2)+4b^2k^2}{(b^2+k^2)^2+2d^2(b^2-k^2)+d^4}\right]Y_2^{\mathscr{D}}(q),$$

(27c)

and

(27c)

$$\rho_2(k,q)=\left[1+\frac{\gamma^2}{2\alpha(a^2+k^2)}-\frac{\gamma^2\alpha}{(\alpha^2+k^2)^2}\right]Y_2^{\mathscr{D}}(q)$$

$$+\frac{\gamma\beta}{(b^2+k^2)^2+2d^2(b^2-k^2)+d^4}\left[\frac{\alpha k^2}{\alpha^2+k^2}+\frac{(b^2+d^2)(b^2-k^2+d^2)-2\alpha bk^2}{2b((\alpha+b)^2+d^2)}\right]Y_1^{\mathscr{D}}(q).$$

(27d)

IV. SUMMARY AND DISCUSSION

We have adapted the wavefunction approach of paper I to K matrix calculations. Expressions for off-shell and half-off-shell K matrices $K(p,q,k^2)$ and $K(k,q,k^2)$ are derived in terms of elementary transcendental functions for Tabakin, Beregi, and Mongan case IV potentials.

For a short-range local interaction the relation between half-off-shell K and T matrices is given by

$$K(k,q,k^2)=\frac{T(k,q,k^2)}{1-(i\pi k/2)T(k)}.$$

The results for nonlocal potentials presented in paper I and here are also found to satisfy this relation.

[1]B. Talukdar, U. Das, and S. Mukhopadhyaya, J. Math. Phys. **20**, 372 (1979) (cited as paper I).

[2]C. Block and V. Gillett, Phys. Lett. **16**, 62 (1965); W.M. Mac Donald and A. Makjian, Phys. Rev. **160**, 730 (1967); F. Schmittroth and W. Tobocman, Phys. Rev. C **3**, 1010 (1971); D. Rodijak, W. Tobocman, and D.K. Tandon, Phys. Rev. C **6**, 1192 (1972).

[3]C.S. Warke and M.K. Srivastava, Phys. Rev. C **15**, 561 (1977).

[4]B. Talukdar, N. Mallick, and S. Mukhopadhyaya, Phys. Rev. C **15**, 1252 (1977).

[5]R.G. Newton, *Scattering Theory of Waves and Particles* (McGraw-Hill New York, 1966).

Lie algebras associated with motion in axisymmetric electromagnetic fields

Thomas P. Mitchell

University of California, Santa Barbara, California 93106
(Received 26 May 1978)

The existence and analytical form of a vector constant of the classical relativistic planar motion of a point charge in an arbitrary time-independent axisymmetric electromagnetic field are established. The components of the vector are utilized, in conjunction with the angular momentum, to construct realizations of the Lie algebras of the Euclidean group E(2), of the special unitary group SU(2), and of the Ladder operators of the harmonic oscillator. The charge is assumed to move in an externally prescribed field. The formulation is gauge invariant.

I. INTRODUCTION

The Hamiltonian canonical formalism of classical mechanics enables one to construct the dynamics of a system in terms of dynamical variables which are functions of the phase space coordinates. This infinite set of variables, together with the operations of the bilinear Poisson bracket and ordinary addition, forms a dynamical Lie algebra, the Poisson bracket being the realization of the abstract Lie bracket. The algebra on the one hand is enveloped by that which allows for ordinary multiplication of the variables in addition to the two operations already stated, and on the other hand it contains the set of constants of the motion as a subalgebra. The outstanding dynamical variable is, of course, the Hamiltonian. Each variable generates a one-parameter group of canonical transformations—an automorphism—of the dynamical algebra into itself. In particular, the transformation generated by the Hamiltonian, with the time as the group parameter, describes the temporal evolution of the system.

The phase space formulation is especially appropriate when one seeks to establish symmetry properties of a system. Symmetries may be found in phase space which are not at all evident from a study of configuration space alone. Such higher symmetries of nongeometrical origin, generally called dynamical symmetries, are of considerable interest, being responsible for hidden symmetries or accidental degeneracies. The symmetry of the hydrogen atom under the operations of the orthogonal rotation group O(4) is a well-known example. McIntosh[1] reviews the relationship of symmetry to degeneracy and provides an extensive bibliography.

Although any dynamical variable can be utilized to generate a canonical transformation, only those which are constants of the motion, and hence whose Poisson brackets with the Hamiltonian are zero, generate transformations that leave the Hamiltonian invariant. The Lie algebra consisting of constants of the motion is thus a symmetry algebra, and the associated group a symmetry group of the system. The role of Lie algebras and groups in classical mechanics is comprehensively presented by Sudarshan and Mukunda[2] and is concisely treated by Rosen.[3] It has also been discussed in several specific applications, e.g., by Fradkin,[4] Mukunda,[5,6] Stehle and Han,[7] Maiella and Vitale,[8] and Mitchell.[9]

The construction of the symmetry algebras associated with the classical motion of a structureless point charge in an externally prescribed time-independent axisymmetric electromagnetic field is the subject of this paper. To be exact, it is assumed that the motion is confined to the equatorial plane and, of course, that the field configuration admits such planar trajectories. It is shown that a vector constant of the motion which lies in the orbital plane always exists, can be explicitly determined, and can be used to construct realizations of three symmetry algebras. These are algebras of the Euclidean group E(2), the unitary unimodular group SU(2), and the algebra of creation and annihilation operators. The construction can be carried through without either prescribing the functional form of the electromagnetic field or choosing a particular gauge. The analysis is relativistic throughout and is readily reduced to a nonrelativistic form if required. It includes as a special case the work of Dulock and McIntosh[10] which is restricted to a field configuration consisting of a uniform magnetic field in the presence of an electric field originating in a harmonic oscillator potential.

II. THE VECTOR CONSTANT

The Hamiltonian

$$H = c(m^2c^2 + \mathbf{P}^2)^{1/2} + q\phi \tag{1}$$

is the total energy, and describes the motion of a zero-spin structureless point charge in an externally prescribed field whose scalar potential is ϕ. In this expression the mechanical linear momentum is represented by $\mathbf{P}$, the rest mass by m and the charge by q. On introducing the vector potential, $\mathbf{A}$, of the magnetic field and the conjugate kinetic momentum $\mathbf{p}$, one can express the Hamiltonian in the form[11]

$$H = c\left[m^2c^2 + \left(\mathbf{p} - \frac{q}{c}\mathbf{A}\right)^2\right]^{1/2} + q\phi. \tag{2}$$

To ensure the admissibility of trajectories confined to the equatorial plane $z = 0$, the potentials $\mathbf{A}$ and ϕ must satisfy the conditions

$$\frac{\partial A_\theta}{\partial z} = \frac{\partial A_r}{\partial z} - \frac{\partial A_z}{\partial r} = \frac{\partial \phi}{\partial z} = 0, \quad z = 0, \tag{3}$$

in which (r, θ, z) are the usual cylindrical coordinates and subscripts denote components. The axisymmetric

 0022-2488/79/030382-03$01.00

nature of the electromagnetic field is guaranteed by omitting any dependence on the azimuthal angle θ from the potentials. It is not necessary to subject them to any particular choice of gauge. For the remainder of the analysis the Hamiltonian adopted is the two-dimensional form

$$H = c\left[m^2c^2 + \left(p_r - \frac{q}{c}A_r\right)^2 + \left(\frac{1}{r}p_\theta - \frac{q}{c}A_\theta\right)^2\right]^{1/2} + q\phi(r), \tag{4}$$

which is obtained by setting $z = p_z = A_z = 0$ in Eq. (2).

Owing to the symmetry of the Hamiltonian under the orthogonal rotation group O(2) the kinetic orbital angular momentum p_θ, conjugate to the azimuthal angle, is a constant of the motion which will be represented by l. Accordingly the vector $\mathbf{l} = \mathbf{r}\times\mathbf{p}$, which generates rotations about the axis of symmetry, is a constant vector normal to the equatorial (orbit) plane. The two-dimensional Hamiltonian given by Eq. (4) can be written as a function of the two variables r and p and the constant l only. A second vector constant $\mathbf{F}$ can now be determined as follows. Let

$$\mathbf{F} = f\mathbf{r} + g\mathbf{p}, \tag{5}$$

in which the two dynamical variables f and g are functions also of $r, p,$ and l alone. A straightforward calculation shows that the Poisson bracket

$$[l, \mathbf{F}] = \mathbf{l}\times\mathbf{F}/l \tag{6}$$

and hence that $\mathbf{F}$ behaves like a vector under rotations. Thus $\mathbf{F}$, which clearly lies in the orbit plane, is a vector constant of the motion if its Poisson bracket with the Hamiltonian

$$[\mathbf{F}, H] = 0. \tag{7}$$

This condition can be rendered in a simple form by first changing the representation (5) to

$$\mathbf{F} = U(r, p, l)\mathbf{e}_r + V(r, p, l)\mathbf{e}_\theta. \tag{8}$$

The unit basis vectors in the (r, θ) coordinate system are symbolized by $(\mathbf{e}_r, \mathbf{e}_\theta)$. The condition for the constancy of $\mathbf{F}$ then becomes

$$\left\{[U, H] - V\frac{\partial H}{\partial p_\theta}\right\}\mathbf{e}_r + \left\{[V, H] + U\frac{\partial H}{\partial p_\theta}\right\}\mathbf{e}_\theta = 0 \tag{9}$$

or, equivalently,

$$[W, H] + iW\frac{\partial H}{\partial p_\theta} = 0, \tag{10}$$

where

$$W(r, p, l) = U(r, p, l) + iV(r, p, l), \quad i = \sqrt{-1}. \tag{11}$$

However, since both H and W are functions of $r, p,$ and l, we have the Poisson bracket

$$[W, H] = \frac{(\mathbf{r}\cdot\mathbf{p})}{rp}\left(\frac{\partial w}{\partial r}\frac{\partial H}{\partial p} - \frac{\partial w}{\partial p}\frac{\partial H}{\partial r}\right), \tag{12}$$

and Eq. (10) is a first-order linear partial differential equation which determines the function $W(r, p, l)$ when H is prescribed. By introducing the independent variables (r, H) in place of (r, p) in Eq. (10) it is reduced to the ordinary differential equation

$$\frac{dW}{dr} + \frac{iL}{r(\mathbf{r}\cdot\mathbf{P})}W = 0, \tag{13}$$

in which the function L is defined by

$$L = \frac{r^2p}{(\mathbf{r}\cdot\mathbf{p})}(\mathbf{r}\cdot\mathbf{P})\left(\frac{\partial H}{\partial p_\theta}\right)\left(\frac{\partial H}{\partial p}\right)^{-1}. \tag{14}$$

Hence, as a solution of Eq. (13) one may take

$$W(r, H, l) = \exp(i\psi), \tag{15}$$

using the notation

$$\psi(r, H, l) = \int_{r_0}^{r}\frac{L}{r(\mathbf{r}\cdot\mathbf{P})}dr \tag{16}$$

in which the lower limit is arbitrary. On combining Eqs. (16), (15), (11), (8), and (5) it is found that

$$\mathbf{F} = \frac{\mathbf{r}}{r}\cos\psi + \frac{\mathbf{r}\times(\mathbf{r}\times\mathbf{P})}{rL}\sin\psi \tag{17}$$

is a vector constant of the motion with unit magnitude. It is similar in form to that previously established[9] for spherically symmetric Hamiltonians. The function L is identified, on using Eqs. (14) and (4) to be the magnitude of the orbital angular momentum

$$L = |\mathbf{r}\times\mathbf{P}| = rP_\theta = p_\theta - \frac{q}{c}rA_\theta. \tag{18}$$

Consequently, ψ which is the angular position of $\mathbf{r}$ relative to $\mathbf{F}$ is found for Eqs. (18) and (16) to be

$$\psi = \int_{r_0}^{r}\left(\frac{l}{r} - \frac{q}{c}A_\theta\right)\left[(H - q\phi)^2/c^2 - m^2c^2 - \left(\frac{l}{r} - \frac{q}{c}A_\theta\right)^2\right]^{-1/2}dr \tag{19}$$

and is recognized as the angular position of the charge in its orbit. If the lower limit of integration is chosen to correspond to a turning point, $\mathbf{r}\cdot\mathbf{P} = 0$, then the vector $\mathbf{F}$ is always directed towards that point. The equation of the orbit is found by taking the inner product of $\mathbf{r}$ and $\mathbf{F}$. In the particular case, e.g., in which a uniform magnetic field $B\mathbf{k}$ acts alone along the axis of symmetry, Eqs. (16), (17), and (19) serve to determine

$$\mathbf{F} = (m\omega\mathbf{r} - \mathbf{k}\times\mathbf{P})(\mathbf{P}^2 + 2m\omega l)^{-1/2}. \tag{20}$$

This well-known vector constant establishes the direction of the position vector of the center of the circle of gyration relative to the coordinate axes. The Larmor frequency qB/mc is denoted by ω. The quantum-mechanical properties of the vector defined in equation (20) are discussed by Johnson and Lippman.[12] Other closely allied quantum mechanical results are given by Malkin and Man'ko,[13] by Katriel and Adam,[14] and by Malkin, Man'ko, and Trifonov.[15]

III. THE LIE ALGEBRAS

On multiplying the vector $\mathbf{F}$ by any scalar dynamical variable $R(r, p, l)$ one obtains a new vector constant

$$\mathbf{G} = R\mathbf{F} \tag{21}$$

if

$$[R, H] = 0. \tag{22}$$

It is possible therefore to adjust the magnitude of the vector by an appropriate choice of the scalar constant of the motion R. It follows from Eq. (7) that

$$[l, G_1] = G_2, \tag{23}$$

$$[l, G_2] = -G_1, \tag{24}$$

and in addition the bracket relation

$$G_2[G_1, G_2] = G[G_1, G] \tag{25}$$

can be readily derived. The subscripts 1 and 2 have been introduced to denote the Cartesian components of **G** whose magnitude is $G = R$. Three choices of the function R are now considered.

The first is to choose it to be the Hamiltonian itself. Then Eq. (25) becomes

$$[G_1, G_2] = 0 \tag{26}$$

and thus through Eqs. (23), (24), and (25) the quantities G_1, G_2, and l provide a realization of the Lie algebra of the group $E(2)$ of motions in two-dimensional Euclidean space. If as the second choice R is taken to be

$$R = il, \quad i = \sqrt{-1}, \tag{27}$$

then Eq. (25) reduces to

$$[G_1, G_2] = l. \tag{28}$$

The identification $I_1 = G_1$, $I_2 = G_2$, $I_3 = l$ makes it possible to write the bracket relations (23), (24), and (28) in the compact form

$$[I_j, I_k] = \sum_m \epsilon_{jkm} I_m, \tag{29}$$

in which ϵ_{jkm} is the permutation tensor. The Lie algebra[16] of the special unitary group SU(2) can accordingly be realized by G_1, G_2, and il defined through Eqs. (17), (21), and (27). As the final choice one may take

$$R = (H - 2\alpha l)^{1/2} \tag{30}$$

in which α represents an arbitrary constant independent of the phase space coordinates. The bracket relations are then

$$[G_1, G_2] = \alpha, \tag{31}$$

$$[G_2, l] = G_1, \tag{32}$$

$$[l, G_1] = G_2, \tag{33}$$

which are those of the creation and annihilation operators of the quantum harmonic oscillator.

[1] H. V. McIntosh, in *Group Theory and Its Applications, Vol. II*, edited by E. M. Loebl (Academic, New York, 1971).
[2] E. C. G Sudarshan and N. Mukunda, *Classical Dynamics: A Modern Perspective* (Wiley, New York, 1974).
[3] G. Rosen, *Formulations of Classical and Quantum Dynamical Theory* (Academic, New York, 1969).
[4] D. M. Fradkin, Prog. Theor. Phys. **37**, 798—812 (1967).
[5] N. Mukunda, Phys. Rev. **155**, 1383—86 (1967).
[6] N. Mukunda, J. Math. Phys. **8**, 1069—72 (1967).
[7] P. Stehle and M. Y. Han, Phys. Rev. **159**, 1076—82 (1967).
[8] G. Maiella and B. Vitale, Nuovo Cimento A **42**, 330—33 (1967).
[9] T. P. Mitchell, "Vector Constants and Their Algebras for Classical Hamiltonians," J. Math. Phys. (in press).
[10] V. A. Dulock and H. V. McIntosh, J. Math. Phys. **7**, 1401—12 (1966).
[11] H. C. Corben and P. Stehle, *Classical Mechanics* (Wiley, New York, 1965).
[12] M. H. Johnson and B. A. Lippmann, Phys. Rev. **76**, 828—32 (1949).
[13] I. A. Malkin and V. I. Man'ko, Yad. Fiz. **8**, 1264—72 (1968) [Sov. J. Nucl. Phys. **8**, 731—5 (1969)].
[14] J. Katriel and G. Adam, Phys. Rev. Lett. **23**, 1310—11 (1969).
[15] I. A. Malkin, V. I. Man'ko and D. A. Trifonov, Nuovo Cimento A **4**, 773—91 (1971).
[16] L. Fonda and G. C. Ghirardi, *Symmetry Principles in Quantum Physics* (Marcel Dekker, New York, 1970).

Extension theorems for operator-valued measures

Mark J. Christensen

School of Mathematics, Georgia Institute of Technology, Atlanta, Georgia 30332
(Received 18 May 1978)

Benioff [J. Math. Phys. **13** (1972)] has shown that every consistent, countable family of positive, normalized operator-valued (PNOV) measures μ_n over R^n can be extended to a PNOV measure over R^∞. In this paper we show that the same result holds for arbitrary, consistent families of PNOV measures over complete, separable metric spaces. Further we show that, while there may be no extension if the topological conditions are relaxed, it is always possible to construct a related family of PNOV measure spaces which: (1) Are measures theoretically indistinguishable from the original spaces; (2) Have an extending PNOV measure. These results use developments in the theory of algebraic models of measures as initiated by Dinculeanu and Foias [Ill. J. Math. **12** (1968)] and applied to stochastic processes by Schreiber, Sun, and Bharucha-Reid [Trans. AMS 158 (1973)].

1. INTRODUCTION

In mathematical physics operator-valued measures over R^n and their extension to R^∞ have been used to study quantum mechanical measurement and decision procedures involving an infinite number of steps (cf. Benioff[1,2]). For example, consider a situation in which a collection of observables is measured repeatedly. This will yield a sequence of operator-valued measures $\{\mu_n\}$ over R^n. Furthermore, they will be consistent in the sense of Kolmogorov. That is, for every n and every measurable subset A of R^n we shall have

$$\mu_n(A) = \mu_{n+1}(A \times R).$$

Hence it is natural to consider extension theorems for operator-valued measures; that is, to investigate the existence of an operator-valued measure μ on R^∞ which reduces to μ_n on R^n. Benioff[1] has investigated the extensions of families of operator-valued measured defined over countable products of $\mathbb{R}$, and obtained extension theorems of Kolmogorov type for such families.

In Sec. 2 of this paper we define operator-valued measures. Section 3 is devoted to a generalization of Benioff's theorem to arbitrary families of Polish spaces; i.e., uncountable collections of complete separable metric spaces. Finally, in Sec. 4, we extend the idea of an algebraic model for a scalar measure (due to Dinculeanu and Foias[3]) to operator-valued measures, obtain an operator-valued version of the Bochner–Neumark theorem, and obtain an analog of the main result of Schreiber, Sun, and Bharucha-Reid.[4] This permits the elimination of the topological conditions on the spaces over which the measures are defined. For while there may not be an extension over non-Polish spaces we shall see that there is a canonical way to construct a family of spaces measure-theoretically indistinguishable from the originals over which there is an extension.

2. OPERATOR-VALUED MEASURES

Let H be a fixed complex Hilbert space with inner product $\langle \cdot, \cdot \rangle$; and, as before, let $\mathcal{L}(H)$ denote the Banach algebra of bounded linear operators on H equipped with the norm topology, i.e., the topology induced by the operator norm

$$\|L\|_{op} = \sup \|Lx\|_H, \quad x \in H, \quad \langle x, x \rangle = 1.$$

If X is a set and $\mathcal{A}$ is a σ-algebra of subsets of X, then a function $\mu: \mathcal{A} \to \mathcal{L}(H)$ is called an *operator-valued measure* if

(i) $\mu(\phi) = \theta$ (the null operator);

(ii) if $\{A_i\}$ is a pairwise disjointed sequence of sets from $\mathcal{A}$, then

$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i)$$

in the strong sense; i.e., for every $x \in H$ we have

$$\left\| \sum_{i=1}^{N} \mu(A_i)x - \mu\left(\bigcup_{i=1}^{\infty} A_i\right)x \right\|_H \to 0$$

as $N \to \infty$. The measure μ is said to be *positive* if, for every $A \in \mathcal{A}$, $\mu(A)$ is a positive operator, i.e., for every $x \in H$, $\langle \mu(A)x, x \rangle \geq 0$. Furthermore, μ is said to be *normalized* if $\mu(X) = I$, and *complete* if $\mu(A) = \theta$ implies that every subset of A is a member of $\mathcal{A}$.

We remark that if $\mu(A) = \theta$ and $B \subseteq A$ is $\mathcal{A}$-measurable then $\mu(B) = \theta$ for positive, operator-valued measures. Hence, the statement "μ-a.s." is well defined. Finally, if μ is also normalized then, for every $x \in H$ with $\langle x, x \rangle = 1$ we have that $\mu_x(A) = \langle \mu(A)x, x \rangle$ is a probability measure and $\mu(A) = \theta$ if and only if $\mu_x(A) = 0$ for every unit vector x of H.

In view of the above we see that a positive, normalized operator-valued measure (called henceforth a PNOV *measure*) is a natural generalization of a probability measure. It is, therefore, natural to ask if extension theorems of Kolmogorov type can be proved for PNOV measures. We have already noted that this is a problem of interest in quantum theory.

3. AN EXTENSION THEOREM FOR PNOV MEASURES

For each α in a directed set I, let Y_α be a nonempty set with σ-algebra $\mathcal{A}_\alpha$. Assume that the Y_α are increasing and the $\mathcal{A}_\alpha$-measurable sets, when imbedded in Y_β, for $\beta \geq \alpha$, are $\mathcal{A}_\beta$-measurable. We may then ask: Under what conditions will a family of PNOV measure μ_α extend to a PNOV measure μ over $\cup_{\alpha \in I} A_\alpha$.

 0022-2488/79/030385-05$01.00

We now state and prove the following result as an answer to the above.

Theorem 1: For each $t\in T$ let X_t be a Polish space (i.e., a complete separable metric space) and let $\mathcal{A}_t$ denote the σ-algebra of Borel sets of X_t. Assume that for each nonempty finite subset $V\subseteq T$ there is given a PNOV measure μ^V on $\mathcal{A}_V=\Pi_{t\in V}\mathcal{A}_t$. Also assume the μ^V are consistent; that is, if $U\subseteq V$ let $\pi_{U,V}:\mathcal{A}_U\to\mathcal{A}_V$ be the canonical injection so that for every $A\in\mathcal{A}_U$

$$\mu^U(A)=\mu^V(\pi_{U,V}(A)).$$

Then there exists a PNOV measure μ on $\Pi_{t\in T}\mathcal{A}_t$ which agrees with each μ^V on $\mathcal{A}_V$.

In order to prove the above theorem the following two lemmas are needed.

Lemma 1: Theorem 1 holds if the expression "PNOV measure" is replaced by "probability measure."

For the proof of this lemma we refer to Ash.[5]

Lemma 2: Let the sets and σ-algebras X_t and $\mathcal{A}_t$ be as in Theorem 1. Let m be a probability measure on $\mathcal{A}=\Pi_{t\in T}\mathcal{A}_t$. Then, for each $A\in\mathcal{A}$ and $\epsilon>0$ there exists a finite subset U of T and a set $B\in\mathcal{A}_U=\Pi_{t\in U}\mathcal{A}_t$ such that,

$$|m(A)-m^U(B)|<\epsilon.$$

Proof of Lemma 2: First we observe (Ash,[5] p. 194) that for each fixed $A\in\mathcal{A}$ there exists a countable subset $T_A=(t_1,t_2,t_3,\cdots)$ of T and a set $\tilde{A}\in\mathcal{A}_{T_A}$ such that

$$(a(t),t\in T)\in A \quad \text{if and only if } (a(t),t\in T_A)\in\tilde{A}.$$

Hence $m(A)=m^{T_A}(\tilde{A})$, where, for any subset S of T we define the measure m^S on $\Pi_{t\in S}\mathcal{A}_t$ by the formula

$$m^S(C)=m(C\times\Pi_{t\in T\setminus S}X_t).$$

But m^{T_A} is a finite measure and hence regular from above. Therefore, if we define for each $n=1,2,\cdots$, the set

$$\tilde{A}_n=\cup_{a\in\tilde{A}}[(a(t_1),\ \ldots,\ a(t_n))\times\prod_{i=n+1}^{\infty}X_{t_i}],$$

then we see that $\tilde{A}_n\supseteq\tilde{A}_{n+1}$ and $\bigcap_{n=1}^{\infty}\tilde{A}_n=\tilde{A}$. Hence

$$\lim_{n\to\infty}m^{T_A}(\tilde{A}_n)=m^{T_A}(\tilde{A}),\ \text{and thus}$$
$$|m^{T_A}(\tilde{A})-m^{T_A}(\tilde{A}_n)|<\epsilon,$$

for large enough n. By the construction of the sets $\tilde{A}_n$, if we set $U=(t_1,t_2,\cdots,t_n)$ and $B=\cup_{a\in\tilde{A}}(a(t_1),a(t_2),\cdots,a(t_n))$, then $m^U(B)=m^{T_A}(\tilde{A}_n)$ and the result is proved.

Lemma 3: Let $\mathcal{A}$ be a σ-algebra of subsets of X; and assume that for each $x\in H$ there is given a finite positive measure μ_x on $\mathcal{A}$. Then, there exists a positive operator-valued measure μ on $\mathcal{A}$ such that $\langle\mu(A)x,x\rangle=\mu_x(A)$ for every $A\in\mathcal{A}$ and $x\in H$ if and only if

(i) $[\mu_{x+y}(A)]^{1/2}\le[\mu_x(A)]^{1/2}+[\mu_y(A)]^{1/2}$,

(ii) $\mu_{\alpha x}(A)=|\alpha|^2\mu_x(A)$,

(iii) $\mu_{x+y}(A)+\mu_{x-y}(A)=2\mu_x(A)+2\mu_y(A)$

for all x, $y\in H$, complex α, and $A\in\mathcal{A}$. Also, for every $A\in\mathcal{A}$ there is a constant $\kappa_A\ge 0$ such that

(iv) $\mu_x(A)\le\kappa_A\|x\|^2$.

For the proof of this lemma we refer to Willansky.[6]

Proof of Theorem: Let $x=H$ and assume $\|x\|=1$. For each nonempty, finite subset V of T define the probability measure $\mu_x^V(\cdot)=\langle\mu^V(\cdot)x,x\rangle$ on $\mathcal{A}_V$. Then for the family $\{\mu_x^V\}_V$ Lemma 1 is applicable; hence there exists a probability measure μ_x on $\Pi_{t\in T}\mathcal{A}_t$ such that $\mu_x^V(A)=\mu_x(\pi_{V,\infty}(A))$, where $\mu_{V,\infty}$ is the canonical injection of $\mathcal{A}_V$ into $\mathcal{A}$.

Now, let x be any element of H. If $x=\theta$ set $\mu_\theta=0$, and if $x\neq\theta$ set $\mu_x=\|x\|^2\mu_{(x/\|x\|)}$. Thus, applying Lemma 2 we have that for every $\alpha\in Z$, x and $y\in H$, A in $\mathcal{A}$ and $\epsilon>0$ there exists a finite subset U of T for which,

$$(*)\quad\begin{aligned}&|\mu_x(A)-\mu_x^U(F)|<\epsilon,\\&|\mu_\alpha(A)-\mu_{\alpha x}^U(F)|<\epsilon,\\&|\mu_y(A)-\mu_y^U(F)|<\epsilon,\\&|\mu_{x+y}(A)-\mu_{x+y}^U(F)|<\epsilon,\ \text{and}\\&|\mu_{x-y}(A)-\mu_{x-y}^U(B)|<\epsilon,\end{aligned}$$

where F is as in Lemma 2.

Hence we see that

(i) $$\begin{aligned}\mu_{x+y}(A)&\le\mu_{x+y}^U(E)+2\epsilon\\&\le\{[\mu_x^U(F)]^{1/2}+[\mu_y^U(E)]^{1/2}\}^2+2\epsilon\\&\le\{[\mu_x(A)+\epsilon]^{1/2}+[\mu_y(A)+\epsilon]^{1/2}\}^2+2\epsilon\end{aligned}$$

and so, letting $\epsilon\to 0$,

$$\text{i.e.,}\qquad\mu_{x+y}(A)\le\{[\mu_x(A)]^{1/2}+[\mu_y(A)]^{1/2}\}^2,$$
$$[\mu_{x+y}(A)]^{1/2}\le[\mu_x(A)]^{1/2}+[\mu_y(A)]^{1/2}.$$

(ii) $$\begin{aligned}&|\mu_{\alpha x}(A)-|\alpha|^2\mu_x(A)|\\&\le|\mu_{\alpha x}(A)-\mu_{\alpha x}^U(F)|+|\mu_{\alpha x}^U(B)-|\alpha|^2\mu_x^U(F)|\\&\quad+|\alpha|^2\mu_x^U(F)-|\alpha|^2\mu_x(A)|\le\epsilon+0+|\alpha|^2\epsilon\end{aligned}$$

and, as $\epsilon\to 0$ the right-hand side goes to zero. Thus,

$$\mu_{\alpha x}(A)=|\alpha|^2\mu_x(A).$$

(iii) $$\begin{aligned}&|\mu_{x+y}(A)+\mu_{x-y}(A)-[2\mu_x(A)+2\mu_y(A)]|\\&\le|\mu_{x+y}(A)-\mu_{x+y}^U(B)|+|\mu_{x-y}(A)-\mu_{x-y}^U(F)|\\&\quad+|\mu_{x+y}^U(B)+\mu_{x-y}^U(B)-[2\mu_x^U(B)+2\mu_y^U(B)]|\\&\quad+2|\mu_x^U(B)-\mu_x(A)|+2|\mu_y^U(F)-\mu_y(A)|\le 6\epsilon,\end{aligned}$$

and as $\epsilon\to 0$, we have

$$\mu_{x+y}(A)+\mu_{x-y}(A)-[2\mu_x(A)+2\mu_y(A)]=0.$$

Finally,

(iv) $$\|x\|^2-\mu_x(A)\ge\|x\|^2-(\mu_x^U(B)-\epsilon)\ge-\epsilon$$

so $\|x\|^2-\mu_x(A)\ge 0$.

Thus, by Lemma 3, there exists a unique PNOV measure μ for which $\langle\mu(\cdot)x,x\rangle=\mu_x(\cdot)$. The topological conditions on the X_t were used in two places: First in Lemma 1, the scalar-valued extension theorem and then in the approximations $(*)$. As is well known (cf. Wegner[7]) these conditions are essential to the existence of the extending measure in the scalar-valued case.

4. ALGEBRAIC MODELS FOR PNOV MEASURES

In this section we (i) define algebraic models for PNOV measures, thereby extending the idea of Dincu-

leanu and Foias to this case, (ii) give a harmonic-analytic result for PNOV measures, and (iii) give an extension theorem for PNOV measures which is an analog of the main result of Schreiber *et al.*

Let $(X, \mathcal{A}, \mu)$ be a PNOV measure space. In order to carry out the formulation of an algebraic model for $(X, \mathcal{A}, \mu)$, or μ, we define the space $L_2(\mu)$. A function $f: X \to Z$, $\mathcal{A}$-measurable, is said to belong to $L_2(\mu)$ is there is some finite number M such that for every $x \in H$

$$\int |f|^2 d\mu_x \leq M\|x\|^2.$$

The smallest value of M which satisfies the above will be denoted by $\|f\|_2^2$. We remark that the simple functions are dense in $L_2(\mu)$. Furthermore, $\|f\|_2 = (\|f\|_2^2)^{1/2}$ is a norm for $L_2(\mu)$.

Proposition 1: $L_2(\mu)$ is a Banach space.

Proof: Suppose $\{f_n\}$ is Cauchy in $L_2(\mu)$. Then, there is a subsequence $\{f_{n_k}\}$ such that $\|f_{n_{k+1}} - f_{n_k}\|_2 < 2^{-k}$.

Define

$$g_K = \sum_{i=1}^{K} |f_{n_{i+1}} - f_{n_i}|,$$

and

$$g = \sum_{i=1}^{\infty} |f_{n_{i+1}} - f_{n_i}|.$$

Then

$$\|g_K\|_2 \leq \sum_{i=1}^{K} \|f_{n_{i+1}} - f_{n_i}\|_2 < 1.$$

Hence, for any $x \in H$,

$$\int |g_K|^2 d\mu_x \leq \|x\|^2,$$

where, as before, $\mu_x = \langle \mu x, x\rangle$. We now apply Fatou's lemma to the measures μ_x;

$$\int |g|^2 d\mu_x \leq \|x\|^2.$$

Therefore,

$$f_{n_1} + \sum_{i=1}^{\infty} (f_{n_{i+1}} - f_{n_i})$$

converges almost surely $[\mu]$. Call the limit element f. Then, given any $\epsilon > 0$, there is an N for which $m, n > N$ implies $\|f_n - f_m\|_2 < \epsilon$. Hence, for each $x \in H$,

$$\int |f - f_m|^2 d\mu_x \leq \lim_{i \to \infty} \inf \int |f_{m_i} - f_m|^2 d\mu_x < \epsilon^2 \|x\|^2.$$

Thus $f \in L_2(\mu)$, and $f_n \to f$ in the L_2-norm.

We remark that the space $L_2(\mu)$ is not, even if H is two-dimensional, a Hilbert space.

To prove Proposition 3, in which the prototype algebraic model for a PNOV measure is given, we shall need a fact about the structure of $L_2(\mu)$.

Proposition 2: The simple functions are dense in $L_2(\mu)$.

The proof is omitted, being exactly as in the case of a scalar-valued measure.

Definition 1: A function $\phi: \Gamma \to \mathcal{L}(H)$ is called a PNOV *function* if

(i) $\phi(\gamma) = I$, if $\gamma = e$ – the identity of Γ, and

(ii) for every N and choice of $\gamma_1, \cdots, \gamma_N \in \Gamma$, $z_1, \cdots, z_n \in Z$, the operator $\sum_{i,j=1}^{N} z_i \bar{z}_j \phi(\gamma_i \gamma_j^{-1})$ is positive. Then for each $x \in H$ the function $\phi_x(\gamma) = \langle \phi(\gamma) x, x\rangle$ is positive-definite.

Proposition 3: Let $\Gamma^\mu = \{f; X \to Z \mid f \text{ is } \mathcal{A}\text{-measurable}, |f| = 1\}$ and define $\phi^\mu: \Gamma^\mu \to \mathcal{L}(H)$ by the formula

$$\phi^\mu_{x,y}(f) = \int f d\mu_{x,y}.$$

Then

(i) the linear span of Γ^μ is dense in $L_2(\mu)$

(ii) ϕ^μ is a PNOV function.

Again, the proofs are immediate, that of (i) following from Proposition 2.

We can now consider the idea of an algebraic model for an operator-valued measure.

Definition 2: An ordered pair (Γ, ϕ) where Γ is an Abelian group and ϕ is a PNOV function defined on Γ will be called an *algebraic measure system*.

Definition 3: An algebraic measure system (Γ, ϕ) is said to be an *algebraic model* for the PNOV measure space $(X, \mathcal{A}, \mu)$ if there is an injective homomorphism h: $\Gamma \to \Gamma^\mu$ such that

(i) the linear span of $h(\Gamma)$ is dense in $L_2(\mu)$, and

(ii) $\phi(\gamma) = \int_X h(\gamma)\, d\mu$.

Thus every PNOV measure space possesses at least one algebraic model, namely (Γ^μ, ϕ^μ). Furthermore, every algebraic measure system is the algebraic model for *some* PNOV measure space. Indeed, we have:

Theorem 2: Let Γ be an Abelian group, and let ϕ be a PNOV function on Γ into $\mathcal{L}(H)$. Then, when Γ is endowed with the discrete topology there exists a PNOV measure ν on G, the (compact) dual of Γ, for which ϕ is its Fourier–Stieltjes transform; that is for every $\gamma \in \Gamma$ and $x, y \in H$

$$\langle \phi(\gamma)x, y\rangle = \int_G (\gamma, g)\, d\nu_{x,y},$$

where $(\cdot, \cdot)$ denotes the duality relation on $\Gamma \times G$.

Proof: Fix $x \in H$, and assume $\|x\| = 1$. Then ϕ_x is a positive-definite function on Γ and $\mu_x(\gamma) = 1$ if $\gamma = e$. Hence, there exists a probability measure ν_x on the Borel σ-algebra of G such that (cf. Rudin[8])

$$\phi_x(\gamma) = \int_G (\gamma, g)\, d\nu_x.$$

Now, for any $x \in H$, define

$$\nu_x = \begin{cases} 0, & \text{if } x = \theta, \\ \|x\|^2 \nu_{(x/\|x\|)}, & \text{if } x \neq \theta. \end{cases}$$

We claim the conditions of Lemma 3, Sec. 3, are satisfied. Clearly conditions (ii) and (iv) are satisfied. Condition (iii) follows directly from the uniqueness of the Fourier–Stieltjes transform in the scalar-valued measure case. To establish (i) we proceed as follows: Define the complex measure (for each $x, y \in H$)

$$\nu_{x,y} = \tfrac{1}{4}[\nu_{x+y} - \nu_{x-y} + i(\nu_{x+iy} - \nu_{x-iy})].$$

Taking the Fourier–Stieltjes transform of $\nu_{x,y}$ we have

$$\mathcal{J}(\nu_{x,y}) = \tfrac{1}{4}[\phi_{x+y} - \phi_{x-y} + i(\phi_{x+iy} - \phi_{x-iy})],$$

where $\mathcal{J}$ denotes the Fourier–Stieltjes transform, and therefore

$$\phi_{x+y} = \phi_x + \phi_y + \mathcal{J}(\nu_{x,y}) + \mathcal{J}(\nu_{y,x}).$$

Hence upon inversion,

$$\nu_{x+y} = \nu_x + \nu_y + \nu_{x,y} + \nu_{y,x}.$$

Now, $\nu_{x+y}(A)$ is a nonnegative number for every measurable set A. So, replacing $x+y$ by $x+\alpha y$, we have

$$0 \leq \nu_{x+\alpha y} = \nu_x + |\alpha|^2 \nu_y + \bar{\alpha}\nu_{x,y} + \alpha\nu_{x,y};$$

and putting $\alpha = \nu_{x,y}/\nu_y$, we obtain $|\nu_{x,y}|^2 \leq \nu_x \nu_y$. Therefore, $(\nu_{x+y})^{1/2} \leq \nu_x^{1/2} + \nu_y^{1/2}$.

From the above we conclude that there exists a PNOV measure ν on G for which

$$\langle \nu(A)x, x\rangle = \nu_x(A).$$

Therefore,

$$\langle \phi(\gamma)x, x\rangle = \int_G (\gamma, g)\, d\nu_x;$$

and

$$\langle \phi(\gamma)x, y\rangle = \int_G (\gamma, g)\, d\nu_{x,y}.$$

We remark that Neumark[9] proved a similar result using a different definition of ϕ, namely that of an operator of positive type. Indeed, the two theorems together show that the definitions are equivalent. Neumark's paper, which appeared in 1943, uses unitary dilations of one-parameter families of operators. It seems that his paper marks the first appearance of this method of proof. This impression is reinforced by Mlak,[10] who credits Neumark with the introduction of this important idea.

We now see that associated with every PNOV measure space $(X, \mathcal{A}, \mu)$ is the model (Γ^μ, ϕ^μ) which is likewise a model for $(G, \mathcal{B}, \nu)$. The relationship between $(X, \mathcal{A}, \mu)$ and $(G, \mathcal{B}, \nu)$ is given by Lemma 4.

Lemma 4: There is a measurable map $\psi: X \to G$ and a mapping $\lambda: L_2(\mu) \to L_2(\nu)$ for which $\lambda(f)\circ\psi = f$ $(\mu - \text{a. s.})$ for each $f \in L_2(\mu)$.

Proof: First the mapping λ is constructed using the fact that Γ_μ is a subset of both $L_2(\nu)$ and $L_2(\mu)$. For each $\gamma \in \Gamma_\mu$, regarded as a function on X, define the function $\lambda(\gamma) = (\cdot, \gamma)$ on G. Since $|\lambda(\gamma)| \equiv 1$ it lies in $L_2(\nu)$ and in fact, since (Γ^μ, ϕ^μ) is a model for $(G, \mathcal{B}, \nu)$, the set $\lambda(\Gamma)$ will generate $L_2(\nu)$. Hence, for any $L_2(\mu)$ sequence of linear combinations $f_n = \sum_{i=1}^\infty C_i^n \gamma_i^n$ approximating f we define $\lambda(f_n) = \sum_{i=1}^\infty C_i^n(\cdot, \gamma_i^n)$. That $\{\lambda(f_n)\}$ is Cauchy in $L_2(\nu)$ is immediate from the fact that,

$$\int_X \gamma\, d\mu = \phi^\mu(\gamma) = \int_G (\cdot, \gamma)\, d\nu.$$

Hence we set $\lambda(f) = \lim_{n\to\infty} \lambda(f_n)$ in $L_2(\nu)$. Now, define $\psi: X \to G$ via $(\psi(x), \gamma) = \gamma(x)$ for every $\gamma \in \Gamma$ and $x \in X$. To see that $\lambda(f)\circ\psi = f(\mu - \text{a. s.})$ we set, as before,

$$f_n = \sum_{i=1}^\infty C_i^n \gamma_i^n \text{ with } f = \lim_{n\to\infty} f_n.$$ Then we have

$$\lambda(f_n)\circ\psi(x) = \sum_{i=1}^\infty C_i^n(\cdot, \gamma_i^n)\circ\psi(x)$$

$$= \sum_{i=1}^\infty C_i^n(\lambda(x), \gamma_i^n) = \sum_{i=1}^\infty C_i^n \gamma_i^n(x).$$

So $\lambda(f_n)\circ\psi(x)$ and f_n lie in the same $L_2(\mu)$ equivalence class, so that $\lambda(f_n)\circ\psi = f_n(\mu - \text{a. s.})$. Since f is the $L_2(\mu)$ limit of the f_n we see that likewise $\lambda(f)\circ\psi = f(\mu - \text{a. s.})$.

Before considering the extension theorem, we make one additional remark. Suppose $f = \chi_A$, the characteristic function of the $\mathcal{A}$-measurable set A. Then, since $\lambda(f)\circ\psi = f$ we see that $\lambda(\chi_A)$ is a. s. $[\mu]$ the characteristic function of some set $F \in \mathcal{B}$, provided μ is complete. Hence $\psi^{-1}(F)$ is a measurable subset of X and $\mu(A \Delta \psi^{-1}(F)) = 0$ since

$$\chi_A = \chi_{\psi^{-1}(B)}.$$

Now let T be a directed set indexing an increasing, consistent family of PNOV measure spaces $\{(X, \mathcal{A}_t, \mu_t)\}$, $t \in T$. Then there is a corresponding family of models $\{(\Gamma_t, \phi_t)\}$, $t \in T$, where $\Gamma_t = \Gamma^{\mu_t}$ and $\phi_t(\gamma) = \int \gamma\, d\mu_t$ for all $\gamma \in \Gamma_t$. Further, the family $\{(\Gamma_t, \phi_t)\}$ is increasing: $\Gamma_s \subseteq \Gamma_t$ if $s \leq t$, since $\mathcal{A}_s \subseteq \mathcal{A}_t$ for the same pair (s, t). In addition the family is consistent in the sense that $\phi_s(\gamma) = \phi_t(\gamma)$ if $s \leq t$ and $\gamma \in \Gamma_s$. Thus the set $\Gamma = \cup_{t\in T}\Gamma_t$ is a group and the function $\phi: \Gamma \to L(H)$, defined by $\phi(\gamma) = \phi_t(\gamma)$ for $\gamma \in \Gamma_t$, is a PNOV function, since each ϕ_t is a PNOV function. Furthermore, by Theorem 2, the algebraic measure system (Γ, ϕ) is an algebraic model for the PNOV-measure space $(G, \mathcal{B}, \nu)$, where, as before, G is the dual of Γ and ϕ is the Fourier–Stieltjes transform of ν.

Now, each of the algebraic measure systems (Γ_t, ϕ_t) also yields a dual group G_t with Borel σ-algebra $\mathcal{B}(G_t)$ (where Γ_t was given the discrete topology) and a PNOV measure ν_t, whose Fourier–Stieltjes transform is ϕ_t, so that $(X, \mathcal{A}_t, \mu_t)$ and $(G_t, \mathcal{B}(G_t), \nu_t)$ are conjugate as in Lemma 4 by means of the mapping $\psi = \psi_t$.

Furthermore, since $\Gamma_s \subseteq \Gamma_t$ for $s \leq t$ we have that G_t is topologically isomorphic to a quotient group of G_s. Likewise, G is topologically isomorphic to a quotient of each G_t. Denoting by ϵ_t the canonical mapping of G_t onto G we can define, for each t, a σ-algebra of G

$$\mathcal{B}_t = \{\epsilon_t(E) \mid E \in \mathcal{B}(G_t)\}.$$

The family of σ-algebras $\mathcal{B}_t$ of G is directed in the sense that $\mathcal{B}_s \subseteq \mathcal{B}_t$ for $s \leq t$. Furthermore, for each t we can define a probability measure ν_t over $\mathcal{B}_t$ by choosing ν_t to be that unique measure having ϕ_t as its Fourier–Stieltjes transform. The family of measures ν_t will, like μ_t, be consistent.

We summarize the above remarks in Theorem 3, which is a direct generalization of the development in Ref. 4.

Theorem 3: Let $(X, \mathcal{A}_t, \mu_t)$, $t \in T$, be an increasing family of PNOV-measure spaces indexed by a directed set T. Assume the family of PNOV measures $\{\mu_t\}$ is consistent. Then there exists an increasing sequence of PNOV-measure spaces $\{(G, \mathcal{B}_t, \nu_t)\}$ over a group G for which

(i) the family $\{(G, \mathcal{B}_t, \nu_t)\}$ has an extension,

and

(ii) if for each $t \in T$ μ_t is complete then there exists a

mapping $\psi_t: X \to G$ which is $\mathcal{B}_t$-measurable, and for each $A \in \mathcal{A}_t$ there is a $B \in \mathcal{B}_t$ for which $\mu_t(A \Delta \psi_t^{-1}(B)) = 0$.

ACKNOWLEDGMENTS

The author would like to extend his thanks to Professor A. T. Bharucha-Reid for bringing the works of Paul Benioff to his attention and to the referee for his helpful comments.

[1]P. A. Benioff, "Operator valued measures in quantum mechanics: Finite and infinite processes," J. Math. Phys. **13**, 231—42 (1972).
[2]P. A. Benioff, "Procedures in Quantum Mechanics without von Neumann's Projection Axiom," J. Math. Phys. **13**, 1347—55 (1972).
[3]N. Dinculeanu and C. Foias, "Algebraic models for measures," Ill. J. Math. **12**, 340—51 (1968).
[4]B. M. Schreiber, T. C. Sun, and A. T. Bharucha—Reid, "Algebraic models for probability measures associated with stochastic processes," Trans. Am. Math. Soc. **158**, 93—105.
[5]R. Ash, *Real Analysis and Probability* (Academic, New York, 1972).
[6]A. Willansky, *Functional Analysis* (Blaisdell, New York, 1964).
[7]H. W. Wegner, "On Consistency of Probability Measures," Zeit. Wahrsh. Ver. Geb. **27**, 335—8 (1973).
[8]W. Rudin, *Four Analysis on Groups* (Wiley, New York, 1962).
[9]M. Neumark, "Positive definite functions on a commutative group (Russian)," Izv. Akad. Nauk. SSR **7**, 237—44 (1943).
[10]W. Mlak, "Unitary dilations of contractions operators," Rozpr. Mat. **46**, Instytut Matematyczny, PAN Warsaw (1965).

Representations of a para-Bose algebra using only a single Bose field

R. Jagannathan

Department of Physics, St. Joseph's College, Tiruchirapalli-620002, India

R. Vasudevan

Matscience, Institute of Mathematical Sciences, Madras-600020, India
(Received 24 May 1978)

Representations of a para-Bose algebra are given entirely in terms of the operators of a single Bose field. Incidentally some types of boson representations of a nonsemisimple graded Lie algebra of the three-dimensional Lorentz group are obtained.

1. INTRODUCTION

In a recent contribution[1] we have produced the representations of para-Fermi operators of any order of the statistics using only the operators of a single Fermi field. Representations of para-Fermi fields has been obtained earlier by Ramakrishnan and coworkers[2] employing the elements of Clifford algebra and by Kalnay and others[3] utilizing the boson operators and suitably restricted subspaces of the boson vector space. It is the aim of this paper to obtain the representations of para-Bose operators of any order of the statistics using only the operators of a single Bose field.

Green[4] was the first to give an ansatz for representing the para-Fermi and para-Bose algebras by forming linear combinations of several different commuting Fermi fields and anticommuting Bose fields respectively. Using Green's ansatz here we obtain the representations of a para-Bose algebra by forming the required different types of anticommuting Bose fields using certain combinations of suitable number of operators of the same basic Bose field.

In Sec. 2 we show how one could get two sets of anticommuting Bose operators having n operators each and hence construct para-Bose operators of second order of the statistics. In Sec. 3 we generalize this approach and demonstrate how one could manufacture any required number of para-Bose operators of any order of the statistics starting from a single basic Bose field. In Sec. 4 the results of Secs. 2 and 3 are shown to provide some types of boson representations of a nonsemisimple graded Lie algebra of the three-dimensional Lorentz group.

2. PARA-BOSE OPERATORS OF ORDER 2

Let us take $2n+1$ operators $\{b_j \mid j=1, 2, \ldots, 2n+1\}$ and their Hermitian conjugates of a single Bose field obeying the well known commutation relations

$$\begin{aligned} &[b_j, b_k]=[b_j^\dagger, b_k^\dagger]=0,\\ &[b_j, b_k^\dagger]=\delta_{jk},\\ &j,k=1, 2, \ldots, 2n+1. \end{aligned} \tag{1}$$

Let us now construct the unitary operators U_1 and V_1 as

$$\begin{aligned} U_1&=\exp\left[i\left(\frac{\pi}{2}\right)^{1/2}(b_1^\dagger+b_1)\right],\\ V_1&=\exp\left[\left(\frac{\pi}{2}\right)^{1/2}(b_1^\dagger-b_1)\right]. \end{aligned} \tag{2}$$

It is easily seen that U_1 and V_1 satisfy the relations

$$\{U_1, V_1\}=\{U_1, V_1^\dagger\}=\{U_1^\dagger, V_1\}=\{U_1^\dagger, V_1^\dagger\}=0. \tag{3}$$

Then the set of operators $\{b_j^{(\alpha)} \mid j=1, 2, \ldots, n;\ \alpha=1, 2\}$ defined by

$$\begin{aligned} &b_j^{(1)}=U_1 b_{j+1}, \qquad b_j^{(1)\dagger}=U_1^\dagger b_{j+1}^\dagger,\\ &b_j^{(2)}=V_1 b_{n+j+1}, \qquad b_j^{(2)\dagger}=V_1^\dagger b_{n+j+1}^\dagger,\\ &j=1, 2, \ldots, n, \end{aligned} \tag{4}$$

form two anticommuting sets of boson operators such that

$$\begin{aligned} &[b_j^{(\alpha)}, b_k^{(\alpha)}]=[b_j^{(\alpha)\dagger}, b_k^{(\alpha)\dagger}]=0,\\ &[b_j^{(\alpha)}, b_k^{(\alpha)\dagger}]=\delta_{jk},\\ &\{b_j^{(1)}, b_k^{(2)}\}=\{b_j^{(1)}, b_k^{(2)\dagger}\}\\ &\qquad=\{b_j^{(1)\dagger}, b_k^{(2)}\}=\{b_j^{(1)\dagger}, b_k^{(2)\dagger}\}=0,\\ &j,k=1, 2, \ldots, n;\quad \alpha=1, 2. \end{aligned} \tag{5}$$

Let us now follow Green's ansatz[4] to arrive at the desired operators of the para-Bose algebra of second order statistics. Thus defining

$$\begin{aligned} \beta_j^{(2)}&=b_j^{(1)}+b_j^{(2)}=\exp\left[i\left(\frac{\pi}{2}\right)^{1/2}(b_1^\dagger+b_1)\right]b_{j+1}\\ &\quad+\exp\left[\left(\frac{\pi}{2}\right)^{1/2}(b_1^\dagger-b_1)\right]b_{n+j+1},\\ \beta_j^{(2)\dagger}&=b_j^{(1)\dagger}+b_j^{(2)\dagger}=\exp\left[-i\left(\frac{\pi}{2}\right)^{1/2}(b_1^\dagger+b_1)\right]b_{j+1}^\dagger\\ &\quad+\exp\left[\left(\frac{\pi}{2}\right)^{1/2}(b_1-b_1^\dagger)\right]b_{n+j+1}^\dagger,\\ j&=1, 2, \ldots, n, \end{aligned} \tag{6}$$

it is easily seen that these satisfy the general para-Bose commutation relations

$$\begin{aligned} &[\beta_j, \{\beta_k^\dagger, \beta_l\}]=2\delta_{jk}\beta_l,\\ &[\beta_j, \{\beta_k, \beta_l\}]=0,\\ &j,k,l=1, 2, \ldots, n. \end{aligned} \tag{7}$$

Green[4] had suggested that such para-Bose operators $\{\beta\}$ can be expressed as suitable direct products of boson operators and Pauli spin matrices. It is emphasized that here we are constructing the para-Bose operators using only the elements of a single Bose field

 0022-2488/79/030390-03$01.00

without any recourse to other types of entities like the Pauli matrices. Also we do not have any restriction on the boson vector space on which the above operators $\{\beta\}$ act, unlike the situation noticed in the works of Kalnay and others.[3]

3. PARA-BOSE OPERATORS OF ANY ORDER

Now we shall generalize the above procedure to construct a set of n para-Bose operators of order p.

In the last section we constructed two anticommuting operators U_1 and V_1 from b_1 and $b_1^\dagger$. As it is well known a third unitary operator anticommuting with both U_1 and V_1 can be found as U_1V_1.

If we now consider operators $\{b_j, b_j^\dagger \mid j=1, 2, \ldots, m\}$ we can define the operators analogous to those of Eq. (2) given by

$$U_j = \exp\left[i\left(\frac{\pi}{2}\right)^{1/2}(b_j^\dagger + b_j)\right],$$
$$V_j = \exp\left[\left(\frac{\pi}{2}\right)^{1/2}(b_j^\dagger - b_j)\right], \tag{8}$$
$$j=1, 2, \ldots, n,$$

such that

$$\{U_j, V_j\} = \{U_j, V_j^\dagger\} = \{U_j^\dagger, V_j\} = \{U_j^\dagger, V_j^\dagger\} = 0,$$
$$[U_j, U_k] = [U_j, U_k^\dagger] = [U_j, V_k] = [U_j, V_k^\dagger] = [V_j, V_k] = [V_j, V_k^\dagger] = 0, \quad j \neq k. \tag{9}$$
$$j, k = 1, 2, \ldots, m.$$

From these operators of Eq. (8) we can construct the following mutually anticommuting unitary operators:

$$A_1 = U_1, \quad A_2 = V_1,$$
$$A_{2r-1} = \prod_{l=1}^{r-1}(U_lV_l)\, U_r, \tag{10}$$
$$A_{2r} = \prod_{l=1}^{r-1}(U_lV_l)\, V_r,$$
$$r = 2, 3, \ldots, m,$$

and

$$A_{2m+1} = \prod_{l=1}^{m}(U_lV_l). \tag{11}$$

These and their Hermitian conjugates obey the relations

$$\{A_\alpha, A_\mu\} = \{A_\alpha, A_\mu^\dagger\} = \{A_\alpha^\dagger, A_\mu\} = \{A_\alpha^\dagger, A_\mu^\dagger\} = 0, \quad \alpha \neq \mu, \tag{12}$$
$$\alpha, \mu = 1, 2, \ldots, 2m+1.$$

To construct a set of n para-Bose operators of order $p = 2m$ or $2m+1$ we have to take the set of operators

$$\{b\}_n^p = \{b_j \mid j = 1, 2, \ldots, [p/2] + pn\} \tag{13}$$

of the basic single Bose field, where $[p/2]$ denotes m, the integer part of $p/2$. Any m operators of the set $\{b\}_n^p$ may be used to construct the required p mutually anticommuting operators $\{A\}$ of Eqs. (10) and (11). Choosing the first m operators $\{b_j \mid j=1, 2, \ldots, m\}$ for this purpose as in Eqs. (10) and (11) let us further define

$$b_j^{(\alpha)} = A_\alpha b_{m+(\alpha-1)n+j},$$
$$b_j^{(\alpha)\dagger} = A_\alpha^\dagger b_{m+(\alpha-1)n+j}^\dagger, \tag{14}$$
$$j=1, 2, \ldots, n; \quad \alpha = 1, 2, \ldots, p.$$

As it can be seen easily these $\{(b_j^{(\alpha)} \mid j=1, 2, \ldots n) \mid \alpha = 1, 2, \ldots, p\}$ constitute a set of p anticommuting sets of boson operators obeying

$$[b_j^{(\alpha)}, b_k^{(\alpha)}] = [b_j^{(\alpha)\dagger}, b_k^{(\alpha)\dagger}] = 0,$$
$$[b_j^{(\alpha)}, b_k^{(\alpha)\dagger}] = \delta_{jk},$$
$$\{b_j^{(\alpha)}, b_k^{(\mu)}\} = \{b_j^{(\alpha)}, b_k^{(\mu)\dagger}\} = \{b_j^{(\alpha)\dagger}, b_k^{(\mu)}\} = \{b_j^{(\alpha)\dagger}, b_k^{(\mu)\dagger}\} = 0, \quad \alpha \neq \mu,$$
$$j, k = 1, 2, \ldots, n; \quad \alpha, \mu = 1, 2, \ldots, p, \tag{15}$$

analogous to Eq. (5). Hence by Green's ansatz we can easily find the para-Bose operators of order p as

$$\beta_j^{(p)} = \sum_{\alpha=1}^{p} b_j^{(\alpha)},$$
$$\beta_j^{(p)\dagger} = \sum_{\alpha=1}^{p} b_j^{(\alpha)\dagger}, \tag{16}$$
$$j = 1, 2, \ldots, n.$$

Then using Eqs. (8), (10), (11), and (14) the explicit expressions for the para-Bose operators of order $2m$ or $2m+1$ are found to be

$$\beta_j^{(2m)} = \sum_{r=1}^{m} i^{r-1}\left[\prod_{l=1}^{r-1}\exp(zb_l^\dagger - z^*b_l)\right] \times \{\exp[i(\pi/2)^{1/2}(b_r^\dagger + b_r)]b_{m+(2r-2)n+j} + \exp[(\pi/2)^{1/2}(b_r^\dagger - b_r)]b_{m+(2r-1)n+j}\}, \tag{17}$$

and

$$\beta_j^{(2m+1)} = \sum_{r=1}^{m} i^{r-1}\left[\prod_{l=1}^{r-1}\exp(zb_l^\dagger - z^*b_l)\right] \times \{\exp[i(\pi/2)^{1/2}(b_r^\dagger + b_r)]b_{m+(2r-2)n+j} + \exp[(\pi/2)^{1/2}(b_r^\dagger - b_r)]b_{m+(2r-1)n+j}\} + i^m\left[\prod_{l=1}^{m}\exp(zb_l^\dagger - z^*b_l)\right]b_{m+2mn+j}, \tag{18}$$
$$j = 1, 2, \ldots, n,$$

where $z = (\pi/2)^{1/2}(1+i)$. These operators $\{\beta_j^{(p)} \mid j = 1, 2, \ldots, n\}$ are easily seen to obey the para-Bose commutation relations (7). Thus we have obtained a set of n para-Bose operators of pth order statistics using only the operators of a single Bose field.

4. REPRESENTATIONS OF A GSO(2,1)

It is known[5] that the generators $\{Q_1, Q_2, Q_3\}$ of the three-dimensional Lorentz group [SO(2, 1)] obeying the Lie algebra relations

$$[Q_1, Q_2] = -iQ_3, \quad [Q_2, Q_3] = iQ_1, \quad [Q_3, Q_1] = iQ_2, \tag{19}$$

Can be represented in terms of any pair of conjugate para-Bose operators $(\beta, \beta^\dagger)$ of any order. Following Ref. 5 we have then

$$Q_1 = \frac{i}{4}(\beta^\dagger\beta^\dagger - \beta\beta),$$
$$Q_2 = \tfrac{1}{4}(\beta^\dagger\beta^\dagger + \beta\beta), \tag{20}$$
$$Q_3 = \tfrac{1}{4}(\beta^\dagger\beta + \beta\beta^\dagger).$$

Now it is evident that each of the representations of $\{\beta\}$ obtained above provides a representation of SO(2, 1) Lie algebra.

A graded Lie algebra involving both commutators and anticommutators has the general form given by

$$\begin{aligned}
&[Q_m, Q_n] = \sum_l f^l_{mn} Q_l, \\
&[Q_m, W_\lambda] = \sum_\mu F^\mu_{m\lambda} W_\mu, \\
&\{W_\lambda, W_\mu\} = \sum_m S^m_{\lambda\mu} Q_m, \\
&m, n = 1, 2, \ldots, D; \quad \lambda, \mu = D+1, D+2, \ldots, D+d.
\end{aligned} \tag{21}$$

Let us now observe that the set $\{Q_1, Q_2, Q_3, W_4 = \beta^\dagger, W_5 = \beta\}$ generates such a graded Lie algebra of SO(2, 1) [GSO(2, 1)] with the structure constants $\{f^l_{mn}, F^\mu_{m\lambda}, S^m_{\lambda\mu}\}$ given as follows. The only nonvanishing elements of $\{f^l_{mn} \mid l, m, n = 1, 2, 3\}$ are

$$\begin{aligned}
&f^1_{23} = -f^1_{32} = i, \\
&f^2_{31} = -f^2_{13} = i, \\
&f^3_{12} = -f^3_{21} = -i,
\end{aligned} \tag{22}$$

The constants $\{(F^\mu_{m\lambda} \mid \mu, \lambda = 4, 5) \mid m = 1, 2, 3\}$ represented for each n as a 2×2 matrix are given by

$$F_1 = \begin{pmatrix} 0 & -i/2 \\ -i/2 & 0 \end{pmatrix}, \quad F_2 = \begin{pmatrix} 0 & i/2 \\ i/2 & 0 \end{pmatrix}, \quad F_3 = \begin{pmatrix} -1/2 & 0 \\ 0 & 1/2 \end{pmatrix}. \tag{23}$$

Similarly the constants $\{(S^m_{\lambda\mu} \mid \lambda, \mu = 4, 5) \mid m = 1, 2, 3\}$ can be represented by the matrices

$$S^1 = \begin{pmatrix} -4i & 0 \\ 0 & 4i \end{pmatrix}, \quad S^2 = \begin{pmatrix} 4 & 0 \\ 0 & 4 \end{pmatrix}, \quad S^3 = \begin{pmatrix} 0 & 4 \\ 4 & 0 \end{pmatrix}. \tag{24}$$

Following Ref. 6 one can introduce a metric tensor $\{g_{kl} \mid k, l = 1, 2, \ldots, D+d\}$ of the graded Lie algebra along with metric tensor $\{h_{mn} \mid m, n = 1, 2, \ldots, D\}$ of the underlying Lie algebra in the following manner:

$$\begin{aligned}
&h_{mn} = \sum_q \sum_l f^l_{mq} f^q_{nl} = h_{nm}, \\
&g_{mn} = h_{mn} - \sum_\lambda \sum_\mu F^\mu_{m\lambda} F^\lambda_{n\mu} = g_{nm}, \\
&g_{\mu\lambda} = \sum_\sigma \sum_m (F^\sigma_{m\mu} S^m_{\lambda\sigma} - F^\sigma_{m\lambda} S^m_{\mu\sigma}) = -g_{\lambda\mu}, \\
&g_{m\mu} = g_{\mu m} = 0, \\
&m, n = 1, 2, \ldots, D; \ \lambda, \mu = D+1, \ldots, D+d.
\end{aligned} \tag{25}$$

Since in our present case of GSO(2, 1) defined by Eqs. (21)—(24), $D = 3$ and $d = 2$, g is a 5×5 matrix and h is a 3×3 matrix. A graded Lie algebra has been called semisimple[6] if

$$\det \| h \| \neq 0, \quad \det \| g \| \neq 0. \tag{26}$$

Then according to what is called the C-theorem in Ref. 6 for each semisimple graded Lie algebra obeying (26) there exists necessarily an antisymmetric C such that

$$\begin{aligned}
&S^m = CF_m, \\
&m = 1, 2, \ldots, D.
\end{aligned} \tag{27}$$

From Eqs. (23) and (24) it is easily seen that there does not exist any such matrix C in the present case of the given GSO(2, 1). Hence the above GSO(2, 1) is non-semisimple. One can also see this directly by computing the determinants $\| g \|$ and $\| h \|$ using Eqs. (22)—(25) and noting that $\det \| g \| = 0$ contrary to the required condition (26). Thus in view of the representations of $\{\beta\}$ considered in previous sections we find that we have here examples of some types of boson representations of a nonsemisimple GSO(2, 1).

In conclusion it can be recalled that when the order of the statistics of the para-Bose system reaches infinity we arrive at a Fermi system.[7,8] Hence it would be interesting to speculate on the fact that one could arrive at a Fermi field by utilizing a single Bose field to create para-Bose fields as above and allowing the order of the statistics tend to infinity. A recent review of other so-called boson representations of fermions is found in Ref. 9.

ACKNOWLEDGMENTS

The authors are grateful to Professor Alladi Ramakrishnan for kind encouragement. They also like to thank Professor T.S. Santhanam for useful discussions.

[1]R. Jagannathan and R. Vasudevan, J. Math. Phys. **19**, 1493 (1978).
[2]A. Ramakrishnan, R. Vasudevan, and P.S. Chandrasekharan, J. Math. Anal. Appl. **35**, 249 (1971).
[3]See A.J. Kalnay, Prog. Theor. Phys. **54**, 1848 (1975).
[4]H.S. Green, Phys. Rev. **90**, 270 (1953).
[5]T.F. Jordan, N. Mukunda, and S.V. Pepper, J. Math. Phys. **4**, 1089 (1963).
[6]A. Pais and V. Rittenberg, J. Math. Phys. **16**, 2062 (1975).
[7]O.W. Greenberg and A.M.L. Messiah, J. Math. Phys. **6**, 500 (1965).
[8]S. Kamefuchi, "Lectures on Parastatistics", Matsci. Rep. 24 (1965). (Published by Inst. Math. Sci. Madras, India.)
[9]P. Garbaczewski, Preprint IC/77/51 of the International Centre for Theoretical Physics, Trieste, Italy (1977) (to appear in Phys. Rep. C).

Computation of a class of 3*j*-coefficients

Martin L. Sage

Chemistry Department, Syracuse University, Syracuse, New York 13210

Kameshwar C. Wali

Physics Department, Syracuse University, Syracuse, New York 13210
(Received 15 May 1978)

We give an analytical procedure to calculate 3*j*-coefficients $\begin{pmatrix} l & l & 2j \\ m & -m & 0 \end{pmatrix}$ for arbitrary integer l.

I. INTRODUCTION

The purpose of this note is to give a procedure to evaluate the 3-*j* coefficients $\begin{pmatrix} l & l & 2j \\ m & -m & 0 \end{pmatrix}$ for *arbitrary* integer l. j is also an integer such that $(l, l, 2j)$ satisfy the standard triangular equality. The tables[1] for 3-*j* coefficients give values only up to $l=8$, $j=4$. Although the standard recursion formula can be used to evaluate the above mentioned 3-*j* symbol in a specific case, the technique for evaluation we shall present is often more convenient in both analytical work as well as for practical computations.

In the next section we discuss the basis of the general method. In Sec. III, we give the necessary formulas. Finally, in the Appendix, we discuss the results for $l=4$, which is sufficiently nontrivial to illustrate the power of the method.

II. THE GENERAL METHOD

We begin by considering the product $X_{l,j}(m)$ given by

$$X_{l,j}(m)=\begin{pmatrix} l & l & 2j \\ m & -m & 0 \end{pmatrix}\begin{pmatrix} l & l & 2j \\ 0 & 0 & 0 \end{pmatrix}, \tag{2.1}$$

where the 3-*j* coefficient $\begin{pmatrix} l & l & 2j \\ 0 & 0 & 0 \end{pmatrix}$ can be evaluated by elementary methods,[2]

$$\begin{pmatrix} l & l & 2j \\ 0 & 0 & 0 \end{pmatrix}=(-1)^{l+j}\left(\frac{(2l-2j)!}{(2l+2j+1)!}\right)^{1/2}\frac{2j!}{(j!)^2}\frac{(l+j)!}{(l-j)!}. \tag{2.2}$$

Using the relationship between the spherical harmonics and 3-*j* coefficients, we find

$$X_{l,j}(m)=\frac{(-1)^m}{(2l+1)}\int d\Omega\, Y^*_{lm}(\theta,\phi)P_{2j}(\cos\theta)\,Y_{lm}(\theta,\phi). \tag{2.3}$$

Now, we define

$$X_{l,j}(m,\alpha)=\frac{(-1)^m}{(2l+1)}\int d\Omega\, Y^*_{lm}(\theta,\phi)P_{2j}(\cos\theta)\times Y_{lm}(\theta,\phi-\alpha), \tag{2.4}$$

$$X_{l,j}(m,\alpha)=\exp(-m\alpha)X_{l,j}(m), \tag{2.5}$$

and

$$\frac{\partial}{\partial\alpha}[X_{l,j}(m,\alpha)]=-imX_{l,j}(m,\alpha). \tag{2.6}$$

From the addition theorem for the spherical harmonics[2]

$$\frac{4\pi}{(2l+1)}\sum_m Y^*_{lm}(\theta,\phi)\,Y_{lm}(\theta,\phi-\alpha)=P_l(\cos\omega),$$

where

$$\cos\omega=\cos^2\theta+\sin^2\theta\cos\alpha. \tag{2.7}$$

We then have

$$\sum_{m=-l}^{l}(-1)^mX_{l,j}(m,\alpha)=\frac{1}{4\pi}\int d\Omega\,P_l(\cos\omega)P_{2j}(\cos\theta)\equiv I_{lj}(\alpha). \tag{2.8}$$

By differentiating the equation (2.8) with respect to α $0, 2, 4, \ldots, 2(l-1)$ times and then setting $\alpha=0$, we obtain l linearly independent equations for $X_{l,j}(m)$. The corresponding equations involving the odd derivatives vanish since $X_{l,j}(-m)=X_{l,j}(m)$. We can also easily verify that equations involving derivatives of higher order than $2(l-1)$ do not yield new equations.

The linearly independent equations can be written in the compact matrix form

$$\Delta X=\tfrac{1}{2}V, \tag{2.9}$$

where X and V are $l\times1$ column matrices whose nth elements are given by

$$X_n=(-1)^{n+1}X_{l,j}(n), \tag{2.10}$$

$$V_n=\delta_{n1}X_{l,j}(0)+(-1)^n\left(\frac{\partial^{2(n-1)}}{\partial\alpha^{2(n-1)}}\;I_{l,j}(\alpha)\right)_{\alpha=0} \tag{2.11}$$

and Δ is a $l\times l$ matrix with

$$\Delta_{kn}=(n)^{2(k-1)},\quad k,n=1,2,\ldots,l. \tag{2.12}$$

The solutions for the required 3-*j* coefficients are contained in the equations

$$X=\tfrac{1}{2}\Delta^{-1}V. \tag{2.13}$$

III. INVERSION OF Δ AND THE EVALUATION OF V

The matrix Δ is a special case of the Vandermode matrix A defined by

$$A_{kn}=x_n^{k-1}, \tag{3.1}$$

where x_i are arbitrary.

If we identify

$$x_n=n^2,\qquad n=1,\ldots,l \tag{3.2}$$

we have

$$A=\Delta. \tag{3.3}$$

 0022-2488/79/030393-03$01.00

Since det$A$ is a homogeneous, antisymmetric polynomial of order $l(l-1)/2$ in the x_i's, we can easily verify that

$$\det A = \prod_{\substack{k,n \\ k>n}} (x_k - x_n). \tag{3.4}$$

The cofactor matrix $C_{nk}(A)$ of A is given by

$$C_{nk}(A) = (-1)^{n+k} \prod_{\substack{i,j \\ l \geq i \geq j \geq 1 \\ i,j \neq k}} (x_i - x_j) \times \sum_{\substack{\{\nu\} \\ l \geq \nu_1 \geq \nu_2 \cdots \geq \nu_{l-j} \geq 1 \\ i \notin \{\nu\}}} x_{\nu_1} x_{\nu_2} \cdots x_{\nu_{l-n}}, \tag{3.5}$$

where the summation in (3.5) is a simple homogeneous polynomial of order $(l-n)$ in the x's that has no multiple factors and does not contain x_i. If we denote these polynomials by $S_{l-n,k}$,

$$S_{l-n,k} = \sum_{\substack{\{\nu\} \\ l \geq \nu_1 \geq \nu_2 \cdots \geq \nu_{l-j} \geq 1 \\ k \notin \{\nu\}}} x_{\nu_1} x_{\nu_2} \cdots x_{\nu_{l-n}}, \tag{3.6}$$

the inverse of A or Δ is given by

$$(\Delta^{-1})_{kn} = (A^{-1})_{kn} = (-1)^{k+n} \left[\prod_{\substack{i,j \\ l \geq i \geq k \geq j \geq 1}} (x_i - x_k)^{-1} (x_k - x_j)^{-1} \right] S_{l-n,k}. \tag{3.7}$$

The polynomials $S_{l-n,k}$ can be obtained from the symmetric polynomials $S^{(m)}$

$$S^{(m)} = \sum_{i=1}^{l} (x_i^m) \tag{3.8}$$

by noting that

$$S_{l-n} = (l-n)^{-1} \sum_{i=1}^{l-n} (-1)^{i+1} S^{(i)} S_{l-n-i},$$

where S_{l-n} are homogeneous polynomials of order $l-n$ like the $S_{l-n,k}$ but containing x_k, and then

$$S_{l-n,k} = \sum_{i=0}^{l-n} (-1)^i x_k^i S_{l-n-i}. \tag{3.9}$$

Further for $x_n^m = (n^2)^m$, the symmetric polynomials $S^{(m)}$ are given by the Bernouilli numbers. The product term in (3.7) which depends on k is given by

$$\prod_{\substack{i,j \\ l \geq i \geq k \geq j \geq 1}} (x_i - x_k)^{-1} (x_k - x_j)^{-1} = \frac{2k^2}{(l+k)!(l-k)!}. \tag{3.10}$$

In order to evaluate V, we need to evaluate the derivatives

$$\frac{\partial^{2k}}{\partial \alpha^{2k}} P_l(\cos\omega)\Big|_{\alpha=0}. \quad \text{Letting } z = \cos\omega,$$

$$\frac{\partial^{2m}}{\partial \alpha^{2m}} P_l(z)\Big|_{\alpha=0} = \sum_{\{\mu\}} \frac{(2m)!}{\mu_1!\mu_2!\cdots\mu_q!} \frac{d^s P_l(z)}{dz^s} \times \left(\frac{1}{2!}\frac{\partial^2 z}{\partial \alpha^2}\right)^{\mu_1} \left(\frac{1}{4!}\frac{\partial^4 z}{\partial \alpha^4}\right)^{\mu_2} \cdots \times \left(\frac{1}{(2q)!}\frac{\partial^2 q_z}{\partial \alpha^{2q}}\right)^{\mu_q}\Bigg|_{\alpha=0}, \tag{3.11}$$

where the summation $\{\mu\}$ extends over all nonnegative integers μ_j such that $\sum j\mu_j = m$ and $s = \sum \mu_j \leq l$. In (3.11) we have retained only the even derivatives of z with respect to α, since the odd derivatives vanish when $\alpha = 0$. Substituting

$$\frac{\partial^{2r} z}{\partial \alpha^{2r}}\Big|_{\alpha=0} = (-1)^r \sin^2\theta \tag{3.12}$$

and

$$\frac{d^s P_l(z)}{dz^s}\Big|_{z=1} = \frac{(l+s)!}{2^s (l-s)!\, s!} \tag{3.13}$$

in (3.11), we find

$$\frac{\partial^{2m}}{\partial \alpha^{2m}} P_l(z)\Big|_{\alpha=0} = \sum_{\{\mu\}} \frac{(-1)^m (2m)!}{\mu_1!\mu_2!\cdots\mu_q!} \frac{(l+s)!}{2^s (l-s)!\, s!} \sin^{2s}\theta \times [2!]^{-\mu_1} [4!]^{-\mu_2} \cdots [(2q)!]^{-\mu_q}. \tag{3.14}$$

When we use (3.13) to evaluate $\{[\partial^{2(n-1)}/\partial\alpha^{2(n-1)}] I_{l,j}(\alpha)\}_{\alpha=0}$ occurring in (2.11), we obtain terms involving

$$\int \sin^{2s}\theta\, P_{2j}(\cos\theta)\, d\Omega \equiv 2\pi J_{2s,2j}. \tag{3.15}$$

The integrals $J_{2s,2j}$ are nonvanishing only when $s \geq j$ and they can be evaluated as follows,

$$J_{2s,2j} = \sum_{t=0}^{s} \frac{(-1)^t s!}{t!(s-t)!} I_{2t,2j}, \tag{3.16}$$

where $I_{2t,2j}$ are the well known integrals[3]

$$I_{2t,2j} = \int_{-1}^{1} x^{2t} p_{2j}(z)\, dz = \frac{2^{2j+1}(2t)!(t+j)!}{(t-j)!(2t+2j+1)!}. \tag{3.17}$$

Thus, finally we obtain

$$V_n = \delta_{n1} X_{l,j}(0) - \frac{(2n-2)!}{2} \sum_{\{\mu\}} \frac{[2!]^{-\mu_1}[4!]^{-\mu_2}\cdots[(2q)!]^{-\mu_q}}{\mu_1!\mu_2!\cdots\mu_q!} \times \frac{(l+s)!}{2^s (l-s)!\, s!} J_{2s,2j}. \tag{3.18}$$

From (3.17) and (3.18), it is straightforward to calculate X and hence the required 3-j coefficients $\left(\begin{smallmatrix} l & l & 2j \\ m & -m & 0 \end{smallmatrix}\right)$.

ACKNOWLEDGMENTS

Work is supported in part by the U.S. Department of Energy. One of the authors would like to thank Mr. G. Trahern for checking the numerical calculation in the Appendix.

APPENDIX

We consider the special case $l=4$ and compute all the 3-j coefficients. In this case $j=0,1,2,3,4$. From (3.7) and (3.18),

TABLE I. Values for $J_{2s,2j}$.

s \ j	0	1	2	3	4
0	2	0	0	0	0
1	$\frac{4}{3}$	$-\frac{4}{15}$	0	0	0
2	$\frac{16}{5}$	$-\frac{32}{105}$	$\frac{16}{315}$	0	0
3	$\frac{32}{35}$	$-\frac{32}{105}$	$\frac{32}{385}$	$-\frac{32}{3003}$	0

TABLE II.

m \ j	0	1	2	3	4
1	$-\frac{1}{3}$	$-\frac{17}{30}\sqrt{5/77}$	$-\frac{3}{2}\sqrt{2/1001}$	$\frac{1}{30}\sqrt{5/143}$	$\frac{28}{15}\sqrt{10/2431}$
2	$\frac{1}{3}$	$\frac{4}{15}\sqrt{5/77}$	$-\frac{11}{6}\sqrt{2/1001}$	$-\frac{11}{15}\sqrt{5/143}$	$\frac{14}{15}\sqrt{10/2431}$
3	$-\frac{1}{3}$	$\frac{7}{30}\sqrt{5/77}$	$\frac{7}{2}\sqrt{2/1001}$	$-\frac{17}{30}\sqrt{5/143}$	$\frac{4}{15}\sqrt{10/2431}$
4	$\frac{1}{3}$	$-\frac{14}{15}\sqrt{5/77}$	$\frac{7}{3}\sqrt{2/1001}$	$-\frac{2}{15}\sqrt{5/143}$	$\frac{1}{30}\sqrt{10/2431}$

$$\Delta^{-1}=\begin{bmatrix} \frac{8}{5} & -\frac{61}{90} & \frac{29}{360} & -\frac{1}{360} \\ -\frac{4}{5} & \frac{169}{180} & -\frac{13}{90} & \frac{1}{180} \\ \frac{8}{35} & -\frac{3}{10} & \frac{3}{40} & -\frac{1}{280} \\ -\frac{1}{35} & \frac{7}{180} & -\frac{1}{90} & \frac{1}{1260} \end{bmatrix}, \tag{A1}$$

$$V_n=\delta_{n1}X_{4,j}(0)-\tfrac{1}{2}(2n-2)!\sum_{(\mu)}\frac{[2!]^{-\mu_1}[4!]^{-\mu_2}\cdots[(2q)!]^{-\mu_q}}{\mu_1!\mu_2!\cdots\mu_q!}$$

$$\times\frac{(4+s)!}{2^s(4-s)!s!}\,j_{2s,2j} \tag{A2}$$

and hence

$$V=\begin{bmatrix} X_{4,j}-\frac{1}{2}J_{0,2j} \\ -5J_{2,2j} \\ -\frac{135}{2}J_{4,2j}-5J_{2,2j} \\ -\frac{1575}{2}J_{6,2j}-\frac{675}{2}J_{4,2j}-5J_{2,2j} \end{bmatrix}. \tag{A3}$$

The required $J_{2s,2j}$ are easily computed from (3.16), and (3.17). These are given in Table I.

Finally using

$$\begin{matrix} & j & 0 & 1 & 2 & 3 & 4 \end{matrix}$$

$$\begin{bmatrix} 4 & 4 & 2j \\ 0 & 0 & 0 \end{bmatrix}=\tfrac{1}{3}\ \tfrac{2}{3}\sqrt{5/77}\ \tfrac{3}{1}\sqrt{2/1001}\ \tfrac{2}{3}\sqrt{5/143}\ \tfrac{7}{3}\sqrt{10/2431}$$

and the equations (2.9) and (2.10), we find the values for $\left(\begin{smallmatrix} 4 & 4 & 2j \\ m & -m & 0 \end{smallmatrix}\right)$ which are given in Table II.

[1]M. Rotenberg, R. Bivins, N. Metropolis, and John K. Wooten, Jr., *The 3-j and 6-j Symbols* (Technology, Cambridge, Massachusetts, 1959).

[2]See for instance, A. Edmonds, *Angular Momentum in Quantum Mechanics* (Princeton U. P., Princeton, New Jersey).

[3]E. Whittaker and G. Watson, *A Course in Modern Analysis* (Macmillan, New York, 1945), Am. ed. Cambridge U. P., p. 310.

Semiclassical perturbation expansion of the multichannel scattering matrix

S. Bosanac[a)]

The Hebrew University of Jerusalem, Physical Chemistry, Jerusalem, Israel
(Received 20 February 1978; revised manuscript received 11 May 1978)

A semiclassical perturbation method for the inelastic S matrix is described. The channels are transformed into a set of the eigenstates and it is assumed that the transition between them is small. The perturbation is the matrix which diagonalizes the coupling matrix. It is shown that such a series is independent of $\hbar$; hence the limit $\hbar \to 0$ of the S matrix can be calculated. A special case of the weak coupling is also discussed.

INTRODUCTION

The semiclassical limit of the elastic S matrix is obtained by expanding the solution of the Schrödinger equation in a power series of the Planck constant.[1,2] It was therefore natural to generalize the idea to more than two particles[3] and to the set of multichannel equations.[4-6] However, the method becomes combersome in such applications. For example, in the simplest case of only two coupled equations (two state approximation) Stückelberg[4] obtained the semiclassical limit of the S matrix after making several assumptions on the potential. With such assumptions the system of two coupled second order differential equations can be transformed into a single equation of the fourth order. The semiclassical S matrix is then obtained from the asymptotic solution of this equation. However, the difficult part in such a procedure is tracing the asymptotic solution across the Stokes lines. It is obvious that if we have three or more channels, the analysis is even more difficult since it would involve, under the most favorable conditions on the potential, the equations of the sixth or higher orders.

To overcome the difficulty, different models were proposed all having one underlying idea in common: Instead of solving a set of the second order equations, a physical model of collision is developed which is described by the equations of the first order. For a two state problem, the system of equations can then be transformed into a single equation of the second order for which the theory of asymptotic solution is well developed. For example, Landau[7] and Zener[8] assumed that in the inelastic collision of atoms the motion of nucleus is classical while at each instant of time the electrons form an eigenstate of the time independent electronic Hamiltonian (adiabatic approximation). The trajectory of the nucleus is linear in the region of the largest transition probability (near the crossing point). From the time dependent nonrelativistic Schrödinger equation they obtained a set of the first order equations in the time variable, the solution of which gives the transition probabilities. Since then there has been different derivations of the basic equations, briefly summarized by Delos *et al.*[9] Among them is also the momentum space formulation of the inelastic collisions.[10]

We can make two criticisms of this approach:

(a) For a two channel problem the method gives the semiclassical limit of the S matrix from the set of the first order equations. However, the sequence of approximations leading to the equations is largely based on the physical intuition. In other words, by making approximations it is difficult to estimate the contribution of the neglected parts and subsequently making this contribution smaller in a systematic manner. (For comparison we recall that in the ordinary WKB method the index of quality[1] provides the estimate of accuracy of the approximation.) This is essential when a non-typical transition occurs, e.g., transition in the case of no crossing.

(b) Generalization to more than two channels is not straightforward. A typical example of a many channel problem is atom-rigid rotor or atom—diatom reactive collision.

In this paper we describe another approach for obtaining the semiclassical solution of the multichannel equations. Let us first look at the simplest way to solve the multichannel equations: the distorted wave expansion. The coupled set of differential equations is transformed into an integral equation which can be solved by iteration[11] and the series is absolutely convergent. However, such a series is of little use in the semiclassical limit since, as it is shown in Sec. 1, it is essentially a power series in $\hbar^{-1}$. (It should be emphasized that the coefficients of such a series are also functions of $\hbar$ but they are finite in the limit $\hbar \to 0$.) Since the distorted wave expansion is also a power series in the coupling constant, the rate of convergency is essentially determined by the relative magnitude of $\hbar$ and the off-diagonal elements of potential matrix, but in general, because of such form of expansion, the first few terms of the series are not accurate representation of the solution of multichannel equations.

The idea is now to diagonalize the coupling matrix and use the diagonalization matrix to transform the wavefunction into a new function, which we call the eigenchannel wavefunctions. Since the diagonalization matrix is a function of the coordinate, the eigenchannel wavefunction satisfies a differential equation which in-

[a)]On leave of absence from R. Boskovic Institute, Zagreb, Croatia, Yugoslavia.

 0022-2488/79/030396-10$01.00

volves explicitly the diagonalization matrix and its derivatives. As shown in Sec. 1, we can solve the new equations by iteration of the appropriate integral equation. These are two essential features of the new series: (a) We obtain a series in which the individual terms are finite in the limit $\hbar \to 0$; (b) the coefficients of the series are entirely determined by the diagonalization matrix. The last feature is important since it implies that the coupling matrix is no longer explicitly present in the solution of the multichannel equations but is represented through the matrix which diagonalized it. For this reason we prove in Sec. 3 that the diagonalization matrix is uniquely determined by the coupling matrix.

Since the coefficients of the series are independent of $\hbar$, we can obtain the S matrix in the limit $\hbar \to 0$ (the semiclassical limit). In Sec. 2 we discuss this limit for the first iteration of the integral equation. An explicit form of the semiclassical S matrix is obtained in which a new matrix η, defined in Sec. 1, plays the role of the coupling matrix. The matrix η is related to the coupling matrix through the diagonalization matrix, as shown in Sec. 3, but very often it can be calculated by the perturbation method directly from the potential matrix. In Sec. 5 we show how this is done for the case of degeneracy in two channels with the other channels being well separated.

It is essential now to point out that in the derivation of the semiclassical S matrix we did not make any *a priori* assumption on the coupling matrix and the number of channels. We have only restricted the limit to the first Born approximation; therefore, we can systematically improve the result by calculating higher order Born terms knowing that all of them are finite in the limit $\hbar \to 0$.

By this method we calculate in Sec. 4 the S matrix for a two channel case. Besides taking only the first Born approximation of the eigenchannel wavefunctions, we also make some approximations to the η matrix which do not essentially alter the final result for the S matrix. It is shown that the diagonal matrix elements of η are exactly zero while the off diagonal are negligible except in a small vicinity of the crossing point. Because of such behavior of η, the integrals are not difficult to calculate. We find a general agreement with the usual LSZ formula; however, even on the level of the first Born approximation several new results are obtained: (a) Corrections due to the repulsion of the potential curves is taken into account; (b) reflection amplitude due to the change in slope near the crossing point is predicted (this only effects the elastic elements of the S matrix); (c) change in the normalization factor of the LSZ inelastic amplitude, i.e., in the extreme semiclassical limit the factor 2 which appears in the LSZ formula for S_{12} is replaced by $2\pi/3$.

The method described in this paper is essentially a perturbation method and is analogous to the distorted wave expansion. As such it has some drawbacks; it fails for large transition amplitudes, but it also fails in the case of weak coupling. This case is discussed in Sec. 6, and it is shown that the two methods are complementary. A special case of the "forbidden" transitions is discussed, and it is shown that the transition amplitude can be calculated from the first order Born approximation.

Solving the system of multichannel equations by diagonalizing the coupling matrix has been previously used for the systems where the coupling matrix is constant, e.g., for the square well potentials,[12] or where the diagonalization matrix is constant, e.g., spin-$\frac{1}{2}$–spin-$\frac{1}{2}$ coupling.[13] The method was also used, under the name of "the sudden approximation," as a model for describing the atom-rigid rotor collision.[14] In this model, the spacing between the energy levels of the rigid rotor is neglected and the centrifugal term is assumed to be degenerate in all channels. The diagonalization matrix is constant in such a case. Therefore, developing a method which is based on a coordinate dependent diagonalization matrix can be regarded as a natural extension of the just described applications. However, the new result is that the perturbation series obtained by solving the set of equations for the eigenchannel wavefunction is semiclassical, i.e., the coefficients of the series are finite in the limit $\hbar \to 0$.

1. THE THEORY

A system of n multichannel equations is in the matrix notation

$$\psi'' = p\psi, \tag{1.1}$$

where

$$p = V + \frac{l(l+1)}{r^2} - k^2 \tag{1.2}$$

The channel energy matrix k^2 and the centrifugal matrix are diagonal while the potential V is nondiagonal. In many cases the off-diagonal elements of V are small compared to the diagonal ones, i.e.,

$$p_{i,j} \ll p_{m,m}; \tag{1.3}$$

therefore, we can use a perturbation method for solving the system (1.1). By defining the perturbation

$$p = p_0 + \epsilon p_1, \tag{1.4}$$

where p_0 is diagonal, the solution of (1.1) can be obtained from the integral equation,[11]

$$\psi = \psi_0 + \frac{\epsilon}{2i} k^{-1} \int_0^r K(r, r') p_1(r') \psi(r')\, dr' \tag{1.5}$$

being a power series in ϵ. From the series (1.5) we calculate the Jost functions, and their ratio gives the S matrix.[11] The Jost functions and the S matrix are given as a power series in ϵ.

There is an inherent weakness of the power expansion of the type (1.5). By putting $\hbar$ dependence of p into ϵ we notice that (1.5) is a power series of the form

$$\psi = a_0 + \hbar^{-1} a_1 + \hbar^{-2} a_2 + \cdots, \tag{1.6}$$

and in the semiclassical limit $\hbar \to 0$ the leading coefficients in (1.6) are very large; therefore, the series for ψ (hence for the Jost functions and the S matrix) is poorly convergent. Approximation of the S matrix with only the first few terms from (1.6), therefore, has no meaning.

Another approach for solving (1.1) is proposed here. It is based on the perturbation expansion but with the coefficients independent of $\hbar$. To obtain such a series, we define a matrix A which diagonalizes p

$$A^{-1}(r)\,p(r)A(r)=\lambda(r), \tag{1.7}$$

where λ is a diagonal matrix. Let us define

$$\varphi=A^{-1}\psi A \tag{1.8}$$

and find the equation which φ satisfies. By taking derivatives of φ we obtain

$$\begin{aligned}\varphi''&=\lambda\varphi+2[\varphi', A^{-1}A']+[\varphi, A^{-1}A'']\\&\quad+2[A^{-1}A', \varphi A^{-1}A']\equiv\lambda\varphi+\epsilon U\end{aligned} \tag{1.9}$$

where

$$[X, Y]=XY-YX \tag{1.10}$$

If we define a new matrix η by

$$A'=A\eta, \tag{1.11}$$

the system (1.9) simplifies to

$$\varphi''=\lambda\varphi+2[\varphi', \eta]+[\varphi, \eta^2]+[\varphi, \eta']+2[\eta, \varphi\eta]\equiv\lambda\varphi+\epsilon U. \tag{1.12}$$

The matrix A is independent of $\hbar$. We also notice that $\lambda\sim\hbar^{-2}$ and $\varphi'\sim\hbar^{-1}$; hence, ϵU, being of the order $\hbar^{-1}$, can be treated as a perturbation. In an analogy to (1.5) we obtain the integral equation for φ,

$$\varphi=\varphi_0+\frac{\epsilon}{2i}k^{-1}\int_0^r K(r, r')\,U(r')\,dr', \tag{1.13}$$

where

$$K(r, r')=f_0^-(r)f_0^+(r')-f_0^-(r')f_0^+(r). \tag{1.14}$$

The functions in (1.13) and (1.14), with an index zero, are the solutions of Eq. (1.9) with $\epsilon=0$, i.e.,

$$\varphi_0''=\lambda\varphi_0, \tag{1.15}$$

and they define the eigenstates of the multichannel problem. The matrix η, defined by (1.11), can be regarded as the coupling between such eigenstates.

The unperturbed solutions in (1.13) and (1.14) are different according to their boundary conditions

$$\begin{aligned}&\varphi_0(r)\to 0, \quad r\to 0,\\&f_0^{\pm}(r)\to\exp(\mp ikr), \quad r\to\infty,\end{aligned} \tag{1.16}$$

where k is a diagonal matrix of the channel wavenumbers. It is now straightforward to show that (1.13) is a power series with the coefficients independent of $\hbar$, i.e.,

$$\varphi=a_0+(\epsilon\hbar)a_1+(\epsilon\hbar)^2a_2+\cdots; \tag{1.17}$$

therefore the series is more suitable for the semiclassical analysis of the S matrix than (1.6) and is converging much faster. The rate of convergency is in fact entirely determined by η and its derivatives.

The Jost functions are determined from

$$\psi=\exp(ikr)\,j^+ + \exp(-ikr)\,j^-, \quad r\to\infty, \tag{1.18}$$

and if we assume that $k_i\neq k_j$, $i\neq j$ for all i and j, then we can find A with the property

$$\lim_{r\to\infty}A(r)=I, \tag{1.19}$$

and we have

$$\varphi=f^-j^+ + f^+j^-\sim f_0^-\,j^+ + f_0^+\,j^-. \tag{1.20}$$

By replacing φ in (1.13) by (1.20) we find

$$j^{\pm}=j_0^{\pm}\mp\frac{\epsilon}{2i}\,k^{-1}\int_0^{\infty} f_0^{\pm}(r)U(r)\,dr \tag{1.21}$$

and again this is a power series with the coefficients independent of $\hbar$.

The S matrix, defined by

$$S=\exp[i\pi(l+1)/2]\,k^{1/2}\,j^+(j^-)^{-1}k^{-1/2}\exp[i\pi(l+1)/2], \tag{1.22}$$

can now be calculated using (1.21).

In the derivation of the S matrix (1.20) we have assumed (1.19). As it will be shown in Sec. 3, the matrix elements of A, which connect the channels with degeneracy, are constant in the limit $r\to\infty$; hence

$$\lim_{r\to\infty}A(r)=C, \tag{1.23}$$

where C is some constant matrix. The asymptotic wavefunction ψ is then related to the eigenchannel wavefunction φ by

$$\lim_{r\to\infty}\psi=C\lim_{r\to\infty}\varphi C^{-1} \tag{1.24}$$

from which we obtain the Jost functions

$$j^{\pm}=\exp(\mp ikr)C\exp(\pm ikr)\,j_c^{\pm}C^{-1} \tag{1.25}$$

where $j_c^{\pm}$ are given by (1.21). The S matrix is now

$$S=\exp[i\pi(l+1)/2]k^{1/2}Cj_c^+(j_c^-)^{-1}C^{-1}k^{-1/2}\exp[i\pi(l+1)/2], \tag{1.26}$$

where we have formally replaced

$$\exp(ikr)\to I. \tag{1.27}$$

For simplicity, we assume from now that $C=I$.

2. DISCUSSION OF THE EXPANSION SERIES

As we have shown in the previous section, the wavefunction is given as a power series in $\epsilon\hbar$; therefore, the S matrix (1.22) can be estimated in the limit $\hbar\to 0$.

Let us first notice that in the limit $\hbar\to 0$, the function U simplifies. We notice from (1.9) that ϵU contains terms with φ and φ'. Since $\varphi\sim\hbar^0$ and $\varphi'\sim\hbar^{-1}$, we have

$$\epsilon U\sim 2[\varphi', \eta], \quad \hbar\to 0 \tag{2.1}$$

and the Jost functions are

$$j^{\pm}=j_0^{\pm}\pm\frac{1}{i}\,k^{-1}\int_0^{\infty} f_0^{\pm}(r)[\eta, \varphi']\,dr. \tag{2.2}$$

The first integral is finite; however, the second needs little attention. For $r\to 0$, the function $f_{0m}^{\pm}$ diverges as r^{-l_m} and, for the off-diagonal element, the integrand is

$$f_{0m}^{\pm}\varphi'_{0n}\sim r^{l_n-l_m}, \quad r\to 0. \tag{2.3}$$

To offset the infinity, we assume

$$\eta_{m,n}\sim r^{|l_m-l_n|}, \quad r\to 0, \tag{2.4}$$

which ensures the finite value of (2.2).

Taking the limit $\hbar \to 0$ of the S matrix can be much simplified in the first Born approximation of (1.13). This can be done without loss of generality since A is independent of $\hbar$. In the Jost function (2.2) we formally replace φ by φ_0, and, because ϵU is small, the inverse of j^- is approximately

$$(j^-)^{-1} \sim \{I - i(kj_0^-)^{-1} \int_0^\infty f_0^-[\eta, \varphi_0'] \, dr\}(j_0^-)^{-1}. \tag{2.5}$$

The S matrix (1.20) is now up to the second order in ϵ

$$S \sim \exp[i\pi(l+1)/2] k^{1/2} j_0^+ \{I - i(kj_0^+ j_0^-)^{-1} \int_0^\infty \varphi_0[\eta, \varphi_0'] \, dr\} \times (j_0^-)^{-1} k^{-1/2} \exp[i\pi(l+1)/2] \tag{2.6}$$

The first integral can be neglected because

$$k^{-1} \int_0^\infty \varphi_0 \varphi_0' \eta \, dr = \tfrac{1}{2} k^{-1} \varphi_0^2 \eta \Big|_0^\infty + O(\hbar) = O(\hbar), \tag{2.7}$$

where we have used (1.19), and the S matrix is

$$S \sim \exp[i\pi(l+1)/2] k^{1/2} j_0^+ [I - i(kj_0^+ j_0^-)^{-1} \int_0^\infty \varphi_0 \eta \varphi_0' \, dr] \times (j_0^-)^{-1} k^{-1/2} \exp[i\pi(l+1)/2]. \tag{2.8}$$

The solution of (1.15) is arbitrary up to a constant matrix so that $j_0^\pm$ can always be defined with modulus 1. In that case

$$S \sim S_0^{1/2} k^{1/2} (I - ik^{-1} \int_0^\infty \varphi_0 \eta \varphi_0' \, dr) k^{-1/2} S_0^{1/2}, \tag{2.9}$$

where S_0 is the S matrix for the eigenstates (1.15).

The last result is the semiclassical approximation of the S matrix in the first Born approximation. However, the equation needs further discussion.

Let us first show that in the semiclassical limit the diagonal elements of the integral in (2.9) goes to zero as $\hbar$. We first replace $\eta \varphi_0'$ by

$$-2\eta\varphi_0' \to 2[\varphi_0', \eta] + [\varphi_0, \eta^2] + [\varphi_0, \eta'] + 2[\eta, \varphi_0 \eta], \tag{2.10}$$

In other words, instead of taking the limit (2.1), we derive (2.9) in the most general form. Since φ_0 is diagonal, it is easy to show that the diagonal elements of (2.10) are all zero, except the last one. We have

$$\epsilon U_{n,n} = 2 \sum_j (\varphi_j^0 - \varphi_n^0) \eta_{nj} \eta_{jn} \tag{2.11}$$

and the diagonal elements of the S matrix are

$$S_{n,n} \sim S_n^0 [1 + ik^{-1} \sum_j \int_0^\infty \varphi_n^0 (\varphi_j^0 - \varphi_n^0) \eta_{nj} \eta_{jn} \, dr]. \tag{2.12}$$

The second term goes to zero as $\hbar$ because η_{nj} are independent of $\hbar$.

The S matrix (2.9) is not unitary. It also appears that it is not symmetric, since the integral in (2.9) contains a derivative of only one φ_0. However, as it will be shown in the next section, the matrix η is arbitrary up to n constants, where n is the number of channels. From this fact we show that, in the semiclassical limit, (2.9) is symmetric.

Any off-diagonal element of the integral in (2.9) is

$$M_{lm} = -k_l^{-1} \int_0^\infty \varphi_l^0 \eta_{lm} \varphi_m^{0'} \, dr. \tag{2.13}$$

By partially integrating (2.13) we obtain

$$M_{lm} = -k_l^{-1} \varphi_l^0 \eta_{lm} \varphi_m^0 \Big|_0^\infty + k_l^{-1} \int_0^\infty (\varphi_l^{0'} \eta_{lm} \varphi_m^0 + \varphi_l^0 \eta_{lm}' \varphi_m^0) \, dr. \tag{2.14}$$

The first term is zero since $A' \to 0$ for $r \to \infty$ and $\varphi_l^0 \to 0$ for $r \to 0$. The last term does not involve derivatives of φ^0 and is of the order $\hbar$. We have approximately

$$M_{lm} \sim k_l^{-1} \int_0^\infty \varphi_l^{0'} \eta_{lm} \varphi_m^0 \, dr. \tag{2.15}$$

Since η is arbitrary, we can always define it with the property (as will be shown in the next section)

$$\eta_{lm} = -\frac{k_l}{k_m} \eta_{ml}; \tag{2.16}$$

hence we have

$$M_{lm} \sim -k_m^{-1} \int_0^\infty \varphi_m^0 \eta_{ml} \varphi_l^{0'} \, dr = M_{ml} \tag{2.17}$$

The S matrix is therefore symmetric in the semiclassical limit, under the assumption (2.16). Formally, the S matrix can now be made unitary by

$$S = S_0^{1/2} \left(I + \frac{i}{2} M\right)\left(I - \frac{i}{2} M\right)^{-1} S_0^{1/2}. \tag{2.18}$$

3. DISCUSSION OF THE DIAGONALIZATION MATRIX A

The essential role in the semiclassical expansion (1.11) is played by the matrix A, defined as

$$A^{-1} p A = \lambda, \tag{3.1}$$

where p is symmetric and in the limit $r \to \infty$ is diagonal.

The matrix A is not uniquely determined by Eq. (3.1). This can be easily checked if we replace $A \to AB = C$ and require that the new matrix also diagonalize p. We get

$$A^{-1} p A = \lambda \to C^{-1} p C = B^{-1} \lambda B = \lambda; \tag{3.2}$$

hence the matrix is undetermined to a matrix which commutes with λ, i.e.,

$$B\lambda = \lambda B \to B_{ij}(\lambda_j - \lambda_i) = 0, \quad i, j = 1, 2, \ldots, n. \tag{3.3}$$

In (3.2) we neglected the case whereby B interchanges the order of eigenvalues.

For $\lambda_j \neq \lambda_i$ the matrix B is diagonal with arbitrary matrix elements; therefore, n matrix elements of A are arbitrary. However, for certain values of r, two or more eigenvalues can be identical in which case B is nondiagonal and more than n matrix elements of A are arbitrary. Since degeneracy usually occurs for isolated points of r, we can fix A for an arbitrary r and continue it over the whole range. In such a case, A is uniquely determined if only n of its values are arbitrarily chosen. Let us therefore prove that the continuity is a sufficient condition for determining A uniquely, up to n constants.

Before giving the details of the proof, let us show on one example how A can be obtained. The matrix p is

$$p = V + \frac{l(l+1)}{r^2} - k^2, \tag{3.4}$$

where $l(l+1)$ and k^2 are diagonal and V is nondiagonal with the elements that go to zero for $r \to \infty$ faster than r^{-2}. In general, the off-diagonal elements go to zero faster than the diagonal ones, but there are cases where[15]

$$V_{ij} \sim \alpha_{ij} f(r), \quad r \to \infty, \tag{3.5}$$

where $f(r)$ is some function independent of i and j.

The matrix k^2 is constant and in general nondegenerate. Under these assumptions we can find A for $r \to \infty$ by a perturbation method. The matrix p can be written as

$$p = p_0 + \epsilon p_1, \tag{3.6}$$

where p_0 includes also the diagonal elements of V. The eigenvalues of p are then approximately

$$\lambda = p_0 + O(\epsilon^2), \tag{3.7}$$

which is easily proved by noticing that

$$\frac{d\lambda}{d\epsilon} = -\frac{\partial D/\partial \epsilon}{\partial D/\partial \lambda} \to 0, \quad \epsilon \to 0, \tag{3.8}$$

where $D = \det(p - \lambda)$. If we take (1.17) into account, we can write

$$A = I + \epsilon A_1, \tag{3.9}$$

and, using (3.1), we get an equation for A_1,

$$p_0 A_1 + p_1 = A_1 p_0, \tag{3.10}$$

the solution being

$$[A_1]_{ij} = \frac{p_{ij}}{p_{jj} - p_{ii}}, \quad [A_1]_{i,i} = 0, \tag{3.11}$$

which goes to zero at least as $p_{i,j}$. In the case of degeneracy, say $k_n^2 = k_m^2$ and $l_n \neq l_m$, the element $[A_1]_{m,n}$ goes to zero,

$$r^2 p_{m,n} \to 0, \quad r \to \infty, \tag{3.12}$$

because of our previous assumption on the behavior of $p_{m,n}$ for $r \to \infty$. If $l_n = l_m$, then

$$[A_1]_{m,n} = \frac{p_{mn}}{V_{nn} - V_{mm}} \tag{3.13}$$

and goes to a constant value if we assume (3.5). It goes to zero if $p_{m,n}$ vanishes more rapidly than either of the diagonal elements of V. For $V_{m,m} = V_{n,n}$ the perturbation method fails and a more general procedure should be used for finding A.

In this simple example, we have shown how the matrix elements of A are determined if n parameters are arbitrarily chosen. In our case, this was the choice (3.11). It is now necessary to show that by fixing A at some point r_0 we can continue A to all other values of r. In other words, by assuming we know $A(r_0)$, we must show that $A(r_0 + \Delta)$ can be uniquely determined.

We can approximately write

$$A(r_0 + \Delta) \sim A(r_0) + \Delta A'(r_0), \tag{3.14}$$

and, to prove our statement, we must show that $A'(r_0)$ is entirely determined by $A(r_0)$.

To obtain A', we first take the derivative of (3.1)

$$p'A + pA' = A'\lambda + A\lambda', \tag{3.15}$$

and assume that

$$A' = A\eta, \tag{3.16}$$

where η is a matrix to be determined. Replacing A' in (3.15) by (3.16), we find

$$A^{-1}p'A + \lambda\eta = \eta\lambda + \lambda', \tag{3.17}$$

for which we obtain

$$\eta_{ij}(\lambda_i - \lambda_j) = \lambda_i' \delta_{i,j} - [A^{-1}p'A]_{i,j}. \tag{3.18}$$

For $i \neq j$

$$\eta_{i,j} = -\frac{[A^{-1}p'A]_{i,j}}{\lambda_i - \lambda_j} \tag{3.19}$$

hence the off-diagonal elements of η are uniquely determined by $A(r_0)$.

We still have to obtain the diagonal elements of η. For $i = j$ in (3.18) the left-hand side is zero; hence $\eta_{i,i}$ are arbitrary. However, we can find them by using the fact that n elements of A are arbitrary. For example, we can choose such A that (2.16) is satisfied. Let us show that such a matrix A satisfies

$$\tilde{A} = k^{-1}A^{-1} \tag{3.20}$$

and defines η, which ensures that the S matrix is symmetric. $\tilde{A}$ in (3.20) is the transpose of A.

If we assume (3.20) the matrix elements of η, given by (3.19), are

$$\eta_{i,j} = \frac{k_i[\tilde{A}p'A]_{i,j}}{\lambda_j - \lambda_i}, \tag{3.21}$$

and by interchanging i and j we find

$$\eta_{j,i} = k_j \frac{[\tilde{A}p'A]_{j,i}}{\lambda_i - \lambda_j} = -\frac{k_j}{k_i}\eta_{i,j} \tag{3.22}$$

which is the required condition (2.16). Therefore, the matrix A must satisfy (3.20) if the S matrix should be symmetric. It follows from (3.20) that

$$\tilde{A}A = k^{-1}, \tag{3.23}$$

and, since A is diagonal in the limit $r \to \infty$, we obtain

$$A \to k^{-1/2}, \quad r \to \infty, \tag{3.24}$$

being the limit of A of the type (1.23). Hence, the symmetric S matrix is

$$S = \exp[i\pi(l+1)/2]\, j^{+}(j^{-})^{-1} \exp[i\pi(l+1)/2]. \tag{3.25}$$

We can now obtain the diagonal elements of η. By taking derivative of (3.23) we find

$$\tilde{A}'A + \tilde{A}A' = \tilde{\eta}k^{-1} + k^{-1}\eta = 0 \tag{3.26}$$

from which it follows

$$\eta_{i,i} = 0, \quad i = 1, 2, \ldots, n, \tag{3.27}$$

thus concluding the calculation of the η matrix.

With this we also conclude our proof that the continuity is a sufficient condition for determining $A(r)$ for all r if $A(r_0)$ is fixed.

The set of equations

$$A' = A\eta, \tag{3.28}$$

where the matrix elements of η are given by (3.19) and (3.27), and the equations

$$\lambda_i' = k_i[\tilde{A}p'A]_{i,i}, \tag{3.29}$$

obtained from (3.18) by setting $i = j$, form a system of nonlinear differential equations which can be integrated numerically. The condition (3.27) ensures that the con-

dition (3.23) is satisfied for all r if it is defined for the initial point $r=r_0$.

In practice, it is more useful to define a unitary A matrix, designated by U. The matrix A, defined by (3.20), is then related to U as

$$A=Uk^{-1/2}. \tag{3.30}$$

4. ONE EXAMPLE

Let us show on a simple example of a two-channel problem how the theory is applied. It is assumed that $p_{11}>p_{22}$ for $r\to\infty$.

The eigenvalues of p are given by

$$\lambda_{1,2}=\frac{p_{11}+p_{22}}{2}\pm\tfrac{1}{2}[(p_{11}-p_{22})^2+4p_{12}^2]^{1/2}, \tag{4.1}$$

and the matrix elements of A are

$$\begin{aligned} a_{ii}&=p_{12}[(\lambda_i-p_{ii})^2+p_{12}^2]^{-1/2}, \quad i=1,2,\\ a_{ij}&=(\lambda_j-p_{jj})[(\lambda_i-p_{ii})^2+p_{12}^2]^{-1/2} \quad i=1,2, \quad j=2,1, \end{aligned} \tag{4.2}$$

where we have for simplicity assumed the unitary A matrix.

The elements of the coupling matrix η are

$$\eta_{ii}=0, \quad i=1,2, \tag{4.3}$$

and

$$\eta_{12}=-\eta_{21}=\frac{(p_{22}-p_{11})^2}{(p_{22}-p_{11})^2+4p_{12}^2}\,\frac{d}{dr}\,\frac{p_{12}}{p_{22}-p_{11}}. \tag{4.4}$$

From (4.1) we deduce that unless $p_{12}=0$, the eigenvalues of p cannot be degenerate. They are approximately degenerate if for $r=r_s$ we have $p_{11}=p_{22}$ and $p_{12}\sim 0$. In this case

$$\lambda_{1,2}\sim p_{11}, \quad \Delta\lambda=\lambda_1-\lambda_2=2p_{12}. \tag{4.5}$$

Near $r=r_s$ the elements of A change rapidly. In the region $r>r_s$ we have approximately

$$\begin{aligned} &a_{11}\sim a_{22}\sim 1,\\ &|a_{12}|\sim|a_{21}|\sim\left|\frac{p_{12}}{p_{22}-p_{11}}\right|\ll 1, \quad r>r_s \end{aligned} \tag{4.6}$$

while for $r<r_s$ they are

$$\begin{aligned} &-a_{12}\sim a_{21}\sim 1,\\ &|a_{11}|\sim|a_{22}|\sim\left|\frac{p_{12}}{p_{22}-p_{11}}\right|\ll 1, \quad r<r_s. \end{aligned} \tag{4.7}$$

In other words, from $r>r_s$ to $r<r_s$, the matrix changes from the unit matrix to an exchange matrix, i.e., the matrix which interchanges the order of the eigenvalues.

The appropriate η matrix, which measures the coupling between the two eigenchannels, is approximately for $r>r_s$

$$|\eta_{12}|\sim\left|\frac{d}{dr}\,\frac{p_{12}}{p_{22}-p_{11}}\right|\ll 1 \tag{4.8}$$

and for $r<r_s$

$$|\eta_{12}|\sim\left|\frac{d}{dr}\,\frac{p_{12}}{p_{22}-p_{11}}\right|\ll 1. \tag{4.9}$$

In both cases it is small. However, for $r=r_s$ it becomes

$$|\eta_{12}|=\left|\frac{p'_{11}-p'_{22}}{4p_{12}}\right|, \tag{4.10}$$

having a large value for a small p_{12}. Therefore, all the transition between the eigenchannels occurs in a small neighborhood of $r=r_s$.

If p'_{12} can be neglected, then $r=r_s$ coincide with the maximum of $|\eta_{12}|$ and (4.10) is its value. The shape of η_{12} is therefore a Lorentzian curve with the width

$$\Delta r_w=8\left|\frac{p_{12}}{p'_{11}-p'_{22}}\right|. \tag{4.11}$$

In the vicinity of $r=r_s$ the matrix element η_{12} can be approximated by

$$\eta_{12}\sim\frac{p'_{11}-p'_{22}}{4p_{12}}\,\frac{1}{1+\epsilon^2}, \tag{4.12}$$

where

$$\epsilon=\frac{p'_{22}-p'_{11}}{2p_{12}}(r-r_s). \tag{4.13}$$

We have neglected p'_{12}.

The S matrix can now be calculated. We will assume approximation (4.12), and for the eigenvalues (4.1) we obtain

$$\lambda_{1,2}\sim\bar{p}\pm p_{12}(1+\epsilon^2)^{1/2}, \tag{4.14}$$

where

$$\bar{p}=(p_{11}+p_{22})/2. \tag{4.15}$$

In our discussion we take a more general form of the S matrix than that given by (2.9). Instead of taking the limit $\hbar\to 0$ as described in Sec. 2, we calculate the most general S matrix in the first Born approximation and then obtain the limit $\hbar\to 0$. This procedure will give more insight into the nature of the semiclassical approximation.

The matrix M define by (2.13) is

$$M=k^{-1}\int_0^\infty\varphi_0([\varphi'_0,\eta]+\tfrac{1}{2}[\varphi_0,\eta^2]+\tfrac{1}{2}[\varphi_0,\eta']+[\eta,\varphi_0\eta])\,dr' \tag{4.16}$$

It can easily be shown that, for a two-channel problem,

$$[\varphi_0,\eta^2]=0, \tag{4.17}$$

and, using (2.14), we obtain for (4.16)

$$M=k^{-1}\int_0^\infty[\tfrac{1}{2}\varphi'_0\eta\varphi_0-\tfrac{1}{2}\varphi_0\eta\varphi'_0+\varphi_0(\eta\varphi_0\eta-\varphi_0\eta^2)]\,dr. \tag{4.18}$$

The off-diagonal element M_{12} is therefore

$$M_{12}=\tfrac{1}{2}k_1^{-1}\int_0^\infty(\varphi_1^{0'}\varphi_2^0-\varphi_1^0\varphi_2^{0'})\eta_{12}\,dr, \tag{4.19}$$

and the diagonal one,

$$M_{jj}=k_j^{-1}\sum_{l=1}^{2}\int_0^\infty\varphi_j^0\eta_{jl}\eta_{lj}(\varphi_l^0-\varphi_j^0)\,dr, \quad j=1,2. \tag{4.20}$$

To calculate the matrix elements (4.19) and (4.20) we use the WKB solution of (1.15), which is for a one turning point problem[2]

$$\varphi_0=2k^{1/2}\lambda^{-1/4}\cos\left(\int_{r_0}^r\lambda^{1/2}\,dr'-\pi/4\right); \tag{4.21}$$

hence the matrix element M_{12} is

$$M_{12}=2(k_2/k_1)^{1/2}\int_0^\infty dr\,\eta_{12}\sin\Big(\int_{r_2}^{r}\lambda_2^{1/2}\,dr'-\int_{r_1}^{r}\lambda_1^{1/2}\,dr'\Big), \tag{4.22}$$

where we have assumed

$$\lambda_1=\lambda_2 \tag{4.23}$$

since most of the contribution comes from the vicinity of $r=r_s$.

With the approximation (4.12) and (4.14) we finally obtain

$$M_{12}=2\sin\phi_0\int_0^\infty\frac{d\epsilon}{1+\epsilon^2}\cos\left(\frac{2p_{12}^2}{|p'_{11}-p'_{22}|\,\bar p^{1/2}}\times\int_0^\epsilon d\epsilon'(1+\epsilon'^2)^{1/2}\right) \tag{4.24}$$

where

$$\phi_0=\int_{r_2}^{r_s}\lambda_2^{1/2}\,dr-\int_{r_1}^{r_s}\lambda_1^{1/2}\,dr. \tag{4.25}$$

We have assumed (2.16), in order that M matrix be symmetric.

The integral can be integrated in the complex ϵ plane. By defining the cut along the imaginary ϵ axes, starting at $\epsilon=\pm i$, we find

$$M_{12}=2\exp(-\pi C/2)[\pi/2+\sigma(C)]\sin\phi_0, \tag{4.26}$$

where

$$C=p_{12}^2/|p'_{11}-p'_{22}|\,\bar p^{1/2} \tag{4.27}$$

and

$$\sigma=\int_1^\infty\frac{d\epsilon}{\epsilon^2-1}\sin\big(C\{\ln[\epsilon+(\epsilon^2-1)^{1/2}]-\epsilon(\epsilon^2-1)^{1/2}\}\big). \tag{4.28}$$

Since

$$C\sim\hbar^{-1}, \tag{4.29}$$

most of the contribution of σ comes from $\epsilon\sim0$ so that we find the limit

$$\lim_{\hbar\to0}\sigma=-\pi/6. \tag{4.30}$$

The function σ is shown in Fig. 1.

We can now calculate the diagonal elements of M. Starting from (4.20) and using (4.21), we obtain

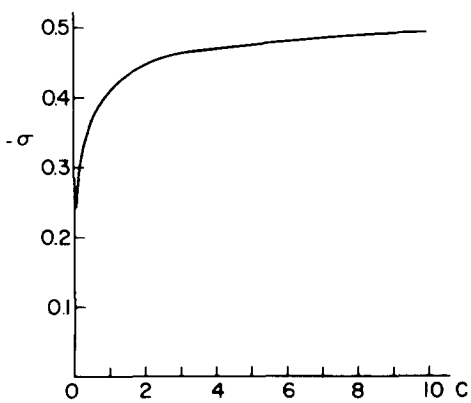


FIG. 1. The correction function σ for the inelastic amplitude S_{12}.

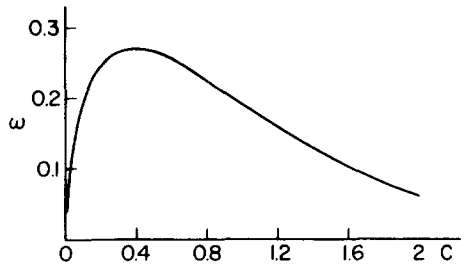


FIG. 2. The correction function ω for the elastic amplitude S_{ii}.

$$M_{11}=-\frac14\frac{|p'_{22}-p'_{11}|}{p_{12}\lambda_1^{1/2}}\int_{-\infty}^{\infty}\frac{d\epsilon}{(1+\epsilon^2)^2}[\cos(\alpha_2-\alpha_1)-1] \tag{4.31}$$

where

$$\alpha_2-\alpha_1=\phi_0-\frac{2p_{12}^2}{|p'_{22}-p'_{11}|\,\bar p^{1/2}}\int_0^\epsilon dx(1+x^2)^{1/2}. \tag{4.32}$$

In (4.31) we have neglected the oscillatory terms $\alpha_1+\alpha_2$ and $2\alpha_1$. The remaining integrals can be calculated as in the case for M_{12}, with the result

$$M_{11}=\frac14\frac{|p'_{22}-p'_{11}|}{p_{12}\lambda_1^{1/2}}\left[\frac{\pi}{2}-\exp(-\pi C/2)\left(\frac{\pi}{2}+\sigma\right)\cos\phi_0-2\omega(c)\cos\phi_0\right], \tag{4.33}$$

where C and σ are defined by (4.27) and (4.28) respectively. The function ω is defined as

$$\omega(C)=C\int_1^\infty dx\sin\big(C\{x(x^2-1)^{1/2}+\ln[x+(x^2-1)^{1/2}]\}\big) \tag{4.34}$$

and is shown in Fig. 2.

Similarly, it can be shown that

$$M_{22}=M_{11} \tag{4.35}$$

under the assumption that $\alpha_1+\alpha_2$ and $2\alpha_2$ can be neglected.

The results for the M matrix, given by (4.26) and (4.33), were obtained under several assumptions; it is assumed that (a) the WKB solution of (1.15) is accurate (which is not the case when r_s is near the turning point of λ_i), (b) p'_{12} can be neglected, and (c) we can use linear approximation of p_{11} and p_{22} near $r=r_s$. To improve results, we can use more accurate (e.g., numerical) solution of (1.15) and take exact η matrix instead of (4.12). However, even the approximate results (4.26) and (4.33) describe the essential features of the scattering matrix. In particular, they give correct oscillatory behavior of S_{12}, first noted by Stückelberg[4] and the exponential decay.[7,8] In addition, corrections due to the repulsion of the eigenchannels near $r=r_s$ are also included through the functions σ and ω.

As has been noted in Sec. 3, the diagonal elements of M are small compared to the off-diagonal ones. From (4.26) and (4.33) we find that this is true provided C is small, in which case

$$|M_{11}/M_{12}|\sim h \tag{4.36}$$

However, in the semiclassical limit, the constant C is large, hence the matrix element M_{12} is small, being of the order

$$|M_{12}| \sim \exp(-\pi C/2) \sim \exp(-\hbar^{-1}) \tag{4.37}$$

while the diagonal elements M_{ii} are of the order

$$|M_{ii}| \sim \hbar. \tag{4.38}$$

Therefore, the elastic matrix elements of M are much larger then the inelastic ones, although the former are also small in the limit $\hbar \to 0$. This is associated with the two physical phenomena. The semiclassical particle approaching the target in the channel 1 notices only the potential λ_1. Where the two potential curves λ_1 and λ_2 come close, the change in the slope of both is large, due to their repulsion. The particle can either be reflected, because of the sudden change of slope or it can tunnel to the channel 2, because of the proximity of the two channels. Since the tunneling amplitude is of the order (4.37), the most dominant will be the reflection amplitude, being of the order $\hbar$. Therefore, the elastic amplitude M_{11} is given as a sum of the amplitude for a particle to be reflected from the "crossing" point and the amplitude for a particle to tunnel to the channel 2 and subsequently tunnel to the channel 1.

5. PERTURBATION EXPANSION OF η

As we have shown, the matrix elements of η, apart from the diagonal ones which are exactly zero, have a Lorentzian shape near the point of degeneracy of two eigenchannels. The exact shape cannot be simply determined from p, and we either have to integrate the set of equations (3.28) and (3.29) or use the perturbation expansion. The second choice applies only if the the off-diagonal elements of p are small compared to the difference between the channel energies. In this section we show how this fact is used to develop the perturbation expansion of η.

The off-diagonal elements of η are

$$\eta_{ij} = \frac{[Ap'A]_{ij}}{\lambda_j - \lambda_i} \tag{5.1}$$

and if we use (3.6) together with (3.9), we easily find

$$[\tilde{A}p'A]_{ij} \sim [p_0']_{ij} + \epsilon[p_1' + p_0'A_1 - A_1p_0']_{ij} \tag{5.2}$$

Since we assume $\lambda_j \neq \lambda_i$ and taking into account (3.7), the matrix element η_{ij} is

$$\eta_{ij} \sim -\frac{d}{dr}\frac{p_{ij}}{p_{jj} - p_{ii}}. \tag{5.3}$$

The diagonal elements are zero.

However, we are not interested in the region of r where the eigenvalues of p are well separated. As we have shown in the previous section, this region does not contribute to the transition probability. A more interesting case is around such r where two or more eigenvalues of p are either degenerate of nearly degenerate. We assume the simplest case; in the vicinity of $r = r_s$ two diagonal elements p_{11} and p_{22} are equal and all the other elements are well separated (the choice of p_{11} and p_{22} is not significant). In such a case, we formulate a more general perturbation expansion of η, based on the perturbation

$$p = p_0 + \epsilon p_1, \tag{5.4}$$

where, in addition to all diagonal elements of p, the unperturbed matrix p_0 containes p_{12} and p_{21}.

As in (3.7) we can show that

$$\lambda_i \sim \lambda_i^0 + O(\epsilon^2), \tag{5.5}$$

where

$$\lambda_i^0 = p_{ii}, \quad i > 2, \tag{5.6}$$

and λ_1^0 and λ_2^0 are given by (4.1).

The matrix A can be represented as

$$A = \mathbf{a}B, \tag{5.7}$$

where $\mathbf{a}$ is a unitary matrix containing units on the diagonal and the elements a_{11}, a_{12}, a_{21}, and a_{22} are given by (4.2). All the other elements of $\mathbf{a}$ are zero. The matrix B can be written as

$$B = I + \epsilon B_1 \tag{5.8}$$

where I is the unit matrix. From the equation for A we find

$$B_1\lambda_0 - \lambda_0 B_1 = \tilde{\mathbf{a}}p_1\mathbf{a} \tag{5.9}$$

where we have used the fact that $\mathbf{a}$ diagonalizes p_0.

From (5.9) we obtain

$$B_{ij}^1 = \frac{[\tilde{\mathbf{a}}p_1\mathbf{a}]_{ij}}{\lambda_j - \lambda_i}, \tag{5.10}$$

with the diagonal elements still to be determined. We find them by requiring that B is unitary, which gives

$$B_{ii}^1 = 0. \tag{5.11}$$

The matrix elements of B^1 for $i, j = 1, 2$ are all zero. For $i, j > 2$, it can easily be shown that

$$B_{ij}^1 = \frac{p_{ij}}{p_{jj} - p_{ii}} \tag{5.12}$$

while for $i = 1, 2$ and $j > 2$ we have

$$B_{ij}^1 = \sum_{k=1}^{2} \frac{a_{ki}p_{kj}}{p_{jj} - \lambda_i}. \tag{5.13}$$

It can also be shown that B^1 is antisymmetric, i.e., $B^1 = -B^1$.

We can now calculate the matrix η. From (5.1) and assuming

$$\eta_{ij} = \eta_{ij}^0 + \epsilon\eta_{ij}^1 \tag{5.14}$$

we obtain

$$\eta_{ij}^0 = \frac{[\tilde{\mathbf{a}}p_0'\mathbf{a}]_{ij}}{\lambda_{jj} - \lambda_{ii}}, \quad i \neq j, \tag{5.15}$$

and

$$\eta_{ij}^1 = \frac{[\tilde{\mathbf{a}}p_1'\mathbf{a}]_{ij} + [\tilde{\mathbf{a}}p_0'\mathbf{a}B^1]_{ij} - [B^1\tilde{\mathbf{a}}p_0'\mathbf{a}]_{ij}}{\lambda_j - \lambda_i}, \quad i \neq j. \tag{5.16}$$

In the derivation of (5.15) and (5.16) we have used the fact that B^1 is antisymmetric, in which case

$$\tilde{A} = I - \epsilon B^1 \tag{5.17}$$

The matrix elements of η^0_{ij} are all zero except η^0_{12} and η^0_{21} which are given by (4.4). On the other hand, the matrix elements of η^1_{ij} are all different than zero except the diagonal ones and η^1_{12} (η^1_{21}). Therefore, for $i,j=1,2$ the matrix η is given by

$$\eta=\eta^0+O(\epsilon^2) \tag{5.18}$$

while if any index of η is greater than 2, we have

$$\eta=\epsilon\eta^1+O(\epsilon^2). \tag{5.19}$$

More specifically, for $i,j>2$ the matrix elements of η are given by (5.1) but for $i=1,2$, and $j>2$ we obtain

$$\eta^1_{ij}=-\frac{B^1_{ij}p'_{jj}}{p_{jj}-\lambda_i}+\sum_{k=1}^{2}\frac{a_{ki}p'_{kj}}{p_{jj}-\lambda_i}+\sum_{k,l,m=1}^{2}\frac{a_{ki}p'_{kl}a_{lm}B'_{mj}}{p_{jj}-\lambda_i}. \tag{5.20}$$

Let us discuss the last result in two limiting cases, i.e., to the right and left of the point $r=r_s$ for which $p_{11}=p_{22}$. In both cases we use the asymptotic values for $a_{i,j}$ given by (4.4). If we specify $i=1$, the matrix element η^1_{1j} is

$$\eta^1_{1j}\sim\frac{d}{dr}\frac{p_{1j}}{p_{jj}-p_{11}}+\frac{p'_{12}p_{2j}}{(p_{jj}-p_{11})(p_{jj}-p_{22})},\quad r\gg r_s, \tag{5.21}$$

being equal to (5.3) if p'_{12} is neglected. However, to the left of $r=r_s$ we obtain

$$\eta^1_{1j}\sim\frac{p'_{2j}}{p_{jj}-p_{11}}-\frac{p_{jj}-p_{22}}{(p_{jj}-p_{11})^2}p_{2j}+\frac{p_{21}p_{1j}}{(p_{jj}-p_{11})(p_{jj}-p_{22})},\quad r\ll r_s, \tag{5.22}$$

indicating that η^1_{1j} changes rapidly in the interval Δr_W. If the derivatives of the off-diagonal elements of p are neglected the change in η^1_{1j} is

$$|\Delta\eta^1_{1j}|=\left|\frac{(p'_{jj}-p'_{11})p_{1j}-(p'_{jj}-p'_{22})p_{1j}}{(p_{jj}-p_{11})^2}\right|. \tag{5.23}$$

In contrast to the element η_{12}, which has a Lorentzian shape, the elements η_{1j} are a steplike functions.

The elements η^1_{2j} are obtained from (5.21) and (5.22) by formally replacing $1\longleftrightarrow 2$. However, (5.22) also changes the sign.

6. WEAK COUPLING LIMIT

The perturbation expansion of the S matrix becomes inadequate for large transition amplitudes. In this section we discuss another case where the expansion (2.9) is not only inadequate, but it actually fails to give correct scattering matrix. For simplicity we take the two-channel example.

The width of η_{12} is

$$\Delta r_W=8\left|\frac{p_{12}}{p'_{11}-p'_{22}}\right|. \tag{6.1}$$

In the derivation of (4.26) and (4.33) we have implicitly assumed that over this distance the eigenchannel wavefunctions φ_i are oscillatory. In other words, we assumed

$$\Delta r_W\gg 2\pi\bar{p}^{-1/2}, \tag{6.2}$$

where $\bar{p}$ is defined by (4.15). However, there are cases where this condition is not satisfied, e.g., p_{12} is small or the derivatives of p_{11} and p_{22} are large and of the opposite sign. In these cases the constant C, defined by (4.27), is small and the transition amplitude M_{12} is large. Furthermore, the elastic amplitude (4.33), being of the order

$$M_{11}\sim\frac{|p'_{22}-p'_{11}|}{p_{22}\bar{p}^{1/2}} \tag{6.3}$$

is infinite for $p_{12}\to 0$. In reality, this is not the case since for $p_{12}\to 0$, the S matrix is diagonal, i.e., the transition matrix M is zero.

There are two ways to solve this problem: numerical integration of the set of equations (1.12) in the interval $[r_s-\Delta r_W,\ r_s+\Delta r_W]$ or perturbation expansion of the type (1.5). There is no practical difficulty in using the numerical integration since the solution is not oscillatory in this interval. Outside the interval η_{12} is negligible therefore the numerical solution can be connected to the solutions of (1.15) and the S matrix is obtained from (1.22) and (1.20).

The perturbation method to be used is the distorted wave expansion of the S matrix. For small p_{12} the series quickly converges, but we must show that is can also be used under a more general condition. To obtain this condition, let us calculate S_{12} in the first Born approximation of the distorted wave expansion. It can be shown that S_{12} is[16]

$$S_{12}\sim\frac{2}{(k_1k_2)^{1/2}}\int_0^\infty dr\,\psi_1 p_{12}\psi_2, \tag{6.4}$$

where ψ_1 and ψ_2 are the solutions of the uncoupled set of equations (1.1). To estimate (6.4), we replace the wavefunctions with the WKB solution and integrate by the steepest descent method. We obtain

$$S_{12}\sim\frac{2p_{12}\sqrt{\pi}}{\bar{p}^{1/4}(p'_{11}-p'_{22})^{1/2}}\sim 2\sqrt{\pi C} \tag{6.5}$$

The distorted wave expansion is therefore valid even when p_{12} is not small provided the derivatives of p_{11} and p_{22} are large and of the opposite sign. As we have shown, this is exactly the case when the semiclassical expansion of the S matrix breaks down. More specifically, if $C^{1/2}$ is small, we can use the distorted wave expansion for calculating the S-matrix. It can be easily shown that this condition is equivalent with the width of η_{12} being smaller than the local wave length of the wavefunction [the condition (6.2) with the reversed inequality sign] and $p_{12}/\bar{p}\ll 1$.

In the case of more than two channels we can use the perturbation expansion of η, described in Sec. 5, to find the behavior of η_{12} in the vicinity of degeneracy in the channels 1 and 2. If the width of η_{12} is smaller than the local wave length, we use the distorted wave expansion to obtain the S-matrix element S_{12}. Since we have assumed that all the other channels are well separated, the transition amplitude S_{12} is independent of the presence of the other channels. However, this is not always the case. For example, p_{12} can be so small that the direct transition amplitude for $1\to 2$ is of the same magnitude of the higher order transitions. An obvious

example is when $p_{12}=0$ in which case the direct transition amplitude is exactly zero. Only the higher order transitions give nonzero contributions. We will refer to such a process as the "forbidden transitions."

Let us show how, in the framework of the η matrix, the transition amplitudes for the "forbidden process" can be calculated. For simplicity we assume that the forbidden transition is between channels 1 and 2.

The matrix p can be written as the block matrix

$$p=\begin{vmatrix} p_{11}, & p' \\ p', & P \end{vmatrix}. \tag{6.6}$$

Let us define a matrix $\mathbf{a}$ which diagonalizes P, so that we have

$$\begin{vmatrix} 1, & 0 \\ 0, & \tilde{\mathbf{a}} \end{vmatrix} \begin{vmatrix} p_{11}, & p' \\ p', & P \end{vmatrix} \begin{vmatrix} 1, & 0 \\ 0, & \mathbf{a} \end{vmatrix} = \begin{vmatrix} p_{11}, & p'\mathbf{a} \\ \tilde{\mathbf{a}}p', & \lambda \end{vmatrix} = \bar{p}. \tag{6.7}$$

The new matrix $\bar{p}$ replaces p in the calculation of η. Therefore, we have new parameters for the transition $1 \to 2$,

$$\bar{p}_{11}=p_{11}, \quad \bar{p}_{22}=\lambda_2 \tag{6.8}$$

and

$$\bar{p}_{21}=\bar{p}_{12}=\sum_{k=2}^{n} p_{1k}\, a_{k2}; \tag{6.9}$$

hence η_{12} is

$$\eta_{12}=\frac{(\lambda_2-p_{11})^2}{(\lambda_2-p_{11})^2+4\bar{p}_{12}^2}\,\frac{d}{dr}\,\frac{\bar{p}_{12}}{\lambda_2-p_{11}}+O(\epsilon^2), \tag{6.10}$$

where $\epsilon=\|p'\mathbf{a}\|$.

Since we assume that p_{jj}, $j\geq 2$, are well separated, we can calculate λ_2 and a_{k2} by the perturbation method. The eigenvalue λ_2 is to the second order in perturbation

$$\lambda_2=p_{22}+\sum_{j=3}^{n}\frac{p_{2j}^2}{p_{22}-p_{jj}} \tag{6.11}$$

while a_{k2} are given by (3.9) and (3.11), i.e.,

$$a_{k2}\sim\delta_{k2}+\frac{p_{k2}}{p_{22}-p_{kk}}, \quad k\geq 3. \tag{6.12}$$

Therefore, we have

$$\bar{p}_{22}\sim p_{22}+\sum_{j=3}^{n}\frac{p_{2j}^2}{p_{22}-p_{jj}} \tag{6.13}$$

and

$$\bar{p}_{12}\sim p_{12}+\sum_{k=3}^{n}\frac{p_{1k}p_{k2}}{p_{22}-p_{kk}}, \tag{6.14}$$

showing that even when p_{12} is strictly zero, we can define an effective $\bar{p}_{12}$ so that the transition amplitude S_{12} can be calculated from the first order distorted wave approximation.

CONCLUSION

The semiclassical solution of the multichannel equations described in this paper is a simple generalization of the usual distorted wave expansion. It is a perturbation method with the diagonalization matrix having the role of perturbation. Since this matrix is independent of $\hbar$, we were able to prove that the resulting series is also independent of $\hbar$. In the weak coupling case the method fails, but, nevertheless, it can be used as a suitable decoupling scheme.

ACKNOWLEDGMENTS

The author wishes to thank Professor D. Milicic of the University of Zagreb for a helpful discussion. We would also wish to thank the Lady Davis Fellowship for financial support.

[1]E.C. Kemble, Phys. Rev., **48**, 549 (1935).
[2]N. Fröman and P.O. Fröman, *JWKB Approximation* (North-Holland, Amsterdam, 1965).
[3]W.H. Bassichis and A. Dar, Ann. Phys. **36**, 130 (1966).
[4]E.C.G. Stückelberg, Helv. Phys. Acta **5**, 370 (1932).
[5]D.J. Vezzetti and S.I. Rubinow, Ann. Phys. **35**, 373 (1965).
[6]B.C. Eu, J. Chem. Phys. **55**, 5600 (1971).
[7]L.D. Landau, Zh. Phys. Sov. Un. **2**, 46 (1932).
[8]C. Zener, Proc. Roy. Soc. A **137**, 696 (1932).
[9]J.B. Delos, W.R. Thorson, and S.K. Knudsen, Phys. Rev. A **6**, 709 (1972).
[10]J.B. Delos and W.R. Thorson, Phys. Rev. A **6**, 720 (1972).
[11]R.G. Newton, Ann. Phys. (N.Y.) **4**, 29 (1958).
[12]R.G. Newton, *Scattering Theory of Waves and Particles* (McGraw-Hill, New York, 1966), Chap. 17.3.
[13]Reference 12, Chap. 15.
[14]R.B. Gerber, Chem. Phys. **16**, 19 (1976).
[15]For example, the coupling matrix for spin–spin transitions or atom-rigid rotor collisions.
[16]Reference 12, Chap. 17.

On the diagonalization of the general quadratic Hamiltonian for coupled harmonic oscillators

Yoel Tikochinsky

Racah Institute of Physics, The Hebrew University of Jerusalem, Jerusalem, Israel
(Submitted 24 April 1978)

It is shown that the general quadratic Hamiltonian for coupled harmonic oscillators can be diagonalized, provided the matrix of the quadratic form is positive definite. This condition is also necessary if the frequencies of the resulting uncoupled oscillators are to be positive. The construction of a diagonalizing matrix follows the usual procedure as in the Hermitian case; the only difference being a change of the metric from

$I=\begin{pmatrix}1 & \\ & 1\end{pmatrix}$ to $J=\begin{pmatrix}1 & \\ & -1\end{pmatrix}$.

1. INTRODUCTION

It is well known that the eigenvalues for a system of n coupled harmonic oscillators can be found by means of a principal axis transformation. Thus, given the Hamiltonian

$$H=(a^\dagger a)/\!/\begin{pmatrix}a\\ a^\dagger\end{pmatrix}, \tag{1}$$

with $(a^\dagger a)=(a_1^\dagger\cdots a_n^\dagger, a_1\ldots a_n)$ and $/\!/=\begin{pmatrix}\alpha & \beta\\ \beta^* & \alpha^*\end{pmatrix}=/\!/^\dagger$, a $2n$ by $2n$ regular matrix A is sought such that

$$(A^{-1})^\dagger/\!/(A^{-1})=/\!/' \tag{2}$$

is diagonal. [Notation: Hermitian conjugation of matrices and operators is denoted by a dagger. Thus $\begin{pmatrix}a\\ a^\dagger\end{pmatrix}^\dagger=(a^\dagger a)$. A star denotes complex conjugation and a tilde stands for the transpose.] Once A has been found, the eigenvalues can be read from the uncoupled form

$$H=(b^\dagger b)/\!/'\begin{pmatrix}b\\ b^\dagger\end{pmatrix}=\sum\hbar\omega_i b_i^\dagger b_i+\text{const}, \tag{3}$$

where

$$\begin{pmatrix}b\\ b^\dagger\end{pmatrix}=A\begin{pmatrix}a\\ a^\dagger\end{pmatrix} \tag{4}$$

are the new boson operators. In order for the b's to satisfy the boson commutation rules, A must fulfill[1,2]

$$A=\begin{pmatrix}\lambda & \mu\\ \mu^* & \lambda^*\end{pmatrix}, \tag{5a}$$

and

$$AJA^\dagger=J, \tag{5b}$$

where $J=\begin{pmatrix}1 & \\ & -1\end{pmatrix}$. The question now arises: What conditions on the Hamiltonian matrix $/\!/$ are sufficient to ensure the existence of a matrix A satisfying (2) and (5), and how is one going to construct A? It seems that such conditions have not been spelled out in the literature, though a set of operational rules to facilitate the computation of A (when it exists), has recently been formulated.[1] Reference 1 also contains a fairly broad list of interesting applications to physical problems.

The purpose of this note is to show that A exists and can be calculated by following routine diagonalization procedures, provided $/\!/$ is positive definite, that is,

$$(x)^\dagger/\!/(x)>0, \quad \forall(x)\neq(0). \tag{6}$$

Equivalently, since $/\!/$ is Hermitian, the eigenvalues of $/\!/$ should be positive. [Note that, in general, the eigenvalues of $/\!/$ are not the diagonal elements of $/\!/'$ in Eq. (2).] As we shall presently show, Condition (6) is also necessary, if the frequencies of the uncoupled Hamiltonian (3) are to be positive. Indeed, let $/\!/'=\begin{pmatrix}\gamma & \\ & \gamma'\end{pmatrix}$. Since $/\!/$ is Hermitian, γ and γ' are real diagonal n by n matrices with diagonal elements γ_i and γ_i'. Let $A=\begin{pmatrix}\lambda & \mu\\ \mu^* & \lambda^*\end{pmatrix}$. Then, by Eq. (5),

$$A^{-1}=JA^\dagger J=\begin{pmatrix}\lambda^\dagger & -\tilde{\mu}\\ -\mu^\dagger & \tilde{\lambda}\end{pmatrix}.$$

Equation (2), therefore, reads

$$\begin{pmatrix}\lambda & -\mu\\ -\mu^* & \lambda^*\end{pmatrix}\begin{pmatrix}\alpha & \beta\\ \beta^* & \alpha^*\end{pmatrix}\begin{pmatrix}\lambda^\dagger & -\tilde{\mu}\\ -\mu^\dagger & \tilde{\lambda}\end{pmatrix}=\begin{pmatrix}\gamma & \\ & \gamma'\end{pmatrix}. \tag{7}$$

Performing the matrix multiplication, we find $\gamma=\gamma'$, that is, $/\!/'=\begin{pmatrix}\gamma & \\ & \gamma\end{pmatrix}$. Thus, the uncoupled Hamiltonian (3) is explicitly given by

$$H=\sum_{i=1}^{n}2\gamma_i(b_i^\dagger b_i+\tfrac{1}{2}). \tag{8}$$

Accordingly, if the frequencies $\omega_i=2\gamma_i/\hbar$ are to be positive, the Hamiltonian matrix

$$/\!/=A^\dagger/\!/'A=(/\!/'^{1/2}A)^\dagger(/\!/'^{1/2}A) \tag{9}$$

must be positive definite.

Before turning to the discussion of diagonalization, let us recast Eq. (2), with the aid of (5), into the usual eigenvalue form, namely,

$$(/\!/J)A^\dagger=A^\dagger(/\!/'J). \tag{10}$$

Note that $/\!/J$ is, in general, not Hermitian. Another useful form is

$$AJ/\!/JA^\dagger=/\!/'. \tag{11}$$

In Sec. 2 we shall first establish the existence of a diagonalizing matrix A, and then show how to actually construct it. As it turns out, the construction follows exactly the same procedure as in the Hermitian case, the only difference being the change of the metric from $I=\begin{pmatrix}1 & \\ & 1\end{pmatrix}$ to $J=\begin{pmatrix}1 & \\ & -1\end{pmatrix}$. Finally, in the Appendix, we amend a shortcoming in Ref. 2, wherein a related problem is discussed, namely, given the diagonalizing matrix A, find the transformation brackets connecting base states of the old operators $(a_i, a_i^\dagger)$ with those of the new operators $(b_i, b_i^\dagger)$.

2. DIAGONALIZATION

A. Existence

Let $/\!/$ be a positive definite $2n$ by $2n$ Hermitian

 0022-2488/79/030406-03$01.00

matrix. Since JHJ shares the same properties, there exists a unitary matrix U such that

$$U^\dagger JHJU = E \tag{12}$$

is a diagonal positive matrix. Define

$$V = UE^{-1/2}, \tag{13}$$

Then V is regular and satisfies

$$V^\dagger JHJV = 1. \tag{14}$$

Let P be a unitary matrix diagonalizing the Hermitian matrix $V^\dagger JV$, that is,

$$P^\dagger V^\dagger JVP = \eta, \qquad \eta_{ij} = \delta_{ij}\eta_i. \tag{15}$$

Then

$$W = VP \tag{16}$$

satisfies

$$W^\dagger JW = \eta, \tag{17}$$

and

$$W^\dagger JHJW = 1. \tag{18}$$

Equation (17) is a congruence transformation on the matrix J conserving the number of positive and negative elements in J (Sylvester's law of inertia[3]). Hence, n of the η_i's are positive and n are negative. Let θ be a diagonal matrix with elements

$$\theta_{ij} = \delta_{ij}|\eta_i|. \tag{19}$$

Then

$$\zeta = ((\zeta_1)(\zeta_2)\cdots(\zeta_{2n})) \equiv W\theta^{-1/2} \tag{20}$$

satisfies

$$(\zeta_i)^\dagger J(\zeta_j) = \delta_{ij}\eta_i/|\eta_i|, \tag{21}$$

and

$$(\zeta_i)^\dagger JHJ(\zeta_j) = \delta_{ij}|\eta_i|^{-1}. \tag{22}$$

Now rearrange the column vectors (ζ_i) into a new matrix

$$Z = ((z_1\cdots(z_n)(\hat{z}_1)\cdots(\hat{z}_n)) \tag{23}$$

such that the vectors (z_i) have positive norm, namely,

$$(z_i)^\dagger J(z_j) = \delta_{ij}, \qquad i,j = 1, \ldots, n, \tag{24a}$$

and the vectors $(\hat{z}_i)$ have negative norm

$$(\hat{z}_i)^\dagger J(\hat{z}_j) = -\delta_{ij}, \qquad i,j = 1, \ldots, n. \tag{24b}$$

Since Z satisfies

$$Z^\dagger JZ = J, \tag{25}$$

and

$$Z^\dagger JHJZ = H', \tag{26}$$

we see that $A = Z^\dagger$ is a solution to the diagonalization problem [Eqs. (5) and (11)]. We note, in passing, that the same proof applies to more general metrics: the diagonal matrix J could have any number $0 \le k \le 2n$ of positive elements and, correspondingly, $2n-k$ negative elements. Equation (26), with the aid of Eq. (25), can be rewritten in the form

$$(HJ)Z = Z(JH') = Z\begin{pmatrix}\gamma & \\ & -\gamma\end{pmatrix}. \tag{27}$$

Thus, (z_i) are eigenvectors of HJ with eigenvalues γ_i, and similarily, $(\hat{z}_i)$ are eigenvectors of HJ with eigenvalues $-\gamma_i$.

B. Construction

The actual construction of a diagonalizing matrix $Z = A^\dagger$ follows closely the procedure employed in the Hermitian case, the only difference being the change of the metric from $I = \begin{pmatrix}1 & \\ & 1\end{pmatrix}$ to $J = \begin{pmatrix}1 & \\ & -1\end{pmatrix}$. We shall make use of the following observations:

(a) The eigenvectors of J have nonvanishing norm. Let

$$HJ(x_i) = \gamma_i(x_i). \tag{28}$$

Since H is positive definite and J is regular,

$$(x_i)^\dagger JHJ(x_i) = \gamma_i(x)^\dagger J(x_i) > 0. \tag{29}$$

Thus $(x_i)^\dagger J(x_i) \gtrless 0$ according to $\gamma_i \gtrless 0$, and (x_i) can be normalized to satisfy

$$(x_i)^\dagger J(x_i) = \pm 1, \quad \gamma_i \gtrless 0. \tag{30}$$

(b) Eigenvectors belonging to different eigenvalues are orthogonal with respect to J. Indeed, if

$$HJ(x_i) = \gamma_i(x_i) \quad \text{and} \quad HJ(x_j) = \gamma_j(x_k), \tag{31}$$

then

$$(x_j)^\dagger JHJ(x_i) = \gamma_i(x_j)^\dagger J(x_i) = \gamma_j(x_j)^\dagger J(x_i). \tag{32}$$

(c) If γ_i is a multiple root of $\det(HJ - \gamma I)$ with multiplicity k, there exist exactly k linearly independent eigenvectors having γ_i as their common eigenvalue. Indeed, let Z satisfy Eqs. (25) and (26). Since the ranks of the two matrices $(HJ - \gamma_i I)$ and $Z^\dagger J(HJ - \gamma_i I)Z = H' - \gamma_i J$ are equal, and the rank of the latter is, obviously, $2n-k$, the homogeneous system $(HJ - \gamma_i I)(x) = (0)$ has exactly $2n - (2n-k) = k$ independent solutions.

(d) The k independent eigenvectors $(x_1), \ldots, (x_k)$ belonging to the degenerate eigenvalue γ_i, can be made orthonormal (with respect to J) via the Schmidt process. Assume, for definiteness, $\gamma_i > 0$. Normalize (x_1) and denote the resulting vector by (y_1). Now subtract from (x_2) its component along (y_1), that is,

$$(x_2') = (x_2) - (y_1)^\dagger J(x_2)\cdot(y_1). \tag{33}$$

Clearly, $(x_2') \neq (0)$ [or else (x_2) and (y_1) would be linearly dependent], and $(x_2')^\dagger J(y_1) = 0$. Since (x_2') is also an eigenvector of HJ (with eigenvalue $\gamma_i > 0$), we have by (a), $(x_2')^\dagger J(x_2') > 0$. Normalize (x_2') and denote the resulting vector by (y_2). The process can be continued to obtain $(y_1), \ldots, (y_k)$ satisfying

$$HJ(y_l) = \gamma_i(y_l), \quad \text{and} \quad (y_l)^\dagger J(y_m) = \delta_{lm}. \tag{34}$$

(e) If $(x_i) = \begin{pmatrix}u_i\\ v_i\end{pmatrix}$ is an eigenvector of HJ with eigenvalue γ_i, then $(\hat{x}_i) = \begin{pmatrix}v_i^*\\ u_i^*\end{pmatrix}$ is an eigenvector of HJ with eigenvalue $-\gamma_i$. Indeed, using the explicit form

$$\begin{pmatrix}\alpha & -\beta\\ -\beta^* & \alpha^*\end{pmatrix}\begin{pmatrix}u_i\\ v_i\end{pmatrix} = \gamma_i\begin{pmatrix}u_i\\ u_i\end{pmatrix},$$

the assertion is obvious. Since both γ_i and $-\gamma_i$ are eigenvalues,

$$\det(HJ - \gamma I) = \prod_{i=1}^{n}(\gamma^2 - \gamma_i^2). \tag{35}$$

We can now summarize the construction of the diagonalizing matrix $A^\dagger$.

(i) Find the n positive roots of the characteristic polynomial (35).

(ii) Obtain n independent eigenvectors $(x_1), \ldots, (x_n)$ belonging to the positive roots $\gamma_1, \ldots, \gamma_n$.

(iii) If necessary, use the Schmidt process to complete the orthonormalization, and denote the resulting vectors by (z_i). Thus, $(z_i)^\dagger J(z_j) = \delta_{ij}$.

(iv) The first half of $A^\dagger$, namely, $\binom{\lambda^\dagger}{\mu^\dagger}$ is given by

$$\binom{\lambda^\dagger}{\mu^\dagger} = ((z_1) \cdots (z_n)) \tag{36}$$

(v) Complete

$$A^\dagger = \begin{pmatrix} \lambda^\dagger & \tilde{\nu} \\ \mu^\dagger & \tilde{\chi} \end{pmatrix} = ((z_1) \cdots (z_n)(\hat{z}_1) \cdots (\hat{z}_n)), \tag{37}$$

where $(\hat{z}_i) = \binom{v_i^*}{u_i^*}$ for $(z_i) = \binom{u_i}{v_i}$.

ACKNOWLEDGMENT

The use of Sylvester's law of inertia in establishing the existence of a diagonalizing matrix was suggested by J. Stein. I am indebted to him for many helpful discussions.

APPENDIX

The purpose of the Appendix is to amend a shortcoming in a related work by the same author. In Ref. 2 the following problem is discussed: Given the diagonalizing matrix $A = \begin{pmatrix} \lambda & \mu \\ \mu^* & \lambda^* \end{pmatrix}$, find the transformation brackets connecting base states of the old operators $(a_i, a_i^\dagger)$ with those of the new operators $(b_i, b_i^\dagger)$. In order to evaluate a multidimentional normalization integral [Ref. 2, Eq. (76)], I have assumed, in addition to Eq. (5), the reality condition

$$(\lambda^\dagger \lambda) = (\lambda^\dagger \lambda)^*. \tag{A1}$$

It turned out that (A1) is too strong in the sense that it cannot always be met. Thus, for example, for $n = 2$, the two matrices

$$\lambda = \begin{pmatrix} \alpha & -i\alpha \\ -i\beta & \beta \end{pmatrix} \quad \text{and} \quad \mu = \begin{pmatrix} a & ia \\ b & -ib \end{pmatrix},$$

with α, β, a, b real, and $\beta^2 \neq \alpha^2$, $a^2 = \alpha^2 - \frac{1}{2}$, and $b^2 = \beta^2 - \frac{1}{2}$, fulfill Conditions (5) but violate (A1). Now the normalization integral can be written as

$$\text{Integral} = \int \frac{d^{2n}x}{\pi^n} \exp[-(\tilde{x})F(x)], \tag{A2}$$

where

$$F = \begin{pmatrix} 1 + \tau_R & \tau_I \\ \tau_I & 1 - \tau_R \end{pmatrix} \tag{A3}$$

is a real $2n$ by $2n$ symmetric matrix, with

$$\tau_R = \mathrm{Re}(\lambda^{-1}\mu) \quad \text{and} \quad \tau_I = \mathrm{Im}(\lambda^{-1}\mu). \tag{A4}$$

Let Q be a real orthogonal matrix diagonalizing F, that is,

$$Q\tilde{F}Q = f, \quad f_{ij} = \delta_{ij} f_i. \tag{A5}$$

Since the Jacobian of the transformation $(y) = Q(x)$ is equal to 1, the integral (A2) reduces to

$$\text{Integral} = \int \frac{d^{2n}y}{\pi^n} \exp\left(-\sum_{i=1}^{2n} f_i y_i^2\right) = \prod_{i=1}^{2n} \frac{1}{\pi^{1/2}} \int e^{-f_i y_i^2} dy_i$$
$$= \prod_{i=1}^{2n} f_i^{-1/2} = (\det F)^{-1/2}, \tag{A6}$$

provided all the eigenvalues f_i are positive. We shall now show that this is indeed the case. Consider the matrix

$$G = F - I = \begin{pmatrix} \tau_R & \tau_I \\ \tau_I & -\tau_R \end{pmatrix}, \tag{A7}$$

with eigenvalues $g_i = f_i - 1$. The assertion will be established if $g_i^2 < 1$ for $i = 1, \ldots, 2n$. Let $\binom{x}{y}$ be a normalized eigenvector of G with eigenvalue g. Then

$$G^2 \binom{x}{y} = \begin{pmatrix} B & C \\ -C & B \end{pmatrix} \binom{x}{y} = g^2 \binom{x}{y}, \tag{A8}$$

where

$$F = \tau_R^2 + \tau_I^2 \quad \text{and} \quad C = \tau_R \tau_I - \tau_I \tau_R. \tag{A9}$$

But $(z) = (x) + i(y)$ also satisfies

$$\tau\tau^\dagger(z) = (F - iC)(z) = g^2(z), \tag{A10}$$

where

$$\tau = \tau_R + i\tau_I. \tag{A11}$$

Hence, by Eq. (5),

$$g^2 = (z)^\dagger \tau\tau^\dagger (z) = 1 - (z)^\dagger (\lambda^\dagger \lambda)^{-1} (z) < 1. \tag{A12}$$

The result (A6) is, therefore, always valid (when A exists), and should replace Eq. (80) of Ref. 2.

[1] C. Tsallis, J. Math. Phys. **19**, 277 (1978).
[2] Y. Tikochinsky, J. Math. Phys. **19**, 270 (1978).
[3] See, for example, M. Böcher, *Introduction to Higher Algebra* (Macmillan, New York, 1915), p. 144.

Conformal Killing horizons

C. C. Dyer

Scarborough College and David Dunlap Observatory, University of Toronto, West Hill, Ontario, Canada

E. Honig

Scarborough College and Department of Physics, University of Toronto, West Hill, Ontario, Canada
(Received 7 March 1978)

We introduce the concept of a conformal stationary limit surface as the boundary where a conformal Killing vector admitted by a spacetime becomes null. This hypersurface is an infinite frequency shift surface for conformal Killing observers and sources. We derive the local conditions under which it can be an event horizon in the sense that it be a null geodesic hypersurface. In particular, we show that it is null and geodesic if and only if the rotation of the conformal Killing congruence has a vanishing norm on this hypersurface. Finally, we discuss the surface gravity associated with such a conformal Killing horizon.

1. INTRODUCTION

It is over twenty years since Rindler first differentiated between particle and event horizons in a classic paper.[1] While he defined a horizon quite generally as a boundary between things that are visually observable and things that are not, his analysis was based entirely on the Robertson–Walker metric. As such it includes all homogeneous and isotropic model universes. With the advent of black hole physics, the concept of event horizon was adapted to include those horizons occurring in asymptotically flat spacetimes.[2,3] The most important spacetimes of this type are the well-studied solutions of Schwarzschild and Kerr, both of which admit a timelike Killing vector that becomes null on a totally geodesic null hypersurface, the event horizon.

In this paper we consider the properties of spacetimes admitting a timelike conformal Killing vector (CKV), ξ^a, which then satisfies the conformal Killing equations[4]

$$\mathcal{L}_\xi g_{ab} \equiv \nabla_a \xi_b + \nabla_b \xi_a = 2\phi g_{ab}. \tag{1}$$

Here $\mathcal{L}_\xi$ is the Lie derivative, $\phi(x^c)$ is the scalar function defined by $4\phi = \nabla_c \xi^c$, g_{ab} is the metric tensor, and ∇_a is its associated covariant derivative. Consider now a null geodesic with tangent vector k^a, so that $k^a k_a = 0$ and $k^a \nabla_a k^b = 0$. It then follows from Eq. (1) that $k^a \nabla_a (\xi^b k_b) = 0$, so that $\xi^a k_a$ is constant along the null geodesic.

It is useful to define a "conformal Killing observer," with 4-velocity $u^a = e^{-\psi}\xi^a$, in regions where ξ^a is timelike, having defined $e^{2\psi} = \xi^a \xi_a$. We use a signature -2, in which case $u^a u_a = +1$ when ξ^a is timelike. From Eq. (1), it then follows that the function ϕ is just ψ' when ψ is well defined, where we write $' \equiv \xi^a \nabla_a$, the covariant derivative along the CK trajectory defined by ξ^a. Thus ϕ just tells us the rate of scale change as we move along the CK trajectory. In the case of a Killing vector, for which $\phi = 0$, it follows that $\xi^a \xi_a$ remains constant along the Killing trajectory.

Consider now a CK source at S, emitting photons at frequency ν_s, joined by a null geodesic, Γ, to a CK observer at 0 who receives the photons at frequency ν_0. Since the measured frequency is proportional to $u^a k_a$, and $\xi^a k_a$ is constant along Γ, it follows that the photons are frequency shifted by the factor

$$\frac{\nu_s}{\nu_0} = \frac{(\xi^a \xi_a)_0^{1/2}}{(\xi^a \xi_a)_s^{1/2}}. \tag{2}$$

Hence the hypersurface defined by $\xi^a \xi_a = 0$, which we shall designate Σ_0, is a surface of infinite frequency shift for CK sources and observers, for S on Σ_0 implies $\nu_0 = 0$ while 0 on Σ_0 implies $\nu_0 \to \infty$ for finite ν_s.[5]

The existence of the CKV, ξ^a, implies that the metric can be written in the form $g_{ab} = f(x^c)\, h_{ab}$, where $f(x^c)$ is a function of spacetime position that is related to $\phi(x^c)$, while h_{ab} is a function of position with the restriction that it be constant on any CK trajectory.[4] When a CK observer can be defined, i.e., when ξ^a is timelike, this observer experiences a purely conformal change in g_{ab} as he moves along the CK trajectory. If the CK congruence is irrotational, he concludes that the spacetime is "conformally static"; otherwise he concludes that the spacetime is "conformally stationary" for he can detect rotation. In regions where the CKV is not timelike, such an observer cannot exist. We are then led to term the boundary, where the CKV becomes null, to be a "conformal stationary limit surface." Even for spacetimes having an ordinary Killing vector, it appears appropriate to distinguish between static and stationary limit surfaces.

In the following sections we consider the circumstances under which Σ_0, the conformal stationary limit surface, can form part of an event horizon, in the sense that it be a null geodesic hypersurface. It should be noted that some of our results could be obtained by means of a conformal transformation since a null geodesic hypersurface is invariant under such a transformation and the work of Takeno[6] directly implies that given a CKV in one space, there is generally a large class of spaces that admit this vector as a KV.[7] On the other hand, the concept of surface gravity is not easily developed with this latter approach because timelike geodesics do not transform to timelike geodesics. The direction that we follow here, however, allows us to introduce this concept in a clear physical manner. Furthermore, it demonstrates how significantly different the CK case is from the ordinary Killing case in terms of horizon degeneracy, surface gravity, and surface temperature.

 0022-2488/79/030409-04$01.00

2. CONFORMAL KILLING CONGRUENCES

To study the geometry of a congruence of conformal Killing trajectories, it is useful to expand the covariant derivative, $\nabla_a u_b$, of the unit tangent vector in the usual form[8]

$$\nabla_a u_b = u_a \dot{u}_b + \omega_{ab} + \sigma_{ab} + \theta h_{ab}, \tag{3}$$

where $\theta = \frac{1}{3}\nabla_a u^a$ is the expansion, $\dot{u}_a = u^c \nabla_c u_a$ is the acceleration, $\omega_{ab} = \nabla_{[a} u_{b]} - u_{[a}\dot{u}_{b]}$ is the vorticity, $\sigma_{ab} = \nabla_{(a} u_{b)} - u_{(a}\dot{u}_{b)} - \theta h_{ab}$ is the traceless shear, and $h_{ab} = g_{ab} - u_a u_b$ is the projection tensor. Using Eq. (1), it is straightforward to show that the CK congruence is shearless[9] and has expansion $\theta = \dot{\psi} = \phi e^{-\psi}$, where in general $\dot{} \equiv u^a \nabla_a$. The acceleration can be written as $\dot{u}_a = \theta u_a - \nabla_a \psi$. We can then write Eq. (3) in the form

$$\nabla_a u_b = e^{-\psi}\phi g_{ab} + \omega_{ab} - u_a \nabla_b \psi, \tag{4}$$

from which we also have

$$\nabla_a \xi_b = \phi g_{ab} - \xi_a \nabla_b \psi + \xi_b \nabla_a \psi + e^{\psi}\omega_{ab}. \tag{5}$$

Forming $\xi_a \nabla_b \xi_c$, permuting and summing we find

$$\xi_{(a}\nabla_b \xi_{c)} = \phi \xi_{(a} g_{bc)} + e^{\psi}\xi_{(a}\omega_{bc)}, \tag{6}$$

where $\langle \cdots \rangle$ indicates the sum of distinct cyclic permutations. In the case that $\omega_{ab} = 0$, this equation is just the usual hypersurface orthogonality condition, $\xi_{[a}\nabla_b \xi_{c]} = 0$, combined with the conformal Killing Eq. (1).

From Eq. (1), we can obtain a particularly useful expression for the acceleration

$$\dot{u}^a = F^{ab} u_b, \quad \text{where } F_{ab} = e^{-\psi}(\phi g_{ab} - \nabla_a \xi_b). \tag{7}$$

Since we can also write F_{ab} as $e^{-\psi}\nabla_{[a}\xi_{b]}$, it follows that F_{ab} is a bivector, whose dual, $\hat{F}_{ab}$ is given by

$$\hat{F}_{ab} = \tfrac{1}{2}\eta_{abrs} F^{rs}, \tag{8}$$

where η_{abrs} is the alternating tensor. The rotation vector associated with u^a is

$$\omega^r = \tfrac{1}{2}\eta^{rabc} u_a \nabla_b u_c, \tag{9}$$

which is related to the vorticity through $\omega_{ab} = \eta_{abrs} u^r \omega^s$. In terms of the bivector F_{ab}, we have the more convenient form of the rotation, $\omega^a = -\hat{F}^{ab} u_b$.

Since we shall be interested in regions where ξ^a may become null, so that u^a is not defined, it is useful to introduce a renormalized rotation vector, $\bar{\omega}^r$, given by

$$\bar{\omega}^r = \tfrac{1}{2}\eta^{rabc}\xi_a \nabla_b \xi_c, \tag{10}$$

which remains defined on Σ_0. It follows that $\bar{\omega}^a = e^{2\psi}\omega^a$ when ξ^a is not null, and that $\bar{\omega}^a = -e^{2\psi}\hat{F}^{ab}u_b$, as above.

The squared norm of this renormalized rotation vector is

$$\bar{\omega}^a \bar{\omega}_a = e^{4\psi}\hat{F}^{ab}\hat{F}_{ac} u_b u^c. \tag{11}$$

We can now use the identity[10]

$$A^{ab}B_{bc} - \hat{B}^{ab}\hat{A}_{bc} = \tfrac{1}{2}\delta^a_c A_{rs} B^{sr}, \tag{12}$$

where A_{ab} and B_{ab} are any two bivectors with duals $\hat{A}_{ab}$ and $\hat{B}_{ab}$ in a four-dimensional Riemannian space. Taking both A_{ab} and B_{ab} to be F_{ab}, we then have

$$\hat{F}^{ab}\hat{F}_{ac} = F^{ab}F_{ac} - \tfrac{1}{2}\delta^b_c F_{rs}F^{rs}, \tag{13}$$

and from Eq. (7) we have

$$F_{rs}F^{rs} = e^{-2\psi}(\nabla^a \xi^b \nabla_a \xi_b - 4\phi^2). \tag{14}$$

Equations (7) and (11) then yield

$$\bar{\omega}^a \bar{\omega}_a e^{-2\psi} = \dot{u}^a \dot{u}_a e^{2\psi} + 2\phi^2 - \tfrac{1}{2}\nabla^a \xi^b \nabla_a \xi_b. \tag{15}$$

We now consider a hypersurface, Σ, defined by $\xi^a \xi_a = \text{constant}$. A normal to Σ can be written as $n_a = \frac{1}{2}\nabla_a(\xi^c \xi_c)$, so that $n_a = \xi^c \nabla_a \xi_c$. It then follows from Eq. (1) that $\xi^a n_a = \phi \xi^a \xi_a$, and also that

$$n_a = e^{\psi}(\phi u_a - e^{\psi}\dot{u}_a). \tag{16}$$

Of course $\dot{u}^a u_a = 0$, since $u^a u_a = 1$, so that we have for the norm of the normal vector

$$n^a n_a = e^{2\psi}(\phi^2 + e^{2\psi}\dot{u}^a \dot{u}_a). \tag{17}$$

Then using Eq. (15) to eliminate $\dot{u}^a \dot{u}_a$, we have the fundamental equation

$$\bar{\omega}^a \bar{\omega}_a - n^a n_a = \xi^a \xi_a (\phi^2 - \tfrac{1}{2}\nabla^a \xi^b \nabla_a \xi_b). \tag{18}$$

This is the generalization to CKV's of a well-known equation for KV's, obtained first by Vishveshwara[2] and later in a somewhat different manner by the application of the Frenet–Serret formalism to Killing trajectories.[11] The corresponding equation for ξ^a a homothetic Killing vector has recently been obtained, where the Frenet–Serret equations are integrated explicitly,[12] similarly to Ref. 11. Equation (18) states that the conformal stationary limit surface, Σ_0, where ξ^a becomes null, is a null hypersurface if and only if $\bar{\omega}^a$ is also null (or the zero vector) on Σ_0.

3. THE CONFORMAL KILLING HORIZON

For any hypersurface to constitute a portion of an event horizon, it is necessary that it be both null and geodesic. Of course there are further global conditions to be considered, but these are not of direct interest here. In the case of Schwarzschild spacetime, the static limit surface for time-translational Killing observers, where ξ^a becomes null, coincides with the event horizon. This is not the case for the Kerr spacetime due to the nonvanishing rotation associated with the time translational Killing vector.[2,3] As we have already pointed out, a similar conclusion is true for a CKV with rotation.

Hence we are faced with two options, the first being to assume that the rotation of ξ^a vanishes. Failing this, we shall assume that there exists another CKV from which, with ξ^a, we can form a "mixed CKV," say $\tilde{\xi}^a$, which has null or vanishing rotation on the stationary limit surface defined by $\tilde{\xi}^a \xi_a = 0$. This is analogous to the "mixed KV" introduced in the Kerr spacetime by Vishveshwara[2] and Carter.[3] In the following, we shall drop the tilde and simply denote the mixed CKV by ξ^a.

From the definition of the normal, n_a to Σ, and Eq. (1) we can write

$$n_a = 2\phi \xi_a - \xi'_a. \tag{19}$$

On the hypersurface Σ_0, we have both ξ^a and n^a null, with $\xi^a n_a = 0$ as well. Since two nonzero orthogonal null vectors must be parallel, it follows that $n_a = \beta \xi_a$ for some scalar function $\beta(x^c)$. (In a similar way, it is ob-

vious that, on Σ_0, $\bar{\omega}^a$ is also parallel to ξ_a, unless it is the zero vector.) Equation (19) then implies that $\xi'_a = h\xi_a$ for some scalar function $h(x^c)$, in which case $n_a = (2\phi - h)\xi_a$. Hence both ξ^a and n^a are parallel to a null geodesic tangent vector, and it follows that the hypersurface where both $\xi^a\xi_a$ and $\bar{\omega}^a\bar{\omega}_a$ vanish is null and geodesic.

The hypersurface Σ_0 then satisfies the local conditions that it constitute a portion of an event horizon. That more, global, conditions are necessary is obvious from a consideration of the Minkowski and Robertson–Walker spacetimes, both having the 15 parameter full conformal group. Some of their CKV's certainly define hypersurfaces which are null and geodesic, forming the past and future null cones of any typical observer. However they do not define a global horizon, which is consistent with our notion of the homogeneity properties of these spacetimes. Of course certain limiting null cones do define particle and event horizons, and these have been discussed by Rindler.[1]

4. SURFACE GRAVITY

We have shown that $\xi'_a = h\xi_a$ on the hypersurface Σ_0. Of course this may also be so away from Σ_0, so that the CK trajectory is a nonnull geodesic, and for the timelike case represents free fall. This is illustrated by the HKV of Minkowski spacetime which has $h = \phi = 1$. Since $\xi'_a = h\xi_a$ implies $(h - \phi)\xi^a\xi_a = 0$, then imposing $\xi'_a = h\xi_a$ with $h \neq \phi$ implies that ξ^a, and hence n^a and $\bar{\omega}^a$, are null and parallel. In our consideration of horizons we are only interested in Σ_0 and restrict ourselves to this hypersurface.

Let λ be the parameter along the CK trajectory such that $\xi^a = dx^a/d\lambda$. (λ is the group parameter related to ξ^a.) If μ is the affine parameter along this curve, i.e., such that $(dx^a/d\mu)\nabla_a(dx^c/d\mu) = 0$, then the function h is just $(d^2\lambda/d\mu^2)/(d\lambda/d\mu)$, relating the two parameters by a "relative acceleration." To derive an expression for h, we form $\xi'_a\xi'_b = h^2\xi_a\xi_b$, and noting that $\xi'_a\xi_b = \xi_a\xi'_b$ on Σ_0, we obtain

$$h^2\xi_b = (\nabla^a\xi^c)\xi_a\nabla_c\xi_b \quad \text{on } \Sigma_0. \tag{20}$$

Using Eqs. (1) and (6), we arrive at the relation on Σ_0

$$(h - \phi)^2 = 2\phi^2 - \tfrac{1}{2}\nabla^a\xi^b\nabla_a\xi_b. \tag{21}$$

To introduce the notion of surface gravity for the horizon, we must consider an acceleration. Since $\dot{u}^a$ becomes undefined on Σ_0, we consider a renormalized acceleration, $e^{\psi}\dot{u}^a$, which remains well behaved on Σ_0, analogous to the renormalized rotation $\bar{\omega}^a$. In general, we have the relation

$$\xi^{a'} = \phi\xi^a + e^{2\psi}\dot{u}^a, \tag{22}$$

so that $e^{2\psi}\dot{u}^a\dot{u}_a = (h - \phi)^2$ on Σ_0. Defining g, the surface gravity, to be the norm of the renormalized acceleration, it follows that

$$g^2 = 2\phi^2 - \tfrac{1}{2}\nabla^a\xi^b\nabla_a\xi_b \tag{23}$$

on the hypersurface Σ_0.[13] We can also use this as a definition of g off Σ_0, and write Eq. (18) in the form

$$\bar{\omega}^a\bar{\omega}_a - n^a n_a = (g^2 - \phi^2)\xi^a\xi_a. \tag{24}$$

It is of some interest to consider the variation of g along each CK trajectory generating the null hypersurface Σ_0. We make use of the identity for CKV's

$$\nabla_c\nabla_a\xi_b = \xi^d R_{dcab} + g_{ab}\nabla_c\phi + g_{bc}\nabla_a\phi - g_{ac}\nabla_b\phi, \tag{25}$$

where R_{dcab} is the Riemann curvature tensor. This is derived in the same manner as the more familiar identity for Killing vectors. Transvecting with ξ^c, it follows that

$$(\nabla_a\xi_b)' = \phi' g_{ab} - \xi_a\nabla_b\phi + \xi_b\nabla_a\phi, \tag{26}$$

where the Reimann term vanishes due to the pair antisymmetry of R_{dcab}. By straightforward differentiation of Eq. (23) along the CK trajectory, using Eq. (26), we find

$$(g^2 + \phi^2)' = 2\xi^{a'}\nabla_a\phi. \tag{27}$$

On the hypersurface Σ_0, where $\xi^{a'} = h\xi^a$ and $g^2 = (h - \phi)^2$, we then have the results

$$(h - 2\phi)' = 0 \quad \text{and} \quad (g - \phi)' = 0 \quad \text{on } \Sigma_0, \tag{28}$$

where we adopt the sign convention $g = h - \phi$. For ϕ constant, i.e. ξ^a is a homothetic Killing vector, it follows that the surface gravity remains constant along each of the null HK curves generating Σ_0, while for a CKV, the surface gravity scales up or down with ϕ along the CK curves. In the latter case ϕ, and hence g, is not necessarily constant on Σ_0. This can be seen most directly from a consideration of Minkowski spacetime, where there are four CKV's with ϕ being proportional to t, x, y, or z (Cartesian coordinates). These ϕ's are obviously not constant on Σ_0, which is in this case just the null cone. Of course this Σ_0 does not define a global horizon in this spacetime. Since $g - \phi$ (or $h - 2\phi$) is the conserved quantity along the null geodesic CK generators of Σ_0, one would also like to consider the variation of $g - \phi$ from one generator of Σ_0 to another, in analogy with the generalized Hawking–Lichnerowicz theorem,[14] where the surface gravity was found to be constant over any connected component of a Killing horizon. The constancy of $g - \phi$, suggests that it is $g - \phi$, and not g, to which the temperature is proportional for such possible horizons.[15]

Following Boyer,[16] it is of interest to consider the possibility of the horizon becoming degenerate in the sense that $\nabla_a(\xi^c\xi_c)$, i.e., n_a, becomes the zero vector. Since $n^a = (\phi - g)\xi^a$, and $(g - \phi)' = 0$, it follows that if the horizon is degenerate at any point along a CK generator, then it must be degenerate along the entire generator. Unlike the Killing horizon case, the surface gravity is not zero, being just ϕ, which itself might not be constant on the generator.

Equations (18), (23), and (28) are generalizations of now familiar results for KV's and their associated horizons, and should be of use in considering a wider range of models for dynamic problems, such as black or white holes or for models of locally inhomogeneous universes. One simple application of these results is to the self-similar spherically symmetric spacetimes considered by Carr and Hawking,[17] and other authors.[18] The existence of the horizon structure in these spacetimes is most easily understood in terms of the existence of the similarity homothetic Killing vector.[19]

ACKNOWLEDGMENTS

We wish to thank Dr. H. C. Corben, Dr. K. Lake, Dr. R. G. McLenaghan, Dr. R. C. Roeder, Dr. C. V. Vishveshwara, and D. Toms for stimulating discussion. We acknowledge the financial support of the National Research Council of Canada.

[1]W. Rindler, Mon. Not. R. Astron. Soc. **116**. 662 (1956).
[2]C. V. Vishveshwara, J. Math. Phys. **9**, 1319 (1968).
[3]B. Carter, J. Math. Phys. **10**. 70 (1969). See also B. Carter, *Black Hole Physics*, *Les Houches* 1972, edited by C. DeWitt and B. S. DeWitt (Gordon and Breach, New York, 1973).
[4]K. Yano, *The Theory of Lie Derivatives and its Applications* (North-Holland, Amsterdam, 1957).
[5]The constancy of $\xi^a k_a$ along a null geodesic, as well as its application to frequency shift was noted by P. Jordan, J. Ehlers, and W. Kundt, Akad. Wiss. Lit., Mainz, Abh. Math. Naturwiss. Kl. **2** (1960). The question of infinite frequency shift surfaces was not raised in their work.
[6]H. Takeno, *Theory of Spherically Symmetric Space—Times* (Hiroshima, 1966), p. 174, revised ed.
[7]The fundamental papers of D. M. Eardley, Commun. Math. Phys. **37**, 287 (1974) and L. Defrise—Carter, Commun. Math. Phys. **40**, 273 (1975) are more general than necessary for a conformal transformation approach to our problem since they discuss the full group of conformal (homothetic) motions and its relation to the isometry group of a conformally related spacetime. While Takeno' results (Ref. 6) would suffice, it should be noted that none of these authors mentions the possibility of horizons being related to the existence of a CKV or an HKV in a physical spacetime. Furthermore, since Eardley primarily restricted himself to spacelike slices admitting a spacelike HKV, this question would not have arisen in his work.
[8]See for example, S. Hawking and G. Ellis, *The Large Scale Structure of Spacetime* (Cambridge U. P., Cambridge, 1973).
[9]D. R. Oliver, Jr. and W. R. Davis, Gen. Rel. Grav. **8**, 905 (1977).
[10]H. S. Ruse, Proc. London Math. Soc. **41**. 302 (1936).
[11]E. Honig, E. L. Schücking, and C. V. Vishveshwara. J. Math. Phys. **15**, 774 (1974).
[12]C. C. Dyer and E. Honig, J. Math. Phys. **20**, 1 (1979).
[13]In the general n-dimensional case Eq. (23) should be replaced by $g^2 = \frac{1}{2}(n\phi^2 - \nabla^a\xi^b\nabla_a\xi_b)$.
[14]B. Carter, *Black Hole Physics* (Gordon and Breach, New York, 1973), pp. 155 and 167.
[15]Work in progress with K. Lake.
[16]R. H. Boyer, Proc. Roy. Soc. A **311**, 245 (1969). See also Ref. 14, p. 148.
[17]B. Carr and S. Hawking, Mon. Not. R. Astron. Soc. **168**, 399 (1974).
[18]M. Cahill and A. Taub, Commun. Math. Phys. **21**, 1 (1971); D. Lin, B. Carr, and S. Fall, Mon. Not. R. Astron. Soc. **177**, 51 (1977); R. Henriksen and P. Wesson, Astrophys. Space Sci. **53**, 429 (1978).
[19]C. C. Dyer, submitted to Mon. Not. R. Astron. Soc.

Linearized analysis of inhomogeneous plasma equilibria: General theory

H. Ralph Lewis

Los Alamos Scientific Laboratory, Los Alamos, New Mexico 87545

Keith R. Symon

Department of Physics, University of Wisconsin, Madison, Wisconsin 53706
(Received 23 March 1978)

A generalized framework is presented for analyzing the linearized equations for perturbations of inhomogeneous plasma equilibria in which there is a collisionless species, some properties of the solutions of the linearized equations are described, and a basis is provided for numerical computations of the linearized properties of such equilibria. It is useful to expand the perturbation potentials in eigenfunctions of the field operator which appears in the linearized equations, and to define a dispersion matrix whose analytical properties determine the nature of the solutions of the initial-value problem. It is also useful to introduce auxiliary functions to replace the usual perturbation distribution functions, and to expand the auxiliary functions in eigenfunctions of the equilibrium Liouville operators. By introducing the auxiliary functions, great freedom is achieved in the choice of the field operator which appears in the linearized equations. This freedom can be used in some problems to define expansion functions for the potentials that are particularly suitable for studying specific normal modes.

I. INTRODUCTION

Although plasmas produced in the laboratory are always inhomogeneous and finite in extent, practical techniques have not been well developed for the theoretical determination of the linear response of a collisionless plasma to a small deviation from an inhomogeneous equilibrium. There is general interest in the solutions of the linearized equations for the evolution of inhomogeneous collisionless plasmas, and in useful techniques for approximating those solutions. There is also considerable practical interest in understanding kinetic effects in some current experiments that involve quite inhomogeneous plasmas at least one of whose particle species can be taken as collisionless. Examples of such experiments are highly diamagnetic pinches containing hot ions.[1,2]

The objective of this paper is to present a general framework for linearized analysis of inhomogeneous collisionless plasma equilibria, to describe some properties of the solutions of the linearized equations, and to provide a basis for obtaining approximate numerical solutions in a practicable way.

Generally speaking, the perturbations of the potentials or of the field functions appropriate to the problem at hand are represented as a linear combination of some basis functions of position with time-dependent coefficients. Then the evolution of the system can be described in terms of a dispersion matrix whose rows and columns are associated with the various basis functions. For example, in the case of a spatially uniform equilibrium of a one-dimensional electrostatic plasma, it is usual and appropriate to express the scalar potential as a Fourier series. Then the dispersion matrix is diagonal, with the result that each wavenumber in the perturbation can be analyzed separately. However, were the equilibrium spatially nonuniform, using a Fourier series for the potential would not yield a diagonal dispersion matrix and each wavenumber could not be analyzed separately. It would be desirable to find a set of basis functions with which the dispersion matrix would be diagonal, or nearly so for the most important eigenfrequencies of the problem. The necessity of choosing a suitable set of basis functions for expanding the perturbation potentials is a central aspect of a linearized analysis for any inhomogeneous equilibrium. The optimal choice is a matter of physical intuition and experience, and it depends on what information is desired in the specific problem at hand. In the treatment that we present here, great freedom is allowed in the choice of basis functions for the perturbation potentials. This is achieved by replacing the perturbation distribution function for each particle species with an auxiliary function, which is the perturbation distribution function for that species plus a linear functional of the perturbation potentials. By this means, an arbitrary operator is introduced into the equation satisfied by the perturbation potentials, and the basis functions are taken as eigenfunctions of that operator.

If it is possible to find the auxiliary functions or, equivalently, the perturbation distribution functions as functionals of the perturbation potentials exactly, then it may not be useful to expand the auxiliary functions. However, this is usually not the case. Generally, it is also necessary to introduce basis functions for the phase space, functions of position and velocity or of position and canonical momentum, which are used to expand the auxiliary functions. To expand the auxiliary function for a particular particle species, we have found it advantageous for both analytical and numerical purposes to take basis functions for the phase space that are simultaneous eigenfunctions of the equilibrium Liouville operator and the equilibrium constants of the motion for that species. This is standard in the case of a spatially uniform equilibrium of an electrostatic plasma, although with that problem the procedure is usually described in different terms. By introducing the eigenfunctions of the equilibrium Liouville operator

 0022-2488/79/030413-24$01.00

in the general nonuniform case, we achieve a unified point of view that is very useful.

Techniques and results that we describe in this paper are being applied to the so-called Vlasov-fluid model, in which the ions are treated as collisionless and the electrons are approximated as a massless fluid[3,4]; they are also being used for investigating the stability of one-dimensional BGK equilibria.[5] The ideas have evolved from these applications and from earlier computational work on linearized behavior of perturbations of spatially uniform equilibria of collisionless plasmas in one dimension.[6] Buneman and Holdren[7] have used similar methods independently. Sedlaček[8] has investigated the spectral properties of the Liouville operator.

In addition to its application to specific problems discussed in the preceding paragraph, our method gives an insight into some of the general properties of the linearized solutions of the initial-value problem in the neighborhood of an inhomogeneous equilibrium. We give a general formula, Eq. (IV.24) or Eq. (IV.26), for the dispersion matrix $D(\omega)$ which determines the character of the linearized solutions. The normal modes occur at complex frequencies ω where $D(\omega)$ has a zero eigenvalue; the form of the normal mode is determined by the corresponding eigenvector. For a completely Hamiltonian system, the form (IV.26) of $D(\omega)$ shows that it is a formally Hermitian matrix function of ω. From this it follows that the normal frequencies ω occur in conjugate pairs. The Hermitian property also implies a variational theorem for the normal frequencies. The form (IV.24) or (IV.26) shows that the elements of $D(\omega)$ are analytic except for branch points on the real axis at multiples of extreme frequencies in the unperturbed particle motions. Because of the cut along the real axis, the general linearized solution consists of a continuum of van Kampen modes with real frequencies in addition to possible discrete normal modes. In the stable case, where there are only van Kampen modes, the asymptotic form of the solution after long times is dominated by the character of $D(\omega)$ in the neighorhood of the branch points. This is in contrast to the homogeneous problem where the long-time solution may be dominated by a Landau-damped mode. Landau-damped modes can occur also in the inhomogeneous case, but there are always real branch points which dominate the asymptotic behavior. The present paper provides an initial approach to the study of general features of this sort for the inhomogeneous initial-value problem.

In Sec. II we give the fundamental linearized equations for the perturbation distribution functions and potentials in the general case, define the class of auxiliary functions that we shall consider, and give the basic equations that determine the auxiliary functions and the perturbation potentials. These basic equations are also specialized to the important case in which the particle equations of motion and the equations for the potentials are derivable from the same Hamiltonian. The Laplace transform solutions for the auxiliary functions and potentials are obtained in Sec. III in two forms, one of which is of particular significance for numerical computations. Expansions of the auxiliary functions and perturbation potentials are introduced in Sec. IV, and general properties of the dispersion matrix that enters the Laplace-transform solutions are derived. The important special case in which there is at most one nonignorable coordinate in the equilibrium is treated in more detail. Properties of the solution of the initial-value problem are discussed in Sec. V.

In Appendix A we discuss Hamilton's principle and give the Lagrange and Hamilton equations in a vector form. Three examples of systems to which the formalism of this paper can be applied are displayed in Appendix B: a one-dimensional electron gas, a three-dimensional completely collisionless multispecies plasma, and the Vlasov-fluid model. We present a simple general proof of the anti-Hermiticity of the equilibrium Liouville operator in Appendix C. We discuss the eigenfunctions of the equilibrium Liouville operator in Appendix D, where we also obtain the eigenfunctions and eigenvalues explicitly for the case in which there is at most one nonignorable coordinate in the equilibrium. Finally, in Appendix E, we consider some properties of analytic matrix functions.

II. THE FUNDAMENTAL EQUATIONS

In order to encompass the variety of situations to which the formalism in this paper is applicable, we begin by writing the basic equations in a general form. We consider a plasma which can be described by a linearized Boltzmann equation for each collisionless species s,

$$\left[\frac{\partial}{\partial t}+L_s\right] f_s^{(1)}(t)=\sum_i U_{si}(t)\,\phi_i^{(1)}(t) = U_s(t)\,\phi^{(1)}(t), \tag{II.1}$$

where t denotes time, and a set of field equations of the form

$$K(t)\,\phi^{(1)}(t)=\sum_s J_s^\dagger f_s^{(1)}(t). \tag{II.2}$$

The quantities L_s, U_s, K, and $J_s^\dagger$ are linear operators, and † denotes adjoint. $f_s^{(1)}$ is the perturbation of a single-particle distribution function f_s for species s about an equilibrium distribution function $f_s^{(0)}$:

$$f_s=f_s^{(0)}+f_s^{(1)}. \tag{II.3}$$

$\phi^{(1)}$ is the perturbation of an array ϕ of potential functions about an equilibrium array $\phi^{(0)}$:

$$\phi=\phi^{(0)}+\phi^{(1)}. \tag{II.4}$$

For example, the elements $\phi_i^{(1)}$ of $\phi^{(1)}$ could be the perturbation of the scalar potential and components of the perturbation of the vector potential for the electromagnetic field; or they might be any linear functionals of the perturbation electromagnetic potentials. The equilibrium quantities $f_s^{(0)}$ and $\phi^{(0)}$ do not depend on time.

In the general theory it often does not matter exactly what independent variables are chosen in addition to time, and it is often convenient to leave them unspecified. Accordingly, we use the notational convention that any given quantity will be denoted by a single symbol regardless of which independent variables are used and regardless of how the quantity is represented. When

we wish to specify a definite set of independent variables in terms of which a quantity is to be expressed, we shall include those variables as arguments with the symbol for the quantity. When an argument is indicated for an operator, it may mean that the operator involves one or more functions of the argument; or it may mean that the operator is a differential or integral operator with respect to the argument; or both of these meanings may be implied. It is frequently useful to represent a function as a column matrix whose elements are the coefficients for the expansion of the function as a linear combination of linearly independent basis functions. The corresponding matrix representation of an operator is a rectangular matrix. We shall use the same symbol to denote the matrix representation of a function or an operator that we use to denote the function or operator itself. These notational conventions will expedite our presentation. Sometimes, as in Eqs. (II.1)–(II.4), we shall indicate the argument t explicitly but suppress all other arguments.

The quantity f_s may be a distribution function for the phase space of position and canonical momentum or for the phase space of position and velocity. For most cases of interest, the two kinds of distribution function are proportional with a constant numerical factor of proportionality. However, this is not always true. Whichever kind of distribution function f_s is, it can be expressed in terms of position coordinates and velocity components or in terms of position coordinates and canonical momenta equally well.

The operator $\mathcal{L}_s$ is the equilibrium Liouville operator for species s. That is, the operator $[\partial/\partial t + \mathcal{L}_s]$ acting on an explicit function of time and phase-space variables gives the total time rate of change of the function assuming that the phase-space variables change with time according to the equilibrium particle equations of motion for species s. Expressed in terms of position and velocity, $\mathcal{L}_s$ is

$$\mathcal{L}_s(\mathbf{r},\mathbf{v}) = \mathbf{v}\cdot\nabla_{\mathbf{r}} + \frac{Q_s}{M_s}\left[\mathbf{E}^{(0)}(\mathbf{r}) + \frac{1}{c}\mathbf{v}\times\mathbf{B}^{(0)}(\mathbf{r})\right]\circ\nabla_{\mathbf{v}} + \frac{1}{M_s}\mathbf{F}_s^{(0)}(\mathbf{r},\mathbf{v})\circ\nabla_{\mathbf{v}}, \tag{II.5}$$

where $\mathbf{E}^{(0)}(\mathbf{r})$ and $\mathbf{B}^{(0)}(\mathbf{r})$ are the equilibrium electric and magnetic fields, Q_s and M_s are the charge and mass per particle of species s, c is the speed of light, and $\mathbf{F}_s^{(0)}(\mathbf{r},\mathbf{v})$ is any force acting in equilibrium on a particle of species s in addition to the Lorentz force. We assume that $\mathbf{F}_s^{(0)}(\mathbf{r},\mathbf{v})$ is derivable from a generalized potential. We also assume that the array of potential functions ϕ has been so chosen that $\mathbf{E}^{(0)}$, $\mathbf{B}^{(0)}$, and $\mathbf{F}_s^{(0)}$ can be determined from $\phi^{(0)}$. The operator U_s is an array, of dimension equal to that of $\phi^{(1)}$, whose elements, U_{si}, are linear operators which act on functions of configuration space and time to produce functions of the phase-space variables and time. In Eq. (II.1), U_s acts on the array of perturbation potentials to give the total time derivative of the perturbation of the distribution function for species s. We refer to K as a *field operator*; it is a linear operator on the array $\phi^{(1)}$, and its elements are linear operators in the space of position and time. The operator $J_s^\dagger$ acts on functions of the phase-space variables and time to produce functions of configuration space and time. It is an integral operator with respect to velocity or canonical momentum, but most of our development does not depend on that fact. In Eq. (II.2), $J_s^\dagger f_s^{(1)}$ is the contribution to the source for $\phi^{(1)}$ due to the perturbation of the distribution function for species s. (The operators U_s and J_s provide a connection between configuration space and the phase space; they can be represented by rectangular matrices. The adjoint of an operator that produces a function in the phase space by operating on a function in configuration space is an operator that does the reverse; it produces a function in configuration space by operating on a function in the phase space. We adopt the convention that an operator which produces a function in the phase space by operating on a function in configuration space be denoted by a symbol without a dagger.) In terms of position and velocity, we take $J_s^\dagger f_s^{(1)}$ to be of the form

$$J_s^\dagger(\mathbf{r},\mathbf{v}) f_s^{(1)}(\mathbf{r},\mathbf{v},t) = \int d^3\mathbf{v}\, \mathcal{J}_s(\mathbf{r},\mathbf{v}) f_s^{(1)}(\mathbf{r},\mathbf{v},t). \tag{II.6}$$

The operators K and U_s may involve the time differentiation operator $\partial/\partial t$, but we do not allow the time t to enter these operators in any other way. Each of the examples that we have considered is consistent with this restriction; and it is required if the equations are to be invariant under time translation. Although invariance under time translation would not rule out the possibility that even $J_s^\dagger$ involve time differentiation, nevertheless we do not allow that possibility. We shall assume that $J_s^\dagger$ is completely time-independent. However, it should be noted that our subsequent development could be modified easily to include time differentiation in $J_s^\dagger$ if that situation were to occur in a problem of interest.

If all of the particle species are collisionless, then the linearized equations certainly have the form of Eqs. (II.1) and (II.2). Sometimes equations of the form of Eqs. (II.1) and (II.2) can describe a plasma in which one or more species are not collisionless. This is the case with the Vlasov-fluid model,[3,4] in which the effect of the electrons is manifested in the form of the coupled equations for the perturbed distribution function of the collisionless ions and the perturbed vector potential. In order to render the meaning of the quantities in Eqs. (II.1) and (II.2) more concrete, in Appendix B we give examples of three systems whose governing equations are of this form. For the present, let it suffice that we do consider systems whose governing equations are of the form of Eqs. (II.1) and (II.2), and that many plasmas are of that type.

We now modify Eqs. (II.1) and (II.2) by introducing an auxiliary function g_s, depending on the phase-space variables and time, to replace $f_s^{(1)}$; the definition of g_s is

$$g_s(t) = f_s^{(1)}(t) - \sum_i P_{si}(t)\,\phi_i^{(1)}(t) = f_s^{(1)}(t) - P_s(t)\,\phi^{(1)}(t), \tag{II.7}$$

where $P_s(t)$ is an array, of dimension equal to that of $\phi^{(1)}$, whose elements, $P_{si}(t)$, are linear operators in the space of the phase-space variables and time. The equations for g_s and $\phi^{(1)}$, obtained by using Eq. (II.7) to eliminate $f_s^{(1)}$ in Eqs. (II.1) and (II.2), are

$$\left[\frac{\partial}{\partial t}+\mathcal{L}_s\right] g_s(t)=\sum_i W_{si}(t)\,\phi_i^{(1)}(t)$$

$$=W_s(t)\,\phi^{(1)}(t), \tag{II. 8}$$

$$\Lambda\phi^{(1)}(t)=\sum_s J_s^\dagger g_s(t), \tag{II. 9}$$

where

$$W_s(t)=U_s(t)-\left[\frac{\partial}{\partial t}+\mathcal{L}_s\right] P_s(t), \tag{II. 10}$$

$$\Lambda=K(t)-\sum_s J_s^\dagger P_s(t). \tag{II. 11}$$

There is an important restriction that we impose on the time dependence of $P_s(t)$. The new field operator, Λ, which replaces the old field operator, $K(t)$, is *required not to involve time* in any way; it is not a differential operator with respect to time. The motivation for defining the functions g_s as in Eq. (II. 7) is that the new field operator Λ in Eq. (II. 9) can be chosen arbitrarily, by appropriate choice of the linear operators P_s. If, for a specific problem, $\phi^{(1)}$ can be adequately represented by a linear combination of a few of the eigenfunctions of Λ, then Eqs. (II. 8) and (II. 9) can be solved conveniently by expanding $\phi^{(1)}$ in terms of the eigenfunctions of Λ and retaining only a few terms in the expansion. This can be of considerable practical importance. Of course, the question still remains how to choose Λ such that it will be useful in a given problem, and that is a very important question of physics. The purpose of this paper is to provide a framework for treating Eqs. (II. 8) and (II. 9) once Λ has been chosen, and to point out the value of seeking appropriate new field operators that are generated by the auxiliary functions g_s. These ideas have served well in a numerical application of the Vlasov-fluid model to screw pinches[4] and in a study of the stability properties of one-dimensional BGK equilibria.[5]

Some of the systems to which the theory in this paper applies are completely Hamiltonian; that is, the particle equations of motion for each collisionless species and the field equations can be derived from Hamilton's principle by using a single Lagrangian. We assume that the equations of motion for the particles can always be derived from a Lagrangian. However, for some of the systems that we consider, like the Vlasov-fluid model,[3,4] the field equations and the particle equations of motion cannot both be derived from the same Lagrangian; in fact, it may be that the field equations for a given problem cannot be derived from any Lagrangian. In the remainder of this section, we indicate how the operators $\mathcal{L}_s$ and U_s are related to the particle Hamiltonian, and how the operator $J_s^\dagger$ is related to the particle Hamiltonian for systems that are completely Hamiltonian. Some further details and background information can be found in Appendix A.

Let $L_s[\mathbf{r}, \dot{\mathbf{r}}, \phi(\mathbf{r}, t)]$ be the Lagrangian for a particle of species s with position vector $\mathbf{r}$ and velocity vector $\dot{\mathbf{r}}$. Considering L_s to be an explicit function of the arguments $\mathbf{r}$, $\dot{\mathbf{r}}$, and ϕ, it is convenient to define a *canonical momentum vector* $\boldsymbol{\rho}$ by

$$\boldsymbol{\rho}=\nabla_{\dot{\mathbf{r}}} L_s[\mathbf{r}, \dot{\mathbf{r}}, \phi(\mathbf{r}, t)], \tag{II. 12}$$

where $\nabla_{\dot{\mathbf{r}}}$ is the gradient with respect to the velocity variables holding $\mathbf{r}$ and ϕ fixed. The components of $\boldsymbol{\rho}$ are simply related to the usual canonical momenta p_j as follows. Suppose that we use an orthogonal coordinate system whose unit vectors are $\hat{\mathbf{e}}_j(\mathbf{r})$, that the coordinate associated with $\hat{\mathbf{e}}_j$ is denoted by q_j, and that the derivative of $\mathbf{r}$ with respect to q_j is expressed as usual in terms of a function $h_j(\mathbf{r})$ by

$$\frac{\partial \mathbf{r}}{\partial q_j}=h_j(\mathbf{r})\,\hat{\mathbf{e}}_j(\mathbf{r}). \tag{II. 13}$$

If we specify the dependence of L_s on $\dot{\mathbf{r}}$ by means of the variables $\dot{q}_j$, then the usual canonical momentum p_j is defined by

$$p_j=\frac{\partial L_s}{\partial \dot{q}_j}; \tag{II. 14}$$

it is related to the jth component of $\boldsymbol{\rho}$ by

$$\rho_j=\hat{\mathbf{e}}_j\cdot\boldsymbol{\rho}=p_j/h_j. \tag{II. 15}$$

Thus, if the coordinate system is Cartesian, the components of $\boldsymbol{\rho}$ are the usual canonical momenta. The particle Hamiltonian corresponding to L_s is defined by

$$H_s[\mathbf{r}, \rho, \phi(\mathbf{r}, t)]=\boldsymbol{\rho}\cdot\dot{\mathbf{r}}-L_s[\mathbf{r}, \dot{\mathbf{r}}, \phi(\mathbf{r}, t)]; \tag{II. 16}$$

and Hamilton's equations, written in vector form, are

$$\dot{\mathbf{r}}=\nabla_\rho H_s,$$

$$\dot{\boldsymbol{\rho}}=-\nabla_{\mathbf{r}} H_s-\sum_i(\nabla\phi_i)\frac{\partial H_s}{\partial \phi_i}=-\nabla_{\mathbf{r}} H_s-(\nabla\phi)\frac{\partial H_s}{\partial \phi}. \tag{II. 17}$$

In these equations H_s is considered to be an explicit function of $\mathbf{r}$, $\boldsymbol{\rho}$, and ϕ; $\nabla_{\mathbf{r}}$ is the position gradient holding $\boldsymbol{\rho}$ *and* ϕ fixed; ∇_ρ is the canonical momentum gradient holding $\mathbf{r}$ and ϕ fixed; and ∇ is the usual gradient for functions of position. The quantity $\partial H_s/\partial\phi$ is an array whose elements are $\partial H_s/\partial\phi_i$. The vector form of Hamilton's equations is equivalent to the usual scalar form

$$\dot{q}_j=\frac{\partial H_s}{\partial p_j},$$

$$\dot{p}_j=-\frac{\partial H_s}{\partial q_j}-\sum_i\frac{\partial \phi_i}{\partial q_j}\frac{\partial H_s}{\partial \phi_i}. \tag{II. 18}$$

A more detailed discussion of the vector form of the Hamilton and Lagrange equations is given in Appendix A.

If f_s is the distribution function for the phase space of position and canonical momentum, it satisfies the following collisionless Boltzmann equation:

$$\frac{\partial f_s}{\partial t}+(\nabla_{\mathbf{r}} f_s)\cdot(\nabla_\rho H_s)-(\nabla_\rho f_s)\cdot\left(\nabla_{\mathbf{r}} H_s+(\nabla\phi)\frac{\partial H_s}{\partial \phi}\right)=0, \tag{II. 19}$$

where, as in Eq. (II. 17), a sum over the elements ϕ_i of ϕ is implied in the last term. The equilibrium distribution function is a solution of the zeroth-order part of this equation:

$$\mathcal{L}_s(\mathbf{r}, \boldsymbol{\rho})f_s^{(0)}(\mathbf{r}, \boldsymbol{\rho})=0. \tag{II. 20}$$

The definition of the equilibrium Liouville operator in terms of position and canonical momentum is

$$\mathcal{L}_s(\mathbf{r},\boldsymbol{\rho})\,G(\mathbf{r},\boldsymbol{\rho}) = (\nabla_{\mathbf{r}} G)\cdot(\nabla_{\boldsymbol{\rho}} H_s) - (\nabla_{\boldsymbol{\rho}} G)\cdot\left[\nabla_{\mathbf{r}} H_s + (\nabla\phi^{(0)})\,\frac{\partial H_s}{\partial\phi^{(0)}}\right], \quad \text{(II.21)}$$

where $G(\mathbf{r},\boldsymbol{\rho})$ is any function. [When we work to zeroth or first order in the perturbed quantities, as in Eqs. (II.20) and (II.21), we need only be concerned with particle motion in the equilibrium potentials. See Ref. 12 with regard to this.]

It is frequently useful to use the canonical coordinates and momenta explicitly, as we shall in Sec. IV. Let q denote the set of coordinates q_j, and p denote the set of their conjugate momenta p_j. If we consider $G(q,p)$ to be any explicit function of q and p, then $\mathcal{L}_s G$ is given by

$$\mathcal{L}_s G(q,p) = \sum_j \left[\frac{\partial G}{\partial q_j}\frac{\partial H_s}{\partial p_j} - \frac{\partial G}{\partial p_j}\left(\frac{\partial H_s}{\partial q_j} + \frac{\partial\phi^{(0)}}{\partial q_j}\frac{\partial H_s}{\partial\phi^{(0)}}\right)\right]. \quad \text{(II.22)}$$

(If G depends explicitly on $\phi^{(0)}$, the corresponding terms must be added.) This expresses the well-known fact that $\mathcal{L}_s G$ is the Poisson bracket $\{G, H_s\}$. The equivalence of Eqs. (II.21) and (II.22) can be demonstrated directly by deriving expressions for $\nabla_{\mathbf{r}}$ and $\nabla_{\boldsymbol{\rho}}$ as indicated in the discussion following Eq. (A.12), and by using the identity

$$\begin{aligned}(h_l\hat{\mathbf{e}}_l)\cdot[\nabla(h_j\hat{\mathbf{e}}_j)]\cdot\left(\frac{1}{h_i}\hat{\mathbf{e}}_i\right) - (h_j\hat{\mathbf{e}}_j)\cdot[\nabla(h_l\hat{\mathbf{e}}_l)]\cdot\left(\frac{1}{h_i}\hat{\mathbf{e}}_i\right)&\\ = -(h_l\hat{\mathbf{e}}_l)\cdot\left[\nabla\left(\frac{1}{h_i}\hat{\mathbf{e}}_i\right)\right]\cdot(h_j\hat{\mathbf{e}}_j)&\\ + (h_j\hat{\mathbf{e}}_j)\cdot\left[\nabla\left(\frac{1}{h_i}\hat{\mathbf{e}}_i\right)\right]\cdot(h_l\hat{\mathbf{e}}_l)&\\ = -(h_l\hat{\mathbf{e}}_l)\cdot\left\{(h_j\hat{\mathbf{e}}_j)\times\left[\nabla\times\left(\frac{1}{h_i}\hat{\mathbf{e}}_i\right)\right]\right\} = 0.&\end{aligned}$$

The perturbation distribution function $f_s^{(1)}$ satisfies the first-order part of Eq. (II.19),

$$\left[\frac{\partial}{\partial t} + \mathcal{L}_s(\mathbf{r},\boldsymbol{\rho})\right] f_s^{(1)} = -\left(\frac{\partial\mathcal{L}_s}{\partial\phi^{(0)}} f_s^{(0)}\right)\phi^{(1)} + (\nabla_{\boldsymbol{\rho}} f_s^{(0)})\cdot(\nabla\phi^{(1)})\,\frac{\partial H_s}{\partial\phi^{(0)}}, \quad \text{(II.23)}$$

which is of the form (II.1). The quantity $[(\partial\mathcal{L}_s/\partial\phi^{(0)})f_s^{(0)}]$ is defined by

$$\begin{aligned}\left(\frac{\partial\mathcal{L}_s}{\partial\phi^{(0)}} f_s^{(0)}\right)&\\ = (\nabla_{\mathbf{r}} f_s^{(0)})\cdot\left(\nabla_{\boldsymbol{\rho}}\frac{\partial H_s}{\partial\phi^{(0)}}\right)&\\ - (\nabla_{\boldsymbol{\rho}} f_s^{(0)})\cdot\left[\left(\nabla_{\mathbf{r}}\frac{\partial H_s}{\partial\phi^{(0)}}\right) + (\nabla\phi^{(0)})\frac{\partial^2 H_s}{\partial\phi^{(0)2}}\right]&\\ = \sum_j\left\{\frac{\partial f_s^{(0)}}{\partial q_j}\frac{\partial^2 H_s}{\partial p_j\partial\phi^{(0)}}\right.&\\ \left. - \frac{\partial f_s^{(0)}}{\partial p_j}\left[\frac{\partial^2 H_s}{\partial q_j\partial\phi^{(0)}} + \frac{\partial\phi^{(0)}}{\partial q_j}\frac{\partial^2 H_s}{\partial\phi^{(0)2}}\right]\right\}.&\end{aligned} \quad \text{(II.24)}$$

The large parentheses on the left side of the equation are used to indicate that the differential operators in $\partial\mathcal{L}_s/\partial\phi^{(0)}$ act only on $f_s^{(0)}$. Comparing Eqs. (II.1) and (II.23), we see that U_s can be written as

$$\begin{aligned}U_s &= -\left[\frac{\partial\mathcal{L}_s}{\partial\phi^{(0)}} f_s^{(0)}\right] + \frac{\partial H_s}{\partial\phi^{(0)}}(\nabla_{\boldsymbol{\rho}} f_s^{(0)})\cdot\nabla\\ &= -\left[\frac{\partial\mathcal{L}_s}{\partial\phi^{(0)}} f_s^{(0)}\right] + \frac{\partial H_s}{\partial\phi^{(0)}}\sum_j\left[\frac{\partial f_s^{(0)}}{\partial p_j}\frac{\partial}{\partial q_j}\right].\end{aligned} \quad \text{(II.25)}$$

It must be remembered throughout that $\nabla_{\mathbf{r}}$ and $\nabla_{\boldsymbol{\rho}}$ are gradient operators *holding the array of potentials fixed.* Note that U_s does not contain the operator $\partial/\partial t$ if the variables $\mathbf{r}$ and $\boldsymbol{\rho}$ are used; if the variables $\mathbf{r}$ and $\mathbf{v}$ are used, U_s may contain $\partial/\partial t$. See the note added in proof, Ref. 13.

In terms of the variables $\mathbf{r}$ and $\boldsymbol{\rho}$, we take the quantity $J_s^\dagger f_s^{(1)}$ that appears in Eq. (II.2) to be of the form

$$J_s^\dagger(\mathbf{r},\boldsymbol{\rho}) f_s^{(1)}(\mathbf{r},\boldsymbol{\rho},t) = \int d^3\rho\, \mathcal{J}_s(\mathbf{r},\boldsymbol{\rho}) f_s^{(1)}(\mathbf{r},\boldsymbol{\rho},t), \quad \text{(II.26)}$$

quite analogously to Eq. (II.6). For completely Hamiltonian systems $\mathcal{J}_s$ can be written as

$$\mathcal{J}_s = \frac{\partial H_s}{\partial\phi^{(0)}}. \quad \text{(II.27)}$$

When relation (II.27) applies, there are some special features of the solution of the linearized problem. These will be developed later in the paper. A proof of Eq. (II.27) can be found in Appendix A. Further discussion of Hamiltonian aspects of the systems treated in this paper is given in Appendices A and C.

The equilibrium distribution function can be expressed as an explicit function of constants of the motion for the equilibrium Hamiltonian,

$$f_s^{(0)}(\mathbf{r},\boldsymbol{\rho}) = \mathcal{J}_s[H_s(q,p,\phi^{(0)}), I_s(q,p)]. \quad \text{(II.28)}$$

One of the constants of the motion is $H_s[q,p,\phi^{(0)}]$ itself. In addition, the equilibrium distribution function may depend on other constants of the motion, which we label with an index k and denote by $I_s^{(k)}(q,p)$; we denote the set of invariants $I_s^{(k)}(q,p)$ by $I_s(q,p)$. As is indicated by the notation, we are now considering H_s to be an explicit function of q, p, and $\phi^{(0)}$. We have required in Eq. (II.20) that $f_s^{(0)}$ be an exact equilibrium solution. However, we note that only a rather small modification of the discussion in this paper would be required to include equilibria which are approximate due to at least one of the constants of the motion $I_s^{(k)}$ being an adiabatic invariant instead of exact. In that case, we would allow the right hand side of Eq. (II.20) to be $O(\epsilon)$, where ϵ is a smallness parameter of the order of the perturbation quantities. If an approximate constant of the motion were a second-order adiabatic invariant, so that the right-hand side of Eq. (II.20) were $O(\epsilon^2)$, then our formalism could be applied without change, as if the constant of the motion were exact. In any event, if we express $f_s^{(0)}$ explicitly in terms of H_s and I_s as in Eq. (II.28), then U_s can be written as

$$\begin{aligned}U_s = \frac{\partial\mathcal{J}_s}{\partial H_s}\left[\left(\mathcal{L}_s\frac{\partial H_s}{\partial\phi^{(0)}}\right) + \sum_j\frac{\partial H_s}{\partial\phi^{(0)}}\frac{\partial H_s}{\partial p_j}\frac{\partial}{\partial q_j}\right]&\\ - \sum_k\frac{\partial\mathcal{J}_s}{\partial I_s^{(k)}}\sum_j\left\{\frac{\partial I_s^{(k)}}{\partial q_j}\frac{\partial^2 H_s}{\partial p_j\partial\phi^{(0)}}\right.&\\ \left. - \frac{\partial I_s^{(k)}}{\partial p_j}\left[\frac{\partial^2 H_s}{\partial q_j\partial\phi^{(0)}} + \frac{\partial\phi^{(0)}}{\partial q_j}\frac{\partial^2 H_s}{\partial\phi^{(0)2}} + \frac{\partial H_s}{\partial\phi^{(0)}}\frac{\partial}{\partial q_j}\right]\right\}.&\end{aligned} \quad \text{(II.29)}$$

A. Specialization to global invariants

A case of great interest and applicability is that in which the invariant $I_s^{(k)}(q,p)$ is precisely p_k, the momentum conjugate to an ignorable coordinate q_k, for each value of k. Then $f_s^{(0)}$ is expressible completely as a function of the Hamiltonian and of the canonical momenta associated with spatial symmetries of the Hamiltonian. In Sec. IV, we shall restrict attention to this case and make use of special properties associated with it. Specializing to this case, we take

$$I_s^{(k)}(q,p) = p_k, \tag{II.30}$$

where p_k is the momentum conjugate to an *ignorable* coordinate q_k, and write

$$f_s^{(0)}(\mathbf{r},\mathbf{p}) = \mathcal{J}_s[H_s(q,p,\phi^{(0)}),\{p_k\}], \tag{II.31}$$

indicating that we consider $\mathcal{J}_s$ to be an explicit function of H_s and of conserved momenta p_k. In order to simplify subsequent formulas, henceforth we shall use the index k exclusively for ignorable coordinates and their conjugate momenta, and sums over k should be understood to run only over the ignorable variables. The symbols q and p will continue to denote the sets of all the canonical coordinates and momenta, nonignorable as well as ignorable.

With this specialization, the operator U_s is simplified considerably. The derivatives $\partial I_s^{(k)}/\partial q_j$ obviously vanish. In addition, because q_k is ignorable, neither $\phi^{(0)}$ nor $\partial H_s/\partial\phi^{(0)}$ depends on q_k, and their derivatives with respect to q_k vanish. Therefore U_s can be written simply as

$$U_s = \frac{\partial\mathcal{J}_s}{\partial H_s}\left[\left(\mathcal{L}_s\frac{\partial H_s}{\partial\phi^{(0)}}\right) + \sum_j \frac{\partial H_s}{\partial\phi^{(0)}}\frac{\partial H_s}{\partial p_j}\frac{\partial}{\partial q_j}\right] + \sum_k \frac{\partial\mathcal{J}_s}{\partial p_k}\frac{\partial H_s}{\partial\phi^{(0)}}\frac{\partial}{\partial q_k}. \tag{II.32}$$

An interesting choice of the quantity P_s in the transformation (II.7), discovered independently by Holdren,[7] is

$$P_s = \frac{\partial\mathcal{J}_s}{\partial H_s}\frac{\partial H_s}{\partial\phi^{(0)}}. \tag{II.33}$$

With this choice, the operator W_s defined by Eq. (II.10) becomes

$$\begin{aligned} W_s &= U_s - P_s\frac{\partial}{\partial t} - P_s\mathcal{L}_s - (\mathcal{L}_s P_s) \\ &= U_s - \frac{\partial\mathcal{J}_s}{\partial H_s}\left[\frac{\partial H_s}{\partial\phi^{(0)}}\frac{\partial}{\partial t} + \frac{\partial H_s}{\partial\phi^{(0)}}\sum_j\frac{\partial H_s}{\partial p_j}\frac{\partial}{\partial q_j} + \left(\mathcal{L}_s\frac{\partial H_s}{\partial\phi^{(0)}}\right)\right] \\ &= \frac{\partial H_s}{\partial\phi^{(0)}}\left(\sum_k\frac{\partial\mathcal{J}_s}{\partial p_k}\frac{\partial}{\partial q_k} - \frac{\partial\mathcal{J}_s}{\partial H_s}\frac{\partial}{\partial t}\right). \end{aligned} \tag{II.34}$$

In obtaining this expression, we have employed the fact that W_s acts only on field functions—that is, on functions which do not depend on the momenta p_j. Formally, this choice of P_s has explicitly brought out in W_s the analogy between time and energy on the one hand and conjugate pairs of coordinates and momenta on the other. This will prove especially convenient in Sec. IV for making the Hermitian character of the dispersion matrix D explicit. Physically,

$$P_s\phi^{(1)} = \frac{\partial\mathcal{J}_s}{\partial H_s}\frac{\partial H_s}{\partial\phi^{(0)}}\phi^{(1)}$$

would be the perturbation distribution function $f_s^{(1)}$ that solves Eq. (II.1) if $\phi^{(1)}$ were time-independent and also independent of the ignorable coordinates q_k. This is evident because the condition for $P_s\phi^{(1)}$ to solve Eq. (II.1) is $W_s\phi^{(1)} = 0$, which is satisfied if $(\partial\mathcal{J}_s/\partial p_k)(\partial\phi^{(1)}/\partial q_k) = 0$ and $(\partial\mathcal{J}_s/\partial H_s)(\partial\phi^{(1)}/\partial t) = 0$. Thus the auxiliary function g_s defined by Eq. (II.7) with this choice of P_s is the nonadiabatic part of $f_s^{(1)}$, and this choice may be expected to be useful for studying the low-frequency behavior of some systems.

III. THE INITIAL VALUE PROBLEM

We now turn to the solution of Eqs. (II.8) and (II.9) for the evolution of the auxiliary functions g_s and the potential functions $\phi_i^{(1)}$. The solution will be in terms of Laplace transforms $\hat{g}_s(\omega)$ and $\hat{\phi}^{(1)}(\omega)$ of $g_s(t)$ and $\phi^{(1)}(t)$:

$$g_s(t) = (2\pi)^{-1}\int_C d\omega\, e^{-i\omega t}\hat{g}_s(\omega), \quad \hat{g}_s(\omega) = \int_0^\infty dt\, e^{i\omega t}g_s(t), \tag{III.1}$$

$$\phi^{(1)}(t) = (2\pi)^{-1}\int_C d\omega\, e^{-i\omega t}\hat{\phi}^{(1)}(\omega), \quad \hat{\phi}^{(1)}(\omega) = \int_0^\infty dt\, e^{i\omega t}\phi^{(1)}(t) \tag{III.2}$$

where C is a suitable Bromwich contour. The operator W_s may contain the time differentiation operator $\partial/\partial t$, but it does not involve time in any other way. Therefore, the Laplace-transform representation of $W_s\phi^{(1)}$ may be written as

$$W_s(t)\phi^{(1)}(t) = (2\pi)^{-1}\int_C d\omega\, e^{-i\omega t}[\hat{W}_s(\omega)\hat{\phi}^{(1)}(\omega) + \Phi_s(\omega)], \tag{III.3}$$

where $\hat{W}_s(\omega)$ equals the result of substituting $-i\omega$ for $\partial/\partial t$ in $W_s(t)$, and where $\Phi_s(\omega)$ can be constructed from $\phi^{(1)}(t)$ and its time derivatives evaluated at $t=0$. The operators Λ and $J_s^\dagger$ do not involve time at all.

The equations for the Laplace transforms $\hat{g}_s(\omega)$ and $\hat{\phi}^{(1)}(\omega)$ that are implied by Eqs. (II.8) and (II.9) are

$$[\mathcal{L}_s - i\omega]\hat{g}_s(\omega) = \hat{W}_s(\omega)\hat{\phi}^{(1)}(\omega) + [\Phi_s(\omega) + g_s(0)], \tag{III.4}$$

and

$$\Lambda\hat{\phi}^{(1)}(\omega) = \Sigma_s J_s^\dagger\hat{g}_s(\omega). \tag{III.5}$$

Equations (III.4) and (III.5) can be solved by first eliminating $\hat{g}_s$ to obtain an equation for $\hat{\phi}^{(1)}$, and then substituting the solution for $\hat{\phi}^{(1)}$ into Eq. (III.4) to obtain an equation for $\hat{g}_s$. The result is

$$\hat{\phi}^{(1)}(\omega) = D^{-1}(\omega)\{\Sigma_s J_s^\dagger[\mathcal{L}_s - i\omega]^{-1}[\Phi_s(\omega) + g_s(0)]\}, \tag{III.6}$$

$$\begin{aligned}\hat{g}_s(\omega) = {}& [\mathcal{L}_s - i\omega]^{-1}\hat{W}_s(\omega)D^{-1}(\omega)\{\Sigma_{s'}J_{s'}^\dagger[\mathcal{L}_{s'} - i\omega]^{-1} \\ & \times[\Phi_{s'}(\omega) + g_{s'}(0)]\} + [\mathcal{L}_s - i\omega]^{-1}[\Phi_s(\omega) + g_s(0)],\end{aligned} \tag{III.7}$$

where

$$D(\omega) = \Lambda - \Sigma_s J_s^\dagger[\mathcal{L}_s - i\omega]^{-1}\hat{W}_s(\omega). \tag{III.8}$$

The contour C for the Laplace transforms must be above all singularities of $\hat{\phi}^{(1)}(\omega)$ and $\hat{g}_s(\omega)$. In this connection we note that $\mathcal{L}_s$ is anti-Hermitian, so that its eigenvalues are purely imaginary. A simple general proof of the anti-Hermiticity of $\mathcal{L}_s$ is presented in Appendix C. We

call $D(\omega)$ the *dispersion operator*. It is independent of the operators $P_s(t)$ in terms of which Λ and $W_s(t)$ were defined by Eqs. (II.10) and (II.11). That is, $D(\omega)$ can also be written as

$$D(\omega)=\hat{K}(\omega)-\Sigma_s J_s^\dagger[\mathcal{L}_s-i\omega]^{-1}\hat{U}_s(\omega), \tag{III.9}$$

where $\hat{K}(\omega)$ and $\hat{U}_s(\omega)$ equal the result of substituting $-i\omega$ for $\partial/\partial t$ in $K(t)$ and $U_s(t)$, respectively. For completely Hamiltonian systems, it is possible to choose the operators P_s in such a way that the dispersion operator $D(\omega)$ is manifestly a Hermitian function of ω. This is discussed in Sec. IV.

Although the Laplace transforms $\hat{\phi}^{(1)}$ and $\hat{g}_s$ given by Eqs. (III.6)—(III.8) are written quite compactly, the expressions are in fact very complicated. The dispersion matrix and some features of the solution of the initial-value problem are discussed in Secs. IV and V, with particular reference to the case in which there is at most one nonignorable coordinate in the equilibrium. Typically (see Sec. V), because of the factor $[\mathcal{L}_s-i\omega]^{-1}$ in the second term in $D(\omega)$, and since the operator $\mathcal{L}_s$ has a continuous spectrum of imaginary eigenvalues, the solution can be written in terms of a continuum of generalized van Kampen modes along the real ω axis, plus contributions from poles (and other singularities if any) in the complex ω plane. If we approximate the problem by choosing an N-dimensional representation of configuration space, then the matrix $D(\omega)$ becomes $N\times N$, and the contribution to the solution from the N-fold degenerate continuum of generalized van Kampen modes can be written down in principle. If, further, we use an approximate finite-dimensional representation of the phase space for each species, then if the operators $\hat{W}_s(\omega)$ are polynomial functions of ω, the singularities of the Laplace transforms will be a finite number of poles that must be located and whose orders must be determined. [An assumption that $\hat{W}_s(\omega)$ be a polynomial function of ω would not pose a serious restriction. For the examples in Appendix B, $\hat{W}_s(\omega)$ is at most a cubic function of ω.] We shall introduce such a finite-dimensional approximation scheme at the beginning of the next section. In that case, there is another form of the Laplace transform $\hat{g}_s(\omega)$ that allows the poles to be located and their orders determined by finding the zeros of a *polynomial* in ω. The degree of the polynomial may be large for any particular approximation; but the polynomial can be written in a form that allows it to be evaluated practically even when the degree is large indeed. This other form of the Laplace transform is also interesting theoretically, apart from its practical applicability to any particular approximation scheme.

In order to obtain the new form of the Laplace transform, we solve Eqs. (III.4) and (III.5) differently. Instead of first eliminating $\hat{g}_s$, we first solve Eq. (III.5) for $\hat{\phi}^{(1)}$ in terms of $\hat{g}_s$, and substitute it into Eq. (III.4) to obtain an equation for $\hat{g}_s$; then the solution for $\hat{g}_s$ can be used in Eq. (III.5) to obtain $\hat{\phi}^{(1)}$. In order to carry out this method of solution, we must concern ourselves with the possibility that Λ has a zero eigenvalue.

Only the part of $\hat{\phi}^{(1)}$ lying outside the nullspace of Λ can be determined from Eq. (III.5). We shall see shortly that Λ can be chosen not to have a zero eigenvalue. This is the situation that we shall develop fully; and, in the process, we shall obtain a final result that does not require that Λ^{-1} exist. For comparison, we first describe *briefly* how $\hat{\phi}^{(1)}$ could be found in terms of the functions $\hat{g}_s$ when the nullspace of Λ is *not* empty.

Let $\mathcal{P}$ be the projection operator for the entire nullspace of Λ:

$$\mathcal{P}\Lambda=\Lambda\mathcal{P}=0,\quad (1-\mathcal{P})\Lambda=\Lambda(1-\mathcal{P})=\Lambda,\quad \mathcal{P}^2=\mathcal{P}.$$

The part of $\hat{\phi}^{(1)}$ lying outside the nullspace of Λ can be determined by projecting out the part of Eq. (III.5) outside the nullspace to obtain

$$\Lambda[(1-\mathcal{P})\hat{\phi}^{(1)}(\omega)]=\Sigma_s(1-\mathcal{P})J_s^\dagger\hat{g}_s(\omega). \tag{III.10}$$

This equation is in terms of quantities that are outside the nullspace of Λ, and Λ possesses an inverse within that complementary space. Now project out the part of Eq. (III.5) in the nullspace of Λ to obtain

$$0=\Sigma_s\mathcal{P}J_s^\dagger\hat{g}_s(\omega). \tag{III.11}$$

This is a condition on the functions $\hat{g}_s(\omega)$ which can be used to determine $\mathcal{P}\hat{\phi}^{(1)}$ from Eq. (III.4). By operating on Eq. (III.4) from the left with $\mathcal{P}J_s^\dagger$, summing over s, and using Eq. (III.11), we can obtain

$$\begin{aligned}&[\Sigma_s\mathcal{P}J_s^\dagger\hat{W}_s(\omega)\mathcal{P}][\mathcal{P}\hat{\phi}^{(1)}(\omega)]\\&=\Sigma_s\mathcal{P}J_s^\dagger\mathcal{L}_s\hat{g}_s(\omega)-\Sigma_s\mathcal{P}J_s^\dagger[\Phi_s(\omega)+g_s(0)]\\&\quad-[\Sigma_s\mathcal{P}J_s^\dagger\hat{W}_s(\omega)(1-\mathcal{P})][(1-\mathcal{P})\hat{\phi}^{(1)}(\omega)].\end{aligned} \tag{III.12}$$

This equation determines $\mathcal{P}\hat{\phi}^{(1)}$ in terms of the functions $\hat{g}_s$ and the function $(1-\mathcal{P})\hat{\phi}^{(1)}$ which can be found from Eq. (III.10). In this way can $\hat{\phi}^{(1)}$ be expressed in terms of the functions $\hat{g}_s$ when Λ has a nonempty nullspace.

The procedure that was just described leads to an unnecessarily complicated result and it is circuitous. Let us see how to avoid that procedure by arranging that Λ have no zero eigenvalue. The real motivation for introducing the operator Λ was to have a field operator whose eigenfunctions are particularly suitable for expanding $\hat{\phi}^{(1)}$ for some specific problem. Suppose that a suitable field operator, which we shall call Λ', has been found. Now note that $(\Lambda'-\delta)$, where δ is an arbitrary complex constant, has the same eigenfunctions as Λ'. Therefore, from the standpoint of representing $\hat{\phi}^{(1)}$, $(\Lambda'-\delta)$ is as good a field operator as Λ'. However, the eigenvalue spectrum of $(\Lambda'-\delta)$ is the eigenvalue spectrum of Λ' shifted by the constant $-\delta$. If Λ' has a zero eigenvalue, then δ can be chosen such that $(\Lambda'-\delta)$ does not. (If δ is complex, Λ' may not be Hermitian; this, however, does not affect our development.) We now ask whether operators P_s can be found such that

$$\Sigma_s J_s^\dagger P_s=\delta, \tag{III.13}$$

in order that we can introduce

$$\Lambda=(\Lambda'-\delta)=\Lambda'-\Sigma_s J_s^\dagger P_s,$$

in analogy with Eq. (II.11). The operators P_s are underdetermined by Eq. (III.13) because, although P_s is an operator that connects the configuration space and the phase space, the equation imposes a condition only in

the configuration space. Thus, Eq. (III.13) will generally have infinitely many solutions. Some operators P_s that satisfy Eq. (III.13) for examples discussed in Appendix B are given explicitly by Eqs. (B28) and (B42) of Appendix B.

Now assume that Λ does not possess a zero eigenvalue and, therefore that Λ^{-1} exists. The solution of Eq. (III.5) for $\hat{\phi}^{(1)}$ is

$$\hat{\phi}^{(1)}(\omega)=\sum_s \Lambda^{-1}J_s^\dagger \hat{g}_s(\omega); \tag{III.14}$$

and, by substituting this into Eq. (III.4), we obtain

$$\sum_r\{[\not{L}_s - i\omega]\delta_{sr} - \hat{W}_s(\omega)\Lambda^{-1}J_r^\dagger\}\hat{g}_r(\omega)=[\Phi_s(\omega)+g_s(0)] \tag{III.15}$$

as the equation for determining the functions $\hat{g}_s(\omega)$. For studying these equations it is helpful to introduce a matrix notation in which the rows and columns are labelled by the species index. A function with a species subscript will be replaced by a column matrix denoted by the unsubscripted symbol for the function. An operator with a species subscript which operates on a function in configuration space to produce a function in the phase space for the species will be replaced by a column matrix denoted by the unsubscripted symbol for the operator. (Our convention is that such an operator does not have a superscript dagger.) The adjoint of such an operator will be replaced by the adjoint of the associated column matrix (a row matrix). An operator with a species subscript which operates on a function in the phase space for the species to produce another function in the phase space for the species will be replaced by a diagonal square matrix denoted by the unsubscripted symbol for the operator. For example, element s of $\hat{g}(\omega)$ is $\hat{g}_s(\omega)$, element s of $\hat{W}(\omega)$ is $\hat{W}_s(\omega)$, element s of $J^\dagger$ is $J_s^\dagger$, and the element of $\not{L}$ in row s and column r is $\not{L}_s\delta_{sr}$.

With this notation we rewrite Eqs. (III.14) and (III.15) as

$$\hat{\phi}^{(1)}(\omega)=\Lambda^{-1}J^\dagger\hat{g}(\omega) \tag{III.16}$$

and

$$\hat{g}(\omega)=S^{-1}(\omega)[\Phi(\omega)+g(0)], \tag{III.17}$$

where

$$S(\omega)=[\not{L} - i\omega\mathrm{I}] - \hat{W}(\omega)\Lambda^{-1}J^\dagger, \tag{III.18}$$

and I is the unit matrix. For the sake of simplicity, we now assume that $\hat{W}(\omega)$ is a *polynomial function* of ω, which implies that $\Phi(\omega)$ also is a polynomial function of ω. This assumption includes the examples in Appendix B as special cases. The form of $\hat{g}(\omega)$ given by Eq. (III.17) is different than that given by Eq. (III.7). Nevertheless, the two are equivalent, and it is apparent that the singularities of $\hat{g}(\omega)$ occur at those values for which $S^{-1}(\omega)$ is singular. We now demonstrate the equivalence of the two forms of $\hat{g}(\omega)$ by a method which will indicate a practical means of finding the zeros of finite-dimensional representations of $S(\omega)$. In order to do that we prove the following theorem.

Theorem: Let σ and ρ be any operators that operate on a function in a space R_1 to produce a function in a space R_2. The adjoints $\sigma^\dagger$ and $\rho^\dagger$ operate on functions in R_2 to produce functions in R_1. Let I_1 be the identity operator in R_1, and let I_2 be the identity operator in R_2. If $[I_2 - \sigma\rho^\dagger]^{-1}$ is nonsingular, then

$$[I_2 - \sigma\rho^\dagger]^{-1}=[I_2 + \sigma A\rho^\dagger],$$

where

$$A=(I_1 - \rho^\dagger\sigma)^{-1}. \tag{III.19}$$

The operator A operates on functions in R_1 to produce functions that are also in R_1.

Proof:

Since we are interested in applications to problems involving functions which can be adequately approximated with a finite dimensional basis, we shall assume that the spaces R_1 and R_2 are finite-dimensional and, therefore, that the operators can be represented by matrices. Beginning with

$$(I_2 - \sigma\rho^\dagger)(I_2 + \sigma A\rho^\dagger)=I_2 + \sigma[(I_1 - \rho^\dagger\sigma)A - I_1]\rho^\dagger,$$

we see that the theorem will be proved if

$$(I_1 - \rho^\dagger\sigma)A = I_1.$$

This equation has the solution given by Eq. (III.19) if $(I_1 - \rho^\dagger\sigma)^{-1}$ is nonsingular. That it is nonsingular can be seen by introducing finite-dimensional bases for R_1 and R_2, so that σ and $\rho^\dagger$ are rectangular matrices and A is a square matrix. We now use the determinantal identity

$$\det(I_1 - \rho^\dagger\sigma)=\det(I_2 - \sigma\rho^\dagger). \tag{III.20}$$

[This identity is of crucial importance for numerical solution of the initial-value problem, it will be discussed in a separate paper dealing with numerical procedures.[3]] Because $(I_2 - \sigma\rho^\dagger)^{-1}$ was assumed to be nonsingular, and because the identity (III.20) is valid with *any* finite-dimensional bases for R_1 and R_2, we conclude that $(I_1 - \rho^\dagger\sigma)^{-1}$ is nonsingular. Thus, the theorem is proved.

The operator $S^{-1}(\omega)$ is nonsingular on the Bromwich contour, if it were not, then the contour for the Laplace transforms would have been chosen incorrectly. Therefore, it can be evaluated by using the theorem that was just proved. The result is

$$S^{-1}(\omega)=[\mathrm{I} + (\not{L} - i\omega\mathrm{I})^{-1}\hat{W}(\omega)D^{-1}(\omega)J^\dagger][\not{L} - i\omega\mathrm{I}]^{-1}, \tag{III.21}$$

which shows that Eq. (III.17) and Eq. (III.7) are equivalent. It is straightforward to show that Eqs. (III.16) and (III.6) for $\hat{\phi}^{(1)}$ are also equivalent. Summarizing our results so far for $\hat{\phi}^{(1)}$ and $\hat{g}$, we may write

$$\hat{g}(\omega)=S^{-1}(\omega)[\Phi(\omega)+g(0)], \tag{III.17}$$

$$\begin{aligned}\hat{\phi}^{(1)}(\omega) &= \Lambda^{-1}J^\dagger\hat{g}(\omega)\\ &= D^{-1}(\omega)J^\dagger[\not{L} - i\omega\mathrm{I}]^{-1}[\Phi(\omega)+g(0)],\end{aligned} \tag{III.22}$$

where

$$D(\omega)=\Lambda - J^\dagger[\not{L} - i\omega\mathrm{I}]^{-1}\hat{W}(\omega), \tag{III.23}$$

and where $S(\omega)$ is given by Eq. (III.18) and $S^{-1}(\omega)$ is given by Eq. (III.21).

Now suppose that we have a finite-dimensional representation for $S(\omega)$. Then $\det S(\omega)$ is a *polynomial* in ω, and the singularities of $S^{-1}(\omega)$ are poles located at the roots of the determinantal equation

$$\det S(\omega)=0.$$

Writing

$$S(\omega)=[\mathcal{L}-i\omega\mathbf{I}_2][\mathbf{I}_2-(\mathcal{L}-i\omega\mathrm{I}_2)^{-1}\hat{W}(\omega)\Lambda^{-1}J^\dagger],$$

where I_2 is the identity matrix with respect to the finite-dimensional basis for the composite phase space of all of the particle species, and using the determinantal identity given by Eq. (III. 20), we can rewrite the determinantal equation as

$$\begin{aligned}\det S(\omega)&=(\det[\mathcal{L}-i\omega\mathbf{I}_2])(\det[I_1-\Lambda^{-1}J^\dagger(\mathcal{L}-i\omega\mathbf{I}_2)^{-1}\hat{W}(\omega)])\\&=(\det[\mathcal{L}-i\omega\mathbf{I}_2])(\det[\Lambda^{-1}D(\omega)])\\&=(\det\Lambda^{-1})(\det[\mathcal{L}-i\omega\mathbf{I}_2])(\det D(\omega))=0,\end{aligned}$$

where I_1 is the identity matrix with respect to the finite-dimensional basis for configuration space. Therefore, the roots of the determinantal equation are the same as those of the *polynomial* equation

$$(\det[\mathcal{L}-i\omega\mathbf{I}_2])(\det D(\omega))=0. \tag{III. 24}$$

Note that this equation does not involve Λ^{-1}, nor do Eqs. (III. 17) and (III. 22) for the Laplace transforms. Thus, we no longer have the restriction that Λ^{-1} exist. Equation (III. 24) is computationally significant because it provides a practical means of computing the roots of $\det S(\omega)$, even when the degree of the polynomial is large, as long as the dimension of the matrix $D(\omega)$ is not large. If Λ has been chosen appropriately, so that a few eigenfunctions of Λ suffice to approximate $\hat{\phi}^{(1)}$, then those eigenfunctions can be used as the finite-dimensional basis for configuration space, and the dimension of $D(\omega)$ will be small. The matrix $[\mathcal{L}-i\omega\mathbf{I}_2]$ has the dimension of the finite-dimensional basis for the phase space, which may be large. However, by choosing a finite subset of the eigenfunctions of $\mathcal{L}$ as the basis, we diagonalize $[\mathcal{L}-i\omega\mathbf{I}_2]$ and can compute its determinant easily. Once the zeros of $\det S(\omega)$ have been found, the coefficients of the inverse powers in the Laurent expansions of $\hat{g}(\omega)$ and $\hat{\phi}^{(1)}(\omega)$ about each of the zeros can be evaluated and used to invert the Laplace transformation to obtain $g(t)$ and $\phi^{(1)}(t)$. An effective scheme for finding the zeros of $\det S(\omega)$ numerically is a global Newton's method. Computational aspects of the procedure that was just outlined for finding a discrete approximation to the eigenfrequency spectrum associated with $\hat{g}(\omega)$ and $\hat{\phi}^{(1)}(\omega)$ will be presented elsewhere.[9]

IV. THE DISPERSION MATRIX

A. Expansions in eigenfunctions of Λ and $\mathcal{L}$

As we noted earlier, it is useful for numerical purposes to take eigenfunctions of Λ as a basis for configuration space, and to take eigenfunctions of $\mathcal{L}$ as a basis for the phase space. It is also useful for analytical purposes to choose these bases.

The introduction of a basis for a linear vector space requires the specification of a weight function with respect to which an inner product in the space is defined. In configuration space we define the inner product (α,β) with respect to a real weight function $\Gamma^{(c)}$ of position by

$$(\alpha,\beta)=\int d^3\mathbf{r}\,\Gamma^{(c)}\alpha^*\beta=(\beta,\alpha)^*, \tag{IV. 1}$$

where α and β are arbitrary functions of position. Note that in general the symbol ϕ stands for an array of potential functions ϕ_i; correspondingly, α,β are in general arrays of functions α_i, β_i, and the product $\alpha^*\beta$ stands for $\Sigma_i\alpha_i^*\beta_i$. By a function in configuration space, we mean a set of functions like $\phi_i(\mathbf{r})$. In the phase space for species s we define the inner product (α,β) with respect to a real weight function $\Gamma_s^{(p)}$ of the phase-space variables for species s by

$$(\alpha,\beta)=\int d\Omega\,\Gamma_s^{(p)}\alpha^*\beta=(\beta,\alpha)^*, \tag{IV. 2}$$

where α and β are arbitrary functions of the phase-space variables for species s and $d\Omega$ is the differential volume element in the phase space for species s. Equations (IV. 1) and (IV. 2) define different scalar products, one for functions in configuration space and one for functions in the phase space for each species. We use the same notation for each; it will be evident from the context which is meant. We require that $\Gamma^{(c)}$ commute with Λ and that $\Gamma_s^{(p)}$ commute with $\mathcal{L}_s$, but the weight functions are otherwise arbitrary. Useful choices for numerical approximations are

$$\Gamma_s^{(p)}=f_s^{(0)}\quad\text{and}\quad\Gamma^{(c)}=1. \tag{IV. 3}$$

We define eigenfunctions and eigenvalues of Λ and $\mathcal{L}_s$ by

$$\Lambda\eta_n=\lambda_n\eta_n \tag{IV. 4}$$

and

$$\mathcal{L}_s w_{sr}=i\mu_{sr}w_{sr}, \tag{IV. 5}$$

where, because of the anti-Hermiticity of $\mathcal{L}_s$, μ_{sr} is real. The indices n and r stand for whatever sets of labels are needed to specify the eigenfunctions η_n and w_{sr}. Again, in general, η_n is an array of functions. Some labels may be discrete and others may be continuous. For numerical work it is appropriate from the outset to define functions in terms of a representation of the phase or configuration space that is spanned by some discrete set of functions. That will ensure that all operators can be represented by matrices and that the labels for the eigenfunctions of Λ and $\mathcal{L}_s$ are discrete. We choose the eigenfunctions w_{sr} to be orthonormal,

$$(w_{sr},w_{sr'})=\delta_{rr'}, \tag{IV. 6}$$

where here $\delta_{rr'}$ stands for a product of Kronecker deltas and Dirac delta functions—one Kronecker delta for each pair of discrete labels, and one Dirac delta function for each pair of continuous labels. In Appendix D, the eigenfunctions of $\mathcal{L}_s$ are discussed further, and they are found explicitly for the case in which there is at most one nonignorable coordinate in the equilibrium for species s.

We assume that the eigenfunctions of Λ and $\mathcal{L}_s$ are complete for the problem. However, Λ *need* not be Hermitian, although it can be chosen to be Hermitian for most problems. Because Λ may not be Hermitian, we introduce a set of functions ζ_n which are dual to the set of functions η_n:

$$(\zeta_n,\eta_{n'})=\delta_{nn'}, \tag{IV. 7}$$

where again $\delta_{nn'}$ stands for a product of Kronecker deltas and Dirac delta functions.

The functions $\phi^{(1)}$ and g_s are expanded in terms of the eigenfunctions of Λ and $\mathcal{L}_s$ according to

$$\phi^{(1)}(t)=\sum_n \alpha_n(t)\eta_n \tag{IV.8}$$

and

$$g_s(t)=\sum_r \gamma_{sr}(t)w_{sr}, \tag{IV.9}$$

where $\alpha_n(t)$ and $\gamma_{sr}(t)$ are coefficients that do not depend on the configuration-space or phase-space variables. The summation symbols here represent sums over all discrete labels and an integral with respect to each continuous label. With respect to these basis functions, $\phi^{(1)}$ is represented by the coefficients $\alpha_n(t)$, and $g_s(t)$ is represented by the coefficients $\gamma_{sr}(t)$:

$$\alpha_n(t)=(\zeta_n, \phi^{(1)}(t)), \quad \gamma_{sr}(t)=(w_{sr}, g_s(t)). \tag{IV.10}$$

In these representations, the operators $\mathcal{L}_s$ and Λ become diagonal matrices:

$$\langle r'|\mathcal{L}_s|r\rangle=(w_{sr'}, \mathcal{L}_s w_{sr})=i\mu_{sr}\delta_{rr'}, \tag{IV.11}$$

$$\langle n'|\Lambda|n\rangle=(\zeta_{n'}, \Lambda\eta_n)=\lambda_n\delta_{nn'}. \tag{IV.12}$$

The dispersion matrix corresponding to the dispersion operator, Eq. (III.8), is

$$\langle n'|D|n\rangle=\lambda_n\delta_{nn'}+\sum_{sr}\frac{i\langle n'|J_s^\dagger|r\rangle\langle r|\hat{W}_s(\omega)|n\rangle}{\mu_{sr}-\omega}, \tag{IV.13}$$

where we have used the notation

$$\langle n'|J_s^\dagger|r\rangle=(\zeta_{n'}, J_s^\dagger w_{sr}), \tag{IV.14}$$

$$\langle r|\hat{W}_s(\omega)|n\rangle=(w_{sr}, \hat{W}_s(\omega)\eta_n). \tag{IV.15}$$

(Recall that the operator $J_s^\dagger$ operates on a function in phase space to produce a function in configuration space, i.e., a field function, while $\hat{W}_s$ does the converse.)

We now specialize to the case of an equilibrium in which the distribution function $f_s^{(0)}$ is written as a function of the global invariants associated with the unperturbed Hamiltonian $H_s(q,p,\phi^{(0)})$, as was discussed in the latter part of Sec. II. The weight function $\Gamma_s^{(p)}$ in Eq. (IV.2) must also be a function of the global invariants. Since the operators Λ and $\mathcal{L}_s$ are not functions of the ignorable coordinates q_k, they commute with the operators $\partial/\partial q_k$. (As discussed in the latter part of Sec. II, the index k will always be associated with ignorable coordinates.) Therefore the eigenfunctions η_n and w_{sr} can be taken to be also eigenfunctions of the (anti-Hermitian) operators $\partial/\partial q_k$. $\mathcal{L}_s$ also commutes with the constants of the motion p_k and H_s. The eigenfunctions η_n and w_{sr} can therefore be written in the form

$$\eta_n=y_n(Q)\Pi_k N_k^{-1/2}e^{i\kappa_k q_k}, \tag{IV.16}$$

$$w_{sr}=u_{sr}(Q,P)\ \delta(H_s-E_s)\Pi_k\delta(p_k-p_k^0)N_k^{-1/2}e^{i\kappa_k q_k}, \tag{IV.17}$$

where Q and P stand for sets of nonignorable coordinates and corresponding momenta, $i\kappa_k$ are eigenvalues of $\partial/\partial q_k$, p_k^0 and E_s are values of p_k and H_s, and $N_k^{1/2}$ is a normalizing factor for $\exp(i\kappa_k q_k)$. The index n stands for a list which includes the eigenvalues κ_k; the index r stands for a list which includes the eigenvalues κ_k, p_k^0, and E_s. The dual functions can be written likewise:

$$\zeta_n=z_n(Q)\Pi_k N_k^{-1/2}e^{i\kappa_k q_k}. \tag{IV.18}$$

If Λ is Hermitian, $z_n(Q)=y_n(Q)$. If we make the choice of P_s given by Eq. (II.33), and use Eqs. (II.34), (IV.16) and (IV.17), then the matrix element (IV.15) becomes

$$\langle r|\hat{W}_s(\omega)|n\rangle=i\delta_{\kappa\kappa'}K_{sr}(\omega)H'_{srn}, \tag{IV.19}$$

where

$$K_{sr}(\omega)=\Gamma_s^{(p)}(E_s,p_k^0)\left[\omega\frac{\partial \mathcal{J}_s}{\partial H_s}+\sum_k\kappa_k\frac{\partial \mathcal{J}_s}{\partial p_k}\right]_{H_s=E_s,\ p_k=p_k^0}, \tag{IV.20}$$

$$H'_{srn}=\int dQ\,dP\frac{\partial H_s}{\partial\phi^{(0)}}\delta(H_s-E_s)u_{sr}^*(Q,P)y_n(Q)\big|_{p_k=p_k^0}, \tag{IV.21}$$

and $\delta_{\kappa\kappa'}$ is a product of Kronecker and Dirac delta functions, where κ is the set of eigenvalues included in r, and κ' is the set included in n. In the same way, using Eq. (II.26), the matrix element (IV.14) becomes

$$\langle n'|J_s^\dagger|r\rangle=\delta_{\kappa\kappa'}J^*_{srn'}, \tag{IV.22}$$

where

$$J^*_{srn'}=\int dQ\,dP\,\Gamma^{(c)}\mathcal{J}_s(Q,P,p_k)\delta(H_s-E_s)\times z^*_{n'}(Q)u_{sr}(Q,P)\big|_{p_k=p_k^0}, \tag{IV.23}$$

and where we have assumed that $\mathcal{J}_s$ does not depend upon the ignorable coordinates q_k.

The dispersion matrix (IV.13) can now be written

$$\langle n'|D|n\rangle=\lambda_n\delta_{nn'}-\delta_{\kappa\kappa'}\sum_{s(r)}\frac{K_{sr}(\omega)J^*_{srn'}H'_{srn}}{\mu_{sr}-\omega}, \tag{IV.24}$$

where the notation (r) means the sum is only over those labels in r other than the κ_k, and where κ refers to the κ_k included in n, and κ' refers to those in n'.

For a completely Hamiltonian system in which Λ is Hermitian, $z_n=y_n$ and $\mathcal{J}_s=\partial H_s/\partial\phi^{(0)}$ [Eq. (II.27)]; if we also assume that $\Gamma^c=1$, then

$$J_{srn}=H'_{srn}. \tag{IV.25}$$

In this case Eq. (IV.24) becomes

$$\langle n'|D|n\rangle=\lambda_n\delta_{nn'}-\delta_{\kappa\kappa'}\sum_{s(r)}\frac{K_{sr}(\omega)H'^*_{srn'}H'_{srn}}{\mu_{sr}-\omega}. \tag{IV.26}$$

We remark that with the choice (II.33) for P_s, Λ defined by Eq. (II.11) is manifestly Hermitian and λ_n is real. [See Eq. (II.26), and note that, for a completely Hamiltonian system, the field operator $K(t)$ is Hermitian.]

In the representation we have chosen, all matrices are diagonal in the indices κ_k corresponding to the ignorable coordinates. We may therefore reduce the dimensions of all matrices by fixing the values of the eigenvalues κ_k, and omitting κ_k from the indices. This is the advantage of choosing to diagonalize simultaneously the operators $\partial/\partial q_k$.

B. Hermitian property of $D(\omega)$

The matrix $D(\omega)$ given by Eq. (IV.26) is evidently a Hermitian matrix function in the sense defined in Appendix E,

$$D^\dagger(\omega)=D(\omega), \tag{IV.27}$$

where $D^\dagger(\omega)$ is the adjoint matrix function. [The adjoint matrix function $D^\dagger(\omega)$ is the adjoint of the matrix $D(\omega)$ calculated as if ω were real.] The properties of such a

matrix are discussed in Appendix E and are used in the next section. A particular property is that the roots of the dispersion relation

$$\det D(\omega) = 0 \tag{IV.28}$$

are either real or occur in conjugate pairs. It can also be shown that the equation

$$\langle n | D(\omega_n) | n \rangle = 0 \tag{IV.29}$$

gives a variational formula for the frequency ω_n of the normal mode whose potential is given by the function η_n.[10] By this we mean that, if η_n is an approximation of order ϵ to the normal mode potential, then the solution of Eq. (IV.29) for ω_n will be accurate to order ϵ^2.

The dispersion *matrix* (IV.26) is a representation of the dispersion *operator* corresponding to a particular choice of orthonormal basis functions. A unitary transformation [even a transformation by a unitary matrix function $U(\omega)$] preserves adjoints. It follows that the Hermitian property (IV.27) holds under any change of orthonormal basis functions η_n in which we expand the potential $\phi^{(1)}$. In particular, according to Eq. (III.9), the *operator* $D(\omega)$ is independent of the choice of P_s in the transformation (II.7). The eigenfunctions η_n of Λ [Eq. (II.11)] depend on P_s, and hence different choices of P_s (including $P_s = 0$!) amount to different choices of representation of the same operator $D(\omega)$. Therefore, the property (IV.27) must hold for the *matrix* $D(\omega)$ resulting from any choice of P_s. The choice (II.33) of P_s makes the form of the matrix $D(\omega)$ especially transparent.

C. Case of one nonignorable coordinate

The matrix $D(\omega)$ [Eq. (IV.24)] is diagonal in the indices referring to the ignorable coordinates. If all coordinates are ignorable, the eigenfunctions (IV.16) and (IV.17) reduce to

$$\eta_\kappa = \Pi_k N_k^{-1/2} e^{i\kappa_k q_k}, \tag{IV.30}$$

$$w^0_{s\kappa p} = (\Gamma_s^{(p)})^{-1/2} \Pi_k \delta(p_k - p_k^0) N_k^{-1/2} e^{i\kappa_k q_k}, \tag{IV.31}$$

where we have replaced the indices n and r by κ and (κ, p^0). Note that in this case the energy is not an independent constant of the motion, and we may take $f_s^{(0)}$ to depend just on the ignorable momenta p_k.

The Liouville operator is

$$\mathcal{L}_s = \sum_k \frac{\partial H_s}{\partial p_k}\frac{\partial}{\partial q_k}, \tag{IV.32}$$

and

$$\mu_{\kappa p^0} = \sum_k \kappa_k \frac{\partial H_s}{\partial p_k}\bigg|_{p=p^0}. \tag{IV.33}$$

All matrices are diagonal in the indices κ; and the matrix $D(\omega)$ is completely diagonal, with diagonal elements

$$\langle \kappa | D(\omega) | \kappa \rangle = \lambda_\kappa - \sum_s \int dp\, \mathcal{J}_s \frac{\partial H_s}{\partial \phi^{(0)}} \sum_k \kappa_k \left(\frac{\partial f_s^{(0)}}{\partial p_k}\right) \Big/ \left[\sum_k \kappa_k \left(\frac{\partial H_s}{\partial p_k}\right) - \omega\right]. \tag{IV.34}$$

The dispersion relation (III.24) requires for each κ that the matrix element (IV.34) vanish if ω is not real; this is the familiar form for the dispersion relation for a homogeneous plasma.

We are interested in inhomogeneous equilibria, in which case at least one coordinate is not ignorable. If there is only one nonignorable coordinate, we can still find the Liouville eigenfunctions w_{sr}, as shown in Appendix D. Since all matrices are diagonal in the indeces κ, we simplify the notation by holding the indices κ fixed throughout the remainder of the discussion, and omitting them from the index lists n and r. We consider submatrices corresponding to fixed values of κ. When we substitute Eq. (IV.16) in Eq. (IV.4), it becomes an equation in one independent variable Q, and the index n is a single discrete or continuous label for the eigenvalues λ_n. The label r includes the ignorable momenta p, the energy E, and an index labeling the eigenvalue μ_{sr}, as discussed in Appendix D. If there is more than one orbit corresponding to the same (p, E), we need another index, a, to distinguish them. To simplify the notation, we omit the index a, understanding that an integral over E always implies also a sum over a.

The eigenfunctions u_{sr} are given by Eqs. (D.17) – (D.20). For periodic orbits,

$$u_{spEl} = T_{spE}^{-1/2} (\Gamma_s^{(p)})^{-1/2} \exp[-i\tilde{G}_{spE}(\tau)] \exp(il\Omega_{spE}\tau), \quad l = 0, \pm 1, \pm 2, \ldots, \tag{IV.35}$$

where

$$\Omega_{spE} = 2\pi / T_{spE}, \tag{IV.36}$$

T_{spE} is the period of the motion in the (Q, P) plane, and $\tilde{G}_{spE}(\tau)$ is given by Eq. (D23). The matrix elements (IV.21) become

$$H'_{spE\,ln} = T_{spE}^{-1/2} (\Gamma_s^{(p)})^{-1/2} \int d\tau\, y_n \frac{\partial H_s}{\partial \phi^{(0)}} \exp[i\tilde{G}_{spE}(\tau)] \times \exp(-il\Omega_{spE}\tau), \tag{IV.37}$$

where Q in $y_n \partial H_s / \partial \phi^{(0)}$ is to be replaced by $Q(\tau)$ for the orbit spE. The matrix element (IV.37) is essentially the Fourier transform of $y_n \partial H_s / \partial \phi^{(0)} \exp(i\tilde{G}_{spE})$. A similar expression can be written for J^*_{spEln}; in the completely Hamiltonian case, it is the conjugate of H'_{spEln} given by Eq. (IV.37), as noted previously. For aperiodic orbits,

$$u_{spE\mu} = (2\pi)^{-1/2} (\Gamma_s^{(p)})^{-1/2} \exp[-iG_{spE}(\tau)] e^{i\mu\tau}, \tag{IV.38}$$

where μ is a continuous index, and $G_{spE}(\tau)$ is given by Eq. (D.20). The matrix elements (IV.21) are

$$H'_{spE\,\mu n} = (2\pi)^{-1/2} (\Gamma_s^{(p)})^{-1/2} \int d\tau\, y_n \frac{\partial H_s}{\partial \phi^{(0)}} \times \exp[iG_{spE}(\tau)] e^{-i\mu\tau}. \tag{IV.39}$$

The submatrix $D(\omega)$ can now be written as

$$\langle n | D(\omega) | n' \rangle = \lambda_n \delta_{nn'} - \sum_{sl} \int dp dE \frac{K_{spE}(\omega) J^*_{spEln} H'_{spEln'}}{l\Omega_{spE} + \beta_{spE} - \omega} - \sum_s \int dp dE d\mu \frac{K_{spE}(\omega) J^*_{spE\mu n} H'_{spE\mu n'}}{\mu - \omega}, \tag{IV.40}$$

where $\Omega_{spE} = 2\pi / T_{spE}$ is the frequency of motion in the (Q, P) phase plane for the periodic orbit (E, p), and β_{spE} is the quantity defined by Eq. (D.21). Each element includes a sum over l for periodic orbits and an integral over μ for aperiodic orbits; the integrals are under-

stood to run over the corresponding intervals in E and p. When there is only one nonignorable coordinate Q, orbits which are aperiodic in the (Q,P) plane must extend to $|Q|\to\infty$. ($|P|\to\infty$ would be a pathological case.) Hence if the equilibrium orbits are confined to a finite region in the (Q,P) plane, they are necessarily periodic and the last term in Eq. (IV.40) does not occur. Examination of Eq. (IV.20) shows that in the last term in Eq. (IV.40), if it occurs, the integrals over p and E and the sum over s can be carried out independently of the value of ω, leading to a particularly simple form involving only an integral over μ; unfortunately that simplification does not arise in the second term.

We see that the elements $\langle n|D(\omega)|n'\rangle$ are analytic in ω except along the real axis. The last term, if present, leads to a cut extending over the entire real ω axis. Each term in the sum over l and s leads in general to a cut along that part of the real ω axis where ω/l corresponds to a frequency of periodic orbits present in the equilibrium distribution. In the special case of simple harmonic oscillations, when $\Omega_s=\Omega_{spE}$ is independent of E and p, the singularity reduces to a pole at $\omega=l\Omega_s+\beta_{spE}$. If $f_s^{(0)}(E,p)$ is constant over some region in (E,p) space, Eq. (IV.20) shows that the cut along the ω axis for the second term in Eq. (IV.40) is not present for the corresponding interval in $l\Omega_{spE}$. There are other special cases. For example, if $f_s^{(0)}(E,p)$ is piecewise constant in E, and if Ω_{spE} depends only on E, the second term in Eq. (IV.40) leads to simple poles in ω. If the integrals are approximated by sums, as for numerical purposes, the resulting matrix elements again have only simple poles.

V. PROPERTIES OF THE SOLUTION OF THE INITIAL-VALUE PROBLEM

The solution of the initial-value problem is obtained by inverting the Laplace transforms given by Eqs. (III.6) and (III.7). We shall focus our attention on the solution for the potential $\phi^{(1)}$. The solution for the auxiliary function g_s could be obtained also; it leads to somewhat greater algebraic complexity.

If we rewrite Eq. (III.6) in the matrix form developed in the last section, we may obtain the following equation for the potential coefficients $\hat{\alpha}_n$,

$$\sum_{n'}\langle n|D(\omega)|n'\rangle\hat{\alpha}_{n'}(\omega)=I_n(\omega), \tag{V.1}$$

where the vector $I_n(\omega)$ is determined by the initial conditions

$$I_n(\omega)=\sum_{s(\tau)}\frac{J^*_{s\tau n}[(w_{s\tau},\Phi_s(\omega))+\gamma_{s\tau}(0)]}{i(\mu_\tau-\omega)}. \tag{V.2}$$

In order to bring out the character of the solutions of Eq. (V.1) for the potential coefficients $\alpha_n(t)$, without involving ourselves in needless algebraic complexity, we shall consider only the case treated at the end of the last section of a completely Hamiltonian system with one nonignorable coordinate, and assume that all equilibrium orbits are periodic in the (Q,P) plane. It will be clear that most of our results apply also in the general case. For systems which are not completely Hamiltonian, those results which depend on the Hermitian character of $D(\omega)$ may not apply. Using the notation introduced at the end of the last chapter, we hold the ignorable indices κ_k fixed and restrict our attention to the corresponding submatrices. The vector I_n then becomes

$$I_n(\omega)=\sum_{sl}\int dp\,dE\,\frac{J^*_{spEln}A_{spEl}(\omega)}{i(l\Omega_{spE}+\beta_{spE}-\omega)}, \tag{V.3}$$

where

$$A_{spEl}(\omega)=T^{-1/2}_{spE}\int d\tau\exp[i\tilde{G}_{spE}(\tau)]\exp(-il\Omega_{spE}\tau)\times[\Phi_s(\omega)+g_s(0)], \tag{V.4}$$

where Q and P in the functions $\Phi_s(\omega)$ and $g_s(0)$ are to be expressed as functions of p,E, and τ. The solution for $\alpha(t)$ is obtained by solving Eq. (V.1) and inverting the Laplace transform,

$$\alpha_n(t)=(2\pi)^{-1}\int_C d\omega\sum_{n'}\langle n|D^{-1}(\omega)|n'\rangle I_{n'}(\omega)e^{-i\omega t}, \tag{V.5}$$

where the integral is over a Bromwich contour C parallel to the real ω axis, above all singularities of the integrand in the ω plane.

A. Van Kampen and Landau solutions

Because of the integrals over p and E in Eqs. (IV.40) and (V.3), the quantities $\langle n|D(\omega)|n'\rangle$ and $I_n(\omega)$ have cuts on the real ω axis, although they are not singular there. In order to close the contour in Eq. (V.5) without having to cross these cuts, we proceed as follows. Add a contour C' antiparallel to the real axis at $\mathrm{Im}\omega=-\infty$. The contribution to the integral from this contour evidently vanishes. Now depress the contour C to the real axis, while raising the contour C' to the real axis. We are left with residues from the zeros of $\det D(\omega)$, which occur at conjugate points in the upper and lower half-plane, and an integral over the real ω axis of the discontinuity in the integrand across the axis. The resulting solution has the general form,

$$\alpha_n(t)=\sum_b A_b C_{nb}e^{-i\omega_b t}+\int d\omega\sum_a A_a(\omega)C_{na}(\omega)e^{-i\omega t}. \tag{V.6}$$

The terms in the sum represent the (discrete) normal modes of the system. The coefficients A_b are complex amplitudes depending on the initial conditions through $I_n(\omega)$. The coefficients C_{nb} determine the character of each normal mode b; that is, they determine the relative amplitudes and phases by which each potential component η_n participates in the mode b; they do not depend on the initial conditions. Explicit formulas for the coefficients A and C will be given later. In the degenerate case when two or more roots ω_b of Eq. (IV.28) coincide, the corresponding term in the sum in Eq. (V.6) may involve a polynomial in t. The integral is over the cut along the real axis, and represents the van Kampen modes, which are singular modes with real frequencies. The real frequencies ω are in general degenerate; that is, there are many van Kampen modes at each frequency ω. The sum over a corresponds to this degeneracy. The amplitudes $A_a(\omega)$ again depend on the initial conditions, while the coefficients $C_{na}(\omega)$ which describe the character of the mode do not. Instabilities can only come from the (discrete) normal modes. Phase mixing of the van Kampen modes can lead to damping (e.g., Landau damping).

If, for some value of the index n, the nth row and

column of the matrix $D(\omega)$ have no off-diagonal elements, then it is easy to show that, corresponding to this index n, there can be no (discrete) normal mode on the real axis with ω coinciding with the frequency of any van Kampen mode. [The element $\langle n|D(\omega \pm i0)|n'\rangle$ has a finite imaginary part if $A_a(\omega) \neq 0$.] This is the case, for example, for a homogeneous plasma. In such cases, the solution for a stable plasma consists only of van Kampen modes. We have not yet proved the corresponding result when the nth row or column of the matrix $D(\omega)$ contains off-diagonal terms.

An alternative way to close the contour C, adopted by Landau, is simply to depress the contour all the way to negative imaginary infinity in the ω plane. As the contour C crosses the cut on the real axis, it enters a different Riemann sheet in the lower half-plane. On that sheet, the integrals over E in Eqs. (V.3) and (IV.40) are defined by analytic continuation in ω; that is, the E-integration is carried out over a contour in the E plane which is deformed so as to keep $l\Omega_{s\not b E} + \beta_{s\not b E}$ below the point ω. The resulting analytically continued D matrix is no longer a Hermitian function of ω. The resulting expression for $\alpha_n(t)$ has still the form (V.6). It contains a sum over normal modes, generally including Landau-damped modes in the lower half-plane. The integral over continuum modes is now to be carried out along vertical cuts extending to negative imaginary infinity from branch points on the real axis.

The Landau and van Kampen forms of solution are of course equivalent. It is a matter of taste and convenience which is chosen. In the nonrelativistic, homogeneous case, if the distribution functions are analytic functions of p and E, there are no branch points on the real axis, and the Landau solution is simpler, the integral over van Kampen modes being replaced in general by a sum over a few discrete Landau-damped normal modes. However, in the inhomogeneous case, there are in general branch points, so that the Landau solution also contains integrals around the vertical cuts, which dominate the solution in the stable case. The asymptotic form of the solution for a stable, inhomogeneous equilibrium will be presented in a later paper.[10] The Landau form has the advantage of exhibiting explicitly the long-time behavior of the solution. The van Kampen form has the advantage that it preserves the Hermitian character of $D(\omega)$ and hence the symmetry between the upper and lower half ω planes.

When these solutions are approximated in numerical calculations, the integrals in Eq. (V.3) and in formula (IV.40) for the matrix elements $\langle n|D(\omega)|n'\rangle$ are replaced by sums. There are no longer any cuts in the ω plane, and the integral in Eq. (V.6) is replaced by a sum over discrete normal modes. (Refer to Sec. III for some additional remarks about numerical approximations.) This sum is an approximation to the integral; if the numerical approximations to the integrals in Eqs. (IV.40) and (V.3) are adequate, the resulting solution will be an adequate approximate solution to the initial-value problem for some finite time interval. In such a numerical approximation, the Landau sheet in the lower half ω plane is no longer directly accessible. The problem of identifying the Landau-damped modes in the numerical solution will not be addressed in the present paper.

B. Discrete normal mode–typical (nondegenerate) case

We assume in what follows that the potentials can be represented with sufficient accuracy by some finite number N of eigenfunctions of Λ, so that the matrix $\langle n|D(\omega)|n'\rangle$ is of finite dimension $N \times N$. We could evaluate the reciprocal matrix $D^{-1}(\omega)$ in the standard way as the matrix of cofactors of the transpose divided by the determinant. The normal modes would then arise from the residues at the roots of the dispersion relation (IV.28).

It is more convenient to invert the matrix $D(\omega)$ by first diagonalizing it,

$$U^{-1}(\omega)\, D(\omega)\, U(\omega) = d(\omega), \tag{V.7}$$

where $d(\omega)$ is diagonal, with diagonal elements $d_m(\omega)$. The matrix $U(\omega)$ which diagonalizes $D(\omega)$ is a unitary function in the sense defined in Appendix E. The column $U^m(\omega)$ of the matrix $U(\omega)$ is an eigenvector of $D(\omega)$ corresponding to the eigenvalue $d_m(\omega)$. If ω_b is a simple root of the dispersion relation (IV.28), then one of the eigenvalues, say $d_b(\omega)$ will vanish at $\omega = \omega_b$. The reciprocal $D^{-1}(\omega)$ can be written, by making use of Eq. (V.7), in the form

$$\langle n|D^{-1}(\omega)|n'\rangle = \sum_m \langle n|U(\omega)|m\rangle\langle m|U^{-1}(\omega)|n'\rangle/d_m(\omega). \tag{V.8}$$

It is now evident from Eqs. (V.5) and (V.8) that the normal mode has the form given by a term in the sum in Eq. (V.6), and that the coefficients are given by

$$C_{nb} = \langle n|U(\omega_b)|b\rangle/d_b'(\omega_b), \tag{V.9}$$

$$A_b = -i\sum_n \langle b|U^{\dagger}(\omega_b)|n\rangle I_n(\omega_b), \tag{V.10}$$

where

$$d_b'(\omega) = \frac{\partial d_b(\omega)}{\partial \omega}. \tag{V.11}$$

C. The continuum–van Kampen modes

The integral over the real ω axis in Eq. (V.6) comes from the integral in Eq. (V.5) taken just above the cut along the real axis minus the same integral taken just below the cut. If we write the matrix $D^{-1}(\omega)$ in the form (V.8), the coefficients A_a and C_{na} are

$$C_{nm} = (2\pi)^{-1}\langle n|U(\omega)|m\rangle/d_m(\omega), \tag{V.12}$$

$$C_{n,-m} = (C_{nm})^*, \tag{V.13}$$

$$A_m = \sum_n \langle m|U^{\dagger}(\omega)|n\rangle I_n(\omega), \tag{V.14}$$

$$A_{-m} = \sum_n \langle n|U(\omega)|m\rangle I_n(\omega - i0), \tag{V.15}$$

where all functions of ω are to be evaluated at the real axis just above the cut, except for $I_n(\omega - i0)$ in Eq. (V.15), which is to be evaluated just below the cut. The index m runs over the values $1, 2, \ldots, N$ corresponding to the eigenfunctions of Λ. We have chosen to let the index $a = m$ [in Eq. (V.6)] correspond to the integral just above the cut, and to let $a = -m$ correspond to the integral below the cut. In Eqs. (V.13) and (V.15), we have used the fact that $D(\omega)$ is a Hermitian function (hence $d_m(\omega)$ is a real function), and that $U(\omega)$ is a unitary function. The necessary modifications when this is not the case are obvious.

We could equally well have taken the inverse of the matrix $D(\omega)$ in the standard way. The resulting formulas would then have been

$$C_{n\alpha} = (2\pi)^{-1}\langle n|M(\omega)|\alpha\rangle/\det D(\omega), \tag{V.16}$$

$$A_\alpha = I_\alpha(\omega), \tag{V.17}$$

where $\langle n|M(\omega)|m\rangle$ is the matrix of cofactors of the transpose of $D(\omega)$, and the quantities on the right are to be evaluated at $\omega + i0$ if $\alpha = m$, and at $\omega - i0$ if $\alpha = -m$. Equations (V.16) and (V.17) correspond to another choice of the degenerate modes.

When the Landau procedure for shrinking the contour C in the integral (V.5) is chosen, the continuum part of the solution plus the contributions from the poles below the real ω axis are rewritten as a sum of Landau-damped modes, plus integrals over vertical cuts extending into the lower half-plane from branch points on the real axis. The latter integrals can be written in a form analogous to that given above for the van Kampen modes. The result is a sum of integrals over a continuum of Landau-damped modes, each integral being over modes with a discrete real frequency and a continuum of damping constants from zero to infinity.

D. Discrete normal modes–degenerate case

By *degenerate* here, we refer to a mode or set of modes corresponding to a multiple root ω of the dispersion relation (IV.28). When the dispersion relation (IV.28) has an n-fold root ω_b, the integrand in Eq. (V.5) will have an nth-order pole. The corresponding term in the solution (V.6) will then generally have an $(n-1)$st-order polynomial in t as the coefficient C_{nb}. An exception occurs when the dispersion matrix $D(\omega)$ has n independent eigenvectors whose n eigenvalues each have a simple zero at ω_b. There are then n independent normal modes of the form shown in the first term in Eq. (V.6), which happen to have the same frequency ω_b. Otherwise, that is, if some eigenvalue of $D(\omega)$ has a multiple zero at ω_b, or if the corresponding eigenvectors are not independent, a polynomial in t will in general occur in the solution. We shall not carry through the corresponding calculations in this paper.

Degeneracies in the eigenvalues of the dispersion matrix $D(\omega)$ are discussed in Appendix E. They do not ultimately affect the character of the solution given by Eq. (V.5), as we see if we note that the inverse $D^{-1}(\omega)$ can be written in the standard way without introducing the eigenvectors of $D(\omega)$ at all. However, if we write $D^{-1}(\omega)$ in the form (V.8), then degenerate eigenvalues must be taken into account. If the eigenvalue $d_m(\omega)$ is n-fold degenerate at ω_b, then $d_m(\omega)$ has in general an n-fold branch point at ω_b, and the transformation matrix $U(\omega)$ has a $2n$-fold branch point. (See Appendix E.) These branch points do not occur in the complete sum over m in Eq. (V.8), but they do occur in the individual terms. We omit the details here. As pointed out in Appendix E, the N-valued function $d_m(\omega)$ has typically double branch points at discrete points in the ω plane, wherever two eigenvalues $d_m(\omega)$ coincide and the corresponding two Riemann sheets are stitched together. Higher-order branch points are exceptional cases, as are cases when a branch point happens to coincide with a zero eigenvalue. Since the sum in Eq. (V.6) is over points ω_b where $d_m(\omega_b) = 0$, we do not ordinarily need to worry about this case. Note however that certain methods of searching for zero eigenvalues $d_m(\omega)$ may involve encircling branch points. (Such methods are discussed by Holdren,[7] who, however, does not mention the problem of branch points.)

ACKNOWLEDGMENTS

We wish to acknowledge helpful discussions related to this work with B. L. Buzbee, R. C. Davidson, J. P. Freidberg, R. A. Gerwin, G. H. Golub, J. L. Schwarzmeier, B. A. Shrauner, and L. Turner.

Part of this work was done while one of us (HRL) was a Visiting Professor at the University of Wisconsin—Madison. He thanks the university for its support during that period.

This work was supported by the United States Energy Research and Development Administration at the Los Alamos Scientific Laboratory and at the University of Wisconsin—Madison.

APPENDIX A: THE LAGRANGE AND HAMILTON EQUATIONS IN VECTOR FORM AND HAMILTON'S PRINCIPLE

In this appendix we give some basic relations that pertain to the vector form of the Lagrange and Hamilton equations of motion for a particle; and we write Hamilton's principle for a completely Hamiltonian system in the form that is particularly relevant to the formulation in Secs. III and IV of the linearized initial-value problem.

For writing the Lagrange and Hamilton equations in vector form, as well as for writing the Liouville operator or for discussing Hamilton's principle, it is useful to introduce gradient operators with respect to velocity or canonical momentum variables. The detailed meaning of any one of the various gradient operators depends on what variables have been chosen as formally independent arguments for the function on which the gradient operates. We consider functions which may depend on the position vector $\mathbf{r}$ and on a vector that we denote generically by $\boldsymbol{\sigma}$. In our discussions, the vector $\boldsymbol{\sigma}$ may be a velocity vector, or it may be a canonical momentum vector as defined in Sec. II by Eq. (II.12) or in this appendix by Eq. (A7). We assume that dependence on $\boldsymbol{\sigma}$ is explicit. However, in addition to explicit dependence on $\mathbf{r}$, the function may also depend implicitly on $\mathbf{r}$ through explicit dependence on the array of potential functions ϕ. There also may be dependence on time, but it is not important for our discussion of the gradient operators and we suppress it.

Let $G[\mathbf{r}, \boldsymbol{\sigma}, \phi(\mathbf{r})]$ be an arbitrary function of formally independent arguments $\mathbf{r}$, $\boldsymbol{\sigma}$, and $\phi(r)$. We define the operators $\nabla_{\mathbf{r}}$ and $\nabla_{\boldsymbol{\sigma}}$ such that the change δG due to infinitesimal changes $\delta\mathbf{r}$ and $\delta\boldsymbol{\sigma}$ is given by

$$\delta G[\mathbf{r}, \boldsymbol{\sigma}, \phi(\mathbf{r})]$$

$$= \sum_i \delta\mathbf{r}\cdot(\nabla\phi_i)\frac{\partial G}{\partial\phi_i} + \delta\mathbf{r}\cdot\nabla_{\mathbf{r}}G[\mathbf{r}, \boldsymbol{\sigma}, \phi] + \delta\boldsymbol{\sigma}\cdot\nabla_{\boldsymbol{\sigma}}G[\mathbf{r}, \boldsymbol{\sigma}, \phi]$$

$$= \delta\mathbf{r} \circ (\nabla\phi) \frac{\partial G}{\partial \phi} + \delta\mathbf{r} \circ \nabla_{\mathbf{r}} G[\mathbf{r}, \sigma, \phi] + \delta\sigma \circ \nabla_{\sigma} G[\mathbf{r}, \sigma, \phi]. \tag{A1}$$

Thus $\nabla_{\mathbf{r}}$ is the gradient operator with respect to $\mathbf{r}$ holding the array ϕ and the vector σ fixed, and ∇_{σ} is the gradient operator with respect to σ holding the array ϕ and the vector $\mathbf{r}$ fixed; ∇ is the usual gradient operator with respect to all occurrences, explicit or implicit, of $\mathbf{r}$.

For convenience, we choose an orthogonal coordinate system whose unit vectors are $\hat{\mathbf{e}}_j(\mathbf{r})$, we denote the coordinate of $\mathbf{r}$ associated with $\hat{\mathbf{e}}_j$ by q_j, and we express the derivative of $\mathbf{r}$ with respect to q_j as usual in terms of a function $h_j(\mathbf{r})$ by

$$\frac{\partial \mathbf{r}}{\partial q_j} = h_j(\mathbf{r})\, \hat{\mathbf{e}}_j(\mathbf{r}). \tag{A2}$$

A result of this expression for $\partial\mathbf{r}/\partial q_j$ is

$$\delta q_j = \delta\mathbf{r} \cdot \nabla q_j = \delta\mathbf{r} \cdot (\hat{\mathbf{e}}_j / h_j). \tag{A3}$$

1. The equations of motion

Each particle of species s moves according to an equation of motion derivable from a Lagrangian $L_s[\mathbf{r}, \dot{\mathbf{r}}, \phi(\mathbf{r}, t)]$, depending explicitly on some potential functions ϕ_i. These potentials are in turn determined by the particle currents and densities. For example, a particle acted on by electromagnetic forces only has the Lagrangian

$$L_s[\mathbf{r}, \dot{\mathbf{r}}, \phi(\mathbf{r}, t)] = \tfrac{1}{2} M_s \dot{\mathbf{r}}^2 - Q_s \Phi(\mathbf{r}, t) + (Q_s/c)\, \dot{\mathbf{r}} \cdot \mathbf{A}(\mathbf{r}, t), \tag{A4}$$

where the electromagnetic potentials Φ and $\mathbf{A}$ comprise the array ϕ. Our formalism also allows L_s to contain other forces (e.g., external forces) that are derivable from a generalized potential. The Lagrange equation of motion can be written as

$$\frac{d}{dt} \nabla_{\dot{r}} L_s - \nabla_{\mathbf{r}} L_s - (\nabla\phi) \frac{\partial L_s}{\partial \phi} = 0, \tag{A5}$$

where the gradient operators are as defined by Eq. (A1).

The Hamiltonian associated with the Lagrangian L_s is defined by

$$H_s[\mathbf{r}, \rho, \phi(\mathbf{r}, t)] = \rho \cdot \dot{\mathbf{r}} - L_s[\mathbf{r}, \dot{\mathbf{r}}, \phi(\mathbf{r}, t)], \tag{A6}$$

where the *canonical momentum vector* ρ is defined by

$$\rho = \nabla_{\dot{r}} L_s[\mathbf{r}, \dot{\mathbf{r}}, \phi(\mathbf{r}, t)], \tag{A7}$$

and where the arguments $\mathbf{r}$, ρ, and $\phi(\mathbf{r}, t)$ of H_s and L_s are considered to be formally independent. The canonical momentum vector is a vector generalization of the scalar canonical momenta. Its components are closely related to the usual generalized momenta p_j defined by

$$p_j = \frac{\partial L_s}{\partial \dot{q}_j}, \tag{A8}$$

where the variables q_j, $\dot{q}_j$, and ϕ_i are considered to be formally independent. The canonical momentum vector can be calculated from Eqs. (A7) and (A1); the result is

$$\rho = \sum_j \hat{\mathbf{e}}_j \left(\frac{1}{h_j}\right) \frac{\partial L_s}{\partial \dot{q}_j} = \sum_j \hat{\mathbf{e}}_j \left(\frac{1}{h_j}\right) p_j, \tag{A9}$$

so that

$$p_j = (h_j\, \hat{\mathbf{e}}_j) \cdot \rho. \tag{A10}$$

Hamilton's equations for a particle of species s with position vector $\mathbf{r}$ and canonical momentum vector ρ can be written in the vector form

$$\begin{aligned} \dot{\mathbf{r}} &= \nabla_{\rho} H_s, \\ \dot{\rho} &= -\nabla_{\mathbf{r}} H_s - (\nabla\phi) \frac{\partial H_s}{\partial \phi}. \end{aligned} \tag{A11}$$

The gradient operators are as defined by Eq. (A1).

The form of Hamilton's equations given by Eqs. (A11) is completely equivalent to the scalar form

$$\begin{aligned} \dot{q}_j &= \frac{\partial H_s}{\partial p_j}, \\ \dot{p}_j &= -\frac{\partial H_s}{\partial q_j} - \sum_i \frac{\partial \phi_i}{\partial q_j} \frac{\partial H_s}{\partial \phi_i}, \end{aligned} \tag{A12}$$

where H_s is considered to be an explicit function of the variables q_j, p_j, and ϕ_i. This can be seen by writing out the components of Eqs. (A11). A convenient way to obtain expressions for the gradients of H_s is to write the variation of H_s as

$$\delta H_s = \sum_i \delta\dot{\mathbf{r}}\, (\nabla\phi_i) \frac{\partial H_s}{\partial \phi_i} + \sum_j \left(\delta q_j \frac{\partial H_s}{\partial q_j} + \delta p_j \frac{\partial H_s}{\partial p_j} \right),$$

substitute expressions for δq_j and δp_j in terms of $\delta\mathbf{r}$ and $\delta\rho$ from Eqs. (A3) and (A10), and compare with Eq. (A1). The resulting expressions for $\nabla_{\mathbf{r}} H_s$ and $\nabla_{\rho} H_s$ are

$$\begin{aligned} \nabla_{\mathbf{r}} H_s &= \sum_j \left[\hat{\mathbf{e}}_j \left(\frac{1}{h_j}\right) \frac{\partial H_s}{\partial q_j} + (\nabla h_j \hat{\mathbf{e}}_j) \cdot \rho \frac{\partial H_s}{\partial p_j} \right] \\ &= \sum_j \left[\hat{\mathbf{e}}_j \left(\frac{1}{h_j}\right) \frac{\partial H_s}{\partial q_j} + (\nabla h_j \hat{\mathbf{e}}_j) \cdot \sum_i \hat{\mathbf{e}}_i \left(\frac{1}{h_i}\right) p_i \dot{q}_j \right], \\ \nabla_{\rho} H_s &= \sum_j \hat{\mathbf{e}}_j h_j \frac{\partial H_s}{\partial p_j}. \end{aligned}$$

Using this expression for $\nabla_{\rho} H_s$, the first of Eqs. (A11) can be written as

$$\sum_j \hat{\mathbf{e}}_j h_j \dot{q}_j = \sum_j \hat{\mathbf{e}}_j h_j \frac{\partial H_s}{\partial p_j},$$

which shows that the jth component of the first of Eqs. (A11) is h_j times the first of Eqs. (A12). In order to show the equivalence of the second of Eqs. (A11) and (A12), we write $\dot{\rho}$ as

$$\begin{aligned} \dot{\rho} &= \sum_j \left[\hat{\mathbf{e}}_j \left(\frac{1}{h_j}\right) \dot{p}_j + p_j \dot{\mathbf{r}} \cdot \nabla \left(\frac{\hat{\mathbf{e}}_j}{h_j}\right) \right] \\ &= \sum_j \left[\hat{\mathbf{e}}_j \left(\frac{1}{h_j}\right) \dot{p}_j + p_j \sum_i h_i \dot{q}_i \hat{\mathbf{e}}_i \cdot \nabla \left(\frac{\hat{\mathbf{e}}_j}{h_j}\right) \right], \end{aligned}$$

and use the above expression for $\nabla_{\mathbf{r}} H_s$ in the second of Eqs. (A11). We then use the identity

$$\begin{aligned} &\sum_j p_j \sum_i h_i \dot{q}_i \hat{\mathbf{e}}_i \cdot \nabla \left(\frac{\hat{\mathbf{e}}_j}{h_j}\right) + \sum_j (\nabla h_j \hat{\mathbf{e}}_j) \circ \sum_i \hat{\mathbf{e}}_i \left(\frac{1}{h_i}\right) p_i \dot{q}_j \\ &\quad = -\sum_{i,j} p_i \dot{q}_j (h_j \hat{\mathbf{e}}_j) \times \left[\nabla \times \left(\frac{\hat{\mathbf{e}}_i}{h_i}\right) \right] = 0 \end{aligned}$$

to obtain that the jth component of the second of Eqs. (A.11) is $1/h_j$ times the second of Eqs. (A.12).

2. Hamilton's principle for a completely Hamiltonian system

The Langrangian for a completely Hamiltonian system consistes of two parts,

$$L = L_{\text{field}} + L_{\text{particle}}, \tag{A13}$$

where L_{particle} includes the interaction of the field with the particles. The quantity L_{field} has the form

$$L_{\text{field}} = \int d^3\mathbf{r}\, \mathcal{L}_{\text{field}}, \tag{A14}$$

where the Lagranian density $\mathcal{L}_{\text{field}}$ is a functional of the potential functions $\phi_i(\mathbf{r}, t)$ and their derivatives, and, possibly, an explicit function of r and t in addition. Varying the time integral of L_{field} with respect to ϕ gives rise to the exact, perhaps nonlinear, field operator on ϕ that corresponds to the linear operator K in Eq. (II. 2). That is,

$$\begin{aligned}\delta\int_{t_1}^{t_2} L_{\text{field}}\,dt &= \delta\int_{t_1}^{t_2} dt \int d^3\mathbf{r}\,\mathcal{L}_{\text{field}} \\ &= \sum_i \int_{t_1}^{t_2} dt \int d^3\mathbf{r}[\bar{K}\phi]_i \delta\phi_i,\end{aligned} \tag{A15}$$

where $\bar{K}$ is the exact field operator.

In order to construct L_{particle}, we introduce a function $\mathbf{R}_s(\mathbf{r}', \boldsymbol{\rho}', t)$ as the position vector and a function $\mathbf{P}_s(\mathbf{r}', \boldsymbol{\rho}', t)$ as the canonical momentum vector of a particle of species s at time t whose initial position and canonical momentum vectors are $\mathbf{r}'$ and $\boldsymbol{\rho}'$, respectively. The quantity L_{particle} is then given by

$$\begin{aligned}L_{\text{particle}} = \sum_s \int d^3\mathbf{r}'\, d^3\boldsymbol{\rho}'\, f_s(\mathbf{r}', \boldsymbol{\rho}', 0) \\ \times\{\mathbf{P}_s(\mathbf{r}', \boldsymbol{\rho}', t)\cdot\dot{\mathbf{R}}_s(\mathbf{r}', \boldsymbol{\rho}', t) \\ - H_s[\mathbf{R}_s, \mathbf{P}_s, \phi(\mathbf{R}_s, t)]\},\end{aligned} \tag{A16}$$

where $f_s(\mathbf{r}', \boldsymbol{\rho}', 0)$ is the initial canonical phase-space distribution function. The quantity in curly brackets is just the single-particle Langranian expressed in terms of canonical variables. The quantities to be varied independently are $\mathbf{P}_s, \mathbf{R}_s$, and ϕ. Variation of the time integral of L_{particle} with respect to $\mathbf{P}_s$ gives

$$\begin{aligned}&\delta_{P_s}\int_{t_1}^{t_2} L_{\text{particle}}\,dt \\ &= \int_{t_1}^{t_2} dt \int d^3\mathbf{r}'\, d^3\boldsymbol{\rho}'\, f_s(\mathbf{r}', \boldsymbol{\rho}', 0)\{\dot{\mathbf{R}}_s - \nabla_{\mathbf{P}_s} H_s\}\cdot\delta\mathbf{P}_s.\end{aligned} \tag{A17}$$

Variation with respect to $\mathbf{R}_s$, requiring that $\delta\mathbf{R}_s$ vanish for $t = t_1$ or t_2, gives

$$\begin{aligned}&\delta_{\mathbf{R}_s}\int_{t_1}^{t_2} L_{\text{particle}}\,dt \\ &= -\int_{t_1}^{t_2} dt \int d^3\mathbf{r}'\, d^3\boldsymbol{\rho}'\, f_s(\mathbf{r}', \boldsymbol{\rho}', 0) \\ &\times\left(\dot{\mathbf{P}}_s + \nabla_{\mathbf{R}_s} H_s + \sum_i (\nabla_{\mathbf{R}_s}\phi_i)\frac{\partial H_s}{\partial\phi_i}\right)\cdot\delta\mathbf{R}_s.\end{aligned} \tag{A18}$$

Variation with respect to ϕ gives

$$\begin{aligned}&\delta_\phi\int_{t_1}^{t_2} L_{\text{particle}}\,dt \\ &= -\sum_s \int d^3\mathbf{r}'\, d^3\boldsymbol{\rho}'\, f_s(\mathbf{r}', \boldsymbol{\rho}', 0)\sum_i \frac{\partial H_s}{\partial\phi_i}\,\delta\phi_i(\mathbf{R}_s, t) \\ &= -\sum_s \int d^3\mathbf{r}\, d^3\boldsymbol{\rho} f_s(\mathbf{r}, \boldsymbol{\rho}, t)\sum_i \frac{\partial H_s}{\partial\phi_i}\,\delta\phi_i(\mathbf{r}, t).\end{aligned} \tag{A19}$$

In the second line, the arguments of H_s are $\mathbf{R}_s$, $\mathbf{P}_s$, and $\phi(\mathbf{R}_s, t)$; in the last line, the arguments are $\mathbf{r}$, $\boldsymbol{\rho}$, and $\phi(\mathbf{r}, t)$. In order to obtain that form of the variation, Liouville's theorem was used to make a change of variables of integration. The reason for expressing the Lagrangian in terms of position and canonical momentum, instead of in terms of the more usual position and velocity, is to ensure that Liouville's theorem can be used to change the variables of integration in this way. In most cases of interest, the analogous change of variables could in fact be made in terms of position and velocity. However, because we allow the possibility of generalized potentials in L_{particle}, there are cases for which position and velocity variables are unsuitable. Some further discussion of the difference between position and canonical momentum variables on the one hand, and position and velocity variables on the other, can be found in Appendix C.

By virtue of Eqs. (A15) and (A17)–(A19), we see that Hamilton's principle.

$$\delta\int_{t_1}^{t_2} L\,dt = 0, \tag{A20}$$

used with the Lagrangian specified by Eqs. (A13), (A14), and (A16), implies Hamilton's equations of motion (A11) or (II. 17) for the particles. It also implies the equation

$$\bar{K}\phi(\mathbf{r}, t) = \sum_s \int d^3\boldsymbol{\rho}\,\frac{\partial H_s}{\partial\phi}\, f_s(\mathbf{r}, \boldsymbol{\rho}, t), \tag{A21}$$

in which the arguments of H_s are $\mathbf{r}$, $\boldsymbol{\rho}$, and $\phi(\mathbf{r}, t)$. It is apparent that linearization of Eq. (A21) will lead to an equation of the form (II. 2), the basic linearized field equation with which we began, and that the quantity $J_s^1 f_s^{(1)}$ is correctly expressed by Eqs. (II. 26) and (II. 27), in which the arguments of H_s are $\mathbf{r}$, $\boldsymbol{\rho}$, and $\phi^{(0)}(\mathbf{r})$.

APPENDIX B: THREE EXAMPLES OF THE FORMULATION

In this appendix we formulate the basic equations for three plasma systems as examples of the fundamental equations that were presented in Sec. II. Gaussian units for electromagnetic quantities are used throughout.

1. One-dimensional electron gas

The exact, nonlinear Vlasov–Poisson equations for a a one-dimensional electron gas in a fixed ion background of number density $n_0(x)$ are

$$\frac{\partial f}{\partial t} + v\frac{\partial f}{\partial x} + \frac{e}{m}\frac{\partial\phi}{\partial x}\frac{\partial f}{\partial v} = 0, \tag{B1}$$

$$\frac{\partial^2\phi}{\partial x^2} = 4\pi e[\int f(x, v, t)\,dv - n_0(x)], \tag{B2}$$

where f is the electron distribution function in the phase space of position and velocity, ϕ is the scalar potential, m is the electron mass, e is the ion charge, and $-e$ is the electron charge. The linearized equations for small perturbations $f^{(1)}$ and $\phi^{(1)}$ about an equilibrium distribution $f^{(0)}$ and potential $\phi^{(0)}$ are

$$\frac{\partial f^{(1)}}{\partial t} + v\frac{\partial f^{(1)}}{\partial x} + \frac{e}{m}\frac{d\phi^{(0)}}{dx}\frac{\partial f^{(1)}}{\partial v} = -\frac{e}{m}\frac{\partial f^{(0)}}{\partial v}\frac{\partial\phi^{(1)}}{\partial x}, \tag{B3}$$

$$-\frac{1}{4\pi}\frac{\partial^2\phi^{(1)}}{\partial x^2}=-e\int f^{(1)}(x,v,t)\,dv. \tag{B4}$$

These equations correspond to Eqs. (II.1) and (II.2). Because the plasma system is completely Hamiltonian, the right-hand side of Eq. (B4) corresponds to Eqs. (II.26) and (II.27).

As an illustration of the introduction of an auxiliary function g to replace the perturbed distribution function, we choose the auxiliary function in the particular way discussed in Sec. II in connection with Eq. (II.33). Letting the equilibrium distribution function be a function of the total energy, $f^{(0)}=f^{(0)}[\epsilon(x,v)]$, where

$$\epsilon=\tfrac{1}{2}mv^2-e\phi^{(0)}(x), \tag{B5}$$

we define g by

$$f^{(1)}(x,v,t)=g(x,v,t)-e\phi^{(1)}(x,t)f^{(0)\prime}(\epsilon), \tag{B6}$$

where $f^{(0)\prime}(\epsilon)=\partial f^{(0)}/\partial\epsilon$. The equations for g and $\phi^{(1)}$ are then

$$\left(\frac{\partial}{\partial t}+\mathcal{L}\right)g=ef^{(0)\prime}(\epsilon)\frac{\partial\phi^{(1)}}{\partial t}, \tag{B7}$$

$$\Lambda\phi^{(1)}=-e\int g(x,v,t)\,dv, \tag{B8}$$

where

$$\mathcal{L}(x,v)=v\frac{\partial}{\partial x}+\frac{e}{m}\frac{d\phi^{(0)}}{dx}\frac{\partial}{\partial v}, \tag{B9}$$

$$4\pi\Lambda=-\frac{\partial^2}{\partial x^2}+k_D^2(x), \tag{B10}$$

and

$$k_D^2(x)=-4\pi e^2\int f^{(0)\prime}(\epsilon)\,dv. \tag{B11}$$

Notice that $k_D^2(x)=1/\lambda_D^2(x)$, where $\lambda_D(x)$ is the local one-dimensional Debye length for the equilibrium.

2. Three-dimensional multispecies plasma

In this example we use the position vector $\mathbf{r}$ and the canonical momentum vector $\boldsymbol{\rho}$ as variables, and write the equations in terms of the canonical phase-space distribution functions $f_s(\mathbf{r},\boldsymbol{\rho},t)$. In addition to the electromagnetic scalar potential, $\phi(\mathbf{r},t)$, we now include the electromagnetic vector potential, $\mathbf{A}(\mathbf{r},t)$. The canonical momentum vector for a particle of species s is defined in terms of the velocity $\mathbf{v}$ according to Eq. (II.12) by

$$\boldsymbol{\rho}=M_s\mathbf{v}+\frac{Q_s}{c}\mathbf{A}(\mathbf{r},t). \tag{B12}$$

The exact equations for the system in the Coulomb gauge, $\nabla\cdot\mathbf{A}=0$, are

$$\frac{\partial f_s}{\partial t}+\frac{1}{M_s}\left(\boldsymbol{\rho}-\frac{Q_s}{c}\mathbf{A}\right)\cdot\nabla_{\mathbf{r}}f_s-Q_s(\nabla\phi)\cdot\nabla_{\boldsymbol{\rho}}f_s+\frac{Q_s}{M_sc}(\nabla_{\boldsymbol{\rho}}f_s)\cdot(\nabla\mathbf{A})\cdot\left(\boldsymbol{\rho}-\frac{Q_s}{c}\mathbf{A}\right)=0, \tag{B13}$$

$$-\nabla^2\phi=4\pi\textstyle\sum_s Q_s\int d^3\rho f_s(\mathbf{r},\boldsymbol{\rho},t), \tag{B14}$$

$$\nabla^2\mathbf{A}-\frac{1}{c^2}\frac{\partial^2\mathbf{A}}{\partial t^2}-\frac{1}{c}\nabla\frac{\partial\phi}{\partial t}=-4\pi\textstyle\sum_s\frac{Q_s}{M_sc}\int d^3\rho\left(\boldsymbol{\rho}-\frac{Q_s}{c}\mathbf{A}\right)f_s(\mathbf{r},\boldsymbol{\rho},t). \tag{B15}$$

The particle Hamiltonian for species s is

$$H_s=\frac{1}{2M_s}\left(\boldsymbol{\rho}-\frac{Q_s}{c}\mathbf{A}\right)^2+Q_s\phi. \tag{B16}$$

We choose the array of potential functions to be

$$\phi\to\begin{pmatrix}\phi\\ \mathbf{A}\end{pmatrix},$$

so that

$$\frac{\partial H_s}{\partial\phi}\to\begin{pmatrix}Q_s\\ -\frac{Q_s}{M_sc}\left(\boldsymbol{\rho}-\frac{Q_s}{c}\right)\mathbf{A}\end{pmatrix},\qquad \frac{\partial^2H_s}{\partial\phi^2}\to\begin{pmatrix}0 & 0\\ 0 & \frac{Q_s^2}{M_sc^2}\,1\end{pmatrix},$$

where $\mathbf{1}$ is the unit dyad. The linearized equations for this completely Hamiltonian system, corresponding to Eqs. (II.1), (II.2), (II.25), and (II.27), are

$$\left(\frac{\partial}{\partial t}+\mathcal{L}_s\right)f_s^{(1)}=U_s\begin{pmatrix}\phi^{(1)}\\ \mathbf{A}^{(1)}\end{pmatrix}, \tag{B17}$$

$$K\begin{pmatrix}\phi^{(1)}\\ \mathbf{A}^{(1)}\end{pmatrix}=\textstyle\sum_s\int d^3\rho\begin{bmatrix}Q_s\\ -\frac{Q_s}{M_sc}\left(\boldsymbol{\rho}-\frac{Q_s}{c}\mathbf{A}^{(0)}(\mathbf{r})\right)\end{bmatrix}f_s^{(1)}(\mathbf{r},\boldsymbol{\rho},t)=\textstyle\sum_s J_s^{\dagger}f_s^{(1)}, \tag{B18}$$

where

$$\mathcal{L}_sf_s^{(1)}=\frac{1}{M_s}\left(\boldsymbol{\rho}-\frac{Q_s}{c}\mathbf{A}^{(0)}\right)\cdot\nabla_{\mathbf{r}}f_s^{(1)}-Q_s(\nabla\phi^{(0)})\cdot\nabla_{\boldsymbol{\rho}}f_s^{(1)}+\frac{Q_s}{M_sc}(\nabla_{\boldsymbol{\rho}}f_s^{(1)})\cdot(\nabla\mathbf{A}^{(0)})\cdot\left(\boldsymbol{\rho}-\frac{Q_s}{c}\mathbf{A}^{(0)}\right), \tag{B19}$$

$$U_s\begin{pmatrix}\phi^{(1)}\\ \mathbf{A}^{(1)}\end{pmatrix}=\frac{Q_s}{M_sc}\left[(\nabla_{\mathbf{r}}f_s^{(0)})+\frac{Q_s}{c}(\nabla_{\boldsymbol{\rho}}f_s^{(0)})\cdot(\nabla\mathbf{A}^{(0)})\right]\cdot\mathbf{A}^{(1)}+Q_s(\nabla_{\boldsymbol{\rho}}f_s^{(0)})\cdot\nabla\phi^{(1)}-\frac{Q_s}{M_sc}(\nabla_{\boldsymbol{\rho}}f_s^{(0)})\cdot(\nabla\mathbf{A}^{(1)})\cdot\left(\boldsymbol{\rho}-\frac{Q_s}{c}\mathbf{A}^{(0)}\right), \tag{B20}$$

$$K=\frac{1}{4\pi}\begin{bmatrix}-\nabla^2 & 0\\ -\frac{1}{c}\nabla\frac{\partial}{\partial t} & \nabla^2-\frac{1}{c^2}\frac{\partial^2}{\partial t^2}-\frac{1}{c^2}\sum_s\omega_{ps}^2\end{bmatrix}, \tag{B21}$$

and

$$\omega_{ps}^2(\mathbf{r})=\frac{4\pi Q_s^2}{M_s}\int d^3\rho f_s^{(0)}(\mathbf{r},\boldsymbol{\rho}).$$

Note that the operator K given by Eq. (B19) involves t because of the operators $\partial/\partial t$ and $\partial^2/\partial t^2$. We now replace K by a completely *time-independent* operator Λ by introducing a particular set of auxiliary functions g_s as in Eq. (II.7),

$$g_s(\mathbf{r},\boldsymbol{\rho},t)=f_s^{(1)}(\mathbf{r},\boldsymbol{\rho},t)-P_s\begin{pmatrix}\phi^{(1)}\\ \mathbf{A}^{(1)}\end{pmatrix}, \tag{B22}$$

where

$$P_s\begin{pmatrix}\phi^{(1)}\\ \mathbf{A}^{(1)}\end{pmatrix}=F_s(v^2)\mathbf{v}\cdot\frac{\partial}{\partial t}\left(\nabla\phi^{(1)}+\frac{1}{c}\frac{\partial\mathbf{A}^{(1)}}{\partial t}\right)$$

$$= -F_s(v^2)\mathbf{v}\cdot\frac{\partial \mathbf{E}^{(1)}}{\partial t}, \tag{B23}$$

$$\mathbf{v} = \frac{1}{M_s}\left(\boldsymbol{\rho} - \frac{Q_s}{c}\mathbf{A}^{(0)}(\mathbf{r})\right),$$

and F_s is an arbitrary function normalized such that

$$\mathbf{1} = 4\pi N M_s^3 Q_s \int d^3\mathbf{v}\, F_s(v^2)\mathbf{v}\mathbf{v}, \tag{B24}$$

where N is the total number of particle species. Note that the auxiliary functions differ from the one defined by Eq. (B6). The resulting equations for g_s and the potentials are

$$\left(\frac{\partial}{\partial t} + \mathcal{L}_s\right) g_s = U_s\begin{pmatrix}\phi^{(1)}\\ \mathbf{A}^{(1)}\end{pmatrix} - \left(\frac{\partial}{\partial t} + \mathcal{L}_s\right) P_s\begin{pmatrix}\phi^{(1)}\\ \mathbf{A}^{(1)}\end{pmatrix}, \tag{B25}$$

$$\Lambda\begin{pmatrix}\phi^{(1)}\\ \mathbf{A}^{(1)}\end{pmatrix} = \sum_s \int d^3\boldsymbol{\rho} \begin{bmatrix} Q_s \\ -\frac{Q_s}{M_s c}\left(\boldsymbol{\rho} - \frac{Q_s}{c}\mathbf{A}^{(0)}(\mathbf{r})\right)\end{bmatrix} g_s(\mathbf{r}, \boldsymbol{\rho}, t), \tag{B26}$$

where

$$\Lambda = \frac{1}{4\pi}\begin{bmatrix} -\nabla^2 & 0 \\ 0 & \nabla^2 - \frac{1}{c^2}\sum_s \omega_{ps}^2(\mathbf{r})\end{bmatrix}. \tag{B27}$$

In Sec. III we asserted that operators P_s could be found whose effect on Λ would be to subtract an arbitrary constant δ. For this example, subtracting a constant δ requires

$$\sum_s J_s^\dagger P_s = \delta\begin{pmatrix}1 & 0\\ 0 & \mathbf{1}\end{pmatrix},$$

where $J_s^\dagger$ is defined by Eq. (B18). A set of operators P_s that satisfy this equation are given by

$$P_s\begin{pmatrix}\phi^{(1)}\\ \mathbf{A}^{(1)}\end{pmatrix} = 4\pi\delta M_s^3 c^2 F_s(v^2)\left[\frac{1}{3}\left(\frac{v}{c}\right)^2\phi^{(1)} - \left(\frac{\mathbf{v}}{c}\right)\cdot\mathbf{A}^{(1)}\right], \tag{B28}$$

where F_s is an arbitrary function normalized according to Eq. (B24).

3. Vlasov-fluid model

The Vlasov-fluid model is a low-frequency model for an ion–electron plasma in which the ions are treated as collisionless and the electrons are treated as a massless, pressureless fluid.[3] The equations for the model are

$$\mathbf{E} + \frac{1}{c}\mathbf{u}_e\times\mathbf{B} = 0, \tag{B29}$$

$$\nabla\times\mathbf{E} = -\frac{1}{c}\frac{\partial \mathbf{B}}{\partial t}, \tag{B30}$$

$$\frac{\partial f}{\partial t} + \mathbf{v}\cdot\nabla_{\mathbf{r}} f + \frac{Q}{M}\left(\mathbf{E} + \frac{1}{c}\mathbf{v}\times\mathbf{B}\right)\cdot\nabla_{\mathbf{v}} f = 0, \tag{B31}$$

$$(\nabla\times\mathbf{B})\times\mathbf{B} = 4\pi Q\int d^3\mathbf{v}\left(\mathbf{E} + \frac{1}{c}\mathbf{v}\times\mathbf{B}\right) f, \tag{B32}$$

where $\mathbf{u}_e$ is the electron fluid velocity, $f(\mathbf{r}, \mathbf{v}, t)$ is the ion distribution function in position-velocity space, Q and M are the ion charge and mass, $\mathbf{E} = -\nabla\phi - (1/c)\,\partial\mathbf{A}/\partial t$, and $\mathbf{B} = \nabla\times\mathbf{A}$. We assume an equilibrium in which the ion distribution function is a function only of the particle energy ϵ,

$$f(\mathbf{r}, \mathbf{v}) = f^{(0)}(\epsilon),$$
$$\epsilon = \tfrac{1}{2}Mv^2 + Q\phi^{(0)}(\mathbf{r}). \tag{B33}$$

Then the linearized equations for the Vlasov-fluid model can be reduced to

$$\left(\frac{\partial}{\partial t} + \mathcal{L}\right) f^{(1)} = -Qf^{(0)\prime}(\epsilon)\,\mathbf{v}\cdot\mathbf{E}^{(1)} \tag{B34}$$

and

$$\frac{1}{4\pi}\left[(\nabla\times\mathbf{B}^{(0)})\times\mathbf{B}^{(1)} + (\nabla\times\mathbf{B}^{(1)})\times\mathbf{B}^{(0)}\right] - Qn^{(0)}\mathbf{E}^{(1)}$$
$$= Q\int d^3\mathbf{v}\left(\mathbf{E}^{(0)} + \frac{1}{c}\mathbf{v}\times\mathbf{B}^{(0)}\right) f^{(1)}$$
$$= J^\dagger f^{(1)}, \tag{B35}$$

where

$$\mathcal{L}(\mathbf{r}, \mathbf{v}) = \mathbf{v}\cdot\nabla_{\mathbf{r}} + \frac{Q}{M}\left(\mathbf{E}^{(0)} + \frac{1}{c}\mathbf{v}\times\mathbf{B}^{(0)}\right)\cdot\nabla_{\mathbf{v}}, \tag{B36}$$
$$n^{(0)}(\mathbf{r}) = \int d^3\mathbf{v}\, f^{(0)}(\epsilon),$$

and

$$f^{(0)\prime}(\epsilon) = \frac{df^{(0)}}{d\epsilon}.$$

We now choose a gauge defined by

$$\mathbf{B}^{(0)}\cdot\mathbf{A}^{(1)} = 0. \tag{B37}$$

(This gauge restricts the type of perturbation that is allowed, but the restriction is not serious.[3]) Equations (B29) and (B37) imply that the perturbation potentials $\phi^{(1)}$ and $\mathbf{A}^{(1)}$, from which $\mathbf{E}^{(1)}$ and $\mathbf{B}^{(1)}$ are derived, can be expressed in terms of a displacement vector $\boldsymbol{\xi}$ by the equations

$$\mathbf{A}^{(1)} = \boldsymbol{\xi}\times\mathbf{B}^{(0)} \text{ and } \phi^{(1)} = \mathbf{E}^{(0)}\cdot\boldsymbol{\xi},$$

where

$$\boldsymbol{\xi}\cdot\mathbf{B}^{(0)} = 0.$$

We take the array ϕ of potential functions to be the vector $\boldsymbol{\xi}$, and consider Eqs. (B34) and (B35) to be simultaneous equations for $f^{(1)}$ and $\boldsymbol{\xi}$ corresponding to Eqs. (II. 1) and (II. 2). The left-hand side of Eq. (B35) is a linear operator acting on $\boldsymbol{\xi}$ which involves $\partial/\partial t$.

In order to transform the linearized equations to the form that has been used[4] for studying certain screw-pinch equilibria, we introduce an auxiliary function g by

$$f^{(1)} = g - \left(M\mathbf{v}\cdot\frac{\partial\boldsymbol{\xi}}{\partial t} - Q\phi^{(1)}\right) f^{(0)\prime}(\epsilon). \tag{B38}$$

The equations for g and $\boldsymbol{\xi}$ are

$$\left(\frac{\partial}{\partial t} + \mathcal{L}\right) g = Mf^{(0)\prime}(\epsilon)\,\mathbf{v}\cdot\left(\frac{\partial}{\partial t} + \mathbf{v}\cdot\nabla\right)\frac{\partial\boldsymbol{\xi}}{\partial t} \tag{B39}$$

and

$$\mathbf{F}(\boldsymbol{\xi}) = Q\int d^3\mathbf{v}\left(\mathbf{E}^{(0)} + \frac{1}{c}\mathbf{v}\times\mathbf{B}^{(0)}\right) g, \tag{B40}$$

where

$$\mathbf{F}(\xi) \equiv (1/4\pi)[(\nabla\times\mathbf{B}^{(0)})\times\mathbf{B}^{(1)} + (\nabla\times\mathbf{B}^{(1)})\times\mathbf{B}^{(0)}] + \nabla[\xi\cdot\nabla p^{(0)}] \tag{B41}$$

and

$$p^{(0)}(\mathbf{r}) = \tfrac{1}{3}\int d^3\mathbf{v}\, Mv^2 f^{(0)}(\epsilon).$$

The relation

$$Qn^{(0)}\mathbf{E}^{(0)} = \nabla p^{(0)}$$

and the identity

$$M\mathbf{v}\mathbf{v}f^{(0)\prime}(\epsilon) = (\nabla_{\mathbf{v}}f^{(0)})\mathbf{v} = \nabla_{\mathbf{v}}(f^{(0)}\mathbf{v}) - f^{(0)}\mathbf{1}$$

are useful in deriving Eq. (B40). Notice that the Vlasov-fluid model is not completely Hamiltonian; the right-hand side of Eq. (B35) cannot be expressed in the form of Eqs. (II.26) and (II.27).

As was stated in Sec. III in connection with the invertibility of the field operator Λ (which we have denoted by $\mathbf{F}$ is this example), it is possible to construct an operator P whose effect on the field operator is to subtract an arbitrary constant δ. This will be true if P satisfies

$$J^{\dagger}P = \delta\left(1 - \frac{\mathbf{B}^{(0)}(\mathbf{r})\mathbf{B}^{(0)}(\mathbf{r})}{B^{(0)2}(\mathbf{r})}\right)\cdot$$

$$= -\delta\frac{1}{B^{(0)2}(\mathbf{r})}[\mathbf{B}^{(0)}(\mathbf{r})\times 1\times\mathbf{B}^{(0)}(\mathbf{r})]\cdot.$$

The right-hand side of this equation is δ times the projection operator onto the plane perpendicular to $\mathbf{B}^{(0)}(\mathbf{r})$. It is well defined for all values of $\mathbf{r}$ because $\mathbf{B}^{(0)}(\mathbf{r})$ must not vanish anywhere in the Vlasov-fluid model. An operator P which satisfies the above equation is given by

$$P\xi(\mathbf{r},t) = -\delta G(v^2)\frac{1}{\mathbf{B}^{(0)2}(\mathbf{r})}[\mathbf{B}^{(0)}(\mathbf{r})\times\mathbf{v}]\cdot\xi(\mathbf{r},t), \tag{B42}$$

where $G(v^2)$ is an arbitrary function normalized such that

$$1 = \frac{Q}{c}\int d^3\mathbf{v}\, G(v^2)\mathbf{v}\mathbf{v}. \tag{B43}$$

APPENDIX C: PROOF OF THE ANTI-HERMITICITY OF THE LIOUVILLE OPERATOR

The Liouville operator $\mathcal{L}$ is used to calculate the time rate of change of a function of phase-space variables and time due to variation of the phase-space variables according to the particle equations of motion. If we let G be an arbitrary function of phase-space variables and time, and if the phase-space variables change with time according to the equations of motion, then the total time derivative of G is given by

$$\frac{dG}{dt} = \dot{G} + \mathcal{L}G = \left(\frac{\partial}{\partial t} + \mathcal{L}\right)G, \tag{C1}$$

where $\dot{G} \equiv \partial G/\partial t$.

We consider a region Ω in $(\mathbf{r},\boldsymbol{\rho})$ space which is transformed into itself under the equations of motion. The region Ω could be the entire phase space, or any invariant region thereof. An important case occurs when the Hamiltonian is periodic in the phase space. We may then restrict our attention to functions $G(\mathbf{r},\boldsymbol{\rho})$ which have the same periodicity. We choose one fundamental period Ω in the phase space, and we reduce the phase space to a "cylinder" by identifying every other period point by point with the fundamental period. The region Ω thus defined satisfies the above conditions. We now prove the following theorem:

Theorem. The Liouville operator $\mathcal{L}$ is anti-Hermitian in Ω with respect to any weight function that can be expressed as a function of constants of the particle motion that do not depend explicitly on time.

This theorem is purely a result of the fact that Liouville's theorem applies to the particle motion, as is guaranteed because the motion is Hamiltonian. In order to prove the anti-Hermiticity in Ω, we begin by defining a transformation from variables $(\mathbf{r}',\boldsymbol{\rho}')$ to variables $(\mathbf{r},\boldsymbol{\rho})$ by

$$\mathbf{r} = \mathbf{R}(\mathbf{r}',\boldsymbol{\rho}',t), \quad \boldsymbol{\rho} = \mathbf{P}(\mathbf{r}',\boldsymbol{\rho}',t), \tag{C2}$$

where $\mathbf{R}(\mathbf{r}',\boldsymbol{\rho}',t)$ is the position vector and $\mathbf{P}(\mathbf{r}',\boldsymbol{\rho}',t)$ is the canonical momentum vector of a particle at time t whose initial position and canonical momentum vectors are $\mathbf{r}'$ and $\boldsymbol{\rho}'$, respectively. This transformation exists and possesses an inverse because $\mathbf{R}$ and $\mathbf{P}$ are the solution of a Hamiltonian system of equations. Another result of the Hamiltonian character, a crucial result, is that the Jacobian of the transformation is unity. Now consider a function $I(t)$ which is the integral over Ω of a function $G(\mathbf{r},\boldsymbol{\rho},t)$,

$$I(t) = \int_{\Omega} d^3\mathbf{r}\, d^3\boldsymbol{\rho}\, G(\mathbf{r},\boldsymbol{\rho},t). \tag{C3}$$

The only further restrictions on G are that this integral exist, and that the gradients of G with respect to $\mathbf{r}$ and $\boldsymbol{\rho}$ exist and be integrable. The time derivative of $I(t)$ is

$$\frac{dI}{dt} = \int_{\Omega} d^3\mathbf{r}\, d^3\boldsymbol{\rho}\, \dot{G}(\mathbf{r},\boldsymbol{\rho},t). \tag{C4}$$

However, dI/dt can be expressed differently by changing variables of integration in the definition of $I(t)$ from $(\mathbf{r},\boldsymbol{\rho})$ to $(\mathbf{r}',\boldsymbol{\rho}')$ according to Eq. (C2),

$$I(t) = \int_{\Omega} d^3\mathbf{r}'\, d^3\boldsymbol{\rho}'\, G[\mathbf{R}(\mathbf{r}',\boldsymbol{\rho}',t), \mathbf{P}(\mathbf{r}',\boldsymbol{\rho}',t), t]. \tag{C5}$$

Thus, we can use Eq. (C1) to write dI/dt as

$$\frac{dI}{dt} = \int_{\Omega} d^3\mathbf{r}'\, d^3\boldsymbol{\rho}'\, \dot{G}[\mathbf{R},\mathbf{P},t] + \int_{\Omega} d^3\mathbf{r}'\, d^3\boldsymbol{\rho}'\, \mathcal{L}[\mathbf{R},\mathbf{P}]G[\mathbf{R},\mathbf{P},t].$$

Transforming back to variables $(\mathbf{r},\boldsymbol{\rho})$, we then have

$$\frac{dI}{dt} = \int_{\Omega} d^3\mathbf{r}\, d^3\boldsymbol{\rho}\, \dot{G}(\mathbf{r},\boldsymbol{\rho},t) + \int_{\Omega} d^3\mathbf{r}\, d^3\boldsymbol{\rho}\, \mathcal{L}(\mathbf{r},\boldsymbol{\rho})G(\mathbf{r},\boldsymbol{\rho},t). \tag{C6}$$

Comparing Eqs. (C4) and (C6) we find

$$\int_{\Omega} d^3\mathbf{r}\, d^3\boldsymbol{\rho}\, \mathcal{L}(\mathbf{r},\boldsymbol{\rho})G(\mathbf{r},\boldsymbol{\rho},t) = 0. \tag{C7}$$

We now let the arbitrary function $G(\mathbf{r},\boldsymbol{\rho},t)$ be given as

$$G(\mathbf{r},\boldsymbol{\rho},t) = w(\mathbf{r},\boldsymbol{\rho})f^*(\mathbf{r},\boldsymbol{\rho},t)g(\mathbf{r},\boldsymbol{\rho},t),$$

where f and g are arbitrary functions and $w(\mathbf{r},\boldsymbol{\rho})$ is any real constant of the particle motion that does not depend explicitly on t, so that $\mathcal{L}(\mathbf{r},\boldsymbol{\rho})w(\mathbf{r},\boldsymbol{\rho}) = 0$. Because $\mathcal{L}(\mathbf{r},\boldsymbol{\rho})$ is a first-order differential operator we have

$$\mathcal{L}(\mathbf{r},\boldsymbol{\rho})G(\mathbf{r},\boldsymbol{\rho},t) = w(\mathbf{r},\boldsymbol{\rho})[f^*(\mathbf{r},\boldsymbol{\rho},t)\mathcal{L}(\mathbf{r},\boldsymbol{\rho})g(\mathbf{r},\boldsymbol{\rho},t) + g(\mathbf{r},\boldsymbol{\rho},t)\mathcal{L}(\mathbf{r},\boldsymbol{\rho})f^*(\mathbf{r},\boldsymbol{\rho},t)].$$

Substituting this into Eq. (C7), we finally obtain

$$\int_\Omega d^3\mathbf{r}\, d^3\rho\, w(\mathbf{r},\rho) f^*(\mathbf{r},\rho,t) \mathcal{L}(\mathbf{r},\rho) g(\mathbf{r},\rho,t)$$

$$= -\int_\Omega d^3\mathbf{r}\, d^3\rho\, w(\mathbf{r},\rho) g(\mathbf{r},\rho,t) \mathcal{L}(\mathbf{r},\rho) f^*(\mathbf{r},\rho,t)$$

$$= -\left[\int_\Omega d^3\mathbf{r}\, d^3\rho\, w(\mathbf{r},\rho) g^*(\mathbf{r},\rho,t) \mathcal{L}(\mathbf{r},\rho) f(\mathbf{r},\rho,t)\right]^*, \quad \text{(C8)}$$

which proves the anti-Hermiticity of $\mathcal{L}$.

The above general theorem applies in particular to the equilibrium Liouville operator $\mathcal{L}_s$ introduced in Sec. II. The operator $\mathcal{L}_s$ is the same operator whether G is considered to be a function of position, velocity and time or a function of position, canonical momentum and time. The reason is that the equilibrium transformation between position and velocity variables on the one hand, and position and canonical momentum variables on the other hand, does not depend explicitly on time, so that $\partial G/\partial t$ has the same meaning with either set of variables. Expressed in terms of the position and canonical momentum vectors $\mathbf{r}$ and ρ, the operator $\mathcal{L}_s$ is

$$\mathcal{L}_s(\mathbf{r},\rho) = [\nabla_\rho H_s]\cdot\nabla_\mathbf{r} - \left(\nabla_\mathbf{r} H_s + (\nabla\phi^{(0)})\frac{\partial H_s}{\partial\phi^{(0)}}\right)\cdot\nabla_\rho, \quad \text{(C9)}$$

where H_s is the equilibrium particle Hamiltonian for species s and is considered to be a function of formally independent arguments $\mathbf{r}, \rho$, and $\phi^{(0)}(\mathbf{r})$. The notation is explained in the latter part of Sec. II and in Appendix A. The expression for $\mathcal{L}_s$ in terms of the position and velocity vectors is given by Eq. (II. 5).

$\mathcal{L}_s$ is usually anti-Hermitian in $(\mathbf{r},\mathbf{v})$ space as well — but not necessarily. In analogy with Eqs. (C2), we define a transformation from variables $(\mathbf{r}',\mathbf{v}')$ to variables $(\mathbf{r},\mathbf{v})$ by

$$\mathbf{r} = \mathbf{R}_s^{(0)}(\mathbf{r}',\mathbf{v}',t), \quad \mathbf{v} = \dot{\mathbf{R}}_s^{(0)}(\mathbf{r}',\mathbf{v}',t), \quad \text{(C10)}$$

where $\mathbf{R}_s^{(0)}$ is now considered to be a function of the initial position and velocity vectors and time. Although this transformation exists and possesses an inverse, the Jacobian of the transformation may not be independent of t. In order to examine the Jacobian more closely, we represent the transformation as the result of three successive transformations: first a transformation from $(\mathbf{r}',\mathbf{v}')$ to $(\mathbf{r}',\rho')$, followed by a transformation from $(\mathbf{r}',\rho')$ to $(\mathbf{r},\rho)$, followed by a transformation from $(\mathbf{r},\rho)$ to $(\mathbf{r},\mathbf{v})$. The Jacobian of the total transformation is the product of the Jacobians of the three successive transformations. The transformation from $(\mathbf{r}',\rho')$ to $(\mathbf{r},\rho)$ is canonical and its Jacobian is unity. Let $T_s(\mathbf{r},\mathbf{v},t)$ be the Jacobian of the transformation from $(\mathbf{r},\mathbf{v})$ to $(\mathbf{r},\rho)$ at time t. Then the Jacobian of the total transformation is

$$\begin{aligned}\frac{\partial(\mathbf{r},\mathbf{v})}{\partial(\mathbf{r}',\mathbf{v}')} &= \frac{\partial(\mathbf{r},\mathbf{v})}{\partial(\mathbf{r},\rho)}\frac{\partial(\mathbf{r},\rho)}{\partial(\mathbf{r}',\rho')}\frac{\partial(\mathbf{r}',\rho')}{\partial(\mathbf{r}',\mathbf{v}')} \\ &= \frac{\partial(\mathbf{r},\mathbf{v})}{\partial(\mathbf{r},\rho)}\frac{\partial(\mathbf{r}',\rho')}{\partial(\mathbf{r}',\mathbf{v}')} \\ &= \frac{T_s(\mathbf{r}',\mathbf{v}',0)}{T_s(\mathbf{r},\mathbf{v},t)}.\end{aligned} \quad \text{(C11)}$$

The Jacobian $T_s(\mathbf{r},\mathbf{v},t)$ is the determinant of the matrix of coefficients of the dyad $\nabla_\mathbf{v}\rho = \nabla_\mathbf{v}\nabla_\mathbf{v} L_s$ with respect to the unit vectors of a Cartesian coordinate system, where L_s is the particle Lagrangian. This is easy to show. If the Lagrangian has the form

$$L_s = \mathbf{v}\cdot\mathbf{A}\cdot\mathbf{v} + \mathbf{v}\cdot\mathbf{B}(\mathbf{r},t) + C(\mathbf{r},t), \quad \text{(C12)}$$

where $\mathbf{A}$ is a constant dyad, and $\mathbf{B}(\mathbf{r},t)$ and $C(\mathbf{r},t)$ are, respectively, arbitrary vector and scalar functions of $\mathbf{r}$ and t, then $T_s(\mathbf{r},\mathbf{v},t)$ will be independent of $\mathbf{r}$, $\mathbf{v}$ and t, and $\partial(\mathbf{r},\mathbf{v})/\partial(\mathbf{r}',\mathbf{v}')$ will be unity. The Langrangian is of this form if the forces are all electromagnetic. However, with more general velocity-dependent potentials, the Lagrangian may not be of this form, and $\partial(\mathbf{r},\mathbf{v})/\partial(\mathbf{r}',\mathbf{v}')$ could depend on time.

If the Jacobian $\partial(\mathbf{r},\mathbf{v})/\partial(\mathbf{r}',\mathbf{v}')$ is time-independent, then the proof of anti-Hermiticity of $\mathcal{L}_s$ in $(\mathbf{r},\rho)$ space that led to Eq. (C8) can be carried through in a quite analogous way for $(\mathbf{r},\mathbf{v})$ space. The result is

$$\int_\Omega d^3\mathbf{r}\, d^3\mathbf{v}\, w(\mathbf{r},\mathbf{v}) f^*(\mathbf{r},\mathbf{v},t)\mathcal{L}_s(\mathbf{r},\mathbf{v}) g(\mathbf{r},\mathbf{v},t)$$

$$= -\left[\int_\Omega d^3\mathbf{r}\, d^3\mathbf{v}\, w(\mathbf{r},\mathbf{v}) g^*(\mathbf{r},\mathbf{v},t)\mathcal{L}_s(\mathbf{r},\mathbf{v}) f(\mathbf{r},\mathbf{v},t)\right]^*, \quad \text{(C13)}$$

where $\mathcal{L}_s(\mathbf{r},\mathbf{v})w(\mathbf{r},\mathbf{v}) = 0$, and Ω is the domain of integration in $(\mathbf{r},\mathbf{v})$ space.

APPENDIX D: THE EIGENFUNCTIONS OF $\mathcal{L}_s$

As indicated at the end of Sec. III, it is useful for numerical and analytical purposes to expand functions of the phase-space variables for species s in terms of eigenfunctions of the equilibrium Liouville operator $\mathcal{L}_s$. In Secs. IV and V, the structure and properties of the dispersion matrix for the case in which there is at most one nonignorable coordinate in the equilibrium are discussed by examining expressions for the matrix elements that are obtained from exact expressions for the eigenfunctions and eigenvalues of $\mathcal{L}_s$. In this appendix we first demonstrate the existence and completeness of the eigenfunctions of $\mathcal{L}_s$, and we indicate a possible scheme for approximating them in the general case. Following that we derive expressions for the eigenfunctions and eigenvalues of $\mathcal{L}_s$ when there is at most one nonignorable coordinate in the equilibrium.

In accordance with Eqs. (IV. 5) and (IV. 6), we write the eigenvalue equation as

$$\mathcal{L}_s w_{sr} = i\mu_{sr} w_{sr}, \quad \text{(D1)}$$

where μ_{sr} is real, and we impose the orthonormality condition

$$(w_{sr}, w_{sr'}) = \delta_{rr'}. \quad \text{(D2)}$$

The label r on w_{sr} and μ_{sr} stands for the set of whatever discrete and continuous labels are required for specifying the eigenfunction and eigenvalue; the symbol $\delta_{rr'}$ stands for a product of Kronecker deltas and Dirac delta functions— one Kronecker delta for each pair of discrete labels, and one Dirac delta function for each pair of continuous labels.

Elsewhere in this paper, particularly in Sec. II and Appendix A, we have considered the equilibrium particle Hamiltonians to be explicit functions of the equilibrium potentials, and we also have allowed the possibility of additional explicit dependence on the coordinates. This was done to facilitate the discussion of completely Hamiltonian systems. However, in this appendix, the concept of a completely Hamiltonian system is unimportant, and it is more convenient simply to regard the

equilibrium particle Hamiltonian for species s as an explicit function of the particle coordinates q_j and their canonical momenta p_j. In order to avoid confusion, we use the notation $\mathcal{H}_s(q,p)$ for the Hamiltonian expressed explicitly in terms of the variables q_j and p_j,

$$\mathcal{H}_s(q,p) \equiv H_s[\mathbf{r}, \rho, \phi^{(0)}(\mathbf{r})], \tag{D3}$$

where H_s is defined by Eq. (II. 16) or (A6). The equilibrium equations of motion are the usual

$$\dot{q}_j = \frac{\partial \mathcal{H}_s}{\partial p_j}, \quad \dot{p}_j = -\frac{\partial \mathcal{H}_s}{\partial q_j} \tag{D4}$$

In order to demonstrate the existence and completeness of the eigenfunctions w_{sr}, we show how they could be determined in principle from a complete knowledge of the equilibrium particle trajectories. The Liouville operator can be written

$$\mathcal{L}_s = \sum_j \left(\frac{\partial \mathcal{H}_s}{\partial p_j} \frac{\partial}{\partial q_j} - \frac{\partial \mathcal{H}_s}{\partial q_j} \frac{\partial}{\partial p_j} \right). \tag{D5}$$

Define a set of $2n-1$ independent functions $\eta_i(q,p)$ satisfying

$$\mathcal{L}_s \eta_i(q,p) = 0, \tag{D6}$$

where n is the number of coordinates q_j. The functions η_i are constants of the particle motion which do not depend explicitly on t. An arbitrary function of the set of $2n-1$ functions is the most general solution of $\mathcal{L}_s f(q,p) = 0$. The phase-space trajectories of the equilibrium particle motion are given by the equations

$$\eta_i(q,p) = c_i, \tag{D7}$$

where the quantities c_i are constants. They can be labeled by the values of the $2n-1$ constants c_i along with the values of an integer variable a. The integer a is used to distinguish disjoint curves that correspond to identical values of the constants c_i. For example, with one-dimensional motion, there may be more than one trajectory associated with the same value of the total energy. The location of a point on a trajectory can be specified conveniently by the time, τ, that it would take for an equilibrium particle to move to that point from a reference point on the trajectory. The sign of τ determines whether the particle would move in a positive or negative sense along the trajectory. The time τ is related to the arc length, σ, between a point and the reference point by

$$\left(\frac{d\sigma}{d\tau}\right)^2 = \sum_j (\hat{q}_j^2 + \hat{p}_j^2) = \sum_j \left[\left(\frac{\partial \mathcal{H}_s}{\partial p_j}\right)^2 + \left(\frac{\partial \mathcal{H}_s}{\partial q_j}\right)^2 \right]. \tag{D8}$$

We now change variables from (q,p) to τ, the integer variable a, and $2n-1$ variables $y_i = \eta_i(q,p)$. In terms of the new variables, $\mathcal{L}_s$ is

$$\mathcal{L}_s = \frac{\partial}{\partial \tau}, \tag{D9}$$

and its eigenfunctions, the solutions of Eq. (D1), are

$$w_{sr}(a, y_1, \ldots, y_{2n-1}, \tau) = d_{sr} \delta_{aa'} \delta(y_1 - c_1) \cdots \delta(y_{2n-1} - c_{2n-1}) \exp(i\mu_{sr}\tau), \tag{D10}$$

where the label r stands for the set of labels $(a', c_1, \ldots, c_{2n-1}, \mu')$, and d_{sr} is a constant chosen to normalize the eigenfunction according to Eq. (D2). The eigenvalue μ_{sr} is real; its value is determined by the nature of the phase-space trajectory. If the trajectory is closed, then the eigenfunction must be periodic in τ, so that only certain values of μ_{sr} are allowed. The allowed values may depend on some or all of the constants a and c_i, and on an integer index μ'. If the trajectory is open, then there is no condition on the eigenfunction, and consequently all values of μ_{sr} are allowed. The label μ' is continuous in this case, and may be taken equal to μ_{sr}.

The eigenfunctions of $\mathcal{L}_s$ defined by Eq. (D10) could be constructed in principle from a complete knowledge of the equilibrium particle trajectories, and they are evidently complete for functions defined in that part of the phase space covered by the equilibrium particle trajectories (in the sense that delta functions can form a complete set).

It should be noted that these eigenfunctions are highly pathological functions of the original variables q and p in the general case. The functions given by Eq. (D10) are acceptable as functions of the variables y_i and τ. However, the variables y_i are expressed in terms of q and p by the functions $\eta_i(q,p)$. There do not in general exist $2n-1$ proper constants of the motion $\eta_i(q,p)$ which are "isolating" in the sense of Wintner.[11] When the unperturbed orbits are ergodic in some f-dimensional region of phase space, then, throughout that region, at least f of the constants $n_i(q,p)$ will be discontinuous at every point! Nevertheless, the existence of the "eigenfunctions" (D10) is useful, both for purposes of the general formalism, and as a basis for various approximations. If we expand the perturbation distribution functions in terms of the eigenfunctions (D10), our formulation will correspond to the method of integration over unperturbed orbits. Moreover, the eigenfunctions (D10) may be highly degenerate, since the eigenvalue μ_{sr} may not depend on all of the constants a' and c_i. We may then construct new eigenfunctions by superposition. It may happen, as in the special case discussed below, that the new eigenfunctions are not so pathological. Their use in the expansion of the distribution functions does not directly correspond to the method of integration over unperturbed orbits. Such functions are often more useful. They may even form the basis for a perturbation calculation of approximate eigenfunctions in the pathological case. Another practical method for approximating a finite subset of the eigenfunctions and eigenvalues may be to convert the eigenvalue equation, Eq. (D1) to a relatively small matrix eigenvalue equation by expanding each of the eigenfunctions in terms of a relatively small set of eigenfunctions of a similar problem for which the eigenfunctions are known.

1. Case of at most one nonignorable coordinate

If all of the coordinates are ignorable, then $\mathcal{L}_s$ is simply

$$\mathcal{L}_s = \sum_k \frac{\partial \mathcal{H}_s}{\partial p_k} \frac{\partial}{\partial q_k}$$

and, because $\partial H_s/\partial p_k$ is independent of the coordinates, the eigenfunctions are proportional to

$$\prod_k \delta(p_k - p_k') \exp(i\kappa_k q_k),$$

in which κ_k is chosen to satisfy any boundary conditions associated with coordinate q_k. This simple case requires no further comment.

We now consider the case in which there is exactly one nonignorable coordinate. Denote the nonignorable coordinate by Q and its conjugate momentum by P; and denote the ignorable coordinates and their constant conjugate momenta by q_k and p_k. The set of coordinates q_k will be denoted by q and the set of momenta p_k will be denoted by p. The Hamiltonian $H_s(Q, P, p)$ is a constant of the motion, and the curves in the (Q, P) phase plane defined by

$$H_s(Q, P, p) = \epsilon, \tag{D11}$$

where ϵ is a constant, are the equilibrium particle trajectories in the phase space projected onto the (Q, P) plane. They are labeled by ϵ and p along with the value of an integer variable a. The integer a is used to distinguish disjoint curves that correspond to identical values of ϵ and p. We specify the location of a point on a trajectory by the time, τ, that it would take an equilibrium particle to move to that point from a reference point on the trajectory, the sense of the motion being determined by the sign of τ. The Hamiltonian equations for motion along a trajectory are

$$\frac{dQ}{d\tau} = \left.\frac{\partial H_s}{\partial P}\right|_{P=P(Q,a,\epsilon,p)} \tag{D12}$$

and

$$\frac{dP}{d\tau} = -\left.\frac{\partial H_s}{\partial Q}\right|_{P=P(Q,a,\epsilon,p)}, \tag{D13}$$

where $P = P(Q, a, \epsilon, p)$ is obtained by solving Eq. (D11) for P as a function of Q, ϵ, and p on the curve specified by a. Equation (D12) can be integrated to give τ as a function of Q, a, ϵ, and p,

$$\tau = \int^Q dQ' \Big/ \left.\frac{\partial H_s}{\partial P}\right|_{P=P(Q',a,\epsilon,p)}. \tag{D14}$$

We now change variables from (Q, P, q, p) to $(\tau, a, \epsilon, q, p)$. For each value of a, the Jacobian of the transformation $(Q, P) \to (\tau, \epsilon)$ is unity,

$$\frac{\partial(Q, P)}{\partial(\tau, \epsilon)} = 1, \tag{D15}$$

as is easily verified by differentiating Eq. (D11) with respect to ϵ holding τ fixed and then using the equations of motion (D12) and (D13). The Liouville operator in terms of the new variables is

$$\mathcal{L}_s = \frac{\partial}{\partial \tau} + \sum_k \frac{\partial H_s}{\partial p_k} \frac{\partial}{\partial q_k}. \tag{D16}$$

This expression for $\mathcal{L}_s$ differs from Eq. (D9) because here the derivative with respect to τ is to be performed holding (ϵ, q, p) fixed, and the coordinates q_k are not constants of the motion. In this expression $\partial H_s/\partial p_k$ is to be calculated considering H_s to be an explicit function of Q, P, and the momenta p_k. Then $\partial H_s/\partial p_k$ is to be expressed as a function of τ, a, ϵ, and the momenta p_k.

As noted in Sec. IV, when there are ignorable coordinates, we may take advantage of the fact that $\mathcal{L}_s$ commutes with $\partial/\partial q_k$, as well as with the constants p_k and H_s. We may choose eigenfunctions of $\mathcal{L}_s$ which are simultaneously eigenfunctions of the operators $\partial/\partial q_k$, p_k, and H_s. The eigenfunction is then a product of the (acceptably behaved) eigenfunctions of each of these operators times a function of the remaining variables. When there is only one nonignorable coordinate, the latter function is a function of τ alone,

$$w_{sr}(\tau, a, \epsilon, q, p) = \delta_{aa'}\, \delta(\epsilon - \epsilon') \prod_k \delta(p_k - p_k')(N_k)^{-1/2} \exp(i\kappa_k q_k)\, u_{sr}(\tau). \tag{D17}$$

The quantity κ_k is chosen to satisfy any boundary condition associated with coordinate q_k. The label r stands for the set of labels $(a', \epsilon', p', \kappa, \mu')$, where p' stands for the set of quantities p_k' and κ stands for the set of quantities κ_k. The quantity $(N_k)^{-1/2}$ is a normalizing factor for $\exp(i\kappa_k q_k)$. The nature of the label μ', discussed below, is determined by whether or not the motion in the (Q, P) plane is periodic. By substituting this form for w_{sr} into the eigenvalue equation, Eq. (D1), we obtain the following eigenvalue equation for $u_{sr}(\tau)$:

$$\frac{du_{sr}}{d\tau} + i\left(\sum_k \kappa_k \frac{\partial H_s}{\partial p_k} - \mu_{sr}\right) u_{sr}(\tau) = 0. \tag{D18}$$

The solution of the equation is

$$u_{sr}(\tau) = d_{sr} \exp\{i[\mu_{sr}\tau - G_{sr}(\tau)]\}, \tag{D19}$$

where

$$G_{sr}(\tau) = \int_{\tau_0}^{\tau} \sum_k \kappa_k \frac{\partial H_s}{\partial p_k}\, d\tau', \tag{D20}$$

τ_0 is any convenient lower limit for the integral, and d_{sr} is a constant chosen to normalize w_{sr} according to Eq. (D2).

The eigenvalues μ_{sr} are determined from the boundary conditions with respect to τ appropriate to the particular phase-space trajectory. There are two cases, according as the motion in the (Q, P) plane is periodic or not.

2. Periodic case

The motion in the (Q, P) plane is periodic if the trajectory in the (Q, P) plane is closed (trapped particles). The motion will also be periodic if the trajectory is open and if, in addition, the (Q, P) phase plane itself is periodic in Q (untrapped particles with a periodic boundary condition with respect to Q). Denote the period of the motion in the (Q, P) plane by T_{sr}. Because both Q and P are periodic functions of τ for fixed ϵ, the eigenfunction u_{sr} must also be a periodic function of τ in order that it can be expressed as a single-valued function of Q and P. Let us denote the mean value of the integrand in Eq. (D20) by

$$\beta_{sr} = \frac{1}{T_{sr}} \int_0^{T_{sr}} \sum_k \kappa_k \frac{\partial H_s}{\partial p_k}\, d\tau. \tag{D21}$$

The function $G_{sr}(\tau)$ may be separated into a periodic part and a part linear in τ,

$$G_{sr}(\tau) = \tilde{G}_{sr}(\tau) + \beta_{sr}\tau, \tag{D22}$$

where the periodic part is given by

$$\tilde{G}_{sr}(\tau) = \int_{\tau_0}^{\tau} \left(\sum_k \kappa_k \frac{\partial \mathcal{H}_s}{\partial p_k} - \beta_{sr} \right) d\tau'. \tag{D23}$$

The condition that $u_{sr}(\tau)$ be periodic then requires that μ_{sr} have one of the values

$$\mu_{sr} = \beta_{sr} + \frac{2\pi\mu'}{T_{sr}}, \quad \mu' = 0, \pm 1, \pm 2, \cdots. \tag{D24}$$

Note that, if $\beta_{sr} = 0$, μ_{sr} does not depend on the quantities κ_k.

3. Aperiodic case

If the trajectory in the (Q, P) plane is open and extends to infinite values of Q, without any periodic boundary conditions with respect to Q, then there is no additional condition which must be satisfied by u_{sr}, and hence no restriction on the values of μ_{sr}. All real values of $\mu_{sr} = \mu'$ are permitted.

APPENDIX E: PROPERTIES OF ANALYTIC MATRIX FUNCTIONS

We wish to study some of the properties of an *analytic matrix function* $A(\omega)$—that is, a matrix whose elements are analytic functions of a variable ω in some region (or regions) of the complex ω plane. We begin with some definitions.

Given an analytic function $f(\omega)$, we define the *conjugate function* $f^*(\omega)$ as follows,

$$f^*(\omega) = [f(\omega^*)]^*. \tag{E1}$$

The conjugate function $f^*(\omega)$ is an analytic function of ω; it may be obtained by analytically continuing $[f(\omega)]^*$ off the real axis if $f(\omega)$ is analytic on the real axis. Loosely speaking, $f^*(\omega)$ is obtained by conjugating $f(\omega)$, treating ω as a real variable. We define a *real function* $f(\omega)$ as a function with the property

$$f(\omega) = f^*(\omega). \tag{E2}$$

A real function takes on conjugate values at conjugate points ω and ω^*. It is the analytic continuation of a function which is real for real ω, if the function is analytic on the real axis.

Given an analytic matrix function $A(\omega)$, we define the *adjoint matrix function* as the matrix

$$A^\dagger(\omega) = [A(\omega^*)]^\dagger. \tag{E3}$$

If $A(\omega)$ is analytic on the real axis, then $A^\dagger(\omega)$ is the adjoint matrix to $A(\omega)$ on the real axis. In that case, it is obtained off the real axis by analytic continuation. Note that this is the only way to define an adjoint matrix which is an analytic function of ω; $[A(\omega)]^\dagger$ is not an analytic function of ω. The matrix elements $A^\dagger_{ij}(\omega)$ of the adjoint matrix are

$$A^\dagger_{ij}(\omega) = A^*_{ji}(\omega). \tag{E4}$$

We can now define a *Hermitian matrix function* as a matrix $D(\omega)$ with the property

$$D^\dagger(\omega) = D(\omega). \tag{E5}$$

Note that the dispersion matrix $D(\omega)$ defined by Eq. (IV. 24) is not generally analytic on the real axis, but has a cut there if the sum over r includes an integral over a continuous variable (for example, E), and if μ_{sr} is a function of that variable. The cut extends over the region of the real axis covered by values of μ_{sr}. In any event, the definitions (E3) and (E5) still apply. [If we analytically continue $D(\omega)$ across the real axis from above according to Landau's prescription, as discussed in Sec. V, then of course $D(\omega)$ is analytic on the real axis, but the definition (IV. 24) is effectively altered by deforming the path of integration, and $D(\omega)$ is no longer Hermitian.] We define a *unitary matrix function* as a matrix $U(\omega)$ with the property

$$U^{-1}(\omega) = U^\dagger(\omega), \tag{E6}$$

where $U^{-1}(\omega)$ is the inverse matrix.

The following theorems can now easily be proved:

Theorem 1: The eigenvalues of a Hermitian matrix function are real functions.

Theorem 2: A Hermitian matrix function can be diagonalized by a unitary matrix function, except possibly at certain values of the variable ω where two or more eigenvalues become degenerate.

Theorem 3: If a Hermitian matrix function $D(\omega)$ is analytic on the real axis, then on the real axis it is a Hermitian matrix in the usual sense; its eigenvectors form a complete orthonormal set. If $D(\omega)$ is any Hermitian matrix function, then, for complex ω, the eigenvectors of $D(\omega^*)$ form a dual set to those of $D(\omega)$.

The proofs are straightforward. For example, to prove Theorem 2, we note that the matrix $D(\omega)$ can certainly be diagonalized by a similarity transformation, except possibly at certain values of ω where two or more eigenvalues become degenerate. [When ω is not real, $D(\omega)$ is not a Hermitian *matrix*, and therefore at a point ω where $D(\omega)$ has a degenerate eigenvalue, $D(\omega)$ may not be diagonalizable, though it will have a Jordan normal form.] We may therefore write, for almost all values of ω,

$$U^{-1}(\omega)\, D(\omega)\, U(\omega) = d(\omega), \tag{E7}$$

for some matrix $U(\omega)$, where $d(\omega)$ is a diagonal matrix. If $D(\omega)$ is a Hermitian matrix function, then using the above definitions and Theorem 1, we may write the adjoint of Eq. (E7) in the form

$$U^\dagger(\omega)\, D(\omega)\, U^{-1\dagger}(\omega) = d(\omega). \tag{E8}$$

Comparing Eqs. (E7) and (E8), if none of the eigenvalues are degenerate, we see that $U(\omega)$ satisfies Eq. (E6).

Let the matrix $D(\omega)$ be of dimensions $N \times N$. Then the matrix $D(\omega)$ has N eigenvalues $d_j(\omega)$. When there is a double degeneracy, the condition

$$d_i(\omega) = d_j(\omega),$$

for two particular values i and j, represents two conditions on ω [real and imaginary parts of $d_i(\omega)$ and $d_j(\omega)$ must coincide], and there will be in general only certain discrete values of ω at which the two eigenvalues $d_i(\omega)$ and $d_j(\omega)$ coincide. Only in exceptional cases will three

or more eigenvalues coincide at a single value of ω. Hence, typically the eigenvalues $d_j(\omega)$ will form an N-valued function $d(\omega)$ on a Riemann surface whose N sheets are stitched together in pairs at discrete branch points. We shall see below that the unitary matrix function $U(\omega)$ which diagonalizes $D(\omega)$ is quadruple-valued at such a branch point; that is, ω must circle the branch point four times before each element of $U(\omega)$ returns to its original value. At a complex value of ω where two or more eigenvalues become degenerate, the corresponding eigenvectors in general also become degenerate (i.e., linearly dependent), and the matrix $U(\omega)$ becomes singular.

We now investigate the behavior of the eigenvalues and eigenvectors in the neighborhood of a double branch point ω_0 of $d(\omega)$. In a region of the ω plane around ω_0 which excludes any other branch points, we can find a unitary matrix function $U(\omega)$ which diagonalizes $D(\omega)$ except for the two rows and columns associated with the degenerate eigenvalue $d(\omega_0)$. Let $D(\omega)$ be thus partially diagonalized, and consider for the remainder of the present discussion only the 2×2 matrix

$$D(\omega)=\begin{pmatrix} 2a & g+ih \\ g-ih & 2b \end{pmatrix}, \tag{E9}$$

where a, b, g, and h are real functions of ω in the sense of definition (E2). The eigenvalues of this matrix are

$$d(\omega)=a+b\pm Q^{1/2}, \tag{E10}$$

where $Q(\omega)$ is the function

$$Q=(a-b)^2+g^2+h^2. \tag{E11}$$

Note that $Q(\omega)$ is a real function and is real and nonnegative at any point ω where a, b, g, and h are real (for example, on the real axis). Hence $d(\omega)$ as given by Eq. (E10) is a real function. At the branch point, $Q(\omega_0)=0$. Evidently ω_0 cannot be real unless $a=b$ and $g=h=0$. If ω is made to circle the branch point ω_0, the two values (E10) of $d(\omega)$ are interchanged. Suitably normalized eigenvectors are readily found. They are the columns of the matrix

$$U(\omega)=2^{-1/2}Q^{-1/4}\begin{bmatrix} \dfrac{g+ih}{[Q^{1/2}+b-a]^{1/2}} & \dfrac{g+ih}{[Q^{1/2}-b+a]^{1/2}} \\ [Q^{1/2}+b-a]^{1/2} & -[Q^{1/2}-b+a]^{1/2} \end{bmatrix}. \tag{E12}$$

The terms in Eq. (E12) are written so that all expressions are real functions except for the factors i written explicitly. (Note that $Q^{1/2}\geq|b-a|$ if a, b, g, and h are real.) The reciprocal matrix $U^{\dagger}(\omega)$ is obtained by interchanging rows and columns in Eq. (E12), and replacing the explicit factors i with $-i$. It can be readily verified that

$$U^{\dagger}DU=\begin{pmatrix} d_+ & 0 \\ 0 & d_- \end{pmatrix}. \tag{E13}$$

A careful examination of the behavior of the terms in Eq. (E12), on the four-sheeted Riemann surface associated with the functions $Q^{1/2}$, $Q^{1/4}$, and $[Q^{1/2}\pm(b-a)]^{1/2}$, in the neighborhood of the branch point ω_0 where $Q(\omega_0)=0$, will show that each circling of the branch point interchanges the two eigenvectors and reverses the sign of one of them. Consequently the unitary matrix function $U(\omega)$ is quadruple-valued around the branch point, as noted in the preceding paragraph. We note also that the eigenvectors become degenerate and the elements of $U(\omega)$ become infinite as $\omega\to\omega_0$ and $Q\to 0$, except for ω_0 on the real axis, in which case $a(\omega_0)=b(\omega_0)$.

We will not in this paper analyze the behavior of the eigenvalues and eigenvectors of $D(\omega)$ in the exceptional cases when more than two eigenvalues become degenerate, or when a degeneracy occurs at a root of the dispersion relation (IV.28).

[1] R. R. Bartsch, C. J. Buchenauer, E. L. Cantrell, J. N. Downing, B. L. Freeman, K. B. Freese, R. F. Gribble, I. Henins, A. R. Jacobson, F. C. Jahoda, T. R. Jarboe, R. Kristal, G. Miller, W. E. Quinn, F. L. Ribe, A. R. Sherwood, R. E. Siemon, and K. S. Thomas, *Proceedings of the Sixth International Conference on Plasma Physics and Controlled Nuclear Fusion Research*, Berchtesgaden, 1976 (IAEA, Vienna, 1977) (Nuclear Fusion—Supplement 1977), Vol. II, p. 193; F. Herrnegger, M. Kaufmann, D. Lortz, J. Neuhauser, J. Nührenberg, W. Schneider, and G. Schramm, *ibid.*, Vol. II, p. 183.

[2] D. A. Baker, L. C. Burkhardt, J. N. DiMarco, A. Haberstich, R. L. Hagenson, R. B. Howell, H. J. Karr, and A. E. Schofield, *Proceedings of the Sixth International Conference on Plasma Physics and Controlled Nuclear Fusion Research*, Berchtesgaden, 1976 (IAEA, Vienna, 1977) (Nuclear Fusion-Supplement 1977), Vol. I, p. 419; C. W. Gowers, D. C. R D. C. Robinson, A. Sykes, A. J. L. Verhage, J. A. Wesson, M. R. C. Watts, and H. A. B. Bodin, *ibid.*, Vol. I, p. 429; A. Buffa, S. Costa, R. Giannella, G. Malesani, G. F. Nalesso, and S. Ortolani, *ibid.*, Vol. I, p. 447.

[3] J. P. Freidberg, Phys. Fluids **15**, 1102 (1972).

[4] H. R. Lewis and L. Turner, Nucl. Fusion **16**, 993 (1976).

[5] J. L. Schwarzmeier, H. R. Lewis, B. A. Shrauner, and K. R. Symon (to be published in Phys. Fluids).

[6] H. R. Lewis, Phys. Fluids **15**, 103 (1972).

[7] O. Buneman and J. P. Holdren (private communication); J. P. Holdren, Ph. D. thesis, Stanford University (1970) (Xerox University Microfilms, Ann Arbor, Michigan 48106, Thesis number 71-02-773).

[8] Z. Sedlaček, Czech. J. Phys. B **22**, 439 (1972); **23**, 892 (1973).

[9] B. L. Buzbee, G. H. Golub, and H. R. Lewis (to be published).

[10] J. L. Schwarzmeier and K. R. (to be published in J. Math. Phys.)

[11] A. Wintner, *The Analytical Foundations of Celestial Mechanics*, (Princeton U. P., Princeton, N. J., 1941), p. 96.

[12] In that case, I_s, H_s, and their derivatives with respect to ϕ are to be evaluated at $\phi(\mathbf{r},t)=\phi^{(0)}(\mathbf{r})$. In order to make our notation less cumbersome, when derivatives of L_s and H_s with respect to ϕ are to be evaluated at $\phi^{(0)}$, we denote this by simply writing the derivatives as if they were with respect to $\phi^{(0)}$. For example, $\partial H_s/\partial\phi^{(0)}$ and $\partial^2 H_s/\partial\phi^{(0)2}$ mean, respectively, $\partial H_s/\partial\phi$ and $\partial^2 H_s/\partial\phi^2$ evaluated at $\phi(\mathbf{r},t)=\phi^{(0)}(\mathbf{r})$.

[13] The reason is that Hamilton's equations for the particle motion do not involve a time derivative of the array of potentials, whereas the Lagrange equations may, and usually do, involve such a derivative. This is related to the fact that the transformation between $(\mathbf{r},\boldsymbol{\rho})$ and $(\mathbf{r},\mathbf{v})$ depends on the array of exact, time-dependent potentials. It is important to choose either the $(\mathbf{r},\boldsymbol{\rho})$ or the $(\mathbf{r},\mathbf{v})$ variables before linearizing the equations.

Invariants in enveloping algebras under the action of Lie algebras of derivations

L. Abellanas and L. Martínez Alonso

Departamento de Métodos Matemáticos de la Física, Facultad de Ciencias Físicas, Universidad Complutense de Madrid, Madrid-3, Spain
(Received 1 March 1978)

An upper bound on the number of algebraically independent invariants in an enveloping algebra U under the action of a Lie algebra G_0 of derivations is obtained. We are able to determine the exact number of invariants for the case $[G_0, G_0] = G_0$. This generalizes previous results about Casimir invariants.

1. INTRODUCTION

In a previous paper[1] we investigated the Casimir invariants of any finite-dimensional complex Lie algebra G. This work gives an upper bound to the maximal number of algebraically independent Casimir invariants. In particular for the case $[G, G] = G$ the bound became a strict equality. Besides for their intrinsic mathematical content, results of this kind play, as well known, a crucial role in the applications of Lie group theory in physics.[2]

Sometimes, one has to deal with Lie group action over noncommutative algebras which are not covered by the above scheme. Such is the case whenever symmetry breaking appears. For instance under the presence of interaction the symmetry group $\mathcal{G}$ of a free Hamiltonian H_0 reduces to some subgroup $\mathcal{G}_0$ of $\mathcal{G}$. Hence the final Hamiltonian H must be invariant only under the action of $\mathcal{G}_0$.

Situations like the one mentioned above naturally lead to the analysis of the invariants in the enveloping algebra $\mathcal{U}$ of a Lie algebra G, with respect to a family of derivations. In this paper we generalize the results of Ref. 1. The tools involved in the present analysis are the elementary aspects of the theory of enveloping algebras.[3] Section 2 is devoted to state some preliminary facts concerning the notion of algebraic independence in enveloping algebras. In Sec. 3 we find an explicit upper bound of the maximal number of algebraically independent invariants in an enveloping algebra, under the action of a Lie algebra G_0 of derivations. In the case $[G_0, G_0] = G_0$ the bound becomes an strict equality. These results are applied to the adjoint action. As an illustrative example, we compute the admissible interaction terms for a Hamiltonian in nonrelativistic quantum mechanics. Finally we show that the general bound derived in this paper is better than another one which is implicit in the work of Joseph.[4]

2. ALGEBRAIC INDEPENDENCE IN $\mathcal{U}$

Let G denote any complex Lie algebra of dimension $|G| = n$, with basis $\{A_\alpha\}_1^n$. Let $\mathcal{U}$ be the universal enveloping algebra of G, consisting of all (noncommutative, in general) polynomials in $A_1, \ldots, A_n$. The linear subspace of $\mathcal{U}$ generated by the monomials $A_1^{\alpha_1} A_2^{\alpha_2} \ldots A_n^{\alpha_n}$ of degree $\alpha_1 + \alpha_2 + \cdots + \alpha_n \leq m$ will be denoted by $\mathcal{U}_m$. Given $u \in \mathcal{U}$, we will refer to the integer $d(u) \equiv \inf\{m : u \in \mathcal{U}_m\}$ as the degree of u.

On the other hand, we will have to consider the symmetric algebra S of G, which is isomorphic to the polynomial algebra $\mathbb{C}[a_1, \ldots, a_n]$ in n commutative variables over the complex field $\mathbb{C}$. More precisely, if we take $S^m \equiv \mathcal{U}_m / \mathcal{U}_{m-1}$ for $m \geq 1$ (by definition $S^0 \equiv \mathbb{C}$), the symmetric algebra is nothing but $S = \oplus_0^\infty S^m$. The canonical projection $j_m : u \in \mathcal{U}_m \to j_m(u) \equiv u + \mathcal{U}_{m-1} \in S^m$ allows us to define the so-called principal part of $u \in \mathcal{U}$ as $[u] \equiv j_{d(u)}(u)$. In this context, the isomorphism mentioned above between S and $\mathbb{C}[a_1, \ldots, a_n]$ can be realized by the identification $a_\alpha \equiv [A_\alpha]$.

There is an interesting mapping which links S and $\mathcal{U}$, namely the canonical isomorphism defined as the unique linear isomorphism $\phi : S \to \mathcal{U}$ verifying

$$\phi(a_{\alpha_1} \cdots a_{\alpha_r}) = \frac{1}{r!} \sum_{\sigma \in \Pi_r} A_{\alpha_{\sigma(1)}} \cdots A_{\alpha_{\sigma(r)}}, \tag{1}$$

where Π_r stands for the permutation group of r objects. Given any integer $m \geq 0$, we will denote it as $\mathcal{U}^m \equiv \phi(S^m)$.

In order to study invariants we need to define some notion of algebraic independence in S and $\mathcal{U}$. In S it is an easy task, because we have at hand the usual notion of algebraically independent sets in commutative algebras, which will be referred to here as S independence.

It seems natural to translate this notion to $\mathcal{U}$ in such a way that it turns out to be compatible with ϕ, in some sense.

Definition: $\{u_j\}_1^r \subset \mathcal{U}$ is said to be $\mathcal{U}$ independent if for all nonzero ordered polynomials $P(x_1, \ldots, x_r) = \sum_I \lambda_{\alpha, \ldots, \alpha_r} x_1^{\alpha_1} \cdots x_r^{\alpha_r}$ we have that $P(u_{\sigma(1)}, \ldots, u_{\sigma(r)}) \neq 0$ in $\mathcal{U}$, for all $\sigma \in \Pi_r$. Sometimes, we will refer to the elements $u_1, \ldots, u_r$ as $\mathcal{U}$ independent.

Lemma 1: Let $\{u_j\}_1^r \subset \mathcal{U}$. Then

$$\{[u_j]\}_1^r \ S \text{ independent} \Longrightarrow \{u_j\}_1^r \ \mathcal{U} \text{ independent}.$$

Proof: Suppose that $\{u_j\}_1^r$ is not $\mathcal{U}$ independent, so that (by previous reordering of the u_j if necessary) we can find P such that $P(u_1, \ldots, u_r) = \sum_I \lambda_{\alpha_1, \ldots, \alpha_r} u_1^{\alpha_1} \cdots u_r^{\alpha_r} = 0$ with $\lambda_{\alpha_1, \ldots, \alpha_r} \neq 0$ for all $(\alpha_1, \ldots, \alpha_r) \in I$. By applying ϕ we get

$$\phi\left\{\sum_{I_{\max}} \lambda_{\alpha_1, \ldots, \alpha_r} [u_1]^{\alpha_1} \cdots [u_r]^{\alpha_r}\right\} \in \mathcal{U}_{d(I_{\max})-1},$$

where $I_{\max} \equiv \{(\alpha_1, \ldots, \alpha_r) \in I : \sum_1^r \alpha_j d(u_j) = \text{maximum}\}$ and $d(I_{\max}) \equiv \sum_{I_{\max}} \alpha_i d(u_i)$.

However, we know $\phi\{\sum_{I_{\max}} \lambda_{\alpha_1, \ldots, \alpha_r} [u_1]^{\alpha_1} \cdots [u_r]^{\alpha_r}\} \in \mathcal{U}^{d(I_{\max})}$.

 0022-2488/79/030437-04$01.00

Since $\mathcal{U}^m \cap \mathcal{U}_{m-1} = \{0\}$, we conclude that this element is zero. But ϕ being a linear bijection, this would mean that $\{[u_j]\}_1^r$ is not S independent. (Q. E. D.)

The next corollary checks the required compatibility.

Corollary 1: Let $p_1, \ldots, p_r$ be homogeneous elements of S. Then

$$\{p_j\}_1^r \ S \text{ independent} \Longrightarrow \{\phi(p_j)\}_1^r \ \mathcal{U} \text{ independent}.$$

Proof: It follows from the fact that $[\phi(p)] = p$, for any homogeneous $p \in S$. (Q. E. D.)

At this point, we want to relate our definition to the explicit notion of dimension proposed by Gel'fand and Kirillov.[5] Given a subalgebra V of $\mathcal{U}$ and a finite set $\overline{v} = \{v_j\}_1^N \subset V$, let us denote by $d(\overline{v}, N)$ the linear dimension of the linear subspace of V generated by the monomials of degree less than or equal to N over the variables $\{v_j\}_1^N$ taken in any order. Gel'fand and Kirillov define

$$\text{Dim } V \equiv \sup_{\overline{v}} \varlimsup_{N\to\infty} \frac{\log d(\overline{v}, N)}{\log N}.$$

(In what follows the symbol dim will denote linear dimensions.) In spite of the apparent complexity of this definition it turns out that if V is commutative, Dim V coincides with the transcedence degree of V over the field $\mathbb{C}$.

In the general case of a subalgebra $V \subset \mathcal{U}$, let gr V denote the graded algebra of V. This algebra is defined to be the (commutative) polynomial subalgebra of S generated by the set $\{[u]: u \in V\}$. Also, let Dim gr V be the maximal number of S-independent elements of gr V. Joseph[4] has proved the following result.

Theorem 1: Dim V = Dim gr V.

The compatibility of our Definition 1 with the Gel'fand–Kirillov dimension derives from the following Proposition.

Proposition 1: Dim V equals the maximal number of $\mathcal{U}$-independent elements of V.

Proof: Let us take a maximal S-independent set $\{p_j\}_1^r$ of homogeneous elements in gr V. We can choose $\{u_j\}_1^r \subset V$ such that $p_j = [u_j]$. From Lemma 1 it follows that $\{u_j\}_1^r$ is $\mathcal{U}$ independent in V. Since Theorem 1 tells us that Dim $V = r$, we conclude that Dim V is less than or equal to the maximal number of $\mathcal{U}$-independent elements of V. It remains to prove that strict inequality yields a contraction.

In fact, let us suppose the existence of a $\mathcal{U}$-independent set $\overline{v} = \{v_j\}_1^R \subset V$ with $R >$ Dim V. In that case the ordered monomials $V_1^{\alpha_1} V_2^{\alpha_2} \cdots V_R^{\alpha_R}$ would be linearly independent. Therefore,

$$\varlimsup_{N\to\infty} \frac{\log d(\overline{v}, N)}{\log N} \geq \varlimsup_{N\to\infty} \frac{\log \binom{N+R}{R}}{\log N} = R.$$

This would imply Dim $V \geq R$ which is absurd. (Q. E. D.)

3. INVARIANTS UNDER THE ACTION OF DERIVATIONS

A. Derivations on enveloping algebras

Definition 2: By a derivation on G we shall mean a linear map $\partial: G \to G$ verifying the property

$$\partial([A, A']) = [\partial(A), A'] + [A, \partial(A')], \qquad A, A' \in G \tag{3}$$

We may extend the derivation ∂ to a linear map on the enveloping algebra $\mathcal{U}$ of G in such a way that it verifies

$$\partial(u_1 u_2) = \partial(u_1) u_2 + u_1 \partial(u_2), \qquad u_1, u_2 \in \mathcal{U}. \tag{4}$$

One easily finds that $\partial(\mathcal{U}^m) \subset \mathcal{U}^m$ for all integers $m \geq 0$. [By definition we take $\partial(1) = 0$]. We associate with the map $\partial: \mathcal{U} \to \mathcal{U}$ another linear map

$$\hat{\partial}: S \to S, \quad \hat{\partial}(p) \equiv \phi^{-1}[\partial(\phi(p))]. \tag{5}$$

It follows at once that $\hat{\partial}(S^m) \subset S^m$ for all integers $m \geq 0$ and we have

$$\hat{\partial}(p) = [\partial(\phi(p))], \quad \approx p \in S^m. \tag{6}$$

Moreover we obtain

$$\hat{\partial}(p_1 p_2) = \hat{\partial}(p_1) p_2 + p_1 \hat{\partial}(p_2), \qquad p_1, p_2 \in S \tag{7}$$

Definition 3: The elements of the sets

$$\mathcal{U}^{I(\partial)} \equiv \{u \in \mathcal{U}: \partial(u) = 0\} \tag{8}$$

$$S^{I(\hat{\partial})} \equiv \{p \in S: \hat{\partial}(p) = 0\} \tag{9}$$

will be called the invariants in $\mathcal{U}$ and S, respectively under the derivation ∂. From (4) and (7), we see that $\mathcal{U}^{I(\partial)}$ and $S^{I(\hat{\partial})}$ are subalgebras of $\mathcal{U}$ and S, respectively. On the other hand, (5) implies

$$\phi(S^{I(\hat{\partial})}) = \mathcal{U}^{I(\partial)}. \tag{10}$$

Lemma 2: $S^{I(\hat{\partial})} = \text{gr}\,\mathcal{U}^{I(\partial)}$.

Proof: Since $\hat{\partial}(S^m) \subset S^m$, each element $p \in S^{I(\hat{\partial})}$ may be written as a sum $p = \sum_0^N p_m$, where $p_m \in S^{I(\hat{\partial})} \cap S^m$. On the other hand one obtains from (10) that $u_m \equiv \phi(p_m) \in \mathcal{U}^{I(\partial)} \cap \mathcal{U}^m$ and therefore $p_m = [u_m]$. Then we deduce $S^{I(\hat{\partial})} \subset \text{gr}\,\mathcal{U}^{I(\partial)}$.

Since $\partial(\mathcal{U}^m) \subset \mathcal{U}^m$, each element $u \in \mathcal{U}^{I(\partial)}$ is of the form $u = \sum_0^N u_m$, where $u_m \in \mathcal{U}^{I(\partial)} \cap \mathcal{U}^m$. If we suppose $u_N \neq 0$, we have $[u] = [u_N]$, then (6) implies

$$\hat{\partial}([u]) = \hat{\partial}([u_N]) = [\partial(u_N)] = 0,$$

that is, $[u] \in S^{I(\hat{\partial})}$. Therefore, $\text{gr}\,\mathcal{U}^{I(\partial)} \subset S^{I(\hat{\partial})}$. (Q. E. D.)

Corollary 2: Dim $\mathcal{U}^{I(\partial)}$ = Dim $S^{I(\hat{\partial})}$.

Proof: It follows at once from Theorem 1 and Lemma 2.

B. Invariants under the action of Lie algebras of derivations

Let G_0 be a complex Lie algebra of dimension $|G_0| = n_0$ and let $X \to \partial_X$ be a representation of G_0 by derivations over G. Let us consider the extension of their derivations to the enveloping algebra $\mathcal{U}$ and to the symmetric algebra S of G. We define the following sets of invariants under the action of G:

$$\mathcal{U}^{I(\partial(G_0))} \equiv \{u \in \mathcal{U}: \partial_X(u) = 0, \quad X \in G_0\}, \tag{11}$$

$$S^{I(\hat{\partial}(G_0))} \equiv \{p \in S: \hat{\partial}_X(p) = 0, \quad X \in G_0\}. \tag{12}$$

From the above results it is straightforward to verify that

$$\phi(S^{I(\hat{\partial}(G_0))}) = \mathcal{U}^{I(\partial(G_0))}$$

and that

$$S^{I(\hat{\partial}(G_0))} = \text{gr}\,\mathcal{U}^{I(\partial(G_0))}.$$

Moreover Corollary 2 generalizes to

$$\mathrm{Dim}\,\mathcal{U}^{I(\hat{\mathfrak{d}}(G_0))} = \mathrm{Dim}\,S^{I(\hat{\mathfrak{d}}(G_0))}. \tag{13}$$

This number will be denoted by $\tau(G, \partial(G_0))$.

Let $\{X_i, i=1,\ldots,n_0\}$ be a basis of G_0 with commutation rules

$$[X_i, X_j] = \sum_k d_{ij}^k X_k. \tag{14}$$

Let us denote $\partial_i \equiv \partial_{X_i} (i=1,\ldots,n_0)$. They act on the basis $\{A_\alpha\}_1^n$ of G in the following form,

$$\partial_i(A_\alpha) = \sum_{\alpha,\nu} \gamma_{i\alpha}^\nu A_\nu, \tag{15}$$

where $\gamma_{i\alpha}^\nu \in \mathbb{C} (i=1,\ldots,n_0;\ \alpha,\nu=1,\ldots,n)$.

From (7) we obtain at once

$$\hat{\partial}_i(p) = \sum_{\alpha,\nu} \gamma_{i\alpha}^\nu a_\nu \frac{\partial p}{\partial a_\alpha}, \qquad p \in S. \tag{16}$$

Let $r(G, \partial(G_0))$ be the maximal rank of the $n_0 \times n$ matrix function defined by $M_{i\alpha}(a) = \sum_\nu \gamma_{i\alpha}^\nu a_\nu$. We have

Theorem 2: $\tau(G, \partial(G_0)) \leq |G| - r(G, \partial(G_0))$ (17)

Proof: From (16) we have that $S^{I(\hat{\mathfrak{d}}(G_0))}$ coincides with the set of polynomial solutions of the following system of differential equations,

$$\hat{\partial}_i(f) = \sum_{\alpha,\nu} \gamma_{i\alpha}^\nu a_\nu \frac{\partial f}{\partial a_\alpha} = 0, \quad i=1,\ldots,n_0. \tag{18}$$

Since $[\hat{\partial}_i, \hat{\partial}_j] = \sum_k d_{ij}^k \hat{\partial}_k$, these differential operators define an integrable distribution of vector fields in $\mathbb{C}^n$. From a classical result due to Frobenius the number of functionally independent solutions is $n - r(G, \partial(G_0))$.

(Q. E. D.)

Generally speaking equality will not be accessible in (17), because of the existence of nonpolynomial solutions of the systems (18). Nevertheless, there is an important exception.

Theorem 3: If $[G_0, G_0] = G_0$, then $\tau(G, \partial(G_0)) = |G| - r(G, \partial(G_0))$.

Proof: Since the property $[G_0, G_0] = G_0$ forces G_0 to be an algebraic Lie algebra,[6] a result of Dixmier[7] implies the existence of $n - r(G, \partial(G_0))$ algebraically independent rational solutions of the system (18). On the other hand, given a rational solution $h = p_1/p_2$, if we take p_1 and p_2 relatively prime polynomials we easily obtain

$$\lambda(X) \equiv \frac{\hat{\partial}_X(p_1)}{p_1} = \frac{\hat{\partial}_X(p_2)}{p_2} \in \mathbb{C}, \qquad X \in G_0.$$

Then $\lambda = \lambda(X)$ is a weight of G_0. But $[G_0, G_0] = G_0$ implies $\lambda = 0$. Therefore, $p_1, p_2 \in S^{I(\hat{\mathfrak{d}}(G_0))}$, and we deduce that $S^{I(\hat{\mathfrak{d}}(G_0))}$ has $n - r(G, \partial(G_0))$ algebraically independent elements. (Q. E. D.)

Remark: In practice one may find a maximal set of homogeneous algebraically independent elements $\{p_r : r=1,\ldots,N \leq \tau(G, \partial(G_0))\}$ in $S^{I(\hat{\mathfrak{d}}(G_0))}$ by solving system (18). An analogous set may be achieved in $\mathcal{U}^{I(\hat{\mathfrak{d}}(G_0))}$ by taking $u_r \equiv \phi(p_r)$ as Corollary 1 shows.

C. Applications

Let G_0 be a Lie subalgebra of G; under the adjoint action $(X, A) \in G_0 \times G \to [X, A] \in G$, G_0 acts over G as a Lie algebra of derivations. The set of invariants in $\mathcal{U}$ under the action of G_0 is given by

$$\mathcal{U}^{I(G_0)} \equiv \{u \in \mathcal{U} : [X, u] = 0,\ \forall\ X \in G_0\}. \tag{19}$$

If $\{A_\alpha : \alpha = 1,\ldots,n\}$ is a basis of G such that their n_0 first terms $\{A_i : i=1,\ldots,n_0\}$ form a basis of G_0, the above results imply that the dimension $\tau(G, G_0)$ of $\mathcal{U}^{I(G_0)}$ verifies

$$\tau(G, G_0) \leq |G| - r(G, G_0), \tag{20}$$

where $r(G, G_0)$ is the maximal rank of the $n_0 \times n$ matrix function $M_{i\alpha}(a) = \sum_\nu C_{i\alpha}^\nu a_\nu$ $(i=1,\ldots,n_0;\ \alpha,\nu=1,\ldots,n$. From Theorem 3, equality in (20) holds if $[G_0, G_0] = G_0$.

In the case $G_0 = G$ the set $\mathcal{U}^{I(G_0)}$ reduces to the set of Casimir invariants[1] of G and we arrive at well-known results[1] about the number of algebraically independent Casimir invariants.

Let us look at an example to illustrate the way in which these questions may be relevant to physics.

Example: Let us consider a system of two particles in nonrelativistic quantum mechanics. Let $\{\mathbf{Q}^{(i)}, \mathbf{P}^{(i)}, \mathbf{S}^{(i)}\}$ be the position, momentum, and spin observables for the particle i $(i=1,2)$. Together with the identity operator I, these observables generate a Lie algebra G of dimension $n=19$. The action, of the groups of translations, rotations, and Galilei transformations are generated respectively by the operators

$$\mathbf{P} = \sum_{i=1,2} \mathbf{P}^{(i)}, \quad \mathbf{J} = \sum_{i=1,2} \mathbf{Q}^{(i)} \times \mathbf{P}^{(i)} + \mathbf{S}^{(i)}, \quad \mathbf{G} = \sum_{i=1,2} m_i \mathbf{Q}^{(i)},$$

where m_i $(i=1,2)$ are the masses of the particles. The Hamiltonian of the system will be of the form

$$H = \sum_{i=1,2} (2m_i)^{-1} \mathbf{P}^{(i)2} + V,$$

where V represents the interaction term and depends upon the generators of the Lie algebra G. Galilean invariance requires V to satisfy the commutation relations

$$[\mathbf{P}, V] = [\mathbf{J}, V] = [\mathbf{G}, V] = 0.$$

Thus, the admissible solutions in the enveloping algebra $\mathcal{U}$ of G will be the invariant elements in $\mathcal{U}$ under the adjoint action of the subalgebra G_0 of G generated by $\{\mathbf{G}, \mathbf{P}, \mathbf{J}, I\}$. An easy computation yields $r(G, G_0) = 9$. Since $[G_0, G_0] = G_0$, we deduce $\tau(G, G_0) = 10$. A maximal algebraically independent set is given by

$$I, \quad \mathbf{S}^{(1)2}, \quad \mathbf{S}^{(2)2}, \quad \mathbf{q}^2, \quad \mathbf{p}^2, \quad \mathbf{l}^2,$$
$$\mathbf{lS}^{(1)}, \quad \mathbf{lS}^{(2)}, \quad \mathbf{pS}^{(1)}, \quad \mathbf{pS}^{(2)},$$

where $\mathbf{q}, \mathbf{p}$, and $\mathbf{l}$ are the relative position, momentum, and angular momentum observables defined by:

$$\mathbf{q} = \mathbf{Q}^{(1)} - \mathbf{Q}^{(2)}, \quad \mathbf{p} = (m_2 \mathbf{P}^{(1)} - m_1 \mathbf{P}^{(2)})/m_1 + m_2, \quad \mathbf{l} = \mathbf{q} \times \mathbf{p}.$$

Final Remark: From a result of Joseph[4] we may deduce the following upper bound for $\tau(G, G_0)$,

$$\tau(G, G_0) \leq 2|G| - |G_0| - r(G), \tag{21}$$

where $r(G)$ is the maximal rank of the $n \times n$ matrix function $M_{\alpha\beta}(a) = \sum_\nu c_{\alpha\beta}^\nu a_\nu (\alpha, \beta, \nu = 1,\ldots,n)$. An elementary application of Rouche's theorem yields $r(G) - r(G, G_0)$

$\leq |G| - |G_0|$. Then we obtain

$$|G| - r(G, G_0) \leq 2|G| - |G_0| - r(G).$$

Therefore, the bound (20) is stronger than (21).

[1]L. Abellanas and L. Martínez Alonso, J. Math. Phys. **16**, 1580 (1975).
[2]J. Patera *et al.*, J. Math. Phys. **17**, 977, 985 (1976); see, also, H. Zassenhaus, "On the invariants of a Lie group. I," in *Computers in Nonassociative Rings and Algebras* (Academic, New York, 1977).
[3]J. Dixmier, *Algebres enveloppantes* (Gauthier–Villars, Paris, 1974).
[4]A. Joseph, Bull. Soc. Math. France **102**, 75 (1974); "Sur les algebres de Weyl," Cours Université de Paris VI (1974).
[5]I. M. Gel'fand and A. A. Kirillov, IHES Publ. Math. **31**, 5 (1966).
[6]C. Chevalley, *Théorie des groupes de Lie. II* (Hermann, Paris, 1951).
[7]J. Dixmier, Bull. Soc. Math. France **85**, 336 (1957).

Constrained Hamiltonian formulation for interacting fields: Stable particlelike solutions[a)]

Gerald Rosen, Oleg Jakubowicz, and Vahe Tatoian

Department of Physics, Drexel University, Philadelphia, Pennsylvania 19104
(Received 16 May 1978)

It is observed that singularity-free localized particlelike solutions to certain essentially nonlinear classical field equations are dynamically stable in a constrained free-field Hamiltonian formulation. Being rather novel, the variational principle for such a theory pertains to an arbitrary pair of neighboring solutions to the field equations, with $\delta\phi$ specified as the difference between neighboring solutions.

PACS numbers: 03.50.Kk

1. INTRODUCTION

Inasmuch as a prefactor coupling constant ordinarily undergoes renormalization with quantization of the theory, it is permissible to view "bare" coupling constants as undetermined Lagrange multipliers in the context of prequantized classical theory. Interaction between fields, or nonlinear self-interaction of a field, arises from a linear-theoretic free-field Hamiltonian if the fields and their conjugate momentum densities are constrained by a subsidiary variational condition. What is particularly interesting about this alternative way of formulating interacting classical field theory is that it can engender stability for singularity-free localized solutions[1-10]—solutions that would otherwise be dynamically unstable with respect to evolution generated by the conventional interacting-field Hamiltonian without subsidiary constraints.

2. GENERAL FORMULATION

To put it precisely, let $H_0 = H_0[\phi, \pi]$ denote a generic free-field Hamiltonian in the canonically conjugate variables $\phi = (\phi_1(\mathbf{x}, t), \ldots, \phi_n(\mathbf{x}, t))$ and $\pi = (\pi_1(\mathbf{x}, t), \ldots, \pi_n(\mathbf{x}, t))$; H_0 is homogeneously quadratic or bilinear in the latter canonical variables. Suppose that (ϕ, π) and $(\phi + \delta\phi, \pi + \delta\pi)$ are *both* physically admissible neighboring solutions to the nonlinear (interacting) classical field theory if and only if

$$\delta H_0 = \int_{R_3} \left(\frac{\partial \phi}{\partial t} \circ \delta\pi - \frac{\partial \pi}{\partial t} \cdot \delta\phi \right) d^3x$$

$$\text{subject to } \delta C = 0, \tag{1}$$

where $C = C[\phi, \pi]$ is a prescribed functional that transforms like energy under Lorentz transformations. From (1) it follows that any admissible solution satisfies the field equations

$$\frac{\partial \phi}{\partial t} = \frac{\delta H_0}{\delta \pi} - \lambda \frac{\delta C}{\delta \pi}, \quad \frac{\partial \pi}{\partial t} = -\frac{\delta H_0}{\delta \phi} + \lambda \frac{\delta C}{\delta \phi} \tag{2}$$

with λ a Lagrange multiplier, while the difference between a generic pair of neighboring solutions satisfies the linear homogeneous constraint equation

$$\int_{R_3} \left(\frac{\delta C}{\delta \phi} \cdot \delta\phi + \frac{\delta C}{\delta \pi} \circ \delta\pi \right) d^3x = 0. \tag{3}$$

In (1)–(3) all functional derivatives are taken in the three-dimensional sense [e.g., $\delta/\delta\phi \equiv \delta/\delta\phi(\mathbf{x})$] and a dot is used to denote n-tuple contraction; otherwise the notation is standard.[11] The invariance of (1) under time translations (with no explicit time dependence in either H_0 or C) implies that the Lagrange multiplier λ is a constant independent of t. Clearly, Eqs. (2) are identical to conventional Hamiltonian field equations with $H \equiv H_0 - \lambda C$. However, the constraint equation (3) on the difference between neighboring solutions (ϕ, π) and $(\phi + \delta\phi, \pi + \delta\pi)$ provides a mild restriction on the manifold of admissible solutions to (2), a single global[12] condition at any instant of time on the $2n$ space–time functions $(\delta\phi, \delta\pi)$.

3. CONSTRAINED ϕ^6 MODEL

As a means of illustrating the effect of (3) on the manifold of solutions to (2), consider the model[1] with $n = 1$,

$$H_0 \equiv \tfrac{1}{2} \int_{R_3} (\pi^2 + |\nabla\phi|^2) d^3x, \quad C \equiv \tfrac{1}{6} \int_{R_3} \phi^6 d^3x. \tag{4}$$

Putting (4) into (2) and (3), we obtain

$$\frac{\partial^2 \phi}{\partial t^2} = \frac{\partial \pi}{\partial t} = \nabla^2\phi + \lambda\phi^5, \tag{5}$$

$$\int_{R_3} \phi^5 \delta\phi \, d^3x = 0. \tag{6}$$

The static spherically-symmetric singularity-free particlelike solution to (5) is $\phi = \bar{\phi} \equiv Z(r^2 + \frac{1}{3}\lambda Z^4)^{-1/2}$ for $\lambda > 0$, in which $r \equiv |\mathbf{x}|$ and the "size parameter" Z is a free (positive or negative) real constant of integration. For solutions in the neighborhood of the latter particlelike solution, (5) and (6) yield

$$\left(\frac{\partial^2}{\partial t^2} - \nabla^2 - 5\lambda\bar{\phi}^4 \right) \delta\phi = 0, \tag{7}$$

$$\int_{R_3} \bar{\phi}^5 \delta\phi \, d^3x = 0. \tag{8}$$

In terms of complex-valued radial functions $\xi_{klm} = \xi_{klm}(r)$ and the complex spherical harmonics Y_l^m, the general solution to (7) is expressible as

$$\delta\phi = r^{-1} \mathrm{Re} \left(\sum_{\text{eigen-}k} \sum_{l=0}^{\infty} \sum_{m=-l}^{l} \xi_{klm} Y_l^m e^{ikt} \right). \tag{9}$$

For $l > 0$, only real eigenvalues k are admitted by (7),[1] and the constraint condition (8) is satisfied automatically as a consequence of the angular integration. For $l = 0$ with $\xi_{k00} \equiv \xi_k$, (7) and (8) yield

[a)]Work supported in part by NASA grant NSG 7491.

 0022-2488/79/030441-02$01.00

$$\left(k^2+\frac{d^2}{dr^2}+5\lambda\bar{\phi}^4\right)\xi_k=0,\quad \int_0^\infty \bar{\phi}^5\xi_k r dr=0 \tag{10}$$

as conditions on admissible radial functions in (9). Subject to the boundary condition $\xi_k(0)=0$ implied by (9) at $r=0$, the eigenvalue equation in (10) admits a "ground state" with $k\cong\pm i(1.9)Z^{-2}\,g^{-1/2}$ and a continuum of "free states" for all real values of k.[1] Being nodeless, the "ground state" is precluded by the subsidary constraint condition in (10), and because the associated perturbation terms are thereby exluded from (9), the constraint formulation engenders dynamical stability for the particlelike solution. Elementary analysis also shows that the subsidiary condition in (10) restrict $l=0$ "free states" to the single eigenmode corresponding to the value $k=0$,

$$\xi_0\propto r\frac{\partial\bar{\phi}}{\partial Z}=(r^3-\tfrac{1}{3}\lambda Z^4 r)(r^2+\tfrac{1}{3}\lambda Z^4)^{-3/2}.$$

4. TRIAD-SCALAR ϕ^6 EXTENSION

Let us now consider the triad-scalar $n=3$ extension of (4) with

$$H_0\equiv\tfrac{1}{2}\sum_{j=1}^{3}\int_{R_3}(\pi_j^2+|\nabla\phi_j|^2)d^3x,\quad C\equiv\tfrac{1}{2}\int_{R_3}(\prod_{j=1}^{3}\phi_j^2)d^3x, \tag{11}$$

for which (2) and (3) produce

$$\frac{\partial^2\phi_i}{\partial t^2}=\frac{\partial\pi_i}{\partial t}=\nabla^2\phi_i+\lambda(\prod_{j(\neq i)}\phi_j^2)\phi_i, \tag{12}$$

$$\sum_{\substack{\text{perm}\\1,2,3}}\int\phi_1^2\phi_1^2\phi_3\,\delta\phi_3\,d^3x=0. \tag{13}$$

The static spherically-symmetric singularity-free particlelike solution to (12) is $\phi_1=\phi_2=\phi_3=\bar{\phi}$ $\equiv Z(r^2+\tfrac{1}{3}\lambda Z^4)^{-1/2}$ for $\lambda>0$, as in the case of (5). For the solutions in the neighborhood of the latter particlelike solution, (12) and (13) require

$$\left(\frac{\partial^2}{\partial t^2}-\nabla^2-\lambda\bar{\phi}^4\right)\delta\phi_i=2\lambda\bar{\phi}^4\sum_{j(\neq i)}\delta\phi_j \tag{14}$$

and (8) with $\delta\phi\equiv\delta\phi_1+\delta\phi_2+\delta\phi_3$. In terms of the latter normal coordinate and $\delta\phi'\equiv\delta\phi_1-\delta\phi_2$, $\delta\phi''\equiv\delta\phi_2-\delta\phi_3$, Eq. (14) resolves to (7) and

$$\left(\frac{\partial^2}{\partial t^2}-\nabla^2+\lambda\bar{\phi}^4\right)\begin{pmatrix}\delta\phi'\\ \delta\phi''\end{pmatrix}=0. \tag{15}$$

Since the effective potential in (15) is repulsive, only "free state" oscillatory modes appear in the general solutions for $\delta\phi'$ and $\delta\phi''$, and hence the constraint formulation engenders dynamical stability for the particlelike solution.

[1] G. Rosen, J. Math. Phys. **6**, 1269 (1965); **7**, 2066 (1966); **9**, 996, 999 (1968); **13**, 595 (1972).
[2] F. H. Roffman, Bull. Amer. Math. Soc. **76**, 70 (1970).
[3] M. S. Berger J. Func. Anal. **9**, 249 (1972).
[4] M. Creutz, Phys. Rev. D **12**, 3126 (1975).
[5] D. E. L. Pottinger and R. J. Rivers, Nucl. Phys. **117**, 189 (1976).
[6] Z. Horváth and L. Palla, Phys. Rev. D **14**, 1711 (1976).
[7] M. Dresden and S. F. Chen, Physica A **83**, 1 (1976).
[8] J. P. Hsu, Phys. Rev. Lett. **36**, 646 (1976).
[9] P. B. Burt, Lett. Nuovo Cimento **18**, 547 (1977).
[10] J. Hietarinta, J. Math. Phys. **18**, 65 (1977).
[11] G. Rosen, *Formulations of Classical and Quantum Dynamical Theory* (Academic, New York, 1969).

[12] This is to contrasted with the dynamical consistency problem solved in principle by P. A. M. Dirac [Proc. Roy. Soc. A **246**, 333 (1958)] for much stronger *local* constraints involving the components of ϕ and π. Required merely to be *stationary* with respect to neighboring solutions, C is *not* required to equal a certain fixed value for any solution, and thus λ is *not* an implicit functional of ϕ in our constrained Hamiltonian formulation.

Linearization stability of gravitational and gauge fields

Judith M. Arms

Department of Mathematics, University of Utah, Salt Lake City, Utah 84112
(Received 7 July 1978)

Conditions are given for the linearization stability of the Yang–Mills and the Einstein–Yang–Mills equations on a spacetime with a compact Cauchy surface. There are sufficient conditions on the Cauchy surface, and necessary and sufficient conditions on the spacetime; the latter are identified with global infinitesimal symmetries of the principal fiber bundle associated with the Yang–Mills (gauge) field. For each system a splitting theorem for the initial data is given and the Cauchy problem is discussed.

I. INTRODUCTION

This paper extends the linearization stability results of Fischer and Marsden[1] and Moncrief[2] (on vacuum spacetimes) and Arms[3] (on the coupled Einstein–Maxwell system) to the case of gravity coupled with a sourceless gauge field. For a spacetime with a compact Cauchy surface and a gauge field with matrix gauge group whose Lie algebra has a metric invariant under the adjoint action of the group, we obtain conditions for linearization stability of the coupled Einstein–Yang–Mills field equations. These conditions are sufficient conditions on the Cauchy surface and necessary and sufficient conditions on the spacetime. Roughly speaking, the results state that linearization stability can be guaranteed when the fields are not "too symmetrical." Thus a generic solution, which lacks symmetry, is linearization stable.

As a step in proving these results, we must show that the Cauchy problem for the Einstein–Yang–Mills system is well posed. In addition, Cauchy problem and linearization stability results are stated for Yang–Mills fields alone. The latter, which like the results for the coupled case state that linearization stability is correlated to lack of symmetry, are very similar to results of Moncrief[4] on Yang–Mills–Higgs fields.

Linearization stability concerns the validity of linear perturbation theory. A solution to the exact equations is said to be linearization stable if any solution to the linearized equations (relative to the exact solution) approximates to first order a curve of exact solutions. This property is not trivial, as is demonstrated by the example of Brill and Deser.[5]

We shall model a Yang–Mills (gauge) field as a 2 form F with values in the Lie algebra $\mathfrak{g}$ of a Lie group G and satisfying certain field equations. F comes from a Lie-algebra-valued 1 form A called the vector potential,

$$F = dA + [A, A]. \tag{1.1a}$$

(In coordinates,

$$F^a_{\mu\nu} = A^a_{\nu,\,u} - A^a_{\mu,\nu} + C^a_{bc}A^b_\mu A^c_\nu, \tag{1.1b}$$

where $[\ ,\ ]$ is the Lie bracket and C^a_{bc} are the structure constants of $\mathfrak{g}$.) [Note: In most of the physics literature, F and A take values in $i\mathfrak{g}$. This factor of i and factors of -1 and 2 coming from variations in several conventions lead to variations in (1.1) and subsequent equations. We detail our conventions in sec. II.] The Yang–Mills field equations are

$$\mathrm{div}F + [A, F] = 0. \tag{1.2a}$$

(In coordinates,

$$F^{a\nu}_{\mu;\nu} + C^a_{bc}A^b_\nu F^{c\nu}_\mu = 0, \tag{1.2b}$$

where the semicolon indicates the covariant derivative.) (1.1) and (1.2) may be derived from the action $\int F^a_{\mu\nu}F_a^{\mu\nu}d(^4\mathrm{vol.})$ by varying with respect to the potential A. (See, e.g., Ref. 6.)

The field and its potential may be multiple-valued globally. We may think of the field as a collection of locally defined single-valued fields. Two locally defined fields are part of the same global field if on the intersection U of their domains there is a gauge transformation $S: U \to G$, such that

$$\tilde{A} = \mathrm{ad}(S^{-1})A + S^*\theta, \quad \tilde{F} = \mathrm{ad}(S^{-1})F, \tag{1.3}$$

where $S^*\theta$ indicates the pullback to U of the canonical $\mathfrak{g}$-valued 1 form θ on G (Ref. 7, p. 41). [For G a matrix group, (1.3) becomes

$$\tilde{A} = S^{-1}AS + S^{-1}dS, \quad \tilde{F} = S^{-1}FS.]$$

The most natural geometric interpretation of the gauge field is as the curvature Ω of a connection ω on a principal fiber bundle P over spacetime with group G. (Cf. Ref. 8 and references therein.) Ω and ω are a 2 form and a 1 form, respectively, with values in $\mathfrak{g}$ and defined on all of P; we require that

$$\mathrm{div}\Omega + [\omega, \Omega] = 0.$$

Given a local section $i: U \to P$, U a neighborhood in spacetime, we can define $A = i^*\omega$ and $F = i^*\Omega$ on U; A and F will fit all the requirements of a Yang–Mills field.

The coupled Einstein–Yang–Mills system consists of the Yang–Mills field equations (1.1) and (1.2) and the Einstein field equations with a Yang–Mills source. The energy-momentum tensor for a gauge field is the obvious generalization of the energy–momentum of electromagnetism,

$$T_{\mu\nu} = F^a_{\mu\lambda}F^\lambda_{a\nu} - \tfrac{1}{4}g_{\mu\nu}F^a_{\lambda\sigma}F_a^{\lambda\sigma}, \tag{1.4}$$

where g is the spacetime metric. (Cf. Ref. 9, p. 68; the energy–momentum tensor above differs from that of Ref. 9 in that following Ref. 10 we use units such that $c = 16\pi$ (gravitational constant) $= 1$ and electric charge is in rationalized units. All our conventions are detailed in sec. II.) The Einstein–Yang–Mills field

 0022-2488/79/030443-11$01.00

equations come from the action

$$\int L = \int (R - \tfrac{1}{4} F^a_{\mu\nu} F_a^{\mu\nu})(-\det g)^{1/2} d^4x, \qquad (1.5)$$

where R is the scalar curvature of g, by varying with respect to g and A.

II. STATEMENT OF THE MAIN RESULTS

Before giving a precise statement of the results, we give some definitions and assumptions used throughout the rest of this paper. Notation and definitions concerning general relativity for the most part will follow Ref. 11, especially Chap. 21, except for units and universal constants which, as noted above, follow Ref. 10; some global definitions missing from these sources are drawn from Ref. 9 or 12. Our definitions for principal fiber bundles and related structures follow Ref. 7, except that we use the Bourbaki wedge product convention ($u \wedge v = u \otimes v - v \otimes u$) as in Ref. 11, thus eliminating factors of $\frac{1}{2}$ that appear in many of the formulas in Ref. 7.

Let $({}^4S, {}^4g)$ be a spacetime with a compact spacelike Cauchy surface M. That is, ${}^4S$ is a four-dimensional, C^∞, paracompact, connected, oriented, time-oriented Lorentz manifold with metric ${}^4g$ (signature $-+++$) and a compact C^∞ hypersurface M such that $g = {}^4g$ restricted to M is positive definite and every inextendible timelike curve in ${}^4S$ intersects M exactly once. Roughly speaking, the last condition means that appropriate data on M and the field equations will determine all the fields everywhere on ${}^4S$. (See sec. III.) In general, tensors on ${}^4S$ have a preceding superscript 4, while tensors on M do not. We work in coordinates $(x^0 = t, x^1, x^2, x^3)$ such that $\{t = 0\} \subset M$. Lower case Greek indices range from 0 to 3, inclusive, and are raised and lowered by ${}^4g$; lower case Latin indices from the middle of the alphabet $(i, j,$ etc.) range from 1 to 3 and are manipulated by g. A semicolon represents covariant differentiation with respect to ${}^4g$, a vertical bar, with respect to g. The volume element $(-\det {}^4g)^{1/2}\, d^4x$ is written ${}^4\mu$; $(\det g)^{1/2}\, d^3x$ is written μ. The second fundamental form of M is $k_{ij} = -Z_{i;j}$, where Z is the future pointing unit normal to M. The canonical momentum for g (see sec. III) is $\pi^{ij} = (k^m{}_m g^{ij} - k^{ij})\mu$, a tensor density.

Let G be a Lie group which is representable as a matrix group and has a Lie algebra $\mathfrak{g}$ with positive definite real inner product γ which is invariant under the adjoint action of G on $\mathfrak{g}$. Such a metric exists when, for example, G is the direct product of an abelian group and a compact semisimple group; the metric can be taken to be minus the Killing form on the semisimple factor.[13] All component work with respect to $\mathfrak{g}$ is written in terms of a left-invariant frame rather than a coordinate frame. Let $n =$ dimension of G and $m = n + 4$. Lower case Latin indices from the beginning of the alphabet range from 1 to n and are raised and lowered by γ. If $\{X_a\}$ is a basis for $\mathfrak{g}$, that is, a basis for the left-invariant vector fields, $[X_a, X_b] = C^c_{ab} X_c$ defines the structure constants C^c_{ab} of $\mathfrak{g}$.

Let P be a principal fiber bundle over ${}^4S$ with group G (acting on the *right*—see Ref. 7, Chap. I, Sec. 5) and connection one form ω. Thus the dimension of P as a manifold is m. Let $Q = P$ restricted to M. Given a local section i: $U \subset {}^4S \hookrightarrow P$, $\tau \circ i =$ identity where τ: $P \to {}^4S$ is the projection map, let ${}^4A = i^*\omega$ and ${}^4F = i^*\Omega$, where Ω is the curvature of ω. (See Ref. 7, Chap. II.) Then $A =$ restriction of ${}^4A$ to $M = i^*$ (ω restricted to Q). For $F =$ restriction of ${}^4F$ to M, it is useful to define $\beta = {}^*F$, where $*$ is the Hodge star operator, or, in indices, $\beta^i_a = \frac{1}{2}\gamma_{ab}[ijk]F^b_{jk}$, where $[ijk]$ is the completely antisymmetric tensor density with $[123] = 1$. If G is the circle, so that the Yang–Mills field ${}^4F$ is an electromagnetic field, then β is the magnetic field. We also define a generalized electric field $\epsilon^i_a = {}^4F^{0i}_a\, {}^4\mu$. β and ϵ are $\mathfrak{g}^*$-valued vector densities, where $\mathfrak{g} \approx \mathfrak{g}^*$ via the adjoint invariant metric γ.

All $\mathfrak{g}$- and $\mathfrak{g}^*$-valued tensors on M and ${}^4S$ throughout this paper, including those defined above, are *tensorial*, except for A and ${}^4A$ which are pseudotensorial. By this we mean that they come, via pullback by a local section, from tensorial forms on Q and P (or in the case of the potential, from the pseudotensorial form ω). (See Ref. 7 p. 75.) This property can also be expressed in terms of behavior under gauge transformations. A gauge transformation is a change of section from i to $\tilde{i}$, $\tilde{i}(x) = R(S(x))\,(i(x))$, where $R(S(x))$ indicates the action of $S(x) \subset G$ on the fiber $\tau^{-1}(x)$ and S: $U \to G$. A tensorial $\mathfrak{g}$-valued object T transforms by $\tilde{T} = \mathrm{ad}(S^{-1})T$, and the pseudotensorial potentials transform by $\tilde{A} = \mathrm{ad}(S^{-1})A + S^*\theta$. [Cf. (1.3) above.] $\mathfrak{g}^*$-valued objects transform so that gauge transformations commute with raising and lowering of indices. Note that if P is nontrivial so that there are no global sections, a $\mathfrak{g}$-valued function, such as V in Theorems 1A and 1B below, may be defined only locally, and its values will depend on the choice of gauge (i.e., section).

For (pseudo) tensorial forms the wedge product is the ordinary wedge product with respect to space or spacetime indices and a Lie bracket with respect to group indices; e.g.,

$$(A \wedge F)^a_{ijk} = C^a_{bc}(A^b_i F^c_{jk} + A^b_j F^c_{ki} + A^b_k F^c_{ij}).$$

The usual exterior derivative d on M or ${}^4S$ is replaced by the covariant exterior derivative D; e.g.,

$$D\phi = d\phi + A \wedge \phi \quad \text{for a tensorial form } \phi,$$

$$D\,{}^4A = d\,{}^4A + \tfrac{1}{2}[{}^4A \wedge {}^4A] = d\,{}^4A + [{}^4A, {}^4A] = {}^4F.$$

[Cf. (1.1) above. The potential has a different formula for its exterior covariant derivative because it is only pseudotensorial; cf. Ref. 7, Chap. II, Sec. 5, Theorem 5.2, p. 77, and Proposition 5.5, p. 79.]

In the following theorems, all tensor fields (including the linear and nonlinear perturbations implicit in a statement about linearization stability) may be taken to be C^∞; or all may be taken to be in suitable Sobolev spaces. Similarly throughout the paper "smooth" means C^∞ or belonging to a suitable Sobolev space.

Theorem 1A: The Yang–Mills field equations are linearization stable at a solution ${}^4A$ if and only if for all tensorial V: $M \to \mathfrak{g}$ such that $[\epsilon, V] = 0$ and $DV = 0$, the image of V lies in the center of $\mathfrak{g}$. [See Notes (1) and (2) below.]

Theorem 1B: The coupled Einstein–Yang–Mills sys-

tem is linearization stable at the solution $({}^4g, {}^4A)$ if the following three conditions hold on M:

(i) $\mathrm{tr}\pi = \pi^k{}_k$ is a constant multiple of the volume element μ;

(ii) at least one of g, π, ϵ, and β is nontrivial; that is, at least one of π, ϵ, and β is not identically zero, or g is not flat; and

(iii) if X is a vector field and V a $\mathfrak{g}$-valued function on M such that $\pounds_X g = 0$, $\pounds_X \pi = 0$, $\pounds_X \epsilon + [A(X), \epsilon] = [\epsilon, V]$, and $F(X, \ ) = DV$, then $X = 0$ and the image of V lies in the center of $\mathfrak{g}$.

Notes: (1) If the image of V is contained in the center of $\mathfrak{g}$, then V is a function in the usual sense because it is invariant under gauge transformations and thus is globally defined.

(2) For Theorem 1A (or if $X = 0$ in Theorem 1B), if the image of V lies in the center of $\mathfrak{g}$, then $DV = 0$ [or $F(X, \ ) = DV$] becomes $dV = 0$, so V is constant.

(3) $\pounds_X T$ is the Lie derivative of a tensor T with respect to X; e.g., $\pounds_X g_{ij} = X_{i|j} + X_{j|i}$.

(4) $\pounds_X \epsilon + [A(X), \epsilon]$ is the gauge covariant analog of $\pounds_X \epsilon$, and $F(X, \ )$ is the gauge covariant analog of $\pounds_X A$. Condition (iii) can be made more symmetric by replacing $F(X, \ ) = DV$ by $\pounds_X \beta + [A(X), \beta] = [\beta, V]$, but this weakens the theorem slightly.

Theorem 2A: The Yang–Mills equations are linearization stable at a solution ${}^4A$ if and only if there are no symmetries of P that preserve ${}^4S$ and the connection ω except the action of the center of G on P. That is, if X is a vertical vector field on P such that $\pounds_X \omega = 0$, then X is a generator of the action of the center of G on P, i.e., $\omega(X) =$ constant in the center of $\mathfrak{g}$.

Theorem 2B: The coupled Einstein–Yang-Mills system is linearization stable at the solution $({}^4g, {}^4A)$ if and only if there are no simultaneous symmetries on the bundle P of the connection 1 form ω and the pullback of the spacetime metric $\tau^{*4}g$ except the action of the center of G on P. That is, if X is a vector field on P such that $\pounds_X \omega = 0$ and $\pounds_X \tau^{*4} g = 0$, then X is a generator of the action of the center of G on P, i.e., X is vertical and $\omega(X) =$ const in the center of $\mathfrak{g}$.

Theorem 3: For the Yang–Mills and Einstein–Yang–Mills field equations, perturbations of the initial data split into three orthogonal components. At points of linearization stability, these are the perturbations violating the (linearized) constraint equations; the gauge perturbations; and the "true degrees of freedom," those preserving the constraints modulo the gauge freedom.

III. THE CAUCHY PROBLEM

Linearization stability of a hyperbolic system that forms a well-posed Cauchy problem is equivalent to linearization stability of any constraint equations on the initial data. (See Proposition 3 at the end of this section.) We outline below the proof that the Yang–Mills and Einstein–Yang–Mills systems are well-posed Cauchy problems. In the process each system will be broken up into constraint equations on the Cauchy surface M and Hamiltonian evolution equations. This Hamiltonian formulation will prove useful in the proofs of the main theorems.

A. The Hamiltonian formalism

Following the Arnowitt, Deser, and Misner formulation of general relativity,[10,14,15] from the Lagrangian $\mathcal{L}$ in (1.5) we derive the canonical momenta for gravity and a Yang–Mills field, either separately or coupled:

$$\pi^{ij} = \frac{\partial \mathcal{L}}{\partial g_{ij,0}} = (k^m_m g^{ij} - k^{ij})\,\mu$$

and

$$\eta^i_a = \frac{\partial \mathcal{L}}{\partial A^a_{i,0}} = {}^4F^{i0}_a\,{}^4\mu = -\epsilon^i_a.$$

(*Note*: This η^i_a is Moncrief's $E^i_{(a)}$[4] except for a factor of i since he considers $i\mathfrak{g}$-valued tensors instead of $\mathfrak{g}$-valued tensors.) The variables ${}^4g_{0\mu}$ and ${}^4A^a_0$ are degenerate in the sense that their canonical momenta are zero. It turns out to be most convenient to use as degenerate variables the lapse N, the shift X, and a tensorial $\mathfrak{g}$-valued function V, defined by

$${}^4g_{00} = -(N^2 - X^i X_i), \quad {}^4g_{0i} = X_i, \text{ and } {}^4A^a_0 = -V^a.$$

X is a vector field on M and its indices are manipulated by g rather than by ${}^4g$. When $G = S^1$ and the gauge field is thus a Maxwell field, V is the usual scalar potential; in the general case, V is still a scalar with respect to coordinate changes on M, although it is a vector in the Lie algebra.

For the Hamiltonian formalism, the action is expressed as

$$\int \mathcal{L} = \int \{\pi g_{,0} + \eta A_{,0} - NH - X\mathcal{J} - VK\}, \tag{3.1}$$

where

$$\begin{aligned} H &= \{-\tfrac{1}{2}(\pi^j{}_j)^2 + \pi^{ij}\pi_{ij} + \tfrac{1}{2}(\eta^i_a\eta^a_i + \beta^i_a\beta^a_i)\}g^{-1/2} - Rg^{1/2} \\ &= \{-\tfrac{1}{2}(\mathrm{tr}\bar{\pi})^2 + \bar{\pi}:\bar{\pi} + \tfrac{1}{2}(\bar{\epsilon}^2 + \bar{\beta}^2) - R\}\mu, \\ \mathcal{J}_i &= -2\pi^j_{i|j} + \eta^j_a(A^a_{j,i} - A^a_{i,j} + C^a_{bc}A^b_i A^c_j) \\ &= -2\nabla\cdot\pi + \eta\times\hat{\cdot}\,\beta, \\ K_a &= \eta^j_{a|j} + C^c_{ab}A^b_j\eta^j_c = \hat{\nabla}\cdot\eta. \end{aligned} \tag{3.2}$$

Here R is the scalar curvature of g; $g^{\pm 1/2} = (\det g_{ij})^{\pm 1/2}$; tr indicates the trace of a tensor with respect to g; a tensor density $T = \bar{T}\mu$, so $\bar{T}$ is an ordinary tensor; $\cdot$ and $\hat{\cdot}$ indicate an inner product on one pair of indices with respect to g and γ, respectively, so $\pi\cdot\pi = \pi^{ik}\pi_k{}^j$ $\pi:\pi = \pi^{ij}\pi_{ij}$, $\epsilon\cdot\epsilon = \epsilon^a_i\epsilon_{aj}$, $\epsilon\hat{\cdot}\epsilon = \epsilon^a_i\epsilon^a_i$, etc.; $\times$ is the usual cross product of vectors on a three-dimensional Riemannian space, so $\eta\times\hat{\cdot}\,\beta = [ijk]\eta^j_a\beta^{ak}$; ∇ and $\hat{\nabla}$ are the covariant and doubly covariant derivatives, and $\nabla\cdot$ and $\hat{\nabla}\cdot$ are the covariant and doubly covariant divergences, so $\nabla\cdot\pi = \pi^{ij}{}_{|j}$ and $\hat{\nabla}\cdot\eta = g^{ij}(\hat{\nabla}\eta)_{ij} = g^{ij}(\eta^a_{i|j} + C^a_{bc}A^b_i\eta^c_j)\,(= *D*\eta)$; and $\bar{\epsilon}^2 = \bar{\epsilon}\,\hat{:}\,\bar{\epsilon} = \bar{\eta}\,\hat{:}\,\bar{\eta}$ and $\bar{\beta}^2 = \bar{\beta}\,\hat{:}\,\bar{\beta}$. [The inner product signs $\cdot$ and $\hat{\cdot}$ in (3.1) are omitted because the products there are natural, i.e., do not require a metric. Note that because π, η, H, $\mathcal{J}$, and K are densities, it is not necessary to multiply the right-hand side of (3.1) by the volume element μ.] When $G = S^1$, so that the gauge field is an electromagnetic field, $\mathcal{J}$

differs from the $\mathcal{J}$ of Ref. 3 by a term $A^\circ(\hat{\nabla}\cdot\eta)$; this change makes $\mathcal{J}$ tensorial.

Let $\Phi=(\mathcal{H},\mathcal{J},\mathcal{K})$; then varying the action $\int\mathcal{L}=\{\pi g_{,0}+\eta A_{,0}-(N,X,V)\Phi\}$ with respect to the degenerate variables N, X, and V gives the constraint equations

$$\Phi=\begin{bmatrix}\mathcal{H}\\ \mathcal{J}\\ \mathcal{K}\end{bmatrix}=0; \tag{3.3}$$

varying with respect to g, A, π, and η gives the evolution equations. Let a prime stand for the functional derivative with respect to g, A, π, and η, and let (h,b,ρ,θ) be a variation in (g,A,π,η). $(h,b,\rho,\theta)\in$ domain of Φ',

$$\Phi':(S^2,\Lambda^1\otimes\mathfrak{g},S^2{}^*,\ (\Lambda^1\otimes\mathfrak{g})^*)\to(\Lambda^0{}^*,\chi^*,((\Lambda^0\times\mathfrak{g})^\sim)^*),$$

where

$S^2=\{$smooth symmetric covariant 2 tensors on $M\}$,

$\Lambda^k=\{$smooth k forms on $M\}$,

$\Lambda^k\otimes\mathfrak{g}=\{$smooth tensorial $\mathfrak{g}$-valued k forms on $M\}$,

$\chi=\{$smooth vector fields on $M\}$,

$(\Lambda^0\otimes\mathfrak{g})^\sim=$ quotient of $\Lambda^0\otimes\mathfrak{g}$ by the constant functions with values in the center of $\mathfrak{g}$,

$S^2{}^*=\{$smooth symmetric contravariant 2 tensor densities on $M\}$,

$\Lambda^0{}^*=\{$smooth scalar densities on $M\}$,

$(\Lambda^1\otimes\mathfrak{g})^*=\{$smooth tensorial $\mathfrak{g}^*$-valued vector densities on $M\}$,

$\chi^*=\{$smooth 1 form densities on $M\}$,

and

$((\Lambda^0\otimes\mathfrak{g})^\sim)^*=\{$smooth doubly covariant divergences on $M\}$,

$=$ image of $*D*=\hat{\nabla}\cdot:(\Lambda^1\otimes\mathfrak{g})^*\to(\Lambda^0\otimes\mathfrak{g})^*$.

[* indicates the natural L_2 dual, usually formed by raising or lowering each index by g or γ and considering tensor densities instead of tensors. For $(\Lambda^0\otimes\mathfrak{g})^\sim$, observe the following. Let $\mathfrak{g}_1=$ center of $\mathfrak{g}$ and let $\mathfrak{g}_2$ be its γ-orthogonal complement. Suppose that a basis $\{X_a\}_{a=1}^n$ for $\mathfrak{g}$ has been chosen so that $X_1,\ldots,X_{n_1}$ spans $\mathfrak{g}_1$ and $X_{n_1+1},\ldots,X_n$ spans $\mathfrak{g}_2$. Now $K_a=\eta^j_{a|j}+C^b_{ab}A^b_j\eta^j_c=\eta^j_{a|j}$ for $a\le n_1$. Let $V\in\Lambda^0\otimes\mathfrak{g}$, $V=V_1+V_2$, where V_i takes values in $\mathfrak{g}_i$. Consider $\int_m VK$. The constant part of V_1 does not contribute to the integral, because it is integrated against a pure divergence. Thus $V\in(\Lambda^0\otimes\mathfrak{g})^\sim$ is the natural dual to K.] Note that as usual the perturbation b of a connection A is tensorial even though A itself is only psuedotensorial.

$$\begin{aligned}(\textstyle\int\mathcal{L})'(h,b,\rho,\theta)&=\int\{-\pi_{,0}h-\eta_{,0}b+\rho g_{,0}+\theta A_{,0}\\&\quad-(N,X,V)\Phi'(h,b,\rho,\theta)\}\\&=\int\{(-\pi_{,0},-\eta_{,0},g_{,0},A_{,0})(h,b,\rho,\theta)\\&\quad-\Phi'^*(N,X,V)(h,b,\rho,\theta)\},\end{aligned} \tag{3.4}$$

where Φ'^* is the formal L_2 adjoint of Φ'. From (3.4) it is immediate that the evolution equations have the form

$$\frac{\partial}{\partial t}\begin{bmatrix}g\\A\\\pi\\\eta\end{bmatrix}=J\circ\Phi'^*\begin{bmatrix}N\\X\\V\end{bmatrix}, \tag{3.5}$$

where J is the antisymmetric matrix

$$\begin{bmatrix}0&\mathrm{Id}\\-\mathrm{Id}&0\end{bmatrix},\quad \mathrm{Id}=\text{identity}.$$

Note that the left-hand side of (3.5) is in the domain of Φ', while the image of Φ'^* is in the dual space to this; the rotation J matches the two sides of (3.5) correctly. We remark that (3.5) is in Hamiltonian form and that J is the symplectic form for this Hamiltonian system; the symplectic manifold is the "L_2 cotangent bundle" of the space of (g,A)'s; see Ref. 14, sec. 2 for a discussion of this symplectic structure in the case of vacuum spacetimes, and Ref. 16 for a general discussion of evolution equations having the form of (3.5).

B. Computations

We now state the derivatives $\Phi'=(\mathcal{H}',\mathcal{J}',\mathcal{K}')$ and their adjoints and the evolution equations (3.5) in detail. See Ref. 14 and references therein for more detailed calculations.

$$\begin{aligned}\mathcal{H}'(h,b,\rho,\theta)&=(\{-\pi^k_k\pi^{ij}+2\pi^{ik}\pi^j_k+\tfrac12(\eta^i_a\eta^{aj}+\beta^i_a\beta^{aj})\}h_{ij}\\&\quad+\{\tfrac14(\pi^j_j)^2-\tfrac12\pi^{ij}\pi_{ij}-\tfrac14(\eta^i_a\eta^a_i+\beta^i_a\beta^a_i)\}h^k_k)g^{-1/2}\\&\quad+\{-\tfrac12(h^j_j)R+(h^k_k)|^j_j-h_{ij}|^{ij}+R^{ij}h_{ij}\}g^{1/2}\\&\quad+[ijk](b^a_{k|j}+C^a_{bc}A^b_jb^c_k)\beta_{ai}g^{-1/2}\\&\quad+(-\rho^k_k\pi^j_j+2\pi^{ij}\rho_{ij})g^{-1/2}+\theta^i_a\eta^a_ig^{-1/2},\end{aligned} \tag{3.6}$$

$$\begin{aligned}\mathcal{J}'_i(h,b,\rho,\theta)&=-2\pi^{kj}_{|j}h_{ik}-2\pi^{jk}(h_{ij|k}-\tfrac12h_{jk|i})\\&\quad+\eta^j_a\{b^a_{j|i}-b^a_{i|j}+C^a_{bc}(A^b_ib^c_j+b^b_iA^c_j)\}-2\rho^j_{i|j}\\&\quad+\theta^j_a(A^a_{j|i}-A^a_{i|j}+C^a_{bc}A^b_iA^c_j),\end{aligned}$$

$$\mathcal{K}'_a(h,b,\rho,\theta)=C^c_{ab}\eta^j_cb^b_j+\theta^j_{a|j}+C^c_{ab}\theta^j_cA^b_j. \tag{3.7}$$

Letting Φ'_g be the derivative of Φ with respect to g, etc.,

$$\begin{aligned}\Phi'^*_g(N,X,V)&=\mathcal{H}'^*_gN+\mathcal{J}'^*_gX+\mathcal{K}'^*_gV\\&=(-\pi^k_k\pi^{ij}+2\pi^{ik}\pi^j_k+\tfrac12(\eta^i_a\eta^{aj}+\beta^i_a\beta^{aj})\\&\quad+\{\tfrac14(\pi^k_k)^2-\tfrac14\pi^{km}\pi_{km}-\tfrac14(\eta^k_a\eta^a_k+\beta^k_a\beta^a_k)\}g^{ij})g^{-1/2}N\\&\quad-\{(\tfrac12NR-N|^k_k)g^{ij}-NR^{ij}+N^{|ij}\}g^{1/2}\\&\quad+X^i_{|k}\pi^{kj}+X^j_{|k}\pi^{ik}-X^k\pi^{ij}{}_{,k}-X^k_{|k}\pi^{ij},\end{aligned} \tag{3.9}$$

$$\begin{aligned}\Phi'^*_A(N,X,V)&=\mathcal{H}'^*_AN+\mathcal{J}'^*_AX+\mathcal{K}'^*_AV\\&=[ijk]\{(N\bar\beta_{ak})_{|j}+C_{abc}N\bar\beta^c_kA^b_j\}-X^j\eta^i_{a|j}-X^j_{|j}\eta^i_a\\&\quad+X^i_{|j}\eta^j_a+X^i(\eta^j_{a|j}+C^c_{ab}A^b_j\eta^j_c)-X^jC^c_{ab}A^b_j\eta^i_c+C^b_{ca}\eta^i_bV^c,\end{aligned} \tag{3.10}$$

where $C_{abc}=\gamma_{ad}C^d_{bc}=C_{[abc]}$, because γ is adjoint invariant.

$$\Phi'^*_\pi(N,X,V)=(2\bar\pi_{ij}-\bar\pi^k_kg_{ij})N+X_{i|j}+X_{j|i}, \tag{3.11}$$

$$\begin{aligned}\Phi'^*_\eta(N,X,V)&=N\bar\eta^a_i+X^j(A^a_{i|j}-A^a_{j|i}+C^a_{bc}A^b_jA^c_i)\\&\quad-V^a_{,i}-C^a_{bc}A^b_iV^c.\end{aligned} \tag{3.12}$$

Applying J and switching to invariant notation, we obtain the evolution equations:

$$g_{,0} = N\{2\bar{\pi} - (\mathrm{tr}\bar{\pi})g\} + \mathcal{L}_X g, \tag{3.13}$$

$$A_{,0} = N\eta + F(X, \ ) - DV, \tag{3.14}$$

$$\begin{aligned}\pi_{,0} = {} & N\{(\mathrm{tr}\bar{\pi})\pi - 2\pi\cdot\pi - \tfrac{1}{2}(\bar{\eta}\hat{\cdot}\eta + \bar{\beta}\hat{\cdot}\beta) \\ & + (\{-\tfrac{1}{4}(\mathrm{tr}\bar{\pi})^2 + \tfrac{1}{2}\bar{\pi}\cdot\bar{\pi} + \tfrac{1}{4}(\bar{\epsilon}^2 + \bar{\beta}^2) + \tfrac{1}{2}R\}g^{\#} - \mathrm{Ric})\mu\} \\ & + \{(\Delta N)g^{\#} + \mathrm{Hess}N\}\mu + \mathcal{L}_X\pi,\end{aligned} \tag{3.15}$$

where Ric is the Ricci curvature, Δ is the Laplace–deRham operator, $\Delta N = -N^{|i}_{|i}$, $\mathrm{Hess}N = N^{|ij}$, $g^{\#} = g^{ij}$, and

$$\eta_{,0} = -\hat{\nabla}\times(N\beta) + \mathcal{L}_X\eta + [A(X), \eta] - (\hat{\nabla}\circ\eta)X - [\eta, V], \tag{3.16}$$

where $\hat{\nabla}\times$ is the doubly covariant curl [cf. first term in (3.10); $\hat{\nabla}\times$ can also be written $*D$]. These equations are coordinate and gauge covariant. A computation shows that $F(X, \ ) = \mathcal{L}_X A - D(A(X))$ is the "gauge covariant Lie derivative of A," i^*(Lie derivative of ω with respect to the horizontal lift of X). Similarly $\mathcal{L}_X\eta + [A(X), \eta]$ is the "gauge covariant Lie derivative of η," i^*[Lie derivative of $\Omega(\ ,$ lift of $Z)$ with respect to the horizontal lift of X].

For the Yang–Mills field alone, uncoupled to gravity, the constraint on the initial data is

$$K(A, \eta) \overset{\text{def}}{=} K(g, A, \pi, \eta) = 0,$$

where g and π are induced by the given background spacetime. The evolution equations are given by (3.14) and (3.16).

C. The nonlinear problem

We are now ready to state precisely the sense in which the Yang–Mills and Einstein–Yang–Mills systems are well-posed Cauchy problems. These propositions, and the linearized versions which follow, are stated for Sobolev spaces W^s defined using L_2 norms as in Ref. 9, Chap. 7.

Proposition 1A: Suppose $A \in W^{4+r}(M)$ and $\eta \in W^{3+r}(M)$, $r \geq 0$, satisfy the constraint $K(A, \eta) = 0$. Then there is a gauge field ${}^4A$ satisfying the Yang–Mills field equations on ${}^4S$ that induces A and η on M. ${}^4A \in W^{4+r}(U)$ for every relatively compact $U \subset {}^4S$ [by abuse of notation we write this ${}^4A \in W^{4+r}({}^4S)$], depends continuously on the given initial data (A, η), and is unique up to gauge transformation.

Proposition 1B: Suppose g and $A \in W^{4+r}(M)$ and π and $\eta \in W^{3+r}(M)$, $r \geq 0$, satisfy the constraint equations $\Phi(g, A, \pi, \eta) = 0$. Then there exists a maximal spacetime ${}^4S$ with metric ${}^4g$ and gauge field ${}^4A$ satisfying the Einstein–Yang–Mills system and an embedding $e: M \to {}^4S$ such that ${}^4g$ and ${}^4A$ induce the initial data (g, A, π, η) on M. ${}^4g$ and ${}^4A \in W^{4+r}({}^4S)$ (notation as in Proposition 1A) and depend continuously on the given initial data. $({}^4S, {}^4g, {}^4A)$ is unique in the sense that any two such developments $({}^4S, {}^4g, {}^4A)$ and $({}^4\tilde{S}, {}^4\tilde{g}, {}^4\tilde{A})$ can be transformed into each other by a simultaneous coordinate and gauge transformation. That is, suppose P and $\tilde{P}$ are the G-bundles with projections τ and $\tilde{\tau}$ to ${}^4S$ and ${}^4\tilde{S}$ and connections ω and $\tilde{\omega}$ that give rise to ${}^4A$ and ${}^4\tilde{A}$. Then there is a bundle isomorphism $\Psi: P \to \tilde{P}$ such that $\tilde{\tau}\circ\Psi\circ i\circ e = \tilde{e}: M \to {}^4\tilde{S}$ for any section $i: e(M) \to P$, $\Psi^*\tilde{\tau}^*{}^4\tilde{g} = \tau^*{}^4g$, and $\Psi^*\tilde{\omega} = \omega$ (equivalently $\Psi^*\tilde{\tau}^*{}^4\tilde{A} = \tau^*{}^4A$).

It is probably possible to reduce the required degree of differentiability by one (i.e., replace W^{4+r} and W^{3+r} by W^{3+r} and W^{2+r}, respectively); cf. Ref. 17 for treatment of the vacuum case.

The proof of these propositions is a generalization of the work of Choquet–Bruhat on the Cauchy problem for gravitational and electromagnetic fields; see, e.g., Ref. 9, Chap. 7 or Ref. 18. The Yang–Mills field equations are a quasilinear second-order hyperbolic system. In a generalized "Lorentz gauge," the principal part of the system decouples to give n copies of the principal part of Maxwell's equations; the existence argument for Maxwell fields and Maxwell fields coupled to gravity generalizes to Yang–Mills fields, and we have existence and uniqueness in Lorentz gauge (and harmonic coordinates).

By considering first other Yang–Mills gauges and then other coordinates, one can show that there is a unique gauge and coordinate transformation from any solution to the unique solution in Lorentz gauge and harmonic coordinates. If the bundle P is nontrivial (i.e., if ${}^4A$ can not be defined globally), the argument becomes somewhat involved, but uses no new ideas; see Ref. 19 for details.

The continuous dependence on initial conditions and maximality follow as usual. (See Ref. 9, Chap. 7.)

The following remarks show that Proposition 1 implies that the Hamiltonian evolution equations (3.5) have unique solutions for every choice of N, X, and V with ${}^4X = NZ + X$ timelike. First note that given a solution $({}^4S, {}^4g, {}^4A)$ to the Einstein–Yang–Mills system, the degenerate variables N, X, and V determine ${}^4g_{0\mu}$ and ${}^4A^a_0$, and thus determine the "timelike curves in P" through $Q = (P$ restricted to $M)$. When $N = 1$ and $X = 0$, these project to unit speed geodesics in ${}^4S$ normal to M; when $V = 0$, the curves in P are horizontal with respect to the connection ω. Any N, X, and V given on $\mathbb{R}\times M$ define a bundle map from $\mathbb{R}\times Q$ into P by picking out the timelike curves that should be the image of the curves $\mathbb{R}\times p$, $p \in Q$; the pullback of the metric and connection on P to $\mathbb{R}\times Q$ gives a solution to (3.5) which is clearly unique.

D. The linearized problem

We shall also need a similar result for the linearized equations. This is essentially a corollary of the proof of Proposition 1, which involves solving the linearized equations and then using an iteration argument to get the nonlinear result. However, there is a minor complication involving change of variables. In the proof, the equations are linearized with respect to the Lagrangian variables ${}^4g$ and ${}^4A$; we need a result for linearization with respect to the Hamiltonian variables g, A, π, η and the degenerate variables N, X, and V. Linearization (and linearization stability) depend on which variables are linearized. For example, the nonlinear equation

$$F(x_1, x_2) = x_1^2 + x_2^2$$

is not linearization stable at the origin, because its linearization

$$2x_1h_1 + 2x_2h_2 = 0h_1 + 0h_2 = 0,$$

is trivial there. If we let $y_i = x_i^2$, then the original equation is already linear: $y_1 + y_2 = 0$. Clearly the problem here is a singular change of variables. If there is a nonsingular change of variables, linearizing with respect to one set of variables is equivalent to linearizing with respect to the other. For suppose $x = f(y)$, $y = g(x)$, $f \circ g =$ identity $= g \circ f$. Then linearization of $F(x) = 0$ is equivalent to linearization of $G(y) \stackrel{\text{def}}{=} F(f(y)) = 0$ because h solves $F'(x) \cdot h = 0$ if and only if $k \stackrel{\text{def}}{=} g'(x) \cdot h$ solves $G'(y) \cdot k = 0$: $G'(y) \circ k = F'(f(y)) \cdot f'(y) \cdot g'(x) \circ h = F'(x) \cdot (\text{identity})' \cdot h = F'(x) \cdot h$. The variables g, A, π, η, N, X, and V are uniquely and differentiably defined by the variables ${}^4g$ and ${}^4A$ (and their derivatives), and vice versa, so that the linearized theorem for the Lagranian variables which falls naturally out of the proof of Proposition 1 is equivalent to a result for the linearized Hamiltonian equations,

$$\frac{\partial}{\partial t}\begin{bmatrix} h \\ b \\ \rho \\ \theta \end{bmatrix} = J \circ \Phi'^* \begin{bmatrix} L \\ Y \\ W \end{bmatrix} + J \circ \Phi'^{*\prime} \begin{bmatrix} N \\ X \\ V \end{bmatrix} \begin{bmatrix} h \\ b \\ \rho \\ \theta \end{bmatrix}; \qquad (3.17)$$

both results are combined in Proposition 2 below. Here (L, Y, W) is a linear perturbation of (N, X, V), so $L \in \Lambda^\circ$, $X \in \chi$, and $W \in \Lambda^\circ \otimes \mathfrak{g}$.

For initial data for the linearized theorem, we have (h, b, ρ, θ) on M satisfying the linearized constraint.

$$\Phi'(h, b, \rho, \theta) = 0. \qquad (3.18)$$

To construct initial data for the Lagrangian version of the results, we need in addition to specify (L, Y, W) and their first derivatives on M. From these we define ${}^4h$ and ${}^4b$ on M by:

$$\begin{aligned} {}^4h_{ij} &= h_{ij}, \\ {}^4h_{0i} &= h_{ij}X^j + Y_i, \\ {}^4h_{00} &= -2NL + 2X_iY^i + h_{ij}X^iX^j, \\ {}^4b_i^a &= b_i^a, \\ {}^4b_0^a &= -W^a. \end{aligned} \qquad (3.19)$$

The background metric and potential $({}^4g, {}^4A)$ determine (N, X, V) everywhere. Using (3.17) we find $h_{ij,0}$ and $b^a_{i,0}$ on M; we now know the first derivatives of h, b, N, X, L, Y, and W on M, so by differentiating (3.19) we can determine ${}^4h_{\mu\nu,0}$ and ${}^4b^a_{\mu,0}$ on M.

Proposition 2A: Suppose ${}^4A$ is a solution of the Yang–Mills field equations on ${}^4S$ and we have initial data b, θ, and W on M for the linearized system as described above. Then there is a solution ${}^4b$ to the linearized (Lagrangian) field equations with the given initial data, and any two such solutions differ by a linearized gauge transformation. [See (3.20) below.] Equivalently, for each choice of gauge V and linearized gauge W, the linearized Yang–Mills evolution equations [(3.17) with h, ρ, L, and Y set $= 0$] have a unique solution (b, θ).

Proposition 2B: Suppose $({}^4g, {}^4A)$ is a solution of the coupled Einstein–Yang–Mills system on ${}^4S$, and we have initial data on M for the linearized system as described above. Then there is a solution $({}^4h, {}^4b)$ on ${}^4S$ to the linearized (Lagrangian) field equations with the given initial data. Any two such solutions differ from each other by a linearized coordinate and gauge transformation:

$$\begin{aligned} {}^4\tilde{h} - {}^4h &= L_{4_{\tilde{Y}}}{}^4g, \\ {}^4\tilde{b} - {}^4b &= L_{4_{\tilde{Y}}}{}^4A - D({}^4\tilde{Y}) - D^4\tilde{W} = F({}^4\tilde{Y}, \ ) - D^4\tilde{W} \end{aligned} \qquad (3.20)$$

with ${}^4\tilde{Y}_\mu$, ${}^4\tilde{Y}_{\mu;0}$, ${}^4\tilde{W}$, and ${}^4\tilde{W}_{,0}$ all zero on M. Equivalently, for each choice of (gravity and Yang–Mills) gauge (N, X, V) and linearized gauge (L, Y, W) on ${}^4S = \mathbb{R} \times M$, the linearized equations (3.17) have a unique solution (h, b, ρ, θ).

Remark: The linearized coordinate and gauge transformation can be interpreted as Lie derivatives on the bundle P: $L_{4_{\tilde{Y}}}{}^4g = i^*(L_{m_{\tilde{Y}}}\tau^{*4}g)$ and $L_{4_{\tilde{Y}}}{}^4A - D({}^4A({}^4\tilde{Y}) + {}^4W) = i^*(L_{m_{\tilde{Y}}}\omega)$, where ${}^m\tilde{Y}$ = the horizontal lift of ${}^4\tilde{Y}$ – (generator of the G action on P corresponding to ${}^4W$). Cf. proof of Theorem 2 in Sec. IV.

Proof: We give the proof of 2B; 2A follows by setting the perturbations of the metric to zero throughout.

It suffices to consider the case where $({}^4g, {}^4A)$ are in harmonic coordinates and Lorentz gauge, as the other cases follow by a change of coordinates on the bundle P. Thus, throughout the proof, (N, X, V) in (3.17) are fixed. By the details of the proof of Proposition 1 (given in, e.g., Ref. 9), there exists a unique solution $({}^4h, {}^4b)$ to the linearized equations satisfying given linearized coordinate and gauge conditions. Now as in Ref. 14, sec. 4, we can decompose any symmetric 2 tensor ${}^4h$ on ${}^4S$ uniquely as

$${}^4h_{\mu\nu} = {}^4\hat{h}_{\mu\nu} + L_{4_{\hat{Y}}}{}^4g_{\mu\nu}, \qquad (3.21)$$

with ${}^4\hat{Y}_\mu$ and ${}^4\hat{Y}_{\mu;0}$ equal to zero on M and $({}^4\hat{h}_\mu{}^\alpha - {}^4\hat{h}^\beta{}_\beta{}^4g_\mu{}^\alpha)_{;\alpha} = 0$. Similarly, given ${}^4\hat{Y}$, any $\mathfrak{g}$-valued 1 form ${}^4b$ can be uniquely decomposed as

$${}^4b^a_\mu = {}^4\hat{b}^a_\mu + L_{4_{\hat{Y}}}{}^4A^a_\mu - D_\mu({}^4A({}^4\hat{Y}) + {}^4\hat{W}), \qquad (3.22)$$

where ${}^4\hat{W}$ and ${}^4\hat{W}_{,0}$ are zero on M and $\hat{\nabla} \cdot {}^4\hat{b} = {}^4b^{;\alpha}_\alpha + C^a_{bc}{}^4A^b_\alpha{}^4\hat{b}^{c\alpha} = 0$ [since ${}^4\hat{W}$ satisfies an equation of the form (7.20), p. 237, Ref. 9]. For any ${}^4\hat{Y}$ and ${}^4\hat{W}$, the quantities $L_{4_{\hat{Y}}}{}^4g$ and $L_{4_{\hat{Y}}}{}^4A - D({}^4A({}^4Y) + {}^4\hat{W})$, which are simply the derivatives of ${}^4g$ and ${}^4A$ under a simultaneous coordinate and Yang–Mills gauge transformation (i.e., a coordinate change on the bundle), will solve the linearized equations. Thus, if $({}^4h, {}^4b)$ is a solution to the linearized equations, by linearity $({}^4\hat{h}, {}^4\hat{b})$ will also be a solution. But $({}^4\hat{h}, {}^4\hat{b})$ satisfy linearized harmonic coordinate and Lorentz gauge conditions, and so must be the unique solution given by the linearized result in the proof of Proposition 1. Let $({}^4\hat{h}, {}^4\hat{b})$ be that unique solution; then $({}^4h, {}^4b)$ as in (3.21) and (3.22), for all possible ${}^4\hat{Y}$'s and ${}^4\hat{W}$'s with ${}^4\hat{Y}_\mu$, ${}^4\hat{Y}_{\mu;0}$, ${}^4\hat{W}$, and ${}^4\hat{W}_{,0}$ equal to zero on M, give all the possible solutions to the linearized equations. This completes the proof of the "Lagrangian half" of the theorem.

By the remarks before the theorem on the equivalence of linearizing the Lagrangian and Hamiltonian equations, any solution $({}^4h, {}^4b)$ to the linearized Lagrangian equations gives rise to (h, b, ρ, θ) and (L, Y, W) (as functions of t) solving (3.17). To complete the proof of the theorem, we need to show that we can specify (L, Y, W) arbitrarily and still get a unique (h, b, ρ, θ). It will suffice to prove the following lemma.

Lemma: The choice of linearized gauge (L, Y, W) is equivalent to the choice of $({}^4\hat{Y}, {}^4\hat{W})$ in (3.21) and (3.22); that is, for every choice of (L, Y, W) on $\mathbb{R}\times M = {}^4S$, there is a unique ${}^4\hat{Y}$ and ${}^4\hat{W}$ on ${}^4S$ so that (3.19) holds for $({}^4h, {}^4b)$ as in (3.21) and (3.22).

Proof: Suppose the solution $({}^4\hat{h}, {}^4\hat{b})$ is known and (L, Y, W) is given. Using (3.21) to eliminate ${}^4h$ in (3.19), we find that

$${}^4\hat{h}_{0i} + \mathcal{L}_{{}^4\hat{Y}}{}^4g_{0i} = ({}^4\hat{h}_{ij} + L_{{}^4\hat{Y}}{}^4g_{ij})X^j + Y_i$$

and

$$\begin{aligned}{}^4\hat{h}_{00} + \mathcal{L}_{{}^4\hat{Y}}{}^4g_{00} &= -2NL + X_iY^i + ({}^4\hat{h}_{ij} + L_{{}^4\hat{Y}}{}^4g_{ij})X^iX^j \\ &= -2NL + X_iY^i + ({}^4\hat{h}_{0i} + L_{{}^4\hat{Y}}{}^4g_{0i})X^i.\end{aligned}$$

We can rewrite this as

$$\mathcal{L}_{{}^4\hat{Y}}{}^4g_{0\mu} - (\mathcal{L}_{{}^4\hat{Y}}{}^4g_{j\mu})X^j = \begin{cases} -2NL + X_iY^i + {}^4\hat{h}_{0i}X^i - {}^4\hat{h}_{00} & \text{for } \mu = 0 \\ Y_i + {}^4\hat{h}_{ij}X^j - {}^4\hat{h}_{0i} & \text{for } \mu = i. \end{cases}$$

Change coordinates so that $\partial/\partial x^0 \to \partial/\partial x^0 - X$; then the equations become

$$\left.\begin{matrix} {}^4\hat{Y}_{0;0} \\ {}^4\hat{Y}_{0;i} + {}^4\hat{Y}_{i;0} \end{matrix}\right\} = \text{function independent of } {}^4\hat{Y}.$$

This is a first-order linear hyperbolic system with the x^0 coordinate curves as multiple characteristics, and has a unique solution ${}^4\hat{Y}$ defined for all time.

Now consider the equation for ${}^4\hat{W}$ obtained by using (3.22) to eliminate ${}^4b$ in (3.19): ${}^4\hat{b}_0 + \mathcal{L}_{{}^4\hat{Y}}{}^4A_0 - D_0({}^4A({}^4Y) + {}^4\hat{W}) = -W$. This first-order linear hyperbolic system with multiple characteristic tangent to $\partial/\partial x^0$ also has a unique solution ${}^4\hat{W}$ for all time. Thus given an arbitrary (L, Y, W), we can produce the unique desire ${}^4\hat{Y}$ and ${}^4\hat{W}$.

From Propositions 1 and 2 we can conclude, just as in Theorem 5.4 in Ref. 14, the following:

Proposition 3: The Yang-Mills or Einstein–Yang–Mills equations are linearization stable at ${}^4A$ or $({}^4g, {}^4A)$ if and only if the corresponding constraint equations are linearization stable at (A, η) or (g, A, π, η).

IV. PROOF OF THE MAIN RESULTS

We first show that the stated conditions are sufficient for linearization stability, and then show that some are also necessary. To show sufficiency, we use the implicit function theorem (IFT):

IFT: Suppose X and Y are Banach spaces and $\Phi: X \to Y$ is a C^r map, $r \geq 1$, $\Phi(x_0) = y_0$, $\ker\Phi'(x_0)$ splits [i.e., has a closed complement X_1 so $X = \ker\Phi'(x_0) \oplus X_1$], and $\Phi'(x_0)$ is surjective. Then $\{x: \Phi(x) = y_0\}$ is a C^r manifold near x_0 with tangent space $\ker\Phi'(x_0) = \{k: \Phi'(x_0)\cdot k = 0\}$ at x_0.

Comparing the conclusion of this theorem to the definition of linearization stability shows that the hypotheses of the IFT give sufficient conditions for linearization stability of $\Phi = y_0$ at x_0.

The key to the application of the IFT to the Yang–Mills and Einstein–Yang–Mills systems is the fact that the adjoint operators K'^* and Φ'^* defined in Sec. III are elliptic. There are several reasons for this. The applies to Banach spaces, so we must use Sobolev spaces in the proof of the theorems. (Specifically, we use W^s spaces with s sufficiently large to make K and Φ C^r maps and take care of other technical details; cf. discussion of such details in Ref. 14, Sec. 1, and Ref. 19.) The ellipticity of K'^* and Φ'^* gives a regularity lemma that allows generalization of the results to C^∞ spaces. The technical condition in the IFT—that $\ker\Phi'$ splits—follows immediately from the fact that Φ'^* (or K'^*) is elliptic. Finally, the ellipticity of Φ'^* implies that $Y = \ker\Phi'^* \oplus \mathrm{Im}\Phi'$, so K' or Φ' is surjective if and only if its adjoint is injective.

Thus conditions under which the adjoint operators are elliptic and injective are sufficient for linearization stability. To show necessity, we will show that linearization stability is inconsistent with a nontrivial kernel for the adjoint operator.

A. The adjoint operators are elliptic

We first consider the coupled case. Φ'^* is a linear partial differential operator

$$\Phi'^*: (\Lambda^0, \chi, (\Lambda^0\otimes\mathfrak{g})^\sim) \to (S^{2*}, (\Lambda^1\otimes\mathfrak{g})^*, S^2, \Lambda^1\otimes\mathfrak{g})$$
$$(N, X^i, V^a) \mapsto (\rho^{ij}, \theta^i_a, h_{ij}, h^a_i).$$

It is elliptic in the sense of Douglis and Nirenberg as extended by Hormander.[20] Let Φ'^*_{ij} represent the operator Φ'^* restricted to the jth "component bundle" of its domain and projected on the ith "component bundle" of its codomain, $j = 1, 2, 3$, $i = 1, 2, 3, 4$. Douglas and Nirenberg introduced the idea of weights s_i and t_j such that the order of Φ'^*_{ij} is less than or equal to $t_j - s_i$. The principal part of Φ'^*_{ij} is defined to be the part of Φ'^*_{ij} that is of order $t_j - s_i$ exactly, and the principal part of Φ'^* is the matrix of principal parts of the Φ'^*_{ij}'s. The principal symbol of the operator is now defined in the usual way.[14,20] Let $t_1 = t_2 = 2$, $t_3 = 1$, $s_1 = s_2 = s_4 = 0$, and $s_3 = 1$; then the only nonzero principal parts of Φ'^* are Φ'^*_{11}, Φ'^*_{32}, and Φ'^*_{43}. [Cf. (3.9)–(3.12)]. For each vector $\xi \in T^*M$, the principal symbol of Φ'^* is

$$\begin{aligned}\sigma(\xi): (N, X^i, V^a) &\mapsto (N(|\xi|^2 g^{ij} - \xi^i\xi^j), 0, -(\xi_iX_j + \xi_jX_i), V^a\xi_i). \end{aligned} \tag{4.1}$$

Φ'^* is said to be elliptic if $\sigma(\xi)$ is injective for every $\xi \neq 0$. But if $\xi \neq 0$ and $\sigma(\xi)(N, X, Y) = 0$, we see immediately from (4.1) that $X = 0$ and $V = 0$. Contracting the first slot in (4.1) gives $N(3|\xi|^2 - |\xi|^2) = 2|\xi|^2 N = 0$ because M is three dimensional, so $N = 0$. Thus Φ'^* is elliptic.

For Yang–Mills fields alone, we consider

$$\begin{aligned}K'^*: (\Lambda^0\otimes\mathfrak{g})^\sim &\to ((\Lambda^1\otimes\mathfrak{g})^*, \Lambda^1\otimes\mathfrak{g}), \\ V &\mapsto ([\eta, V], -DV)\end{aligned} \tag{4.2}$$

[cf. (3.10) and (3.12)], which, as above, has injective symbol and is therefore elliptic.

B. The Cauchy surface conditions imply that the adjoint operators are injective

K'^* is injective exactly when the conditions of Theorem 1A hold. Comparing the conditions and (4.2) shows that $V \in \ker K'^*$ implies that V takes values in the center of $\mathfrak{g}$ and $DV = dV = 0$, so V is a constant in the center of $\mathfrak{g}$, i.e., $V = 0$ in $(\Lambda^0 \otimes \mathfrak{g})^\sim$. Conversely, if $\ker K'^* = \{0\}$, the conditions of the theorem hold. Half of Theorem 1A is now proved.

To prove Theorem 1B it suffices to show that the three conditions stated in the theorem imply that Φ'^* is injective. So suppose that $(N, X, V) \in \ker \Phi'^*$ and conditions (i)–(iii) of the theorem hold. Then equations (3.9)–(3.12) are equal to zero. The trace of (3.11) is

$$N\bar{\pi}^k_k = 2X^k{}_{|k}. \tag{4.3}$$

Combining this equation and (3.11), one can compute that

$$(\mathcal{L}_X \pi)^k_k = X^j \pi^k_{k|j} + \{2N\pi^{jk}\pi_{jk} - \tfrac{1}{2}N(\pi^k_k)^2\} g^{-1/2}. \tag{4.4}$$

The trace of (3.9) is

$$0 = N\{-\tfrac{1}{2}(\pi^k_k)^2 + 2\pi^{jk}\pi_{jk} + \tfrac{1}{2}(\eta^k_a \eta^a_k + \beta^k_a \beta^a_k)\} g^{-1/2} - 2N^{|k}{}_{|k} g^{1/2} + X^k \pi^j{}_{j|k}, \tag{4.5}$$

where we have used (4.4) and the constraint $H = 0$ to eliminate the $\mathcal{L}_X \pi$ and curvature terms. By condition (i) the last term in (4.5) is zero. Now multiply (4.5) by N and integrate over M to get

$$0 = 2\int N^2 k^{ij} k_{ij} \mu + \frac{1}{2}\int N^2 \bar{\eta}^k_a \bar{\eta}^a_k \mu + \frac{1}{2}\int N^2 \bar{\beta}^k_a \bar{\beta}^a_k \mu + 2\int N^{|k} N_{|k} \mu \tag{4.6}$$

recalling that $k_{ij} = \tfrac{1}{2}\bar{\pi}^k{}_k g_{ij} - \bar{\pi}_{ij} = 0$ if and only if $\pi^{ij} = 0$. Observe that (4.6) implies that each integrand is zero, so N is constant, $Nk = 0$, $N\eta = 0$, and $N\beta = 0$. Condition (ii) now forces $N = 0$ by the following argument. If one of k (equivalently, π), $\eta = -\epsilon$, or β is nonzero, $N = 0$ is immediate from (4.6). If all three are identically zero, then $H = R = 0$, and $H'^*(N, X, V) = 0$, the first equation of $\Phi'^*(N, X, V) = 0$, reduces to

$$(NR^{ij} + N^{|k}{}_{|k} g^{ij}) g^{1/2} = 0. \tag{4.7}$$

However, (4.6) implies that N is constant, so (4.7) reduces further to $NR^{ij} = 0$. M is three dimensional, so the Ricci curvature $R^{ij} = 0$ if and only if g is flat; by condition (ii), g is not flat, so N must be identically zero.

With $N = 0$, $\Phi'^*(N, X, V) = 0$ reduces to

$$0 = X^i{}_{|k}\pi^{kj} + X^j{}_{|k}\pi^{ik} - X^k \pi^{ij}{}_{|k} - X^k{}_{|k}\pi^{ij} = -\mathcal{L}_X \pi, \tag{4.8}$$

$$\begin{aligned} 0 &= -X^j \eta^i_{a|j} - X^i_{|j}\eta^j_a + X^i_{|j}\eta^j_a + X^i(\eta^j_{a|j} + C^c_{ab} A^b_j \eta^j_c) \\ &\quad - X^j C^c_{ab} A^b_j \eta^i_c - C^c_{ab} V^b \eta^i_c \\ &= -\mathcal{L}_X \eta + (\hat{\nabla}\cdot\eta)X - [A(X) + V, \eta] = \mathcal{L}_X \epsilon + [A(X), \epsilon] - [\epsilon, V], \end{aligned} \tag{4.9}$$

because $K = \hat{\nabla}\cdot\eta = 0$,

$$0 = X_{i|j} + X_{j|i} = \mathcal{L}_X g, \tag{4.10}$$

$$\begin{aligned} 0 &= X^j(A^a_{i|j} - A^a_{j|i} + C^a_{bc} A^b_j A^c_i) - V^a_{|i} - C^a_{bc} A^b_i V^c \\ &= \mathcal{L}_X A - D(A(X) + V) = F(X, \,) - DV. \end{aligned} \tag{4.11}$$

By (iii) this implies that $X = 0$ and V has values in the center of $\mathfrak{g}$. As in the proof of Theorem 1A, this implies $V = 0$ in $(\Lambda^0 \otimes \mathfrak{g})^\sim$. This completes the proof of Theorem 1B. For the variation of the theorem mentioned in note (4) following the statement of the theorem, we need only show that (4.11) implies that $0 = \mathcal{L}_X \beta + [A(X), \beta] - [\beta, V]$. This follows from a tedious but straightforward calculation using the Bianchi identity for g, the Jacobi identity for $\mathfrak{g}$, and (4.10).

C. Identification of symmetries and the kernels of the adjoint operators

For Theorems 2A and 2B, $(0, 0, V) \in$ domain K'^* and $(N, X, V) \in$ domain Φ'^* must be identified as vector fields on the bundle and then $K'^*(0, 0, V)$ and $\Phi'^*(N, X, V)$ must be interpreted geometrically. Let ${}^4X = NZ + X$, a section of $(T^4S$ restricted to $M)$, and let ${}^mX = (\text{horizontal lift of } {}^4X) - V_p$, where V_p is the tangent to the fiber corresponding to $V \in \mathfrak{g}$. Because V is tensorial, ${}^mX$ is a right-invariant vector field on Q. (Note: the minus sign corresponds to the sign in $V = -{}^4A_0$.) Suppose ${}^mX$ is extended to all of P so that it is right invariant. Then the flow of ${}^mX$ is a one-parameter family of bundle automorphisms, or, in other words, of simultaneous coordinate and gauge transformations. Let $\tau^*{}^4g(\lambda)$ and $\omega(\lambda)$ be the pullback under this flow of $\tau^*{}^4g$ and ω. For small λ, M is spacelike in the metric ${}^4g(\lambda) = i^* \tau^*{}^4g(\lambda)$, so $g(\lambda)$ and $\omega(\lambda)$ induce $(g, A, \pi, \eta)(\lambda)$ on M. If $N^2 - X^i X_i \geq 0$ so that X is timelike, this is the evolution of (g, A, π, η) in ${}^4S$ with lapse N, shift X, and gauge V, so

$$\begin{bmatrix} h \\ b \\ \rho \\ \theta \end{bmatrix} \overset{\text{def}}{=} \frac{\partial}{\partial \lambda} \begin{bmatrix} g \\ A \\ \pi \\ \eta \end{bmatrix} = J \circ \Phi'^* \begin{bmatrix} N \\ X \\ V \end{bmatrix} \tag{4.12}$$

from the derivation of the evolution equations in Sec. III. Now (h, b, ρ, θ) are the "Hamiltonian variables," which depend linearly on the corresponding "Lagrangian variables,"

$$\begin{aligned} {}^4h &= \mathcal{L}_{{}^4X}\, {}^4g = i^* \mathcal{L}_{{}^mX} \tau^*{}^4g \\ \text{and} \qquad & \\ {}^4b &= \mathcal{L}_{{}^4X}\, {}^4A - D({}^4A({}^4X) + V) = i^* \mathcal{L}_{{}^mX} \omega. \end{aligned} \tag{4.13}$$

(Cf. discussion of linearization before Proposition 2 in Sec. III.) Note that $\mathcal{L}_{{}^mX}\omega$ is horizontal, i.e., tensorial, and so is completely specified by ${}^4b$.) ${}^4h$ and ${}^4b$ in turn depend linearly on ${}^mX$. Thus by linearity, (4.12) holds for all (N, X, V), because the timelike vectors span the tangent space.

(4.12) and (4.13) suggest that the elements of $\ker \Phi'^*$ are symmetries of the fields. (This idea for gravity alone is due to Moncrief.[2]) So suppose

$$\mathcal{L}_{{}^mX} \tau^*{}^4g = 0 \quad \text{and} \quad \mathcal{L}_{{}^mX}\omega = 0 \tag{4.14}$$

for a vector field ${}^mX$ on P. Let ${}^mU$ be any generator of the group action on P and let ${}^mY$ be the horizontal lift of any vector field on ${}^4S$ to P. Then

$$\begin{aligned} 0 &= L_{{}^mX}\omega({}^mU) = d\omega({}^mX, {}^mU) + {}^mU(\omega({}^mX)) \\ &= {}^mX(\omega({}^mU)) - \omega([{}^mX, {}^mU]) = -\omega([{}^mX, {}^mU]) \end{aligned}$$

and

$$0 = L_{m_X}\tau^{*4}g({}^mY, {}^mU)$$
$$= {}^mX(\tau^{*4}g({}^mY, {}^mU)) - \tau^{*4}g([{}^mX, {}^mY], {}^mU)$$
$$- \tau^{*4}g({}^mY, [{}^mX, {}^mU])$$
$$= -\tau^{*4}g({}^mY, [{}^mX, {}^mU])$$

using the fact that ${}^mU$ is vertical and ${}^mY$ is horizontal. This implies that $[{}^mX, {}^mU] = 0$, i.e., ${}^mX$ is right invariant. Thus ${}^4X = \tau_*{}^mX$ is well defined and the vertical component of ${}^mX$ is a tensorial function $-V: P \to \mathfrak{g}$. Let NZ and X be the normal and tangential components of ${}^4X$ on M. From (4.12)–(4.14), we see that $(N, X, V) \in \ker \Phi'^*$.

(Recall that V is actually a tensorial function modulo constant functions with values in the center of $\mathfrak{g}$. Those constant functions correspond to the action of the center of G on P, which is always a symmetry for any $\tau^{*4}g$ and ω. We will call such symmetries and their generators, ${}^mX = -V_p = \text{const}$, trivial, because they correspond to the trivial element of $\ker\Phi'^*$.)

Conversely, suppose we have an element of $\ker\Phi'^*$. We shall need to apply Proposition 2 in sec. III, so to avoid confusion with the gauge (N, X, V) and the perturbed gauge (L, Y, W) in the statement of that theorem, let $(\hat{L}, \hat{Y}, \hat{W})$ be the given element of $\ker\Phi'^*$. Let ${}^4\hat{Y} = \hat{L}Z + \hat{Y}$ on M and ${}^m\hat{Y} =$ (horizontal lift of ${}^4\hat{Y}$) $- \hat{W}_p$ on Q. We will extend ${}^m\hat{Y}$ uniquely to all of P, so that

$${}^4h = \mathcal{L}_{4\hat{Y}}{}^4g \quad \text{and} \quad {}^4b = \mathcal{L}_{4\hat{Y}}{}^4A - D({}^4A({}^4\hat{Y}) + \hat{W}) = i^*\mathcal{L}_{m\hat{Y}}\omega \tag{4.15}$$

are identically zero. Suppose the gauge (N, X, V) is fixed. Consider the lemma in the proof of Proposition 2. Let ${}^4\hat{h} = 0$ and ${}^4\hat{b} = 0$ on ${}^4V$. By the lemma, for any choice of (L, Y, W) we get existence of a unique ${}^4\hat{Y}$ and ${}^4\hat{W}$ on ${}^4S$ so that (4.15) [which corresponds to (3.21) and (3.22)] holds. We choose $(L, Y, W) \equiv 0$ so that by (3.19) ${}^4h_{0\mu} = 0$ and ${}^4b_0 = 0$. (In Sec. III, ${}^4\hat{Y}$ and ${}^4\hat{W}$ were zero on M, but the lemma clearly applies for any initial data.) Now with ${}^4h$ and ${}^4b$ as in (4.15), the corresponding Hamiltonian variables $(h, b, \rho, \theta) = J \circ \Phi'^*(\hat{L}, \hat{Y}, \hat{W}) = 0$ on M [cf. discussion of the geometric meaning of $\Phi'^*(N, X, V)$ above], and satisfy the linearized evolution equations (3.17) with perturbed gauge $(L, Y, W) = 0$ by the way ${}^m\hat{Y}$ was extended to P. By Proposition 2, the obvious solution $(h, b, \rho, \theta) \equiv 0$ is the unique solution, so ${}^4h \equiv 0$ and ${}^4b \equiv 0$. Thus we have established a one-to-one correspondence between $\ker\Phi'^*$ and the set of nontrivial simultaneous symmetries of $\tau^{*4}g$ and ω; by setting all objects related to gravity (e.g., N, X, L, Y, h) equal to zero in the above argument, we have a one-to-one correspondence between $\ker K'^*$ and the nontrivial symmetries of ω which preserve ${}^4S$. This proves the sufficiency of the conditions in Theorems 2A and 2B.

D. The second-order conditions

It has been shown above that $\ker K'^*$ or $\ker\Phi'^*$ is trivial exactly when the conditions of Theorems 1A, 2A, or 2B, as appropriate, hold. To show that these conditions are necessary as well as sufficient for linearization stability, we assume stability and nontriviality of the kernels. From these assumption come second-order conditions on the linear perturbations. In the following sections, we show the existence of linear perturbations violating these second-order conditions. This contradiction shows that the assumption was false, and linearization stability requires that the kernels be trivial; i.e., the conditions of the theorems are necessary, and when they are violated we have linearization instability.

Assume that $K(A, \eta) = 0$ is linearization stable at (A_0, η_0). Then for any solution (b, θ) of the linearized equation

$$K'(A_0, \eta_0)(b, \theta) = 0,$$

there is a one-parameter family $(A(\lambda), \eta(\lambda))$ such that

$$(A(0), \eta(0)) = (A_0, \eta_0), \quad (A'(0), \eta'(0)) = (b, \theta),$$

and

$$K(A(\lambda), \eta(\lambda)) = 0 \tag{4.16}$$

Differentiating (4.16) twice and evaluating at $\lambda = 0$ yields

$$K'(A_0, \eta_0) \cdot (A''(0), \eta''(0))$$
$$+ K''(A_0, \eta_0) \cdot ((b, \theta), (b, \theta)) = 0. \tag{4.17}$$

Now assume that there is $V \neq 0$ in $\ker K'^*$. Contract (4.17) against V and integrate over M to get

$$0 = \int_M V \cdot \{K'(A''(0), \eta''(0)) + K''((b, \theta), (b, \theta))\}$$
$$= \int_M \{K'^*(V) \cdot (A''(0), \eta''(0)) + V \cdot K''((b, \theta), (b, \theta))\}$$
$$= \int_M V \cdot K''((b, \theta), (b, \theta)), \tag{4.18}$$

a second-order condition on the linear perturbation (b, θ). Similarly, a nontrivial $(N, X, V) \in \ker\Phi'^*$ gives rise to a second-order condition on linear Einstein–Yang–Mills perturbations (h, b, ρ, θ),

$$\int_M (N, X, V) \cdot \Phi''((h, b, \rho, \theta), (h, b, \rho, \theta)) = 0. \tag{4.19}$$

We note that the second-order condition is analogous to the quadratic algebraic condition on the moduli space in the theorem of Kuranishi on small deformations of complex structures (See, e.g., Ref. 21 and references therein). This theorem has been used by Atiyah, Hitchin and Singer[22] in the study of the space of self-dual Yang–Mills fields on Riemannian manifolds. In the cases considered by Atiyah, Hitchin, and Singer, the second cohomology of a certain elliptic complex (analogous to $\ker K'^*$ or $\ker\Phi'^*$ in the present work) is trivial, so that the theorem of Kodaira, Nirenberg, and Spencer (which the Kuranishi theorem generalizes) gives a linearization stability argument. In the case that the second cohomology is not trivial, the Kuranishi theorem says that the space of nonlinear deformations of complex structures is (locally) parameterized by linear deformations h satisfying

$$H(Q(h, h)) = 0, \tag{4.20}$$

where Q is a certain quadratic function and H is projection onto the second cohomology. If H is projection onto $\ker K'^*$ or $\ker\Phi'^*$ and Q is K'' or Φ'', then (4.20) is our second-order condition. This suggests, as has been suggested before[4,14,23] that at points of linearization instability, the second-order condition is sufficient to

select exactly those linear perturbations which are tangent to curves of nonlinear perturbations. This appears to be true in the case of gravity alone.[24]

E. Instability for symmetric Yang-Mills fields

Differentiating (3.8) and substituting into (4.18),

$$2\int_M C^c_{ab}\, V^a b^b_i \theta^i_c = 2\int_M [V, b]\hat{:}\theta = 0 \tag{4.21}$$

for each $V \in \ker K'^*$. We wish to find (b, θ) satisfying

$$K'(b,\theta) = \hat{\nabla}\cdot\theta + [b,\eta] = 0 \tag{4.22}$$

and violating (4.21). This is most easily done using a particular gauge for the background fields (A, η) that corresponds to a reduction of the principal bundle P. From (4.2), $V \in \ker K'^*$ means

$$DV = 0 \quad \text{and} \quad [\eta, V] = 0. \tag{4.23}$$

Recall that $\mathfrak{g} = \mathfrak{g}_1 \otimes \mathfrak{g}_2$, where $\mathfrak{g}_1$ is the center of $\mathfrak{g}$ and $\mathfrak{g}_2$ its orthogonal complement. Let $V = V_1 + V_2$, $V_i \in \mathfrak{g}_i$. Now $DV_1 = dV_1 \in \mathfrak{g}_1$, and $DV_2 \in \mathfrak{g}_2$ because $C_{[abc]} = 0$. Therefore, by (4.23), $DV_1 = dV_1 = 0$, and so $V_1 = 0$ in $(\Lambda^0 \otimes \mathfrak{g})^{\sim}$. Thus, it suffices to consider $V = V_2$. Let $\mathfrak{g}^V$ be the centralizer of $V(p)$ for any one $p \in P$; $\mathfrak{g}^V$ is a subalgebra of $\mathfrak{g}$, and let G^V be the corresponding subgroup of G. By the holonomy theorem of Ambrose and Singer combined with a reduction theorem due to Cartan, Ehresmann, Kobayashi, and Nomizu (Ref. 7, Theorems 7.1 and 8.1, pp. 83–90, and note 2, p. 288), the bundle and its connection may be reduced to a bundle and connection with group G^V. The reduced bundle may be identified with $\{q: p$ and q are joined by a horizontal curve$\}$, a subbundle embedded in P. By choosing sections of the subbundle, we can assume that the range of A, β and $\eta \subset \mathfrak{g}^V$, and that $V = \text{const}$ because $0 = DV = dV$ on horizontal curves. Similarly, if there are any other functions W such that $DW = 0$, we may assume that the bundle has been reduced and the gauge chosen so that A and β take values in $\mathfrak{g}^V \cap \mathfrak{g}^W$ and $W = \text{const}$.

Let $\mathfrak{g}^{V\perp}$ be the γ orthogonal complement of $\mathfrak{g}^V$ in $\mathfrak{g}$. Note that D and D^* map $\Lambda^k \otimes \mathfrak{g}^{V\perp}$ into $\Lambda^m \otimes \mathfrak{g}^{V\perp}$ (for appropriate k and m) because for $W \in \mathfrak{g}^{V\perp}$

$$[A, W]\hat{:}\mathfrak{g}^V = [\mathfrak{g}^V, A]\hat{:}W \subset \mathfrak{g}^V\hat{:}W = 0 \tag{4.24}$$

(again using $C_{[abc]} = 0$). Consider

$$D^*: \Lambda^1 \otimes \mathfrak{g}^{V\perp} \to \Lambda^0 \otimes \mathfrak{g}^{V\perp},$$

which has surjective but not injective symbol, just like the ordinary divergence d^*. By Theorem 1 of Ref. 25, this implies that $\ker D^*$ is infinite dimensional. Now consider

$$T: \ker D^* \to \mathfrak{g}^{V\perp},$$
$$W \mapsto \int_M [W;\eta].$$

T has nontrivial kernel because $\mathfrak{g}^{V\perp}$ is finite dimensional while $\ker D^*$ is infinite dimensional; let b be any nonzero element of $\ker T$.

Let $\theta = \theta^t + D\mu$, where $\hat{\nabla}\cdot\theta^t = -D^*\theta^t = 0$. We will choose μ so that the linearized constraint equation (4.22) is satisfied and choose θ^t so that the second-order condition (4.21) is violated. (4.22) becomes

$$-\hat{\nabla}\cdot\theta = D^*D\mu = \hat{\Delta}\mu = [b;\eta], \tag{4.25}$$

where $\hat{\Delta} = D^*D$ is the "gauge-covariant Laplace–deRham operator" on $\mathfrak{g}^{V\perp}$-valued functions. Just as for the usual Laplacian on a compact manifold, (4.25) has a solution, unique up to $\ker\hat{\Delta} = \ker D = \{$covariant constant functions$\}$ if and only if $[b;\eta]$ is orthogonal to $\ker D$. But the gauge was chosen so that if $W \in \ker D$, then W is constant; then $\int([b;\eta]\hat{:}W) = (\int[b;\eta])\hat{:}W = 0$, because $b \in \ker T$. Thus we can solve (4.25) for $\mu \in \Lambda^0 \otimes \mathfrak{g}^{V\perp}$. (4.21) becomes

$$2\int_M [V, b]\hat{:}\theta = 2\int_M [V, b]\hat{:}\theta^t = 0, \tag{4.26}$$

because

$$\begin{aligned}\int_M [V, b]\hat{:}D\mu &= \int_M C^c_{ab}\, V^a b^b_i (\mu^d_{,j} + C^d_{ef} A^e_j \mu^f)\gamma_{cd} g^{ij} \\ &= -\int_M C^c_{ab} V^a (b^b_{i\,|j} + C^b_{ef} A^e_j b^f_i)\mu^d \gamma_{cd} g^{ij} \\ &= \int_M [V, D^*b]\hat{:}\mu = 0,\end{aligned}$$

using integration by parts, $V = \text{const}$, $C_{[abc]} = 0$, the Jacobi identity in $\mathfrak{g}$, $[V, A] = 0$, and $b \in \ker D^*$. It suffices to show that there is at least one element $\theta^t \in \ker D^*$ such that (4.26) does not hold. Assume not; i.e., assume that (4.26) holds for all θ^t. Then $[V, b]$ is orthogonal to $\ker D^*$, so $[V, b] = DW$ for some $W \in \Lambda^0 \otimes \mathfrak{g}^{V\perp}$, and $\hat{\Delta}W = D^*[V, b] = d^*[V, b] + [A;[V, b]] = [V, d^*b + [A;b]] = [V, D^*b] = 0$, using $V = \text{constant}$, the Jacobi indentity, $[V, A] = 0$, and $b \in \ker D^*$. Thus $W \in \ker\hat{\Delta} = \ker D$, so $[V, b] = DW = 0$. But b takes values in $\mathfrak{g}^{V\perp}$, so this forces b to be zero. This contradiction proves that there is a θ^t so (4.26) does not hold. Thus, a Yang–Mills symmetry $V \in \ker K'^*$ is inconsistent with linearization stability. (We thank Kobayashi for suggesting the final argument, and Singer for inspiring the approach to gauge fields used in the above proof.) Theorems 1A and 2A are now proved.

F. Instability for symmetric Einstein-Yang-Mills fields

To complete the proof of Theorem 2B, we show the existence of $(h, b, \rho, \theta) \in \ker\Phi'$ and violating (4.19). First consider the case where $(N, X, V) = (0, 0, V) \in \ker\Phi'^*$. Then ${}^mX$ is a vertical vector field on P, that is an infinitesimal change of Yang–Mills gauge only, and (4.19) reduces to (4.18). For the (b, θ) shown to exist in Sec. IV.E, $(0, b, 0, \theta) \in \ker\Phi'^*$, by the following argument. β and η take values in $\mathfrak{g}^V$ while b and θ were chosen to take values in $\mathfrak{g}^{V\perp}$. $\hat{\nabla}\times b = {}^*Db$ will also take values in $\mathfrak{g}^{V\perp}$ by (4.24), so

$$H'(0, b, 0, \theta) = \hat{\nabla}\times b\,\hat{:}\,\overline{\beta} + \theta\hat{:}\overline{\eta} = 0$$

and

$$\mathcal{J}'(0, b, 0, \theta) = \eta\times\hat{:}(\hat{\nabla}\times b) + \theta\times\hat{:}\beta = 0.$$

When $(N, X) \neq 0$, then the argument for the existence of a linear perturbation violating (4.19) goes exactly as in the case of gravity alone.[26] (The addition of the Yang–Mills fields adds purely algebraic terms to the differential operators involved, so that the analysis is not affected.) We briefly sketch the argument. Since nonzero elements of $\ker\Phi'^*$ correspond to symmetries, the existence of such elements is hypersurface invariant and we may, by choosing an appropriate hypersurface, assume $N \neq 0$, say $N > 0$, on $U \subset M$. We construct operators $P: S^2 \to (\Lambda^1, \Lambda^0)$ and $T: S^2 \to S^{2*}$ such that (4.27) and (4.28) below hold and

$$K = \{(h, 0, \rho, 0): P(h) = 0,\ \rho = T(h),$$

and $\text{supp}h \subset U\} \subset \ker\Phi'^*$,

and again using Theorem 1 of Ref. 25, K is infinite dimensional. For elements of K, (4.19) becomes

$$\int_M \{N(-\tfrac{1}{2}(\nabla h)^2 + \tfrac{1}{2}(\nabla \operatorname{tr} h)^2)$$
$$+ \text{(lower order terms in } h \text{ and } \rho)\} = 0. \tag{4.27}$$

P is chosen so that

$$\|\nabla \operatorname{tr} h\|_0 \leq \text{const}\|h\|_0. \tag{4.28}$$

From (4.27), (4.28), and $P(h)=0$, integration by parts, and the Schwarz inequality, it can be calculated that

$$\|h\|_1 \leq \text{const}\|h\|_0,$$

which by Rellich's theorem gives a compact embedding of $K \cap W^1$ in W^0. This contradicts the infinite dimensionality of K, thus showing the existence of perturbations violating the second-order condition and completing the proof of Theorems 1A, 1B, 2A, and 2B.

G. The splitting theorem

This theorem, which in the Einstein–Yang–Mills case generalizes that of Moncrief[27] for gravity alone, follows immediately from application of a general splitting in sympletic geometry given in Ref. 28. K and Φ are moments for the actions of the group of bundle automorphisms of Q which fix M and the dynamical group of coupled gravitational and gauge fields, respectively. The generators of the group actions are given by the ranges of $J \circ K'^*$ and $J \circ \Phi'^*$. Applying the splitting theorem in Ref. 28, we have

$$(\Lambda^1 \otimes \mathfrak{g},\ (\Lambda^1 \otimes \mathfrak{g})^*) = \text{Im}K'^* \oplus \text{Im}J \circ K'^* \oplus \ker(-K' \circ J) \cap \ker K'$$

and

$$(S^2,\ \Lambda^1 \otimes \mathfrak{g},\ S^{2*},\ (\Lambda^1 \otimes \mathfrak{g})^*)$$
$$= \text{Im}\Phi'^* \oplus \text{Im}J \circ \Phi'^* \oplus \ker(-\Phi'^* \circ J) \cap \ker\Phi'.$$

In each splitting, the sums are L_2 orthogonal, and at points of linearization stability the first, second, and third summands are as stated in the theorem, respectively.

ACKNOWLEDGMENTS

It is a pleasure to thank R. Jantzen, S. Kobayashi, J.E. Marsden, M.H. Protter, I.M. Singer, and A.H. Taub for many helpful discussions.

[1] A.E. Fischer and J.E. Marsden, Bull. Am. Math. Soc. **79**, 997–1003 (1973); Proc. Symp. Pure Math. AMS **27**, Part 2, 219–63 (1975).
[2] V. Moncrief, J. Math. Phys. **16**, 493–98 (1975).
[3] J.M. Arms, J. Math. Phys. **18**, 830–33 (1977).
[4] V. Moncrief, Ann. Phys. (N.Y.) **108**, 387–400 (1977).
[5] D. Brill and S. Deser, Commun. Math. Phys. **32**, 291–304 (1973).
[6] C.N. Yang and R.L. Mills, Phys. Rev. **96**, 191–95 (1954).
[7] S. Kobayashi and K. Nomizu, *Foundations of Differential Geometry* (Interscience, New York, 1963), Vol. 1, pp. 63–112.
[8] T.T. Wu and C.N. Yang, Phys. Rev. D **12**, 3845–57 (1975).
[9] S.W. Hawking and G.F.R. Ellis, *The Large Scale Structure of Space–Time* (Cambridge U.P., Cambridge, England, 1973).
[10] R. Arnowitt, S. Deser, and C.W. Misner, *Gravitation: An Introduction to Current Research*, edited by L. Witten (Wiley, New York, 1962), pp. 227–65.
[11] C.W. Misner, K.S. Thorne, and J.A. Wheeler, *Gravitation* (Freeman, San Francisco, 1973).
[12] R.K. Sachs and H.H. Wu, *General Relativity for Mathematicians* (Springer–Verlag, New York, 1977).
[13] S. Helgason, *Differential Geometry and Symmetric Spaces* (Academic, New York, 1962), p. 122.
[14] A.E. Fischer and J.E. Marsden, "Topics in the Dynamics of General Relativity," in *Proceedings of the Summer School of the Italian Physical Society*, edited by J. Ehlers (to appear).
[15] A. Hanson, T. Regge, and C. Teitelboim, *Constrained Hamiltonian Systems* (Accademia Nazionale dei Lincei, Rome, 1976); P. Codero and C. Teitelboim, Ann. Phys. 100, 607–31 (1976).
[16] A.E. Fischer and J.E. Marsden, "A new Hamiltonian structure for the dynamics of general relativity," Gen. Rel. Grav. **7**, 915–20 (1976).
[17] T.J.R. Hughes, T. Kato, and J.E. Marsden, Arch. Rational Mech. Anal. **63**, 273–94 (1977).
[18] Y. Choquet-Bruhat, *Gravitation: An Introduction to Current Research*, edited by L. Witten (Wiley, New York, 1962), pp. 130–68.
[19] J.M. Arms, "Linearization stability of coupled gravitational and gauge fields" (Ph.D. thesis, Univ. California, Berkeley, 1977).
[20] L. Hörmander, *Linear Partial Differential Operators* (Springer–Verlag, Berlin, 1963), pp. 267–68; Ann. Math. **83**, 129–209 (1966).
[21] J. Morrow and K. Kodaira, *Complex Manifolds* (Holt, Reinhart, and Winston, New York, 1971), pp. 155–73.
[22] M.F. Atiyah, N.J. Hitchin, and I.M. Singer, Proc. Nat. Acad. Sci. USA **74**, 2662–663 (1977); "Self-duality in four-dimensional Riemannian geometry," Proc. R. Soc. Lond. A **362**, 425–61 (1978).
[23] V. Moncrief, J. Math. Phys. **17**, 1893–902 (1976).
[24] A.E. Fischer, J.E. Marsden, and V. Moncrief, "Symmetry breaking in general relativity," to appear in Gen. Rel. Grav.; "The structure of the solution space of Einstein's equations" (in preparation).
[25] J.-P. Bourguignon, D.G. Ebin, and J.E. Marsden, C.R. Acad. Sci. Paris Sér. A **282**, 867–70 (1976).
[26] J.M. Arms and J.E. Marsden, "The absence of Killing fields is necessary for linearization stability of Einstein's equations," Ind. Math. J. (Jan. 1979).
[27] V. Moncrief, J. Math. Phys. **16**, 1556–60 (1975).
[28] J.M. Arms, A.E. Fischer, and J.E. Marsden, C.R. Acad. Sci. Paris, Ser. A **281**, 517–20 (1975).

Conservation laws for the classical Maxwell–Dirac and Klein–Gordon–Dirac equations[a)]

R. T. Glassey
Department of Mathematics, Indiana University, Bloomington, Indiana 47401
W. A. Strauss
Department of Mathematics, Brown University, Providence, Rhode Island 02912
(Received 24 July 1978)

Solutions to the classical coupled Maxwell–Dirac and Klein–Gordon–Dirac equations in a space–time of dimension four are considered. These equations are invariant under the 15-dimensional conformal group, in the case of zero mass. The resulting conservation laws are explicitly exhibited in terms of the Cauchy data at a fixed time in a form suitable for analysis by the techniques of partial differential equations.

1. INTRODUCTION

The Maxwell–Dirac equations are the basic equations of relativistic quantum electrodynamics. In the notation of Ref. 1 they take the form

$$(i\not\partial - M)\psi = g\not A\psi,$$
$$\frac{\partial}{\partial x^\nu} F^{\mu\nu} = g\bar\psi\gamma^\mu\psi, \tag{1}$$

where

$$F^{\mu\nu} = \frac{\partial A^\mu}{\partial x_\nu} - \frac{\partial A^\nu}{\partial x_\mu}.$$

Mathematical as well as physical interest centers on the questions of global existence of solutions to the Cauchy problem and their asymptotic behavior.

We consider solutions of the Cauchy problem for the classical versions of these equations in a space–time of dimension four. Global existence of such solutions is an open problem. Only in the case of one space dimension has it been established and the asymptotic behavior of solutions analyzed.[2–4]

A similar situation prevails for solutions to the classical Klein–Gordon–Dirac equations

$$(i\not\partial - M)\psi = -g\phi\psi, \tag{2}$$
$$(-\Box + m^2)\phi = g\bar\psi\psi$$

although certain "special" global solutions are known in three space dimensions.[5] These special solutions are a consequence of one of the identities we give in Sec. 3.

For those wave equations for which an adequate theory exists, certain explicit *a priori* estimates have played a central role.[6–12] Mathematicians collectively call them "energy estimates" although only one of them represents the physical energy. Many of them are due to Morawetz.[6] Some of them are a direct consequence of the conformal invariance of the equations and Noether's Theorem.[13] Indeed, if an equation is given by a Lagrangian L and if L is invariant under a one-parameter family of transformations, the solutions of the equation satisfy a conservation law. The infinitesimal generator $\mathcal{H}$ of the family of transformations can be used directly with the equation to write the conservation law. An exposition of these ideas, appropriate for the present context, may be found in Ref. 7.

It is well known that the M–D and the K–G–D systems possess a Lagrangian formulation and, if the mass is zero, are invariant under the 15-dimensional conformal group C.[14,15] This fact is (indirectly) reproved in this paper. The multipliers $\mathcal{H}$ depend linearly on the ψ and A^μ (or ψ and ϕ) and their first derivatives. Thus there are 15 independent conservation laws, one for each generator of C. From the four translations come the energy and momenta and from the six Lorentz transformations come the angular momenta in space–time. The other five generators of C, the dilation and four inversions, provide five more conservation laws. They were explicitly calculated[7] for the case of the nonlinear Klein–Gordon equation

$$u_{tt} - \Delta u + m^2 u + F(u) = 0$$

and were applied to the existence and scattering theory of solutions.

The purpose of this paper is to explicitly write the conservation laws for the classical M–D and K–G–D equations, in the hope that they can be similarly applied to the mathematical theory.

What makes our calculations nontrivial is the form in which the conservation laws are to be written. The usual form in tensor notation[16] is not suitable for our purposes because the algebraic signs of the various terms must be made explicit. For the Cauchy problem the data are prescribed at a fixed time, while in scattering problems the data at times $+\infty$ and $-\infty$ are studied. Thus we separate the time from the space variables. The resulting identities are written in terms of the Cauchy data at fixed times. The inversional identities are quite complicated; so far as we know, this is the first time they have been explicitly written in such a form.

In Sec. 2 we list the notational conventions and then briefly discuss the existence question. The conservation laws for the K–G–D and M–D equations are derived in Secs. 3 and 4, respectively. For each of these systems we exhibit two additional conservation laws, giving a total of 17. We also derive some new bounds which must be satisfied by any finite-energy solution. Of course, for the M–D equations we must also append a gauge condition. In Sec. 4 we first give

a)Research supported in part by NSF grants MCS 77-01340 and MCS 75-08827.

 0022-2488/79/030454-05$01.00

the results in the Lorentz gauge $\partial A^\mu/\partial x^\mu = 0$ and then in the Coulomb (or radiation) gauge $\nabla \cdot \mathbf{A} = 0$. The latter is better suited to our purposes, because the major terms in the energy density are positive.

The actual calculations, although elementary, are prohibitively lengthy and therefore some of the derivations will only be briefly sketched.

We have also carried out this program for the Yang–Mills equations in Minkowski space, and can draw some conclusions about the asymptotic behavior of classical gauge fields. The details will appear elsewhere.

2. EXPLICIT FORM OF THE EQUATIONS AND NOTATION

The Dirac spinor $\psi = \psi(x, t)$ $(x \in \mathbb{R}^3)$ is a function on space–time into spin space, a complex four-dimensional space. $\psi^\dagger$ denotes the conjugate transpose of ψ, and $\bar{\psi}$ is defined by $\bar{\psi} = \psi^\dagger \gamma^0$, where $\gamma^0 = \left[\begin{smallmatrix} I & 0 \\ 0 & -I \end{smallmatrix}\right]$ is a 4×4 matrix. (I here is the 2×2 identity matrix). The A^μ $(\mu = 0, 1, 2, 3)$ are real-valued electromagnetic potentials defined on $\mathbb{R}^3 \times \mathbb{R}$. In standard notation[1] the Dirac matrices γ^μ $(\mu = 0, 1, 2, 3)$ are given by

$$\gamma^k = \begin{bmatrix} 0 & \sigma^k \\ -\sigma^k & 0 \end{bmatrix} \quad (k = 1, 2, 3),$$

where the σ^k's are the Pauli matrices

$$\sigma^1 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad \sigma^2 = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}, \quad \sigma^3 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}.$$

Thus $(\gamma^0)^* = \gamma^0$, $(\gamma^k)^* = -\gamma^k$ $(k = 1, 2, 3)$ and $\gamma^\mu \gamma^\nu + \gamma^\nu \gamma^\mu = 2g^{\mu\nu} I$, where $g^{00} = +1$, $g^{kk} = -1$ $(k = 1, 2, 3)$, $g^{\mu\nu} = 0$ for $\mu \neq \nu$.

The M–D equations have the form (1), where

$$\not{A} = \gamma^0 A^0 - \sum_{k=1}^{3} \gamma^k A^k,$$

$$i\not{\partial} = i\gamma^0 \frac{\partial}{\partial t} + i \sum_{k=1}^{3} \gamma^k \frac{\partial}{\partial x^k},$$

where $x^\mu = (t, x^1, x^2, x^3)$ denote the physical variables, and $x^\mu = g^{\mu\nu} x_\nu$ (summation on ν). In order to write the Dirac operator in symmetric hyperbolic form, we introduce the matrices

$$\beta = -i\gamma^0, \quad \alpha_k = \gamma^k \gamma^0 \quad (k = 1, 2, 3).$$

[These are *not* the same α_k, β as used in Ref. 1]. Then we have $\alpha_k^* = \alpha_k$, $\beta^* = -\beta$, $\alpha_k^2 = I$, $\beta^2 = -I$, $\alpha_k \beta + \beta \alpha_k = 0$, $\alpha_k \alpha_j + \alpha_j \alpha_k = 0$, $j \neq k$.

The M–D equations can then be written in the Lorentz gauge as

$$\psi_t = \sum_k \alpha_k \frac{\partial \psi}{\partial x^k} + M\beta\psi - igA^0\psi - ig\sum_k A^k \alpha_k \psi, \tag{3a}$$

$$A^0_{tt} - \Delta A^0 = g\psi^\dagger \psi, \tag{3b}$$

$$A^k_{tt} - \Delta A^k = -g\psi^\dagger \alpha_k \psi \quad (k = 1, 2, 3), \tag{3c}$$

$$\frac{\partial A^0}{\partial t} + \sum_k \frac{\partial A^k}{\partial x^k} = 0.$$

Here Δ denotes the Laplacian $\sum_k (\partial/\partial x^k)^2$, and all sums here and throughout the paper will be taken over the indices 1, 2, 3 unless otherwise specified. We also write $\mathbf{A} = (A^1, A^2, A^3)$.

We now consider the effect of a gauge transformation. Let $(\psi, A^0, \mathbf{A})$ be a solution of the M–D equations (3) in the Lorentz gauge. Let $\Phi = \Phi(x, t)$ be a sufficiently smooth real-valued function. The change of variables

$$\begin{aligned} &\psi = \exp(ig\Phi)\psi', \\ &A^0 = A^{0\prime} - \Phi_t, \\ &A^k = A^{k\prime} + \Phi_{x^k} \quad (k = 1, 2, 3) \end{aligned} \tag{4}$$

converts the M–D equations to the system

$$\begin{aligned} &\frac{\partial \psi'}{\partial t} = \sum_k \alpha_k \frac{\partial \psi'}{\partial x^k} + M\beta\psi' - igA^{0\prime}\psi' - ig\sum_k A^{k\prime}\alpha_k \psi', \\ &A^{0\prime}_{tt} - \Delta A^{0\prime} + \frac{\partial}{\partial t}(\Delta\Phi - \Phi_{tt}) = g|\psi'|^2, \\ &A^{k\prime}_{tt} - \Delta A^{k\prime} + \frac{\partial}{\partial x^k}(\Phi_{tt} - \Delta\Phi) = -g\psi'^\dagger \alpha_k \psi' \quad (k = 1, 2, 3) \end{aligned} \tag{5}$$

with gauge condition

$$\frac{\partial A^{0\prime}}{\partial t} + \sum_k \frac{\partial A^{k\prime}}{\partial x^k} - \Phi_{tt} + \Delta\Phi = 0.$$

Suppose we want the "primed" equations to be in the Coulomb (or radiation) gauge:

$$\sum_k \frac{\partial A^{k\prime}}{\partial x^k} = 0.$$

Then Φ should be chosen so that

$$\Phi_{tt} - \Delta\Phi = \frac{\partial A^{0\prime}}{\partial t},$$

which, by the second of Eqs. (4), is the same as requiring

$$-\Delta\Phi = \frac{\partial A^0}{\partial t}. \tag{6}$$

With this choice of Φ, the M–D equations in the Coulomb gauge take the form

$$\begin{aligned} &\frac{\partial \psi'}{\partial t} = \sum_k \alpha_k \frac{\partial \psi'}{\partial x^k} + M\beta\psi' - igA^{0\prime}\psi' - ig\sum_k A^{k\prime}\alpha_k \psi', \\ &-\Delta A^{0\prime} = g|\psi'|^2, \\ &A^{k\prime}_{tt} - \Delta A^{k\prime} + \frac{\partial^2 A^{0\prime}}{\partial t \partial x^k} = -g\psi'^\dagger \alpha_k \psi' \quad (k = 1, 2, 3), \\ &\sum_k \frac{\partial A^{k\prime}}{\partial x^k} = 0. \end{aligned} \tag{7}$$

In a similar way the K–G–D equations can be put into the form

$$\psi_t = \sum_k \alpha_k \frac{\partial \psi}{\partial x^k} + (M - g\phi)\beta\psi, \tag{8a}$$

$$\phi_{tt} - \Delta\phi + m^2\phi = g\bar{\psi}\psi, \tag{8b}$$

where, of course, ϕ is a real-valued (scalar) function.

We conclude this section with some brief comments on the existence of solutions. First we must introduce the norms

$$\|u\|_p = \left(\int |u(x)|^p dx\right)^{1/p}, \quad 1 \leq p < \infty,$$

$$\|u\|_\infty = \operatorname{ess\,sup} |u(x)|.$$

The integration is taken over all space $x \in \mathbb{R}^3$ here as elsewhere (unless otherwise specified). A pair of standard inequalities, associated with the name of

Sobolev, are

$$\|u\|_6 \leq \text{const}\,\|\nabla u\|_2$$

and[19]

$$\|u\|_\infty \leq \text{const}\,(\|\nabla^2 u\|_2 + \|u\|_2).$$

Denote by H^n the usual Sobolev space of functions which are square-integrable over $\mathbb{R}^3$ together with their derivatives up to order n. It is a Hilbert space with the norm

$$\|u\|_{H^n}^2 = \sum_{|\alpha|\leq n} \|D^\alpha u\|_2^2,$$

where D^α is a derivative of order $|\alpha|$.

Consider the Cauchy (initial-value) problem for the K–G–D equations (8). Following a procedure[17,3] which is by now very well known, we can solve (8) by iteration, locally in time. The basic space is the space in which ψ has (spinor) values in H^n, ϕ has (real) values in H^n, and ϕ_t has (real) values in H^{n-1} at each time, where $n \geq 2$. The basic nonlinear mapping takes (ψ, ϕ, ϕ_t) into $(\phi\psi, 0, \bar{\psi}\psi)$. By the Sobolev inequality, this mapping is locally Lipschitz on this space. It follows[17,3] that the usual iterates (well known to mathematicians and physicists alike) converge in nontrivial time interval to a solution of (8). Because n can be taken arbitrarily large, the solutions of (8) are smooth (have square-integrable derivatives of all orders). Thus we are dealing with *bona fide* solutions of (8). The only trouble is the extension of the solutions to all times. This is a problem of great difficulty which has not yet been solved. To solve it, it would be sufficient to show that the solutions of (8) do not blow up (become unbounded) in the norm of the above space at finite times. Certain weaker *a priori* estimates on solutions would also be sufficient. Below we prove an estimate (13), which appears to be too weak for this purpose, but only slightly so. For the M–D equations, (3) or (7), exactly the same local existence theorem is valid,[18] and the general remarks on the global question remain applicable.

3. CONSERVATION LAWS FOR THE K-G-D EQUATIONS

The simplest conservation law results from multiplying (8a) by $\psi^\dagger$ and taking the real part. Thus

$$\frac{\partial}{\partial t}|\psi|^2 = \sum_k \frac{\partial}{\partial x^k}\psi^\dagger \alpha_k \psi.$$

If this identity is integrated over all space at a fixed time, we obtain the conservation of *charge*

$$\int |\psi|^2\,dx = \text{const}. \tag{9a}$$

If, on the other hand, we integrate it over the interior of a light cone and use the divergence theorem, we obtain

$$\int_K \psi^\dagger \left(I \pm \sum \frac{x^k}{r}\alpha_k\right)\psi\,dS \leq \sqrt{2}\int |\psi|^2\,dx \tag{9b}$$

By (9a) the right side is a constant. Here I is the identity matrix, $r = |x|$, dS is the usual three-dimensional surface measure on K, and K is any characteristic light cone $|x| = |t| + \text{const}$. The plus sign is to be used on forward light cones and the minus sign on backward light cones. Notice that the 4×4 matrix $I \pm \sum (x^k/r)\alpha_k$ has eigenvalues 2, 2, 0, 0. Hence (9b) is an estimate on half of the components of ψ (those in the 2-eigenspace) on light cones.

Another quadratic invariant was discovered in Ref. 5. In the representation of the α_k's, β given in Sec. 2 it takes the form

$$\int [|\psi_1 - \bar{\psi}_4|^2 + |\psi_2 + \bar{\psi}_3|^2]\,dx = \text{const}. \tag{10}$$

If initially the constant in (10) vanishes, then $\bar{\psi}\psi(x, t) \equiv 0$ for all t and global existence for such initial conditions follows easily from this.[5]

To derive energy conservation, we multiply (8b) by ϕ_t and then multiply (8a) by $2\psi_t^\dagger$ on the left and take the imaginary part of the result. Adding these two expressions, we obtain the identity

$$\frac{\partial}{\partial t} e(\psi, \phi) = \sum_k \frac{\partial}{\partial x^k}[\text{Im}\psi^\dagger \alpha_k \psi_t - \phi_t \phi_{x^k}], \tag{11}$$

where the energy density is given by

$$e(\psi, \phi) = \text{Im}\sum_k \psi^\dagger \alpha_k \psi_{x^k} - (M - g\phi)\bar{\psi}\psi - \tfrac{1}{2}(\phi_t^2 + |\nabla\phi|^2 + m^2\phi^2).$$

This gives the conservation of *energy*

$$\int e(\psi, \phi)\,dx = \text{const}. \tag{12}$$

Even though the energy density is not positive or negative, we can still obtain an estimate which could prove useful in the global existence question. It is

$$\int (\phi_t^2 + |\nabla\phi|^2 + m^2\phi^2)\,dx \leq \text{const}\{\int |\nabla\psi|^2\,dx + 1\}^{1/2} \tag{13}$$

with a constant independent of time. Inequality (13) is derived by estimating each of the first three terms in (12) separately. The mass term is bounded because of (9a). By (9a) and the Schwarz inequality,

$$|\text{Im}\sum_k \int \psi^\dagger \alpha_k \psi_{x^k}\,dx| \leq \text{const}\,\|\nabla\psi\|_2.$$

The third term in (12) is estimated using the Hölder inequality and then the Sobolev inequality as

$$\left|\int \phi\bar{\psi}\psi\,dx\right| \leq \|\phi\|_6 \|\psi\|_6^{1/2}\|\psi\|_2^{3/2}$$

$$\leq \text{const}\,\|\nabla\phi(t)\|_2 \|\nabla\psi(t)\|_2^{1/2}\|\psi(0)\|_2^{3/2}.$$

The resulting estimate implies (13).

To derive momentum conservation, we multiply (8a) by $\psi_{x^j}^\dagger$, take the imaginary part, and then multiply (8b) by ϕ_{x^j}. The following identity results:

$$\frac{\partial}{\partial t} P_j(\psi, \phi) = \frac{1}{2}\frac{\partial}{\partial x^j}(\phi_t^2 - |\nabla\phi|^2 - m^2\phi^2) \tag{14}$$

$$+ \sum_k \frac{\partial}{\partial x^k}(\phi_{x^j}\phi_{x^k} - \text{Im}\psi^\dagger \alpha_k \psi_{x^j}),$$

where the "momentum density" $P_j(\psi, \phi)$ is defined by

$$P_j(\psi, \phi) = \text{Im}\psi_{x^j}^\dagger \psi + \phi_t \phi_{x^j} \quad (j = 1, 2, 3). \tag{15}$$

Thus the *conserved momenta* are

$$\int P_j(\psi, \phi)\,dx = \text{const} \quad (j = 1, 2, 3). \tag{16}$$

Identities (11) and (14) are the basic expressions from which the other conservation laws follow. We leave out their derivations, all of which are elementary but tedious. The *angular momenta* are

$$\int [x^j e(\psi, \phi) - t P_j(\psi, \phi)]\,dx = \text{const} \quad (j = 1, 2, 3) \tag{17}$$

and

$$\int [x^k P_j(\psi,\phi) - x^j P_k(\psi,\phi) + \tfrac{1}{2}\,\mathrm{Im}\psi^\dagger \alpha_j \alpha_k \psi]\,dx = \text{const} \quad (k \neq j). \tag{18}$$

This completes the list of the ten identities following from the Lorentz invariance of the K–G–D equations [identities (12), (16), (17), (18)].

The following "dilational" and "inversional" identities are valid under the assumption that $M = m = 0$.

The *dilational identity* takes the form

$$\int [\mathrm{Im} r\psi_r^\dagger \psi + \phi_t(\phi + r\phi_r) - te(\psi,\phi)]\,dx = \text{const}, \tag{19}$$

where $r = |x|$. It follows essentially by multiplying (14) by x^j and summing on j.

The first *inversional identity* is

$$\int [(r^2 + t^2)\,e(\psi,\phi) - 2\mathrm{tr}\ \mathrm{Im}\psi_r^\dagger \psi - 2t\phi_t(\phi + r\phi_r) + \phi^2]\,dx = \text{const}. \tag{20}$$

This follows by multiplying (11) by $r^2 + t^2$, multiplying (14) by $-2tx^j$, and summing. It is this inversional identity that has proven to be quite useful in analogous situations.[6,7,12]

Finally, the three remaining *inversional identities* have the form

$$\int \Big[tx^j e(\psi,\phi) + \tfrac{1}{2}(t^2 + 2(x^j)^2 - r^2) P_j(\psi,\phi) + x^j \phi\phi_t + \sum_{m\neq j} x^j x^m P_m(\psi,\phi) + \tfrac{1}{2}\,\mathrm{Im}\sum_{m\neq j} x^m \psi^\dagger \alpha_m \alpha_j \psi \Big] dx = \text{const} \quad (j = 1,2,3). \tag{21}$$

This follows from repeated use of (11) and (14).

4. CONSERVATION LAWS FOR THE M-D EQUATIONS

The *charge* is

$$\int |\psi|^2\,dx = \text{const}, \tag{22}$$

exactly as before. The cone estimate is also the same. When $M = 0$, we have an additional quadratic invariant integral. The 4×4 symmetric matrix

$$C = \begin{bmatrix} 0 & I \\ I & 0 \end{bmatrix}$$

commutes with each α_k as given in Sec. 2. Thus we find

$$\int \psi^\dagger C \psi\,dx = \text{const}. \tag{23}$$

Subtracting (22) and (23), we find

$$\int \psi^\dagger (I - C)\psi\,dx = \int \{|\psi_1 - \psi_3|^2 + |\psi_2 - \psi_4|^2\}\,dx$$

to be conserved as well. In particular, if $\psi_1(x,t) = \psi_3(x,t)$ and $\psi_2(x,t) = \psi_4(x,t)$ at a specific time, then the same is valid at any other time so long as the solution exists, provided $M = 0$.

The energy identity, derived similarly as in Sec. 3, takes the form

$$\frac{\partial}{\partial t} e(\psi,A) = \mathrm{Im}\sum_k \frac{\partial}{\partial x^k}(\psi^\dagger \alpha_k \psi_t) + \sum_j \frac{\partial}{\partial x^j}\Big(A_t^0 A_{x^j}^0 - \sum_k A_t^k A_{x^j}^k\Big), \tag{24}$$

where the energy density $e(\psi,A)$ is defined by

$$e(\psi,A) = \mathrm{Im}\sum_k \psi^\dagger \alpha_k \psi_{x^k} - M\bar{\psi}\psi - gA^0|\psi|^2 - g\sum_k A^k \psi^\dagger \alpha_k \psi + \frac{1}{2}\Big[\Big(\frac{\partial A^0}{\partial t}\Big)^2 + |\nabla A^0|^2\Big] - \frac{1}{2}\sum_k \Big[\Big(\frac{\partial A^k}{\partial t}\Big)^2 + |\nabla A^k|^2\Big]. \tag{25}$$

Energy conservation is therefore expressed by

$$\int e(\psi,A)\,dx = \text{const}. \tag{26}$$

Note that the last two terms in (25) have opposite signs. This shows clearly that the Lorentz gauge is not well suited to obtaining *a priori* estimates.

The other fundamental identity is obtained essentially by multiplying (3b) by $A_{x^j}^0$, (3c) by $A_{x^j}^k$, and (3a) by $\psi_{x^j}^\dagger$ and then taking the imaginary part of this last expression. There results the identity

$$\frac{\partial}{\partial t} P_j(\psi,A) = \frac{1}{2}\frac{\partial}{\partial x^j}\Big[-\Big(\frac{\partial A^0}{\partial t}\Big)^2 + |\nabla A^0|^2 + \sum_k \Big(\Big(\frac{\partial A^k}{\partial t}\Big)^2 - |\nabla A^k|^2\Big)\Big] - \mathrm{Im}\sum_k \frac{\partial}{\partial x^k}(\psi^\dagger \alpha_k \psi_{x^j}) - \sum_l \frac{\partial}{\partial x^l}\Big(A_{x^j}^0 A_{x^l}^0 - \sum_k A_{x^j}^k A_{x^l}^k\Big), \tag{27}$$

where the "momentum density" $P_j(\psi,A)$ is given by

$$P_j(\psi,A) = \mathrm{Im}\psi_{x^j}^\dagger \psi - A_t^0 A_{x^j}^0 + \sum_k A_t^k A_{x^j}^k \quad (j = 1,2,3). \tag{28}$$

Hence the *conserved momenta* are given by

$$\int P_j(\psi,A)\,dx = \text{const} \quad (j = 1,2,3).$$

Using (24) and (27), we can next establish the conservation of *angular momenta*

$$\int [x^j e(\psi,A) - tP_j(\psi,A) + A^0 A_t^j - A^j A_t^0]\,dx = \text{const} \quad (j = 1,2,3) \tag{29}$$

and

$$\int [x^k P_j(\psi,A) - x^j P_k(\psi,A) - A^k A_t^j + A^j A_t^k + \tfrac{1}{2}\,\mathrm{Im}\psi^\dagger \alpha_j \alpha_k \psi]\,dx = \text{const} \quad (j \neq k). \tag{30}$$

The remaining identities are valid under the hypothesis that $M = 0$.

The *dilational identity* (with $r = |x|$)

$$\int [\mathrm{Im} r\psi_r^\dagger \psi - te(\psi,A) - A_t^0(A^0 + rA_r^0) + \sum_k A_t^k(A^k + rA_r^k)]\,dx = \text{const} \tag{31}$$

follows essentially from multiplying (27) by x^j and summing on j.

The first *inversional identity* has the form

$$\int \{(r^2 + t^2)\,e(\psi,A) + 2tr\,\mathrm{Im}\psi^\dagger \psi_r + 2t[A_t^0(A^0 + rA_r^0) - \sum_k A_t^k(A^k + rA_r^k)] + 2\sum_k x^k(A^0 A_t^k - A^k A_t^0) - 3(A^0)^2 + \sum_k (A^k)^2\}\,dx = \text{const}. \tag{32}$$

This results from the fairly obvious use of (24) and (27). The last two terms in (32) appear via use of the gauge condition.

The last three *inversional identities* are derived from repeated use of (27), with multipliers apparent from their explicit form:

$$\int\{-tx^j e(\psi,A)+\tfrac{1}{2}(t^2+2(x^j)^2-r^2)P_j(\psi,A)$$
$$+\sum_{m\neq j}x^m[x^j P_m(\psi,A)+\tfrac{1}{2}\operatorname{Im}\psi^\dagger\alpha_m\alpha_j\psi+A^mA_t^j-A^jA_t^m] \quad (33)$$
$$+x^j\Big(-A^0A_t^0+\sum_k A^kA_t^k\Big)-t(A^0A_t^j-A^jA_t^0)+A^0A^j\}dx$$
$$=\text{const}\quad (j=1,2,3).$$

We are not aware of any physical interpretation of (33).

Finally, we consider the gauge transformation (4) under the choice (6) of Φ. Thus our equations take the form (7) in the *Coulomb gauge*. We give below the analog of identities (26), (31), (32) in the "primed" variables. These new invariants can be obtained in two ways: (i) They can be derived directly from Eqs. (7) as above, or (ii) they can be rewritten from (26), (31), (32), using (4) and (6). In these calculations we use the relation

$$\int|\nabla A^{0\prime}|^2dx=g\int A^{0\prime}|\psi'|^2dx,$$

which follows from the second of Eqs. (7). We obtain in this manner the following:

Energy conservation:

$$\frac{1}{2}\int|\nabla A^{0\prime}|^2dx+\frac{1}{2}\sum_k\int\Big[\Big(\frac{\partial A^{k\prime}}{\partial t}\Big)^2+|\nabla A^{k\prime}|^2\Big]dx$$
$$+\int\Big(M\bar\psi'\psi'-\operatorname{Im}\sum_k\psi'^\dagger\alpha_k\psi'_{x^k}\Big)dx$$
$$+g\sum_k\int A^{k\prime}\psi'^\dagger\alpha_k\psi'dx\equiv\int e'(\psi',A')\,dx=\text{const.} \quad (26')$$

Dilational identity ($M=0$):

$$\int\Big[te'(\psi',A')+\operatorname{Im}r\psi_r'^\dagger\psi'+\sum_k A_t^{k\prime}(A^{k\prime}+rA_r^{k\prime})\Big]dx=\text{const.} \quad (31')$$

Inversional identity ($M=0$):

$$\int\Big[(r^2+t^2)e'(\psi',A')-2\sum_k x^kA^{0\prime}A_t^{k\prime}+2t\operatorname{Im}r\psi_r'^\dagger\psi'$$
$$+2t\sum_k A_t^{k\prime}(A^{k\prime}+rA_r^{k\prime})+3(A^{0\prime})^2-\sum_k(A^{k\prime})^2\Big]dx=\text{const.} \quad (32')$$

Notice that the first two terms in (26′) now have the *same* sign. It follows just as in Sec. 3 that for a solution $\psi', A^{0\prime}, \mathbf{A}'$ of the M–D equations (7) in the Coulomb gauge, we have the estimate

$$\|\nabla A^{0\prime}(t)\|_2^2+\sum_k\Big[\|A_t^{k\prime}(t)\|_2^2+\|\nabla A^{k\prime}(t)\|_2^2\Big]$$
$$\leq\text{const}\ (1+\|\nabla\psi'(t)\|_2).$$

[1] J. D. Bjorken and S. D. Drell, *Relativistic Quantum Mechanics* and *Relativistic Quantum Fields* (McGraw-Hill, New York, 1964 and 1965).
[2] J. M. Chadam, J. Funct. Anal. **13**, 173 (1973).
[3] J. M. Chadam and R. T. Glassey, Proc. Am. Math. Soc. **43**, 373 (1974).
[4] M. Reed, *Abstract Non-Linear Wave Equations*, Springer Lecture Notes in Math. No. 507 (Springer, New York, 1976).
[5] J. M. Chadam and R. T. Glassey, Arch. Ratl. Mech. Anal. **54**, 223 (1974).
[6] C. S. Morawetz, *Notes on Time Decay and Scattering for Some Hyperbolic Problems* (Soc. Ind. Appl. Math., Philadephia, 1975).
[7] W. A. Strauss, *Invariant Wave Equations*, Erice 1977, Lecture Notes in Physics No. 73 (Springer, New York, 1978), p. 197.
[8] L. Tartar, Univ. of Wisconsin, Math. Res. Ctr. Report 1584 (1976).
[9] J. Cooper and W. A. Strauss, Indiana U. Math. J. **25**, 671 (1976).
[10] R. T. Glassey, Comm. Math. Phys. **53**, 9 (1977).
[11] R. T. Glassey, J. Math. Phys. **18**, 1794 (1977).
[12] J. Ginibre and G. Velo, preprint.
[13] E. Noether, Nach. Ges. Göttingen Math. Phys. Kl., 235 (1918).
[14] E. Bessel-Hagen, Math. Ann. **84**, 258 (1921).
[15] J. A. McLennan, Nuovo Cimento **3**, 1360 (1956). These references consider only the free equations without interaction.
[16] A. O. Barut, *Electrodynamics and Classical Theory of Fields and Particles* (MacMillan, New York, 1964).
[17] I. E. Segal, Ann. Math. **78**, 339 (1963).
[18] L. Gross, Comm. Pure Appl. Math. **19**, 1 (1966).
[19] A. Friedman, *Partial Differential Equations* (Holt, New York, 1969).

The Cauchy problem in general relativity. III. On locally imbedding a family of null hypersurfaces

J. Smallwood

Faculty of Mathematical Studies, University of Southampton, S09 5NH, England
(Received 26 May 1978)

This paper is concerned with the problem of locally imbedding a null hypersurface in a Riemannian manifold. More precisely, on a one-parameter family of null hypersurfaces, rigged by an arbitrary null vector field, in a four-dimensional space–time manifold, a particular symmetric affine connection is used to derive the corresponding generalized Gauss–Codazzi equations. In addition, expressions are obtained for the projections of the Ricci tensor, which are relevant to the characteristic initial-value problem of general relativity.

INTRODUCTION

There is a generally satisfactory and well developed theory for the problem of locally imbedding hypersurfaces in a Riemannian space, when such hypersurfaces are themselves Riemannian spaces.[1] [Following Schouten, we use the term "Riemannian space" to denote a space endowed with a non-degenerate metric tensor, and if in addition, the metric is positive (negative) definite, the space is called ordinary.] However, the same is not true when the imbedded hypersurfaces are characteristic (null, isotropic) hypersurfaces.

From an intrinsic point of view, an n-dimensional null manifold is one which has defined on it a degenerate metric of rank $n-1$. Due to this degeneracy, the metric of such a manifold does not induce any affine structure on the manifold, in contradistinction to the case of Riemannian spaces. In order to overcome this problem, one is forced to choose, in a more or less arbitrary way, a linear connection which will depend, in general, upon quantities other than the metric alone. There are many such choices, and in the case where a null manifold is considered as a hypersurface in a Riemannian space, the quantities may be extrinsic to the hypersurface. The choice of connection is the crucial step in the subsequent imbedding theory.

The main objects of this paper are two-fold: After some preliminaries in Secs. 1 and 2, we shall study some aspects of the problem of locally imbedding a one-parameter family of rigged null hypersurfaces, ${}^{p}N_3$ in a four-dimensional space–time manifold V_4. We shall do this by, in Sec. 3, making a particular choice of linear connection for ${}^{p}N_3$ and then, in Sec. 4, calculating the generalized Gauss–Codazzi equations. These may be considered as necessary conditions for ${}^{p}N_3$, with our choice of connection, to be imbedded in V_4, or simply as expressions for various projections into ${}^{p}N_3$ of the Riemann tensor of V_4. Previous attempts at this problem have been unsatisfactory for various reasons. Either the choice of connection has been such that it depends on quantities extrinsic to the hypersurfaces,[2,3,4,5] in which case the geometrical interpretation of the resulting Gauss–Codazzi equations has been unclear, or else it has been such that the Gauss–Codazzi equations can only be derived for a restricted class of null hypersurfaces.[5,6] With our choice, neither of these problems arises.

In Sec. 5, we pursue our second objective, which is to complete the decomposition of the Riemann tensor of V_4 into components transverse and normal to ${}^{p}N_3$, and then apply the same procedure to the Ricci tensor of V_4. The resulting expressions are independent of the metric structure of V_4, and these results will thus enable us, in a subsequent paper, to study the local characteristic initial-value problem in general relativity from a geometrical viewpoint. (This paper is, in fact, one in an ongoing series relating to a larger program concerned with the Cauchy problem in general relativity—Papers I and II are Refs. 7 and 8, respectively.)

In what follows, we shall adopt, in the main, the conventions and notation of Schouten.[9] In particular, an X_n denotes an n-dimensional bare manifold; an L_n is an X_n with a linear connection; in an A_n the connection is symmetric, and in a V_n, the connection is, in addition, the metric connection. Greek indices run from 0 to n, and Latin indices from 1 to n, where, after Sec. 1, we take $n=4$. For our purposes, a V_4 will refer specifically to a Riemannian manifold, signature $(+---)$. Main differences and additions to Schouten's notation are as follows: partial derivatives are denoted by a comma; covariant derivatives in $L_n(A_n,V_n)$ by a semicolon and in $L_{n-1}(A_{n-1})$ by a colon; an N_3 denotes a characteristic hypersurface, signature $(0--)$ in V_4, and a one-parameter family of X_{n-1}'s(L_{n-1}'s,A_{n-1}'s) foliating an $X_n(L_n,A_n,V_n)$ will be denoted by ${}^{p}X_{n-1}({}^{p}L_{n-1},{}^{p}A_{n-1})$.

Finally, we re-emphasize that all our considerations are purely local; we make no attempt to study the global problems of the geometry of null hypersurfaces in V_4.

1. GEOMETRIC PRELIMINARIES

A. A rigged ${}^{p}X_{n-1}$ in an X_n

In this subsection we shall give a brief résumé of the results concerning a rigged ${}^{p}X_{n-1}$ in X_n. Further details may be found in Schouten.[10]

An X_{n-1}, with coordinates $'x^a$, in an X_n, with coordinates x^α may be given either by its null form $C^n(x^\alpha)=0$ (where n takes a single value), or by its parametric form $x^\alpha=B^\alpha('x^a)$. An X_{n-1} given by its null form induces a one-parameter family ${}^{p}X_{n-1}$ of X_{n-1}'s in X_n, given by

 0022-2488/79/030459-06$01.00

$$C^n(x^\alpha)=p. \tag{1.1}$$

The parametric equations of ${}^pX_{n-1}$ are

$$x^\alpha = B^\alpha({}'x^a_{(p)},p), \tag{1.2}$$

where ${}'x^a_{(p_0)}$ are the coordinates of a particular member ${}^{p_0}X_{n-1}$ of ${}^pX_{n-1}$. The one-parameter family given by (1.1) foliates X_n; that is, any point in X_n belongs to precisely one member of ${}^pX_{n-1}$.

The connecting quantities B^α_a, where

$$B^\alpha_a \overset{\text{def}}{=} B^\alpha{}_{,a}, \tag{1.3}$$

are used as projection operators. For example if w_α is a covariant vector of X_n, then $B^\alpha_a w_\alpha \overset{\text{def}}{=} {}'w_a$ are the components of the projection of w_α into X_{n-1}, in the coordinates of ${}^pX_{n-1}$. Similarly, if v^a is a contravariant vector field in ${}^pX_{n-1}$, then $B^\alpha_e {}'v^e \overset{\text{def}}{=} {}'v^\alpha$ are the components of v^a, considered as a vector of X_n, in the coordinates of X_n.

To make our notation absolutely consistent, we should refer to object fields Φ (indices suppressed) in ${}^pX_{n-1}$ as either ${}^p\Phi$, which would denote the one-parameter family of Φ's defined throughout X_n, or as ${}^{p_0}\Phi$, which would denote Φ defined only on one particular ${}^{p_0}X_{n-1}$. However, since it will always be made clear from the text which we mean, we shall omit this superscript, for ease of presentation.

In order to project contravariant vector fields of X_n into ${}^pX_{n-1}$, or to form a vector field in X_n corresponding to a covariant vector field in ${}^pX_{n-1}$, we must first rig ${}^pX_{n-1}$. This means that we must define a direction at each point of X_n which does not lie in the member of ${}^pX_{n-1}$ through that point. In practice, this is done by defining a contravariant vector field $\overline{C}^\alpha_n$ in X_n, which nowhere lies in ${}^pX_{n-1}$. Such a vector field is said to rig, or transvect ${}^pX_{n-1}$. The covariant normal to ${}^pX_{n-1}$ is given by $C^n_\alpha \overset{\text{def}}{=} C^n{}_{,\alpha}$. In general, we have $\overline{C}^\alpha_n C^n_\alpha = \lambda(x^\alpha)\neq 0$. If we define $C^\alpha_n \overset{\text{def}}{=} \lambda^{-1}\overline{C}^\alpha_n$, then C^α_n is also a rigging field, in the direction of $\overline{C}^\alpha_n$, which satisfies the so-called first normalizing condition,

$$C^\alpha_n C^n_\alpha = 1. \tag{1.4}$$

By virtue of the rigging field C^α_n, we may form the connecting quantities (projection operators) B^a_α. The projection operators satisfy

$$B^\alpha_a C^n_\alpha = B^a_\alpha C^\alpha_n = 0, \quad B^a_\alpha B^\alpha_b = \delta^a_b. \tag{1.5}$$

We define further the projection operators

$$B^\alpha_\beta \overset{\text{def}}{=} B^e_\beta B^\alpha_e, \quad C^\alpha_\beta \overset{\text{def}}{=} C^n_\beta C^\alpha_n. \tag{1.6}$$

Under a transformation $C^\alpha_n \to \sigma C^\alpha_n$, $C^n_\alpha \to \sigma^{-1} C^n_\alpha$ $[\sigma=\sigma(x^\alpha)]$, which preserves (1.4), the projection operators B^α_a, B^a_α, B^α_β, and C^α_β are invariant.

We have the fundamental relationship

$$B^\alpha_\beta + C^\alpha_\beta = \delta^\alpha_\beta, \tag{1.7}$$

and using (1.7), we may write any contravariant vector v^α of X_n in the form

$$v^\alpha = B^\alpha_\beta v^\beta + C^\alpha_\beta v^\beta,$$

where ${}'v^\alpha \overset{\text{def}}{=} B^\alpha_\beta v^\beta$ is the ${}^pX_{n-1}$, or transverse part of v^α, and ${}''v^\alpha$ is the component of v^α in the direction of the rigging, or normal part. Similar remarks apply to covariant vectors and higher order tensors in X_n. For example, any symmetric contravariant tensor $T^{\alpha\beta}$ of order 2 may be written in the form

$$T^{\alpha\beta} = B^{\alpha\beta}_{\gamma\delta} T^{\gamma\delta} + 2B^{(\alpha}_{\gamma} C^{\beta)}_{\delta} T^{\gamma\delta} + C^{\alpha\beta}_{\gamma\delta} T^{\gamma\delta}, \tag{1.8}$$

where $B^{\alpha\beta}_{\gamma\delta}$ is shorthand for $B^\alpha_\gamma B^\beta_\delta$, etc. Then ${}'T^{\alpha\beta} \overset{\text{def}}{=} B^{\alpha\beta}_{\gamma\delta} T^{\gamma\delta}$ is the transverse–transverse (or t–t) part of $T^{\alpha\beta}$, $B^{(\alpha}_{\gamma} C^{\beta)}_{\delta} T^{\gamma\delta}$ is the transverse–normal (or t–n) part, and $C^{\alpha\beta}_{\gamma\delta} T^{\gamma\delta}$ is the normal–normal (or n–n) part. Any tensor written in the form (1.8) is said to be decomposed into its transverse and normal parts, or projections, relative to ${}^pX_{n-1}$.

B. Lie derivatives in a rigged ${}^pX_{n-1}$[11]

The Lie derivative of a tensor field defined on ${}^pX_{n-1}$ w.r.t. a vector field v^α in X_n is not, *a priori*, well defined. However, by means of the projection operators B^α_a and B^a_α, we may form the tensor field in X_n corresponding to the given tensor field in ${}^pX_{n-1}$. We may take the Lie derivative of the former in the usual way. We then define the Lie derivative of a tensor field in ${}^pX_{n-1}$ as the components in ${}^pX_{n-1}$ of the Lie derivative of the corresponding tensor field in X_n. That is, for any tensor field ${}'T^{a\cdots}_{\cdots b}$ in ${}^pX_{n-1}$, and any contravariant vector field v^α in X_n, we define $\pounds'_v T^{a\cdots}_{\cdots b}$ by

$$\pounds'_v T^{a\cdots}_{\cdots b} \overset{\text{def}}{=} B^{a\cdots\beta}_{\alpha\cdots b} \pounds_v B^{\alpha\cdots f}_{e\cdots\beta} {}'T^{e\cdots}_{\cdots f}. \tag{1.9}$$

Should v^α itself lie in ${}^pX_{n-1}$, i.e., if $v^\alpha = B^\alpha_\beta v^\beta$, then the Lie derivative of ${}'T^{a\cdots}_{\cdots b}$ will be the same whether computed by the definition (1.9) or by taking the Lie derivative in the usual way, as is clearly possible in this case.

C. Covariant derivatives in a rigged ${}^pL_{n-1}$ in an L_n[12]

Suppose that we have an ${}^pL_{n-1}$ in L_n with connections ${}'\Gamma^a_{bc}$ and $\Gamma^\alpha_{\beta\gamma}$, respectively. Let $\mathscr{C}: x^\alpha = x^\alpha(s)$ be some arbitrary curve, parametrized by s, within ${}^{p_0}L_{n-1}$, for some $p=p_0$. For any tensor field $T^{\alpha\cdots}_{\cdots\beta}$ of L_n, the absolute derivative along $\mathscr{C}$ is given, using (1.3), by

$$\frac{DT^{\alpha\cdots}_{\cdots\beta}}{Ds} \overset{\text{def}}{=} T^{\alpha\cdots}_{\cdots\beta;\gamma}\frac{dx^\gamma}{ds} = T^{\alpha\cdots}_{\cdots\beta;\gamma} B^\gamma_e \frac{d'x^e_{(p_0)}}{ds}. \tag{1.10}$$

We now define the covariant derivative of $T^{\alpha\cdots}_{\cdots\beta}$ w.r.t. ${}'\Gamma^a_{bc}$ by

$$\frac{DT^{\alpha\cdots}_{\cdots\beta}}{Ds} \overset{\text{def}}{=} T^{\alpha\cdots}_{\cdots\beta:e}\frac{d'x^e_{(p_0)}}{ds}. \tag{1.11}$$

Since $\mathscr{C}$ and p_0 are arbitrary, (1.10) and (1.11) give

$$T^{\alpha\ldots}_{\ldots\beta:c} \overset{\text{def}}{=} T^{\alpha\ldots}_{\ldots\beta;\gamma} B^{\gamma}_{c}. \tag{1.12}$$

Suppose that we have some arbitrary vector field $'v^{\alpha} = B^{\alpha}_{e}{}' v^{e}$ in L_n which lies in ${}^{p}L_{n-1}$. From (1.12) we obtain

$$'v^{e} B^{\alpha}_{e:c} + B^{\alpha}_{e} v^{e}{}_{,c} + B^{\alpha}_{e}{}' \Gamma^{e}_{cf}{}' v^{f} \overset{\text{def}}{=} (B^{\alpha}_{e}{}' v^{e})_{,c} + \Gamma^{\alpha}_{\gamma\beta} B^{\beta}_{e}{}' v^{e} B^{\gamma}_{c}.$$

Since $'v^{a}$ is arbitrary, the above equation defines $B^{\alpha}_{b:c}$ by

$$B^{\alpha}_{b:c} \overset{\text{def}}{=} B^{\alpha}_{b,c} - B^{\alpha}_{e}{}' \Gamma^{e}_{cb} + B^{\beta\gamma} \Gamma^{\alpha}_{\gamma\beta}. \tag{1.13}$$

We emphasize that the above results are simply definitions, and entirely independent of the specific choice of connection in L_n or ${}^{p}L_{n-1}$.

Let us now suppose that $'\Gamma^{a}_{bc}$ and $\Gamma^{\alpha}_{\beta\gamma}$ are both symmetric, that is, ${}^{p}L_{n-1}$ and L_n are an ${}^{p}A_{n-1}$ and A_n, respectively. Then, remembering (1.3), we see that

$$B^{\alpha}_{b:c} = B^{\alpha}_{(b:c)}. \tag{1.14}$$

Covariantly differentiating (1.12), with $'v^{\alpha}$ substituted for $T^{\alpha\ldots}_{\ldots\beta}$ gives

$$'v^{e} B^{\alpha}_{e:cd} + 2B^{\alpha}_{e:(c}{}' v^{e}_{\ :d)} + B^{\alpha}_{e}{}' v^{e}_{\ :cd} \overset{\text{def}}{=} \beta^{\delta\gamma}_{dc}{}' v^{\alpha}_{\ ;\gamma\delta} + {}' v^{\alpha}_{\ ;\gamma} B^{\gamma}_{c:d}.$$

Alternating over c and d, and remembering (1.14) we obtain

$$'v^{e} B^{\alpha}_{e:[cd]} + \tfrac{1}{2} B^{\alpha}_{e}{}' R_{dcb}{}^{e\prime} v^{b} \overset{\text{def}}{=} B^{\delta\gamma\beta}_{dce} R_{\delta\gamma\beta}{}^{\alpha\prime} v^{e}.$$

where $R_{\delta\gamma\beta}{}^{\alpha}$ is the Riemann tensor of A_n and $'R_{dcb}{}^{a}$ that of ${}^{p}A_{n-1}$. Since $'v^{a}$ is arbitrary, we may use the last equation as the definition of $B^{\alpha}_{b:[cd]}$, giving

$$B^{\alpha}_{b:[cd]} \overset{\text{def}}{=} \tfrac{1}{2} B^{\delta\gamma\beta}_{dcb} R_{\delta\gamma\beta}{}^{\alpha} - \tfrac{1}{2} B^{\alpha}_{a}{}' R_{dcb}{}^{a}. \tag{1.15}$$

2. INDUCED METRIC STRUCTURE IN ${}^{p}N_3$

In a V_4, let an ${}^{p}X_3$ be given by the equation $\phi(x^{\alpha}) = p$, where ϕ satisfies $g^{\alpha\beta}\phi_{,\alpha}\phi_{,\beta} = 0$, $g_{\alpha\beta}$ being the metric tensor of V_4. Then ${}^{p}X_3$ is a one-parameter family of null hypersurfaces foliating V_4. These hypersurfaces are not in themselves Riemannian spaces, since, as we shall demonstrate shortly, each member of ${}^{p}X_3$ has induced upon it a degenerate metric tensor of rank 2, signature $(0--)$, which makes ${}^{p}X_3$ an ${}^{p}N_3$.

The covariant normal to ${}^{p}N_3$ is $\phi_{,\alpha}$ and $k^{\alpha} \propto g^{\alpha\beta}\phi_{,\beta}$ is any tangent vector field to the congruence of null geodesics ruling ${}^{p}N_3$, Using k_{α} as a basis vector in V_4, we may complete the basis by introducing the real null vector n_{α}, and the complex null vector m_{α}, defined by

$$g_{\alpha\beta} = 2k_{(\alpha}n_{\beta)} - 2m_{(\alpha}\bar{m}_{\beta)}, \tag{2.1}$$

$$k_{\alpha}n^{\alpha} = -m_{\alpha}\bar{m}^{\alpha} = 1, \tag{2.2}$$

all other scalar products zero.

The null tetrad $(k_{\alpha}, n_{\alpha}, m_{\alpha}, \bar{m}_{\alpha})$ is not uniquely defined by (2.1) and (2.2). The most general transformations of the tetrad preserving the direction of k^{α} are the so-called null rotations about k^{α}, given by

$$k^{\alpha} \to A k^{\alpha}, \tag{2.3a}$$

$$n^{\alpha} \to A^{-1} n^{\alpha} - D m^{\alpha} - \bar{D}\bar{m}^{\alpha} + A D \bar{D} k^{\alpha}, \tag{2.3b}$$

$$m^{\alpha} \to e^{iE}(m^{\alpha} - A\bar{D}k^{\alpha}), \tag{2.3c}$$

where A and E are real, $A > 0$, and D a complex function of x^{α}.

The necessary and sufficient condition that any null vector q^{α}, say, lies in ${}^{p}N_3$, is that $q^{\alpha} \propto k^{\alpha} \Longleftrightarrow q^{\alpha}k_{\alpha} = 0$. Hence it is clear from (2.2) that n^{α} transvects ${}^{p}N_3$. Furthermore, it is easy to show that (2.3) allows n^{α} to be transformed into any null vector field with the same time sense as n^{α}, and not parallel to k^{α}. Hence n^{α} is the most general null vector field transvecting ${}^{p}N_3$. Since $k_{\alpha} \propto \phi_{,\alpha}$ and n^{α} and k_{α} satisfy the first normalizing condition (1.4), we may, as described in Sec. 1A, use these vectors to form the projection operators $B^{\alpha}_{a}, B^{a}_{\alpha}, B^{\alpha}_{\beta}$, and C^{α}_{β}. Equations (1.5), (1.6), and (1.7) yield, in particular,

$$B^{\alpha}_{b}B^{a}_{\alpha} = \delta^{a}_{b}, \quad B^{\alpha}_{a}k_{\alpha} = B^{a}_{\alpha}n^{\alpha} = 0, \quad B^{\alpha}_{\beta} = \delta^{\alpha}_{\beta} - n^{\alpha}k_{\beta},$$

relationships which will be used frequently and without reference in the remainder of this paper.

If we project the null tetrad spanning V_4 in its contravariant and covariant forms into ${}^{p}N_3$, we see using (2.2) that the contravariant and covariant triads

$$T^{\alpha}_{(a)} \overset{\text{def}}{=} (k^{\alpha}, m^{\alpha}, \bar{m}^{\alpha}), \quad T^{(a)}_{\alpha} \overset{\text{def}}{=} (n_{\alpha}, m_{\alpha}, \bar{m}_{\alpha}),$$

respectively, span ${}^{p}N_3$. These triads in hypersurface coordinates are given by

$$T^{a}_{(a)} \overset{\text{def}}{=} B^{a}_{\alpha} T^{\alpha}_{(a)} \overset{\text{def}}{=} (k^{a}, m^{a}, \bar{m}^{a}),$$

$$T^{(a)}_{a} \overset{\text{def}}{=} B^{\alpha}_{a} T^{(a)}_{\alpha} \overset{\text{def}}{=} (n_{a}, m_{a}, \bar{m}_{a}),$$

respectively. The scalar products between the triad members are from (2.2) clearly

$$k^{e}n_{e} = -m^{e}\bar{m}_{e} = 1, \tag{2.4}$$

all others zero.

The covariant metric tensor of ${}^{p}N_3$, induced by virtue of its imbedding in V_4, is given by[13]

$$'g_{ab} \equiv B^{\alpha\beta}_{ab} g_{\alpha\beta}; \tag{2.5}$$

$'g_{ab}$ possesses a single eigendirection of eigenvalue zero, defined by k^{a}, since from (2.1), (2.4), and (2.5) we have

$$'g_{ab} = -2m_{(a}\bar{m}_{b)} \Rightarrow {}'g_{ae}k^{e} = 0. \tag{2.6}$$

Equation (2.6) implies that $'g_{ab}$ is degenerate, with determinant of rank 2. Due to this degeneracy, $'g_{ab}$ cannot be inverted to give a contravariant metric $'^{*}g^{ab}$ satisfying $'^{*}g^{ae\prime}g_{be} = \delta^{a}_{b}$. Instead we introduce a substitute contravariant metric given by

$$'g^{ab} \equiv B^{ab}_{\alpha\beta} g^{\alpha\beta} = -2m^{(a}\bar{m}^{b)}. \tag{2.7}$$

Hence $'g^{ab}$ is simply the projection of $g^{\alpha\beta}$ into ${}^{p}N_3$. $'g^{ab}$ and $'g_{ab}$ satisfy

$$'g_{ae}{}'g_{bf}{}'g^{ef} = {}'g_{ab}, \tag{2.8a}$$

$$'g^{eb}n_{e} = 0, \tag{2.8b}$$

$$'g^{ae\prime}g_{be} = \delta^{a}_{b} - k^{a}n_{b}. \tag{2.8c}$$

In fact, from an intrinsic point of view, given $'g_{ab}$ and n_a on ${}^pN_3$, (2.8) defines $'g^{ab}$ uniquely.

The tetrad transformations (2.3) induce the following transformations in the triad vectors:

$$k^a \to Ak^a, \tag{2.9a}$$

$$m^a \to e^{iE}(m^a - A\bar{D}k^a), \tag{2.9b}$$

$$n_a \to A^{-1}n_a - Dm_a - \bar{D}\bar{m}_a, \tag{2.9c}$$

$$m_a \to e^{iE}m_a, \tag{2.9d}$$

where A, D, and C are now functions of $'x^a_{(p)}$ and p. The covariant metric $'g_{ab}$ of ${}^pN_3$ is invariant under (2.9); $'g^{ab}$ is invariant only under the subgroup of (2.9) given by $D=0$.

3. AN INTRINSIC AFFINE CONNECTION FOR ${}^pN_3$

Ideally, we should like to construct on ${}^pN_3$ a symmetric, metric connection dependent only on quantities intrinsic to ${}^pN_3$. Unfortunately, this is, in general, impossible. In particular, it can be shown that except for a very narrow class of null manifolds, no linear connection which is both metric and symmetric can be constructed.[14] A number of authors have considered the problem of constructing a connection on an N_3, considering variously either the symmetric[2,3,4,5,15] or metric[6] property to be more fundamental. Of these connections, only two (Refs. 6 and 15) are intrinsic to N_3. The point of view we shall adopt is that it is necessary for the connection to be intrinsic if it is to be geometrically meaningful. To this end, we shall choose the intrinsic, symmetric, nonmetric connection first introduced by Dautcourt[15]; our motivation being simply that it turns out, with this choice, that we can derive a set of generalized Gauss–Codazzi equations whose geometrical interpretation is clear, and which also have useful physical applications in general relativity.

Dautcourt's connection is given by[15]

$$'\Gamma^a_{bc} \overset{\text{def}}{=} \tfrac{1}{2}{}'g^{ae}('g_{be,c} + {}'g_{ce,b} - {}'g_{bc,e}) + k^a n_{(b,c)}. \tag{3.1}$$

From its definition, it is clear that $'\Gamma^a_{bc}$ is not invariant under (2.9), although it is invariant under the subgroup $A = 1$, $D = 0$. This is unfortunate and differs from the case of the metric connection in V_4 which is independent of the related transformations (2.3). However, it is the case, from an intrinsic point of view, given $'g_{ab}$ and n_a, that $'\Gamma^a_{bc}$ is uniquely determined [by (2.4), (2.6), and (2.8)]. This is as much as we can really hope for; in a V_3, the metric connection is determined by the six independent functions contained in the metric of V_3. In an N_3, the metric contains only three independent functions, and thus one might expect that more information would be required to determine a connection. This is indeed the case; giving n_a is equivalent to specifying another three independent functions.

By projecting $'\Gamma^a_{bc}$ into V_4, one can readily show, using (2.5) and (2.7), that

$$'\Gamma^a_{bc} = B^{a\beta\gamma}_{\alpha bc}\Gamma^\alpha_{\beta\gamma} + B^a_\alpha B^\alpha_{b,c} + B^{a\beta\gamma}_{\alpha bc}k^\alpha n_{(\beta;\gamma)}, \tag{3.2}$$

where $\Gamma^\alpha_{\beta\gamma}$ is the metric connection of V_4. From (3.2), we see that the connection $'\Gamma^a_{bc}$ of ${}^pN_3$ in V_4 is determined by the metric of V_4, and the rigging field n^α. For the rest of the paper, we shall assume that a rigged ${}^pN_3$ in V_4 is in addition an ${}^pA_3$ with connection given by (3.2).

At this stage it is convenient to define certain tensor fields induced in ${}^pN_3$ by virtue of its imbedding in V_4 and its rigging. We define (cf. Schouten[2]):

$$l_{cb} \overset{\text{def}}{=} -B^{\gamma\beta}_{cb}n_{\beta;\gamma}, \tag{3.3a}$$

$$l_c{}^a \overset{\text{def}}{=} -B^{\gamma a}_{c\alpha}n^\alpha{}_{;\gamma}, \tag{3.3b}$$

$$h_c{}^a \overset{\text{def}}{=} -B^{\gamma a}_{c\alpha}k^\alpha{}_{;\gamma}, \tag{3.3c}$$

$$h_{cb} = h_{(cb)} \overset{\text{def}}{=} -B^{\gamma\beta}_{cb}k_{\beta;\gamma}. \tag{3.3d}$$

Substituting (3.3a) in (3.2) gives

$$'\Gamma^a_{bc} = B^{a\beta\gamma}_{\alpha bc}\Gamma^\alpha_{\beta\gamma} + B^a_\alpha B^\alpha_{b,c} - k^a l_{(bc)}. \tag{3.4}$$

The quantities defined by (3.3) are apparently implicitly dependent on $\Gamma^\alpha_{\beta\gamma}$. This is undesirable, since for example, (3.4) does not give $'\Gamma^a_{bc}$ explicitly in terms of $\Gamma^\alpha_{\beta\gamma}$. However, remembering $k_\alpha \propto \phi_{,\alpha}$, and writing

$$k_\alpha = e^\rho\phi_{,\alpha}, \tag{3.5}$$

where ρ is some arbitrary scalar function, we may show in a straightforward manner using (1.9), (2.1), (2.2), (2.5), (2.7), and (3.5), that the definitions (3.3) lead to

$$l_{cb} = -\tfrac{1}{2}\pounds'_n g_{cb} + \rho_{,(c}n_{b)} + n_{[c,b]}, \tag{3.6a}$$

$$l_c{}^a = {}'g^{ae}l_{ce}, \tag{3.6b}$$

$$h_c{}^a = -\tfrac{1}{2}{}'g^{ae}\pounds'_k g_{ce} - k^e l_{ce}k^a, \tag{3.6c}$$

$$h_{cb} = {}'g_{eb}h_c{}^e = -\tfrac{1}{2}\pounds'_k g_{cb}. \tag{3.6d}$$

We see that (3.6) are independent of $\Gamma^\alpha_{\beta\gamma}$ and $'\Gamma^a_{bc}$. In particular, substituting (3.6a) in (3.4) gives us an expression for $'\Gamma^a_{bc}$ explicitly in terms of $\Gamma^\alpha_{\beta\gamma}$. (This result was first due to Dautcourt[16].)

In contrast to the case of an ordinary V_3 in V_4[1], we see that h_{cb} is now intrinsic to ${}^pN_3$, and l_{cb} (equivalently $\pounds'_n g_{cb}$) now defines the extrinsic curvature of ${}^pN_3$. We shall refer to l_{cb} therefore, and not h_{cb}, as the second fundamental form of ${}^pN_3$.

The Appendix contains some relations which are a direct consequence of the definition of $'\Gamma^a_{bc}$ in (3.1), and the definitions (3.3), and which are used in some of the ensuing calculations in Secs. 4 and 5.

4. THE GAUSS–CODAZZI EQUATIONS

Suppose we are given a rigged ${}^pN_3$ in V_4, as described in Sec. 3. The imbedding of ${}^pN_3$ in V_4 is determined by the functions B^α, the covariant normal $\phi_{,\alpha}$ (or equivalently k_α) and the rigging field n^α. Using (1.13) and (3.3d), we may rewrite (3.4) as

$$B^\alpha_{b:c} \overset{\text{def}}{=} B^\alpha_{b,c} + B^{\alpha'}_e\Gamma^e_{bc} - B^{\beta\gamma}_{bc}\Gamma^\alpha_{\beta\gamma} = n^\alpha h_{bc} + B^\alpha_e k^e l_{(bc)}. \tag{4.1a}$$

Similarly, we may rewrite (3.3b) and (3.3d), using (1.12), as

$$n^{\alpha}{}_{:c} \overset{\text{def}}{\equiv} n^{\alpha}{}_{;\gamma} B^{\gamma}_{c} = -B^{\alpha}_{e} l_{c}{}^{e} - n^{\alpha} l_{ce} k^{e}, \tag{4.1b}$$

$$B^{\beta}_{b} k_{\beta:c} \overset{\text{def}}{=} B^{\beta\gamma}_{bc} k_{\beta;\gamma} = -h_{cb}. \tag{4.1c}$$

We may formally regard Eq. (4.1) as differential equations for the quantities B^{α}, n^{α}, and k_{α} (cf. Schouten for the imbedding of an A_{n-1} in E_n[16] and a V_{n-1} in R_n[17]), however, the significance of (4.1) in the present case is that the integrability conditions for these equations lead to relationships between quantities of ${}^{p}N_3$ and V_4.

The integrability conditions for (4.1a) are, using (1.15),

$$B^{\alpha}_{b:[cd]} \overset{\text{def}}{=} \tfrac{1}{2} B^{\delta\gamma\beta}_{dcb} K_{\delta\gamma\beta}{}^{\alpha} - \tfrac{1}{2} B^{\alpha\prime}_{e} R_{dcb}{}^{e}$$

$$= (n^{\alpha} h_{b[c})_{:d]} + (B^{\alpha}_{e} k^{e} l_{(b[c)})_{:d]},$$

where $K_{\delta\gamma\beta}{}^{\alpha}$ is the Riemann tensor of V_4, and all symmetrizations are to be performed before antisymmetrizations. Using (4.1) in the above equation, we obtain

$$2B^{\alpha}_{b:[cd]} \overset{\text{def}}{=} B^{\delta\gamma\beta}_{dcb} K_{\delta\gamma\beta}{}^{\alpha} - B^{\alpha\prime}_{e} R_{dcb}{}^{e}$$

$$= 2n^{\alpha}(h_{b[c:d]} - h_{b[c} l_{d]e} k^{e})$$

$$+ 2B^{\alpha}_{e}(k^{e} l_{(b[c):d]} - h_{b[c} l_{d]}{}^{e} - l_{(b[c)} h_{d]}{}^{e}), \tag{4.2}$$

where, from (1.12) and (3.3c) we have used the fact that

$$k^{\alpha}{}_{:d} \overset{\text{def}}{=} B^{\delta}_{d} k^{\alpha}{}_{;\gamma} = -B^{\alpha}_{e} h_{d}{}^{e}.$$

Transvecting (4.2) with B^{a}_{α} gives

$$2B^{a}_{\alpha} B^{\alpha}_{b:[cd]} \overset{\text{def}}{=} B^{\delta\gamma\beta a}_{dcb\alpha} K_{\delta\gamma\beta}{}^{\alpha} - {}'R_{dcb}{}^{a}$$

$$= 2k^{a} l_{(b[c):d]} - 2h_{b[c} l_{d]}{}^{a} - 2l_{(b[c)} h_{d]}{}^{a}. \tag{4.3}$$

Ostensibly, we could obtain another set of integrability conditions for (4.1a) by transvecting (4.2) with k_{α}. However, we may show that

$$k_{\alpha} B^{\alpha}_{b:[cd]} = {}'g_{eb} k^{f} B^{e}_{\alpha} B^{\alpha}_{f:[cd]} - {}'g_{eb} k^{e}{}_{:[cd]}$$

and hence (4.3) are the only independent equations in (4.2). We may also obtain the integrability conditions for (4.1b) and (4.1c) but since we may show that

$$B^{a}_{\alpha} n^{\alpha}{}_{:[cd]} = -{}'g^{ae} n_{f} B^{f}_{\alpha} B^{\alpha}_{e:[cd]} - {}'g^{ae} n_{e:[cd]},$$

$$k_{\alpha} n^{\alpha}{}_{:[cd]} = -n_{e} k^{e}{}_{:[cd]} - n_{e} B^{e}_{\beta} B^{\beta}_{f:[cd]} k^{f},$$

$$(B^{\beta}_{b} k_{\beta:[c})_{d]} = -k_{\beta} B^{\beta}_{b:[cd]} - h_{b[c} l_{d]e} k^{e},$$

we see that the integrability conditions for (4.1b) and (4.1c) may be derived from (4.3). That is, (4.3) is the only independent set of integrability conditions for (4.1).

Equation (4.3) gives the totally transverse projection of $K_{\delta\gamma\beta}{}^{\alpha}$ relative to ${}^{p}N_3$. Let us now consider this tensor in its completely covariant form. There are precisely three independent projections of this tensor, namely the totally transverse (t–t–t–t) projection, $B^{\delta\gamma\beta\alpha}_{dcba} K_{\delta\gamma\beta\alpha}$, the three times transverse, once normal (t–t–t–n) projection, $B^{\delta\gamma\beta}_{dcb} K_{\delta\gamma\beta\alpha} n^{\alpha}$, and the twice transverse, twice normal (n–t–t–n) projection, $B^{\gamma\beta}_{cb} K_{\delta\gamma\beta\alpha} n^{\delta} n^{\alpha}$. It is straightforward to show, using (2.5) and (1.8), that

$$B^{\delta\gamma\beta\alpha}_{dcba} K_{\delta\gamma\beta\alpha} = (\delta^{f}_{b}\, {}'g_{ae} - n_{a}\, {}'g_{be} k^{f}) B^{\delta\gamma\beta e}_{dcf\alpha} K_{\delta\gamma\beta}{}^{\alpha},$$

$$B^{\delta\gamma\beta}_{dcb} K_{\delta\gamma\beta\alpha} n^{\alpha} = n_{e} B^{\delta\gamma\beta e}_{dcb\alpha} K_{\delta\gamma\beta}{}^{\alpha}.$$

Hence both the (t–t–t–t) and (t–t–t–n) projections are determined from (4.3), and using this latter equation we obtain, using (2.6), (2.8c), (3.6), (A1), (A2), and (A3),

$$B^{\delta\gamma\beta\alpha}_{dcba} K_{\delta\gamma\beta\alpha} = {}'g_{ae}\, {}'R_{dcb}{}^{e} - 2h_{b[c} l_{d]a} - 2l_{(b[c)} h_{d]a} + 2h_{b[c:d]} n_{a}, \tag{4.4a}$$

$$B^{\delta\gamma\beta}_{dcb} K_{\delta\gamma\beta\alpha} n^{\alpha} = 2l_{[c|b|:d]} + 2l_{(b[c)} l_{d]e} k^{e}. \tag{4.4b}$$

In analogy with the case of a V_{n-1} in V_n[1], we shall refer to (4.4a) as the generalized equation of Gauss, and (4.4b) as the generalized equation of Codazzi.

5. DECOMPOSITION OF THE RIEMANN AND RICCI TENSORS OF V_4

In the last section, we saw that two of the three independent projections of $K_{\delta\gamma\beta\alpha}$ are given by the generalized Gauss–Codazzi equations (4.4). With (3.6), (4.4) gives these projections in terms of quantities that are independent of the metric structure of V_4. An expression with the same property for the remaining independent (n–t–t–n) projection is obtained by taking the Lie derivative w.r.t. n^{α} of (3.3a). This gives immediately

$$\pounds_{n} l_{cb} = -B^{\gamma\beta}_{cb} \pounds_{n} n_{\beta;\gamma} - n_{\beta;\gamma} B^{\alpha\delta}_{cb} \pounds_{n} B^{\gamma\beta}_{\alpha\delta}.$$

Expanding the R.H.S. of the above, using (1.9), (1.12), (3.3), (3.5), (3.6), and (4.1a) yields, eventually,

$$B^{\gamma\beta}_{cb} K_{\delta\gamma\beta\alpha} n^{\delta} n^{\alpha} = \pounds_{n} l_{(cb)} + l_{ce} l_{b}{}^{e} + (\pounds_{n} n_{(b)})_{:c)} - k^{e} \pounds_{n} n_{e} l_{(cb)} + \rho_{,(c} \pounds_{n} n_{b)}. \tag{5.1}$$

From (1.8), (2.2), and (2.7) we may write

$$K_{\gamma\beta} \overset{\text{def}}{=} g^{\alpha\delta} K_{\delta\gamma\beta\alpha} = {}'g^{ef} B^{\alpha\delta}_{fe} K_{\delta\gamma\beta\alpha} + 2k^{e} B^{\delta}_{e} K_{\delta(\gamma\beta)\alpha} n^{\alpha}.$$

The three independent projections of $K_{\gamma\beta}$ are then given by

$$B^{\gamma\beta}_{cb} K_{\gamma\beta} = {}'g^{ef} B^{\delta\gamma\beta\alpha}_{fcbe} K_{\delta\gamma\beta\alpha} + 2k^{e} B^{\delta\,\gamma\beta}_{e(cb)} K_{\delta\gamma\beta\alpha} n^{\alpha},$$

$$B^{\gamma}_{c} K_{\gamma\beta} n^{\beta} = -{}'g^{ef} B^{\delta\gamma\beta}_{fce} K_{\delta\gamma\beta\alpha} n^{\alpha} - k^{e} B^{\gamma\beta}_{ce} K_{\delta\gamma\beta\alpha} n^{\delta} n^{\alpha},$$

$$K_{\gamma\beta} n^{\gamma} n^{\beta} = {}'g^{ef} B^{\gamma\beta}_{fe} K_{\delta\gamma\beta\alpha} n^{\delta} n^{\alpha},$$

the (t–t), (t–n), and (n–n) components, respectively. Using (2.8c), (3.6), (4.4), (5.1), (A1), (A2), and (A3) in the above,

we obtain

$$B^{\gamma\beta}_{cb}K_{\gamma\beta} = {}'R_{(cb)} + 2\pounds_k \tilde{l}_{cb} - 2(k^e \tilde{l}_{e(c)}{}_{:b)} + 4'g^{ef}h_{e(c}\tilde{l}_{b)f} - {}'g^{ef}h_{ef}\tilde{l}_{cb}$$
$$- g^{ef}\tilde{l}_{ef}h_{cb} - 2k^e \tilde{l}_{e(c}\pounds_k n_{b)} - 2\tilde{l}_{ce}k^e \tilde{l}_{bf}k^f + 2\tilde{l}_{ef}k^e k^f \tilde{l}_{cb}$$
$$+ 2'g^{ef}h_{f(c}n_{b):e} + \tfrac{1}{2}(\pounds_k n_{(c)}{}_{:b)} - \tfrac{1}{4}\pounds_k n_b \pounds_k n_c, \qquad (5.2a)$$

where ${}'R_{cb} \overset{\text{def}}{=} {}'R_{ecb}{}^{e}$, $\tilde{l}_{cb} \overset{\text{def}}{=} l_{(cb)}$;

$$B^{\gamma}_{c}K_{\gamma\beta}n^{\beta} = -k^e \pounds_n l_{(ce)} - 2l_{[c}{}^{e}{}_{:e]} + l_e{}^e l_{cf}k^f - 2l^e_{[c;e]}k^e l_{cf} - l_e{}^f k^e n_{c:f}$$
$$- \tfrac{3}{2}l_c{}^e \pounds_k n_e - k^e(\pounds_n n_{(e)}{}_{:c)} + k^e \pounds_n n_e l_{(cf)}k^f - k^e \rho_{,(e}\pounds_n n_{c)}; \qquad (5.2b)$$

$$K_{\gamma\beta}n^{\gamma}n^{\beta} = \pounds_n l_e{}^e - l_e{}^f l_f{}^e - l_e{}^e k^f \pounds_n n_f + 2k^e l_e{}^f \pounds_n n_f$$
$$+ {}'g^{ef}(\pounds_n n_e)_{:f} + \tfrac{1}{2}{}'g^{ef}\pounds_n n_e \pounds_k n_f + {}'g^{ef}\rho_{,e}\pounds_n n_f, \qquad (5.2c)$$

CONCLUSION

Equations (4.4) reveal that the Gauss–Codazzi Equations in our case are in fact not a great deal more complicated than their counterparts for a V_3 in V_4.[1] It must also be remembered that (4.4) holds for an arbitrary choice of null rigging field n^α. The essential feature of the expressions we have obtained for the projections of the Ricci tensor in (5.2) is that they are independent of the metric of V_4. Although these expressions are quite complex, especially when written out fully with the aid of (3.6), in any particular application one invariably makes a specific choice of n^α, and this leads to considerable simplification.

In a subsequent paper, we shall, by suitable choice of n^α, analyze the double null initial-value problem in general relativity,[17] and give a more geometric interpretation than has yet been possible. We also hope to apply the results of this paper to a geometrization of the null-timelike initial-value problem[18,19,20] and eventually to provide a unification of these two versions of the characteristic initial-value problem in general relativity.

ACKNOWLEDGMENTS

I am grateful to Dr. R.A. d'Inverno, both for his initial suggestions which instigated this work, and for many helpful discussions thereafter. I am indebted to the S.R.C. for financial support.

APPENDIX

Using (2.2) in the definitions (3.3), we immediately get

$$h_{ce}k^e = l_c{}^e n_e = 0, \quad l_{ce}k^e = -h_c{}^e n_e. \qquad (A1)$$

From the form of ${}'\Gamma^a_{bc}$ given by (3.1), some straightforward calculations lead to the expressions (of which all but the second are given in Ref. 15):

$$n_{c:b} = n_{[c:b]} = n_{[c,b]},$$
$$k^a{}_{:c} = -{}'g^{ae}h_{ce} + k^e n_{c:e}k^a, \qquad (A2)$$
$${}'g_{bc:d} = 2n_{(b}h_{c)d},$$
$${}'g^{ab}{}_{.c} = 2'g^{e(a}k^{b)}n_{c:e}.$$

Equations (A1) and (A2) lead directly to the following:

$$k^\alpha{}_{:e}k^e = 0,$$
$${}'g_{bc:e}k^e = 0, \qquad (A3)$$
$${}'g_{ec:d}k^e = h_{cd},$$
$$k^e{}_{:e} = -{}'g^{ef}h_{ef}.$$

[1]J.A. Schouten, *Ricci-Calculus* (Springer, Berlin, 1954) pp. 242–53.
[2]J.A. Schouten, *Ricci-Calculus* (Springer, Berlin, 1954) pp. 235–37.
[3]J.B. Kammerer, C. R. Acad. Sci. **264**, 86 (1967).
[4]G. Lemmer, Nuovo Cimento **37**, 1659 (1965).
[5]P. Hajicek, Commun. Math. Phys. **34**, 37 (1973).
[6]L.M. Sokolowski, Acta Phys. Pol. B **6**, 529 (1975).
[7]R.A. d'Inverno, J. Math. Phys. **16**, 670 (1975).
[8]R.A. d'Inverno, J. Math. Phys. **16**, 674 (1975).
[9]J.A. Schouten, *Ricci-Calculus* (Springer, Berlin, 1954).
[10]J.A. Schouten, *Ricci-Calculus* (Springer, Berlin, 1954), pp. 72–9.
[11]J.A. Schouten, *Ricci-Calculus* (Springer, Berlin, 1954), p. 354.
[12]P.K. Rashewsky, *Riemannische Geometrie und Tensoranalysis* (VEB Deutscher Verlag der Wissenschaften, Berlin, 1959), pp. 515–22.
[13]L.P. Eisenhart, *Riemannian Geometry* (Princeton U. P., Princeton, 1960). p. 146.
[14]W.O. Vogel, Archiv. Math. **16**, 106 (1965).
[15]G. Dautcourt, J. Math. Phys. **8**, 1492 (1967).
[16]G. Dautcourt, private communication.
[17]R.K. Sachs, J. Math. Phys. **3**, 908 (1962).
[18]H. Bondi, M.G.J. Van der Burg, and A.W.K. Metzner, Proc. R. Soc. A **269**, 21 (1962).
[19]R.K. Sachs, Proc. R. Soc. A **270**, 103 (1962).
[20]L. Tamburino and J.M. Winicour, Phys. Rev. **150**, 1039 (1966).

Geometric derivation of the kinetic energy in collective models[a)]

D. J. Rowe

Department of Physics, University of Toronto, Toronto, Ontario, Canada M5S 1A7

G. Rosensteel

Department of Physics, Tulane University, New Orleans, Louisiana 70118

(Received 10 August 1978)

A separation of the many-particle kinetic energy into collective and intrinsic components is shown to result simply from a general form of the Laplace–Beltrami operator. The geometric structure of the decomposition is thereby clearly exhibited and the intricate computations previously necessary are eliminated.

1. INTRODUCTION

The very considerable successes of the Bohr–Mottelson collective model[1,2] have motivated a number of publications in recent years addressed to the problem of effecting a change of coordinates on N-particle configuration space $\mathbb{R}^{3N}$ from Cartesian to collective plus intrinsic coordinates. In essence, the various authors consider a Lie group G which acts on $\mathbb{R}^{3N}$ and decompose $\mathbb{R}^{3N}$ into orbits of G and a smooth transversal. The collective coordinates are taken to be a chart for the generic orbits while the intrinsic variables are a set of coordinates for the transversal. Thus, Villars[3,4] and Rowe[5] took $G = \mathrm{SO}(3)$ and obtained microscopic expressions of the rotational model. The full Bohr–Mottelson model for rotations and vibrations, given by the choice $G = \mathrm{GL}_+(3)$, has been investigated by Zickendraht,[6] Dzyublik *et al.*,[7] Rosensteel,[8] Gulshani and Rowe,[9] and Weaver, Cusson, and Biedenharn.[10]

The major technical problem is the expression of observables, given in Cartesian coordinates, in terms of the new coordinates. Among these the kinetic energy $T = -(1/2m)\Sigma_{n=1}^{N}\Sigma_{i=1}^{3}\partial^2/\partial x_{ni}^2$ has been computed and found to have relatively simple forms. This comes initially as a surprise in view of the complicated manipulations involved in the transformations. The purpose of this note is to exhibit the geometric origin of these simple forms and to show that they emerge directly from a very general form of the Laplace–Beltrami operator.

Observe that the many-particle kinetic energy is proportional to the Laplace–Beltrami operator Δ on $\mathbb{R}^{3N}$. In terms of an arbitrary coordinate chart (y_μ),

$$T = -\frac{1}{2m}\Delta = -\frac{1}{2m}\frac{1}{\sqrt{g}}\sum_{\mu,\nu}\frac{\partial}{\partial y_\mu}\sqrt{g}g^{\mu\nu}\frac{\partial}{\partial y_\nu}, \tag{1}$$

where $g^{\mu\nu}$ is the inverse to the metric $g_{\mu\nu} = g(\partial/\partial y_\mu, \partial/\partial y_\nu)$ and $g = \det(g_{\mu\nu})$. However, to avoid the unwieldy coordinate change implicit in (1), it is convenient to express Δ in terms of an arbitrary basis $\{\pi_\mu\}$ of vector fields. For an arbitrary Riemannian manifold one readily finds, from the invariant definition of Δ,[11]

$$\Delta = \frac{1}{\sqrt{g}}\sum_{\mu,\nu}\pi_\mu\sqrt{g}g^{\mu\nu}\pi_\nu + \sum_{\mu,\nu}\Big(\sum_{\lambda}C_{\lambda\mu}^{\lambda}\Big)g^{\mu\nu}\pi_\nu, \tag{2}$$

where $g_{\mu\nu} = g(\pi_\mu,\pi_\nu)$ and $[\pi_\mu,\pi_\nu] = \Sigma_\lambda C_{\mu\nu}^{\lambda}\pi_\lambda$. This expression clearly coincides with (1) for the special case $\pi_\mu = \partial/\partial y_\mu$.

For our case of the Euclidean space $\mathbb{R}^{3N}$, there is a still more convenient expression which eliminates the $\sqrt{g}$ dependence. On $\mathbb{R}^{3N}$, an arbitrary vector field π_μ can be expressed in terms of Cartesian derivatives

$$\pi_\mu = \sum_{n=1}^{N}\sum_{i=1}^{3}A_\mu^{ni}(x)\frac{\partial}{\partial x^{ni}}. \tag{3}$$

Thus, if the set of vector fields $\{\pi_\mu\}$ is a basis, Eq. (3) can be inverted to give

$$\frac{\partial}{\partial x^{ni}} = \sum_{\mu}(A^{-1})_{ni}^{\mu}\pi_\mu. \tag{4}$$

The metric and its inverse in the $\{\pi_\mu\}$ basis are therefore

$$g_{\mu\nu} \equiv g(\pi_\mu,\pi_\nu) = \sum_{n,i}A_\mu^{ni}A_\nu^{ni}, \tag{5}$$

$$g^{\mu\nu} = \sum_{n,i}(A^{-1})_{ni}^{\mu}(A^{-1})_{ni}^{\nu}. \tag{6}$$

Substituting Eq. (4) into the Cartesian expression $T = -(1/2m)\Sigma_{n,i}\partial^2/\partial x_{ni}^2$ for the kinetic energy, we obtain:

$$T = -\frac{1}{2m}\sum_{\mu,\nu}\sum_{n,i}(A^{-1})_{ni}^{\mu}\pi_\mu(A^{-1})_{ni}^{\nu}\pi_\nu$$

$$= -\frac{1}{2m}\sum_{\mu,\nu}\sum_{n,i}\pi_\mu(A^{-1})_{ni}^{\mu}(A^{-1})_{ni}^{\nu}\pi_\nu$$

a)Work supported in part by the National Research Council of Canada.

 0022-2488/79/030465-04$01.00

$$+\frac{1}{2m}\sum_{\mu,\nu}\sum_{n,i}[\pi_\mu,(A^{-1})^\mu_{ni}](A^{-1})^\nu_{ni}\pi_\nu$$

$$=-\frac{1}{2m}\sum_{\mu,\nu}\pi_\mu g^{\mu\nu}\pi_\nu-\frac{1}{2m}\sum_{\mu,\nu}\left(\sum_{mj}\frac{\partial}{\partial x_{mj}}A^{mj}_\mu\right)g^{\mu\nu}\pi_\nu. \tag{7}$$

This is the desired form used in this note.

To separate T into collective and intrinsic components, a set of vector fields $\{\pi_\mu\}$ must be chosen to contain a subset of collective momenta, i.e., vector fields on the G orbit, plus a complementary set of intrinsic momenta. A natural choice for the collective momenta, for example, is obtained from the Lie algebra of the group G.

If we impose the requirement that the intrinsic momenta be everywhere orthogonal to the collective momenta, it follows that the metric assumes block diagonal form and that the kinetic energy splits into collective and intrinsic parts,

$$T=T_{\rm coll}+T_{\rm int}. \tag{8}$$

We now apply the above results to evaluate $T_{\rm coll}$ for the cases $G=$SO(3) and $G=$GL$_+$(3). For the latter case we also evaluate $T_{\rm int}$.

2. SO(3) ORBITS

The tangent space to an orbit surface of SO(3) is spanned by the angular momentum L_k,

$$L_k=-i\sum_{i,j=1}^{3}\epsilon_{kij}\sum_{n=1}^{N}x_{ni}\frac{\partial}{\partial x_{nj}},\quad k=1,2,3. \tag{9}$$

The metric on the orbit is therefore

$$g_{k'k}\equiv g(L_{k'},L_k)=-(\delta_{k'k}\,\mathrm{tr}Q-Q_{k'k}), \tag{10}$$

where $Q_{k'k}=\Sigma_{n=1}^{N}x_{nk'}x_{nk}$ is the quadrupole moment. But, this metric is recognized to be proportional to the rigid body inertia tensor, $g_{k'k}=-(1/m)\mathscr{I}_{k'k}$. Moreover, the second term in Eq. (7) is zero, since in this case $\Sigma_{mj}(\partial/\partial x_{mj})A^{mj}_k=0$. Therefore, the collective kinetic energy is given by

$$T_{\rm coll}=\tfrac{1}{2}\sum_{k',k}L_{k'}(\mathscr{I}^{-1})_{k'k}L_k. \tag{11}$$

This expression is identical to that obtained in Ref. 5 by the much more complicated chain rule method. However, it differs from the expression of Villars[3,4] who used intrinsic momenta which were not orthogonal to the angular momenta. Consequently his expression for T also contains a coupling term.

3. GL$_+$(3) ORBITS

For GL$_+$(3) the construction is entirely similar except that there are now two bases of interest. Both cases are computed below.

(1) One natural basis of vector fields tangent to the GL$_+$(3) orbit surfaces is given by the nine operators τ_{ij}, $1\leqslant i,j\leqslant 3$,

$$\tau_{ij}=-i\sum_{n=1}^{N}x_{ni}\frac{\partial}{\partial x_{nj}}. \tag{12}$$

The metric is

$$g_{i'j',ij}\equiv g(\tau_{i'j'},\tau_{ij})=-\delta_{j'j}Q_{i'i} \tag{13}$$

with inverse

$$g^{i'j,ij}=-\delta_{j'j}(Q^{-1})_{i'i}. \tag{14}$$

For the second term in the expression for the Laplace–Beltrami operator, Eq. (7), observe that $A^{nk}_{ij}=-i\delta^k_j x_{ni}$. Hence,

$$\sum_{n,k}\frac{\partial}{\partial x_{nk}}(A^{nk}_{ij})=-iN\delta_{ij}. \tag{15}$$

Therefore, the collective kinetic energy is evaluated to be

$$T_{\rm coll}=\frac{1}{2m}\sum_{i'ij}\tau_{i'j}Q^{-1}_{i'i}\tau_{ij}-\frac{i}{2m}N\sum_{ij}Q^{-1}_{ji}\tau_{ij}. \tag{16}$$

This expression is similar, but not quite identical, to that given by Weaver, Cusson, and Biedenharn[10] who restricted their consideration to orbits of the subgroup SL(3)$\subset$GL$_+$(3).

(b) There is a second basis for the vector fields on the GL$_+$(3) orbits, which is preferred since it partially separates the rotational from the vibrational degrees of freedom. This basis is defined as follows[9]:

For each $\mathbf{x}\in\mathbb{R}^{3N}$, let $R(\mathbf{x})\in$SO(3) denote the orthogonal matrix which diagonalizes the quadrupole moment $Q_{ij}=\Sigma_{n=1}^{N}x_{ni}x_{nj}$. The corresponding eigenvalues are denoted by $(\lambda_1^2(\mathbf{x}),\lambda_2^2(\mathbf{x}),\lambda_3^2(\mathbf{x}))$. Thus, $R\cdot Q\cdot{}^tR=\mathrm{diag}(\lambda_1^2,\lambda_2^2,\lambda_3^2)$. Define a basis of vector fields for the GL$_+$(3) orbit surfaces by $\{L_A,t_A,\mathscr{L}_A,A=1,2,3\}$, where

$$L_A=\sum_{B,C}\sum_{ij}\epsilon_{ABC}R_{Bi}R_{Cj}\tau_{ij}, \tag{17a}$$

$$t_A=\sum_{ij}R_{Ai}R_{Aj}\tau_{ij}, \tag{17b}$$

$$\mathscr{L}_A=\sum_{B,C}\sum_{ij}\epsilon_{ABC}\frac{\lambda_C}{\lambda_B}R_{Bi}R_{Cj}\tau_{ij}. \tag{17c}$$

In order to determine the kinetic energy, it is first necessary to compute the metric with respect to this basis. It is found that the only nonzero components to the metric tensor are given by the diagonal entries

$$g(t_A,t_A)=-\lambda_A^2, \tag{18a}$$

$$g(L_A,L_A)=-(\lambda_B^2+\lambda_C^2), \tag{18b}$$

$$g(\mathscr{L}_A,\mathscr{L}_A)=-(\lambda_B^2+\lambda_C^2),\quad A,B,C,\text{cyclic}, \tag{18c}$$

and the off-diagonal components

$$g(L_A,\mathscr{L}_A)=-2\lambda_B\lambda_C. \tag{18d}$$

The inverse to the metric is easily evaluated, since at most a 2×2 block needs to be inverted in the subspace spanned by $\{L_A,\mathscr{L}_A\}$. The inverse is given by

$$\begin{pmatrix} \lambda_B^2+\lambda_C^2 & 2\lambda_B\lambda_C \\ 2\lambda_B\lambda_C & \lambda_B^2+\lambda_C^2 \end{pmatrix}^{-1} = \begin{pmatrix} \frac{\lambda_B^2+\lambda_C^2}{(\lambda_B^2-\lambda_C^2)^2} & \frac{-2\lambda_B\lambda_C}{(\lambda_B^2-\lambda_C^2)^2} \\ \frac{-2\lambda_B\lambda_C}{(\lambda_B^2-\lambda_C^2)^2} & \frac{\lambda_B^2+\lambda_C^2}{(\lambda_B^2-\lambda_C^2)^2} \end{pmatrix}. \tag{19}$$

The calculation of the Laplace–Beltrami operator will be completed if we can evaluate the second term in Eq. (7). This requires the computation of the derivatives of the coefficients of the vector fields. If we define these coefficients by

$$L_A = \sum_{mj} L_A^{mj} \frac{\partial}{\partial x^{mj}}, \quad t_A = \sum_{mj} t_A^{mj} \frac{\partial}{\partial x^{mj}},$$

$$\mathscr{L}_A = \sum_{mj} \mathscr{L}_A^{mj} \frac{\partial}{\partial x_{mj}},$$

and observe the identity,

$$\sum_{ij} \tau_{ij}(R_{Ai}R_{Bj}) = -i\delta_{AB} \sum_{C\neq A} [(\lambda_A^2+\lambda_C^2)/(\lambda_A^2-\lambda_C^2)],$$

then one can show that

$$\sum_{mj} \frac{\partial}{\partial x_{mj}} (L_A^{mj}) = \sum_{mj} \frac{\partial}{\partial x_{mj}} (\mathscr{L}_A^{mj}) = 0 \tag{20a}$$

and

$$\sum_{mj} \frac{\partial}{\partial x_{mj}} (t_A^{mj}) = -i(N-2) - 2i \sum_{B\neq A} \frac{\lambda_A^2}{\lambda_A^2-\lambda_B^2}. \tag{20b}$$

Therefore, the collective kinetic energy is given in this basis by

$$T_{\text{coll}} = \frac{1}{2m} \sum_A \left(\frac{\lambda_B^2+\lambda_C^2}{(\lambda_B^2-\lambda_C^2)^2} (L_A^2 + \mathscr{L}_A^2) - \frac{4\lambda_B\lambda_C}{(\lambda_B^2-\lambda_C^2)^2} \mathscr{L}_A L_A \right) + \frac{1}{2m} \sum_A t_A \frac{1}{\lambda_A^2} t_A - \frac{i}{2m} \sum_A \left(\frac{N-2}{\lambda_A^2} + 2 \sum_{B\neq A} \frac{1}{(\lambda_A^2-\lambda_B^2)} \right) t_A, \tag{21}$$

where we have used the fact that $[L_A,\lambda_B] = [\mathscr{L}_A,\lambda_B] = 0$. This expression is identical to that obtained by Gulshani and Rowe[8] using the much more complicated chain rule method.

4. THE INTRINSIC KINETIC ENERGY

To find T_{int} to complement the T_{coll} of Eq. (21) in Eq. (8), it is necessary to first find a set of $3N-9$ intrinsic momenta which are everywhere orthogonal to the GL$_+$(3) collective momenta.

First observe that $\mathscr{L}_A$ of Eq. (17c) can be written

$$\mathscr{L}_A = -\mathscr{J}_{BC} = -\sum_{m,n=1}^{N} \mathscr{D}_{Bm}\mathscr{D}_{Cn} j_{mn}, \quad A,B,C \text{ cyclic}, \tag{22}$$

where

$$j_{mn} = -i \sum_{i=1}^{3} \left(x_{mi} \frac{\partial}{\partial x_{ni}} - x_{ni} \frac{\partial}{\partial x_{mi}} \right), \tag{23}$$

$$\mathscr{D}_{Bm} = \sum_i \frac{1}{\lambda_B} R_{Bi} x_{mi}. \tag{24}$$

Thus j_{mn} is an angular momentum in N-dimensional particle index space. Furthermore, since

$$\sum_{n=1}^{N} \mathscr{D}_{An}\mathscr{D}_{Bn} = \delta_{AB}, \tag{25}$$

it follows that the elements $\mathscr{D}_{An}$, $A=1,2,3$, $n=1,...,N$, may be identified with the first three rows of an SO(N) matrix.

These observations suggest that one considers, as intrinsic momenta, the $3N-9$ vector fields

$$\mathscr{J}_{\alpha A} = \sum_{m,n=1}^{A} \mathscr{D}_{\alpha m}\mathscr{D}_{An} j_{mn}, \quad \alpha = 4,...,N, A = 1,2,3, \tag{26}$$

where the elements $\mathscr{D}_{\alpha m}$, $\alpha = 4,...,N$, $m = 1,...,N$ complete an SO(N) matrix.

We immediately find

$$g(\mathscr{J}_{\alpha A},L_B) = g(\mathscr{J}_{\alpha A},t_B) = g(\mathscr{J}_{\alpha A},\mathscr{L}_B) = 0, \tag{27}$$

which confirms that the intrinsic momenta and the collective momenta are orthogonal. The intrinsic components of the metric are

$$g(\mathscr{J}_{\alpha A},\mathscr{J}_{\beta B}) = \delta_{\alpha\beta}\delta_{AB}\lambda_A^2 \tag{28}$$

and hence, since $\Sigma_{mj}(\partial/\partial x_{mj})(\mathscr{J}_{\alpha A}^{mj}) = 0$, one obtains

$$T_{\text{int}} = -\frac{1}{2m} \sum_{\alpha,A} \frac{1}{\lambda_A^2} \mathscr{J}_{\alpha A}^2. \tag{29}$$

This T_{int} plus the T_{coll} of Eq. (21) give the total many-particle kinetic energy T.

Since the Euclidean space $\mathbb{R}^{3N}$ is the direct sum of the Euclidean subspaces $\mathbb{R}^{3N}$, the center-off-mass space, and $\mathbb{R}^{3(N-1)}$, the configuration space of relative coordinates, it follows immediately that the kinetic energy can be further decomposed,

$$T = T_{\text{cm}} + T_{\text{coll}} + T_{\text{int}},$$

where T_{cm} is the center-of-mass kinetic energy and T_{coll} and T_{int} are respectively the collective and intrinsic kinetic energies for $\mathbb{R}^{3(N-1)}$. This is the decomposition of T obtained by Zickendraht[5] and by Dzyublik *et al.*[7]

5. CONCLUDING REMARKS

The contruction outlined above is very general and can be employed to yield other decompositions of the kinetic energy, e.g., that appropriate for the K-harmonic coordinate system of Simonov.[12]

In the following paper we shall consider the explicit construction of collective and intrinsic coordinate charts and the corresponding decompositions of the many-particle Hilbert space into collective and intrinsic sub-Hilbert spaces.

ACKNOWLEDGMENTS

The authors are pleased to acknowledge helpful discussions with E.V. Vanagas.

[1]A. Bohr, Dan. Mat. Fys. Medd. **26**, No. 14 (1952).
[2]A. Bohr, B. Mottelson, and J. Rainwater, Rev. Mod. Phys. **48**, 365 (1976).
[3]F. Villars, Ann. Rev. Nucl. Sci. **7**, 185 (1957).
[4]F. Villars and G. Cooper, Ann. Phys. **56**, 224 (1970).
[5]D.J. Rowe, Nucl. Phys. A **152**, 273 (1970).
[6]W. Zickendraht, J. Math. Phys. **12**, 1663 (1971).
[7]A.Y. Dzyublik, V.I. Ovcharenko, A.I. Steshenko, and G.V. Filipov, Yad. Fiz. **15**, 869 (1972) [Sov. J. Nucl. Phys. **15**, 487 (1972)].
[8]G. Rosensteel, *On the Algebraic Formulation of Collective Models*, Ph.D. thesis, University of Toronto, 1975.
[9]P. Gulshani and D. J. Rowe, Can. J. Phys. **54**, 970 (1976).
[10]O.L. Weaver, R.Y. Cusson, and L.C. Biedenharn, Ann. Phys. (N.Y.) **102**, 493 (1976).
[11]N.J. Hicks *Notes on Differential Geometry* (Van Nostrand, London, 1971).
[12]Y.A. Simonov, Yad. Fiz. **3**, 630 (1966); **7**, 1210 (1968) [Sov. J. Nucl. Phys. **3**, 461 (1966); **7**, 722 (1968)].

The quantum harmonic oscillator revisited: A new look from stochastic electrodynamics

L. de la Peña[a] and A. M. Cetto

Instituto de Física, UNAM, Apdo. Postal 20-364, México 20, D.F. Mexico
(Received 28 June 1978)

We apply the theory of stochastic electrodynamics to the study of the (nonrelativistic) harmonic oscillator by using the Fokker–Planck method. It is demonstrated that the equilibrium distribution in phase space is exactly equal to that given by quantum statistical mechanics, i.e., the corresponding Wigner distribution, and that analysis of this distribution by means of a decomposition in terms of canonical densities leads automatically to the usual description of quantum mechanics in terms of excited states. All fundamental equations of quantum mechanics are recovered as approximations to zero order in the radiation terms; the first-order terms lead to the radiative corrections predicted by quantum electrodynamics, namely, the decay of states and the Lamb shift of the energy levels. The necessary differences between both treatments of the oscillator and their implications are briefly discussed.

I. INTRODUCTION

Fifteen years ago, Marshall published a remarkable paper[1] concerning the quantum-mechanical harmonic oscillator as seen from the standpoint of stochastic electrodynamics (SED). This paper had been preceded by a small series of independent attempts to study certain properties of linear systems as predicted by SED,[2] inspired mainly by Welton's phenomenological account of the Lamb shift as a result of vacuum fluctuations[3] and by the spirit of quantum electrodynamics (QED) as reflected, e.g., in Weisskopf's beautiful review paper,[4] in which the vacuum is considered a real field, the fluctuations of which are the source of some significant properties of the electron.

In his pioneering paper, Marshall shows that a pointlike oscillator subject to a random zero-point radiation field may reach a stationary phase-space distribution as a result of the interplay between the vacuum-field force and the radiation-reaction force ($2e^2/3c^3\ddot{x}$); then the system is found to behave essentially as a quantum-mechanical harmonic oscillator if the field has a spectral energy density $\rho_0(\omega) = \hbar|\omega|^3/2\pi^2c^3$ and if h is identified with Planck's constant. Moreover, Marshall showed that this similarity in behavior is extended to temperatures higher than zero if the spectral density is taken to vary according to Planck's law.

The above mentioned work[1,2] gave birth to SED, a theory which in the course of the last 15 years has received increased attention from some of its founders[5,6] as well as several newcomers.[7–16] Though SED is far from being a finished theory, it has been shown to furnish a basis for the explanation of several typically quantum-mechanical phenomena in particularly clear physical terms. Accounts of some of the main aspects of this work may be found in various review papers.[17]

Many important questions remain, however, to be answered before concluding that SED describes and explains correctly the behavior of quantum systems at a more fundamental level than the usual theory does, and that it may therefore represent a good alternative to the latter. Some of the main open questions are related to the well-known problem of defining a phase-space description for quantum mechanics (QM), to the status of QM as a statistical theory, and to the origin of quantization, to mention but a few. Such questions, which have troubled many—if not all—of us at least at some instant, reappear at their full strength when trying to endow the usual formalism of quantum theory with a more fundamental basis, as SED proposes to do. The results obtained up to now show that SED may not only revive these and other questions, but may also be in a good position to answer them satisfactorily.

It is within this spirit that we return—once more—to the harmonic oscillator and exploit its physical and mathematical simplicity to obtain some old and some new results. In view of both the novelty and the economy of the approach, which produces from a single picture the detailed behavior of the oscillator as predicted by quantum mechanics, quantum statistics, and quantum electrodynamics, the work presented here may hopefully be valuable also for physicists working in more orthodox fields. We should be pleased if this paper can be considered a retake of Marshall's work of 1963, in the light of the more recent developments of the theory.

The structure of the paper is as follows. In Sec. II we recall and discuss the Fokker–Planck equation which has been derived more than once for the harmonic oscillator of SED. When the system is in thermodynamic equilibrium with the incoherent background radiation field at temperature T, the solution of the Fokker–Planck equation coincides with the Wigner distribution for the quantum oscillator, as is shown in Sec. III. In Secs. IV and V the temperature is eliminated from the description by a formal manipulation in order to generate a statistical temperature-independent description of the mechanical system in phase space, which is mathematically complete and leads in a unique way to the time-independent Schrödinger equation for all eigenstates—ground and excited—as is shown in Sec. VI.[19]

The Hilbert-space formalism emerges therefore as a natural means of description of the harmonic oscillator of SED, even in phase space. That this, however, is neither necessary nor possible for a general stochastic oscillator is evidenced in Sec. VII, where it is shown that even though the Fokker–Planck equation for a Brownian oscillator has ex-

[a]Consultant, Instituto Nacional de Energía Nuclear, Mexico.

 0022-2488/79/030469-15$01.00

actly the same form as the corresponding Fokker–Planck equation of SED, the formalism developed in the preceding sections is not only superfluous, but even misleading in the Brownian case. The reason for this is traced to the Boltzmann statistics governing the thermal bath which causes Brownian motion. To disclose the Planck statistics of the background radiation field together with the analysis of the phase-space density in terms of Boltzmann factors as the source of quantization in SED, is one of the main points of the present paper. If we add to this the fact that only a spectral density $\rho_0 \sim \omega^3$ at $T = 0$—as assumed here—is able to generate the correct equilibrium solution for a harmonic oscillator with radiation reaction, as shown by Marshall,[1] we obtain a fundamental conclusion: SED is the *only* physically consistent stochastic theory capable of reproducing—to the extent to which a truly statistical theory can do so—the results of QM; at least for the harmonic oscillator this conclusion is inescapable.

In Sec. VIII, the physical meaning of pure states is discussed within the frame of our phase-space description, and the temperature dependence, left out in Sec. IV, is brought back to retrieve Bloch's equation for the canonical density matrix. Therefore, SED describes correctly both the mechanical and the thermodynamic aspects of the quantum oscillator in equilibrium. Also, analysis of the nonstationary problem leads to the time-dependent Schrödinger equation as a zero-order approximation (in τ) to the equation of evolution of the mechanical system (Sec. IX).

In Sec. X, a calculation of the radiative effects to first order in τ leads to the Lamb shift and the decay rate of excited states, both as functions of temperature. By setting the temperature equal to zero, the well-known results predicted by nonrelativistic QED are recovered.

A discussion of the full meaning of these results and their bearing on the problem of the existence of divergencies in nonrelativistic QED is left for the concluding section.

II. FOKKER–PLANCK EQUATION FOR THE SED OSCILLATOR

Under the assumption that the (otherwise classical) electron is continuously scattered by a background radiation field, and in turn contributes to this background field by radiating when accelerated, SED leads to a generalized Fokker–Planck equation for the phase-space distribution, namely[12,15,16,20] (to first order in $\tau = 2e^2/3mc^3$)

$$\frac{\partial Q}{\partial t} + \frac{\mathbf{p}}{m}\cdot\nabla Q + \nabla_p\cdot\left(\mathbf{F} + \frac{\tau}{m}\mathbf{p}\cdot\nabla\mathbf{F}\right)Q - \nabla_p\nabla_p : D^{pp}Q - \nabla\nabla_p : D^{xp}Q = 0, \tag{II.1}$$

where F is the external force acting on the electron and the diffusion operators are given by

$$D_{ik}^{xp}Q(x,p,t) = e^2\int_0^t dt'\langle E_i^m(t)E_j^m(t')\rangle \times\frac{\partial x_k}{\partial \bar{p}_j}Q(\bar{x},\bar{p},t') \tag{II.2}$$

and

$$D_{ik}^{pp}Q(x,p,t) = e^2\int_0^t dt'\langle E_i^m(t)E_j^m(t')\rangle \times\frac{\partial p_k}{\partial \bar{p}_j}Q(\bar{x},\bar{p},t'). \tag{II.3}$$

Here, if the particle passed through $\bar{x}(t';t)$ and $\bar{p}(t';t)$ at $t' < t$, a classical trajectory with the stochastic force turned off would bring it to $x(t)$ and $p(t)$ at $t' = t$. $\mathbf{E}^m$ is the electric component of the stochastic field, modified (to order τ^2) by the presence of the particle. For a Planck-distributed field we have

$$\langle E_j(t)E_j(t')\rangle = \frac{2\pi}{3}\int_{-\infty}^{\infty}\rho_T(\omega)e^{i\omega(t-t')}d\omega\delta_{ij} \tag{II.4}$$

with the spectral energy density given by

$$\rho_T(\omega) = \frac{\hbar|\omega|^3}{2\pi^2c^3}\frac{1+e^{-\beta\hbar\omega}}{1-e^{-\beta\hbar\omega}} = \rho_0(\omega)\frac{1+\epsilon}{1-\epsilon}, \tag{II.5}$$

where $\rho_0(\omega)$ refers to the zero-point field,

$$\rho_0(\omega) = \frac{\hbar|\omega|^3}{2\pi^2c^3}, \tag{II.6}$$

while the factor $(1-\epsilon)^{-1}(1+\epsilon)$, with

$$\epsilon = e^{-\beta\hbar\omega} = e^{-\hbar\omega/kT} \tag{II.7}$$

expresses the thermal dependence of $\rho_T(\omega)$.

In eliminating the run away solutions generated by the radiation reaction force $m\tau\ddot{\mathbf{x}}$, the spectral density is modified by the introduction of an extra factor into Eq. (II.5), which can be absorbed in $\rho_0(\omega)$,

$$\rho_0^m(\omega) = \frac{1}{1+\tau^2\omega^2}\rho_0(\omega). \tag{II.8}$$

The details of this modification will be found in Ref. 14.

For the harmonic oscillator the operators D acquire a particularly simple form: To first order in τ and under the assumption that the system is close to equilibrium, the Q in Eqs. (II. 2) and (II. 3) may be extracted from the integral sign and both D^{xp} and D^{pp} become constant factors. We call this procedure the Markovian approximation (for a more thorough discussion see Refs. 13 and 16). For a general external force, the process can of course be far from Markovian. For the calculation of D^{pp} we have $\partial p/\partial\bar{p} = \cos\omega(t-t')\cdot\exp[-\frac{1}{2}\tau\omega^2(t-t')]$. In the Markovian approximation, all terms of order τ within the integral can be neglected, Q may be taken out of the integral, and the limit of integration can be extended to infinity; Eq. (II. 3) reduces then to

$$D_{ik}^{pp} = e^2\delta_{ik}\int_0^{\infty}dt'\langle E_i(t)E_j(t')\rangle\cos\omega(t-t')$$

and this gives, when introducing Eqs. (II. 4) and (II. 5),

$$D_{ik}^{pp} = \delta_{ik}\frac{\hbar\tau m\omega^3}{2}\frac{1+\epsilon}{1-\epsilon} \equiv \delta_{ik}D_0\frac{1+\epsilon}{1-\epsilon}, \tag{II.9}$$

where $D_0 = \frac{1}{2}\hbar\tau m\omega^3$.

The calculation of D^{xp} is somewhat complicated by the appearance of a divergence. Since this term is related to a radiative correction, we postpone its calculation until Sec. X, where it will be shown to be also a constant factor proportional to τ [see Eq. (X.15)].

We now rewrite Eq. (II. 1) for the one-dimensional harmonic oscillator in the form

$$\frac{\partial Q}{\partial t} + (L_0 + L_2)Q + \frac{\partial}{\partial p}L_1 Q = 0, \tag{II.10}$$

where L_0 is the classical Liouville operator,

$$L_0 = \frac{p}{m}\frac{\partial}{\partial x} - m\omega^2 x\frac{\partial}{\partial p}, \tag{II.11}$$

L_2 is a correction of order τ to L_0,

$$L_2 = D^{xp}\frac{\partial^2}{\partial x\partial p}, \tag{II.12}$$

and L_1 contains the diffusive and dissipative terms,

$$L_1 = -\tau\omega^2 p - D^{pp}\frac{\partial}{\partial p}. \tag{II.13}$$

The grouping of terms in Eq. (II.10) has been made according to their properties of transformation under an inversion of the momentum p: L_0 and L_2 both change their signs, while $\partial_p L_1$ does not, due to the fact that the first nonvanishing contributions to D^{pp} and D^{xp} are even in p. This means that if the equilibrium distribution is assumed to be an even function of p, i.e.,

$$Q(x, -p) = Q(x,p), \tag{II.14}$$

then Eq. (II.10) separates into two equations:

$$(L_0 + L_2)Q = 0, \tag{II.15}$$

$$\frac{\partial}{\partial p}L_1 Q = 0. \tag{II.16}$$

By integrating Eq. (II.16) to $L_1 Q = h(x)$ and integrating once more over the whole range of p we get $h(x)\int_{-\infty}^{\infty} dp = -\int(\tau\omega^2 p + D^{pp}\partial_p)Q dp = 0$, due to the condition (II.14); we therefore must take $h(x) = 0$ and hence

$$L_1 Q = 0. \tag{II.17}$$

Equations (II.15) and (II.17), obtained under hypothesis (II.14), serve to determine completely the equilibrium distribution. Since L_2 is but a radiative correction to L_0, the equilibrium distribution to zero order in τ is the solution of the system of equations

$$L_0 Q = 0, \tag{II.18}$$

$$L_1 Q = 0. \tag{II.19}$$

Within this approximation, the equilibrium distribution satisfies the principle of detailed balance. To prove this explicitly, we recall that this principle, when applied to the (stationary) Fokker–Planck equation

$$LQ = 0, \tag{II.20}$$

may be expressed in terms of the conditions[21]

$$Q = \bar{Q}, \tag{II.21}$$

$$LQf = QL^{\dagger}f, \tag{II.22}$$

for any real, well-behaved function $f(x,p)$, where the bar denotes a reversal of the classical motion (such that $\bar{x} = x$, $\bar{p} = -p$) and $L^{\dagger}$ is the adjoint of L. For

$$L = L_0 + \frac{\partial}{\partial p}L_1 \tag{II.23}$$

Eq. (II.22) yields, after minor transformations,

$$fLQ = -2(L_1 Q)\frac{\partial f}{\partial p}. \tag{II.24}$$

With $f = 1$ we recover Eq. (II.20) and hence Eq. (II.24) reduces to $(L_1 Q)(\partial_p f) = 0$; since f is an arbitrary function of x and p, we thus obtain

$$L_1 Q = 0$$

and hence

$$L_0 Q = 0.$$

These are precisely Eqs. (II.18) and (II.19), which we have obtained by imposing only the condition (II.14). Hence our hypothesis (II.14), though weaker than the conditions for detailed balance, leads to the same separation of the Fokker–Planck equation into a Liouville equation for the stationary solution, Eq. (II.18), and a diffusion-dissipation equation, (II.19). It should be added that detailed balance breaks down when the radiative correction to L_0, namely L_2, is included, which shows that our hypothesis (II.14) is less stringent than detailed balance. In other words, the one-dimensional SED system satisfies the hypothesis of detailed balance only when radiative corrections are disregarded. It must be stressed that the above results apply quite generally to any one-dimensional SED system, the harmonic oscillator being only a particular case of this situation.

III. EQUILIBRIUM SOLUTION IN PHASE SPACE: THE WIGNER DISTRIBUTION

The equations determining the stationary phase-space distribution of the oscillator in equilibrium with the radiation field, in the Markov approximation, i.e., Eqs. (II.18) and (II.19), read, when written in disclosed form,

$$\frac{p}{m}\frac{\partial Q}{\partial x} - m\omega^2 x\frac{\partial Q}{\partial p} = 0, \tag{III.1}$$

$$\tau\omega^2 pQ + D^{pp}\frac{\partial Q}{\partial p} = 0, \tag{III.2}$$

with D^{pp} given by Eq. (II.9). Before proceeding to solve these equations, let us introduce the dimensionless variables:

$$p' = \frac{1}{\sqrt{m\omega\hbar}}p, \qquad x' = \left(\frac{m\omega}{\hbar}\right)^{1/2}x, \tag{III.3a}$$

and

$$\xi = \frac{2}{\hbar\omega}\left(\frac{p^2}{2m} + \frac{1}{2}m\omega^2 x^2\right) = p'^2 + x'^2. \tag{III.3b}$$

Then Eqs. (III.1) and (III.2) read

$$p'\frac{\partial Q}{\partial x} - x'\frac{\partial Q}{\partial p'} = 0, \tag{III.4}$$

$$p'Q + \frac{1+\epsilon}{1-\epsilon}\frac{\partial Q}{\partial p'} = 0. \tag{III.5}$$

From now on, however, we will simply drop the primes on x and p, since no confusion should arise.

Equation (III.4) tells us that Q is a function of the energy variable only,

$$Q = Q(\xi).$$

By denoting $\partial_\xi Q = Q'$, we then obtain from Eq. (III.5),

$$Q + \frac{1+\epsilon}{1-\epsilon}Q' = 0, \tag{III.6}$$

the solution of which is

$$Q = N\exp\left(-\frac{1-\epsilon}{1+\epsilon}\xi\right).$$

The factor N is determined from the normalization condition,

$$1 = \int\int Q dx dp = \pi\hbar\int_0^\infty Q(\xi)d\xi = \pi\hbar N\frac{1+\epsilon}{1-\epsilon}. \tag{III.7}$$

Hence the equilibrium solution is

$$Q = \frac{1}{\pi\hbar}\frac{1-\epsilon}{1+\epsilon}\exp\left(-\frac{1-\epsilon}{1+\epsilon}\xi\right). \tag{III.8}$$

This phase-space distribution for the harmonic oscillator coincides exactly with the Wigner distribution for the problem,[22a] i.e., it is the result obtained from QM by applying the Weyl transformation to the canonical density matrix for the harmonic oscillator, which in dimensionless variables may be written as follows[22b]:

$$\rho(x_1,x_2) = \frac{1}{\hbar\sqrt{\pi}}\frac{1-\epsilon}{\sqrt{1-\epsilon^2}} \times\exp\left(-\frac{1}{2(1-\epsilon^2)}[(1+\epsilon^2)(x_1^2+x_2^2) - 4\epsilon x_1x_2]\right). \tag{III.9}$$

To prove this, we recall that the Weyl transformation implies first performing the change of variables

$$x = \frac{x_1+x_2}{2}, \quad z = \frac{x_2-x_1}{2} \tag{III.10}$$

in Eq. (III.9), to obtain

$$\tilde{Q}(x,z) \equiv \rho(x-z,x+z) = \frac{1}{\hbar\sqrt{\pi}}\frac{1-\epsilon}{\sqrt{1-\epsilon^2}} \times\exp-\left(\frac{1-\epsilon}{1+\epsilon}x^2 + \frac{1+\epsilon}{1-\epsilon}z^2\right)$$

and then taking the Fourier transform of $\tilde{Q}(x,z)$ with respect to z to obtain

$$Q(x,p) \equiv \frac{1}{2\pi}\int\tilde{Q}(x,z)e^{-2ipz}dz = \frac{1}{\hbar\pi}\frac{1-\epsilon}{1+\epsilon}\exp\left(-\frac{1-\epsilon}{1+\epsilon}[x^2+p^2]\right), \tag{III.11}$$

which is just Eq. (III.8). Below we shall have occasion to see in a more suggestive form the origin of this transformation. For the time being it suffices to accept the fact that the equilibrium distribution predicted by SED coincides with the one predicted by statistical QM. The rest of the paper is devoted to a more careful study of the connection between both theories.

The mean energy corresponding to this distribution is

$$\bar{E} = \frac{\hbar\omega}{2}\bar{\xi} = \frac{\hbar\omega}{2}\frac{1+\epsilon}{1-\epsilon} \tag{III.12}$$

in agreement with Planck's law—as was expected, since the oscillator is in thermodynamic equilibrium with a radiation field with this same energy per normal mode. This result, though well known, is worth some comments. As temperature goes to zero ($\epsilon\to 0$), the mean energy tends to its asymptotic value

$$\bar{E}_0 = \tfrac{1}{2}\hbar\omega,$$

i.e., a value different from zero, as a result of the existence of unfreezeable fluctuations in the field, corresponding to its zero-point energy. The oscillator in equilibrium with this field then also acquires an unfreezeable motion; hence, according to SED the zero-point energy of matter is but a result of the existence of the zero-point energy of the radiation field. In the case of purely mechanical systems, such as, e.g., Brownian motion, the heat bath is totally frozen at $T = 0$ and with it also the stochastic motion. We see here an aspect of the essential role played by the stochastic field of SED in explaining the results of QM: No mechanical model will do.[23]

IV. SEPARATION OF MECHANICAL AND THERMAL VARIABLES: THE ENERGY SPECTRUM

We now explore the possibility of excluding the temperature from the description, in order to extract and analyze the purely mechanical problem of the oscillator, as is done in QM when dealing with pure states. We will see that an immediate result of this procedure is the emergence of discrete energy levels. For this purpose, let us express the partition function of the system $Z(\beta)$ as an integral over states of the oscillator characterized by a given value of an abstract variable E; we *assume* that the relative probability with which each of these states contributes to $Z(\beta)$ is properly described in terms of canonical densities and the spectral density $g(E)$; thus, we write $Z(\beta)$ in the form

$$Z(\beta) = \int_0^\infty g(E)e^{-\beta E}dE. \tag{IV.1}$$

The parameter E plays clearly the role of an energy and we will call it the energy variable; so $g(E)$ is the spectral density

of states of the oscillator with energy E. More generally, we may define the mean value of any physical quantity A depending on the energy by means of the formula

$$\bar{A}(\beta)=\frac{1}{Z(\beta)}\int_0^\infty A(E)g(E)e^{-\beta E}dE. \tag{IV.2}$$

Equation (IV.2) and its inverse—assuming that they exist—allow us to construct from a function of β—i.e., temperature—$a(\beta)\equiv\bar{A}(\beta)$ a new function $\check{a}(E)\equiv A(E)$ depending now on E, whenever we know $Z(\beta)$ or its inverse, $g(E)$.

Equations (IV.1) and (IV.2) taken together show that the mean value of $A(E)=1$ is unity, as it must be; taking $A(E)=E$ we get for the mean energy the well-known formula

$$\bar{E}=\frac{1}{Z}\int_0^\infty Eg(E)e^{-\beta E}dE=-\frac{\partial}{\partial\beta}\ln Z(\beta)$$

whence

$$Z(\beta)=\exp\left[-\int^\beta \bar{E}(\beta')d\beta'\right]. \tag{IV.3}$$

For the harmonic oscillator the mean energy is given by Eq. (III.12), with $\epsilon=e^{-\beta\hbar\omega}$; hence we can in fact calculate the partition function by means of a simple integration,

$$Z(\beta)=\exp\left(-\frac{1}{2}\int^{\beta\hbar\omega}\frac{1+e^{-y}}{1-e^{-y}}dy\right)=\frac{\exp(-\frac{1}{2}\beta\hbar\omega)}{1-\exp(-\beta\hbar\omega)}. \tag{IV.4}$$

According to Eq. (IV.1) the spectral density of states is given by the inverse Laplace transform of the partition function; to calculate it we develop the denominator of Eq. (IV.4) in a power series of $e^{-\beta\hbar\omega}$, thus getting

$$g(E)=\mathcal{L}^{-1}Z(\beta)=\sum_{k=0}^\infty \mathcal{L}^{-1}\exp[-\beta\hbar\omega(k+\tfrac{1}{2})]$$

or

$$g(E)=\sum_{k=0}^\infty \delta(E-\hbar\omega(k+\tfrac{1}{2})). \tag{IV.5}$$

Thus we see that the density of states is zero for all values of the energy E that do not belong to the discrete spectrum E_k, where

$$E_k=\hbar\omega(k+\tfrac{1}{2}). \tag{IV.6}$$

Stated in more usual terms, we have obtained the surprising result that the energy E is quantized.[24] Two comments are worth mentioning at this point. The first is connected with the use of Planck's law; it might be alleged that the quantization of energy has been introduced through the back door, since it is already contained in the assumed density for the field; this, however, is not the case, since Planck's distribution may be derived without any quantum postulate from the assumption of SED, namely, the postulated existence of the zero-point field.[25] The second comment is that the emergence of the right discrete spectrum of E depends crucially on our (arbitrary) selection of Gibbs' densities in the definition of our linear transform, Eqs. (IV.1) and (IV.2). Since this transform will lead us to the usual Schrödinger description of the oscillator in a unique and natural way (though in a rather abstract treatment), the justification on *physical grounds* of the need of just this transformation, which here is simply postulated, will close the gap to a rigorous derivation of Schrödinger's theory from SED. In other words, to explain why we *must* go—and not only why we *may* go—to the usual quantum mechanical description of the oscillator, we must derive Eqs. (IV.1) and (IV.2) for the harmonic oscillator from first principles; this will obviously imply an elucidation of the meaning of the variable E, introduced here in an entirely formal way.

V. SEPARATION OF MECHANICAL AND THERMAL VARIABLES: PSEUDOSTATES OF THE OSCILLATOR

With $g(E)$ given by Eq. (IV.5), Eq. (IV.2) reduces to [with $a(\beta)\equiv\bar{A}(\beta)$ and $\check{a}(E)\equiv A(E)$]

$$a(\beta)=\frac{1}{Z(\beta)}\sum_0^\infty \check{a}(E_k)e^{-\beta E_k}, \tag{V.1}$$

where the partition function is given by [see Eq. (IV.4)]

$$Z(\beta)=\sum_0^\infty e^{-\beta E_k}=\frac{\epsilon^{1/2}}{1-\epsilon}. \tag{V.2}$$

We now use these results to analyze the phase-space density $Q(x,p;\beta)$ in terms of energy states; for this purpose we first define the energy pseudostates $Q_k(x,p)$ as follows,

$$Q_k(x,p)\equiv\check{Q}(x,p;E_k). \tag{V.3}$$

The reason for the name "pseudostate" will become apparent below. From Eq. (V.1) we get

$$Q(\xi;\beta)=\frac{1}{Z}\sum_0^\infty Q_k(\xi)e^{-\beta E_k}. \tag{V.4}$$

By virtue of Eq. (V.2), we see from Eq. (V.4) that the Q_k also are normalized to unity,

$$\int Q_k(\xi)dxdp=1. \tag{V.5}$$

We can now proceed to determine the components $Q_k(\xi)$ by introducing Eq. (V.4) into (III.6),

$$\sum_0^\infty Q_k\epsilon^k+\frac{1+\epsilon}{1-\epsilon}\sum_0^\infty Q'_k\epsilon^k=0,$$

or since $(1+\epsilon)(1-\epsilon)^{-1}=\Sigma_0^\infty(\epsilon^m+\epsilon^{m+1})$,

$$\sum_k \epsilon^k\left[Q_k+Q'_k\sum_m(\epsilon^m+\epsilon^{m+1})\right]=0.$$

Regrouping terms with equal power of ϵ,

$$\sum_n \epsilon^n\left[Q_n+\sum_{m=0}^n Q'_m+\sum_{m=0}^{n-1}Q'_m\right]=0,$$

and noticing that this equality must hold at any temperature (i.e., any value of ϵ), we obtain

$$Q_n+Q'_n+2\sum_{m=0}^{n-1}Q'_m=0 \tag{V.6}$$

for any $n\geqslant 0$ [with $Q_m=0$ for $m<0$, since the energy was assumed nonnegative from the beginning; see Eq. (IV.2)]. By writing this equation for $n-1$,

$$Q_{n-1} + Q'_{n-1} + 2\sum_{m=0}^{n-2} Q'_m = 0,$$

and subtracting from Eq. (V.6), one is left with a simple recurrence relation among the functions Q_n,

$$Q'_n + Q_n = Q_{n-1} - Q'_{n-1}. \tag{V.7}$$

For $n = 0$, this yields

$$Q'_0 + Q_0 = 0, \quad Q_0 = C_0 e^{-\xi}.$$

To obtain the subsequent solutions Q_n, we use the ansatz

$$Q_n = C_n e^{-\xi} P_n(\xi); \tag{V.8}$$

substitution of Eq. (V.8) in (V.7) gives

$$P'_n = 2P_{n-1} - P'_{n-1},$$

which upon the change of variable $y = 2\xi$ is easily recognizable as a recurrence relation for the Laguerre polynomials $L_n(y)$,

$$\frac{dL_n}{dy} = L_{n-1} - \frac{dL_{n-1}}{dy}.$$

We therefore find that Eq. (V.8) can be written as

$$Q_n = C_n e^{-\xi} L_n(2\xi)$$

where L_n is the Laguerre polynomial of order n (irregular solutions of Laguerre's equation are obviously eliminated on physical grounds),

$$L_n(2\xi) = \sum_{s=0}^{n} \frac{n!(-2\xi)^s}{(n-s)!(s!)^2}, \tag{V.9}$$

and C_n is determined from the normalization condition [see Eq. (5)]

$$\frac{1}{\pi\hbar}\int_0^\infty Q_n d\xi = C_n \int_0^\infty e^{-\xi} L_n(2\xi) d\xi$$

whence

$$C_n = \frac{(-1)^n}{\pi\hbar}.$$

The normalized solution is therefore

$$Q_n = \frac{(-1)^n}{\pi\hbar} e^{-\xi} L_n(2\xi) \tag{V.10}$$

and the mean energy assigned to Q_n is correctly given by Eq. (IV.6) as can be proved by calculating the integral

$$\int_0^\infty \xi Q_n(\xi) d\xi$$

with the use of Eqs. (V.9) and (V.10), or else, by multiplying Eq. (V.7) by ξ and integrating over ξ. From the differential equation satisfied by the nth Laguerre polynomial

$$\xi L''_n + (1 - 2\xi)L'_n + 2nL_n = 0 \tag{V.11}$$

we obtain the equation for Q_n,

$$\xi Q''_n + Q'_n + (2n + 1 - \xi)Q_n = 0, \tag{V.12}$$

which is, by the way, of the same form as the radial part of the Schrödinger equation for the s states of the hydrogen atom.

The Q_n form an orthonormal set of functions in ξ space, as follows from Eq. (V.12), and therefore only one of them (namely Q_0) is positive definite throughout. This means that the Q_n for $n \neq 0$ do not represent real distributions for physical states; it is their combination in nonnegative forms such as Eq. (V.4) that represents a true distribution. It is to stress this fact that we say that the Q_n represent pseudostates of the system. In the limit $\beta \to \infty$ ($T \to 0$) Q tends to Q_0, i.e., Q_0 is the equilibrium distribution at temperature $T = 0$.

Even though the isolated Q_n correspond to real physical states only for $n = 0$, the fact of their forming a complete basis in ξ space renders them particularly useful for the analysis of the harmonic oscillator even outside thermodynamic equilibrium. In the following we shall analyze some of their mathematical properties, mainly those which will help us establish the connection with the usual (Schrödinger) description in configuration space.

We first note that Eq. (V.7) suggests the introduction of raising and lowering operators. Let B and its adjoint $B^\dagger$ be such that

$$Q_n = \frac{1}{n} B^\dagger Q_{n-1}, \tag{V.13a}$$

$$Q_{n-1} = \frac{1}{n} B Q_n. \tag{V.13b}$$

Introducing the latter into Eq. (V.7) one gets

$$n(Q'_n + Q_n) = BQ_n - \partial_\xi (BQ_n)$$

and introducing (V.13a) into Eq. (V.7) one gets after making $n \to n + 1$,

$$(n + 1)(-Q'_n + Q_n) = B^\dagger Q_n + \partial_\xi (B^\dagger G_n).$$

Addition of these two equations gives

$$(2n + 1)Q_n = LQ_n + \partial_\xi (MQ_n) + \partial_\xi Q_n, \tag{V.14}$$

where

$$L = L^\dagger = B + B^\dagger \tag{V.15a}$$

and

$$M = -M^\dagger = B^\dagger - B. \tag{V.15b}$$

Equations (V.12) and (V.14) can be combined to obtain an equation for the operators themselves,

$$-2\partial_\xi + \xi - \xi\partial_\xi^2 = L + \partial_\xi M. \tag{V.16}$$

By taking the adjoint of this equation and subtracting one from the other, one gets

$$(\partial_\xi M) = -4\partial_\xi + \partial_\xi^2 \xi - \partial_\xi^2 = -2\partial_\xi$$

whence

$$M = -2\xi\partial_\xi + M_0,$$

where M_0 is an integration constant. By introducing this into Eq. (V.16) one obtains

$$L = \xi(1 + \partial_\xi^2) - M_0\partial_\xi.$$

$B^\dagger$ and B are then, according to Eqs. (V.15),

$$B^\dagger = \tfrac{1}{2}(L + M) = \frac{\xi}{2}(\partial_\xi - 1)^2 - \frac{M_0}{2}(\partial_\xi - 1), \tag{V.17a}$$

$$B = \tfrac{1}{2}(L - M) = \frac{\xi}{2}(\partial_\xi + 1)^2 - \frac{M_0}{2}(\partial_\xi + 1). \quad \text{(V.17b)}$$

To determine M_0 we notice that at $\xi = 0$, Eq. (V.12) reduces to

$$2nQ_n = -Q'_n - Q_n$$

but the term on the lhs is equal to $2B^\dagger Q_{n-1}$, according to Eq. (V.13a), and hence, by applying (V.17a) at $\xi = 0$ and comparing with Eq. (V.7) we obtain $M_0 = -1$, and therefore,

$$B^\dagger = \tfrac{1}{2}[\xi(\partial_\xi - 1)^2 + \partial_\xi - 1] \quad \text{(V.18a)}$$

$$B = \tfrac{1}{2}[\xi(\partial_\xi + 1)^2 + \partial_\xi + 1]. \quad \text{(V.18b)}$$

The commutator of these operators is

$$[B^\dagger, B] = \xi\partial_\xi^2 + \partial_\xi - \xi, \quad \text{(V.19)}$$

so that the differential equation (V.12) can be written simply as

$$[B, B^\dagger]Q_n = (2n + 1)Q_n. \quad \text{(V.20)}$$

The operators $B^\dagger$, B are helpful in determining statistical properties of the pseudostates Q_n. Let us take for instance the average of an arbitrary real function of energy $F(\xi)$,

$$\langle n \mid F \mid n \rangle = \pi\hbar \int FQ_n d\xi = \frac{\pi\hbar}{n} \int FB^\dagger Q_{n-1} d\xi$$

$$= \frac{\pi\hbar}{n} \int Q_{n-1} BF d\xi = \frac{1}{n}\langle n - 1 | BF | n - 1 \rangle.$$

By applying this formula to $F(\xi) = \xi^r$ one obtains

$$\langle n \mid \xi^r \mid n \rangle = \frac{1}{2n}\langle n - 1 \mid \xi^{r+1} + (2r + 1)\xi^r + r^2\xi^{r-1} \mid n - 1 \rangle.$$

For $r = 0$, this gives

$$\bar{\xi}_n = \langle n | \xi | n \rangle = 2n + 1, \quad \text{(V.22)}$$

whereas for $r = 1$ one gets

$$\overline{\xi^2}_n - (\bar{\xi}_n)^2 = 1, \quad \text{(V.23)}$$

and so on.

Equation (V.22) gives once more the mean energy of the pseudostate Q_n as $E_n = \frac{1}{2}\hbar\omega(n + \frac{1}{2})$, just as in QM, but Eq. (V.23) shows that the energy has a nonzero dispersion; in fact, for all $k > 1$ $\langle H^k \rangle \neq \langle H \rangle^k$ and $\langle n | H^2 | n \rangle = E_n^2 + (\frac{1}{2}\hbar\omega)^2$, compared to the QM value $\langle n | \hat{H}^2 | n \rangle = E_n^2$. We see that in SED (and, as we have previously remarked, in any stochastic theory developed as a foundation to QM) the energy and all other physical quantities are not dispersionless, even in "pure" states such as Q_n. Here, then, we uncover a rather fundamental difference between SED and QM, in the shape of a particular instance of a general result due to Cohen,[26] according to which no phase-space description can exactly reproduce all results of usual QM. From the point of view of SED, this is an almost obvious result, since no stochastic theory can generate dispersionless (physical) states of the dynamical variables. The reason for this difference between SED and QM is that QM is not a strictly statistical theory, whereas SED is. In our particular instance this may be seen by realizing that the *definition* of the dispersion of the energy used in QM in terms of eigenvalues of the Hamiltonian operator differs from that used in our statistical phase space description. Hence a seemingly terrible problem (for SED, of course) reduces when closely analyzed to a simple difference of definitions, from the formal point of view. Of course, physically speaking both theories predict different results for the dispersion of the energy; it is important, however, to note that this basic difference cannot be tested experimentally with spectroscopic studies, for instance, since these only reveal energy differences occurring during transitions between pseudostates and we will have occasion below to show that the transition matrix elements are the same in both theories (see Sec. IX). We may try to explain these results by saying that in SED, where no sharp energy levels exist, the transition must be understood as a resonance phenomenon involing the oscillations of the background field.

VI. RETRIEVAL OF USUAL QM

As already recalled in Sec. III, application of the inverse Weyl transformation to the phase space density leads to the description of the statistical behavior of the system in terms of the density matrix in the coordinate representation. In this section we try to look at the origin of this connection in a more direct way by using a method suggested by previous work.[12,13,15,16,19]

For this purpose we return to the description in phase space, i.e., to Eqs. (III.4) and (III.5):

$$p\partial_x Q - x\partial_p Q = 0, \quad \text{(III.4)}$$

$$pQ + \frac{1}{2}\frac{1 + \epsilon}{1 - \epsilon}\partial_p Q = 0. \quad \text{(III.5)}$$

In eliminating the temperature dependence by a procedure entirely similar to that of Sec. IV and V, one arrives at the equations

$$p\partial_x Q_k - x\partial_p Q_k = 0, \quad \text{(VI.1a)}$$

$$(x\partial_p + p\partial_x + 4xp)Q_k = (4xp - p\partial_x - x\partial_p)Q_{k-1}. \quad \text{(VI.1b)}$$

Equation (VI.1a) expresses the fact that each Q_k is a function of $\xi = p^2 + x^2$; Eq. (VI.1b) can also be obtained more directly from Eq. (V.7).

By combining Eqs. (VI.1) one obtains the following set of equations:

$$(\partial_x + 2x)Q_k = (2x - \partial_x)Q_{k-1},$$

$$(\partial_p + 2p)Q_k = (2p - \partial_p)Q_{k-1}.$$

Just as in going from Eq. (III.8) to Eq. (III.9) we Fourier transform these equations with respect to the variable p; then for

$$\tilde{Q}_k(x,z) = \int dp e^{2ipz} Q_k(x,p) \quad \text{(VI.2)}$$

we obtain the set of equations

$$(\partial_x + 2x)\tilde{Q}_k = (2x - \partial_x)\tilde{Q}_{k-1}$$

$$(\partial_z + 2z)\tilde{Q}_k = -(2z - \partial_z)\tilde{Q}_{k-1}.$$

These equations are identical except for the overall sign of the rhs. This suggests introducing a new pair of variables x_1 and x_2, namely,

$$x_1 = x - z, \quad x_2 = x + z \tag{VI.3}$$

in terms of which they transform into

$$(\partial_1 + x_1)\tilde{Q}_k = (-\partial_2 + x_2)\tilde{Q}_{k-1}, \tag{VI.4a}$$

$$(\partial_2 + x_2)\tilde{Q}_k = (-\partial_1 + x_1)\tilde{Q}_{k-1}. \tag{VI.4b}$$

In particular, for $k = 0$ we obtain

$$(\partial_1 + x_1)\tilde{Q}_0 = 0, \quad (\partial_2 + x_2)\tilde{Q}_0 = 0, \tag{VI.5}$$

whence

$$Q_0 = N_0 \exp[-\tfrac{1}{2}(x_1^2 + x_2^2)]. \tag{VI.6}$$

By the symmetry of Eqs. (VI.4), if $\tilde{Q}_{k-1}$ is a product of functions of the type

$$\tilde{Q}_{k-1} = \varphi_{k-1}(x_1)\varphi_{k-1}(x_2),$$

then also $\tilde{Q}_k$ can be written as such. Since Q_0 is such a product, according to Eq. (VI.6), we look for solutions of the form

$$\tilde{Q}_k = \varphi_k(x_1)\varphi_k(x_2). \tag{VI.7}$$

According to Eq. (VI.2), $\tilde{Q}_k(x,0)$ is the configuration-space density function $\rho_k(x)$,

$$\rho_k(x) \equiv \int dp Q_k(x,p) = \tilde{Q}_k(x,0).$$

Since, on the other hand, $x_1 = x_2 = x$ at $z = 0$, as seen from Eqs. (VI.3), we obtain

$$\rho_k(x) = \varphi_k(x_1)\varphi_k(x_2)|_{z=0} = \varphi_k^2(x). \tag{VI.8}$$

Notice that, even though the Q_k are not positive definite for $k \neq 0$, the ρ_k are; therefore these marginal distributions have the formal properties of true densities.

To construct the set of φ_k, we propose to write

$$\varphi_k(x_i) = \exp(-\tfrac{1}{2}x_i^2)\theta_k(x_i), \quad i = 1,2. \tag{VI.9}$$

where $\theta_0 = N_0^{1/2}$ is constant, according to Eq. (VI.5). From Eqs. (VI.4) we then obtain

$$\theta_k(x_1)\theta'_k(x_2) = \theta_{k-1}(x_2)[-\theta'_{k-1}(x_1) + 2x_1\theta_{k-1}(x_1)], \tag{VI.10a}$$

$$\theta_k(x_2)\theta'_k(x_1) = \theta_{k-1}(x_1)[-\theta'_{k-1}(x_2) + 2x_2\theta_{k-1}(x_2)]. \tag{VI.10b}$$

Combination of Eqs. (VI.10) gives

$$\theta''_k - 2x_i\theta'_k + C_k\theta_k = 0 \tag{VI.11}$$

for every $\theta_k(x_i)$, where the C_k are the separation constants. From Eqs. (VI.10) it also follows that $\theta_{k-1} \sim \theta'_k$; from this and the fact that θ_0 is constant, it follows that θ_k is a polynomial or order k. By requiring the coefficient of x_i^k in Eq. (VI.11) to vanish we obtain $C_k = 2k$ and hence

$$\theta''_k - 2x_i\theta'_k + 2k\theta_k = 0. \tag{VI.12}$$

The θ_k are therefore the Hermite polynomials,

$$\theta_k(x_i) = H_k(x_i) \tag{VI.13}$$

and the φ_k defined in Eq. (VI.9) satisfy the eigenvalue equation which follows from Eqs. (VI.9) and (VI.12),

$$\varphi''_k(x_i) + (2k + 1 - x_i^2)\varphi_k(x_i) = 0. \tag{VI.14}$$

At $z = 0$ both Eqs. (VI.14) reduce to

$$\varphi''_k(x) - x^2\varphi_k(x) = (2k + 1)\varphi_k(x) \tag{VI.15}$$

which is the Schrödinger equation for the probability amplitude [in the sense of Eq. (VI.8)], with the correct eigenvalues already inserted.

The above results may be summarized as follows: We started from the energy representation of the phase-space density

$$Q(\xi;\beta) = \frac{1}{Z}\sum_k Q_k(\xi)e^{-\beta E_k}$$

with $E_k = \hbar\omega(k + \frac{1}{2})$ and $Z = \Sigma_k e^{-\beta E_k}$, which we subjected first to a Fourier transformation with respect to p, obtaining

$$\tilde{Q}(x,z;\beta) = \frac{1}{Z}\sum_k \tilde{Q}_k(x,z)e^{-\beta E_k}$$

and then transferred to the x_1x_2 space defined in Eqs. (VI.3), to get

$$\rho(x_1,x_2) \equiv \tilde{Q}\left(\frac{x_1 + x_2}{2}, \frac{x_2 - x_1}{2};\beta\right)$$

$$= \frac{1}{Z}\sum_k \tilde{Q}_k(x_1,x_2)e^{-\beta E_k}$$

or finally, using Eq. (VI.7),

$$\rho(x_1,x_2) = \frac{1}{Z}\sum_k \varphi_k(x_1)\varphi_k(x_2)e^{-\beta E_k}. \tag{VI.16}$$

Here we recognize the usual expression for the canonical density matrix in the x representation. On the other hand, the Q_k may be explicitly written with the help of the inverse of Eq. (VI.2) and Eqs. (VI.3) and (VI.7), as follows,

$$Q_k(x,p) = \frac{1}{2\pi}\int \tilde{Q}_k(x,z)e^{-2ipz}dZ$$

$$= \frac{1}{2\pi}\int \varphi_k(x - z)\varphi_k(x + z)e^{-2ipz}dZ.$$

We recognize here the usual expression for Wigner's distrubution for the k th pseudostate,[22] which happens to be positive definite only for the ground state and for the equilibrium state at temperature $T = 1/k\beta$,

$$Q(x,p;\beta) = \frac{1}{2\pi Z}\int dz e^{-2ipz} \times \sum_k e^{-\beta E_k}\varphi_k(x - z)\varphi_k(x + z). \tag{VI.18}$$

The above results are developments of Eqs. (III.9) and (III.11) in terms of the basis φ_k.

Here also, the structure of the equations suggests the introduction of raising and lowering operators, now in x_1x_2 space. In fact, by introducing Eq. (VI.7) into Eqs. (VI.4) we obtain

$$\varphi_k(x_1)(\partial_2 + x_2)\varphi_k(x_2) = \varphi_{k-1}(x_2)(-\partial_1 + x_1)\varphi_k(x_1),$$
$$\varphi_k(x_2)(\partial_1 + x_1)\varphi_k(x_1) = \varphi_{k-1}(x_1)(-\partial_2 + x_2)\varphi_k(x_2).$$

This system of equations suggests a separation of the form

$$\varphi_k(x_i) = \frac{1}{d_k}(-\partial_i + x_i)\varphi_{k-1}(x_i),$$
$$\varphi_k(x_i) = \frac{1}{d_k}(\partial_i + x_i)\varphi_k(x_i); \quad i = 1,2$$

and comparison with Eq. (VI.14) gives $d_k^2 = 2k$. We therefore obtain

$$\varphi_k(x_i) = \frac{1}{\sqrt{k}} A_i^\dagger \varphi_{k-1}(x_i), \tag{VI.19a}$$

$$\varphi_{k-1}(x_i) = \frac{1}{\sqrt{k}} A_i \varphi_k(x_i), \tag{VI.19b}$$

where the lowering and raising operators are

$$A_i = \frac{1}{\sqrt{2}}(\partial_i + x_i), \tag{VI.20a}$$

$$A_i^\dagger = \frac{1}{\sqrt{2}}(-\partial_i + x_i), \tag{VI.20b}$$

respectively. These operators obey the commutation relations:

$$[A_1, A_2] = [A_1^\dagger, A_2^\dagger] = 0, \tag{VI.21}$$

and

$$[A_i, A_i^\dagger] = 1. \tag{VI.22}$$

When going over to configuration space ($z = 0$), the two sets of operators reduce to the well-known creation and annihilation operators:

$$A_1 = A_2 \to \frac{1}{\sqrt{2}}(\partial + x) \equiv a, \tag{VI.23a}$$

$$A_1^\dagger = A_2^\dagger \to \frac{1}{\sqrt{2}}(-\partial + x) \equiv a^\dagger, \tag{VI.23b}$$

and Eqs. (VI.19) take on the usual form:

$$\varphi_k(x) = \frac{1}{\sqrt{k}} a^\dagger \varphi_{k-1}(x), \tag{VI.24a}$$

$$\varphi_{k-1}(x) = \frac{1}{\sqrt{k}} a \varphi_k(x). \tag{VI.24b}$$

The commutation rules (VI.21) and (VI.22) are not altered in taking $z = 0$.

The usual formalism in configuration space arises therefore as a projection of the operator formalism in x_1x_2 space, related to phase space through Eqs. (VI.2) and (VI.3).

VII. COMPARISON WITH BROWNIAN MOTION

A classical theory applied to the harmonic oscillator embedded in a classical Planck-distributed background radiation field thus leads in a natural way to the quantum description of the harmonic oscillator.

This is by no means a trivial results; for instance, it suffices to make a small but arbitrary change in the temperature dependence of the equilibrium field, to destroy a good deal of the theory. To stress the remarkable properties of the theory and at the same time exhibit the source of the differrences between more familiar stochastic systems and their QM (i.e., SED) counterpart—a point on which still some confusion prevails—it is illustrative to examine the Brownian harmonic oscillator. The interest of this digression is enforced by the fact that the Brownian oscillator obeys a Fokker–Planck equation which has exactly the same form as the quantum mechanical one, as far as the phase-space dependence goes[27]:

$$\frac{\partial Q}{\partial T} + \frac{p}{m}\frac{\partial Q}{\partial x} + \frac{\partial}{\partial p}(-m\omega^2 x - \gamma p)Q - D\frac{\partial^2 Q}{\partial p^2} = 0 \tag{VII.1}$$

[compare with Eq. (II.10), with $L_2 = 0$ and $D^{pp} = \text{const.}$)]. The diffusion coefficient D can be calculated with the same formula (II.2), and the force $eE^m(t)$ being replaced by a delta-correlated function, $\langle F(t)F(t')\rangle = D_b\delta(t' - t)$, corresponding to white noise,

$$D \equiv D^{pp} = \int_{-\infty}^{\infty} \langle F(t)F(t')\rangle \frac{\partial p}{\partial \bar{p}} dt'$$

$$= \int D_b\delta(t' - t)\frac{\partial p}{\partial \bar{p}} dt' = D_b. \tag{VII.2}$$

We notice incidentally an important difference between the Brownian system and the SED system; whereas in the latter D^{pp} is in general a phase function and becomes independent of x and p for the harmonic oscillator only, in the classical case, due to the infinitely small correlation time of the stochastic force (i.e., due to its spectral density being independent of frequency), D is the same constant D_b for any external potential. This causes any equilibrium distribution in the Brownian case to have the canonical form $\sim e^{-\beta H}$, whereas in SED this form is characteristic of linear systems only.

The equilibrium solution is obtained also in this case by separating the Liouville part,

$$L_0 Q = \frac{p}{m}\frac{\partial Q}{\partial x} - m\omega^2 x\frac{\partial Q}{\partial p} = 0 \tag{VII.3}$$

from the diffusive-dissipative part,

$$-L_1 Q = \gamma p Q + D_b\frac{\partial Q}{\partial p} = 0, \tag{VII.4}$$

in the stationary Fokker–Planck equation. Notice that the equilibrium distribution satisfies the principle of detailed balance. The problem seems therefore very similar to that of SED, as can be seen by comparing with Eqs. (III.1) and (III.2), the only difference being the specific values of D_b and the friction coefficient γ, which are related through Einstein's formula

$$\frac{D_b}{m\gamma} = kT = \frac{1}{\beta}. \tag{VII.5}$$

The (normalized) equilibrium solution is therefore the Maxwell–Boltzmann distribution

$$Q = \frac{\beta\omega}{2\pi}\exp\left[-\beta\left(\frac{p^2}{2m} + \frac{m\omega^2x^2}{2}\right)\right] = \frac{\beta\omega}{2\pi}e^{-\beta H} \tag{VII.6}$$

and the mean energy is given by

$$\bar{E} = \frac{1}{\beta} = kT. \tag{VII.7}$$

From Eqs. (IV.3) and (VI.7) we obtain now

$$Z = \exp\left[-\int^{\beta}\bar{E}(\beta)d\beta\right] = \frac{1}{\beta}$$

and hence, from Eq. (IV.1),

$$g(E) = 1$$

i.e., the spectral density of states is constant. Equation (IV.2) applied to this case,

$$Q(H;\beta) = \beta\int\check{Q}(H;E)e^{-\beta E}dE,$$

together with Eq. (VII.6) give

$$\check{Q}(H;E) = \frac{\omega}{2\pi}\delta(E - H). \tag{VII.8}$$

No "energy levels" are obtained: Quite on the contrary, the distribution of $\check{Q}(E)$ is uniform throughout the energy range ($g = 1$). Instead of a discrete set of orthogonal functions Q_n, we get a continuous, complete set of orthogonal functions $\check{Q}(E)$, such that

$$\int\check{Q}(H;E)\check{Q}(H;E')dH = \left(\frac{\omega}{2\pi}\right)^2\delta(E - E').$$

The inverse Weyl transform [see Eqs. (VI.2), (VI.3), and (VI.7)] of Eq. (VII.8) does obviously not satisfy a second-order Schrödinger-like equation; only the complete $Q(\beta)$, Eq. (VII.6), is amenable to such a description. Indeed, by defining

$$\tilde{Q}(x,z) = \int dp\, Q(x,p;\beta)e^{2ipz/\alpha},$$

where α is a parameter to be determined below, and taking the Fourier transform of Eqs. (VII.3) and (VII.4) one obtains

$$\partial_x\tilde{Q} = -m\beta\omega^2x\tilde{Q}, \quad \partial_z\tilde{Q} = -\frac{4m}{\alpha^2\beta}z\tilde{Q},$$

the solution of which can be written as a product of the form

$$\tilde{Q}(x,z) = \varphi(x + z)\varphi(x - z)$$

only if

$$\alpha = \frac{2}{\beta\omega} = \frac{2kT}{\omega},$$

i.e., if the Weyl transformation is temperature dependent. In that case, φ is the solution of the eigenvalue equation

$$-\frac{2}{m\omega^2\beta^2}\varphi'' + \frac{m\omega^2}{2}x^2\varphi = \frac{1}{\beta}\varphi = \bar{E}\varphi \tag{VII.9}$$

with a unique eigenvalue; the normalized solution is

$$\varphi = \left(\frac{m\omega^2\beta}{2\pi}\right)^{1/4}\exp\left(-\frac{m\omega^2\beta}{4}x^2\right) \tag{VII.10}$$

and the mean energy is correctly given by $1/\beta = kT$. There is, of course, no sense in constructing a whole Hilbert space for only one vector.

So we see that, though the equilibrium solutions for the Brownian and the SED oscillators, Eqs. (VII.6) and (VII.8), respectively, have exactly the same x and p dependence, their temperature dependence is essentially different, and this gives rise, in its turn, to a difference in their mechanical behavior. The most striking difference is perhaps that related to the energy spectrum—continuous against discrete—but there exist other important ones. For instance, a Brownian oscillator can be frozen by reducing the temperature of the heat bath to zero; the SED oscillator never does come to a standstill.

We can now trace the origin of these striking physical and mathematical differences to the correlation function of the stochastic force, which by Eq. (II.3) determines the diffusion coefficient D^{pp}. In the Brownian case, the stochastic force is provided by an immensely large number of classical particles colliding with the oscillator: These obey the Maxwell–Botlzmann distribution when in equilibrium. It is the background radiation field obeying Planck's law which "quantizes" the electron: An electron immersed in a classical mechanical heat bath would, of course, behave essentially as a Brownian particle.

Below we will see that the smallness of the parameter $\tau = 2e^2/3mc^3$, the value of which is fixed by classical electrodynamics, also plays an important role in defining the quantum behavior of the SED system in nonequilibrium conditions. For an electron τ is of the order of 10^{-22} s, much smaller than the usual values of γ/ω^2 for a Brownian particle.

VIII. THERMODYNAMIC BEHAVIOR: BLOCH'S EQUATION

From the point of view of SED, the Schrödinger description is incomplete because it concerns only the isolated, temperature-independent components of the complete distribution, without regard to the way in which they combine in a real, physical situation. Neither the state of thermodynamic equilibrium nor the way in which this state is attained, are described by the Schrödinger equation. Hence, to describe the thermodynamic behavior it is necessary to resort to statistical mechanics; to study the time variation of a nonequilibrium distribution it is necessary to resort to quantum electrodynamics: Statistical mechanics and QED are brought in to restore the information which was dropped in order to retrieve the Schrödinger description in terms of pure states. We shall now recover this lost information contained in the full Fokker–Planck equation (II.10). In the next two sections we shall pay attention to the non-Liouvillian terms in the Fokker–Planck equation, which refer explicitly to the interaction of the oscillator with the radiation field, and study their effects on the time variation of what in QM are

stationary eigenstates. In this section we are concerned with the state of thermal equilibrium, which is assumed to be reached after a long time of such interaction between oscillator and field.

We recall from Sec. VI that the canonical density matrix corresponding to the thermal equilibrium distribution in phase space is

$$\rho(x_1,x_2)=\frac{1}{Z}\sum_k \varphi_k(x_1)\varphi_k(x_2)e^{-\beta E_k}. \tag{VIII.1}$$

Pure states (or Schrödinger) QM works with the φ_k as separate entities; as we have seen above, this is perfectly admissible also for SED as long as they are not supposed to represent on their own real physical states. The pure states represented by the amplitudes φ_k of the pseudostate Q_k are just about as physical as, e.g., the points of the reciprocal lattice of a crystal, or the Fourier components of a sound wave; and mathematically they are as useful, because they can be used as a basis to describe any real state.

Equation (VIII.1) is a hybrid expression, in the sense that it contains both statistical and quantum mechanical information. The latter has already been isolated and is fully contained (to zero order in τ) in the Schrödinger equation for the amplitudes φ_k. With the purpose of extracting now the statistical information and trying to preserve simplicity, we shall take the unnormalized density matrix

$$\rho_{\rm un}(x_1,x_2)\equiv Z\rho(x_1,x_2)$$

$$=\sum_k \varphi_k(x_1)\varphi_k(x_2)e^{-\beta E_k} \tag{VIII.2}$$

as is usual in quantum statistics, and derive it with respect to β,

$$\frac{\partial\rho_{\rm un}}{\partial\beta}=-\sum_k E_k\varphi_k(x_1)\varphi_k(x_2)e^{-\beta E_k}.$$

Introducing here the Schrödinger equation (VI.15) written in terms of the Hamiltonian operator, as usual,

$$\hat H_1\varphi_k(x_1)=\left(-\frac{\hbar^2}{2m}\partial_1^2+V_1\right)\varphi_k(x_1)=E_k\varphi_k(x_1),$$

we obtain

$$\frac{\partial\rho_{\rm un}}{\partial\beta}=-\sum_k \hat H_1\varphi_k(x_1)\varphi_k(x_2)e^{-\beta E_k}$$

$$=-\hat H_1\sum_k \varphi_k(x_1)\varphi_k(x_2)e^{-\beta E_k},$$

whence

$$\frac{\partial\rho_{\rm un}}{\partial\beta}=-\hat H_1\rho_{\rm un}(x_1,x_2), \tag{VIII.3}$$

The unnormalized canonical density matrix defined in Eq. (VIII.2) is thus found to satisfy the Bloch equation, with the usual quantum mechanical Hamiltonian operator. We must stress that this equation, which gives a complete statistical acount of the equilibrium state in phase space, has been derived here from the postulates of SED.

IX. TIME DEPENDENCE NEAR EQUILIBRIUM

Though up to now we have studied only the state of thermodynamic equilibrium, the same mathematical apparatus developed for arriving at the usual time independent quantum results can be used to study the nonstationary solutions of the Fokker–Planck equation near equilibrium. For this purpose we shall work in x_1x_2 space, in which the stationary solution was written in terms of the basis φ_k as

$$\rho(x_1,x_2)=\frac{1}{Z}\sum_k \varphi_k(x_1)\varphi_k(x_2)e^{-\beta E_k}. \tag{VIII.1}$$

Any ρ of the form

$$\rho(x_1,x_2)=\sum_{k,l} C_kC_l^*\varphi_k(x_2)\varphi_l(x_1) \tag{IX.1}$$

which does not coincide with Eq. (VIII.1) will, of course, not be an equilibrium solution and hence, the coefficients C_k will in general depend on time. In order to determine the time dependence of C_k we must demand $\rho(x_1,x_2)$ to be a solution of the complete Fokker–Planck equation (II.10) written in x_1x_2 space. As a first step we take the Fourier transform of this equation [see Eq. (VI.2)], to get

$$\frac{\partial\tilde Q}{\partial t}+(\tilde L_0+\tilde L_2)\tilde Q-2iz\tilde L_1\tilde Q=0,$$

where

$$\tilde L_0(x,z)=-\frac{i\hbar}{2m}\partial_z\partial_x+2im\omega^2xz,$$

$$\tilde L_1(x,z)=\frac{i}{2}\tau\omega^2\partial_z+2iD^{pp}z,$$

and

$$\tilde L_2(x,z)=2iD^{xp}z\partial_x.$$

To complete this transformation we introduce the variables x_1 and x_2, thus obtaining for $\rho(x_1,x_2)=\tilde Q(x,z)$,

$$\frac{\partial\rho}{\partial t}+(\tilde L_0+\tilde L_2)\rho-\frac{i}{\hbar}(x_1-x_2)\tilde L_1\rho=0, \tag{IX.2}$$

where

$$\tilde L_0=-\frac{i\hbar}{2m}(\partial_1^2-\partial_2^2)+\frac{im\omega^2}{2\hbar}(x_1^2-x_2^2), \tag{IX.3}$$

$$\tilde L_1=\frac{i\hbar\tau m\omega^2}{2}(\partial_1-\partial_2)+\frac{iD^{pp}}{\hbar}(x_1-x_2), \tag{IX.4}$$

and

$$\tilde L_2=\frac{i}{\hbar}D^{xp}(x_1-x_2)(\partial_1+\partial_2). \tag{IX.5}$$

We recall that both $\tilde L_1$ and $\tilde L_2$ are of order τ compared with $\tilde L_0$, and that τ is a very small parameter. Hence the time dependence of ρ is determined to lowest order in τ through the equation

$$\frac{\partial\rho}{\partial t}+\tilde L_0\rho=0, \tag{IX.6}$$

which can be written out explicitly as

$$\frac{\partial\rho}{\partial t}=\frac{i\hbar}{2m}(\partial_1^2-\partial_2^2)\rho-\frac{i}{\hbar}[V(x_1)-V(x_2)]\rho, \tag{IX.7}$$

where $V=\frac{1}{2}m\omega^2x^2$. This is just the von Neumann equation for the time evolution of the density matrix (in the x representation).

By introducing Eq. (IX.1) into Eq. (IX.7) we obtain

$$\begin{aligned}i\hbar\sum(\dot C_kC_l^*+C_k\dot C_l^*)\varphi_k(x_1)\varphi_l(x_2)&\\ =\sum C_kC_l^*\Big[-\frac{\hbar^2}{2m}(\partial_1^2-\partial_2^2)+V(x_1)-V(x_2)\Big]\varphi_k(x_1)\varphi_l(x_2)&\\ =\sum C_kC_l^*(E_k-E_l)\varphi_k(x_1)\varphi_k(x_2).&\end{aligned}$$

The second equality follows from the stationary Schrödinger equation

$$-\frac{\hbar^2}{2m}\partial^2\varphi_k+V\varphi_k=E_k\varphi_k. \tag{IX.8}$$

By multiplying by $\varphi_k(x_1)\varphi_l(x_2)$ and integrating over x_1 and x_2 we obtain

$$i\hbar(\dot C_kC_l^*+C_k\dot C_l^*)=C_kC_l^*(E_k-E_l)$$

whence

$$i\hbar\dot C_k=C_kE_k$$

the solution of which is

$$C_k(t)=C_k(0)\exp\Big(-\frac{iE_k}{\hbar}t\Big). \tag{IX.9}$$

We can therefore write for the density matrix to zero order in τ

$$\begin{aligned}\rho(x_1,x_2)&=\sum C_k(0)C_l^*(0)\exp\Big(-\frac{i}{\hbar}(E_k-E_l)t\Big)\varphi_k(x_1)\varphi_l(x_2)\\ &=\sum C_k(0)C_l^*(0)\psi_k(x_1)\psi_l^*(x_2),\end{aligned} \tag{IX.10}$$

where $\psi_k=\exp[-(iE_k/\hbar)t]\varphi_k$ satisfies the time dependent Schrödinger equation

$$i\hbar\frac{\partial\psi_k}{\partial t}=-\frac{\hbar^2}{2m}\frac{\partial^2\psi_k}{\partial x^2}+V\psi_k \tag{IX.11}$$

as follows directly from Eq. (IX.8). The fundamental equations of QM for the time evolution, namely, von Neumann's equation for the density matrix and Schrödinger's equation for the probability amplitude, are therefore deduced from SED, but only as approximate results to zero order in τ. They hold good when the system is close enough to equilibrium for the quantity $\hbar^{-1}|(x_1-x_2)\tilde L_1\rho|$ to be small compared with $|\tilde L_0\rho|$ (recall that in equilibrium, $\tilde L_1\rho=0$); under these circumstances, also $|\tilde L_2\rho|$ is small compared with $|\tilde L_0\rho|$.

X. THE RADIATIVE CORRECTIONS: LIFETIME AND LAMB'S SHIFT

The diffusion-dissipation term $\tilde L_1\rho$ is responsible for the attainment of equilibrium, as follows from the fact that it is the equation $\tilde L_1\rho=0$ which singles out the right equilibrium solution from the infinitely many solutions of $\tilde L_0\rho=0$. We therefore expect this term to give rise, near equilibrium, to a slow time variation of the population of the pseudostates described by ψ_k, which to zero order is constant in time,

$$\rho_k^0=|\psi_k|^2=|C_k(0)|^2\varphi_k^2=|C_k(0)|^2\tilde Q_k(x,0).$$

The term $\tilde L_1\rho$ should therefore fix the lifetime of the pseudostates Q_k.

The term $\tilde L_2\rho$, on the other hand, is a correction to the Liouvillian term, and we must therefore expect it to alter the energy values corresponding to the pseudostates Q_k, which to zero order have been found to be $E_k=\hbar\omega(k+\frac{1}{2})$.

In brief, we expect to obtain a real (coming from $\tilde L_1$) and an imaginary (from $\tilde L_2$) correction to the exponent of $C_k(t)$, both due to the interaction of the oscillator with the background radiation field. To calculate these radiative corrections, we use the method of variation of parameters and rewrite Eq. (IX.10) in the form

$$\rho=\sum c_kc_l^*e^{-i\omega_{kl}t}\varphi_k(1)\varphi_l(2), \tag{X.1}$$

where $\hbar\omega_{kl}=E_k-E_l$ and the c_k's are slowly varying functions of time, to be determined. By introducing this into Eq. (IX.2)—which is exact to first order in τ—and taking into account Eq. (IX.7), we are left with

$$\begin{aligned}\sum(c_k\dot c_l^*+\dot c_kc_l^*)e^{-i\omega_{kl}t}\varphi_k(x_1)\varphi_k(x_2)&\\ =\Big(-\tilde L_2+\frac{i}{\hbar}(x_1-x_2)\tilde L_1\Big)\sum c_kc_l^*e^{-i\omega_{kl}t}\varphi_k(x_1)\varphi_l(x_2)&\\ =-K\sum c_kc_l^*e^{-i\omega_{kl}t}\varphi_k(x_1)\varphi_l(x_2),&\end{aligned} \tag{X.2}$$

where the operator K is

$$K=K_1+K_2^*+K_{12}+K_{21}^* \tag{X.3a}$$

with

$$K_i=\frac{\tau\omega^2}{2}x_i\partial_i+\frac{D^{pp}}{\hbar^2}x_i^2+\frac{i}{\hbar}D^{xp}x_i\partial_i \tag{X.3b}$$

and

$$\begin{aligned}K_{ij}=&-\frac{\tau\omega^2}{2}x_i\partial_j-\frac{D^{pp}}{\hbar^2}x_ix_j\\ &+\frac{i}{\hbar}D^{xp}x_i\partial_j\quad(i,j=1,2,\quad i\neq j).\end{aligned} \tag{X.3c}$$

By multiplying Eq. (X.2) by $\varphi_{k'}(x_1)\varphi_{l'}(x_2)$ and integrating over x_1 and x_2 we get (interchanging k,l and k',l')

$$\begin{aligned}(c_k\dot c_l^*+\dot c_kc_l^*)e^{-i\omega_{kl}t}&\\ =-\sum c_{k'}c_{l'}^*e^{-i\omega_{k'l'}t}\langle kl|K|k'l'\rangle,&\end{aligned} \tag{X.4}$$

where the indices k,k' (l,l') refer to the variable x_1 (x_2).

In order to determine $c_k(t)$, we must separate it from c_l^*,

as was done in obtaining Eq. (IX.9). To perform this separation we note that the time variation of c_k must be due to operators acting on $\varphi_k(x_1)$. But the operators K_{ij} act on both $\varphi_k(x_1)$ and $\varphi_l(x_2)$; hence the separation is not so straightforward.

In a first-order calculation, however—which is all we can do, since K is already written to first order in τ—the problem becomes simple. Let us assume, to simplify the procedure, that the initial state of the system is described by

$$\rho(t=0)=\varphi_n(x_1)\varphi_n(x_2),$$

i.e.,

$$c_k(0)=\delta_{nk} \quad \text{and} \quad c_l^*(0)=\delta_{nl}. \tag{X.5}$$

Since near equilibrium the radiative corrections are small, we may safely assume that for short times, c_n remains of order 1, whereas the c_k ($k\neq n$) remain small; then to first order, Eq. (X.4) reduces to

$$(c_k\dot{c}_l^*+\dot{c}_kc_l^*)e^{-i\omega_{kl}t}=c_nc_n^*\langle kl\,|K\,|nn\rangle.$$

For $l=n$ we have

$$(c_k\dot{c}_n^*+\dot{c}_kc_n^*)e^{-i\omega_{kn}t}=-c_nc_n^*\langle kn|K\,|nn\rangle. \tag{X.6}$$

From Eqs. (X.3) and the properties of the φ functions for the harmonic oscillator, we have

$$\langle kn|K_{12}|nn\rangle=\langle kn|K_{21}^*|nn\rangle=0,$$

$$\langle kn|K_1|nn\rangle=\langle k\,|K_1|n\rangle,$$

and

$$\langle kn|K_2^*|nn\rangle=\delta_{kn}\langle n|K_2^*|n\rangle$$

whence Eq. (X.6) becomes

$$c_k\dot{c}_n^*+\dot{c}_kc_n^*$$

$$=-c_nc_n^*e^{i\omega_{kn}t}[\langle k\,|K_1|n\rangle+\delta_{kn}\langle n|K_2^*|n\rangle].$$

The action of K over the variables x_1 and x_2 has thus become separated: The first term within brackets gives rise to a variation of $\psi_k(x_1)$ and the second one is responsible for the change of $\psi_n^*(x_2)$ and hence of the time variation of c_k or c_n^*, respectively. We may therefore write

$$c_k\dot{c}_n^*=-c_nc_n^*\langle n|K_2^*|n\rangle\delta_{kn}$$

and

$$\dot{c}_kc_n^*=-c_nc_n^*e^{i\omega_{kn}t}\langle k\,|K_1|\,n\rangle.$$

Taking $k=n$ in the second of these equations we get

$$\dot{c}_n=-\langle n|K_1|n\rangle c_n$$

whence

$$c_n(t)=\exp(-\langle n|K_1|n\rangle t).$$

We shall write this result in the more convenient form

$$c_n(t)=\exp\left(-\tfrac{1}{2}\Gamma_nt-\frac{i\delta E_n}{\hbar}t\right), \tag{X.7}$$

where

$$\Gamma_n\equiv\frac{1}{T_n}=2\mathrm{Re}\langle n|K_1|n\rangle \tag{X.8}$$

is the decay constant (such that $\rho_n\sim e^{-\Gamma_nt}$) and

$$\delta E_n=\hbar\mathrm{Im}\langle n|K_1|n\rangle \tag{X.9}$$

is the correction to the energy E_n [see Eq. (IX.9)].

To calculate Γ_n we take the average of the real part of K_1 given by Eq. (X.3b),

$$\mathrm{Re}\langle n|K_1|n\rangle=-\frac{\tau\omega^2}{4}+\frac{D^{pp}}{m\omega\hbar}(n+\tfrac{1}{2}),$$

and substitute Eq. (II.9) for D^{pp}, thus obtaining

$$\Gamma_n=\frac{\tau\omega}{\hbar}\left(\frac{1+\epsilon}{1-\epsilon}E_n-E_0\right), \tag{X.10}$$

where $\epsilon=e^{-\beta\hbar\omega}=e^{-\hbar\omega/kT}$.

According to this result, the radiative decay of excited states is accelerated by raising the temperature of the background field. It is interesting to note that at $T>0$, we have $\Gamma_0>0$, which shows that the ground state is not a stable state at positive temperatures. At low temperatures Γ_0 decreases exponentially towards zero: $\Gamma_0\simeq\tau\omega^2e^{-\hbar\omega/kT}$, whereas at very high temperatures it grows linearly with T: $\Gamma_0\simeq(\tau\omega/2\hbar)kT$.

At $T=0$, Eq. (X.10) reduces to

$$\Gamma_n=\frac{\tau\omega}{\hbar}(E_n-E_0)=\tau\omega^2n=n\Gamma_1 \tag{X.11}$$

or, since for a harmonic oscillator

$$\sum_{k<n}|\langle n|x|k\rangle|^2=|\langle n|x|n-1\rangle|^2=\frac{\hbar n}{2m\omega},$$

we may rewrite this expression as

$$\Gamma_n=\frac{4e^2\omega^3}{3\hbar c^3}\sum_{k<n}|\langle n|x|k\rangle|^2. \tag{X.12}$$

This result coincides with the first-order prediction of quantum electrodynamics for the radiative decay of excited states in the dipole approximation.[28]

To calculate δE_n we take the average value of the imaginary part of K_1. For this purpose we must first calculate D^{xp} from Eq. (II.2); we recall that in the Markovian approximation this equation reduces to

$$D^{xp}=e^2\int_0^t dt'\,\langle E^m(t)E^m(t')\rangle\frac{\partial x(t)}{\partial\bar{p}(t';t)}. \tag{X.13}$$

From the solution of the nonstochastic equation of motion for the oscillator we get

$$\frac{\partial x(t)}{\partial\bar{p}(t';t)}=\frac{1}{m\omega}\exp[-\tfrac{1}{2}\tau\omega^2(t-t')]\sin\omega(t-t')$$

which introduced into Eq. (X.13), together with Eqs. (II.4) to (II.8), yields

$$D^{xp}=\frac{2\pi e^2}{m\omega}\int_{-\infty}^{\infty}d\omega'\,\tfrac{1}{3}\rho_T^m(\omega')\int_0^t dt'$$

$$\times\exp[i\omega'(t-t')-\tfrac{1}{2}\tau\omega^2(t-t')]\sin\omega(t-t')$$

$$=\frac{\tau\hbar}{\pi}\int_0^{\infty}d\omega'\,\frac{\omega'^3}{1+\tau^2\omega'^2}$$

$$\times\frac{\omega^2-\omega'^2}{(\omega^2-\omega'^2)^2+\tau^2\omega^4\omega'^2}\,\frac{1+e^{-\hbar\omega'\beta}}{1-e^{-\hbar\omega'\beta}}, \tag{X.14}$$

where the factor $\frac{1}{3}$ was introduced to take account of the one-dimensional character of our problem.

The above integral diverges logarithmically: We are meeting here the well-known divergence encountered by QED in the calculation of the Lamb shift. The origin of this divergence lies in the spectral density assumed for the zero-point field, $\rho_0(\omega)\sim\omega^3$, which being nonintegrable implies an infinite energy density for the vacuum. We may in principle make the integral convergent by introducing a cutoff to account for the relativistic effects left out in our simplified treatment (pair creation, finite velocity of the particle, etc.); but since we lack a reasonable theory to evaluate these effects, we would rather resort to the customary renormalization procedure typical of QED calculations. First note that the integral is divergent even for the free particle, i.e., in the limit $\omega=0$, which implies an (infinite) correction to the self-energy (to the mass, in our nonrelativistic version) of the particle. Since this is not a contribution to the Lamb shift proper, because it must be already included in the experimental mass value, we subtract it and thus obtain

$$D_L^{xp}=\frac{\tau\hbar}{\pi}\int_0^\infty d\omega'\frac{\omega'^3}{1+\tau^2\omega'^2}\left[\frac{\omega^2-\omega'^2}{(\omega^2-\omega'^2)^2+\tau^2\omega^4\omega'^2}-\frac{-\omega'^2}{\omega'^4+\tau^2\omega^4\omega'^2}\right]\frac{1+\epsilon'}{1-\epsilon'}$$

$$=\frac{\tau\hbar}{2\pi}\omega^2\int_0^\infty d\omega'\frac{\omega'^3}{1+\tau^2\omega'^2}\times\frac{\omega^2(1+\tau^2\omega^2)-\omega'^2}{(\omega'^2+\tau^2\omega^4)[(\omega^2-\omega'^2)^2+\tau^2\omega^4\omega'^2]}\,\frac{1+\epsilon'}{1-\epsilon'},$$

which is already convergent. In evaluating this integral we take account of the resonance at $\omega'=\omega$, neglect the unimportant terms of order τ^2 and simultaneously introduce a cutoff at the frequency ω_c, which we do not specify further but may estimate to be of the order of $2mc^2$, the onset of pair creation effects. We thus obtain

$$D_L^{xp}=-\frac{\tau\hbar}{2\pi}\omega^2\left(\ln\frac{\omega_c}{\omega}\right)\frac{1+\epsilon}{1-\epsilon}$$

$$=\frac{\tau\hbar}{2\pi m}F'\left(\ln\frac{\omega_c}{\omega}\right)\frac{1+\epsilon}{1-\epsilon}$$

$$=-\frac{\tau\hbar}{2\pi m}\left(\ln\frac{\omega_c}{\omega}\right)\frac{1+\epsilon}{1-\epsilon}\cdot V'', \tag{X.15}$$

where $\epsilon=\exp(-\beta\hbar\omega)$ is to be evaluated at the frequency of the oscillator. We can then write for the energy correction

$$\delta_L E_n\equiv\delta E_n-\delta E_n(\omega=0)=\langle n|D_L^{xp}x\partial_x|n\rangle$$

$$=-\frac{\tau\hbar}{2\pi m}\left(\ln\frac{\omega_c}{\omega}\right)\langle n|V''x\partial_x|n\rangle\frac{1+\epsilon}{1-\epsilon}$$

or

$$\delta_L E_n=\frac{\tau\hbar}{2\pi m}\left(\ln\frac{\omega_c}{\omega}\right)\langle n|\nabla^2V|n\rangle\frac{1+e^{-\beta\hbar\omega}}{1-e^{-\beta\hbar\omega}}. \tag{X.16}$$

We have written ∇^2V instead of the one-dimensional expression V'', to give to the final result a form which is correct regardless of the number of dimensions. Equation (X.16) shows that each level of the harmonic oscillator suffers a shift due to the interaction with the background field,[29] and that this shift grows with temperature. Since the level shift is the same for all pseudostates, because $\langle n|\nabla^2V|n\rangle=-3m\omega^2$ (in three dimensions) is independent of n, the Lamb shift does not alter the radiation spectrum.

At $T=0$, Eq. (X.16) reduces to

$$\delta_L E_n=\frac{\tau\hbar}{2\pi m}\left(\ln\frac{\omega_c}{\omega}\right)\langle n|\nabla^2V|n\rangle. \tag{X.17}$$

On comparing this result with the renormalized result given by QED for any bounded problem[30]:

$$\delta E_n=\frac{\tau\hbar}{2\pi m}\left(\ln\frac{E_c}{\langle\Delta E\rangle}\right)\langle n|\nabla^2V|n\rangle, \tag{X.18}$$

where E_c is the cutoff energy and $\langle\Delta E\rangle$ is an average transition energy, we find a most satisfactory agreement. On the other side, at very high temperatures ($\beta\to0$) Eq. (X.16) predicts any energy correction proportional to temperature.

XI. CONCLUDING REMARKS

In the preceding sections we have seen that all fundamental equations of quantum theory, namely, Schrödinger's equation for ψ, von Neumann's equation for the time evolution of the density matrix, Bloch's equation for its temperature dependence; the relationship between these equations, together with Wigner's distribution and, on top of it all, the corrections predicted by (nonrelativistic) QED, can be obtained in a quite simple and intuitive form from the postulates of SED for the case of the harmonic oscillator. This has been accomplished without having to adjust any free parameter, since the two parameters contained in the theory, namely τ and $\hbar$, are fixed, the first one by classical electrodynamics and the second one by the zero-point energy of the background field. The theory predicts results in remarkable agreement with quantum theory, even the structure of the radiation spectrum and the lifetimes of excited states, despite the necessary existence of discrepancies between the two theories; we are in front of a theory which differs from the usual one in both its physical content and its extension, but which reproduces all experimentally confirmed results and gives a well-defined physical interpretation of the formalism of quantum theory. It is most remarkable that a seemingly classical theory which starts with the equation of motion

$$m\ddot{x} = F(x) + m\tau\dddot{x} + eE(t) \qquad \text{(XI.1)}$$

with an appropriate stochastic field $E(t)$, in fact describes a quantum system, including the QED corrections.

As it is, however, the theory is neither finished nor free from difficulties. Perhaps the two most important points which remain to be solved for the theory to be considered a serious alternative to the usual treatment of the oscillator—not to mention the extension to more general atomic system—are those related with the analysis of the density in terms of pseudostates and with the divergences caused by the correlation function of the field. Let us conclude this paper by discussing them briefly.

In connection with the development of $Q(\beta)$ in terms of $\check{Q}(E_k)$, our problem is to find the physical justification for the decomposition in terms of Boltzmann factors; as pointed out above, the whole quantum mechanical description in terms of pseudostates (or excited states, in the usual language) rests upon this decomposition, allowing for the introduction of the formal energy parameter E which happens to be restricted to discrete values. Only after assigning a physical meaning to this parameter and justifying Eq. (V.1) on first principles will we be able to say why one must use the usual formalism of quantum theory instead of the more direct one of SED. At present, we know that SED contains quantum theory in a well-defined limit, but we still ignore why.

As to the second point, we already remarked that the divergences occurring in the calculation of D^{xp} (leading to the Lamb shift) come from the structure adopted for $\rho_0(\omega)$, Eq. (II.6), which implies an infinite energy density for the field. Since this spectral density is derived from first principles (essentially from the demand of Lorentz covariance for the field[25]), it is not a simple matter to alter it; in fact, this problem remains totally open. However, the advantage of SED over more conventional treatments of the Lamb shift is twofold: First there is the conceptual advantage of having clearly identified the source of the divergence; second we have obtained an UV divergence which is less severe than in usual nonrelativistic QED; in fact, in SED a single mass renormalization suffices to yield a finite (though somewhat large) result. The reason for this lies in the extra ω^2-dependent denominator introduced in going from ρ_0 to ρ_0^m, Eqs. (II.6) and (II.8), when eliminating the runaway solutions of the equation of motion. Thus SED seems to open a new path for the study of one of the most important problems of quantum theory, namely, its divergences.

ACKNOWLEDGMENTS

We acknowledge valuable comments and suggestions from T.A. Brody, P. Claverie, and S. Diner.

[1]T.W. Marshall, Proc. R. Soc. A **276**, 475 (1963).
[2]N.S. Kalitsin, JETP **25**, 407 (1953); E. Adirovich and M. Podgoretskii, JETP **26**, 150 (1954); P. Braffort, M. Spighel, and C. Tzara, C.R. Acad. Sci. (Paris) **239**, 157 (1954); P. Braffort and C. Tzara, C.R. Acad. Sci. (Paris) **239**, 1779 (1954); A.A. Sokolov and V.S. Tumanov, JETP **30**, 802 (1956). Baraffort *et al.* recognized the need of adding a stochastic component to the motion of any charged particle, from the absorber theory of J.A. Wheeler and R.P. Feynman, Rev. Mod. Phys. **17**, 157 (1945).
[3]T.A. Welton, Phys. Rev. **74**, 1157 (1948). The paper inspired also some earlier theories, different from SED but coinciding with it in the recognition of the zero point field as a real, physical field giving rise to characteristic phenomena. See, e.g., D. Park and H.T. Epstein, Am. J. Phys. **17**, 301 (1949).
[4]V.F. Weisskopf, Rev. Mod. Phys. **21**, 305 (1949).
[5]P. Braffort, M. Surdin, and A. Taroni, C.R. Acad. Sci. Paris B **261**, 4339 (1965); P. Braffort and A. Taroni, C.R. Acad. Sci. Paris B **264**, 1437 (1967); P. Braffort, C.R. Acad. Sci. Paris B **270**, 12 (1970).
[6]T.W. Marshall, Proc. Camb. Philos. Soc. **61**, 537 (1965), Nuovo Cimento **38**, 206 (1965).; L.L. Henry and T.W. Marshall, Nuovo Cimento **41**, 188 (1966), Isv. Vyssh. Uchebn. Zaved. Fiz. **12**, 34 (1968).
[7]T.H. Boyer, Phys. Rev. D **11**, 790 (1975) and references to his previous work therein; Phys. Rev. D **11**, 809 (1975); **13**, 2832 (1976).
[8]O. Theimer, Phys. Rev. D **4**, 1597 (1971); O. Theimer and P.R. Peterson, Phys. Rev. D **10**, 3962 (1974), Phys. Rev. A **16**, 2055 (1977).
[9]M. Surdin, C.R. Acad. Sci. B Paris **270**, 193 (1970); Int. J. Theor. Phys. **4**, 117 (1971); **9**, 185 (1974).
[10]E. Santos, An. R. Soc. Esp. Fis. Quim. **64**, 317 (1968); Lett. Nuovo Cimento **4**, 497 (1972); Nuovo Cimento B **22**, 201 (1974); J. Math. Phys. **15**, 1954 (1974); An. Fis. **71**, 329 (1975).
[11]E. Santos, Nuovo Cimento B **19**, 57 (1974).
[12]P. Claverie and S. Diner, Ann. Fond. L. de Broglie **1**, 73 (1976).
[13]P. Claverie and S. Diner, in Int. J. Quantum Chem. **2**, Suppl. 1, 41 (1977).
[14]L. de la Peña and A.M. Cetto, Phys. Lett. A **47**, 183 (1974), Rev. Mex. Fis. **25**, 1 (1976).
[15]L. de la Peña and A.M. Cetto, J. Math. Phys. **18**, 1612 (1977); A.M. Cetto and L. de la Peña, Ann. Fond. L. de Broglie **3**, 15 (1978); L. de la Peña and A.M. Cetto, Int. J. Quantum Chem. **12**, Suppl. 1, 23 (1977).
[16]L. de la Peña and A.M. Cetto, Phys. Lett. A **62**, 389 (1977), Found. Phys. **8**, 191 (1978).
[17]A review of Boyer is cited in Ref. 7 above. A complementary review paper by A.M. Cetto, P. Claverie, S. Diner, L. de la Peña, and E. Santos, is under preparation. Some aspects of SED are incidentally reviewed by Milonni in Ref. 18.
[18]P.W. Milonni, Phys. Rep. **25**, 1 (1976).
[19]A derivation of the Schrödinger equation from SED which applies only for the harmonic oscillator, has been given also by Surdin in the 1971 paper cited in Ref. 9 above.
[20]In Refs. 15 and 16, a small extra force term $\hat{f}$ appears in Eq. (II.1). It may be shown, however, that this term is cancelled exactly by another contribution which was mistakenly omitted. We are grateful to P. Claverie for having called our attention to this point.
[21]H. Haken, Rev. Mod. Phys. **47**, 67 (1975).
[22a]Wigner's distribution is constructed and discussed in S. de Groot, *La transformation de Weyl et la fonction de Wigner: une forme alternative de la Mécanique quantique* (Presses Universitaires de Montréal, Canada, 1974). A brief account of its application to the harmonic oscillator may be found in Ref. 22b, Sec. 2.7. Some aspects of the relation between SED and the Wigner distribution in the more general case can be found in Refs. 12, 13, 15, and 16.
[22b]R.P. Feynman, *Statistical Mechanics* (Benjamin, New York, 1972).
[23]The point of view expressed in the paper is not precisely popular among physicists; for a recent discussion of the subject see, e.g., Ch. P. Enz, in *Physical Reality and Mathematical Description*, edited by J. Mehra (Reidel, Dordrecht, 1974).
[24]A similar method was previously employed in an entirely different context by P. Smet and J. Tillieu, J. Phys. Radium **23**, 299 (1962).
[25]T.H. Boyer, Phys. Rev. **182**, 1374 (1969); **1**, 1526 (1970).
[26]L. Cohen, J. Math. Phys. **7**, 781 (1966).
[27]See, e.g., the standard reference: the papers by M. Ch. Wang and G.E. Uhlenbeck, *et al.* in *Selected Papers on Noise and Stochastic Processes*, edited by N. Wax (Dover, New York, 1954).
[28]In the derivation of Eq. (II.1) the long-wavelength approximation for the electric field, $E(x,t) = E(t)$, was used;[15] since this implies $e^{ikx} \simeq 1$, it is equivalent to the dipole approximation in the evaluation of T_n.
[29]The calculation of the Lamb shift for the ground states of the harmonic oscillator has received attention since the earliest works on SED; see, e.g., Ref. 2. More recent calculations can be found in Refs. 11, 14, and 16. In the latter it is shown that the Lamb shift is a direct result of the coupling of the electric dipole of the system to the background field.
[30]See, e.g., J.J. Sakurai, *Advanced Quantum Mechanics* (Addison-Wesley, Reading, Mass., 1967), Sec 2.8.

Szegö's theorem for Hankel determinants

J. S. Geronimo

Georgia Institute of Technology, School of Mathematics, Atlanta, Georgia 30332
(Received 22 June 1978)

An analog of Szegö's formula for asymptotic Töeplitz determinants is proved for Hankel determinants. The proof uses a set of recurrence formulas developed for polynomials orthogonal on a finite segment of the real line and some properties of the Jost function associated with those polynomials. The techniques of inverse scattering theory are used to calculate the correction terms to the asymptotic formula. The results are valid for weight functions that have a finite number of jump points and an absolutely continuous part that is unbounded.

I. INTRODUCTION

In this article an analog of Szegö's theorem on Töeplitz determinants is proved for Hankel determinants. The theorem on Töeplitz determinants was first proved by Szegö[1] in 1952. Since then much work has been done to weaken the assumptions needed to prove the theorem and to prove it from different points of view.[2-8] Here we wish to prove the theorem for Hankel determinants. This was first done by Hirshman[3] using Banach algebra techniques. Our proof is much different and allows one to extend the theorem to spectral functions that are not absolutely continuous. The proof closely follows the one given in Ref. 8. The theorem is proved using a new set of recurrence formulas developed for polynomials orthogonal on a finite segment of the real line and some properties of the Jost function. Using the techniques of inverse scattering theory, one is easily able to calculate the correction terms to the asymptotic formula.

We proceed as follows; First (Sec. II) the theory of orthogonal polynomials is briefly reviewed and the equations of inverse scattering theory are introduced. Then, in Sec. III, a special case of the theorem is stated and proved. Following this (Sec. IV) we find the correction terms. Finally in Sec. V the theory is extended to spectral functions that are not absolutely continuous.

II. Preliminaries

Suppose $\rho(\lambda)$[9] is a nondecreasing function with infinitely many points of increase on a segment $[a,b]$ of the real line such that

$$s_n=\int_a^b \lambda^n d\rho(\lambda) \tag{II.1}$$

exists for all n. We are to find polynomials $p(\lambda,n)$, such that

(i) $p(\lambda,n)$ is a polynomial of precise degree n in which the coefficient of λ^n is positive,

(ii) $\int_a^b p(\lambda,n)p(\lambda,m)d\rho(\lambda)=\delta(n,m),\ m,n=0,1,2,\cdots.$

Using standard orthogonalization procedures, one finds

$$p(\lambda,n)=[H_{n-1}H_n]^{-1/2}$$

$$\times\begin{vmatrix} s_0 & s_1 & s_2 & \cdot & \cdot & \cdot & s_n \\ s_1 & s_2 & s_3 & \cdot & \cdot & \cdot & s_{n+1} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ s_{n-1} & s_n & \cdot & \cdot & \cdot & \cdot & s_{2n-1} \\ 1 & \lambda & & & & & \lambda^n \end{vmatrix}, \quad n\geqslant 0, \tag{II.2}$$

where

$$H_n=\begin{vmatrix} s_0 & s_1 & s_2 & \cdot & \cdot & \cdot & s_n \\ s_1 & s_2 & \cdot & \cdot & \cdot & \cdot & s_{n+1} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ s_n & s_{n+1} & \cdot & \cdot & \cdot & \cdot & s_{2n} \end{vmatrix}, \quad n\geqslant 0. \tag{II.3}$$

[Here $H_{-1}\equiv 1$; thus $p(\lambda,0)=H_0^{1/2}=s_0^{-1/2}$.] The coefficient of λ^n in $p(\ ,n)$ can be determined from Eq. (II.2) and is

$$K(n)=\left(\frac{H_{n-1}}{H_n}\right)^{1/2}, \quad n=0,1,\cdots. \tag{II.4}$$

In Ref. 10 it was shown that the polynomials satisfy the following two term recurrence formulas:

$$p(\lambda,n)=\frac{a(\infty)}{a(n)}\{[Z-B(n-1)]p(\lambda,n-1) + \frac{\psi(Z,n-1)}{Z}\}, \quad n=1,2,\cdots, \tag{II.5}$$

and

$$\psi(Z,n)=\frac{a(\infty)}{a(n)}\left\{\frac{\psi(Z,n-1)}{Z} + \left[\left(1-\frac{a(n)^2}{a(\infty)^2}\right)Z-B(n-1)\right]p(\lambda,n-1)\right\}, \quad n=1,2,\cdots, \tag{II.6}$$

where

$$a(n)=\frac{K(n-1)}{K(n)}, \tag{II.7}$$

 0022-2488/79/030484-08$01.00

$$b(n)=\int_a^b \lambda p(\lambda,n)^2 d\rho(\lambda), \tag{II.8}$$

$$B(n)=\frac{b(n)-b(\infty)}{a(\infty)}, \tag{II.9}$$

and

$$\lambda=a(\infty)(Z+1/Z)+b(\infty). \tag{II.10}$$

Here it is assumed that

$$\lim_{n\to\infty} a(n)=a(\infty)>0, \quad \lim_{n\to\infty} b(n)=b(\infty), \quad b(n) \text{ real}, \tag{II.11}$$

and

$$\psi(Z,0)=p(\lambda,0)=K(0)>0. \tag{II.12}$$

Considering $p(\lambda,n)$ and $\psi(Z,n)$ as two components of a function Φ defined by

$$\Phi(Z,n)=\begin{pmatrix} p(\lambda,n) \\ \psi(Z,n) \end{pmatrix}. \tag{II.13}$$

Equations (II.5) and (II.6) can be condensed to

$$\Phi(Z,n)=C(n)\Phi(Z,n-1) \tag{II.14}$$

where

$$C(n)=\frac{a(\infty)}{a(n)}\begin{bmatrix} Z-B(n-1) & 1/Z \\ [1-a(n)^2/a(\infty)^2]Z-B(n-1) & 1/Z \end{bmatrix}. \tag{II.15}$$

Two other useful solutions of Eq. (II.14) are

$$\Phi_+(Z,n)=\begin{pmatrix} p_+(Z,n) \\ \psi_+(Z,n) \end{pmatrix}$$

and

$$\Phi_-(Z,n)=\begin{pmatrix} p_-(Z,n) \\ \psi_+(Z,n) \end{pmatrix} \tag{II.17}$$

where the components of these vectors satisfy the following boundary conditions:

$$\lim_{n\to\infty} |p_\pm(Z,n)-Z^{\pm n}|=0, \quad |Z|\leqslant 1, \tag{II.18}$$

$$\lim_{n\to\infty} |\psi_+(Z,n)|=0, \quad |Z|\leqslant 1, \tag{II.19}$$

$$\lim_{n\to\infty} |\psi_-(Z,n)-(1-Z^2)Z^{-n}|=0, \quad |Z|\geqslant 1, \tag{II.20}$$

Notice from the boundary conditions and the recurrence formulas that

$$\overline{p_+(Z,n)}=p_+(1/Z,n)=p_-(Z,n), |Z|=1. \tag{II.21}$$

It can be shown[10] that $\Phi_+(Z,n)$ and $\Phi_-(Z,n)$ are linearly independent for $|Z|=1, Z\neq\pm 1$. Therefore,

$$\Phi(Z,n)=\frac{K(0)}{a(\infty)(Z-1/Z)} \times[f_-(Z)\Phi_+(Z,n)-f_+(Z)\Phi_-(Z,n)],^{10}$$

$$|Z|=1, \quad Z\neq\pm 1. \tag{II.22}$$

Here

$$f_-(Z)=\overline{f_+(Z)}=f_+(1/Z), \quad |Z|=1, \tag{II.23}$$

and

$$f_+(Z)=\frac{a(\infty)}{K(0)Z}\lim_{n\to\infty} Z^n\psi(Z,n)=\frac{a(\infty)}{K(0)Z}\lim_{n\to\infty}\psi^*(Z,n) \tag{II.24}$$

with

$$\psi^*(Z,n)=Z^n\psi(Z,n). \tag{II.25}$$

$f_+(Z)$ is the discrete analog of the Jost function.

In order to proceed further, it is convenient at this point to introduce the techniques of Banach algebras. Thus, let A denote the class of functions integrable on $-\pi\leqslant\theta\leqslant\pi$ such that if g is an element of A then

$$g(\theta)\approx\sum_{K=-\infty}^{\infty} g(K)e^{iK\theta} \tag{II.26}$$

with

$$||g||=\sum_{K=-\infty}^{\infty}|g(K)|<\infty. \tag{II.27}$$

Let A^+ and A^- denote those functions in A of the form

$$g(\theta)\approx\sum_{K=0}^{\infty} g(K)e^{iK\theta} \tag{II.28}$$

and

$$h(\theta)\approx\sum_{K=-\infty}^{0} h(K)e^{iK\theta} \tag{II.29}$$

respectively.

Let $\|g\|$ be the norm on A, A^+, and A^-, then A, A^+, and A^- are Banach algebras.[11]

If

$$\sum_{n=1}^{\infty} n\left[\left|1-\frac{a(n)^2}{a(\infty)^2}\right|+|B(n-1)|\right]<\infty, \tag{II.30}$$

then the following properties can be proved[10]

(1) $Zf_+(Z)$ and $p_+(Z,n)$ are analytic inside the unit circle and continuous on it,

(2) $Zf_+(Z)$ and $p_+(Z,n)$ are elements of A^+,

(3) If $f_+(Z)=0$ for $|Z|<1$, then the zeros are

(a) simple,

(b) real,

(c) finite in number. Finally

(4) If $f_+(Z)=0$ for $|Z|=1$ then the zeros

(a) must occur at $Z=+1$ and/or -1,

(b) $Zf_+(Z)/d(Z)\in A^+$.

Here $d(Z)$ is

(a) equal to 1 if $f_+(Z)\neq 0$ for $|Z|=1$,

(b) equal to $1-Z$ if $f_+(Z)=0$ for $Z=+1$, (II.31)

(c) equal to $1+Z$ if $f_+(Z)=0$ at $Z=-1$,

(d) equal to $1-Z^2$ if $f_+(Z)=0$ at $Z=-1$ and $Z=+1$.

It follows from the above conditions that the spectral function with respect to which the polynomials are orthogonal can be written as[10]

$$d\rho(\lambda)\approx\begin{cases}\sigma(\lambda)d\lambda, & b(\infty)-2a(\infty)\leqslant\lambda\leqslant b(\infty)+2a(\infty),\\ \sum_{i=1}^{n}\rho_i\delta(\lambda-\lambda_i)d\lambda, & \lambda \text{ not as above}\end{cases}$$

with

$$\sigma(\lambda)d\lambda=\frac{a(\infty)\sin\theta d\lambda}{\pi K(0)^2|f_+|^2} \tag{II.32}$$

and

$$\rho_i=\frac{p_+(Z_i,0)}{K(0)^2 f'_+(\lambda_i)}. \tag{II.33}$$

Here λ_i denote the roots of $f_+(Z)$ for $|Z|<1$. It follows from property 4*b*, Eq. (II.33), and the Wiener–Levy theorem that[10]

$$\ln\left(\frac{\sigma(\lambda)|d(Z)|^2}{\sin\theta}\right)\in A. \tag{II.34}$$

Often one is given the spectral function instead of the Jost function. Using a modification of the Poisson-integral formulas, one can construct the Jost function from the spectral function in the following manner[10]

$$f_+(Z)=d(Z)\frac{\pi_+(Z_i-Z)\pi_-(Z-Z_i)}{Z\prod_{i=1}^{n}(1-ZZ_i)}$$

$$\times\exp\left[-\frac{1}{4\pi}\int_{-\pi}^{\pi}\ln\left(\frac{\sigma(\lambda')|d(z')|^2K(0)^2}{a(\infty)\sin\theta'}\right)\times\left(\frac{\exp i\theta'+Z}{\exp i\theta'-Z}\right)d\theta'\right], \tag{II.35}$$

$$Z'=e^{i\theta'},\quad |Z|<1,$$

$$\lambda'=a(\infty)(Z'+1/Z')+b(\infty),$$

where the Z_i's are the jump points of the spectral function and are determined from Eq. (II.12).[12] $\pi_\pm$ means the product over the roots of $f_+(Z)$ for $Z_i\gtrless 0$ respectively.

From Eqs. (II.14), (II.18), and (II.19)

$$p_+(Z,n)=\sum_{i=n}^{\infty}A(n,i)Z^i. \tag{II.36}$$

If one defines[13]

$$\omega(n,m)=1/2\pi\oint_{|Z|=1}[(1-s(Z))]Z^{n+m}\frac{dZ}{Z}+\sum_{i=1}^{n}\rho_i Z_i^{n+m}, \tag{II.37}$$

$$s(Z)=\frac{f_-(Z)}{f_+(Z)}=\frac{\overline{f_+(Z)}}{f_+(Z)},\quad |Z|=1, \tag{II.38}$$

where the $Z_{i's}$ are the zeros of the $f_+(Z)$ inside the unit circle, then it can be shown that[14,15]

$$A(n,n)=[\det(1+\omega)_{n+1}^{\infty}/\det(1+\omega)_n^{\infty}]^{1/2}. \tag{II.39}$$

Here

$$\det[1+\omega]_{n+1}^{\infty}=\begin{vmatrix}1+\omega(n+1,n+1) & \omega(n+1,n+2) & \omega(n+1,n+3) & \cdots\\ \omega(n+2,n+1) & 1+\omega(n+2,n+2) & \omega(n+2,n+3) & \cdots\\ \omega(n+3,n+1) & \omega(n+3,n+3) & 1+\omega(n+3,n+3) & \cdots\\ \cdot & \cdot & \cdot & \cdots\end{vmatrix} \tag{II.40}$$

Now from the recurrence formula (II.14) and boundary conditions (II.18)[10,16]

$$a(n)/a(\infty)=A(n,n/A(n-1,n-1), \tag{II.41}$$

Another formula we will need connects two different sets of orthogonal polynomials. More precisely, given two infinite systems of orthogonal polynomials $\{p^0(\lambda,n)\}$ and $\{p(\lambda,n)\}$ with associated spectral functions $d\rho^0(\lambda)$ and $d\rho(\lambda)$ respectively, and writing

$$p(\lambda,n)=\sum_{i=0}^{n}K(n,i)p^0(\lambda,i), \tag{II.42}$$

what is $K(n,n)$? Taking

$$q(n,m)=\int p^0(\lambda,n)p^0(\lambda,m)[d\rho(\lambda)-d\rho^0(\lambda)], \tag{II.43}$$

one finds[9,16]

$$K(n,n)=\{\det[1+q]_0^{n-1}/\det[1+q]_0^n\}^{1/2}, \tag{II.44}$$

where

$$\det[1+q]_0^n=\begin{vmatrix} 1+q(0,0) & q(0,1) & (0,2) & \cdots & q(0,n) \\ q(1,0) & 1+q(1,1) & q(1,2) & \cdots & q(1,n) \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ q(n,0) & q(n,1) & \cdot & \cdots & 1+q(n,n) \end{vmatrix}. \tag{II.45}$$

From the recurrence formulas and the orthogonality property, $a(n)$ is related to $a^0(n)$ in the following manner:

$$a^0(n)/a(n)=K(n+1,n+1)/K(n,n), \tag{II.46}$$

where it is assumed that

$$a^0(\infty)=a(\infty). \tag{II.47}$$

III. STATEMENT AND PROOF OF THE THEOREM

In this section we prove Szegö's theorem on Hankel determinants. Here it will be assumed that $Zf_+(Z)\neq 0$ for $|Z|\leqslant 1$. This is equivalent to assuming that the spectral function with respect to which the polynomials are orthogonal has only a bounded absolutely continuous part [see Eqs. (II.32) and (II.33)]. This assumption will be removed in Sec. V.

Theorem: If

$$(1)\quad \sum_{n=1}^{\infty} n\left\{\left|1-\frac{a(n)^2}{a(\infty)^2}\right|+|B(n-1)|\right\}<\infty \tag{III.1}$$

and

$$(2)\quad Zf_+(Z)\neq 0 \quad \text{for} \quad |Z|\leqslant 1, \tag{III.2}$$

then

$$\sum_{n=1}^{\infty} 2n\ln\left(\frac{a(\infty)^2}{a(n)^2}\right) = \ln\left(\frac{f_+(1)[-f_+(-1)]}{C^2}\right)+\sum_{m=1}^{\infty} m|\gamma(m)|^2, \tag{III.3}$$

where

$$\gamma(m)=(1/2\pi)\int_{-\pi}^{\pi}\ln[Zf_+(Z)]e^{-im\theta}d\theta, \quad m\geqslant 1, Z=e^{i\theta}, \tag{III.4}$$

and

$$C=a(\infty)\prod_{i=1}^{\infty}\frac{a(\infty)}{a(i)} =\exp-(1/4\pi)\int_{-\pi}^{\pi}\ln\left(\frac{\pi K(0)^2\sigma(\lambda')}{a(\infty)\sin\theta'}\right)d\theta'. \tag{III.5}$$

A. Remarks

From the properties of $Zf_+(Z)$ listed following Eq. (II.30) and the assumption (2), we find

$$-(1/2)\int_{-\pi}^{\pi}\ln[(Zf_+(Z)]e^{im\theta}d\theta=0, \quad m\geqslant 1.$$

Taking the complex conjugate of the above equation, adding it to Eq. (III.4), and then using Eq. (II.32) gives an expression for $\gamma(m)$ in terms of the spectral function

$$\gamma(m)=-1/2\pi\int_{-\pi}^{\pi}\ln\left(\frac{\sigma(\theta)}{\sin\theta}\right)e^{-im\theta}d\theta, m\geqslant 1. \tag{III.6}$$

The right-hand side of Eq. (III.3) can completely be written in terms of the spectral function if one uses Eq. (II.35). The more familiar form of the theorem

$$\lim_{n\to\infty} H_n\left[\frac{C^2}{H_0a^2(\infty)}\right]^{n+1}=\left(\frac{f_+(1)[-f_+(-1)]}{C^2}\right)^{1/2} \exp\sum_{m=1}^{\infty} m[1/2|\gamma_m|^2+\ln a(\infty)^2] \tag{III.7}$$

can be obtained by exponentiating both sides of Eq. (III.3) and using Eqs. (II.4) and (II.6).

B. Proof of the theorem

Since $Zf_+(Z)$ is an element of A^+ and $Zf_+(Z)\neq 0$ for $|Z|\leqslant 1$, the Wiener–Levy theorem says that

$$ln[Zf_+(Z)]=\sum_{m=0}^{\infty}\gamma(m)Z^m, \quad |Z|\leqslant 1, \tag{III.8}$$

where

$$\sum_{m=0}^{\infty}|\gamma(m)|<\infty. \tag{III.9}$$

It is a consequence of Eq. (III.1) that[18]

$$(1/2\pi)\int_{-\pi}^{\pi}\ln[Zf_+(Z)]\,\overline{\ln(Zf_+(Z))'}\,Z^{-1}d\theta =\sum_{n=1}^{\infty} n|\gamma(n)|^2, \quad Z=e^{i\theta}. \tag{III.10}$$

In Reference 10 it was shown that $[a(\infty)/K(0)]\psi^*(Z,n)\to Zf_+(Z)$ uniformly in norm, and in Appendix A it is shown that Eq. (III.2) implies that $\psi^*(Z,n)\neq 0$ for $|Z|\leqslant 1$ for all n. Therefore,

$$\lim_{n\to\infty}\frac{1}{2\pi}\int_{-\pi}^{\pi}\ln\left[\frac{a(\infty)}{K(0)}\psi^*(Z,n)\right]\overline{\ln\left(\frac{a(\infty)}{K(0)}\psi^*(Z,n)\right)}\,Z^{-1}d\theta =\frac{1}{2\pi}\int_{-\pi}^{\pi}\ln[Zf_+(Z)]\overline{\ln[Zf_+(Z)]'}Z^{-1}d\theta, \quad Z=e^{i\theta}. \tag{III.11}$$

The left-hand side of the above equation can be rewritten in the following way:

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}\ln\left[\frac{a(\infty)^2}{K(0)}\psi^*(Z,n)\right]\overline{\ln\left(\frac{a(\infty)}{K(0)}\psi^*(Z,n)\right)}Z^{-1}d\theta$$

$$=\frac{1}{2\pi}\int_{-\pi}^{\pi}\left[\frac{\overline{\psi^*(Z,n)'}}{\psi^*(Z,n)}\right]\ln\left[\frac{a(\infty)^2}{K(0)}\psi^*(Z,n)\right]d\theta\,. \quad \text{(III.12)}$$

Adding and subtracting $\ln(1-Z)^2$ to the above equations,

$$=\frac{1}{2\pi}\int_{-\pi}^{\pi}\left[\frac{\overline{\psi^*(Z,n)'}}{\psi^*(Z,n)}\right]\ln\left[\frac{\psi^*(Z,n)}{(1-Z)}\right]d\theta$$

$$+\frac{1}{2\pi}\int_{-\pi}^{\pi}\left[\frac{\overline{\psi^*(Z,n)'}}{\psi^*(Z,n)}\right]\ln(1-Z^2)d\theta, \quad Z=e^{i\theta} \quad \text{(III.13)}$$

Now let

$$\left\{\frac{1}{Z_{n,l}}\right\} \text{ be the zeros of } \frac{\psi^*(Z,n)}{\alpha(n)},$$

$$\left\{\frac{1}{Z_{n+1,j}}\right\} \text{ be the zeros of } \frac{\psi^*(Z,n+1)}{\alpha(n+1)},$$

$$\{Z'_{n,k}\} \text{ be the zeros of } \frac{Z^n p(\lambda,n)}{\alpha(n)}, \quad \text{(III.14)}$$

and

$$\alpha(n)=K(0)\prod_{i=1}^{n}\frac{a(\infty)}{a(i)}. \quad \text{(III.15)}$$

From the recurrence formulas we see that

$$\left.\frac{\psi^*(Z,n)}{\alpha(n)}\right|_{Z=0}=\left.\left(\frac{Z^n p(\lambda,n)}{\alpha(n)}\right)\right|_{Z=0}=1. \quad \text{(III.16)}$$

Now

$$\frac{\psi^*(Z,n)'}{\psi^*(Z,n)}=\sum_{i=1}^{2n}\frac{1}{Z-1/Z_{n,i}}, \quad |Z|=1, \quad \text{(III.17)}$$

and

$$\left[\frac{\overline{\psi^*(Z,n)'Z}}{\psi^*(Z,n)}\right]=\sum_{i=1}^{2n}\frac{Z_{n,i}}{Z_{n,i}-Z}, \quad |Z|=1, \quad \text{(III.18)}$$

where we have used the fact that the coefficients of $\psi^*(Z,n)$ are real[10] and consequently the complex roots must come in conjugate pairs. The second term in Eq. (III.13) can now be written as

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}\left[\frac{\overline{\psi^*(Z,n)'Z}}{\psi^*(Z,n)}\right]\ln(1-Z^2)d\theta$$

$$=\frac{1}{2\pi}\int_{-\pi}^{\pi}\sum_{i=1}^{2n}\left[\frac{Z_{n,i}}{Z_{n,i}-Z}\right]\ln(1-Z^2)d\theta\,,$$

$$|Z|=1, \quad \text{(III.19)}$$

which is, from Eq. (III.14),

$$=-\ln\left(\frac{\psi^*(1,n)}{\alpha(n)}\,\frac{\psi^*(-1,n)}{\alpha(n)}\right). \quad \text{(III.20)}$$

We now show that

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}\left[\frac{\overline{\psi^*(Z,n+1)'Z}}{\psi^*(Z,n+1)}\right]\ln\left[\frac{Z^n p(\lambda,n)}{\alpha(n)}\right]d\theta$$

$$=\frac{1}{2\pi}\int_{-\pi}^{\pi}\left[\frac{\overline{\psi^*(Z,n)'Z}}{\psi^*(Z,n)}\right]\ln\left[\frac{\psi^*(Z,n)}{\alpha(n)(1-Z^2)}\right]d\theta\,,$$

$$|Z|=1. \quad \text{(III.21)}$$

Since $Zf_+^R(Z)$ has no zeros for $|Z|\leqslant 1$ neither does $Z^n p(\lambda,n)$[10] for all n. Thus, using Eq. (III.18), the left-hand side of Eq. (III.21) is

$$-\sum_{j=1}^{2n+2}\ln\left[Z^n_{(n+1,j)}\frac{p(\lambda_{(n+1,j)},n)}{\alpha(n)}\right]. \quad \text{(III.22)}$$

Likewise the right hand side of Eq. (III.21) is

$$-\sum_{i=1}^{2n}\ln\left[\frac{\psi^*(Z_{(n,i)},n)}{\alpha(n)(1-Z^2_{(n,i)})}\right]. \quad \text{(III.23)}$$

In Appendix B it is shown that Eq. (III.22) = Eq. (III.23). The first term in Eq. (III.13) can be recast into the following form:

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}\left[\frac{\overline{\psi^*(Z,n)'Z}}{\psi^*(Z,n)}\right]\ln\left[\frac{\psi^*(Z,n)}{\alpha(n)(1-Z^2)}\right]d\theta$$

$$=\frac{1}{2\pi}\sum_{j=1}^{n}\int_{-\pi}^{\pi}\left[\frac{\overline{\psi^*(Z,j)'Z}}{\psi^*(Z,j)}\right]$$

$$\times\ln\left[\frac{\alpha(j)}{\alpha(j-1)}\,\frac{(1-Z^2)Z^{j-1}p(\lambda j-1)}{\psi^*(Z,j)}\right]d\theta. \quad \text{(III.24)}$$

Solving Eq. (II.5) for $1/Z\psi(Z,n-1)$ and substituting the result into Eq. (II.6) yields

$$\psi(Z,n)=p(\lambda,n)-\frac{a(n)}{a(\infty)}Zp(\lambda,n-1). \quad \text{(III.25)}$$

Letting $Z\to 1/Z$ in the above equation and subtracting it from the original equation gives

$$\psi(Z,n)-\psi(1/Z,n)$$

$$=(1-Z^2)\frac{a(n)}{a(\infty)}\frac{1}{Z}p(\lambda,n-1); \quad \text{(III.26)}$$

multiplying the above equation by $Z^n/\alpha(n)$ and then using Eqs; (III.25) and (III.15) yields

$$\frac{\psi^*(Z,n)}{\alpha(n)}-Z^{2n}\frac{\psi^*(1/Z,n)}{\alpha(n)}$$

$$=(1-Z^2)\frac{a(n)^2}{a(\infty)^2}Z^{n-1}\frac{p(\lambda,n-1)}{\alpha(n-1)}. \quad \text{(III.27)}$$

Now dividing the above equation by $\psi^*(Z,n)/\alpha(n)$ and substituting the result into Eq. (III.24)

$$=-\frac{1}{2\pi}\sum_{j=1}^{n}\int_{-\pi}^{\pi}\left[\frac{\overline{\psi^*(Z,j)'Z}}{\psi^*(Z,j)}\right]$$

$$\times\ln\frac{a^2(n)}{a^2(\infty)}\left[1-Z^{2n}\frac{\psi^*(1/Z,j)}{\psi^*(Z,j)}\right]d\theta \quad |Z|=1. \quad \text{(III.28)}$$

It is clear from Eq. (III.18) that only the pole at $Z=0$ contributes, and from Eq. (III.27)

$$Z^{2n}\frac{\psi^*(1/Z,n)}{\alpha(n)}\bigg|_{Z=0}=1-\frac{a(n)^2}{a(\infty)^2}. \tag{III.29}$$

Substituting this into Eq. (III.28) yields

$$=-\sum_{j=1}^{n}2j\ln\left\{1-\left[1-\frac{a(j)^2}{a(\infty)^2}\right]\right\}$$

$$=-\sum_{j=1}^{n}2j\ln\frac{a(n)^2}{a(\infty)^2}. \tag{III.30}$$

Letting $n\to\infty$ in the above equation and Eq. (III.20) and then using Eqs. (III.15) and (II.23) gives the desired result:

$$2\sum_{j=1}^{\infty}j\ln\frac{a(\infty)^2}{a(j)^2}$$

$$=\ln\left[\frac{f_+(1)(-f_+(-1))}{C^2}\right]+\sum_{m=1}^{\infty}m|\gamma(m)|^2.$$

IV. COMPUTATION OF THE CORRECTION TERMS

If one wishes to include the corrections terms, then Eq. (III.8) becomes

$$\ln H_n\left(\frac{C^2}{H_0a(\infty)^2}\right)^{n+1}$$

$$=\ln\det[1+\omega]_{n+1}^{\infty}+\tfrac{1}{2}\ln\left[\frac{f_+(1)[-f(-1)]}{C^2}\right]$$

$$+\sum_{m=1}^{\infty}\frac{m}{2}|\gamma(m)|^2+\sum_{m=1}^{n}m\ln[a(\infty)^2], \tag{IV.1}$$

where $\det[1+\omega]_{n+1}^{\infty}$ is defined in Eq. (II.40). The above formula is most easily derived by returning to Eq. (III) and using Eq. (II.41):

$$\sum_{j=1}^{n}j\ln\frac{a(\infty)^2}{a(j)^2}$$

$$=-\sum_{j=n+1}^{\infty}j\ln\left[\frac{A(j-1,j-1)}{A(j,j)}\right]^2+\sum_{m=1}^{\infty}m\frac{1}{2|\gamma(m)|^2}$$

$$+\tfrac{1}{2}\ln\left[\frac{f+(1)[-f+(-1)]}{C^2}\right]. \tag{IV.2}$$

Equations (II.4) and (II.7) show us that

$$\prod_{j=1}^{n}\frac{1}{a(j)^{2j}}=\prod_{j=1}^{n}\left[\frac{K(j)}{K(j-1)}\right]^{2j}=H_nK(n)^{2n+2} \tag{IV.3}$$

and from Eq. (II.39)

$$\prod_{j=n+1}^{\infty}\frac{A(j-1,j-1)^{2j}}{A(j,j)}=\frac{A(n,n)2n+2}{\det[1+\omega]_{n+1}^{\infty}}. \tag{IV.4}$$

Therefore,

$$\ln H_n+\ln[K^2(n)A^2(n,n)a^n(\infty)]^{n+1}$$

$$=\tfrac{1}{2}\ln\left[\frac{f_+(1)[-f_+(-1)]}{C^2}\right]+\ln\det[1+\omega]_{n+1}^{\infty}$$

$$+\sum_{m=1}^{\infty}\frac{m}{2}|\gamma(m)|^2. \tag{IV.5}$$

Now, using Eqs. (II.7) and (II.41), we have

$$K^2(n)=K(0)^2\prod_{i=1}^{n}\frac{1}{a(i)^2}=\frac{1}{H_0}\prod_{i=1}^{n}\frac{1}{a(i)^2} \tag{IV.6}$$

and

$$A(n,n)^2=\prod_{i=n+1}^{\infty}\frac{a(\infty)^2}{a(i)^2}. \tag{IV.7}$$

Therefore, from Eq. (III.6),

$$K(n)^2A^2(n,n)a^n(\infty)=\frac{C^2}{H_0}a(\infty)^{-n-2}. \tag{IV.8}$$

Substituting this into Eq. (IV.5) gives the desired result.

V. EXTENSION OF THE THEOREM

In this section the requirement that $Zf_+(Z)\neq0$ for $|Z|\leqslant1$ is removed. The final theorem is: If

$$\sum n\left\{\left|1-\frac{a(n)^2}{a(\infty)^2}\right|+|B(n-1)|\right\}<\infty, \tag{V.1}$$

then

$$\sum_{n=1}^{\infty}2n\ln\left(\frac{a(\infty)^2}{a(n)^2}\right)=\sum_{n=1}^{\infty}2n\ln\frac{K(n+1,n+1)}{K(n,n)}$$

$$+\ln\left[(-1)^{n+1}\left(\frac{f_+(1)}{d(1)}\frac{f_+(-1)}{d(-1)}\right)\frac{1}{c_0^2}\right]$$

$$+\sum_{m=1}^{\infty}m|\hat{\gamma}(m)|^2, \tag{V.2}$$

where

$$\hat{\gamma}(m)=-\frac{1}{2}\int_{-\pi}^{\pi}\ln\left(\frac{\sigma(\lambda')|d(Z')|^2}{\sin\theta'}\right)e^{-im\theta'}d\theta',$$

$$m\geqslant1, Z'=e^{i\theta'} \tag{V.3}$$

$$c_0=\exp\left[-\frac{1}{4\pi}\int_{-\pi}^{\pi}\ln\left(\frac{\pi K(0)^2\sigma(\lambda')|d(Z')|^2}{a(\infty)\sin\theta'}\right)d\theta'\right] \tag{V.4}$$

and n is the number of jump points in the spectral function (the number of zeros of $f_+(Z)$ for $|Z|<1$). $K(n,n)$ is given by formula (II.42) with

$$q(n,m)=\int p^0(\lambda,n)p^0(\lambda,m)d(\rho(\lambda)-\rho^0(\lambda)). \tag{V.5}$$

Here

$$d\rho^0(\lambda)=\sigma(\lambda)|d(Z)|^2d\lambda. \tag{V.6}$$

To prove the theorem, first define

$$f_+^0(Z)=\frac{f_+(Z)}{d(Z)}\prod_{i=1}^{n}\frac{(1-ZZ_i)}{\pi_+(Z_i-Z)\pi_-(Z-Z_i)} \tag{V.7}$$

where the Z_i's are the zeros of $f_+(Z)$ for $|Z|<1$, $d(Z)$ is defined in Eq. (II.31), and the $\pi_\pm$ are defined following Eq. (II.35). Notice that $Zf_+^0(Z)|_{Z=0}$ is positive; therefore, using Eq. (II.32) and the fact that

$$\left|\frac{\Pi_{i=1}^{n}(1-ZZ_i)}{\pi_+(Z_i-Z)\pi_-(Z-Z_i)}\right|=1\quad\text{for}\quad|Z|=1, \tag{V.8}$$

one finds

$$\sigma(\lambda)|d(Z)|^2=\frac{a(\infty)\sin\theta}{\pi K(0)^2|f_+^0(Z)|^2}\equiv\sigma^0(\lambda). \tag{V.9}$$

Let $\{P^0(\lambda,n)\},\{a^0(n)\}$, and $\{b^0(n)\}$ be the orthogonal polynomials and coefficients in the recurrence relation generated by the spectral function $\sigma^0(\lambda)d\lambda$. Since the interval of orthogonality is the same, $a^0(\infty)=a(\infty)$ and $b^0(\infty)=b(\infty)$. It can be shown[19] that

$$\sum_{n=1}^{\infty} n\left\{\left|1-\frac{a(n)^2}{a(\infty)^2}\right|+|B(n-1)|\right\}$$
$$\Rightarrow\sum_{n=1}^{\infty} n\left\{\left|1-\frac{a^0(n)^2}{a(\infty)^2}\right|+|B^0(n-1)|\right\}<\infty \tag{V.10}$$

Therefore, the proof from Sec. III applies and

$$\sum_{n=1}^{\infty} 2n\ln\frac{a^0(\infty)^2}{a^0(n)^2}$$
$$=\ln\frac{f_+^0(1)[-f_+^0(-1)]}{c_0^2}+\sum_{m=1}^{\infty} m|\gamma^0(m)|^2. \tag{V.11}$$

It follows from Eq. (V.7) that

$$f_+^0(1)f_+^0(-1)=(-1)^n\frac{f_+(1)}{d(1)}\frac{f_+(-1)}{d(-1)}. \tag{V.12}$$

Substituting the above equation plus Eq. (II.46) into Eq. (V.10) yields the desired result. To find the correction terms, one follows the procedure presented in Sec. IV.

VI. CONCLUSIONS

We have proved an analog of Szegö's theorem for Hankel determinants. The theorem was proved using only some new recurrence formulas satisfied by polynomials orthogonal on a segment of the real line and certain properties of the Jost function. The proof was motivated by a similar proof for polynomials orthogonal on the unit circle. The theorem has been extended to Hankel determinants derived from measures that are not absolutely continuous.

ACKNOWLEDGMENTS

I would like to thank Professor K. M. Case for many useful discussions.

APPENDIX A

Here we show that if $Zf_+(Z)\neq 0$ for $|Z|\leqslant 1$, then $\psi^*(Z,n)\neq 0$ for $|Z|\leqslant 1$. This is done by showing that

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{Z^{-n}Z^m\psi(1/Z,m)}{|f_+(Z)|^2K(0)^2/a(\infty)^2}d\theta=0,$$
$$n\leqslant 2m,\quad Z=e^{i\theta}, \tag{A1}$$

and

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{\psi(Z,n)\psi(1/Z,n)d\theta}{|f_+(Z)|^2K(0)^2/a(\infty)^2}=1,\quad Z=e^{i\theta}. \tag{A2}$$

Note that $Z^m\psi(1/Z,m)$ is a polynomial in Z of degree $2m$. From the recurrence formulas (II.5) and (II.6), it is easy to show that the coefficient of Z^{2m} $[K(2m,2m)]$ is

$$K(2m,2m)=K(0)\prod_{i=1}^{2m}\frac{a(\infty)}{a(i)}, \tag{A3}$$

which is positive. Equations (A1), (A2), and (A3) imply that $Z^m\psi(1/Z,m)$ is a polynomial orthogonal on the unit circle and that

$$Z^{2n}Z^{-n}\psi(Z,n)=\psi^*(Z,n) \tag{A4}$$

is not equal to zero for $|Z|\leqslant 1$.[20]

To prove Eq. (A1) let $Z\to 1/Z$ in Eq. (II.22), then substituting the result into Eq. (A1) using Eq. (II.23) yields

$$=\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{a(\infty)Z^{m-n}}{K(0)(1/Z-Z)}$$
$$\times\frac{[f_+(Z)\psi_+(1/Z,m)-f_-(Z)\psi_-(1/Z,m)]}{|f_+(Z)|^2}d\theta, \tag{A5}$$

which is

$$=\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{a(\infty)Z^{m-n}}{K(0)(1/Z-Z)}\frac{\psi_+(1/Z,m)}{f_-(Z)}d\theta$$
$$-\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{a(\infty)Z^{m-n}}{K(0)(1/Z-Z)}\frac{\psi_-(1/Z,m)}{f_+(Z)}d\theta. \tag{A6}$$

Letting $\theta\to-\theta$ in the first integral gives

$$=\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{a(\infty)Z^{n-m}}{K(0)(Z-1/Z)}\frac{\psi_+(Z,m)}{f_+(Z)}d\theta$$
$$-\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{a(\infty)Z^{m-n}}{K(0)(1/Z-Z)}\frac{\psi_-(1/Z,m)}{f_+(Z)}d\theta. \tag{A7}$$

Using the recurrence relation (II.5) and (II.6), Eqs. (II.17), (II.18), and (II.19), and the properties of $p_+(Z,n)$ listed following Eq. (II.30), we find

$$\psi_+(Z,n)=B(n)\prod_{i=n+1}^{\infty}\frac{a(\infty)}{a(i)}Z^{n+1}+O(Z^{n+2}) \tag{A8}$$

and

$$\psi_-(1/Z,n)=\prod_{i=n+1}^{\infty}\frac{a(\infty)}{a(i)}(1-1/Z^2)Z^n+O(Z^{n-1}). \tag{A9}$$

Substituting these results into Eq. (A7) and using the fact that $Zf_+(Z)\neq 0$ for $|Z|\leqslant 1$ yields Eq. (A1). To prove Eq (A2), substitute in $Z^{-n}\psi(Z,n)$ in place of Z^{-n} in Eq. (A1). Then, following the same steps that led to Eq. (A7), we find

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{\psi(Z,n)\psi(1/Z,n)}{|f_+(Z)|^2K(0)^2/a(\infty)^2}d\theta$$
$$=\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{a(\infty)}{K(0)(Z-1/Z)}\frac{\psi(1/Z,n)\psi_+(Z,n)}{f_+(Z)}d\theta$$
$$-\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{a(\infty)\psi(Z,n)\psi_-(1/Z,n)}{K(0)(1/Z-Z)f_+(Z)}d\theta,$$
$$n\geqslant m,\quad Z=e^{i\theta}. \tag{A10}$$

Equation (A2) can now be obtained using Eqs. (A8), and (A9) and the fact that near zero

$$\psi(Z,n)\sim K(0)\prod_{i=1}^{2n}\frac{a(\infty)}{a(i)}Z^{-n} \tag{A11}$$

and from Eq. (II.23)

$$Zf_*(Z)\sim a(\infty)\prod_{i=1}^{\infty}\frac{a(\infty)}{a(i)}. \tag{A12}$$

APPENDIX B

We now prove that Eq. (III.23) is equal to Eq. (III.24). Letting $Z\to 1/Z$ in Eq. (II.5) and then subtracting it from the original equation yields

$$\frac{1}{Z}\psi(Z,n)-Z\psi(1/Z,n)=(1-Z^2)\frac{p(\lambda,n)}{Z}. \tag{B1}$$

Now multiplying by $Z^n/\alpha(n)(1-Z^2)$ gives

$$\frac{\psi_*(Z,n)}{\alpha(n)(1-Z^2)}-\frac{Z^{2n+2}\psi^*(1/Z,n)}{\alpha(n)(1-Z^2)}=\frac{Z^n p(\lambda,n)}{\alpha(n)}. \tag{B2}$$

Using the notation of Eq. (III.15) we see from Eq. (III.16) that

$$\prod_{i=1}^{2n}\frac{\psi^*(Z_{(n,i)},n)}{\alpha(n)(1-Z^2_{(n,i)})}=\prod_{i=1}^{2n}Z^n{}_{(n,i)}\frac{p(\lambda_{(n,i)},n)}{\alpha(n)}. \tag{B3}$$

Since the leading coefficient of $p(\lambda,n)/\alpha(n)$ is 1, the right-hand side of Eq. (B3) can be written as

$$=\prod_{i=1}^{2n}\prod_{K=1}^{2n}(Z_{(n,i)}-Z'_{(n,K)}). \tag{B4}$$

Interchanging the order of the product

$$=\prod_{K=1}^{2n}\prod_{i=1}^{2n}(Z'_{(n,K)}-Z_{(n,i)}), \tag{B5}$$

which is, using Eq. (III.15),

$$=\prod_{K=1}^{2n}Z'^{2n}_{(n,K)}\frac{\psi^*(1/Z'_{(n,K)},n)}{\alpha(n)}. \tag{B6}$$

Letting $Z\to 1/Z$ in Eq. (II.6) and then multiplying by $Z^{2n+2}/\alpha(n+1)$ yields

$$\begin{aligned}&Z^{2n+2}\frac{\psi^*(1/Z,n+1)}{\alpha(n+1)}\\&=Z^{2n+2}\frac{\psi^*(1/Z,n)}{\alpha(n)}+\left\{\left[1-\frac{a(n+1)^2}{a(\infty)^2}\right]Z^{-1}-B(n)\right\}\\&\times Z^{2n+2}\frac{p(\lambda,n)}{\alpha(n)};\end{aligned} \tag{B7}$$

thus at a zero of $p(\lambda,n)/\alpha(n)$

$$\begin{aligned}&Z'^{2n+2}_{(n,K)}\frac{\psi^*(1/Z'_{(n,K)},n+1)}{\alpha(n+1)}\\&=Z'^{2n+2}_{(n,K)}\frac{\psi^*(1/Z'_{(n,K)},n)}{\alpha(n)},\end{aligned} \tag{B8}$$

and Eq. (B6) becomes[21]

$$=\prod_{K=1}^{2n}Z'^{2n+2}_{(n,K)}\frac{\psi^*(1/Z'_{(n,K)},n+1)}{\alpha(n+1)}. \tag{B9}$$

Writing Eq. (B9) out as in Eq. (B4) and then changing the order of the product and again using the fact that the leading coefficient of $p(\lambda,n)/\alpha(n)$ is 1 gives the desired result:

$$\prod_{i=1}^{2n}\frac{\psi^*(Z_{(n,i)},n)}{\alpha(n)(1-Z^2_{(n,i)})}=\prod_{j=1}^{2n+2}Z^n{}_{(n+1,j)}\frac{p(\lambda_{(n+1,j)},n)}{\alpha(n)}.$$

[1] V. Grenander and G. Szegö, *Toeplitz Forms and Their Applications* (University of California Press, Berkeley, 1959).
[2] M. Kac, Duke Math. J. **21**, 501 (1954).
[3] I. I. Hirshman Jr., J. Anal. Math., **14**, 225 (1965).
[4] A. Devinatz, Illinois J. Math **11**, 160 (1967).
[5] M. E. Fisher and R. E. Hartwig, Adv. Chem. Phys. **15**, 333 (1968).
[6] H. Widom, Am. J. Math. **95**, 333 (1973).
[7] H. Widom, Indiana University Math. J. **21**, 333 (1973).
[8] J. S. Geronimo and K. M. Case, J. Math. Phys., submitted for publication.
[9] G. Szegö, *Orthogonal Polynomials*, American Mathematical Society, Vol 23, (Amer. Math. Soc., New York, 1939).
[10] J. S. Geronimo and K. M. Case, Trans Am. Math. Soc., submitted for publication.
[11] G. Baxter, J. Math. Anal. Appl. **2**, 223 (1961).
[12] When using Eq. (II.10) to calculate Z given λ, one uses the branch of the square root that makes $Z=0$ at $\lambda=\infty$.
[13] K. M. Case, J. Math. Phys. **14**, 916 (1973).
[14] G. S. Guseinov, Sov. Math. Dokl. **17**, 596 (1976).
[15] K. M. Case, J. Math. Phys. **17**, 1703 (1976).
[16] K. M. Case and S. C. Chiu, J. Math. Phys. **14**, 1643 (1973).
[17] K. M. Case, J. Math. Phys. **15**, 2166 (1974).
[18] J. S. Geronimo, Trans. Am. Math. Soc., submitted for publication.
[19] J. S. Geronimo and P. G. Neval, in preparation.
[20] Ya. L. Geronimus, *Polynomials Orthogonal on a Circle and Interval* translated by D. E. Brown, edited by I. N. Sneddon (Pergamon, New York, 1960).
[21] Note that Eq. (III.16) and the fact that the leading coefficient of $p(\lambda,n)/\alpha(n)$ is 1 implies $\prod_{K=1}^{2n}Z'(n,K)=1$.

The master analytic function and the Lorentz group. III. Coupling of continuous representations of O(2,1)

Debabrata Basu

Department of Physics, Indian Institute of Technology, Kharagpur, India

S. Datta Majumdar

Department of Mathematics, Visva-Bharati, Santiniketan, India
(Received 16 November 1977)

The Clebsch–Gordan problem for continuous representations belonging to the principal series of O(2,1) is treated by a method developed previously for the coupling of a discrete and a continuous representation. The values of the complex variable x occurring in the fundamental differential equation of the problem are restricted to lie on the unit circle, and the Clebsh–Gordan coefficients are identified with the Fourier coefficients of solutions of this equation. If j belongs to the discrete class there is only one acceptable solution of the second order equation. But, if j_1,j_2,j all belong to the continuous class any two independent solutions of the equation give possible Clebsch–Gordan series. The problem of orthogonalizing the solutions in the latter case is solved and the normalization factor is determined using the Sturm–Liouville theory of differential equations. The Clebsch–Gordan coefficients generated by an orthogonal pair of solutions become automatically orthogonal. To determine the j values appearing in the reduction, a product state $x^m{}_2$ is expanded in a series of the coupled states $g_{jm}(x)$ by means of the Burchnall–Chaundy formula followed by the Sommerfeld–Watson transformation.

1. INTRODUCTION

The problem of constructing the unitary irreducible representations (UIR's) of the three-dimensional Lorentz Group O(2,1) or its covering group SU(1,1) was solved by Bargmann[1] many years ago. Following this work a number of authors[2-9] investigated the Clebsch–Gordan (CG) problem for the group. In the first two papers[10,11] of this series (I and II), we have shown how a single "Master Analytic Function" can be used to solve two apparently unconnected problems, finding the O(2,1) content of O(3,1), and the CG series and coefficients (CGC's) for O(2,1). In the second paper[11] the CG problem for two discrete representation or one discrete and one continuous (nonexceptional) representations was solved by a novel method based on the theory of hypergeometric function (HGF). This method is now applied to the more difficult problem of the coupling of two continuous representations. We shall consider, in this paper, only continuous nonexceptional representations, i.e., representations belonging to the principal series. The CGC's obtained by us for continuous j_1j_2j are absolutely convergent for all values of the magnetic quantum numbers and exhibit a strong resemblance to those of the rotation group. There are also other good features not posessed by the two other expressions available in the literature. While Holman and Biedenharn's[4] (HB's) expression for the normalization factor (NF) is too complicated for the use, that of Wang[7] involves unsightly fourth roots of factorials instead of the usual square roots. These may be partly due to their evaluation procedure which essentially consists in summing a divergent series with the aid of a convergence factor. HB's expression, besides, is not analytically continuable to other cases of coupling. The three expressions for the CGC (including ours) are all different and are not transformable into one another.

From general considerations and also from the analysis of Basu[12] it appears that for the extension to be possible the hypothetical variable x should lie on the unit circle. The CGC will then become identical with the Fourier coefficient (FC) of an appropriate solution of Eq. (2.4) of II, namely,

$$\Big\{x(1-x)^2\frac{d^2}{dx^2}+(1-x)\times[j_1+j_2+1-m+x(j_1+j_2+m-1)]\times\frac{d}{dx}+\frac{(j_1-m)j_2}{x}+(j_1+m)j_2x+j_1(j_1+1)+j_2(j_2+1)-j(j+1)\Big\}g_{jm}=0 \tag{1.1}$$

Unlike the cases considered in II, the solutions of the above equation for continuous j_1,j_2 are not square-integrable and cannot normally be expanded in Fourier series. A rigorous justification of the procedure adopted can be obtained by using the theory of generalized functions.[13] But as this is not very relevant in the present context we shall only present the main results of our investigation here.

Although the coupling of continuous representations is mathematically more involved, it has the great advantage that the orthonormality (ON) of CGC can be established by using the familiar Sturm–Liouville theory of differential equtions. The simplification arises from the form of the normalization factor (NF) for states belonging to the continuum.[14] For the principal series of representations this is a phase and drops out completely when a CGC is multiplied by its complex conjugate in the orthonormality condition

$$\sum_{m_2}\bar{C}_r(j_1j_2j;m_1,m_2)C_s(j_1j_2j';m_1,m_2)=\delta_{rs}\delta(Imj-Imj'). \tag{1.2}$$

The ON of CGC thus becomes identical with the ON of solutions of Eq. (1.1) on the unit circle. The NF for states of

 0022-2488/79/030492-07$01.00

the principal series, as pointed by Barut and Fronsdal,[14] can also be taken to be unity. We have achieved some simplification by using the alternative value throughout. This is not possible when one (or both) of the representations coupled belongs to the discrete class.

We first find the solutions of Eq. (1.1) which give the correct CG series for a particular final j. The solution of Eq. (1.1) to be taken is

$$g^{(1)}_{jm} = x^{-j_2} e^{(1)}_{jm}$$

$$= x^{-j_2}(1-x)^{j_1+j_2-j}$$

$$\times F(-j-m, j_1-j_2-j; j_1-j_2-m+1; x) \tag{1.3}$$

if j belongs to the positive discrete class, and

$$g^{(5)}_{jm} = x^{-j_1+m} e^{(5)}_{jm}$$

$$= x^{-j_1+m}(1-x)^{j_1+j_2-j}$$

$$\times F(-j+m, j_2-j_1-j; j_2-j_1-j+m+1; x) \tag{1.4}$$

if j belongs to the negative discrete class. This conforms with HB's observations that if any of the three j's belongs to the discrete class D^+ or D^-, then there is only one set of CGC for a given m. If on the other hand j belongs to the continuous class D^c, then any two independent solutions of Eq. (1.1) correspond to possible CG series, and one is led to the problem of choosing from the infinite possibilities a pair of solutions which are orthogonal on the unit circle. The CGC's generated by this pair of orthonormalized basis functions are then automatically orthonormal. The problem of orthogonalization has been solved in Sec. 4 by a simple analytical method. The exact agreement between the expressions of NF determined in Secs. 3 and 4 by two widely different methods provides a good test of the correctness of our results.

After the selection of the appropriate solution of Eq. (1.1) pertaining to a given j, we have to determine the NF for the CGC and the spectrum of j values. These are obtained by expanding $x^{j_1+m_2}$ in a series of functions of the type e_{jm} by means of the Burchnall and Chaundy formula[15] and then applying the Sommerfeld–Watson transformation (SWT).

2. Construction of the Basis Functions and the Fourier Expansion

The group SU(1,1) [covering group of O(2,1)] consists of all 2×2 complex matrices of the form,

$$u = \begin{pmatrix} \alpha & \beta \\ \bar{\beta} & \bar{\alpha} \end{pmatrix}; \quad \det u = 1. \tag{2.1}$$

The UIR's of the group can be considered in the space D_χ of homogeneous functions of two complex variables (ξ_1, ξ_2) transforming according to the fundamental representation of SU(1,1) when D_χ is realized as the space of functions of a single complex variable $z = (\xi_1/\xi_2)$, the operator T_u of the representation is given by,

$$T_u f(z) = (\beta z + \bar{\alpha})^{2j} f\left(\frac{\alpha z + \bar{\beta}}{\beta z + \bar{\alpha}}\right). \tag{2.2}$$

The generators of the group can then be represented as differential operators of the form,

$$J_+ = iz\left(2j - z\frac{\partial}{\partial z}\right), \quad J_- = i\frac{\partial}{\partial z}, \quad J_3 = \left(z\frac{\partial}{\partial z} - j\right). \tag{2.3}$$

The generators (2.3) can be exponentiated to the principal series of UIR's only if the complex numbers z_1, z_2 or equivalently $x = z_2/z_1$ are defined on the boundary of the unit circle.[12]

In the O(2) basis the product states are monomials,

$$f^{j_1,j_2}_{m_1,m_2}(z_1,z_2) = z_1^{j_1,m_1} z_2^{j_2,m_2} = z_1^{j_1,m} z_2^{j_2} x^{m_2}$$

$$m_1, m_2 = 0, \pm\tfrac{1}{2}, \pm 1, \cdots. \tag{2.4}$$

As j_1, j_2 are in the principal series of UIR's $\mathrm{Re}j_1 = \mathrm{Re}j_2 = -\frac{1}{2}$ and the numbers $\sigma = j_1 + j_2 + 1$ and $j_0 = j_1 - j_2$ are, therefore, pure imaginary.

The coupled states g_{jm}, on the other hand, are the bases of UIR's of the group appearing in the reduction,

$$T^{j_1}_u \times T^{j_2}_u = \sum_j \oplus \sigma_j T^j_u. \tag{2.5}$$

and these are of the form,

$$f_{jm} = z_1^{j_1+m} z_2^{j_2} g_{jm}(x). \tag{2.6}$$

The function g_{j_m} is a solution of Eq. (1.1) which on substitution $g = x^{-j_2}(1-x)^{\sigma-j-1}F$ reduces to the standard HG equation,

$$x(1-x)F'' + [2j + (1-x)(j_0 - m + 1 - 2j)]F' + (j+m)(j_0 - j)F = 0. \tag{2.7}$$

Since j can be complex, there are, in general, two linearly independent solutions of Eq. (2.7) which must be chosen so that they satisfy the requirement of orthonormality. The pair of solutions,

$$g^{(1)}_{jm} = x^{-j_2}(1-x)^{\sigma-j-1} F(-j-m, j_0-j; j_0-m+1; x) \tag{2.8a}$$

$$g^{(2)}_{jm} = x^{-j_2}(-x)^{j-j_0}(1-x)^{\sigma-j-1},$$

$$\times F(-j+m, j_0-j; j_0+m+1; x^{-1}) \tag{2.8b}$$

as shown in Sec. 4 satisfy

$$(g^{(r)}_{jm}, g^{(s)}_{jm}) = N^{(r)}_{jm}\delta_{rs}\delta(\mathrm{Im}j - \mathrm{Im}j'). \tag{2.9}$$

The Hilbert space H under the reduction (2.5), therefore, decomposes into two mutually orthogonal subspaces leading to a pair of orthogonal CGC's.

The solution $g^{(2)}_{jm}$ can be expressed as a linear combination of the first and second solution of the HG equation,

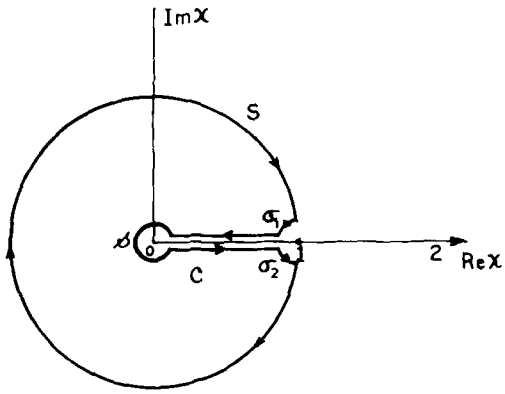


FIG. 1. The contour Σ.

$$g^{(2)}_{jm} = \Gamma\begin{bmatrix} m - j_0, j_0 + m + 1 \\ j + m + 1, -j + m \end{bmatrix} g^{(1)}_{jm}$$

$$- e^{i\pi(j_0 - m)}\Gamma\begin{bmatrix} j_0 - m + 1, m - j_0, j_0 + m + 1 \\ m - j_0 + 1, j_0 - j, j_0 + j + 1 \end{bmatrix} g^{(5)}_{jm} \tag{2.10}$$

where $g^{(5)}_{jm}$ stands for the second solution of the HG equation and is given by Eq. (1.4).

However, when j belongs to $D^{\pm}$, the values of m are bounded above or or below and Eq. (2.7) admits only one independent solution which is given by Eq. (2.8a) for D^{+} and by Eq. (1.4) for D^{-}. There is, therefore, no duplicity for representations belonging to $D^{\pm}$.

Since, in Eq. (2.4), $j_1 j_2$ and m are fixed, the monomials x^{m_2} can be regarded as representing the bases in the product representation. Further, m_2 (or m_1) takes all integer or half-integer values on the real line. The nonnormalized CGC's are, therefore, the coefficients of Fourier expansion of $g^{(1)}_{jm}$ and $g^{(2)}_{jm}$

$$g^{(r)}_{jm} = \sum_{m_2 = -\infty}^{\infty} a^{(r)}_{m_2} x^{m_2}; \quad r = 1,2. \tag{2.11}$$

The FC's can be expressed as an integral over the unit circle S (see Fig. 1) in the clockwise direction. Thus,

$$a^{(1)}_{m_2} = -\frac{1}{2\pi i}\int_S dx\; x^{-j_2 - m_2 - 1}(1-x)^{\sigma - j - 1}$$

$$\times F(-j - m, j_0 - j; j_0 - m + 1; x). \tag{2.12}$$

The singularities of the integrand are the branch points at $x = 0$ and at $x = 1$. With the standard choice of the cut for the HGF the integrand in Eq. (2.12) is single valued and analytic in the entire x-plane assumed cut along the positive real axis from zero to infinity. If, therefore, we choose a closed contour Σ, as shown in Fig. 1, then by Cauchy's theorem,

$$\oint_\Sigma x^{-j_2 - m_2 - 1}(1-x)^{\sigma - j - 1}$$

$$\times F(-j - m, j_0 - j; j_0 - m + 1; x)dx = 0. \tag{2.13}$$

Since the integrals over the circular arcs σ_1 and σ_2 near $x = 1$ vanish, we have

$$a^{(1)}_{m_2} = \frac{1}{2\pi i}\int_C x^{-j_2 - m_2 - 1}(1-x)^{\sigma - j - 1}$$

$$\times F(-j - m, j_0 - j; j_0 - m + 1; x)dx, \tag{2.14}$$

where C stands for the "*Pochhammer part*" of Σ formed by the small circle of radius ϵ around the origin and the part of the branch cut from ϵ to 1-0.

Comparing Eq. (2.14) with the standard single loop Pochhamer contour integral representation[16] of the generalized HGF, we obtain

$$a^{(1)}_{m_2} = e^{-i\pi(j_2 + m_2)}$$

$$\times\Gamma\begin{bmatrix} \sigma - j \\ j_2 + m_2 + 1, j_1 - m_2 - j + 1 \end{bmatrix}$$

$$\times {}_3F_2\begin{bmatrix} -j_2 - m_2, j_0 - j, -j - m \\ j_0 - m + 1, j_1 - m_2 - j + 1 \end{bmatrix}. \tag{2.15}$$

Similarly,

$$a^{(2)}_{m_1, m_2} = e^{-2i\pi m_2} a^{(1)}_{-m_1, -m_2}. \tag{2.16}$$

The ${}_3F_2{}^{(1)}$ functions appearing in Eqs. (2.15) and (2.16) are absolutely convergent for all values of m_1, m_2, m.

We shall also require the function $g^{(5)}_{jm}$ as given by Eq. (1.4) and its FC,

$$a^{(5)}_{m_2} = a^{(1)\mathscr{I}}_{m_2}, \tag{2.17}$$

where the superscript $\mathscr{I}$ corresponds to the simultaneous interchange $j_1 \leftrightarrow j_2, m_1 \leftrightarrow -m_2$ ($\mathscr{I}$-transform).

3. THE CLEBSCH-GORDAN SERIES

For positive m we expand $x^{j_2 + m_2}$ in a series of the form[15]

$$x^{j_2 + m_2} = \sum_{n=0}^{\infty} (-)^n \frac{(a)_n (b)_n}{(c + n - 1)_n n!}$$

$$\times {}_3F_2\begin{bmatrix} -j_2 - m_2, c + n - 1, -n \\ a, b \end{bmatrix}(1-x)^n$$

$$\times F(a + n, b + n; c + 2n; 1 - x), \tag{3.1}$$

with $a = 1 - \sigma - m$, $b = -2j_2$, $c = 2(1 - \sigma)$. This can be written as a sum of residues at $j = \sigma - n - 1$ of the analytic function

$$\chi(j) = \Gamma\begin{bmatrix} -j-m,j_0-j,-\sigma+j+1,-\sigma-j \\ -2j-1,1-\sigma-m,-2j_2 \end{bmatrix} {}_3F_2^{(0)} f_{jm}(x), \tag{3.2}$$

where

$${}_3F_2^{(0)} = {}_3F_2\begin{bmatrix} -j_2-m_2,-\sigma+j+1,-\sigma-j \\ -2j_2,1-\sigma-m \end{bmatrix}, \tag{3.3}$$

$$f_{jm}(x) = (1-x)^{\sigma-j-1}F(-j-m,j_0-j;-2j;1-x).$$

Let us now enclose the poles at $j=\sigma-n-1$ by a contour C consisting of the infinite semicircle on the left half of the j-plane and the line $\mathrm{Re}\, j = -\frac{1}{2}$. Since C contains additional poles at $j=-m+k,(k=0,1,2,\cdots)$ and since the integral over the semicircular part vanishes, we have,

$$x^{j_2+m_2} = \frac{1}{2\pi i}\int_{-1/2-i\infty}^{-1/2+i\infty}\chi(j)dj - \sum_{j=-1(\mathrm{or}-3/2)}^{m} D_j^{+} \tag{3.4}$$

where D_j^{+} stands for the residue of $\chi(j)$ at $j+m=k$, $(k=0,1,\cdots,m-1)$.

when j_1 and j_2 both belong to D^c the duplicity phenomenon discussed in the preceding section makes it necessary to transform the rhs of Eq. (3.4) into a form involving the basis functions $g^{(1)},g^{(2)}$ and the corresponding FC's (or CGC's). Using the appropriate Slater identity[16] to express ${}_3F_2^{(0)}$ occurring in Eq. (3.4) in terms of the complex conjugate of the FC's of $g^{(1)}$ and $g^{(2)}$, we obtain after some calculations,

$$\begin{aligned} x^{j_2+m_2} &= \frac{e^{i\pi j_1}\sin\pi j_1}{\sin\pi j_0}\sum_j D_j^{+} + \frac{1}{2\pi i}\int_{-1/2}^{-1/2+i\infty} dj\,\Gamma\begin{bmatrix} -j_0-j,-j_0+j+1,j_0-j,j_0+j+1,-j-m,j+1-m \\ j_0-m+1,-j_0-m+1,-2j-1,2j+1 \end{bmatrix} e^{i\pi(2j_2+1)}\bar{a}_{m_2}^{(1)}x^j g_{jm}^{(1)} \\ &+ \frac{1}{2\pi i}\int_{-1/2}^{-1/2+i\infty} dj\,\Gamma\begin{bmatrix} -j_0-j,-j_0+j+1,j_0-j,j_0+j+1,-j+m,j+m+1 \\ 1+m-j_0,1+m+j_0,-2j-1,2j+1 \end{bmatrix} e^{i\pi(2j_2+1)}\bar{a}_{m_2}^{(2)}x^j g_{jm}^{(2)}. \end{aligned} \tag{3.5}$$

By introducing $\eta(j,m)$ which is 1 for discrete j and $-\cot\pi(j+m)$ for continuous j, the discrete and the continuous parts can be given the same analytic form:

$$x^{j_2+m_2} = \left(\sum_{j=-(\mathrm{or}-3/2)}^{-m} + i\int_{-1/2}^{-1/2+i\infty} dj\, b^{(1)}\right)(j_1j_2j;m_1,m_2)x^j g_{jm}^{(1)} + i\int_{-1/2}^{-1/2+i\infty} dj\, b^{(2)}(j_1j_2j;m_1,m_2)x^j g_{jm}^{(2)}, \tag{3.6}$$

where

$$\begin{aligned} b^{(1)}(j_1j_2j;m_1,m_2) &= \bar{\eta}(j,m)e^{i\pi(j_2+m_2)}\Gamma\begin{bmatrix} -j_0-j,-j_0+j+1,j_0-j,j_0+j+1,-2j,-\sigma+j+1 \\ -j_2+m_2,-j_0-m+1,j_0-m+1,-2j-1,-j+m,j+m+1,-j_1+j+1-m_2 \end{bmatrix} \\ &\times {}_3F_2\begin{bmatrix} j_2-m_2+1,-j_0+j+1,j+1-m \\ -j_0-m+1,-j_1+j+1-m_2 \end{bmatrix}, \end{aligned}$$

$$b^{(2)}(j_1j_2j;m_1,m_2) = e^{2i\pi m_2}b^{(1)}(j_1j_2j;-m_1,-m_2). \tag{3.7}$$

The NF's for the CG series and coefficients are obtained by comparing the coefficients of direct and inverse expansions. If $A_{jm}^{(r)},(r=1,2)$, is the NF, then

$$A_{jm}^{(1)} = \frac{1}{2\pi}e^{i\pi(j_2+1/2)}\left\{\Gamma\begin{bmatrix} -j_0-j,j_0+j+1,j_0-j,-j_0+j+1,-j-m,j+1-m \\ -j_0-m+1,j_0-m+1,-2j-1,2j+1 \end{bmatrix}\right\}^{1/2}, \tag{3.8}$$

$$A_{jm}^{(2)} = A_{j-m}^{(1)}.$$

These are in complete agreement with the NF's determined in Sec. 4 by a different method. Using Eqs. (3.8), (2.15), (2.16), we finally obtain the normalized CGC's,

$$C_r(j_1j_2j;m_1,m_2) = A^{(r)}_{jm}a^{(r)}_{m_2}, \tag{3.9}$$

for j lying in the continuous spectrum. For the discrete spectrum the CGC's are givem by C_1 when j belongs to D^+ and by C_2 when j belongs to D^-.

4. THE ORTHOGONALITY OF THE EIGENFUNCTIONS AND THE CGC's

The first step in the derivation of the orthogonality condition for the CGC's consists in constructing a suitable scalar product (f,g) with respect to which the UIR's T^j_u appearing in the reduction (Eq. (3.6)) are required to be unitary. This is easily achieved by adopting Gelfand's definition[13] of the Kronecker product of two UIR's and constructing the corresponding invariant bilinear functional satisfying,

$$(T_uf_1,T_uf_2) = (f_1,f_2) \tag{4.1}$$

where

$$T_uf(z_1,z_2) = (\beta z_1 + \bar{\alpha})^{2j_1}(\beta z_2 + \bar{\alpha})^{2j_2}f(z_1',z_2'),,$$
$$z_r' = \frac{\alpha z_r + \bar{\beta}}{\beta z_r + \bar{\alpha}}, \quad r = 1,2. \tag{4.2}$$

When j_1j_2 are in the principal series of UIR's and accordingly z_1,z_2 are allowed to vary on the unit circle ($z_r = e^{i\phi r}$) a bilinear functional satisfying Eq. (4.1) exists and is of the form,

$$(f_1,f_2) = \int d\phi_1 d\phi_2 \bar{f}_1(z_1,z_2)f_2(z_1,z_2)z_1^{-2j_1-1}z_2^{-2j_2-1} \tag{4.3}$$

According to our choice of a basis of the Hilbert space

$$f_1 = z_1^{j_1+m}z_2^{j_2}g_{jm}(x) \tag{4.4}$$

Using Eqs. (4.3) and (4.4) it is now easy to verify that the functions g_{jm} in the coupled representation should be orthonormalized according to the scalar product,

$$(g_{jm},g_{j'm}) = \int_0^{2\pi}\bar{g}_{jm}(x)g_{j'm}(x)d\phi, \tag{4.5}$$

where $x = e^{i(\phi_2-\phi_1)} = e^{i\phi}$. The orthogonality of the eigenfunctions $g^{(r)}_{jm}(x)$ of Sec. 2 can now be established in the traditional way by explicit evaluations of this scalar product with the help of Sturm–Liouville theory of the second order differential equation.

We shall first show that if j is in the principal series the functions $g^{(1)}_{jm},g^{(2)}_{jm}$ constitute two independent orthogonal sets.

Using Eqs. (2.8a), (4.5) and the standard formula

$$F(a,b;c;z) = \Gamma\begin{bmatrix} c,b-a \\ b,c-a \end{bmatrix}(-z)^{-a}F(a,1-c+a;1-b+a;z^{-1}) + \Gamma\begin{bmatrix} c,a-b \\ a,c-b \end{bmatrix}(-z)^{-b}F(b,1-c+b;1-a+b;z^{-1}), \tag{4.6}$$

and following the method of Sec. 2 the integral over S can now be expressed as a single loop Pochhammer-integral over C (Fig. 1). The Eq. (4.5) then takes the following form,

$$(g^{(1)}_{jm},g^{(1)}_{j'm}) = e^{i\pi(m-\sigma-1)}\Gamma\begin{bmatrix} 1-j_0-m, m-j_0 \\ -j_0-j, -j_0+j+1 \end{bmatrix} \times(-i)\int_C \phi_{jm}\phi_{j'm}x^{j_0-m}dx + e^{i\pi(j_0-\sigma-1)}\Gamma\begin{bmatrix} 1-j_0-m, j_0-m \\ -j-m, j+1-m \end{bmatrix} (-i)\int_C \phi_{jm}(x)\phi_{j'm}(x)dx, \tag{4.7}$$

where

$$\phi_{jm}(x) = (1-x)^{-j-1}F(-j-m,j_0-j;j_0-m+1;x) \tag{4.8}$$

The integrand in the second term on the rhs of Eq. (4.7) which is regular at $x = 0$ is continuous across the branch cut and the integral therefore vanishes. The only contribution to the scalar product therefore comes from the first term and we have

$$(g^{(1)}_{jm},g^{(1)}_{j'm}) = e^{i\pi(m-\sigma-1)}\Gamma\begin{bmatrix} -j_0-m+1, m-j_0 \\ -j_0-j, -j_0+j+1 \end{bmatrix}G(j',j;m), \tag{4.9}$$

where

$$G(j',j;m) = -i\lim_{\epsilon\to 0}\left[(e^{2\pi ij_0}-1)\int_\epsilon^1 dx\, x^{j_0-m}\phi_{jm}(x)\phi_{j'm}(x) + \int_s x^{j_0-m}\phi_{jm}(x)\phi_{j'm}(x)dx\right]. \tag{4.10}$$

Here "s" stands for the small circle of radius ϵ centered at the origin.

The integrals appearing in the rhs of the above equation can be evaluated by using the differential equation satisfied by the eigenfunctions ϕ_{jm} and $\phi_{j'm}$, and we have

$$\int_\epsilon^1 x^{j_0-m}\phi_{jm}(x)\phi_{j'm}(x)dx$$

$$= \frac{\lim_{x\to 1}}{(j-j')(j+j'+1)}\left[(1-x)^2\left(\phi_{j'm}\frac{d\phi_{jm}}{dx}-\phi_{jm}\frac{d\phi_{j'm}}{dx}\right)\right]$$

$$-\frac{e^{j_0-m+1}}{(j_0-m+1)}, \tag{4.11}$$

$$\int_s x^{j_0-m}\phi_{jm}(x)\phi_{j'm}(x)dx = \frac{(e^{2\pi i j_0}-1)e^{j_0-m+1}}{j_0-m+1} \tag{4.12}$$

Combining Eqs. (4.10), (4.11) and (4.12), we finally obtain

$G(j'j;m)$

$$= \frac{2e^{i\pi j_0}\sin\pi j_0}{(j-j')(j+j'+1)}\lim_{x\to 1}\left[(1-x)^2\left(\phi_{j'm}\frac{d\phi_{jm}}{dx}-\phi_{jm}\frac{d\phi_{j'm}}{dx}\right)\right]. \tag{4.13}$$

Following I, the rhs of Eq. (4.13) can now be evaluated and we obtain after some calculations

$(g^{(1)}_{jm},g^{(1)}_{j'm})$

$$= 4\pi^2 e^{i\pi(j_0-\sigma)} \times \Gamma\begin{bmatrix} -j_0-m+1, j_0-m+1, -2j-1, 2j+1 \\ -j_0-j, -j_0+j+1, j_0-j, j_0+j+1, -j-m, j+1-m \end{bmatrix} \times \delta(\mathrm{Im}j-\mathrm{Im}j') \text{ for } \mathrm{Im}j, \mathrm{Im}j'>0. \tag{4.14}$$

The NF as given by Eq. (4.14) agrees exactly with that obtained in the previous section [Eq. (3.8)] in an entirely different manner. Similarly,

$$(g^{(2)}_{jm},g^{(1)}_{j'm}) = -ie^{i\pi(j_0-\sigma-1)}\int_C dx\phi^{\mathcal{F}}_{jm}(x)\phi_{j'm}(x). \tag{4.15}$$

Although $g^{(2)}$ and $g^{(1)}$ have branch points at the origin, the integrand as before [Eq. (4.7)] is regular at $x=0$. The integral therefore vanishes,

$$(g^{(1)}_{jm},g^{(1)}_{j'm}) = 0.$$

The CGC's $C_r(j_1j_2j;m_1,m_2)$ of the previous section are then the coefficients of Fourier expansion of the orthonormalized eigenfunctions,

$$f^{(r)}_{jm} = A^{(r)}_{jm}g^{(r)}_{jm} \tag{4.16}$$

satisfying

$$(f^{(r)}_{jm}f^{(s)}_{j'm}) = \delta_{rs}\delta(\mathrm{Im}j'-\mathrm{Im}j). \tag{4.17}$$

The orthogonality condition for the CGC's now follows immediately from that of the product states and Eqs. (4.16) and (4.17):

$$\sum_{m_2}\bar{C}_r(j_1j_2j;m_1,m_2)C_s(j_1j_2j';m_1,m_2) = \delta_{rs}\delta(\mathrm{Im}j'-\mathrm{Im}j). \tag{4.18}$$

When j lies in the discrete spectrum, the orthonormality condition can be derived essentially in the same way, and we have

$$\sum_{m_2}\bar{C}(j_1j_2j;m_1,m_2)C(j_1j_2j';m_1,m_2) = \delta_{jj'}. \tag{4.19}$$

APPENDIX

To derive Eq. (3.5), we transform the rhs of Eq. (3.4) by using the identity

$$F_p(0.45) = \Gamma\begin{bmatrix}\alpha_{035}, & \alpha_{045}, & \alpha_{015}\\ \alpha_{235}, & \alpha_{245}, & \alpha_{125}\end{bmatrix}F_p(2.01) - \Gamma\begin{bmatrix}\alpha_{035}, & \alpha_{045}, & \alpha_{015}\\ \alpha_{143}, & \beta_{02}, & 1-\beta_{02}\end{bmatrix}F_n(2.14). \tag{A 1}$$

This gives

$${}_3F_2^{(0)} = \Gamma\begin{bmatrix} -j_1+m_1, j_1-m_1+1, -j_0+j+1, -j_0-j, -2j_2, 1-\sigma-m \\ j+1-m, -j_2-m_2, -j-m, -\sigma+j+1, 1+m-j_0, -j_1-j+m_2\end{bmatrix}{}_3F_2^{(2)}$$

$$-\Gamma\begin{bmatrix} -j_1+m_1, j_1-m_1+1, -j_0+j+1, -j_0-j, -2j_2, 1-\sigma-m \\ -j_2+m_2, 1+m-j_0, j_0-m, -\sigma-j, -j_1+j+1-m_2, 1-j_0-m\end{bmatrix}{}_3F_2^{(1)}, \tag{A2}$$

where

$$ {}_3F_2^{(1)} = {}_3F_2\begin{bmatrix} j_2 - m_2 + 1, -j_0 + j + 1, j + 1 - m \\ -j_0 - m + 1, -j_1 + j + 1 - m_2 \end{bmatrix}, \quad {}_3F_2^{(2)} = {}_3F_2\begin{bmatrix} j_2 + m_2 + 1, -j_0 - j, -j + m \\ -j_0 + m + 1, -j_1 - j + m_2 \end{bmatrix}. \tag{A3} $$

The ${}_3F_2(1)$ functions appearing in Eq. (A2) are invariant under the Legendre transformation $j \to -j - 1$, for instance,

$$ \Gamma\begin{bmatrix} -\sigma + j + 1 \\ -j_1 + j + 1 - m_2 \end{bmatrix} {}_3F_2^{(1)}(j) = \Gamma\begin{bmatrix} -\sigma - j \\ -j_1 - j - m_2 \end{bmatrix} {}_3F_2^{(1)}(-j-1). \tag{A4} $$

Substituting Eq. (A2) in Eq. (3.4), folding the integral about the real axis, and using the relation (2.10), we have

$$ \begin{aligned} x^{j_2 + m_2} = {} & \sum_j D^j{}_+ - \frac{1}{2\pi i} \int_{-1/2}^{-1/2 + i\infty} dj \\ & \Gamma\begin{bmatrix} -j_1 + m_1, j_1 - m_1 + 1, -\sigma + j + 1, -j_0 - j, -j_0 + j + 1, j_0 - j, j_0 + j + 1, -j - m, j + 1 - m \\ -j_2 + m_2, -j_0 + m + 1, j_0 - m, -j_0 - m + 1, -j_1 + j + 1 - m_2, -2j - 1, 2j + 1, j_0 - m + 1 \end{bmatrix} {}_3F_2^{(1)} x^j g_{jm}^{(1)} \\ & + \frac{1}{2\pi i} \int_{-1/2}^{-1/2 + i\infty} dj \\ & \Gamma\begin{bmatrix} -j_1 + m_1, j_1 - m_1 + 1, -j_0 - j, -j_0 + j + 1, j_0 - j, j_0 + j + 1, -\sigma - j, -j + m, j + m + 1 \\ -j_2 - m_2, -j_0 + m + 1, j_0 + m + 1, j_0 - m + 1, -j_0 + m, -j_1 - j + m_2, -2j - 1, 2j + 1 \end{bmatrix} {}_3F_2^{(2)} x^j g_{jm}^{(2)} + Y, \end{aligned} \tag{A5} $$

where

$$ Y = e^{i\pi(j_0 - m)} \frac{1}{2\pi i} \int_{-1/2}^{-1/2 + i\infty} dj \Gamma\begin{bmatrix} -j_1 + m_1, j_1 + m_1 + 1, -j_0 - j, -j_0 + j + 1, -\sigma - j, -j + m, j + m + 1 \\ -j_2 - m_2, -j_0 + m + 1, -j_0 + m + 1, -j_1 - j + m_2, -2j - 1, 2j + 1 \end{bmatrix} {}_3F_2^{(2)} x^j g_{jm}^{(5)} $$

$$ = e^{i\pi(j_0 - m)} \frac{1}{2\pi i} \int_{-1/2 - i\infty}^{-1/2 + i\infty} dj \Gamma\begin{bmatrix} -j_1 + m_1, j_1 - m_1 + 1, -\sigma - j, -j + m, -j_0 - j \\ -j_2 - m_2, -j_0 + m + 1, -j_1 - j + m_2, -2j - 1 \end{bmatrix} {}_3F_2^{(2)} f_{jm}(x). \tag{A6} $$

Replacing $f_{jm}(x)$ by $x^{-j_0 + m} f_{jm}(x)^{\mathscr{F}}$ and the ${}_3F_2(1)$ function occurring in Eq. (A6) by

$$ \Gamma\begin{bmatrix} -j_0 + m + 1, -j_1 - j + m_2, -\sigma + j + 1 \\ -2j_1, -\sigma + m + 1, j_2 + m_2 + 1 \end{bmatrix} {}_3F_2\begin{bmatrix} -j_1 + m_1, -\sigma - j, -\sigma + j + 1 \\ -2j_1, -\sigma + m + 1 \end{bmatrix}, $$

we now have

$$ Y = \frac{e^{i\pi(j_0 - m)}}{2\pi i} \Gamma\begin{bmatrix} -j_1 + m_1, j_1 - m_1 + 1 \\ -j_2 - m_2, j_2 + m_2 + 1 \end{bmatrix} \int_{-1/2 - i\infty}^{-1/2 + i\infty} dj \Gamma\begin{bmatrix} -j + m, -j_0 - j, -\sigma - j, -\sigma + j + 1 \\ -2j_1, -\sigma + m + 1, -2j - 1 \end{bmatrix} {}_3F_2^{(0)} x^{-j_0 + m} f_{jm}(x)^{\mathscr{F}}. \tag{A7} $$

The integrand in this expression is seen to be the $\mathscr{F}$-transform of $\chi(j)$ of Eq. (3.2) having simple poles at $j = \sigma - n - 1$ but at no other points inside C. Therefore,

$$ Y = e^{i\pi j_0} \frac{\sin \pi j_2}{\sin \pi j_1} x^{-j_0 + m} x^{j_1 - m_1} = e^{i\pi j_0} \frac{\sin \pi j_2}{\sin \pi j_1} x^{j_2 + m_2}. \tag{A8} $$

Using this value of Y and the expressions (2.15) and (2.16) for the FC's, we obtain Eq. (3.5).

[1] V. Bargmann, Ann. Math. **48**, 568 (1947).
[2] L. Pukanszky, Trans. Am. Math. Soc. **100**, 116 (1961).
[3] M. Andrews and J. Gunson, J. Math. Phys. **5**, 1391 (1964).
[4] W.J. Holman and L.C. Biedenharn, Ann. Phys. **39**, 1 (1966); **47**, 205 (1968).
[5] S.S. Sannikov, Sov. Phys.-Doklady **11**, 1045 (1967).
[6] I. Ferretti and M. Verde, Nuovo Cimento **55**, 110 (1968).
[7] K.H. Wang, J. Math. Phys. **11**, 2077 (1970).
[8] N. Mukunda and B. Radharkrishnan, J. Math. Phys. **15**, 1320 (1974); **15**, 1332 (1974); **15**, 1643 (1974); **15**, 1656 (1974).
[9] A.O. Barut and R. Wilson, J. Math. Phys. **17**, 900 (1976).
[10] D. Basu and S.D. Majumdar, J. Math. Phys. **17**, 185 (1976) -I.
[11] S.D. Majumdar, J. Math. Phys. **17**, 193 (1976) -II; this paper contains a brief summary of the methods used and the results obtained by the previous authors.
[12] D. Basu, J. Math. Phys. **18**, 742 (1977).
[13] I.M. Gelfand, M.I. Graev and N. Ya. Vilenkin, *Generalized Functions* (Academic, New York, 1966), Vol. 5.
[14] A.O. Barut and C. Fronsdal, Proc. R. Soc. London Ser. A **287**, 532 (1965).
[15] A. Erde'lyi (Ed.), *Higher Transcendental Functions* (McGraw-Hill, New York, 1953), Vol. 1.
[16] L.J. Slater, *Generalized Hypergeometric Functions* (Cambridge U.P. Cambridge, 1966).

Remarks on the Hamiltonian path integral in polar coordinates

Wolfgang Langguth[a] and Akira Inomata

Department of Physics, State University of New York at Albany, Albany, New York 12222
(Received 7 July 1978)

Problems associated with the derivation of the Hamiltonian path integral in polar coordinates are examined. First the use of the ill-defined asymptotic formula of the modified Bessel function is pointed out. A procedure is proposed to justify its practical use, in which the mass m is complexified and the limit $\mathrm{Im}\, m \to 0$ is taken after path integrations. Hereby a restriction is imposed on the class of allowed potentials. The difference between the Hamiltonian path integral so obtained and the phase space path integral formally defined is also discussed.

1. INTRODUCTION

Feynman's path integral approach to quantum mechanics[1] has been considered unique in that the propagator is described by a path integral,

$$K(\mathbf{r}'',\mathbf{r}';\tau) = \int \exp\left(\frac{i}{\hbar}\int^{\tau} L(\mathbf{r},\dot{\mathbf{r}})dt\right)\mathscr{D}\mathbf{r}, \tag{1.1}$$

which rests on the Lagrangian of the system rather than the Hamiltonian. In recent years, more attention has been placed on the Hamiltonian path integral formulated in phase space,[2]

$$K\mathbf{r}'',\mathbf{r}';\tau) = \iint \exp\left(\frac{i}{\hbar}\int^{\tau} [\mathbf{p}\cdot\dot{\mathbf{r}} - H(\mathbf{r},\mathbf{p})]dt\right)\mathscr{D}\mathbf{r}\mathscr{D}\mathbf{p}, \tag{1.2}$$

whose equivalence to the Lagrangian path integral (1.1) has been shown only in Cartesian coordinates. Since there is no general proof[3,4] for the canonical invariance of the Hamiltonian integral (1.2), it is as yet unknown to what extent the Hamiltonian formulation is valid. The first question we have to clarify would be as to whether the equivalence between the Lagrangian and Hamiltonian integrals can be maintained under point canonical transformations. A general expectation is that the Hamiltonian path integral (1.2) on any coordinate basis would yield a desired propagator if an appropriate effective Hamiltonian, which may or may not be the same as the classical one, is chosen.[5,6] While the effective Hamiltonian coincides with the classical one in the Cartesian formulation, the same is not true, as we shall see shortly, on the polar coordinate basis. The transformation from Cartesian to polar coordinates is certainly an important example of point transformations. In the present paper we shall be concerned with problems in deriving the Hamiltonian path integral of the form (1.2) from the Lagrangian path integral (1.1) specifically in polar coordinates.

Edwards and Gulyaev[7] calculated the Lagrangian path integral (1.1) for a free particle in polar coordinates and obtained the radial propagator for the l wave,

$$K_l(\mathbf{r}'',\mathbf{r}';\tau) = \int \exp\left[\frac{i}{\hbar}\int^{\tau}\left(\tfrac{1}{2}m\dot{r}^2 - \frac{l(l+1)\hbar^2}{2mr^2}\right)dt\right]\mathscr{D}r, \tag{1.3}$$

by using the asymptotic form of the modified Bessel function for $|z| \gg 1$,

$$I_\nu(z) \sim (2\pi z)^{-1/2}\exp[z - \tfrac{1}{2}(\nu^2 - \tfrac{1}{4})z^{-1}]. \tag{1.4}$$

Following the recipe Garrod[2] used in the Cartesian formulation, Peak and Inomata[8] derived from (1.3) the Hamiltonian path integral for the l wave in the presence of a central potential $V(r)$,

$$K_l(r'',r';\tau) = \iint \exp\left(\frac{i}{\hbar}\int^{\tau}[p\dot{r} - H_l(r,p)]dt\right)\mathscr{D}r\mathscr{D}p \tag{1.5}$$

with the effective Hamiltonian,

$$H_l(r,p) = \frac{1}{2m}\left(p^2 + \frac{l(l+1)\hbar^2}{r^2}\right) + V(r). \tag{1.6}$$

Furthermore, without expanding in partial waves, but using the asymptotic formula (1.4) repeatedly, Peak and Inomata[8] expressed, after a step of formal replacements, the total propagator (1.1) in the form,

$$K(r'',\theta'',\phi'';r',\theta',\phi';\tau)$$
$$= \lim_{N\to\infty}(2\pi\hbar)^{-3N}\int \exp\left(\frac{i}{\hbar}\int^{\tau}(p_r\dot{r} + p_\theta\dot{\theta} + p_\phi\dot{\phi} - H_{\mathrm{eff}})dt\right)$$
$$\times \prod^{N-1}(dr d\theta d\phi)\prod^{N}(dp_r dp_\theta dp_\phi) \tag{1.7}$$

with the effective Hamiltonian,

$$H_{\mathrm{eff}} = p_r^2/(2m) + (p_\theta^2 - \tfrac{1}{4}\hbar^2)/(2mr^2)$$
$$+ (p_\phi^2 - \tfrac{1}{4}\hbar^2)/(2mr^2\sin^2\theta) + V(r). \tag{1.8}$$

The path integrals (1.3) and (1.5) have been proven useful for such simple systems as a harmonic oscillator, a charged particle in a uniform magnetic field, a rigid rotator, and a particle in an inverse square potential,[8] the three body problem[9,10] and the Aharonov–Bohm effect.[11] An expression similar to (1.7) had been used earlier by Gutzwiller[12] for the hydrogen atom problem. Indeed, the Hamiltonian path integral (1.7) appears to be the correct form in polar coordinates.[4] A careful examination, however, shows that the asymptotic expan-

[a]Present Address: Physikalisches Institut der Universität Würzburg, D – 8700 Würzburg, Germany.

 0022-2488/79/030499-06$01.00

sion formula (1.4), on which the derivation of (1.3), (1.5), and (1.7) is based, is not valid for the range of the argument pertinent to quantum mechanics. Accordingly, the expressions (1.3), (1.5), and (1.7) will be groundless. This odd situation does not occur in evaluating the statistical density matrix. The problem stems from the fact that the imaginary unit is involved in the exponent of the path integral (1.1). It is indeed the same fact that makes Feynman's formulation of quantum mechanics pathological.[13] In Sec. 2 and 3, we attempt to justify the practical use of the asymptotic formula (1.4) in quantum mechanics and to save most of the results previously obtained via the formula (1.4).

There have been proposed a number of methods to provide a mathematically sound basis for the Feynman path integral. Naturally we expect that some of them would be helpful to resolve our problem as well. The recent proposal of DeWitt–Morette[14] to reformulate the Feynman path integral by means of the notion of prodistribution seems most promising and must be applicable to our problem. However, since we follow the conventional time-division method of path integration, we borrow Nelson's idea[15] to derive the Hamiltonian path integral in polar coordinates within the framework of Nelson's theorem, as extended by Faris.[16] An unrigorous justification for the case of an inverse-squared potential is given in the Appendix. In Sec. 4, we discuss the difference between the Hamiltonian path integral so derived and the phase space path integral formally defined, emphasizing the importance of the role of symmetry involved.

2. THE HAMILTONIAN PATH INTEGRAL IN POLAR COORDINATES

Feynman's path integral (1.1) is customarily defined by the limiting process

$$K(\mathbf{r}'',\mathbf{r}';\tau) = \lim_{N\to\infty} A_N \int \exp\left[\frac{i}{\hbar}\sum_{j=1}^{N} \epsilon L\left(\mathbf{r}_j,\frac{\Delta\mathbf{r}_j}{\epsilon}\right)\right] d\mathbf{r}_1 d\mathbf{r}_2 \cdots d\mathbf{r}_{N-1}, \tag{2.1}$$

where $A_N=(m/2\pi i\hbar\epsilon)^{(3/2)N}$, $\epsilon=t_j-t_{j-1}=\tau/N$, $\Delta\mathbf{r}_j=\mathbf{r}_j-\mathbf{r}_{j-1}$, $\mathbf{r}_j=\mathbf{r}(t_j)$, $\mathbf{r}_0=\mathbf{r}(0)$ and $\mathbf{r}_N=\mathbf{r}(\tau)$. For a particle of mass m in a central potential the Lagrangian in the j-th interval can be given in polar coordinates as

$$L(\mathbf{r}_j,\Delta\mathbf{r}_j/\epsilon) = \tfrac{1}{2}m(r_j^2+r_{j-1}^2)/\epsilon^2 - (mr_jr_{j-1}/\epsilon^2)\cos\Theta_j - V(r_j) \tag{2.2}$$

with

$$\cos\Theta_j = \cos\theta_j\cos\theta_{j-1} + \sin\theta_j\sin\theta_{j-1}\cos(\phi_j-\phi_{j-1}). \tag{2.3}$$

Substitution of (2.2) into (2.1) yields the path integral expressed on the polar coordinate basis,

$$K(r'',\theta'',\phi'';r',\theta',\phi';\tau) = \lim_{N\to\infty} A_N \int \exp\left[\frac{i}{\hbar}\sum_{j=1}^{N}\left(\frac{1}{2}\frac{m}{\epsilon}(r_j^2+r_{j-1}^2) - \frac{m}{\epsilon}r_jr_{j-1} \times \cos\Theta_j - \epsilon V(r_j)\right)\right] \prod_{j=1}^{N-1}(r_j^2\sin\theta_j dr_j d\theta_j d\phi_j). \tag{2.4}$$

Use of the partial wave expansion formula[17]

$$\exp(z\cos\delta) = (\pi/2z)^{1/2}\sum_{l=0}^{\infty}(2l+1)P_l(\cos\delta)I_{l+1/2}(z) \tag{2.5}$$

given in terms of the Legendre functions $P_l(\cos\delta)$ and the modified Bessel functions $I_{l+1/2}(z)$, enables us to express (2.4) in the form,

$$K(r'',\theta'',\phi'';r',\theta',\phi';\tau) = \sum_{l=0}^{\infty}\sum_{n=-l}^{l} K_l(r'',r';\tau)Y_l^{n*}(\theta',\phi')Y_l^n(\theta'',\phi'') \tag{2.6}$$

with the radial propagator of the l wave

$$K_l(r'',r';\tau) = \lim_{N\to\infty}(4\pi\hbar)^N A_N \int \prod_{j=1}^{N}\left(\frac{i\hbar\pi\epsilon}{2mr_jr_{j-1}}\right)^{1/2} \times \exp\left(\frac{im}{2\hbar\epsilon}(r_j^2+r_{j-1}^2) - \frac{i\epsilon}{\hbar}V(r_j)\right) \times I_{l+1/2}\left(\frac{m}{i\hbar\epsilon}r_jr_{j-1}\right)\prod_{j=1}^{N-1}(r_j^2dr_j). \tag{2.7}$$

On the other hand, employing another expansion formula,[17]

$$\exp(z\cos\delta) = \sum_{\nu=-\infty}^{\infty}\exp(i\nu\delta)I_\nu(z), \tag{2.8}$$

we can also express the total propagator (2.4) in a different fashion,

$$K(r'',\theta'',\phi'';r',\theta',\phi';\tau) = \lim_{N\to\infty}(4\pi\hbar)^N A_N \int \prod_{j=1}^{N}\exp\left(\frac{im}{2\hbar\epsilon}(r_j-r_{j-1})^2 - \frac{i\epsilon}{\hbar}V(r_j)\right) \times \sum_{\mu_j}\exp[\mu_j(\theta_j-\theta_{j-1})] \times \exp\left(\frac{im}{\hbar\epsilon}r_jr_{j-1}\right)I_{\mu_j}\left(\frac{m}{i\hbar\epsilon}r_jr_{j-1}\right) \times \sum_{\nu_j}\exp[\nu_j(\phi_j-\phi_{j-1})]\exp\left(\frac{im}{\hbar\epsilon}r_jr_{j-1}\sin\theta_j\sin\theta_{j-1}\right) \times I_{\nu_j}\left(\frac{m}{i\hbar\epsilon}r_jr_{j-1}\sin\theta_j\sin\theta_{j-1}\right)\prod^{N-1}(r^2\sin\theta\, dr\, d\theta\, d\phi). \tag{2.9}$$

If use is made of the asymptotic formula (1.4), the results (2.7) and (2.9) can be brought to the forms (1.5) and (1.7), respectively. However, it must be pointed out that the asymptotic formula (1.4) is valid only if[17]

$$\mathrm{Re}\,z>0, \quad \text{or } |\arg z|<\pi/2. \tag{2.10}$$

The argument of the modified Bessel functions $I_\nu(z)$ involved in either (2.7) or (2.9) is purely imaginary, i.e., $\mathrm{Re}\,z=0$, or $\arg z=\pi/2$. Evidently the use of the asymptotic formula (1.4) for the polar coordinate calculation of the quantum mechanical propagator is improper. The asymptotic formula of $I_\nu(z)$ for large $|z|$ in the range $-3\pi/2<\arg z<\pi/2$ is available in the form,[17]

$$I_\nu(z)\sim(2\pi z)^{-1/2}\exp[z-\tfrac{1}{2}(\nu^2-\tfrac{1}{4})z^{-1}] + (2\pi z)^{-1/2} \times\exp[-z+\tfrac{1}{2}(\nu^2-\tfrac{1}{4})z^{-1}+i\pi(\nu+\tfrac{1}{2})]. \tag{2.11}$$

Under the restriction (2.10), the second term of (2.11) can be neglected, and the formula (2.11) is readily reduced to (1.4). Therefore, it appears to be a proper manner to use (2.11) in-

stead of (1.4) for the integrations of (2.7) and (2.9), where $\mathrm{Re}z = 0$. Unfortunately the asymptotic formula (2.11) does not lead us to any quantum-mechanically significant propagator.

To resolve this problem, we wish to suggest the following procedure. First we introduce a complex mass M in the place of the particle mass m so that the argument of the modified Bessel function may acquire a positive real part; namely, we write

$$M = m(1 + i\eta), \tag{2.12}$$

where $m = \mathrm{Re}M > 0$, and $m\eta = \mathrm{Im}M$ such that

$$\mathrm{Re}z = m\eta|z/M| > 0. \tag{2.13}$$

As a result, the second term of the asymptotic formula (2.11) drops off for large $|z|$, and the use of the formula (1.4) becomes proper. After completing the path integrations, we take the limit $\eta \to 0$ in order to recover the propagator for a real particle. Whether the exchange of the orders of the limiting processes $\eta \to 0$ and $N \to \infty$ is legitimate or not will be discussed in the next section. Here, conjecturing that it is so, we proceed to apply the proposed procedure to the path integrations in (2.7) and (2.9).

In calculating the path integral (2.7), we choose the imaginary part of the mass to be positive, i.e., $\eta > 0$. Consequently, the argument of the modified Bessel function in (2.7) has a positive real part,

$$\mathrm{Re}z = m\eta r_j r_{j-1}/(\hbar\epsilon), \tag{2.14}$$

which enables us to use the asymptotic formula (1.4). Until the form of the potential is specified, the path integrations cannot be explicitly carried out. However, with the help of the asymptotic formula (1.4), we can immediately rewrite (2.7) as

$$K_l(r'',r';\tau,\eta) = (r r')^{-1} \lim_{N\to\infty} \left(\frac{M}{2\pi i\hbar\epsilon}\right)^{N/2} \int \exp\left[\frac{i}{\hbar}\int^{\tau}\left(\tfrac{1}{2}M\dot{r}^2 - \frac{l(l+1)\hbar^2}{2Mr^2} - V(r)\right)dt\right] \prod_{j=1}^{N-1}(dr), \tag{2.15}$$

where $\mathrm{Im}M = m\eta > 0$. In the limit $\eta \to 0$, we would arrive at the l wave propagator (1.5) in quantum mechanics,

$$K_l(r'',r';\tau) = \lim_{\eta\to 0} K_l(r'',r';\tau;\eta). \tag{2.16}$$

The path integral (2.9) involves two modified Bessel functions, but the arguments of both functions can be made simultaneously to carry positive real parts insofar as the imaginary part of the mass is taken to be positive. Thus, by choosing $\eta > 0$, we can utilize the asymptotic formula (1.4) to expand both of the modified Bessel functions in (2.9). Using the following relation,

$$\exp\left(\frac{iM}{2\hbar\epsilon}(\Delta r)^2\right) = \left(\frac{i\hbar\epsilon}{2\pi M}\right)^{1/2} \int_{-\infty}^{\infty} \exp\left[\frac{i}{\hbar}\left(p\Delta r - \frac{\epsilon}{2M}p^2\right)\right]dp \tag{2.17}$$

in addition to the asymptotic formula (1.4) in (2.9) leads to

$$K(r'',\theta'',\phi'';r',\theta',\phi';\tau;\eta) = (r'^2 r''^2 \sin\theta' \sin\theta'')^{-1/2} \lim_{N\to\infty} (2\pi\hbar)^{-3N} \times \sum_{\mu_1\mu_2\cdots\mu_N} \sum_{\nu_1\nu_2\cdots\nu_N} \int\int \prod_{j=1}^{N} \exp\left\{\frac{i}{h}(p_j\Delta r_j + \mu_j\hbar\theta_j + \nu_j\hbar\Delta\phi_j - H_j\epsilon)\right\} \prod_{j=1}^{N-1}(dr\,d\theta\,d\phi) \prod_{j=1}^{N}(dp), \tag{2.18}$$

where

$$\Delta r_j = r_j - r_{j-1},\ \Delta\theta_j = \theta_j - \theta_{j-1},\ \Delta\phi_j = \phi_j - \phi_{j-1},$$

and

$$H_j = p_j^2/(2M) + (\mu_j^2 - \tfrac{1}{4})\hbar^2/(2Mr_j r_{j-1}) + (\nu_j^2 - \tfrac{1}{4})\hbar^2/(2Mr_j r_{j-1}\sin\theta_j \sin\theta_{j-1}) + V(r_j). \tag{2.19}$$

Again we expect that the above path integral (2.18) will yield in the limit $\eta_1 \to 0$ the correct propagator for a real particle,

$$K(r'',\theta'',\phi'';r',\theta',\phi';\tau) = \lim_{\eta\to 0} K(r'',\theta'',\phi'';r',\theta',\phi';\tau;\eta). \tag{2.20}$$

3. CONVERGENCE TO THE PHYSICAL PROPAGATORS

In proposing a way out of the difficulty associated with the use of the asymptotic formula (1.4), we have interchanged the order of the limiting processes $\eta \to 0$ and $N \to \infty$ in path integrations. A question remains to be answered as to whether the limit could really converge to the physical propagator for a real mass. An answer to this question has been prepared by Nelson[15] in establishing the connection between the Feynman and Wiener integrals by means of an analytic continuation in the mass parameter. We would merely like to exploit his result. Let us first summarize briefly Nelson's arguments in a way directly applicable to our problem.

Consider the Schrödinger equation,

$$-i\hbar\partial_t u(\mathbf{r},t) = \{(\hbar^2/2m)\nabla^2 - V(\mathbf{r})\}u(\mathbf{r},t). \tag{3.1}$$

Here the potential $V(\mathbf{r})$ is a real function of the coordinate variable $\mathbf{r}$ alone, and $u(\mathbf{r},t)$ is square-integrable for any time t, i.e., $u \in L^2(\mathbf{R}^3)$. Let $\psi(\mathbf{r})$ be $u(\mathbf{r},0)$. Define

$$X(t;m) = \exp\{(i\hbar t/2m)\nabla^2\} \tag{3.2}$$

on the domain $D(\nabla^2)$ of all ψ, where ∇^2 is self-adjoint, and

$$Y(t;V) = \exp(-itV/\hbar) \tag{3.3}$$

on the domain $D(V)$ of all ψ, where V is the self-adjoint operator of multiplication by the function V.

If V is a Kato potential,[18] that is, if V fulfills the following conditions:

(i) $D(V) \supset D(\nabla^2)$,

(ii) for all positive a, there always exists an a-dependent b such that for all $\psi \in D(\nabla^2)$

$$\|V\psi\| \leq a\|\nabla^2\psi\| + b\|\psi\|,$$

then the operator $(\hbar^2/2m)\nabla^2 - V$ is self-adjoint and according to a theorem of Trotter[19]

$$U(t;m,V)\psi = \lim_{N\to\infty} \{X(t/N;m)Y(t/N;V)\}^N \psi, \tag{3.4}$$

where

$$U(t;m,V)=\exp[(i\hbar t/2m)\nabla^2-itV/\hbar]. \tag{3.5}$$

In particular, in the absence of V, the solution of (3.1) is given by

$$u(\mathbf{r},t)=U(t;m,0)\psi(\mathbf{r})=X(t,m)\psi(\mathbf{r}), \tag{3.6}$$

or, more explicitly,

$$u(\mathbf{r},t)=(m/2\pi i\hbar t)^{3/2}\int\exp\{\tfrac{1}{2}im[(\mathbf{r}-\mathbf{r}')^2/\hbar]t\}\psi(\mathbf{r}')d\mathbf{r}'. \tag{3.7}$$

Similarly, in the presence of a Kato potential which is sufficiently regular, (3.4) can be written as

$$U(t;m,V)\psi(\mathbf{r}'')=\int K(\mathbf{r}'',\mathbf{r}';t)\psi(r')dr', \tag{3.8}$$

where

$$K(\mathbf{r}'',\mathbf{r}';t)=\lim_{N\to\infty}\left(\frac{m}{2\pi i\hbar\epsilon}\right)^{3N/2}\times\int\exp\left(\frac{i}{\hbar}\sum_{j=1}^{N}\frac{1}{2}\frac{m}{\epsilon}[\Delta\mathbf{r}_j^2-\epsilon V(\mathbf{r}_j)]\right)\prod_{j=1}^{N-1}a\mathbf{r}_j, \tag{3.9}$$

which coincides with the definition of the Feynman path integral.

If the mass parameter m in (3.1) is a complex number, the results of Kato and Trotter are not applicable. However, Nelson devises a means to show that whenever $\mathrm{Im}m>0$ and ψ is square-integrable, the limit (3.4) exists for a potential is real and continuous on the complement of a closed set F in $\mathbf{R}^3$ of capacity 0. In particular, if m is purely imaginary with $\mathrm{Im}m>0$, then $U(t;m,V)$ (3.4) can be expressed in terms of the Wiener integral defined on the space Ω of all paths, where $V(\omega(t))$ is a continuous function of t for $0\leqslant t<\infty$ for almost every $\omega\in\Omega$. In general, the solution is not defined by a path integral if m is not purely imaginary. A path integral may be formed for a complex mass system, but it is not assured to be convergent. We are interested in knowing whether $U(t;M,V)\psi$ with a complex mass $M=m(1+i\eta)$ has a limit for $\eta\to 0$.

Nelson proves for the above-mentioned almost-everywhere-continuous potentials the following theorem: There is a set Q of real numbers of Lebesque measure 0 such that for all $m\notin Q$, $\psi\in L^2$ and $t\geqslant 0$

$$U(t;m,V)\psi(\mathbf{r})=\lim_{\eta\to 0}V(t;M,V)\psi(\mathbf{r}) \tag{3.10}$$

exists when M approaches m nontangentially from the upper half-plane; the operators $U(t;m,V)$ so defined have the semigroup property,

$$U(t;m,V)U(s;m,V)=U(t+s;m,V) \tag{3.11}$$

for $0\leqslant t,s<\infty$, and for all ψ in L^2, $t\to U(t;m,V)$ is continuous from $[0,\infty]$ to L^2.

As a consequence, we see that if V is a Kato potential,

$$\lim_{\eta\to 0}\lim_{N\to\infty}[X(t/N;M)Y(t/N;V)]^N\psi(\mathbf{r}), \tag{3.12}$$

has a limit for all M with $\eta\geqslant 0$ *and* $m\neq 0$, which is expressed in the form of the Feynman path integral (3.8). This certainly justifies our proposed procedure applied to (2.16) and (2.21) provided that the potential involved is of the Kato class. A recent work of Faris[16] has indicated that the Kato condition on the potential can be relaxed. Faris has found that the expression (3.12) converges to the Feynman path integral (3.8) even if V is a Rollnik potential defined by the condition

$$\int\int\frac{V(r)V(r')}{|\mathbf{r}-\mathbf{r}'|^2}d\mathbf{r}d\mathbf{r}'<\infty. \tag{3.13}$$

Correspondingly our propagators (2.16) and (2.21) may include the Rollnik potentials.

In fact, Nelson's theorem deals with a wider class of potentials. However, special care has to be given to each singular potential concerning the restriction on the mass of the strength of coupling, since the path integral does not always converge for all real m. As an example, Nelson has examined the inverse-square potential which does not belong to the Rollnik class, and suggested that the solution of the Schrödinger equation, involving a highly attractive singular potential, obtained by a Feynman integral, is physically relevant even though the operator may not be unitary. In the Appendix, we shall also discuss the inverse-square potential in a different manner and show that the proposed limiting procedure works for the singular potential.

4. IS THE HAMILTONIAN PATH INTEGRAL A PHASE SPACE PATH INTEGRAL?

At the end of Sec. 2, we have obtained what we have called the Hamiltonian path integral in polar coordinates (2.18) for the particle with a complex mass M. This result, as has been observed in Sec. 3, is valid for $\mathrm{Im}M>0$ and has a limit for $\mathrm{Im}M\to 0$ insofar as a restricted class of potentials is involved. We may safely suggest that the polar coordinate formulation of the Hamiltonian path integral is possible at least for the Rollnik potential and the inverse square potential although there is no guarantee that the path integrations can actually be carried out.

It is certainly interesting to ask if the Hamiltonian path integral that can be reached from (2.18) after the limiting procedure is equivalent to a phase space path integral obtainable from (1.2) by a formal substitution of polar variables.

The Hamiltonian path integral (2.18) with a real mass may be expressed as

$$\begin{aligned}K(r'',\theta'',\phi'';r',\theta',\phi';\tau)&=(r'^2r''^2\sin\theta'\sin\theta'')^{-1/2}\lim_{\eta\to 0}\lim_{N\to\infty}(2\pi\hbar)^{-3N}\\&\times\sum_{\mu_1\cdots\mu_N}\sum_{\nu_1\cdots\nu_N}\int\int\prod_{j=1}^{N}\delta(\mu-p_\theta/\hbar)\delta(\nu-p_\phi/\hbar)\\&\times\exp\left\{\frac{i}{\hbar}(p_r\Delta r+p_\theta\Delta\theta+p_\phi\Delta\phi-H_{\mathrm{eff}}\epsilon)\right\}\\&\times\prod^{N-1}(dr d\theta d\phi)\prod^{N}(dp_r dp_\theta dp_\phi)\end{aligned} \tag{4.1}$$

with

$$H_{\text{eff}} = p_r^2/(2M) + (p_\theta^2 - \tfrac{1}{4}\hbar^2)/(2Mr^2)$$

$$+(p_\phi^2 - \tfrac{1}{4}\hbar^2)/(2Mr^2\sin^2\theta) + V(r). \tag{4.2}$$

On the other hand, the path integral (1.2), when expressed formally in terms of polar coordinates and their conjugate momenta, takes on the form

$$P(r'',\theta'',\phi'';r',\theta',\phi';\tau)$$

$$= (r'^2 r''^2 \sin\theta'\sin\theta'')^{-1/2} \lim_{\eta\to 0}\lim_{N\to\infty} (2\pi\hbar)^{-3N}$$

$$\times \int\int \prod_{j=1}^{N} \exp\left(\frac{i}{\hbar}(p_r\Delta r + p_\theta\Delta\theta + p_\phi\Delta\phi - H\epsilon)\right)$$

$$\times \prod^{N-1} (dr d\theta d\phi) \prod^{N} (dp_r dp_\theta dp_\phi), \tag{4.3}$$

where H is the classical Hamiltonian,

$$H = p_r^2/(2M) + p_\theta^2/(2Mr^2) + p_\phi^2/(2Mr^2\sin^2\theta) + V(r). \tag{4.4}$$

This phase space integral differs from (4.1) in the following two points. First, the Hamiltonian (4.4) differs from (4.2) by an additive term $(1+\csc^2\theta)\hbar^2/(8Mr^2)$. Secondly, the values of the momenta conjugate to the angular variables are discrete in (4.1), whereas they are continuous in (4.3).

The midpoint prescription used by McLaughlin and Schulman[5] and by Gervais and Jevicki[6] in dealing with point canonical transformations in path integrals leads us to an additional potential term which can account for the difference of the Hamiltonians (4.2) and (4.4). The phase space path integral one can derive from (1.2) on the polar coordinate basis by following the same prescription is of the form of (4.3) with the effective Hamiltonian (4.2) instead of (4.4). The result appears to be identical[20] with the formal expression (1.7) but differs from it since (1.7) actually means the expression (4.1). Formal integrations of (1.7) over the continuous angular momenta give us an effective Lagrangian,

$$L_{\text{eff}} = L_{\text{cl}} + (1+\csc^2\theta)\hbar^2/(8mr^2) - i\hbar(\dot{r}/r + \tfrac{1}{2}\dot{\theta}\cot\theta) \tag{4.5}$$

which is not our starting Lagrangian. However if we take the contributions only from discrete values of the angular momenta, we can get back to the starting Lagrangian L_{cl}.

The appearance of the quantized angular momenta in (4.1) is due to the rotational symmetry the system carries. The second gap must be filled by a proper account of constraints due to the symmetry.[21] In order to obtain the path integral of the form (4.3) with (4.2) via the midpoint prescription, one has to assume that the angular intervals $\Delta\theta$ and $\Delta\phi$ range from $-\infty$ to ∞ although the normal ranges are $0\leqslant\Delta\theta\leqslant\pi$ and $0\leqslant\Delta\phi\leqslant 2\pi$. In fact, because of the underlying spherical symmetry, one can introduce new variables ranging from $-\infty$ to ∞ by $\hat{\theta}=\Delta\theta+2\pi\mu$ and $\hat{\phi}=\Delta\phi+2\pi\nu$ ($\mu,\nu=0,\pm1,\pm2,\cdots$). However, the set of indices (μ,ν) specifies a class of homotopically equivalent paths, and the total propagator is the sum of all propagators belonging to different homotopy classes, i.e., $K=\sum_{\mu,\nu} K(\mu,\nu)$. Now we can understand the Hamiltonian path integral (4.1) with (4.2) as a consequence of such constraints. The phase space integral that can be reached on the polar basis by the midpoint prescription describes only a partial progagator for a central potential problem.

In the following, we wish to illustrate, by taking the two-dimensional rotator as an example, how the quantized angular momentum comes about. Since the validity of the η limit is obvious in this example, we simply use a real mass. The Lagrangian for the rotator is

$$L = \tfrac{1}{2}I\dot{\phi}^2. \tag{4.6}$$

If $-\infty<\phi<\infty$ is assumed, one can get the propagator,

$$K(\Delta\phi;\tau) = (I/2\pi i\hbar\tau)^{1/2}\exp[iI(\Delta\phi)^2/2\hbar\tau)], \tag{4.7}$$

from the one-dimensional free particle propagator,

$$K(\Delta x;\tau) = (mr^2/2\pi i\hbar\tau)^{1/2}\exp[im(\Delta x)^2/2\hbar\tau)],$$

by setting $\Delta x = R\Delta\phi$ and $I = mR^2$. However the result is not quite correct for the rotator. We must notice that the polar variable ϕ ordinarily ranges from 0 to 2π and that the space is doubly connected. Therefore, by retaining $\Delta\phi$ in the range $0\leqslant\Delta\phi\leqslant 2\pi$, we replace $\Delta\phi$ by $\hat{\phi} = \Delta\phi + 2\pi n$, and write (4.7) as

$$K(\hat{\phi};\tau) = (I/2\pi i\hbar\tau)^{1/2}\exp[iI\hat{\phi}^2/(\hbar\tau)], \tag{4.8}$$

or

$$K_n(\Delta\phi;\tau) = (I/2\pi i\hbar\tau)^{1/2}\exp[iI(\Delta\phi+2\pi n)^2/2\hbar\tau)], \tag{4.9}$$

For $n=0$, (4.9) reduces to the form of (4.7), describing the propagator for a rotation by an angle $\Delta\phi = |\Delta\phi|\leqslant 2\pi$. For $n\neq 0$, (4.9) represents n full rotations plus a rotation by the angle $\Delta\phi$. Namely, the integral number n classifies all homotopically inequivalent rotations with a fixed $\Delta\phi$. Therefore, the total propagator must be given by

$$K(\Delta\phi;\tau) = (I/2\pi i\hbar\tau)^{1/2} \sum_{n=-\infty}^{\infty}$$

$$\times \exp[iI(\Delta\phi+2\pi n)^2/2\hbar\tau)], \tag{4.10}$$

which can easily be transformed with the aid of the property of the theta function[17] to the standard form,

$$K(\Delta\phi;\tau) = (2\pi)^{-1} \sum_{\nu=-\infty}^{\infty} \exp(i\hbar\nu^2\tau/2I)\exp(i\nu\Delta\phi). \tag{4.11}$$

We can also get the result (4.10) directly from the path integral for the Lagrangian (4.6) subjected to the periodic constraint,[11] $\delta(\hat{\phi}-\int\dot{\phi}dt)$.

Next we rewrite (4.10) as

$$K(\Delta\phi;\tau)=(I/2\pi i\hbar\tau\omega)^{1/2}\int\sum_{n=-\infty}^{\infty}$$

$$\times\exp[iI\hat{\phi}^2/2\hbar\tau)]\delta(\Delta\phi-\hat{\phi}+2\pi n)d\hat{\phi}. \tag{4.12}$$

Furthermore, using the Poisson formula,

$$\sum_{\nu=-\infty}^{\infty}\exp(i\nu\theta)=2\pi\sum_{n=-\infty}^{\infty}\delta(\theta+2\pi n),$$

and the integral formula,

$$\int_{-\infty}^{\infty}\exp[i(ap-bp^2)]dp=(\pi/ib)^{1/2}\exp(ia^2/4b),$$

we put (4.12) in the form

$$K(\Delta\phi;\tau)=\frac{1}{2\pi\hbar}\iint\sum_{\nu=-\infty}^{\infty}\exp[i\nu(\Delta\phi-\hat{\phi})]$$

$$\times\exp\left(\frac{ip}{\hbar}\hat{\phi}-\frac{i\tau\hbar}{2I}p^2\right)dpd\hat{\phi}. \tag{4.13}$$

Integrating over $\hat{\phi}$, we arrive at the expression

$$K(\Delta\phi;\tau)=\sum_{\nu}\int\delta(\nu-p/\hbar)\exp\left(\frac{i}{\hbar}(p\Delta\phi-H\tau)\right)dp \tag{4.14}$$

with

$$H=\frac{1}{2I}p^2. \tag{4.15}$$

This relation holds true for any short time propagator $K(\Delta\phi;\epsilon)$. On the other hand, from the semigroup property of the propagator we have

$$K(\Delta\phi;\tau)=\lim_{N\to\infty}\int\prod_{j}^{N}K(\Delta\phi_j;\epsilon)\prod_{j}^{N-1}(d\phi_j). \tag{4.16}$$

Therefore, substitution of (4.14) for each short time propagator in (4.16) yields

$$K(\Delta\phi;\tau)=\lim_{N\to\infty}\sum_{\nu_1\cdots\nu_N}\int\prod_{j}^{N}\delta(\nu-p/\hbar)$$

$$\times\exp\left(\frac{i}{\hbar}(p\Delta\phi-H\epsilon)\right)\prod^{N-1}(d\phi)\prod^{N}(dp). \tag{4.17}$$

This is the Hamiltonian path integral which gives us the correct propagator (4.11), and is a special case of (4.1). The additive term $-\hbar^2/(8I)$, which is missing in (4.15), is a constant in this case and gives rise only to a constant phase difference which may be eliminated in the normalization process.

APPENDIX

Here we wish to present an unrigorous argument concerning the validity of the proposed limiting procedure for a inverse-square potential.

For the harmonic oscillator which belongs to the Rollnik class, the path integral (2.7) has been calculated directly without resort to the asymptotic formula (1.4), the result being[8]

$$K_l(r'',r';\tau)=(r\dot{r}'')^{-1/2}(m\omega/i\hbar)\csc(\omega\tau)$$

$$\times\exp\{(im\omega/2\hbar)(r'^2+r''^2)\cot(\omega\tau)\}$$

$$\times I_{l+1/2}[(m\omega/i\hbar)r\dot{r}''\csc(\omega\tau)] \tag{A1}$$

for $\mathrm{Re}m>0$. Thus the validity of the limiting process (2.16) is obvious in this particular potential. By comparing (2.15) and (A1), we find the following formula,[22]

$$\lim_{N\to\infty}\left(\frac{m}{2\pi i\hbar\epsilon}\right)^{N/2}\int\exp\left[\frac{i}{\hbar}\int^{\tau}\left(\tfrac{1}{2}m\dot{r}^2\right.\right.$$

$$\left.\left.-\frac{(\lambda^2-\frac{1}{4})\hbar^2}{2mr^2}-\tfrac{1}{2}m\omega^2r^2\right)dt\right]\prod^{N-1}(dr)$$

$$=(r\dot{r}'')^{-1/2}(m/i\hbar\tau)\csc(\omega\tau)\exp[(im\omega/2\hbar)(r'^2+r''^2)$$

$$\times\cot(\omega\tau)]\,I_{\lambda}[(m\omega/i\hbar)r\dot{r}''\csc(\omega\tau)], \tag{A2}$$

valid for $\mathrm{Re}\lambda>-1$, $\mathrm{Re}m>0$, and $\mathrm{Im}m\geqslant 0$. Now we set $\omega=0$ and

$$\lambda(l)=[(l+\tfrac{1}{2})^2+2mk/\hbar^2]^{1/2} \tag{A3}$$

to obtain

$$\int\exp\left[\frac{i}{\hbar}\int^{\tau}\left(\tfrac{1}{2}m\dot{r}^2-\frac{l(l+1)\hbar^2}{2mr^2}-\frac{k}{r^2}\right)dt\right]\mathscr{D}r$$

$$=(r\dot{r}'')^{-1/2}(m/i\hbar\tau)\exp[im(r'^2+r''^2)/(2\hbar\tau)]$$

$$\times I_{\lambda(l)}[(m/i\hbar\tau)r\dot{r}''], \tag{A4}$$

which is valid for $\mathrm{Re}\lambda>-1$. Although the physical range is $\lambda^2>0$, i.e., $k\geqslant-\frac{1}{2}0l+\frac{1}{2})^2\hbar^2/m$, the identity (A4) is valid for any $l\geqslant 0$ and any k real. This is consistent with Nelson's result.[15]

[1]R.P. Feynman, Rev. Mod. Phys. **20**, 367 (1948); R.P. Feynman and A.R. Hibbs, *Quantum Mechanics and Path Integrals* (McGraw-Hill, New York, 1965).
[2]C. Garrod, Rev. Mod. Phys. **38**, 483 (1966) and references therein.
[3]L.D. Faddeev, Theor. Math. Phys. **1**, 3 (1969).
[4]D.I. Blokhintsev and B.M. Barbashov, Sov. Phys. Usp. **15**, 193 (1972).
[5]D.W. McLanghlin and L.S. Schulman, J. Math. Phys. **12**, 2520 (1971).
[6]J.L. Gervais and A. Jevicki, Nucl. Phys. B **110**, 93 (1976).
[7]S.F. Edwards and Y.V. Gulyaev, Proc. R. Soc. London Ser. A **279**, 229 (1964).
[8]D. Peak and A. Inomata, J. Math. Phys. **10**, 1422 (1969).
[9]M.J. Goovaerts, J. Math. Phys. **16**, 720 (1975).
[10]D. Khandekar and S.V. Lawande, J. Math. Phys. **18**, 712 (1977).
[11]A. Inomata and V.A. Singh, J. Math. Phys. **19**, 2318 (1978).
[12]M. Gutzwiller, J. Math. Phys. **8**, 1970 (1967).
[13]See, for example, I.M. Gel'fand and A.M. Yaglom, J. Math. Phys. **1**, 48 (1960).
[14]C. DeWitt-Morette and A. Maheshwari, "Path Integrals in Non-Relativistic Quantum Mechanics", a monograph to appear in Phys. Rep.
[15]E. Nelson, J. Math. Phys. **5**, 332 (1963).
[16]W. Faris, a work unpublished but cited by B. Simon, *Quantum Mechanics for Hamiltonians Defined as Quadratic Forms* (Princeton U. P., Princeton, N.J., 1971), p. 52.
[17]See for example, E.T. Whittacker and G.N. Watson, *A Course of Modern Analysis* (Cambridge U. P., Cambridge, 1962), 4th ed.
[18]T. Kato, Trans. Am. Math. Soc. **70**, 195 (1951).
[19]H.F. Trotter, Proc. Am. Math. Soc. **10**, 345 (1959); Pac. J. Math. **8**, 887 (1958); T. Kato, Proc. Jpn. Acad. **35**, 467 (1959).
[20]A.M. Arthurs, Proc. R. Soc. London Ser. A **318**, 523 (1970), in which a similar result is obtained in a different way and claimed to be equivalent to (2.18).
[21]Simple systems with a periodic constraint are discussed in Ref. 11.
[22]The formula (56) given in Ref. 8 erroneously contains the ω-dependent singular potential.

Tensor harmonics in the helicity basis

D. A. Akyeampong

Department of Mathematics, University of Ghana, Legon, Ghana
(Received 5 April 1978)

Instead of the spherical basis vectors used to study tensor harmonics, we propose the use of the helicity basis. The rank-j tensor harmonics in this basis are explicitly worked out and their properties exhibited.

INTRODUCTION

The basis vectors frequently used in the study of tensor harmonics are the spherical ones $\mathbf{e}_\mu(\mu = \pm 1,0)$ which are related to the unit vectors $\mathbf{e}_x$, $\mathbf{e}_y$, and $\mathbf{e}_z$ of the three-dimensional Euclidean space $E(3)$ by the equations

$$\mathbf{e}_\pm = \mp \frac{1}{\sqrt{2}}(\mathbf{e}_x \pm i\mathbf{e}_y), \quad \mathbf{e}, \mathbf{e}_0 = \mathbf{e}_z. \tag{1.1a}$$

The vectors $\mathbf{e}_\mu$ satisfy the identity

$$\mathbf{e}_\mu^* = (-1)^\mu \mathbf{e}_{-\mu}, \tag{1.1b}$$

and the orthonormality condition

$$\mathbf{e}_\mu^* \cdot \mathbf{e}_\nu = \delta_{\mu\nu}, \tag{1.1c}$$

the star denoting complex conjugation.

Using these basis vectors and the scalar spherical harmonics $Y^1_m(\theta,\phi)$, one then obtains the vector spherical harmonics $\mathbf{Y}_{JlM}(\theta,\theta)(\mathrm{J} = l \pm 1,1)$ by coupling Y^1_m and $\mathbf{e}_\mu$ with the O(3) Clebsch–Gordon coefficients[1],

$$\mathbf{Y}_{JlM}(\theta,\phi) = \sum_\mu C(l\,1J;M-\mu,\mu) Y^l_{M-\mu}\mathbf{e}_\mu. \tag{1.2}$$

The vector spherical harmonic so constructed transforms according to the irreducible representation D^J of the O(3) group. Higher rank tensor spherical harmonics are also built up in the same way, namely by the coupling of the lower ones with the irreducible basis vectors in the product space through Clebsch–Gordan coefficients so as to transform according to a given representation of the O(3) group. Tensor multipoles needed, for example, for the study of electromagnetic and gravitational radiations, are then obtained by operating the unit vector $\mathbf{r}/r$, the gradient vector ∇, and the angular momentum vector $\mathbf{L}$ on the appropriate spherical harmonics. This is often a tedious procedure especially if higher rank harmonics are involved and it is our belief that it can be circumvented if a suitable choice of basis vectors is used.

For this and other reasons, therefore, we wish to study in this paper tensor harmonics using the helicity basis of Jacob and Wick.[2] Its chief merit, as we shall see, is that the tensor harmonics and tensor multipoles are treated on an equal footing, thereby eliminating those tedious algebraic manipulations referred to earlier.

In Sec. 2 we introduce the helicity basis states of Jacob and Wick and prove the important result that under a rotation the basis state is invariant up to a phase factor. The vector harmonic in this basis is discussed in Sec. 3 while tensor harmonics of rank $j(j \geqslant 2)$ are studied in Sec. 4. The general result is particularly simple; the tensor harmonics in the helicity basis are the product of an O(3)D-matrix and the basis vectors in the product space. Relations between these and the spherical harmonics are also given.

2. THE HELICITY BASIS

The operator Λ, defined by

$$\Lambda = \mathbf{S}\cdot\mathbf{n}$$

where $\mathbf{S}$ is the intrinsic angular momentum vector, and $\mathbf{n} = \mathbf{p}/|\mathbf{p}|$ is the unit vector in the direction of the linear momentum $\mathbf{p}$, is called the helicity operator. Its eigenvalues, which shall be denoted by λ, are either integers or half odd integers: In this study we shall only be interested in integer values of λ. In the helicity basis, therefore, the orientation of the spin $\mathbf{S}$ refers to a coordinate system whose z axis is in the direction of the momentum vector $\mathbf{p}$. Let G_z be the group consisting of all rotations $H_\mathbf{k}$ which leave invariant the unit vector $\mathbf{k}$ along the z axis.

$$H_\mathbf{k}\mathbf{k} = \mathbf{k}. \tag{2.1}$$

If we define a rotation $L_\mathbf{p}$ which takes the unit vector $\mathbf{k}$ into the direction of $\mathbf{p}$

$$L_\mathbf{p}\mathbf{k} = \mathbf{p}, \tag{2.2}$$

then, in terms of the Euler angles θ,ϕ and ψ

$$L_\mathbf{p} = \exp(-i\phi J_z)\exp(-i\theta J_y)\exp(-i\psi J_z), \tag{2.3}$$

where J_y and J_z are the angular momentum operators. Consider now a general rotation R which takes $\mathbf{P}$ into $\mathbf{p}'$, so that

$$\mathbf{p}' = R\mathbf{p} \tag{2.4}$$

and $p^2 = p'^2$. It then follows from Eqs. 2.2 and 2.4. that

$$L_{p'}^{-1} R L_\mathbf{p}\mathbf{k} = \mathbf{k},$$

so that $\mathbf{L}_{p'}^{-1} R L_\mathbf{p} \in G_z$, the group of rotations around the z-axis and for some $h \in G_z$, we have

$$h = L_{R\mathbf{p}}^{-1} R L_\mathbf{p}, \tag{2.5}$$

a rotation representing a special direction along the z-axis.

Denoting the helicity states by $\chi_{J\lambda}(\mathbf{p})$, the relation between $\chi_{J\lambda}$ and $\phi_{JM}(\mathbf{p})$, the angular momentum states with momentum $\mathbf{p}$, spin $\mathbf{J}$, and $J_z = M$, is given by

$$\chi_{J\lambda}(\mathbf{p}) = \sum_{MM'} D^J_{\lambda M'}(L_\mathbf{p}^{-1})\phi_{JM'}(\mathbf{p}), \tag{2.6}$$

where $D^J_{\lambda M}(L_\mathbf{p}^{-1}) = D^{J*}_{M\lambda}(L_\mathbf{p})$

In the rotated frame, we have

 0022-2488/79/030505-04$01.00

$$\chi'_{J\lambda}(\mathbf{p}') = \sum_{M'} D^{J}_{\lambda M'}(L_{\mathbf{p}}^{-1})\phi'_{JM'}(\mathbf{p}')$$

$$= \sum_{M'M''} D^{J}_{\lambda M'}(L_{R\mathbf{p}}^{-1}) D^{J}_{M'M''}(R)\phi_{JM''}(\mathbf{p})$$

$$= \sum_{\lambda'} D^{J}_{\lambda\lambda'}(h)\chi_{J\lambda'}(\mathbf{p}), \tag{2.7}$$

on using Eqs. (2.4) and (2.5). Since h is diagonal, the result

$$\chi'_{J\lambda}(\mathbf{p}') = [\exp i\alpha(\lambda)]\chi_{J\lambda}(\mathbf{p}), \tag{2.8}$$

α real, then follows. Thus under rotation $\chi_{J\lambda}$ is invariant up to a phase factor. It is this property which makes the helicity basis, in our view, a natural one to use.

The rotation of the coordinate system in such a way that the z axis is in the direction of $\mathbf{p}$ will require three Euler angles, the two polar angles θ, ϕ and the angle ψ which is arbitrary, since we are working on the unit sphere S^2. In the following, therefore, ψ will be chosen to be zero.

3. THE VECTOR HARMONICS

The vector harmonics in the helicity basis are easily obtained from Eq. (2.6) by observing that the spherical basis vectors $\mathbf{e}_\mu$ are related to the helicity basis vectors $\boldsymbol{\epsilon}_\lambda$ by the equation

$$\mathbf{e}_\mu = \sum_\lambda D^{1*}_{\mu\lambda}(\phi,\theta,0)\ \boldsymbol{\epsilon}_\lambda, \tag{3.1}$$

where

$$D^{J}_{MN}(\phi,\theta,\psi)$$
$$= \langle JM|\exp(-i\phi J_z)\exp(-i\theta j_y)\exp(-i\psi J_z)|JN\rangle. \tag{3.2}$$

From the orthonormality properties of the rotation matrices and of $\mathbf{e}_\mu$, we get

$$\boldsymbol{\epsilon}^*_\lambda = (-1)^\lambda \boldsymbol{\epsilon}_{-\lambda} \quad \text{and} \quad \boldsymbol{\epsilon}^*_\lambda \cdot \boldsymbol{\epsilon}_{\lambda'} = \delta_{\lambda\lambda'}. \tag{3.3}$$

Starting from the definition of a vector spherical harmonic, Eq. (1.2), we obtain the relation

$$\mathbf{Y}_{JlM}(\theta,\phi) = \sum_\lambda (-1)^{1+\lambda} C(Jl1; -\lambda\lambda)\mathbf{H}^{J}_{M\lambda}(\phi,\theta), \tag{3.4a}$$

or

$$(-1)^{l+\lambda}\mathbf{H}^{J}_{M\lambda}(\phi,\theta) = \sum_l C(Jl1;-\lambda\lambda)\,\mathbf{Y}_{JlM}(\theta,\phi), \tag{3.4b}$$

where

$$\mathbf{H}^{J}_{M\lambda}(\phi,\theta) = \left(\frac{2J+1}{4\pi}\right)^{1/2} D^{J*}_{M\lambda}(\phi,\theta,0)\ \boldsymbol{\epsilon}_\lambda \tag{3.5}$$

shall be referred to as the vector harmonic in the helicity basis, the scalar one being, of course, the usual

$$Y^{J}_{M}(\theta,\phi) = \left(\frac{2J+1}{4\pi}\right)^{1/2} D^{J*}_{M0}(\phi,\theta,0). \tag{3.6}$$

$\mathbf{H}^{J}_{M\lambda}$ is the eigenstate of J^2 and J_3 but not of L^2, a small price to pay considering other advantages. It satisfies the reality condition

$$\mathbf{H}^{J*}_{M\lambda} = (-1)^M \mathbf{H}^{J}_{-M-\lambda}, \tag{3.7}$$

and under the parity operation P we have

$$P{:}\mathbf{H}^{J}_{M\lambda} = (-1)^{J+1} H^{J}_{M-\lambda}. \tag{3.8}$$

The harmonics are also orthogonal in the vector space $E(3)$

$$\int d\Omega\ \mathbf{H}^{J*}_{M'\lambda'} \cdot \mathbf{H}^{J'}_{M'\lambda'} = \delta_{JJ'}\delta_{MM'}\delta_{\lambda\lambda'}, \tag{3.9}$$

with $d\Omega = \sin\theta d\theta d\phi$, $0 \leqslant \theta \leqslant \pi$, and $0 \leqslant \phi \leqslant 2\pi$. Hence any vector field $\mathbf{A}(\mathbf{r},t)$ can be expanded in terms of the vector harmonics as

$$\mathbf{A}(\mathbf{r},t) = \sum_{JM\lambda} A^{J}_{M\lambda}(r,t)\mathbf{H}^{J}_{M\lambda}(\phi,\theta), \tag{3.10}$$

where $A^{J}_{M\lambda}$, the scalar functions of $r(=|\mathbf{r}|)$ and t only, are given by

$$A^{J}_{M\lambda}(r,t) = \int d\Omega\ \mathbf{H}^{J}_{M\lambda}(\phi,\theta)\cdot\mathbf{A}(\mathbf{r},t). \tag{3.11}$$

This shows that vector equations with rotational symmetry such as Maxwell's equations can be solved in terms of these harmonics.

We mention in passing that the scalar function $\boldsymbol{\epsilon}^*_\lambda \cdot \mathbf{H}^{J}_{M\lambda}$ is the same as a function introduced some time ago by Newman and Penrose[3] and called the spin-λ spherical harmonic.

One obvious advantage of the use of the helicity basis is the ease with which the vector multipoles are obtained. In this formalism they are straightforwardly given as

$$\mathbf{H}^{J}_{M1} = \left(\frac{2J+1}{4\pi}\right)^{1/2} D^{J*}_{M1}\boldsymbol{\epsilon}_1,$$
$$\mathbf{H}^{J}_{M0} = \left(\frac{2J+1}{4\pi}\right)^{1/2} D^{J*}_{M0}\boldsymbol{\epsilon}_0, \tag{3.12}$$

and

$$\mathbf{H}^{J}_{M-1} = \left(\frac{2J+1}{4\pi}\right)^{1/2} D^{J*}_{M-1}\boldsymbol{\epsilon}_{-1},$$

to be compared with the usual method of deriving them via the application of the operators $\hat{\mathbf{r}}, r\nabla$, and $\mathbf{L}$ on the scalar harmonics.[4]

By an *ad hoc* procedure of introducing a new definition for the vector multipoles, Daumens and Minnaert[5] obtain results equivalent to ours. Our notations $\mathbf{H}_{\pm1}$, $\mathbf{H}_0$ are related to the transverse electric $\mathbf{X}_{-1}$, transverse magnetic $\mathbf{X}_{+1}$, and the longitudinal electric $\mathbf{X}_0$ of these authors by the equations

$$\mathbf{H}^{J}_{M0} = \mathbf{X}^{J}_{0M}$$
$$\mathbf{H}^{J}_{M\pm1} = \pm\frac{1}{\sqrt{2}}(\mathbf{X}^{J}_{+1M} \pm \mathbf{X}^{J}_{1M}) \tag{3.13}$$

It should be stated that although we obtain similar results the two approaches are very different.

4. TENSOR HARMONICS.

Because of their applications to problems involving gravitational radiation, rank-two tensor spherical harmon-

ics have also been studied by Regge and Wheeler,[6] Zerilli[7] and more recently by Daumens and Minnaert. These were constructed by coupling the tensors of the spherical basis and the spherical harmonics through Clebsch–Gordan coefficients. We evaluate in this section the second-rank tensor harmonics in the helicity basis. Using Eq. (3.5), it is a simple matter to see that the rank-two harmonics in this basis are

$$\mathbf{H}^{JS}_{M\lambda} = \left(\frac{2J+1}{4\pi}\right)^{1/2} D^{J*}_{M\lambda}(\phi,\theta,0)\ \mathbf{t}^{S}_{\lambda} \tag{4.1}$$

where $\mathbf{t}^{S}_{\lambda}(S=0,1,2;\lambda = \pm 2, \pm 1, 0)$, the nine basis vectors in the product space $E(3)\otimes E(3)$, are

$$\mathbf{t}^{S}_{\lambda} = \sum_{\lambda'} C(11S;\lambda-\lambda',\lambda')\boldsymbol{\epsilon}_{\lambda'}\otimes\boldsymbol{\epsilon}_{\lambda-\lambda'}, \tag{4.2}$$

with

$$(\mathbf{t}^{S}_{\lambda})^{*} = (-1)^{S+\lambda}\mathbf{t}^{S}_{-\lambda} \tag{4.3}$$

and

$$(\mathbf{t}^{S}_{\lambda})^{*}\cdot\mathbf{t}^{S'}_{\lambda'} = \delta_{SS'}\delta_{\lambda\lambda'}. \tag{4.4}$$

The harmonic (4.1) is related to the rank-two tensor spherical harmonics $\mathbf{Y}^{lS}_{JM}(\theta,\phi)$ by

$$\mathbf{Y}^{lS}_{JM}(\theta,\phi) = \sum_{\lambda}(-1)^{\lambda}C(JSl;-\lambda\lambda)\mathbf{H}^{JS}_{M\lambda}(\phi,\theta) \tag{4.5}$$

or

$$(-1)^{\lambda}\mathbf{H}^{JS}_{M\lambda}(\phi,\theta) = \sum_{l} C(JSl;-\lambda\lambda)\mathbf{Y}^{lS}_{JM}(\theta,\phi) \tag{4.6}$$

The H-functions satisfy the condition

$$\mathbf{H}^{JS\ *}_{M\lambda} = (-1)^{M+S}\mathbf{H}^{JS}_{M-\lambda}, \tag{4.7}$$

transform under the parity operation P as

$$P\colon \mathbf{H}^{JS}_{M\lambda} = (-1)^{J+S}\mathbf{H}^{JS}_{M-\lambda} \tag{4.8}$$

The tensors are orthogonal in the product space $E(3)\otimes E(3)$,

$$\mathbf{H}^{JS*}_{M\lambda}\cdot\mathbf{H}^{J'S'}_{M'\lambda'} = \frac{1}{4\pi}[(2J+1)(2J'+1)]^{1/2}D^{J}_{M\lambda}D^{J'}_{M'\lambda}\delta_{SS'}\delta_{\lambda\lambda'}, \tag{4.9}$$

and are also orthonormal,

$$\int d\Omega\ \mathbf{H}^{JS*}_{M\lambda}\cdot\mathbf{H}^{J'S'}_{M'\lambda'} = \delta_{JJ'}\delta_{MM'}\delta_{SS'}\delta_{\lambda\lambda'}. \tag{4.10}$$

Therefore in this basis, a tensor field $\overleftrightarrow{\mathbf{A}}(\mathbf{r},t)$ (using the dyadic notation) can be expanded in terms of $\mathbf{H}^{JS}_{M\lambda}(\phi,\theta)$ as

$$\overleftrightarrow{\mathbf{A}}(\mathbf{r},t) = \sum_{JMS\lambda} A^{JS}_{M\lambda}(r,t)\mathbf{H}^{JS}_{M\lambda}(\phi,\theta), \tag{4.11}$$

which gives, in general, nine quantities of helicities $\lambda = \pm 2, \pm 1, 0$; $S = 0,1,2$. Using Eq. (4.10), the expansion coefficients $A^{JS}_{M\lambda}(r,t)$ are obtained as

$$A^{JS}_{M\lambda} = \int d\Omega\ \mathbf{H}^{JS*}_{M\lambda}\cdot\overleftrightarrow{\mathbf{A}}. \tag{4.12}$$

The above analysis can thus clearly be used to study, for example, the weak-field radiative solutions of the Einstein equations.

The nine tensor multipoles are given in this basis to be simply

$$\mathbf{H}^{J2}_{M\pm 2},\quad \mathbf{H}^{J2}_{M\pm 1},\quad \mathbf{H}^{J2}_{M0},$$
$$\mathbf{H}^{J1}_{M\pm 1},\quad \mathbf{H}^{J1}_{M0},\quad \text{and}\quad \mathbf{H}^{J0}_{M0},$$

to be compared with the rather tedious way of obtaining them in the spherical basis.[7] Again, our results are related to those in Ref. 5 by means of the identifications[8]

$$(-1)^{S}\mathbf{X}^{SJ}_{0M} = \mathbf{H}^{JS}_{M0},$$
$$(-1)^{\lambda}\mathbf{X}^{SJ}_{\pm\lambda M} = \frac{1}{\sqrt{2}}(\mathbf{H}^{JS}_{M\mp\lambda} \pm (-1)^{\lambda}\mathbf{H}^{JS}_{M\pm\lambda}). \tag{4.13}$$

Construction of rank j tensors ($j>2$) is now easy. The problem clearly reduces to constructing a set of basis vectors in the product space $E(3)\otimes E(3)\otimes\cdots\otimes E(3)$ (j times), since the rotation matrix remains formally unchanged. By a simple inductive argument we can show that the rank-j tensor harmonic in the helicity basis is given by the expression

$$\mathbf{H}^{JS_1S_2\cdots S_j}_{M\lambda}(\phi,\theta) = \left(\frac{2J+1}{4\pi}\right)^{1/2} D^{J*}_{M\lambda}\mathbf{t}^{S_1S_2\cdots S_j}_{\lambda}, \tag{4.14}$$

where the basis vectors are

$$\begin{aligned}\mathbf{t}^{S_1S_2\cdots S_j}_{\lambda} &= \sum_{\lambda_1\lambda_2\cdots\lambda_j} C(11S_1;\lambda_1,\lambda_2-\lambda_1)C(S_1 1S_2;\lambda_2,\lambda_3-\lambda_2)\times\cdots\\ &\quad\times C(S_{j-2}1S_{j-1};\lambda_{j-1},\lambda_j-\lambda_{j-1})\\ &\quad\times C(S_{j-1}1S_j;\lambda_j,\lambda-\lambda_j)\\ &\quad\times\boldsymbol{\epsilon}_{\lambda_1}\otimes\boldsymbol{\epsilon}_{\lambda_2-\lambda_1}\otimes\boldsymbol{\epsilon}_{\lambda_3-\lambda_2}\otimes\cdots\otimes\boldsymbol{\epsilon}_{\lambda_j-\lambda_{j-1}}\otimes\boldsymbol{\epsilon}_{\lambda-\lambda_j}\\ &= \sum_{\lambda_1\lambda_2\cdots\lambda_j}\prod_{i=0}^{j} C(S_{i-1}1S_i;\lambda_i,\lambda_{i+1}-\lambda_i)\boldsymbol{\epsilon}_{\lambda_{i+1}-\lambda_i},\end{aligned}$$

with $S_{-1}=0$, $S_0=1$, $\lambda_0=0$, $\lambda_{j+1}=\lambda$. (4.15)

The orthogonality properties of the above harmonic then follow from those of the rotation matrix and of the helicity basis vectors $\boldsymbol{\epsilon}_{\lambda}$, as shown earlier.

Although a rank j tensor has 3^j components, it often happens that in certain physical situations, such as in general relativity, only the fully symmetric and traceless ones may be needed. For such particular cases, only the maximal coupling ($S_{j-1}=j$) are needed and the completely symmetric basis vectors are given by

$$\mathbf{t}^{j}_{\lambda} = \left[\frac{2^j(j+\lambda)!(j-\lambda)!}{(2j)!}\right]^{1/2}$$

$$\times \sum_{\lambda_i} \prod_{i=1}^{j} [(1+\lambda_i)!(1-\lambda_i)!]^{1/2} \epsilon_{\lambda_i}, \tag{4.16}$$

with $\lambda = \sum_{i=1}^{j} \lambda_i$.

5. CONCLUSIONS

We suggest in this paper the use of the helicity basis vectors for the study of the tensor harmonics and have derived the rank-j tensor harmonics in this basis. The tensor multipoles are then shown to be special cases of these harmonics. Equations possessing rotational invariance can be solved using this basis. The applications of these results to the Maxwell's equations and the weak-field approximations to the Einstein's equations are presently being studied.

ACKNOWLEDGMENTS

The author is indebted to Professor M. A. Rashid and Professor J. Strathdee for useful discussions. He is also indebted to Professor Abdus Salam for the hospitality extended to him during his stay at the International Center for Theoretical Physics, Trieste, Italy, where part of this work was done.

[1]See, for example, A. R. Edmonds, *Angular Momentum in Quantum Mechanics* (Princeton U. P., Princeton, New Jersey, 1960).
[2]M. Jacob and G. C. Wick, Ann. Phys. (N.Y.) **7**, 404 (1959).
[3]E. T. Newman and R. Penrose, J. Math. Phys. **7**, 863 (1966); Proc. R. Soc. A **305**, 175 (1968). See also H. E. Moses, Ann. Phys. (N.Y.) **41**, 166 (1967).
[4]M. E. Rose, *Multipole Fields* (Wiley, New York, 1955), p. 23.
[5]M. Daumens and P. Minnaert, J. Math. Phys. **17**, 1903 (1976).
[6]T. Regge and J. A. Wheeler, Phys. Rev. **108**, 1063 (1957).
[7]F. J. Zerilli, J. Math. Phys. **11**, 2203 (1970).
[8]Note the following corrections in Ref. 5. In Eq. (43a), $\hat{j}$ should read $\hat{J}$ and in Eq. (43b) a factor $\hat{l}/\hat{J}$ is missing under the summation sign.

On the problems of object restoration and image extrapolation in optics

M. Bertero

Istituto di Scienze Fisiche dell'Universitá and Istituto Nazionale di Fisica Nucleare, Genova, Italy

C. De Mol[a)]

Département de Mathématique, Université Libre de Bruxelles, Belgium

G. A. Viano

Istituto di Scienze Fisiche dell'Universitá and Istituto Nazionale di Fisica Nucleare, Genova, Italy
(Received 22 November 1977)

In this paper we consider the problems of object restoration and image extrapolation, according to the regularization theory of improperly posed problems. In order to take into account the stochastic nature of the noise and to introduce the main concepts of information theory, great attention is devoted to the probabilistic methods of regularization. The kind of the restored continuity is investigated in detail; in particular we prove that, while the image extrapolation presents a Hölder type stability, the object restoration has only a logarithmic continuity.

1. INTRODUCTION

In Fourier optics a vast amount of literature has been devoted to the problem of object restoration. The interested reader is referred to the review papers by Frieden[1] and Goodman.[2]

One of the points, which has been largely debated, concerns restoration beyond the diffraction limit in the presence of noise. As already stated[2,3] a significant improvement in resolution can be accomplished if the object is very poorly resolved by the optical system at the start; on the other hand, if the object is extremely complex at the start, improvement of resolution requires signal-to-noise ratios that are unrealistically high. One might argue that these limitations are due to the fact that the number of degrees of freedom of an image is finite.[3] In other words the image is "ambiguous" in the sense that many different objects can have one and the same image (within a given accuracy) and therefore the observer must necessarily make use of some *a priori* knowledge.[3]

This remark is in accordance with the fact that the object restoration belongs to a large class of linear inverse problems (relevant in many fields like identification of targets, probing of media, geophysical exploration, and in any other field of remote sensing) which are ill posed in the sense of Hadamard, since the solutions do not depend continuously on the data. Therefore, as it has been emphasized by many authors,[4-6] these problems must be reconsidered according to the regularization theory of improperly posed problems.[7,8] This is a means for introducing *a priori* knowledge under the form of precise *a priori* constraints (as far as possible of physical origin) which restrict the class of admitted solutions. Then regularization gives a continuous dependence of the solution on the data in the sense that a small variation of the data leads to a small change in the solution.

In this paper we consider an ideal, diffraction limited, space-invariant, imaging system. For one-dimensional, coherent objects, identically zero outside the interval $[-1,1]$, one has

$$y(t) = \int_{-1}^{1} \frac{\sin[c(t-s)]}{\pi(t-s)} x(s)ds + z(t), \quad |t| \leqslant 1, \tag{1.1}$$

where y is the image (assumed to be known on the interval $[-1,1]$), x is the object, and z is the noise; the quantity $d = \pi/c$ is the Rayleigh resolution distance.

We shall analyze the following problems:

(1) object restoration, i.e., to estimate the object x given y;

(2) image extrapolation, i.e., to estimate a band-limited function whose restriction to the interval $[-1,1]$ approximates y.

The main reason for our attention to the optical system described above is that, in this simple case, one knows the eigenfunctions of the integral operator of Eq. (1.1) [they are the linear prolate spheroidal wavefunctions[1] $\psi_k(c,x)$, with the associated eigenvalues $\lambda_k(c)$] and their asymptotic behvior when $k \to +\infty$. From this fact it derives the possibility of obtaining very precise results on problems (1) and (2).

Now, as we said above, the main prescription in order to restore the continuity in the case of ill-posed problems, consists in restricting the admitted solutions by imposing a supplementary global bound[7,8]; next one must investigate the kind of continuity which has been restored.[9] In some problems it is possible to have a fairly satisfactory Hölder type stability; i.e., if we denote by ϵ the data accuracy, then the solution accuracy is dominated by a term proportional to ϵ^{λ} $(0 < \lambda < 1)$. On the other hand there are problems where the solution accuracy is at best proportional to $|\ln\epsilon|^{-1}$ (logarithmic continuity). One of the main results of this paper consists in showing that the restored continuity is of the Hölder type in the extrapolation of a given image piece beyond its

[a)]Aspirant du Fonds National Belge de la Recherche Scientifique.

 0022-2488/79/030509-13$01.00

borders [Problem (2)], while in the object restoration [Problem (1)] it is only logarithmic.

In the usual regularization theory of ill-posed problems, one supposes that the error and the solution lie in some bounded sets of the respective Hilbert spaces. These conditions could appear too rigid, since it should be preferable, from a physical point of view, to regard the noise as a stochastic process. Furthermore, after the advent of the Shannon information theory, many concepts which were originally elaborated in that context, were successively largely applied to the analysis of optical systems.[10] These are the principal motivations for introducing probabilistic methods of regularization of ill-posed problems. We devote Sec. 3 to probabilistic methods in a Hilbert space setting, along the lines introduced by two of the authors;[11] furthermore a comparison of the two procedures is discussed in detail.

The paper is organized as follows. In Sec. 2 we analyze the regularization method proposed by Tikhonov and Miller,[7,8] with attention to the specific questions which are of interest in the optical problems here considered. In Sec. 3 we develop the probabilistic method. Section 4 is devoted to the problem of object restoration; here we prove that the restored continuity is only logarithmic. In Sec. 5 we analyze the problem of image extrapolation, showing that in this case the restored continuity is of the Hölder type. Finally in Sec. 6 we try some conclusions.

2. TIKHONOV–MILLER METHOD

In this Section we sketch a method first proposed by Tikhonov[7] for Fredholm equations of the first kind and independently developed by Miller[8] for improperly posed problems in a Hilbert space setting. Here we follow the paper of Miller[8] for the presentation of the general method; we also give some results on stability estimates which, to our knowledge, are not available in the literature.

Let X, Y be separable Hilbert spaces and $A: X \to Y$ a linear continuous operator such that the inverse operator A^{-1} exists; if $A^{-1}: Y \to X$ is not continuous, the problem: find $x \in X$ such that $y = Ax$, where $y \in Y$ is a given vector, is improperly posed in the sense of Hadamard. In practice this pathology is very serious, since the data are always affected by errors. More precisely, the data vector y can be viewed as the sum of two terms: $y = Ax + z$, where $x \in X$ is an unknown vector which has to be approximately determined and $z \in Y$ is an unknown vector describing errors or noise. Since A^{-1} is not continuous, the knowledge of y and of a bound on the error z is not sufficient in order to find an approximation to x. Further "*a priori*" knowledge on x is required.

A. The general method

The basic point is to assume for the error $z = Ax - y$ and for the unknown vector x the following prescribed bounds: $\|z\|_Y = \|Ax - y\|_Y \leqslant \epsilon$, $\|Bx\|_Z \leqslant E$, where $B: X \to Z$ (Z is a Hilbert space) is a linear operator (the so-called *constraint operator*[8]), densely defined in X, which has a bounded inverse; ϵ, E are given positive numbers (we do not consider the case where only one of the two numbers ϵ and E is known[8]). Then, any vector $\bar{x} \in X$ satisfying the conditions:

$$\|A\bar{x} - y\|_Y \leqslant \epsilon \tag{2.1}$$

$$\|B\bar{x}\|_Z \leqslant E \tag{2.2}$$

can be called an approximation to the unknown vector x. Such a definition is reasonable if:

(i) there exists at least one vector $\bar{x}$ satisfying (2.1), (2.2) (in such a case the pair ϵ, E is called *permissible*[8]);

(ii) there exists a norm or seminorm $\langle\cdot\rangle$ such that

$$M(\epsilon,E) = \sup\{\langle x\rangle \mid x \in X, \|Ax\|_Y \leqslant \epsilon, \|Bx\|_Z \leqslant E\} \tag{2.3}$$

tends to zero, as $\epsilon \to 0$, for fixed E.

When (ii) holds true, we say that Problem (2.1), (2.2), i.e., the problem of finding a vector $\bar{x}$ satisfying the conditions (2.1), (2.2) is stable with respect to the norm or seminorm $\langle\cdot\rangle$. Any upper bound for $M(\epsilon,E)$ is called a *stability estimate* and $M(\epsilon,E)$ itself is called the *best possible stability estimate*. $M(\epsilon,E)$ gives the size, in the sense of the norm or seminorm $\langle\cdot\rangle$, of the "packet" of the vectors $\bar{x}$ satisfying (2.1), (2.2); indeed if $\bar{x}_1$, $\bar{x}_2$ satisfy (2.1), (2.2), then $\|A(\bar{x}_1 - \bar{x}_2)\|_Y \leqslant 2\epsilon$, $\|B(\bar{x}_1 - \bar{x}_2)\|_Z \leqslant 2E$, so that $\langle x_1 - x_2\rangle \leqslant 2M(\epsilon,E)$. Now, if $\{\epsilon,E\}$ is a permissible pair, one has to find a method in order to exhibit explicitly a vector which satisfies the constraints (2.1), (2.2).

Let $K^y_{\epsilon,E}$ be the set of all the vectors $\bar{x}$ satisfying (2.1), (2.2). $K^y_{\epsilon,E}$ is convex and bounded; besides the unknown vector x belongs to $K^y_{\epsilon,E}$. If $\bar{x} \in K^y_{\epsilon,E}$, then $\|\bar{x} - x\|_X$ is a measure, in the sense of the norm of X, of the error which is done by taking $\bar{x}$ as an approximation to x. Since x is unknown, a pessimistic estimate of this error is given by

$$\mathscr{E}(\bar{x}) = \sup\{\|\bar{x} - x\|_X \mid x \in K^y_{\epsilon,E}\}. \tag{2.4}$$

It is not difficult to prove that there always exists a vector $\bar{x}_0$ which minimizes the functional $\mathscr{E}(\cdot)$ and belongs to $K^y_{\epsilon,E}$; then we can take such a vector as a "*best-possible*" approximation to x. Besides, if there exists a center of symmetry of $K^y_{\epsilon,E}$, then it coincides with $\bar{x}_0$. Unfortunately the set $K^y_{\epsilon,E}$ usually does not have a center of symmetry; therefore it can be quite hard to find a vector which minimizes $\mathscr{E}(\cdot)$. However we can combine the two constriants (2.1), (2.2) into a single one and introduce the convex, bounded set

$$\tilde{K}^y_{\epsilon,E} = \left\{ x \,\middle|\, x \in X, \frac{1}{2\epsilon^2}\|Ax - y\|_Y^2 + \frac{1}{2E^2}\|Bx\|_Z^2 \leqslant 1 \right\}. \tag{2.5}$$

Then $K^y_{\epsilon,E} \subset \tilde{K}^y_{\epsilon,E} \subset K^y_{\sqrt{2}\epsilon,\sqrt{2}E}$. We do not make a large error if we consider $\tilde{K}^y_{\epsilon,E}$ instead of $K^y_{\epsilon,E}$ and corresponding $\tilde{\mathscr{E}}(\cdot)$ instead of $\mathscr{E}(\cdot)$, $\tilde{\mathscr{E}}(\cdot)$ being defined as in Eq. (2.4), where $K^y_{\epsilon,E}$ is replaced by $\tilde{K}^y_{\epsilon,E}$. Since $\tilde{K}^y_{\epsilon,E}$ has a center of symmetry which is given by

$$\tilde{x} = \left[A^*A + \left(\frac{\epsilon}{E}\right)^2 B^*B\right]^{-1} A^*y, \tag{2.6}$$

$\tilde{x}$ minimizes $\tilde{\mathscr{E}}(\cdot)$ and it can be taken as a "*nearly-best-possible*" approximation to x. Indeed, recalling that $\tilde{K}^{y}_{\epsilon,E} \subset K^{y}_{\sqrt{2}\epsilon,\sqrt{2}E}$ we can conclude that the approximation $\tilde{x}$ is best-possible but for a factor of $\sqrt{2}$. Finally, it is important to note that this method gives an approximation to x, which is satifactory independently of the choice of norm (or seminorm) for measuring the error. Indeed, one can prove that[8]

$$\langle \tilde{x} - x \rangle \leqslant \sqrt{2} M(\epsilon,E), \tag{2.7}$$

where $M(\epsilon,E)$ is defined by Eq. (2.3).

The previous method suggests the introduction of the following stability estimate,

$$\tilde{M}(\epsilon,E) = \sup\left\{ \langle x \rangle \middle| x \in X, \frac{1}{2\epsilon^2} \|Ax - y\|_Y^2 + \frac{1}{2E^2} \|Bx\|_Z^2 \leqslant 1 \right\}. \tag{2.8}$$

It is clear that $\tilde{M}(\epsilon,E)$ is an upper bound for $M(\epsilon,E)$; more precisely one has

$$M(\epsilon,E) \leqslant \tilde{M}(\epsilon,E) \leqslant \sqrt{2} M(\epsilon,E) \tag{2.9}$$

and, instead of inequality (2.7), one gets

$$\langle \tilde{x} - x \rangle \leqslant \tilde{M}(\epsilon,E). \tag{2.10}$$

B. Stability estimates

Here we give results on the stability of Problem (2.1), (2.2) with respect to those norms and seminorms which are more appropriate for the problems considered in this paper; the stability can be achieved by properly choosing the constraint operator B.

(i) Let us assume that the operator B has a bounded inverse, without any further property; then Problem (2.1), (2.2) is stable with respect to the family of seminorms: $\langle x \rangle = |(x,u)_X|$, $\|u\|_X = 1$ (where u is a fixed but arbitrary vector on the unit sphere).

The following result is proved in the paper of Miller[8]:

Theorem 2.1: If $C = A^*A + (\epsilon/E)^2 B^*B$, then

$$\tilde{M}_u(\epsilon,E) = \sqrt{2}\epsilon (C^{-1}u,u)_X^{1/2}, \tag{2.11}$$

$\tilde{M}_u(\epsilon,E)$ being defined by Eq. (2.8) with $\langle\cdot\rangle$ replaced by $|(\cdot,u)_X|$.

From Theorem 2.1 we can derive the following result:

Theorem 2.2: $\tilde{M}_u(\epsilon,E)$ tends to zero, as $\epsilon \to 0$, for fixed E; besides $\tilde{M}_u(\epsilon,E) = O(\epsilon)$ iff $u \in$ range (A^*).

Proof: If we write $|B| = (B^*B)^{1/2}$ and if we introduce the operator $S = A|B|^{-1}$ (S^{-1} exists since A^{-1} exists), then Eq. (2.11) can be rewritten as follows,

$$\tilde{M}_u(\epsilon,E) = \sqrt{2}\epsilon \left(\left[S^*S + \left(\frac{\epsilon}{E}\right)^2 I \right]^{-1} |B|^{-1}u, |B|^{-1}u \right)_X^{1/2}. \tag{2.12}$$

For the sake of simplicity we suppose that the operator A is compact (however the Theorem holds true even without this assumption). Then the operator S is compact and we denote by $\{s_k^2\}_{k=0}^{+\infty}$ ($s_k > 0$; $s_0 \geqslant s_1 \geqslant s_2 \geqslant \cdots$) the set of the eigenvalues of the operator S^*S and by $\{u_k\}_{k=0}^{+\infty}$ the set of the corresponding eigenvectors. $\{u_k\}_{k=0}^{+\infty}$ is an orthonormal basis in X so that we can write

$$\tilde{M}_u(\epsilon,E) = \sqrt{2}\epsilon \left(\sum_{k=0}^{+\infty} \frac{|(|B|^{-1}u,u_k)_X|^2}{s_k^2 + (\epsilon/E)^2} \right)^{1/2}. \tag{2.13}$$

From this equation it is clear that $\tilde{M}_u(\epsilon,E)$ tends to zero, as $\epsilon \to 0$, for fixed E. Besides, let us assume that $u \in$ range (A^*); then there exists a vector $v \in Y$ such that $u = A^*v$. If we introduce the vectors $v_k = s_k^{-1} S u_k$ [$\{v_k\}_{k=0}^{+\infty}$ is an orthonormal basis in range (S)[12]], then Eq. (2.13) becomes

$$\tilde{M}_u(\epsilon,E) = \sqrt{2}\epsilon \left(\sum_{k=0}^{+\infty} \frac{s_k^2}{s_k^2 + (\epsilon/E)^2} |(v,v_k)_Y|^2 \right)^{1/2} \tag{2.14}$$

so that $\tilde{M}_u(\epsilon,E) \leqslant \sqrt{2}\|v\|_Y \epsilon$ and $\tilde{M}_u(\epsilon,E) = O(\epsilon)$.

Now, let us assume $\tilde{M}_u(\epsilon,E) = O(\epsilon)$; then there exists a constant δ_u such that

$$\sum_{k=0}^{+\infty} \frac{|(|B|^{-1}u,u_k)_X|^2}{s_k^2 + (\epsilon/E)^2} \leqslant \delta_u. \tag{2.15}$$

If K is the maximum integer such that $s_k \geqslant \epsilon/E$, then $K \to +\infty$ when $\epsilon \to 0$; therefore

$$\sum_{k=0}^{K} s_k^{-2} |(|B|^{-1}u,u_k)_X|^2 \leqslant 2 \sum_{k=0}^{K} \frac{|(|B|^{-1}u,u_k)_X|^2}{s_k^2 + (\epsilon/E)^2} \leqslant 2\delta_u \tag{2.16}$$

for any K. We get

$$\sum_{k=0}^{+\infty} s_k^{-2} |(|B|^{-1}u,u_k)_X|^2 \leqslant 2\delta_u \tag{2.17}$$

or, in other words, $|B|^{-1}u \in$ range (S^*). Then there exists a vector $v \in Y$ such that $|B|^{-1}u = S^*v$ and it follows that $u = A^*v$. The Theorem is proved.

(ii) Let us assume now that the operator B has a compact inverse. Then the set $K_E = \{x | x \in X, \|Bx\|_Z \leqslant E\}$ is a compact subset of X; analogously the set $J_E = AK_E$ is a compact subset of Y and the operator A defines a continuous, one to one mapping of K_E onto J_E. From a well known topological lemma[13] it follows that the inverse mapping from J_E onto K_E is continuous; such a result implies that $M(\epsilon,E)$ $= \sup\{\|x\|_X | x \in X, \|Ax\|_Y \leqslant \epsilon, \|Bx\|_Z \leqslant E\}$ $= \sup\{\|A^{-1}y\|_X | y \in Y, \|y\|_Y \leqslant \epsilon, \|BA^{-1}y\|_Z \leqslant E\}$ tends to zero, as $\epsilon \to 0$, for fixed E. Therefore, one can conclude that: *when the operator B has a compact inverse, Problem* (2.1), (2.2) *is stable with respect to the norm of X.*

The previous result is essentially qualitative; we can have a more precise result (see Theorem 2.3) in the following case. Let us assume that $A : X \to Y$ is a compact operator (such an assumption is satisfied for the problems considered in Secs. 4 and 5); then, let $\{\alpha_k^2\}_{k=0}^{+\infty}$ ($\alpha_k > 0$; $\alpha_0 \geqslant \alpha_1 \geqslant \alpha_2 \geqslant \cdots$)

be the set of the eigenvalues of the operator A^*A and $\{u_k\}_{k=0}^{+\infty}$ be the set of the corresponding eigenvectors,

$$A^*Ax = \sum_{k=0}^{+\infty} \alpha_k^2 x_k u_k, \quad x_k = (x,u_k)_X. \tag{2.18}$$

In such a case a considerable simplification is introduced if one considers a constraint operator B such that B^*B and A^*A commute, i.e.,

$$B^*Bx = \sum_{k=0}^{+\infty} \beta_k^2 x_k u_k, \quad x_k = (x,u_k)_X. \tag{2.19}$$

Then *the operator B has a compact inverse iff* $\lim_{k\to+\infty}\beta_k^2 = +\infty$. It is self evident that the condition is necessary. In order to show that it is also sufficient, one must prove that any set $K_E = \{x \in X \mid \|Bx\|_Z \leqslant E\}$ is conditionally compact. Such a result follows easily from the proposition[14]: *A subet K of* l_p, $p \geqslant 1$, *is conditionally compact iff it is bounded and* $\lim_{n\to\infty}\Sigma_{k=n}^{+\infty}|x_k|^p = 0$ *uniformly for* $\{x_k\}_{k=0}^{+\infty} \in K$.

Now, let $I_{\epsilon/E}$ be the set defined as follows,

$$I_{\epsilon/E} = \{k \mid \alpha_k \geqslant (\epsilon/E)\beta_k\}. \tag{2.20}$$

Since A^*A is compact, the eigenvalues α_k^2 accumulate to zero; on the other hand, the eigenvalues β_k^2 of B^*B tend to infinity. Therefore, the set $I_{\epsilon/E}$ contains only a finite number of points.

Besides, if we write

$$\alpha\left(\frac{\epsilon}{E}\right) = \min\{\alpha_k \mid k \in I_{\epsilon/E}\},$$

$$\beta\left(\frac{\epsilon}{E}\right) = \min\{\beta_k \mid k \notin I_{\epsilon/E}\}, \tag{2.21}$$

we have the following result:

Theorem 2.3: The inequality holds,

$$\frac{1}{2}\left(\frac{\epsilon}{\alpha(\epsilon/E)} + \frac{E}{\beta(\epsilon/E)}\right) \leqslant M(\epsilon,E) \leqslant \frac{\epsilon}{\alpha(\epsilon/E)} + \frac{E}{\beta(\epsilon/E)}, \tag{2.22}$$

where $M(\epsilon,E)$ is given by Eq. (2.3) with $\langle\cdot\rangle = \|\cdot\|_X$.

Proof: Inequality (2.22) is a consequence of a lemma (three norm lemma) proved by Miller[15] which states that

$$\tfrac{1}{2}[L(\epsilon,E) + H(\epsilon,E)] \leqslant M(\epsilon,E) \leqslant L(\epsilon,E) + H(\epsilon,E), \tag{2.23}$$

where:

$$L(\epsilon,E) = \sup\left\{\left(\sum_{k\in I_{\epsilon/E}} |x_k|^2\right)^{1/2} \mid x \in X, \left(\sum_{k\in I_{\epsilon/E}} \alpha_k^2 |x_k|^2\right)^{1/2} \leqslant \epsilon\right\}, \tag{2.24}$$

$$H(\epsilon,E) = \sup\left\{\left(\sum_{k\notin I_{\epsilon/E}} |x_k|^2\right)^{1/2} \mid x \in X, \left(\sum_{k\notin I_{\epsilon/E}} \beta_k^2 |x_k|^2\right)^{1/2} \leqslant E\right\}. \tag{2.25}$$

Since the functionals involved in Eqs. (2.24), and (2.25), are linear in the parameters $|x_k|^2$, it is easy to show that

$$L(\epsilon,E) = \frac{\epsilon}{\alpha(\epsilon/E)}, \quad H(\epsilon,E) = \frac{E}{\beta(\epsilon/E)}, \tag{2.26}$$

and the Theorem is proved.

In the particular case $B = (A^*)^{-1}$ we can obtain a more precise result:

Theorem 2.4: If $B = (A^*)^{-1}$, then

$$M(\epsilon,E) \leqslant \sqrt{\epsilon E}, \tag{2.27}$$

where $M(\epsilon,E)$ is defined by Eq. (2.3) with $\langle\cdot\rangle = \|\cdot\|_X$. Besides the equality holds when the ratio ϵ/E is equal to one of the eigenvalues of the operator A^*A.

Proof: From the Schwarz inequality we have, for any $x \in$ range (A^*)

$$\|x\|_X^2 = (A^{-1}Ax,x)_X = (Ax,(A^*)^{-1}x)_Y \leqslant \|Ax\|_Y \|(A^*)^{-1}x\|_Y, \tag{2.28}$$

and the inequality (2.27) follows from the constraints $\|Ax\|_Y \leqslant \epsilon$, $\|(A^*)^{-1}x\|_Y \leqslant E$. Besides, if the pair $\{\epsilon,E\}$ is such that $\epsilon/E = \alpha_k^2$, where α_k^2 is an eigenvalue of A^*A and if we take $x = E\alpha_k u_k$ (u_k is the eigenvector corresponding to α_k^2), then $\|Ax\|_Y = E\alpha_k^2 = \epsilon$, $\|(A^*)^{-1}x\|_Y = E$; on the other hand, $\|x\|_X = E\alpha_k = \sqrt{\epsilon E}$ and we can conclude that, in such a case, the equality holds in (2.27).

(iii) Let us assume now that the inverse of B^*B is an operator of the trace class. In other words, there exists an orthonormal basis $\{u_k\}_{k=0}^{+\infty}$ in X such that the operator B^*B has the spectral representation (2.19) and

$$\sum_{k=0}^{+\infty} \beta_k^{-2} < +\infty. \tag{2.29}$$

In the domain of B we can introduce the following norm (which shall be used in Sec. 5),

$$|||x||| = \sum_{k=0}^{+\infty} |x_k|, \quad x_k = (x,u_k)_X; \tag{2.30}$$

indeed, by the Schwarz inequality

$$|||x||| \leqslant \left(\sum_{k=0}^{+\infty} \beta_k^{-2}\right)^{1/2}\left(\sum_{k=0}^{+\infty} \beta_k^2 |x_k|^2\right)^{1/2} = \left(\sum_{k=0}^{+\infty} \beta_k^{-2}\right)^{1/2} \|Bx\|_Z. \tag{2.31}$$

Now, the set $K_E\{x \in X \mid \|Bx\|_Z \leqslant E\}$, is compact with respect to the norm (2.30); indeed, by the Schwartz inequality and condition (2.29),

$$\sum_{k=n}^{+\infty} |x_k| \leqslant E\left(\sum_{k=n}^{+\infty} \beta_k^{-2}\right)^{1/2} \to 0, \quad n \to +\infty \tag{2.32}$$

uniformly for $x \in K_E$. The compactness criterion in l_p, $p \geqslant 1$, already recalled[14] implies that K_E is a conditionally compact

subset of l_1. Besides it is easy to see that the set K_E is closed with respect to the norm (2.30).

If we put $\beta = \min_k \{\beta_k\} > 0$ and we remark that from the constraint $\|Bx\|_Z \leqslant E$ it follows that $|x_k| \leqslant E/\beta$, we get, for any $x \in K_E$, $\|x\|_X \leqslant (E/\beta)^{1/2} |||x|||$, so that

$$\|Ax\|_Y \leqslant \|A\| \, \|x\|_X \leqslant \sqrt{\frac{E}{\beta}} \, \|A\| \, |||x|||. \tag{2.33}$$

Therefore, if we set $J_E = AK_E$, the operator A defines a continuous mapping of K_E onto J_E, when K_E is normed with the norm (2.30) and J_E is normed with the norm of Y. The topological lemma already recalled,[13] implies that the inverse mapping from J_E onto K_E is continuous. Therefore, $M(\epsilon,E)$, defined by Eq. (2.3) with $\langle\cdot\rangle = |||\cdot|||$, tends to zero, as $\epsilon \to 0$, for fixed E; we can conclude that: *When condition* (2.29) *is satisfied, Problem* (2.1), (2.2) *is stable with respect to the norm* (2.30).

When A^*A and B^*B are given by Eqs. (2.18), (2.19), we can get the following more precise result.

Theorem 2.5: If $M(\epsilon,E)$ is defined by Eq. (2.3) with $\langle\cdot\rangle = |||\cdot|||$, then

$$\frac{\epsilon}{2}\left(\sum_{k \in I_{\epsilon/E}} \alpha_k^{-2}\right)^{1/2} + \frac{E}{2}\left(\sum_{k \notin I_{\epsilon/E}} \beta_k^{-2}\right)^{1/2} \leqslant M(\epsilon,E) \leqslant \epsilon \times \left(\sum_{k \notin I_{\epsilon/E}} \alpha_k^{-2}\right)^{1/2} + E\left(\sum_{k \notin I_{\epsilon/E}} \beta_k^{-2}\right)^{1/2}, \tag{2.34}$$

where $I_{\epsilon/E}$ is the set defined in Eq. (2.20).

Proof: By the three norm lemma[15] we have for $M(\epsilon,E)$ an inequality like (2.23) where now

$$L(\epsilon,E) = \sup\left\{\sum_{k \in I_{\epsilon/E}} |x_k| \,\middle|\, \left(\sum_{k \in I_{\epsilon/E}} \alpha_k^2 |x_k|^2\right)^{1/2} \leqslant \epsilon\right\}, \tag{2.35}$$

$$H(\epsilon,E) = \sup\left\{\sum_{k \notin I_{\epsilon/E}} |x_k| \,\middle|\, \left(\sum_{k \notin I_{\epsilon/E}} \beta_k^2 |x_k|^2\right)^{1/2} \leqslant E\right\}, \tag{2.36}$$

Then, by the Schwarz inequality (which is precise) we have

$$L(\epsilon,E) = \epsilon\left(\sum_{k \in I_{\epsilon/E}} \alpha_k^{-2}\right)^{1/2}, \quad H(\epsilon,E) = E\left(\sum_{k \notin I_{\epsilon/E}} \beta_k^{-2}\right)^{1/2} \tag{2.37}$$

and the Theorem is proved.

We conclude with a result which is a modified version of Theorem 2.5.

Theorem 2.6: If $\tilde{M}(\epsilon,E)$ is defined by Eq. (2.8) with $\langle\cdot\rangle = |||\cdot|||$, then

$$\tilde{M}(\epsilon,E) = \sqrt{2}\epsilon\left(\sum_{k=0}^{+\infty} \frac{1}{\alpha_k^2 + (\epsilon/E)^2\beta_k^2}\right)^{1/2}. \tag{2.38}$$

Proof: Indeed, by the Schwarz inequality, we have

$$\sum_{k=0}^{+\infty} |x_k| = \sum_{k=0}^{+\infty} \left(\frac{\alpha_k^2 + (\epsilon/E)^2\beta_k^2}{\alpha_k^2 + (\epsilon/E)^2\beta_k^2}\right)^{1/2} |x_k|$$

$$\leqslant \left(\|Ax\|_Y^2 + \left(\frac{\epsilon}{E}\right)^2 \|Bx\|_Z^2\right)^{1/2} \times \left(\sum_{k=0}^{+\infty} \frac{1}{\alpha_k^2 + (\epsilon/E)^2\beta_k^2}\right)^{1/2}. \tag{2.39}$$

Since the Schwarz inequality is precise, we get Eq. (2.38).

C. Methods of eigenfunction expansions

In the case where the operator $A: X \to Y$ is compact and the operators A^*A and B^*B commute [see Eqs. (2.18), (2.19)], Eq. (2.6) takes the form

$$\tilde{x} = \sum_{k=0}^{+\infty} \frac{\alpha_k}{\alpha_k^2 + (\epsilon/E)^2\beta_k^2} y_k u_k, \quad y_k = (y, v_k)_Y, \tag{2.40}$$

where $v_k = \alpha_k^{-1} A u_k$ [$\{v_k\}_{k=0}^{+\infty}$ is an orthonormal basis in $\overline{\text{range}(A)}$[12]]. An approximation, even more simple than (2.40), is given by

$$\tilde{\tilde{x}} = \sum_{k \in I_{\epsilon/E}} \frac{y_k}{\alpha_k} u_k, \tag{2.41}$$

where the set $I_{\epsilon/E}$ is defined by Eq. (2.20) and y_k, v_k are as in Eq. (2.40).

If $\bar{x}$ is any vector satisfying the conditions (2.1), (2.2), then

$$\|A(\tilde{\tilde{x}} - \bar{x})\|_Y \leqslant \sqrt{2}\epsilon, \quad \|B(\tilde{\tilde{x}} - \bar{x})\|_Z \leqslant \sqrt{2}E; \tag{2.42}$$

indeed, if $\bar{x}_k = (\bar{x}, u_k)$,

$$\|A(\tilde{\tilde{x}} - \bar{x})\|_Y^2 = \sum_{k \in I_{\epsilon/E}} |y_k - \alpha_k \bar{x}_k|^2 + \sum_{k \notin I_{\epsilon/E}} \alpha_k^2 |\bar{x}_k|^2$$

$$\leqslant \sum_{k \in I_{\epsilon/E}} |y_k - \alpha_k \bar{x}_k|^2 + \left(\frac{\epsilon}{E}\right)^2 \sum_{k \notin I_{\epsilon/E}} \beta_k^2 |x_k|^2 \tag{2.43}$$

$$\leqslant \|A\bar{x} - y\|_Y^2 + \left(\frac{\epsilon}{E}\right)^2 \|B\bar{x}\|_Z^2 \leqslant 2\epsilon^2.$$

In a similar way one can prove the second inequality (2.42). Therefore, we get

$$\langle \tilde{\tilde{x}} - \bar{x} \rangle \leqslant \sqrt{2} M(\epsilon,E), \tag{2.44}$$

$M(\epsilon,E)$ being defined by Eq. (2.3). We can conclude that, if Problem (1), (2) is stable with respect to the norm or seminorm $\langle\cdot\rangle$, both $\tilde{x}$ and $\tilde{\tilde{x}}$ converge (in the sense of $\langle\cdot\rangle$) to the "true" solution of the problem, as $\epsilon \to 0$, for fixed E.

3. PROBABILISTIC METHOD

Let us assume, as in Sec. 2, that the data vector y is the sum of two terms, $y = Ax + z$, where x is the vector which has to be approximately determined and z is an unknown vector describing errors or noise; then, the basic point in the probabilistic method is to consider the vectors x, y, z as the values of random variables ξ, η, ζ.[11,16]

More precisely, let $(\Omega,\mathcal{F},P)$ be a probability space, i.e., Ω is an abstract point set, $\mathcal{F}$ is a σ-algebra of subsets of Ω, and P is a measure on $\mathcal{F}$ with $P(\Omega)=1$; besides, let X, Y be separable Hilbert spaces and let $\xi:\Omega\to X$, $\zeta:\Omega\to Y$ be weak random variables[17] (hereafter shortened to w.r.v.), i.e., the mappings ξ, ζ induce cylinder measures μ_ξ, μ_ζ on X, Y, respectively; then we consider the w.r.v. $\eta:\Omega\to Y$ given by

$$\eta = A\xi + \zeta, \tag{3.1}$$

where $A:X\to Y$ is a linear continuous operator whose inverse A^{-1} exists but is not continuous.

If we assume that the joint measure of the pair of w.r.v. ξ,ζ is know, then the problem is the following: Given a value y of the w.r.v. η, find the best-possible mean square estimate of the w.r.v. ξ.

Now, we assume (as it is reasonable in many practical applications) that *the w.r.v. ξ, ζ are Gaussian and independent and that they have zero mean*; in such a case their joint measure is uniquely characterized by the covariance operators R_ξ, R_ζ. Then the covariance operator R_η is expressed in terms of R_ξ, R_ζ by,[11,16]

$$R_\eta = AR_\xi A^* + R_\zeta, \tag{3.2}$$

Besides, the w.r.v. ξ, η are not independent and their cross-covariance operator is given by

$$R_{\xi\eta} = R_\xi A^*. \tag{3.3}$$

In the following we shall also assume that *the inverse of the operator R_ζ exists*. Indeed, if this condition is not satisfied, there exists a vector v_0 such that $R_\zeta v_0 = 0$; then, the random variable $(\zeta,v_0)_Y$ takes the value zero with probability one or, in other words, the component of the data vector in the direction of v_0 is not affected by error or noise. Since such a situation is not realistic, we assume that the inverse of R_ζ exists. From Eq. (3.2) it follows that the inverse of the operator R_η also exists; besides, range (R_ζ) and range (R_η) are both dense in Y.

The covariance operators R_ξ, R_ζ have a role, in the probabilistic method, similar to that of the bounds ϵ, E and of the constraint operator B in the Tikhonov–Miller method. For this reason we assume that: $R_\zeta = \epsilon^2 N$, where ϵ is a "small" positive number and $N:Y\to Y$ is a linear bounded operator independent of ϵ.

A. The general method

If $L:Y\to X$ is any linear continuous operator, we call the w.r.v. $\tilde{\xi}_L = L\eta$ a *linear estimate* of the w.r.v. ξ. Then, for any $u\in X$, the reliability of the estimate for the random variable $(\xi,u)_X$ is measured by the mean square error (here E denotes the expectation)

$$\begin{aligned} Q(u;L,\epsilon) &= E\{|(\xi - L\eta,u)_X|^2\} \\ &= ([R_\xi - R_\xi A^* L^* - LAR_\xi + LR_\eta L^*]u,u)_X. \end{aligned} \tag{3.4}$$

If there exists a unique linear continuous operator $L_0:Y\to X$ which minimizes $Q(u;L,\epsilon)$ for any $u\in X$, then the w.r.v. $\tilde{\xi} = L_0\eta$ is said to be the *best linear estimate* of the w.r.v. ξ.

We have the following result[17,11]:

Theorem 3.1: *There exists a unique best linear estimate $\tilde{\xi}$ of the w.r.v. ξ iff the operator $R_{\xi\eta}R_\eta^{-1}$ is bounded on the range (R_η); in such a case*

$$L_0 = R_{\xi\eta}R_\eta^{-1} = R_\xi A^*(AR_\xi A^* + \epsilon^2 N)^{-1}. \tag{3.5}$$

Remark 3.1: If R_ζ has a bounded inverse, then the operator R_η also has a bounded inverse. In such a case the product $R_{\xi\eta}R_\eta^{-1}$ is a bounded operator and therefore there is no question in the interpretation of Eq. (3.5). On the other hand, if the inverse of R_ζ is not continuous, then $R_{\xi\eta}R_\eta^{-1}$ is the product of two operators where $R_{\xi\eta}$ is bounded while R_η^{-1} is unbounded and densely defined. [see Eq. (3.2)–we recall that A^{-1} is not continuous.] If $R_{\xi\eta}R_\eta^{-1}$ is bounded on range (R_η), then Theorem 3.1 states that L_0 is the usual extension (by continuity) of $R_{\xi\eta}R_\eta^{-1}$ to the closure of range (R_η), i.e., to the whole Hilbert space Y. For simplicity we still denote by $R_{\xi\eta}R_\eta^{-1}$ [as in Eq. (3.5)] such an extension.

Remark 3.2: If we consider the problem formulated at the beginning of this section, then the solution is: Given a value y of the w.r.v. η, the best possible mean-square estimate of the w.r.v. ξ is

$$\tilde{x} = R_\xi A^*(AR_\xi A^* + \epsilon^2 N)^{-1} y. \tag{3.6}$$

It is interesting to compare formally Eq. (3.6) with Eq. (2.6). If we put $N = I$ and $R_\xi = E^2(B^*B)^{-1}$ in Eq. (3.6) and we use the identity

$$(A^*R_\zeta^{-1}A + R_\xi^{-1})R_\xi A^* = A^*R_\zeta^{-1}(AR_\xi A^* + R_\zeta), \tag{3.7}$$

we see that, in such a case, Eqs. (3.6) and (2.6) coincide.

Remark 3.3: In the case $N = I$ (i.e., in the case where ζ is the so-called "white noise") there is an interesting interpretation of the least-square estimate (3.6). We introduce the operator $T = R^{1/2}A^*$ (recall that a covariance operator is a bounded, nonnegative, self-adjoint operator), so that we can write $L_0 = R_\xi^{1/2}T(T^*T + \epsilon^2 I)^{-1}$; besides we denote by $\{\tau_k\}_{k=0}^{+\infty}$ the set of the nonzero singular values of T, i.e., $T^*u_k = \tau_k v_k$, $Tv_k = \tau_k u_k$, where $\{u_k\}_{k=0}^{+\infty}$ is an orthonormal basis in $\overline{\text{range}\,(T)}\subset X$ and $\{v_k\}_{k=0}^{+\infty}$ is an orthonormal basis in $\overline{\text{range}\,(T^*)}\subset Y$;[12] then, from Eq. (3.6), by means of the relation $R_\xi^{1/2}u_k = \tau_k A^{-1}v_k$, we get

$$\begin{aligned} \tilde{x} &= \sum_{k=0}^{+\infty} \frac{\tau_k}{\tau_k^2 + \epsilon^2} y_k R_\xi^{1/2} u_k \\ &= \sum_{k=0}^{+\infty} \frac{\tau_k^2}{\tau_k^2 + \epsilon^2} y_k A^{-1} v_k, \end{aligned} \tag{3.8}$$

where $y_k = (y,v_k)_Y$. Now, the random variables $\xi_k = (\xi,R_\xi^{-1/2}u_k)_X$, $\eta_k = (\eta,v_k)_Y$ (k fixed but arbitrary) are Gaussian and: $E\{|\xi_k|^2\} = 1$, $E\{|\eta_k|^2\} = \tau_k^2 + \epsilon^2$, $E\{\xi_k\bar{\eta}_k\} = \tau_k$; therefore their correlation coefficient is

$r_k = \tau_k(\tau_k^2 + \epsilon^2)^{-1}$ and their average mutual information is[18]:

$$J_k = J(\xi_k,\eta_k) = -\tfrac{1}{2}\ln(1 - r_k^2) = \tfrac{1}{2}\ln(1 + \tau_k^2/\epsilon^2). \quad (3.9)$$

From Eq. (3.9) we get $\tau_k^2(\tau_k^2 + \epsilon^2)^{-1} = 1 - \exp(-2J_k)$, so that we can write Eq. (3.8) as follows

$$\tilde{x} = \sum_{k=0}^{+\infty} (1 - e^{-2J_k}) y_k A^{-1} v_k. \quad (3.10)$$

From this expression we see that the best linear estimate $\tilde{x}$ can be obtained, from the formal solution $A^{-1}y = \Sigma_{k=0}^{+\infty} y_k A^{-1} v_k$, by means of a penalty on the coefficients in terms of the amount of information on the random variable ξ_k contained in the random variable η_k.

Let us assume now that the w.r.v. ξ has a finite variance, i.e., $E\{\|\xi\|_X^2\} < +\infty$ (recall that ξ has a zero mean); as is known, this condition is satisfied iff R_ξ is an operator of the trace class.[17] In such a case, we can try to define a "global" mean square error in the estimate of ξ (by means of $\tilde{\xi}_L = L\eta$) as the variance of the w.r.v. $\xi - L\eta$, i.e.,

$$Q(L,\epsilon) = E\{\|\xi - L\eta\|_X^2\} = \mathrm{tr}(P_\xi - R_\xi A^* L^* - LAR_\xi + LR_\eta L^*). \quad (3.11)$$

$Q(L,\epsilon)$ is finite iff the w.r.v. $L\eta$ has also a finite variance. Next we can define a best linear estimate $\tilde{\xi} = L_0\eta$, in the sense of the mean square error (3.11), if there exists a bounded operator L_0 which minimizes $Q(L,\epsilon)$.

We must distinguish two cases. The first is when the noise ζ has a finite variance; then, from Eq. (3.2) it follows that R_η is of the trace class and therefore the w.r.v. $L\eta$ has also a finite variance for any bounded linear operator $L: Y \to X$. In such a case, if there exists a bounded operator L_0 which minimizes $Q(u;L,\epsilon)$ then it also minimizes $Q(L,\epsilon)$ and vice versa. The second case is when ζ (and therefore also η) does not have a finite variance. Now, since the covariance operator of $L\eta$ is $LR_\eta L^*$, it follows that $L\eta$ has a finite variance iff the operator L is of the Schmidt class. Therefore, there exists a best linear estimate in the sense of the mean-square error (3.11) iff the operator $R_{\xi\eta}R_\eta^{-1}$ has a bounded extension of the Schmidt class.

B. Least mean-square errors

We can define the *least mean-square error* in the estimate of $(\xi,u)_X$ as

$$\delta(u,\epsilon) = \inf_L [Q(u;L,\epsilon)]^{1/2}, \quad (3.12)$$

where $Q(u;L,\epsilon)$ is defined by Eq. (3.4). If the best linear estimate exists, then (3.12) gives a measure of the reliability of the estimate; indeed, we have $\delta(u,\epsilon) = [Q(u;L_0,\epsilon)]^{1/2}$, L_0 being given by Eq. (3.5). After some simple calculations we get

$$\delta(u,\epsilon) = ([R_\xi - L_0 R_\eta L_0^*]u,u)_X^{1/2}. \quad (3.13)$$

In the general case (i.e., even if the best linear estimate does not exist) we have the following result,[11]

$$\delta(u,\epsilon) = ([1 - V^*V]R_\xi^{1/2}u, R_\xi^{1/2}u)_X^{1/2}, \quad (3.14)$$

where $V: X \to Y$ is the unique linear continuous operator (with $\|V\| \leq 1$) such that $R_{\eta\xi} = R_\eta^{1/2} V R_\xi^{1/2}$.

Remark 3.4: $\delta(u,\epsilon)$ has a role, in the probabilistic method, similar to that of the best stability estimate (2.3), with $\langle\cdot\rangle = |(\cdot,u)|_X$, in the Tikhonov–Miller method. Indeed, if we put $N = I$ and $R_\xi = E^2(B^*B)^{-1}$ in Eq. (3.5), from Eqs. (3.7) and (3.13) we get

$$\delta(u,\epsilon) = \epsilon\left(\left[A^*A + \left(\frac{\epsilon}{E}\right)^2 B^*B\right]^{-1} u,u\right)_X^{1/2}; \quad (3.15)$$

this expression coincides, except for a factor of $\sqrt{2}$, with Eq. (2.11).

The following result[11] gives conditions which guarantee that $\delta(u,\epsilon)$ tends to zero when $\epsilon \to 0$:

Theorem 3.2: If the operator $R_\zeta = \epsilon^2 N$ has a bounded inverse, then, for any $u \in X$

$$\lim_{\epsilon \to 0} \delta(u,\epsilon) = 0 \quad (3.16)$$

iff the problem

$$x \in \mathrm{range}(R_\xi^{1/2}), \quad Ax = 0 \quad (3.17)$$

has only the trivial solution.

Remark 3.5: In the problems considered in this paper, the inverse of the operator A exists; therefore problem (3.17) has only the trivial solution.

The following result is the analog of Theorem 2.2:

Theorem 3.3: If the operator $R_\zeta = \epsilon^2 N$ has a bounded inverse and if the inverse of the operator R_ξ exists, then

$$\delta(u,\epsilon) = O(\epsilon), \quad \epsilon \to 0 \quad (3.18)$$

iff $u \in \mathrm{range}\,(A^* N^{-1/2})$.

Proof: From the identity (3.7) and Eq. (3.13) we get

$$\delta(u,\epsilon) = \epsilon([A^* N^{-1} A + \epsilon^2 R_\xi^{-1}]^{-1} u,u)_X^{1/2} \quad (3.19)$$

and, if we introduce the operator $S = N^{-1/2} A R_\xi^{1/2}$, we can write

$$\delta(u,\epsilon) = \epsilon([S^*S + \epsilon^2 I]^{-1} R_\xi^{1/2}u, R_\xi^{1/2}u)_X^{1/2}. \quad (3.20)$$

At this point the proof proceeds like the proof of Theorem 2.2.

When the inverse of the operator $R_\zeta = \epsilon^2 N$ is not bounded, it seems that no general result on the behavior of $\delta(u,\epsilon)$, when $\epsilon \to 0$, can be derived; assumptions on the ranges of the operators $(AR_\xi A^*)^{1/2}$ and $R_\xi^{1/2} = \epsilon N^{1/2}$ have to be done. However there exists an important case where assumptions of this type can be proposed in a quite natural way: We mean when both the operators R_ξ, R_ζ are of the trace class. In such a case the covariance operator R_η belongs to the same class and one can introduce the average mutual information $J(\xi,\eta)$ of the w.r.v. ξ,η.[18,19] Our prescription is to require that $J(\xi,\eta)$ is finite [in the problem of Sec. 4, $J(\xi,\eta)$ is the amount of information about the object contained in the optical image and the finite variances of ξ and η are the average energies of the object and of the image, respectively].

According to the results of Baker[19] one can give necessary and sufficient conditions on the covariance operators R_ξ, R_ζ in order that $J(\xi,\eta) < +\infty$; if we take into account these results, then the previous assumptions can be formulated as follows:

(i) $\mathrm{tr}(R_\xi) < +\infty$, $\mathrm{tr}(N) < +\infty$;

(ii) $AR_\xi A^* = N^{1/2} S N^{1/2}$ where $S: Y \to Y$ is a linear operator of the trace class.

The following result holds[11]: *if the assumptions (i), (ii) are satisfied, then Eq.* (3.16) *holds for any* $u \in X$, *iff the problem* (3.17) *has only the trivial solution.*

When the w.r.v. ξ has a finite variance, we can consider the mean-square error (3.11) and define a least mean-square error as follows,

$$\delta(\epsilon) = \inf_L Q(L,\epsilon). \tag{3.21}$$

If the assumptions (i), (ii) are satisfied, then it is possible to prove[11] that $\delta(\epsilon) \to 0$ when $\epsilon \to 0$.

C. Methods of eigenfunction expansions

We consider now the case where the operator $A: X \to Y$ is compact. We assume that the operators A^*A and R_ξ commute; the same assumption is done on the operators AA^* and $R_\zeta = \epsilon^2 N$.

If A and A^* both have an inverse operator (such a condition is satisfied in the problems considered in this paper), then all the singular values α_k of A are nonzero; besides, $\{u_k\}_{k=0}^{+\infty}$ (the set of the eigenvectors of A^*A) and $\{v_k\}_{k=0}^{+\infty}$ (the set of the eigenvectors of AA^*; $v_k = \alpha_k^{-1} A u_k$) are an orthonormal basis in X, Y, respectively.

Thanks to the previous assumption we can write

$$R_\xi x = \sum_{k=0}^{+\infty} \rho_k x_k u_k, \quad R_\zeta y = \epsilon^2 \sum_{k=0}^{+\infty} \nu_k y_k v_k, \tag{3.22}$$

where $x_k = (x,u_k)_X$, $y_k = (y,v_k)_Y$.

Now, the operator L_0, Eq. (3.5), is bounded iff $\sup(\alpha_k \rho_k \nu_k^{-1}) < +\infty$;[11] in such a case Eq. (3.6) takes the form

$$\tilde{x} = \sum_{k=0}^{+\infty} \frac{\alpha_k \rho_k}{\alpha_k^2 \rho_k + \epsilon^2 \nu_k} y_k u_k \tag{3.23}$$

or a form similar to (3.10),

$$\tilde{x} = \sum_{k=0}^{+\infty} (1 - e^{-2J_k}) \frac{y_k}{\alpha_k} u_k, \tag{3.24}$$

where now $J_k = \frac{1}{2}\ln[1 + (\alpha_k^2 \rho_k)/(\epsilon^2 \nu_k)]$ is the average mutual information of the random variables $\xi_k = (\xi,u_k)_X$ and $\eta_k = (\eta,v_k)_Y$.

As regards the least mean-square error (3.12) we have

$$\delta(u,\epsilon) = \epsilon \left(\sum_{k=0}^{+\infty} \frac{\rho_k \nu_k}{\alpha_k^2 \rho_k + \epsilon^2 \nu_k} |(u,u_k)_X|^2 \right)^{1/2} \tag{3.25}$$

and we see that: $\delta(u,\epsilon)$ *tends to zero, when* $\epsilon \to 0$, *for any* $u \in X$, without any further condition on the covariance operators R_ξ, R_ζ. Such a peculiarity is a consequence of the fact that, now, the operators $(AR_\xi A^*)^{1/2}$ and $R_\zeta^{1/2}$ commute.

In the case where the w.r.v. ξ has a finite variance ($\sum_{k=0}^{\infty} \rho_k < +\infty$), we have, for the least mean-square error (3.21), the expression

$$\delta(\epsilon) = \epsilon \left(\sum_{k=0}^{+\infty} \frac{\rho_k \nu_k}{\alpha_k^2 \rho_k + \epsilon^2 \nu_k} \right)^{1/2} \tag{3.26}$$

and we find that $\delta(\epsilon)$ *tends to zero when* $\epsilon \to 0$. If we take $N = I$ (i.e., $\nu_k = 1$) and $R_\xi = E^2 (B^* B)^{-1}$ (i.e., $\rho_k = E^2 \beta_k^{-2}$) we get

$$\delta(\epsilon) = \epsilon \left(\sum_{k=0}^{+\infty} \frac{1}{\alpha_k^2 + (\epsilon/E)^2 \beta_k^2} \right)^{1/2}, \tag{3.27}$$

i.e., we obtain formally Eq. (2.38) except for a factor of $\sqrt{2}$.

An estimation, even more simple that (3.23) and similar to (2.41) can be obtained as follows: We consider the class of the linear bounded operators $\tilde{L}: Y \to X$ defined by $\tilde{L}y = \sum_{k \in I} \alpha_k^{-1} y_k u_k$, where I is an arbitrary finite set and $y_k = (y,v_k)_Y$; then we look for the operator $\tilde{L}_0$ in such a class, which minimizes the mean square error (3.4). Now let us consider the set

$$I_\epsilon = \{k \,|\, \alpha_k^2 \geqslant \epsilon^2 (\nu_k/\rho_k)\}; \tag{3.28}$$

I_ϵ contains only a finite number of points for any $\epsilon > 0$ iff $\alpha_k^2 \rho_k \nu_k^{-1}$ tends to zero when $k \to +\infty$. When this condition is satisfied, it is easy to see that $\tilde{L}_0$ exists and it is the operator corresponding to I_ϵ. Therefore, given the value y of η, we have the following estimate of ξ,

$$\tilde{\tilde{x}} = \sum_{k \in I_\epsilon} \frac{y_k}{\alpha_k} u_k. \tag{3.29}$$

Equation (3.29) can be obtained from Eq. (3.24) replacing the factor $1 - \exp(-2J_k)$ by 1 when $J_k \geqslant \frac{1}{2}\ln 2$ and by 0 when $J_k < \frac{1}{2}\ln 2$.

The minimum of the mean-square error (3.4) over the class of the operators $\tilde{L}$ is

$$\tilde{\delta}^2(u,\epsilon) = \inf_{\tilde{L}} Q(u;\tilde{L},\epsilon)$$

$$= \sum_{k \in I_\epsilon} \frac{\epsilon^2 \nu_k}{\alpha_k^2} |(u,u_k)_X|^2 + \sum_{k \notin I_\epsilon} \rho_k |(u,u_k)_X|^2; \tag{3.30}$$

besides, when ξ has a finite variance, the minimum of the mean square error (3.11) is

$$\tilde{\delta}^2(\epsilon) = \inf_{\tilde{L}} Q(\tilde{L},\epsilon)$$

$$= \sum_{k \in I_\epsilon} \frac{\epsilon^2 \nu_k}{\alpha_k^2} + \sum_{k \notin I_\epsilon} \rho_k. \tag{3.31}$$

Since the condition $\alpha_k^2 \rho_k \nu_k^{-1} \to 0$, $k \to +\infty$ is assumed, the

expressions (3.30) and (3.31) are finite and we have the following results: (a) $\tilde{\delta}(u,\epsilon)\to 0$, when $\epsilon\to 0$, *for any* $u\in X$; (b) *if the w.r.v.* ξ *has a finite variance, then* $\tilde{\delta}(\epsilon)\to 0$, *when* $\epsilon\to 0$.

In order to prove (a), we remark that Eq. (3.30) can be written as follows:

$$\tilde{\delta}^2(u,\epsilon)=\sum_{k=0}^{+\infty}\delta_k(\epsilon)\,|(u,u_k)_X|^2, \tag{3.32}$$

where $\delta_k(\epsilon)=\rho_k$, $k\notin I_\epsilon$ and $\delta_k(\epsilon)=\epsilon^2\nu_K\alpha_k^{-2}$, $k\in I_\epsilon$. Then statement (a) follows easily observing that $\delta_k(\epsilon)\leqslant\rho_k$ for any $\epsilon>0$ and $\delta_k(\epsilon)\to 0$, when $\epsilon\to 0$. Analogously statement (b) can be proved recalling that, by assumption, $\Sigma_{k=0}^{+\infty}\rho_k<+\infty$.

4. OBJECT RESTORATION

In this section we consider the problem of object restoration for the optical system described in the introduction—see Eq. (1.1); we recall that the problem is to estimate the object x, given the image y on the interval $[-1,1]$. If we assume that the object and the image have a finite energy, then both x and y belong to $L^2(-1,1)$. Besides, the integral operator A, defined as follows

$$(Ax)(t)=\int_{-1}^{1}\frac{\sin[c(t-s)]}{\pi(t-s)}x(s)ds,\qquad |t|\leqslant 1 \tag{4.1}$$

is a linear, compact, self-adjoint, nonnegative operator in $L^2(-1,1)$; its inverse A^{-1} exists but is not continuous, i.e., the problem of object restoration is an improperly posed problem. Therefore, we can apply the general methods of Secs. 2 and 3 by setting $X=Y=Z=L^2(-1,1)$.

We give here a list of the main properties of the operator A which are useful for our analysis:

(i) A is an operator of the trace class; indeed it is the first iterate of the finite Fourier transform operator,[1] which is of the Schmidt class.

(ii) The eigenvectors of the operator A are given by

$$u_k(t)=\lambda_k^{-1/2}\psi_k(c,t),\quad k=0,1,2,\ldots,\quad |t|\leqslant 1, \tag{4.2}$$

where $\psi_k(c,t)$ are the linear prolate spheroidal wavefunctions[1] and $\lambda_k=\lambda_k(c)$ are the corresponding eigenvalues; $\{u_k\}_{k=0}^{+\infty}$ is an orthonormal basis in $X=L^2(-1,1)$.

(iii) The eigenvalues $\lambda_k=\lambda_k(c)$ form a decreasing sequence: $\lambda_0>\lambda_1>\lambda_2>\cdots$, bounded away from 1 and approaching 0 as $k\to+\infty$. More precisely we have the following asymptotic behavior

$$\lambda_k=O\left[\frac{1}{k}\left(\frac{ce}{k}\right)^{2k}\right],\quad k\to+\infty, \tag{4.3}$$

which can be derived from the power series expansion of $\psi_k(c,t)$ as a function of c.[20,21] Besides we have

$$\sum_{k=0}^{+\infty}\lambda_k=\mathrm{tr}(A)=2c/\pi. \tag{4.4}$$

We recall also that the eigenvalues λ_k have a step function behavior; i.e., the λ_k remain approximately unity for $k<2c/\pi$ while they fall off to zero very rapidly for $k>2c/\pi$.

(iv) The eigenfunctions u_k, for $k\to+\infty$, tend to the corresponding Legendre polynomials, uniformly with respect to $t\in[-1,1]$; more precisely[21,22]

$$\frac{u_k(t)}{(k+\frac{1}{2})^{1/2}}=P_k(t)+a_{k,1}^0P_{k+2}(t)+a_{k,-1}^0P_{k-2}(t)+\cdots, \tag{4.5}$$

where $a_{k,r}^0=O(k^{-|r|})$, $k\to+\infty$, $2r\geqslant -k$.

(v) The linear prolate spheroidal wavefunctions $\psi_k(c,t)$, $-\infty<t<+\infty$, form an orthonormal basis in the space of the L^2-band-limited functions (i.e., the L^2-functions whose Fourier transform has a support contained in the interval $[-c/2\pi,c/2\pi]$). In particular we have the following expansion,

$$\frac{\sin[c(t-s)]}{\pi(t-s)}=\sum_{k=0}^{+\infty}\psi_k(c,t)\psi_k(c,s) \tag{4.6}$$

and, by means of Parseval equality, we get

$$\sum_{k=0}^{+\infty}|\psi_k(c,t)|^2=c/\pi. \tag{4.7}$$

Then it follows the uniform bound

$$|\psi_k(c,t)|\leqslant\left(\frac{c}{\pi}\right)^{1/2},\quad -\infty<t<+\infty \tag{4.8}$$

(this property will be used in the next section).

We proceed now to discuss the problem of object restoration according to the methods outlined in Secs. 2 and 3.

First we analyse the Tikhonov–Miller method and the case of "weak" stability and of "strong" stability [the cases (i) and (ii) of Sec. 2B, respectively].

In the case of "weak" stability we can reconstruct only some weighted averages of x, i.e., the functionals

$$(x,u)_X=\int_{-1}^{1}x(t)\overline{u(t)}dt,\quad \|u\|_X=1. \tag{4.9}$$

From Theorem 2.2 it follows that we have a very good reconstruction of the functional (4.9) if $u\in\mathrm{range}\,(A^*)$, i.e., if (we recall that now $A=A^*$) there exists a vector $v\in X=L^2(-1,1)$ such that

$$u(t)=\int_{-1}^{1}\frac{\sin[c(t-s)]}{\pi(t-s)}v(s)ds,\quad |t|\leqslant 1. \tag{4.10}$$

A function u as (4.10) is very smooth; indeed it is the restriction to the interval $[-1,1]$ of an entire function whose Fourier transform has a support contained in the interval $[-c/2\pi,c/2\pi]$. Moreover, if the operator B has a compact inverse, then we can say that $M_u(\epsilon,E)$ tends to zero uniformly with respect to u (on the unit sphere) since we have $M_u(\epsilon,E)\leqslant M(\epsilon,E)$, where $M(\epsilon,E)$ is defined by Eq. (2.3) with $\langle\cdot\rangle=\|\cdot\|_X$.

Next we come to the case (ii), i.e., to the "strong" stability.

First let us observe that, for any vector $x \in X$ we can always find a constraint operator B like (2.19), with $\lim_{k\to\infty}\beta_k = +\infty$, and a positive number E such that the vector x belongs to the compact subset of $X: K_E = \{\bar{x}\in X \,|\, \|B\bar{x}\|_X \leqslant E\}$. Besides, if we recall the properties of the linear prolate spheroidal wavefunctions, we can interpret the constraint (2.19) as a condition on the concentration of the finite Fourier transform of x.[23] Then one can easily understand that the goodness of the reconstruction of x is as much satisfactory, as far as the finite Fourier transform of x is concentrated in the interval $[-c/2\pi, c/2\pi]$, or equivalently as rapidly as the β_k tend to infinity.

Now we take $B = A^{-1}$, i.e, we assume that the object x belongs to range(A)—see Eq. (4.10); therefore, condition (2.19) holds with $\beta_k = \lambda_k^{-1}$ and Theorem 2.4 can be applied. We see that in this case we have very satisfactory behavior of the stability estimate when $\epsilon \to 0$ (Hölder type stability). We can understand this result if we note that we are considering a class of objects whose finite Fourier transform is rather strongly concentrated in the interval $[-c/2\pi, c/2\pi]$. However, usually, it is too restrictive to assume that the object belongs to this class. It is more interesting to suppose that the eigenvalues β_k grow like a power of k for $k \to +\infty$. Let us spend a few words in order to justify this assertion.

From property (iv)—Eq. (4.5)—of the linear prolate spheroidal wavefunctions it follows that

$$x_k = (x, u_k)_X = O(x_k^0), \quad k \to +\infty, \tag{4.11}$$

where

$$x_k^0 = (k + \tfrac{1}{2})^{1/2} \int_{-1}^{1} x(t) P_k(t)\, dt. \tag{4.12}$$

In other words the asymptotic behavior of the coefficients of the expansion of x as a series of linear prolate spheroidal wavefunctions is given by the asymptotic behavior of the coefficients of the expansion of x as a series of Legendre polynomials.

Now suppose that x has continuous derivatives up to the order m and vanishes outside the interval $[-1+\delta, 1-\delta]$ ($\delta > 0$, fixed but arbitrary), then *the coefficients x_k decrease, for $k \to +\infty$, at least as k^{-m}*. Indeed, by means of a well-known asymptotic formula of Legendre polynomials[24] (valid in the interval $[-1+\delta, 1-\delta]$), integrating m times by parts, it is easy to show that

$$\int_{-1+\delta}^{1-\delta} x(t) P_k(t)\, dt = O(K^{-(m+1/2)}), \tag{4.13}$$

so that, thanks to Eqs. (4.11) and (4.12), $x_k = O(k^{-m})$. As a consequence of this remark, we can conclude that, if we take $\beta_k = k^\mu$ in Eq. (2.19), then the set $K_E = \{\bar{x}\in X \,|\, \|B\bar{x}\|_X \leqslant E\}$ contains functions which vanish outside an interval $[-1+\delta, 1-\delta]$ and whose derivatives are continuous up to the order $m > \mu + \frac{1}{2}$.

We have now the following result:

Theorem 4.1: If, for $k > 2c/\pi$, $\beta_k = \gamma k^\mu$ ($\gamma, \mu > 0$), then there exists functions F_1, F_2 such that

$$EF_1\left(\frac{\epsilon}{E}\right) \leqslant M(\epsilon, E) \leqslant EF_2\left(\frac{\epsilon}{E}\right) \tag{4.14}$$

[where $M(\epsilon, E)$ is defined by Eq. (2.3) with $\langle\cdot\rangle = \|\cdot\|_X$] and, for $\epsilon \to 0$

$$F_1(\epsilon) = O(|\ln\epsilon|^{-\mu}), \quad F_2(\epsilon) = O(|\ln\epsilon|^{-\mu(1-\eta)}), \tag{4.15}$$

where η may be taken as small as desired.

Proof: If we recall the step function behavior of the eigenvalues λ_k and the increasing behavior of the constraints β_k for $k > 2c/\pi$, we can assert that the set $I_{\epsilon/E}$, Eq. (2.20), contains all the values of k until a certain $k_0 > 2c/\pi$, provided that the ratio ϵ/E is sufficiently small. Therefore, in Eq. (2.22), Theorem 2.3, we have $\alpha(\epsilon/E) = \lambda_{k_0}$, $\beta(\epsilon/E) = \gamma(k_0+1)^\mu$ and, from the inequalities

$$\lambda_{k_0} \geqslant \frac{\epsilon}{E}\gamma k_0^\mu, \quad (k_0+1)^\mu > k_0^\mu,$$

we get

$$\frac{E}{2\gamma}(k_0+1)^{-\mu} \leqslant M(\epsilon, E) \leqslant \frac{2E}{\gamma} k_0^{-\mu}. \tag{4.16}$$

k_0 is a function of the ratio ϵ/E. In order to prove the theorem we have to show that there exists an upper bound for k_0 which is $O(|\ln(\epsilon/E)|)$, and a lower bound which is $O(|\ln(\epsilon/E)|^{1-\eta})$, where η may be made as small as desired.

First we find an upper bound for k_0. Property (iii), and more precisely, Eq. (4.3), implies that there exists, for fixed c, a constant γ_1 such that $\lambda_k \leqslant \gamma_1 \exp(-k)$, $k > 2c/\pi$; then, if we set $k_1 = \sup\{k \,|\, \gamma_1\exp(-k) \geqslant (\gamma\epsilon/E)k^\mu\}$, we have $k_0 \leqslant k_1$. On the other hand, $k_1 = [\chi_1]$; here $[\chi_1]$ denotes the greatest integer such that $[\chi_1] \leqslant \chi_1$ and χ_1 is the solution of the equation: $\chi_1^{-\mu}\exp(-\chi_1) = (\gamma_1\epsilon)/(\gamma E)$. It is not difficult to see that, when $\epsilon \to 0$, $\chi_1 = O(|\ln(\epsilon/E)|)$ so that, from the first half of inequality (4.16) we get the first half of inequality (4.14).

Then we find a lower bound for k_0. Equation (4.3) implies also that there exists (for fixed c) constants γ_2, δ such that $\lambda_k \geqslant \gamma_2 \exp[-(k)^{1+\delta}]$, $k > 2c/\pi$, where δ may be made as small as desired; then, if we set $k_2 = \sup\{k \,|\, \gamma_2 \exp[-(k)^{1+\delta}] \geqslant (\gamma\epsilon/E)k^\mu\}$ we have $k_0 \geqslant k_2$. On the other hand, $k_2 = [\chi_2]$, where χ_2 is the solution of the equation $\chi_2^{-\mu}\exp(-\chi_2^{1+\delta}) = (\gamma\epsilon)/(\gamma_2 E)$. It is not difficult to see that, when $\epsilon \to 0$, $\chi_2 = O(|\ln(\epsilon/E)|^{1-\eta}$, where $\eta = \delta(1+\delta)^{-1}$ and, from the second half of inequality (4.16) we get the second half of inequality (4.14).

Theorem 4.1 can be interpreted by saying that, when we do not make too restrictive assumptions on the unknown vector x, then *the restored continuity in the problem of object restoration is, at most, logarithmic.*

Now we come to the probabilistic method. As regards the least mean-square error $\delta(u,\epsilon)$, it is enough to observe that it has properties very similar to the best stability estimate $M_u(\epsilon, E)$ (see, for instance Theorem 2.2 and Theorem 3.3).

It is more interesting to analyze the least mean-square error $\delta(\epsilon)$. If we consider the case where all the operators A,

R_ξ, R_ζ are diagonal, then $\delta(\epsilon)$ is given by Eq. (3.26) with $\alpha_k = \lambda_k$. now $\delta(\epsilon)$ is finite if R_ξ is an operator of the trace class, i.e., if ξ has a finite variance. This assumption means that we are considering a class of optical objects whose average energy is finite. The same assumption is reasonable for ζ; if we normalize the operator N in such a way that $\mathrm{tr}(N) = 1$, then ϵ^2 is just the average energy of the noise. We can easily find upper and lower bounds for $\delta(\epsilon)$ if we further assume that $\lambda_k^2 \rho_k \nu_k^{-1} \to 0$, $k \to +\infty$ [i.e., we suppose that the amount of information on the component $\xi_k = (\xi, u_k)_X$ of the object, contained in the corresponding component $\eta_k = (\eta, u_k)_X$ of the image, tends to zero for $k \to +\infty$—see Eq. (3.24)]. Now we can say that the set I_ϵ, Eq. (3.28), contains only a finite number of points and we have the following result,

$$\tfrac{1}{2}\delta_0(\epsilon) \leqslant \delta(\epsilon) \leqslant \delta_0(\epsilon), \tag{4.17}$$

where

$$\delta_0(\epsilon) = \epsilon \left(\sum_{k \in I_\epsilon} \frac{\nu_k}{\lambda_k^2} \right)^{1/2} + \left(\sum_{k \notin I_\epsilon} \rho_k \right)^{1/2}. \tag{4.18}$$

The inequalities (4.17) follow from Eq. (3.26) when we observe that:

$$k \in I_\epsilon, \quad \frac{1}{2} \frac{\nu_k}{\lambda_k^2} \leqslant \frac{\rho_k \nu_k}{\lambda_k^2 \rho_k + \epsilon^2 \nu_k} \leqslant \frac{\nu_k}{\lambda_k^2}, \tag{4.19}$$

$$k \notin I_\epsilon, \quad \frac{1}{2} \frac{\rho_k}{\epsilon^2} \leqslant \frac{\rho_k \nu_k}{\lambda_k^2 \rho_k + \epsilon^2 \nu_k} \leqslant \frac{\rho_k}{\epsilon^2}. \tag{4.20}$$

Now, the eigenvalues λ_k decrease for increasing k; moreover, if the signal to noise ratios ρ_k/ν_k also decrease (at least for $k > S$), then, for ϵ sufficiently small, we have

$$\delta_0(\epsilon) = \epsilon \left(\sum_{k=0}^{k_0} \frac{\nu_k}{\lambda_k^2} \right)^{1/2} + \left(\sum_{k=k_0+1}^{+\infty} \frac{\nu_k}{\beta_k^2} \right)^{1/2}$$

$$\leqslant \frac{\epsilon}{\lambda_{k_0}} + \frac{1}{\beta_{k_0+1}}, \tag{4.21}$$

where $\beta_k^2 = \nu_k \rho_k^{-1}$. As a consequence we see that, if the signal to noise ratios decrease like an inverse power of k, $\rho_k \nu_k^{-1} = (\text{constant}) \cdot k^{-2\mu}$, then $\delta_0(\epsilon)$ is bounded by a function which is $O(|\ln\epsilon|^{-\mu(1-\eta)})$, for $\epsilon \to 0$ (η may be taken as small as desired; the proof is just the same as in Theorem 4.1).

5. IMAGE EXTRAPOLATION

The problem of extrapolation of optical image data has been already considered;[6] in that paper a stabilization condition was imposed by requiring the boundedness of the total energy of the image. Here we prefer to reinforce the stabilization constraint, requiring explicity that the functions, which we want to extrapolate, are optical images corresponding to objects whose energy is finite and bounded by E. Moreover, observing that the magnification of the optical system described by Eq. (1.1) is one, we assume that these optical images are measured in the interval $[-1, +1]$. In this case we can show that the restored continuity, in the problem of extrapolation of an image piece beyond its borders, is of the Hölder type. Let us remark that this result holds true only if the images, which we want to extrapolate, are approximately known in an interval which is equal to the support of the object. Otherwise, if the images are measured in a smaller interval, it is reasonable to conjecture that the restored continuity is not of the Hölder type, but weaker.

The problem of image extrapolation can be formulated as follows: Let $x = x(t)$, $-\infty < t < +\infty$, be a squareintegrable band-limited function, i.e.,

$$x \in L^2(-\infty, +\infty), \quad x(t) = \int_{-c/2\pi}^{c/2\pi} e^{2\pi i t\omega} \hat{x}(\omega) d\omega. \tag{5.1}$$

Let Ax be the restriction of x to the interval $[-1,1]$, and $y = Ax + z$ the image measured in the interval$[-1,1]$; z describes the experimental errors. Then the problem of image extrapolation is to estimate x given y.

Now, in order to apply the methods of Secs. 2 and 3, we must specify the spaces X, Y, and the main properties of the operator A.

In $L^2(-\infty, +\infty)$ let X be the subspace of the band-limited functions (5.1); since X is closed, then it is a separable Hilbert space with respect to the inner product of $L^2(-\infty, +\infty)$. Besides, let $Y \subset L^2(-\infty, +\infty)$ be the (closed) subspace of those functions whose support is contained in the interval $[-1,1]$; Y is also a separable Hilbert space and is isomorphic to $L^2(-1,1)$. Then $A: X \to Y$ is the operator defined by

$$(Ax)(t) = x(t), \quad |t| \leqslant 1; \quad (Ax)(t) = 0, \quad |t| > 1. \tag{5.2}$$

A is a compact operator of the Schmidt class; in other words the operator AA^* is of the trace class. At this purpose, let us remark that the adjoint $A^*: Y \to X$ is given by

$$(A^* y)(t) = \int_{-1}^{1} \frac{\sin[c(t-s)]}{\pi(t-s)} y(s) ds, \quad y \in Y, \tag{5.3}$$

as easily follows from the identity for band-limited functions

$$x(t) = \int_{-\infty}^{+\infty} \frac{\sin[c(t-s)]}{\pi(t-s)} x(s) ds, \quad x \in X. \tag{5.4}$$

Indeed, from Eqs. (5.3) and (5.4), we get, for any $x \in X$, $y \in Y$, $(Ax, y)_Y = (x, A^* y)_X$. Then, from Eqs. (5.3) and (5.2) we obtain that the operator AA^* coincides with the operator (4.1) which is of the trace class, as it was shown in Sec. 4.

Next we want to show that all the singular values of both the operators A and A^* are nonzero. In other words we want to prove that: *The equations $Ax = 0$ and $A^* y = 0$ have only the trivial solutions $x = 0$ and $y = 0$, respectively.* Indeed, $Ax = 0$ means that x is a.e. zero in the interval $[-1,1]$; since x is an entire function, we conclude that it is zero on the whole $(-\infty, +\infty)$. Analogously $A^* y = 0$ means that the Fourier transform of y, i.e., $\hat{y}$, is a.e. zero in the interval $[-c/2\pi, c/2\pi]$; since $\hat{y}$ is an entire function, they y is zero on the whole $(-\infty, +\infty)$, i.e., $y = 0$.

Now, let us denote by $\{u_k\}_{k=0}^{+\infty}$ the set of the eigenvectors of the operator A^*A and by $\{v_k\}_{k=0}^{+\infty}$ the set of the eigenvectors of the operator AA^*; $\{u_k\}_{k=0}^{+\infty}$ is an orthonormal basis in X and $\{v_k\}_{k=0}^{+\infty}$ is an orthonormal basis in Y. If we recall that AA^* coincides with the operator (4.1) we have

$$u_k(t) = \psi_k(c,t), \quad |t| < +\infty;$$

$$v_k(t) = \lambda_k^{-1/2}\psi_k(c,t), \quad |t| \leqslant 1; \qquad (5.5)$$
$$k = 0,1,2,\cdots,$$

$\psi_k(c,t)$ being normalized to one with respect to the norm of $L^2(-\infty,+\infty)$—see also property (v) of Sec. 4. The vectors u_k, v_k satisfy the relations

$$Au_k = \sqrt{\lambda_k}\, v_k, \quad A^*v_k = \sqrt{\lambda_k}\, u_k, \quad k = 0,1,2,\cdots. \qquad (5.6)$$

At this point we can specify the stabilization constraint. We assume that the unknown vector x belongs to a class of optical images corresponding to optical objects whose energy is finite and bounded by E,

$$x(t) = \int_{-1}^{1} \frac{\sin[c(t-s)]}{\pi(t-s)} v(s)ds, \quad |t| < +\infty,$$
$$\int_{-1}^{1} |v(s)|^2 ds \leqslant E^2, \qquad (5.7)$$

i.e., $x = A^*v$, $\|v\|_Y \leqslant E$. This condition can be written in the form (2.2) if we take $B = (A^*)^{-1}$. Since the operator $(B^*B)^{-1} = A^*A$ is of the trace class, the problem is stable with respect to the norms or seminorms considered in Sec. 2. Besides, from Theorem 2.4, it follows that

$$M(\epsilon,E) \leqslant \sqrt{\epsilon E}\,, \qquad (5.8)$$

where $M(\epsilon,E)$ is given by Eq. (2.3) with

$$\langle x \rangle = \|x\|_X = \left(\int_{-\infty}^{+\infty} |x(t)|^2 dt \right)^{1/2}. \qquad (5.9)$$

The approximation for the unknown vector x is then given by

$$\tilde{x}(t) = \sum_{k=0}^{+\infty} \frac{\lambda_k^{3/2}}{\lambda_k^2 + (\epsilon/E)^2} y_k \psi_k(c,t), \qquad (5.10)$$

where

$$y_k = (y,v_k)_Y = \frac{1}{\sqrt{\lambda_k}} \int_{-1}^{1} y(t)\psi_k(c,t)dt. \qquad (5.11)$$

Thanks to Eq. (2.7), $\tilde{x}$ gives an approximation to the unknown image x [in the sense of the norm of $L^2(-\infty,+\infty)$] with an error which is of the order of $\sqrt{\epsilon}$ (Hölder continuity).

Now, if we consider the best stability estimate $M_u(\epsilon,E)$, given by Eq. (2.3) with

$$\langle x \rangle = |(x,u)_X| = \left| \int_{-\infty}^{+\infty} x(t)\overline{u(t)}dt \right|,$$
$$\int_{-\infty}^{+\infty} |u(t)|^2 dt = 1, \qquad (5.12)$$

then, from inequality $|(\cdot,u)_X| \leqslant \|\cdot\|_X$ and Eq. (5.8) we obtain

$$M_u(\epsilon,E) \leqslant M(\epsilon,E) \leqslant \sqrt{\epsilon E}\,. \qquad (5.13)$$

Inequality (5.13) means that, if one pretends to reconstruct only some weighted averages of x, then the error is, at most, of the order of $\sqrt{\epsilon}$. The situation can be more favorable if one takes $u \in \text{range}(A^*)$—see Theorem 2.2.

Finally we come to the norm $|||\cdot|||$ defined in Eq. (2.30). Such a norm is interesting in the problem of image extrapolation, since it is stronger that the uniform norm. Indeed, from inequality (4.8) we have

$$\|x\|_\infty = \sup_t |x(t)| \leqslant \sum_{k=0}^{+\infty} |x_k| \sup_t |\psi_k(c,t)|$$
$$\leqslant \sqrt{c/\pi}\, |||x|||. \qquad (5.14)$$

As a consequence, any stability estimate for $M(\epsilon,E)$ with $\langle\cdot\rangle = |||\cdot|||$ is a stability estimate for the uniform convergence. Now, from Theorem 2.5, we have

$$\tilde{M}(\epsilon,E) = \sqrt{2\epsilon}\left(\sum_{k=0}^{+\infty} \frac{\lambda_k}{\lambda_k^2 + (\epsilon/E)^2} \right)^{1/2}. \qquad (5.15)$$

Then, from the inequality

Then, from the inequality $t^{2\alpha} \leqslant C_\alpha(t^2+1)$ $[t > 0, 0 \leqslant \alpha \leqslant 1, C_\alpha = \alpha^\alpha(1-\alpha)^{1-\alpha}]$, with $0 < \alpha < \frac{1}{2}$, we have

$$\tilde{M}(\epsilon,E) \leqslant \sqrt{2}\epsilon C_\alpha^{1/2}\left(\sum_{k=0}^{+\infty} \lambda_k^{1-2\alpha} \right)^{1/2}\left(\frac{\epsilon}{E}\right)^{\alpha-1}$$
$$= \sqrt{2}\Gamma_\alpha E\left(\frac{\epsilon}{E}\right)^\alpha, \qquad (5.16)$$

where $\Gamma_\alpha < +\infty$ is $\alpha < \frac{1}{2}$—see Eq. (4.3). If we denote by $M_\infty(\epsilon,E)$ the best stability estimate (2.3) with $\langle\cdot\rangle = \|\cdot\|_\infty$, then we have the result

$$M_\infty(\epsilon,E) \leqslant (\text{constant})E\left(\frac{\epsilon}{E}\right)^\alpha, \qquad (5.17)$$

where α is any number between 0 and $\frac{1}{2}$.

Now we come to the probabilistic method. The stabilization constraint (5.7) can be translated in a probabilistic language by saying that there exists a Y-valued w.r.v. w such that $\xi = A^*w$; therefore, one has: $R_\xi = A^*R_wA$ and the w.r.v. ξ has a finite variance.

We have the simplest case assuming for R_ξ and R_ζ the following expressions,

$$R_w = E^2 I, \quad R_\zeta = \epsilon^2 I, \qquad (5.18)$$

where I is the identity operator in Y. Then it is easy to verify that, in this case, Eq. (3.6) gives Eq. (5.10). Moreover, since $\text{tr}(R_\xi) = E^2 \cdot \text{tr}(A^*A) = E^2 \sum_{k=0}^{+\infty} \lambda_k = (2c/\pi)E^2 < \infty$—see Eq. (4.4)—both the least mean-square errors (3.25) and (3.26) can be analyzed; more precisely we obtain

$$\delta(u,\epsilon)=\epsilon\left(\sum_{k=0}^{+\infty}\frac{\lambda_k}{\lambda_k^2+(\epsilon/E)^2}\,|(u,u_k)_X|^2\right)^{1/2} \tag{5.19}$$

and

$$\delta(\epsilon)=\epsilon\left(\sum_{k=0}^{+\infty}\frac{\lambda_k}{\lambda_k^2+(\epsilon/E)^2}\right)^{1/2}. \tag{5.20}$$

Then we have $\delta(u,\epsilon)=2^{-1/2}\tilde{M}_u(\epsilon,E)$ and $\delta(\epsilon)=2^{-1/2}\tilde{M}(\epsilon,E)$—see Eq. (5.15)—and therefore the results previously obtained for the best stability estimates hold true also for the least mean-square errors.

Next we assume in Eq. (3.26), $\rho_k=E^2\lambda_k w_k$, with the condition $\Sigma_{k=0}^{+\infty}w_k=1$; furthermore we assume, for the noise, the condition $\Sigma_{k=0}^{+\infty}\nu_k=1$. Then $\delta(\epsilon)$ reads as follows,

$$\delta(\epsilon)=\epsilon\left(\sum_{k=0}^{+\infty}\frac{w_k\nu_k\lambda_k}{w_k\lambda_k^2+(\epsilon/E)^2\nu_k}\right)^{1/2}. \tag{5.21}$$

From the inequality used in Eq. (5.16) we get

$$\delta(\epsilon)\leqslant C_\alpha^{1/2}\epsilon^\alpha E^{1-\alpha}\left(\sum_{k=0}^{+\infty}\frac{w_k^{1-\alpha}\nu_k^\alpha}{\lambda_k^{2\alpha-1}}\right)^{1/2} \tag{5.22}$$

with $0\leqslant\alpha\leqslant 1$. The series at the rhs of Eq. (5.22) is certainly convergent for $0<\alpha\leqslant\frac{1}{2}$ (in the specific case $\alpha=\frac{1}{2}$ it becomes $\Sigma_{k=0}^{+\infty}\sqrt{w_k\nu_k}\leqslant 1$). The series can also converge for $\alpha>\frac{1}{2}$ if the sequence $\{(\nu_k/w_k)^\alpha\lambda_k^{1-2\alpha}\}_{k=0}^{+\infty}$ is bounded; in the latter case we have a stronger Hölder continuity.

6. CONCLUSIONS

The main result of our paper is the proof that in the object restoration problem, the continuity, in most realistic cases, is at best logarithmic. This result derives essentially from the rapid exponential fall of the eigenvalues λ_k of the operator A. Now we can try to extend this result. In fact an asymptotic behavior of the same kind [more precisely $\lambda_k=O(\exp(-Dk\ln k))$ where D is a constant] holds true for every integral operator whose kernel is an entire analytic function of finite order.[25] So we may argue that we get at best logarithmic continuity whenever, for inverting such an integral operator, we impose *a priori* bounds on a finite number of derivatives of the solutions. Analytic kernels are involved in some inverse problems such as the near-field reconstruction,[26] or the Bojarski–Lewis inverse scattering method when one has information only over a finite frequency band (bandpass kernels).[5]

ACKNOWLEDGMENTS

The authors are indebted to Professor K. Miller and Professor G. Talenti for many useful discussions on the mathematical aspects of the problem, and to Professor A. Consortini, Professor L. Ronchi, and Professor G. Toraldo di Francia for discussing the questions of optics of this paper. They want also to thank Prof. W.M. Boerner for many suggestions and for a list of references.

[1] B.R. Frieden, Prog. Opt. **9**, 311 (1971).
[2] J.W. Goodman, Prog. Opt. **8**, 1 (1970).
[3] G. Toraldo di Francia, J. Opt. Soc. Am. **59**, 799 (1969).
[4] G.A. Deschamps and H.S. Cabayan, IEEE Trans. Antennas Propag. **AP-20**, 268 (1972).
[5] W.L. Perry, IEEE Trans. Antennas Propag. **AP-22**, 826 (1974).
[6] G.A. Viano, J. Math. Phys. **17**, 1160 (1976).
[7] A. Tikhonov, Sov. Math. Dokl. **4**, 1624 (1963).
[8] K. Miller, SIAM J. Math. Anal. **1**, 52 (1970).
[9] F. John, Commun. Pure Applied Math. **13**, 551 (1956).
[10] P.B. Fellgett and E.H. Linfoot, Philos. Trans. R. Soc. A **247**, 369 (1955).
[11] M. Bertero and G.A. Viano, to appear in Bollettino U.M.I.
[12] T. Kato, *Perturbation Theory for Linear Operators* (Springer-Verlag, Berlin, 1966), p. 261.
[13] A. Tikhonov and V. Arsenine, *Méthodes de Résolution de Problémes mal posés* (Moscou, 1976), p. 29.
[14] N. Dunford and J.L. Schwartz, *Linear Operators* (Interscience, New York, 1958), Vol I, p. 338.
[15] K. Miller, Arch. Ratl. Mech. Anal. **16**, 126 (1964).
[16] J.N. Franklin, J. Math. Anal. Appl. **31**, 682 (1970).
[17] A.V. Balakrishnan, *Applied Functional Analysis* (Springer, New York, 1976), Chapter 6.
[18] J.M. Gel'fand and A.M. Yaglom, Am. Math. Soc. Trans., Ser. 2 **12**, 199 (1959).
[19] C.R. Baker, in *Lecture Notes in Mathematics*, Vol. 526 (Springer-Verlag, Berlin, 1976), p. 1.
[20] D. Slepian, J. Math. Phys. (Cambridge, Mass.) **44**, 99 (1965).
[21] J. Meixner and F.W. Schäfke, *Mathieusche Funktionen und Sphäroidfunktionen* (Berlin, 1954).
[22] A. Erde'lyi *et al.*, *Higher Transcendental Functions* (McGraw-Hill, New York, 1955), Vol. 3.
[23] H.J. Landau and H.O. Pollak, Bell System Tech. **40**, 65 (1961).
[24] E.T. Copson, *Theory of Functions of a Complex Variable*, (Oxford U.P., London, 1962), p. 285.
[25] E. Hille and J.D. Tamarkin, Acta Math. **57**, 1 (1931).
[26] H.S. Cabayan, R.C. Murphy, and T.J.F. Pavlásek, IEEE Trans. Antennas Propag. **AP-21**, 346 (1973).

Higher Euler operators and some of their applications

S. J. Aldersley

Department of Applied Mathematics, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1
(Received 7 December 1977)

A set of operators which are associated with the Euler–Lagrange operator is introduced. An analysis of the commutation properties of these new operators, which will be referred to as the higher Euler operators, leads to a generalization of the necessary conditions for an expression to be an Euler–Lagrange expression. A product rule is derived for the higher Euler operators. In the special case of the Euler–Lagrange operator this product rule is basic to simple proofs of sufficiency theorems for the existence of a Lagrangian given the potential Euler–Lagrange expressions. By considering a certain homogeneity property, a characterization of Lagrangians in terms of their Euler–Lagrange expressions is established. Examples of applications of this characterization are given. A general procedure is given for constructing equivalent (not necessarily scalar density) Lagrangians when the field functions are tensorial and the Euler–Lagrange expressions are tensor densities. These results give particular significance to one of the higher Euler operators.

INTRODUCTION

The purpose of this paper is to introduce and examine a collection of operators which arose in the consideration of necessary and sufficient conditions for the existence of a Lagrangian in field theoretic variational problems. These operators can be viewed as generalizations of the Euler–Lagrange operator and will be called higher Euler operators. In addition, a constructive approach to the inverse problem in the calculus of variations is developed and applied to particular examples. We shall concentrate on the local form of the various operators and functions involved, obtaining results within the domain of an arbitrary chart of some manifold.

In Sec. 1, the notation used throughout the paper is presented and the higher Euler operators are defined and briefly discussed. The following section is a detailed analysis of the commutation properties of these higher Euler operators. It leads to identities for Euler–Lagrange expressions and an interesting relation between the first Euler operator and the Euler–Lagrange operator. An identity, based upon the Euler–Lagrange expression of a product, is obtained in the third section. It gives rise to very simple proofs of certain sufficiency conditions for the existence of a Lagrangian when given expressions which are possibly Euler–Lagrange expressions. The analysis of the inverse problem in the calculus of variations is continued in Sec. 4 where a constructive procedure for obtaining its solution is formulated in special cases. In the final section, methods of obtaining Lagrangians equivalent to a given Lagrangian are presented, giving added significance to the first Euler operator.

1. Definitions and Notation

Our considerations in this paper will be based upon an n-dimensional C^∞ manifold M. If (U,x) is a chart of M, then x^i, $i = 1,...,n$ will denote the coordinates and the notation $,i$ will be used to denote partial differentiation with respect to x^i. The major concern of this paper is real valued functions of M which on the domain U of a chart of M take the form

$$F = F(\rho_A;\rho_{A,i};\cdots;\rho_{A,i_1\cdots i_\alpha};\lambda_\Omega;\lambda_{\Omega,i};\cdots;\lambda_{\Omega,i_1\cdots i_\alpha}) \tag{1.1}$$

In (1.1), the ρ_A's and the λ_Ω's are field functions on M ($\rho_A = \rho_A(x^i), A = 1,...,N$ and $\lambda_\Omega = \lambda_\Omega(x^i), \Omega = 1,...,N'$) and the function F is assumed to be a differentiable concomitant of the field variables (e.g., for all $\tau = 0,1,2,...$, $\partial F/\partial\rho_{A,i_1\cdots i_\tau}$ exists and is continuous). It will *not* be assumed that the field functions are tensorial or that any relation exists between field functions defined on two charts which intersect (unless specified otherwise). We demand that given any particular chart (U,x), there corresponds a set $\mathscr{P}(U,x)$ consisting of all field functions of the form $\{\rho_A(x^i);\lambda_\Omega(x^i)\}$. The set of all concomitant functions F of the form (1.1) (for all values of α and α'), will be denoted by $\mathscr{F}(\mathscr{P}(U,x))$. Note that any element of $\mathscr{F}(\mathscr{P}(U,x))$ is of *finite order* in the derivatives of the field functions.

To illustrate the above notation, we consider the following example. On a chart (U,x), take $\rho_A(x^a) = (g_{ij}(x^a),\Gamma^i_{jk}(x^a))$ and $\lambda_\Omega(x^a) = (\phi_i(x^a))$, where $g_{ij}(x^a)$, $\Gamma^i_{jk}(x^a)$, and $\phi_i(x^a)$ take on x-component values of metric tensor fields, linear connections, and vector fields, respectively, on M. Then the x-components T^{ij} of a tensorial concomitant of the form

$$T^{ij} = T^{ij}(g_{rs};\Gamma^r_{st};\Gamma^r_{st,u};\phi_r),$$

constitute n^2 elements of $\mathscr{F}(\mathscr{P}(U,x))$.

If $F\in\mathscr{F}(\mathscr{P}(U,x))$ [some (U,x)] then the notations $F^{;A,i_1\cdots i_\tau}: = \partial F/\partial\rho_{A,i_1\cdots i_\tau}$ for $\tau = 1,2,\cdots$ and $F^{;A,i_1\cdots i_0}: = F^{;A}: = \partial F/\partial\rho_A$ will be used. The summation convention will be invoked for repeated capital and small Latin indices and repeated capital Greek indices in any term of an expression (e.g., $T^{A,i_1\cdots i_\alpha}\rho_{A,i_1\cdots i_\alpha}: = \Sigma^N_{A=1}\Sigma^n_{i_1,\ldots,i_\alpha=1}T^{A,i_1\cdots i_\alpha}\rho_{A,i_1\cdots i_\alpha}$.) Defining the differential operator D_i on $\mathscr{F}(\mathscr{P}(U,x))$ by

$$D_iF: = \sum_{\alpha=0}^{\infty}(F^{;A,i_1\cdots i_\alpha}\rho_{A,i_1\cdots i_\alpha i} + F^{;\Omega,i_1\cdots i_\alpha}\lambda_{\Omega,i_1\cdots i_\alpha i}), \tag{1.2}$$

for $F\in\mathscr{F}(\mathscr{P}(U,x))$, we see that in (U,x), $F(\rho_A;\lambda_\Omega)_{,i} = (D_iF)(\rho_A;\lambda_\Omega)$. The sum to infinity in (1.2) must stop at

 0022-2488/79/030522-10$01.00

some finite α since by definition F is of finite order in the derivatives of ρ_A and λ_Ω. This technical convenience will be used repeatedly in the paper. Note that for each i, D_i maps $\mathscr{F}(\mathscr{P}(U,x))$ into itself.

The primary concern of this paper is the Euler–Lagrange operator which takes the form

$$E^A(F):=\sum_{\alpha=0}^{\infty}(-1)^\alpha D_{i_1\cdots i_\alpha}F^{;A,i_1\cdots i_\alpha}, \tag{1.3}$$

for $F\in\mathscr{F}(\mathscr{P}(U,x))$, where $D_{i_1\cdots i_0}F:=F$ and $D_{i_1i_2\cdots i_\alpha}:=D_{i_1}D_{i_2}\cdots D_{i_\alpha}$ $(\alpha=1,2,\cdots)$. Clearly if $F\in\mathscr{F}(\mathscr{P}(U,x))$, then for each A, $E^A(F)$ is also in $\mathscr{F}(\mathscr{P}(U,x))$. One obtains useful generalizations of the Euler–Lagrange operator by defining the αth Euler operator $(\alpha=1,2,\cdots)$ by

$$E^{A,i_1\cdots i_\alpha}(F):=\sum_{\beta=\alpha}^{\infty}(-1)^\beta\binom{\beta}{\alpha}D_{i_{\alpha+1}\cdots i_\beta}F^{;A,i_1\cdots i_\beta}, \tag{1.4}$$

where $F\in\mathscr{F}(\mathscr{P}(U,x))$, $D_{i_{\alpha+1}\cdots i_\alpha}F:=F$, and $\binom{\beta}{\alpha}:=\beta!/(\alpha!(\beta-\alpha)!)$. We extend (1.4) to $\alpha=0$ by setting[1] $E^{A,i_1\cdots i_0}(F):=E^A(F)$.

Clearly, knowledge of the partial derivatives $F^{;A,i_1\cdots i_\alpha}$ for $\alpha=0,1,2,\cdots$ serves to determine $E^{A,i_1\cdots i_\beta}(F)$ for any $\beta=0,1,2,\cdots$ via (1.4). An interesting property of the set of Euler operators is that the converse also holds. In the following proposition and the remainder of the paper the convention $\binom{\beta}{\alpha}:=0$ for $\alpha>\beta$ will be employed.

Proposition 1.1: For any $F\in\mathscr{F}(\mathscr{P}(U,x))$ and all $\alpha=0,1,2,\cdots$

$$F^{;A,i_1\cdots i_\alpha}=\sum_{\beta=\alpha}^{\infty}(-1)^\beta\binom{\beta}{\alpha}D_{i_{\alpha+1}\cdots i_\beta}E^{A,i_1\cdots i_\beta}(F). \tag{1.5}$$

Proof: Substituting from (1.4) for the βth Euler operator in the right-hand side of (1.5) we obtain the expression (let F be of order μ in the derivatives of ρ_A)

$$\sum_{\beta=\alpha}^{\mu}(-1)^\beta\binom{\beta}{\alpha}D_{i_{\alpha+1}\cdots i_\beta}\sum_{\gamma=\beta}^{\mu}(-1)^\gamma\binom{\gamma}{\beta}D_{i_{\beta+1}\cdots i_\gamma}F^{;A,i_1\cdots i_\gamma}.$$

Upon interchanging the order of summation in the above we deduce the expression

$$\sum_{\gamma=\alpha}^{\mu}\left[\sum_{\beta=\alpha}^{\gamma}(-1)^{\gamma+\beta}\binom{\beta}{\alpha}\binom{\gamma}{\beta}\right]D_{i_{\alpha+1}\cdots i_\gamma}F^{;A,i_1\cdots i_\gamma}. \tag{1.6}$$

Due to the fact that

$$\sum_{\beta=\alpha}^{\gamma}(-1)^{\gamma+\beta}\binom{\beta}{\alpha}\binom{\gamma}{\beta}=\begin{cases}0, & \text{if } \gamma\neq\alpha,\\ 1, & \text{if } \gamma=\alpha,\end{cases} \tag{1.7}$$

(1.6) equals the left-hand side of (1.5), proving the proposition. The identity (1.7) is a special case $(\gamma=0)$ of the following lemma which will be useful later. □

Lemma 1.1:

$$(-1)^\tau\binom{\gamma}{\tau-\alpha}=\sum_{\beta=\gamma}^{\tau}(-1)^\beta\binom{\beta}{\alpha}\binom{\tau-\gamma}{\beta-\gamma},\quad \text{for}$$

$\tau-\gamma\geqslant 0$.

Proof. Expanding the expression $[1+x(1+y)]^{\tau-\gamma}\times(1+y)^\gamma$ via the binomial theorem and rearranging the order of summation we obtain

$$[1+x(1+y)]^{\tau-\gamma}(1+y)^\gamma=\sum_{\alpha=0}^{\tau}y^\alpha\sum_{\beta=\gamma}^{\tau}x^{\beta-\gamma}\binom{\tau-\gamma}{\beta-\gamma}\binom{\beta}{\alpha}. \tag{1.8}$$

Setting $x=-1$ in (1.8) implies, upon expanding the left-hand side of (1.8), that

$$(-1)^{\tau-\gamma}\sum_{\alpha=\tau-\gamma}^{\tau}y^\alpha\binom{\gamma}{\tau-\alpha}=\sum_{\alpha=0}^{\tau}y^\alpha\sum_{\beta=\gamma}^{\tau}(-1)^{\beta-\gamma}\binom{\beta}{\alpha}\binom{\tau-\gamma}{\beta-\gamma}. \tag{1.9}$$

Comparison of the coefficients of y^α, $\alpha=0,1,\ldots,\tau$ in (1.9) establishes the lemma. □

Thus the set of Euler operators can be used to completely determine an element of $\mathscr{F}(\mathscr{P}(U,x))$ (up to a constant function) in the same sense that the various partial derivatives do. (For this interpretation we take $\mathscr{P}(U,x):=\{\rho_A\}$.) The advantage of the Euler operators is the fact that the Euler–Lagrange operator $E^A(\mathscr{L})$ is a tensor density whenever the ρ_A's are tensorial (or partly linear connections) and when $\mathscr{L}$ is a scalar density.[2] The higher order Euler operators do not possess this property and do not appear to be tensor valued (of any type) except in a few special cases. (E.g., $T^{\cdots}_{\cdots}$ a tensor of αth order in derivatives of ρ_A implies $E^{A,i_1\cdots i_\alpha}(T^{\cdots}_{\cdots})$ is tensor valued if ρ_A is a tensor or linear connection.)

2. Commutation Properties of the Euler Operators

One of the main theorems of the paper is proven in this section. We begin with two lemmas which are fundamental to its topics.

Lemma 2.1: If $F\in\mathscr{F}(\mathscr{P}(U,x))$, then[3]

$$(D_jF)^{;A,i_1\cdots i_\alpha}=D_j(F^{;A,i_1\cdots i_\alpha})+F^{;A,(i_1\cdots i_{\alpha-1}}\delta_j^{i_\alpha)}$$

$\alpha=0,1,2,\cdots$. [In particular $(D_jF)^{;A}=D_j(F^{;A})$.] □

The proof of Lemma 2.1 follows easily from the definition of D_j.

Lemma 2.2: If $F\in\mathscr{F}(\mathscr{P}(U,x))$, then

$$[D_{j_1\cdots j_\beta}(F)]^{;A,i_1\cdots i_\alpha}=\sum_{\gamma=0}^{\alpha}\binom{\beta}{\gamma}[D_{(j_1\cdots j_{\beta-\gamma}}\delta_{j_\beta}^{(i_1}\cdots\delta_{j_\beta)}^{i_\gamma}F^{;A,i_{\gamma+1}\cdots i_\alpha)}] \tag{2.1}$$

for all $\alpha,\beta=0,1,2,\cdots$.

Proof: We use induction on β. The case $\beta=1$ is given by Lemma 2.1. Suppose (2.1) holds for all $\beta=0,1,\ldots,\tau-1$. For $\beta=\tau$ we have, from (2.1), that

$$[D_{j_1\cdots j_\tau}(F)]^{;A,i_1\cdots i_\alpha}=\sum_{\gamma=0}^{\alpha}\binom{\tau-1}{\gamma}D_{(j_1\cdots j_{\tau-\gamma-1}}\delta_{j_\tau}^{(i_1}\cdots\delta_{j_\tau}^{i_\gamma}(D_{j_\tau)}F)^{;A,i_{\gamma+1}\cdots i_\alpha)}.$$

Applying Lemma 2.1 we deduce from the above equation that

$$[D_{j_1\cdots j_\tau}(F)]^{;A,i_1\cdots i_\alpha}$$
$$= D_{j_1\cdots j_\tau}(F^{;A,i_1\cdots i_\alpha}) + \sum_{\gamma=1}^{\alpha}\left[\binom{\tau-1}{\gamma} D_{j_\tau(j_1\cdots j_{\tau-\gamma})}\right.$$
$$\times \delta^{(i_1}_{j_\tau}\cdots\delta^{i_\gamma}_{j_{\tau-1}})(F^{;A,i_{\gamma+1}\cdots i_\alpha}) + \binom{\tau-1}{\gamma-1} D_{(j_1\cdots j_{\tau-1})j_\tau}$$
$$\left.\times \delta^{(i_1}_{j_{\tau-1}}\cdots\delta^{i_\gamma}_{j_{\tau-\gamma}})\delta^{i_{\gamma+1}}_{j_\tau}(F^{;A,i_{\gamma+1}\cdots i_\alpha})\right], \tag{2.2}$$

where we have relabeled the summation index in the last term. Equation (2.1) with $\beta=\tau$ now follows from (2.2) after noting that the term in brackets under the summation sign in (2.2) is equal to

$$\binom{\tau}{\gamma} D_{(j_1\cdots j_{\tau-\gamma})} \delta^{(i_1}_{j_{\tau-\gamma+1}}\cdots\delta^{i_\gamma}_{j_\tau)}(F^{A,i_{\gamma+1}\cdots i_\alpha}) \square.$$

With the aid of Lemma 2.2 we may establish the following proposition:

Proposition 2.1: For any $F\in\mathscr{F}(\mathscr{P}(U,x))$ and any $\alpha,\beta=0,1,2,\cdots$ the identity

$$E^{A,j_1\cdots j_\beta}(F)^{;B,i_1\cdots i_\alpha}$$
$$= \sum_{\gamma=0}^{\alpha}\binom{\beta+\gamma}{\beta} E^{A,j_1\cdots j_\beta(i_1\cdots i_\gamma}(F^{;B,i_{\gamma+1}\cdots i_\alpha)}, \tag{2.3}$$

holds.

Proof: The definition of the βth Euler operator along with Lemma 2.2 imply that

$$E^{A,j_1\cdots j_\beta}(F)^{;B,i_1\cdots i_\alpha}$$
$$= \sum_{\tau=\beta}^{\beta+\alpha}(-1)^\tau\binom{\tau}{\beta}\sum_{\gamma=0}^{\tau-\beta}\binom{\tau-\beta}{\gamma} D_{j_{\beta+1}\cdots j_{\tau-\gamma}}$$
$$\times(F^{;Aj_1\cdots j_{\tau-\gamma}(i_1\cdots i_\gamma;B,i_{\gamma+1}\cdots i_\alpha)}) + \sum_{\tau=\beta+\alpha+1}^{\infty}(-1)^\tau\binom{\tau}{\beta}$$
$$\times\sum_{\gamma=0}^{\alpha}\binom{\tau-\beta}{\gamma} D_{j_{\beta+1}\cdots j_{\tau-\gamma}(i_1\cdots i_\gamma}(F^{;Aj_1\cdots j_{\tau-\gamma}(i_1\cdots i_\gamma;B,i_{\gamma+1}\cdots i_\alpha)}).$$

Upon changing the order of the sums in the first term on the right-hand side of the above equation we can combine the two sums in that equation to find that

$$E^{A,j_1\cdots j_\beta}(F)^{B,i_1\cdots i_\alpha}$$
$$= \sum_{\gamma=0}^{\alpha}\sum_{\tau=\beta+\gamma}^{\infty}(-1)^\tau\binom{\tau}{\beta}\binom{\tau-\beta}{\gamma} D_{j_{\beta+1}\cdots j_{\tau-\gamma}}$$
$$\times(F^{;Aj_1\cdots j_{\tau-\gamma}(i_1\cdots i_\gamma;B,i_{\gamma+1}\cdots i_\alpha)}). \tag{2.4}$$

In view of the identity

$$\binom{\tau}{\beta}\binom{\tau-\beta}{\gamma} = \binom{\gamma+\beta}{\beta}\binom{\tau}{\beta+\gamma}$$

and the definition of the μth Euler operation, (2.3) follows from (2.4). □

It is well known[2] that the Euler–Lagrange operator annihilates expressions of the form D_iF^i, where $F^i\in\mathscr{F}(\mathscr{P}(U,x))$. This result can be generalized to yield the next proposition.

Proposition 2.2: If $F^i\in\mathscr{F}(\mathscr{P}(U,x))$, then for $\tau=1,2,3,\cdots$

$$E^{A,j_1\cdots j_\tau}(D_iF^i) = -E^{A,(j_1\cdots j_{\tau-1}}(F^{j_\tau)}).$$

Proof: From the definition of the τth Euler operator (1.4) and Lemma 2.1 it follows that

$$E^{A,j_1\cdots j_\tau}(D_iF^i)$$
$$= \sum_{\alpha=\tau}^{\infty}(-1)^\alpha\binom{\alpha}{\tau} D_{ij_{\tau+1}\cdots j_\alpha}F^{i;Aj_1\cdots j_\alpha}$$
$$+ \sum_{\alpha=\tau}^{\infty}(-1)^\alpha\binom{\alpha}{\tau} D_{j_{\tau+1}\cdots j_\alpha}F^{(j_1;Aj_2\cdots j_\alpha)}.$$

Expanding the second summation in the above and rearranging terms we find that

$$E^{A,j_1\cdots j_\tau}(D_iF^i)$$
$$= \sum_{\alpha=\tau}^{\infty}(-1)^\alpha\binom{\alpha}{\tau} D_{ij_{\tau+1}\cdots j_\alpha}F^{i;Aj_1\cdots j_\alpha}$$
$$+ \sum_{\alpha=\tau+1}^{\infty}\left[(-1)^\alpha\binom{\alpha-1}{\tau} D_{j_{\tau+1}\cdots j_\alpha}F^{j_\alpha;Aj_1\cdots j_{\alpha-1}}\right]$$
$$+ \sum_{\alpha=\tau}^{\infty}\left[(-1)^\alpha\binom{\alpha-1}{\tau-1} D_{j_{\tau+1}\cdots j_\alpha}F^{(j_1;Aj_2\cdots j_\tau)j_{\tau+1}\cdots j_\alpha}\right].$$

In the above, the first and second terms on the right-hand side cancel while the last term gives $-E^{A,(j_1\cdots j_{\tau-1}}(F^{j_\tau)})$, proving the proposition. Note that the proof also works for $\tau=0$ upon defining $\binom{\gamma}{-1}=0$, $\gamma=0,1,\cdots$. This gives $E^A(D_iF^i)\equiv 0$. □

Proposition 2.2 forms the basis of the following:

Proposition 2.3: If $F\in\mathscr{F}(\mathscr{P}(U,x))$, then for all $\alpha,\beta=0,1,2,\cdots$,

$$E^{A,i_1\cdots i_\alpha}[E^{Bj_1\cdots j_\beta}(F)]$$
$$= (-1)^\beta\sum_{\gamma=0}^{\alpha}\binom{\beta+\gamma}{\beta} E^{A,(i_1\cdots i_{\alpha-\gamma}|}(F^{;Bj_1\cdots j_\beta|i_{\alpha-\gamma+1}\cdots i_\alpha)}).$$

Proof: The proof is inductive in nature. From Proposition 2.2 we deduce that

$$E^{A,i_1\cdots i_\alpha}[E^{Bj_1\cdots j_\beta}(F)]$$
$$= (-1)^\beta\left[E^{A,i_1\cdots i_\alpha}(F^{;Bj_1\cdots j_\beta}) - \sum_{\mu=\beta+1}^{\infty}(-1)^{\mu-\beta}\binom{\mu}{\beta}\right.$$
$$\left.\times E^{A,(i_1\cdots i_{\alpha-1}|}(D_{j_{\beta+1}\cdots j_\mu}F^{Bj_1\cdots j_{\mu-1}|i_\alpha)})\right].$$

Now suppose that for $1 \leqslant \tau < \alpha$,

$$E^{A,i_1\cdots i_\alpha}[E^{B j_1\cdots j_\beta}(F)]$$
$$= (-1)^\beta \left[\sum_{\mu=0}^{\tau-1} \binom{\beta+\mu}{\beta} E^{A,(i_1\cdots i_{\alpha\ \mu}|}(F^{;B j_1\cdots j_\beta|i_{\alpha\ \mu+1}\cdots i_\alpha)} \right]$$
$$+ (-1)^\tau \left[\sum_{\mu=\beta+\tau}^{\infty} (-1)^\mu \binom{\mu}{\beta} E^{A,(i_1\cdots i_{\alpha\ \tau}|}(D_{j_{\beta+1}\cdots j_{\mu\ \tau}} \right.$$
$$\left. \times F^{;B j_1\cdots j_{\mu\ \tau}|i_{\alpha\ \tau+1}\cdots i_\alpha)} \right]. \tag{2.5}$$

Proposition 2.2 now implies that

$$E^{A,i_1\cdots i_\alpha}[E^{B j_1\cdots j_\beta}(F)]$$
$$= (-1)^\beta \left[\sum_{\mu=0}^{\tau-1} \binom{\beta+\mu}{\beta} E^{A,(i_1\cdots i_{\alpha\ \mu}|}(F^{;B j_1\cdots j_\beta|i_{\alpha\ \mu+1}\cdots i_\alpha)} \right.$$
$$\left. + \binom{\beta+\tau}{\beta} E^{A,(i_1\cdots i_{\alpha\ \tau}|}(F^{;B j_1\cdots j_\beta|i_{\alpha\ \tau+1}\cdots i_\alpha)} \right]$$
$$+ \left[(-1)^{\tau+1} \sum_{\mu=\beta+\tau+1}^{\infty} (-1)^\mu \binom{\mu}{\beta} E^{A,(i_1\cdots i_{\alpha\ \tau}|} \right.$$
$$\left. \times (D_{j_{\beta+1}\cdots j_{\mu\ \tau}} F^{B j_1\cdots j_{\mu\ \tau}|i_{\alpha\ \tau+1}\cdots i_\alpha)}) \right].$$

We have shown that (2.5) holds for $\tau + 1$ whenever it holds for $\tau (< \alpha)$ and since it holds for $\tau = 1$ it holds for all $\tau \leqslant \alpha$. When $\tau = \alpha$ the above equation implies the Proposition. □

With the aid of the previous propositions we may establish the main theorem of this section.

Theorem 2.1: For any $F \in \mathscr{F}(\mathscr{P}(U,x))$ and any integer $0 \leqslant \alpha \leqslant \tau$, where $\tau = 0,1,2,\cdots$, the identity

$$E^{A,(i_1\cdots i_\alpha}[E^{B,i_{\alpha+1}\cdots i_\tau)}(F)]$$
$$= \sum_{\gamma=\alpha}^{\tau} (-1)^{\gamma-\alpha} \binom{\gamma}{\alpha} E^{A,(i_{\gamma+1}\cdots i_\tau}(F)^{;B,i_1\cdots i_\gamma)} \tag{2.6}$$

holds.

Proof: From Proposition 2.3 it follows, upon relabeling the summation indices, that the left-hand side of (2.6) becomes

$$(-1)^{\tau-\alpha} \sum_{\gamma=\tau-\alpha}^{\tau} \binom{\gamma}{\tau-\alpha} E^{A,(i_{\gamma+1}\cdots i_\tau}(F^{;B,i_1\cdots i_\gamma)}). \tag{2.7}$$

From Proposition 2.1 we deduce that

$$E^{A,(i_{\gamma+1}\cdots i_\tau}(F)^{;B,i_1\cdots i_\gamma)}$$
$$= \sum_{\mu=\tau-\gamma}^{\tau} \binom{\mu}{\tau-\gamma} E^{A,(i_{\tau\ \mu+1}\cdots i_\tau}(F^{;B,i_1\cdots i_{\tau\ \mu})},$$

and thus the right-hand side of (2.6) becomes

$$\sum_{\gamma=\alpha}^{\tau} \sum_{\mu=\tau-\gamma}^{\tau} (-1)^{\gamma-\alpha} \binom{\gamma}{\alpha} \binom{\mu}{\tau-\gamma} E^{A,(i_{\tau-\mu+1}\cdots i_\tau}(F^{;B,i_1\cdots i_{\tau-\mu})}.$$

In order to compare this last expression with (2.7) we relabel the summations to obtain the expression

$$\sum_{\gamma=\alpha}^{\tau} \sum_{\nu=0}^{\gamma} (-1)^{\gamma-\alpha} \binom{\gamma}{\alpha} \binom{\tau-\nu}{\tau-\gamma} E^{A,(i_{\nu+1}\cdots i_\tau}(F^{;B,i_1\cdots i_\nu)}).$$

Changing the order of summation in the above expression we deduce that the right-hand side of (2.6) is equal to the expression

$$\sum_{\nu=0}^{\tau} \left[\sum_{\gamma=\nu}^{\tau} (-1)^{\gamma-\alpha} \binom{\gamma}{\alpha} \binom{\tau-\nu}{\tau-\gamma} \right] E^{A,(i_{\nu+1}\cdots i_\tau}(F^{;B,i_1\cdots i_\nu)}). \tag{2.8}$$

The equality of expressions (2.7) and (2.8) now follows from Lemma 1.1 and the fact that

$$\binom{\tau-\nu}{\tau-\gamma} = \binom{\tau-\nu}{\gamma-\nu},$$

proving the theorem. □

An important corollary to Theorem 2.1 is the the following result due to Anderson[4,5]:

Corollary 2.1: If $F \in \mathscr{F}(\mathscr{P}(U,x))$, then $E^A(F)$ satisfies

$$E^{A,i_1\cdots i_\alpha}[E^B(F)] = E^A(F)^{;B,i_1\cdots i_\alpha},$$

for all $\alpha = 0,1,2,3\cdots$. □

Theorem 2.1 can be bypassed in proving Corollary 2.1 by taking the cases $\beta = 0$ in Propositions 2.1 and 2.3. The importance of Corollary 2.1 is that it gives necessary conditions for a collection of functions $T^A \in \mathscr{F}(\mathscr{Q}(U,x))$ $[(\mathscr{Q}(U,x)) = \{\rho_A\}]$, to satisfy in order for there to exist a function $L \in \mathscr{F}(\mathscr{Q}(U,x))$ for which $T^A = E^A(L)$, namely $E^{A,i_1\cdots i_\alpha}(T^B) = T^{A;B,i_1\cdots i_\alpha}$ for all $\alpha = 0,1,2,\cdots$. These conditions can be applied to concomitant problems related to gravitation and other field theories.[6,7] They are also basic to the study of conservation laws related to tensorial Euler–Lagrange expressions.[4,5,7,8]

An interesting consequence of Theorem 2.1 is based upon the identity (set $\tau = \alpha + 1$ in Theorem 2.1)

$$E^A(F)^{;B,i_1\cdots i_\alpha}$$
$$= \frac{1}{\alpha}\{E^{A,(i_1}(F)^{;B,i_2\cdots i_\alpha)} - E^{A,(i_1\cdots i_{\alpha\ 1}}[E^{B,i_\alpha)}(F)]\}, \tag{2.9}$$

where $\alpha = 1,2,\cdots$. If $T^{A,i} \in \mathscr{F}(\mathscr{Q}(U,x))$ is a set of functions for which there exists an $F \in \mathscr{F}(\mathscr{Q}(U,x))$ such that $T^{A,i} = E^{A,i}(F)$, then the equations

$$T^{A;B,i_1\cdots i_\alpha} = \frac{1}{\alpha}[T^{A,(i_1;B,i_2\cdots i_\alpha)} - E^{A,(i_1\cdots i_{\alpha\ 1}}(T^{B,i_\alpha)})] \tag{2.10}$$

(for $\alpha = 1,2,\cdots$), form an integrable system of partial differential equations for some functions $T^A \in \mathscr{F}(\mathscr{Q}(U,x))$. It is clear from (2.9) that T^A must take the form

$$T^A = E^A(F) + T_0^A, \tag{2.11}$$

where $T_0^A = T_0^A(\rho_B)$. Hence knowledge of $E^{A,i}(F)$ essentially determines $E^A(F)$ via (2.10) and (2.11). The properties of the first Euler operator $E^{A,i}$ will be examined further in Sec. 5.

We conclude this section with the following alternate

form of Theorem 2.1 which is easily obtained (from the theorem) using Lemma 1.1 with $\gamma = 0$.

Corollary 2.2: For any $F \in \mathcal{F}(\mathcal{P}(U,x))$ and any integer α, where $0 \leqslant \alpha \leqslant \tau$ and $\tau = 0,1,2,\cdots$, the identity

$$E^{A,(i_{\alpha+1}\cdots i_\tau)}(F)^{;B,i_1\cdots i_\alpha)} = \sum_{\beta=\alpha}^{\tau} \binom{\beta}{\alpha} E^{A,(i_1\cdots i_\beta}[E^{B,i_{\beta+1}\cdots i_\tau)}(F)],$$

holds. □

3. A PRODUCT RULE AND ITS CONSEQUENCES

In the previous section, necessary conditions for a collection of functions $T^A \in \mathcal{F}(\mathcal{P}(U,x))$ to be an Euler–Lagrange expression were obtained. In certain cases these conditions are also sufficient as has been shown by Horndeski,[9] Atherton and Homsy,[10] Anderson,[5] and Ahner and Moose.[11] In this section we study the result of applying the Euler operators to products of elements of $\mathcal{F}(\mathcal{P}(U,x))$. As a special case of this we obtain a product rule for the Euler–Lagrange operator which leads to very simple derivations of the above mentioned sufficiency conditions.

Proposition 3.1: For $F,G \in \mathcal{F}(\mathcal{P}(U,x))$,

$$E^{A,i_1\cdots i_\alpha}(FG) = \sum_{\beta=\alpha}^{\infty} \binom{\beta}{\alpha}[D_{i_{\alpha+1}\cdots i_\beta}(G)E^{A,i_1\cdots i_\beta}(F) + D_{i_{\alpha+1}\cdots i_\beta}(F)E^{A,i_1\cdots i_\beta}(G)] \tag{3.1}$$

Proof: To prove (3.1) we require the following Leibniz rule from calculus,

$$D_{i_1\cdots i_\gamma}(FG) = \sum_{\tau=0}^{\gamma} \binom{\gamma}{\tau} D_{(i_1\cdots i_\tau}(F)D_{i_{\tau+1}\cdots i_\gamma)}(G). \tag{3.2}$$

From the definition of the αth Euler operator (1.4) it follows that

$$E^{A,i_1\cdots i_\alpha}(FG) = \sum_{\gamma=\alpha}^{\infty} (-1)^\gamma \binom{\gamma}{\alpha} D_{i_{\alpha+1}\cdots i_\gamma} \times (GF^{;A,i_1\cdots i_\gamma} + FG^{;A,i_1\cdots i_\gamma}).$$

Applying (3.2) to the above we deduce that

$$E^{A,i_1\cdots i_\alpha}(FG) = \sum_{\gamma=\alpha}^{\infty} (-1)^\gamma \binom{\gamma}{\alpha} \sum_{\tau=0}^{\gamma-\alpha} \binom{\gamma-\alpha}{\tau} \times [D_{(i_{\alpha+1}\cdots i_{\alpha+\tau}}(G)D_{i_{\alpha+\tau+1}\cdots i_\gamma)}F^{;A,i_1\cdots i_\gamma} + D_{(i_{\alpha+1}\cdots i_{\alpha+\tau}}(F)D_{i_{\alpha+\tau+1}\cdots i_\gamma)}G^{;A,i_1\cdots i_\gamma}].$$

Redefining the range of the second summation on the right-hand side of the above equation so that $\tau = \alpha,\ldots,\gamma$, then using the fact that

$$\binom{\gamma}{\alpha}\binom{\gamma-\alpha}{\tau-\alpha} = \binom{\tau}{\alpha}\binom{\gamma}{\tau},$$

we interchange the summations in that equation to obtain

$$E^{A,i_1\cdots i_\alpha}(FG) = \sum_{\tau=\alpha}^{\infty} \binom{\tau}{\alpha} D_{i_{\alpha+1}\cdots i_\tau}(G) \times \left[\sum_{\gamma=\tau}^{\infty} (-1)^\gamma \binom{\gamma}{\tau} D_{i_{\tau+1}\cdots i_\gamma}F^{;A,i_1\cdots i_\gamma}\right] + \sum_{\tau=\alpha}^{\infty} \binom{\tau}{\alpha} D_{i_{\alpha+1}\cdots i_\tau} \times (F)\left[\sum_{\gamma=\tau}^{\infty} (-1)^\gamma \binom{\gamma}{\tau} D_{i_{\tau+1}\cdots i_\gamma}G^{;A,i_1\cdots i_\gamma}\right]$$

noting the symmetry of $i_1\cdots i_\gamma$. Equation (3.1) now follows from this last equation and Eq. (1.4). □

A useful special case of Proposition 3.1 occurs when $\alpha = 0$ in (3.1) yielding the product rule

$$E^A(FG) = \sum_{\beta=0}^{\infty} [D_{i_1\cdots i_\beta}(G)E^{A,i_1\cdots i_\beta}(F) + D_{i_1\cdots i_\beta}(F)E^{A,i_1\cdots i_\beta}(G)], \tag{3.3}$$

where $F,G \in \mathcal{F}(\mathcal{P}(U,x))$. Equation (3.3) is the basis of the remaining results of this section.

A set of field functions $\mathcal{P}(U,x) = \{\rho_A;\lambda_\Omega\}$ will be called ρ-star-shaped with respect to ${}^0\rho$ if for all $t \in [0,1]$, $({}^t\rho_A;\lambda_\Omega) \in \mathcal{P}(U,x)$, where ${}^t\rho_A := t(\rho_A - {}^0\rho_A) + {}^0\rho_A$ and $(\rho_A;\lambda_\Omega) \in \mathcal{P}(U,x)$.

Theorem 3.1[5,9,10,11]: Suppose $\mathcal{P}(U,x)$ is ρ-star-shaped with respect to ${}^0\rho$ and let $T^A \in \mathcal{F}(\mathcal{P}(U,x))$ satisfy

$$E^{A,i_1\cdots i_\alpha}(T^B) = T^{A;B,i_1\cdots i_\alpha}, \tag{3.4}$$

for all $\alpha = 0,1,2,\cdots$. Then the function

$$L := \int_0^1 {}^tT^A(\rho_A - {}^0\rho_A)dt, \tag{3.5}$$

where

$${}^tT^A := T^A({}^t\rho_B;{}^t\rho_{B,i};\cdots;{}^t\rho_{B,i_1\cdots i_\beta};\lambda_\Omega;\lambda_{\Omega,i};\cdots;\lambda_{\Omega,i_1\cdots i_\gamma}),$$

satisfies

$$T^A = E^A(L). \tag{3.6}$$

Proof: Equation (3.3) implies that

$$E^B[{}^tT^A(\rho_A - {}^0\rho_A)] = {}^tT^B + \sum_{\alpha=0}^{\infty} [E^{B,i_1\cdots i_\alpha}({}^tT^A)(\rho_A - {}^0\rho_A)_{,i_1\cdots i_\alpha}].$$

In view of the fact that

$$\frac{\partial}{\partial\rho_{B,i_1\cdots i_\alpha}} = t\frac{\partial}{\partial{}^t\rho_{B,i_1\cdots i_\alpha}}, \quad \alpha = 0,1,2,\ldots,$$

we deduce from (3.4) that $E^{B,i_1\cdots i_\alpha}({}^tT^A) = t\partial{}^tT^B/\partial{}^t\rho_{A,i_1\cdots i_\alpha}$. Thus

$$E^B[{}^tT^A(\rho_A - {}^0\rho_A)] = {}^tT^B + t\sum_{\alpha=0}^{\infty}\left[\frac{\partial{}^tT^B}{\partial{}^t\rho_{A,i_1\cdots i_\alpha}}(\rho_{A,i_1\cdots i_\alpha} - {}^0\rho_{A,i_1\cdots i_\alpha})\right],$$

from which we deduce the formula $(dt\,{}^tT^B/dt) = E^B[{}^tT^A(\rho_A - {}^0\rho_A)]$. Equations (3.6) and (3.5) follow upon integrating with respect to t from 0 to 1. □

While Theorem 3.1 is a major step towards a solution to the inverse problem for field theories, it is still deficient in certain aspects. Foremost among these is the possibility that the given functions $T \in \mathscr{F}(\mathscr{P}(U,x))$ do not satisfy (3.4) yet there exists a transformation to an equivalent[12] set of functions $T^A \in \mathscr{F}(\widetilde{\mathscr{P}}(U,x))$ which do. The author does not know of any progress towards resolving this problem except in the case of single integral problems in the calculus of variations.[13]

Another difficulty with Theorem 3.1 which cannot be resolved is the nature of the space of field functions $\mathscr{P}(U,x)$. The theorem applies only when the field functions are star-shaped and, for example, is inapplicable to Lorentz metric tensor fields. Furthermore, whenever the function ${}^0\rho_A \neq 0$, the resulting Lagrangian (3.5) depends upon ${}^0\rho_A$ in addition to ρ_A and hence is an element of $\mathscr{F}(\widetilde{\mathscr{P}}(U,x))$, where $\widetilde{\mathscr{P}}(U,x) = \{\rho_A;\lambda_\alpha;{}^0\rho_A\} \neq \mathscr{P}(U,x)$. (Note that ${}^0\rho_A$ is fixed but ρ_A and λ_Ω may range over different field function values.) This feature of Theorem 3.1 leads to some peculiar consequences as will be seen in the following section. Note that when ${}^0\rho_A \neq 0$ we must extend the definition of E^A in (3.6) to include $\mathscr{F}(\widetilde{\mathscr{P}}(U,x))$.

A partial solution to these difficulties can be obtained for functions $T^A \in \mathscr{F}(\mathscr{P}(U,x))$ which are homogeneous of certain degrees in the field functions ρ_A and their derivatives. Since many functions T^A can be decomposed into a sum of homogeneous terms, the following result, which is a trivial consequence of (3.3), is of some use:

Proposition 3.2[5]: Let $T^A \in \mathscr{F}(\mathscr{P}(U,x))$ be homogeneous of degree m in the ρ_A's and their derivatives. If T^A satisfies

$$E^{B,i_1\cdots i_\alpha}(T^A) = T^{B;A,i_1\cdots i_\alpha}, \tag{3.7}$$

$\alpha = 0,1,2,\cdots$, then

$$(m+1)T^A = E^A(T^B\rho_B). \tag{3.8}$$

Proof: From Eq. (3.3) we find that

$$E^A(T^B\rho_B) = T^A + \sum_{\alpha=0}^{\infty} E^{A,i_1\cdots i_\alpha}(T^B)\rho_{B,i_1\cdots i_\alpha}.$$

Applying (3.7) to the above yields the equation

$$E^A(T^B\rho_B) = T^A + \sum_{\alpha=0}^{\infty} T^{A;B,i_1\cdots i_\alpha}\rho_{B,i_1\cdots i_\alpha}$$

from which (3.8) follows using Euler's theorem on homogeneous functions. □

Corollary 3.1: If $m \neq -1$, then $T^A = E^A(L)$, where

$$L = \frac{1}{m+1} T^B\rho_B. \quad \square$$

To handle the case $m = -1$ a different approach is required. This is the topic of the next section.

4. APPLICATIONS OF A CERTAIN HOMOGENEITY PROPERTY

Let A^{ij} denote the second order tensor density concomitant of a metric tensor which is locally given by

$$A^{ij}(g_{ab};g_{ab,c};g_{ab,cd}) = \sqrt{g}G^{ij}, \tag{4.1}$$

where G^{ij} is the Einstein tensor and $g = |\det(g_{ij})|$. (For further notation see Ref. 14.) If we restrict attention to 4-spaces M, then clearly at any point P of M,

$$A^{ij}(tg_{ab};tg_{ab,c};tg_{abcd}) = A^{ij}(g_{ab};g_{ab,c};g_{ab,cd}), \tag{4.2}$$

where $t \in \mathbb{R}^+$. (A^{ij} is homogeneous of degree zero in g_{ab} and its derivatives.) It is well-known that A^{ij} is derivable from a variational principle and hence satisfies Eqs. (3.7) with $\mathscr{P}(U,x) = \{g_{ij}\}$. Therefore, by Proposition 3.2, $A^{ij} = E^{ij}(g_{ab}A^{ab}) = E^{ij}(-\sqrt{g}R)$ as we expect. The scalar density concomitant $A := -\sqrt{g}R$ is also homogeneous in g_{ab} and its derivatives, but of degree one.

Since A^{ij} is a tensor density there is another homogeneity condition which it satisfies. Applying the transformation $x^i = t\bar{x}^i$ ($t \in \mathbb{R}^+$) to A^{ij} we deduce the relation

$$t^2A^{ij}(g_{ab};g_{ab,c};g_{ab,cd}) = A^{ij}(t^2g_{ab};t^3g_{ab,c};t^4g_{ab,cd})$$

at any point P of M, which simplifies using (4.2) to

$$t^2A^{ij}(g_{ab};g_{ab,c};g_{ab,cd}) = A^{ij}(g_{ab};tg_{ab,c};t^2g_{ab,cd}). \tag{4.3}$$

Since $P \in M$ is arbitrary (as is the metric g_{ab}) it follows that (4.3) holds throughout M for any metric tensor on M.

An important feature of (4.3) is that any tensor density concomitant T^{ij} of the metric and its first two derivatives which satisfies (4.3) and in addition satisfies $T^{ij}|_j = 0$ must be $a\sqrt{g}G^{ij}$ where a is a constant.[14] A significant observation is that the scalar density A exhibits similar properties, namely,

$$t^2A(g_{ab};g_{ab,c};g_{ab,cd}) = A(g_{ab};tg_{ab,c};t^2g_{ab,cd}), \tag{4.4}$$

for all $t \in \mathbb{R}^+$. Furthermore, any scalar density concomitant of the metric and its first two derivatives which satisfies (4.4) must be $a\sqrt{g}R$, where a is a constant.[14] The natural problem to consider here is the extent to which the property (4.3) of A^{ij} determines, through the relation $A^{ij} = E^{ij}(A)$, the property (4.4) of A.

To formulate the relevant theorems we shall require some additional notation. Let $F \in \mathscr{F}(\mathscr{P}(U,x))$ and define for γ and γ_Ω nonnegative integers,

$$F(t) := F(t^\gamma\rho_B;t^{\gamma+1}\rho_{B,i};\cdots;t^{\gamma+\alpha}\rho_{B,i_1\cdots i_\alpha};$$
$$t^{\gamma_\Omega}\lambda_\Omega;t^{\gamma_\Omega+1}\lambda_{\Omega,i};\cdots;t^{\gamma_\Omega+\alpha'}\lambda_{\Omega,i_1\cdots i_{\alpha'}}), \tag{4.5}$$

(no sums over Ω), where $t \in \mathbb{R}^+$. We shall be concerned with those $F \in \mathscr{F}(\mathscr{P}(U,x))$ for which $F(t)$ is defined for all $t \in \mathbb{R}^+$ and which satisfy the condition [generalizing (4.3) and (4.4)],

$$t^\beta F = F(t), \tag{4.6}$$

for some β, γ, γ_Ω and all $t \in \mathbb{R}^+$. One of the important consequences of (4.6) is that generally one can "reconstruct" the function F in the same sense that $\sqrt{g}G^{ij}$ was reconstructed from (4.3). This reconstruction process can best be demonstrated by examining a special case.

Lemma 4.1: Let $\mathcal{Q}(U,x) = \{\rho_A\}$ and let $G \in \mathcal{F}(\mathcal{Q}(U,x))$ satisfy (4.6) with $\gamma = 0$ and β a positive integer. If $\mathcal{Q}(U,x)$ is such that $\lim_{t\to 0^+} H(t) \in \mathcal{F}(\mathcal{Q}(U,x))$ whenever $H \in \mathcal{F}(\mathcal{Q}(U,x))$ and if G is of class C^β, then G is a polynomial in the derivatives of ρ_A. Furthermore G is a sum of terms of the form

$$\phi(\rho_A)^{A_1 i_1 \cdots i_{\mu_1} A_2 j_1 \cdots j_{\mu_2} \cdots A_\nu l_1 \cdots l_{\mu_\nu}} \rho_{A_1 i_1 \cdots i_{\mu_1}} \rho_{A_2 j_1 \cdots j_{\mu_2}} \cdots \rho_{A_\nu l_1 \cdots l_{\mu_\nu}},$$

where

$$\sum_{k=1}^{\nu} \mu_k = \beta. \square$$

The proof of Lemma 1.1 is basically the same as that used to deduce Eq. (2.5) in Ref. 14. Differentiate β times with respect to t in (4.6) (with G instead of F) and then take the limit as $t \to 0^+$.

The main result of this section can now be derived.

Theorem 4.1: Suppose $\mathcal{P}(U,x)$ is such that $\lim_{t\to 0^+} H(t) \in \mathcal{F}(\mathcal{P}(U,x))$ whenever $H \in \mathcal{F}(\mathcal{P}(U,x))$. Let $T^A \in \mathcal{F}(\mathcal{P}(U,x))$ satisfy

$$t^\beta T^A = T^A(t), \tag{4.7}$$

where $t \in \mathbb{R}^+$ and β, γ, and γ_Ω [cf. (4.5)] are all nonnegative integers.

If there exists an $L \in \mathcal{F}(\mathcal{P}(U,x))$ (of order α in derivatives of ρ_A), defined throughout $\mathcal{P}(U,x)$ and of class (at least) $C^{\beta+\gamma+\alpha+1}$ for which $T^A = E^A(L)$, then

$$T^A = E^A(\tilde{L}), \tag{4.8}$$

where

$$\tilde{L} := \frac{1}{(\beta+\gamma)!} \lim_{t\to 0^+} \frac{d^{\beta+\gamma}}{dt^{\beta+\gamma}} [L(t)]. \tag{4.9}$$

Proof: By direct computation using the chain rule we deduce that for all $t \in \mathbb{R}^+$,

$$E^A[L(t)] = t^\gamma E^A(L)(t).$$

Thus

$$t^\gamma T^A(t) = E^A[L(t)],$$

and by (4.7) we deduce that

$$t^{\beta+\gamma} T^A = E^A[L(t)]. \tag{4.10}$$

Due to the fact that $L(t)$ is a differentiable function of the set of real variables $\{t; \rho_A; \rho_{A,i}; \cdots; \rho_{A,i_1 \cdots i_\alpha}; \lambda_\Omega; \lambda_{\Omega,i}; \cdots; \lambda_{\Omega,i_1 \cdots i_\alpha}\}$ the derivative d/dt commutes with the derivatives $\partial/\partial\rho_{A,i_1 \cdots i_\mu}$. If we consider $L(t)$ as a function on $x(U) \times \mathbb{R}^+$ [$\rho_A = \rho_A(x^i)$, $\lambda_\Omega = \lambda_\Omega(x^i)$ on $x(U)$], then the differentiability requirements on L imply that the derivatives d/dt and d/dx^i ($= D_i$) commute. Therefore, upon differentiating $\beta + \gamma$ times with respect to t in (4.10) we deduce that

$$(\beta+\gamma)! T^A = E^A\left(\frac{d^{\beta+\gamma} L(t)}{dt^{\beta+\gamma}}\right).$$

Equations (4.8) and (4.9) now follow from the above equation upon taking the limit as $t \to 0^+$. $\square$

Theorem 4.1 can often be used to solve the inverse problem considered in Sec. 3 when Theorem 3.1 and Proposition 3.2 do not apply. As an example consider the tensor density concomitant of a metric tensor given by

$$T^{ij} := \sqrt{g} R (\tfrac{1}{2} g^{ij} R - 2R^{ij}) + 2\sqrt{g} R^{|ij} - 2\sqrt{g} g^{ij} R^{|k}{}_{|k},$$

in a 4-space. The problem is to determine if $T^{ij} = E^{ij}(\mathcal{L})$, where $\mathcal{L}$ is a scalar density. Evidently, even if T^{ij} can be shown to satisfy the conditions in (3.7), T^{ij} is homogeneous of degree -1 in g_{ab} and its derivatives so that Proposition 3.2 will not apply. If we require that g_{ab} be a Lorentzian metric tensor then Theorem 3.1 is also inapplicable. We can employ Theorem 4.1 by noting that under (4.5) (with $\gamma = 0$) we obtain the relation $t^4 T^{ij} = T^{ij}(t)$. If $\mathcal{L}$ is a scalar density concomitant of the metric tensor and its derivatives, then by demanding that $T^{ij} = E^{ij}(\mathcal{L})$ we find that the scalar density

$$\tilde{\mathcal{L}} := \frac{1}{4!} \lim_{t\to 0^+} (d^4/dt^4) \mathcal{L}(t)$$

must also satisfy $T^{ij} = E^{ij}(\tilde{\mathcal{L}})$. We show that $\tilde{\mathcal{L}}$ is a scalar density by construction. Applying the replacement theorem[15] to $\mathcal{L}(t)$ we find that

$$\mathcal{L}(t) = \mathcal{L}(g_{ab}; 0; t^2 g_{abcd}; \cdots; t^\alpha g_{abc_1 \cdots c_\alpha}),$$

where the $g_{abc_1 \cdots c_\delta}$, $\delta = 2, \ldots, \alpha$ are the corresponding metric normal tensors[15] of g_{ab}. Hence upon applying the derivatives with respect to t and then taking the limit as $t \to 0^+$ in the result we find that

$$\tilde{\mathcal{L}} = \eta^{abcdrstu}(g_{kl}) g_{abcd} g_{rstu} + \eta^{abcdrs}(g_{kl}) g_{abcdrs}.$$

Since the metric normal tensors may be rewritten in terms of covariant derivatives of the curvature tensor,[15] $\tilde{\mathcal{L}}$ can be written in the form

$$\tilde{\mathcal{L}} = \zeta^{abcdrstu}(g_{kl}) R_{abcd} R_{rstu} + \zeta^{abcdrs}(g_{kl}) R_{abcd|rs}, \tag{4.11}$$

where the $\zeta^{\cdots}$ coefficients are tensor densities built from g_{kl} only. The last term in (4.11) is clearly a divergence and it can be shown[16] that the first term must take the form

$$\mathcal{L}_1 = a\sqrt{g} R^2 + b\sqrt{g} R_{ij} R^{ij} + c\sqrt{g} R_{ijkl} R^{ijkl} + d\epsilon^{ijkl} R^{ab}{}_{ij} R_{abkl}$$

where a, b, c, and d are constants. The coefficients of d in $\mathcal{L}_1$ has an identically vanishing Euler–Lagrange expression.[17] Since the Euler–Lagrange expression corresponding to $\sqrt{g}(R^2 - 4R_{ij}R^{ij} + R_{ijkl}R^{ijkl})$ also vanishes identically[17] it is clear that $\mathcal{L}_1$ will have the same Euler–Lagrange expressions as

$$\mathcal{L}_2 = (a-c)\sqrt{g} R^2 + (b+4c)\sqrt{g} R_{ij} R^{ij}.$$

Demanding that $E^{ij}(\mathcal{L}) = T^{ij}$ implies that $E^{ij}(\mathcal{L}_2) = T^{ij}$ which[18] yields the result $a - c = 1$ and $b + 4c = 0$, i.e.,

$$T^{ij} = E^{ij}(\sqrt{g} R^2).$$

In summary, we have answered the question of whether or not T^{ij} was an Euler–Lagrange tensor by explicitly constructing all possible effective Lagrangians which were sca-

lar densities. Then by direct computation we found a suitable Lagrangian.

Another application of Theorem 4.1 is the problem of determining a scalar density $\mathscr{L}$ which yields the Cotton tensor density defined by

$$C^{ij} = \epsilon^{iab} R^{j}{}_{a|b} + \epsilon^{jab} R^{i}{}_{a|b}$$

(in a 3-space), as an Euler–Lagrange expression. Horndeski[9] has shown that C^{ij} satisfies Eqs. (3.4) ($\rho_A = g_{ij}$) and has constructed a scalar density $\mathscr{V}$ of the form

$$\mathscr{V} = \mathscr{V}(g_{ab};\cdots;g_{ab,cde};h_{ab};\cdots;h_{ab,cde})$$

for which $E^{ij}(\mathscr{V}) = C^{ij}$, where g_{ab} and h_{ab} are both positive (for negative) definite metric tensors. This result is basically a consequence of Theorem 3.1. The case of indefinite metrics has not been solved at present but if we seek scalar densities built from g_{ab} and its derivatives (only), Theorem 4.1 can be used to prove the following:

Theorem 4.2: There does not exist a class $C^{\alpha+4}$ scalar density $\mathscr{L}$ of the form $\mathscr{L} = \mathscr{L}(g_{ab};g_{ab,c};\cdots;g_{ab,c_1\cdots c_\alpha})$ for which $E^{ij}(\mathscr{L}) = C^{ij}$ in a 3-space. □

The proof of Theorem 4.2 can be found in Ref. 19. If we relax the demand that $\mathscr{L}$ be a scalar density, then the Lagrangian $\bar{\mathscr{L}}$ of (4.9) can be written, after subtracting out a divergence, in the equivalent form

$$L' = \xi^{abcpqrstu} g_{ab,c} g_{pq,r} g_{st,u} + \xi^{abcpqrs} g_{ab,c} g_{pq,rs},$$

where the $\xi^{\cdots}$ coefficients are functions of g_{ab} only. The question of the existence of coefficients $\xi^{\cdots}$ for which $E^{ij}(L') = C^{ij}$ is unresolved at present.

In spaces of dimension $(2n+1)$ for $n = 2,3,\cdots$, one can define tensor densities which exhibit properties similar to those of C^{ij} tensor. It can be shown that Theorem 4.2 generalizes to incorporate these tensors[20] so that C^{ij} cannot be considered unique (in the sense that it does not follow from a scalar density Lagrangian built from one metric).

5. EQUIVALENT LAGRANGIANS AND THE FIRST EULER OPERATOR

In this section we consider the field functions $\mathscr{T}(U,x) := \{\rho_{A_\mu}\}_{\mu=1}^{\mu_0}$ where for each μ, ρ_{A_μ} is a tensor. Define the integer κ_μ to be the number of covariant indices on ρ_{A_μ} minus the number of contravariant indices on ρ_{A_μ}. The following lemma is basic to our considerations here:

Lemma 5.1: Suppose T^{A_ν} is a tensor density where for each A_ν, $T^{A_\nu} \in \mathscr{F}(T(U,x))$. (The number of covariant indices on T^{A_ν} equals the number of contravariant indices on ρ_{A_ν} and similarly for the contravariant indices on T^{A_ν}.) Then

$$(n-\kappa_\nu) T^{A_\nu} = \sum_{\mu=1}^{\mu_0} \left(\kappa_\mu \sum_{\alpha=0}^{\infty} T^{A_\nu;B_\mu i_1\cdots i_\alpha} \rho_{B_\mu i_1\cdots i_\alpha} + \sum_{\alpha=1}^{\infty} \alpha T^{A_\nu;B_\mu i_1\cdots i_\alpha} \rho_{B_\mu i_1\cdots i_\alpha} \right) \tag{5.1}$$

where n is the dimension of the space. □

The proof of Lemma 5.1 follows from Theorem 2.1 of Ref. 21 upon contracting over s and r in (2.14) of that paper.

Theorem 5.1: If for some $L \in \mathscr{F}(\mathscr{T}(U,x))$, $E^{A_\nu}(L)$ is a tensor density for each $\nu = 1,\ldots,\mu_0$, then

$$E^{A_\nu}(L) = E^{A_\nu}(L'),$$

where

$$L' = \frac{1}{n} \sum_{\mu=1}^{\mu_0} (\kappa_\mu E^{A_\mu}(L) \rho_{A_\mu} - E^{A_\mu,k}(L) \rho_{A_\mu,k}). \tag{5.2}$$

Proof: Applying Proposition 3.1 we find that

$$E^{A_\nu}(L') = \frac{1}{n} \Bigg(\Bigg\{ \sum_{\mu=1}^{\mu_0} \kappa_\mu \sum_{\alpha=0}^{\infty} E^{A_\nu,i_1\cdots i_\alpha} [E^{B_\mu}(L)] \rho_{B_\mu,i_1\cdots i_\alpha} \Bigg\} + \kappa_\nu E^{A_\nu}(L) - \Bigg\{ \sum_{\mu=1}^{\mu_0} \sum_{\alpha=0}^{\infty} E^{A_\nu,i_1\cdots i_\alpha} [E^{B_\mu j}(L)] \rho_{B_\mu,ji_1\cdots i_\alpha} \Bigg\} + D_j E^{A_\nu,j}(L) \Bigg).$$

Theorem 2.1 and Corollary 2.1 imply that the above equation becomes

$$E^{A_\nu}(L') = \frac{1}{n} \Bigg\{ \Bigg[\sum_{\mu=1}^{\mu_0} \kappa_\mu \sum_{\alpha=0}^{\infty} E^{A_\nu}(L)^{;B_\mu i_1\cdots i_\alpha} \rho_{B_\mu i_1\cdots i_\alpha} \Bigg] + \kappa_\nu E^{A_\nu}(L) - \Bigg[\sum_{\mu=1}^{\mu_0} \sum_{\alpha=0}^{\infty} E^{A_\nu,j}(L)^{;B_\mu i_1\cdots i_\alpha} \rho_{B_\mu,ji_1\cdots i_\alpha} \Bigg] + D_j E^{A_\nu,j}(L) + \Bigg[\sum_{\mu=1}^{\mu_0} \sum_{\alpha=1}^{\infty} \alpha E^{A_\nu}(L)^{;B_\mu i_1\cdots i_\alpha} \rho_{B_\mu i_1\cdots i_\alpha} \Bigg] \Bigg\}.$$

However,

$$D_j E^{A_\nu,j}(L) = \sum_{\mu=1}^{\mu_0} \sum_{\alpha=0}^{\infty} E^{A_\nu,j}(L)^{;B_\mu i_1\cdots i_\alpha} \rho_{B_\mu,ji_1\cdots i_\alpha}$$

and thus by Lemma 5.1 applied to $E^{A_\nu}(L)$ we find that

$$E^{A_\nu}(L') = E^{A_\nu}(L),$$

proving the theorem. □

Since for tensorial ρ_{A_μ} and scalar densities $\mathscr{L} \in \mathscr{F}(\mathscr{T}(U,x))$, $E^{A_\mu}(\mathscr{L})$ is always a tensor density,[2] the following is an obvious consequence of Theorem 5.1:

Corollary 5.1: If $\mathscr{L} \in \mathscr{F}(\mathscr{T}(U,x))$ is a scalar density, then for any $\nu = 1,\cdots,\mu_0$,

$$E^{A_\nu}(\mathscr{L}) = E^{A_\nu}(\mathscr{L}'),$$

where

$$\mathscr{L}' = \frac{1}{n} \sum_{\mu=1}^{\mu_0} [\kappa_\mu E^{A_\nu}(\mathscr{L}) \rho_{A_\mu} - E^{A_\mu,k}(\mathscr{L}) \rho_{A_\mu,k}]. \square$$

The above results indicate that the first Euler operators ($E^{A_\nu}{}^k(L)$) and the Euler–Lagrange expressions determine a Lagrangian (provided the Euler–Lagrange expressions are tensor densities). In view of the analysis of Sec. 2 [cf. Eqs. (2.10) and (2.11)], the Euler–Lagrange expressions are almost determined by the first Euler operators. Thus, the first Euler operators contain most of the information contained in the Lagrangian. In some cases the Lagrangian (and therefore the Euler–Lagrange expressions) is completely determined by the first Euler operators, as is shown in the following:

Proposition 5.1: Let $L\in\mathcal{F}(\mathcal{T}(U,x))$ and suppose for each ν, $E^{A_\nu}(L)$ is a tensor density, is homogeneous of degree m_ν in ρ_{A_μ} and homogeneous of degree $m_\mu+1$ in ρ_{A_μ} for all $\mu\neq\nu$. If $n-\sum_{\mu=1}^{\mu_0}\kappa_\mu(m_\mu+1)\neq 0$, then for all ν,

$$E^{A_\nu}(L)=E^{A_\nu}(\tilde{L}), \tag{5.3}$$

where

$$\tilde{L}:=-\left[n-\sum_{\mu=1}^{\mu_0}\kappa_\mu(m_\mu+1)\right]^{-1}\sum_{\sigma=1}^{\mu_0}E^{B_\sigma k}(L)\rho_{B_\sigma k}. \tag{5.4}$$

Proof: By Proposition 3.2 we have that

$$E^{A_\nu}\left(\frac{\kappa_\nu}{n}E^{B_\nu}(L)\rho_{B_\nu}\right)=\frac{\kappa_\nu(m_\nu+1)}{n}E^{A_\nu}(L). \tag{5.5}$$

Equation (3.3) and Corollary 2.1 imply that

$$E^{A_\nu}\left(\frac{\kappa_\nu}{n}E^{B_\mu}(L)\rho_{B_\mu}\right)=\frac{\kappa_\mu}{n}\sum_{\tau=0}^{\infty}E^{A_\nu}(L)^{;B_\mu i_1\cdots i_\tau}\rho_{B_\mu i_1\cdots i_\tau}$$

for all $\mu\neq\nu$. The homogeneity of $E^{A_\nu}(L)$ in ρ_{B_μ} then yields the equation

$$E^{A_\nu}\left(\frac{\kappa_\mu}{n}E^{B_\mu}(L)\rho_{B_\mu}\right)=\frac{\kappa_\mu(m_\mu+1)}{n}E^{A_\nu}(L). \tag{5.6}$$

Theorem 5.1 together with Eqs. (5.2), (5.5), and (5.6) imply that for all $\nu=1,\cdots,\mu_0$,

$$E^{A_\nu}(L)=\sum_{\mu=1}^{\mu_0}\left(\frac{\kappa_\mu(m_\mu+1)}{n}E^{A_\nu}(L)\right)$$

$$-\frac{1}{n}E^{A_\nu}\left[\sum_{\mu=1}^{\mu_0}E^{B_\mu k}(L)\rho_{B_\mu k}\right]$$

Equations (5.3) and (5.4) now follow from this last equation, proving the proposition. □

Given an $L\in\mathcal{F}(\mathcal{T}(U,x))$ which is homogeneous of degree $m_\nu+1$ in ρ_{A_ν} for all $\nu=1,\ldots,\mu_0$, the expressions $E^{A_\mu}(L)$ are homogeneous of degree m_μ in ρ_{A_μ} and homogeneous of degree $m_\nu+1$ in ρ_{A_ν} for all $\nu\neq\mu$. This leads to the following direct consequence of Proposition 5.1 which is similar to Theorem 6 of Ref. 2:

Corollary 5.1: If $\mathscr{L}\in\mathcal{F}(\mathcal{T}(U,x))$ is a scalar density which is homogeneous of degree $m_\mu+1$ in ρ_{A_μ} for all $\mu=1,\cdots,\mu_0$, then provided $n-\sum_{\mu=1}^{\mu_0}\kappa_\mu(m_\mu+1)\neq 0$, $E^{A_\nu}(\mathscr{L})=E^{A_\nu}(\tilde{\mathscr{L}})$ for each $\nu=1,\ldots,\mu_0$, where $\tilde{\mathscr{L}}$ is given by (5.4). □

As an application of the above analysis, consider the scalar density $\mathscr{L}=\sqrt{g}R$ in a 4-space. This scalar density is homogeneous of degree one in g_{ab} and its derivatives. Here $\kappa=2$ and therefore the Euler–Lagrange expression corresponding to $\mathscr{L}$ is the same as that corresponding to the function $\tilde{\mathscr{L}}$, where

$$\tilde{\mathscr{L}}=-\tfrac{1}{2}E^{rs,t}(\mathscr{L})g_{rs,t}.$$

By the definition of the first Euler operator we find that

$$\tilde{\mathscr{L}}=\tfrac{1}{2}g_{rs,t}(\sqrt{g}R)^{;rs,t}-g_{rs,t}(\sqrt{g}R)^{;rs,tu}{}_{,u}$$

and thus

$$\tilde{\mathscr{L}}=\tfrac{1}{2}g_{rs,t}(\sqrt{g}R)^{;rs,t}+(\sqrt{g}R)^{;rs,tu}g_{rs,tu}-[(\sqrt{g}R)^{;rs,tu}_{\ \ g_{rs,t}}]_{,u}.$$

Carrying out the indicated operations in the above equation we deduce that[22]

$$\tilde{\mathscr{L}}=\sqrt{g}R-(\sqrt{g}R^{;rs,tu}g_{rs,t})_{,u}$$

or explicitly,

$$\tilde{\mathscr{L}}=\sqrt{g}\left(g^{jr}\begin{Bmatrix}k\\l\ r\end{Bmatrix}\begin{Bmatrix}l\\j\ k\end{Bmatrix}-g^{mk}\begin{Bmatrix}r\\k\ m\end{Bmatrix}\begin{Bmatrix}l\\l\ r\end{Bmatrix}\right)$$

which is the usual first order Lagrangian in general relativity.[23]

ACKNOWLEDGMENTS

I would like to express my appreciation to G.W. Horndeski for several discussions on the topics of this paper. I would also like to express my gratitude to H.F. Ahner, I.M. Anderson, and D. Lovelock for informative conservations. This work was supported in part by the National Research Council of Canada.

[1]These operators have arisen previously in a number of applications without a specific definition. See, e.g., Refs. 4, 5, 9, 10, and 11.
[2]D. Lovelock, J. Austral. Math. Soc. **14**, 482 (1972).
[3]Round brackets about a collection of small Latin indices (all covariant or all contravariant) denotes total symmetrization of the contained indices. Vertical bars inside a pair of round brackets denote the exclusion of the (small Latin) indices they contain from the symmetrization process.
[4]I.M. Anderson, "Tensorial Euler–Lagrange Expressions and Conservation Laws," Aeq. Math. (to be published).
[5]I.M. Anderson, "Mathematical Foundations of the Einstein Field Equations," Ph.D. thesis, University of Arizona, 1976.
[6]S.J. Aldersley, Phys. Rev. D. **15**, 3507 (1977).
[7]G.W. Horndeski, J. Math. Phys. **17**, 1980 (1976).
[8]G.W. Horndeski, "Gauge Invariance and Charge Conservation," Tensor (to appear).
[9]G.W. Horndeski, Tensor **28**, 303 (1974).
[10]R.W. Atherton and G.M. Homsy, Stud. Appl. Math. **54**, 31 (1975).
[11]H.F. Ahner and A.E. Moose, J. Math. Phys. **18**, 1367 (1977). Also see comment on "Covariant inverse problem in the calculus of variations," J. Math. Phys. **19**, 1799 (1978).
[12]Equivalent here means that any solution of the field equations $T^A=0$ transforms into a solution of the equations $\tilde{T}^A=0$, and conversely. Atherton and Homsy (Ref. 6) give an example of this situation.

[13]J. Douglas, Trans. Am. Math. Soc. **50**, 71 (1941).
[14]S.J. Aldersley, Phys. Rev. D. **15**, 370 (1977).
[15]T.Y. Thomas, *The Differential Invariants of Generalized Spaces* (Cambridge U.P., Cambridge, 1934), pp. 109 and 130.
[16]Use is made of H. Weyl, *The Classical Groups* (Princeton U.P., Princeton, 1939) (page 52 noting page 65) to construct $\zeta^{abcdrstu}$ from which the formula for $\mathscr{L}_1$ follows.
[17]D. Lovelock and H. Rund, *Tensors, Differential Forms, and Variational Principles* (Wiley, New York, 1975), p. 322.
[18]Use is made of Eq. (3.43) on page 312 of Ref. 17 in this calculation, noting the notational differences.
[20]S.J. Aldersley and G.W. Horndeski, Conformally Invariant Tensorial Concomitants of a Pseudo-Riemannian Metric (Submitted for publication).
[21]G.W. Horndeski, Utilitas Mathematica **9**, 3 (1976).
[22]This calculation can be facilitated by noting that $\sqrt{g}R$ is of the form $\zeta^{abcrst}g_{ab,c}g_{rs,t}+\zeta^{abcd}g_{ab,cd}$, where the $\zeta^{\cdots}$'s are tensor density concomitants of the metric.
[23]R. Adler, M. Bazin, and M. Schiffer, *Introduction to General Relativity* (McGraw-Hill, New York, 1965), p. 310.

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.*