Path integral for coherent states of the dynamical group SU(1,1)

Christopher C. Gerry
Division of Science and Mathematics, University of Minnesota at Morris, Morris, Minnesota 56267

Steven Silverman
Department of Physics, Seton Hall University, South Orange, New Jersey 07079

(Received 8 December 1981; accepted for publication 4 June 1982)

Path integrals over coherent states of the dynamical group (noninvariance group) SU(1,1) are constructed. From the continuous limit the relevant classical dynamics is extracted and is shown to take place in a curved phase space of the form of a Lobachevskii plane. Applications are made to the harmonic oscillator, a model of superfluid helium, the Morse oscillator, and the hydrogen atom. It is shown that when SU(1,1) is the relevant dynamical group the motion will appear oscillator-like on the Lobachevskii plane.

PACS numbers: 02.20. + b, 02.30. + g, 03.65.Fd

I. INTRODUCTION

Coherent states were originally constructed to follow the classical motion of a one dimensional harmonic oscillator.[1] These states, $\{|z\rangle\}$, can be constructed in three ways: (i) as states which minimize the uncertainty relation $\Delta x \Delta p \geqslant \hbar/2$, (ii) as states which diagonalize the boson annihilation operator $\hat{a}$, i.e., $\hat{a}|z\rangle = z|z\rangle$, and (iii) from the application of the displacement operator $D(z) = e^{z\hat{a}^+ - z^*\hat{a}}$ (an element of the Heisenberg–Weyl group) on the ground state $|0\rangle$

$$|z\rangle = D(z)|0\rangle.$$

As it happens, all three definitions turn out to be equivalent, yielding states which do follow the classical motion of the oscillator.

The notion of generalized coherent states arises in connection with the attempt to construct quasi-classical states for dynamical systems other than the oscillator. Considerable effort by Nieto and collaborators[2] has gone into the construction of analytic coherent states for generalized potentials. In their work, the approach is to take the canonical variables (x,p) of the classical problem and find a map to a set of variables (X,P) which makes the Hamiltonian *look* like a harmonic oscillator. The quantum analoges of the variables are then used to construct a generalized uncertainty relation, the generalized coherent states being those states which minimize this relation.

An alternate method of creating generalized coherent states is associated with irreducible representations of Lie groups. Spin coherent states based on SU(2) have been discussed by Radcliffe and others[3] while Perelomov[4] has given a general method for constructing coherent states for an arbitrary Lie group. This involves defining a unitary operator with elements of the Lie algebra in analogy to the displacement operator for the Heisenberg–Weyl group, and letting this operator act on some conveniently chosen vector out of the irreducible representation. Such systems of generalized coherent states might be expected to be relevant for a system that has an associated dynamical group. By this we mean a group whose Lie algebra is a spectrum generating algebra (SGA). That is, the energy eigenvalue problem can be written in terms of the group generators and the irreducible representations yield the spectrum. For instance, it is well known that the group SU(1,1) ~ SO(2,1) is an SGA for such systems as the harmonic oscillator, the hydrogen atom, the Morse potential,[5] and some many-body problems.[6]

Functional integrals in the ordinary coherent state representation are well known to be useful for many-body systems of bosons.[7] Recently path integrals in the SU(2) coherent state representations have been discussed.[8] This type of functional integral yields a corresponding classical dynamics whose phase space is curved, in fact, the two-dimensional sphere S^2. Starting from the quantum mechanical Hamiltonian of the rigid rotator the corresponding canonical equations of motion reproduce the Euler equations for a rigid body.

In this paper we propose to construct the path integral in the SU(1,1) coherent state [SU(1,1)CS] representation. In doing so we are first faced with a choice of two nonequivalent definitions of SU(1,1)CS. There is, of course, the definition of Perelomov already alluded to, but there is also a definition given by Barut and Girardello.[9] However, for these latter states, technical difficulties, such as the fact that the overlap of two states and matrix elements of operators cannot be expressed as closed forms, prohibits us from readily forming a useful path integral. We thus use the Perelomov definition.

The actual construction of the propagator for SU(1,1)CS closely follows that for SU(2)CS.[8] There are some differences, however. In general the physical states of the system are not identical to the basis vectors of the group representation. They are related though by rotation about an axis in the abstract space of the group. For systems where this transformation is unitary, appropriate physical coherent states are defined and the propagator is given in terms of these states. The model of superfluid helium II discussed in Sec. IV is an example of this. When the transformation is not unitary, as for the hydrogen atom (Sec. V B), we deal with a functional integral expression of a resolvent whose poles give the bound state energies. We obtain a propagator for the coherent group states whose path integral expression contains not the usual Hamiltonian but an auxiliary zero-energy "Hamiltonian". We show that the classical mechanics of the SU(1,1)CS path integral is again a curved phase space, this time a Lobachevskii plane. For systems where the Hamiltonian (or an auxiliary one) is linear in SU(1,1) generators, the

 0022-2488/82/111995-09$02.50 © 1982 American Institute of Physics

corresponding classical motion will look oscillator-like on the Lobachevskii plane.

The plan of the paper is as follows: In Sec. II we briefly review SU(1,1) and the relevant irreducible representations. Perelomov's definition of SU(1,1)CS is presented and various matrix elements of SU(1,1) generators are given. We can immediately use the SU(1,1)CS to determine the propagator for a one-dimensional harmonic oscillator (whose physical states are composed of two irreducible representations of SU(1,1)). We follow through the calculation of the partition function to demonstrate that the total propagator must include the parts from all representations needed to cover the physical states. [The sum occurs since SU(1,1) is not a symmetry group in contrast to SU(2)[10]]. In Sec. III we derive a path integral expression of the propagator in SU(1,1)CS and derive the relevant classical limit. In Sec. IV we apply the formalism to a superfluid in the Bogoliubov approximation, where SU(1,1) has been shown to be the relevant dynamical group. In Sec. V applications are made to the Morse oscillator and the hydrogen atom. A brief discussion concludes the paper.

II. SU(1,1) LIE ALGEBRA AND DEFINITION OF COHERENT STATES

The Lie algebra of SU(1,1) consists of three generations K_1, K_2, and K_0 which satisfy the commutation relations

$$[K_1,K_2] = -iK_0, \tag{2.1a}$$

$$[K_2,K_0] = iK_1, \tag{2.1b}$$

$$[K_0,K_1] = iK_2. \tag{2.1c}$$

Defining the raising and lowering operators

$$K_\pm = K_1 \pm iK_2,$$

the above algebra becomes

$$[K_0,K_\pm] = \pm K_\pm, \tag{2.2a}$$

$$[K_-,K_+] = 2K_0. \tag{2.2b}$$

The Casimir operator is

$$\hat{C}_2 = K_0^2 - K_1^2 - K_2^2 = K_0^2 - \tfrac{1}{2}(K_+K_- + K_-K_+). \tag{2.3}$$

By Schur's lemma we have

$$\hat{C}_2 = k(k-1)\hat{I}, \tag{2.4}$$

where $\hat{I}$ is the identity operator. There are four distinct types of unitary irreducible representations, as has been discussed by Barut and Fronsdal.[11] We are primarily interested in the representation known as $\mathcal{D}^+(k)$, the positive discrete series. We denote the basis vectors as $|k,n\rangle$. For the series $\mathcal{D}^+(k)$, K_0 (the generator of a compact subgroup) is diagonal such that

$$K_0|k,n\rangle = (k+n)|k,n\rangle, \tag{2.5}$$

where $n = 0,1,2,\ldots$ and $k > 0$.

A realization of the SU(1,1) algebra can be given in terms of the creation and annihilation operators $\hat{a}^+,\hat{a}$ satisfying the canonical algebra $[\hat{a},\hat{a}^+] = 1$.
We have

$$K_+ = \tfrac{1}{2}(\hat{a}^+)^2,$$

$$K_- = \tfrac{1}{2}(\hat{a})^2,$$

$$K_0 = \tfrac{1}{4}(\hat{a}\hat{a}^+ + \hat{a}^+\hat{a}). \tag{2.6}$$

Using this realization and the eigenstate of the harmonic oscillator, $\hat{C}_2$ becomes

$$\hat{C}_2 = -\tfrac{3}{16}\hat{I}$$

so that $k = 1/4, 3/4$. Using the commutation relations for $\hat{a}$ and $\hat{a}^+$ it is easy to see that we can write Eq. (2.6c) as

$$K_0 = \tfrac{1}{2}(\hat{a}^+\hat{a} + \tfrac{1}{2}). \tag{2.7}$$

Then the Hamiltonian for the $1-d$ harmonic oscillator can be written as

$$H = 2\hbar\omega K_0. \tag{2.8}$$

Note that the two representations of SU(1,1) (for $k = \frac{1}{4}$ and $\frac{3}{4}$) are needed to cover all the states of the oscillator. For $k = \frac{1}{4}$ we obtain the energy levels $\frac{1}{2}\hbar\omega, \frac{5}{2}\hbar\omega, \frac{9}{2}\hbar\omega$, etc., and for $k = \frac{3}{4}$ we get the levels $\frac{3}{2}\hbar\omega, \frac{7}{2}\hbar\omega, \frac{11}{2}\hbar\omega$, etc.

In the case of the oscillator the physical states are essentially identical to the group states [basis vectors of $\mathcal{D}^+(k)$ representation]. There are many examples where this is not the case. An explicit example will be given in Sec. IV where we discuss a simple model of superfluid helium. For now let us just state that the Hamiltonian can be written in the form

$$\tilde{H} = aK_0 + bK_1 + c,$$

where a, b, and c depend on parameters of the potential. The eigenvalue problem is

$$\tilde{H}|\widetilde{\psi_n}\rangle = E_n|\widetilde{\psi_n}\rangle. \tag{2.9}$$

Now perform the "tilting" transformation on Eq. (2.9), i.e., write

$$e^{-i\theta K_2}\tilde{H}e^{i\theta K_2}e^{-i\theta K_2}|\widetilde{\psi_n}\rangle = E_n e^{-i\theta K_2}|\widetilde{\psi_n}\rangle. \tag{2.10}$$

From the Baker–Hausdorf–Campbell formula we have

$$e^{-i\theta K_2}K_0\,e^{+i\theta K_2} = K_0\cosh\theta + K_1\sinh\theta, \tag{2.11a}$$

$$e^{-i\theta K_2}K_1\,e^{+i\theta K_2} = K_0\sinh\theta + K_1\cosh\theta, \tag{2.11b}$$

so that

$$e^{-i\theta K_2}\tilde{H}e^{+i\theta K_2} = K_0(a\cosh\theta + b\sinh\theta) + K_1(a\sinh\theta + b\cosh\theta) + c. \tag{2.12}$$

With the appropriate choice of θ, the coefficient of K_1 will vanish and we obtain the discrete part of the spectrum from the new eigenvalue problem

$$H|\psi_n\rangle = E_n|\psi_n\rangle, \tag{2.13}$$

where

$$H = K_0(a\cosh\theta + b\sinh\theta) + c \tag{2.14}$$

and

$$|\psi_n\rangle = e^{-i\theta K_2}|\widetilde{\psi_n}\rangle. \tag{2.15}$$

$|\widetilde{\psi_n}\rangle$ is the physical state and $|\psi_n\rangle$ is the corresponding group state. Note that here θ depends only on potential parameters so that the tilting operation is a unitary transformation. For some problems where SU(1,1) is relevant, this is not the case, e.g., the Coulomb problem (see Wulfman[12]).

We now describe the SU(1,1) coherent states as defined by Perelomov.[4] These states are obtained by operating on the fidicial vector $|\psi_0\rangle = |k,0\rangle$ with the group element

$$\exp(-i\phi K_0)\exp(-i\tau K_2)\exp(-i\psi K_0).$$

Up to a phase this element can be written as

$$D(\alpha) = \exp(\alpha K_+ - \alpha^* K_-), \tag{2.16}$$

where $\alpha = -(\tau/2)e^{-i\phi}$, in analogy to the displacement operator of the ordinary coherent states. Using the Lie algebra (2.2), Eq. (2.16) can be written

$$D(\alpha) = e^{\zeta K_+} e^{\beta K_0} e^{\gamma K_-}, \tag{2.17}$$

where $\zeta = -\tanh(\tau/2)e^{-i\phi}, \beta = 2\ln\cosh(\tau/2)$ $= \ln(1-|\zeta|^2)$, and $\gamma = -\zeta^*$. Thus the coherent states are

$$|\zeta,k\rangle = D(\alpha)|\psi_0\rangle$$

or

$$|\zeta,k\rangle = (1-|\zeta|^2)^k e^{\zeta K_+}|k,0\rangle. \tag{2.18}$$

Since

$$|k,n\rangle = \left(\frac{\Gamma(2k)}{n!\Gamma(n+2k)}\right)^{1/2}(K_+)^n|k,0\rangle, \tag{2.19}$$

we have, on expanding the exponential in Eq. (2.17),

$$|\zeta,k\rangle = (1-|\zeta|^2)^k \sum_{n=0}^{\infty}\left(\frac{\Gamma(n+2k)}{n!\Gamma(2k)}\right)^{1/2}\zeta^n|k,n\rangle. \tag{2.20}$$

The overlap of two states $|\zeta,n\rangle$ and $|\zeta' k\rangle$ is given as

$$\langle\zeta',k|\zeta,k\rangle = (1-|\zeta'|^2)^k(1-|\zeta|^2)^k(1-\zeta'^*\zeta)^{-2k}. \tag{2.21}$$

Unity is resolved as

$$I = \int d\mu_k(\zeta)|\zeta,k\rangle\langle\zeta,k|, \tag{2.22}$$

where

$$d\mu_k(\zeta) = \frac{2k-1}{\pi}\frac{d^2\zeta}{(1-|\zeta|^2)^2} \tag{2.23}$$

for $k \neq \frac{1}{2}$ and, for $k = \frac{1}{2}$,[13]

$$du_{1/2}(\zeta) = \frac{1}{\pi}\frac{d^2\zeta}{(1-|\zeta|^2)^2}. \tag{2.24}$$

In the event that physical and group states are related in the manner of Eq. (2.15) it would seem prudent to define a corresponding physical coherent state $|\zeta,k\rangle$. We take this simply as

$$|\widetilde{\zeta,k}\rangle = e^{+i\theta K_2}|\zeta,k\rangle. \tag{2.25}$$

The resolution of unity and the overlap are unchanged by the replacement $|\zeta,k\rangle \to |\widetilde{\zeta,k}\rangle$. However, matrix elements related to the physical quantities must be calculated with $|\widetilde{\zeta,k}\rangle$ and in general are not the same as those calculated with $|\zeta,k\rangle$.

For later use we present matrix elements of K_+, K_-, and K_0. These are

$$\langle\zeta'|K_+|\zeta\rangle/\langle\zeta'|\zeta\rangle = \frac{2k\zeta'^*}{(1-\zeta'^*\zeta)}, \tag{2.26a}$$

$$\langle\zeta'|K_-|\zeta\rangle/\langle\zeta'|\zeta\rangle = \frac{2k\zeta}{(1-\zeta'^*\zeta)}, \tag{2.26b}$$

$$\langle\zeta'|K_0|\zeta\rangle/\langle\zeta'|\zeta\rangle = k(1+\zeta'^*\zeta)/(1-\zeta'^*\zeta). \tag{2.26c}$$

These are easily derived using Eqs. (2.11), (2.18), and (2.21).

We can define a propagator in terms of SU(1,1)CS as

$$K(\zeta',\zeta;T) = \sum_k \langle\widetilde{\zeta,k}|\exp\left[-\frac{i}{\hbar}\tilde{H}T\right]|\widetilde{\zeta,k}\rangle \tag{2.27}$$

or

$$K(\zeta',\zeta;T) = \sum_k \langle\zeta',k|\exp\left[-\frac{i}{\hbar}HT\right]|\zeta,k\rangle, \tag{2.28}$$

where $H = e^{i\theta K_2}\tilde{H}e^{+i\theta K_2}$. For the harmonic oscillator $\theta = 0$ and using Eqs. (2.8), (2.20), and (2.21) it is easy to show that

$$\begin{aligned} K(\zeta',\zeta;T) &= \sum_k \langle\zeta',k|\exp(-i2\omega K_0 T)|\zeta,k\rangle \\ &= \sum_{k=1/4}^{3/4} e^{-2i\omega Tk}\langle\zeta'|\zeta e^{-2i\omega T}\rangle \\ &= \sum_{k=1/4}^{3/4} e^{-2i\omega Tk}(1-|\zeta'|^2)^k(1-|\zeta|^2)^k \\ &\quad\times(1-\zeta'^*\zeta e^{-2i\omega T})^{-2k}, \end{aligned} \tag{2.29}$$

where we have summed over the representations required to cover the states. As a check we can determine the partition function

$$Z = \mathrm{Tr}K(\zeta',\zeta,-i\hbar\beta) = \sum_{k=1/4}^{3/4} Z_k, \tag{2.30}$$

where

$$\begin{aligned} Z_k &= e^{-2k\hbar\omega\beta}\int\frac{|2k-1|}{\pi}d^2\zeta(1-|\zeta|^2)^{|2k-1|-1} \\ &\quad\times(1-e^{-2k\omega\beta|\zeta|^2})^{-|2k-1|-1} \end{aligned} \tag{2.31}$$

(see Ref. 13).

We obtain

$$Z_{1/4} = \frac{e^{-\hbar\omega\beta/2}}{1-e^{-2\beta\hbar\omega}}, \tag{2.32}$$

$$Z_{3/4} = \frac{e^{-3\hbar\omega\beta/2}}{1-e^{-2\beta\hbar\omega}}, \tag{2.33}$$

so that

$$Z = Z_{1/4} + Z_{3/4} = \frac{e^{-\hbar\omega\beta/2}}{1-e^{-\beta\hbar\omega}}, \tag{2.34}$$

which is, of course, the correct result.

III. PATH INTEGRAL AND CLASSICAL DYNAMICS

We now consider the path integral expression for the propagator given by Eq. (2.27). Defining ϵ as T/N and using Eq. (2.22) we write in the usual manner

$$K(\zeta',\zeta;T) = \sum_k \langle\widetilde{\zeta',k}|\exp\left[-\frac{iT}{\hbar}\tilde{H}\right]|\widetilde{\zeta,k}\rangle \tag{3.1}$$

$$\begin{aligned} &= \sum_k \lim_{N\to\infty}\int\cdots\int\langle\widetilde{\zeta_N,k}|e^{-(i/\hbar)\epsilon\tilde{H}}|\widetilde{\zeta_{N-1},k}\rangle \\ &\quad\times\langle\widetilde{\zeta_{N-1},k}|e^{-(i/\hbar)\epsilon\tilde{H}}|\widetilde{\zeta_{N-2},k}\rangle \\ &\quad\cdots\langle\widetilde{\zeta,k}|e^{-(i/\hbar)\epsilon\tilde{H}}|\widetilde{\zeta_0,k}\rangle\prod_{j=1}^{N-1}du_k(\zeta_j), \end{aligned} \tag{3.2}$$

or

$$K(\zeta',\zeta;T)=\lim_{N\to\infty}\sum_k\int\cdots\int\prod_{j=1}^{N}\langle\widetilde{\zeta_j,k}\,|e^{-(i/\hbar)\epsilon\tilde{H}}|\widetilde{\zeta_{j-1},k}\rangle \times\prod_{j=1}^{N-1}du_k(\zeta_j), \tag{3.3}$$

where $\zeta_0=\zeta$ and $\zeta_N=\zeta'$. For small ϵ we have

$$\langle\widetilde{\zeta_j k}\,|e^{-(i/\hbar)\epsilon\tilde{H}}|\widetilde{\zeta_{j-1}k}\rangle\cong\langle\zeta_j k\,|e^{-i\theta K_2}(1-\frac{i}{\hbar}\epsilon\tilde{H})e^{+i\theta K_2} \times|\zeta_{j-1}k\rangle=\langle\zeta_j k\,|(1-\frac{i}{\hbar}\epsilon H)|\zeta_{j-1}k\rangle. \tag{3.4}$$

Then with

$$\langle\zeta_j,k\,|(1-\frac{i}{\hbar}\epsilon H)|\zeta_{j-1},k\rangle =\langle\zeta_j,k\,|\zeta_{j-1},k\rangle\Big[1-\frac{i}{\hbar}\epsilon\mathscr{H}_k(\zeta_j,\zeta_{j-1})\Big] \approx\langle\zeta_j,k\,|\zeta_{j-1},k\rangle\exp\Big[-\frac{i}{\hbar}\epsilon\mathscr{H}_k(\zeta_j,\zeta_{j-1})\Big], \tag{3.5}$$

where

$$\mathscr{H}_k(\zeta_j,\zeta_{j-1})=\langle\zeta_j,k\,|H\,|\zeta_{j-1},k\rangle/\langle\zeta_j,k\,|\zeta_{j-1},k\rangle, \tag{3.6}$$

Eq. (3.3) becomes

$$K(\zeta',\zeta;T)=\lim_{N\to\infty}\sum_k\int\cdots\int\prod_{j=1}^{N}\exp\{\ln\langle\zeta_j,k\,|\zeta_{j-1},k\rangle -\frac{i}{\hbar}\epsilon\mathscr{H}_k(\zeta_j,\zeta_{j-1})\}\prod_{j=1}^{N-1}d\mu_k(\zeta_j). \tag{3.7}$$

With $\zeta_{j-1}=\zeta_j-\Delta\zeta_j$ it can be shown, using (2.21), that to order $\Delta\zeta_j$

$$\ln\langle\zeta_j,k\,|\zeta_{j-1},k\rangle=\frac{k}{(1-|\zeta_j|^2)}(\zeta_j^*\Delta\zeta_j-\zeta_j\Delta\zeta_j^*).$$

Thus we have

$$K(\zeta',\zeta;T)=\lim_{N\to\infty}\sum_k\int\cdots\int\prod_{j=1}^{N}\exp\Big\{\frac{\epsilon i}{\hbar}\Big[\frac{i\hbar k}{(1-|\zeta_j|^2)} \times\Big(\zeta_j\frac{\zeta_\alpha^*}{\epsilon}-\zeta_j^*\frac{\Delta\zeta_\alpha}{\epsilon}\Big) -H_k(\zeta_j;\zeta_{j-1})\Big]\Big\}\prod_{j=1}^{N-1}d\mu_k(\zeta_j). \tag{3.8}$$

In the continuous limit this may be written as

$$K(\zeta',\zeta;T)=\sum_k\int\mathscr{D}\mu_k(\zeta)\exp\Big\{\frac{i}{\hbar}\int_0^T\mathscr{L}_k(\zeta,\dot\zeta,\zeta^*,\dot\zeta^*)dt\Big\}, \tag{3.9}$$

where

$$\mathscr{L}(\zeta,\dot\zeta,\zeta^*,\dot\zeta^*)=\frac{i\hbar k}{(1-|\zeta|^2)}[\zeta(t)\dot\zeta^*(t)-\zeta^*(t)\dot\zeta(t)] -\mathscr{H}_k(\zeta,\zeta^*). \tag{3.10}$$

With $S=\int_T^0\mathscr{L}_k dt$ and setting $\delta S=0$, we obtain the Euler–Lagrange equation

$$\frac{d}{dt}\Big(\frac{\partial\mathscr{L}_k}{\partial\dot\zeta}\Big)-\frac{\partial\mathscr{L}_k}{\partial\zeta}=0 \tag{3.11}$$

and its complex conjugate. Using (3.10) this can be shown to yield the analog of Hamilton's equations

$$\dot\zeta=i\frac{(1-|\zeta|^2)}{2\hbar k}\frac{\partial\mathscr{H}_k}{\partial\zeta^*},$$
$$\dot\zeta^*=-i\frac{(1-|\zeta|^2)}{2\hbar k}\frac{\partial\mathscr{H}_k}{\partial\zeta}. \tag{3.12}$$

Poisson brackets can be defined as

$$\{A,B\}=\frac{(1-|\zeta|^2)}{2i\hbar k}\Big(\frac{\partial A}{\partial\zeta^*}\frac{\partial B}{\partial\zeta}-\frac{\partial A}{\partial\zeta}\frac{\partial B}{\partial\zeta^*}\Big) \tag{3.13}$$

and Eqs. (3.12) can be written as

$$\dot\zeta=\{\zeta,\mathscr{H}_k\},$$
$$\dot\zeta^*=\{\zeta^*,\mathscr{H}_k\}. \tag{3.14}$$

Comparing Eq. (3.13) with Berezin[14] we see that we are dealing with a curved phase space in the form of a Lobachevskii plane. In terms of the group parameters τ and ϕ,

$$\{A,B\}=\frac{1}{\hbar k\sinh\tau}\Big\{\frac{\partial A}{\partial\tau}\frac{\partial B}{\partial\phi}-\frac{\partial A}{\partial\phi}\frac{\partial B}{\partial\tau}\Big\}. \tag{3.15}$$

For the harmonic oscillator we have $H=2\hbar\omega K_0$ so that

$$\mathscr{H}(\zeta,\zeta^*)=2\hbar\omega k\cosh\tau \tag{3.16}$$

and the equations of motion are

$$\dot\tau=\{\tau,\mathscr{H}(\tau)\}=0, \tag{3.17}$$
$$\dot\phi=\{\phi,\mathscr{H}(\tau)\}=-2\omega. \tag{3.18}$$

The solutions to (3.17) are simply $\tau=$ const, $\phi=-2\omega t+\phi_0$, which shows that for the oscillator the parameters τ and ϕ are closely related to action and angle variables in this curved phase space.

Now since the tilted form of $\tilde H$ will by design depend linearly only on the operator K_0 it is easy to see that for almost any system where SU(1,1) is the SGA, that system will classically appear as some sort of harmonic oscillator. Explicit examples will be presented in the next two sections.

IV. APPLICATION TO A MODEL OF SUPERFLUID HELIUM

As an application of the formalism developed in the previous section, let us consider the SU(1,1) treatment of a superfluid Bose system as discussed by Solomon.[6] The Hamiltonian for this system can be written as $\tilde H=\Sigma_p\tilde H_p$, where

$$\tilde H_p=\epsilon_p\hat a_p^+\hat a_p+\tfrac12\sum_{q,r}V_p\hat a_{q+p}^+\hat a_{r-p}^+\hat a_q\hat a_r, \tag{4.1}$$

where $\epsilon_p=p^2/2m$. Limiting ourselves to the case where there are only three states such that p,q,r take on values $(-1,0,+1)$ and where $\epsilon_\pm=\epsilon,\epsilon_0=0,V_\pm=V,V_0=0$, along with the assumption of a macroscopically occupied ground state $(\hat a_0,\hat a_0^+\sim N_0^{1/2})$, we have the reduced Hamiltonian

$$\tilde H_R=(\epsilon+N_0V)(\hat a_+^+\hat a_++\hat a_-^+\hat a_-) +N_0V(\hat a_+^+\hat a_-^++\hat a_+\hat a_-). \tag{4.2}$$

This Hamiltonian can be cast into a form which is linear in SU(1,1) generators by noting that

$$K_1=-\tfrac12(\hat a_+^+\hat a_-^++\hat a_+\hat a_-),$$
$$K_2=\tfrac12 i(\hat a_+^+\hat a_-^+-\hat a_+\hat a_-),$$

and

$$K_0 = (\hat{a}_+^+ \hat{a}_+ + \hat{a}_-^+ \hat{a}_- + 1) \tag{4.3}$$

provide a realization of SU(1,1). Thus

$$\tilde{H}_R = 2N_0V(\gamma K_0 - K_1 - \tfrac{1}{2}\gamma), \tag{4.4}$$

where $\gamma = 1 + \epsilon/N_0V$. Discrete energy eigenvalues are found via diagonalization through the tilting operation

$$\begin{aligned} H_R &= e^{-i\theta K_2}\tilde{H}_R e^{i\theta K_2} \\ &= 2N_0V[K_0(\gamma\cosh\theta - \sinh\theta) \\ &\quad - K_1(\cosh\theta - \gamma\sinh\theta) - \tfrac{1}{2}\gamma]. \end{aligned} \tag{4.5}$$

If $V > 0$ then $\gamma > 1$, so choosing $\theta = \coth^{-1}\gamma$ we obtain

$$H = 2N_0V(\operatorname{csch}\theta K_0 - \tfrac{1}{2}\gamma). \tag{4.6}$$

(This is the group theoretical version of the Bogoliubov transformation.) The Casimir operator for this system is

$$C_2 = K_0^2 - K_1^2 - K_2^2 = \tfrac{1}{4}(\Delta^2 - 1), \tag{4.7}$$

where $\Delta = \hat{a}_+^+ \hat{a}_+ - \hat{a}_-^+ \hat{a}_-$ so that $k = \frac{1}{2} + \frac{1}{2}|\Delta|$.

Since K_0 has eigenvalues $(n + \frac{1}{2} + \frac{1}{2}|\Delta|)$ the energy spectrum is

$$E_n = (2n + 1 + |\Delta|)E - N_0V - \epsilon, \tag{4.8}$$

where

$$E = (2\epsilon N_0V + \epsilon^2)^{1/2}.$$

The propagator for this system can be written as a sum over all possible $|\Delta|$.

$$K(\zeta',\zeta;T) = \sum_{|\Delta|} K_{|\Delta|}(\zeta',\zeta;T), \tag{4.9}$$

where

$$\begin{aligned} K_{|\Delta|}(\zeta',\zeta;T) &= \langle \widetilde{\zeta',k_p} \left| \exp\left[-\frac{iT}{\hbar}\tilde{H}_R \right] \right| \widetilde{\zeta,k_p} \rangle \\ &= \lim_{N\to\infty} \int\cdots\int \prod_{j=1}^{N-1} d\mu_{|\Delta|}(\zeta_j) \prod_{j=1}^{N} \\ &\quad \times \exp\left\{ \frac{\epsilon i}{\hbar}\left[\frac{i\hbar k}{(1-|\zeta_j|^2)}\left(\zeta_j \frac{\Delta\zeta_j^*}{\epsilon} - \zeta_j^* \frac{\Delta\zeta_j}{\epsilon} \right) \right.\right. \\ &\quad - 2N_0V[k\operatorname{csch}\theta(1 + \zeta_j^*\zeta_{j-1})/ \\ &\quad \left.\left. (1 - \zeta_j^*\zeta_{j-1}) - \tfrac{1}{2}\gamma] \right\}. \end{aligned} \tag{4.10}$$

Thus we see that on the Lobachevskii plane the classical motion will appear like a harmonic oscillator with frequency $\omega = N_0V\operatorname{csch}\theta/\hbar$.

The results here can easily be generalized to the multilevel system where the relevant dynamical group is $\Pi_p \otimes \mathrm{SU}(1,1)_p$. In this case the $\mathrm{SU}(1,1)_p$ generators are realized as

$$\begin{aligned} K_1^{(p)} &= -\tfrac{1}{2}(\hat{a}_p^+ \hat{a}_{-p}^+ + \hat{a}_p \hat{a}_{-p}), \\ K_2^{(p)} &= \tfrac{1}{2}i(\hat{a}_p^+ \hat{a}_{-p}^+ - \hat{a}_p \hat{a}_{-p}), \\ K_0^{(p)} &= \tfrac{1}{2}(\hat{a}_p^+ \hat{a}_p + \hat{a}_{-p}^+ \hat{a}_{-p} + 1), \end{aligned} \tag{4.11}$$

so that the Hamiltonian is now (to second order in $\hat{a}_p$)

$$\tilde{H} = \tfrac{1}{2}N^2V_0 + \sum_p NV_p(\gamma_p K_0^{(p)} - K_1^{(p)} - \tfrac{1}{2}\gamma_p), \tag{4.12}$$

where $\gamma_p = 1 + \epsilon_p/NV_p$ and where

$$\tilde{N} = N_0 + \sum_{p>0}(\hat{a}_p^+ \hat{a}_p + \hat{a}_{-p}^+ \hat{a}_{-p}) \tag{4.13}$$

is the total number of particles. The group states for this Hamiltonian will be given as the representation $\Pi_p \otimes \mathscr{D}^+(k_p)$, where $k_p = \frac{1}{2} + \frac{1}{2}|\Delta|_p$ and where $\Delta_p = \hat{a}_p^+ \hat{a}_p - \hat{a}_{-p}^+ \hat{a}_{-p}$. The physical SU(1,1)CS are

$$|\widetilde{\{\zeta\}}\rangle = \prod_p |\widetilde{\zeta,k_p}\rangle = \prod_p e^{i\theta_p K_2^{(p)}} |\zeta,k_p\rangle. \tag{4.14}$$

The propagator can be written as a product of sums of the form of Eq. (4.9)

$$K(\zeta',\zeta;T) = e^{-i(NV/2)T} \prod_p \sum_{|\Delta_p|} K_{|\Delta_p|}(\zeta',\zeta;T), \tag{4.15}$$

where

$$\begin{aligned} K_{|\Delta_p|}(\zeta,'\zeta;T) &= \int \mathscr{D}\mu_{|\Delta_p|}(\zeta) \\ &\quad \times \exp\left\{ \frac{i}{\hbar}\int_0^T dt \left[\frac{i\hbar k}{(1-|\zeta|^2)} \right.\right. \\ &\quad \times (\zeta\dot{\zeta}^* - \zeta^*\dot{\zeta}) - 2NV_p[k_p \operatorname{csch}\theta_p \\ &\quad \left.\left. \times (1 + |\zeta|^2)/(1 - |\zeta|^2) - \tfrac{1}{2}\gamma_p] \right] \right\}. \end{aligned} \tag{4.16}$$

The grand canonical partition function (GCPF) can be evaluated quite easily in the SU(1,1)CS formulation. For this we need the number operator in terms of the SU(1,1) generators. From Eqs. (4.13) and (4.11) we have

$$N = N_0 + \sum_{p>0}(2K_0^{(p)} - 1). \tag{4.17}$$

This is diagonal in the group states $|k,n\rangle$ having eigenvalues $N_0 + (2(k_p + n) - 1 = N_0 + 2n + |\Delta|_p$. The corresponding number operator in the physical basis $|\widetilde{k,n}\rangle$ must therefore be

$$\tilde{N} = N_0 + \sum_{p>0}(2e^{i\theta_p K_2^{(p)}} K_0^{(p)} e^{-i\theta_p K_2^{(p)}} - 1). \tag{4.18}$$

Thus the GCPF can be given as

$$Z = \mathrm{Tr}K'(\zeta',\zeta; -i\hbar\beta),$$

where

$$\begin{aligned} K'(\zeta',\zeta; -i\hbar\beta) &= \langle \widetilde{\{\zeta'\}} | \exp\left[\frac{-iT}{\hbar}(\tilde{H} - \mu\tilde{N}) \right] | \widetilde{\{\zeta\}} \rangle |_{T=-i\hbar\beta} \\ &= e^{-\frac{i}{2\hbar}NV_0T} e^{\frac{i}{\hbar}\mu N_0 T} \\ &\quad \times \prod_p \sum_{|\Delta_p|} K'_{|\Delta_p|}(\zeta',\zeta;T)\Big|_{T=-i\hbar\beta} \end{aligned} \tag{4.19}$$

and where

$$\begin{aligned} K'_{|\Delta_p|}(\zeta',\zeta;T) \\ = \langle \widetilde{\zeta',k_p} | \exp\left[-\frac{iT}{\hbar}(\tilde{H}'_p - \mu\tilde{N}_p) \right] | \widetilde{\zeta,k_p} \rangle. \end{aligned}$$

For $N_p = 2K_0^{(p)} - 1$ and $\tilde{H}'_p = NV_p(\gamma_p K_0^{(p)} - K_1^{(p)} - \frac{1}{2}\gamma_p)$ we get

$$K'_{|\Delta_p|}(\zeta',\zeta;T) = \exp\left[\frac{i}{\hbar}T\left(\frac{\gamma_p}{2} - \mu\right)\right]\langle\zeta',k_p| \times \exp\left\{-\frac{iT}{\hbar}[2NV_p \operatorname{csch}\theta_p - 2\mu]K_0^{(p)}\right\}|\zeta,k_p\rangle. \tag{4.20}$$

Using Eqs. (2.20) and (2.21) we obtain an exact expression for $K'_{|\Delta_p|}$, namely

$$K'_{|\Delta_p|}(\zeta',\zeta;T) = \exp\left[\frac{i}{\hbar}T\left(\frac{\gamma_p}{2} - \mu\right)\right]\exp\left(\frac{i}{\hbar}\alpha_p k_p T\right) \times (1 - |\zeta'|^2)^{k_p}(1 - |\zeta|^2)^{k_p} \times \left(1 - \zeta'^*\zeta\exp\left(-\frac{iT}{\hbar}\alpha_p\right)\right)^{-2k_p}, \tag{4.21}$$

where $\alpha_p = 2NV_p \operatorname{csch}\theta_p - 2\mu$. Letting $T = -i\hbar\beta$ and taking the trace we have

$$Z_{|\Delta_p|} = \exp\beta\left(\frac{\gamma_p}{2} - \mu - \alpha_p k_p\right)\int d\mu_{kp}(\zeta)(1 - |\zeta|^2)^{2k_p} \times (1 - e^{-\beta\alpha_p}|\zeta|^2)^{-2k_p}, \tag{4.22}$$

which can be integrated to yield

$$Z_{|\Delta_p|} = e^{\beta(\gamma_p/2 - \mu)}\frac{e^{-\alpha_p k_p \beta}}{1 - e^{-\beta\alpha_p}}. \tag{4.23}$$

So finally, the GCPF has the following decomposition in terms of SU(1,1)CS:

$$Z = e^{-\beta(NV_0/2 - \mu N_0)}\prod_p (e^{-\beta\mu})\sum_{|\Delta_p|}\frac{e^{\beta(\gamma_p - \alpha_p(\frac{1}{2} + \frac{1}{2}|\Delta_p|))}}{1 - e^{-\beta\alpha_p}}. \tag{4.24}$$

V. APPLICATION TO THE MORSE OSCILLATOR AND THE HYDROGEN ATOM

A. Morse oscillator

We consider here the problem of a particle bound by the Morse potential,[15]

$$V(q) = D\{\exp[-2a(q - q_0)] - 2b\exp[-a(q - q_0)]\}. \tag{5.1}$$

For the S states ($\vec{L} = 0$) the Hamiltonian can almost, but not quite, be written in terms of SU(1,1) operators. The algebra in this case can be realized as[16]

$$K_0 = \frac{1}{a^2 h(q)}[\vec{p}^2 - 2ME] + \tfrac{1}{4}h(q), \tag{5.2a}$$

$$K_1 = \frac{1}{a^2 h(q)}[\vec{p}^2 - 2ME] - \tfrac{1}{4}h(q), \tag{5.2b}$$

$$K_2 = \frac{i}{aq}(i\vec{q}\cdot\vec{p} + 1) - \frac{i}{2}, \tag{5.2c}$$

where

$$h(q) = \frac{(8mD)^{1/2}}{a}\exp[-a(q - q_0)]. \tag{5.3}$$

The Hamiltonian is thus

$$\tilde{H} = \frac{a^2 h(q)}{2m}\left(K_0 - \frac{b}{a}(2mD)^{1/2}\right) - E\ . \tag{5.4}$$

The energy levels can be found from the eigenvalue problem

$$(\tilde{H} - E)|\psi_n\rangle = \frac{a^2 h(q)}{2m}\left(K_0 - \frac{b}{a}(2mD)^{1/2}\right)|\psi_n\rangle = 0. \tag{5.5}$$

Since the terms in the parentheses in (5.5) depend only on K_0, we may take $|\psi_n\rangle$ as a group state, i.e., the tilting angle is zero. To find the energy levels take note that

$$C_2 = K_0^2 - K_1^2 - K_2^2 = -\frac{2mE}{a^2} - \frac{1}{4} \tag{5.6}$$

so that $k = \frac{1}{2}(C_2 + \frac{1}{4})^{1/2}$. Thus with $|\psi_n\rangle = |k,n\rangle$ we have the energy levels[15]

$$E_n = -\frac{a^2}{2m}\left[\frac{b}{a}(2mD)^{1/2} - (n + \tfrac{1}{2})\right]^2, \tag{5.7}$$

where

$n = 0,1,2,\ldots,n_{\max}$, where $n_{\max} + \frac{1}{2} \leqslant b(2nD/a)^{1/2}$.

The path integral for the Morse oscillator can be constructed as

$$K(\zeta',\zeta;T) = \lim_{N\to\infty}\sum_k\int\cdots\int\prod_{j=1}^{N-1}d\mu_k(\zeta_j) \times\prod_{j=1}^{N}\exp\left\{\frac{i\epsilon}{\hbar}\left[\frac{\hbar k}{(1 - |\zeta_j|^2)}\left(\zeta_j\frac{\Delta\zeta_j^*}{\epsilon} - \zeta_j^*\frac{\Delta\zeta_j}{\epsilon}\right) - \mathscr{H}(\zeta_j,\zeta_{j-1},t_j)\right]\right\}, \tag{5.8}$$

where

$$\mathscr{H}(\zeta_j,\zeta_{j-1}t_j) = \frac{a^2 h(q(t_j))}{2m} \times\left[k\frac{(1 + \zeta_j^*\zeta_{j-1})}{(1 - \zeta_j^*\zeta_{j-1})} - \frac{b}{a}(2mD)^{1/2}\right] - E. \tag{5.9}$$

Let us now follow Duru and Kleinert[17] in reparametrizing time by making the transformation $\lambda s(t) = \int^t dt'\, h(q(t))$. If we choose $\lambda = (8mD/a)^{1/2}$, s will have the dimensions of time. With $\sigma = s_j - s_{j-1}$ and $\lambda\sigma = h(q(t_j))\epsilon$ we have

$$K(\zeta',\zeta;T) = e^{-\frac{i}{\hbar}ET}\sum_k\int\cdots\int\prod_{j=1}^{N-1}du_k(\zeta_j)\prod_{j=1}^{N} \times\exp\left[\frac{i\sigma}{\hbar}\frac{\hbar k}{(1 - |\zeta_j|^2)}\left(\zeta_j\frac{\Delta\zeta_j^*}{\sigma} - \zeta_j^*\frac{\Delta\zeta_j}{\sigma}\right) - \frac{a^2\lambda}{2m}\left[k\frac{(1 + \zeta_j\zeta_{j-1})}{(1 - \zeta_j^*\zeta_{j-1})} - \frac{b}{a}(2mD)^{1/2}\right]\right]. \tag{5.10}$$

Inserting a factor

$$\lambda\int_s^\infty ds\,\delta\left(T - \int_s^{s'}\frac{ds}{h(s)}\right)$$

to allow s to be treated as an independent variable, we have in the continuous limit

$$K(\zeta',\zeta;T) = e^{-iET/\hbar\lambda}\int_{s'}^{\infty} ds\,\delta\left(T - \int ds \frac{1}{h(s)}\right)\frac{1}{h(s')} \times \sum_k \int \mathscr{D}u_k(\zeta) \times \exp\left\{\frac{i}{h}\int_s^{s'} ds\,\mathscr{L}(\zeta,\zeta',\zeta^*,\zeta^{*\prime})\right\}, \quad (5.11)$$

where

$$\mathscr{L}(\zeta,\zeta',\zeta^*,\zeta^{*\prime}) = \frac{i\hbar k}{(1-|\zeta|^2)}(\zeta\zeta^{*\prime} - \zeta^*\zeta') - \frac{a^2\lambda}{2m}\left[k\frac{(1+|\zeta|^2)}{(1-|\zeta|^2)} - \frac{b}{a}(2mD)^{1/2}\right] \quad (5.12)$$

and where the prime means derivative with respect to the new time-like variable. The equations of motion will simply be those of Eq. (3.12) with the replacement $t \to s$. It is straightforward to show that the motion will be oscillator-like with frequency $\omega = a^2\lambda/4\hbar m$. Thus, using a reparametrization of time in the SU(1,1)CS path integral we can make the Morse oscillator appear as a harmonic oscillator on the Lobachevskii plane.

B. The hydrogen atom

Finally we consider the case of the nonrelativistic hydrogen atom. As is well known, SO(2,1)($\sim$SU(1,1)) is the relevant spectrum generating group and is in fact a subgroup of the much larger dynamical group SO(4,2).[18] The SO(2.1) subalgebra is realized as

$$K_0 = (rp^2 + mr)/2m, \quad (5.13a)$$

$$K_1 = (rp^2 - mr)/2m, \quad (5.13b)$$

$$K_2 = (\vec{r}\cdot\vec{p} - i). \quad (5.13c)$$

Unfortunately the Hamiltonian of the H atom

$$\tilde{H} = \frac{p^2}{2m} - \frac{e^2}{r} \quad (5.14)$$

cannot be written as a polynomial in the SO(2,1) generators of Eq. (5.13). Instead one writes the energy eigenvalue problem as

$$r(\tilde{H} - E)\widetilde{|\psi\rangle} = 0 \quad (5.15)$$

and then, using Eq. (5.13), obtain

$$\tilde{\Omega}(E)\widetilde{|\psi\rangle} = 0, \quad (5.16)$$

where

$$\tilde{\Omega}(E) = r(\tilde{H} - E) = \left\{\frac{1}{2}(K_0 + K_1) - \frac{E}{m}(K_0 - K_1) - e^2\right\}. \quad (5.17)$$

Applying the tilting operation to (5.16) with the tilting angle θ chosen as

$$\theta(E) = \operatorname{arctanh}\{(2E + m)/(2E - m)\}, \quad (5.18)$$

we have

$$e^{-i\theta(E)K_2}\tilde{\Omega}(E)e^{i\theta(E)K_2}e^{-i\theta(E)K_2}\widetilde{|\psi\rangle} = \Omega(E)|\psi\rangle = 0, \quad (5.19)$$

where

$$\Omega(E) = e^{-i\theta(E)K_2}\tilde{\Omega}(E)e^{i\theta(E)K_2} = (-2E/m)^{1/2}K_0 - e^2 \quad (5.20)$$

and where

$$|\psi\rangle = e^{-i\theta(E)K_2}\widetilde{|\psi\rangle}. \quad (5.21)$$

Now for the H atom bound states, a particular set of basis states for SO(4,2) are labeled $|nlm\rangle$ where l and m are the usual angular momentum quantum numbers and n is the principal quantum number. These labels arise from the decomposition $SO(4,2) \supset \underset{lm}{SO(3)} \otimes \underset{nl}{SO(2,1)}$ which corresponds to a separation of variables in spherical polar coordinates. The number l is attached to both SO(3) and SO(2,1) because of the fact that the groups are "coupled" through the equality of their Casimer operators; i.e., if $L_{ij} = \epsilon_{ijk}(\vec{r}x\vec{p})^k$ generate SO(3) then using Eq. (5.13) we have

$$C_2 = K_0^2 - K_1^2 - K_2 = L^2 = \tfrac{1}{2}L_{ij}L_{ij}. \quad (5.22)$$

Thus $k(k-1) = l(l+1)$ so k has possible values $k = -l$ or $k = l + 1$. Only the latter choice leads to a unitary irreducible representation. In this case the action of K_0 on $|nlm\rangle$ is to give the principal quantum number n.

$$K_0|nlm\rangle = n|nlm\rangle, \quad (5.23)$$

where $n - l - 1 = n_r$ is the radial quantum number. We can relabel the states as $|l+1,n_r\rangle$ in accordance with the convention used for SU(1,1) states throughout this paper. (The azimuthal quantum number has been suppressed). With $|\psi\rangle = |l+1,n_r\rangle$ in Eq. (5.19) we obtain, from Eqs. (5.20) and (5.23), the energy levels of hydrogen

$$E_n = -me^4/2n^2 \quad (\hbar = 1). \quad (5.24)$$

Note that since the tilting angle depends on E_n, the transformation (5.21) gives a one-to-one relationship between the physical states and the group and is therefore nonunitary.

Now in constructing the SU(1,1) functional integral for the H atom two problems arise. The first is that $\tilde{H}$ is not polynomial in the SU(1,1) generators and the second is nonunitarity of the tilting transformation. One could write $\tilde{H}$ as

$$\tilde{H} = \frac{1}{r}(r\tilde{H}) = \frac{m}{K_0 - K_1}\left[\frac{1}{2}(K_0 + K_1) - e^2\right] \quad (5.25)$$

and then write a path integral expression as in Eq. (3.9) with

$$\mathscr{H}(\zeta,\zeta^*) = m\int d\mu_l(\zeta')\widetilde{\langle\zeta|}(K_0 - K_1)^{-1}\widetilde{|\zeta'\rangle} \times \widetilde{\langle\zeta'|}[\tfrac{1}{2}(K_0 + K_1) - e^2]\widetilde{|\zeta\rangle}. \quad (5.26)$$

However this expression is awkward because of the presence of $(K_0 - K_1)^{-1}$ and in any case the physical coherent states are not yet defined. In fact, it is not clear how they would be defined since, unlike the case for superfluid helium, the tilting is nonunitary.

We avoid these difficulties by writing a functional integral for a resolvent whose poles yield the bound state energy spectrum. We write

$$G(E) = \operatorname{Tr} 1/\Omega(E), \quad (5.27)$$

where $\Omega(E)$ is given by Eq. (5.20). In terms of the state $|l+1,n_r\rangle$

$$G(E)=\sum_l\sum_{n_r}\frac{|l+1,n_r\rangle\langle l+1,n_r|}{(-2E/m)^{1/2}(n_r+l+1)-e^2}. \quad (5.28)$$

This obviously has poles when E is given by Eq. (5.24).

We write $G(E)$ as

$$G(E)=\frac{1}{i}\int_0^\infty dT\,\mathrm{Tr}\,e^{-i\Omega(E)T} \quad (5.29)$$

and evaluate the trace with SU(1,1)CS to get

$$G(E)=\sum_l G_l(E),$$

where

$$G_l(E)=\frac{1}{i}\int_0^\infty dT\,d\mu_{l+1}(\zeta)\langle\zeta,l+1|e^{-i\Omega(E)T}|\zeta,l+1\rangle, \quad (5.30)$$

and where

$$|\zeta,l+1\rangle=(1-|\zeta|^2)^{l+1}\sum_{n_r=0}^\infty\left[\frac{\Gamma(n_r+2l+2)}{n_r!\Gamma(2l+2)}\right]^{1/2}\times\zeta^{n_r}|l+1,n_r\rangle. \quad (5.31)$$

We now define

$$G_l(\zeta',\zeta;T)=\langle\zeta',l+1|e^{-i\Omega(E)T}|\zeta,l+1\rangle, \quad (5.32)$$

which by the usual methods can be written as the functional integral

$$G_l(\zeta',\zeta;T)=\int\mathcal{D}\mu_{l+1}(\zeta(t))\exp\left\{i\int_0^T dt\,\mathscr{L}'\right\}, \quad (5.33)$$

where

$$\mathscr{L}'=\frac{i(l+1)}{(1-|\zeta|^2)}(\zeta\dot\zeta^*-\zeta^*\dot\zeta)-\Omega_l(\zeta,\zeta^*) \quad (5.34)$$

and where

$$\Omega(\zeta,\zeta^*)=\langle\zeta,l+1|\Omega(E)|\zeta,l+1\rangle=(-2E/m)^{1/2}(l+1)\frac{(1+|\zeta|^2)}{(1-|\zeta|^2)}-e^2. \quad (5.35)$$

$\Omega_l(\zeta,\zeta^*)$ may be thought of as a sort of auxiliary zero energy "Hamiltonian" which is oscillator-like on the curved phase space.

$G_l(E)$ can be evaluated exactly. From Eq. (2.21) we have

$$G_l(\zeta',\zeta;T)=e^{-i[(-2E/m)^{1/2}(l+1)-e^2]T}\times[(1-|\zeta'|^2)(1-|\zeta|^2)]^{l+1}\times[1-\zeta'^*\zeta e^{-i(-2E/m)^{1/2}T}]^{-2(l+1)}. \quad (5.36)$$

Upon taking the trace we find that

$$G_l(E)=\tfrac{1}{2}\int_0^\infty dT\,e^{i[e^2-(-2E/m)^{1/2}(l+\frac{1}{2})]T}\times\left[\sin\left(\left(\frac{-\alpha E}{m}\right)^{1/2}\frac{T}{2}\right)\right]^{-1}. \quad (5.37)$$

Using the identity

$$\frac{1}{2i}\left[\sin\left(\frac{\alpha t}{2}\right)\right]^{-1}=\exp\left(-\frac{i\alpha t}{2}\right)\sum_{n_r=0}^\infty\exp(-in_r\alpha t), \quad (5.38)$$

we obtain

$$G_l(E)=i\sum_{n_r=0}^\infty\int_0^\infty dT\times\exp(i[e^2-(-2E/m)^{1/2}(l+1+n_r)]T), \quad (5.39)$$

which upon integration yields

$$G_l(E)=\sum_{n_r=0}^\infty\frac{1}{[+(-2E/m)^{1/2}(n_r+l+1)-e^2]}, \quad (5.40)$$

which obviously has poles at the bound state energies given by Eq. (5.24). It is interesting that the identity (5.38) is the one used to obtain the bound state energies of the one-dimensional harmonic oscillator.

VI. CONCLUDING REMARKS

In this work we have, in a sense, done the converse of the work of Nieto and collaborators[2] in establishing coherent states for potentials other than the oscillator. As we said earlier, their approach is to map the classical problem onto another classical set of variables such that the Hamiltonian has the appearance of an oscillator; the coherent states being a set of states which minimize a new uncertainty relation constructed from the quantum analogs of the new variables. What we have done is to take coherent states defined for the dynamical group SU(1,1) and through a path integral determined the equivalent classical problem. We have found that the relevant phase space is curved—the Lobachevskii plane—and that for problems where SU(1,1) is a dynamical group, the motion will look oscillator-like in that space.

We have also attempted to show the utility of SU(1,1)CS in calculating the partition function of a simple model of superfluidity. It should be remarked that if the ordinary coherent states are used to calculate Z from the Hamiltonian of Eq. (4.2), one immediately encounters the problems of the mixing of modes + and −. This problem is circumvented with the SU(1,1)CS, which have the further advantage that the Bogoliubov transformation is automatic.

Finally we mention another possible application. For an N-dimensional oscillator SU(1,1) is the dynamical group and the representations which cover the states are given by $k=\frac{1}{2}(l+N/2)$, where l is the usual angular momentum quantum number. Indeed SO(2,1) has recently been shown to be useful in forming a large N expansion in quantum mechanics.[19] In that work the method is perturbation theoretic. However, if we look at Eq. (3.8) we see that $\hbar$ cancels out of the exponential altogether if $\mathscr{H}$ is given by Eq. (3.16), indicating that $1/k$ could play the role of $\hbar$. This leads us to believe that we might construct a Bohr–Sommerfeld type quantization rule and determine the energy levels for such things as the anharmonic oscillator without the need for taking the limit $N\to\infty$. Similar considerations have been used for pseudospin Hamiltonians.[20] This and related problems will be discussed elsewhere.

[1] See for example J. R. Klauder and E. Sudashan, *Fundamentals of Quantum Optics* (Benjamin, New York, 1968).

[2] M. M. Nieto and L. M. Simmons, Jr., Phys. Rev. D**20**, 1321, 1332, and 1342 (1979).

[3]J. M. Radcliffe, J. Phys. **A4**, 313 (1971).
[4]A. M. Perelomov, Commun. Math. Phys. **26**, 222 (1972).
[5]See B. G. Wybourne, *Classical Groups for Physicists* (Wiley, New York, 1974), Chap. 18.
[6]A. I. Solomon, J. Math. Phys. **12**, 390 (1969).
[7]A. Casher, D. Lurie, and M. Revzen, J. Math. Phys. **9**, 1312 (1968).
[8]H. Kuratsuji and T. Suzuki, J. Math. Phys. **21**, 472 (1980).
[9]A. O. Barut and L. Girardello, Commun. Math. Phys. **21**, 41 (1971).
[10]It must be remembered that SU(1,1) is not a symmetry group so that in contrast to SU(2), the quantum number labeling the representations is not a conserved quantity.
[11]A. O. Barut and C. Fronsdal, Proc. R. Soc. (London) **287A**, 532 (1965).
[12]C. E. Wulfman in *Group Theory and its Applications*, Vol. II, edited by E. M. Loebl (Academic, New York, 1971).
[13]There is a difficulty, apparently unnoticed by Perelomov, for the case $k = \frac{1}{2}$. A straightforward evaluation of Eq. (2.22) with $k = \frac{1}{4}$ gives a divergence. We suggest that one shoud always write the integral as:

$$\langle 0|\hat{I}|0\rangle = 1 = \int \frac{|2k-1|}{\pi} \frac{d^2\zeta}{(1-|\zeta|^2)^2}(1-|\zeta|^2)^{|2k-1|+1},$$

i.e., when doing the integrals we always make the replacement

$2k \to |2k-1|+1$.
[14]F. A. Berezin, Commun. Math. Phys. **40**, 153 (1975).
[15]P. M. Morse, Phys. Rev. **34**, 57 (1929).
[16]P. Cordero and S. Hojman, Lett. Nuovo Cimento **4**, 1123 (1970).
[17]I. H. Duru and H. Kleinert, Phys. Lett. **84B**, 185 (1979).
[18]A. O. Barut, *Dynamical Groups and Generalized Symmetries in Quantum Theory* (U. of Canterbury P., Christchurch, New Zealand, 1971).
[19]L. D. Mlodinow and N. Papanicolaou, Ann. Phys. (N.Y.) **128**, 314 (1980).
[20]R. Shankar, Phys. Rev. Lett. **45**, 1088 (1980).

Dynamical group of microscopic collective states. II. Boson representations in d dimensions

J. Deenen and C. Quesne[a)]
Physique Théorique et Mathématique CP 229, Université Libre de Bruxelles, Bd. du Triomphe, B 1050 Brussels, Belgium

(Received 17 February 1982; accepted for publication 7 May 1982)

The present series of papers deals with various realizations of the dynamical group $\mathcal{S}p_c(2d,R)$ of microscopic collective states for an A nucleon system in d dimensions, defined as those A particle states invariant under the orthogonal group $O(n)$ associated with the $n = A - 1$ Jacobi vectors. In the present paper, we derive two boson representations of $\mathcal{S}p_c(2d,R)$, namely the Dyson representation and the Holstein–Primakoff (HP) one. Our starting point is a representation of microscopic collective states, as introduced in the first paper of the present series, in a Barut Hilbert space $\mathcal{F}_c$ of analytic functions in $\nu = \frac{1}{2}d(d+1)$ complex variables. Basis functions in $\mathcal{F}_c$, classified according to the chain $\mathcal{S}p_c(2d,R) \supset \mathcal{U}_c(d)$, can be put into one-to-one correspondence with basis functions, classified according to the chain $u(\nu) \supset u(d)$, in a Bargmann Hilbert space $\mathcal{B}$ of analytic functions in ν complex variables representing ν-dimensional boson states. By equating the complex variables of $\mathcal{F}_c$ and their conjugate momenta with those of $\mathcal{B}$, we get the non-Hermitian Dyson representation of $\mathcal{S}p_c(2d,R)$. We then go from the latter to the Hermitian HP representation by means of a canonical transformation that restores the Hermiticity properties of the variables and conjugate momenta. The inverse of the HP representation gives the unitary representation in quantum mechanics of the classical canonical transformation relating the oscillator Hamiltonians of the microscopic collective model and the boson macroscopic one. From the ν boson creation and annihilation operators, it is possible to build the generators of a $u(\nu)$ group, which in the physical three-dimensional case reduces to $u(6)$. The latter is finally compared with the U(6) group appearing in the interacting boson model.

PACS numbers: 02.20. + b, 03.65.Fd, 21.60.Fw

1. INTRODUCTION

In the present series of papers, we study various realizations of the dynamical group of microscopic collective states for an A nucleon system. As was suggested by some authors,[1,2] the collective subspace of the A nucleon space is assumed to be spanned by all the A nucleon states invariant under the transformations of the orthogonal group $O(n)$ associated with the $n = A - 1$ Jacobi vectors. This subspace carries a single irreducible representation (IR) of the dynamical group of collective states we are concerned with.

In the first paper[3] of the present series (henceforth referred to as I), we showed that in d dimensions the dynamical group of microscopic collective states is the restriction to the collective subspace of a symplectic group in $2d$ dimensions, $\mathcal{S}p(2d,R)$, and we denoted it by $\mathcal{S}p_c(2d,R)$. In the physical three-dimensional case, the latter reduces to $\mathcal{S}p_c(6,R)$. We then proposed various realizations of $\mathcal{S}p_c(2d,R)$ that we studied in detail only in the one-dimensional case.

In the present paper, we wish to establish a relationship between the $O(n)$ invariant microscopic collective model and the interacting boson model[4] (IBM) of Arima and Iachello through the study of boson representations of $\mathcal{S}p_c(6,R)$. As the IBM, which is based upon a U(6) group, is quite successful in describing many nuclear collective states, there have been recently various attempts[5] to microscopically explain the U(6) symmetry first introduced from a macroscopic viewpoint.

In the $O(n)$ invariant part of the A-body problem, the possible appearance of a $u(6)$ group was first noted by Vanagas[1] in an indirect way through the branching rules for the IR's of the groups characterizing the collective states. Later on, Moshinsky and Seligman[6] determined in two dimensions the classical canonical transformation relating the oscillator Hamiltonians of the microscopic collective model and the boson macroscopic one. The latter has a U(3) symmetry group, which is the two-dimensional analog of U(6). The generalization of their work to three dimensions seemed, however, difficult to carry out due to the explicit use of the Dzublik[7]–Zickendraht[8] coordinate transformation.

In I, we proposed to realize the dynamical group in d dimensions, $\mathcal{S}p_c(2d,R)$, in terms of $\nu = \frac{1}{2}d(d+1)$ boson creation and annihilation operators through a generalized Holstein–Primakoff[9] (HP) representation and to invert such representation in order to express these boson operators in terms of the generators of $\mathcal{S}p_c(2d,R)$. From the ν boson creation and annihilation operators, it is then straightforward to construct the generators of a $u(\nu)$ group, which in the physical three-dimensional case reduces to $u(6)$. The inverse of the HP representation of $\mathcal{S}p_c(2d,R)$ gives the unitary representation in quantum mechanics of the classical canonical transformation relating the oscillator Hamiltonians of the microscopic collective model and the boson macroscopic one.[10,11] We have therefore found a way to generalize the work of Ref. 6 to any number of dimensions.

As in I detailed calculations were performed only for $\mathcal{S}p_c(2,R)$, we carry them out for the d-dimensional case in

[a)] Maître de recherches F.N.R.S.

 0022-2488/82/112004-12$02.50 © 1982 American Institute of Physics

the present paper. Actually we shall not only study the HP representation of $\mathcal{S}p_c(2d,R)$, but also the Dyson[12] one. For such purpose, it will turn out that the Barut representation of microscopic collective states, introduced in I, is a good starting point. By Barut representation we mean a representation in a Barut Hilbert space[13] $\mathcal{F}_c$ of analytic functions in ν complex variables. Such Barut Hilbert space is the O(n) invariant subspace of a Bargmann Hilbert space[14] $\mathcal{F}$ of analytic functions in dn complex variables.

The function in ν complex variables which represents a given microscopic collective state in the Barut Hilbert space $\mathcal{F}_c$ can be put into one-to-one correspondence with an analytic function in ν complex variables representing a definite ν-dimensional boson state in a Bargmann Hilbert space $\mathcal{B}$. Basis functions in $\mathcal{F}_c$ are classified according to the chain $\mathcal{S}p_c(2d,R)\supset\mathcal{U}_c(d)$ while those in $\mathcal{B}$ are classified according to the chain $\mathrm{U}(\nu)\supset\mathrm{U}(d)$. The one-to-one mapping between $\mathcal{F}_c$ and $\mathcal{B}$ is such that functions transforming in a given way under $\mathcal{U}_c(d)$ are mapped onto functions transforming in the same way under $\mathrm{U}(d)$. Any microscopic collective state has therefore a boson image; from this mapping one can then derive the boson image of any collective operator, and in particular that of the generators of $\mathcal{S}p_c(2d,R)$.

The boson representation of $\mathcal{S}p_c(2d,R)$ obtained in equating the complex variables of $\mathcal{F}_c$ and their conjugate momenta with those of $\mathcal{B}$ is the non-Hermitian Dyson representation, which is therefore essentially equivalent to the Barut representation. We can also go from the latter to the Hermitian HP representation by means of a canonical transformation relating both sets of complex variables and conjugate momenta, which preserves their commutation relations and restores their Hermiticity properties.

In Sec. 2, we start to review some results obtained in I for the Barut representation of microscopic collective states and then establish the forms of the $\mathcal{S}p_c(2d,R)$ generators which will be relevant for the present study. In Sec. 3, we explicitly construct the representation in $\mathcal{F}_c$ of the microscopic collective states classified according to the chain $\mathcal{S}p_c(2d,R)\supset\mathcal{U}_c(d)$ and of highest weight with respect to $\mathcal{U}_c(d)$, and that in $\mathcal{B}$ of ν-dimensional boson states classified according to the chain $\mathrm{U}(\nu)\supset\mathrm{U}(d)$ and of highest weight with respect to $\mathrm{U}(d)$. In Sec. 4, we turn to the general form of irreducible tensors $T[2\dot{0}]$ with respect to $\mathrm{U}(d)$ and determine some selected basis for them. We are then in a position to explicitly derive the Dyson and HP representations of $\mathcal{S}p_c(2d,R)$ in Secs. 5 and 6, respectively. For the latter both a compact form and a finite expansion are obtained. Finally, in Sec. 7, we compare the U(6) group obtained in the present picture and the U(6) group of the IBM and briefly discuss some possible extensions of the method of getting boson representations developed in the present paper.

2. BARUT REPRESENTATION OF $\mathcal{S}p_c(2d,R)$

A translationally invariant system of A fermions in d dimensions can be described alternatively in terms of its Jacobi coordinates x_{is}, $i=1,...,d$, $s=1,...,n=A-1$, and their conjugate momenta $p_{is}=-i\partial/\partial x_{is}$, or in terms of the corresponding boson creation and annihilation operators η_{is} and ξ_{is}. To the n Jacobi vectors is associated an orthogonal group O(n), under whose transformations the A particle collective states are assumed to be invariant.[1]

In I, we considered an oscillator basis for the collective subspace of the A particle space, spanned by the states

$$|\phi_{N_{11}N_{12}\cdots N_{dd}}\rangle = a_{N_{11}N_{12}\cdots N_{dd}}\prod_{i\leqslant j=1}^{d}(\mathcal{D}_{ij}^{\dagger})^{N_{ij}}|0\rangle, \tag{2.1}$$

depending upon $\nu=\frac{1}{2}d(d+1)$ quantum numbers N_{ij}, $1\leqslant i\leqslant j\leqslant d$, which can take all non-negative integer values. In Eq. (2.1), $a_{N_{11}N_{12}\cdots N_{dd}}$ is some normalization coefficient, $|0\rangle$ is the boson vacuum state, and the operators $\mathcal{D}_{ij}^{\dagger}$ are defined by

$$\mathcal{D}_{ij}^{\dagger}=\mathcal{D}_{ji}^{\dagger}=\sum_{s=1}^{n}\eta_{is}\eta_{js},\quad 1\leqslant i\leqslant j\leqslant d. \tag{2.2a}$$

With the operators

$$\mathcal{D}_{ij}=\mathcal{D}_{ji}=\sum_{s=1}^{n}\xi_{is}\xi_{js},\quad 1\leqslant i\leqslant j\leqslant d, \tag{2.2b}$$

and

$$\mathcal{C}_{ij}=\mathcal{C}_{ij}+\frac{n}{2}\delta_{ij},\quad \mathcal{C}_{ij}=\sum_{s=1}^{n}\eta_{is}\xi_{js},\quad i,j=1,...,d, \tag{2.2c}$$

they generate an $\mathcal{S}p(2d,R)$ group. The states (2.1) belong to the single IR $\langle(n/2)^d\rangle$ of $\mathcal{S}p(2d,R)$, so that the dynamical group of microscopic collective states is the restriction $\mathcal{S}p_c(2d,R)$ of $\mathcal{S}p(2d,R)$ to the collective subspace. Its generators are denoted by $\mathcal{D}_{ij}^{c\dagger}$, $\mathcal{D}_{ij}^{c}$ and $\mathcal{C}_{ij}^{c}=\mathcal{C}_{ij}^{c}+(n/2)\delta_{ij}$, respectively.

The A particle states can be represented in a Bargmann Hilbert space[14] of analytic functions in dn complex variables z_{is}, $i=1,...,d$, $s=1,...,n$. In $\mathcal{F}$, the boson operators η_{is} and ξ_{is} are represented by z_{is} and $\partial/\partial z_{is}$, respectively. The collective subspace of the A particle space is then mapped onto the subspace $\mathcal{F}_c$ of $\mathcal{F}$ which is left invariant by O(n). It was shown in I that any collective state $|\psi\rangle$ is represented in $\mathcal{F}_c$ by an analytic function $\bar{\psi}(w_{11},w_{12},...,w_{dd})=\langle w_{11}w_{12}\cdots w_{dd}|\psi\rangle$ in ν complex collective variables

$$w_{ij}=w_{ji}=\sum_{s=1}^{n}z_{is}z_{js},\quad 1\leqslant i\leqslant j\leqslant d. \tag{2.3}$$

In particular the basis states (2.1) are represented by the functions

$$\langle w_{11}w_{12}\cdots w_{dd}|\phi_{N_{11}N_{12}\cdots N_{dd}}\rangle = a_{N_{11}N_{12}\cdots N_{dd}}\prod_{i\leqslant j=1}^{n}w_{ij}^{N_{ij}}. \tag{2.4}$$

The space $\mathcal{F}_c$ can be equipped with a scalar product which is preserved in the one-to-one mapping between microscopic collective states and analytic functions of $\mathcal{F}_c$,

$$\langle\phi|\psi\rangle=\int d\sigma(w_{11},w_{12},...,w_{dd})\times\langle\phi|w_{11}w_{12}\cdots w_{dd}\rangle\langle w_{11}w_{12}\cdots w_{dd}|\psi\rangle. \tag{2.5}$$

The measure $d\sigma(w_{11},w_{12},...,w_{dd})$ directly derives from the measure[14] of the Bargmann space $\mathcal{F}$, $\Pi_{i=1}^{d}\Pi_{s=1}^{n}d\mu(z_{is})$,

where

$$d\mu(z_{is}) = \pi^{-1} \exp(-z_{is}z_{is}^{*}) d \operatorname{Re} z_{is}\, d \operatorname{Im} z_{is}, \tag{2.6}$$

by integration over the $dn - \nu$ noncollective variables. In the present study, we shall not need its explicit form, which was determined in I only in the one-dimensional case. We shall therefore leave its determination in the d-dimensional case for a subsequent paper.

$\mathcal{F}_c$ is called the Barut subspace of $\mathcal{F}$ and the representation of microscopic collective states in $\mathcal{F}_c$ their Barut representation because a scalar product similar to the one defined in Eq. (2.5) for $d = 1$ was already considered by Barut and Girardello.[13] Instead of the basis (2.4) of $\mathcal{F}_c$, we could use as well a basis of coherent states associated with the Lie algebra of $\mathcal{S}\mathcal{p}_c(2d,R)$, as these authors did and we explicitly showed in the $d = 1$ case.

In the Barut representation, the generators of the dynamical group $\mathcal{S}\mathcal{p}_c(2d,R)$ are represented by the following differential operators:

$$\mathcal{D}_{ij}^{c\dagger} = w_{ij},$$

$$\mathcal{D}_{ij}^{c} = \sum_{kl} (1+\delta_{ki})(1+\delta_{lj}) w_{kl} \frac{\partial^2}{\partial w_{ik} \partial w_{jl}} + n(1+\delta_{ij}) \frac{\partial}{\partial w_{ij}} \tag{2.7}$$

and

$$\mathcal{E}_{ij}^{c} = \mathcal{C}_{ij}^{c} + \frac{n}{2}\delta_{ij}, \quad \mathcal{C}_{ij}^{c} = \sum_{k} (1+\delta_{kj}) w_{ik} \frac{\partial}{\partial w_{jk}},$$

which with respect to the scalar product (2.5) of $\mathcal{F}_c$ satisfy the same Hermiticity properties as the original operators, i.e.,

$$(\mathcal{D}_{ij}^{c\dagger})^{\dagger} = \mathcal{D}_{ij}^{c} \quad \text{and} \quad (\mathcal{E}_{ij}^{c})^{\dagger} = \mathcal{E}_{ji}^{c}. \tag{2.8}$$

They also satisfy the usual commutation relations of the generators of an $\mathcal{S}\mathcal{p}(2d,R)$ group:

$$\begin{aligned}
&[\mathcal{E}_{ij}^{c}, \mathcal{E}_{kl}^{c}] = \delta_{jk}\mathcal{E}_{il}^{c} - \delta_{il}\mathcal{E}_{kj}^{c},\\
&[\mathcal{E}_{ij}^{c}, \mathcal{D}_{kl}^{c\dagger}] = \delta_{jk}\mathcal{D}_{il}^{c\dagger} + \delta_{jl}\mathcal{D}_{ik}^{c\dagger},\\
&[\mathcal{E}_{ij}^{c}, \mathcal{D}_{kl}^{c}] = -\delta_{ik}\mathcal{D}_{jl}^{c} - \delta_{il}\mathcal{D}_{jk}^{c},\\
&[\mathcal{D}_{ij}^{c\dagger}, \mathcal{D}_{kl}^{c\dagger}] = [\mathcal{D}_{ij}^{c}, \mathcal{D}_{kl}^{c}] = 0,\\
&[\mathcal{D}_{ij}^{c}, \mathcal{D}_{kl}^{c\dagger}] = \delta_{ik}\mathcal{E}_{lj}^{c} + \delta_{il}\mathcal{E}_{kj}^{c} + \delta_{jk}\mathcal{E}_{li}^{c} + \delta_{jl}\mathcal{E}_{ki}^{c}.
\end{aligned} \tag{2.9}$$

In the present study, it is convenient to introduce a matrix notation for the generators of $\mathcal{S}\mathcal{p}_c(2d,R)$. Let $\mathcal{D}^{c\dagger}$ denote the $d \times d$ matrix whose elements are $\mathcal{D}_{ij}^{c\dagger}$, $i,j = 1,\ldots,d$. In the same way, we define the matrices $\mathcal{D}^{c}$, $\mathcal{E}^{c}$, and $\mathcal{C}^{c}$. In $\mathcal{F}_c$, the complex variables w_{ij} and the corresponding differential operators Δ_{ij}, defined by

$$\Delta_{ij} = (1+\delta_{ij})\partial/\partial w_{ij}, \tag{2.10}$$

can also be considered as the elements of $d \times d$ matrices $\mathbf{w}$ and $\boldsymbol{\Delta}$, respectively. All the matrices $\mathcal{D}^{c\dagger}$, $\mathcal{D}^{c}$, $\mathbf{w}$, and $\boldsymbol{\Delta}$ are symmetrical. The representation (2.7) of the generators of $\mathcal{S}\mathcal{p}_c(2d,R)$ in $\mathcal{F}_c$ can be rewritten in matrix notation as

$$\mathcal{D}^{c\dagger} = \mathbf{w}, \tag{2.11a}$$

$$\mathcal{D}^{c} = \boldsymbol{\Delta}[\mathcal{C}^{c} + (n-d-1)\mathbf{I}] = [\tilde{\mathcal{C}}^{c} + n\mathbf{I}]\boldsymbol{\Delta} \tag{2.11b}$$

and

$$\mathcal{E}^{c} = \mathcal{C}^{c} + \frac{n}{2}\mathbf{I}, \quad \mathcal{C}^{c} = \mathbf{w}\boldsymbol{\Delta}, \tag{2.11c}$$

where $\mathbf{I}$ denotes the $d \times d$ unit matrix and $\tilde{\mathcal{C}}^{c}$ the transpose of $\mathcal{C}^{c}$, i.e., $\tilde{\mathcal{C}}_{ij}^{c} = \mathcal{C}_{ji}^{c}$. To obtain Eq. (2.11b), we have explicitly used the commutation relation

$$[\Delta_{ij}, w_{kl}] = \delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk}, \tag{2.12}$$

which results from the definition (2.10) of Δ_{ij}.

In Secs. 5 and 6, Eq. (2.11) will prove to be the appropriate starting point to get boson representations of $\mathcal{S}\mathcal{p}_c(2d,R)$. However, before beginning to study such representations, we first introduce in the next section still another basis of $\mathcal{F}_c$, spanned by all the states classified according to IR's of the chain

$$\mathcal{S}\mathcal{p}_c(2d,R) \supset \mathcal{U}_c(d), \tag{2.13}$$

where $\mathcal{U}_c(d)$ is the unitary group in d dimensions generated by the operators $\mathcal{C}_{ij}^{c}$, $i,j = 1,\ldots,d$. The basis states (2.1), although characterized by a given weight in $\mathcal{U}_c(d)$, do not indeed belong to a definite IR of $\mathcal{U}_c(d)$ in general. Basis states classified according to (2.13) will then be put into one-to-one correspondence with representations of boson states in a Bargmann space $\mathcal{B}$.

3. BASIS FOR IRREDUCIBLE REPRESENTATIONS OF THE $\mathcal{S}\mathcal{p}_c(2d,R) \supset \mathcal{U}_c(d)$ AND $\mathfrak{U}(\nu) \supset \mathfrak{U}(d)$ CHAINS OF GROUPS

Let us denote microscopic collective states classified according to the chain (2.13) by

$$\left| \begin{matrix} h_1 h_2 \cdots h_d \\ (h) \end{matrix} \right\rangle,$$

where $[h_1 h_2 \cdots h_d]$ is a Young diagram specifying the IR of $\mathcal{U}_c(d)$, and (h) is the row index of such IR. As in the following we shall be concerned only with the highest weight state (hws) of the IR $[h_1 h_2 \cdots h_d]$,

$$\left| \begin{matrix} h_1 h_2 \cdots h_d \\ (h)_{\max} \end{matrix} \right\rangle,$$

that we shall denote in short by $|h_1 h_2 \cdots h_d\rangle$, we do not have to make any explicit choice for (h), which might therefore be a Gel'fand pattern[15] associated with the canonical chain of $\mathcal{U}_c(d)$ or any other set of quantum numbers corresponding to a noncanonical chain of $\mathcal{U}_c(d)$. The collective states

$$\left| \begin{matrix} h_1 h_2 \cdots h_d \\ (h) \end{matrix} \right\rangle \quad \text{and} \quad |h_1 h_2 \cdots h_d\rangle$$

are represented in $\mathcal{F}_c$ by

$$\langle w_{11} w_{12} \cdots w_{dd} \left| \begin{matrix} h_1 h_2 \cdots h_d \\ (h) \end{matrix} \right\rangle$$

and $\langle w_{11} w_{12} \cdots w_{dd} | h_1 h_2 \cdots h_d \rangle$, respectively.

Let us determine the explicit form of the representation $\langle w_{11} w_{12} \cdots w_{dd} | h_1 h_2 \cdots h_d \rangle$ of the hws of the IR $[h_1 h_2 \cdots h_d]$. By definition, it must satisfy the following system of first-order differential equations:

$$\sum_k w_{ik}\Delta_{ki}\langle w_{11}w_{12}\cdots w_{dd}|h_1h_2\cdots h_d\rangle = h_i\langle w_{11}w_{12}\cdots w_{dd}|h_1h_2\cdots h_d\rangle, \quad i=1,\ldots,d, \tag{3.1}$$

and

$$\sum_k w_{ik}\Delta_{kj}\langle w_{11}w_{12}\cdots w_{dd}|h_1h_2\cdots h_d\rangle = 0, \quad 1\leqslant i<j\leqslant d. \tag{3.2}$$

To find the solution of such a system, it is convenient to first replace the set of ν independent variables w_{ij}, $1\leqslant i\leqslant j\leqslant d$, by another set of ν variables, defined by

$$w_{1,2,\ldots,i-1,i;1,2,\ldots,i-1,j} = \sum_p (-1)^p w_{1p(1)}w_{2p(2)}\cdots w_{i-1\,p(i-1)}w_{ip(j)}, \quad 1\leqslant i\leqslant j\leqslant d, \tag{3.3}$$

where the summation is carried out over the $i!$ permutations of the indices $1,2,\ldots,i-1,j$. The new variables are minors of order i of det $\mathbf{w}$. They are functionally independent and may therefore be used instead of the w_{ij} because their Jacobian with respect to the w_{ij} is equal to

$$\prod_{i=1}^{d-1}(w_{1,2,\ldots,i;1,2,\ldots,i})^{d-i}\neq 0.$$

Let us look for a solution of Eqs. (3.1) and (3.2) of the following form:

$$\langle w_{11}w_{12}\cdots w_{dd}|h_1h_2\cdots h_d\rangle = \sum_{n_{11}n_{12}\cdots n_{dd}} a_{n_{11}n_{12}\cdots n_{dd}}\prod_{i=1}^{d}\prod_{j=i}^{d}(w_{1,2,\ldots,i-1,i;1,2,\ldots,i-1,j})^{n_{ij}}, \tag{3.4}$$

where n_{ij} are some non-negative integers and $a_{n_{11}n_{12}\cdots n_{dd}}$ some constants to be determined. By noting that

$$\Big(\sum_k w_{ik}\Delta_{kj}\Big)w_{1,2,\ldots,l-1,l;1,2,\ldots,l-1,m} = \delta_{jm}w_{1,2,\ldots,l-1,l;1,2,\ldots,l-1,i}, \quad 1\leqslant i<j\leqslant d, \tag{3.5}$$

and by successively imposing that Eq. (3.2) is satisfied for $i=1$, $i=2,\ldots$, and $i=d-1$, it can be proved that the state (3.4) does not depend upon the variables $w_{1j}(j>1)$, $w_{1,2;1,j}(j>2),\ldots,w_{1,2,\cdots,d-2,d-1;1,2,\cdots,d-2,d}$. It then reduces to

$$\langle w_{11}w_{12}\cdots w_{dd}|h_1h_2\cdots h_d\rangle = \sum_{n_{11}n_{12}\cdots n_{dd}} a_{n_{11}n_{22}\cdots n_{dd}}\prod_{i=1}^{d}(w_{1,2,\ldots,i;1,2,\ldots,i})^{n_{ii}}. \tag{3.6}$$

It is then straightforward to show that Eq. (3.1) is satisfied if

$$n_{ii} = \tfrac{1}{2}(h_i - h_{i+1}), \quad i=1,\ldots,d-1, \tag{3.7}$$

and

$$n_{dd} = \tfrac{1}{2}h_d.$$

Equation (3.7) implies that $h_1,h_2,\ldots,h_d$ must be non-negative even integers.

We have established that $\mathscr{F}_c$ is a carrier space for the direct sum of IR's $[h_1h_2\cdots h_d]$ of $\mathscr{U}_c(d)$, for which $h_1,h_2,\ldots,h_d$ may take any even values such that $h_1\geqslant h_2\geqslant\cdots\geqslant h_d\geqslant 0$. Each one of these IR's appears with a multiplicity one in the direct sum. Due to the one-to-one mapping between $\mathscr{F}_c$ and the collective subspace of the A particle space, the same is true for the latter. Moreover the representation in $\mathscr{F}_c$ of the hws of the IR $[h_1h_2\cdots h_d]$ is given by

$$\langle w_{11}w_{12}\cdots w_{dd}|h_1h_2\cdots h_d\rangle = A_{h_1h_2\cdots h_d}\prod_{i=1}^{d}(w_{1,2,\ldots,i;1,2,\ldots,i})^{(1/2)(h_i-h_{i+1})}, \tag{3.8}$$

where h_{d+1} is assumed to be equal to zero and $A_{h_1h_2\cdots h_d}$ is some normalization coefficient, determined by the condition

$$\int d\sigma(w_{11},w_{12},\ldots,w_{dd})\langle h_1h_2\cdots h_d|w_{11}w_{12}\cdots w_{dd}\rangle \times\langle w_{11}w_{12}\cdots w_{dd}|h_1h_2\cdots h_d\rangle = 1. \tag{3.9}$$

The representation in $\mathscr{F}_c$ of the other basis states of the IR $[h_1h_2\cdots h_d]$ could be obtained from Eq. (3.8) by applying some appropriate lowering operators.

We shall now prove that the functions (3.8) can be put into one-to-one correspondence with functions representing boson states classified according to some group $\mathfrak{U}(d)$. For that purpose, let us introduce ν independent boson creation and annihilation operators $a_{ij}^\dagger = a_{ji}^\dagger$ and $a_{ij} = a_{ji}$, $1\leqslant i\leqslant j\leqslant d$ (not to be confused with the operators η_{is} and ξ_{is}!). They satisfy the usual commutation relations of boson operators, which in the present case take the following form:

$$[a_{ij}^\dagger,a_{kl}^\dagger] = [a_{ij},a_{kl}] = 0,$$
$$[a_{ij},a_{kl}^\dagger] = \delta_{(ij),(kl)} = (1+\delta_{ij})^{-1}(\delta_{ik}\delta_{jl}+\delta_{il}\delta_{jk}). \tag{3.10}$$

From the operators $a_{ij}^\dagger$ and a_{ij}, we can build the generators

$$\mathfrak{G}_{ij,kl} = a_{ij}^\dagger a_{kl}, \quad 1\leqslant i\leqslant j\leqslant d, \ 1\leqslant k\leqslant l\leqslant d, \tag{3.11}$$

of a $\mathfrak{U}(\nu)$ group, whose commutation relations are

$$[\mathfrak{G}_{ij,kl},\mathfrak{G}_{pq,rs}] = \delta_{(kl),(pq)}\mathfrak{G}_{ij,rs} - \delta_{(ij),(rs)}\mathfrak{G}_{pq,kl}. \tag{3.12}$$

Let us now define non-normalized boson operators by the following relations:

$$\bar{a}_{ij}^\dagger = (1+\delta_{ij})^{1/2}a_{ij}^\dagger,$$
$$\bar{a}_{ij} = (1+\delta_{ij})^{1/2}a_{ij} = (\bar{a}_{ij}^\dagger)^\dagger. \tag{3.13}$$

From Eq. (3.10), their commutation relations can be written as

$$[\bar{a}_{ij}^\dagger,\bar{a}_{kl}^\dagger] = [\bar{a}_{ij},\bar{a}_{kl}] = 0,$$
$$[\bar{a}_{ij},\bar{a}_{kl}^\dagger] = \delta_{ik}\delta_{jl}+\delta_{il}\delta_{jk}. \tag{3.14}$$

It is now straightforward to show that the operators $\mathfrak{G}_{ij}$, defined by

$$\mathfrak{G}_{ij} = \sum_k \bar{a}_{ik}^\dagger\bar{a}_{jk} = \sum_k[(1+\delta_{ik})(1+\delta_{jk})]^{1/2}\mathfrak{G}_{ik,jk}, \tag{3.15}$$

satisfy the following commutation relations:

$$[\mathfrak{G}_{ij},\mathfrak{G}_{kl}] = \delta_{jk}\mathfrak{G}_{il} - \delta_{il}\mathfrak{G}_{kj}, \tag{3.16}$$

and therefore generates a $\mathfrak{U}(d)$ subgroup of $\mathfrak{U}(\nu)$.

N boson states can be classified according to IR's of the group chain

$$\mathfrak{U}(\nu)\supset\mathfrak{U}(d). \tag{3.17}$$

They are characterized by the symmetrical IR $[N]$ of $\mathfrak{U}(\nu)$, the IR $[h_1h_2\cdots h_d]$ of $\mathfrak{U}(d)$, and the row (h) of the latter. As the

first-order Casimir operators of $\mathfrak{U}(\nu)$ and $\mathfrak{U}(d)$ are related by the equation

$$\sum_i \mathbb{G}_{ii} = 2\sum_{i<j} \mathbb{G}_{ij,ij}, \tag{3.18}$$

which directly derives from Eq. (3.15), N is entirely specified by the IR of $\mathfrak{U}(d)$ through the relation

$$2N = \sum_i h_i. \tag{3.19}$$

Boson states classified according to IR's of the chain (3.17) may therefore be denoted by

$$\left| \begin{matrix} h_1 h_2 \cdots h_d \\ (h) \end{matrix} \right),$$

where we have used a round bracket in order to distinguish them from the collective states

$$\left| \begin{matrix} h_1 h_2 \cdots h_d \\ (h) \end{matrix} \right\rangle.$$

The hws of the IR $[h_1 h_2 \cdots h_d]$ of $\mathfrak{U}(d)$ is written in short as $|h_1 h_2 \cdots h_d)$.

As is well known, the boson operators $a_{ij}^\dagger$ and a_{ij} are represented in a Bargmann space[14] $\mathscr{B}$ of analytic functions in ν complex variables $\zeta_{ij} = \zeta_{ji}$, $1 \leqslant i \leqslant j \leqslant d$, by ζ_{ij} and $\partial/\partial\zeta_{ij}$, respectively. The non-normalized boson operators $\bar{a}_{ij}^\dagger$ and $\bar{a}_{ij}$ are then represented by

$$\begin{aligned} \bar{\zeta}_{ij} &= (1+\delta_{ij})^{1/2}\zeta_{ij} \\ \text{and} \qquad & \\ \chi_{ij} &= (1+\delta_{ij})^{1/2}\frac{\partial}{\partial\zeta_{ij}} = (1+\delta_{ij})\frac{\partial}{\partial\bar{\zeta}_{ij}}, \end{aligned} \tag{3.20}$$

respectively. The generators of $\mathfrak{U}(\nu)$ and $\mathfrak{U}(d)$ become the first-order differential operators

$$\mathbb{G}_{ij,kl} = \zeta_{ij}\,\partial/\partial\zeta_{kl} \tag{3.21}$$

and

$$\mathbb{G}_{ij} = \sum_k \bar{\zeta}_{ik}\chi_{kj}, \tag{3.22}$$

and the boson states

$$\left| \begin{matrix} h_1 h_2 \cdots h_d \\ (h) \end{matrix} \right)$$

are represented by N th degree polynomials in ζ_{ij}, $1 \leqslant i \leqslant j \leqslant d$, or equivalently in $\bar{\zeta}_{ij}$, $1 \leqslant i \leqslant j \leqslant d$, that we shall denote by

$$\left(\bar{\xi}_{11}\bar{\xi}_{12}\cdots\bar{\xi}_{dd} \left| \begin{matrix} h_1 h_2 \cdots h_d \\ (h) \end{matrix} \right. \right).$$

The representation $(\bar{\xi}_{11}\bar{\xi}_{12}\cdots\bar{\xi}_{dd}|h_1 h_2\cdots h_d)$ of the hws of the IR $[h_1 h_2 \cdots h_d]$ of $\mathfrak{U}(d)$ must satisfy the following system of first-order differential equations:

$$\begin{aligned} &\sum_k \bar{\xi}_{ik}\chi_{ki}(\bar{\xi}_{11}\bar{\xi}_{12}\cdots\bar{\xi}_{dd}|h_1 h_2\cdots h_d) \\ &\quad = h_i(\bar{\xi}_{11}\bar{\xi}_{12}\cdots\bar{\xi}_{dd}|h_1 h_2\cdots h_d), \quad i = 1,\ldots,d, \end{aligned} \tag{3.23}$$

and

$$\sum_k \bar{\xi}_{ik}\chi_{kj}(\bar{\xi}_{11}\bar{\xi}_{12}\cdots\bar{\xi}_{dd}|h_1 h_2\cdots h_d) = 0, \quad 1 \leqslant i < j \leqslant d. \tag{3.24}$$

This system has exactly the same form as the corresponding system of Eqs. (3.1) and (3.2), that we solved for the representation of collective states in $\mathscr{F}_c$, if we make $\bar{\xi}_{ij}$ and χ_{ij} correspond to w_{ij} and Δ_{ij}, respectively. Such a correspondence makes sense because χ_{ij} is defined in terms of $\bar{\xi}_{ij}$ in the same way as Δ_{ij} in terms of w_{ij}; as a consequence, the commutation relation of χ_{ij} with $\bar{\xi}_{kl}$ is just the same as that of Δ_{ij} with w_{kl}, given in Eq. (2.12), i.e.,

$$[\chi_{ij}, \bar{\xi}_{kl}] = \delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk}. \tag{3.25}$$

In close analogy with the corresponding assertion for $\mathscr{F}_c$, we may therefore state without proof the following result: $\mathscr{B}$ is the carrier space for the direct sum of IR's $[h_1 h_2 \cdots h_d]$ of $\mathfrak{U}(d)$, for which $h_1, h_2, \ldots, h_d$ may take any even values such that $h_1 \geqslant h_2 \geqslant \cdots \geqslant h_d \geqslant 0$. Each one of these IR's appears with a multiplicity one in the direct sum. The representation in $\mathscr{B}$ of the hws of the IR $[h_1 h_2 \cdots h_d]$ is given by

$$\begin{aligned} &(\bar{\xi}_{11}\bar{\xi}_{12}\cdots\bar{\xi}_{dd}|h_1 h_2\cdots h_d) \\ &\quad = B_{h_1 h_2 \cdots h_d} \prod_{i=1}^{d} (\bar{\xi}_{1,2,\ldots,i;1,2,\ldots,i})^{(1/2)(h_i - h_{i+1})}, \end{aligned} \tag{3.26}$$

where $h_{d+1} = 0$, $\bar{\xi}_{1,2,\ldots,i;1,2,\ldots,i}$ is defined in terms of the $\bar{\xi}_{jk}$ by a relation similar to Eq. (3.3), and $B_{h_1 h_2 \cdots h_d}$ is some normalization coefficient. The latter is determined by the condition

$$\begin{aligned} &\int \prod_{i<j=1}^{d} d\mu(\zeta_{ij})(h_1 h_2\cdots h_d|\bar{\xi}_{11}\bar{\xi}_{12}\cdots\bar{\xi}_{dd}) \\ &\quad \times(\bar{\xi}_{11}\bar{\xi}_{12}\cdots\bar{\xi}_{dd}|h_1 h_2\cdots h_d) = 1, \end{aligned} \tag{3.27}$$

where $d\mu(\zeta_{ij})$ is the usual Bargmann measure defined in Eq. (2.6), and Eq. (3.20) is used to express the $\bar{\xi}_{ij}$ in terms of the ζ_{ij}. Note that $B_{h_1 h_2 \cdots h_d}$ in general differs from the normalization coefficient $A_{h_1 h_2 \cdots h_d}$ of Eq (3.8). Equation (3.26) generalizes a well-known three-dimensional result[16–18] to an arbitrary number of dimensions.

We have established that there exists a one-to-one correspondence between the representation of collective states in $\mathscr{F}_c$ and that of ν-dimensional boson states in $\mathscr{B}$ (and therefore also between the collective and boson states themselves): Each basis function

$$\left\langle w_{11}w_{12}\cdots w_{dd} \left| \begin{matrix} h_1 h_2 \cdots h_d \\ (h) \end{matrix} \right\rangle \right.$$

of $\mathscr{F}_c$ is mapped onto the corresponding basis function

$$\left(\bar{\xi}_{11}\bar{\xi}_{12}\cdots\bar{\xi}_{dd} \left| \begin{matrix} h_1 h_2 \cdots h_d \\ (h) \end{matrix} \right. \right)$$

of $\mathscr{B}$. The existence of such a mapping will enable us to derive boson representations of $\mathscr{S}\!\!\!\!\mathscr{p}_c(2d,R)$. As functions transforming in a given way under $\mathscr{U}_c(d)$ are mapped onto functions transforming in the same way under $\mathfrak{U}(d)$, the generators of $\mathscr{U}_c(d)$ must be mapped onto those of $\mathfrak{U}(d)$:

$$\mathscr{C}^c = \mathbb{C}. \tag{3.28}$$

In Eq. (3.28), we have introduced a matrix notation for the generators of $\mathfrak{U}(d)$ similar to the one used for the $\mathscr{S}\!\!\!\!\mathscr{p}_c(2d,R)$ generators (in the same way, we shall use hereafter symmetrical $d \times d$ matrices $\bar{a}^\dagger$, $\bar{a}$, $\bar{\xi}$, and χ). It now remains to determine the boson representations of $\mathscr{D}^{c\dagger}$ and $\mathscr{D}^c$. Some results useful for that purpose will be established in the next section.

4. IRREDUCIBLE TENSORS $T[20]$ WITH RESPECT TO $\mathfrak{U}(d)$

The transformation properties of the operators $\mathscr{D}^{c\dagger}_{ij}$ with respect to $\mathscr{U}_c(d)$ follow from Eqs. (2.7) and (2.9). The relations

$$\mathscr{D}^{c\dagger}_{ij} = \mathscr{D}^{c\dagger}_{ji} \tag{4.1a}$$

and

$$[\mathscr{C}^{c}_{ij}, \mathscr{D}^{c\dagger}_{kl}] = \delta_{jk}\mathscr{D}^{c\dagger}_{il} + \delta_{jl}\mathscr{D}^{c\dagger}_{ik} \tag{4.1b}$$

show that the operators $\mathscr{D}^{c\dagger}_{ij}$ are the components of a symmetrical tensor transforming according to the IR $[2\dot{0}]$ of $\mathscr{U}_c(d)$. Due to Eq. (3.28), in the fermion–boson mapping they will be mapped onto the corresponding components of a tensor transforming in the same way under $\mathfrak{U}(d)$. We are thus led to study the most general form of an irreducible tensor $T[2\dot{0}]$ with respect to the group $\mathfrak{U}(d)$. Such a tensor, whose components will be denoted by T_{ij}, $i,j = 1,...,d$, has to satisfy relations similar to Eqs. (4.1a) and (4.1b), i.e.,

$$T_{ij} = T_{ji} \tag{4.2a}$$

and

$$[\mathfrak{C}_{ij}, T_{kl}] = \delta_{jk}T_{il} + \delta_{jl}T_{ik}. \tag{4.2b}$$

Irreducible tensors $T[2\dot{0}]$ with respect to $\mathfrak{U}(d)$ are easy to find. From their definition and commutation relations, it can be shown that the boson operators $\bar{a}^{\dagger}_{ij}$ satisfy Eqs. (4.2a) and (4.2b). More generally, the set of operators

$$T^{(m)}_{ij} = (\mathfrak{C}^m a^{\dagger})_{ij}, \quad m = 0,1,2,..., \tag{4.3}$$

form the components of a $T[2\dot{0}]$ tensor for any non-negative integer m value. The proof that they obey Eqs. (4.2a) and (4.2b) is most easily carried out by induction: Starting from the relation

$$T^{(m)}_{ij} = \sum_k \mathfrak{C}_{ik}T^{(m-1)}_{kj}, \tag{4.4}$$

the symmetry and commutation properties of $T^{(m)}_{ij}$ can be shown to derive from those of $T^{(m-1)}_{kj}$. With the set of operators $T^{(m)}_{ij}$, we have almost exhausted all the possible $T[2\dot{0}]$ tensors because it can be proved that any irreducible tensor $T[2\dot{0}]$ is a linear combination of the tensors $T^{(m)}_{ij}$, $m = 0,1,...,$

$$T_{ij} = \sum_{m=0}^{\infty} p_m(\Phi_1,...,\Phi_d)T^{(m)}_{ij}, \tag{4.5}$$

whose coefficients $p_m(\Phi_1,...,\Phi_d)$ are analytic functions of the $\mathfrak{U}(d)$ Casimir operators

$$\Phi_k = \operatorname{tr}\mathfrak{C}^k = \sum_{i_1\cdots i_k} \mathfrak{C}_{i_1i_2}\mathfrak{C}_{i_2i_3}\cdots\mathfrak{C}_{i_ki_1}, \quad k = 1,...,d. \tag{4.6}$$

The Eq. (4.5) demonstration is detailed in Appendix A.

Equation (4.5) gives the most general form of a $T[2\dot{0}]$ irreducible tensor as an infinite series. It may be converted into a sum of a finite number of terms by noting that the first d operators $T^{(m)}_{ij}$, corresponding to $m = 0,1,...,d-1$, form a basis for the $T[2\dot{0}]$ tensors. This means that they are linearly independent and that any $T^{(k)}_{ij}$ operator, for which $k \geqslant d$, can be written as their linear combination. The coefficients of the linear combinations are functions of the Casimir operators $\Phi_1,...,\Phi_d$ of $\mathfrak{U}(d)$.

Before starting to prove the above-mentioned result, one has to remember that any boson operator is entirely determined whenever one knows its matrix elements between arbitrary boson states classified according to the chain (3.17). When applied to a boson state transforming according to the IR $[h_1\cdots h_i\cdots h_d]$ of $\mathfrak{U}(d)$, any $T[2\dot{0}]$ tensor can only lead to boson states transforming according to an IR $[h_1\cdots h_i + 2\cdots h_d]$ for some $1 \leqslant i \leqslant d$. The Kronecker product $[h_1h_2\cdots h_d] \otimes [2\dot{0}]$ is indeed given by

$$[h_1h_2\cdots h_d] \otimes [2\dot{0}] = \sum_{i=1}^{d} \oplus [h_1\cdots h_i + 2\cdots h_d], \tag{4.7}$$

where the multiplicity of each IR appearing on the right-hand side is equal to 1. According to the Wigner–Eckart theorem,[19] all the matrix elements of the components of $T[2\dot{0}]$ between boson states transforming according to $[h_1\cdots h_i\cdots h_d]$ and $[h_1\cdots h_i + 2\cdots h_d]$ are determined by a single reduced matrix element. Alternatively, one may use the matrix element of an appropriate component of the tensor between hws of the IR's.[18] In the present case, the appropriate component is T_{ii} because it has to increase the eigenvalue of $\mathfrak{C}_{ii}$ by two and to leave those of $\mathfrak{C}_{jj}$, $j \neq i$, unchanged. We therefore conclude that any $T[2\dot{0}]$ tensor is entirely determined by the knowledge of the matrix elements $(h_1\cdots h_i + 2\cdots h_d | T_{ii} | h_1\cdots h_i\cdots h_d)$, $i = 1,...,d$, corresponding to an arbitrary IR $[h_1\cdots h_d]$ of $\mathfrak{U}(d)$.

Now let us show the linear independence of the operators $T^{(m)}_{ij}$, $m = 0,1,...,d-1$. This amounts to proving that the relation

$$\sum_{m=0}^{d-1} \alpha_m(\Phi_1,...,\Phi_d)T^{(m)}_{ij} = 0, \tag{4.8}$$

where α_m is some function of the Casimir operators, implies that

$$\alpha_m(\Phi_1,...,\Phi_d) = 0, \quad m = 0,1,...,d-1. \tag{4.9}$$

From the above Eq. (4.8) is equivalent to the following system of d equations in the d unknowns α_m, $m = 0,1,...,d-1$:

$$\sum_{m=0}^{d-1} \alpha_m(\phi_1,...,\phi_d) \times (h_1\cdots h_i\cdots h_d | T^{(m)}_{ii} | h_1\cdots h_i - 2\cdots h_d) = 0, \quad i = 1,...,d, \tag{4.10}$$

where $\phi_1,...,\phi_d$ are the eigenvalues of $\Phi_1,...,\Phi_d$ corresponding to the IR $[h_1\cdots h_i\cdots h_d]$.

The matrix elements appearing on the left-hand side of Eq. (4.10) assume a very simple form because both bra and ket are of highest weight. Starting from the relation

$$(h_1\cdots h_i\cdots h_d | T^{(m)}_{ii} | h_1\cdots h_i - 2\cdots h_d) = \left(h_1\cdots h_i\cdots h_d \middle| \sum_j \mathfrak{C}_{ij}T^{(m-1)}_{ji} \middle| h_1\cdots h_i - 2\cdots h_d\right), \tag{4.11}$$

we note that on the right-hand side the terms with $j < i$ disappear because $\mathfrak{C}_{ij}$ gives zero when acting on the bra, while those with $j > i$ can be transformed into

$$\left(h_1\cdots h_i\cdots h_d \middle| \sum_{j>i} \{T^{(m-1)}_{ji}\mathfrak{C}_{ij} + [\mathfrak{C}_{ij}, T^{(m-1)}_{ji}]\} \middle| h_1\cdots h_i - 2\cdots h_d\right), \tag{4.12}$$

where $\mathbb{C}_{ij}$ gives zero when acting on the ket. After a straightforward calculation, we get the relation

$$(h_1\cdots h_i\cdots h_d|T_{ii}^{(m)}|h_1\cdots h_i-2\cdots h_d)$$
$$=(h_1\cdots h_i\cdots h_d|(\mathbb{C}_{ii}+d-i)T_{ii}^{(m-1)}|h_1\cdots h_i-2\cdots h_d), \tag{4.13}$$

from which

$$(h_1\cdots h_i\cdots h_d|T_{ii}^{(m)}|h_1\cdots h_i-2\cdots h_d)$$
$$=\lambda_i^m(h_1\cdots h_i\cdots h_d|\bar{a}_{ii}^{\dagger}|h_1\cdots h_i-2\cdots h_d), \tag{4.14}$$

where

$$\lambda_i=h_i+d-i. \tag{4.15}$$

Taking into account Eq. (4.14) and the fact that $(h_1\cdots h_i\cdots h_d|\bar{a}_{ii}^{\dagger}|h_1\cdots h_i-2\cdots h_d)\neq 0$, the system of equations (4.10) becomes

$$\sum_{m=0}^{d-1}\alpha_m(\phi_1,\ldots,\phi_d)\lambda_i^m=0,\quad i=1,\ldots,d. \tag{4.16}$$

Since $\lambda_1>\lambda_2>\cdots>\lambda_d$ as a consequence of the inequalities $h_1\geqslant h_2\geqslant\cdots\geqslant h_d$, the determinant of the coefficients α_m,

$$D(\lambda_1,\ldots,\lambda_d)=\begin{vmatrix}1&\lambda_1&\lambda_1^2&\cdots&\lambda_1^{d-1}\\1&\lambda_2&\lambda_2^2&\cdots&\lambda_2^{d-1}\\\vdots&\vdots&\vdots&&\vdots\\1&\lambda_d&\lambda_d^2&\cdots&\lambda_d^{d-1}\end{vmatrix}$$
$$=\prod_{i>j}(\lambda_i-\lambda_j), \tag{4.17}$$

is different from zero. The solution of the system (4.16) is therefore $\alpha_m(\phi_1,\ldots,\phi_d)=0$, $m=0,1,\ldots,d-1$, for any IR $[h_1h_2\cdots h_d]$. Consequently, Eq. (4.9) is satisfied.

It remains to show that for any $k\geqslant d$, the $T_{ij}^{(k)}$ operator is linearly dependent upon $T_{ij}^{(m)}$, $m=0,1,\ldots,d-1$, i.e.,

$$T_{ij}^{(k)}=\sum_{m=0}^{d-1}\alpha_m^{(k)}(\Phi_1,\ldots,\Phi_d)T_{ij}^{(m)},\quad k\geqslant d, \tag{4.18}$$

where $\alpha_m^{(k)}$ is some function of the Casimir operators. The proof of Eq. (4.18), which is quite similar to that of the linear independence of $T_{ij}^{(m)}$, $m=0,1,\ldots,d-1$, is given in Appendix B together with the explicit form of the $\alpha_m^{(k)}$ coefficients.

Finally, by combining Eqs. (4.5) and (4.18), any irreducible tensor $T[20]$ can be written in terms of $T^{(m)}$, $m=0,1,\ldots,d-1$, as

$$T_{ij}=\sum_{m=0}^{d-1}f_m(\Phi_1,\ldots,\Phi_d)(\mathbb{C}^m a^{\dagger})_{ij}, \tag{4.19}$$

where f_m is some function of the $\mathfrak{U}(d)$ Casimir operators.

Other expansions of T_{ij} may be obtained by considering other bases for the $T[20]$ tensors. Two of them are worth mentioning. The first one uses as basis the set of d operators

$$(a^{\dagger}\tilde{\mathbb{C}}^m)_{ij},\quad m=0,1,\ldots,d-1, \tag{4.20}$$

where $\tilde{\mathbb{C}}$ is the transpose of $\mathbb{C}$. In a way quite similar to the proof of Eq. (4.19), it can be shown that any irreducible tensor $T[20]$ may be written as

$$T_{ij}=\sum_{m=0}^{d-1}(a^{\dagger}\tilde{\mathbb{C}}^m)_{ij}\,g_m(\tilde{\Phi}_1,\ldots,\tilde{\Phi}_d), \tag{4.21}$$

where g_m is some function of the operators

$$\tilde{\Phi}_k=\operatorname{tr}\tilde{\mathbb{C}}^k=\sum_{i_1i_2\cdots i_k}\mathbb{C}_{i_2i_1}\mathbb{C}_{i_3i_2}\cdots\mathbb{C}_{i_1i_k},\quad k=1,\ldots,d, \tag{4.22}$$

which are known[20] to form a possible choice for the $\mathfrak{U}(d)$ Casimir operators. In the derivation of Eq. (4.21), use is made of the following equations:

$$(h_1\cdots h_i+2\cdots h_d|(\bar{a}^{\dagger}\tilde{\mathbb{C}}^m)_{ii}|h_1\cdots h_i\cdots h_d)$$
$$=\tilde{\lambda}_i^m(h_1\cdots h_i+2\cdots h_d|a_{ii}^{\dagger}|h_1\cdots h_i\cdots h_d) \tag{4.23}$$

and

$$\tilde{\lambda}_i=h_i-i+1, \tag{4.24}$$

which are the counterparts of Eqs. (4.14) and (4.15), respectively.

The second alternative expansion for T_{ij} uses as basis the set of d operators

$$[\Phi_{m+1},\bar{a}_{ij}^{\dagger}],\quad m=0,1,\ldots,d-1, \tag{4.25}$$

already considered in Ref. 10. This is briefly discussed in Appendix C and will not be used in the following sections for reasons detailed in the same appendix.

We can now proceed to the boson representations of $\mathscr{S}\!\mathscr{p}_c(2d,R)$, beginning in the next section with Dyson representation.

5. DYSON REPRESENTATION OF $\mathscr{S}\!\mathscr{p}_c(2d,R)$

The simplest way to explicitly map the functional Hilbert spaces $\mathscr{F}_c$ and $\mathscr{B}$ onto one another consists in equating the complex variables and differential operators defined in one space with the corresponding ones defined in the other space, i.e.,

$$\mathbf{w}=\bar{\xi},$$
$$\Delta=\chi. \tag{5.1}$$

From Eqs. (2.12) and (3.25), it is clear that such a procedure does preserve the commutation relations. However, it violates Hermiticity since χ_{ij} is the Hermitian conjugate of $\bar{\xi}_{ij}$ in $\mathscr{B}$ where as $\{\Delta[\mathbf{w}\Delta+(n-d-1)\mathbf{I}]\}_{ij}$ (instead of Δ_{ij}) is the Hermitian conjugate of w_{ij} in $\mathscr{F}_c$ as it can be seen from Eq. (2.11).

Let us perform the transformation (5.1) followed by the replacement

$$\bar{\xi}\rightarrow a^{\dagger},$$
$$\chi\rightarrow\bar{a}, \tag{5.2}$$

on the representation (2.11) of the $\mathscr{S}\!\mathscr{p}_c(2d,R)$ generators in $\mathscr{F}_c$. We obtain the following boson representation of $\mathscr{S}\!\mathscr{p}_c(2d,R)$:

$$\mathscr{D}^{c\dagger}=a^{\dagger}, \tag{5.3a}$$
$$\mathscr{D}^{c}=\bar{a}[\mathbb{C}+(n-d-1)\mathbf{I}]=[\tilde{\mathbb{C}}+n\mathbf{I}]\bar{a}, \tag{5.3b}$$
$$\mathscr{C}^{c}=\mathscr{C}^{c}+(n/2)\mathbf{I},\quad \mathscr{C}^{c}=\mathbb{C}, \tag{5.3c}$$

which is nothing else than the finite non-Hermitian Dyson[12] expansion. It can be easily checked that the right-hand sides of Eqs. (5.3a), (5.3b), and (5.3c) do indeed satisfy the commutation relations (2.9) of the $\mathscr{S}\!\mathscr{p}_c(2d,R)$ generators, but not

their Hermiticity properties, given in Eq. (2.8), since the condition $(\mathscr{D}^{c\dagger})^{\dagger} = \mathscr{D}^{c}$ is not preserved.

In the present section, we have established through the use of Eqs. (5.1) and (5.2) that the Barut representation of $\mathscr{S}\!\!\!/_c(2d,R)$ is essentially equivalent to its Dyson representation. By restoring the Hermiticity properties of the $\mathscr{S}\!\!\!/_c(2d,R)$ generators, we can now go from the Dyson representation to the HP one. For such a purpose, we need to consider a more complicated mapping of $\mathscr{B}$ onto $\mathscr{F}_c$ than that defined in Eq. (5.1). We proceed to study it in detail in the next section.

6. HOLSTEIN–PRIMAKOFF REPRESENTATION OF $\mathscr{S}\!\!\!/_c(2d,R)$

Since the mapping $\bar{\xi}\to\mathbf{w}$, $\chi\to\Delta$ preserves the commutation relations, it is a canonical transformation. There exists therefore an operator U such that

$$\mathbf{w} = U\bar{\xi}U^{-1}, \tag{6.1a}$$

$$\Delta = U\chi U^{-1}. \tag{6.1b}$$

Let us impose the two following conditions on the operator U: (i) U maps $\mathfrak{U}(d)$ onto $\mathscr{U}_c(d)$, so that Eq. (3.28) is satisfied or in other words

$$\mathbf{w}\Delta = \bar{\xi}\chi; \tag{6.2}$$

(ii) U preserves the Hermiticity properties, so that

$$\mathbf{w}^{\dagger} = \Delta[\mathbf{w}\Delta + (n-d-1)\mathbf{I}], \tag{6.3}$$

whenever

$$\bar{\xi}^{\dagger} = \chi. \tag{6.4}$$

Consequently U is not a unitary operator.

When introducing Eq. (6.1) into Eq. (6.2), we obtain that U must fulfill the following condition:

$$U\bar{\xi}\chi U^{-1} = \bar{\xi}\chi. \tag{6.5}$$

It must therefore commute with the representation in $\mathscr{B}$ of all the generators $\mathfrak{C}_{ij}$ of $\mathfrak{U}(d)$. Consequently it must be a function

$$U = U(\Phi_1,\ldots,\Phi_d) \tag{6.6}$$

of the $\mathfrak{U}(d)$ Casimir operators Φ_k (or more exactly of their representation in $\mathscr{B}$ that we denote by the same symbol). We shall assume that this function is real so that U is a Hermitian operator, i.e.,

$$U = U^{\dagger}. \tag{6.7}$$

Actually we shall show hereafter by explicitly constructing U that such an assumption is consistent with the conditions (6.2) and (6.3).

It now remains to combine Eq. (6.1) with Eq. (6.3). By taking Eqs. (6.2), (6.4), (6.5), and (6.7) into account, we obtain

$$U^{-1}\chi U = U\chi[\bar{\xi}\chi + (n-d-1)\mathbf{I}]U^{-1} \tag{6.8}$$

or

$$U^{-2}\chi U^{2} = \chi[\bar{\xi}\chi + (n-d-1)\mathbf{I}]. \tag{6.9}$$

Alternatively we may consider the Hermitian conjugate of Eq. (6.9), which can be written as

$$U^{2}\bar{\xi}U^{-2} = [\bar{\xi}\chi + (n-d-1)\mathbf{I}]\bar{\xi}. \tag{6.10}$$

It is obvious that Eq. (6.10) does not entirely determine the function $U(\Phi_1,\ldots,\Phi_d)$: In the equation U^2 occurs at the same times as U^{-2} so that it will be determined up to a constant factor; moreover, as the equation involves U^2 instead of U, the latter will be determined up to a sign factor. We shall take advantage of this freedom hereafter so as to fix the normalization of U in a convenient way.

Before looking for the explicit expression of U, let us first find how the representation in $\mathscr{B}$ of the $\mathscr{S}\!\!\!/_c(2d,R)$ generators is transformed under the canonical transformation (6.1). From Eqs. (2.11), (6.1a), (6.2), and (6.8), we obtain the following relations:

$$\mathscr{D}^{c\dagger} = U\bar{\xi}U^{-1},$$

$$\mathscr{D}^{c} = U^{-1}\chi U, \tag{6.11}$$

and

$$\mathscr{E}^{c} = \mathscr{C}^{c} + (n/2)\mathbf{I}, \quad \mathscr{C}^{c} = \bar{\xi}\chi.$$

$\mathscr{D}^{c\dagger}$ and $\mathscr{D}^{c}$ may therefore be obtained from $\bar{\xi}$ and χ by acting with U and U^{-1}, respectively.

When we carry out the replacement (5.2) in Eq. (6.11), we obtain the following boson representation of $\mathscr{S}\!\!\!/_c(2d,R)$:

$$\mathscr{D}^{c\dagger} = U\bar{a}^{\dagger}U^{-1}, \tag{6.12a}$$

$$\mathscr{D}^{c} = U^{-1}\bar{a}U, \tag{6.12b}$$

$$\mathscr{E}^{c} = \mathscr{C}^{c} + (n/2)\mathbf{I}, \quad \mathscr{C}^{c} = \mathfrak{C} = \bar{a}^{\dagger}\bar{a}, \tag{6.12c}$$

which is nothing else than the Hermitian HP[9] representation. In Eq. (6.12), U is a function of the $\mathfrak{U}(d)$ Casimir operators satisfying the condition

$$U^{2}a^{\dagger}U^{-2} = [\mathfrak{C} + (n-d-1)\mathbf{I}]a^{\dagger}. \tag{6.13}$$

Let us now turn to the determination of the operator U explicit form. From (6.6) it follows that U is diagonal in the basis

$$\begin{pmatrix} h_1h_2\cdots h_d \\ (h) \end{pmatrix}$$

and that its eigenvalues only depend upon the $\mathfrak{U}(d)$ IR labels:

$$U\begin{pmatrix} h_1h_2\cdots h_d \\ (h) \end{pmatrix} = u(h_1,h_2,\ldots,h_d)\begin{pmatrix} h_1h_2\cdots h_d \\ (h) \end{pmatrix}. \tag{6.14}$$

Both sides of Eq. (6.13) are irreducible tensors $T[20\dot{0}]$ with respect to $\mathfrak{U}(d)$. By taking the matrix element of their ii component between the bra $(h_1\cdots h_i\cdots h_d|$ and the ket $|h_1\cdots h_i - 2\cdots h_d)$ and using Eqs. (4.14) and (4.15), we obtain the following recursion relation for $u(h_1,h_2,\ldots,h_d)$:

$$u^{2}(h_1,h_2,\ldots,h_d)[u(h_1,\ldots,h_i-2,\ldots,h_d)]^{-2}$$
$$= \lambda_i + n - d - 1 = h_i + n - i - 1. \tag{6.15}$$

Its solution can be written as

$$u(h_1,h_2,\ldots,h_d) = \left\{\prod_{i=1}^{d} 2^{h_i/2}\Gamma\left(\frac{h_i+n-i+1}{2}\right)\times\left[\Gamma\left(\frac{n-i+1}{2}\right)\right]^{-1}\right\}^{1/2}, \tag{6.16}$$

if we choose $u(h_1,h_2,\ldots,h_d)$ to be positive and $u(0,0,\ldots,0)$ to be equal to 1 (so that the vacuum state is left invariant by the operator U). By expressing $h_1,h_2,\ldots,h_d$ in terms of $\phi_1,\phi_2,\ldots,\phi_d$, we can in principle convert u into a function of the ϕ_i's and then replace the latter by the corresponding

operators Φ_i so as to obtain the explicit form of the operator U.

Equation (6.12) does not correspond to the usual form of the HP representation,[9] which involves either an expansion into boson operators or the square root of some boson operator. Let us now proceed to derive such equivalent expressions. From Eq. (4.19) $\mathscr{D}^{c\dagger}$ may be written as a linear combination of the operators $\mathbb{C}^m a^\dagger$, $m = 0,1,...,d-1$,

$$\mathscr{D}^{c\dagger} = \sum_{m=0}^{d-1} x_m(\Phi_1,...,\Phi_d)\mathbb{C}^m a^\dagger. \tag{6.17a}$$

The coefficients $x_m(\Phi_1,...,\Phi_d)$ of this linear combination must satisfy the equation

$$\sum_{m=0}^{d-1} x_m(\Phi_1,...,\Phi_d)\mathbb{C}^m a^\dagger = U a^\dagger U^{-1}, \tag{6.18}$$

from which we obtain the following system of linear equations:

$$\sum_{m=0}^{d-1} x_m(\phi_1,...,\phi_d)\lambda_i^m = (\lambda_i + n - d - 1)^{1/2}, \quad i = 1,...,d, \tag{6.19}$$

by taking matrix elements between hws and using Eqs. (4.14), (4.15), and (6.16). The solution of such a system is given by

$$x_m = [D(\lambda_1,...,\lambda_d)]^{-1}\begin{vmatrix} 1 & \lambda_1 & \cdots & \lambda_1^{m-1} & (\lambda_1+n-d-1)^{1/2} & \lambda_1^{m+1} & \cdots & \lambda_1^{d-1} \\ 1 & \lambda_2 & & \lambda_2^{m-1} & (\lambda_2+n-d-1)^{1/2} & \lambda_2^{m+1} & & \lambda_2^{d-1} \\ \vdots & & & & & & & \vdots \\ 1 & \lambda_d & & \lambda_d^{m-1} & (\lambda_d+n-d-1)^{1/2} & \lambda_d^{m+1} & \cdots & \lambda_d^{d-1} \end{vmatrix}$$
$$= (-)^m \sum_{i=1}^{d} \left\{ (\lambda_i + n - d - 1)^{1/2} \sum_{\substack{j_1<j_2<\cdots<j_m \\ j_1,...,j_m \neq i}} \left[\frac{1}{(\lambda_{j_1}-\lambda_i)\cdots(\lambda_{j_m}-\lambda_i)} \prod_{k\neq i,j_1,...,j_m} \frac{\lambda_k}{\lambda_k-\lambda_i} \right] \right\}, \tag{6.20}$$

where $D(\lambda_1,...,\lambda_d)$ is defined in Eq. (4.17). When we combine Eq. (6.17a) with the following two equations,

$$\mathscr{D}^c = \sum_{m=0}^{d-1} \bar{a}\mathbb{C}^m x_m(\Phi_1,...,\Phi_d) \tag{6.17b}$$

and

$$\mathscr{C}^c = \mathscr{C}^c + (n/2)\mathbf{I}, \quad \mathscr{C}^c = \mathbb{C} = \bar{a}^\dagger \bar{a}, \tag{6.17c}$$

we obtain the HP representation of $\mathscr{S}\!\!\!/_c(2d,R)$ as a finite expansion into boson operators.

An alternative compact form of Eq. (6.17) may be found by noting that the matrix $\mathbf{X}$, defined by

$$\mathbf{X} = \sum_{m=0}^{d-1} x_m(\Phi_1,...,\Phi_d)\mathbb{C}^m, \tag{6.21}$$

satisfies the following equation:

$$\mathbf{X}^2 = \mathbb{C} + (n-d-1)\mathbf{I}. \tag{6.22}$$

To demonstrate Eq. (6.22), let us start from the relation

$$\mathbf{X}\bar{a}^\dagger = U\bar{a}^\dagger U^{-1}, \tag{6.23}$$

which results from Eqs. (6.12a), (6.17a), and (6.21), and let us multiply it from the left by U and from the right by U^{-1}. When using Eqs. (6.6) and (6.23) the left-hand side is transformed into

$$U\mathbf{X}\bar{a}^\dagger U^{-1} = \mathbf{X}U\bar{a}^\dagger U^{-1} = \mathbf{X}^2\bar{a}^\dagger, \tag{6.24}$$

while the right-hand side is simply given by Eq. (6.13). The resulting relation can therefore be written as

$$\mathbf{X}^2 a^\dagger = [\mathbb{C} + (n-d-1)\mathbf{I}]a^\dagger. \tag{6.25}$$

By taking the square of Eq. (6.21), we can write $\mathbf{X}^2$ as

$$\mathbf{X}^2 = \sum_{m=0}^{2d-2} \bar{x}'_m(\Phi_1,...,\Phi_d)\mathbb{C}^m. \tag{6.26}$$

Since for any positive integer m value, $\mathbb{C}^m$ is the sum of a $\mathfrak{u}(d)$ scalar, $d^{-1}(\mathrm{tr}\,\mathbb{C}^m)\mathbf{I} = d^{-1}\Phi_m\mathbf{I}$, and an irreducible tensor $T[1\dot{0}-1]$, $\mathbb{C}^m - d^{-1}(\mathrm{tr}\mathbb{C}^m)\mathbf{I}$, we can decompose $\mathbf{X}^2$ into a scalar

$$d^{-1}(\mathrm{tr}\,\mathbf{X}^2)\mathbf{I} = d^{-1}\sum_{m=0}^{2d-2} \bar{x}'_m(\Phi_1,...,\Phi_d)\Phi_m\mathbf{I}, \tag{6.27}$$

and a $T[1\dot{0}-1]$ tensor

$$\mathbf{X}^2 - d^{-1}(\mathrm{tr}\,\mathbf{X}^2)\mathbf{I} = \sum_{m=1}^{2d-2} \bar{x}'_m(\Phi_1,...,\Phi_d)[\mathbb{C}^m - d^{-1}(\mathrm{tr}\,\mathbb{C}^m)\mathbf{I}]. \tag{6.28}$$

From Ref. 21, it is known that only $d-1$ independent irreducible $T[1\dot{0}-1]$ tensors can be constructed from the $\mathfrak{u}(d)$ generators. In Eq. (6.28), the operators $[\mathbb{C}^m - d^{-1}](\mathrm{tr}\,\mathbb{C}^m)\mathbf{I}]$, $m = d,...,2d-2$, can therefore be expressed as linear combinations of $[\mathbb{C}^m - d^{-1}(\mathrm{tr}\mathbb{C}^m)\mathbf{I}]$, $m = 1,..,d-1$, with coefficients depending upon the $\mathfrak{u}(d)$ Casimir operators. Consequently, Eqs. (6.28) and (6.26), respectively, become

$$\mathbf{X}^2 - d^{-1}(\mathrm{tr}\,\mathbf{X}^2)\mathbf{I} = \sum_{m=1}^{d-1} \bar{x}_m(\Phi_1,...,\Phi_d)[\mathbb{C}^m - d^{-1}(\mathrm{tr}\,\mathbb{C}^m)\mathbf{I}] \tag{6.29}$$

and

$$\mathbf{X}^2 = \sum_{m=0}^{d-1} \bar{x}_m(\Phi_1,...,\Phi_d)\mathbb{C}^m, \tag{6.30}$$

where

$$\bar{x}_0(\Phi_1,...,\Phi_d) = d^{-1}\sum_{m=0}^{2d-2} \bar{x}'_m(\Phi_1,...,\Phi_d)\Phi_m - d^{-1} \times \sum_{m=1}^{d-1} \bar{x}_m(\Phi_1,...,\Phi_d)\Phi_m. \tag{6.31}$$

When one introduces Eq. (6.30) into Eq. (6.25), the linear independence of the irreducible tensors $\mathbb{C}^m a^\dagger$, $m = 0,1,...,d-1$, proved in Sec. 4, enables one to determine the values of the $\bar{x}_m$ coefficients. For $d > 1$, one obtains

$$\bar{x}_0 = n - d - 1,$$
$$\bar{x}_1 = 1, \tag{6.32}$$
$$\bar{x}_m = 0, \quad m = 2,...,d-1,$$

while for $d = 1$

$$\bar{x}_0 = \mathbb{C}_{11} + n - d - 1 = \Phi_1 + n - d - 1. \tag{6.33}$$

This completes the proof of Eq. (6.22).

As a consequence of Eq. (6.22), $\mathbf{X}$ may be represented by the square root of $\mathbb{C} + (n-d-1)\mathbf{I}$, which must be understood as a compact form for the finite expansion given in Eq. (6.21). The HP representation of $\mathcal{S}p_c(2d,R)$ may therefore be written as

$$\mathcal{D}^{c\dagger} = [\mathbb{C} + (n-d-1)\mathbf{I}]^{1/2}\bar{a}^\dagger,$$
$$\mathcal{D}^{c} = \bar{a}[\mathbb{C} + (n-d-1)\mathbf{I}]^{1/2}, \tag{6.34}$$
$$\mathcal{C}^c = \mathcal{C}^c + (n/2)\mathbf{I}, \quad \mathcal{C}^c = \mathbb{C} = \bar{a}^\dagger \bar{a}.$$

Instead of starting from an expansion of $\mathcal{D}^{c\dagger}$ in terms of $\mathbb{C}^m\bar{a}^\dagger$, it is also possible to use Eq. (4.21) in order to expand $\mathcal{D}^{c\dagger}$ in terms of $\bar{a}^\dagger\tilde{\mathbb{C}}^m$. We then obtain the following alternative form of Eqs. (6.17) and (6.34):

$$\mathcal{D}^{c\dagger} = \sum_{m=0}^{d-1} \bar{a}^\dagger \tilde{\mathbb{C}}^m y_m(\tilde{\Phi}_1,...,\tilde{\Phi}_d) = \bar{a}^\dagger [\tilde{\mathbb{C}} + n\mathbf{I}]^{1/2}$$
$$\mathcal{D}^{c} = \sum_{m=0}^{d-1} y_m(\tilde{\Phi}_1,...,\tilde{\Phi}_d)\tilde{\mathbb{C}}^m\bar{a} = [\tilde{\mathbb{C}} + n\mathbf{I}]^{1/2}\bar{a}, \tag{6.35}$$
$$\mathcal{C}^c = \mathcal{C}^c + \frac{n}{2}\mathbf{I}, \quad \mathcal{C}^c = \mathbb{C} = \bar{a}^\dagger\bar{a},$$

where the eigenvalues of $y_m(\tilde{\Phi}_1,...,\tilde{\Phi}_d)$ are given by

$$y_m = (-1)^m \sum_{i=1}^{d} \left\{ (\tilde{\lambda}_i + n)^{1/2} \sum_{\substack{j_1<j_2<\cdots<j_m \\ j_1,\ldots,j_m \neq i}} \times \left[\frac{1}{(\tilde{\lambda}_{j_1} - \tilde{\lambda}_i)\cdots(\tilde{\lambda}_{j_m} - \tilde{\lambda}_i)} \prod_{k \neq i, j_1,\ldots,j_m} \frac{\tilde{\lambda}_k}{\tilde{\lambda}_k - \tilde{\lambda}_i} \right] \right\}. \tag{6.36}$$

Equations (6.17), (6.34), and (6.35) generalize to any number of dimensions the results obtained in Refs. 3 and 22–24 for the one-dimensional case, and in Ref. 25 for the two-dimensional one.

As can be seen in Eqs. (6.17), (6.34), and (6.35), the $\mathfrak{U}(\nu)$ group makes its appearance in a highly nonlinear way in the $\mathcal{S}p_c(2d,R)$ framework. However, when the number of particles becomes very large, this nonlinear dependence becomes much simpler. In this case indeed, the term $n\mathbf{I}$ becomes dominant in the square root $[\mathbb{C} + (n-d-1)\mathbf{I}]^{1/2}$ or $[\tilde{\mathbb{C}} + n\mathbf{I}]^{1/2}$, so that $n^{-1/2}\mathcal{D}^{c\dagger}$ and $n^{-1/2}\mathcal{D}^c$ behave as $\bar{a}^\dagger$ and $\bar{a}$, respectively:

$$n^{-1/2}\mathcal{D}^{c\dagger} \overset{n\to\infty}{\longrightarrow} \bar{a}^\dagger, \tag{6.37}$$
$$n^{-1/2}\mathcal{D}^{c} \overset{n\to\infty}{\longrightarrow} \bar{a}.$$

When the number of particles is not very large, it is still possible to express the boson operators, and consequently the $\mathfrak{U}(\nu)$ generators, in terms of the generators of the dynamical group $\mathcal{S}p_c(2d,R)$ by inverting Eq. (6.34) or (6.35). In compact form, we obtain

$$\bar{a}^\dagger = [\mathcal{C}^c + (n-d-1)\mathbf{I}]^{-1/2}\mathcal{D}^{c\dagger} = \mathcal{D}^{c\dagger}[\tilde{\mathcal{C}}^c + n\mathbf{I}]^{-1/2}, \tag{6.38}$$
$$\bar{a} = \mathcal{D}^c[\mathcal{C}^c + (n-d-1)\mathbf{I}]^{-1/2} = [\tilde{\mathcal{C}}^c + n\mathbf{I}]^{-1/2}\mathcal{D}^c,$$

and

$$\mathbb{C} = \bar{a}^\dagger\bar{a} = \mathcal{C}^c, \tag{6.39}$$

where $[\mathcal{C}^c + (n-d-1)\mathbf{I}]^{-1/2}([\tilde{\mathcal{C}}^c + n\mathbf{I}]^{-1/2})$ is defined as the inverse of $[\mathcal{C}^c + (n-d-1)\mathbf{I}]^{1/2}([\tilde{\mathcal{C}}^c + n\mathbf{I}]^{1/2})$, for which an expansion in terms of $(\mathcal{C}^c)^m$, $m = 0,1,...,d-1$ $[(\tilde{\mathcal{C}}^c)^m$, $m = 0,1,...,d-1]$ could be easily found.

Equation (6.38) can be interpreted in the following way. Since the first-order Casimir operator of $\mathcal{U}_c(d)$ can be considered as the collective part $\mathcal{H}^c_{\rm osc}$ of a harmonic oscillator Hamiltonian in the microscopic model and that of $\mathfrak{U}(d)$ as a ν-dimensional harmonic oscillator Hamiltonian $\mathfrak{H}_{\rm osc}$ in a boson macroscopic model (neglecting the zero point energy in both cases), it follows from Eqs. (3.18) and (6.39) that in the fermion–boson mapping (6.38), $\mathcal{H}^c_{\rm osc}$ is mapped onto $2\,\mathfrak{H}_{\rm osc}$. Equation (6.38) then gives the unitary representation in quantum mechanics of the classical canonical transformation relating $\mathcal{H}^c_{\rm osc}$ and $\mathfrak{H}_{\rm osc}$. It generalizes to an arbitrary number of dimensions the two-dimensional result which was derived in Ref. 6 starting from the Dzublik[7]–Zickendraht[8] transformation.

7. CONCLUDING REMARKS

In the present paper, we have shown through the study of the Dyson and HP representations of their dynamical group that O(n) invariant microscopic collective states in d dimensions can be described in terms of $\nu = \frac{1}{2}d(d+1)$ boson creation and annihilation operators which can be combined to form the generators of a $\mathfrak{U}(\nu)$ group. In three dimensions, a $\mathfrak{U}(6)$ group therefore makes its appearance in the O(n) invariant microscopic model, and consequently it is possible to establish a relationship between the latter and the IBM.

However, as has already been noticed by Vanagas,[26] the microscopic model does not reduce to the IBM because, in spite of the presence of a U(6) group in both pictures, there remain some important differences between them. In the IBM, the U(6) group is a symmetry group, which implies that the total number of bosons is conserved. In the microscopic model $\mathfrak{U}(6)$ is a symmetry group only when the nucleons interact through harmonic oscillator forces as the canonical transformation relating the oscillator Hamiltonians of the microscopic model and the boson model enables us to carry the U(6) symmetry from the latter to the former; such a procedure is not possible for more general Hamiltonians. In the microscopic model, the total number of bosons is therefore not a good quantum number for an arbitrary interaction.

Moreover, the most general Hamiltonian of the IBM only contains one- and two-body terms whereas, in the microscopic model, the collective Hamiltonian is a function of the $\mathcal{S}p_c(2d,R)$ generators—and therefore of the boson oper-

ators—which is not restricted to a low degree polynomial.

More importantly, there is a marked difference between the microscopic model and the IBM coming from the physical interpretation of the bosons. In the IBM, they are assumed to represent coupled pairs of nucleons or holes outside closed shells so that their number is fixed by that of the active nucleons. In the microscopic model, the microscopic structure of the bosons is more complicated and arises from the consideration of the whole set of A nucleons, including both the active ones and those belonging to the closed shell core. Their number is not fixed and might even grow to infinity, although for practical purposes one would be restricted to small values.

The boson expansions considered in the present paper are exact ones, which means that starting from fermion states satisfying the Pauli principle, one gets boson images still satisfying it. It is, however, well known that A particle states invariant under O(n) violate the Pauli principle except for s-shell nuclei. In the case of other shells, the collective states belong to the IR $(\lambda_1\lambda_2\lambda_3)$ of O(n) obtained by filling compactly with the A nucleons all the single-particle states in an oscillator well up to a given level and by considering the most symmetrical IR $[\lambda_1\lambda_2\lambda_3]$ of U(3) in the last unfilled level.[1,26–29] Due to the complementarity of the groups O(n) and $\mathcal{S}\not{h}(2d,R)$,[30] the collective states still belong to a single IR of the latter. Therefore, their dynamical group is the restriction of $\mathcal{S}\not{h}(2d,R)$ to this single IR.

An open question deserving further study is whether boson representations of the dynamical group can be obtained for an arbitrary IR $(\lambda_1\lambda_2\lambda_3)$ of O(n) by a procedure similar to the one used in the present paper in the case where $(\lambda_1\lambda_2\lambda_3) = (000)$. An argument in favor of such a generalization comes from a recent work by Dobaczewski.[31] He showed that by representing fermion states by functions of complex variables, it is possible to derive their boson representations whenever they form the carrier space of any IR of any semisimple compact Lie subgroup of SO($2N+1$), where N is the number of single-fermion states. We use a similar approach to his in the present paper for the noncompact $\mathcal{S}\not{h}(2d,R)$ group. The only difference between both procedures comes from the explicit use of coherent states in the sense of Perelomov[32] in the work of Dobaczewski and the underlying presence of coherent states in the sense of Barut and Girardello[13] in the present work (as was explicitly shown in I in the one-dimensional case). Which type of coherent states should be used in the case of an arbitrary IR of O(n) remains to be investigated.

APPENDIX A: PROOF OF EQUATION (4.5)

In this appendix, we wish to prove that any irreducible tensor $T[20]$ may be written as a linear combination of the tensors $T^{(m)}_{ij}$, $m = 0,1,...$, whose coefficients are analytic functions of the $\mathfrak{U}(d)$ Casimir operators $\Phi_1,...,\Phi_d$.

We start by noting that any $T[20]$ tensor being a rank 2 covariant tensor is a sum of terms, each one of these terms having a number of creation operators exceeding by one that of the annihilation ones. Let us restrict ourselves to one such term containing $p+1$ creation and p annihilation operators and assume that the proposition we wish to demonstrate is true for all values less than p. The $2p$ contravariant indices must be contracted with $2p$ covariant ones in order to leave two uncontracted covariant indices, respectively, equal to i and j. In such a procedure, the order of the creation and annihilation operators does not matter as their commutation only introduces lower order terms for which the theorem is assumed to be true. We may therefore put contracted operators together. In such a way we obtain chains of indices made up of cycles: one open cycle beginning with index i and ending with index j, and several closed cycles. As a result, the term considered gives rise to a product of various factors: (i) one factor corresponding to the open cycle and containing $q+1$ creation and q annihilation operators $(0 \leqslant q \leqslant p)$,

$$\sum_{k_1\cdots k_{2q}} \bar{a}^{\dagger}_{ik_1}\bar{a}_{k_1k_2}\bar{a}^{\dagger}_{k_2k_3}\cdots\bar{a}_{k_{2q-1}k_{2q}}\bar{a}^{\dagger}_{k_{2q}j} = (\mathfrak{C}^q\bar{a}^{\dagger})_{ij}; \qquad \text{(A1)}$$

(ii) several factors corresponding to the closed cycles and containing equal numbers of creation and annihilation operators,

$$\sum_{k_1\cdots k_{2r}} \bar{a}^{\dagger}_{k_1k_2}\bar{a}_{k_2k_3}\bar{a}^{\dagger}_{k_3k_4}\cdots\bar{a}^{\dagger}_{k_{2r-1}k_{2r}}\bar{a}_{k_{2r}k_1} = \Phi_r. \qquad \text{(A2)}$$

Of course, all these factors satisfy the relation $p = q + \Sigma\, r$. The proof of Eq. (4.5) is thus completed by induction over p.

APPENDIX B: PROOF OF EQUATION (4.18)

In this appendix, we wish to show the possibility of finding some functions $\alpha^{(k)}_m$ of the $\mathfrak{U}(d)$ Casimir operators such that

$$T^{(k)}_{ij} = \sum_{m=0}^{d-1} \alpha^{(k)}_m(\Phi_1,...,\Phi_d)T^{(m)}_{ij}, \quad k \geqslant d. \qquad \text{(B1)}$$

By taking the matrix element of both sides of Eq. (B1) between the hws of the IR's $[h_1\cdots h_i\cdots h_d]$ and $(h_1\cdots h_i - 2\cdots h_d]$ and using Eq. (4.14), we can transform Eq. (B1) into the following system of d equations in the d unknowns $\alpha^{(k)}_m$, $m = 0,1,...,d-1$:

$$\lambda_i^{\,k} = \sum_{m=0}^{d-1} \alpha^{(k)}_m(\phi_1,...,\phi_d)\lambda_i^{\,m}, \qquad \text{(B2)}$$

where $\phi_1,...,\phi_d$ are as before the eigenvalues of $\Phi_1,...,\Phi_d$ corresponding to the IR $[h_1h_2\cdots h_d]$.

The solution of such a system is given by

$$\alpha^{(k)}_m = [D(\lambda_1,...,\lambda_d)]^{-1}\begin{vmatrix} 1 & \lambda_1 & \cdots & \lambda_1^{m-1} & \lambda_1^{k} & \lambda_1^{m+1} & \cdots & \lambda_1^{d-1} \\ 1 & \lambda_2 & \cdots & \lambda_2^{m-1} & \lambda_2^{k} & \lambda_2^{m+1} & \cdots & \lambda_2^{d-1} \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ 1 & \lambda_d & \cdots & \lambda_d^{m-1} & \lambda_d^{k} & \lambda_d^{m+1} & \cdots & \lambda_d^{d-1} \end{vmatrix}, \quad m = 0,1,...,d-1, \qquad \text{(B3)}$$

where $D(\lambda_1,...,\lambda_d)$ is defined in Eq. (4.17). The right-hand side of Eq. (B3) is a function of the λ_i's or, equivalently, of the h_i's. By expressing the latter in terms of the ϕ_i's, we can in principle convert $\alpha_m^{(k)}$ into a function of the ϕ_i's, $\alpha_m^{(k)}(\phi_1,...,\phi_d)$. As this form is valid for any IR $[h_1h_2\cdots h_d]$ of $\mathfrak{U}(d)$, we get the coefficients $\alpha_m^{(k)}(\Phi_1,...,\Phi_d)$ appearing in Eq. (B1) by respectively replacing $\phi_1,...,\phi_d$ by $\Phi_1,...,\Phi_d$.

Alternatively, Eq. (4.18) could be proved by a procedure similar to the one that enables us to go from Eq. (6.26) to Eq. (6.30). In this method, we express the $T[1\dot{0}-1]$ irreducible tensor $\mathbb{C}^m - d^{-1}(\mathrm{tr}\,\mathbb{C}^m)\mathbf{I}$, for $m = k \geqslant d$, in terms of the $d-1$ independent ones, corresponding to $m = 1,...,d-1$.[21]

APPENDIX C: STUDY OF THE SET OF OPERATORS $[\Phi_{m+1},\bar{a}_{ij}^{\dagger}]$

In this appendix, we wish to review some properties of the set of operators $[\Phi_{m+1},\bar{a}_{ij}^{\dagger}]$, $m = 0,1,...,d-1$, where Φ_{m+1} is a $\mathfrak{U}(d)$ Casimir operator, as defined in Eq. (4.6).

It is obvious from their definition that the operators $[\Phi_{m+1},\bar{a}_{ij}^{\dagger}]$ are irreducible tensors $T[2\dot{0}]$ with respect to $\mathfrak{U}(d)$ for any non-negative integer m value. Moreover, we are going to prove that the first d of them, corresponding to $m = 0,1,...,d-1$, form a basis for $T[2\dot{0}]$ tensors. Since in Sec. 4 we did show that the basis for $T[2\dot{0}]$ tensors is made of d elements, it is sufficient to prove that the operators $[\Phi_{m+1},\bar{a}_{ij}^{\dagger}]$, $m = 0,1,...,d-1$, are linearly independent. By explicitly calculating the commutator, one finds that

$$[\Phi_{m+1},\bar{a}_{ij}^{\dagger}] = 2(m+1)(\mathbb{C}^m\bar{a}^{\dagger})_{ij} + \text{terms of lower degree.} \tag{C1}$$

The linear independence of the operators $(\mathbb{C}^m a^{\dagger})_{ij}$ therefore implies that of the operators $[\Phi_{m+1},\bar{a}_{ij}^{\dagger}]$.

Consequently, it is in principle possible to write any $T[2\dot{0}]$ tensor as

$$T_{ij} = \sum_{m=0}^{d-1} f'_m(\Phi_1,...,\Phi_d)[\Phi_{m+1},\bar{a}_{ij}^{\dagger}], \tag{C2}$$

where f'_m is some function of $\Phi_1,...,\Phi_d$. Such an expansion has already been proposed in Ref. 10. However, except for some low d values, this expansion is not so convenient to use as those given in Eqs. (4.19) and (4.21), because the counterpart of Eqs. (4.14) and (4.23) in the case of expansion (C2) assumes a very complicated form due to the complexity of the dependence of the ϕ_i's upon the h_i's.[19]

[1] V. Vanagas, "The Microscopic Nuclear Theory Within the Framework of the Restricted Dynamics," in *Lecture Notes in Physics* (University of Toronto, Toronto, 1977).
[2] E. Chacón, M. Moshinsky, and V. Vanagas, J. Math. Phys. **22**, 605 (1981).
[3] J. Deenen and C. Quesne, J. Math. Phys. **23**, 878 (1982).
[4] A. Arima and F. Iachello, Ann. Phys. (N.Y.) **99**, 253 (1976); **111**, 201 (1978); **123**, 468 (1979).
[5] T. Otsuka, A. Arima, F. Iachello, and I. Talmi, Phys. Lett. B **76**, 139 (1978); T. Otsuka, A. Arima, and F. Iachello, Nucl. Phys. A **309**, 1 (1978); T. Otsuka, Phys. Rev. Lett. **46**, 710 (1981); J. N. Ginocchio, Ann. Phys. (N.Y.) **126**, 234 (1980); J. N. Ginocchio and I. Talmi, Nucl. Phys. A **337**, 431 (1980); A. Arima, N. Yoshida, and J. N. Ginocchio, Phys. Lett. B **101**, 209 (1981).
[6] M. Moshinsky and T. H. Seligman, J. Math. Phys. **22**, 1526 (1981).
[7] A. Ya. Dzublik, V. I. Ovcharenko, A. I. Steshenko, and G. Filippov, Yad. Fiz. **15**, 869 (1972) [Sov. J. Nucl. Phys. **15**, 487 (1972)].
[8] W. Zickendraht, J. Math. Phys. **12**, 1663 (1971).
[9] T. Holstein and H. Primakoff, Phys. Rev. **58**, 1098 (1940).
[10] M. Moshinsky, Physica (Utrecht) A **114**, 322 (1982).
[11] J. Deenen and C. Quesne, Physica (Utrecht) A **114**, 318 (1982).
[12] F. J. Dyson, Phys. Rev. **102**, 1217 (1956).
[13] A. O. Barut and L. Girardello, Commun. Math. Phys. **21**, 41 (1971).
[14] V. Bargmann, Commun. Pure Appl. Math. **14**, 187 (1961).
[15] I. M. Gel'fand and M. L. Tseitlin, Dokl. Akad. Nauk SSSR **71**, 825 (1950).
[16] R. Pérez, "SU(3) Symmetry and Realistic Interactions," thesis, U.N.A.M., Mexico (1967).
[17] V. Vanagas, S. Ališauskas, R. Kalinauskas, and E. Nadjakov, Bulg. J. Phys. **7**, 168 (1980).
[18] C. Quesne, J. Math. Phys. **22**, 1482 (1981).
[19] J. D. Louck, Am. J. Phys. **38**, 3 (1970).
[20] M. C. K. Aguilera-Navarro and V. C. Aguilera-Navarro, J. Math. Phys. **17**, 1173 (1976).
[21] L. C. Biedenharn, J. Math. Phys. **4**, 436 (1963); G. E. Baird and L. C. Biedenharn, *ibid.* **4**, 1449 (1963).
[22] A. O. Barut, Phys. Rev. **139**, B1433 (1965).
[23] P. A. Mello and M. Moshinsky, J. Math. Phys. **16**, 2017 (1975).
[24] L. D. Mlodinow and N. Papanicolaou, Ann. Phys. (N.Y.) **128**, 314 (1980).
[25] L. D. Mlodinow and N. Papanicolaou, Ann. Phys. (N.Y.) **131**, 1 (1981).
[26] V. Vanagas, "The Microscopic Theory of the Collective Motion in Nuclei," in *Group Theory and Its Applications in Physics—1980*, edited by T. H. Seligman, AIP Conference Proceedings No. 71 (American Institute of Physics, New York, 1981).
[27] G. F. Filippov, Fiz. Elem. Chastits At. Yadra **4**, 992 (1973) [Sov. J. Part. Nucl. **4**, 405 (1974)].
[28] L. Sabaliauskas, Liet. fiz. rink. **19**, 5 (1979).
[29] O. Castaños, A. Frank, E. Chacón, P. O. Hess, and M. Moshinsky (to be published).
[30] M. Moshinsky and C. Quesne, J. Math. Phys. **12**, 1772 (1971).
[31] J. Dobaczewski, Nucl. Phys. A **369**, 213, 237 (1981).
[32] A. M. Perelomov, Commun. Math. Phys. **26**, 222 (1972); Usp. Fiz. Nauk **123**, 23 (1977) [Sov. Phys. Usp. **20**, 703 (1977)].

G_2 van der Waerden invariant[a)]

R. Gaskell
Department of Physics, Lafayette College, Easton, Pennsylvania 18042
R. T. Sharp[b)]
Institut des Hautes Études Scientifiques, Bures-sur-Yvette, France

(Received 19 March 1982; accepted for publication 21 May 1982)

The G_2 van der Waerden invariant is given. It solves the external labeling problem connected with direct products of irreducible representations of G_2.

PACS numbers: 02.20. + b, 11.30.Pb, 03.65. − w

I. INTRODUCTION

As the simplest of the exceptional groups, G_2 holds a perhaps disproportionate fascination. Among its physical applications are various tentative elementary particle schemes[1] and its use, in connection with the chain $SU(7) \supset SO(7) \supset G_2 \supset SO(3)$, in classifying f-shell many-particle states.[2]

A practical problem, for any group, is the calculation of its Wigner coefficients which couple states of three irreducible representations (IR's) to a scalar, or equivalently, its Clebsch–Gordan coefficients which couple states of two IR's to give a third.[3] The coefficients are needed for coupling states and tensors, and also, in connection with the Wigner–Eckart theorem, for the calculation of matrix elements of tensor operators.

In Sec. II we give the general van der Waerden invariant for G_2. It consists of a set of products of powers of a finite number of elementary couplings. It provides a complete nonredundant solution of the external labeling problem which is symmetric in the three IR's and which makes no reference to the internal basis states to be used. To calculate Wigner coefficients, one expands the van der Waerden invariant in products of states of the three IR's, using whatever basis states are convenient for the problem at hand.

Section III contains some concluding remarks.

II. THE VAN DER WAERDEN INVARIANT

Long ago van der Waerden[4] wrote down a general invariant in the basis states of three IR's of SU(2) in the form of a product of powers of certain "elementary scalars." Although relatively little used for higher groups,[5,6] it is a textbook way of obtaining SU(2) Wigner coefficients.[7]

With the help of Speiser's[8] graphical methods it is straightforward to compute the Clebsch–Gordan series for the direct product of any two IR's of G_2: the product IR's correspond to basis states of the first factor IR, perhaps with some cancellation due to Weyl reflections. Examinations of couplings of low-lying IR's suggests the following 36 elementary couplings:

$$C_i = A_j A_k, \quad D_i = B_j B_k, \quad E = A_1 A_2 A_3,$$

$$F = B_1 B_2 B_3, \quad G_i = B_i A_j A_k,$$

$$H_i = A^{\,2}_{i} B_j B_k, \quad J_i = A^{\,3}_{i} B_j B_k,$$

$$K_{ij} = A_i B_j A^{\,2}_{k}, \quad L_{ij} = B_i A^{\,2}_{j} A_k B_k,$$

$$M_{ij} = B_i B^{\,2}_{j} A^{\,3}_{k}, \quad N = A_1 B_1 A_2 B_2 A_3 B_3. \tag{1}$$

The notation is such that the product $A^{\,a_1}_{\,1} B^{\,b_1}_{\,1} A^{\,a_2}_{\,2} B^{\,b_2}_{\,2} A^{\,a_3}_{\,3} B^{\,b_3}_{\,3}$ stands for the coupling $(a_1,b_1;a_2,b_2;a_3,b_3)$, where (a_i,b_i) are the representation labels of the ith representation; (10) is the seven-dimensional and (01) the 14-dimensional fundamental IR. The subscripts i, j, k in (1) are 1,2,3 in any order. A coupling $(a_1,b_1;a_2,b_2;a_3,b_3)$ stands for the coupling of the three IR's in question to form a scalar, or, equivalently, the coupling of any two of the IR's to form the third. The designation of the couplings (1) as elementary implies that they constitute an integrity basis for the coupling problem; any coupling can be written as a product of powers of the elementary couplings; this conjecture is justified at the end of this section.

Because of syzygies (polynomial identities) relating the couplings (1), some products of powers must be eliminated to avoid multiple counting. According to Racah's counting of labels, not more than ten at a time of the elementary couplings may appear with arbitrary exponents in the same product when defining general couplings ($10 = \frac{1}{2}(r + 3l)$, where $r = 14$ and $l = 2$ are the order and rank respectively of G_2).

The van der Waerden invariants of SU(4)[5] and SO(5)[6] were determined heuristically with the help of Speiser's correspondence between couplings and basis states. The G_2 problem is sufficiently complicated that we were forced to implement Speiser's method analytically with the help of generating functions. We first give the results, and then describe their justification.

109 types of invariant may be distinguished. Each of the first 108 is characterized by a product of nonnegative integer powers of ten of the elementary scalars. Each set of ten includes the six denoted in (1) by C_i and D_i. The possible choices for the other four are

[a)] Supported in part by the Natural Sciences and Engineering Research Council of Canada and by the Ministère de l'Éducation du Québec.
[b)] On leave from McGill University, Montreal, Quebec, Canada.

$$EK_{ij}K_{ji}M_{ij},\quad J_iK_{jk}K_{kj}M_{jk},\quad EK_{ij}K_{kj}L_{ji},$$
$$EK_{ij}L_{jk}M_{ji},\quad G_iJ_jK_{ki}M_{ki},$$
$$EFK_{ij}L_{ji},\quad EFL_{ij}M_{ik},\quad FG_iK_{ji}M_{ji},$$
$$FG_iK_{ji}L_{ij},\quad FG_iL_{ij}M_{ik},$$
$$EFK_{ij}M_{ij},\quad FG_iH_jM_{ik},\quad FG_iH_jM_{ki},$$
$$G_iK_{ji}L_{ik}M_{ij},\quad G_iK_{ji}K_{ki}L_{ij},$$
$$G_iJ_jK_{ki}M_{ik},\quad G_iH_jJ_jM_{ik},\quad G_iH_jJ_jM_{ki}. \tag{2}$$

Each of the 18 groupings in (2) stands for six, obtained by giving i,j,k the values 1, 2, 3 in any order. The 109th type of invariant is a product of N (appearing only linearly) and nonnegative integer powers of the eight elementary scalars C_i, D_i, E, F.

Each linearly independent G_2 coupling is obtained exactly once by giving the exponents of the elementary couplings in each of the 109 terms all nonnegative integer values (except N which appears at most linearly). Of course, when some exponents vanish, the same product of powers may be obtained from different terms of (2); each such product of powers of elementary couplings is to be counted only once. The ten exponents in a product of powers of C_i and D_i and the four elementary couplings from a term of (2) provide the six representation labels and the four "missing" labels; the "missing" labels may be conveniently taken as the exponents of the four couplings in one term of (2), or of E and F in the term involving N.

Our derivation of the G_2 couplings makes use of Weyl's[9] characteristic function $\xi_{ab}(\eta,\zeta)$, given explicitly for G_2 by Behrends, Dreitlein, Fronsdal, and Lee.[1] According to Weyl the character $\chi_{ab}(\eta,\zeta)$ is given in terms of the characteristic by

$$\chi_{ab}(\eta,\zeta)=\xi_{ab}(\eta,\zeta)/\xi_{00}(\eta,\zeta); \tag{3}$$

η,ζ are dummy variables which carry as exponents the horizontal and vertical components of the weights of the representation (a,b). We shall find it more convenient to use the variables $x=\zeta^2$ and $y=\zeta^3\eta$ instead; the exponents of x and y, when positive, are the labels of the IR for which the weight in question is its highest. For our purposes we need the characteristic generating function, which turns out to be

$$\Xi_{AB}(x,y)=\sum_{s=0}^{11}\frac{(-1)^s x_s y_s}{(1-Ax_s)(1-By_s)}, \tag{4}$$

where (x_s,y_s) for s from 0 to 11 are (x,y), $(y/x,y)$, $(y/x,y^2/x^3)$, $(y/x^2,y^2/x^3)$, $(y/x^2,y/x^3)$, $(1/x,y/x^3)$, $(1/x,1/y)$, $(x/y,1/y)$, $(x/y,x^3/y^2)$, $(x^2/y,x^3/y^2)$, $(x^2/y,x^3/y)$, and $(x,x^3/y)$. When expanded in powers of A and B,

$$\Xi_{AB}(x,y)=\sum_{ab}A^aB^b\xi_{ab}(x,y), \tag{5}$$

the characteristic generating function provides the characteristic functions ξ_{ab} as coefficients. Similarly, the character generator $\mathrm{X}_{AB}(x,y)$ when expanded,

$$\mathrm{X}_{AB}(x,y)=\sum_{ab}A^aB^b\chi_{ab}(x,y), \tag{6}$$

provides, as coefficients, the characters of irreducible representations; the character generator for G_2 has been given by the present authors.[10] Multiplying $\mathrm{X}_{A_1B_1}(x,y)$ by $\Xi_{A_2B_2}(x,y)$, we obtain

$$\mathrm{X}_{A_1B_1}\Xi_{A_2B_2}=\sum_{a_1b_1a_2b_2a_3b_3}A_1^{a_1}B_1^{b_1}A_2^{a_2}B_2^{b_2}\xi_{a_3b_3}(x,y)C_{a_1b_1a_2b_2a_3b_3}, \tag{7}$$

where $C_{a_1b_1a_2b_2a_3b_3}$ is the multiplicity of (a_3b_3) in the direct product $(a_1b_1)\times(a_2b_2)$. In deriving (7) we used Speiser's

$$\chi_{a_1b_1}(x,y)\xi_{a_2b_2}(x,y)=\sum_{a_3b_3}\xi_{a_3b_3}(x,y)C_{a_1b_1a_2b_2a_3b_3}. \tag{8}$$

The characteristic $\xi_{a_3b_3}(x,y)$ contains one term in each of the 12 Weyl sectors of G_2; one of them, the only one with both exponents positive, is $x^{a_3+1}y^{b_3+1}$. Consequently, the multiplicity of the representation (a_3b_3), in a linear combination of characteristics such as (7), may be obtained by multiplying by $x^{-a_3-1}y^{-b_3-1}$ and isolating the x^0y^0 term. This can be done for all (a_3b_3) simultaneously by multiplying (7) by $x^{-1}y^{-1}[(1-A_3/x)(1-B_3/y)]^{-1}$ and picking out the x^0y^0 part. This is equivalent to multiplying (7) by $x^{-1}y^{-1}$, keeping the part whose expansion contains only nonnegative powers of x and y, and then setting $x=A_3$ and $y=B_3$. The projection of nonnegative powers of x and y is simplified by the procedure described in Ref. 11. The result is the G_2 Clebsch–Gordan generating function

$$G(A_1,B_1,A_2,B_2,A_3,B_3)=\sum_{a_1b_1a_2b_2a_3b_3}A_1^{a_1}B_1^{b_1}A_2^{a_2}B_2^{b_2}A_3^{a_3}B_3^{b_3}C_{a_1b_1a_2b_2a_3b_3}, \tag{9}$$

which gives the multiplicities $C_{a_1b_1a_2b_2a_3b_3}$ of all G_2 couplings. Because of its complexity (it contains 108 terms) we refrain from reproducing the generating function here. Each term has ten denominator factors of the form $1-X$, where the X are distinct elementary couplings which include C_1, C_2, C_3, D_1, D_2, D_3, the other four being one of the 108 sets given in (2). The terms include one with numerator unity, 26 whose numerator is a single elementary coupling, 55 whose numerator is a product of two elementary couplings, 25 with a product of three elementary couplings as numerators, and one with a product of four. The numerator factors are always chosen from the couplings in the denominator of the same term (not including C_i and D_i), except for one term which has N as its numerator; when the denominator factors are expanded to obtain higher couplings, N should be interpreted as EF when multiplied by denominator couplings other than E, F, C_i, D_i.

We have verified that the products of elementary couplings obtained by expanding the terms of the generating function correspond to the 108 types given by (2), to the special products involving N, and to no others.

III. CONCLUDING REMARKS

The van der Waerden invariant described in Sec. II not only enumerates all G_2 couplings, but solves the associated labeling problem. To construct the actual couplings, and in particular, to determine Wigner or Clebsch–Gordan coefficients, more computation is required. First one needs to determine each of the 11 distinct types of scalar defined by (1);

that task is relatively simple. More difficult is the projection of the stretched part (representation labels additive) in all IR labels of an allowed product of powers of elementary scalars, and its expansion in products of appropriate basis states of the three IR's involved; such a computation for SO(5) is carried out in Ref. 6.

Other resolutions of the labeling problem that implied by the "compatibility rules" (2) are possible. For each syzygy (polynomial identity relating elementary couplings) one must select one term and eliminate products containing it in order to avoid double counting. The choice of the term to be eliminated is somewhat arbitrary; a different choice leads to a different coupling scheme. The couplings arising from one scheme are linear combinations of those arising from another. We imposed on our solutions the restrictions that C_i and D_i not figure in the term of a syzygy to be eliminated and that the solution be symmetric under interchange of the representations 1,2,3. We give an example of a different solution which respects these conditions. There is a syzygy which involves the terms $M_{ij}M_{ji}$ and H^3_k. Our solution eliminates $M_{ij}M_{ji}$. Consider two terms in the generating function which are identical except for the factors $[(1-M_{ij})(1-H_k)]^{-1}$ and $M_{ji}\,[(1-M_{ji})(1-H_k)]^{-1}$. Combining the two terms and replacing $M_{ij}M_{ji}$ by H^3_k, one gets $(1+H_k+H^2_k)\,[(1-M_{ij})(1-M_{ji})]^{-1}$. H^3_k is now eliminated, and M_{ij} and M_{ji} are compatible; the number of terms is reduced at the expense of complicating the numerators.

The composite states corresponding to our solution of the external labeling problem, although complete and nonredundant, are not in general mutually orthogonal. To obtain orthonormal coupled states, one may use a Schmidt procedure or, alternatively, diagonalize some Hermitian operator such as that defined by the metric (overlap) matrix of our nonorthonormal couplings.

A possible further application for our G_2 Clebsch–Gordan generating function is in obtaining generating functions involving coupled G_2 representations; the SU(3) Clebsch–Gordan generating function was used in that way in Ref. 12 and that for SU(2) in Ref. 13.

By isolating the part of the van der Waerden invariant in which the three coupled IR's are the same, and taking note of the exchange symmetries of the elementary couplings, one can quite easily determine the scalar part of the symmetric, antisymmetric, or mixed-symmetry three-box plethysm based on any IR of G_2. The complete content of two-box plethysms is determined similarly.

ACKNOWLEDGMENT

We thank J. Patera for help in the early stages of our calculation and for useful discussions in the later stages.

[1] R. E. Behrends, J. Dreitlein, C. Fronsdal, and W. Lee, Rev. Mod. Phys. **34**, 1 (1962).
[2] G. Racah, Ergebn. Exact. Naturw. **37**, 28 (1965).
[3] E. P. Wigner, "Reduction of direct products and restrictions of representations to subgroups: the everyday tasks of quantum theorists," in *Lie Algebras: Applications and Computational Methods*, edited by B. Kolman (SIAM, Philadelphia, 1973), p. 1.
[4] B. L. van der Waerden, *Die Gruppentheoretische Methode in der Quantenmechanik* (Springer-Verlag, Berlin, 1932).
[5] R. T. Sharp and David Lee, Rev. Mex. Fis. **20**, 203 (1971).
[6] M. Hongoh, R. T. Sharp, and D. E. Tilley, J. Math. Phys. **15**, 782 (1974).
[7] V. Heine, *Group Theory in Quantum Mechanics* (Pergamon, New York, 1960).
[8] D. Speiser, *Lectures of the Istanbul Summer School of Theoretical Physics* (Gordon and Breach, New York, 1963), Vol. 1.
[9] H. Weyl, Math. Z. **23**, 271 (1925); **24**, 328, 377 (1926).
[10] R. Gaskell and R. T. Sharp, J. Math. Phys. **22**, 2736 (1981).
[11] C. Bodine and R. W. Gaskell, "Generating functions for IR multiplicities," J. Math. Phys. (to appear).
[12] R. Gaskell, G. Rosensteel, and R. T. Sharp, J. Math. Phys. **22**, 2732 (1981).
[13] J. Patera and R. T. Sharp, J. Phys. A: Math. Gen. **13**, 397 (1980).

Cartan–Gram determinants for the simple Lie groups

Alfred C. T. Wu
Department of Physics, University of Michigan, Ann Arbor, Michigan 48109

(Received 31 March 1982; accepted for publication 25 June 1982)

The Cartan–Gram determinants for the simple root systems are evaluated for the simple Lie groups A_n, B_n, C_n, D_n, and E_k ($k = 6,7,8$). The determinants satisfy a linear recursion relation which turns out to be the same for all these groups. For the E_n family, the Cartan–Gram determinant contains an explicit factor of $(9 - n)$ which vanishes for $n = 9$ and is negative for $n > 9$. This gives a simple explanation why the E_n family terminates at E_8. The Cartan–Gram determinant affords a systematic explanation for the nonexistence of the forbidden Dynkin diagrams.

PACS numbers: 02.20.Rt, 02.20.Sv

I. INTRODUCTION

The Cartan–Killing classification of simple Lie groups into the classical groups [namely, the unitary $A_n = \mathrm{SU}(n+1)$, orthogonal $B_n = \mathrm{SO}(2n+1)$, $D_n = \mathrm{SO}(2n)$, and symplectic $\mathrm{Sp}(2n)$] and the five exceptional ones (G_2,F_4,E_6,E_7, and E_8) is well known.[1–6] We shall speak of Lie groups and Lie algebras interchangeably.

The purpose of this note is to advocate the use of the Gram determinant (or, apart from a scale, the determinant of the Cartan matrices[1–5]) as an unambiguous clean test for the linear independence of the set of simple root vectors. For the allowed Dynkin diagrams, the Cartan–Gram determinants are positive definite. For the forbidden configurations, the determinants are negative or zero. The Cartan–Gram determinants are explicitly evaluated for the simple Lie groups A_n, B_n, C_n, D_n, and E_n. The answers are remarkably simple. The determinants are found to satisfy the *same* recursion relation for all these groups. For the E_n family, the Cartan–Gram determinant contains an explicit factor of $(9 - n)$, which vanishes for $n = 9$ and is negative for $n > 9$. This gives a simple explanation why the E_n family terminates at E_8.

For the basic notion and terminology, the reader is referred to the literature.[1–6] The following two paragraphs provide a minimal setting.

An arbitrary Lie group is decomposable into a semi-simple one and a solvable one. A semi-simple group is decomposable into simple groups. The r parameters of a simple Lie group of rank n can be split into n commuting (i.e., simultaneously diagonalizable) operators H_i ($i = 1,\ldots,n$), $\frac{1}{2}(r-n)$ raising operators E_α and $\frac{1}{2}(r-n)$ lowering operators $E_{-\alpha}$ such that, among other things, the composition (commutator) relation reads

$$[H_i, E_\alpha] = \alpha_i E_\alpha. \tag{1}$$

This says that the commutator between H_i and E_α corresponds to an eigenvalue problem in the adjoint representation. The eigenvalues α_i ($i = 1,\ldots,n$) are called components of a root vector α in a n-dimensional Euclidean space. By judicious choice, the root space can be spanned by a set of basis vectors (not necessarily orthogonal) called the simple roots. The set of Dynkin diagrams correspond to all the admissible simple root vectors that satisfy the following three requirements:

(a) angular restriction between two vectors: $\theta = \pi/2$, $2\pi/3$, $3\pi/4$, $5\pi/6$,

(b) relative length restriction: $-$, $1{:}\sqrt{2}{:}\sqrt{3}$, and

(c) linear independence.

II. THE CARTAN DETERMINANT; THE GRAM DETERMINANT

For a rank n Lie group, the Cartan matrix is a $n \times n$ matrix whose elements A_{ij} is defined as

$$A_{ij} = \frac{2(\alpha_i,\alpha_j)}{(\alpha_j,\alpha_j)}, \tag{2}$$

where α_i denotes the ith simple root and (α_i, α_i) is the inner product. The Cartan matrices for simple Lie groups are listed in the literature.[2,4,5]

On the other hand, the Gram matrix for a set of n vectors is defined as

$$G_{ij} = (\alpha_i,\alpha_j). \tag{3}$$

It is well known that the Gram determinant[7] (a) is positive for a set of linearly independent real Euclidean vectors, and (b) vanishes if and only if the set of vectors is linearly dependent.

The determinant of the Cartan matrix will be called the Cartan determinant here. It is simply proportional to the Gram determinant, the proportionality constant depends on the normalization of the simple roots.

With the known simple root system for the simple Lie groups, their Cartan determinants $\Delta \equiv \det A_{ij}$ of (2) and the Gram determinants $g \equiv \det G_{ij}$ of (3) can be easily evaluated. With the nesting structure such that the determinant of the next subgroup corresponds to the first principal minor, we have for the $n \times n$ determinants:

 0022-2488/82/112019-03$02.50 © 1982 American Institute of Physics

$$\Delta_n(A_n) = \begin{vmatrix} 2 & -1 & 0 & \cdot & \cdot & \cdot & \cdot \\ -1 & 2 & -1 & 0 & \cdot & \cdot & \cdot \\ 0 & -1 & 2 & -1 & 0 & \cdot & \cdot \\ \cdot & & & \ddots & & & \\ \cdot & & & \cdot & -1 & 2 & -1 \\ \cdot & & & \cdot & 0 & -1 & 2 \end{vmatrix} = n+1, \tag{4a}$$

$$g_n(A_n) = 2^{-n}(n+1), \tag{4b}$$

$$\Delta_n(B_n) = \begin{vmatrix} 2 & -1 & 0 & \cdot & \cdot & & \cdot \\ -1 & 2 & -1 & 0 & \cdot & & \cdot \\ 0 & -1 & 2 & -1 & 0 & & \cdot \\ \cdot & & & \ddots & & & \\ \cdot & & & & & & \\ & & & \cdot & -1 & 2 & -2 \\ \cdot & & & \cdot & 0 & -1 & 2 \end{vmatrix} = 2, \tag{5a}$$

$$g_n(B_n) = 2^{2-n}, \tag{5b}$$

$$\Delta_n(C_n) = \begin{vmatrix} 2 & -1 & 0 & \cdot & \cdot & & \cdot \\ -1 & 2 & -1 & 0 & \cdot & & \cdot \\ 0 & -1 & 2 & -1 & 0 & \cdot & \cdot \\ \cdot & \cdot & \cdot & \ddots & \cdot & & \\ \cdot & & & \cdot & -1 & 2 & -1 \\ \cdot & & & \cdot & 0 & -2 & 2 \end{vmatrix} = 2, \tag{6a}$$

and

$$g_n(C_n) = 2^{-n}. \tag{6b}$$

Note that the matrices for C_n and B_n are the transpose of each other.

$$\Delta_n(D_n) = \begin{vmatrix} 2 & -1 & 0 & \cdot & \cdot & & \cdot \\ -1 & 2 & -1 & 0 & & & \cdot \\ 0 & -1 & 2 & -1 & 0 & & \cdot \\ \cdot & \cdot & \cdot & \ddots & & & \\ \cdot & \cdot & \cdot & \cdot & 2 & -1 & -1 \\ \cdot & & & \cdot & -1 & 2 & 0 \\ \cdot & \cdot & \cdot & \cdot & -1 & 0 & 2 \end{vmatrix} = 4, \tag{7a}$$

$$g_n(D_n) = 2^{2-n}, \tag{7b}$$

$$\Delta_n(E_n) = \begin{vmatrix} 2 & -1 & 0 & \cdot & \cdot & \cdot & \cdot & & \cdot \\ -1 & 2 & -1 & 0 & \cdot & \cdot & \cdot & & \cdot \\ \cdot & & \ddots & & & & & & \\ \cdot & & -1 & 2 & -1 & 0 & & & \cdot \\ & & \cdot & -1 & 2 & 0 & -1 & 0 & 0 \\ \cdot & & & \cdot & 0 & 2 & -1 & 0 & 0 \\ \cdot & & & \cdot & -1 & -1 & 2 & -1 & 0 \\ & & & \cdot & 0 & 0 & -1 & 2 & -1 \\ \cdot & & & \cdot & 0 & 0 & 0 & -1 & 2 \end{vmatrix} = 9-n, \tag{8a}$$

$$[E_4 \equiv A_4, E_5 \equiv D_5],\ g_n(E_n) = 2^{-n}(9-n). \tag{8b}$$

A simple recursion relation is seen to hold for all these cases. We have

$$\Delta_n = 2\Delta_{n-1} - \Delta_{n-2}, \tag{9a}$$

$$g_n = g_{n-1} - \tfrac{1}{4} g_{n-2}. \tag{9b}$$

The $(9-n)$ factor of the Cartan–Gram determinant for the E_n family gives a simple explanation why the family does not extend beyond E_8.

We state without elaboration that the Cartan–Gram determinant affords a clean systematic test for the admissibility of a simple root system.

[1] E. B. Dynkin, Usp. Mat. Nauk. **2**, No. 4 (20), 59 (1947); Transl. Am. Math. Soc. **9**, 328 (1962).

[2]N. Jacobson, *Lie Algebras* (Wiley, New York, 1962).
[3]R. Gilmore, *Lie Groups, Lie Algebras and Some of Their Applications* (Wiley, New York, 1974).
[4]B. G. Wybourne, *Classical Groups for Physicists* (Wiley, New York, 1974).
[5]R. Slansky, "Group Theory for Unified Model Building," Phys. Rep. **79**, 1 (1981).
[6]A. O. Barut and R. Rączka, *Theory of Group Representations & Applications* (PWN Scientific Publishers, Warsaw, 1980), 2nd rev. ed.
[7]See, e.g., R. Courant and D. Hilbert, *Methods of Mathematical Physics* (Wiley-Interscience, New York, 1953), Vol. 1, p.34.

A closed formula for the product of irreducible representations of SU (3)

Michael F. O'Reilly
Department of Mathematics, Rhodes University, Grahamstown 6140, South Africa

(Received 10 February 1982; accepted for publication 14 May 1982)

We determine a closed formula in terms of p, q, r, and s for the decomposition of the product $[p, q]\,[r, s]$ of finite-dimensional irreducible representations of SU (3). We also determine in terms of p, q, r, s, m, and n necessary and sufficient conditions that a term $[m, n]$ appears in this decomposition and its multiplicity.

PACS numbers: 02.20.Qs, 02.30.Lt

1. INTRODUCTION

Although there are various algorithms for determining the decomposition of the tensor products of representations of SU (3) the author is unaware of any closed formula for this. There are obvious advantages of a closed formula in determining structure and detecting patterns for generalization to the more intractable problem for SU (n).

The formula developed here is a triple summation which includes repeated terms; the multiplicity of these terms is determined separately. The proofs, mostly by induction, are themselves unrevealing and tedious.

2. NOTATION AND PRELIMINARY RESULTS

Let A_2 denote the Lie algebra of infinitesimal operators associated with SU(3). Let e_{ij} denote the 3×3 matrix with entry 1 in the (i,j)th position and zero elsewhere. The operators

$$T_+ = e_{12},\quad T_- = e_{21},\quad U_+ = e_{23},\quad U_- = e_{32},$$
$$V_+ = e_{13},\quad V_- = e_{31},\quad T_c = e_{11} - e_{22},\quad U_c = e_{22} - e_{33}$$

form a basis of A_2. Here $T_\pm$, $U_\pm$, $V_\pm$ are the usual spin operators but $T_c = 2T_3$, $U_c = 2U_3$ in the notation of Gasiorowicz.[1] T_c and U_c will thus have integer eigenvalues. The commutator relations are then

$$[T_+,T_-] = T_c,\quad [T_c,T_\pm] = \pm 2T_\pm,$$
$$[U_+,U_-] = U_c,$$
$$[U_c, U_\pm] = \pm 2U_\pm,\quad [V_+, V_-] = U_c + T_c,$$
$$[T_c, U_\pm] = \mp U_\pm$$
$$[T_c, V_\pm] = \pm V_\pm,\quad [U_c, T_\pm] = \mp T_\pm,$$
$$[U_c, V_\pm] = \pm V_\pm$$
$$0 = [T_c, U_c] = [T_\pm, U_\mp] = [T_\pm, V_\pm] = [U_\pm, V_\pm],$$
$$[V_\pm, U_\mp] = \pm T_\pm,\quad [U_\pm, T_\pm] = \mp V_\pm, [T_\pm, V_\mp] = \mp U_\mp.$$

Let M be a finite-dimensional representation space for A_2 (a left A_2-module). For arbitrary $Z_i \epsilon A_2$, $v \epsilon M$, the product $[Z_1, Z_2]$ acts on v by the rule $[Z_1, Z_2]v = Z_1(Z_2 v) - Z_2(Z_1 v)$ which we use more frequently in the form

$$Z_1(Z_2 v) = [Z_1,Z_2]v + Z_2(Z_1 v) \tag{2.1}$$

M has a basis of common eigenvectors of T_c and U_c; the associated eigenvalues l, m respectively are integers. If M is irreducible it is generated over A_2 by a maximal vector v^+ unique up to scalar multiplication[1]; v^+ is distinguished by the property that it is annihilated by T_+ and U_+ (and therefore by $V_+ = [T_+, U_+]$). The eigenvalues r, s of T_c and U_c, respectively, associated with v^+ are non-negative. Since two irreducible A_2-modules are isomorphic if and only if they have the same ordered pairs of eigenvalues associated with their maximal vectors, the ordered pairs $[r, s]$ may be used to denote the distinct irreducible A_2-modules. v^+ is annihilated by $r + 1$ successive applications of T_- and $s + 1$ of U_-. The dimension of $[r, s]$ is

$$\tfrac{1}{2}(r + 1)(s + 1)(r + s + 2). \tag{2.2}$$

The number of irreducible components of M will equal the dimension of the subspace W of M annihilated by both T_+ and U_+; the components of M may be determined by choosing a basis of W consisting of common eigenvectors of T_c and U_c and finding the eigenvalue pairs corresponding to each element of this basis.

Let $w \epsilon [r, s]$ be an arbitrary common eigenvector of T_c and U_c with eigenvalues l, m, respectively. Then $T_\pm w$, $U_\pm w$, and $V_\pm w$ will also be eigenvectors of T_c and U_c which, if not the zero vector, will have eigenvalue pairs

$$(l \pm 2, m \mp 1),\quad (l \mp 1, m \pm 2),\quad (l \pm 1, m \pm 1), \tag{2.3}$$

respectively.

The simplest A_2-module is [0,0] which has a dimension equal to 1 and is annihilated by A_2. Next simplest are [1,0] and [0,1] each of dimension 3. Let u_0, $\bar{u}_0$ be maximal vectors of [1,0] and [0,1], respectively. Then [1,0] has as basis $\{u, T_- u, U_- T_- u\}$ while $U_- u = 0$. Similarly [0,1] has as basis $\{\bar{u}, U_-\bar{u}, T_- U_-\bar{u}\}$ and $T_-\bar{u} = 0$.

Let M, N be A_2-modules, $v \epsilon M$, $w \epsilon N$. The action of A_2 on the Kronecker product $M \otimes N$ is given by

$$Z(v \otimes w) = Zv \otimes w + v \otimes Zw,\quad Z \epsilon A_2. \tag{2.4}$$

To simplify notation we write $[r, s]\,[p, q]$ for $[r, s] \otimes [p, q]$.

3. GENERATING FORMULAS

In this section we obtain formulas (3.2) and (3.3) for $[r, s]$ [1,0] and $[r, s][0,1]$ which may be used to generate the formula for $[r, s][p, q]$.

Proposition 1: Let v, u be maximal vectors of $[r, s]$ and [1,0], respectively. Then

$$w_1 = v \otimes u,\quad w_2 = T_- v \otimes u - rv \otimes T_- u,$$

and

 0022-2488/82/112022-07$02.50 © 1982 American Institute of Physics

$$w_3 = U_-v \otimes T_-u - sv \otimes U_-T_-u + \frac{s}{r+s+1} U_-T_-v \otimes u$$
$$- \frac{s+1}{r+s+1} T_-U_-v \otimes u,$$

if not zero, are maximal vectors in $[r, s]$ [1,0].

Proof: Apply T_+ and U_+ to each vector. Since T_+ annihilates v and u, $T_+ w_1 = T_+v \otimes u + v \otimes T_+u = 0$. Similarly $U_+w_1 = 0$. So w_1 is a maximal vector. Also,

$$T_+w_2 = T_+T_-v \otimes u + T_-v \otimes T_+u - rT_+v \otimes T_-u$$
$$- rv \otimes T_+T_-u = [T_+,T_-]v \otimes u + T_-T_+v$$
$$\otimes u - r\{v \otimes [T_+,T_-]u + v \otimes T_-T_+u\}$$

[by (2.1) and maximality of u and v]

$$= T_cv \otimes u - rv \otimes T_cu$$
$$= rv \otimes u - rv \otimes u = 0.$$

Similarly,

$$U_+w_2 = U_+T_-v \otimes u - T_-v \otimes U_+u - rU_+v$$
$$\otimes T_-u - rv \otimes U_+T_-u$$
$$= [U_+,T_-]v \otimes u - rv \otimes [U_+,T_-]u = 0$$

since $[U_+,T_-] = 0$. So w_2 is a maximal vector.

Further, omitting some zero terms,

$$T_+w_3 = T_+U_-v \otimes T_-u + U_-v \otimes T_+T_-u - sv$$
$$\otimes T_+U_-T_-u + \frac{s}{r+s+1} T_+U_-T_-v \otimes u$$
$$- \frac{s+1}{r+s+1} T_+T_-U_-v \otimes u.$$

Reducing this term by term using (2.3), (2.4), and the commutator relations,

$$T_+U_-v = [T_+,U_-]v = 0,$$
$$T_+T_-u = [T_+,T_-]u = T_cu = u,$$
$$T_+U_-T_-u = [T_+,U_-]T_-u + U_-T_+T_-u$$
$$= U_-[T_+,T_-]u = U_-T_cu = U_-u = 0,$$
$$T_+U_-T_-v = U_-T_cv = rU_-v,$$
$$T_+T_-U_-v = [T_+,T_-]U_-v + T_-T_+U_-v$$
$$= T_cU_-v + T_-[T_+,U_-]v + T_-U_-T_+v$$
$$= (r+1)U_-v.$$

Thus,

$$T_+w_3 = U_-v \otimes u + \frac{s}{r+s+1} rU_-v \otimes u$$
$$- \frac{s+1}{r+s+1} (r+1)U_-v \otimes u = 0,$$
$$U_+w_3 = U_+U_-v \otimes T_-u + U_-v \otimes U_+T_-u - sv$$
$$\otimes U_+U_-T_-u + \frac{s}{r+s+1} U_+U_-T_-v \otimes u$$
$$- \frac{s+1}{r+s+1} U_+T_-U_-v \otimes u.$$

Here

$$U_+U_-v = [U_+,U_-]v = U_cv = sv,$$
$$U_+T_-u = [U_+,T_-]u = 0,$$
$$U_+U_-T_-u = [U_+,U_-]T_-u + U_-U_+T_-u$$
$$= U_cT_-u + U_-[U_+,T_-]u = T_-u,$$
$$U_+U_-T_-v = (s+1)T_-v, U_+T_-U_-v$$
$$= [U_+,T_-]v + T_-U_+U_-v$$
$$= T_-[U_+,U_-]v = T_-U_cv = sT_-v.$$

So

$$U_+w_3 = sv \otimes T_-u - sv \otimes T_-u$$
$$+ \frac{s}{r+s+1} (s+1)T_-v \otimes u - \frac{s+1}{r+s+1} sT_-v \otimes u,$$

whence w_3 is also a maximal vector. Q.E.D.

Propositon 2:

(i) $[r, s]$ [0,0] = $[r, s]$. (3.1)

(ii) If $r \neq 0$, $s \neq 0$,

$$[r, s][1,0] = [r+1, s] + [r-1, s+1] + [r, s-1]. \quad (3.2a)$$

If $r \neq 0$,

$$[r,0][1,0] = [r+1,0] + [r-1,1]. \quad (3.2b)$$

If $s \neq 0$,

$$[0, s][1,0] = [1, s] + [0, s-1]. \quad (3.2c)$$

Proof:

(i) Trivially if $\{u\}$ is a basis of [0,0] the mapping $v \otimes u \to v$, $v \in [r, s]$ is isomorphism, giving the result.

(ii) If $s = 0$, then $U_-v = 0$ and together these give $w_3 = 0$.

If $r = 0$ then $T_-v = 0$ and so $w_2 = 0$. Applying T_c and U_c to w_1, w_2 and w_3 when these are nonzero, using (2.3), shows that they are common eigenvectors with eigenvalue pairs $(r+1, s)$, $(r-1, s+1)$, and $(r, s-1)$, respectively. Thus, when the corresponding w_i is nonzero, $[r+1, s]$, $[r-1, s+1]$, and $[r, s-1]$ are components of $[r, s][1,0]$. A check on dimensions using (2.2) show that these are the only components in each case.

The linear function $\Theta: A_2 \to A_2$ which maps $T_\pm \to U_\pm$, $U_\pm \to T_\pm$ and $V_\pm \to -V_\pm$ is an automorphism. This can be checked using the commutator relations. Given an A_2-module M the conjugate module $\bar{M}$ is defined by taking the same underlying vector space and defining $Z \cdot v = \theta(Z)v$, $Z \in A_2$, $v \in M$. Since the role of U_c and T_c are interchanged, $\overline{[r, s]} = [s,r]$. Also $\overline{M \otimes N} = \bar{M} \otimes \bar{N}$ (where M and N are A_2-modules). Applying this to Proposition 2(ii) gives an immediate proof of Proposition 3.

Proposition 3: If $r, s \neq 0$,

$$[r, s]\ [0,1] = [r, s+1] + [r+1, s-1] + [r-1, s]. \quad (3.3a)$$

If $r \neq 0$,

$$[r,0]\ [0,1] = [r,1] + [r-1,0], \quad (3.3b)$$

If $s \neq 0$,

$$[0, s][0,1] = [0, s+1] + [1, s-1], \quad (3.3c)$$

Note: To facilitate notation we will use pairs $[r, s]$ with either r or s negative. Such pairs will by definition be zero and termed *redundant*. Using such notation, (3.2a) includes (3.2b) and (3.2c) while (3.3a) includes (3.3b) and (3.3c). However, products with such symbols will not be permitted as the formulas (3.2) and (3.3) do not hold when r or s is negative.

4. PRODUCT FORMULAS

A formula for $[r, s][p, q]$ is obtained in Proposition 6. First, however, it is nesessary to treat the cases where $s = q = 0$ (Proposition 4) and where $q = 0$ (Proposition 5). In all cases the proof is by induction involving (somewhat tedious) combinations of summations.

To simplify notaion, $\sum_{i=e,f}^{g,h}$ is used to denote the sum where i ranges from max $\{e,f\}$ to min $\{g, h\}$.

Proposition 4:

$$[r,0]\,[p,0] = \sum_{i=0}^{r,p} [r+p-2i,i] \tag{4.1}$$

Proof: We use induction on the minimum of p and r (say p). Equation (3.1) gives the result when $p = 0$. The proof given for the inductive step involves products $[r,0]\,[p-3,0]$. To avoid products of redundant terms (see Note) it is necessary to consider the cases $p = 1$ and $p = 2$ separately. When $p = 1$ the result is given by (3.2). For $p = 2$, since $[2,0] = [1,0][1,0] - [0,1]$ from (3.2b) we have (since $r \geqslant 2$)

$$\begin{aligned}[r,0][2,0] &= [r,0]\{[1,0][1,0] - [0,1]\} \\ &= \{[r+1,0] + [r-1,1]\}[1,0] - [r,0][0,1] \\ &= [r+2,0] + [r,1] + [r-2,2],\end{aligned}$$

given the result for $p = 2$. For $p > 2$, again using (3.2) and the induction hypothesis, we have

$$\begin{aligned}[r,0][p,0] &= [r,0]\{[p-1,0][1,0] - [p-2,1]\} \\ &= [r,0]\{[p-1,0][1,0] - [p-2,0][0,1] + [p-3,0]\} \\ &= [1,0]\sum_{i=0}^{p-1}[r+p-2i-1,i] \\ &\quad - [0,1]\sum_{i=0}^{p-2}[r+p-2i-2,i] \\ &\quad + \sum_{i=0}^{p-3}[r+p-2i-3,i] \\ &= \sum_{i=0}^{p-1}\{[r+p-2i,i] + [r+p-2i-2,i+1] \\ &\quad + [r+p-2i-1,i-1]\} \\ &\quad - \sum_{i=0}^{p-2}\{[r+p-2i-2,i+1] + [r+p-2i-1,i-1] \\ &\quad + [r+p-2i-3,i]\} \\ &\quad + \sum_{i=0}^{p-3}[r+p-2i-3,i] \\ &= \left\{\sum_{i=0}^{p-1}[r+p-2i,i]\right\} + [r-p,p] + [r-p+1,p-2] \\ &\quad - [r-p+1,p-2] \\ &= \sum_{i=0}^{p}[r+p-2i,i].\end{aligned}$$

Q.E.D.

Proposition 5:

$$[r,s]\,[p,0] = \sum_{j=0}^{s,p}\sum_{i=0}^{p-j,r}[r+p-j-2i, s-j+i]. \tag{4.2}$$

Proof: We use induction on s.

Case $s = 0$: The result is trivially true for s = 0 since then (4.2) reduces to (4.1).

Case $s = 1$: As in Proposition 4, we must treat the case $s = 1$ separately. For $s = 1$ and $r = 0$ it reduces to (3.3b); for $r \geqslant 1$ $[r,1] = [r,0][0,1] - [r-1,0]$ so by hypothesis

$$\begin{aligned}[p,0][r,1] &= [p,0]\{[r,0][0.1] - [r-1,0]\} \\ &= \sum_{i=0}^{p,r}[r+p-2i,i][0,1] - \sum_{i=0}^{p,r-1}[r+p-2i-1,i] \\ &= \sum_{i=0}^{p,r}[r+p-2i,i+1] + \sum_{i=0}^{p,r}[r+p-2i+1,i-1] \\ &\quad + \sum_{i=0}^{p,r}[r+p-2i,i+1] - \sum_{i=0}^{p,r-1}[r+p-2i-1,i].\end{aligned}$$

The last two sums combine to give $[p-r-1,r]$ whenever $p \geqslant r$ and zero otherwise. Since the term is redundant for $p = r$, it occurs only when $p \geqslant r+1$. It may included in the second sum with $i = r+1$ giving $\sum_{i=1}^{p,r+1}[r+p-2i,i+1]$, since the term for $i = 0$ is also redundant. Replacing the dummy variable i by $i+1$ this becomes $\sum_{i=0}^{p-1,r}$ $[r+p-2i-1,i]$ giving

$$\begin{aligned}[p,0]\,[r,1] &= \sum_{i=0}^{p,r}[r+p-2i,i+1] \\ &\quad + \sum_{i=0}^{p-1,r}[r+p-2i-1,i],\end{aligned}$$

which agrees with (4.2) for $s = 1$.

Case $s > 1$: For $r \geqslant 1, s > 1$ using (3.3a) and the induction hypothesis,

$$\begin{aligned}[r,s][p,0] &= \{[r,s-1][0,1] - [r+1,s-2] \\ &\quad - [r-1,s-1]\}[p,0] \\ &= \sum_{j=0}^{s-1,p}\sum_{i=0}^{p-j,r}[r+p-j-2i, s-j+i-1]\,[0,1] \\ &\quad - [r+1,s-2]\,[p,0] \\ &\quad - [r-1,s-1][p,0] \\ &= \sum_{j=0}^{s-1,p}\sum_{i=0}^{p-j,r}\{[r+p-j-2i, s-j+i] \\ &\quad + [r+p-j-2i+1, s-j+i-2] \\ &\quad + [r+p-j-2i-1, s-j+i-1]\} \\ &\quad - \sum_{j=0}^{s-2,p}\sum_{i=0}^{p-j,r+1}[r+p-j-2i+1, s-j+i-2] \\ &\quad - \sum_{j=0}^{s-1,p}\sum_{i=0}^{p-j,r-1}[r+p-j-2i-1, s-j+i-1],\end{aligned} \tag{4.3}$$

When $r = 0, s > 1$, we obtain a similar formula but lacking the fifth double summation.

Combining the second and fourth double summations gives

$$\left\{\sum_{j=0}^{s-1,p}\sum_{i=0}^{p-j,r}-\sum_{j=0}^{s-2,p}\sum_{i=0}^{p-j,r+1}\right\}$$
$$[r+p-j-2i+1, s-j+i-2]$$
$$=\sum_{j=0}^{s-2,p}\left\{\sum_{i=0}^{p-j,r}-\sum_{i=0}^{p-j,r+1}\right\}[r+p-j-2i+1, s-j+i-2]$$
$$+\left\{\text{if } p>s-1, \sum_{i=0}^{p-s+1,r}[r+p-s-2i+2, i-1]\right\}. \tag{4.4}$$

$\left\{\sum_{i=0}^{p-j,r}-\sum_{i=0}^{p-j,r+1}\right\}$ will only contribute terms when $p-j\geqslant r+1$, that is $j\leqslant p-r-1$; then the contribution will have $i=r+1$ and be $-[p-j-r-1, s-j+r-1]$. The sum of $p\geqslant s-1$ contains redundant terms (for $i=0$ and in particular for $p=s-1$) and simplifies to

$$\left\{\text{if } p\geqslant s, \sum_{i=1}^{p-s+1,r}[r+p-s-2i+2, i-1]\right\}$$

which on replacing i by $i+1$ gives

$$\left\{\text{if } p\geqslant s, \sum_{i=0}^{p-s,r-1}[r+p-s-2i, i]\right\}.$$

So (4.4) becomes

$$-\sum_{j=0}^{s-2,p-r-1}[p-j-r-1, s-j+r-1]$$
$$+\left\{\text{if } p\geqslant s, \sum_{i=0}^{p-s,r-1}[r+p-s-2i, i]\right\}. \tag{4.5}$$

Similarly combining the third and fifth double summations of (4.3) for $r\geqslant 1$, gives

$$\left\{\sum_{j=0}^{s-1,p}\sum_{i=0}^{p-j,p}-\sum_{j=0}^{s-1,p}\sum_{i=0}^{p-j,r-1}\right\}$$
$$[r+p-j-2i-1, s-j+i-1]$$
$$=\sum_{j=0}^{s-1,p-r-1}[p-r-j-1, s-j+r-1].$$

This is also obtained in the case $r=0$ (when the fifth double summation is absent) by putting $r=0$ in the third double summation. Combining this with (4.5) yields

$$\{\text{if } s\leqslant p-r, [p-r-s, r]\}$$
$$+\left\{\text{if } p\geqslant s, \sum_{i=0}^{p-s,r-1}[r+p-s-2i, i]\right\}$$
$$=\left\{\text{if } p\geqslant s, \sum_{i=0}^{p-s,r}[r+p-s-2i, i]\right\}.$$

But this is just the extra term together with the condition under which it will occur, if the range of the first summation on (4.3) is extended to include $j=s$.

So (4.3) becomes $\sum_{j=0}^{s,p}\sum_{i=0}^{p-j,r}$ $[r+p-j-2i, s-j+i]$, which agrees with (4.2).

Q.E.D.

Proposition 6:

$$[r,s][p,q]=\sum_{k=0}^{q,r+s}\sum_{j=0}^{s,p,r+s-k}\sum_{i=0,j-s+k}^{p-j+k,r}$$
$$[r+p-j-2i+k, s+q+i-j-2k]. \tag{4.6}$$

Proof: The result is trivial for $r=s=0$ so we consider the case where $r+s\neq 0$. The proof is by induction on q.

Case $q=0$: For $q=0$ the result is given by (4.2).

Case $q=1$: We must again consider the case $q=1$ separately. For $q=1, p=0$ the result follows from (3.3). For $q=1, p\geqslant 1$, apply (3.3b) and (4.2) to obtain

$$[r,s][p,1]=[r,s]\{[p,0][0,1]-[p-1,0]\}$$
$$=\sum_{j=0}^{s,p}\sum_{i=0}^{p-j,r}[r+p-j-2i, s-j+i][0,1]$$
$$-\sum_{j=0}^{s,p-1}\sum_{i=0}^{p-j-1,r}[r+p-j-2i-1, s-j+i]$$
$$=\sum_{j=0}^{s,p}\sum_{i=0}^{p-j,r}[r+p-j-2i, s-j+i+1]$$
$$+\sum_{j=0}^{s,p}\sum_{i=0}^{p-j,r}[r+p-j-2i+1, s-j+i-1]$$
$$+\sum_{j=0}^{s,p}\sum_{i=0}^{p-j,r}[r+p-j-2i-1, s-j+i]$$
$$-\sum_{j=0}^{s,p-1}\sum_{i=0}^{p-j-1,r}[r+p-j-2i-1, s-j+i]. \tag{4.7}$$

The last two double summations combine to give

$$\sum_{j=0}^{s,p-1}\left\{\sum_{i=0}^{p-j,r}-\sum_{i=0}^{p-j-1,r}\right\}[r+p-j-2i-1, s-j+i]$$
$$+\{\text{if } s\geqslant p, [r-1, s-p]\}. \tag{4.8}$$

For $r=0$ or $s\leqslant p-r-1$, this is zero, since $r\leqslant p-j-1$. Otherwise it equals

$$\sum_{j=0,p-r}^{s,p-1}[r-p+j-1, s+p-2j],$$
$$+\{\text{if } s\geqslant p, [r-1, s-p]\}$$
$$=\sum_{j=0,p-r}^{s,p}[r-p+j-1, s+p-2j].$$

Since the term for $j=p-r$ is redundant (when $s=p-r$, this is the only term), (4.8) becomes

$$\begin{cases}\sum_{j=0,p-r+1}^{s,p}[r-p+j-1, s+p-2j], & \text{if } s\geqslant p-r+1, r>0\\ 0, & \text{otherwise.}\end{cases} \tag{4.9}$$

These are exactly the extra terms obtained by allowing the value of i in the second double sum in (4.7) to range from 0 to min $(r, p-j+1)$. Thus (4.7) becomes

$$[r,s][p,1]=\sum_{j=0}^{s,p}\sum_{i=0}^{p-j,r}[r+p-j-2i, s-j+i+1]$$
$$+\sum_{j=0}^{s,p}\sum_{i=0}^{p-j+1,r}[r+p-j-2i+1, s-j+i-1] \tag{4.10}$$

To see that (4.10) agrees with (4.6) consider the cases $r>0$ and $r=0$ separately. In either case min $(r+s,p)=1$ so in (4.6), k takes only the values 0 and 1. For $r>0$, min $(s,p,r+s-k)=\min(s,p)$ and (4.6) becomes

$$[r,s][p,1] = \sum_{j=0}^{s,p} \sum_{i=0,j-s}^{p-j,r} [r+p-j-2i, s+1+i-j] + \sum_{j=0}^{s,p} \sum_{i=0,j-s+1}^{p-j+1,r} [r+p-j-2i+1, s+i-j-1]. \tag{4.11}$$

Here $\max(0,j-s) = 0$ and

$$\max(0,j-s+1) = \begin{cases} 0 & \text{if } j<s \\ 1 & \text{if } j=s \end{cases}.$$

Since the term for $j=s$ and $i=0$ in the second double summation in (4.10) is redundant, (4.10) agrees with (4.11) and hence (4.6) for $r>0$. For $r=0$ (4.10) becomes

$$\sum_{j=0}^{s,p} [p-j, s-j+1] + \sum_{j=0}^{s,p} [p-j+1, s-j-1] = \sum_{j=0}^{s,p} [p-j, s-j+1] + \sum_{j=0}^{s-1,p} [p-j+1, s-j-1]$$

on removing the redundant term for $j=s$. In this case (4.6) only has terms for $i=0$ and since $\min(s,p,r+s-k) = \min(p, s-k)$, (4.6) becomes

$$\sum_{j=0}^{s,p} [p-j, s+1-j] + \sum_{j=0}^{p,s-1} [p-j+1, s-j-1].$$

So again (4.6) and (4.10) agree, proving (4.6) is valid for $q=1$.

Case $q \geqslant 2$: To prove the inductive step (for $q \geqslant 2$) we use (3.3a) and the inductive hypothesis to obtain (for $p \neq 0$)

$$\begin{aligned}
&[r,s][p,q] \\
&= [r,s]\{[p,q-1][0,1] - [p+1,q-2] - [p-1,q-1]\} \\
&= \sum_{k=0}^{q-1,r+s} \sum_{j=0}^{s,p,r+s-k} \sum_{i=0,j-s+k}^{p-j+k,r} [r+p-j-2i+k, s+q+i-j-2k] \\
&\quad + \sum_{k=0}^{q-1,r+s} \sum_{j=0}^{s,p,r+s-k} \sum_{i=0,j-s+k}^{p-j+k,r} [r+p-j-2i+k+1, s+q+i-j-2k-2] \\
&\quad + \sum_{k=0}^{q-1,r+s} \sum_{j=0}^{s,p,r+s-k} \sum_{i=0}^{p-j+k,r} [r+p-j-2i+k-1, s+q+i-j-2k-1] \\
&\quad - \sum_{k=0}^{q-2,r+s} \sum_{j=0}^{s,p+1,r+s-k} \sum_{i=0,j-s+k}^{p+1-j+k,r} [r+p+1-j-2i+k, s+q+i-j-2k-2] \\
&\quad - \sum_{k=0}^{q-1,r+s} \sum_{j=0}^{s,p-1,r+s-k} \sum_{i=0,j-s+k}^{p-1-j+k,r} [r+p-j-2i+k-1, s+q+i-j-2k-1]
\end{aligned} \tag{4.12}$$

When $p=0$, the last triple sum of (4.12) is absent. We simplify (4.12) by first combining the second and fourth triple sums, then the third and fifth triple sums, and then combining the results and adding this to the first triple sum. We treat the case $p \neq 0$ and $p=0$ separately.

Case $p \neq 0$: Terms in the difference of the second and fourth triple sums will occur because of differences in the upper bound of the ranges of k, j, and i. A difference in the range of k will occur when $q-1 \leqslant r+s$. This difference will be the terms for $k=q-1$ in the second triple sum, that is

$$\left\{\text{if } q-1 \leqslant r+s, \sum_{j=0}^{s,p,r+s-q+1} \sum_{i=0,j-s+q-1}^{p-j+q-1,r} [r+p-j-2i-q, s-q+i-j]\right\}. \tag{4.13}$$

For $k<q-1$, a difference in the range of j will occur whenever $\min(s,p,r+s-k) = p$ and $\min(s,p+1,r+s-k) = p+1$. This will require $s \geqslant p+1$ and $r+s-k \geqslant p+1$, that is $k \leqslant r+s-p-1$. The difference will be the terms for $j=p+1$ in the fourth triple sum,

$$\left\{\text{if } s \geqslant p+1, -\sum_{k=0}^{q-2,r+s-p-1} \sum_{i=0,p+1-s+k}^{k,r} [r-2i+k, s+q+i-p-2k-3]\right\}. \tag{4.14}$$

For $k<q-1$ and $j<p+1$, difference in the range of i will occur when $\min(p+1-j+k,r) = p+1-j+k$ and $\min(p-j+k,r) = p-j+k$. This requires $r \geqslant p+1-j+k$ so $j \geqslant p+1-r+k$. The difference will be the terms for $i=p+1-j+k$ in the fourth triple sum, that is

$$-\sum_{k=0}^{q-2,r+s} \sum_{j=0,p+1-r+k}^{s,p,r+s-k} [r-p+j-k-1, s+q+p-2j-k-1]. \tag{4.15}$$

A similar analysis of the difference of the third and fifth triple sums gives

$$\left\{\text{if } s \geqslant p, \sum_{k=0}^{q-1,r+s-p} \sum_{i=0,p-s+k}^{k,r} [r-2i+k-1, s+q+i-p-2k-1]\right\}, \tag{4.16}$$

and

$$\sum_{k=0}^{q-1,r+s} \sum_{j=0,p-r+k}^{s,p-1,r+s-k} [r-p+j-k-1, s+q+p-2j-k-1]. \tag{4.17}$$

Now in (4.14) replace k by $k-1$ giving

$$\left\{\text{if } s \geqslant p+1, -\sum_{k=1}^{q-1,r+s-p} \sum_{i=0,p+k-s}^{k-1,r} [r-2i+k-1, s+q+i-p-2k-1]\right\} \tag{4.18}$$

and combine this with (4.16). If $s=p$ then $i=k$ giving

$$\sum_{k=0}^{q-1,r} [r-k-1, q-k-1].$$

Since the term here for $k=r$ is redundant we may write this as

$$\sum_{k=0}^{q-1,r-1} [r-k-1, q-k-1]. \tag{4.19a}$$

Similarly if $s \geqslant p+1$, (4.16) and (4.18) combine to give the term for $k=0$ (and hence $i=0$) in (4.16) plus the terms with $i=k$ in (4.16) whenever $r \geqslant k$. These are

$$[r-1, s+q-p-1]$$
$$+\sum_{k=1}^{q-1,r}[r-k-1, s+q-p-k-1]$$
$$=\sum_{k=0}^{q-1,r}[r-k-1, s+q-p-k-1]. \tag{4.19b}$$

In this the term for $k=r$ is redundant. So (4.19a) and (4.19b) may be written commonly as

$$\left\{\text{if } s\geqslant p, \sum_{k=0}^{q-1,r-1}[r-k-1, s+q-p-k-1]\right\}. \tag{4.20}$$

Similarly (4.15) and (4.17) combine giving terms

(i) for $k=q-1$ in (4.17) whenever $r+s\geqslant q-1$ and when $k<q-1$;

(ii) for $j=p$ in (4.15) whenever $\min(s,r+s-k)\geqslant p$; and

(iii) for $j=p-r+k$ in (4.17) whenever $p-r+k\geqslant 0$.

[All these terms under (iii) are, however, redundant].

The terms of (i) are

$$\sum_{j=0,p-r+q-1}^{s,p-1,r+s-q+1}[r-p-q+j, s+p-2j].$$

Here the terms for $j=p-r+q-1$ and $j=r+s-q+1$ (if they occur) are redundant. As the latter is the only term when $r+s=q-1$, the terms of (i) reduce to

$$\left\{\text{if } r+s\geqslant q, \sum_{j=0,p-r+q}^{s,p-1,r+s-q}[r-p-q+j, s+p-2j]\right\}. \tag{4.21}$$

The terms of (ii) will occur when $s\geqslant p$ and $k\leqslant r+s-p$. They will be

$$-\sum_{k=0}^{q-2,r+s-p}[r-k-1, s+q-p-k-1].$$

In this sum the terms are redundant for $k\geqslant r$, so (ii) gives

$$\left\{\text{if } s\geqslant p, -\sum_{k=0}^{q-2,r-1}[r-k-1, s+q-p-k-1]\right\}. \tag{4.22}$$

Equations (4.20) and (4.22) combine to give a single term

$$\{\text{if } r\geqslant q, s\geqslant p, [r-q, s-p]\}. \tag{4.23}$$

But this is exactly the term under the correct conditions for including a term with $j=p$ in (4.21) giving

$$\left\{\text{if } r+s\geqslant q, \sum_{j=0,p-r+q}^{s,p,r+s-q}[r-p-q+j, s+p-2j]\right\}. \tag{4.24}$$

So

$$\begin{aligned}(4.24) &= (4.23)+(4.21)\\ &= (4.20)+(4.22)+(4.21)\\ &= (4.15)+(4.17)+(4.20)\\ &= (4.15)+(4.17)+(4.14)+(4.16)\end{aligned}$$

To have the sum of the last four triple sums of (4.12) we must add (4.13) to (4.24). Note that in (4.13), if $j=r+s-q+1$, the terms are of the form $[p-s+2q-2i-1, i-r-1]$ and are redundant since $i\leqslant r$. In particular if $r+s=q-1$, (4.13) is 0. Also for all j, the terms for $i=j-s+q-1$ are redundant. So (4.13) may be rewritten

$$\left\{\text{if } r+s>q, \sum_{j=0}^{s,p,r+s-q}\sum_{i=0,j-s+q}^{p-j+q-1,r}[r+p-j-2i+q, s-q+i-j]\right\}.$$

Adding this to (4.24) gives

$$\left\{\text{if } r+s>q, \sum_{j=0}^{s,p,r+s-q}\sum_{i=0,j-s+q}^{p-j+q,r}[r+p-j-2i+q, s-q+i-j]\right\}. \tag{4.25}$$

Case $p=0$: Here, the fifth triple sum of (4.12) does not occur and the third triple sum is

$$\sum_{k=0}^{q-1,r+s}\sum_{i=0,-s+k}^{k,r}[r-2i+k-1, s+q+i-2k-1]$$

which equals (4.16) with $p=0$. So the last four triple sums reduce to

$$(4.13)+(4.14)+(4.15)+(4.16)=(4.13)+(4.18)+(4.15)+(4.16)=(4.13)+(4.15)+(4.20) \text{ with } p=0.$$

Equation (4.13) becomes

$$\left\{\text{if } q-1\leqslant r+s, \sum_{i=0,-s+q-1}^{q-1,r}[r-2i+q, s-q+i]\right\}.$$

Here the terms for $i=-s+q-1$ (which is the only term when $r+s=q-1$) is redundant. So (4.13) reduces to

$$\left\{\text{if } q\leqslant r+s, \sum_{i=0,-s+q}^{q-1,r}[r-2i+q, s-q+i]\right\}. \tag{4.26}$$

Also (4.15) becomes $\Sigma_{k=0}^{q-2,r+s}[r-k-1, s+q-k-1]$ in which all terms are redundant for $k\geqslant r$, giving $-\Sigma_{k=0}^{q-2,r-1}[r-k-1, s+q-k-1]$. Combining this with (4.20) which becomes (for $p=0$) $\Sigma_{k=0}^{q-1,r-1}[r-k-1, s+q-k-1]$ gives $\{\text{if } q\leqslant r, [r-q, s]\}$. It is convenient to write this as $\{\text{if } q\leqslant r+s, [r-q, s]\}$ (introducing redundant terms for $r<q\leqslant r+s$). Combining this with (4.26) we get for the sum of the last four triple sums in the case $p=0$,

$$\left\{\text{if } q\leqslant r+s, \sum_{i=0,-s+q}^{q,r}[r-2i+q, s-q+i]\right\}.$$

This is again (4.25) with $p=0$.

All that remains is to combine (4.25) to the first triple sum of (4.12). This has the effect of extending the upper index for k to $\min(q,r+s)$ which give (4.6), as required, for the product $[p,q][r,s]$. Q.E.D.

5. MULTIPLICITY FORMULA

We now determine the conditions necessary for a term $[m,n]$ to occur in the product $[p,q][r,s]$ and the multiplicity with which this term occurs. The equations

$$\begin{aligned}r+p-k-2i+k &= m,\\ s+q+i-j-2k &= n,\end{aligned} \tag{5.1}$$

must be solved for i, j, and k subject to the conditions

$$0\leqslant k\leqslant \min(q,r+s),$$
$$0\leqslant j\leqslant \min(s,p,r+s-k), \tag{5.2}$$
$$\max(0,j-s+k)\leqslant i\leqslant \min(p-j+k,r).$$

For a fixed value of k, (5.1) has a unique solution for i and j, namely,

$$i=\tfrac{1}{3}\{(r-s)+(p-q)-(m-n)\}+k,$$
$$j=\tfrac{1}{3}\{(r+2s)+(p+2q)-(m+2n)\}-k. \tag{5.3}$$

Since i and j are integers this requires that

$$3 \text{ divides } (r-s)+(p-q)-(m-n). \tag{5.4}$$

Write

$$C=\tfrac{1}{3}\{(r-s)+(p-q)-(m-n)\}$$

and

$$D=\tfrac{1}{3}\{(r+2s)+(p+2q)-(m+2n)\}$$

so (5.3) becomes $i=C+k, j=D-k$,

and substituting this is (5.2) gives necessary and sufficient conditions for k, namely,

$$0\leqslant k\leqslant \min(q,r+s), \tag{5.5a}$$
$$0\leqslant D-k\leqslant \min(s,q,r+s), \tag{5.5b}$$
$$\max(0,D-s)\leqslant C+k\leqslant \min(p-D+2k,r). \tag{5.5c}$$

Equation (5.5b) may be rewritten as separate inequalities $k\leqslant D, k\geqslant D-s, k\geqslant D-p, D\leqslant r+s$ and (5.5c) may be rewritten $k\geqslant \max(-C, D-C-s), k\leqslant -C+r, k\geqslant C+D-p$. Collecting these we obtain

$$\max(0, D-s, D-p, -C, D-C-s, D+C-p)\leqslant k\leqslant \min(q, D, r-C) \text{ and } D\leqslant r+s. \tag{5.6}$$

In particular a solution for k requires that

$$\max(0, D-s, D-p, -C, D-C-s, D+C-p)\leqslant \min(q, D, r-C) \text{ and } D\leqslant r+s. \tag{5.7}$$

Expressing (5.7) as separate simultaneous equalities gives

$$D\geqslant 0,\ C\leqslant r,\ D\leqslant s+q,\ D\leqslant r+2s,\ D+C\leqslant r+s,$$
$$D\leqslant p+q,\ D\leqslant p+r+s,$$
$$D+C\leqslant p+r,\ -C\leqslant q,\ -C\leqslant r+s,\ -C\leqslant D,$$
$$D+C\leqslant q+s,\ D-C\leqslant r+2s,$$
$$-C\leqslant s,\ D\leqslant r+s,\ D+C\leqslant p+q,\ D+C\leqslant p+r+s,$$
$$C\leqslant p,\ D+2C\leqslant r+p.$$

Some of these inequalities are implied by others:

$$-C\leqslant s\Rightarrow -C\leqslant r+s,$$
$$D+C\leqslant p+r\Rightarrow D+C\leqslant p+r+s,$$
$$D\leqslant r+s\Rightarrow D\leqslant r+2s,$$
$$D+C\leqslant p+r \text{ and } -C\leqslant s\Rightarrow D\leqslant p+r+s.$$

Further

$$D-C\leqslant q+s\Leftrightarrow s+q-n\leqslant s+q\Leftrightarrow n\geqslant 0$$

and

$$D+2C\leqslant r+p\Leftrightarrow r+p-m\leqslant r+p\Leftrightarrow m\geqslant 0.$$

Collecting these inequalities we get

Proposition 7(a): The necessary and sufficient conditions for $[m,n]$ to appear as a summand in the product $[r,s][p,q]$ are, if $C=\tfrac{1}{3}\{(r-s)+(p-q)-(m-n)\}$ and $D=\tfrac{1}{3}\{(r+2s)+(p+2q)-(m+2n)\}$ that

(i) C (and hence D) are integers satisfying

(ii) $0\leqslant D\leqslant \min(s+q, p+q, r+s)$,
$-\min(q, s)\leqslant C\leqslant \min(r,p)$,
$0\leqslant D+C\leqslant \min(r+s, p+r, p+q)$.

Since each value of k satisfying (5.6) gives a unique solution for i and j, the multiplicity of $[m,n]$ in (4.6) when the conditions of Proposition 7(a) are satisfied is

$$\begin{aligned}&1+\min(q,r+s,D,r-C)-\max(0,D-s,D-p,\\&-C,D-C-s,D+C-p)\\&=1+\min(q,q+s-D,q+p-D,q+C,n,\\&q+p-D-C,\\&D,s,p,D+C,D+n-q,p-C,r-C,\\&r+s-C-D,\\&r+p-C-D,r+n-C-q,r,r+p-D-2C).\end{aligned}$$

Note that $r+p-D-2C=m, r+n-C-q=r+s-D$ and that the terms apart from p, q, r, s, m, n are the differences of D, C, and $D+C$, respectively, and the end points of the intervals in which they lie according to Proposition 7(a).

Proposition 7(b): If the conditions of Proposition 7(a) are satisfied and if a, b, c are the minimum differences of D, C, and $D+C$, respectively, from the end points of the intervals in which they are thus constrained to lie, then the multiplicity of $[m, n]$ is

$$1+\min(p, q, r, s, m, n, a, b, c).$$

An alternative explicit formula for this multiciplicity may be found in Biedenharn and Louck,[2] Eq. (3.5) using the pattern multiciplicity determined in Eq. (2.17) of Louck, Lake, and Biedenharn.[13]

[1] S. Gasiorowicz, *Elementary Particle Physics* (Wiley, New York, 1967), pp. 257 ff.
[2] L. C. Biedenharn and J. D. Louck, J. Math. Phys. **13**, 1992 (1972).
[3] J. D. Louck, M. A. Lake, and L. C. Biedenharn, J. Math. Phys. **16**, 2411 (1975).

Addition formula for the $Z_N \times Z_N$ symmetric solutions of the factorization equations

Mauro M. Doria[a)]
J. W. Gibbs Laboratory, Yale University, New Haven, Connecticut 06511

(Received 11 February 1982; accepted for publication 18 June 1982)

An addition formula is derived which contains the addition relations for theta functions with characteristic proportional to $1/N$, N integer.

PACS numbers: 02.30.Dk

I. INTRODUCTION

Theta functions with characteristic proportional to $1/N$ have now become a subject of interest in physics. They provide the solution of the factorization equations associated with the symmetry $Z_N \times Z_N$. The factorization equations are known to be responsible for major developments in two-dimensional physics. They were first introduced by Yang[1] in the nonrelativistic quantum theory of N boson particles interacting via a two-body delta function potential in one spatial dimension. The factorization equations appeared there as conditions for the scattering matrix of two particles that allow the exact diagonalization of the Hamiltonian by means of the Bethe *ansatz* method. Later the factorization equations were rediscovered by Baxter[2] in his solution of the eight-vertex lattice model. There they correspond to the condition of commutativity of the transfer matrix at different values of the spectral parameter. Factorization equations were also extensively used to compute two-dimensional quantum S-matrices that have the property of factorizability.[3,4] This property means that the S-matrix for the scattering of N particles can be represented as the product of two-body S-matrices. For completely integrable field theories, the use of factorization equations plus unitarity and crossing symmetry of the S-matrix makes it possible to calculate the S-matrix up to CDD ambiguities.[3,4] The quantum inverse scattering method recently proposed by Faddeev *et al.*[5] also relies on these equations for finding the energy spectrum of several quantum theories through its algebraization of the Bethe *ansatz*.

The connection of factorization equations with abelian manifolds was first pointed out by Cherednik.[6] D. V. Chudnovsky and G. V. Chudnovsky,[7] and, independently, Belavin[8] have found that the solutions of the factorization equations with symmetry $Z_N \times Z_N$ are given by ratios of theta functions with characteristic $1/N$. In Ref. 7 it is claimed that these solutions provide immediate generalizations of the XYZ model for spin variables that correspond to the Nth root of unity. The same functions have also made their appearance in G. 't Hooft's analysis of the properties of electric and magnetic fluxes in SU(N) gauge theories.[9]

If one wishes to obtain physical theories from the $Z_N \times Z_N$ symmetric solutions of the factorization equation, it is important to have in advance equations that relate these theta functions for different arguments. The factorization equations themselves involve unknown functions at different points of their arguments. Such relations are known as addition formulas.

In this paper an explicit derivation of these addition formulas for the theta functions with characteristic proportional to $1/N$ is given. One single formula including several addition formulas is obtained. It is conjectured that it contains all possible addition relations for these functions. To the best of our knowledge such a general theorem is not available in the mathematical literature, although special cases of it can be found.

The plan of the paper is the following: the definition of theta functions with characteristic proportional to $1/N$ and some of their properties are introduced in Sec. II, and in Sec. III the addition formula for these functions is derived. Some useful mathematical relations are relegated to the Appendix.

II. GENERAL PROPERTIES

Theta functions of characteristic proportional to $1/N$ (N integer) are functions of two continuous complex variables x and τ that can be defined by the series

$$\Theta^{(N)}_{\epsilon\mu}(x,\tau) = \sum_{n=-\infty}^{+\infty} e^{i\pi\{(n+\epsilon/N)^2\tau + 2(n+\epsilon/N)(x+\mu/N)\}}. \quad (1.1)$$

The condition $\mathrm{Im}\tau \geqslant 0$ is imposed in order to assure convergence. The characteristic is defined as the pair of integer numbers (ϵ,μ) such that each one can take values from zero to $N-1$. Therefore, for the case $N=2$ they correspond to the well-known Jacobi's theta functions.[10] There are N^2 independent functions and except for the case $N=2$ arbitrary (ϵ,μ) or $N>2$ with $\epsilon=\mu=0$, parity is not well defined for these functions. From the definition (1.1) it is straightforward to deduce the following properties:

$$\Theta_{\epsilon+\epsilon',\mu+\mu'}(x,\tau) = \exp\left(i\pi\left\{\frac{\epsilon'^2}{N^2}\tau + \frac{2\epsilon'}{N}\left[x+\frac{1}{N}(\mu+\mu')\right]\right\}\right) \times \Theta_{\epsilon\mu}\left(x+\frac{\epsilon'}{N}\tau+\frac{\mu'}{N},\tau\right), \quad (1.2)$$

$$\Theta_{\epsilon+N,\mu}(x,\tau) = \Theta_{\epsilon\mu}(x,\tau), \quad (1.3)$$

$$\Theta_{\epsilon,\mu+N}(x,\tau) = e^{(2\pi i/N)\epsilon}\Theta_{\epsilon\mu}(x,\tau), \quad (1.4)$$

$$\Theta_{\epsilon\mu}(-x,\tau) = \Theta_{-\epsilon,-\mu}(x,\tau). \quad (1.5)$$

We have suppressed the index N for reasons of economy.

An interesting and well-known property[10] of the function of $\Theta_{00}(x,\tau)$ is that it admits a product expansion. Equation (1.2) makes readily available similar product expansions for the above-defined theta functions if one considers the appropriate shifts on $\Theta_{00}(x,\tau)$;

a) Supported by Conselho Nacional de Desenvolvimento Cientifico e Tecnologico-Brasil.

 0022-2488/82/112029-04$02.50 © 1982 American Institute of Physics

$$\Theta_{\epsilon\mu}(x,\tau)=\exp\left[2\pi i\frac{\epsilon}{N}\left(x+\frac{\mu}{N}\right)\right]q^{\epsilon^2/N^2}\prod_{n=1}^{\infty}(1-q^{2n})$$
$$\times\prod_{k=1}^{\infty}\left[1+2q^{2k-1}\cos 2\pi\left(x+\frac{\epsilon}{N}\tau+\frac{\mu}{N}\right)+q^{4k-2}\right], \tag{1.6}$$

where it is understood $q=e^{i\pi\tau}$.

III. ADDITION FORMULAS

The general addition formula for the theta functions of characteristic proportional to $1/N$ will be obtained in two steps. The first one consists in deriving a particular addition formula that for $N=2$ reduces to Jacobi's fundamental theorem.[10] For $N=3$ it gives a result derived once by Krazer.[11] The second step corresponds to taking appropriate shifts in the arguments of the theta functions present in this particular addition formula and getting a more general result. Property (1.2) is used to achieve this goal.

Consider $2N$ theta functions $\Theta_{\epsilon\mu}(x_i,\tau)$ with the same characteristic (ϵ,μ) and the same theta period τ.

A. Addition formula (1)

$$\sum_{\epsilon=0}^{N-1}\sum_{\mu=0}^{N-1}[\Theta_{\epsilon\mu}(t_1,\tau)\Theta_{\epsilon\mu}(t_2,\tau)\cdots\Theta_{\epsilon\mu}(t_{2N},\tau)]\exp\left(\frac{-2\pi i}{N}\epsilon\mu\right)$$
$$=N[\Theta_{00}(r_1,\tau)\Theta_{00}(r_2,\tau)\cdots\Theta_{00}(r_{2N},\tau)]. \tag{2.1}$$

The vectors $T=(t_1,t_2,\cdots,t_{2N})$ and $R=(r_1,r_2,\cdots,r_{2N})$ are related by the orthogonal transformation $R=(1/N)U_{(2N)}T$, where $U_{(2N)}$ is the matrix defined in the Appendix.

Proof: Consider the product of $2N$ theta functions $\Theta_{\epsilon\mu}(x_i,\tau)$ with the same characteristic (ϵ,μ) proportional to $1/N$ and the same theta period τ, but with independent arguments $x_i, 1\leqslant i\leqslant 2N$;

$$\Theta_{\epsilon\mu}(t_1,\tau)\Theta_{\epsilon\mu}(t_2,\tau)\cdots\Theta_{\epsilon\mu}(t_{2N},\tau)$$
$$=\sum_{n=-\infty}^{\infty}\exp\left(i\pi\left\{\left(n+\frac{\epsilon}{N}I\right)^T\left(n+\frac{\epsilon}{N}I\right)\tau\right.\right.$$
$$\left.\left.+2\left(n+\frac{\epsilon}{N}I\right)^T\left(T+\frac{\mu}{N}I\right)\right\}\right). \tag{2.2}$$

In addition to the vector T, this equation makes use of the vectors $n=(n_1,n_2,\cdots,n_{2N})$ and I which has all of its $2N$ components equal to one. The summation symbol should be understood as meaning

$$\sum_{n=-\infty}^{\infty}\equiv\sum_{n_1=-\infty}^{\infty}\sum_{n_2=-\infty}^{\infty}\cdots\sum_{n_{2N}=-\infty}^{\infty}.$$

Half of the exponential term that is n-independent goes to the left-hand side and the μ-dependent term is made explicit on the right side,

$$[\Theta_{\epsilon\mu}(t_1,\tau)\Theta_{\epsilon\mu}(t_2,\tau)\cdots\Theta_{\epsilon\mu}(t_{2N},\tau)]$$
$$\times\exp\left(-\frac{2\pi i}{N}\epsilon\mu\right)=\sum_{n=-\infty}^{\infty}\left[\exp\frac{2\pi i}{N}(n^TI+\epsilon)\right]^{\mu}$$
$$\times\exp\left(i\pi\left\{\left(n+\frac{\epsilon}{N}I\right)^T\left(n+\frac{\epsilon}{N}I\right)\tau\right.\right.$$
$$\left.\left.+2\left(n+\frac{\epsilon}{N}I\right)^TT\right\}\right) \tag{2.3}$$

The left-hand side becomes a sum over all N theta products with the same ϵ, when we sum over μ in the above equation. One can get rid of the μ-dependence inside the summation. Then a constraint has to be introduced on the integer values that the summation variables n_i can assume.

$$\sum_{\mu=0}^{N-1}\left[\exp\left\{\frac{2\pi i}{N}(n^TI+\epsilon)\right\}\right]^{\mu}$$
$$=\begin{cases}N \text{ if } n^TI+\epsilon=Nk,\\ 0 \text{ if } n^TI+\epsilon\neq Nk,\end{cases} \tag{2.4}$$

where from now on k is any integer real number. Therefore one obtains

$$\sum_{\mu=0}^{N-1}[\Theta_{\epsilon\mu}(t_1,\tau)\Theta_{\epsilon\mu}(t_2,\tau)\cdots\Theta_{\epsilon\mu}(t_{2N},\tau)]\exp\left(\frac{-2\pi i}{N}\epsilon\mu\right)$$
$$=N\sum_{n=-\infty}^{\infty}\exp\left(i\pi\left\{\left(n+\frac{\epsilon}{N}I\right)^T\left(n+\frac{\epsilon}{N}I\right)\tau\right.\right.$$
$$\left.\left.+2\left(n+\frac{\epsilon}{N}I\right)^TT\right\}\right)$$
$$n^TI+\epsilon=Nk. \tag{2.5}$$

Take the vectors $(n+(\epsilon/N)I)$ and T and rotate them by the orthogonal matrix (A5) defined in the Appendix. One gets two new vectors, defined by

$$p^{(\epsilon)}=\frac{1}{N}U_{(2N)}\left(n+\frac{\epsilon}{N}I\right)=\frac{1}{N}U_{(2N)}n+\frac{\epsilon}{N}I \tag{2.6}$$

and

$$R=\frac{1}{N}U_{(2N)}T. \tag{2.7}$$

The vector $p^{(\epsilon)}$ will play the role of the new summation variables. Its components are not always integer numbers, as can be deduced from (A7),

$$p_k{}^{(\epsilon)}=\frac{1}{N}(n_1+n_2+\cdots+n_{2N}+\epsilon)-n_k. \tag{2.8}$$

Nevertheless the constraint imposed on the n-summation, $n^TI+\epsilon=Nk$, is exactly the condition required for $p_k{}^{(\epsilon)}$ to be an integer number. Therefore *noninteger values for* $p^{(\epsilon)}$ *should be simply disregarded*. The constraint has to be rewritten in terms of the new integer variables $p_k{}^{(\epsilon)}$ because it also imposes a condition on them. Take Eq. (2.6) and apply the orthogonal transformation defined in the Appendix. Thanks to properties (A2) and (A3), one gets

$$\frac{1}{N}U_{(2N)}p^{(\epsilon)}=n+\frac{\epsilon}{N}I. \tag{2.9}$$

Transpose the above equation and sandwich it with the unit vector I. Then one obtains the constraint in the new form

$$p^{(\epsilon)T}I=Nk+\epsilon. \tag{2.10}$$

Now by use of the orthogonality property (A6), one gets

$$\sum_{\mu=0}^{N-1}\left[\Theta_{\epsilon\mu}(t_1,\tau)\Theta_{\epsilon\mu}(t_2,\tau)\cdots\Theta_{\epsilon\mu}(t_{2N},\tau)\right]\exp\left(\frac{-2\pi i}{N}\epsilon\mu\right)$$
$$=N\sum_{p^{(\epsilon)}=-\infty}^{\infty}\exp\left(i\pi\{p^{(\epsilon)T}p^{(\epsilon)}\tau+2p^{(\epsilon)T}R\}\right),$$
$$p^{(\epsilon)T}I=Nk+\epsilon. \tag{2.11}$$

The constraint is dropped if we sum over ϵ from zero to $N-1$ because all integer numbers admit one of the forms $Nk, Nk+1, \cdots, Nk+N-1$. Therefore there are no more requirements over the components $p_k{}^{(\epsilon)}$ and the right-hand side becomes a product of $2N$ theta functions of null characteristic

$$\sum_{\epsilon=0}^{N-1}\sum_{p^{(\epsilon)}=-\infty}^{\infty}\exp\left(i\pi\{p^{(\epsilon)T}p^{(\epsilon)}\tau+2p^{(\epsilon)T}R\}\right)$$
$$=\Theta_{00}(r_1,\tau)\Theta_{00}(r_2,\tau)\cdots\Theta_{00}(r_{2N},\tau), p^{(\epsilon)T}I=Nk+\epsilon. \tag{2.12}$$

B. Addition formula (2)

$$N\left[\Theta_{\epsilon_1\mu_1}(y_1)\Theta_{\epsilon_2\mu_2}(y_2)\cdots\Theta_{\epsilon_{2N}\mu_{2N}}(y_{2N})\right]$$
$$=\sum_{\epsilon=0}^{N-1}\sum_{\mu=0}^{N-1}\exp\left(\frac{-2\pi i}{N}(f+\epsilon)\mu\right)\left[\Theta_{\epsilon+\epsilon_1',\mu+\mu_1'}(x_1,\tau)\right.$$
$$\left.\times\Theta_{\epsilon+\epsilon_2',\mu+\mu_2'}(x_2,\tau)\cdots\Theta_{\epsilon+\epsilon_{2N}',\mu+\mu_{2N}'}(x_{2N},\tau)\right], \tag{2.13}$$

where

$$f=\frac{1}{N}(\epsilon_1+\epsilon_2+\ldots+\epsilon_{2N}),$$
$$g=\frac{1}{N}(\mu_1+\mu_2+\ldots+\mu_{2N}), \tag{2.14}$$
$$\epsilon_i'=f-\epsilon_i,\quad \mu_i'=g-\mu_i, \tag{2.15}$$
$$X=\frac{1}{N}U_{(2N)}Y. \tag{2.16}$$

Proof: Make the transformation

$$r_k=y_k+\frac{\epsilon_k}{N}\tau+\frac{\mu_k}{N} \tag{2.17}$$

in the right-hand side of the addition formula (2.1), where (ϵ_k,μ_k) are any integer numbers. Looking at the previously defined vector T, one obtains

$$t_k=x_k+\frac{(f-\epsilon_k)}{N}\tau+\frac{(g-\mu_k)}{N}, \tag{2.18}$$

where x_k is related to y_k by (2.16). Property (1.2) enables us to relate a theta function of general characteristic with the theta function of zero characteristic

$$\Theta_{00}(r_k,\tau)=\exp\left(-i\pi\left\{\frac{\epsilon_k^2}{N^2}\tau+2\frac{\epsilon_k}{N}\left[y_k+\frac{\mu_k}{N}\right]\right\}\right)$$
$$\times\Theta_{\epsilon_k\mu_k}(y_k,\tau). \tag{2.19}$$

The product of $2N$ null-characteristic theta functions where each one is independently shifted by (2.17) is

$$\Theta_{00}(r_1,\tau)\Theta_{00}(r_2,\tau)\cdots\Theta_{00}(r_{2N},\tau)$$
$$=\exp\left(-i\pi\left\{\frac{\tau}{N^2}\left(\sum_{k=1}^{2N}\epsilon_k^2\right)+\frac{2}{N}\left(\sum_{k=1}^{2N}\epsilon_k y_k\right)\right.\right.$$
$$\left.\left.+\frac{2}{N^2}\left(\sum_{k=1}^{2N}\epsilon_k\mu_k\right)\right\}\right)\Theta_{\epsilon_1\mu_1}(y_1,\tau)\Theta_{\epsilon_2\mu_2}(y_2,\tau)\cdots$$
$$\times\Theta_{\epsilon_{2N}\mu_{2N}}(y_{2N},\tau). \tag{2.20}$$

Following the same pattern the use of (1.2) gives

$$\Theta_{\epsilon\mu}(t_k,\tau)=\exp\left(-i\pi\left\{\frac{(f-\epsilon_k)^2}{N^2}\tau+2\frac{(f-\epsilon_k)}{N}\right.\right.$$
$$\left.\left.\times\left[x_k+\frac{g-\mu_k}{N}+\frac{\mu}{N}\right]\right\}\right)$$
$$\times\exp\left(-2\pi i\frac{f\mu}{N}\right)\Theta_{\epsilon+f-\epsilon_k,\mu+g-\mu_k}(x_k,\tau). \tag{2.21}$$

It is straightforward to derive similar results for the product of $2N$ theta functions of characteristic (ϵ,μ) independently shifted.

$$\Theta_{\epsilon\mu}(t_1,\tau)\Theta_{\epsilon\mu}(t_2,\tau)\ldots\Theta_{\epsilon\mu}(t_{2N},\tau)$$
$$=\exp\left(-i\pi\left\{\frac{\tau}{N^2}\left(\sum_{k=1}^{2N}\epsilon_k^2\right)+\frac{2}{N}\right.\right.$$
$$\left.\left.\times\sum_{k=1}^{2N}(f-\epsilon_k)x_k+\frac{2}{N^2}\left(\sum_{k=1}^{2N}\epsilon_k\mu_k\right)\right\}\right)$$
$$\times\left[\Theta_{\epsilon+f-\epsilon_1,\mu+g-\mu_1}(x_1,\tau)\Theta_{\epsilon+f-\epsilon_2,\mu+g-\mu_2}(x_2,\tau)\ldots\right.$$
$$\left.\times\Theta_{\epsilon+f-\epsilon_{2N},\mu+g-\mu_{2N}}(x_{2N},\tau)\right]. \tag{2.22}$$

When expressions (2.20) and (2.22) are introduced in the addition formula (2.1) all the ϵ_k, μ_k, x_k, and y_k dependence in their exponents cancel against each other. This can be checked with the help of properties (A7) and (A8) and so the addition formula (2) is obtained. For the case $N=2$ it coincides with the result of Smith[12] that leads to all possible addition formulas for Jacobi's theta functions. For arbitrary N and special values of the arguments y_k and of the characteristics (ϵ_k,μ_k), this addition formula should degenerate into the one contained in Krazer's book.[13]

ACKNOWLEDGMENTS

The author thanks Hsiung Chia Tze and is specially indebted to Feza Gürsey for helpful discussions. This research is in part supported by the U. S. Department of Energy under contract No. DE-AC-02-76ER03075.

APPENDIX

Consider the $(2N)\times(2N)$ matrix $U_{(2N)}$ defined by its elements

$$U_{ij}=\begin{cases}1 \text{ if } i\neq j,\\ 1-N \text{ if } i=j,\end{cases}\quad 1\leqslant i,j\leqslant 2N. \tag{A1}$$

This matrix has the two properties

$$U_{(2N)}^T=U_{(2N)} \tag{A2}$$

and

$$U_{(2N)}^2=N^2I_{(2N)}, \tag{A3}$$

where $I_{(2N)}$ is the $(2N)\times(2N)$ identity matrix.

For a vector $a=(a_1,a_2,\ldots,a_{2N})$ define A as the sum over all its $2N$ components,

$$A=\sum_{i=1}^{2N}a_i. \tag{A4}$$

Consider the orthogonal transformation

$$a'=\frac{1}{N}U_{(2N)}a. \tag{A5}$$

For any two vectors a and b, we have

$$a'^{T}b' = a^{T}b. \tag{A6}$$

The components of the rotated vector are given by

$$a_i' = \frac{1}{N}A - a_i \tag{A7}$$

and it follows that

$$A' = A. \tag{A8}$$

In particular, for the unit vector I, defined with all its $2N$ components equal to one, we obtain

$$\frac{1}{N}U_{(2N)}I = I. \tag{A9}$$

[1]C. N. Yang, Phys. Rev. Lett. **19**, 1312 (1967).
[2]R. J. Baxter, Ann. Phys. **70**, 193 (1972).
[3]A. B. Zamolodchikov and Al. B. Zamolodchikov, Ann. Phys. **120**, 253 (1979).
[4]M. Karowski, G. Thun, T. Troung, and P. Weiss, Phys. Lett. B **67**, 321 (1977).
[5]L. D. Faddeev, E. K. Sklyanin, and E. A. Takhtajan, Teor. Mat. Fiz. **40**, 107 (1980).
[6]J. Cherednick, Dokl. Acad. Sci. USSR **249**, 1095 (1979).
[7]D. V. Chudnovsky and G. V. Chudnovsky, Saclay Preprint DPh-T/80/131.
[8]A. A. Belavin, Nucl. Phys. B **180**, 189–250 (1981).
[9]G. 't Hooft, Nucl. Phys. B **153**, 141 (1979).
[10]H. Hancock, *Theory of Elliptic Functions* (Dover, New York, 1958).
[11]A. Krazer, Math. Ann. **22**, 416 (1883).
[12]H. J. S. Smith, Proc. London Math. Soc. **I**, 1 (1866).
[13]A. Krazer, *Lehrbuch der Thetafunktionen* (Teubner, Leipzig, 1903), Sec. 6, Chap. 8.

On a unified approach to transformations and elementary solutions of Painlevé equations

A. S. Fokas and M. J. Ablowitz
Department of Mathematics and Computer Science, Clarkson College of Technology, Potsdam, New York 13676

(Received 9 June 1981; accepted for publication 11 August 1981)

An algorithmic method is developed for investigating the transformation properties of second-order equations of Painlevé type. This method, which utilizes the singularity structure of these equations, yields explicit transformations which relate solutions of the Painlevé equations II–VI, with different parameters. These transformations easily generate rational and other elementary solutions of the equations. The relationship between Painlevé equations and certain new equations quadratic in the second derivative of Painlevé type is also discussed.

PACS numbers: 02.30.Dk

I. INTRODUCTION

We say than an equation is of Painlevé type if all its solutions possess the Painlevé property, i.e., their only singularities are poles or nonmovable critical points.[1] The most well known second-order equations of Painlevé type are the so called six Painlevé equations PI–PVI[1] discovered by Painlevé[2] and his school[3] at the turn of the century. They classified all equations of the form $w'' = F(w', w, z)$, where F is rational in w', algebraic in w, and locally analytic in z, which have the Painlevé property. They found that, within a Möbius[4] transformation, there exist fifty such equations. Distinguished among these fifty equations are PI–PVI. Any other of the fifty equations can either be integrated in terms of known functions or can be reduced to one of these six equations. Although PI–PVI were first discovered from strictly mathematical considerations, they have recently appeared in several physical applications (see, for example Refs. 5–7).

Explicit transformations and relevant exact solutions admitted by the Painlevé equations first appeared in the Soviet literature[8] and are summarized in Ref. 9; the main points are as follows:

(i) For certain choices of the parameters, PII–V admit one-parameter families of solutions expressible in terms of the classical transcendental functions: Airy,[3,10] Bessel,[11] Weber–Hermite,[12] and Whittaker,[13] respectively.

(ii) PII–V admit transformations (see Ref. 14–17) which map solutions of a given Painlevé equation to solutions of the same equation but with different values of the parameters.

(iii) Using (ii) one can construct (for certain choices of the parameters) various elementary solutions of PII–V. These solutions are either rational or are functions which are related (through repeated differentiations and multiplications) to the above-mentioned classical transcendental functions.

(iv) For PVI it was only known that, for a certain single choice of its parameters, it admits a one-parameter family of solutions expressible in terms of hypergeometric functions.[18]

However, the above results apparently were obtained by rather ad hoc methods. Moreover, in spite of the extensive amount of research on Painlevé equations, the transformation properties of PVI were not found. Also, no other one-parameter family of solutions of PVI was found, save for the one mentioned above. It is important to note that PVI is in a sense the most general Painlevé equation since it contains the other five as limiting cases.[1]

In this paper:

(1) We develop an algorithmic method for systematically investigating the transformation properties of second-order equations of the Painlevé type. This method yields explicit transformations: (a) Between a given Painlevé equation and the same Painlevé equation but with different values of its parameters. (b) Between two different Painlevé equations (for example PIII and PV). (c) Between a Painlevé equation of the type investigated by Painlevé (i.e., linear in the second derivative) and an equation of the Painlevé type which is quadratic in the second derivative.

(2) As an application of this method we rederive the known transformation properties of PII–V and also derive the transformation properties of PVI. The latter are used to obtain (for various choices of parameters) one-parameter families of solutions of PVI. Among these solutions are rational solutions as well as solutions which are related (through repeated differentiations and multiplications) to hypergeometric functions.

(3) We relate PIII and PVI to certain new equations quadratic in the second derivative and of the Painlevé type.

Some of the results concerning PVI have been announced in Ref. 19.

A. The connection with inverse scattering and monodromy preserving transformations

In recent years considerable interst has developed in Painlevé equations. Ablowitz (Ramani) and Segur,[20] (Ref. 21) have discovered a deep connection between equations of Painlevé type and the PDE's solvable by the inverse scattering transform.[22] For example, PII and special cases of PIII and PIV may be obtained from the similarity reduction of the modified Korteweg-deVries, the sine-Gordon and the nonlinear Schrödinger equations, respectively. It is interesting that proper reductions of the Korteweg–deVries (KdV) equation lead to both PI and PII:

 0022-2488/82/112033-10$02.50 © 1982 American Institute of Physics

(i) KdV and PII. Consider the KdV equation in the form

$$u_t + 6uu_x + u_{xxx} = 0. \tag{1.1}$$

Equation (1.1) is clearly invariant under the group of transformations $x' = \lambda x$, $t' = \lambda^3 t$, $u' = \lambda^{-2} u$, where λ is some arbitrary parameter. The solutions of (1.1) invariant under this group of transformations (the so-called similarity solutions) are characterized by $u = (3t)^{-2/3} U(z)$, $z = x(3t)^{-1/3}$, where $U(z)$ satisfies

$$K_1(U) = U''' + 6UU' - (2U + zU') = 0. \tag{1.2}$$

Whitham[23] has noted that Eq. (1.2) is related to PII. Actually there is a one-to-one correspondence between the integrated form of (1.2) and PII. Equation (1.2) can be integrated once using the following identity:

$$[(2U - z)K_2(U)]' = (2U - z)K_1(U), \tag{1.3}$$

where

$$K_2(U) = U'' + 2U^2 - zU + \frac{\nu + U' - U'^2}{2U - z}. \tag{1.4}$$

Equation (1.4) is essentially equation PXXIV of Ref. 1 and is related by a one-to-one map with PII (see Sec. III).

(ii) KdV and PI. Equation (1.1) is also invariant under the group of transformations

$$x' = x + 6t\lambda, \quad t' = t + \lambda/\alpha, \quad u' = u + \lambda, \tag{1.5}$$

where λ and α are arbitrary parameters. Regarding α fixed, one immediately[24] obtains the following characterization for the solutions of (1.1) invariant under (1.5):

$$u = \alpha t + U(z), \quad z = x - 3\alpha t^2, \tag{1.6}$$

where

$$U''' + 6UU' + \alpha = 0. \tag{1.7}$$

Equation (1.7), upon integration, yields PI.

Using the ideas of Ablowitz and Segur one can characterize a nonelementary one-parameter family of solutions of, say, (1.2) through a Gel'fand–Levitan linear integral equation of the Fredholm type. Recently we have proposed a new method[25] for linearizing the Painlevé equations, using singular integral equations and Riemann–Hilbert theory. In this way we have characterized a three-parameter family of solutions of (1.2). This work suggests that the transformations given in this paper may be useful in obtaining the general solution of, say, PII using our method. It is an important fact that using these transformations one can find the general solution of a given Painlevé equation for arbitrary values of its parameters α, provided one knows the general solution of this equation for only a range of α. For example, for PII one needs to know the general solution only for $-1/2 < \alpha < 1/2$.

We also note that there is a close connection between ODE's of Painlevé type and monodromy-preserving deformations. This was emphasized and used by Flaschka and Newell[26] and by Sato *et al.*[27] In particular, Flaschka and Newell derived a formal system of linear singular integral equations from which the solutions of PII and of a special case of PIII are to be found. However, they did not investigate in general the question of existence of solutions of their integral equations.

From the above comments one sees the richness and broad mathematical content associated with the investigation of Painlevé equations. Undoubtedly, considerable research will continue in this area.

II. A METHOD FOR INVESTIGATING THE TRANSFORMATION PROPERTIES OF SECOND-ORDER EQUATIONS OF THE PAINLEVÉ TYPE

Suppose we are given one of the fifty equations found by Painlevé and his school, which we write in the form

$$v'' = P_1 v'^2 + P_2 v' + P_3, \tag{2.1}$$

where P_1, P_2, P_3 depend on v, z, and a set of parameters denoted here by α.

The first problem is to find the discrete Lie-point symmetries of (2.1), i.e., to find transformations of the form

$$\hat{v}(z; \hat{\alpha}) = F(v(z; \alpha), z), \tag{2.2}$$

where the function F is such that if $v(z; \alpha)$ solves (2.1) with parameters α, then $\hat{v}(z; \hat{\alpha})$ solves (2.1) with parameters $\hat{\alpha}$. Using the singularity structure of (2.1), the procedure for finding such transformations is immensely simplified; since the only transformation of the type (2.2) preserving the Painlevé property is the Möbius transformation, one immediately replaces (2.2) by

$$\hat{v}(z; \hat{\alpha}) = \frac{a_1 v + a_2}{a_3 v + a_4}, \tag{2.3}$$

where $a_1, \ldots, a_4$ are functions of z only. Using (2.3) the Lie-point discrete symmetries of (2.1) are easily obtained.

Having obtained the Lie-point symmetries of (2.1) one may look for generalized discrete symmetries of (2.1),[28] i.e., for transformations of the form

$$\hat{v}(z; \hat{\alpha}) = F(v'(z; \alpha), v(z; \alpha), z). \tag{2.4}$$

However, since we are not only interested in finding transformations relating a Painlevé equation to itself, but also relating two different equations of Painlevé type, we replace (2.4) by

$$u(z; \hat{\alpha}) = F(v'(z; \alpha), v(z; \alpha), z), \tag{2.5}$$

where F is such that u satisfies some second-order equation of the Painlevé type. The only transformation of the type (2.5), *linear in* v',[29] preserving the Painlevé property is the one involving the Riccati equation, i.e.,

$$u(z; \hat{\alpha}) = \frac{v' + av^2 + bv + c}{dv^2 + ev + f}, \tag{2.6}$$

where $a, b, \ldots, f$ depend on z only. Equation (2.6) plays a central role in our analysis.

The algorithm: Given Eq. (2.1) determine $a, \ldots, f$ by requiring that (2.6) define a one-to-one invertible map between solutions v of (2.1) and solutions u of some second-order equation of the Painlevé type. In this process the latter equation is completely determined.

Let us be more specific. Introducing the notation

$$J = dv^2 + ev + f, \quad Y = av^2 + bv + c, \tag{2.7}$$

differentiating (2.6), and using (2.1) to replace v'' and (2.6) to replace v', one obtains

$$Ju' = [P_1J^2 - 2dJv - eJ]u^2 + [-2P_1JY + JP_2 + 2avJ + bJ + 2dvY + eY - (d'v^2 + e'v + f')]u + [P_1Y^2 - P_2Y + P_3 - 2avY - bY + a'v^2 + b'v + c']. \quad (2.8)$$

There are two cases to be distinguished:

(A) Find $a, ..., f$ such that (2.8) reduces to a linear equation for v,

$$A(u', u, z)v + B(u', u, z) = 0. \quad (2.9)$$

Having determined $a, ..., f$ upon substitution of $v = -B/A$ in (2.6) one determines the equation for u, which is of the same type as (2.1) (i.e., it will be one of the fifty equations mentioned above).

(B) Find $a, ..., f$ such that (2.8) reduces to a quadratic equation for v,

$$A(u', u, z)v^2 + B(u', u, z)v + C(u', u, z) = 0. \quad (2.10)$$

Then (2.6) yields an equation for u which is quadratic in the second derivative. These types of equations, having the Painlevé property have not previously been considered in the literature.

Note that (i) It turns out that PII-PV admit transformations of both types (A) and (B). However, PVI does not admit a transformation of the type (A) above. (ii) Utilizing the way that the parameters α enter in the equation for u one can find a transformation relating Eq. (2.1) with different α's. (iii) When equations (2.9) and/or (2.10) break down (i.e., $A = B = C = 0$), they define one-parameter families of solutions of (2.1). Using these solutions and the transformation properties of (2.1) new one-parameter families of solutions can be obtained.

These points will be clarified after applying the above method to PII.

III. PAINLEVÉ II

In this section we use Painlevé II to illustrate: (a) How the transformation (2.6) can be used for investigating the transformation properties of a given equation. (b) How certain of these transformation properties can be used for obtaining elementary solutions.

Note that: (i) Here we look only for transformations of the type (A), i.e., we invoke (2.9). This is only for convenience. We stress that transformations of the type (B), [see (2.10)] exist for all PII–PVI. In this paper we shall consider such transformations by necessity for PVI [since transformations (A) do not exist in this case] and as an aid to the reader for PIII.

(ii) Painlevé and his school found that some of the fifty equations mentioned in Sec. I are related to PI–PVI. For example equations PXXXIV, PXXXV, PXLV, PXLVI, PXLVII of Ref. 1 are related to PII. An exhaustive investigation of transformations (A), not only establishes this relationship, but also gives a one-to-one correspondence between PII and each of the above equations. However, here we only present the relevant result for PXXXIV and also comment on PXXXV. We note that if one is simply interested in finding a transformation mapping PII to PII then any of the above transformations may be used.

Theorem 3.1: Let $v(z; \alpha)$ be a solution of PII

$$v'' = 2v^3 + zv + \alpha. \quad (3.1)$$

Then $\bar{v}(z; \bar{\alpha})$ are also solutions of PII, where

$$\bar{v}(z; \bar{\alpha}) = -v(z; \alpha); \quad \bar{\alpha} = -\alpha, \quad (3.2)$$

$$\bar{v}(z; \bar{\alpha}) = -v(z; \alpha) - \frac{1 + 2\alpha}{2v^2 + 2v' + z};$$

$$\bar{\alpha} = \alpha + 1, \quad \alpha \neq -\tfrac{1}{2}. \quad (3.3)$$

The case $\alpha = -1/2$ is considered in Lemma 3.1.

Theorem 3.2: Let $v(z; \alpha)$ be a solution of PII and let $u(z; \nu)$ be a solution of

$$u'' + 2u^2 - zu + \frac{\nu + u' - (u')^2}{2u - z} = 0; \quad \nu = \alpha(\alpha + 1). \quad (3.4)$$

Then there exists the following one-to-one correspondence between solutions of (3.1) and (3.4)

$$u = -v' - v^2, \quad v = \frac{u' + \alpha}{2u - z}. \quad (3.5)$$

Equation (3.4) under the transformation $w = (u - z/2)/(4\alpha + 1)$ reduces to PXXXIV of Ref. 1.

Lemma 3.1: PII admits a one-parameter family of solutions characterized by

$$v' + v^2 + z/2 = 0, \quad (3.6)$$

iff $\alpha = -1/2$.

A. Derivation of the above results

Comparing (3.1) and (2.1) one finds that $P_1 = P_2 = 0$ and $P_3 = 2v^3 + zv + \alpha$. In considering Eq. (2.8) one has to consider separately the two cases $d = 0$, and $d \neq 0$. Here we only consider $d = 0$. Then (2.8) becomes

$$u' = -eu^2 + \left[\frac{3aev^2 + (2af + 2eb - e')v + (bf + ec - f')}{ev + f}\right]u + \left[\frac{2(1 - a^2)v^3 + (a' - 3ab)v^2 + (z + b' - b^2 - 2ac)v + (\alpha - bc + c')}{ev + f}\right]. \quad (3.7)$$

Our goal now is to choose a, b, c, e, and f in such a way that (3.7) becomes a *linear* equation for v. It is clear that this will be the case if each of the above bracketed expressions (i.e., the coefficient of u and the term independent of u) is linear in v. Then, it is obvious that $a^2 = 1$ and that also $ev + f$ must divide each numerator appearing in the above brackets. Re-

quiring this to be the case with $e \neq 0$, one is led to establish a one-to-one correspondence between PII and PXXXV of Ref. 1. However, a simpler possible case is $e = 0$. Then, it is clear from Eq. (2.6) that one may take, without loss of generality, $f = -1$ [the minus sign is only for relating u directly to (1.4)] and $c = 0$, since one can always "absorb" them in u by a Möbius transformation. Hence, inserting $c = d = e = 0, f = -1$ in (3.7), this equation reduces to

$$u' = (2av + b)u + 3abv^2 - (z + b' - b^2)v - \alpha. \quad (3.8)$$

Thus necessarily $b = 0$. Hence, Eqs. (2.6) and (3.8) imply

$$(u' + \alpha) = v(2au - z),\ u = -(v' + av^2);\ a^2 = 1. \quad (3.9)$$

Taking for convenience $a = 1$ and substituting $v = (u' + \alpha)/(2u - z)$ in the above expression for u, Eq. (3.4) follows.

1. $\alpha = -1/2$

The transformation (3.5b) breaks down iff $u = z/2$. But then $\alpha + u'$ must be zero, or $\alpha = -1/2$. (Actually one can easily check that $u = z/2, \alpha = -1/2$ solve (3.4).) Hence Eq. (3.5a) implies Lemma 3.1.

2. The transformation from PII to PII

Using the above results (i.e., the results of Theorem 3.2) one can easily derive (3.3). The basic idea is to exploit the fact that ν in (3.4) is quadratic in α. Therefore, there exist two values of the parameter α, namely α and $-(\alpha + 1)$, which give the same value of ν and hence the same value of u, i.e., $u(z; \alpha) = u(z; -(\alpha + 1))$. But then

$$\begin{aligned}\bar{v}(z; -(\alpha + 1)) &= \frac{u'(z; -(\alpha + 1)) - (\alpha + 1)}{2u(z; -(\alpha + 1)) - z} \\ &= \frac{u'(z; \alpha) - (\alpha + 1)}{2u(z; \alpha) - z} \\ &= v(z; \alpha) - \frac{(2\alpha + 1)}{2u(z; \alpha) - z}.\end{aligned}$$

Hence, replacing u by $-(v^2 + v')$ and $\bar{v}(z; -(\alpha + 1))$ by $-v(z; (\alpha + 1))$, Eq. (3.3) follows.

B. How to obtain elementary solutions

The transformations (3.2) and (3.3) can be used to obtain all known elementary solutions of PII. Similarly, one can use transformations (3.5) to obtain elementary solutions of Eq. (3.4). First note that (3.3), (3.5) imply that

$$\bar{u}(z; -\alpha) = u(z; \alpha - 1), \quad (3.10)$$

$$\bar{u}(z; \alpha + 1) = -u(z; \alpha) - 2\left(\frac{u'(z; \alpha) - (\alpha + 1)}{2u(z; \alpha) - z}\right)^2. \quad (3.11)$$

1. Rational solutions of PII

It is clear that $v = 0, \alpha = 0$ solve PII. Then using (3.3) one can obtain a rational solution of PII for every positive integer

$$v(z; 0) = 0, \quad v(z; 1) = -\frac{1}{z},$$

$$v(z; 2) = \frac{1}{z} - \frac{3z^2}{z^3 + 4}, \ldots. \quad (3.12)$$

Then Eq. (3.2) generates a rational solution for every negative integer.

2. Rational solutions of (3.4)

$u = 0, \alpha = 0$ solve Eq. (3.4). Then (3.10) and (3.11) imply similar results as (1) above:

$$u(z; 0) = 0, \quad u(z; 1) = -\frac{2}{z^2},$$

$$u(z; 2) = -\frac{6z(z^3 - 8)}{(z^3 + z)^2}, \ldots. \quad (3.13)$$

Note that the hierarchies of solutions (3.12) and (3.13) are related by the transformations (3.5).

3. Airy type solutions of PII

Lemma 3.1 implies that $v(z; -\frac{1}{2}) = y'/y$ is a solution of PII where y is any solution of the Airy equation $y'' + (z/2)y = 0$. One cannot use this solution directly in (3.3) to generate new solutions, because in this case (3.3) breaks down. The trick is to first use (3.2) and then (3.3). In this way one generates the following hierarchy of solutions:

$$v(z; -\tfrac{1}{2}) = y'/y, \quad v(z; \tfrac{1}{2}) = -y'/y,$$
$$v(z; \tfrac{3}{2}) = y'/y - y^2/(2y'^2 + zy^2), \ldots. \quad (3.14)$$

4. Airy type solutions of (3.4)

Similarly as above

$$u(z; -\tfrac{1}{2}) = z/2, \quad u(z; \tfrac{1}{2}) = -2y'^2/y^2 - z/2, \ldots. \quad (3.15)$$

Remarks:

1. The results of Theorem 3.1 and Lemma 3.1 were first given in Refs. 14 and 3, respectively. The result of Theorem 3.1 was rederived later in Refs. 30 and 31 by exploiting the connection with the inverse scattering of the KdV equation.
2. We emphasize that the logical steps used here for (a) deriving (3.4) and (3.5), (b) exploiting the quadratic dependence of ν on α to obtain (3.3), (c) characterizing one-parameter families of solutions of PII when the transformation (3.5) breaks down, and (d) generating elementary solutions, remain valid for considering all Painlevé equations PII–PVI.

IV. PAINLEVÉ III

In this section we consider PIII. Having familiarized the reader with our method we now present both types of transformations (A) and (B).

Theorem 4.1: Let $v(z; \alpha, \beta, \gamma, \delta)$ be a solution of PIII

$$v'' = \frac{v'^2}{v} - \frac{1}{z}v' + \frac{1}{z}(\alpha v^2 + \beta) + \gamma v^3 + \frac{\delta}{v}. \quad (4.1)$$

Then $\bar{v}(z, \bar{\alpha}, \bar{\beta}, \bar{\gamma}, \bar{\delta})$ are also solutions of PIII, where

$$\bar{v}(z;\bar{\alpha},\bar{\beta},\bar{\gamma},\bar{\delta}) = -v(z;\alpha,\beta,\gamma,\delta);\quad \bar{\alpha}=-\alpha,\ \bar{\beta}=-\beta,\ \bar{\gamma}=\gamma,\ \bar{\delta}=\delta, \tag{4.2}$$

$$\bar{v}(z;\bar{\alpha},\bar{\beta},\bar{\gamma},\bar{\delta}) = [v(z;\alpha,\beta,\gamma,\delta)]^{-1};\quad \bar{\alpha}=-\beta,\ \bar{\beta}=-\alpha,\ \bar{\gamma}=-\delta,\ \bar{\delta}=-\gamma, \tag{4.3}$$

$$\bar{v}(z;\bar{\alpha},\bar{\beta},\bar{\gamma},\bar{\delta}) = \frac{\gamma^{1/2}}{\bar{\gamma}^{1/2}}v\left[1+\frac{2+\beta(-\delta)^{1/2}+\alpha(\gamma)^{-1/2}}{z(v'/v+\gamma^{1/2}v+(-\delta)^{1/2}/v)-1-\beta(-\delta)^{-1/2}}\right]; \tag{4.4a}$$

$$\bar{\alpha}=-[2+\beta(-\delta)^{-1/2}]\gamma^{-1/2},\quad \bar{\beta}=-[2+\alpha\gamma^{-1/2}](-\delta)^{1/2}\left(\frac{\gamma^{1/2}}{\bar{\gamma}^{1/2}}\right),\quad (-\bar{\delta})^{1/2}=(-\delta)^{1/2}\frac{\gamma^{1/2}}{\bar{\gamma}^{1/2}}. \tag{4.4b}$$

In (4.4) we have assumed that

$$\gamma\neq 0 \text{ and } 2+\alpha(\gamma)^{-1/2}+\beta(-\delta)^{1/2}\neq 0. \tag{4.4c}$$

If $\gamma=0$ then (4.4) is replaced by

$$\bar{v}(z;\bar{\alpha},\bar{\beta},0,\bar{\delta}) = z\,v'/v^2-(1+\beta(-\delta)^{1/2})v + z(-\delta)^{1/2}/v^2, \tag{4.5a}$$

$$\bar{\alpha}=(-\delta)^{1/2},\ \bar{\beta}=\alpha(2-\beta(-\delta)^{1/2}),\ \bar{\delta}=-\alpha^2. \tag{4.5b}$$

The case $2+\beta(-\delta)^{1/2}+\alpha(\gamma)^{-1/2}=0$ is considered in Lemma 4.1.

Theorem 4.2: Let $v(z;\alpha,\beta,1,\delta)$[32] be a solution of PIII and let $w(x;\bar{\alpha},\bar{\beta},\bar{\gamma},0)$ be a solution of PV,

$$\frac{d^2w}{dx^2}=\frac{3w-1}{2w(w-1)}\left(\frac{dw}{dx}\right)^2-\frac{1}{x}\frac{dw}{dx}+\frac{(w-1)^2}{x^2}\times\left(\frac{\bar{\alpha}}{4}w+\frac{\bar{\beta}}{4w}\right)+\frac{(-\bar{\delta})^{1/2}w}{x}, \tag{4.6a}$$

where

$$x=\frac{z^2}{2},\ \bar{\alpha}=\frac{(1+\alpha-\mu)^2}{32},\ \bar{\beta}=-\frac{(1-\alpha-\mu)^2}{32},\quad \bar{\gamma}=(-\delta)^{1/2},\quad \mu=-[1+\beta(-\delta)^{1/2}]. \tag{4.6b}$$

Then there exists the following one to one correspondence between solutions of (4.1) and (4.6)

$$v'(w-1)+(w+1)v^2+\left[\frac{\mu}{2}v+(-\delta)^{1/2}\right](w-1)=0, \tag{4.7}$$

$$zv=-\frac{w'}{w}+(1+\alpha-\mu)w-\frac{\alpha}{2}+\frac{\mu+\alpha-1}{w}.$$

Theorem 4.3: Let $v(z;\alpha,\beta,\gamma,\delta)$ be a solution of PIII and let $\phi(z;\rho,\sigma,\tau)$ be a solution of

$$\left(\phi''+\rho\phi+\frac{\sigma}{2}\right)^2=\frac{\phi^2}{z^2}(\phi'^2+\rho\phi^2+\sigma\phi+\tau), \tag{4.8a}$$

$$\rho=4\gamma^{1/2}(-\delta)^{1/2},\quad \sigma=4[\alpha(-\delta)^{1/2}-\beta\gamma^{1/2}],\quad \tau=-4(\alpha+\gamma^{1/2})[\beta+(-\delta)^{1/2}]. \tag{4.8b}$$

Then there exists the following one-to-one correspondence between solutions of (4.1) and (4.8):

$$\frac{\phi}{z}=\frac{v'}{v}+\gamma^{1/2}v+\frac{(-\delta)^{1/2}}{v},$$

$$v=\frac{\phi'+(z/\phi)(\phi''+\rho\phi+\sigma/2)}{2(\gamma^{1/2}\phi+\alpha+\gamma^{1/2})}, \tag{4.9}$$

provided that neither $\beta=\delta=0$, nor $\alpha=\gamma=0$, nor $2+\alpha\gamma^{-1/2}+\beta(-\delta)^{-1/2}=0$. These exceptional cases are considered below.

Lemma 4.1: PIII admits a one-parameter family of solutions characterized by

$$(1-\alpha\gamma^{-1/2})/z=v'/v+\gamma^{1/2}v+(-\delta)^{1/2}/v, \tag{4.10}$$

iff

$$2+\alpha\gamma^{-1/2}+\beta(-\delta)^{-1/2}=0.$$

Lemma 4.2: The general solution of PIII in the case that $\beta=\delta=0$ is given by

$$v=\frac{\phi'}{\gamma^{1/2}\phi+\alpha+\gamma^{1/2}};\quad \int\frac{d\phi}{\phi^2/2+\phi+c_1}=\ln z+c_2, \tag{4.11}$$

where c_1, c_2 are arbitrary constants.

Using (4.3) similar results are obtained for the case that $\alpha=\gamma=0$.

Remarks:

1. The results of Theorem 4.2 and the above lemma's were first given in Refs. 15, 11, and 33, respectively.
2. One can clearly combine the transformations (4.2)–(4.4) to obtain new transformations. For example, combining (4.3) and (4.4) one can derive the corresponding result of Ref. 15. Also note that a finite number of products of (4.2) and (4.3) yields the identity.
3. Elementary solutions of PIII can be derived in the same manner as in PII.

Derivation of the above results: In this case $P_1=1/v$, $P_2=-1/z$, $P_3=(1/z)(\alpha v^2+\beta)+\gamma v^3+\delta/v$. Thus the coefficient of u^2 in (2.8) is $-(dv^2+f)/v$.

A. Requiring (2.8) to be linear in v

Then the above coefficient of u^2 implies that either $d=0$ or $f=0$. In what follows we shall consider only the case $f=0$. (The result of Ref. 15 is derivable by considering the case $d=0$). With $f=0$ Eq. (2.8) becomes

$$u'=(-dv)u^2+\left[\frac{2adv^3+(bd+ea-d'-dz^{-1})v^2-(e'+ez^{-1})v-ec}{dv^2+ev}\right]u+[\ldots], \tag{4.12}$$

where the last bracket is independent of u. The coefficient of u in the above will be linear in v iff dv^2+ev is a root of the numerator. This implies that $e=0$. Then, without loss of generality, we can take $d=1$ and $a=0$, since one can always "absorb" them in u with the aid of a Möbius transformation. If $e=a=0$ and $d=1$ then the term in (4.12) inde-

pendent of u is linear in v iff $c^2+\delta=0$ and $b=-(\beta+c)/cz$. Thus, with the above choices of $a, \ldots, f$ Eqs. (2.6) and (4.12) yield

$$u=\frac{v'}{v^2}+\frac{b}{v}+\frac{c}{v^2};\ c=(-\delta)^{1/2},\ b=-\left(\frac{\beta+c}{cz}\right) \tag{4.13}$$

$$v=\frac{u'+(z^{-1}-b)u-\alpha z^{-1}}{\gamma-u^2}. \tag{4.14}$$

Substituting (4.14) into (4.13) one obtains an equation for u, namely,

$$u''=uu'^2/(u^2-\gamma)+\cdots. \tag{4.15}$$

It is then clear that $\gamma=0$ is of special interest.

1. $\gamma=0$

Then (4.15) becomes

$$u''=\frac{u'^2}{u}-\frac{u'}{z}-\frac{\alpha^2}{z^2u}+\frac{\alpha}{z^2}\left(1+\frac{c-\beta}{c}\right)+cu^2.$$

Let $u=\dfrac{\bar v}{z}$ to obtain

$$\bar v''=\frac{\bar v'^2}{\bar v}-\frac{\bar v'}{z}+c\frac{\bar v^2}{z}+\alpha\left(1+\frac{c-\beta}{c}\right)-\frac{\alpha^2}{\bar v}. \tag{4.16}$$

Equations (4.13) (with $u=\bar v/z$) and (4.16) imply (4.5).

2. $\gamma\neq 0$

Without loss of generality take $\gamma=1$. Equation (4.15) must be, within a Möbius transformation, one of the fifty equations mentioned in the introduction. Thus let $u=(Aw+B)/(Cw+D)$ to transform (4.15) to

$$w''=\left[2C+\frac{(Aw+B)(AD-CB)}{[(Aw+B)^2-\gamma(Cw+D)^2](Cw+D)}\right]w'^2+\cdots. \tag{4.17}$$

Hence, if $A=B=z$ and $D=-C=1$ then $w''=(1/2w-1/(w-1))w'^2+\cdots$. Therefore, under the transformation

$$u=-(w+1)/(w-1) \tag{4.18}$$

Eq. (4.15) becomes

$$w''=\frac{(3w-1)}{2w(w-1)}w'^2-\frac{1}{z}w'+\frac{(w-1)^2}{z^2}\left(\bar\alpha w+\frac{\bar\beta}{w}\right)+2cw, \tag{4.19}$$

where $\bar\alpha, \bar\beta$ are defined in (4.6b). Letting $z=(2x)^{1/2}$ in (4.19) and using (4.18) in (4.13) and (4.15) the result of Theorem 4.2 follows.

3. $2+\beta(-\delta)^{1/2}-\alpha\gamma^{-1/2}=0$

The transformation (4.14) breaks down iff $u=\gamma^{1/2}$. Then $u'+(z^{-1}-b)u-\alpha z^{-1}$ must be zero when $u=\gamma^{1/2}$. This implies $2+\beta(-\delta)^{1/2}-\alpha\gamma^{-1/2}=0$. Hence, using (4.13) one obtains the result of Lemma 4.1.

Using Theorem 4.2 one can obtain (4.4) in a similar way as in obtaining the corresponding result for PII. However, we choose to obtain (4.4) using Theorem 4.3, which is now derived.

B. Requiring (2.8) to be quadratic in v

Take, for example, $d=f=0$. (We remind the reader that our investigation is not exhaustive.) Then, without loss of generality, $b=0$ and $e=1$. Hence, (2.8) becomes

$$u'=u^2\left[\frac{av^2-z^{-1}v-c}{\gamma}\right]+[\ ]u+[\ ].$$

The two brackets [ ], in the above equation contain v quadratically iff

$$a^2=\gamma,\ c^2+\delta=0. \tag{4.20}$$

Therefore, with the above choices of $a, \ldots, f$ Eqs. (2.6) and (2.8) yield

$$\phi/z=v'/v+\gamma^{1/2}v+(-\delta)^{1/2}/v, \tag{4.21}$$

$$(\gamma^{1/2}+\alpha+\gamma^{1/2}\phi)v^2-\phi' v+(\beta+(-\delta)^{1/2}-(-\delta)^{1/2}\phi)=0, \tag{4.22}$$

where we have used for convenience the substitution $u=\phi/z$. Equation (4.22) yields

$$v=\frac{\phi'+\Delta^{1/2}}{2(\gamma^{1/2}\phi+\alpha+\gamma^{1/2})},\quad \Delta=\phi'^2+\rho\phi^2+\sigma\phi+\tau. \tag{4.23}$$

where ρ, σ, τ are defined in (4.8b). Noting that $\Delta'=2\phi'\Omega$, where

$$\Omega=\phi''+\rho\phi+\frac{\sigma}{2}, \tag{4.24}$$

and substituting (4.22) into (4.21) one obtains an equation for ϕ. This equation, using the fact that

$$\rho\phi^2+\sigma\phi+\tau=4(\gamma^{1/2}\phi+\alpha+\gamma^{1/2})\times[(-\delta)^{1/2}\phi-\beta-(-\delta)^{1/2}],$$

takes the form $(\Delta^{1/2}+\phi')(\Delta^{1/2}\phi/z-\Omega)=0$, which implies

$$\Delta^{1/2}\phi/z=\Omega. \tag{4.25}$$

Therefore, the transformations (4.21) and (4.23) relate PIII and Eq. (4.25). Using (4.25) in (4.23) one obtains the result of Theorem 4.3.

1. $\beta=\delta=0$

Then, using (4.21), (4.23), and (4.25) (where we pick the positive root of $\Delta^{1/2}$) we have

$$\frac{\phi}{z}=\frac{v'}{v}+\gamma^{1/2}v,\quad v=\frac{\phi'}{\gamma^{1/2}\phi+\alpha+\gamma^{1/2}},$$

$$\phi''-\frac{\phi\phi'}{z}=0. \tag{4.26}$$

However, the equation for ϕ is now very simple and it can be immediately integrated to $z\phi'=\phi^2/2+\phi+c_1$. Thus Lemma 4.2 immediately follows.

2. $2+\alpha\gamma^{-1/2}+\beta(-\delta)^{-1/2}=0$

The transformation (4.23) breaks down iff $\gamma^{1/2}\phi+\alpha+\gamma^{1/2}=0$. But then $\phi'+\Delta^{1/2}=0$ which implies $2+\alpha\gamma^{-1/2}+\beta(-\delta)^{-1/2}=0$. Then, using Eq. (4.21) the result of Lemma 4.1 is again derived.

3. The transformation from PIII to PIII

This transformation is easily obtained by finding two sets of $\{\alpha, \beta, \gamma, \delta\}$ which give the same values for ρ, σ, τ. Solving the equations defining ρ, σ, τ for α, β, δ (keeping γ fixed) one finds

$$(-\bar\delta)^{1/2}=\frac{\rho}{4}(\bar\gamma)^{-1/2},\ \bar\beta=\left[\bar\alpha(-\bar\delta)^{1/2}-\frac{\sigma}{4}\right](\bar\gamma)^{-1/2},$$

$$\bar\alpha=\frac{2}{\rho}\left[\frac{\sigma}{4}-\frac{\rho}{2}\pm\left(\frac{\sigma^2}{16}-\frac{\tau\rho}{4}\right)^{1/2}\right](\bar\gamma)^{1/2}. \tag{4.27}$$

However, using the definitions of ρ, σ, τ it follows that

$$\sigma^2/16-\tau\rho/4=[\alpha(-\delta)^{1/2}+\beta\gamma^{1/2}+2\gamma^{1/2}(-\delta)^{1/2}]^2.$$

Using the positive root of (4.27) one finds the trivial result

$$\bar\alpha=\alpha(\bar\gamma^{1/2}/\gamma^{1/2}),\quad \bar\beta=\beta(\gamma^{1/2}/\bar\gamma^{1/2}),$$
$$(-\bar\delta)^{1/2}=(-\delta)^{1/2}(\gamma^{1/2}/\bar\gamma^{1/2}).$$

However, using the negative root, one obtains the expressions for $\bar\alpha,\bar\beta,\bar\gamma$ appearing in (4.4b). Then using

$$\bar v=\frac{\phi'+\Delta^{1/2}}{2(\bar\gamma^{1/2}\phi+\bar\alpha+\bar\gamma^{1/2})}=\frac{v(\gamma^{1/2}\phi+\alpha+\gamma^{1/2})}{\bar\gamma^{1/2}\phi+\bar\alpha+\bar\gamma^{1/2}}$$

and replacing ϕ by (4.21), Eqs. (4.4) follow.

V. PAINLEVÉ IV AND V

Using our method one can easily find transformations which map PIV and PV to themselves, but with different values of the parameters. These transformations were first given in Refs. 16 and 17, respectively. Here, for completeness, we give these transformations and advise the interested reader to derive them himself as a simple exercise of our method.

Theorem 5.1: Let $v(z;\alpha,\beta)$ be a solutions of PIV

$$v''=\frac{v'^2}{2v}+\frac{3}{2}v^3+4zv^2+2(z^2-\alpha)v+\beta v^{-1}. \tag{5.1}$$

Then $\bar v(z;\bar\alpha,\bar\beta)$ is also a solution of PIV, where

$$\bar v=\frac{v'-v^2-2zv-(-2\beta)^{1/2}}{2v};$$

$$\bar\alpha=\frac{1}{4}[2-2\alpha+3(-2\beta)^{1/2}],$$

$$\bar\beta=-\frac{1}{2}\left[1+\alpha+\frac{(-2\beta)^{1/2}}{2}\right]^2, \tag{5.2}$$

provided that

$$1+\alpha+(-2\beta)^{1/2}/2\neq 0. \tag{5.3}$$

Theorem 5.2: Let $v(z;\alpha,\beta,\gamma,\delta)$ be a solution of PV

$$v''=\frac{3v-1}{2v(v-1)}v'^2-\frac{1}{z}v'+\frac{\alpha}{z^2}v(v-1)^2$$
$$+\frac{\beta}{z^2}\frac{(v-1)^2}{v}+\frac{\gamma}{z}v+\frac{\delta v(v+1)}{v-1}. \tag{5.4}$$

Then $\bar v(z;\bar\alpha,\bar\beta,\bar\gamma,\bar\delta)$ is also a solution of PV, where

$$\bar v=1-\frac{2(-2\delta)^{1/2}zv}{zv'-(2\alpha)^{1/2}v^2+[(2\alpha)^{1/2}-(-2\beta)^{1/2}+(-2\delta)^{1/2}z]v+(-2\beta)^{1/2}}, \tag{5.5a}$$

$$\bar\alpha=-\frac{1}{16\delta}[\gamma+(-2\delta)^{1/2}(1-(-2\beta)^{1/2}-(2\alpha)^{1/2})]^2,\quad \bar\beta=\frac{1}{16\delta}[\gamma-(-2\delta)^{1/2}(1-(-2\beta)^{1/2}+(-2\delta)^{1/2}z)]^2, \tag{5.5b}$$

$$\bar\gamma=(-2\delta)^{1/2}[(-2\beta)^{1/2}-(2\alpha)^{1/2}],\ \bar\delta=\delta,$$

provided that $\delta\neq 0$ and

$$(-2\delta)^{1/2}[1-(-2\beta)^{1/2}-(2\alpha)^{1/2}]\neq\gamma. \tag{5.6}$$

Remarks:

1. One can easily find Lie-point discrete symmetries of the above equations. For example, if $v(z;\alpha,\beta,\gamma,\delta)$ solves PV then $\bar v=v^{-1}(z;-\beta,-\alpha,-\gamma,\delta)$ also solves PV.
2. When the above transformations break down, i.e., when (5.3) and/or (5.6) are violated then PIV and/or PV, just as for PII and PIII, admit one-parameter families of solutions.
3. Elementary solutions of the above equations can be derived in a similar manner as in PII.

VI. PAINLEVÉ VI

Theorem 6.1: Let $v(z;\alpha,\beta,\gamma,\delta)$ be a solution of PVI:

$$v''=\frac{1}{2}\left(\frac{1}{v}+\frac{1}{v-1}+\frac{1}{v-z}\right)v'^2$$
$$-\left(\frac{1}{z}+\frac{1}{z-1}+\frac{1}{v-z}\right)v'+\frac{v(v-1)(v-z)}{z^2(z-1)^2}$$
$$\times\left[\alpha+\frac{\beta z}{v^2}+\frac{\gamma(z-1)}{(v-1)^2}+\frac{\delta z(z-1)}{(v-z)^2}\right]. \tag{6.1}$$

Then $\bar v(z;\bar\alpha,\bar\beta,\bar\gamma,\bar\delta)$ are also solutions of PVI, where

$$\bar v(z;\bar\alpha,\bar\beta,\bar\gamma,\bar\delta)=zv(1/z;\alpha,\beta,\gamma,\delta);$$
$$\bar\alpha=\alpha,\bar\beta=\beta,\bar\gamma=-\delta+\tfrac12,\bar\delta=-\gamma+\tfrac12, \tag{6.2}$$
$$\bar v(z;\bar\alpha,\bar\beta,\bar\gamma,\bar\delta)=1-v(1-z;\alpha,\beta,\gamma,\delta);$$
$$\bar\alpha=\alpha,\bar\beta=-\gamma,\bar\gamma=-\beta,\bar\delta=\delta, \tag{6.3}$$
$$\bar v(z;\bar\alpha,\bar\beta,\bar\gamma,\bar\delta)=1-(1-z)v(1/(1-z);\alpha,\beta,\gamma,\delta);$$
$$\bar\alpha=\alpha,\bar\beta=\delta-\tfrac12,\bar\gamma=-\beta,\bar\delta=-\gamma+\tfrac12, \tag{6.4}$$
$$\bar v=v+2((z+1)v-2z)$$
$$\times\left(-\frac{2z(z-1)}{\kappa}\frac{\Phi'}{\Phi}+\frac{(z-1)I}{\kappa\Phi}-(z+1)\right)^{-1}; \tag{6.5a}$$
$$\bar\alpha=\tfrac12[(-2\beta)^{1/2}-1]^2,\bar\beta=-\tfrac12[(2\alpha)^{1/2}+1]^2,$$
$$\bar\gamma=\gamma+\kappa\mu/4,\bar\delta=\delta+\kappa\mu/4. \tag{6.5b}$$

In (6.5) Φ, I, κ, μ are defined by

$$\Phi=z\frac{v'}{v}+\frac{(\lambda-\kappa-1)}{2(z-1)}v+\frac{(\lambda+\kappa+1)z}{2(z-1)}\frac{1}{v}$$
$$-\frac{\lambda}{2}\frac{(z+1)}{(z-1)}-\left(\frac12+\frac{\mu}{4}\right), \tag{6.6}$$

$$I\doteq\Phi^2+\frac{\mu}{2}\Phi+\nu, \tag{6.7}$$

$$\kappa=(-2\beta)^{1/2}-(2\alpha)^{1/2}-1,\ \lambda=(-2\beta)^{1/2}+(2\alpha)^{1/2},$$

$$\mu=\frac{4}{\kappa}\left(\frac{1}{2}-\gamma-\delta\right),\quad \nu=2\delta-1+\left(\frac{\mu}{4}+\frac{\kappa}{2}\right)^2. \tag{6.8}$$

In (6.5) we have assumed that

$$\Phi\neq 0,\quad \kappa\neq 0,\quad \nu\neq 0. \tag{6.9}$$

Theorem 6.2: Let $v(z;\alpha,\beta,\gamma,\delta)$ be a solution of PVI and let $\Phi(z;\kappa^2,\lambda,\mu,\nu)$ be a solution of

$$(z-1)^2\Omega^2=\frac{1}{z^2}\left(\Phi'^2+\frac{I^2-\kappa^2\Phi^2}{z(z-1)^2}\right)\Psi^2, \tag{6.10}$$

where

$$\Omega\doteq\Phi''+\frac{(3z-1)\Phi'}{2z(z-1)}+\frac{2\Phi I+\mu I/2-\Phi\kappa^2}{z(z-1)^2},$$

$$\Psi\doteq(z+1)\Phi+\frac{\mu}{4}(z+1)+\frac{\lambda}{2}(z-1), \tag{6.11}$$

and I,κ,λ,μ, and ν are defined by (6.7) and (6.8). The Eqs. (6.6) and (6.12) below express a one-to-one correspondence between solutions of (6.1) and (6.10)

$$v=\left(-\frac{(z+1)}{z}\Phi'+\frac{2\kappa}{z(z-1)}\Phi-(z-1)^2\frac{\Omega}{\Psi}\right)\times\left(-\frac{2\Phi'}{z}+\frac{I}{z^2}+\frac{\kappa(z+1)\Phi}{z^2(z-1)}\right)^{-1}. \tag{6.12}$$

We have assumed that (6.9) is valid.

Lemma 3.1: PVI admits a one-parameter family of solutions characterized by

$$zv'+\frac{(\lambda-\kappa-1)}{2(z-1)}v^2-\left[\frac{\lambda}{2}\frac{(z+1)}{(z-1)}+\frac{1}{2}+\frac{\mu}{4}\right]v+\frac{(\lambda+\kappa+1)z}{2(z-1)}=0, \tag{6.13}$$

iff $\nu=0,\ \kappa\neq 0$.

This result, which is an immediate consequence of theorem 6.2, was first given in Ref. 18. Note that if

$$v=-\frac{2z(z-1)}{\lambda-\kappa-1}\frac{w'}{w}, \tag{6.14}$$

then w satisfies a certain hypergeometric equation.

A. Derivation of the above results

In deriving the above results we follow the same logical steps as with PIII.

1. The transformation from PVI to (6.10)

In this case the coefficient of u^2 in (2.8) is

$$\frac{(3/2\,v^2-(z+1)v+z/2)}{v(v-1)(v-z)}J^2-2dvJ-eJ.$$

Therefore, it is impossible to choose $a,\ldots,f$ in such a way that (2.8) reduces to a linear equation for v. However, by choosing $d=f=0$, and then (without loss of generality) $e=1$, $b=0$, Eq. (2.8) reduces to a quadratic equation for v iff

$$a=(2\alpha)^{1/2}/z(z-1),$$
$$c=(-2\beta)^{1/2}/(z-1). \tag{6.15}$$

Then (2.6) and (2.8) become

$$v'+av^2-uv+c=0, \tag{6.16}$$
$$\hat{A}v^2+\hat{B}v+\hat{C}=0, \tag{6.17}$$

where $\hat{A},\hat{B},\hat{C}$ are known functions of u',u,z. Equation (6.17) may be simplified if one uses the transformation

$$u=\frac{\phi}{z}+\frac{\hat{\lambda}}{z-1}+\frac{\hat{\mu}}{z(z-1)};$$
$$\hat{\lambda}=-\frac{(\alpha+(2\alpha)^{1/2}+\beta+\gamma+\delta)}{\kappa},$$
$$\hat{\mu}=\lambda-\hat{\lambda},\ k\neq 0, \tag{6.18}$$

where κ and λ are defined in (6.8). Replacing u in terms of ϕ in (6.16) and (6.17) one obtains

$$v'+\frac{(2\alpha)^{1/2}}{z(z-1)}v^2-\left(\frac{\phi}{z}+\frac{\hat{\lambda}}{z-1}+\frac{\hat{\mu}}{z(z-1)}\right)v+\frac{(-2\beta)^{1/2}}{z-1}=0, \tag{6.19}$$

$$Av^2+Bv+C=0, \tag{6.20a}$$

where

$$A=-\frac{\phi'}{z}+\frac{1}{2z^2}\left(I+\frac{\kappa(z+1)}{(z-1)}\phi\right),$$
$$B=\frac{(z+1)}{z}\phi'-\frac{2\kappa\phi}{z(z-1)}, \tag{6.20b}$$
$$C=-2\phi'+\frac{\kappa(z+1)}{z(z-1)}\phi-Az$$

and I is defined by (6.7). Equations (6.19) and (6.20) are the analogues of Eqs. (4.21) and (4.22). Equation (6.20) yields

$$v=\left(-B+\frac{(z-1)}{z}\Delta^{1/2}\right)\Big/2A;\quad \Delta=\phi'^2+\frac{I^2-\kappa^2\phi^2}{z(z-1)^2}. \tag{6.21}$$

Substituting (6.21) into (6.19) one obtains

$$\frac{\Delta^{1/2}}{z(z-1)}\Psi+\Omega=0, \tag{6.22}$$

where Ψ and Ω are defined by (6.11). (In obtaining this equation it is crucial to note that A and $I^2-\kappa^2\Phi^2$ are common factors.) Therefore, the transformations (6.19) and (6.21) define a one-to-one correspondence between PVI and (6.22). Note that the two different branches of $\Delta^{1/2}$ in (6.21) correspond to the two different branches of $\Delta^{1/2}$ in (6.22). If one wants to get rid of the square root in (6.22), one may replace $\Delta^{1/2}$ in (6.21) by $-z(z-1)\Omega/\Psi$. Then (6.21) becomes (6.12), and using (6.12) in (6.22), Eq. (6.10) follows.

2. $\nu=0$

The transformation (6.21) breaks down iff $A=0$. It then follows (requiring that $-B+(z-1)\Delta^{1/2}/z$ is also zero) that $\phi=0$ and $\nu=0$. Hence, substituting $\phi=0$, $\nu=0$ in (6.19), Lemma 6.1 follows.

3. The transformation from PVI to PVI

The trick again is to find two sets of parameters $\{\alpha,\beta,\gamma,\delta\}$ and $\{\bar{\alpha},\bar{\beta},\bar{\gamma},\bar{\delta}\}$ which give rise to the same κ^2,λ,μ,ν. Hence it is clear that if the set $\{\alpha,\beta,\gamma,\delta\}$ corresponds to κ,λ,μ,ν then the set $\{\bar{\alpha},\bar{\beta},\bar{\gamma},\bar{\delta}\}$ must correspond to $-\kappa,\lambda,\mu,\nu$. Thus

$$\bar{\delta} = \frac{\nu}{2} + \frac{1}{2} - \frac{1}{2}\left(\frac{\mu}{4} - \frac{\kappa}{2}\right)^2$$
$$= \frac{\nu}{2} + \frac{1}{2} - \frac{1}{2}\left(\frac{\mu}{4} + \frac{\kappa}{2}\right)^2 + \frac{\kappa\mu}{4}$$
$$= \delta + \frac{\kappa\mu}{4}. \tag{6.23}$$

Similarly for $\bar{\gamma}$. Also solving the equations $(-2\bar{\beta})^{1/2} + (2\bar{\alpha})^{1/2} = (-2\beta)^{1/2} + (2\alpha)^{1/2}$ and $(-2\bar{\beta})^{1/2} - (2\bar{\alpha})^{1/2} - 1 = -(-2\beta)^{1/2} + (2\alpha)^{1/2} + 1$, one obtains $(2\bar{\alpha})^{1/2} = -1 + (-2\beta)^{1/2}$ and $(-2\bar{\beta})^{1/2} = 1 + (2\alpha)^{1/2}$. Finally, using (6.12) with v replaced by $\bar{v}$ and κ by $-\kappa$, one obtains (6.5a).

B. How to obtain elementary solutions

Using Theorem 6.1, one may obtain infinite hierarchies of elementary solutions of PVI. As with PII, it is important to notice that if one starts with the solution v characterized by Lemma 6.1, one cannot use (6.5a) directly [since in this case (6.5a) breaks down]; one must first use a Lie-point discrete symmetry to obtain a new solution $\hat{v}$ and then use (6.5a). We also note that the Lie-point symmetry (6.2) cannot be used, because for this symmetry $\bar{\nu} = \nu$ (hence if $\nu = 0$, $\bar{\nu} = 0$). If one uses instead the Lie-point symmetry (6.4) one has the following result:

Lemma 6.2: Let $v(z; \alpha, \beta, \gamma, \delta)$ be the one-parameter family of solutions of (6.13), where κ, λ, μ are defined by (6.8) and $\nu = 0$. Use the transformation (6.4) to evaluate $\hat{v}(z; \hat{\alpha}, \hat{\beta}, \hat{\gamma}, \hat{\delta})$. Then apply to this solution the transformation (6.5a) to obtain a new one-parameter family of solutions $\bar{v}$ of PVI, with parameters

$$\bar{\alpha} = \frac{1}{2}(\hat{\beta}_1 - 1)^2, \quad \bar{\beta} = -\frac{1}{2}(\hat{\alpha}_1 + 1)^2,$$
$$\bar{\gamma} = \hat{\gamma} + \frac{\hat{\kappa}\hat{\mu}}{4}, \quad \bar{\delta} = \hat{\delta} + \frac{\hat{\kappa}\hat{\mu}}{4}, \tag{6.24a}$$

where

$$\hat{\alpha} = \alpha, \ \hat{\beta} = \delta - \tfrac{1}{2}, \ \hat{\gamma} = -\beta, \hat{\delta} = -\gamma + \tfrac{1}{2},$$

and

$$\hat{\alpha}_1 = (2\alpha)^{1/2}, \quad \hat{\beta}_1 = (-2\beta)^{1/2}. \tag{6.24b}$$

Example 6.1: The solution of Eq. (6.13) is in general expressible in terms of the hypergeometric functions. Hence, using the above lemma one can obtain an infinite hierarchy of one-parameter family of solutions of PVI, all of which are related [through the repeated application of (6.5a)] to these hypergeometric functions. However, for some special choices of the parameters $\alpha, \beta, \gamma, \delta$ Eq. (6.13) becomes very simple. In this case one may, for example, derive infinite hierarchies of rational solutions. Let us make such a choice in order to illustrate our results. Let $\alpha = \beta = 0$, $\gamma = 1/2$, $\delta = -3/2$. Then, using (6.8), $\kappa = -1, \lambda = 0, \mu = -6$, $\nu = 0$. Therefore, since $\nu = 0$, PVI must admit a one-parameter family of solutions characterized by (6.13). Actually in this case (6.13) reduces to $zv' + v = 0$ and hence $v = \tau/z$, τ some arbitrary constant. Now starting with

$$v = \tau/z; \alpha = \beta = 0, \gamma = 1/2, \delta = -3/2 \tag{6.25}$$

in (6.4), one finds

$$\hat{v} = 1 - \tau(1 - z)^2; \hat{\alpha} = \hat{\gamma} = \hat{\delta} = 0, \hat{\beta} = -2. \tag{6.26}$$

Then, using (6.8), either

$$\hat{\kappa} = 1, \quad \hat{\lambda} = 2, \quad \hat{\mu} = 2, \hat{\nu} = 0$$

or

$$\hat{\kappa} = -3, \quad \hat{\lambda} = -2, \quad \hat{\mu} = -2/3, \quad \hat{\nu} = 16/9.$$

The first choice used in (6.13) rederives $\hat{v}$; however, the second choice [used in (6.5)] yields

$$\bar{v} = \frac{z(\tau z^2 - 2\tau z + \tau - 1)}{2\tau z^3 - 3\tau z^2 + \tau - 1};$$
$$\bar{\alpha} = \frac{9}{2}, \quad \bar{\beta} = -\frac{1}{2}, \quad \bar{\gamma} = \frac{1}{2}, \quad \bar{\delta} = \frac{1}{2}. \tag{6.27}$$

One can verify directly that the functions v, $\hat{v}$, and $\bar{v}$, as defined by Eqs. (6.25), (6.26), and (6.27), respectively, satisfy PVI.

Remarks

(1) The transformation (6.4) is the product of the transformations (6.2) and (6.3). Similarly, one can obtain a transformation as the product of (6.3) and (6.2).

(2) It is worth noting that one cannot use just the Lie-point discrete symmetries [i.e., Eqs. (6.2)–(6.4)] to generate an infinite hierarchy of exact solutions. This, which is consistent with Ref. 28, follows from the fact that a finite number of products of these transformations yields the identity. For example, one obtains the identity after repeating the application of (6.4) three times.

ACKNOWLEDGMENTS

This work was partially supported by the Mathematics Divisions of AFOSR and ONR.

[1] E. L. Ince, *Ordinary Differential Equations* (Dover N. Y., 1947).
[2] P. Painlevé, Bull. Soc. Math. France **28**, 214 (1900). P. Painlevé, Acta Math. **25**, 1 (1902).
[3] B. Gambier, Acta. Math. **33**, 1 (1909).
[4] Two functions u and w are related by a Möbius, or bilinear or projective transformation iff $u = (Aw + B)/(Cw + D)$, where $AD - BC \neq 0$.
[5] E. Barouch, B. M. McCoy, and T. T. Wu, Phys. Rev. Lett. **31**, 1409 (1973); T. T. Wu, B. M. McCoy, C. A. Tracy, and E. Barouch, Phys. Rev. B **13**, 316 (1976).
[6] M. Jimbo, T. Miwa, Y. Mori, and M. Sato, Physica D. **1D**, 80 (1980); M. Jimbo and T. Miwa, Proc. Japan Acad. **56** A (9), 405 (1980).
[7] J. W. Miles, Proc. R. Soc. London A **361**, 277 (1978).
[8] The work done in the Soviet Union up to 1976 is reviewed by the editor of Diff. Urav. in the following two review articles (where he quotes 127 references): N. P. Erugin, Diff. Urav. **12**, 387 (1976); and N. P. Erugin, Diff. Urav. **12**, 579 (1976).
[9] V. I. Gromak, Diff. Urav. **14**, 2131 (1978).
[10] N. P. Erugin, Dokl. Akad. Nark. BSSR **2**, (1958).
[11] N. A. Lukashevich, Diff. Urav. **3**, (11), 994 (1967).
[12] N. A. Lukashevich, Diff. Urav. **3**, 395 (1967).
[13] N. A. Lukashevich, Diff. Urav. **4**, 732 (1968).
[14] N. A. Lukashevich, Diff. Urav. **7**, 1124 (1971).
[15] V. I. Gromak, Diff. Urav. **11**, 373 (1975).
[16] N. A. Lukashevich, *Author's abstract of doctoral dissertation* (in Russian) (Kiev, 1971).
[17] V. Gromak, Diff. Urav. **12**, 740 (1976).

[18]N. A. Lukashevich and A. I. Yablonskii, Diff. Urav. **3**, 264 (1967).
[19]A. S. Fokas and Y. C. Yortsos, Lett. Nuovo Cimento **30**, 539 (1981).
[20]M. J. Ablowitz and H. Segur, Phys. Rev. Lett. **38**, 1103 (1977).
[21]M. J. Ablowitz, A. Ramani, and H. Segur Lett. Nuovo Cimento **23**, 333 (1978) J. Math. Phys. **21**, 715 (1980); J. Math. Phys. **21**, 1006 (1980).
[22]C. Gardner, J. M. Greene, M. D. Kruskal, and R. M. Miura, Phys. Rev. Lett. **19**, 1095 (1967).
[23]G. B. Whitham (private communication).
[24]Given a continuous group of transformations which is admitted by a given equation, there is a standard procedure for finding the solutions of the equation invariant under this group (the so-called invariant solutions). See, for example, L. V. Ovsjannikov, *Group Properties of Differential Equations* (translated by G. W. Bluman), original publication in 1962; G. W. Bluman and J. D. Cole, *Similarity Methods for Differential Equations* (Springer, New York, 1974).
[25]A. S. Fokas and M. J. Ablowitz, *On a linearization of the Korteweg–deVries and Painlevé* II *equations*, Phys. Rev. Lett (1981).
[26]H. Flaschka and A. C. Newell, Commun. Math. Phys. **76**, 65 (1980).
[27]1) M. Sato, T. Miwa, and M. Jimbo, A series of papers entitled "Holonomic Quantum Fields," RIMS, Kyoto Univ. **14**, 223 (1977); *ibid* **15**, 201 (1979); *ibid*, **15**, 871 (1979).
2) K. Ueno, "Monodromy-preserving deformations of linear differential equations with irregular singular points" (RIMS Preprint, 1979), p. 301.
3) M. Jimbo and T. Miwa, "Deformations of linear ODE" (RIMS Preprint 1980, p. 315.
4) M. Jimbo and T. Miwa, "Monodromy-preserving deformation of linear ODE's with rational coefficients" (RIMS Preprint 1980), p. 327.
5) M. Jimbo and T. Miwa, "Deformation of linear ODE's" (RIMS Preprint 1980), p. 329.
[28]It is well known that exactly-solvable PDE's possess infinitely many continuous symmetries. Actually there is a strong indication that this is the defining feature of an exactly-solvable equation. See, for example A. S. Fokas, J. Math. Phys. **21**, 1381 (1980); A. S. Fokas and B. Fuchssteiner, Lett. Al Nuovo Cimento **28**, 299 (1980). These symmetries in general involve derivatives, in contrast to Lie-point symmetries. It is quite interesting that the Painlevé equations possess infinitely many *discrete* symmetries, involving derivatives (i.e., non Lie-point)
[29]It is well known that there exist equations of the form $(w')^n + \sum_{j=1}^{n} P_j(z,w)(w')^{n-j} = 0$, $n > 1$, n integer, which satisfy the Fuchs theorem concerning the absence of movable critical points. One may thus try to use appropriate transformations, related to these equations, instead of (2.6). We have not considered such transformations.
[30]M. Boiti and F. Pempinelli, Nuovo Cimento, **51** B, 70 (1979).
[31]H. Airault, Stud. Appl. Math. **61**, 31 (1979).
[32]If $\gamma \neq 0$, then taking $\gamma = 1$ is without loss of generality, as one may always use a scaling transformation.
[33]N. A. Lukashevich, Diff. Urav. **1**(6), 561 (1965).

The equivalence problem for the heat equation

W. Strampp
Fachbereich 17, Mathematik Gesamthochschule Kassel, Wilhelmshöher Allee 73, 3500 Kassel, West Germany

(Received 8 March 1982; accepted for publication 30 June 1982)

In this note we ask for the classes of equations of the second order which can be transformed into the heat equation $u_t = u_{xx}$. To give a partial answer to the question we express the heat equation by differential forms and prolong it by the Estabrook–Wahlquist method. This is motivated by the fact that our analysis is based upon conservation laws for which ideals of differential forms are a very suitable framework. Necessary conditions are derived for deciding whether a given equation can be transformed by some invertible point transformation into the heat equation or into its prolongation. In particular, the prolongation method enables us to understand the connection of various equations to the heat equation.

PACS numbers: 02.30.Jr, 02.30.Qy, 02.40.Vh

1. INTRODUCTION

The theory of transformation of equations has become a fundamental tool in the study of nonlinear equations. Lie groups of point transformations have been used traditionally for generating solutions from known ones. A prominent example for the generation of solutions by Bäcklund transformations is the sine–Gordon equation. Famous examples for Bäcklund transformations connecting different equations are furnished by Miura's transformation connecting the Korteweg–de Vries with the modified Korteweg–de Vries equation and by the Cole–Hopf transformation linearizing Burgers equation.

More recently several efforts have been undertaken for transforming nonlinear equations into linear ones. We mention the work of Flato and Simon (see for example Ref. 1) on nonlinear representation theory to which the work of Anderson, Harnad, and Winternitz on nonlinear superposition principles seems to be related. See for example Ref. 2.

A different approach has been undertaken by Bluman[3] and Bluman and Kumei.[4] They obtained the following condition: If the equation under consideration has an invariance generator involving an arbitrary solution of some linear equation as well as a certain contact transformation then the equation is linearizable. This is derived from the fact that invariance properties are injected from one equation into another by one-to-one mappings. Examples of this are found also in Ref. 5.

The purpose of this paper is as follows. We know of various second order equations such as Fokker–Planck equations, Burgers', and further remarkable nonlinear equations like $u_t = (u^{-2}u_x)_x$ being transformable to the heat equation. Therefore, we aim to obtain all equations in equivalence with the heat equation from a unified point of view. The following analysis will be concerned with conservation laws while group properties will not appear in our approach.

The heat equation can be expressed by a contact form and a closed two-form representing a conservation law. This is not unique since the heat equation possesses many conservation laws. The heat equation can also be expressed by several contact forms and prolongation forms obtained by the Estabrook–Wahlquist method.[6] The prolongation forms, being pseudopotentials, will then be understood as conservation laws while the heat equation itself appears as integrability or closure condition.

An equation is called equivalent to the heat equation if it is represented by an ideal of differential forms which can be mapped into the ideal of the heat equation or into its prolongation by an invertible contact transformation. The fact that conservation laws are transformed into conservation laws by contact transformations whereas contact conditions are preserved provides us with criteria for equivalence.

2. EXTENDED POINT TRANSFORMATIONS

In this section we shall briefly describe the most important properties of contact transformations which can be found in Ref. 7 where emphasis is made on groups of contact transformations. A concise summary is also given in Ref. 4. When the number of the dependent variables is greater than one, there will be no other contact transformation than extended point transformations. We shall therefore restrict ourselves from the beginning to point transformations.

Consider the spaces of variables

$$s = \{z | z = (x,t,u^1,\ldots,u^n)\}$$

and

$$S = \{Z | Z = (X,T,U^1,\ldots,U^n)\}.$$

We shall study transformations

$$K: s \to S \tag{2.1}$$

given by

$$X = \bar{X}(z), \quad T = \bar{T}(z), \quad U^i = \bar{U}^i(z). \tag{2.2}$$

Let us separate the variables into dependent and independent ones. This enables us to extend the spaces s and S to s_E and S_E by introducing the derivatives

$$u^i_x, u^i_t,$$

and

$$U^i_X, U^i_T.$$

We endow the extended spaces with the contact forms

$$\omega^i = du^i - u^i_x dx - u^i_t dt \tag{2.3}$$

 0022-2488/82/112043-05$02.50 © 1982 American Institute of Physics

and

$$\Omega^i = dU^i - U^i_X dX - U^i_T dT. \tag{2.4}$$

Now we have to answer two questions. How are derivatives transformed and what happens to contact forms under (2.2)? First of all, a vector valued function

$$u(x,t) = (u^1(x,t),...,u^n(x,t)),$$

is transformed into

$$U(X,T) = (U^1(X,T),...,U^n(X,T)),$$

where

$$U^i(X,T) = \overline{U}^i(x,t,u(x,t)), \tag{2.5a}$$

and

$$X = \overline{X}(x,t,u(x,t)), \tag{2.5b}$$

$$T = \overline{T}(x,t,u(x,t)). \tag{2.5c}$$

From differentiating (2.5a) we obtain

$$D_x \overline{U}^i = U^i_X D_x \overline{X} + U^i_T D_x \overline{T},$$
$$D_t \overline{U}^i = U^i_X D_t \overline{X} + U^i_T D_t \overline{T} \tag{2.6}$$

where (2.5b,c) has been used and the operators of total differentiation:

$$D_x = \partial/\partial x + u^i_x \partial/\partial u^i,$$
$$D_t = \partial/\partial t + u^i_t \partial/\partial u^i.$$

Assuming that the determinant

$$D = D_x \overline{X} \cdot D_t \overline{T} - D_t \overline{X} \cdot D_x \overline{T}$$

does not vanish, we obtain from (2.6) the transformation of the derivatives

$$U^i_X = (D_t \overline{T} \cdot D_x \overline{U}^i - D_x \overline{T} \cdot D_t \overline{U}^i)/D,$$
$$U^i_T = (-D_t \overline{X} \cdot D_x \overline{U}^i + D_x \overline{X} \cdot D_t \overline{U}^i)/D. \tag{2.7}$$

The point transformation K has now been extended to a transformation between the extended spaces s_E and S_E. In what follows we shall not distinguish between K and its extension. The extended transformation will respect the contact conditions. This can be seen as follows. Application of the pull-back mapping K^* to the contact form Ω^i yields

$$K^*(\Omega^i) = (\overline{U}^i_{u^j} - U^i_X \overline{X}_{u^j} - U^i_T \overline{T}_{u^j})\omega^j. \tag{2.8}$$

Thus if a mapping i annuls the contact forms ω^i then $K \circ i$ will annul the contact forms Ω^i.

3. CONSERVATION LAWS AND THE ESTABROOK–WAHLQUIST PROLONGATION OF THE HEAT EQUATION

First of all we shall formulate the heat equation as an ideal of differential forms on the manifold s_E generated by

$$du \wedge dx + du_x \wedge dt, \tag{3.1a}$$

and the contact form

$$du - u_x\, dx - u_t\, dt. \tag{3.1b}$$

An embedding i of a two-dimensional submanifold into s_E is called a solution of (3.1a,b) if these forms are annulled by i^*. Embeddings i coincide with solutions of

$$u_t = u_{xx} \tag{3.2}$$

when functional independence of $x \circ i$ and $t \circ i$ is imposed, see Ref. 5.

By a conservation law of (3.1) or (3.2) we understand a one-form on s_E

$$\omega = f\, dx + g\, dt \tag{3.3}$$

whose exterior derivative $d\omega$ is annulled by an embedding i whenever i annuls (3.1a,b). In particular one-forms whose exterior derivatives lie in the ideal generated by (3.1a,b) are conservation laws.

The heat equation has the following conservation laws:

$$pu\, dx + (pu_x - p_x u)dt \tag{3.4}$$

and

$$pu_x\, dx + (pu_t - p_x u_x)dt, \tag{3.5}$$

where $p(x,t)$ satisfies the adjoint equation

$$p_t = -p_{xx}. \tag{3.6}$$

As far as I know these are the only conservation laws of (3.1a,b) on s_E. (Polynomial conservation laws such as are known for equations of Korteweg–de Vries type cannot exist, see Ref. 8). Some solutions of (3.6) are furnished by the so called heat polynomials

$$p_n(x,t) = (x - 2t\partial/\partial x)^n \circ 1, \quad n = 0,1,\cdots.$$

Instead of (3.1a) we could have used the exterior derivative of any conservation law for the formulation of the heat equation by forms.

The use of adjoint equations for obtaining conservation laws can be generalized to arbitrary linear equations. For example the equation

$$u_t = au_{xx} + bu_x + cu$$

has the following conservation law:

$$pu\, dx + (apu_x + (bp - (ap)_x)u)dt,$$

if p satisfies

$$p_t = -(ap)_{xx} + (bp)_x - cp.$$

The analysis of the following sections is therefore not limited to the heat equation.

In order to obtain further conservation laws, we shall submit the heat equation to the Estabrook–Wahlquist prolongation procedure.[6] A similar prolongation for Burgers' equation is carried out in Ref. 9 where, however, forms are not used. Upon introducing new variables y^k we prolong the ideal (3.1a,b) to

$$du \wedge dx + du_x \wedge dt, \tag{3.7a}$$
$$dy^k - A^k dx - B^k dt, \tag{3.7b}$$
$$du - u_x dx - u_t dt, \tag{3.7c}$$
$$dy^k - y^k_x dx - y^k_t dt, \tag{3.7d}$$

(k being undetermined). Following Estabrook–Wahlquist[6] we require dependence of A^k and B^k upon u, u_x and the new variables y^k. Then we particularize A^k and B^k so that the exterior derivative of (3.7b) lies in the prolonged ideal, i.e., (3.7b) is a conservation law. After obvious calculations which proceed in analogy to those given in Ref. 9 we obtain

$$A^k = A_1^k(y) + A_2^k(y)u,$$
$$B^k = A_3^k(y) - A_4^k(y)u + A_2^k(y)u_x, \tag{3.8}$$

where the following commutation relations must be satisfied:

$$[A_1,A_3] = 0, \quad [A_2,A_4] = 0, \tag{3.9a}$$
$$[A_1,A_2] = [A_1,A_4], \tag{3.9b}$$
$$[A_1,A_2] = A_4. \tag{3.9c}$$

The commutator of A_i and A_j is defined by

$$[A_i,A_j]^k = A_i^h A_{j_y}^k - A_j^h A_{i_y}^k.$$

In what follows the A_i will be assumed to be scalar functions, i.e., $k = 1$, whereby equations (3.9) become solvable.

From (3.9a) it follows that

$$A_3 = cA_1, \quad A_4 = eA_2, \quad c,e = \text{constant}.$$

(3.9b) then leads to

$$c = -e,$$

and (3.9c) to

$$-A_1' + A_2'A_1/A_2 = e. \tag{3.10}$$

If A_1 and A_2 are determined according to (3.10) we have a two-dimensional Lie algebra satisfying the commution relations (3.9). By inserting the algebra into (3.8) we obtain

$$A = A_1 + A_2u,$$
$$B = -eA_1 - eA_2u + A_2u_x. \tag{3.11}$$

Let us introduce the prolonged space

$$s_p = \{z_p | z_p = (x,t,u,u_x,u_t,y,y_x,y_t)\}.$$

The heat equation can now be expressed on s_p by

$$(y_x - A)dx + (y_t - B)dt,$$
$$du - u_x\,dx - u_t\,dt, \tag{3.12}$$

and

$$dy - y_x dx - y_t dt,$$

since by annulling (3.12) by an embedding i the heat equation is given back as integrability condition. More precisely, the differential system expressed by (3.12) is

$$y_x = A, \tag{3.13a}$$
$$y_t = B, \tag{3.13b}$$
$$u_t = u_{xx}. \tag{3.13c}$$

Here the close relationship between prolongation and searching for Bäcklund transformations becomes evident, see Refs. 6,9. Equations (3.13a,b) can be considered as a Bäcklund transformation with (3.13c) as integrability condition.

4. EQUIVALENCE OF EQUATIONS TO THE NONPROLONGED HEAT EQUATION

Let us be given a second order equation in the space s_E and the heat equation in the space S_E

$$U_T = U_{XX}. \tag{4.1}$$

Following Bluman and Kumei[4] we shall call both equations equivalent if their solutions are in one-to-one correspondence by an invertible point transformation (2.2):

$$K: s \to S.$$

We shall now derive conditions for equivalence. Let us assume that both equations were given by differential forms. Then a solution i of the equation under consideration is mapped into a solution $K \circ i$ of (4.1). By $(K \circ i)^* = i^* \circ K^*$ it follows that a solution i of the equation we are looking for will annul the forms

$$K^*(dU \wedge dX + dU_X \wedge dT) \tag{4.2a}$$
$$K^*(dU - U_X dX - U_T dT). \tag{4.2b}$$

[Here K means the extended transformation (2.7)].

Equation (4.2a) is the exterior derivative of

$$K^*(U dX + U_X dT),$$

being a one-form on s_E while (4.2b) will be annuled by i whenever i annuls the contact form

$$\omega = du - u_x dx - u_t dt,$$

see (2.8).

This leads us to describe the class of equations in equivalence with (4.1) by a closed two form, i.e., a conservation law $d\omega^1$ and the contact form ω. Conversely we obtain by analoguous reasonings, that a solution i of the heat equation must annul

$$(K^{-1})^* d\omega^1 = d((K^{-1})^* \omega^1).$$

This means that $(K^{-1})^* \omega^1$ is a conservation law of the heat equation (4.1). Therefore, we obtain the equivalent equation as

$$d\omega^1 = d(f\,dx + g\,dt),$$
$$\omega = du - u_x dx - u_t dt. \tag{4.3}$$

A sufficient condition for equivalence can be formulated now by using the conservation laws (3.4) and (3.5). We must have

$$(K^{-1})^* \omega^1 = pU\,dX + (pU_X - p_X U)dT$$

or

$$(K^{-1})^* \omega^1 = pU_X\,dX + (pU_T - p_X U_X)dT$$

meaning that

$$\omega^1 = K^*(pU\,dX + (pU_X - p_X U)dT) \tag{4.4a}$$

or

$$\omega^1 = K^*(pU_X\,dX + (pU_T - p_X U_X)dT). \tag{4.4b}$$

By applying K^* to a conservation law

$$(K^{-1})^* \omega^1 = F\,dX + G\,dT$$

we obtain keeping the contact condition in mind

$$K^*(F\,dX + G\,dT) = (F(K)D_x\overline{X} + G(K)D_x\overline{T})dx + (F(K)D_t\overline{X} + G(K)D_t\overline{T})dt.$$

[Here $F(K)$ stands for the mapping composed of K and F, etc.]. Therefore, equivalence of (4.3) to the heat equation now means that

$$f = F(K)D_x\overline{X} + G(K)D_x\overline{T},$$
$$g = F(K)D_t\overline{X} + G(K)D_t\overline{T}, \tag{4.5}$$

where F and G are given by some conservation law (4.4a,b)

and p solves

$$p_T = -p_{XX}.$$

To sum up, we have obtained the following condition for equivalence. An equation given by (4.3) is equivalent to the heat equation if there exists some conservation law and an invertible point transformation such that (4.5) holds. On the other hand, (4.5) can be used to determine wide classes being in equivalance with the heat equation .

Example 1: we use (4.4a) with $p = 1$, i.e.,

$$F = U, \quad G = U_X.$$

Consider the transformation

$$\overline{X} = x\exp(t), \quad \overline{T} = \exp(2t)/2, \quad \overline{U} = u\exp(-t),$$

and extend it according to (2.8). Then (4.5) yields

$$f = u, \quad g = u_x + xu,$$

giving the Fokker–Planck equation

$$u_t = u_{xx} + (xu)_x$$

as an equivalent equation.

Example 2: We use (4.4b) with $p = X$. This means

$$F = XU_X, \quad G = XU_T - U_X.$$

Let us try the transformation

$$\overline{X} = u, \quad \overline{T} = t, \quad \overline{U} = x.$$

Equation (4.5) together with (2.7) then gives

$$f = u, \quad g = (-1/u_x),$$

whereby

$$u_t = u_x^{-2} u_{xx}$$

becomes equivalent to the heat equation.

5. EQUIVALENCE OF EQUATIONS TO THE PROLONGED HEAT EQUATION

Let us consider the prolonged heat equation in the space S_E yielding

$$(Y_X - A)dX + (Y_T - B)dT, \tag{5.1a}$$

$$dU - U_X\, dX - U_T\, dT, \tag{5.1b}$$

$$dY - Y_X\, dX - Y_T\, dT. \tag{5.1c}$$

By (5.1) we express the differential equation, see (3.12),

$$\begin{aligned} Y_X &= A, \\ Y_T &= B, \\ U_T &= U_{XX}. \end{aligned} \tag{5.2}$$

Let us now find equations in equivalence with (5.2) by the same procedure as in the foregoing section. Here, however the only conservation law at hand is given by (5.1a). We consider a one-to-one point transformation K given by (2.2) with $u^1 = u$, $u^2 = y$. Next, we apply the pullback K^* to the conservation law (5.1a) of (5.2). To the form thus obtained, we add the contact forms on s_E obtaining an equation given by

$$\begin{aligned} &K^*((Y_X - A)dX + (Y_T - B)dT), \\ &du - u_x\, dx - u_t\, dt, \\ &dy - y_x\, dx - y_t\, dt. \end{aligned} \tag{5.3}$$

Let $U_X(K), U_T(K), Y_X(K)$ and $Y_T(K)$ be given by (2.7) then we obtain for (5.3)

$$\begin{aligned} &((Y_X(K) - A(K))D_x\overline{X} + (Y_T(K) - B(K))D_x\overline{T})dx \\ &+ ((Y_X(K) - A(K))D_t\overline{X} + (Y_T(K) - B(K))D_t\overline{T})dt, \\ &du - u_x dx - u_t\, dt, \quad dy - y_x dx - y_t\, dt. \end{aligned} \tag{5.4}$$

Solutions of (5.1) and (5.4) are now in one-to-one correspondence by K. Stated otherwise, solutions of (5.2) are in one-to-one correspondence with the solutions of the equation

$$(Y_X(K) - A(K))D_x\,\overline{X} + (Y_T(K) - B(K))D_x\overline{T} = 0, \tag{5.5a}$$

$$(Y_X(K) - A(K))D_t\,\overline{X} + (Y_T(K) - B(K))D_t\overline{T} = 0, \tag{5.5b}$$

and its integrability condition. Equations (5.5a,b) can again be used for deriving classes of equivalent equations.

Example 3: we use the following transformation K

$$X = x, \quad T = t, \quad U = \overline{U}(u)\,y, \quad Y = y.$$

Extending K by (2.7) and inserting it into (5.5) yields

$$\begin{aligned} &y_x - (A_1(y) + A_2(y)\,\overline{U}(u)\,y) = 0, \\ &y_t - \{ -eA_1(y) - eA_2(y)\,\overline{U}(u)\,y \\ &\qquad + A_2(y)\,D_x(\overline{U}(u)\,y)\} = 0, \end{aligned}$$

where A_1 and A_2 satisfy (3.10). Upon particularizing $A_2 = 1$, $A_1 = 0$, and $e = 0$, we obtain

$$\begin{aligned} y_x &= \overline{U}(u)\,y, \\ y_t &= (\overline{U}'(u)\,u_x + \overline{U}(u)^2)\,y, \\ u_t &= u_{xx} + 2\overline{U}(u)\,u_x + \overline{U}''(u)\,u_x^2/\overline{U}'(u). \end{aligned} \tag{5.6}$$

Thus we have obtained equivalence of (5.6) to the prolonged heat equation by means of a Cole–Hopf transformation

$$x = X, \quad t = T, \quad u = \overline{U}^{-1}(U/Y), \quad y = Y,$$

which appears in our setting as an invertible point transformation. Obviously, (5.6) coincides with Burgers equation if we set $\overline{U}(u) = u$.

Example 4: consider the transformation K

$$X = y, \quad T = t, \quad U = \overline{U}(u), \quad Y = x.$$

From (2.7) it follows that

$$\begin{aligned} U_X &= \overline{U}'u_x/y_x, \\ U_T &= \overline{U}'u_t - \overline{U}'u_x y_t/y_x, \\ Y_X &= 1/y_x, \\ Y_T &= -y_t/y_x. \end{aligned}$$

By inserting into Eqs. (5.5) we obtain the equivalent equation

$$\begin{aligned} y_x &= 1/(A_2(x)\overline{U}(u) + A_1(x)), \\ y_t &= -A_2(x)\overline{U}'(u)u + eA_2(x)\overline{U}(u)/(A_2(x)\overline{U}(u) + A_1(x)) \\ &\quad + eA_1(x)/(A_2(x)\overline{U}(u) + A_1(x)), \\ (-1/(A_2(x)\overline{U}(u) + A_1(x)))_t &= (A_2(x)\overline{U}'(u)u_x)_x \\ &\quad - eA_1'(x)\overline{U}'(u)/(A_2(x)\overline{U}(u) + A_1(x)). \end{aligned}$$

Upon particularizing, we obtain remarkable diffusion equations. Setting $e = 0$ and $A_1 = 0$ we can satisfy (3.10) by arbitrary functions A_2. This gives

$$(-1/A_2(x)\overline{U}(u))_t = (A_2(x)\overline{U}'(u)u_x)_x.$$

Examples important in nonlinear diffusion processes are obtained by setting

$$\overline{U} = \exp(-au),$$

and

$$\overline{U} = -u^{-1}, \quad A_2 = 1,$$

yielding

$$(\exp(au)/A_2(x))_t = (A_2(x)\exp(-au)u_x)_x,$$

and

$$u_t = (u^{-2}u_x)_x.$$

Various other methods for linearizing these equations can be found in Refs. 10–12.

Example 5: we choose again $e = 0$, $A_1 = 0$ and A_2 arbitrary. The transformation

$$X = x, \quad T = y, \quad U = \overline{U}(u), \quad Y = t,$$

is extended by (2.7) to

$$U_X = (y_t\overline{U}'(u)u_x - y_x\overline{U}(u)u_t)/y_t,$$
$$U_T = \overline{U}'(u)u_t/y_t,$$
$$Y_X = -y_x/y_t,$$
$$Y_T = 1/y_t,$$

Upon inserting into (5.5) we obtain

$$y_x = \overline{U}(u)/\overline{U}'(u)(A_2(t)\overline{U}(u)u_t - u_x),$$
$$y_t = 1/A_2(t)\overline{U}(u)(A_2(t)\overline{U}(u)u_t - u_x).$$

By setting $A_2 = 1$ the following integrability condition is obtained

$$-\overline{U}(u)^2\overline{U}'(u)u_{tt} + (\overline{U}(u) + \overline{U}(u)\overline{U}'(u))u_{xt} - \overline{U}'(u)u_{xx} + (\overline{U}(u)\overline{U}'(u)^2 - \overline{U}(u)\overline{U}'(u))u_t^2 + \overline{U}(u)\overline{U}''(u)u_xu_t - \overline{U}''(u)u_x^2 = 0.$$

The restriction $\overline{U} = u$ finally yields the following equation equivalent to the prolonged heat equation:

$$-u^2u_{tt} + 2uu_{xt} - u_{xx} = 0.$$

[1] M. Flato and J. Simon, J. Math. Phys. **21**, 913–917 (1980).
[2] R. L. Anderson, J. Harnad, and P. Winternitz, Physica D (1982).
[3] G. W. Bluman, SIAM J. Appl. Math. **39**, 238–247 (1980).
[4] S. Kumei and G. W. Bluman, preprint (1981).
[5] B. K. Harrison and F. B. Estabrook, J. Math. Phys. **12**, 653–666 (1971).
[6] H. D. Wahlquist and F. B. Estabrook, J. Math. Phys. **16**, 1–7 (1975).
[7] R. L. Anderson and N. H. Ibragimov, "Lie–Bäcklund Transformations in Applications," in *SIAM Studies in Applied Mathematics* (SIAM, Philadelphia, 1979).
[8] A. Galindo, J. Math. Phys. **20**, 1256–1259 (1979).
[9] D. J. Kaup, Physica D **1**, 391–411 (1980).
[10] A. Munier, J. R. Burgan, J. Gutierrez, E. Fijalkow, and M. R. Feix, SIAM J. Appl. Math. **40** (2), 191–207 (1981).
[11] G. W. Bluman and S. Kumei, J. Math. Phys. **21**, 1019–1023 (1980).
[12] A. S. Fokas and Y. C. Yortsos, SIAM J. Appl. Math. **42** (2), 318–332 (1982).

Space–time memory functions and solution of nonlinear evolution equations

Takeo Nishigori
Department of Nuclear Engineering, Osaka University, Yamadaoka 2-1, Suita, Osaka 565, Japan

(Received 1 June 1982; accepted for publication 25 June 1982)

A new approach is presented for solving a certain class of nonlinear partial differential equations. A space–time memory function $\Lambda(\mathbf{r},t)$ is introduced to exactly convert a given nonlinear evolution equation into the following linear form: $(\partial/\partial t)f(\mathbf{r},t) = \Omega(\mathbf{r})f(\mathbf{r},t) + \int_0^t dt' \int d\mathbf{r}' \Lambda(\mathbf{r}-\mathbf{r}',t-t')f(\mathbf{r}',t')$. A Markovian integro-differential operator $\Omega(\mathbf{r})$ and the memory function $\Lambda(\mathbf{r},t)$ reflect the nonlinearity, and are determined depending on a given initial condition. The approach is useful if higher-order memory functions associated with Λ are insensitive to approximation. The Korteweg–de Vries equation is treated as an example. For certain initial profiles the memory function is shown to be identically zero, and we find exact *linear* partial differential equations leading to the single- and the two-soliton solution. In the case of the three-soliton solution, the second-order memory function vanishes exactly, and $\Lambda(\mathbf{r},t)$ is found to be a single exponential function of t.

PACS numbers: 02.30.Jr, 02.60.Nm, 02.50.Ga

I. INTRODUCTION

We have developed in a previous paper[1] a new approach to nonlinear initial-value problems on the basis of the fact that a nonlinear *ordinary* differential equation for a variable $x(t)$ can exactly be converted into a linear non-Markovian equation

$$\frac{d}{dt}x(t) = \omega x(t) + \int_0^t \Lambda(t-t')x(t')dt', \tag{1}$$

where a renormalized frequency ω and a memory kernel $\Lambda(t-t')$ reflect the nonlinearity and are determined in terms of a given initial value $x(0)$. Since the linear equation (1) is easily solved, the problem is reduced to determination of the memory functions $\Lambda(t)$. It has been shown that higher-order memory functions $\Lambda_n(t)$ associated with $\Lambda(t)$ are insensitive to approximation, and a simple and drastic approximation, such as $\Lambda_n(t) \propto \delta(t)$ for a dissipative system[1] and $\Lambda_n(t) \simeq 0$ for an oscillatory system,[2] yields statisfactory results.[3]

In this paper we extend the approach to space-dependent cases, and consider a quantity $f(\mathbf{r},t)$ which evolves in space and time according to a nonlinear *partial* differential equation. The extension to the space-dependent case is, however, not unique. The simplest way would be to regard $\mathbf{r}$ to be merely parametric; then, as mentioned in Ref. 1, $f(\mathbf{r},t)$ obeys a non-Markovian, but spatially local, equation of the form (1) with ω and Λ depending on the parameter $\mathbf{r}$. An early work[4] on the memory function formalism in classical liquids is quite suggestive in this respect; namely, a natural generalization of Eq. (1) is to replace $x(t)$ by a vector $\mathbf{f}(t)$ whose components $f(\mathbf{r},t)$ are indexed by a continuous parameter $\mathbf{r}$. Thus, we have a non-Markovian and spatially nonlocal equation

$$\frac{\partial}{\partial t}f(\mathbf{r},t) = \Omega(\mathbf{r})f(\mathbf{r},t) + \int_0^t dt' \times \int d\mathbf{r}' \Lambda(\mathbf{r}-\mathbf{r}',t-t')f(\mathbf{r}',t'); \tag{2a}$$

$$\Omega(\mathbf{r})f(\mathbf{r},t) = \int \omega(\mathbf{r}-\mathbf{r}')f(\mathbf{r}',t)d\mathbf{r}'. \tag{2b}$$

We have assumed here that the system is spatially homogeneous and infinitely extended. In contrast to the classical kinetic theory, where the frequency "matrix" $\omega(\mathbf{r}-\mathbf{r}')$ is uniquely determined by virtue of Mori's projection operator,[5] there is no such unique way of determining $\omega(\mathbf{r}-\mathbf{r}')$ in the present deterministic systems.[6] Namely, Eq. (2a) gives at $t=0$ a Fredholm integral equation for $\omega(\mathbf{r})$,

$$\frac{\partial}{\partial t}f(\mathbf{r},0) = \int f(\mathbf{r}-\mathbf{r}',0)\omega(\mathbf{r}')d\mathbf{r}', \tag{3}$$

which does not necessarily have a well-behaved unique solution; when the original evolution equation contains linear terms in $f(\mathbf{r},t)$ and in its spatial derivatives, as is often the case, the function $\omega(\mathbf{r})$ should involve a Dirac delta function $\delta(\mathbf{r})$ and its derivatives. We therefore assume a simple linear integro-differential operator for $\Omega(\mathbf{r})$ so that Eq. (3) is fulfilled.

Assuming that an appropriate solution exists to integral equations like Eq. (3), we can apply the formalism presented in the previous paper[1] to the space-dependent cases as well to derive Eq. (2a). We extract the linear part $\Omega(\mathbf{r})f(\mathbf{r},t)$ from $\partial f(\mathbf{r},t)/\partial t$. We then treat the residual

$$f_1(\mathbf{r},t) = \frac{\partial}{\partial t}f(\mathbf{r},t) - \Omega(\mathbf{r})f(\mathbf{r},t),$$

which is nonlinear in f, as a new dependent variable, and consider the nonlinear equation $\partial f/\partial t = \Omega f + f_1$ for f as a linear equation for f and f_1. An evolution equation for f_1 is constructed to define another dependent variable $f_2(\mathbf{r},t)$, and so on. We thus convert the original nonlinear partial differential equation for f into an infinite set of linear integro-differential equations for $f, f_1, f_2, \ldots$. The irrelevant higher variables $f_1, f_2 \ldots$ are then eliminated to obtain the single closed linear integro-differential equation (2a). A second- and higher-order memory functions are associated with Λ, and the Fourier–Laplace transform $\hat{\Lambda}(\mathbf{k},z)$ of the memory function is found to be given by a sum of products of infinite continued fractions in z.

The usefulness of the present approach depends on whether or not the memory function is easily amenable to approximation. As an illustrative example we treat the

 0022-2488/82/112048-05$02.50 © 1982 American Institute of Physics

Korteweg–de Vries (KdV) equation. Several exact methods have been developed to obtain soliton solutions.[7] A bilinearization method has been given by Hirota.[8] We now study a linearization method. The structure of the linear equation (2a) varies according to the initial profile [note that $\Omega(\mathbf{r})$ determined by Eq. (3) depends on $f(\mathbf{r},0)$]. We find a first- and a third-order *linear* partial differential equation leading exactly to the single- and the two-soliton solution, respectively; the memory function Λ vanishes in these cases. Equation (2a) with a single exponential memory function, on the other hand, gives the exact three-soliton solution; the second-order memory function vanishes in this case.

In Sec. II the linear equation (2a) is derived and a general formula for the memory function is given. In Sec. III the KdV equation is treated. The final section is added for remarks.

II. MEMORY FUNCTION AND LINEARIZATION

Consider a nonlinear partial differential equation

$$\Phi(f,f_t,f_x,f_y,f_z,f_{xx},f_{xy},\ldots,f_{tx},\ldots,f_{txx},\ldots,\mathbf{r},t)=0 \tag{4a}$$

subject to an initial condition

$$f(\mathbf{r},0)=a(\mathbf{r})\not\equiv 0 \tag{4b}$$

and to appropriate boundary conditions. Here, Φ is a nonlinear function of f and its partial derivatives $f_t, f_x, \ldots$, and may explicitly depend on the space and time coordinates $\mathbf{r}=(x,y,z)$ and t. We have assumed that $f(\mathbf{r},0)\not\equiv 0$ and that Eq. (4a) does not involve the second- and higher-order derivatives with respect to t. We derive in this section the linear equation (2a) from Eqs. (4a) and (4b) along the line of reasoning of the previous paper.[1]

We extract from f_t the linear part (2b) as follows:

$$f_t(\mathbf{r},t)=\int \omega(\mathbf{r}-\mathbf{r}')f(\mathbf{r}',t)d\mathbf{r}'+f_1(\mathbf{r},t). \tag{5}$$

To determine $\omega(\mathbf{r})$ we require that the residual f_1 vanish at $t=0$. Then,

$$a^{(1)}(\mathbf{r})=\int a(\mathbf{r}-\mathbf{r}')\omega(\mathbf{r}')d\mathbf{r}', \tag{6}$$

where $a(\mathbf{r})$ and $a^{(1)}(\mathbf{r})=f_t(\mathbf{r},0)$ are given by Eqs. (4a) and (4b). We look for a solution to the integral equation (6) of the following form:

$$\omega(\mathbf{r})=\tilde{\omega}(\mathbf{r})+\omega^{(0)}\delta(\mathbf{r})+\boldsymbol{\omega}^{(1)}\cdot\frac{\partial}{\partial\mathbf{r}}\delta(\mathbf{r})+\cdots, \tag{7}$$

where $\tilde{\omega}(\mathbf{r})$ is a well-behaved function. Assuming for simplicity that $\omega^{(0)}, \boldsymbol{\omega}^{(1)},\ldots$ are constant, we rewrite Eq. (6) as

$$\begin{aligned}a^{(1)}(\mathbf{r})&=\int\tilde{\omega}(\mathbf{r}-\mathbf{r}')a(\mathbf{r}')d\mathbf{r}'+\omega^{(0)}a(\mathbf{r})+\boldsymbol{\omega}^{(1)}\cdot\frac{\partial}{\partial\mathbf{r}}a(\mathbf{r})+\cdots\\&=\Omega(\mathbf{r})a(\mathbf{r}),\end{aligned} \tag{8}$$

where the second equality defines the operator $\Omega(\mathbf{r})$. We suppose that the unknowns $\tilde{\omega}(\mathbf{r}),\omega^{(0)},\boldsymbol{\omega}^{(1)},\ldots$ can be determined by Eq. (8) so that the sum of the derivatives in Eq. (7) terminates at a certain low order. Equation (5) then becomes

$$f_t(\mathbf{r},t)=\Omega(\mathbf{r})f(\mathbf{r},t)+f_1(\mathbf{r},t), \tag{9}$$

and the residual f_1 is now defined by

$$f_1(\mathbf{r},t)=f_t(\mathbf{r},t)-\Omega(\mathbf{r})f(\mathbf{r},t). \tag{10}$$

As mentioned already, we regard the nonlinear equation (9) for f as a linear equation for f and the new dependent variable f_1. An evolution equation for f_1 is found by differentiating Eq. (10) with respect to t. Extracting linear terms in f and f_1, we write

$$\begin{aligned}\frac{\partial}{\partial t}f_1(\mathbf{r},t)=&\int\alpha_{10}(\mathbf{r}-\mathbf{r}')f(\mathbf{r}',t)d\mathbf{r}'\\&+\int\omega_1(\mathbf{r}-\mathbf{r}')f_1(\mathbf{r}',t)d\mathbf{r}'+f_2(\mathbf{r},t).\end{aligned} \tag{11}$$

We require that $f_2(\mathbf{r},0)\equiv 0$ to obtain an integral equation for $\alpha_{10}(\mathbf{r})$,

$$a_1^{(1)}(\mathbf{r})=\int a(\mathbf{r}-\mathbf{r}')\alpha_{10}(\mathbf{r}')d\mathbf{r}', \tag{12}$$

where $a_1^{(1)}(\mathbf{r})=\partial f_1(\mathbf{r},0)/\partial t$ is given in terms of the initial profile $a(\mathbf{r})$. We assume that a solution $\alpha_{10}(\mathbf{r})$ can be found in a form similar to (7). The f_1 term in Eq. (11), on the other hand, can uniquely be separated by decomposing $\partial f_1(\mathbf{r},t)/\partial t$ into terms of different orders at the point $t=0$. A term of order 0 has already been extracted. Let $f_1(\mathbf{r},t)$ be of order ν_1 at the zero $t=0$, i.e.,

$$\begin{aligned}&f_1(\mathbf{r},t)=a_1^{(\nu_1)}(\mathbf{r})t^{\nu_1}+a_1^{(\nu_1+1)}(\mathbf{r})t^{\nu_1+1}+\cdots;\\&\nu_1\geqslant 1,\quad a_1^{(\nu_1)}(\mathbf{r})\not\equiv 0.\end{aligned} \tag{13}$$

We require f_2 to have no term proportional to t^{ν_1}, which, if any, should be included in the f_1 term. Comparing the terms of ν_1 th power on both sides of Eq. (11), we obtain

$$\begin{aligned}(\nu_1+1)a_1^{(\nu_1+1)}(\mathbf{r})=&\int\alpha_{10}(\mathbf{r}-\mathbf{r}')a^{(\nu_1)}(\mathbf{r}')\,d\mathbf{r}'\\&+\int a_1^{(\nu_1)}(\mathbf{r}-\mathbf{r}')\omega_1(\mathbf{r}')d\mathbf{r}',\end{aligned} \tag{14}$$

where $a^{(\nu_1)}(\mathbf{r})=[\partial^{\nu_1}f(\mathbf{r},0)/\partial t^{\nu_1}]/\nu_1!$. We again assume that $\omega_1(\mathbf{r})$ can be determined from the integral equation (14). The f and f_1 components of $\partial f_1/\partial t$ have thus been extracted, and the residual now defines the new dependent variable $f_2(\mathbf{r},t)$, which is of order $\nu_2\neq 0, \nu_1$ at the zero $t=0$.

We further construct an evolution equation for f_2 and define another new variable $f_3(\mathbf{r},t)$, and so on. In general, we introduce a higher variable $f_n(\mathbf{r},t)$, which is of order $\nu_n\geqslant 1$ at the zero $t=0$, and which does not contain terms proportional to t^{ν_m}, $1\leqslant m\leqslant n-1$. Note that a higher variable is not necessarily of higher order at the zero than the lower variables. Decomposing $\partial f_n/\partial t$ into terms of different orders at $t=0$, we write

$$\begin{aligned}\frac{\partial}{\partial t}f_n(\mathbf{r},t)=&\sum_{m=0}^{n-1}\int\alpha_{nm}(\mathbf{r}-\mathbf{r}')f_m(\mathbf{r}',t)d\mathbf{r}'\\&+\int\omega_n(\mathbf{r}-\mathbf{r}')f_n(\mathbf{r}',t)d\mathbf{r}'+f_{n+1}(\mathbf{r},t),\\&n=0,1,2,\ldots,\end{aligned} \tag{15}$$

where $f_0=f$, $\alpha_{0m}=0$, and $\omega_0=\omega$. Denoting that $a_m^{(\nu)}(\mathbf{r})=[\partial^{\nu}f_m(\mathbf{r},0)/\partial t^{\nu}]/\nu!$, and comparing like powers of t on both sides of Eq. (15), we have

$$a_n^{(1)}(\mathbf{r}) = \int a(\mathbf{r}-\mathbf{r}')\alpha_{n0}(\mathbf{r}')d\mathbf{r}', \tag{16a}$$

$$(\nu_m+1)a_n^{(\nu_m+1)}(\mathbf{r}) = \sum_{m'=0}^{m-1}\int \alpha_{nm'}(\mathbf{r}-\mathbf{r}')a_{m'}^{(\nu_m)}(\mathbf{r}')d\mathbf{r}' + \int a_m^{(\nu_m)}(\mathbf{r}-\mathbf{r}')\alpha_{nm}(\mathbf{r}')d\mathbf{r}', \quad m=1,2,\ldots,n-1, \tag{16b}$$

$$(\nu_n+1)a_n^{(\nu_n+1)}(\mathbf{r}) = \sum_{m'=0}^{n-1}\int \alpha_{nm'}(\mathbf{r}-\mathbf{r}')a_{m'}^{(\nu_n)}(\mathbf{r}')d\mathbf{r}' + \int a_n^{(\nu_n)}(\mathbf{r}-\mathbf{r}')\omega_n(\mathbf{r}')d\mathbf{r}', \tag{16c}$$

which are integral equations determining $\alpha_{n0},\ldots,\alpha_{nn-1}$, and ω_n successively. Here, we put $\alpha_{nm}=0$ for those m for which

$$\nu_m \leqslant \nu_n - 2 \tag{17a}$$

or

$$\nu_m = \nu_1 - 1, \nu_2 - 1, \ldots, \nu_{n-1} - 1, \tag{17b}$$

since $\partial f_n/\partial t$ does not contain terms of these powers. We again assume that Eqs. (16a)–(16c) have solutions of a form similar to (7).

We now have an infinite set of linear equations (15) with initial conditions

$$f(\mathbf{r},0) = a(\mathbf{r}), \quad f_n(\mathbf{r},0) = 0, \quad n \geqslant 1. \tag{18}$$

In terms of Fourier–Laplace transforms

$$\hat{f}_n(\mathbf{k},z) = \int_0^\infty dt\, e^{-zt}\int d\mathbf{r}\, e^{i\mathbf{k}\cdot\mathbf{r}} f_n(\mathbf{r},t) \quad ,$$

we readily obtain the solution

$$\hat{f}_n(\mathbf{k},z) = \sum_{m=0}^{n-1}\hat{\Lambda}_{nm}(\mathbf{k},z)\hat{f}_m(\mathbf{k},z), \quad n \geqslant 1, \tag{19}$$

where

$$\hat{\Lambda}_{nm}(\mathbf{k},z) = \hat{\Xi}_n(\mathbf{k},z)[\hat{\alpha}_{nm}(\mathbf{k}) + \hat{\Lambda}_{n+1m}(\mathbf{k},z)] = \hat{\Xi}_n\hat{\alpha}_{nm} + \hat{\Xi}_n\hat{\Xi}_{n+1}\hat{\alpha}_{n+1m} + \cdots, \tag{20}$$

$$\hat{\Xi}_n(\mathbf{k},z) = 1/[z - \hat{\omega}_n(\mathbf{k}) - \hat{\Lambda}_{n+1n}(\mathbf{k},z)]. \tag{21}$$

The first equation in Eqs. (15), i.e., Eq. (9), then gives Eq. (2a) with the space-time memory function $\Lambda(\mathbf{r},t) = \Lambda_{10}(\mathbf{r},t)$ given by the inverse transform of $\hat{\Lambda}_{10}(\mathbf{k},z)$, which is a sum of products of infinite continued fractions in z.

Considerable simplification of the above results occurs when, as is often the case, the nth variable f_n is of order n at the zero $t=0$ for every n. Then, since $\nu_n = n$, we see from Eq. (17a) that $\alpha_{nm}=0$ for $m \leqslant n-2$. Hence, the variable f_n is coupled only with f_{n-1}, f_n, and f_{n+1}, and the memory function $\hat{\Lambda}_{10}(\mathbf{k},z) = \hat{\Xi}_1(\mathbf{k},z)\hat{\alpha}_{10}(\mathbf{k})$ is given by the single infinite continued fraction $\hat{\Xi}_1$.

The nonlinear effect can thus be converted exactly into the linear space–time memory effect. The memory function is, however, expressed in terms of the infinite continued fractions, and an approximate truncation is needed in general. We expect that there are many problems in which the higher-order memory functions become insensitive to approximation with increasing order, as we have seen in the space-independent cases.[1,2] In an example discussed in the next section, it is shown that the memory function Λ_{10} or the second-order memory functions Λ_{20} and Λ_{21} can be made to vanish exactly by an appropriate dependent variable transformation.

III. EXAMPLE

Let us consider as an illustrative example the KdV equation

$$u_t + 6uu_x + u_{xxx} = 0, \tag{22a}$$

which exhibits N solitons for the boundary condition $u(\pm\infty,t)=0$ and for the special initial profiles[7]

$$u(x,0) = 2\kappa^2 A\,\mathrm{sech}^2\kappa x, \quad A = \tfrac{1}{2}N(N+1), \tag{22b}$$

where we have assumed for simplicity that $u(x,0)$ contains only one parameter κ. Several exact methods have been found for solving nonlinear equations like Eq. (22a). Hirota used a transformation

$$u(x,t) = 2\frac{\partial^2}{\partial x^2}\ln f(x,t) \tag{23}$$

to bilinearize the KdV equation (22a) and to obtain the exact multisoliton solutions.[8] We now examine the linear equation (2a) for the variable $f(x,t)$ introduced by Hirota.

It is convenient to rewrite Eq. (23) as

$$f(x,t) = \exp\left(\frac{1}{2}\int_{-\infty}^{x}dx'\int_{-\infty}^{x'}dx''u(x'',t)\right), \tag{24}$$

where an arbitrary factor of the form $\exp[C_1(t)x + C_2(t)]$ has been omitted to avoid unnecessary complications. The initial condition for f is

$$f(x,0) = (1+e^{Kx})^A, \tag{25}$$

where $K = 2\kappa$. Our first equation is

$$\frac{\partial}{\partial t}f(x,t) = \Omega(x)f(x,t) + f_1(x,t). \tag{26}$$

It follows from the requirement $f_1(x,0)=0$ that

$$-K^3A[1+(3A-1)e^{Kx} + Ae^{2Kx}]e^{Kx}(1+e^{Kx})^{A-3} = \Omega(x)(1+e^{Kx})^A. \tag{27}$$

It is seen that Ω is a differential operator

$$\Omega(x) = \sum_{\nu=1}^{3}\alpha_\nu\frac{\partial^\nu}{\partial x^\nu}. \tag{28}$$

For $N=1$ $(A=1)$ we have $\alpha_1 = -K^2 - K\alpha_2 - K^2\alpha_3$ with arbitrary α_2 and α_3. The residual f_1 vanishes identically, as shown in the Appendix. Hence, putting, say, $\alpha_2=\alpha_3=0$, we obtain an exact linear equation

$$\frac{\partial}{\partial t}f(x,t) = -K^2\frac{\partial}{\partial x}f(x,t), \tag{29}$$

which is subject to the initial condition $f(x,0) = 1 + \exp Kx$. The unique solution $f(x,t) = 1 + \exp K(x - K^2t)$ gives the well-known single soliton solution $u(x,t) = 2\kappa^2\,\mathrm{sech}^2\kappa(x - 4\kappa^2t)$. Another choice for the arbitrary constants α_2 and α_3 yields the identical result.

For $N \geqslant 2$, on the other hand, we obtain

$$\frac{\partial}{\partial t}f(x,t) = \left(2K^2\frac{A}{A-2}\frac{\partial}{\partial x} - 3K\frac{A}{A-2}\frac{\partial^2}{\partial x^2} + \frac{2}{A-2}\frac{\partial^3}{\partial x^3}\right)f(x,t) + f_1(x,t). \tag{30}$$

It is shown in the Appendix that

$$f_1(x,t)\equiv 0 \tag{31}$$

for $N=2,(A=3)$. We thus obtain the following exact linear equation:

$$\frac{\partial}{\partial t}f(x,t)=\left(6K^2\frac{\partial}{\partial x}-9K\frac{\partial^2}{\partial x^2}+2\frac{\partial^3}{\partial x^3}\right)f(x,t) \tag{32}$$

with the initial condition $f(x,0)=(1+\exp Kx)^3$. The solution is

$$f(x,t)=1+3e^{Kx-K^3t}+3e^{2Kx-(2K)^3t}+e^{3Kx-9K^3t}, \tag{33}$$

which gives the two-soliton solution.

For $N\neq 1,2$, the residual f_1 no longer vanishes, and the memory term appears. For $N=3$, for example, we have

$$\begin{aligned}\frac{\partial}{\partial t}f(x,t)&=\Omega(x)f(x,t)+f_1(x,t),\\ \frac{\partial}{\partial t}f_1(x,t)&=A(x)f(x,t)+\omega_1 f_1(x,t)+f_2(x,t),\end{aligned} \tag{34}$$

corresponding to Eq. (15). Here, $\Omega(x)$ is the differential operator in Eq. (30) with $A=6$, $\omega_1=-18K^3$, and $A(x)$ is a sixth-order differential operator

$$\begin{aligned}&A(x)=27\sum_{\nu=1}^{6}\beta_\nu\frac{\partial^\nu}{\partial x^\nu};\\ &\beta_1=20K^5,\beta_2=-127K^4/3,\beta_3=31K^3,\\ &\beta_4=-121K^2/12,\beta_5=3K/2,\beta_6=-1/12.\end{aligned} \tag{35}$$

It is seen in the Appendix that

$$f_2(x,t)\equiv 0, \tag{36}$$

which truncates the infinite chain of equations (15) exactly. The memory function takes the form

$$A(x,t)=27e^{\omega_1 t}\sum_{\nu=1}^{6}\beta_\nu\delta^{(\nu)}(x), \tag{37}$$

and Eq. (2a) becomes

$$\begin{aligned}&\frac{\partial}{\partial t}f(x,t)\\ &=\Omega(x)f(x,t)+\int_0^t e^{\omega_1(t-t')}A(x)f(x,t')dt',\end{aligned} \tag{38}$$

which is subject to the initial condition $f(x,0)=(1+\exp Kx)$.[6] The solution is easily found with the aid of the Laplace transform in t to be

$$\begin{aligned}f(x,t)=1&+6e^{Kx-K^3t}+15e^{2Kx-8K^3t}+10e^{3Kx}\\ &\times(e^{-9K^3t}+e^{-27K^3t})+15e^{4Kx-28K^3t}+6e^{5Kx-35K^3t}\\ &+e^{6Kx-36K^3t},\end{aligned}$$

which gives the three-soliton solution.[7]

The multisoliton solutions for $N=4,5\ldots$ will be obtained in a similar way. If N is not an integer, however, an approximation is necessary for the memory function. The solution in this case has been found numerically to include an oscillatory tail.[7] Such a solution may well be accounted for in the present formalism by a simple oscillatory memory function.[2] An approximate memory function is also necessary when some perturbation exists to the pure soliton equation (22a). If the perturbation is dissipative corresponding to actual experimental situations, it seems worthwhile to try a simple damping memory kernel.[1] The usefulness of the linear equation (2a) with these modeled memory functions will be discussed in a later paper.

IV. REMARKS

It should be mentioned that the dependent variable transformation (23) plays a crucial role in the analysis of the preceding section. It is thus important in the application of the present formalism to find an appropriate change of the dependent variable so that the memory function A becomes insensitive to approximation [we may even put $A\simeq 0$ in Eq. (2a) to obtain a good zeroth-order approximation]. We hope that such an appropriate transformation can be found in each nonlinear problem.[9]

Several authors[10,11] have recently used a similar approach based on the linear imbedding of Carleman,[12] although they are all concerned with space–independent cases discussed in the previous paper.[1] Their approach differs from ours in that they do not consider the renormalization and the coupling of a variable with the lower-order ones, and hence do not convert an infinite set of linear equations into a single closed equation like Eq. (1), which corresponds in statistical–mechanical systems to Mori's generalized Langevin equation.[5,13] Thus, our approach, if applied to the stochastic process treated in Ref. 11, gives the result identical to that[14] of the Mori formalism.

APPENDIX

Let us first prove $f_1(x,t)\equiv 0$ for the single-soliton case, where f_1 is defined by

$$f_1(x,t)=f_t-\alpha_1 f_x-\alpha_2 f_{xx}-\alpha_3 f_{xxx} \tag{A1}$$

with $\alpha_1=-K^2-K\alpha_2-K^2\alpha_3$. Since $f_1(x,0)=0$, we show that

$$\frac{\partial^n}{\partial t^n}f_1(x,0)=0,\quad n=1,2,\ldots, \tag{A2}$$

to complete the proof. A direct proof of Eq. (A2) from the KdV equation requires all the derivatives $\partial^n u(x,0)/\partial t^n$, which are difficult to calculate. A simple way is to calculate the time derivatives of f, instead of u, by making use of Hirota's bilinear equation[8]

$$\begin{aligned}&f\frac{\partial}{\partial x}(f_t+f_{xxx})\\ &-f_x(f_t+f_{xxx})+3(f_{xx}^2-f_x f_{xxx})=0.\end{aligned} \tag{A3}$$

We now verify by mathematical induction that

$$a_n(x)=\frac{\partial^n}{\partial t^n}f(x,0)=(-K^3)^n e^{Kx}. \tag{A4}$$

For $n=1$ Eq. (A4) follows from Eq. (A1) at $t=0$. Assume that Eq. (A4) holds for $n\leqslant k$. Then, differentiating Eq. (A3) k times with respect to t, and putting $t=0$, we obtain

$$a\frac{d}{dx}(a_{k+1}+a_k''')-a'(a_{k+1}+a_k''')=0, \tag{A5}$$

where $a=f(x,0)$ and the prime denotes differentiation with respect to x. Dividing both sides of this result by a^2, and then integrating from $-\infty$ to x, we have Eq. (A4) with

$n = k + 1$, and thus Eq. (A4) is verified. On substituting Eq. (A4) into

$$\frac{\partial^n}{\partial t^n} f_1(x,0) = a_{n+1} - \alpha_1 a_n' - \alpha_2 a_n'' - \alpha_3 a_n''',$$

we obtain Eq. (A2).

In the case of the two-soliton solution, on the other hand, f_1 is defined by

$$f_1(x,t) = f_t - 6K^2 f_x + 9K f_{xx} - 2f_{xxx}.$$

The left-hand side of Eq. (A5) is now equal to

$$(-K^3)^k \cdot 54 \cdot 9^k K^4 (e^{3Kx} + 2e^{4Kx} + e^{5Kx}),$$

and we find in a way similar to that in Eq. (A4)

$$a_n(x) = (-K^3)^n (3e^{Kx} + 3 \cdot 8^n e^{2Kx} + 9^n e^{3Kx}) \qquad \text{(A6)}$$

for $n = 1,2,\ldots$. Hence, we have $\partial^n f_1(x,0)/\partial t^n = 0$, which implies $f_1(x,t) \equiv 0$.

The identity (36) for

$$f_2(x,t) = f_{tt} - \Omega(x) f_t - A(x) f - \omega_1 (f_t - \Omega f) \qquad \text{(A7)}$$

can be proved similarly by combining

$$a_n(x) = (-K^3)^n [6e^{Kx} + 8^n \cdot 15 e^{2Kx} + (27^n + 9^n) \cdot 10 e^{3Kx} + 28^n \cdot 15 e^{4Kx} + 35^n \cdot 6 e^{5Kx} + 36^n e^{6Kx}], \quad n \geqslant 1, \qquad \text{(A8)}$$

which is found by mathematical induction, with Eq. (A7) to obtain

$$\frac{\partial^n}{\partial t^n} f_2(x,0) = 0, \quad n \geqslant 1.$$

It should be mentioned that in this simple illustrative example Eqs. (A4), (A6), and (A8) directly give the solutions $f(x,t)$. In actual problems, all the higher-order coefficients a_n are unnecessary (and difficult to calculate); we need only a few lower-order ones to obtain the $a_m^{(\nu)}(\mathbf{r})$'s in Eqs. (16a)–(16c).

[1] T. Nishigori, J. Math. Phys. **22**, 2903 (1981).
[2] T. Nishigori, Physica A (to appear).
[3] We recall here the successful memory function approach to classical liquids, in which a simple exponential memory function leads to excellent results for the dynamical structure factor. See, e.g., T. Nishigori, J. Stat. Phys. **20**, 83 (1979).
[4] A. Z. Akcasu and J. J. Duderstadt, Phys. Rev. **188**, 479 (1969); J. L. Lebowitz, J. K. Percus, and J. Sykes, *ibid.* **188**, 487 (1969).
[5] H. Mori, Prog. Theor. Phys. **33**, 423 (1965).
[6] The projection operator involves an inverse matrix $\langle \mathbf{ff} \rangle^{-1}$, where the vector $\mathbf{f}$ consists of dynamical variables f_i and $\langle \ldots \rangle$ denotes a thermal average. In the present deterministic systems, the average is not taken,[1] and since the matrix $\mathbf{ff} = (f_i f_j)$ is obviously singular, the projection opertor cannot be defined.
[7] G. L. Lamb, Jr., *Elements of Soliton Theory* (Wiley, New York, 1980).
[8] R. Hirota, Phys. Rev. Lett. **27**, 1192 (1971); R. Hirota and J. Satsuma, Prog. Theor. Phys. Suppl. No. 59, 64 (1976).
[9] D. J. Kaup, Physica D **1**, 391 (1980).
[10] W. H. Steeb and F. Wilhelm, J. Math. Anal. Appl. **77**, 601 (1980); L. Brenig and V. Fairén, J. Math. Phys. **22**, 649 (1981); R. F. S. Andrade and A. Rauh, Phys. Lett. A **82**, 276 (1981).
[11] R. Graham and A. Schenzle, Phys. Rev. A **25**, 1731 (1982).
[12] T. Carleman, Acta Math. **59**, 63 (1932); E. W. Montroll, AIP Conf. Proc. No. **46**, 337 (1978).
[13] T. Nishigori, Physica A **100**, 266 (1980).
[14] H. Fujisaka and S. Grossmann, Z. Phys. B **43**, 69 (1981).

Discontinuous solutions for first-order systems through a limiting process

Pierre Gravel
Département de mathématiques, Collège militaire royal de Saint-Jean, Saint-Jean, Québec, J0J IR0, Canada

(Received 1 March 1982; accepted for publication 18 June 1982)

The possibility for physically general quasilinear differential systems to have discontinuous solutions or solutions with discontinuous derivatives is investigated using the method of asymptotic regularizations.

PACS numbers: 02.30.Jr, 02.90. + p

1. INTRODUCTION

It has been shown (Ref. 1) that asymptotic regularizations may be used to obtain conditions for infinitesimal shocks in a general relativistic context. A nontrivial example of this has been given in Ref. 2. We show here how the method can be used to obtain discontinuity conditions for a "physically general" quasilinear first-order differential system that is, a first-order system equivalent to a fourth (at most) order system. The similarity with Choquet–Bruhat's work (Ref. 3) will be noted although, as indicated in Ref. 1, it is only fortuitous. We apply the theory to a general system representing a single fourth-order equation for one unknown in two variables.

2. A LOOK AT THE SYSTEMS STUDIED

Let Ω be a neighborhood of an l-dimensional manifold V_l, coordinatized by the real-valued functions x^λ ($\lambda = 0,\ldots,\lambda - 1$). All Greek indices will have this domain of definition.

On this neighborhood we will try to find the conditions to be satisfied by a set of functions u^i ($i = 1,\ldots,N$ as for all other Latin indices) in order for them to be solutions of the first-order quasilinear system

$$L^j(u) \equiv a_i^{j\lambda}(x,u)\frac{\partial u^i}{\partial x^\lambda} + b^j(x,u) = 0, \tag{1}$$

having discontinuities across a hypersurface Σ defined by the equation $\phi(x) = 0$, ϕ a regular function of the coordinates.

It is known that these general systems of first order may be used in lieu of arbitrary order quasilinear systems simply by introducing supplementary dependent variables (Ref. 4). For example, the wave equation in a single dimension

$$\frac{\partial^2 y}{\partial t^2} - C^2\frac{\partial^2 y}{\partial x^2} = 0$$

can be put into the first-order form

$$z = \partial y/\partial t,$$
$$w = \partial y/\partial x,$$
$$\frac{\partial z}{\partial t} - C^2\frac{\partial w}{\partial x} = 0.$$

In that case, we say that $y(t,x)$ is the fundamental (dependent) variable and $z(t,x)$, $w(t,x)$ the supplementary (dependent) variables.

This way of putting things has a radical impact on the way asymptotic regularizations will have to be defined for solutions with discontinuities for systems of the type of Eq. (1) replacing higher-order systems.

Indeed, let us assume that all the regularizations have the form

$$u^i = \sum_{q=0}^{\infty} \omega^{-q} u_q^i(x,\omega\phi) \tag{2}$$

and that the fundamental dependent variables are labeled by the first p indices. Then if the original higher-order system generating (1) is of order higher than 2, there will automatically be an index value $i > p$ such that

$$u^i = \sum_{q=-L}^{\infty} \omega^{-q} v_q^i(x,\omega\phi), \tag{3}$$

L a positive integer.

For example, if the uniform beam equation

$$\frac{\partial^2 u^1}{\partial t^2} - K\frac{\partial^4 u^1}{\partial x^4} = 0 \tag{4}$$

is studied, the corresponding first-order system could be

$$u^2 = \frac{\partial u^1}{\partial t},$$
$$u^3 = \frac{\partial u^1}{\partial x},$$
$$u^4 = \frac{\partial u^3}{\partial x},$$
$$u^5 = \frac{\partial u^4}{\partial x},$$
$$\frac{\partial u^2}{\partial t} - K\frac{\partial u^5}{\partial x} = 0,$$

(although this is not strictly equivalent).

Since both u^1 and u^2 have the form (2) we have

$$\sum_{q=0}^{\infty}\omega^{-q}u_q^2(t,x,\omega\phi) = u^2 = \frac{\partial u^1}{\partial x}$$
$$= \frac{\partial}{\partial x}\sum_{q=0}^{\infty}\omega^{-q}u_q^1(t,x,\omega\phi)$$
$$= \omega u_0^{1'}\phi_{,x} + \sum_{q=0}^{\infty}\omega^{-q}(u_{q,x}^1 + u_{q+1}^1\phi_{,x}). \tag{5}$$

In this example, $L = 1$ and

$$v_q = u_{q,x}^1 + u_{q+1}^1\phi_{,x} \quad \text{for } q \geqslant 0,$$
$$v_{-1} = u_0^1\phi_{,x}.$$

Equations of type (2) are inadmissible in the most general case since they force the v_q^1's, $q < 0$, to be zero, which has nothing to do with the original higher-order equations. This

 0022-2488/82/112053-06$02.50 © 1982 American Institute of Physics

is illustrated in the beam system by the fact that $\omega u_0^{1'}\phi, x$ (i.e., ωv_{-1}) must vanish, that is,

$$u_0^{1'} = 0.$$

This cannot be acceptable since introducing

$$u^1 = \sum_{q=0}^{\infty} \omega^{-q} u_q^1(t, x, \omega\phi)$$

directly in (4) cannot yield this result.

Since we cannot avoid the appearance of expressions of type (3), we must replace (2) by a more general formula

$$u^i = \sum_{q=-L}^{\infty} \omega^{-q} u_q^i(x, \omega\phi), \tag{6}$$

where L is a positive integer (in each particular problem the unused u_q^i are considered to be identically zero).

Once this change is made one realizes that it will not be possible to expand the $a_i^{J\lambda}$ and b^J as infinite integral series around a given $\bar{u}_0^J$ as this is done in Ref. 4. Expressions such as

$$a_i^{J\lambda}(x,u) = a_i^{J\lambda}(x,\bar{u}_0) + a_{ih}^{J\lambda}(u^h - \bar{u}_0^h) + \tfrac{1}{2}a_{ihk}^{J\lambda}(u^h - \bar{u}_0^h)(u^k - \bar{u}_0^k) + \ldots \tag{7}$$

$$b^j(x,u) = b^J(x,\bar{u}_0) + b^J{}_h(u^h - \bar{u}_0^h) + \tfrac{1}{2}b^J{}_{hk}(u^h - \bar{u}_0^h)(u^k - \bar{u}_0^k), \tag{8}$$

where

$$a_{ih}^{J\lambda} = \frac{\partial a_i^{J\lambda}}{\partial u^h}(x,\bar{u}_0) \tag{9}$$

will not be used in the general case since the terms nonlinear in $(u^h - \bar{u}_0^h)$, if these series are not all terminated after a common finite number of terms, introduce in (1) an infinity of expressions starting with ω^P's, the $P'x$ being various positive integers. Since we will later use a limiting process as ω goes to infinity this should not be allowed to be since from (1) one would obtain an infinity of equations from these ω^P terms which would all have to be satisfied independently at the order at which $\bar{u}_0$ is approximated.

On the other hand, truncating (7) and (8) to finite polynomials in $(u^h - \bar{u}_0^h)$ from the outset excludes interesting cases such as the one developed in Ref. 2. This leads us to the following classification: We shall group the u^h's in such a way that those having a development (6) which do not contain positive powers in ω have suffixes h in the range 1–M. Equations (7) and (8) for $a^{J\lambda}$'s and b^J's following these lines are then those for which all the $a_{ihk\ldots n}^{J\lambda}$ and $a_{hk\ldots n}^{J\lambda}$ contain a common maximum set of indices $hk \ldots n$ larger than M.

For example, if we suppose that this number is 2, then the maximum number of indices of the group $(hk\ldots n)$ which will be greater than M in $a_{ihk\ldots n}^{i\lambda}$ will be 2. Since these $a_{ihk\ldots n}^{J\lambda}$ are multiplied by $(u^h - \bar{u}_0^h)$ terms in (7), we then have only two of these for which the representative sum starts at $2L$ instead of ω^L. Hence each is a series starting at most at ω^{2L}. The complete $a_i^{J\lambda}$ has thus as maximum power of ω the integer $2L$, and consequently we do not have an infinity of equations to satisfy [the same is true of the b_J's hence for the system (1) as a whole].

3. TRACTABLE SETS OF EQUATIONS

It would have been desirable, after these changes to have (1) written down in expanded form with respect to powers of ω. However one can easily realize the fact that this is a practical impossibility for the general case, where one finds indices $h, k, \ldots, n$ large than M in all terms of (7) and (8). Indeed, in that case one could not obtain the full form of the coefficients of the positive powers of ω, these being then the sum of an infinity of different terms coming each from a $(u^h - \bar{u}_0^h)(u^k - \bar{u}_0^k)\ldots(u^n - \bar{u}_0^n)$ product.

We are thus led to consider only particular series, more precisely those which do not contain indices $h, k, \ldots, n$ larger than M soon past the first few terms of (7) and (8), this for low values of L. We will be guided in this unavoidable choice by the equations of physics which are, as far as we know, of order not greater than four. The fundamental variables u^h being such that $h \in \{1, \ldots, M\}$, one then obtains $L = 3$, the third derivatives starting with an ω^{+3} at most.

As for the position of the last terms in (7) and (8), the choice is even more arbitrary. Considerations of systems such as the one given in Ref. 2 leads us to limit ourselves to $a_i^{J\lambda}$'s which are polynomials of degree 2 in the u^J's, $J \leqslant M$ and degree 1 in the u^J's, $J > M$. The b^J will be restricted to degree 2 with respect to the u^J's, $J > M$.

Thus we will have

$$u^J = \sum_{q=-3}^{\infty} \omega^{-q} u_q^1(x, \omega\phi), \tag{10}$$

and for $a_J^{J\lambda}$ and b^J

$$a_i^{J\lambda}(x,u) = a_i^{J\lambda}(x,\bar{u}_0) + a_{ih}^{J\lambda}(u^h - \bar{u}_0^h) + \tfrac{1}{2}a_{ihk}^{J\lambda}(u^h - \bar{u}_0^h)(u^k - \bar{u}_0^k) \tag{11}$$

$$b^J(x,u) = b^J(x,\bar{u}_0) + b^J{}_h(u^h - \bar{u}_0^h) + \tfrac{1}{2}b^J{}_{hk}(u^h - \bar{u}_0^h)(u^k - \bar{u}_0^k) + \tfrac{1}{6}b^J{}_{hkm}(u^h - \bar{u}_0^h)(u^k - \bar{u}_0^k)(u^m - \bar{u}_0^m) + \ldots \tag{12}$$

where the undescored indices are less than or equal to M and $\bar{u}_0^n$ is independent of $\omega\phi$.

4. EXPANSION OF THE EQUATIONS

Substituting expressions (11) and (12) into (1), grouping together the terms in the various powers of ω and limiting ourselves to those of order one at most with respect to the norm $\| \|_2$ defined in Ref. 1, we get the following equation in which

$$\bar{\bar{u}}_0^i \equiv u_0^i - \bar{u}_0^i: \tag{13}$$

$$\omega^7\{a_{ih}^{J\lambda}u^h{}_{-3}u'^i{}_3 l_\lambda\} + \omega^6\{a_{ih}^{J\lambda}(u^h{}_{-3}(u^i{}_{-3,\lambda} + u'^i{}_{-2}l_\lambda) + u^h{}_{-2}u'^i{}_{-3}l_\lambda) + \tfrac{1}{2}b^J{}_{hk}u^h{}_{-3}u^k{}_{-3}\}$$

$$+ \omega^5\{a_{ih}^J(u^h{}_{-3}(u^i{}_{-2,\lambda} + u'^i{}_{-1}l_\lambda) + u^h{}_{-2}(u^i{}_{-3,\lambda} + u'^i{}_{-2}l_\lambda) + u^h{}_{-1}u'^i{}_{-3}l_\lambda) + \tfrac{1}{2}b^J{}_{hk}(u^h{}_{-3}u^k{}_{-2} + u^h{}_{-2}u^k{}_{-3})\}$$

$$+ \omega^4\{a_{ih}^{J\lambda}(u^h{}_{-3}(u^i{}_{-1,\lambda} + u'^i{}_0 l_\lambda) + u^h{}_{-2}(u^i{}_{-2,\lambda} + u'^i{}_{-1}l_\lambda) + u^h{}_{-1}(u^i{}_{-3,\lambda} + u'^i{}_{-2}l_\lambda))$$

$$+ u'^{i}_{-3} l_\lambda (a^{J\lambda}_{i0} + a^{J\lambda}_{ih} \bar{u}^h_0 + \tfrac{1}{2} a^{J\lambda}_{ihk} \bar{u}^h_0 \bar{u}^k_0)$$
$$+ \tfrac{1}{2} b^J_{hk} (u^h_{-3} u^k_{-1} + u^h_{-1} u^k_{-3} + u^h_{-2} u^k_{-2})\}$$

$$+ \omega^3 \{ a^{J\lambda}_{ih} (u^h_{-3} (u^i_{0,\lambda} + u'^{i}_{1} l_\lambda) + u^h_{-2} (u^i_{-1,\lambda}$$
$$+ u'^{i}_{0} l_\lambda) + u^h_{-1} (u^i_{-2,\lambda} + u'^{i}_{-1} l_\lambda))$$
$$+ (u^i_{-3,\lambda} + u'^{i}_{-2} l_\lambda)(a^{J\lambda}_{i0} + a^{J\lambda}_{ih} \bar{u}^h_0$$
$$+ \tfrac{1}{2} a^{J\lambda}_{ihk} \bar{u}^h_0 \bar{u}^k_0) + b^J_h u^h_{-3}$$
$$+ u'^{i}_{-3} l_\lambda (a^{J\lambda}_{ih} u^h_1 + \tfrac{1}{2} a^{J\lambda}_{ihk} (u^h_1 \bar{u}^k_0 + u^k_1 \bar{u}^h_0))$$
$$+ \tfrac{1}{2} b^J_{hk} (u^h_{-3} \bar{u}^k_0 + u^k_{-3} \bar{u}^h_0 + u^h_{-2} u^k_{-1} + u^h_{-1} u^k_{-2})\}$$

$$+ \omega^2 \{ a^{J\lambda}_{ih} (u^h_{-3} (u^i_{1,\lambda} + u'^{i}_{2} l_\lambda) + u^h_{-2} (u^i_{0,\lambda} + u'^{i}_{1} l_\lambda)$$
$$+ u^h_{-1} (u^i_{-1,\lambda} + u'^{i}_{0} l_\lambda))$$
$$+ (u^i_{-2,\lambda} + u'^{i}_{-1} l_\lambda)(a^{J\lambda}_{i0} + a^{J\lambda}_{ih} \bar{u}^h_0$$
$$+ \tfrac{1}{2} a^{J\lambda}_{ihk} \bar{u}^h_0 \bar{u}^k_0) + b^J_h u^h_{-2}$$
$$+ (u^i_{-3,\lambda} + u'^{i}_{-2} l_\lambda)(a^{J\lambda}_{ih} u^h_1 + \tfrac{1}{2} a^{J\lambda}_{ihk} (u^h_1 \bar{u}^k_0 + u^k_1 \bar{u}^h_0))$$
$$+ u'^{i}_{-3} l_\lambda (a^{J\lambda}_{ih} u^h_2 + \tfrac{1}{2} a^{J\lambda}_{ihk} (u^h_2 \bar{u}^k_0 + u^k_2 \bar{u}^h_0 + u^h_1 u^k_1))$$
$$+ \tfrac{1}{2} b^J_{hk} (u^h_{-3} u^k_1 + u^h_1 u^k_{-3} + u^h_{-2} \bar{u}^k_0$$
$$+ u^k_{-2} \bar{u}^h_0 + u^h_{-1} u^k_{-1})\}$$

$$+ \omega^1 \{ a^{J\lambda}_{ih} (u^h_{-3} (u^i_{2,\lambda} + u'^{i}_{3} l_\lambda) + u^h_{-2} (u^i_{1,\lambda}$$
$$+ u'^{i}_{2} l_\lambda) + u^h_{-1} (u^i_{0,\lambda} + u'^{i}_{1} l_\lambda))$$
$$+ (u^i_{-1,\lambda} + \bar{u}'^{i}_{0} l_\lambda)(a^{J\lambda}_{i0} + a^{J\lambda}_{ih} \bar{u}^h_0 + \tfrac{1}{2} a^{J\lambda}_{ihk} \bar{u}^h_0 \bar{u}^k_0) + b^J_h u^h_{-1}$$
$$+ (u^i_{-2,\lambda} + u'^{i}_{-1} l_\lambda)(a^{J\lambda}_{ih} u^h_1 + \tfrac{1}{2} a^{J\lambda}_{ihk} (u^h_1 \bar{u}^k_0 + u^k_1 \bar{u}^h_0))$$
$$+ (u^i_{-3,\lambda} + u'^{i}_{-2} l_\lambda)(a^{J\lambda}_{ih} u^h_2 + \tfrac{1}{2} a^{J\lambda}_{ihk} (u^h_2 \bar{u}^k_0$$
$$+ u^k_2 \bar{u}^h_0 + u^h_1 u^k_1))$$
$$+ u'^{i}_{-3} l_\lambda (a^{J\lambda}_{ih} u^h_3 + \tfrac{1}{2} a^{J\lambda}_{ihk} (u^h_3 \bar{u}^k_0$$
$$+ u^k_3 \bar{u}^h_0 + u^h_1 u^k_2 + u^h_2 u^k_1))$$
$$+ \tfrac{1}{2} b^J_{hk} (u^h_{-3} u^k_2 + u^h_2 u^k_{-3} + u^h_{-2} u^k_1$$
$$+ u^h_1 u^k_{-2} + u^h_{-1} \bar{u}^k_0 + u^k_{-1} \bar{u}^h_0)\}$$

$$+ \omega^0 \{ a^{J\lambda}_{ih} (u^h_{-3} (u^i_{3,\lambda} + u'^{i}_{4} l_\lambda) + u^h_{-2} (u^i_{2,\lambda}$$
$$+ u'^{i}_{3} l_\lambda) + u^h_{-1} (u^i_{1,\lambda} + u'^{i}_{2} l_\lambda))$$
$$+ (u^i_{0,\lambda} + u'^{i}_{1} l_\lambda)(a^{J\lambda}_{i0} + a^{J\lambda}_{ih} \bar{u}^h_0 + \tfrac{1}{2} a^{J\lambda}_{ihk} \bar{u}^h_0 \bar{u}^k_0)$$
$$+ (u^i_{-1,\lambda} + \bar{u}'^{i}_{0} l_\lambda)(a^{J\lambda}_{ih} u^h_1 + \tfrac{1}{2} a^{J\lambda}_{ihk} (u^h_1 \bar{u}^k_0 + u^k_1 \bar{u}^h_0))$$
$$+ (u^i_{-2,\lambda} + u'^{i}_{-1} l_\lambda)(a^{J\lambda}_{ih} u^h_2$$
$$+ \tfrac{1}{2} a^{J\lambda}_{ihk} (u^h_2 u^k_0 + \bar{u}^k_2 \bar{u}^h_0 + u^h_1 u^k_1))$$
$$+ (u^i_{-3,\lambda} + u'^{i}_{-2} l_\lambda)(a^{J\lambda}_{ih} u^h_3 + \tfrac{1}{2} a^{J\lambda}_{ihk} (u^h_3 \bar{u}^k_0$$
$$+ u^k_3 \bar{u}^h_0 + u^h_1 u^k_3 + u^h_2 u^k_1))$$
$$+ u'^{i}_{-3} l_\lambda (a^{J\lambda}_{ih} u^h_4 + \tfrac{1}{2} a^{J\lambda}_{ihk} (u^h_4 \bar{u}^k_0 + u^k_4 \bar{u}^h_0$$
$$+ u^h_1 u^k_3 + u^h_3 u^k_1 + u^h_2 u^k_2))$$
$$+ \tfrac{1}{2} b^J_{hk} (u^h_{-3} u^k_3 + u^h_3 u^k_{-3} + u^h_{-2} u^k_2$$
$$+ u^h_2 u^k_{-2} + u^h_{-1} u^k_1 + u^h_1 u^k_{-1} + \bar{u}^h_0 \bar{u}^k_0)$$
$$+ b^J_0 + b^J_h \bar{u}^h_0 + b^J_{hk} \bar{u}^j_0 \bar{u}^k_0 + b^J_{hkm} \bar{u}^h_0 \bar{u}^k_0 \bar{u}^m_0$$
$$+ b^J_{hkmn} \bar{u}^h_0 \bar{u}^k_0 \bar{u}^m_0 \bar{u}^n_0 + \ldots$$
$$O(\omega^{-2}) = 0. \tag{14}$$

Here we cannot assume, as is done in Ref. 4, that $\bar{u}^i_0 = 0$; this would exclude important cases such as the electromagnetic tensor studied in Ref. 2. where

$$F_{\alpha\beta} = G_{\alpha\beta}(x) + H_{\alpha\beta}(x, \omega\phi).$$

If one uses the norm $\| \ \|_2$ defined in Ref. 1 the terms of order n are the coefficients of ω^{n+1}, with the one exception that when u^i_0 is independent $\omega\phi$ (i.e., $\bar{u}^i_0 \equiv 0$ for all i) the expression

$$a^{J\lambda}_{i0} u^i_{0,\lambda} + b^J_0$$

in the coefficient of ω^0 remains part of this coefficient. In that case the full 0-order term is as follows:

$$a^{J\lambda}_{ih} (u^h_{-3} (u^i_{2,\lambda} + u'^{i}_{3} l_\lambda) + u^h_{-2} (u^i_{1,\lambda}$$
$$+ u'^{i}_{2} l_\lambda) + u^h_{-1} (\bar{u}^i_{0,\lambda} + u'^{i}_{1} l_\lambda)$$
$$+ u^h_1 (u^i_{-2,\lambda} + u'^{i}_{-1} l_\lambda) + u^h_2 (u^i_{-3,\lambda}$$
$$+ u'^{i}_{-2} l_\lambda) + u^h_3 u'^{i}_{-3} l_\lambda)$$
$$+ a^{J\lambda}_{i0} (u^i_{-1,\lambda} + \bar{u}^i_{0,\lambda}) + \tfrac{1}{2} a^{J\lambda}_{ihk} (u^h_1 u^k_1 (u^i_{-3,\lambda}$$
$$+ u'^{i}_{-2} l_\lambda) + u'^{i}_{-3} l_\lambda (u^h_1 u^k_2 + u^k_1 u^h_2))$$
$$+ b^J_h u^h_{-1} + \tfrac{1}{2} b^J_{hk} (u^h_{-3} u^k_2 + u^h_2 u^k_{-3}$$
$$+ u^h_{-2} u^k_1 + u^h_1 u^k_{-2}) + b^J_0.$$

5. THE DISCONTINUITIES AND THEIR REGULARIZATION

The combination of (10), (13), and the use of differential systems of the fourth order yield

$$u^i = \sum_{q=-3}^{\infty} \omega^{-q} u^i_q. \tag{15}$$

From Ref. 1 we already know the form we have to give to the regularizations of continuous functions having discontinuous first derivatives. We have

$$u^i = \bar{u}^i_0 + \frac{1}{\omega} u^i_1, \tag{16}$$

where

$$\lim_{\phi \to 0^+} u'^{i}_{1} l_\alpha = [\bar{u}^i_{0,\alpha}]$$

and

$$u^i_1 = 0 \quad \text{if} \quad \phi \leqslant 0.$$

(we use throughout the same notation as in Ref. 1).

One might think that this regularization could be generalized in a direct fashion to higher discontinuities.

For instance, one might believe that putting

$$u^i = \bar{u}^i_0 + \frac{1}{\omega^3} u^i_3, \tag{17}$$
$$u^i_3 = 0 \quad \text{if} \quad \phi \leqslant 0, \quad u^i_3 \in C^2(\Omega),$$

with

$$\lim_{\phi \to 0^+} u'''^i l_\alpha l_\beta l_r = [\bar{u}^i_{0,\alpha\beta r}] = b_{\alpha\beta} l_r$$

(Hadamard's conditions) (18)

would be the right thing to do. The derivatives

$$u^i{}_{,\alpha\beta r} = \bar{u}^i_{0,\alpha\beta r} + u'''^i_3 l_\alpha l_\beta l_r + O(\omega^{-1})$$

would then be continuous for any ω only if (18) held true. However this equation is equivalent to

$$\lim_{\phi \to 0^+} u'''^i_3 l_\alpha l_\beta = b_{\alpha\beta},$$

which would imply that the symmetric tensor $b_{\alpha\beta}$ defined on Σ and at first thought of as arbitrary should be a simple tensor from the start, built with the vector l_α. Hence, (17) and (18) must be discarded and another more suitable kind of regularization found for this type of discontinuity.

In order to achieve this, we will generalize the process of substitution of the fundamental dependent variables by nonfundamental ones that we did at the beginning of this work to functions having only higher-order discontinuous derivatives. Thus for a $C^2(\Omega)$ function $\bar{u}^i_0$ which is only $C^3(\Omega_+)$ and $C^3(\Omega_-)$, one will not have to construct directly an expression of type (17); the regularizing term will be introduced only in the new function representing the second derivative.

Indeed, if $\bar{u}^i_0$ is a fundamental dependent variable, we will put

$$u^i = u^i_0 = \bar{u}^i_0, \tag{19}$$

$$u^{\tilde{i}\tilde{\alpha}} = \bar{u}^{\tilde{i}\tilde{\alpha}}_0 = u^i_{0,\alpha}, \tag{20}$$

and

$$u^{\tilde{i}\tilde{\alpha}\tilde{\beta}} = \bar{u}^{\tilde{i}\tilde{\alpha}\tilde{\beta}}_0 + \frac{1}{\omega} u^{\tilde{i}\tilde{\alpha}\tilde{\beta}}_1 = \bar{u}^{\tilde{i}\tilde{\alpha}}_{0,\beta} + \frac{1}{\omega} u^{\tilde{i}\tilde{\alpha}\tilde{\beta}}_1, \tag{21}$$

where the tilded indices are equivalent to $J > M$ values for which the u^J's would be first and second derivatives of u^J_0. One gets from (21) by taking derivatives

$$u^{\tilde{i}\tilde{\alpha}\tilde{\beta}}_{,\lambda} = \bar{u}^{\tilde{i}\tilde{\alpha}\tilde{\beta}}_{0,\lambda} + u'^{\tilde{i}\tilde{\alpha}\tilde{\beta}}_1 l_\lambda + \frac{1}{\omega} u^{\tilde{i}\tilde{\alpha}\tilde{\beta}}_{1,\lambda}$$

$$= \bar{u}^i_{0,\alpha\beta\lambda} + u'^{\tilde{i}\tilde{\alpha}\tilde{\beta}}_1 l_\lambda + \frac{1}{\omega} u^{\tilde{i}\tilde{\alpha}\tilde{\beta}}_{1,\lambda}.$$

Thus if we want to regularize the third order derivatives of $\bar{u}^i_0$ appearing in the initial higher-order system, we can postulate that we have for the variables in the corresponding quasilinear system

$$\lim_{\phi \to 0^+} u'^{\tilde{i}\tilde{\alpha}\tilde{\beta}}_1 l_\lambda = [\bar{u}^i_{0,\alpha\beta\lambda}],$$

which is in fact equivalent to

$$\lim_{\phi \to 0^+} u'^{\tilde{i}\tilde{\alpha}\tilde{\beta}}_1 = b_{\alpha\beta}. \tag{22}$$

We also have

$$\lim_{\omega \to \infty} u^{\tilde{i}\tilde{\alpha}\tilde{\beta}} = \bar{u}^{i\alpha}_{0,\beta}.$$

We shall repeat this procedure for the derivatives of other orders, designating by the expression "discontinuous functions of order n" those functions for which only the derivatives of order no smaller than n can be discontinuous. This implies that all the underscored terms appearing in (14) vanish. One should note that the coefficients of powers of ω larger than 2 remain unchanged.

6. FOURTH ORDER DISCONTINUITIES

The simplest case is certainly that in which only derivatives of order 4 can be discontinuous. The required regularization will take the form

$$u^i = \bar{u}^i_0(x), \quad i \leqslant M,$$

$$u^{\tilde{i}\tilde{\alpha}} = \bar{u}^{\tilde{i}\tilde{\alpha}}_0(x) = \bar{u}^i_{0,\alpha},$$

$$u^{\tilde{i}\tilde{\alpha}\tilde{\beta}} = \bar{u}^{\tilde{i}\tilde{\alpha}\tilde{\beta}}_0(x) = \bar{u}^i_{0,\alpha\beta},$$

$$u^{\tilde{i}\tilde{\alpha}\tilde{\beta}\tilde{\lambda}} = u^{\tilde{i}\tilde{\alpha}\tilde{\beta}\tilde{\lambda}}_0(x) + \frac{1}{\omega} u^{\tilde{i}\tilde{\alpha}\tilde{\beta}\tilde{\lambda}}_1(x, \omega\phi)$$

$$= \bar{u}^i_{0,\alpha\beta\lambda} + \frac{1}{\omega} u^{\tilde{i}\tilde{\alpha}\tilde{\beta}\tilde{\lambda}}_1,$$

where

$$\lim_{\phi \to 0^+} u'^{\tilde{i}\tilde{\alpha}\tilde{\beta}\tilde{\lambda}}_1 l_\mu = [\bar{u}^i_{0,\alpha\beta\lambda\mu}] = b_{\alpha\beta\lambda} l_\mu$$

(Hadamard's conditions).

Equation (14) is then reduced to

$$a^{J\lambda}_{i0}(\bar{u}^i_{0,\lambda} + u'^i_1 l_\lambda) + b^J_0 + O(\omega^{-1}) = 0.$$

Once the norm is taken, the orders 0 and 1 yield

$$a^{J\lambda}_{i0} \bar{u}^i_{0,\lambda} + b^J_0 = 0 \tag{23}$$

and

$$a^{J\lambda}_{i0} u'^i_1 l_\lambda = 0.$$

The first equation just says that $\bar{u}^i_0$ is the discontinuous solution we are looking for. The second implies that

$$\det(a^{J\lambda}_{i0} l_\lambda) = 0 \tag{24}$$

if we have nonvanishing discontinuities; it is the well-known characteristic hypersurface equation.

7. THIRD ORDER DISCONTINUITIES

In this case we use regularizations

$$u^i = \bar{u}^i_0(x), \quad i \leqslant M,$$

$$u^{\tilde{i}\tilde{\alpha}} = \bar{u}^{\tilde{i}\tilde{\alpha}}_0(x) = \bar{u}^i_{0,\alpha}(x),$$

$$u^{\tilde{i}\tilde{\alpha}\tilde{\beta}} = \bar{u}^{\tilde{i}\tilde{\alpha}\tilde{\beta}}_0(x) + \frac{1}{\omega} u^{\tilde{i}\tilde{\alpha}\tilde{\beta}}_1(x, \omega\phi)$$

$$= \bar{u}^i_{0,\alpha\beta}(x) + \frac{1}{\omega} u^{\tilde{i}\tilde{\alpha}\tilde{\beta}}_1(x, \omega\phi),$$

$$u^{\tilde{i}\tilde{\alpha}\tilde{\beta}\tilde{\lambda}} = \bar{u}^{\tilde{i}\tilde{\alpha}\tilde{\beta}\tilde{\lambda}}_0(x) + \bar{\bar{u}}^{\tilde{i}\tilde{\alpha}\tilde{\beta}\tilde{\lambda}}_0(x, \omega\phi) + \frac{1}{\omega} u^{\tilde{i}\tilde{\alpha}\tilde{\beta}\tilde{\lambda}}_1$$

$$= \bar{u}^i_{0,\alpha\beta\lambda} + u'^{\tilde{i}\tilde{\alpha}\tilde{\beta}}_1 l_\lambda + \frac{1}{\omega} u^{\tilde{i}\tilde{\alpha}\tilde{\beta}}_{1,\lambda},$$

where

$$u^{\tilde{i}\tilde{\alpha}\tilde{\beta}}_1 = 0 \quad \text{if in } \Omega_-$$

$$\lim_{\phi \to 0^+} u'^{\tilde{i}\tilde{\alpha}\tilde{\beta}}_1 = d^1_{i\alpha\beta}, \quad \lim_{\phi \to 0^+} u''^{\tilde{i}\tilde{\alpha}\tilde{\beta}\tilde{\lambda}}_1 = \frac{1}{\omega} d^2_{i\alpha\beta}$$

in which

$$d^{\,1}_{i\bar{\alpha}\bar{\beta}} l_\lambda = [\bar{u}^i_{0,\alpha\beta\lambda}], \quad d^{\,2}_{i\bar{\alpha}\bar{\beta}} l_\lambda l_\delta = [\bar{u}^i_{0,\alpha\beta\lambda\delta}].$$

From these, (14) yields to first order after taking the norm and averaging (cf. Ref. 2)

$$\bar{u}'^i_0 l_\lambda (a^{J\lambda}_{i0} + a^{J\lambda}_{ih}\bar{u}^h_0) = 0 \tag{25}$$

$$\bar{u}^i_{0,\lambda} a^{J\lambda}_{i0} + b^J_0 = 0 \tag{26}$$

$$(\bar{u}^i_{0,\lambda} + u'^i_1 l_\lambda)(a^{J\lambda}_{i0} + a^{J\lambda}_{ih}\bar{u}^h_0) + a^{J\lambda}_{ih}\bar{u}'^i_0 u^h_1 l_\lambda + b^J \bar{u}^h_0 + \frac{1}{2} b^J_{hk}\bar{u}^h_0\bar{u}^k_0 + \ldots = 0. \tag{27}$$

Apart from (26) which is a repetition of (23), Eq. (25) yields for effective discontinuities

$$\det(a^{J\lambda}_{i0} + a^{J\lambda}_{ik\bar{\alpha}\bar{\beta}\mu} \overset{1}{d}{}_{k\bar{\alpha}\bar{\beta}} l_\mu) l_\lambda = 0 \tag{28}$$

which indicates that the form of the discontinuity hypersurface is linked to the values of the discontinuity it bears. Using coordinates adapted to the hypersurface one gets (cf. Ref. 2)

$$\lim_{\phi \to 0^+} \det(a^{J0}_{i0} + a^{J0}_{ih}\bar{u}^h_0) = 0. \tag{29}$$

If the $a^{J\lambda}_i$ are assumed independent of the third-order derivatives of the u^i's, $i \leqslant M$, Eqs. (25) and (27) yield

$$a^{J\lambda}_{i0}\bar{u}^i_0 l_\lambda = 0 \tag{30}$$

and

$$a^{J\lambda}_{i0}(\bar{u}^i_{0,\lambda} + u'^i_1 l_\lambda) + a^{J\lambda}_{ih}\bar{u}'^i_0 u^j_1 l_\lambda + b^J_h \bar{u}^h_0 + \frac{1}{2} b^J_{hk}\bar{u}^h_0\bar{u}^k_0 + \ldots = 0, \tag{31}$$

which give after the $\phi \to 0^+$ limit

$$a^{J}_{i\bar{\alpha}\bar{\beta}} \overset{2}{d}{}_{i\bar{\alpha}\bar{\beta}} l_\lambda = 0$$

and

$$\left[\lim_{\phi \to 0^+} a^{J\lambda}_{i\bar{\alpha}\bar{\beta}\bar{\gamma}0}\right]\left[(\overset{1}{d}{}_{i\bar{\alpha}\bar{\beta}} l_\lambda)_{,\lambda} + \overset{1}{d}{}_{i\bar{\alpha}\bar{\beta}\bar{\gamma}} l_\lambda\right] + \lim_{\phi \to 0^+}\left[a^{J\lambda}_{i\bar{\alpha}\bar{\beta}0} b^J_{i\bar{\alpha}\bar{\beta}\lambda} + b^J_{i\bar{\alpha}\bar{\beta}\lambda\bar{J}\bar{\mu}\bar{\nu}\bar{\gamma}} \overset{1}{d}{}_{\bar{J}\bar{\mu}\bar{\nu}} l_\lambda\right] \overset{1}{d}{}_{i\bar{\alpha}\bar{\beta}} l_\lambda + \ldots = 0,$$

which are propagation equations on Σ for the discontinuities $\overset{1}{d}{}_{i\bar{\alpha}\bar{\beta}}$.

8. SECOND ORDER DISCONTINUITIES

We put

$$u^i = \bar{u}^i_0(x) \quad i \leqslant M,$$

$$u^{i\bar{\alpha}} = \bar{u}^{i\bar{\alpha}}_0(x) + \frac{1}{\omega} u^{i\bar{\alpha}}_1(x,\omega\phi) = \bar{u}^i_{0,\alpha}(x) + \frac{1}{\omega} u^{i\bar{\alpha}}_1(x,\omega\phi),$$

$$u^{i\bar{\alpha}\bar{\beta}} = \bar{u}^{i\bar{\alpha}\bar{\beta}}_0(x) + \bar{\bar{u}}^{i\bar{\alpha}\bar{\beta}}_0(x,\omega\phi) + \frac{1}{\omega} u^{i\bar{\alpha}\bar{\beta}}_1(x,\omega\phi) = \bar{u}^i_{0,\alpha\beta}(x) + u'^{i\bar{\alpha}}_1(x,\omega\phi) l_\beta + \frac{1}{\omega} u^{i\bar{\alpha}}_{1,\beta}(x,\omega\phi), \tag{32}$$

$$u^{i\bar{\alpha}\bar{\beta}\bar{\gamma}} = \omega u^{i\bar{\alpha}\bar{\beta}\bar{\gamma}}_{-1}(x,\omega\phi) + \bar{u}^{i\bar{\alpha}\bar{\beta}\bar{\gamma}}_0(x) + \bar{\bar{u}}^{i\bar{\alpha}\bar{\beta}\bar{\gamma}}_0(x,\omega\phi) + \frac{1}{\omega} u^{i\bar{\alpha}\bar{\beta}\bar{\gamma}}_1(x,\omega\phi) = \omega u''^{i\bar{\alpha}}_1 l_\beta l_\gamma + \bar{u}^i_{0,\alpha\beta\gamma} + (u'^{i\bar{\alpha}}_{1,\gamma} l_\beta + u'^{i\bar{\alpha}}_1 l_{\beta,\gamma} + u'^{i\bar{\alpha}}_{1,\beta} l_\gamma) + \frac{1}{\omega} u^{i\bar{\alpha}}_{1,\beta\gamma},$$

where $u^{i\bar{\alpha}}_1 = 0$ if $\phi < 0$, and

$$\lim_{\phi \to 0^+} u'^{i\bar{\alpha}}_1 = \overset{1}{d}{}_{i\bar{\alpha}} \text{ with } \overset{1}{d}{}_{i\bar{\alpha}} l_\gamma = [\bar{u}^i_{0,\alpha\gamma}],$$

$$\lim_{\phi \to 0^+} u''^{i\bar{\alpha}}_1 = \frac{1}{\omega} \overset{2}{d}{}_{i\bar{\alpha}} \text{ with } \overset{2}{d}{}_{i\bar{\alpha}} l_\beta l_\gamma = [\bar{u}^i_{0,\alpha\beta\gamma}]. \tag{33}$$

$$\lim_{0 \to 0^+} u'''^{i\bar{\alpha}}_1 = \frac{1}{\omega^2} \overset{3}{d}{}_{i\bar{\alpha}} \text{ with } \overset{3}{d}{}_{i\bar{\alpha}} l_\beta l_\gamma l_\lambda = [\bar{u}^i_{0,\alpha\beta\gamma\lambda}].$$

$$\begin{aligned}
&\omega^3\{a^{J\lambda}_{ih} u^h_{-1} u'^i_{-1} l_\lambda\} + \omega^{2-}\{a^{J\lambda}_{ih} u^h_{-1}(u^i_{-1,\lambda} + u'^i_0 l_\lambda) \\
&\quad + u'^i_{-1} l_\lambda (a^{J\lambda}_{i0} + a^{J\lambda}_{ih}\bar{u}^h_0) + \tfrac{1}{2} b^J_{hk} u^h_{-1} u^k_{-1}\} \\
&\quad + \omega^1\{a^{J\lambda}_{ih} u^h_{-1}(u^i_{0,\lambda} + u'^i_1 l_\lambda) + (u^i_{-1,\lambda} + u'^i_0 l_\lambda)(a^{J\lambda}_{i0} \\
&\quad + a^{J\lambda}_{ih}\bar{u}^h_0) + u'^i_{-1} l_\lambda a^{J\lambda}_{ih} u^h_1 + b^J_h u^h_{-1} \\
&\quad + \tfrac{1}{2} b^J_{hk}(u^h_{-1}\bar{u}^k_0 + u^k_{-1}\bar{u}^h_0)\} \\
&\quad + \omega^0\{a^{J\lambda}_{ih} u^h_{-1} u^i_{1,\lambda} + (u^i_{0,\lambda} + u'^i_1 l_\lambda)(a^{J\lambda}_{i0} + a^{J\lambda}_{ih}\bar{u}^h_0) \\
&\quad + (u^i_{-1,\lambda} + u'^i_0 l_\lambda) a^{J\lambda}_{ih} u^h_1 + b^J_0 \\
&\quad + b^J_h \bar{u}^h_0 + \tfrac{1}{2} b^J_{hk}(u^h_{-1} u^k_1 + u^h_1 u^k_{-1} \\
&\quad + \bar{u}^h_0\bar{u}^k_0) + \tfrac{1}{6} b^J_{hkn}\bar{u}^h_0\bar{u}^k_0\bar{u}^n_0 + \ldots\} \\
&\quad + O(\omega^{-2}) = 0.
\end{aligned} \tag{34}$$

Then, averaging and taking the norm and assuming the $a^{J\lambda}_i$ and b^J do not depend on derivatives of order 2 and 3 of the u^i, $i \leqslant M$, all yield

$$a^{J\lambda}_{i0} u'^1_{-1} l_\lambda = 0, \tag{35}$$

$$a^{J\lambda}_{i0}(u^i_{1,\lambda} + \bar{u}'^i_0 l_\lambda) + a^{J\lambda}_{ih} u^h_2 u'^i_{-1} l_\lambda = 0, \tag{36}$$

$$a^{J\lambda}_{i0}\bar{u}^i_{0,\lambda} + b^J_0 = 0, \tag{37}$$

$$a^{J\lambda}_{i0} u'^i_1 l_\lambda + a^{J\lambda}_{ih} u^h_1 (u^i_{-1,\lambda} + \bar{u}'^i_0 l_\lambda) = 0. \tag{38}$$

Equation (37) is not new to us, and neither is (35) [cf. (30)]. On the other hand, both (36) and (38) can be simplified to

$$a^{J\lambda}_{i0}(u^i_{-1,\lambda} + \bar{u}'^i_0 l_\lambda) = 0 \tag{37'}$$

and

$$a^{J\lambda}_{i0} u'^i_1 l_\lambda = 0. \tag{38'}$$

This last result is necessary since it could happen that the third derivatives be definable as continuous across Σ even if the second derivatives are not; since we are looking for conditions on these, Eq. (37') alone would be insufficient to obtain the characteristic equation when

$$\lim_{\phi \to 0^+} u'^i_{-1} = 0.$$

9. AN EXAMPLE

The more arduous cases where discontinuous first and zeroth order derivatives appear in an order-four system will be left aside due to their complexity, which implies results

the interpretation of which is not obvious unless *ad hoc* assumptions are made on the whole system. For example, in the case of order-0 discontinuities, the complete equation is (14) itself; obviously, without extra assumptions one cannot hope to find more than the order-6 condition

$$\det\left[a_{ih}^{J\lambda}u^{h}{}_{-3}l_{\lambda}\right]=0$$

for nonvanishing $\lim_{\phi\to 0^{+}} u'^{i}{}_{-3}$ discontinuities.

We shall illustrate in a later publication the procedure developed here to the case of Yang's third-order equations for the gravitational potentials $g_{\alpha\beta}$, which read

$$R_{\alpha\beta;\gamma}=R_{\alpha\gamma;\beta}+\text{Bianchi's identities.}$$

($R_{\alpha\beta}$ is the contracted Riemann tensor). These equations are an obvious generalization of Einstein's empty space set, which arise when one wants to interpret gravitation as due to a parallel displacement gauge field (Ref. 5).

As a somewhat less involved example, let us have a look at the general fourth-order equation in two variables x and y for one unknown function u:

$$F(x,y;u;p,q;r,s,t;a,b,c,d;e,f,g,h,i)=0, \tag{39}$$

where

$$p=u_{,1},\quad q=u_{,2};$$
$$r=u_{,11},\quad s=u_{,12},\quad t=u_{,22};$$
$$a=u_{,111},\quad b=u_{,112},\quad c=u_{,122},\quad d=u_{,222};$$
$$e=u_{,1111},\quad f=u_{,1112},\quad g=u_{,1122},$$
$$h=u_{,1222},\quad i=u_{,2222}.$$

(1 and 2 as suffixes to the right of a comma denote partial derivatives with respect to x and y respectively). Following Refs. 4 and 6 we can obtain a first-order canonical system equivalent to the fourth-order equation if we make the identifications

$$u_1=x,\quad u_2=y,\quad u_3=u;\quad u_4=p,\quad u_5=q;$$
$$u_6=r,\quad u_7=s,\quad u_8=t;$$
$$u_9=a,\quad u_{10}=b,\quad u_{11}=c,\quad u_{12}=d;\quad u_{13}=e,$$
$$u_{14}=f,\quad u_{15}=g,\quad u_{16}=h,\quad u_{17}=i.$$

and then put, assuming (39) can be formally solved for e in the form

$$e=G(x,y;u;p,q;r,s,t;a,b,c,d;f,g,h,i),$$
$$u_{1,1}=u_{2,2},\quad u_{4,1}=u_6u_{2,2},$$
$$u_{2,1}=0,\quad u_{7,1}=u_{6,2},$$
$$u_{3,1}=u_4u_{2,2},\quad u_{8,1}=u_{7,2},$$
$$u_{5,1}=u_{4,2},\quad u_{9,1}=u_{13}u_{2,2},$$
$$u_{6,1}=u_9u_{2,2},\quad u_{14,1}=u_{13,2},$$
$$u_{10,1}=u_{9,2},\quad u_{15,1}=u_{14,2},$$
$$u_{11,1}=u_{10,2},\quad u_{16,1}=u_{15,2},$$
$$u_{12,1}=u_{11,2},\quad u_{17,1}=u_{16,2},$$
$$u_{13,1}=(G_1+G_3u_4+G_4u_6+G_6u_9)u_{2,2}+G_5u_{4,2}+G_7u_{6,2}+G_8u_{7,2}+G_{14}u_{13,2}+G_{15}u_{14,2}+G_{16}u_{15,2}+G_{17}u_{16,2}$$

It can be seen at once that this form of the equivalent quasilinear system contains no b^J term. The $a_i^{J\lambda}$ terms are more easily written as a list than in matrix form. We have

$$1=a_1^{11}=a_2^{21}=a_3^{31}=a_5^{41}=a_4^{51}=a_7^{61}=a_8^{71}=a_6^{81}$$
$$=a_{10}^{91}=a_{11}^{10,1}=a_{12}^{11,1}=a_9^{12,1}=a_{14}^{13,1}$$
$$=a_{15}^{14,1}=a_{16}^{15,1}=a_{17}^{16,1}$$
$$-1=a_2^{12}=a_4^{42}=a_6^{62}=a_7^{72}=a_9^{92}=a_{10}^{10,2}=a_{11}^{11,2}$$
$$=a_{13}^{13,2}=a_{14}^{14,2}=a_{15}^{15,2}=a_{16}^{16,2}=a_{13}^{17,2}$$
$$a_2^{3,2}=-u_4,\quad a_2^{52}=-u_6,\quad a_2^{82}$$
$$=-u_9,\quad a_2^{12,2}=-u_{13}$$
$$a_{13}^{17,2}=G_{14},\quad a_{14}^{17,2}=G_{15},\quad a_{15}^{17,2}$$
$$=G_{16},\quad a_{16}^{17,2}=G_{17}$$
$$a_4^{17,2}=G_5,\quad a_6^{17,2}=G_7,\quad a_7^{17,2}=G_8$$
$$a_2^{17,2}=G_1+G_3u_4+G_4u_6+G_6u_9.$$

It is rather easy to obtain in this case the characteristic determinant for fourth-order discontinuities. It is

$$l_1^{13}\{l_1^4-G_{17_0}l_2^4-G_{16_0}l_1l_2^3-G_{15_0}l_1^2l_2^2-G_{14_0}l_1^3l_2\}=0.$$

[The subscript 0 has the same meaning here on the G's as it had in (24)].

As noted before, this condition is independent of the values the discontinuities may have along the characteristic "hypersurface" (a curve here). On the other hand, referring to (28) or (30), we know that the conditions for third order discontinuities depend both on the specific form of G *and* the values of the discontinuities; the corresponding equation is

$$l_1^{13}\{l_1^4-(G_{17_0}+G_{17_{(\alpha\beta\gamma)}}[\bar{u}_{3_{0,\alpha\beta\gamma}}])l_2^4-(G_{16_0}+G_{16_{(\alpha\beta\gamma)}}[\bar{u}_{3_{0,\alpha\beta\gamma}}])l_1l_2^3-(G_{15_0}+G_{15_{(\alpha\beta\gamma)}}[\bar{u}_{3_{0,\alpha\beta\gamma}}])l_1^2l_2^2-(G_{14_0}+G_{14_{(\alpha\beta\gamma)}}[\bar{u}_{3_{0,\alpha\beta\gamma}}])l_1^3l_2\}=0,$$

where the sets $(\alpha\beta\gamma)$ being summed refer, successively, to the functions a, b, c, and d and the square brackets denote as usual the discontinuities of the quantity they contain. These can be interpreted as in Sec. 8.

10. CONCLUSION

Although it may seem that the conditions imposed in Sec. 3 are somewhat restrictive, the example of Sec. 9 shows that part of the problem may be avoided if equivalent systems of the form used in the example are used (i.e., systems where the new variables nowhere appear in an nondifferentiated form).

[1] P. Gravel, "Asymptotic Regularizations as an Alternative to Distributions for the Study of Singular Hypersurfaces," J. Math. Phys. **21** (1), 21–3 (1980).

[2] P. Gravel, "Asymptotic Regularizations and Shocks in Non-Inductive Heat-Current-Free Relativistic Fluids," J. Gen. Ref. Grav. **12** (12), 973–87 (1980).

[3] Y. Choquet-Bruhat, "Ondes asymptotiques et approchées pour les systèmes d'équations aux dérivées partielles non-linéaires," J. Math. Phys. Appl. **48**, 117–158 (1969).

[4] R. Courant and D. Hilbert, *Methods of Mathematical Physics* (Interscience, New York, 1961), Vol. 2.

[5] C. N. Yang, "Integral Formalism for Gauge Fields," Phys. Rev. Lett. **33** (7), 445–7 (1974).

Laplace asymptotic expansions of conditional Wiener integrals and generalized Mehler kernel formulas

Ian Davies and Aubrey Truman
Mathematics Department, Heriot-Watt University, Riccarton, Currie, Edinburgh EH14 4AS, Scotland

(Received 23 December 1980; accepted for publication 21 August 1981)

Imitating Schilder's results for Wiener integrals rigorous Laplace asymptotic expansions are proven for conditional Wiener integrals. Applications are given for deriving generalized Mehler kernel formulas, up to arbitrarily high orders in powers of $\hbar$, for $\exp\{-TH(\hbar)/\hbar\}(x,y)$, $T>0$ where $H(\hbar)=[(-\hbar^2/2)\Delta_1+V]$, Δ_1 being the one-dimensional Laplacian, V being a real-valued potential $V\in C^\infty(\mathbb{R})$, bounded below, together with its second derivative.

PACS numbers: 02.30.Mv

1. INTRODUCTION

For $\hbar>0$, let $H(\hbar)$ denote the quantum mechanical Hamiltonian

$$H(\hbar)=\left\{\frac{-\hbar^2}{2}\frac{d^2}{dx^2}+\frac{A^2x^2}{2}\right\}\quad(A\in\mathbb{R}).$$

The Mehler formula for the kernel $\exp\{-tH(\hbar)/\hbar\}(x,y)$ states that

$$\exp\{-tH(\hbar)/\hbar\}(x,y)$$
$$=(A/2\pi\hbar\sinh At)^{1/2}$$
$$\times\exp\{(-A/2\hbar\sinh At)[(x^2+y^2)\cosh At-2xy]\},$$

$x,y\in\mathbb{R}$. It is well known that the validity of the Mehler formula is equivalent to the result that the WKB approximation is exact for harmonic oscillator potentials.

This suggests that for more general potentials V, e.g., $V(x)=A^2x^2/2+B^2x^4/4$ $(A,B\in\mathbb{R})$ one should be able to use quasiclassical expansions to determine the kernel $\exp\{-tH(\hbar)/\hbar\}(x,y)$, $x,y\in\mathbb{R}$, for $H(\hbar)=\{(-\hbar^2/2)(d^2/dx^2)+V(x)\}$, up to arbitrarily high orders in $\hbar$, in terms of the corresponding classical mechanical solutions. Formal results of this kind have been given by Mizrahi,[1] De-Witt,[2] and Truman,[3] but, to date, there does not seem to be any rigorous proof of the validity of these expansions.

For a rigorous proof we require a generalization of the Laplace asymptotic expansions for Lebesgue integrals to integrals over function space. The functional integral required in this context is the conditional Wiener integral. Here we base our main technical result Theorem 1 on a number of new estimates—Lemmas 5, 7, and 14 and the seminal work of Schilder,[4] who considered Laplace asymptotic expansions for the Wiener integral. We learned the method of proof of Lemma 5 of this paper from Simon's[5] excellent book on functional integration in quantum physics. (The underlying idea is the beautiful proof of Kolmogorov's lemma.) There are, however, several alternative ways of obtaining this estimate. Our results do not use the extensive work of Donsker and Varadhan in this area, because, for the conditional Wiener integral, Schilder's methods seem more obvious.

In Theorem 1, using Donsker's "flat" integral,[6] the functional integrand is assumed to have a unique global maximum, but we can allow there to be any finite number of global maxima. In this case Theorem 1 can be used to obtain Bender–Wu type formulae for the x^{2N}-anharmonic oscillator (see Ref. 7). By combining some of the ideas of Schilder with those of Pincus,[8] Theorem 1, with more restrictive assumptions, can be extended to a general Gaussian process, but more general results of this kind are now available in the work of Ellis and Rosen.[9]

The applications of our theorem to establishing a generalized Mehler kernel formula are given in Theorems 2 and 3 in Sec. 2 of this paper. In Theorem 2 general anharmonic oscillator potentials $V(x)\in C^\infty(\mathbb{R})$, which are bounded below and convex $[V''(x)>0]$, e.g., $V(x)=(A^2x^2/2+B^2x^4/4)$ are considered and a generalized Mehler kernel formula is given, up to arbitrarily high orders in $\hbar$, for arbitrary finite times. In Theorem 3 nonconvex potentials are studied and a generalized Mehler kernel formula is given but only for sufficiently small times. As will be seen these results and the Bender–Wu type results are quite easy to prove. The main reason for this is that the requisite conditions for Theorem 1 are expressed in terms of a supnorm and not the more restrictive L^p norms used in some treatments.

We state Theorem 1 and the attendant conditions below. The proof is deferred until Sec. 3. In what follows $C_0[0,T]$ is the Banach space of continuous functions $z:[0,T]\to\mathbb{R}$ with $z(0)=z(T)=0$ equipped with supnorm

$$\|z\|=\sup_{\tau\in[0,T]}|z(\tau)|.$$

$C_0[0,T]$ supports the conditional Wiener measure $d\mu_{0,0;0,T}$, with covariance

$$\int_{C_0[0,T]}z(s)z(t)d\mu_{0,0;0,T}(z)=(2\pi T)^{-1/2}s(1-t/T),\qquad 0\leqslant s<t\leqslant T,$$

with mean zero

$$\int_{C_0[0,T]}z(s)d\mu_{0,0;0,T}(z)=0,\quad 0\leqslant s\leqslant T.$$

For the associated probability measure

$$\mu^{-1}_{0,0;0,T}\{C_0[0,T]\}d\mu_{0,0;0,T}(z)=(2\pi T)^{1/2}d\mu_{0,0;0,T}(z),$$

we use the notation

$$(2\pi T)^{1/2}\int_{C_0[0,T]}F(z)d\mu_{0,0;0,T}(z)=\mathbb{E}_z^T\{F(z)\},$$

 0022-2488/82/112059-12$02.50 © 1982 American Institute of Physics

for suitable functionals F. Slightly abusing notation, for measurable sets A, we shall sometimes write

$$\mathbb{E}_z^T\{\chi_A(z)\} = \mathbb{E}_z^T\{A\},$$

where χ_A is the characteristic function of the set A.
$C_0^*[0,T]$ is the reproducing kernel Sobolev space associated with $C_0[0,T]$; $z \in C_0^*[0,T]$ if z is absolutely continuous with derivative $z'(\tau)$ in $L^2[0,T]$, $\int_0^T [z'(\tau)]^2\, d\tau < \infty$.

Theorem 1: Let $F(z)$ be a real-valued continuous functional defined on $C_0[0,T]$ and suppose that the functional $\{F(z) - 2^{-1}\int_0^T [z'(\tau)]^2\, d\tau\}$ has a unique maximum at $x_0 \in C_0^*[0,T]$ with $\{F(x_0) - 2^{-1}\int_0^T [x_0'(\tau)]^2\, d\tau\} = b$. If F satisfies the conditions 1–6 below, then

$$\exp\{-b\lambda^{-2}\}\mathbb{E}_z^T\{\exp\{\lambda^{-2}F(\lambda z)\}\}$$
$$= \Gamma_0 + \lambda\Gamma_1 + \lambda^2\Gamma_2 + \cdots + \lambda^{n-3}\Gamma_{n-3} + O(\lambda^{n-2}),$$

as $\lambda \to 0$, where the Γ_i are integrals dependent only on the functional F and its Frechet derivatives at x_0.

(1) $F(z)$ is measurable.

(2) $F(z) \leqslant (b + L_1) + L_2\|z\|^2$, $\mu_{0,0;0,T}$ almost everywhere. L_1, L_2 are positive real numbers with $L_2 < \min\{\gamma/2T, 1/4T\}$, γ being the constant in Lemma 5. (See also remark at the end of Theorem 3.)

(3) $F(z)$ is continuous for

$$\|z\| \leqslant \max\{(L_1+1)^{1/2}/|L_2 - 1/2T|^{1/2}, [2T(L_1+1)/\gamma]^{1/2}\}$$

and upper semicontinuous on $C_0[0,T]$.

(4) $F(z)$ has $n \geqslant 3$ continuous Frechet derivatives in a ball of radius δ centered at x_0 in $C_0[0,T]$, $\delta > 0$. We further assume that the Frechet derivatives $D^j F$ satisfy
$D^j F(x_0 + \eta)z^j = O(\|z\|^j)$, if $\|\eta\| < \delta$.

(5) For some $\epsilon > 0$, for $\|\eta\| < \delta$,
$\mathbb{E}_z^T\{\exp\{(1+\epsilon)D^2F(x_0+\eta)z^2/2\}\}$ is uniformly bounded.

(6) $x_0'(\cdot)$ is of bounded variation on $[0,T]$.

Since we can deduce the above result with $b \neq 0$ from the corresponding theorem with $b = 0$ by making the replacement $F(z) \to [F(z) - b]$, we only prove Theorem 1 for $b = 0$. (This incidentally enables us to avoid a *non sequitur* in Schilder's original paper.[10]) The proof of this version of the theorem is given in Sec. 3. As we have already remarked the seminal idea for this proof is Schilder's. Theorems 2 and 3 together with their proofs are given in Sec. 2 of this paper.

For convenience we shall use a notation consistent with that of Schilder's original paper. For instance, if $z \in C_0[0,T]$, $z^n(\cdot) \in C_0^*[0,T]$ denotes its polygonalization defined by

$$z^n(\tau) = z\left(\frac{jT}{n}\right) + \left(\tau - \frac{jT}{n}\right)\left[z\left(\frac{\overline{j+1}\,T}{n}\right) - z\left(\frac{jT}{n}\right)\right]\frac{n}{T},$$
$$jT/n \leqslant \tau \leqslant (j+1)T/n,$$

for $j = 0,1,2,\ldots,(n-1)$. The corresponding n-vector is denoted z^n, where $z^n = (z(T/n), z(2T/n), \ldots, z(T))$. As we have already stated $\|\ \|$ denotes the supnorm throughout so that

$$\|z^n(\cdot)\| = \sup_{\tau\in[0,T]} |z^n(\tau)| = \sup_{j=1,2,\ldots,(n-1)} \left|z\left(\frac{jT}{n}\right)\right| = \|z^n\|.$$

2. SOME APPLICATIONS OF THEOREM 1

We now discuss the kernel $\exp\{-TH(\hbar)/\hbar\}(x,y)$, for the quantum mechanical Hamiltonian $H(\hbar) = [(-\hbar^2/2)\Delta_1 + V]$, where the real-valued potential $V \in C^\infty(\mathbb{R})$ and is bounded below. We are aiming for a generalized Mehler formula for the kernel expressing it in terms of classical mechanical solutions up to arbitrarily high orders in $\hbar$. We begin with a standard result from the direct methods of the calculus of variations.

Lemma 1: Fix $x, y \in \mathbb{R}$ and $T > 0$. Let the real-valued potential $V \in C^\infty(\mathbb{R})$ be bounded below. Let the functional

$$A(z) = 2^{-1}\int_0^T [z'(\tau)]^2\, d\tau + \int_0^T V[z(\tau)]\, d\tau$$

(note sign of V) be defined for $z \in \mathscr{A}$, the space of absolutely continuous functions $z(\tau)$ defined on the interval $[0,T]$ with $z(0) = x$ and $z(T) = y$. Then A attains its global minimum at at least one path $X_{\min} \in \mathscr{A}$, $X_{\min}$ is smooth and satisfies the Euler–Lagrange equation $X_{\min}''(\tau) = V'[X_{\min}(\tau)]$, $\tau \in [0,T]$, with $X_{\min}(0) = x$ and $X_{\min}(T) = y$.

Proof: See for instance Chap. 4 of Ref. 11. □

The next lemma is basic to our results.

Lemma 2: Let the self-adjoint quantum mechanical Hamiltonian $H(\hbar) = [(-\hbar^2/2)\Delta_1 + V]$, where the real-valued potential $V \in C^\infty(\mathbb{R})$ is bounded below. Let the wavefunction $\psi \in \mathscr{S}(\mathbb{R})$. Then, for each fixed $T > 0$,

$$\exp\{-TH(\hbar)/\hbar\}\psi(x)$$
$$= \int \exp\{-TH(\hbar)/\hbar\}(x,y)\psi(y)\, dy,$$

where the kernel is given by

$$\exp\{-TH(\hbar)/\hbar\}(x,y)$$
$$= \hbar^{-1/2}\exp\{-A(X_{\min})/\hbar\}\int_{C_0[0,T]} d\mu_{0,0;0,T}(z)$$
$$\times \exp\left\{-\hbar^{-1}\int_0^T [V[X_{\min}(\tau) + \hbar^{1/2}z(\tau)]$$
$$- V[X_{\min}(\tau)] - \hbar^{1/2}z(\tau)V'[X_{\min}(\tau)]]\right\} d\tau$$

and $X_{\min}$ is the global minimum of A above, $\mu_{0,0;0,T}$ being the conditional Wiener measure on $C_0[0,T]$, the space of continuous functions $z:[0,T] \to \mathbb{R}$ with $z(0) = z(T) = 0$.

Proof: The Feynman–Kac formula[12] gives with $\lambda = \hbar^{1/2}$, for real $V \in C^\infty(\mathbb{R})$ bounded below, $\psi \in \mathscr{S}(\mathbb{R})$,

$$\exp\{-TH(\hbar)/\hbar\}\psi(x)$$
$$= \mathbb{E}_z\left\{\exp\left\{-\lambda^{-2}\int_0^T V[x + \lambda z(\tau)]d\tau\right\}\psi(x + \lambda z(T))\right\},$$

$\mathbb{E}_z$ denoting Wiener integral. We simply translate $z \to (z + a)$ with $a(\cdot) = \lambda^{-1}(X_{c1}(\cdot) - x)$, $X_{c1}(\cdot)$ being a smooth solution of $X_{c1}''(\tau) = V'[X_{c1}(\tau)]$, $\tau \in [0,T]$, with $X_{c1}(0) = x$, $X_{c1}(T)$ as yet unspecified. Using the translation formula,[13] which is valid since the functional integrand is bounded and continuous, we obtain

$$\exp\{-TH(\hbar)/\hbar\}\psi(x)$$
$$= \exp\left\{-\frac{1}{2\lambda^2}\int_0^T [X_{c1}'(\tau)]^2 d\tau\right\}$$
$$\times \mathbb{E}_z\left\{\exp\left\{-\lambda^{-2}\int_0^T V[X_{c1}(\tau) + \lambda z(\tau)]d\tau\right.\right.$$
$$\left.\left. - \lambda^{-1}\int_0^T X_{c1}'(\tau)\, dz(\tau)\right\}\psi(X_{c1}(T) + \lambda z(T))\right\}.$$

Setting

$$A(X_{c1}) = 2^{-1}\int_0^T [X'_{c1}(\tau)]^2\,d\tau + \int_0^T V[X_{c1}(\tau)]\,d\tau,$$

gives after integrating by parts[14] in the exponential,

$$\exp\{-TH(\hbar)/\hbar\}\psi(x)$$
$$= \exp\{-A(X_{c1})/\hbar\}\mathbb{E}_z\Big\{\exp\Big\{-\lambda^{-2}\int_0^T [V[X_{c1}(\tau)+\lambda z(\tau)]$$
$$- V[X_{c1}(\tau)] - \lambda z(\tau)V'[X_{c1}(\tau)]\,]d\tau\Big\}$$
$$\times\exp\{-\lambda^{-1}X'_{c1}(T)z(T)\}\psi(X_{c1}(T)+\lambda z(T))\Big\}.$$

But, using what we hope is an obvious notation,

$$\mathbb{E}_z\{F(z)\} = \int d\mu_{0,0;z(T),T}(z)\,dz(T)F(z),$$

giving for $y = \lambda z(T) + X_{c1}(T)$,

$$\Delta^2 V = V[X_{c1}(\tau)+\lambda z(\tau)] - V[X_{c1}(\tau)] - \lambda z(\tau)V'[X_{c1}(\tau)],$$
$$\exp\{-TH(\hbar)/\hbar\}\psi(x)$$
$$= \hbar^{-1/2}\exp\{-A(X_{c1})/\hbar\}$$
$$\times\int d\mu_{0,0;\lambda^{-1}(y-X_{c1}(T)),T}(z)\,dy\,\psi(y)$$
$$\times\exp\Big\{-\lambda^{-2}\int_0^T \Delta^2 V\,d\tau$$
$$-\lambda^{-2}X'_{c1}(T)(y - X_{c1}(T))\Big\}.$$

Hence we have shown that

$$\exp\{-TH(\hbar)/\hbar\}(x,y)$$
$$= \hbar^{-1/2}\exp\{-A(X_{c1})/\hbar\}\int d\mu_{0,0;\lambda^{-1}(y-X_{c1}(T)),T}(z)$$
$$\times\exp\Big\{-\lambda^{-2}\int_0^T \Delta^2 V\,d\tau$$
$$-\lambda^{-2}X'_{c1}(T)(y - X_{c1}(T))\Big\}.$$

We choose $X_{c1}(\cdot) = X_{\min}(\cdot)$ so that $X_{c1}(T) = y$ and the desired result follows. □

Lemma 2 has the following corollary.

Corollary 1: For real $V\in C^\infty(\mathbb{R})$ bounded below, for the self-adjoint quantum mechanical Hamiltonian $H(\hbar) = [(-\hbar^2/2)\Delta_1 + V]$, and for $\hbar^{1/2} = \lambda$,

$$\exp\{-TH(\hbar)/\hbar\}(x,y)$$
$$= (2\pi T\hbar)^{-1/2}\exp\{-A(X_{\min})/\hbar\}\mathbb{E}_z^T\{\exp\{\lambda^{-2}F(\lambda z)\}\}$$

where the functional

$$F(z) = -\int_0^T \{V[X_{\min}(\tau)+z(\tau)]$$
$$- V[X_{\min}(\tau)] - z(\tau)V'[X_{\min}(\tau)]\}\,d\tau$$

and

$$F(z) - 2^{-1}\int_0^T [z'(\tau)]^2\,d\tau$$
$$= -A(X_{\min}+z) + A(X_{\min}),\quad z\in C_0^*[0,T].$$

Proof: The first part of the corollary is obvious. From the hypothesis of Lemma 1

$$A(z) = 2^{-1}\int_0^T [z'(\tau)]^2\,d\tau + \int_0^T V[z(\tau)]\,d\tau,\quad z\in\mathscr{A}.$$

Thus

$$-A(X_{\min}+z) + A(X_{\min})$$
$$= -\int_0^T [V[X_{\min}(\tau)+z(\tau)] - V[X_{\min}(\tau)]$$
$$+ X'_{\min}(\tau)z'(\tau) + 2^{-1}[z'(\tau)]^2]\,d\tau,$$

since $X_{\min}\in\mathscr{A}$ by definition and $X_{\min}+z\in\mathscr{A}$, $z\in C_0^*[0,T]$. But $X_{\min}$ satisfies the Euler–Lagrange equation $X''_{\min} = V'[X_{\min}]$ and so

$$\int_0^T X'_{\min}(\tau)z'(\tau)\,d\tau = -\int_0^T z(\tau)V'[X_{\min}(\tau)]\,d\tau,$$

giving the required result. □

Evidently for us to be able to apply Theorem 1 we require the maximum of $[-A(X_{\min}+z) + A(X_{\min})]$ to be unique in $C_0^*[0,T]$. By definition of $X_{\min}$ this maximum is zero and it is uniquely attained at $z\equiv 0$ iff $X_{\min}$ the global minimum of A over $\mathscr{A}$ is unique. When V is convex the next lemma assures us that this uniqueness holds.

Lemma 3: When $V\in C^\infty(\mathbb{R})$ is real-valued, bounded below and strictly convex $X_{\min}$ the global minimum of

$$A(X) = 2^{-1}\int_0^T [X'(\tau)]^2\,d\tau + \int_0^T V[X(\tau)]\,d\tau$$

over $\mathscr{A}$, the absolutely continuous paths on $[0,T]$ with $X(0) = x$, $X(T) = y$, is unique.

Proof: Assume there are two global minima X_1 and X_2. Then, since X_1 and X_2 are smooth and each satisfy the Euler–Lagrange equation, if $h(\tau) = X_1(\tau) - X_2(\tau)$, $h(0) = h(T) = 0$, and from Taylor's theorem, for some τ-dependent θ, $0\leqslant\theta\leqslant 1$,

$$h''(\tau) = h(\tau)V''[X_1(\tau) - \theta h(\tau)],\quad \tau\in[0,T].$$

When $h'(0) = 0$, $X'_1(0) = X'_2(0)$, and the usual theory of ordinary differential equations assures us that $X_1(\tau) = X_2(\tau)$, $\tau\in[0,T]$. Thus, there are only two cases to consider $h'(0) > 0$ and $h'(0) < 0$. In the first case $h(0) = h(T) = 0$, $h'(0) > 0$, assures us that $h(\cdot)$ must have a first local maximum ξ, say, with $h(\xi) > 0$, $h'(\xi) = 0$. But from the above equation $h''(\xi) > 0$ and we have a contradiction. The second case is dealt with similarly by looking for a first local minimum. □

We now come to our first application.

Theorem 2: Let $H(\hbar)$ be the self-adjoint quantum mechanical Hamiltonian $H(\hbar) = [(-\hbar^2/2)\Delta_1 + V]$, where $V\in C^\infty(\mathbb{R})$ is strictly convex and bounded below. Let $X_{\min}$ denote the unique minimum of

$$A(z) = 2^{-1}\int_0^T [z'(\tau)]^2\,d\tau + \int_0^T V[z(\tau)]\,d\tau$$

over the absolutely continuous paths $z:[0,T]\to\mathbb{R}$ with $z(0) = x$, $z(T) = y$. Then, for each finite time $T > 0$,

$$\exp\{-TH(\hbar)/\hbar\}(x,y)$$
$$= \exp\{-A(X_{\min})/\hbar\}(2\pi T\hbar)^{-1/2}$$
$$\times\mathbb{E}_z^T\{\exp\{\hbar^{-1}F(\hbar^{1/2}z)\}\},$$

where

$$F(z) = -\int_0^T \{V[X_{\min}(\tau)+z(\tau)] - V[X_{\min}(\tau)] - z(\tau)V'[X_{\min}(\tau)]\}\,d\tau,$$

satisfies the conditions (1)–(6) of Theorem 1.

Hence, for each finite integer $n \geqslant 3$,

$$\exp\{-TH(\hbar)/\hbar\}(x,y)\exp\{A(X_{\min})/\hbar\}$$
$$= (2\pi T\hbar)^{-1/2}\sum_{i=0}^{n-3}(i!)^{-1} \times \mathbb{E}_z^T\left\{\exp\left\{-\frac{1}{2}\int_0^T V''(X_{\min})z^2\,d\tau\right\} \times\left(\hbar^{1/2}(3!)^{-1}\int_0^T V'''(X_{\min})z^3\,d\tau + \cdots + \hbar^{(n-3)/2}[(n-1)!]^{-1}\int_0^T V^{n-1}(X_{\min})z^{n-1}\,d\tau\right)^i\right\} + O(\hbar^{(n-2)/2}), \quad (*)$$

$\int_0^T V''(X_{\min})z^2\,d\tau$ being a shorthand for $\int_0^T V''(X_{\min}(\tau))z^2(\tau)\,d\tau$, etc.

Proof: We need only show that conditions (1)–(6) of Theorem 1 are satisfied for

$$F(z) = -\int_0^T \{V[X_{\min}(\tau)+z(\tau)] - V[X_{\min}(\tau)] - z(\tau)V'[X_{\min}(\tau)]\}\,d\tau.$$

Conditions (1), (3), (4), and (6) are easy to check. Condition (2) is satisfied because, for some η, with $\|\eta\| \leqslant \|z\|$,

$$F(z) = -2^{-1}\int_0^T V''[X_{\min}(\tau)+\eta(\tau)]z^2(\tau)\,d\tau \leqslant 0.$$

Condition (5) is valid because a simple calculation gives

$$D^2F(x_0+\eta)z^2 = D^2F(\eta)z^2 = -\int_0^T V''[X_{\min}(\tau)+\eta(\tau)]z^2(\tau)\,d\tau < 0$$

and $\mathbb{E}_z^T\{\exp\{(1+\epsilon)D^2F(x_0+\eta)z^2/2\}\} \leqslant \mathbb{E}_z^T\{1\} = 1$.

The final identity of Theorem 2 follows from the proof of Theorem 1. □

Remark: The asymptotic series resulting from equation (*) does not contain any odd powers of $\hbar^{1/2}$, because of the symmetry of the Gaussian measure. We now extend our results to a wider class of potentials.

Lemma 4: When $V \in C^\infty(\mathbb{R})$ is real-valued bounded below and $V'' \geqslant -|\beta|$, for some constant β, the global minimum of

$$A(X) = 2^{-1}\int_0^T [X'(\tau)]^2\,d\tau + \int_0^T V[X(\tau)]\,d\tau,$$

over $\mathscr{A}$, the space of absolutely continuous paths on $[0,T]$ with $X(0)=x$, $X(T)=y$ is unique if $T < \pi/|\beta|^{1/2}$.

Proof: Assume there are two global minima X_1 and X_2. Since X_1 and X_2 are smooth and each satisfy the Euler–Lagrange equation, if $h(\tau) = X_1(\tau) - X_2(\tau)$, $h(0) = h(T) = 0$, and, for some point $\xi(\tau)$ intermediate to $X_1(\tau)$ and $X_2(\tau)$,

$$h''(\tau) = h(\tau)V''(\xi(\tau)), \quad \tau \in [0,T].$$

Multiplying both sides of this equation by $h(\tau)$ and integrating by parts gives

$$\int_0^T [h'(\tau)]^2\,d\tau = -\int_0^T h^2(\tau)V''(\xi(\tau))\,d\tau.$$

However,

$$\int_0^T [h'(\tau)]^2\,d\tau \geqslant \frac{\pi^2}{T^2}\int_0^T h^2(\tau)\,d\tau,$$

π^2/T^2 being the least eigenvalue λ of the equation $-h''(\tau) = \lambda h(\tau)$, with $h(0) = h(T) = 0$. By hypothesis $-V''(\xi(\tau)) \leqslant |\beta|$ and so from above

$$\frac{\pi^2}{T^2}\int_0^T h^2(\tau)\,d\tau \leqslant |\beta|\int_0^T h^2(\tau)\,d\tau.$$

Hence, for $T < \pi/|\beta|^{1/2}$, we must have $\int_0^T h^2(\tau)\,d\tau = 0$, so that $X_1(\tau) = X_2(\tau)$, $\tau \in [0,T]$. □

Our second application is contained in Theorem 3.

Theorem 3: Let $H(\hbar)$ be the self-adjoint quantum mechanical Hamiltonian $H(\hbar) = [(-\hbar^2/2)\Delta_1 + V]$, where $V \in C^\infty(\mathbb{R})$ is bounded below and $V'' \geqslant -|\beta|$. Let $A(X)$ as defined above attain its global minimum at the unique point $X_{\min} \in \mathscr{A}$, the space of absolutely continuous functions $z:[0,T] \to \mathbb{R}$ with $z(0) = x$, $z(T) = y$. Then, for sufficiently small time T,

$$\exp\{-TH(\hbar)/\hbar\}(x,y) = \exp\{-A(X_{\min})/\hbar\}(2\pi T\hbar)^{-1/2} \times \mathbb{E}_z^T\{\exp\{\hbar^{-1}F(\hbar^{1/2}z)\}\},$$

where

$$F(z) = -\int_0^T \{V[X_{\min}(\tau)+z(\tau)] - V[X_{\min}(\tau)] - z(\tau)V'[X_{\min}(\tau)]\}\,d\tau,$$

satisfies the conditions (1)–(6) of Theorem 1.

Hence equation (*) is valid for each finite integer $n \geqslant 3$, for sufficiently small time T.

Proof: Lemma 2 establishes the desired identity given in the equation above.

Conditions (1), (3), (4), and (6) are easy to check. Since, for some η with $\|\eta\| \leqslant \|z\|$,

$$F(z) = -2^{-1}\int_0^T V''[X_{\min}(\tau)+\eta(\tau)]z^2(\tau)\,d\tau \leqslant \frac{|\beta|T}{2}\|z\|^2 = L_2\|z\|^2,$$

for $L_2 = |\beta|T/2$, condition (2) is valid if $T < \min\{(\gamma/|\beta|)^{1/2}, (1/2|\beta|)^{1/2}\}$.

Condition (5) is valid because

$$D^2F(x_0+\eta)z^2 = D^2F(\eta)z^2 = -\int_0^T V''[X_{\min}(\tau)+\eta(\tau)]z^2(\tau)\,d\tau \leqslant |\beta|T\|z\|^2$$

and

$$\mathbb{E}_z^T\{\exp\{(1+\epsilon)D^2F(x_0+\eta)z^2/2\}\} \leqslant \mathbb{E}_z^T\{\exp\{(1+\epsilon)(|\beta|T\|z\|^2/2)\}\}$$

for $(1+\epsilon)|\beta|T/2 < \gamma/T$ from Lemma 8. The last condition is valid, for some $\epsilon > 0$, if $T < (2\gamma/|\beta|)^{1/2}$, which is valid by hypothesis since $T < (\gamma/|\beta|)^{1/2}$.

Remark: In the last theorem it is important to have the best possible value for γ. Lemma 5 gives the value of γ, $\gamma = (1-2^{-1/4})^2/4$, an inequality of Doob[15] gives the improved value $\gamma = 2$. Doob's inequality is, however, much more difficult to prove than the simple estimate of Lemma 5.

The justification for calling Theorem 2 and 3 generalized Mehler kernel formulas is given in the final corollary.

Corollary 2: Let the self-adjoint quantum mechanical Hamiltonian $H(\hbar) = [(-\hbar^2/2)\Delta_1 + V]$, where $V \in C^\infty(\mathbb{R})$ satisfies the conditions in Theorem 2 (Theorem 3). Then, setting $A(X_{\min}) = A(x,y,T)$, for each finite time $T>0$ (for each sufficiently small finite time $T>0$),

$$\exp\{-TH(\hbar)/\hbar\}(x,y) = (2\pi\hbar)^{-1/2}\exp\{-A(x,y,T)/\hbar\} \times\left(\frac{\partial^2 A}{\partial x\,\partial y}(x,y,T)\right)^{1/2}\{1+\hbar K_1 + O(\hbar^2)\},$$

where

$$K_1 = -\frac{1}{8}\int_0^T V^{(iv)}(X_{\min}(\tau))G^2(\tau,\tau)\,d\tau + \frac{1}{24}\int_0^T\int_0^T d\tau\,d\sigma\, V'''(X_{\min}(\tau))V'''(X_{\min}(\sigma)) \times[3G(\tau,\tau)G(\tau,\sigma)G(\sigma,\sigma) + 2G^3(\tau,\sigma)]\,d\tau\,d\sigma,$$

$G(\tau,\sigma)$ (the Feynman–Green's function) being the Green's function of the Sturm–Liouville differential operator $[d^2/d\sigma^2 - V''(X_{\min}(\sigma))]$ with zero boundary conditions $G(0,\tau) = G(T,\tau) = 0$.

Proof: The proof of this result (and corresponding results for higher order terms) follows from the identity:

$$I = (2\pi T)^{1/2}\mathbb{E}_z^T\left\{\exp\left\{\sum_{i=1}^n \alpha_i z(t_i)\right\} \times\exp\left\{-\frac{1}{2}\int_0^T V''[X_{\min}(\tau)]z^2(\tau)\,d\tau\right\}\right\} = (2\pi)^{-1/2}\left(\frac{\partial^2 A}{\partial x\,\partial y}(x,y,t)\right)^{1/2}\exp\left\{\frac{1}{2}\sum_{i,j=1}^n \alpha_i\alpha_j G(t_i,t_j)\right\}.$$

To prove this denote $\int_0^T z'(\tau)w'(\tau)\,d\tau$ by (z,w), for $z,w \in C_0^*[0,T]$. Equipped with this inner product $C_0^*[0,T]$ is a reproducing kernel Hilbert space with kernel $\rho(s,t) = s(1-t/T)$, $s<t$. Now defining $L: C_0^*[0,T] \to C_0^*[0,T]$ by $(Lz)''(\tau) = -V''(X_{\min}(\tau))z(\tau)$ and integrating by parts we see that

$$(z,Lz) = \int_0^T V''(X_{\min}(\tau))z^2(\tau)\,d\tau.$$

Integrating we can find (Lz) explicitly. We obtain

$$(Lz)(t) = -\int_0^t dt'\int_0^{t'} V''(X_{\min}(t''))z(t'')\,dt'' + \frac{t}{T}\int_0^T dt'\int_0^{t'} V''(X_{\min}(t''))z(t'')\,dt''.$$

Hence, for θ the Heaviside function,

$$(Lz)(t) = \int_0^T\left[\frac{t}{T} - \theta(t-t')\right]dt' \times\int_0^{t'} V''(X_{\min}(t''))z(t'')\,dt''.$$

Define $L_1: C_0^*[0,T] \to C_0^*[0,T]$ by

$$(L_1 w)(t) = \int_0^T\left(\frac{t}{T} - \theta(t-t')\right)w(t')\,dt'$$

and define

$$(L_2 z)(t') = \int_0^{t'} V''(X_{\min}(t''))z(t'')\,dt'', \qquad L_2: \tilde{C}_0[0,T] \to \tilde{C}_0[0,T],$$

where $\tilde{C}_0[0,T]$ is the Hilbert space of absolutely continuous $z:[0,T]\to\mathbb{R}$ with $z(0)=0$. Then $L = L_1L_2$.

Let $\{\theta_1,\ldots,\theta_n\}$, $\{\phi_1,\ldots,\phi_n\}$ be orthonormal systems in $C_0^*[0,T]$. Then from the Cauchy–Schwarz inequalities

$$\sum_{i=1}^n |(\phi_i, L_1L_2\theta_i)| \leqslant \sum_{i=1}^n \|L_1^*\phi_i\|\|L_2\theta_i\| \leqslant\left\{\sum_{i=1}^n (\phi_i, L_1L_1^*\phi_i)\right\}^{1/2}\left\{\sum_{i=1}^n (\theta_i, L_2^*L_2\theta_i)\right\}^{1/2} \leqslant \|L_1^*\|_2\|L_2\|_2 = \|L_1\|_2\|L_2\|_2 < \infty,$$

$\|\ \|_2$ being the Hilbert–Schmidt norms on $C_0^*[0,T]$ and $\tilde{C}_0[0,T]$. Here the last step is valid because

$$(L_1 w)(t) = \int_0^T\left[\frac{t}{T} - \theta(t-t')\right]\int_0^{t'}\frac{dw}{dt''}(t'')\,dt'' = \int_0^T\frac{dw}{dt''}(t'')\,dt''\int_{t''}^T dt'\left[\frac{t}{T} - \theta(t-t')\right] = \int_0^T\frac{dw}{dt''}(t'')\left[\frac{t}{T}(T-t'') - \theta(t-t'')(t-t'')\right]$$

show that L_1 has an L^2 kernel and $\|L_1\|_2 < \infty$ and similarly for L_2. It follows from the above inequality that $L = L_1L_2$ is trace-class.

Setting

$$z_0(t) = \sum_{i=1}^n \alpha_i\,\rho(t_i,t),$$

and using a standard result for Gaussian processes,[16,17] we now obtain

$$I = (2\pi T)^{-1/2}\mathbb{E}_z^T\{\exp\{(z_0,z)\}\exp\{-\tfrac{1}{2}(z,Lz)\}\} = (2\pi T)^{-1/2}\exp\{\tfrac{1}{2}(z_0,(1+L)^{-1}z_0)\}[\det(1+L)]^{-1/2}.$$

$\det(1+L)$ and $(1+L)^{-1}$ can now be calculated as in Ref. 1. The results of the calculation are detailed above. □

Essentially the above result was derived formally by Mizrahi.[1] Equivalent results were obtained by DeWitt[2] and independently and almost simultaneously by Truman.[3] We believe the above to be the first rigorous proof of this result. Results for the behavior for large arguments of eigenfunctions of quantum mechanical Hamiltonians reminiscent of those above have been published by Carmona and Simon.[18] The calculation of higher order terms, using the Feynman–Green's function, is extremely difficult and the results stated in Theorems 2 and 3 seem more useful.

3. PRELIMINARY LEMMAS AND PROOF OF THEOREM 1

A. Preliminary lemmas

We begin with a crucial estimate which is based on a beautiful argument of Kolmogorov. We learned this method of proof in Simon's excellent book. (See Lemma 18.7 in Ref. 5).

Lemma 5: For some fixed constants $C,\gamma>0$ and for sufficiently large a'

$$\mathbb{E}_z^T\{\|z\|>a'\}<C\exp\{-\gamma a'^2/T\}.$$

Proof: Denote the covariance $\mathbb{E}_z^T\{z(s)z(t)\}$ by $\rho(s,t)$ where $\rho(s,t)=s(1-t/T)$, $0\leqslant s<t\leqslant T$. Setting

$$\begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} = \begin{pmatrix} \rho(s,s) & \rho(s,t) \\ \rho(t,s) & \rho(t,t) \end{pmatrix}^{-1} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}^{-1},$$

the joint probability distribution of $x_1=z(s)$, $x_2=z(t)$ is proportional to $\exp\{-\frac{1}{2}(b_{11}x_1^2+b_{22}x_2^2+2b_{12}x_1x_2\}$. Defining $y=(x_1-x_2)$, we see that the joint probability distribution is proportional to

$$\exp\{-\tfrac{1}{2}[(b_{11}+b_{22}+2b_{12})(x_1-(b_{22}+b_{12})y/(b_{11}+b_{22}+2b_{12}))^2+(b_{22}b_{11}-b_{12}^2)y^2/(b_{11}+b_{22}+2b_{12})]\}.$$

Integrating with respect to x_1, gives

$$\mathbb{E}_z^T\{|z(s)-z(t)|\geqslant a'\}=N\int_a^\infty \exp\left\{\frac{-y^2}{2(a_{11}+a_{22}-2a_{12})}\right\}dy,$$

for some normalization constant N. However, from the above explicit form of ρ,
$a_{11}+a_{22}-2a_{12}=\mathbb{E}_z^T\{|z(t)-z(s)|^2\}\leqslant 2|s-t|$, $s,t\in[0,T]$. And so, substituting for N, we deduce that, for $a>0$,

$$\begin{aligned}\mathbb{E}_z^T\{|z(s)-z(t)|\geqslant a\} &\leqslant N\int_a^\infty \exp\left\{\frac{-y^2}{4|s-t|}\right\}dy \\ &= 2(2\pi)^{-1/2}\int_{a/(2|s-t|)^{1/2}}^\infty \exp\left(\frac{-x^2}{2}\right)dx.\end{aligned}$$

Since $\int_a^\infty \exp(-x^2/2)dx\leqslant a^{-1}\exp(-a^2/2)$, we arrive at

$$\mathbb{E}_z^T\left\{\left|z\left(\frac{kT}{2^n}\right)-z\left(\frac{\overline{k-1}T}{2^n}\right)\right|\geqslant C_0 2^{-n/4}\right\} \leqslant 2(2\pi)^{-1/2}C_0^{-1}T^{1/2}2^{1/2}2^{-n/4}\exp\left\{\frac{-C_0^2 2^{n/2}}{4T}\right\}.$$

It follows that

$$\mathbb{E}_z^T\left\{\left|z\left(\frac{kT}{2^n}\right)-z\left(\frac{\overline{k-1}T}{2^n}\right)\right| \geqslant C_0 2^{-n/4},\ \text{some } 0\leqslant k\leqslant 2^n,\ \text{some } n\right\} \leqslant 2\pi^{-1/2}C_0^{-1}T^{1/2}\sum_{n=0}^\infty 2^{3n/4}\exp\left\{\frac{-C_0^2 2^{n/2}}{4T}\right\}.$$

Denoting the nth term in the above infinite series by u_n, for $n\geqslant 1$, if $C_0^2/4T>2$,

$$u_n/u_{n-1}=\exp\{\tfrac{3}{4}\ln 2-2^{(n-1)/2}(2^{1/2}-1)C_0^2/4T\}\leqslant e^{-\alpha},$$

where $\alpha=2(2^{1/2}-1)-(3/4)\ln 2>0$. Hence, for $C_0^2>8T$, we obtain the estimate

$$\mathbb{E}_z^T\left\{\left|z\left(\frac{kT}{2^n}\right)-z\left(\frac{\overline{k-1}T}{2^n}\right)\right| \geqslant C_0 2^{-n/4},\ \text{some } 0\leqslant k\leqslant 2^n,\ \text{some } n\right\} \leqslant \frac{2\pi^{-1/2}C_0^{-1}T^{1/2}}{(1-e^{-\alpha})}\exp\left\{\frac{-C_0^2}{4T}\right\}.$$

Now, if

$$\left|z\left(\frac{kT}{2^n}\right)-z\left(\frac{\overline{k-1}T}{2^n}\right)\right|\leqslant C_0 2^{-n/4},$$

for $0\leqslant k\leqslant 2^n$, for each n, then, writing $t=T\sum_{j=1}^\infty a_j 2^{-j}$, each a_j being 0 or 1, $t\in[0,T]$, and observing

$$|z(t)-z(0)|<\sum_{j=1}^\infty |z(s_{j+1})-z(s_j)|,$$

s_j being the jth partial sum of $T\sum_{j=1}^\infty a_j 2^{-j}$, we deduce

$$|z(t)-z(0)|\leqslant\sum_{j=0}^\infty a_j C_0 2^{-j/4}\leqslant C_0/(1-2^{-1/4})=a'.$$

Hence, if $|z(t)-z(0)|>a'$, for some $t\in[0,T]$, we know that for some $0\leqslant k\leqslant 2^n$, some n,

$$\left|z\left(\frac{kT}{2^n}\right)-z\left(\frac{\overline{k-1}T}{2^n}\right)\right|>C_0 2^{-n/4},$$

and it follows from the above that, since $z(0)=0$,

$$\mathbb{E}_z^T\{|z(t)|>a',\ \text{for some } t\in[0,T]\} \leqslant \frac{2\pi^{-1/2}T^{1/2}}{a'(1-e^{-\alpha})}\exp\{-a'^2(1-2^{-1/4})^2/4T\},$$

for $a'>(8T)^{1/2}(1-2^{-1/4})^{-1}$, $\alpha=2(2^{1/2}-1)-\frac{3}{4}\ln 2>0$.

Lemma 6 (Schilder): If

$$\max_{0\leqslant j\leqslant n-1}\left[\sup_{jT/n<\tau<(j+1)T/n}\left|z(\tau)-z\left(\frac{jT}{n}\right)\right|\right]<\frac{\delta}{2},$$

then $\|z-z^n(\cdot)\|<\delta$, $z^n(\cdot)$ being the polygonalization of z.

Proof: Let τ_0 be a point where $|z(\tau)-z^n(\tau)|$ attains its maximum. Let j be such that $jT/n<\tau_0\leqslant(j+1)T/n$. Now

$$|z(\tau_0)-z^n(\tau_0)|\leqslant|z(\tau_0)-z(jT/n)|+|z(jT/n)-z^n(\tau_0)|.$$

By hypothesis

$$|z(\tau_0)-z(jT/n)|<\delta/2 \quad\text{and}\quad |z(\overline{j+1}T/n)-z(jT/n)|<\delta/2.$$

From the last inequality every point on the straight line z^n joining $(jT/n, z(jT/n))$ and $(\overline{j+1}T/n, z(\overline{j+1}T/n))$ satisfies $|z(jT/n)-z^n(\tau)|<\delta/2$, $\tau\in[jT/n,(j+1)T/n]$. In particular this is valid for $\tau=\tau_0$. This proves the lemma. □

We also require a more refined estimate than that of lemma 5.

Lemma 7: For each sufficiently large integer m, for any $\delta'>0$,

$$\mathbb{E}_z^T\{\|z-z^m(\cdot)\|\geqslant\delta'\}\leqslant C\delta'^{-1}(mT)^{1/2}\exp\{-Dm\delta'^2/T\},$$

for constants $C,D>0$.

Proof: Repeating the argument of Lemma 5, we obtain, for $C_0^2 m/4T>2$,

$$\mathbb{E}_z^T\left\{\left|z\left(\frac{jT}{m}+\frac{kT}{m2^n}\right)-z\left(\frac{jT}{m}+\frac{\overline{k-1}T}{m2^n}\right)\right| \geqslant C_0 2^{-n/4},\ \text{some } 0\leqslant k\leqslant 2^n,\ \text{some } n\right\} \leqslant \frac{2\pi^{-1/2}}{(1-e^{-\alpha})}C_0^{-1}\left(\frac{T}{m}\right)^{1/2}\exp\left\{-\frac{mC_0^2}{4T}\right\},$$

where $\alpha=2(2^{1/2}-1)-\frac{3}{4}\ln 2>0$.

We now observe that, if

$$\left|z\left(\frac{jT}{m}+\frac{kT}{m2^n}\right)-z\left(\frac{jT}{m}+\frac{\overline{k-1}T}{m2^n}\right)\right|\leqslant C_0 2^{-n/4},$$

for $0\leqslant k\leqslant 2^n$, for each n, then writing

$$\left(t-\frac{jT}{m}\right)=\frac{T}{m}\sum_{k=1}^{\infty}a_k 2^{-k},$$

where $a_k=0$ or 1, as in Lemma 5, we deduce

$$\left|z(t)-z\left(\frac{jT}{m}\right)\right|\leqslant C_0/(1-2^{-1/4})=C_0',$$

for $t\in[jT/m,(j+1)T/m]$. Hence, if $|z(t)-z(jT/m)|\geqslant C_0'$, for some $t\in[jT/m,(j+1)T/m]$, we know that, for some n and some $0\leqslant k\leqslant 2^n$,

$$\left|z\left(\frac{jT}{m}+\frac{kT}{m2^n}\right)-z\left(\frac{jT}{m}+\frac{\overline{k-1}T}{m2^n}\right)\right|\geqslant C_0 2^{-n/4}.$$

From the above inequality then, for $C_0=C_0'(1-2^{-1/4})$, $C_0^2 m/4T>2$,

$$\mathbb{E}_z^T\left\{\left|z(t)-z\left(\frac{jT}{m}\right)\right|\geqslant C_0', \text{ some } t\in\left[\frac{jT}{m},\frac{(j+1)T}{m}\right]\right\}$$
$$\leqslant\frac{2\pi^{-1/2}}{(1-e^{-\alpha})}C_0^{-1}\left(\frac{T}{m}\right)^{1/2}\exp\left\{-\frac{mC_0^2}{4T}\right\}.$$

We now let

$$Q_j^m=\left\{z\in C_0[0,T]\,\middle|\,\sup_{jT/m<\tau<(j+1)T/m}\left|z(\tau)-z\left(\frac{jT}{m}\right)\right|\geqslant\frac{\delta'}{2}\right\},$$

for $j=0,1,2,\ldots,(m-1)$, integer m. Using the above inequality, we see that

$$\mathbb{E}_z^T\left\{\bigcup_{j=0}^{m-1}Q_j^m\right\}\leqslant\frac{2\pi^{-1/2}}{(1-e^{-\alpha})}C_0^{-1}(mT)^{1/2}\exp\left\{-\frac{mC_0^2}{4T}\right\},$$

where $C_0=\delta'(1-2^{-1/4})/2$. Now, if

$$z\notin\bigcup_{j=0}^{m-1}Q_j^m,$$

then

$$\sup_{jT/m<\tau<(j+1)T/m}\left|z(\tau)-z\left(\frac{jT}{m}\right)\right|<\frac{\delta'}{2},$$

for $j=0,1,2,\ldots,(m-1)$ and so by Lemma 6 $\|z-z^m(\cdot)\|<\delta'$.

It follows that, for $\delta'^2 m(1-2^{-1/4})^2/16T>2$,

$$\mathbb{E}_z^T\{\|z-z^m(\cdot)\|\geqslant\delta'\}$$
$$\leqslant\frac{2\pi^{-1/2}}{(1-e^{-\alpha})}C_0^{-1}(mT)^{1/2}$$
$$\times\exp\{-m\delta'^2(1-2^{-1/4})^2/16T\},$$

where $C_0=\delta'(1-2^{-1/4})/2$. □

The functionals on $C_0[0,T]$ which we encounter will be bounded by $\exp\{k\|z\|^2+l\|z\|\}$, $z\in C_0[0,T]$. The next estimate is basic in our theorem.

Lemma 8: for $k,l\in\mathbb{R}$, $\mathbb{E}_z^T\{\exp\{k\|z\|^2+l\|z\|\}\}<\infty$, for $k<\gamma/T$, γ being as given in Lemma 5.

Proof: Let $\chi(n)$ be the characteristic function of $\{z\in C_0[0,T]\,|\,\|z\|\in(n,n+1]\}$. Then

$$\mathbb{E}_z^T\{\exp\{k\|z\|^2+l\|z\|\}\}$$
$$=\sum_{n=0}^{\infty}\mathbb{E}_z^T\{\exp\{k\|z\|^2+l\|z\|\}\chi(n)\}$$
$$\therefore\mathbb{E}_z^T\{\exp\{k\|z\|^2+l\|z\|\}\}$$
$$\leqslant\sum_{n=0}^{\infty}\exp\{k(n+1)^2+l(n+1)\}\mathbb{E}_z^T\{\chi(n)\}$$
$$\therefore\mathbb{E}_z^T\{\exp\{k\|z\|^2+l\|z\|\}\}$$
$$\leqslant\sum_{n=0}^{\infty}\exp\{k(n+1)^2+l(n+1)\}\mathbb{E}_z^T\{\|z\|>n\}.$$

Hence, we deduce from Lemma 5

$$\mathbb{E}_z^T\{\exp\{k\|z\|^2+l\|z\|\}\}$$
$$\leqslant\sum_{n=0}^{\infty}C\exp\{k(n+1)^2+l(n+1)-\gamma n^2/T\}<\infty,$$

for $k<\gamma/T$. □

As has been seen previously, for each $z\in C_0[0,T]$, we can define its polygonalization $z^n(\cdot)$ by

$$z^n(\tau)=z\left(\frac{jT}{n}\right)+\left(\tau-\frac{jT}{n}\right)\left(z\left(\frac{\overline{j+1}T}{n}\right)-z\left(\frac{jT}{n}\right)\right)\frac{n}{T},$$

$$\frac{jT}{n}\leqslant\tau<\frac{(j+1)T}{n},\quad j=0,1,\ldots,(n-1).$$

We denote the associated vector by z^n, where

$$z^n=(z(T/n),z(2T/n),\ldots,z(\overline{n-1}T/n),z(T)),$$

and in what follows, for vectors z^n,

$$\|z^n\|=\sup_{j=1,2,\ldots,n}|z_j^n|,$$

z_j^n being the components of z^n, denotes the supnorm of z^n. The next four lemmas are due to Schilder.

Lemma 9 (Schilder): Let A_n be the following $(n\times n)$ tridiagonal matrix

$$(n/T)\begin{bmatrix}2&-1&0&&&&&&\\-1&2&-1&&&&&&\\0&-1&2&-1&&&&&\\\cdot&\cdot&\cdot&\cdot&\cdot&\cdot&\cdot&\cdot&\cdot\\\cdot&\cdot&\cdot&0&-1&2&-1&&\\\cdot&\cdot&\cdot&\cdot&0&-1&1&&\end{bmatrix}$$

then for s^n, an n-dimensional vector

$$s^n A_n s^n = \sum_{j=1}^{n} (n/T)(s_j^n - s_{j-1}^n)^2, \quad s_0^n = 0.$$

We postulate that

$$\int_0^T [ds^n(\tau)/d\tau]^2 \, d\tau = s^n A_n s^n$$

and

$$z^n A_n z^n \leqslant \int_0^T [z'(\tau)]^2 \, d\tau,$$

where $s^n(\tau)$ is that polygonal function belonging to $C_0[0,T]$ having s^n as its associated vector.

Proof: See Lemma 4 of Ref. 4. □

Lemma 10 (Schilder): If $z \in C_0^*[0,T]$, then, for $\tau_2 > \tau_1$,

$$\sup_{\tau_1 \leqslant \tau \leqslant \tau_2} [z(\tau) - z(\tau_1)]^2 \leqslant (\tau_2 - \tau_1) \int_{\tau_1}^{\tau_2} [z'(\tau)]^2 \, d\tau.$$

Proof: See Lemma 5 of Ref. 4. □

Let us define the function $A(\delta)$ by

$$A(\delta) = \sup_{z \in A_\delta} \left\{ F(z) - \frac{1}{2} \int_0^T [z'(\tau)]^2 \, d\tau \right\},$$

where A_δ is the set

$$\{ z \in C_0^*[0,T] \mid \|z - x_0\| \geqslant \delta \}, \quad \text{for } \delta > 0.$$

Lemma 11 (Schilder): Suppose $F(z)$ satisfies the conditions given in the theorem then $A(\delta) < 0$ for $\delta > 0$.

Proof: See Lemma 6 of Ref. 4. □

Lemma 12 (Schilder): If $\|s^n(\cdot) - x_0^n(\cdot)\| \geqslant \omega$ and $w - \rho_n > 0$ then $F[s^n(\cdot)) - \frac{1}{2} s^n A_n s^n \leqslant A(\omega - \rho_n)$, where $\rho_n = \|x_0 - x_0^n(\cdot)\|$. $x_0^n(\cdot)$ being the n-polygonalization of the unique maximizing path x_0 of

$$F(z) - \frac{1}{2} \int_0^T [z'(\tau)]^2 \, d\tau.$$

Proof: See Lemma 7 of Ref. 4. □

The following lemma is merely an evaluation of one of the constants that would appear in a statement of equivalence of norms. The interesting thing is that the proof does not seem to come from any eigenvalue result.

Lemma 13: Let A_n be the matrix defined in Lemma 9. Then, if w^n is any vector in $\mathbb{R}^n$,

$$w^n A_n w^n \geqslant T^{-1} \|w^n\|^2,$$

$\| \ \|$ being supnorm.

Proof: From Lemma 9 we are required to prove that

$$n \sum_{j=1}^{n} (w_j^n - w_{j-1}^n)^2 \geqslant \|w^n\|^2,$$

$w^n = (w_1^n, w_2^n, \ldots, w_n^n)$, $w_0^n = 0$. Let $P(k)$ be the proposition:

$$k \left[(w_1^n)^2 + (w_2^n - w_1^n)^2 + \cdots + (w_k^n - w_{k-1}^n)^2 \right] \geqslant \|(w_1^n, w_2^n, \ldots, w_k^n, 0, 0, \ldots, 0)\|^2.$$

Assume that $P(k)$ is true for some $k < n$. Then, trivially, by inductive hypothesis,

$$\begin{aligned}(k+1)&\left[(w_1^n)^2 + (w_2^n - w_1^n)^2 + \cdots + (w_k^n - w_{k+1}^n)^2 \right] \\ &\geqslant (k+1)\left[(w_1^n)^2 + (w_2^n - w_1^n)^2 + \cdots + (w_k^n - w_{k-1}^n)^2 \right] \\ &\geqslant k \left[(w_1^n)^2 + (w_2^n - w_1^n)^2 + \cdots + (w_k^n - w_{k-1}^n)^2 \right] \\ &\geqslant \|(w_1^n, w_2^n, \ldots, w_k^n, 0, 0, \ldots, 0)\|^2.\end{aligned}$$

Thus, we need only prove that

$$(k+1)\left[(w_1^n)^2 + (w_2^n - w_1^n)^2 + \cdots + (w_{k+1}^n - w_k^n)^2 \right] \geqslant (w_{k+1}^n)^2.$$

However,

$$\begin{aligned}(k+1)&\left[(w_1^n)^2 + (w_2^n - w_1^n)^2 + \cdots + (w_{k+1}^n - w_k^n)^2 \right] \\ &= (k+1)\left[(w_1^n)^2 + (w_2^n - w_1^n)^2 + \cdots + (w_k^n - w_{k-1}^n)^2 \right] \\ &\quad + (k+1)(w_{k+1}^n - w_k^n)^2\end{aligned}$$

and by inductive hypothesis

$$\begin{aligned}(k+1)&\left[(w_1^n)^2 + \cdots + (w_{k+1}^n - w_k^n)^2 \right] \\ &\geqslant (k+1)(w_k^n)^2/k + (k+1)(w_{k+1}^n - w_k^n)^2 \\ &= f(w_k^n / w_{k+1}^n)(w_{k+1}^n)^2,\end{aligned}$$

where $f(x) = (k+1)x^2/k + (k+1)(x-1)^2$. Trivially $f(w_k^n / w_{k+1}^n) \geqslant 1$ and so $P(k) \Rightarrow P(k+1)$. Since $P(1)$ is true by inspection, $P(n)$ is true proving the lemma. □

If $\beta > 0$ we define the function $\chi(\beta, y, z)$ to be the characteristic function of the set $\{ z \in C_0[0,T] \mid \|y - z\| \leqslant \beta \}$. The following lemma is crucial in the proof of Theorem 1.

Lemma 14: Let $F(z)$ be as given in Theorem 1 and let $\delta > 0$. then, for sufficiently small λ,

$$\begin{aligned}I(\lambda) &= \mathbb{E}_z^T\{(1 - \chi(\delta/\lambda, x_0/\lambda, z)) \exp(\lambda^{-2} F(\lambda z))\} \\ &= O(\exp(\alpha \lambda^{-2})),\end{aligned}$$

for some $\alpha < 0$.

Proof: Although we will proceed to prove this lemma in a manner similar to Schilder's, it can be obtained as a special case of an abstract theorem of Varadhan.[19]

Since $x_0(\tau)$ is continuous on $[0,T]$ we have

$$\lim_{n \to \infty} \|x_0 - x_0^n(\cdot)\| = 0$$

and, from Lemma 9, we have

$$\lim_{n \to \infty} x_0^n A_n x_0^n \leqslant \int_0^t \left| \frac{dx_0}{d\tau}(\tau) \right|^2 d\tau.$$

Therefore both sequence $\{ \|x_0^n(\cdot)\| \}$ and $\{ x_0^n A_n x_0^n \}$ are bounded. Hence there exists a positive constant c sufficiently large so that, for all n,

$$L_1/c + 2L_2 \|x_0^n\| (T/c)^{1/2} + L_2 \|x_0^n\|^2/c + (x_0^n A_n x_0^n / c)^{1/2} < \tfrac{1}{8}$$

and, simultaneously,

$$-A(\delta)/c < \tfrac{1}{8}.$$

From the continuity assumptions on $F(z)$ we can find an $\eta < \delta/4$ such that, for $\|z\| < a$, where $a = [8T(L_1 + 1)]^{1/2} / |2^{-1/4} - 1|$, and $\|z - y\| \leqslant \eta$

$$F(z) - F(y) \leqslant D = -(1 - 2^{-1/4})^2 \left(\frac{\delta}{2}\right)^2 A\left(\frac{\delta}{2}\right) \Big/ 4Tc > 0.$$

Finally we choose n so large that

$$L_1 - n(1 - 2^{-1/4})^2 \eta^2 / 32T < -1,$$

that, for $n = m$, Lemma 7 is applicable for $\delta' = \eta/\lambda$, and so large that $\rho_n = \|x_0 - x_0^n(\cdot)\| < \delta/8$. In what follows we keep these choices of n, η, and c fixed.

Observe that $I(\lambda) = I_2(\lambda) + I_3(\lambda) + I_4(\lambda)$, where

$$\begin{aligned}I_2(\lambda) = \mathbb{E}_z^T\{&[1 - \chi(\delta/\lambda, x_0/\lambda, z)][1 - H(\eta/\lambda, n, z)] \\ &\times \exp\{\lambda^{-2} F(\lambda z)\}\}\end{aligned}$$

$$I_3(\lambda) = \mathbb{E}_z^T\{[1-\chi(\delta/\lambda,x_0/\lambda,z)]H(\eta/\lambda,n,z)\chi(a/\lambda,0,z) \times \exp\{\lambda^{-2}F(\lambda z)\}\},$$

$$I_4(\lambda) = \mathbb{E}_z^T\{[1-\chi(\delta/\lambda,x_0/\lambda,z)]H(\eta/\lambda,n,z) \times [1-\chi(a/\lambda,0,z)]\exp\{\lambda^{-2}F(\lambda z)\}\},$$

where $H(\eta/\lambda,n,z)$ is the characteristic function of $\{z\in C_0[0,T] \mid \|z-z^n(\cdot)\|\leqslant \eta/\lambda\}$. We show that $I_2(\lambda)$, $I_3(\lambda)$, and $I_4(\lambda)$ are each $O(\exp\{\alpha\lambda^{-2}\})$, for some $\alpha<0$.

First consider $I_2(\lambda)$. Clearly,

$$I_2(\lambda)\leqslant \mathbb{E}_z^T\{(1-H(\eta/\lambda,n,z))\exp\{\lambda^{-2}F(\lambda z)\}\}$$

and using the Cauchy–Schwarz inequality

$$I_2(\lambda)\leqslant \mathbb{E}_z^T\{(1-H(\eta/\lambda,n,z))\}^{1/2}\mathbb{E}_z^T\{\exp\{2\lambda^{-2}F(\lambda z)\}\}^{1/2}.$$

From Lemma 7, for an absolute constant K_2,

$$\mathbb{E}_z^T\{1-H(\eta/\lambda,n,z))\} \leqslant K_2\lambda/\eta\exp\{-n(1-2^{-1/4})^2\eta^2/16\lambda^2 T\},$$

and, using the known bounds on F,

$$\mathbb{E}_z^T\{\exp\{2\lambda^{-2}F(\lambda z)\}\} \leqslant \mathbb{E}_z^T\{\exp\{2\lambda^{-2}L_1+2L_2\|z\|^2\}\} = \exp\{2\lambda^{-2}L_1\}\mathbb{E}_z^T\{\exp\{2L_2\|z\|^2\}\}.$$

From Lemma 8, $\mathbb{E}_z^T\{\exp\{2L_2\|z\|^2\}\}<\infty$, if $2L_2<\gamma/T$, which it is by the hypotheses of Theorem 1. Hence, for constants K_2', K_2'', for sufficiently small λ,

$$|I_2(\lambda)|\leqslant K_2'(\lambda/\eta)^{1/2}\exp\{\lambda^{-2}(L_1-n(1-2^{-1/4})^2\eta^2/32T)\} \leqslant K_2''\exp\{-\lambda^{-2}\}.$$

We now consider $I_3(\lambda)$,

$$I_3(\lambda) = \mathbb{E}_z^T\{(1-\chi(\delta/\lambda,x_0/\lambda,z))H(\eta/\lambda,n,z)\chi(a/\lambda,0,z) \times\exp\{\lambda^{-2}F(\lambda z)\}\}.$$

In the integral above the integration takes place over a subset of $\{z\in C_0[0,T] \mid \|\lambda z-\lambda z^n(\cdot)\|\leqslant\eta \text{ and } \|\lambda z\|\leqslant a\}$ and hence by the choice of η it follows that

$$F(\lambda z)\leqslant F(\lambda z^n(\cdot))+D.$$

Therefore

$$I_3(\lambda)\leqslant \mathbb{E}_z^T\{(1-\chi(\delta/\lambda,x_0/\lambda,z))H(\eta/\lambda,n,z) \times\exp\{\lambda^{-2}(F(\lambda z^n(\cdot))+D)\}\}.$$

Now

$$\|\lambda z-\lambda z^n(\cdot)\|<\eta,$$

$$\|\lambda z-x_0\|>\delta \Rightarrow \|\lambda z^n(\cdot)-x_0\|>(\delta-\eta)$$

and, thus,

$$I_3(\lambda)\leqslant \mathbb{E}_z^T\{[1-\chi((\delta-\eta)/\lambda,x_0/\lambda,z^n(\cdot))] \times\exp\{\lambda^{-2}(F(\lambda z^n(\cdot))+D)\}\}.$$

It follows that $I_3(\lambda)$ can be bounded by an $(n-1)$ dimensional Lebesgue integral, viz.,

$$I_3(\lambda)\leqslant(2\pi T)^{1/2}\left(\frac{2\pi T}{n}\right)^{-n/2}\exp\{D\lambda^{-2}\} \times\int_{\mathbb{R}^{n-1}}\{[1-\chi((\delta-\eta)/\lambda,x_0/\lambda,z^n(\cdot))] \times\exp\{\lambda^{-2}F(\lambda z^n(\cdot))-\tfrac{1}{2}z^nA_nz^n\}\}\Big|_{z_n=0}d^{n-1}z,$$

where $d^{n-1}z=dz_1\cdots dz_{n-1}$, $z_1=z(T/n)$, $z_2=z(2T/n)$, etc.

Setting $z^n=\lambda^{-1}(w^n+x_0^n)$; $z_n=w_n=x_n=0$; gives

$$I_3(\lambda)\leqslant\lambda^{-(n-1)}(2\pi T)^{1/2}\left(\frac{2\pi T}{n}\right)^{-n/2}\exp\{D\lambda^{-2}\} \times\int_{\mathbb{R}^{n-1}}[1-\chi(\delta-\eta,x_0,w^n(\cdot)+x_0^n(\cdot))] \times\exp\{\lambda^{-2}F(w^n(\cdot)+x_0^n(\cdot)) -\tfrac{1}{2}\lambda^{-2}(w^n+x_0^n)A_n(w^n+x_0^n)\}d^{n-1}w.$$

Since n was picked to ensure that $\rho_n=\|x_0-x_0^n(\cdot)\|<\delta/8$ and $\eta<\delta/4$, it follows that $\delta-\eta-\rho_n>\delta-\eta-2\rho_n>0$. Clearly $\|w^n(\cdot)\|<(\delta-\eta-\rho_n)$ implies $\|w^n(\cdot)+x_0^n(\cdot)-x_0\|<(\delta-\eta)$ and so

$$\operatorname{supp}[1-\chi(\delta-\eta,x_0,w^n(\cdot)+x_0^n(\cdot))] \subset\operatorname{supp}[1-\chi(\delta-\eta-\rho_n,0,w^n(\cdot))].$$

Therefore

$$I_3(\lambda)\leqslant\lambda^{-(n-1)}(2\pi T)^{1/2}\left(\frac{2\pi T}{n}\right)^{-n/2}\exp\{D\lambda^{-2}\} \times\int_{\|w^n\|>\delta-\eta-\rho_n}\exp\{\lambda^{-2}w^nA_nw^nJ\}d^{n-1}w,$$

where

$$J=\frac{[F(w^n(\cdot)+x_0^n(\cdot))-\tfrac{1}{2}(w^n+x_0^n)A_n(w^n+x_0^n)]}{w^nA_nw^n}.$$

The last integral is taken over those $w^n\in\mathbb{R}^{n-1}$ [recall that $w^n=(w_1,w_2,\ldots,w_{n-1},0)$] having supnorm greater than $(\delta-\eta-\rho_n)$.

By considering separately the cases $w^nA_nw^n<c$, $w^nA_nw^n\geqslant c$, we prove that $J\leqslant A(\delta-\eta-2\rho_n)/c$. For the case $w^nA_nw^n<c$, we can deduce from Lemma 12 that

$$F(w^n(\cdot)+x_0^n(\cdot))-\tfrac{1}{2}(w^n+x_0^n)A_n(w^n+x_0^n) \leqslant A(\delta-\eta-2\rho_n).$$

Since by hypothesis $c>w^nA_nw^n>0$ and we know $A(\cdot)$ is negative, we deduce

$$J\leqslant A(\delta-\eta-2\rho_n)/w^nA_nw^n\leqslant A(\delta-\eta-2\rho_n)/c.$$

For the case $w^nA_nw^n\geqslant c$, we deduce from the hypotheses on F, that

$$\begin{aligned}J&\leqslant[L_1+L_2\|w^n(\cdot)+x_0^n(\cdot)\|^2 -\tfrac{1}{2}(w^n+x_0^n)A_n(w^n+x_0^n)]/w^nA_nw^n\\ &=[L_1+L_2\|w^n+x_0^n\|^2 -\tfrac{1}{2}(w^n+x_0^n)A_n(w^n+x_0^n)]/w^nA_nw^n\\ &\leqslant[L_1+L_2\|w^n\|^2+2L_2\|w^n\|\|x_0^n\|+L_2\|x_0^n\|^2 -\tfrac{1}{2}w^nA_nw^n-w^nA_nx_0^n-\tfrac{1}{2}x_0^nA_nx_0^n]/w^nA_nw^n\\ &\leqslant L_1/w^nA_nw^n+L_2T+2L_2\|x_0^n\|(T/w^nA_nw^n)^{1/2} +L_2\|x_0^n\|/w^nA_nw^n-\tfrac{1}{2}-w^nA_nx_0^n/w^nA_nw^n,\end{aligned}$$

where in the last step we are using Lemma 13 and $-\tfrac{1}{2}x_0^nA_nx_0^n\leqslant 0$. Using now the hypotheses $L_2T<\tfrac{1}{4}$, $w^nA_nw^n\geqslant c>0$, and the Cauchy–Schwarz inequality $w^nA_nx_0^n\leqslant[(w^nA_nw^n)(x_0^nA_nx_0^n)]^{1/2}$, we obtain

$$J\leqslant-\tfrac{1}{4}+L_1/c+2L_2\|x_0^n\|(T/c)^{1/2}+L_2\|x_0^n\|^2/c +(x_0^nA_nx_0^n/c)^{1/2}.$$

From our choice of c then it follows that

$$J \leqslant -\tfrac{1}{8} \leqslant A(\delta)/c \leqslant A(\delta - \eta - 2\rho_n)/c,$$

proving that $J \leqslant A(\delta - \eta - 2\rho_n)/c$, whatever the value of $(w^n A_n w^n)$.

We have now proved that

$$I_3(\lambda) \leqslant \lambda^{-(n-1)}(2\pi T)^{1/2}\left(\frac{2\pi T}{n}\right)^{-n/2} \exp\{D\lambda^{-2}\} \times \int_{\|w^n\| > \delta - \eta - \rho_n} \times \exp\{\lambda^{-2} w^n A_n w^n A(\delta - \eta - 2\rho_n)/c\} d^{n-1}w.$$

We now set $v = [-2A(\delta - \eta - \rho_n)/c]^{1/2}\mu^n = \lambda^{-1} v w^n$, giving the inequality

$$I_3(\lambda) \leqslant v^{-(n-1)}(2\pi T)^{1/2}\left(\frac{2\pi T}{n}\right)^{-n/2} \exp\{D\lambda^{-2}\} \times \int_{\|\mu^n\| > (\delta - \eta - \rho_n)\lambda^{-1} v} \exp\{-\tfrac{1}{2}\mu^n A_n \mu^n\} d^{n-1}\mu,$$

or

$$I_3(\lambda) \leqslant v^{-(n-1)} \exp\{D\lambda^{-2}\} \times \mathbb{E}_z^T\{[1 - \chi((\delta - \eta - \rho_n)\lambda^{-1} v, 0, z^n(\cdot))]\}.$$

However, since $\|z\| \geqslant \|z^n(\cdot)\|$,

$$\operatorname{supp}[1 - \chi((\delta - \eta - \rho_n)\lambda^{-1} v, 0, z^n(\cdot))] \subset \operatorname{supp}[1 - \chi((\delta - \eta - \rho_n)\lambda^{-1} v, 0, z)]$$

and so

$$I_3(\lambda) \leqslant v^{-(n-1)} \exp\{D\lambda^{-2}\} \times \mathbb{E}_z^T\{[1 - \chi(\delta - \eta - \rho_n)\lambda^{-1} v, 0, z]\}.$$

We now use Lemma 5 to deduce, for an absolute constant K_3, for sufficiently small λ,

$$\mathbb{E}_z^T\{[1 - \chi(\delta - \eta - \rho_n) v/\lambda, 0, z)]\} \leqslant K_3 \exp\{-(1 - 2^{-1/4})^2(\delta - \eta - \rho_n)^2 v^2/4\lambda^2 T\}.$$

Finally then, since $\eta < \delta/4$, $\rho_n < \delta/8$ and $A(\cdot)$ is a negative decreasing function, for K_3' an absolute constant

$$\begin{aligned} I_3(\lambda) &\leqslant K_3' \exp\{\lambda^{-2}[D + (1 - 2^{-1/4})^2(\delta - \eta - \rho_n)^2 \times A(\delta - \eta - 2\rho_n)/2Tc]\} \\ &\leqslant K_3' \exp\{\lambda^{-2}[D + (1 - 2^{-1/4})^2(\delta/2)^2 A(\delta/2)/2Tc]\} \\ &= K_3' \exp\{\lambda^{-2}(1 - 2^{-1/4})^2(\delta/2)^2 A(\delta/2)/4Tc\} \end{aligned}$$

It remains for us to consider $I_4(\lambda)$. This term is relatively easy to deal with. Evidently

$$\begin{aligned} I_4(\lambda) &= \mathbb{E}_z^T\{[1 - \chi(\delta/\lambda, x_0/\lambda, z)] H(\eta/\lambda, n, z) \times [1 - \chi(a/\lambda, 0, z)] \exp\{\lambda^{-2} F(\lambda z)\}\} \\ &\leqslant \mathbb{E}_z^T\{[1 - \chi(\delta/\lambda, 0, z)] \exp\{L_1\lambda^{-2} + L_2\|z\|^2\}\} \end{aligned}$$

and by the Cauchy–Schwarz inequality

$$I_4(\lambda) \leqslant \exp\{L_1\lambda^{-2}\}[\mathbb{E}_z^T\{[1 - \chi(a/\lambda, 0, z)]\}]^{1/2} \times [\mathbb{E}_z^T\{\exp\{2L_2\|z\|^2\}\}]^{1/2}.$$

Using Lemma 5 again we obtain for a constant K_4

$$\mathbb{E}_z^T\{[1 - \chi(a/\lambda, 0, z)]\} \leqslant K_4 \exp\{-(1 - 2^{-1/4})^2 a^2/4T\lambda^2\}$$

and, since the second integral is convergent for $L_2 < \gamma/2T$, by the choice of a

$$\begin{aligned} I_4(\lambda) &\leqslant K_4' \exp\{\lambda^{-2}(L_1 - (1 - 2^{-1/4})^2 a^2/8T)\} \\ &= K_4' \exp\{-\lambda^{-2}\}, \end{aligned}$$

K_4' being an absolute constant. This proves Lemma 14 with

$$\alpha = \max\{-1, (1 - 2^{-1/4})^2(\delta/2)^2 A(\delta/2)/4Tc\} < 0. \quad \square$$

Lemma 15 (Schilder): If $x_0'(\tau)$ is of bounded variation on $[0,T]$ and if

$$\int_0^T f(\tau) y(\tau)\, d\tau - \int_0^T x_0'(\tau)\, dy(\tau) = 0,$$

$$f \in L^2[0,T], \quad \text{for all } y \in C_0^*[0,T],$$

then

$$\int_0^T f(\tau) y(\tau)\, d\tau - \int_0^T x_0'(\tau)\, dy(\tau) = 0,$$

for a.e. $y \in C_0[0,T]$.

Proof: If $x_0'(\tau)$ is of bounded absolute variation, imitating a theorem of Nelson,[20] for a.e. $y \in C_0[0,T]$,

$$\int_0^T x_0'(\tau)\, dy(\tau) = -\int_0^T y(\tau) dx_0'(\tau).$$

Hence, for a.e., $y \in C_0[0,T]$,

$$\left[\int_0^T f(\tau) y(\tau)\, d\tau - \int_0^t x_0'(\tau)\, dy(\tau)\right]$$

is realizable as a continuous linear functional, which is zero on the dense subset $C_0^*[0,T]$. $\square$

B. Proof of Theorem 1

For brevity in what follows let $f_j(\eta) z^j$ denote $D^j F(x_0 + \eta)(z,z,\ldots,z)/j!$, where $D^j F$ is the jth Frechet derivative F at the point $(x_0 + \eta)$, the bracket containing j arguments, $j = 0,1,2,\cdots$.

Choose $\delta > 0$ such that the hypotheses on F hold. Let $\mathbb{E}_z^T\{\exp\{\lambda^{-2} F(\lambda z)\}\} = h_1(\lambda) + h_2(\lambda)$, where

$$h_1(\lambda) = \mathbb{E}_z^T\{\chi(\delta/\lambda, x_0/\lambda, z) \exp\{\lambda^{-2} F(\lambda z)\}\},$$

$$h_2(\lambda) = \mathbb{E}_z^T\{[1 - \chi(\delta/\lambda, x_0/\lambda, z)] \exp\{\lambda^{-2} F(\lambda z)\}\}.$$

From Lemma 14 we deduce that $h_2(\lambda) = O(\exp\{\alpha\lambda^{-2}\})$, where $\alpha < 0$, and so for any integer n, $h_2(\lambda) = O(\lambda^{n-2})$.

Using a Cameron–Martin type translation,[21] we arrive at

$$h_1(\lambda) = \exp\left\{-\tfrac{1}{2}\lambda^{-2}\int_0^t [x_0'(\tau)]^2\, d\tau\right\} \times \mathbb{E}_y^T\left\{\chi(\delta/\lambda, 0, y) \exp\left\{-\lambda^{-1}\int_0^T x_0'(\tau) dy(\tau) + \lambda^{-2} F(\lambda y + x_0)\right\}\right\}.$$

From Taylor's theorem for functionals we may write

$$\begin{aligned} F(\lambda y + x_0) &= F(x_0) + \lambda DF(x_0) y + \lambda^2 D^2 F(x_0)(y,y)/2 + k_3(\lambda y) \\ &= f_0(0) + \lambda f_1(0) y + \lambda^2 f_2(0) y^2 + k_3(\lambda y), \end{aligned}$$

where $|k_3(\lambda y)| = O(\lambda^3)$, if $\|\lambda y\| \leqslant \delta$.

$$\therefore h_1(\lambda) = \exp\left\{ -\tfrac{1}{2}\lambda^{-2} \int_0^T [x_0'(\tau)]^2 \, d\tau + \lambda^{-2} f_0(0) \right\}$$
$$\times \mathbb{E}_y^T \left\{ \chi(\delta/\lambda, 0, y) \right.$$
$$\times \exp\left\{ \lambda^{-1} \left(f_1(0) y - \int_0^t x_0'(\tau) dy(\tau) \right) \right\}$$
$$\left. \times \exp\{ f_2(0) y^2 + \lambda^{-2} k_3(\lambda y) \} \right\}.$$

By hypothesis

$$\left(F(z) - \frac{1}{2} \int_0^T [z'(\tau)]^2 \, d\tau \right)$$

has a maximum of 0 at x_0 over $C_0^*[0,T]$, so it and its first Frechet derivative are zero at x_0. Thus,

$$f_1(0) y - \int_0^T x_0'(\tau) \, dy(\tau) = 0,$$

for $y \in C_0^*[0,T]$, and so by Lemma 15

$$h_1(\lambda) = \mathbb{E}_z^T \{ \chi(\delta/\lambda, 0, z) \exp\{ f_2(0) z^2 + \lambda^{-2} k_3(\lambda z) \} \}.$$

The Taylor expansion of $\exp(x)$ reads

$$\exp(x) = \sum_{i=0}^{n-1} x^i / i! + R_n(x),$$

where

$$|R_n|(x) \leqslant \begin{cases} x^n / n! \exp(x), & \text{if } x \geqslant 0, \\ |x|^n / n!, & \text{if } x < 0. \end{cases}$$

We may now write $h_1(\lambda)$ in the form

$$h_1(\lambda) = \sum_{i=0}^{n-3} (1/i!) \, \mathbb{E}_z^T \{ \chi(\delta/\lambda, 0, z)$$
$$\times \exp\{ f_2(0) z^2 \} [\lambda^{-2} k_3(\lambda z)]^i \} + J_{n-2}(\lambda),$$

where denoting the characteristic function of the set $\{ z \in C_0[0,T] \mid k_3(\lambda z) \geqslant 0 \}$ by $B(\lambda, z)$

$$|J_{n-2}(\lambda)| \leqslant 1/(n-2)! \mathbb{E}_z^T \{ \chi(\delta/\lambda, 0, z) |\lambda^{-2} k_3(\lambda z)|^{n-2}$$
$$\times \exp\{ f_2(0) z^2 + \lambda^{-2} k_3(\lambda z) \} B(\lambda, z) \}$$
$$+ 1/(n-2)! \mathbb{E}_z^T \{ \chi(\delta/\lambda, 0, z) |\lambda^{-2} k_3(\lambda z)|^{n-2}$$
$$\times \exp\{ f_2(0) z^2 \} [1 - B(\lambda, z)] \}.$$

From Taylor's theorem for functionals it follows that, if $\|\lambda y\| < \delta$, then

$$\lambda^2 f_2(0) y^2 + k_3(\lambda y) = k_2(\lambda y) = \lambda^2 f_2(\eta) y^2, \eta \in C_0[0,T]$$

with $\|\eta\| < \delta$, where by hypothesis $|k_3(\lambda y)| \leqslant C_3 \lambda^3 \| y \|^3$, C_3 being a constant. Thus,

$$|J_{n-2}(\lambda)| \leqslant 1/(n-2)! \mathbb{E}_z^T \{ \chi(\delta/\lambda, 0, z) (C_3 \lambda)^{n-2} \|z\|^{3(n-2)}$$
$$\times \exp\{ f_2(\eta) z^2 \} B(\lambda, z) \}$$
$$+ 1/(n-2)! \mathbb{E}_z^T \{ \chi(\delta/\lambda, 0, z) (C_3 \lambda)^{n-2} \|z\|^{3(n-2)}$$
$$\times \exp\{ f_2(0) z^2 \} [1 - B(\lambda, z)] \}$$

By using the Cauchy–Schwarz inequality (or Hölder's inequality) and condition 5 of Theorem 1 we have that $J_{n-2}(\lambda) = O(\lambda^{n-2})$.

We have now proved that

$$h_1(\lambda) = \sum_{i=0}^{n-3} (1/i!) \mathbb{E}_z^T \{ \chi(\delta/\lambda, 0, z)$$
$$\times \exp\{ f_2(0) z^2 \} [\lambda^{-2} k_3(\lambda z)]^i \} + O(\lambda^{n-2}).$$

However, $k_3(\lambda z) = \lambda^3 f_3(0) z^3 + \cdots + \lambda^{n-1} f_{n-1}(0) z^{n-1} + k_n(\lambda z)$, where $\lambda^{-2} k_n(\lambda z) = O(\lambda^{n-2} \|z\|^n)$, for $\|\lambda z\| < \delta$, expanding by the binomial theorem, therefore gives, using condition 5 of Theorem 1, and Hölder's inequality

$$h_1(\lambda) = \sum_{i=0}^{n-3} (1/i!) \mathbb{E}_z^T \{ \chi(\delta/\lambda, 0, z)$$
$$\times \exp\{ f_2(0) z^2 \} [\lambda f_3(0) z^3 + \cdots$$
$$+ \lambda^{n-3} f_{n-1}(0) z^{n-1}]^i \} + O(\lambda^{n-2}).$$

It can be seen from the Hölder inequality, Lemmas 5 and 8, and condition 5 of Theorem 1, that, for sufficiently small λ,

$$\sum_{i=0}^{n-3} (1/i!) \mathbb{E}_z^T \{ [1 - \chi(\delta/\lambda, 0, z)] \exp\{ f_2(0) z^2 \} [\lambda f_3(0) z^3 + \cdots$$
$$\cdots \lambda^{n-3} f_{n-1}(0) z^{n-1}]^i \}$$
$$= O(P(\lambda) \exp\{ \beta \lambda^{-2} \}) = O(\lambda^{n-2}),$$

where P is a polynomial and β is a negative constant. Replacing χ by $[1 - (1 - \chi)]$, we finally obtain

$$h_1(\lambda) = \sum_{i=0}^{n-3} (1/i!) \mathbb{E}_z^T \{ \exp\{ f_2(0) z^2 \}$$
$$\times [\lambda f_3(0) z^3 + \cdots \lambda^{n-3} f_{n-1}(0) z^{n-1}]^i \} + O(\lambda^{n-2}),$$

so that

$$h_1(\lambda) = \Gamma_0 + \Gamma_1 \lambda + \cdots + \Gamma_{n-3} \lambda^{n-3} + O(\lambda^{n-2}),$$

where the Γ_i are dependent only on the Frechet derivatives of F at x_0, for $i = 1,2,\ldots,(n-3)$. This completes the proof of the theorem.

ACKNOWLEDGMENTS

We are grateful to David Elworthy and Ken Brown for helpful conversations. One of us (I.D.) is grateful to the SRC for a research studentship. One of us (A.T.) is indebted to Monroe Donsker for reawakening his interest in asymptotic analysis of functional integrals at the Les Houches Workshop on stochastic differential equations in theoretical physics, June 1980.

[1] M. Mizrahi, J. Math. Phys. **20**, 844 (1979).
[2] C. De Witt-Morette, Ann. Phys. **97**, 367 (1976).
[3] A. Truman, "The polygonal path formulation of the Feynman path integral," in *Feynman Path Integrals*, edited by S. Albeverio *et al. Springer Lecture Notes in Physics*, Vol. 106 (Springer, Berlin, 1978), p. 73.
[4] M. Schilder, Trans. Am. Math. Soc. **125**, 63 (1965).
[5] B. Simon, *Functional Integration and Quantum Physics* (Academic, New York, 1979).
[6] H. H. Kuo, *Gaussian Measures in Banach Spaces, Springer Lecture Notes in Mathematics*, Vol. 463 (Springer, Berlin, 1975), see p. 112.
[7] I. Davies and A. Truman, "On Laplace Expansions of conditional Wiener Integrals and Bender–Wu formulae for the x^{2N}-anharmonic oscillator" (to appear in J. Math. Phys.).
[8] M. Pincus, Trans. Am. Math. Soc. **134**, 193 (1968).
[9] R. S. Ellis and J. R. Rosen, Bull. Am. Math. Soc. **3** (1), 705 (1980); "Asymptotic analysis of Gaussian integrals I: Isolated minimal points," to appear in Trans. Am. Math. Soc.; "Asymptotic analysis of Gaussian integrals, II: Manifold of minimal points," to appear in Commun. Math. Phys.; see also H. Doss, Ann. Inst. Henri Poincaré B **16** (1), 17 (1980).
[10] The *non sequitur* occurs in deducing the second displayed inequality from the first displayed inequality on p. 75. The deduction is valid for $b < 0$ but it appears not to be valid for all $b > 0$.
[11] N. I. Akhiezer, *The Calculus of Variations* (Blaisdell, New York, London, 1962), see Chap. 4.

[12]See Ref. 5.
[13]See Ref. 6, p. 113.
[14]E. Nelson, *Dynamical Theories of Brownian Motion* (Princeton U.P., Princeton, 1967).
[15]J. L. Doob, Ann. Math. Stat. **20**, 393 (1949).
[16]See Chap. 1 of Ref. 5.
[17]K. D. Elworthy and A. Truman, "Feynman maps, Cameron–Martin formulae and the anharmonic oscillator" (in preparation).
[18]R. Carmona and B. Simon, "Pointwise Bounds on eigenfunctions and wavepackets in N-body quantum systems, V: lower bounds and path integrals," to be submitted to Commun. Math. Phys.
[19]S. R. S. Varadhan, Commun. Pure Appl. Math. **19**, 261 (1966); see Theorem 3.5 and also Sec. 6 of M. D. Donsker and S. R. S. Varadhan, Commun. Pure Appl. Math. **29**, 389 (1976).
[20]See Ref. 14, p. 40.
[21]See Ref. 6.

Moving frames and prolongation algebras [a)]

Frank B. Estabrook
Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California 91109

(Received 8 December 1981; accepted for publication 12 February 1982)

We consider differential ideals generated by sets of 2-forms which can be written with constant coefficients in a canonical basis of 1-forms. By setting up a Cartan–Ehresmann connection, in a fiber bundle over a base space in which the 2-forms live, one finds an incomplete Lie algebra of vector fields in the fibers. Conversely, *given* this algebra (a *prolongation* algebra), one can derive the differential ideal. The two constructs are thus dual, and analysis of either derives properties of both. Such systems arise in the classical differential geometry of moving frames. Examples of this are discussed, together with examples arising more recently: the Korteweg–de Vries and Harrison–Ernst systems.

PACS numbers: 02.40. + m

I. INTRODUCTION

When Cartan's method of moving frames is applied to classical problems of differential geometry, a quite specific type of differential ideal is often presented for analysis and solution.[1] These ideals are generated by sets of 2-forms, and have a canonical structure inasmuch as they are expressed in an (anholonomic) basis of 1-forms in which all their terms have *constant coefficients*. Such ideals are, in fact, derived by specializing the closure relations that are fulfilled by the left or right invariant 1-forms in the spaces of Lie groups. They can be analyzed systematically for invariances, conservation laws, Bäcklund correspondences, etc., and their integral submanifolds classified by computation of Cartan's local algebraic characters.[2] Depending on how dependent and independent sets of sets of variables are introduced, they give rise to various elegant sets of coupled nonlinear partial differential equations. The sine-Gordon equation is only the most famous of these.

It seems that the sets of nonlinear evolution equations more recently treated with techniques such as inverse scattering and prolongation, in which auxiliary variables are introduced satisfying linear "connection" equations, often can be derived from differential ideals *which are of the same canonical type*. The prolongation structures that result are indeed already known to involve incomplete Lie algebras; these "prolongation algebras" are dual to (and in a certain sense, as we will see, more general than) the canonically expressed ideals.

In the following we illustrate this bringing together of classical and modern problems by discussing some famous examples from a unified standpoint. First, the classical differential geometric problem of immersing a surface of constant negative curvature in Euclidean 3-space has been elegantly expounded by Chern and Terng,[3] as an introduction to some nice generalizations of this problem to affine and higher dimensional spaces: We discuss the analysis of the canonical ideals that arise in the simplest case, staying close to their notation (but not being concerned with any constructions based on ideas of Euclidean metric or parallel transport). It is instructive how prolongation algebras, Bäcklund transformation, and permutation theorems arise.

Second, we give some new results based on the prolongation algebra of the Korteweg–de Vries equation.[4] A canonical ideal for this system is given, in an anholonomic basis in a seven dimensional group space, and a new expression of the Bäcklund transformation results.

As a final example, we briefly discuss Harrison's ideal for the Ernst equation[5] and give its prolongation algebra.

II. IMMERSION IDEALS

We consider first a six parameter Lie group that is built upon the 3-parameter rotation group O^3, and that in fact expresses application of moving orthogonal frames to Euclidean space. But for the most part, we will refrain from such interpretations and simply take the group, and others related to it, as given for analysis. In terms of six basis 1-forms $\theta^i,\omega^i, i = 1,2,3$, the set of 2-forms

$$I: \begin{aligned} &d\theta^1 + \omega^2\wedge\theta^3 - \omega^3\wedge\theta^2,\\ &d\theta^2 + \omega^3\wedge\theta^1 - \omega^1\wedge\theta^3,\\ &d\theta^3 + \omega^1\wedge\theta^2 - \omega^2\wedge\theta^1,\\ &d\omega^1 + \omega^2\wedge\omega^3,\\ &d\omega^2 + \omega^3\wedge\omega^1,\\ &d\omega^3 + \omega^1\wedge\omega^2, \end{aligned} \tag{1}$$

generates a closed differential ideal, I. If, in a space of six dimensions, the forms in Eq. (1) were to vanish identically, the basis 1-forms θ^i,ω^i could be called left-invariant, and the space (locally, at least) identified with the group space. A set of canonical (± 1 and 0) structure constants for the group can be read off from Eq. (1). The Jacobi identies for these follow from the closure property.

A dual vector (Lie) algebra to Eq. (1) is found by prolongation.[4,6–8] Denoting the six basis forms collectively as η^μ ($\eta^1 = \theta^1$, etc., $\eta^4 = \omega^1$, etc.), spanning a base space, we introduce an unspecified number of fiber coordinates y^A, ($A = 1,\ldots$), and an equal number of connection 1-forms

$$dy^A - X^A_\mu(y^B)\,\eta^\mu\,. \tag{2}$$

The X^A_μ are introduced as functions only of y^B, and the product form of the second term in Eq. (2) is characteristic of *principal* connections.[8] By requiring that the exterior derivatives of the connection forms in Eq. (2) be in the ideal gener-

[a)]Research carried out under Contract NAS7-100, sponsored by the National Aeronautics and Space Administration.

 0022-2488/82/112071-06$02.50 © 1982 American Institute of Physics

ated together by Eq. (1) and Eq. (2), a set of partial differential equations in commutator form is found for the functions X^A_μ; these can be written as Lie brackets of vector fields in the fibers, the vector fields being $X_\mu = X^A_\mu\, \partial/\partial y^A$. For the present case we set $X_i = X^A_i(\partial/\partial y^A)$, $i = 1,2,3$, and $U_1 = X^A_4(\partial/\partial y^A)$, $U_2 = X^A_5(\partial/\partial y^A)$, $U_3 = X^A_6(\partial/\partial y^A)$, and find that they must satisfy the algebra

$$A: \quad \begin{aligned} &[X_i, X_j] = \epsilon_{ijk} X_k, \\ &[X_i, U_j] = \epsilon_{ijk} U_k, \\ &[U_i, U_j] = 0. \end{aligned} \tag{3}$$

ϵ_{ijk} is the permutation symbol, and repeated indices denote summation.

Conversely, given the explicit six dimensional Lie algebra A, Eq. (3), one can introduce forms η^μ (that is, θ^i, ω^i), postulate the connection Eq. (2), and derive the set of curvature forms I, Eq. (1). This duality construction is more general than simply using an inner product $X_\mu \cdot \eta^\nu = \delta^\nu_\mu$, as is customary in Lie group theory, for in that case the X_μ are vectors in the same space as the η^μ. This duality construction also generalizes nicely to incomplete Lie algebras, as we will see.

By putting identically equal to zero one or more basis forms in the set of curvature forms of a group, one immediately obtains members of an interesting class of nonlinear systems. Such systems are (1) generated by a 2-forms having (in a certain anholonomic basis) constant coefficients, and are (2) closed under exterior differentiation. Let us call these canonical systems.

We illustrate this with the simple example of putting $\theta^3 = 0$ in Eq. (1). Let us call this the immersion ideal I_I (as defined by a set of generators):

$$I_I: \quad \begin{aligned} &d\omega^1 + \omega^2 \wedge \omega^3, \\ &d\omega^2 + \omega^3 \wedge \omega^1, \\ &d\omega^3 + \omega^1 \wedge \omega^2, \\ &d\theta^1 - \omega^3 \wedge \theta^2, \\ &d\theta^2 + \omega^3 \wedge \theta^1, \\ &\omega^1 \wedge \theta^2 - \omega^2 \wedge \theta^1. \end{aligned} \tag{4}$$

We analyze I_I by noting that it is possible to introduce two auxiliary variables f and g (pseudopotentials) so that there is a closed subset of 2-forms involving exterior derivatives of the five original basis 1-forms. (For anholonomic basis forms such a subset replaces the usually unwritten identities $ddx^i = 0$ for natural basis forms dx^i). The complete enlarged ideal is

$$I_{I'}: \quad \begin{aligned} &d\theta^1 - f(\omega^1 \wedge \theta^2 - \omega^2 \wedge \theta^1) - \omega^3 \wedge \theta^2, \\ &d\theta^2 - g(\omega^1 \wedge \theta^2 - \omega^2 \wedge \theta^1) + \omega^3 \wedge \theta^1, \\ &d\omega^1 + \omega^2 \wedge \omega^3, \\ &d\omega^2 + \omega^3 \wedge \omega^1, \\ &d\omega^3 + \omega^1 \wedge \omega^2, \\ &df - fg\omega^1 + (1+f^2)\omega^2 - g\omega^3, \\ &dg - (1+g^2)\omega^1 + fg\omega^2 + f\omega^3, \\ &\omega^1 \wedge \theta^2 - \omega^2 \wedge \theta^1. \end{aligned} \tag{5}$$

Cartan's characteristic integers[9,10] then follow from considering as generators in a space of seven dimensions just the two 1-forms and the single algebraic 2-form: $s_0 = 2$, $s_1 = 1$, $s_2 = 1$, $g = 3$. The genus, g, of $I_{I'}$, or I_I, is thus 3: this is the number of independent variables in a general solution [or the dimension of a maximal integral submanifold on which all the forms in Eq. (4) pull back to zero]. There is, however, also a Cauchy characteristic vector, which must lie in the maximal integral manifolds: it could be denoted ∂/ω^3, meaning that its contractions with the 1-forms ω^1, ω^2, θ^1, θ^2 vanish and its contraction with ω^3 is 1. So with proper choice of similarity variables the number of independent coordinates in a resulting set of partial differential equations can be just two. This is not surprising to those who remember the Gauss and Codazzi equations resulting from I_I when intrinsic coordinates are used in immersed Riemannian surfaces.

From Eq. (2) and Eq. (4) we find the following prolongation algebra:

$$A_I: \quad \begin{aligned} &[Z, Y] = X, \\ &[X, Z] = Y, \\ &[Y, X] = Z, \\ &[V, Z] = -U, \\ &[U, Z] = V, \\ &[V, Y] = 0, \\ &[U, X] = 0, \\ &[U, V] = 0, \\ &[U, Y] + [V, X] = 0. \end{aligned} \tag{6}$$

This is an incomplete 5-dimensional algebra, dual (and abstractly equivalent) to (4). One explores such algebras by systematically introducing new vectors and using the Jacobi relation. In this case the definition

$$[V, X] = W$$

leads directly to the *closed* Lie algebra Eq. (3). In this way, if I_I were first given, we would find that it can be derived from I by setting a basis form to zero.

Lastly, consider the immersion ideal for surfaces of constant negative curvature,[3] I_N:

$$I_N: \quad \begin{aligned} &d\omega^1 + \omega^2 \wedge \omega^3, \\ &d\omega^2 + \omega^3 \wedge \omega^1, \\ &d\omega^3 + \omega^1 \wedge \omega^2, \\ &d\theta^1 - \omega^3 \wedge \theta^2, \\ &d\theta^2 + \omega^3 \wedge \theta^1, \\ &\omega^1 \wedge \theta^2 - \omega^2 \wedge \theta^1, \\ &\omega^1 \wedge \omega^2 + \theta^1 \wedge \theta^2. \end{aligned} \tag{7}$$

One easily checks that this is closed, and thus is what we have called a canonical system. The Cartan characters, in seven dimensions (again prolonging with f and g) are $s_0 = 2$, $s_1 = 2$, $s_2 = 0$, $g = 3$. It does not, however, appear that I_N can be derived from a finite dimensional Lie group by equating basis forms to zero. In fact, the dual prolongation algebra to I_N is

$$A_N: \quad \begin{aligned} &[X, Z] = Y, \quad [Z, Y] = X, \\ &[U, Z] = V, \quad [Z, V] = U, \\ &[U, X] = 0, \quad [V, Y] = 0, \\ &[U, V] - [X, Y] = Z, \\ &[V, X] + [U, Y] = 0. \end{aligned} \tag{8}$$

By introducing more vectors by definition, and using Jacobi relations, enlarged algebras are found, but closure of the algebra A_N, Eq. (8), in a finite number of such steps does not appear to occur. Dual sets of forms to such enlarged algebras can always be reduced to I_N by setting the additional basis forms to zero, but they still will include purely algebraic 2-forms. Perhaps the canonical differential systems we are dealing with can all be regarded as derived from the closure relations of infinite dimensional Lie groups, by setting all but a finite number of the invariant basis forms to zero, but this remains to be shown. [Peter Gragert (private communication) has found several finite Lie algebras which arise from "closing" A_N with additional relations. For example, one such is $[Y,[Y,U]]=0$; in that case one derives a nine-dimensional algebra, and one finds the ideal I_N generated (or specialized) in a nine-dimensional group space by four (basis)1-forms and two algebraic 2-forms. $s_0=4, s_1=2, g=3$].

III. THE BÄCKLUND CORRESPONDENCE

The next step in the classic analysis is to observe that, because I_N already has a Cauchy characteristic vector, a prolongation 1-form can be added to I_N without introducing any additional prolongation variable. The 1-form used by Chern and Terng is

$$\Psi \equiv \theta^2 + \sin\tau\,\omega^3 + \cos\tau\,\omega^2, \tag{9}$$

where τ is an arbitrary parameter. Let us denote this augmented ideal as I_B (B for Bäcklund):

$$I_B:\quad \begin{aligned} &d\omega^1+\omega^2\wedge\omega^3,\\ &d\omega^2+\omega^3\wedge\omega^1,\\ &d\omega^3+\omega^1\wedge\omega^2,\\ &d\theta^1-\omega^3\wedge\theta^2,\\ &d\theta^2+\omega^3\wedge\theta^1,\\ &\theta^2+\sin\tau\,\omega^3+\cos\tau\,\omega^2,\\ &\omega^1\wedge\theta^2-\omega^2\wedge\theta^1,\\ &\omega^1\wedge\omega^2+\theta^1\wedge\theta^2. \end{aligned} \tag{10}$$

Now (in five dimensions) $s_0=1, s_1=2, g=2$. We assert this since if, as before, we include also variables f, g and the two 1-forms involving them, and call the completed augmented ideal in seven dimensions $I_{B'}$, then for it $s_0=3, s_1=2$, $g=2$.

The vector field ∂/ω^3 is a subcharacteristic[7] of $I_{B'}$, belonging to the closed subideal $I_{N'}$. Remarkably, there is a second, belonging to an isomorphic subideal, which we next write. There will be a Bäcklund correspondence between solutions of these subideals exactly as occurs for the KdV ideal treated in Ref. 7.

The isomorphic subideal emerges upon introduction of new basis forms according to the transformation

$$\begin{aligned} \theta^{1*} &= \theta^1 + \sin\tau f^*\Psi,\\ \theta^{2*} &= -\omega^2 + \Psi^*,\\ \omega^{1*} &= \omega^1,\\ \omega^{2*} &= \cos\tau\,\omega^2 + \sin\tau\,\omega^3,\\ \omega^{3*} &= -\sin\tau\,\omega^2 + \cos\tau\,\omega^3,\\ f^* &= f/(\cos\tau - g\sin\tau),\\ g^* &= (\sin\tau + g\cos\tau)/(\cos\tau - g\sin\tau), \end{aligned} \tag{11}$$

where

$$\Psi^* = f^*\Psi/f. \tag{12}$$

The first of these relations are in Chern and Terng—relating forms at different locations in E^3. Here we rather interpret them as a local change of basis in a 7-space.

It is now readily computed that an algebraically equivalent set of generators for $I_{B'}$, in terms of these new generators, is

$$I_{B'}:\quad \begin{aligned} &d\omega^{1*}+\omega^{2*}\wedge\omega^{3*},\\ &d\omega^{2*}+\omega^{3*}\wedge\omega^{1*},\\ &d\omega^{3*}+\omega^{1*}\wedge\omega^{2*},\\ &d\theta^{1*}-f^*(\omega^{1*}\wedge\theta^{2*}-\omega^{2*}\wedge\theta^{1*})-\omega^{3*}\wedge\theta^{2*},\\ &d\theta^{2*}-g^*(\omega^{1*}\wedge\theta^{2*}-\omega^{2*}\wedge\theta^{1*})+\omega^{3*}\wedge\theta^{1*},\\ &df^*-f^*g^*\omega^{1*}+(1+f^{*2})\omega^{2*}-g^*\omega^{3*},\\ &dg^*-(1+g^{*2})\omega^{1*}+f^*g^*\omega^{2*}+f^*\omega^{3*},\\ &\omega^{1*}\wedge\theta^{2*}-\omega^{2*}\wedge\theta^{1*},\\ &\omega^{1*}\wedge\omega^{2*}+\theta^{1*}\wedge\theta^{2*},\\ &\Psi^*\equiv\theta^{2*}-\sin\tau\,\omega^{3*}+\cos\tau\,\omega^{2*}. \end{aligned} \tag{13}$$

The first nine forms in $I_{B'}$ in Eq. (13) are identical in form to those in $I_{N'}$ as first introduced: they may be written starred or unstarred, and transform into each other through Eqs. (11) and (12)! The 1-forms Ψ and Ψ^* are related by a scalar factor, Eq. (12), and are a connection form. The last 2-forms in Eq. (13) are a closed subideal, in the ideal of Ψ and the last two 2-forms in Eq. (10), and conversely. The second subcharacteristic is ∂/ω^{3*}; written in the unstarred frame it is

$$\sin^2\tau f\frac{\partial}{\theta^1} - \sin\tau\frac{\partial}{\omega^2} + \cos\tau\frac{\partial}{\omega^3} - \sin\tau(\cos\tau - g\sin\tau)\frac{\partial}{\theta^2}. \tag{14}$$

IV. SINE-GORDON EQUATION

Knowing that the search for two-dimensional integral submanifolds of I_B (or $I_{B'}$) is a well-posed problem, one next explores the resulting partial differential equations. Above we have taken the view that the first five forms in I_B (or the first seven in $I_{B'}$) can be set identically equal to zero, so defining a basis set in a space, and then we considered the remaining algebraic forms as fields there. Now it is appropriate to take precisely the opposite tack, and introduce "intrinsic" variables. We choose an algebraically degenerate coordinate representation of the basis forms in terms of two to-be-independent variables u^1 and u^2, writing the autonomous basis forms du^1 and du^2, such that the *second* set of forms in I_B is identically zero, and the p.d.e's arise from annulling the first set.

Since, *inter alia*, Ψ (or Ψ^*) is now taken identically zero, we can use the Bäcklund correspondence of Eq. (11) of Sec. III to introduce the to-be-dependent variables symmetrically while ensuring that the last two algebraic 2-forms also vanish identically. It is not difficult to arrive at something like the following ansatz although, of course, the particular dependent variables and notation that we adopt, α and ψ, conform to Ref. 3.

$$\theta^1 = \theta^{1*} = \cos\psi\cos\alpha du^1 + \sin\psi\sin\alpha du^2,$$
$$-\omega^{2*} = \theta^2 = -\cos\psi\sin\alpha du^1 + \sin\psi\cos\alpha du^2,$$
$$-\theta^{2*} = \omega^2 = -\sin\psi\cos\alpha du^1 + \cos\psi\sin\alpha du^2,$$
$$\omega^{1*} = \omega^1 = -\sin\psi\sin\alpha du^1 - \cos\psi\cos\alpha du^2, \quad (15)$$
$$\sin\tau\omega^3 = (\cos\psi\sin\alpha + \cos\tau\sin\psi\cos\alpha)du^1 - (\sin\psi\cos\alpha + \cos\tau\cos\psi\sin\alpha)du^2,$$
$$\sin\tau\omega^{3*} = (\sin\psi\cos\alpha + \cos\tau\cos\psi\sin\alpha)du^1 - (\cos\psi\sin\alpha + \cos\tau\sin\psi\cos\alpha)du^2.$$

Substituting these into the first five 2-forms in I_B gives just two forms,

$$(d\psi - d\alpha)\wedge(du^1 + du^2) + \frac{1+\cos\tau}{\sin\tau}\sin(\psi+\alpha)du^1\wedge du^2,$$

and (16)

$$(d\psi + d\alpha)\wedge(du^1 - du^2) - \frac{1-\cos\tau}{\sin\tau}\sin(\psi-\alpha)du^1\wedge du^2,$$

from which sine-Gordon equations for ψ and α immediately follow. A final remark: If the forms in Eq. (16) are taken by themselves as a rather elegant ideal for the s-G equation, the characters in 4-space are $s_0 = 0, s_1 = 2, g = 2$. The Bäcklund transformation can be seen as an involutory inversion: either $\theta^1,\theta^2,\omega^1,\omega^2 \to \theta^1, -\omega^2,\omega^1, -\theta^2$, or in terms of coordinates $\alpha,\psi,u^1,u^2 \to \pi-\psi,\pi-\alpha,u^1,u^2$.

V. THE KdV ALGEBRA

Although to one using moving frames in E^3 the above local manipulations of already familiar relations may seem superfluous, it is my contention that these methods are of much more widespread applicability, in contexts far removed from those of metric geometry, in fact whenever canonical ideals of 2-forms arise. In Ref. 4 the following prolongation algebra was derived, beginning with the Kortweg–de Vries equation in its usual coordinate form (the sign of X_7 has been changed from Ref. 4):

$$\begin{aligned}
&[X_1,X_2] = X_7, \quad [X_1,X_7] = -X_5, \quad [X_2,X_7] = -X_6,\\
&[X_1,X_3] = [X_2,X_3] = [X_5,X_3] = [X_6,X_3] = [X_7,X_3] = [X_1,X_4] = [X_2,X_6] = 0\\
A_{KV}:\ &[X_2,X_5] - [X_1,X_6] = 0, \quad [X_6,X_7] - [X_2,X_7] = 0,\\
&[X_1,X_5] + [X_2,X_4] = 0, \quad [X_1,X_6] + [X_5,X_6] = 0,\\
&[X_2,X_1] + [X_3,X_4] + [X_1,X_6] = 0.
\end{aligned}$$

From this incomplete seven vector algebra, and Eq. (2), we find the following canonical (set of generators of an) ideal, in terms of seven basis 1-forms η^i, $i = 1...7$:

$$\begin{aligned}
&\Omega^1 \equiv d\eta^1,\\
&\Omega^2 \equiv d\eta^2,\\
&\Omega^3 \equiv d\eta^3,\\
&\Omega^4 \equiv d\eta^4,\\
&\Omega^5 \equiv d\eta^5 - \eta^1\wedge\eta^7 + \eta^5\wedge\eta^7,\\
&\Omega^6 \equiv d\eta^6 - \eta^2\wedge\eta^7 - \eta^6\wedge\eta^7,\\
I_{KV}:\ &\Omega^7 \equiv d\eta^7 + \eta^1\wedge\eta^2 + \eta^1\wedge\eta^6 + \eta^2\wedge\eta^5 - \eta^5\wedge\eta^6,\\
&\Sigma^1 \equiv \eta^1\wedge\eta^6 + \eta^2\wedge\eta^5 - \eta^3\wedge\eta^4 - \eta^5\wedge\eta^6,\\
&\Sigma^2 \equiv \eta^1\wedge\eta^5 - \eta^2\wedge\eta^4,\\
&\Sigma^3 \equiv \eta^4\wedge\eta^5,\\
&\Sigma^4 \equiv \eta^4\wedge\eta^6,\\
&\Sigma^5 \equiv \eta^4\wedge\eta^7,\\
&\Sigma^6 \equiv \eta^5\wedge\eta^7. \qquad (18)
\end{aligned}$$

We denote the first seven, $\Omega^1\cdots\Omega^7$, *curvature* forms, inasmuch as they are a closed set in themselves. No auxiliary variables $f,g,\ldots$ need be sought as already I_{KV} can be interpreted as a set of algebraic generators $\Sigma^1\cdots\Sigma^6$ in a group space of seven dimensions, spanned by left invariant 1-forms η^i. The Cartan characters are $s_0 = 0, s_1 = 5, g = 2$. Two-dimensional integral submanifolds are maximal.

To find the KdV equation we now, as above, search for a degenerate expression of the basis forms η^i in terms of two to-be-independent variables, say dx and dt, and 14 to-be-dependent variables u^i,v^i: that is, we write $\eta^i = u^i dx + v^i dt$, and then require the u^i,v^i to be related so that $\Sigma^1,\ldots,\Sigma^6$ vanish identically. It then turns out that in the Ω^i either η^1 or η^4 occur in each quadratic term, so a further allowable specialization is to take $\eta^1 \sim dx$ and $\eta^4 \sim dt$. We now are reduced to seven "intrinsic" variables, and assign these without further loss of generality to conform with previous notation (u,z,p,w,v) for the KdV equation:

$$\begin{aligned}
\eta^1 &= -2dx,\\
\eta^2 &= -2udx + 2(p+6u^2)dt,\\
\eta^3 &= -2(v+u^2)dx - wdt,\\
\eta^4 &= -8dt, \qquad (19)\\
\eta^5 &= -8udt,\\
\eta^6 &= -8vdt,\\
\eta^7 &= 4zdt.
\end{aligned}$$

It can be checked that $\Sigma^1\cdots\Sigma^6$ vanishes identically.

Five coupled partial differential equations result from annuling the curvature forms Ω^i:

$$\begin{aligned}
&u_x = z,\\
&z_x = p + 2u^2 - 4v,\\
&v_x = uz, \qquad (20)\\
&w_x - 2(v+u^2)_t = 0,\\
&(p+6u^2)_x + u_t = 0.
\end{aligned}$$

Other choices of variables will give very different equations. η^2 and η^4 also occur such that they can be specialized to dx and dt— this would presumably be a sort of hodograph transformation of the system. The essential point is that all of this is contained in the algebra A_{KV}, Eq. (17).

VI. BÄCKLUND TRANSFORMATION OF THE KdV ALGEBRA

Equation (2) for the KdV system has a 1-dimensional solution

$$dy - \eta^2 - \eta^6 + y\eta^7 - \tfrac{1}{2}y^2(\eta^1 - \eta^5)\,. \tag{21}$$

Specializing this with Eq. (19) one finds the modified KdV for y. Another way of introducing the simple pseudopotential y is to search for one-dimensional representations of A_{KV} by setting $\tilde{X}_1 = X_1(y)(\partial/\partial y)$, etc. The fact that a solution exists shows that SL(2,R), the only nontrivial group on a line, is a homomorphic image of A_{KV}:

$$\begin{aligned}
\tilde{X}_1 &= \tfrac{1}{2}y^2\frac{\partial}{\partial y},\\
\tilde{X}_2 &= \frac{\partial}{\partial y},\\
\tilde{X}_3 &= 0,\\
\tilde{X}_4 &= 0,\\
\tilde{X}_5 &= -\tfrac{1}{2}y^2\frac{\partial}{\partial y},\\
\tilde{X}_6 &= \frac{\partial}{\partial y},\\
\tilde{X}_7 &= -y\frac{\partial}{\partial y}.
\end{aligned} \tag{22}$$

The Bäcklund correspondence previously derived[4,11] takes the general form

$$\begin{aligned}
\eta^{1*} &= -\eta^1,\\
\eta^{2*} &= y^2\eta^1 + \eta^2 - y^2\eta^5 + 2\eta^6 - 2y\eta^7,\\
\eta^{3*} &= -\tfrac{1}{2}y^4\eta^1 - 3y^2\eta^2 - \eta^3 + \tfrac{1}{2}y^4\eta^5 - 3y^2\eta^6 + 3y^3\eta^7,\\
\eta^{4*} &= -\eta^4,\\
\eta^{5*} &= y^2\eta^4 + \eta^5,\\
\eta^{6*} &= -\tfrac{1}{2}y^4\eta^4 - y^2\eta^5 - \eta^6,\\
\eta^{7*} &= -y^3\eta^4 - 2y\eta^5 - \eta^7.
\end{aligned} \tag{23}$$

The ideal I_{KV}, Eq. (18), when augmented with the connection form Eq. (21), is invariant under this linear transformation of basis forms.

The algebra A_{KV}, Eq. (17), is invariant under the inverse linear transformation, which in this case turns out remarkably to be just the transpose (the transformation is involutive!):

$$\begin{aligned}
X_1^* &= -X_1 + y^2X_2 - \tfrac{1}{2}y^4X^3,\\
X_2^* &= X_2 - 3y^2X_3,\\
X_3^* &= -X_3,\\
X_4^* &= -X_4 + y^2X_5 - \tfrac{1}{2}y^4X_6 - y^3X_7,\\
X_5^* &= -y^2X_2 + \tfrac{1}{2}y^4X_3 + X_5 - y^2X_6 - 2yX_7,\\
X_6^* &= 2X_2 - 3y^2X_3 - X_6,\\
X_7^* &= -2yX_2 + 3y^3X_3 - X_7.
\end{aligned} \tag{24}$$

Now one regards y as a function in the space of the X's, with directional derivatives that can be read off from Eq. (22).

There are also two invariance transformations of the prolongation algebra A_{KV}, or of I_{KV}, that have constant factors, and that express scale and Galilean invariance; they can be used to introduce the well-known "eigenvalue" parameter in Eq. (23) or Eq. (24). (These isomorphisms of I_{KV}, or A_{KV}, have independently been found by Shadwick.[4])

VII. THE HARRISON–ERNST SYSTEM

B. K. Harrison[5] has shown that solutions of the Ernst and related equations of general relativity are two-dimensional integral submanifolds of the following ideal:

$$I_{HE}:\quad \begin{aligned}
&d\xi_1 - \xi_1\wedge\xi_4,\\
&d\xi_2 - \xi_2\wedge\xi_5 + \xi_2\wedge\xi_4,\\
&d\xi_3 - \xi_3\wedge\xi_5 + \xi_1\wedge\xi_2,\\
&d\xi_4,\\
&d\xi_5,\\
&d\xi_6 + \xi_5\wedge\xi_6,\\
&\xi_1\wedge\xi_3 + \xi_2\wedge\xi_4,\\
&\xi_2\wedge\xi_5 + \xi_1\wedge\xi_6,\\
&\xi_3\wedge\xi_5 + \xi_4\wedge\xi_6,\\
&\xi_2\wedge\xi_3 + \lambda\xi_1\wedge\xi_4,\\
&\xi_1\wedge\xi_5 + \lambda\xi_2\wedge\xi_6,\\
&\xi_4\wedge\xi_5 + \lambda\xi_3\wedge\xi_6,\quad \lambda \pm 1.
\end{aligned} \tag{25}$$

For the Ernst equation, $\lambda = -1$. This is closed, and indeed of our canonical form. The first six 2-forms are closed by themselves, so we may take $n = 6$. From the rest we calculate $s_0 = 0, s_1 = 4, g = 2$. Using Eq. (2), we find the following prolongation algebra:

$$A_{HE}:\quad \begin{aligned}
&[X_1,X_2] + X_3 = 0,\\
&[X_1,X_3] - [X_2,X_4] - X_2 = 0,\\
&[X_1,X_5] - \lambda\,[X_2,X_6] = 0,\\
&\lambda\,[X_2,X_3] - [X_1,X_4] + X_1 = 0,\\
&[X_1,X_6] - [X_2,X_5] + X_2 = 0,\\
&[X_3,X_5] - [X_4,X_6] - X_3 = 0,\\
&[X_3,X_4] = 0,\\
&[X_4,X_5] - \lambda\,[X_3,X_6] = 0,\\
&[X_5,X_6] + X_6 = 0.
\end{aligned} \tag{26}$$

This algebra has also been derived by Harrison.[12] We present no further analysis here, but remark that Harrison and others have recently found Bäcklund transformations and other elegant properties for this system. We offer in closing the speculation that such elegant properties are always properties of nonlinear systems that can be expressed as incomplete Lie algebras.

[1] L. E. Evtushik, Yu. G. Lumiste, N. M. Ostianov, and A. P. Shirokov, "Differential-Geometric Structures on Manifolds," J. Soviet Math. **14**, 6 (1980).

[2] É. Cartan, *Les Systèmes Différentiels Extérieurs et Leurs Applications Géométriques* (Hermann et Cie, Paris, 1945).

[3] S.-S. Chern and C.-L. Terng, "An Analogue of Bäcklund's Theorem in Affine Differential Geometry," Rocky Mountain J. Math. **10**, 105 (1980).

[4] H. D. Wahlquist and F. B. Estabrook, "Prolongation Structures of Nonlinear Evolution Equations," J. Math. Phys. **16**, 1 (1975); W. F. Shadwick,

"The KdV Prolongation Algebra," *ibid.* **21**, 454 (1980).
[5]B. K. Harrison, "Bäcklund Transformation for the Ernst Equation of General Relativity," Phys. Rev. Lett. **41**, 1197 (1978).
[6]R. Hermann, "The Pseudopotentials of Estabrook and Wahlquist, the Geometry of Solitons, and the Theory of Connections," Phys. Rev. Lett. **36**, 835 (1976).
[7]F. B. Estabrook and H. D. Wahlquist, "Prolongation Structures, Connection Theory and Bäcklund Transformation," in *Nonlinear Evolution Equations Solvable by the Spectral Transform*, edited by F. Calogero (Pittman, London, 1978).
[8]F. A. E. Pirani, D. C. Robinson, and W. F. Shadwick, *Local Jet Bundle Formulation of Bäcklund Transformations* (Reidel, Boston, 1979).
[9]F. B. Estabrook, "Some Old and New Techniques for the Practical Use of Exterior Differential Forms," in *Bäcklund Transformation, the Inverse Scattering Method, Solitons and their Applications, Lecture Notes in Mathematics*, Vol. 515, edited by R. Miura (Springer, Berlin, 1976).
[10]F. B. Estabrook, "Differential Geometry as a Tool for Applied Mathematicians," in *Geometrical Approaches to Differential Equations, Lecture Notes in Mathematics*, Vol. 810, edited by R. Martini (Springer, Berlin, 1980).
[11]H. D. Wahlquist and F. B. Estabrook, "Bäcklund Transformation for Solutions of the Korteweg-de Vries Equation," Phys. Rev. Lett. **31**, 1386–90 (1973).
[12]B. K. Harrison, Proc. Utah Acad. Sci. Arts Lett. **53**, Part 1, 67 (1976).

The differential geometric structure of general mechanical systems from the Lagrangian point of view[a)]

Robert Hermann
Association for Physical and Systems Mathematics, Inc., Brookline, Massachusetts 02146

(Received 24 March 1981; accepted for publication 21 April 1982)

Several differential-geometric points of view on analytical mechanics of systems with a finite number of degrees of freedom are developed in generality, emphasizing Cartan's calculus of differential forms and Ehresmann's theory of jet spaces. The classical theory of Lagrange's equations *with external forces and constraints* ("holonomic" or "nonholonomic") is put into an invariant and coordinatefree form. The relation between this "Lagrangian" and the "Hamiltonian-symplectic" approach, which is that most extensively used in the contemporary mathematical physics literature, is also developed.

PACS numbers: 02.40. + m, 03.20. + i, 02.30.Wd

1. INTRODUCTION

One of the interesting mathematical developments of the last 20 years has been the revival within the context of professional mathematics of the classical subject of *analytical mechanics*. This has been "geometrized" and some classical problems have been profitably examined in a new light. The treatises by Abraham and Marsden[1] and Arnold[2] are now the standard references for this material. Perhaps most important, the links between analytical mechanics and quantum mechanics have been made more precise.

However, this work has only penetrated into the traditional problems of *particle mechanics*. Not as much has been done with the much more complicated problems of general mechanical systems that are encountered in the applied parts of the physical sciences and engineering. There has as yet been little work geometrizing in the same spirit the so-called *nonholonomic* mechanical systems. My aim here is to show how a contemporary version of Elie Cartan's methods[3] (particularly utilizing the work of Ehresmann[4]) may be used to describe such mechanical systems. The modern geometric point of view also suggests certain further extension and application of the work on mechanics that was developed in the framework of classical tensor analysis, especially the work of Gabriel Kron[5] and a group in Japan led by Kondo[6] called the Research Association for Applied Geometry (RAAG). Some of the material presented here has been touched on in my own work.[7,8]

Another motivation is to revitalize the *Lagrangian* (i.e., the calculus of variations) point of view. In the mathematical work cited above[1,2] it has been found convenient to emphasize the *Hamiltonian* picture, whereas the classical treatises[9,11] are primarily Lagrangian. (Differential geometrically, the Hamiltonian involves the *cotangent bundle*, the Lagrangian the *tangent bundle*.) No doubt, this is because of the role that the cotangent bundle (with its "symplectic structure") plays in quantum mechanics and the theory of linear partial differential equations. Now, in working elementary mechanics problems (and handling constraints) it is much easier to use "velocities" rather than "momenta," i.e., the Lagrangian rather than the Hamiltonian point of view. Further, Lagrangian methods are much simpler for introducing certain types of *interactions* between systems. (This is why Lagrangian methods still dominate in elementary particle physics.) Thus, I intend this paper to serve as a foundational treatment of the Lagrangian point of view, with further applications to be developed in later work. Among these applications, those arising from recent work in control theory[12–15] on feedback linearization of nonlinear mechanical systems, should be especially noted.

Since the methods of the theory of Pfaffian systems are basic to this approach to mechanics, and are not at all well known to physical scientists, I begin with a brief explanation of some of the basic concepts, assuming that the reader knows the fundamentals of "calculus on manifolds," i.e., the theory of vector fields (together with associated Lie group theory) and differential forms on finite dimensional C^∞ paracompact manifolds.

2. PFAFFIAN SYSTEMS

The notation of calculus and manifolds are those of Ref. 7. All manifolds, maps, and geometric data will be C^∞, finite dimensional, and paracompact, unless mentioned otherwise. Let X be such a manifold. Here is the notation we shall use for the basic objects of the calculus of manifolds used in this work.

$\mathscr{F}(X)$ = algebra of C^∞, real-valued functions on X.

For $x \in X$, X_x = tangent vector space to X at x.

X_x^d = dual space to X_x
= space of one-covectors at x.

$T(X) = \{(x,v) : x \in X,\ V \in X_x\}$
= tangent vector bundle.

$T^d(X) = \{(x,\theta) : x \in X,\ \theta \in X_x^d\}$
= cotangent bundle.

$\mathscr{V}(X) = C^\infty$ cross sections of $T(X)$
= derivations of $\mathscr{F}(X)$.

[a)] Supported by Ames Research Center (NASA), Grant NSG-2402; U. S. Army Research Office, Contract #ILI61102BH57-05 MATH; NSF MCS8003227.

 0022-2488/82/112077-13$02.50 © 1982 American Institute of Physics

$\mathscr{D}^1(X)$ = cross sections of $T^d(X)$
= dual $\mathscr{F}(X)$ module to $\mathscr{V}(X)$.
≡ space of *one-differential*
or *Pfaffian* forms.

$\mathscr{D}^n(X)$ = exterior differential forms of degree n.

$d:\mathscr{D}^n(X) \to \mathscr{D}^{n+1}(X)$, *exterior derivative*

$(V_1, V_2) \to [V_1, V_2] = V_1V_2 - V_2V_1$, an R-linear map

$[\,,\,]:\mathscr{V}(X)\times\mathscr{V}(X) \to \mathscr{V}(X)$, called *Jacobi bracket.*

If θ is a differential form and V is a vector field, then

$$\mathscr{L}_V(\theta)$$

denotes the *Lie derivative* of θ by V.

$G^m(X_x)$ = Grassmann manifold of m-dimensional linear subspaces of the vector space X_x, $m = 0,1,2,\ldots$.

$G^m(T(X))$ = the fiber bundle over X whose fiber over $x \in X$ is $G^m(X_x)$.

$G^m(T^d(X))$ = bundle of m-dimensional linear subspaces of the cotangent vector spaces.

If $E \to X$ is a fiber space over X, let $\Gamma(E)$ denote the space of C^∞ cross section maps: $X \to E$.

Definition: A *Pfaffian system*[16–26] (of dimension m) on X is a cross-section map

$$\gamma: X \to G^m(T(X)),$$

i.e., an element of $\Gamma[G^m(T(X))]$.

Given such a cross section $\gamma \in \Gamma[G^m(T(X))]$, one can define the "dual" object.

$\gamma^d \in \Gamma[G^{n-m}(T^d(X))]$,

$\gamma^d(X)$ = space of $\theta \in X_x^d$ such that $\theta(\gamma(x)) = 0$
= annihilator of $\gamma(x)$ in the dual space.

γ^d is actually the object Cartan would call a "Pfaffian system," since he preferred to work with differential forms rather than vector fields. However, we will call either γ or γ^d a "Pfaffian system." (When we consider such systems *with singularities*, it is necessary to distinguish one from its dual.) Notice that γ^d is an $\mathscr{F}(X)$ submodule of $\mathscr{D}^1(X)$. It is *locally free* in the sense that each point has a neighborhood in which the module has a basis. Often, it is convenient to use quasialgebraic methods and notation when dealing with these objects, particularly when the methods of the theory of *sheaves* and *schemes* are to be utilized. (However, in this work we shall not use these methods.) It will often be convenient notationally to identify Pfaffian systems with such modules. For this purpose, we shall often use the notation $\mathscr{P}$ for the module of differential forms and $\mathscr{V}$ for the dual module of vector fields.

3. THE DERIVED SYSTEMS OF A REGULAR PFAFFIAN SYSTEM

Keep the notation of Sec. 2. Let

$$\gamma \in \Gamma[G^n(T(M))]$$

be a Pfaffian system interpreted geometrically as a cross section of the Grassmann bundle of m-dimensional linear subspaces of the tangent vector bundle to X.

Set

$\mathscr{V}(\gamma)$ = the set of vector fields $V \in \mathscr{V}(X)$ such that

$$V(x) \in \gamma(x) \text{ for all } x \in X. \tag{3.1}$$

Thus, the orbit curves of the Pfaffian system are the curves in X that are orbit curves of *some* vector field on $\mathscr{V}(\gamma)$. [Note that $\mathscr{V}(\gamma)$ is not a Lie subalgebra of $\mathscr{V}(X)$, unless the Pfaffian system is completely integrable in the Frobenius sense, i.e., defines a foliation.]

Set

$$\mathscr{V}^1(\gamma) = \mathscr{V}(\gamma) + [\mathscr{V}(\gamma), \mathscr{V}(\gamma)]. \tag{3.2}$$

For $x \in X$, set

$$\gamma^1(x) = \mathscr{V}^1(\gamma)(x). \tag{3.3}$$

Thus, $\gamma^1(x)$ is a linear subspace of the tangent space that contains $\gamma(x)$. As x varies, we get a family $x \to \gamma^1(x)$ of tangent spaces. Let us say that γ is *one-regular* if the dimension of these spaces is *constant*. In this case, γ^1 defines another Pfaffian system, which is called (by Cartan) the *first derived system*.

Similarly, set

$$\mathscr{V}^2(\gamma) = \mathscr{V}(\gamma) + [\mathscr{V}(\gamma), \mathscr{V}(\gamma)] + [\mathscr{V}(\gamma), [\mathscr{V}(\gamma), \mathscr{V}(\gamma)]], \tag{3.4}$$

$$\gamma^2(x) = \mathscr{V}^2(\gamma)(x). \tag{3.5}$$

Let us say that γ is two-regular if it is one-regular and if

$$\dim \gamma^2(x)$$

is constant as x ranges over X. In this case, γ^2 defines a Pfaffian system called the *second derived system*. Continue in this way to define the nth derived system. The original system γ is said to be *regular* if all its derived systems are regular. In this paper, we shall consider only regular systems. (One can prove that, in general, there is always an open subset of X on which the system is regular. If the system is real analytic, this subset is also dense in S, and its complement is contained in analytic varieties of lower dimension.)

In many of Cartan's papers of Pfaffian systems, the study of the properties of the derived system is the key feature. We now turn to the study of the *structure tensors*, which live on the vector bundles associated with the derived systems.

4. THE STRUCTURE TENSOR OF A REGULAR PFAFFIAN SYSTEM

Continue with γ as a regular Pfaffian system on a manifold X, considered as a cross section of the Grassmann bundle. Let $\gamma^1, \gamma^2, \ldots$ denote the derived systems. For $x \in X$, we have an increasing sequence of tangent subspaces:

$$\gamma(x) \subset \gamma^1(x) \subset \gamma^2(x) \subset \cdots .$$

Then $\gamma^1, \gamma^2, \ldots$ also define vector bundles over X. These

are the pullback of the standard vector bundles over the Grassmann manifold. For convenience, we will make no notational distinction between the γ's as cross sections of the Grassmann bundles and as vector bundles over X.

Set

$$E^1(x) = \gamma^1(x)/\gamma(x),$$
$$E^2(x) = \gamma^2(x)/\gamma^1(x), \tag{4.1}$$
$$\vdots$$

Each $E^1(x)$, $E^2(x)$,... is a linear vector space. As x varies, they define *vector bundles* E^1, E^2,... over X. They play a basic role in the study of the structure of Pfaffian systems.

We shall now define *structure tensors* $\tau^1,\tau^2,...$, which are tensor fields associated with the vector bundles defined by $\gamma^1,\gamma^2,...$ and E^1, E^2... .

To define τ^1, again work at a point x of X. Pick V_1, $V_2 \in \mathscr{V}(\gamma)$. Then,

$$[V_1, V_2](x)$$

is a tangent vector to X, which lies in $\gamma^1(x)$. Consider its projection mod $\gamma(x)$, i.e., as a vector in $E^1(x)$. We obtain a skew-symmetric map

$$(V_1, V_2) \to \tau^1(V_1, V_2).$$

Note now that τ only depends on the values of V_1 and V_2 at x, not on their derivatives. In this way, τ defines a skew-symmetric, bilinear map

$$\tau^1(x):\gamma^1(x)\times\gamma^1(x) \to E^1(x).$$

Explicitly,

$$\tau^1(x)(V_1(x), V_2(x)) \equiv [V_1, V_2](x) \bmod \gamma(x). \tag{4.2}$$

As x varies, we obtain a bilinear vector bundle map

$$\tau^1:\gamma\times\gamma \to E^1. \tag{4.3}$$

This is called the *first integrability tensor*. Notice that, by the very definition of E^1, τ is onto, hence it is zero if and only if the Pfaffian system with which we began is Frobenius integrable.

We can now continue. For V_1, V_2, $V_3 \in \mathscr{V}(\gamma)$, consider the triple commutator:

$$[V_1[V_2, V_3]](x).$$

It lies in $\gamma^2(x)$. Its nontensorial component can be eliminated by projecting mod $\gamma^1(x)$. We thus obtain a trilinear map

$$\tau^2(x):\gamma(x)\times\gamma(x)\times\gamma(x) \to \gamma^2(x)/\gamma(x) \equiv E^2(x).$$

Now, as x varies, we obtain a tensor field τ^2 as a trilinear bundle map

$$\gamma\times\gamma\times\gamma \to E^2.$$

This procedure can obviously be iterated to obtain nth degree *structure tensors*

$$\tau^n:\gamma\times\cdots\times\gamma \to E^n.$$

We will now discuss in what sense these tensors are "invariants" for the equivalence problem.

5. THE STRUCTURE TENSORS OF A PFAFFIAN SYSTEM AS INVARIANTS FOR THE EQUIVALENCE PROBLEM

Let

$$\gamma: X \to G^m(T(X)),$$
$$\gamma': X' \to G^m(T(X')),$$

define Pfaffian systems (of the same dimension) on manifolds X and X'.

Definition: γ and γ' are *equivalent* if there is a diffeomorphism

$$\phi: X \to X'$$

such that

$$\gamma'(\phi(x)) = \phi_*(\gamma(x)), \quad \text{for all } x \in X. \tag{5.1}$$

In words, the natural actions of ϕ on the tangent bundle and on the associated Grassmann bundles intertwine the cross-section maps γ and γ'.

Of course, this general definition is not how Cartan and Lie would think of it! Cartan, for example, would deal locally with "moving frames," i.e., locally defined vectors of independent one-forms X and X':

$$\omega = \begin{pmatrix} \omega_1 \\ \vdots \\ \omega_{n-m} \end{pmatrix}, \tag{5.2}$$

$$\omega' = \begin{pmatrix} \omega'_1 \\ \vdots \\ \omega'_{n-m} \end{pmatrix} \tag{5.3}$$

such that, for $x \in X$, $x' \in X'$,

$$\gamma(x) = \{v \in X_x : \omega(v) = 0\}, \tag{5.4}$$
$$\gamma'(x') = \{v' \in X'_x : \omega'(v') = 0\}. \tag{5.5}$$

Relation (5.1) now means that there is a $(n-m)\times(n-m)$ matrix M of functions on M such that

$$\omega = M\phi^*(\omega'). \tag{5.6}$$

If ϕ satisfies (5.1), it also satisfies

$$\phi_*(\mathscr{V}(\gamma')) = \mathscr{V}(\gamma'). \tag{5.7}$$

From (5.7) it follows that ϕ acts on the derived systems γ^1, γ^2,..., the associated vector bundles, and intertwines the action of the structure tensors. The "algebraic invariants" of these structure tensors will be "equivalence invariants" of the Pfaffian systems.

6. THE EHRESMANN JET-CALCULUS. THE TWO-JET BUNDLE AS A VECTOR BUNDLE OVER THE ONE-JET BUNDLE

The Ehresmann theory of "jets" of mappings[4] is essentially a formalization of certain ideas of classical differential geometry; for example, "orders of contact," "tangent and osculating curves and surfaces." In modern language, it is a *functor* attaching to a pair (Y, X) of manifolds a sequence

$$J^0(Y,X) = Y \times X,$$
$$J^1(Y,X),$$
$$\vdots$$
$$J^r(Y,X),$$
$$\vdots$$

of manifolds, with a "tower" of fiber space maps naturally defined:

$$J^0(Y,X) \leftarrow J^1(Y,X) \leftarrow \cdots . \tag{6.1}$$

In this paper, which is only concerned with mechanical systems of a *finite* number of degrees of freedom, i.e., systems governed by *ordinary* differential equations, the main cases of interest are

$$Y = T, \text{ an interval of real numbers } t, \text{ say } 0 \leqslant t < \infty, \tag{6.2}$$

$$X = Q, \text{ the configuration space of a mechanical system,}$$
$$r = 0,1,2 . \tag{6.3}$$

Accordingly, I will give the definition and some of the properties of the theory of jets in this case. Let $\mathscr{M}(T,Q)$ be the space of smooth maps $\mathbf{q}: T \to Q$, i.e., the space of curves in Q.

Introduce an equivalence relation in

$$T \times \mathscr{M}(T,Q)$$

as follows:

$$(t, \mathbf{q}) \sim (t', \mathbf{q}')$$

if and only if

$$t = t',$$
$$\mathbf{q}(t) = \mathbf{q}'(t'),$$
$$\frac{d\mathbf{q}}{dt}(t) = \frac{d\mathbf{q}'}{dt}(t),$$
$$\vdots$$
$$\frac{d^r\mathbf{q}}{dt^r}(t) = \frac{d^r\mathbf{q}}{dt^r}(t).$$

In words, $\mathbf{q}$ and $\mathbf{q}'$ meet to the rth order at the value $t = t'$.

$J^r(T,Q)$ is the quotient of $T \times \mathscr{M}(T,Q)$ by this equivalence relation. If $\mathbf{q} \in \mathscr{M}(T,Q)$, $\partial^r\mathbf{q}$ is the map $T \to J^r(T,Q)$, which assigns to $t \in T$ the equivalence class to which it belongs. It is called the rth *prolongation* or r-jet of the map $\mathbf{q}$.

Since contact to rth order implies contact to $(r-1)$st order, there is a "forgetting highest derivatives" map

$$J^r(T,Q) \to J^{r-1}(T,Q).$$

This is readily seen to be a fiber space map, i.e., a submersion which is a local product. Goldschmidt and Sternberg have shown[27] that each of the fibers has an affine structure, which (passing to the quotient) defines a vector bundle structure. We will now directly construct the vector bundle structure.

Let us restrict attention to the case

$$r = 2$$

which, of course, is the most important for mechanics.

Let

$$\mathbf{q}: T \to Q$$

be a curve in Q. Let $\partial^1\mathbf{q}$ be its one-jet. It is a map

$$T \to J^1(T,Q).$$

Assign to (t,q) the tangent vector to the curve $t \to \partial^1\mathbf{q}$. It is an element of

$$J^1(T,Q)_{\partial^1\mathbf{q}(t)},$$

the tangent vector space $J^1(T,Q)$ at the point $\partial^1\mathbf{q}(t)$. Call this

$$\phi(t,\mathbf{q}).$$

This map ϕ is constant on the equivalence classes (for $r = 2$) which define $J^2(T,Q)$, hence ϕ passes to the quotient to define a map

$$\boldsymbol{\phi}: J^2(T,Q) \to T(J^1(T,Q)). \tag{6.4}$$

(The right-hand side is the *tangent bundle* to the one-jet space.)

Another form of the vector bundle structure of $J^2(T,Q)$ will be more convenient for physical purposes. Let

$$X = J^1(T,Q).$$

Let

$$\pi: J^1(T,Q) \to Q \tag{6.5}$$

be the natural projection map. (π is the "forgetting" map

$$J^1 \to J^0 = T \times Q,$$

followed by the Cartesian projection $T \times Q \to Q$). Let E be the tangent vector bundle to Q pulled back to X via the map π. Explicitly, a point of E is a pair

$$(x, v)$$

with

$$x \in X, \quad v \in Q_{\pi(x)}.$$

Theorem 6.1: $J^2(T,Q)$, considered as a vector bundle over $X = J^1(T,Q)$, is naturally isomorphic to the vector bundle E.

Proof: Consider the map $\boldsymbol{\phi}$, introduced in formula (6.4), followed by the map

$$\pi_*: T(X) \to T(Q).$$

It is readily verified (left to the reader) that this is the natural isomorphism required for Theorem 6.1.

Now, let us assume that

$$Q = R^n,$$

so that we can use the notation of mechanics books. For general manifolds, this is equivalent to working within a fixed local coordinate system. One can now define mappings:

$$J^1(T,Q) \to R^n \times R^n \times R,$$
$$j^1(\mathbf{q})(t) \to (q,\dot{q},t),$$
$$J^2(T,Q) \to R^n \times R^n \times R^n \times T,$$
$$j^2(q)(t) \to (q,\dot{q},\ddot{q},t)$$

as follows:

$$q(\mathbf{q},t) = \mathbf{q}(t),$$
$$\dot{q}(\mathbf{q},t) = \frac{d\mathbf{q}}{dt}(t),$$
$$\ddot{q}(q,t) = (d^2\mathbf{q}(dt^2)(t)).$$

We will call $(q,\dot{q},t)$ and $(q,\dot{q},\ddot{q},t)$ the *Newtonian* coordinates for $J^1(T,Q)$, $J^2(T,Q)$. (Physically, $\dot{q}$ is, of course, *velocity*, $\ddot{q}$ *ac-*

celeration.) π is then defined as follows:

$$\pi(q,\dot{q},\ddot{q},t) = (q,\dot{q},t).$$

Two points of J^2 go into the same point of J^1 if and only if their $(q,\dot{q},t)$ coordinates are the same.

Suppose now that $\mathbf{q}$ is a map $\pi \to Q$ given in the classical style by $t \to \mathbf{q}(t)$. $\partial^1\mathbf{q}$ is the map

$$t \to (q(t), dq/dt, t).$$

Now, the basis of tangent vectors to $T(J^1(T,Q))$ corresponds to the coordinates $(q,\dot{q},t)$, and may be labeled

$$\partial_q, \partial_{\dot{q}}, \partial_t.$$

The tangent vector to the curve $\partial^1\mathbf{q}$ is then

$$\left(\frac{dq}{dt}\right)\partial_q + \left(\frac{d^2q}{dt^2}\right)\partial_{\dot{q}} + \partial_t.$$

We see that the fiber of π goes over to the affine subspace

$$\left(\frac{dq}{dt}\right)\partial_q + \partial_t + (\text{multiples of } \partial_{\dot{q}})$$

of $Q_{q(t)}$. When the coordinates (q) are changed, the *second derivatives*, modulo the first derivatives, change in a *linear* homogeneous way, which is the tipoff to the "vector bundle structure."

Finally, here is a more direct way to define this vector bundle structure. Let $(t,\mathbf{q})$ be an element of $T \times \mathscr{M}(T,Q)$. Assign to $(t,\mathbf{q})$ the linear map $\mathscr{F}(Q) \to R$ labeled $\delta^2(t,\mathbf{q})$ as follows:

$$\delta^2(t,\mathbf{q})(f) = d^2(dt^2(f(\mathbf{q}(t)))).$$

The space of linear maps in $\mathscr{F}(Q)$ is a linear space. If $(t,\mathbf{q})$, $(t',\mathbf{q}')$ have the same second order of contact, then

$$\delta^2(t,q) = \delta^2(t',q').$$

Hence, δ passes to the quotient to define a map

$$J^2(T,Q) \to L(\mathscr{F}(Q), R).$$

$[L(\mathscr{F}(Q), R) =$ vector space of known maps $\mathscr{F}(Q) \to R.]$

Theorem 6.2: The fibers of the forgetting map $J^2 \to J^1$ go over under this map to a subspace of $L(\mathscr{F}(Q), R)$.

Proof: Let us use coordinates (q). If $(t,\mathbf{q})$, with $\mathbf{q}: t \to \mathbf{q}(t)$, then $\delta^2(t,\mathbf{q})$ is the map

$$f \to f_{qq}\dot{q}^2 + f_q\ddot{q}.$$

Q.E.D.

To obtain the vector space structure for the fiber of the forgetting map $J^2 \to J^1$, one has only to assign to each point of the fiber the quotient linear vector space corresponding to the affine subspace.

Remark: Here is the general setting. We assigned to each fiber $J^2(T,Q)$ an affine subspace

$$v_0 + S$$

of a vector space V. For two points in the same fiber, the point v_0 is the *same*. Thus, we can map the fiber into the quotient V/S, obtaining the vector bundle structure directly.

Theorem 6.3: Suppose Q is a real vector space. Then the addition of maps $\mathbf{q}: \pi \to Q$ passes to quotient to define vector bundle structures on the fibers of the forgetting map

$$J^1(T,Q) \to J^0(T,Q) = T \times Q,$$

$$J^2(T,Q) \to J^1(T,Q).$$

$J^1(T,Q)$ is naturally identified (via the map)

$$(t,q) \to \left(t,\mathbf{q}(t), \frac{dq}{dt}(t)\right) \equiv (t,q,\dot{q})$$

with

$$T \times Q \times Q.$$

Similarly, $J^2(T,Q)$ is identified via the map

$$(t,q) \to (t,q,\dot{q},\ddot{q})$$

with $T \times Q \times \dot{Q} \times Q$. The vector bundle structures are $J^1(T,Q)$ and $J^2(T,Q)$ and are given as follows:

$$(t,q,\dot{q}) + (t,q,\dot{q}') = (t,q,\dot{q}, + \dot{q}'),$$

$$(t,q,\dot{q},\ddot{q}) + (t,q,\dot{q},\ddot{q}) = (t,q,\dot{q},\ddot{q},\ddot{q}').$$

Note that these are the natural ways velocities and accelerations add.

7. THE NEWTON–LAGRANGE EQUATIONS IN THE LANGUAGE OF DIFFERENTIAL FORMS

In the traditional "analytical mechanics," Lagrange's equations are the basic tool. (Recall that Lagrange's book was called *Mécanique Analytique*.) There are two approaches—directly, via Newton's equations, and indirectly, via the calculus of variations or "Hamilton's principle." In this section I will briefly review the former approach and then show how Lagrange's equations can be formulated using differential forms.

Let us start with Newton's equations (for particles) in the usual form:

$$\begin{aligned} m_1\frac{d^2\mathbf{r}_1}{dt^2} &= -\frac{\partial}{\partial\mathbf{r}_1}V_1(\mathbf{r}_1,\ldots,\mathbf{r}_N) + \mathbf{F}_1(\mathbf{r}_1,\ldots,\mathbf{r}_N,\dot{\mathbf{r}}_1,\ldots,\dot{\mathbf{r}}_N) \\ &\vdots \qquad\qquad \vdots \\ m_N\frac{d^2\mathbf{r}_N}{dt^2} &= -\frac{\partial}{\partial\mathbf{r}_N}V_N + \mathbf{F}_n. \end{aligned} \tag{7.1}$$

Here $\mathbf{r}_1,\ldots,\mathbf{r}_N$ are the position coordinates of N particles in R^3. $V_1,\ldots,V_N$ are the potential functions determining the "internal forces." $\mathbf{F}_1,\ldots,\mathbf{F}_N$ (functions of the positions $\mathbf{r}_1,\ldots,\mathbf{r}_N$ and the velocities $\mathbf{r}_1,\ldots,\mathbf{r}_N$) are the *external forces*.

Following the classical ideas, define variables q and $\dot{q}$ called *configuration* and *velocity variables*.

$$q = (q_1,\ldots,q_{3N}),$$

$$\mathbf{r}_1 = \begin{pmatrix} q_1 \\ q_2 \\ q_3 \end{pmatrix},\ldots,\mathbf{r}_N = \begin{pmatrix} q_{3N-2} \\ q_{3N-1} \\ q_{3N} \end{pmatrix},$$

$$\dot{q} = (\dot{q}_1,\ldots,\dot{q}_{3N}),$$

$$\dot{\mathbf{r}}_1 = \begin{pmatrix} \dot{q}_1 \\ \dot{q}_2 \\ \dot{q}_3 \end{pmatrix},\ldots,\dot{\mathbf{r}}_N = \begin{pmatrix} \dot{q}_{3N-2} \\ \dot{q}_{3N-1} \\ \dot{q}_{3N} \end{pmatrix}.$$

Set

$$T = \tfrac{1}{2}m_1(\dot{\mathbf{r}}_1)^2 + \cdots + \tfrac{1}{2}m_N(\dot{r}_N)^2, \tag{7.2}$$

$$L = T - V. \tag{7.3}$$

L, the *Lagrangian*, then becomes a function of $(q,\dot{q})$.

$$\frac{\partial L}{\partial \dot{\mathbf{r}}_1} = m_1 \dot{\mathbf{r}}_1,$$

$$\frac{\partial L}{\partial \mathbf{r}_1} = -\frac{\partial V}{\partial \mathbf{r}_1},$$

$$\frac{d}{dt}\frac{\partial L}{\partial \dot{\mathbf{r}}_1} - \frac{\partial L}{\partial \mathbf{r}_1} = m_1 \ddot{\mathbf{r}}_1 + \frac{\partial V}{\partial \mathbf{r}_1}$$

$$= , \text{ using Newton's equations (7.1), } \mathbf{F}_1.$$

Introduce indices

$$1 \leqslant a,b,\ldots \leqslant 3N$$

and the summation convention for these indices. Then (using the traditional[9] arguments) Eqs. (7.1) take the form

$$\frac{d}{dt}\frac{\partial L}{\partial \dot{q}^a} - \frac{\partial L}{\partial q^a} = F_a(q,\dot{q}). \tag{7.4}$$

The F_a are just the external forces rewritten so that Eq. (7.4) is equivalent to Eq. (7.1). [For example, $F_1(q,\dot{q})$ is the first component of the "vector" $\mathbf{F}_1(r,\mathbf{r})$.]

Notice the way the indices a occur in Eq. (7.4), i.e., *downstairs*. This indicates that the forces are—in the Lagrangian framework—*differential forms*.

Let us now forget where *Lagrange's equations (7.4)* came from, and study them using techniques of the theory of differential forms. Let X be the space of variables

$$(q^a, \dot{q}^a, t).$$

Let us assume that (q^a) are the coordinates of a manifold Q. We say that $(q^a, \dot{q}^a, t)$ are the natural coordinates (that we defined and called "Newtonian" in Sec. 6) for

$$J^1(T, Q).$$

Thus, we will *identify* X with $J^1(T, Q)$. Consider the following one-forms in X:

$$\theta^a = dq^a - \dot{q}^a dt. \tag{7.5}$$

They are called (for geometric reasons) the *contact forms*. Set

$$\theta \equiv L\,dt + \lambda_a \theta^a. \tag{7.6}$$

The λ_a are functions on X which we shall determine in a moment. Let us calculate $d\theta$:

$$d\theta = dL \wedge dt + d\lambda_a \wedge \theta^a + \lambda_a d\theta^a. \tag{7.7}$$

Now,

$$dL = \frac{\partial L}{\partial q^a} dq^a + \frac{\partial L}{\partial \dot{q}^a} d\dot{q}^a + \frac{\partial L}{\partial \tau} d\tau,$$

$$d\theta^a = d(dq^a) - d\dot{q}^a \wedge dt$$

$$= 0 - d\dot{q}^a \wedge dt.$$

Hence

$$d\theta = \left(\frac{\partial L}{\partial q^a} dq^a + \frac{\partial L}{\partial \dot{q}^a} d\dot{q}^a\right) \wedge dt + d\lambda_a \wedge d\theta^a - \lambda_a d\dot{q}^a \wedge dt.$$

If we now set

$$\lambda_a = \frac{\partial L}{\partial \dot{q}^a}, \tag{7.8}$$

notice that the terms in $d\theta$ involving $d\dot{q}$ cancel, and we are left with

$$d\theta = \frac{\partial L}{\partial q^a} dq^a \wedge dt + d\lambda_a \wedge \theta^a. \tag{7.9}$$

Use (7.5) to express dq^a in terms of θ^a, and substitute into (7.9), using also that $dt \wedge dt = 0$:

$$d\theta = \frac{\partial L}{\partial q^a} \theta^a \wedge dt + d\lambda_a \wedge \theta^a$$

$$= \left(d\left(\frac{\partial L}{\partial \dot{q}^a}\right) - \frac{\partial L}{\partial q^a} dt\right) \wedge \theta^a.$$

Set

$$\theta_a = d\left(\frac{\partial L}{\partial \dot{q}^a}\right) - \frac{\partial L}{\partial q^a} dt. \tag{7.10}$$

Then, we have what one might call the *basic formula* of analytical mechanics:

$$d\theta = \theta_a \wedge \theta^a. \tag{7.11}$$

Let us recapitulate what has been done. We have taken the Lagrangian function L and constructed the one-form θ on X. (θ is called the *Cartan form*, since it first appeared in Cartan's book, *Leçons sur les Invariants Intégraux*.) The contact forms θ^a are purely "kinematical." However, the θ_a involve the Lagrangian L and are "dynamical." Now, set

$$\Omega_a = \theta_a - F_a\,dt, \tag{7.12}$$

where the F_a are the component functions of the force, as they appear in Lagrange's equations (7.4). Also,

$$\Omega_a = d\left(\frac{\partial L}{\partial \dot{q}^a}\right) - \left(\frac{\partial L}{\partial q^a} + F_a\right) dt. \tag{7.13}$$

Let

$$t \longrightarrow \left(q(t), \dot{q}(t) = \frac{dq}{dt}, t\right)$$

be a curve in X that represents a solution of Lagrange's equations. Let T be the one-dimensional manifold whose coordinate is t. This determines a map

$$\phi: T \to X.$$

Theorem 7.1: The curve satisfies the Lagrange equations if and only if

$$\phi^*(\Omega_a) = 0 = \phi^*(\theta^a). \tag{7.14}$$

Now, we give general definitions using general bases for the $\mathcal{F}(Q)$ module of one-forms on Q, i.e., using what Cartan called "moving frames."

8. THE NEWTON–LAGRANGE EQUATIONS IN TERMS OF GENERAL MOVING FRAMES ("QUASICOORDINATES")

Continue with the notation of Sec. 7. Let $\omega^1,\ldots,\omega^n$ be a basis for one-forms on Q. Let T be the interval $0 \leqslant t < \infty$. Consider t as a function, ω^a as one-forms on $T \times Q$. Lift up ω^a and t to $J^1(T, Q)$ via the "forgetting" map

$$J^1(T, Q) \to J^0(T, Q) = T \times Q.$$

[We use the same notation for a form and its lift up to $J^1(T, Q)$.] Define real-valued functions

$$v^a: J^1(T, Q) \to R$$

as follows:

Given a map

$$\mathbf{q}: T \to R$$

for $t \in T$, let

$$\frac{d\mathbf{q}}{dt} \in Q_{\mathbf{q}(t)}$$

be the tangent vector to $\mathbf{q}$ at t. Then,

$$v^a(t,\mathbf{q}) = \omega^a\left(\frac{d\mathbf{q}}{dt}\right). \tag{8.1}$$

The functions

$$v^a: T \times \mathscr{M}(T, Q)$$

defined by (14.1) pass to quotient to "live" on $J^1(T, Q)$.

Theorem 8.1: The one-forms ω^a, dt, dv^a on $J^1(T, Q)$ (defined as explained above) form a basis for one-forms on $J^1(T, Q)$. The forms

$$\theta^a = \omega^a - v^a\, dt \tag{8.2}$$

are the *contact forms*, with the property that

$$j^1(\mathbf{q})^*(\theta^a) = 0 \tag{8.3}$$

for each $\mathbf{q} \in \mathscr{M}(T, Q)$. Let $\mathscr{C}$ be the Pfaffian system of one-forms on $J^1(T, Q)$ spanned by the θ^a. It is called the *contact system* and is invariant when the basis (ω^a) is changed.

The proof follows in a straightforward manner from the definitions.

Now, let

$$L: J^1(T, Q) \to R$$

be a Lagrangian.

Theorem 8.2: There is a unique one-form $\theta(L)$ on $J^1(T, Q)$ which satisfies the following conditions:

$$\theta(L) - L\, dt \in \mathscr{C}\,, \tag{8.4}$$

$$d\theta(L) \in \mathscr{D}^1(J^1(T, Q)) \wedge \mathscr{C}\,. \tag{8.5}$$

Remark: Equation (8.5) means that $d\theta(L)$ belongs to the Grassmann algebra ideal (not the exterior differential system, i.e., the differential ideal) generated by the contact forms.

Proof of Theorem 8.2: In terms of the basis

$$(\omega^a, dt, dv^a)$$

for one-forms on $J^1(T, Q)$ described above, dL can be written in the following form:

$$dL = L_t dt + L_a \omega^a + L_{n+a} dv^a, \tag{8.6}$$

where the (L_t, L_a, L_{n+a}), $1 \leqslant a,b, \leqslant n$, are functions on $J^1(T, Q)$. Let us first suppose that $\theta(L)$ exists, satisfying Eqs. (8.4) and (8.5), and prove that it is unique. In terms of the moving frame, $\theta(L)$ can be written in the following form:

$$\theta(L) = L\, dt + f_a(\omega^a - v^a\, dt) \tag{8.7}$$

$$= (L - f_a v^a)\, dt + f_a \omega^a, \tag{8.8}$$

where (f_a) are functions on $J^1(T, Q)$. Then

$$\begin{aligned} d\theta(L) &= L_a \omega^a \wedge dt + L_{n+a} dv_a \wedge dt + df_a \wedge \theta^a \\ &\quad + f_a(d\omega^a - dv^a \wedge dt) \\ &= (L_{n+a} - f_a) dv^a \wedge dt + L_a \theta^a \wedge dt + df^a \wedge \theta^a \\ &\quad + f_a d\omega^a. \end{aligned} \tag{8.9}$$

The hypothesis, condition (8.5), means that there are one-forms η_a such that

$$d\theta(L) = \eta_a \wedge \theta^a. \tag{8.10}$$

Now, the (θ^a, dv^a, dt) form an $F(J^1(T, Q))$-basis for $\mathscr{D}^1(J^1(T, Q))$. The right-hand side of Eq. (8.10) contains no term in $dv^a \wedge dt$. Thus, equality of Eqs. (8.9) and (8.10) requires that

$$L_{n+a} = f_a, \tag{8.11}$$

i.e.,

$$\theta(L) = L\, dt + L_{n+a} \theta^a. \tag{8.12}$$

This proves uniqueness.

The converse, i.e., that using Eq. (8.12) as the definition of the form $\theta(L)$ satisfies Eq. (8.10), is a direct calculation, left to the reader.

We can now write the Lagrange–Newton equation in a convenient and natural way in terms of the geometry of the one-jet spaces. Let F be a one-form in $\mathscr{C}$. Set

$$\Omega(L, F) = d\theta(L) + F \wedge dt. \tag{8.13}$$

The *characteristic vectors* of the two-form $\Omega(L)$ are the elements of $T(J^1(T, Q))$ such that

$$v \lrcorner \Omega(L, F) = 0. \tag{8.14}$$

A curve in $J^1(T, Q)$ is said to be a *characteristic curve* if $\Omega(L)$, its tangent vector at each point, is a characteristic vector.

Remark: $d\Omega(L)$ is not necessarily zero, hence a v satisfying Eq. (14.13) is not necessarily "Cauchy characteristic" in the sense of Cartan.[3] If there is a possibility of confusion, perhaps a term like "algebraically characteristic" could be used.

Theorem 8.3: A curve $\mathbf{q}: t \to q(t)$ in Q is a solution of the Newton–Lagrange equations (13.4) in the classical sense if and only if the curve

$$t \to \partial^1 \mathbf{q}(t)$$

is a characteristic curve of $\Omega(L, F)$ in the sense defined above.

Proof: To obtain the classical formula, specialize the one-forms ω^a to the differentials dq^a of a coordinate system (q^a) for Q.

9. MECHANICAL SYSTEMS WITH CONSTRAINTS

Continue with the notation of Sec. 9. We have seen that a mechanical system is determined by a "configuration" manifold Q, a time-parameter manifold T, and

a function L on $J^1(T, Q)$,

a one-form F on $J^1(T, Q)$, which lies in the Pfaffian system determined by the contact form.

We now introduce a set of *constraints* as a Pfaffian system $\mathscr{P}$ on $J^1(T, Q)$. We will define sets of curves in Q by imposing conditions on the tangent sectors to their prolongations. The following special notation will be useful here:

If $\mathbf{q}: t \to q(t)$ is a curve in Q, then $\mathbf{a}(\mathbf{q})$ denotes the tangent vector curve to the prolongation $\partial^1 \mathbf{q}$. Thus, $\mathbf{a}(\mathbf{q})$ is a curve on the tangent bundle to $J^1(T, Q)$, which annihilates the one-forms on $\mathscr{C}$, the contact Pfaffian system. (9.1)

Remark: The curve $\mathbf{a}(\mathbf{q})$ can be identified with the two-prolongation $\partial^2\mathbf{q}: T \to J^2(T, Q)$ under the identifications developed in Sec. 8. Physically, it is, of course, the *acceleration*.

Definition: A curve $\mathbf{q}: T \to Q$ is a *trajectory* of the mechanical system-with-constants $(L, F, \mathscr{P})$ if its prolonged curve $\partial^1\mathbf{q}$ satisfies the following conditions:

$$\mathbf{a}(\mathbf{q})(t) \lrcorner \Omega(L, F) \in \mathscr{P}, \tag{9.2}$$

$$\mathscr{P}(\mathbf{a}(\mathbf{q})(t)) = 0. \tag{9.3}$$

Remark: Geometrically, relations (9.2)–(9.3) mean that a $\partial^1\mathbf{q}$ is a solution curve of the constraint Pfaffian system $\mathscr{P}$ (which are a set of second order differential equations for q) and that $t \to \mathbf{a}(q(t))$ are characteristic vectors of the two-form $\Omega(L, F)$ restricted to the tangent vectors of $J^1(T, Q)$, which are annihilated by $\mathscr{P}$.

The constraint system $\mathscr{P}$ encountered in the classical treatises[9,11] seem to be of *first-order* type, namely, there is a Pfaffian system $\mathscr{P}'$ on $T \times Q \equiv J^0(T, Q)$ such that

> $\mathscr{P}$ is generated by the pull-back $\pi^*(\mathscr{P}')$, where $\pi: J^1(T, Q) \to J^0(T, Q)$ is the natural "forgetting" map on the jet spaces. (9.4)

Since it is really no extra mathematical work, we will carry along the more general type. It is not clear whether they do occur in physical situations.

Definition: The constraints are said to be *holonomic* if the Pfaffian system $\mathscr{P}$ is completely integrable in the Frobenius sense.

10. MECHANICAL SYSTEMS WHICH HAVE THE SAME TRAJECTORIES

Given the manifolds Q and T, what we have done is to assign a set of curves satisfying a system of second-order ordinary differential equations [(10.2)–(10.3) in coordinate free terms] to the triple $(L, F, \mathscr{P})$. We ask

> When do two such systems $(L, F, \mathscr{P})$, $(L', F', \mathscr{P}')$ determine the same set of trajectories? (10.1)

In other words, we want to describe the fiber of the mapping

$$\text{(Mechanical systems)} \to \text{(Trajectories)}.$$

Of course, this question can be examined at the simplest classical "calculation of variations" level, i.e., when the forces F and the constraints $\mathscr{P}$ are zero. There it comes down to the following, more classical, question:

> Given two Lagrangians L and L', when do they determine the same differential equations? (10.2)

Since the Lagrange equations

$$\frac{d}{dt}(L_{\dot{q}}) - L_q = 0 \tag{10.3}$$

are linear in L, this is the same as asking the following question:

> When is the Lagrangian $(L - L') = L''$ "exact," i.e., Lagrange equations (10.3) are *identities* for $L'' = L - L'$?

Theorem 10.1[8]**:** Lagrangians L and L' determine the same set of curves in Q if and only if

$$d(\theta(L) - \theta(L')) = 0. \tag{10.4}$$

Relation (1.4) indicates that the Inverse Problem (determining which families of curves in Q are extremals of calculus of variations problems) should involve *topology*, since the existence of forms satisfying conditions on the exterior derivative is intrinsically topological. Related developments have been provided (without using differential forms) by Takens[28] and Anderson and Duchamp.[29]

Let us now consider this question in a broader geometric context. Let X be a manifold with

$$\Omega \in \mathscr{D}^2(X)$$

a two-form on X. [X might be $J^1(T, Q)$, where Q is the configuration space of a mechanical system.] Suppose also that $\mathscr{P}$ and $\mathscr{C}$ are arbitrary Pfaffian systems on X.

Let

$$\mathscr{V}(\Omega, \mathscr{C}, \mathscr{P}) = \{V \in \mathscr{V}(X): V \quad \Omega \in \mathscr{P}, \mathscr{C}(V) = \mathscr{P}(V) = 0\}. \tag{10.5}$$

The *trajectories of the system* $(\Omega, \mathscr{P})$ are then the orbit curves of the vector fields in $\mathscr{V}(\Omega, \mathscr{C}, \mathscr{P})$.

Given $\mathscr{P}$ and $\mathscr{C}$ we are interested in determining the fibers of the mapping

$$\Omega \to \mathscr{V}(\Omega, \mathscr{P}). \tag{10.6}$$

The most immediate question of this type is:

> When is $\mathscr{V}(\Omega, \mathscr{C}, \mathscr{P})$ equal to $\mathscr{V}(\mathscr{P}) \cap \mathscr{V}(\mathscr{C})$, i.e., when does the relation $\mathscr{C}(V) = 0 = \mathscr{P}(V)$ for $V \in \mathscr{V}(X)$ imply $V \lrcorner \Omega \in \mathscr{P}$? (10.7)

Of course, the simplest condition of this sort would be the following one:

> Ω belongs to the Grassmann algebra ideal generated by $\mathscr{P}$. (10.8)

Another related question is the following one:

> Given a mechanical system $(\Omega, \mathscr{C}, \mathscr{P})$, when does there exist another form Ω' which lies in the same fiber as Ω in the mapping (10.6) (i.e., determines the same trajectories), but also satisfies the following condition:
>
> $$d\Omega' = 0. \tag{10.9}$$

In turn, this question is related to the "symplectic" nature of the mechanical system.

We will now investigate this point for mechanical systems associated most directly with traditional analytical mechanics.

11. CAUCHY CHARACTERISTICS OF A LAGRANGIAN SYSTEM WITH EXTERNAL FORCES

Let Q be an n-dimensional manifold, and let

$$X = J^1(T, Q)$$

and

$$L: X \to R$$

be a Lagrangian function for Q. Let

$$\theta(L) \in \mathscr{D}^1(X)$$

be the Cartan form associated with L. Let $\mathscr{C} \subset \mathscr{D}^1(X)$ be the

Pfaffian system on X generated by the contact forms.

Let

$$F \in \mathscr{C}$$

be a *force form*, and let

$$\Omega = d\theta(L) + F \wedge dt. \tag{11.1}$$

The vector fields V on X which satisfy the following conditions:

$$V(t) = 1, \tag{11.2}$$

$$V \lrcorner \Omega = 0, \tag{11.3}$$

$$\mathscr{C}(V) = 0, \tag{11.4}$$

are called the *Newton–Lagrange* vector fields. The orbits of V are then prolongations of the solution of the classical Newton–Lagrange equations.

Let

$$\mathscr{E}(\Omega) \tag{11.5}$$

be the differential ideal in the Grassmann algebra of X generated by Ω. For each $x \in X$, let $\mathscr{C}(\mathscr{E}(\Omega))(x)$ be the Cauchy characteristic vector of this ideal. In this case,

$$\mathscr{C}(\mathscr{E}(\Omega))(x) = \{v \in X_x : v \lrcorner \Omega = 0 = v \quad d\Omega\}. \tag{11.6}$$

Definition: The system is said to be *nonsingular* if the following condition is satisfied:

$$\dim(\mathscr{C}(\mathscr{E}(\Omega))(x)) \tag{11.7}$$

is constant as x ranges over X.

By the fundamental "Cauchy characteristic" theorem[16] of Cartan, the Pfaffian system $\mathscr{C}(\mathscr{E}(\Omega))$ is *completely integrable*. We will say that the Cauchy characteristics [and the mechanical system (L, F)] is *regular* if the foliation $\mathscr{C}(\mathscr{E}(\Omega))$ is *regular* in the sense that there is a fiber space mapping

$$\alpha: X \to Z \tag{11.8}$$

of X onto a manifold Z, whose fibers are the Cauchy characteristic submanifolds of $\mathscr{E}(\Omega)$. Another way of putting this is to say that the space of leaves of the Cauchy characteristic foliation admits a manifold structure.

By Cartan's theorem,[16] there is a two-form Ω' on Z such that

$$\alpha^*(\Omega') = \Omega. \tag{11.9}$$

Thus, the system, in a sense, "lives" on the quotient space Z.

12. INFINITESIMAL SYMMETRIES OF LAGRANGIAN SYSTEMS WITH EXTERNAL FORCES

Continue with the notation of Section 11. Let $X = J^1(T, Q)$, L, F, Ω, $\mathscr{E}(\Omega)$, $\mathscr{C}(\mathscr{E}(\Omega))$,

$$Z = \mathscr{C}(\mathscr{E}(\Omega)) \backslash X \tag{12.1}$$

be as defined there. Continue to assume that Z is a manifold, i.e., the Cauchy characteristic foliation is regular.

Definition: A vector field V on X is said to be an *infinitesimal symmetry* of the mechanical system (L, F) if

$$\mathscr{L}_V(\Omega) = 0. \tag{12.2}$$

Theorem 12.1: If V satisfies (12.2), then it also satisfies the following conditions:

$$\mathscr{L}_V(\mathscr{E}(\Omega)) \subset \mathscr{E}(\Omega), \tag{12.3}$$

$$\mathscr{L}_V(\mathscr{C}(\mathscr{E}(\Omega))) \subset \mathscr{C}(\mathscr{E}(\Omega)). \tag{12.4}$$

In words, V generates an infinitesimal symmetry of the exterior differential system generated by Ω and the Cauchy characteristic foliation.

Proof: This is a general property of exterior differential systems, and follows readily from the standard identities linking the operation of the Lie derivative, contraction, and exterior derivative.

A particularly important symmetry for mechanical systems is, of course, time-translation. This corresponds to a vector field on $X = J^1(T, Q)$ that we will call

$$\partial_t. \tag{12.5}$$

Thus, if (ω^a) is a basis for differential forms on Q (also known as "quasicoordinates" or "moving frame"), and

$$\omega^a, dv^a, dt \tag{12.6}$$

is the associated basis for X, then

$$\omega^a(\partial_t) = 0 = dv^a(\partial_t), \tag{12.7}$$

$$(dt)(\partial_t) = 1. \tag{12.8}$$

Set

$$L_t = \partial_t(L), \tag{12.9}$$

$$F_t = \mathscr{L}_{\partial_t}(F). \tag{12.10}$$

(L_t, F_t) are the Lagrangian and force law of another mechanical system, hence their Cartan form can be constructed. Call it Ω'.

Theorem 12.2:

$$\mathscr{L}_{\partial_t}(\Omega) = \Omega'. \tag{12.11}$$

Proof: Again, this is a slight extension of the material in Chap. 4 of Ref. 7. The proof is left to the reader.

Definition: The system (L, F) is said to be *time-invariant* if

$$L_t = F_t = 0. \tag{12.12}$$

Theorem 12.3: Suppose the system (L, F) is time-invariant. Then, the vector field ∂_t is a symmetry of the system, in the sense defined above. Further, ∂_t passes to the quotient to define a vector field ∂_t' on the quotient manifold Z, i.e.,

$$\alpha_*(\partial_t) = \partial_t'. \tag{12.13}$$

Proof: Again, this follows from general principles. In the simplest case

$$F = 0. \tag{12.14}$$

Theorem 12.3 is the essence of the symplectic structure approach[1,2] to Hamiltonian mechanics. In this case,

$$d\Omega \equiv d(\theta(L)) = 0,$$

hence, Ω' is a two-form Z. It defines a *symplectic structure* on Z, in the sense of Refs. 1 and 2, and a Poisson bracket (Lie algebra) structure on $\mathscr{F}(Z)$. The vector field ∂_t' on Z then satisfies

$$\mathscr{L}_{\partial_t'}(\Omega') = 0. \tag{12.15}$$

Then,

$$d(\partial'_t \lrcorner \Omega') = 0. \tag{12.16}$$

$\partial'_t \lrcorner \Omega'$, as a closed one-form on Z determines an element of

$$H^1(Z, R),$$

the (de Rham) one-cohomology of Z with real coefficients. If this element is zero, there is a function

$$h: Z \to R$$

such that

$$dh = \partial'_t \lrcorner \Omega'.$$

h is then a *Hamiltonian function* for the system. The orbits of ∂'_t are the solutions of Hamilton's equations for this Hamiltonian. In favorable cases (e.g., maximal rank conditions for certain Hessian matrices) Z and its associated symplectic structure can be mapped onto the "canonical" symplectic structure on $T^d(Q)$, the cotangent bundle to Q thus linking up the Lagrangian approach of this paper with the Hamiltonian or symplectic approach, which is more popular in the contemporary literature. However, it should be remarked that the former approach is more general, and better adapted to a program of modernization of the approach in the classical treatises.[9,11]

13. SOME NEWTON–LAGRANGE MECHANICAL SYSTEMS WITH FORCES WHICH ADMIT SYMPLECTIC STRUCTURES

After this review of the relation between the Lagrangian and symplectic approaches to the traditional force-free analytical mechanics problems, let us move on to seek generalization to the case where there are force forms F. Then, the two-form

$$\Omega = d\theta(L) + F \wedge dt \tag{13.1}$$

is not *a priori* closed. There are then two questions to study:

What are the conditions on F that the form Ω satisfy $d\Omega = 0$? (13.2)

How can one modify Ω to obtain a closed two-form Ω' whose characteristic curves determine the solution of the Newton–Lagrange equations? (13.3)

I will treat both of these problems in this section.

Theorem 13.1: Let Ω be a two-form on $J^1(T, Q)$ of the form (13.1) with F a one-form which lies in the Grassmann algebra generated by the contact forms $\mathscr{C}$. Then, Ω satisfies

$$d\Omega = 0 \tag{13.4}$$

if and only if F arises from a time-dependent, exterior-closed one-form on Q in the natural way.

Proof: Since this is a purely local matter, choose coordinates (q^a) for Q, and the corresponding coordinates

$$(q^a, \dot{q}^a, t)$$

for $J^1(T, Q)$.

Then, our assumption that F is a "force-form" means that it can be written in the following form:

$$F = F_a(dq^a - d\dot{q}^a\, dt), \tag{13.5}$$

where the F_a are functions on

$$J^1(T, Q).$$

Then,

$$F \wedge dt = F_a \wedge dq^a \wedge dt. \tag{13.6}$$

Since

$$d(d\theta(L)) = 0, \tag{13.7}$$

the condition (13.6) requires that

$$d(F_a \wedge dq^a \wedge dt) = 0, \tag{13.8}$$

which implies that

$$\frac{\partial(F_a)}{\partial \dot{q}^b} \wedge d\dot{q}^b \wedge dq^a \wedge dt = 0,$$

and hence

$$\frac{\partial F_a}{\partial \dot{q}^b} = 0. \tag{13.9}$$

Then, we can assign to F the closed one-form

$$F' = F_a dq^a \tag{13.10}$$

in $Q \times T$. Again, it follows from (13.6) that

$$dF' = 0. \tag{13.11}$$

It is readily seen that the assignment

$$F \to F' \tag{13.12}$$

is "intrinsic," i.e., is independent of the coordinates.

Conversely, the steps are reversible, i.e., a form F' satisfying (13.11) leads to an F and an Ω.

Condition (13.11) suggests the introduction of a "potential," i.e., *locally* a function V on $Q \times T$ such that

$$dV \wedge dt = dF \wedge dt. \tag{13.13}$$

Of course, de Rham's relation between differential form cohomology and the topology of manifolds leads to interesting (and reasonably well known and explored) relations between "local" and "global" potentials.

Now, suppose that condition (13.6) is not satisfied. Let us modify Ω in the following way:

$$\Omega' = \Omega + f_{ab}(dq^a - \dot{q}^a dt) \wedge (dq^b - \dot{q}^b dt), \tag{13.14}$$

where (f_{ab}) is a skew-symmetric matrix of functions. Notice that Ω and Ω' determine the same trajectories, so the tangent vectors v such that

$$\begin{aligned} v \lrcorner \Omega &= 0, \\ v \lrcorner \mathscr{C} &= 0 \end{aligned} \tag{13.15}$$

and

$$\begin{aligned} v \lrcorner \Omega' &= 0, \\ v \lrcorner \mathscr{C} &= 0 \end{aligned} \tag{13.16}$$

are the *same*.

Theorem 13.2: $d\Omega' = 0$ if and only if the following conditions are satisfied:

$$d(f_{ab} dq^a \wedge dq^b) \wedge dt = 0, \tag{13.17}$$

$$\frac{\partial}{\partial \dot{q}^a}(F_b) = 2f_{ab}, \tag{13.18}$$

$$\frac{\partial}{\partial t}(f_{ab}) = \frac{1}{2}\left(\frac{\partial}{\partial q^a}(F_b - 2f_{cb}\dot{q}^c) - \frac{\partial}{\partial \dot{q}^b}(F_a - 2f_{ca}\dot{q}^a)\right). \tag{13.19}$$

Proof: Follows from applying exterior derivative d to both sides of (13.15), and the relation

$$d\Omega = dF \wedge dt.$$

We can now analyze further relations (13.17)–(13.19). They imply that the force terms are of the following type:

$$F_a = 2f_{ba}\dot{q}^b + f_a, \tag{13.20}$$

where the a's and b's are functions of q and t alone.

Theorem 13.3: With the force law F of form (13.20), the following conditions are necessary and sufficient that the two-differential form Ω' be closed:

$$d(f_a dq^a) = \frac{\mathscr{L}_\partial}{\partial t}(\mathscr{L}_{ab} dq^a \wedge dq^b)$$

$$= \frac{\partial}{\partial t}(f_{ab}) dq^a \wedge dq^b. \tag{13.21}$$

Proof: Substitute the ansatz (13.20) into the other conditions.

An important case physically is that where the force law is time-independent. In this case, the conditions found above simplify and can be written in an elegant global form.

Theorem 13.4: Suppose that we define a mechanical system by means of a Lagrangian function L and a force law F, with $\partial_t(F) = 0$. Suppose that this force law satisfies condition (13.20). Define differential forms α and β on Q so that

$$\alpha = f_{ab} dq^a \wedge dq^b, \tag{13.22}$$

$$\beta = f_a dq^a. \tag{13.23}$$

Then, the condition that the two-differential form Ω' be closed is that the following conditions be satisfied:

$$d\alpha = 0,$$

$$d\beta = 0.$$

If $d\Omega' = 0$, then the trajectories of the mechanical system are Cauchy characteristic curves of Ω'. Assuming that the trajectories define a regular foliation of $J^1(R, Q)$, the quotient space of the foliation has a symplectic structure.[30]

The condition for the existence of a symplectic structure on the space of trajectories is especially important in electromagnetic theory. For example, suppose that q are the coordinates of a charged particle in R^3. It is well known that the force law then has the form described in condition (13.20), where the f_{ab} are the components of the "magnetic" field, the f_a the "electric" field. The conditions $0 = d\alpha$, $0 = d\beta$ are then part of Maxwell's equations. It is well known that to write the charged particle equations in Hamiltonian form requires that these differential forms be exact; the forms whose exterior derivatives give them are the potentials. Thus, we see that the charged particle equations themselves may be given a symplectic structure in a global way without the intervention of the potentials. Such a possibility is very relevant to quantization.

14. NEWTON'S EQUATIONS OF MECHANICS DIRECTLY IN TERMS OF THE VECTOR BUNDLE STRUCTURE OF THE TWO-JET SPACE

I now briefly sketch a direct method of expressing the relation

$$\text{mass} \times \text{acceleration} = \text{force} \tag{14.1}$$

without the intervention of concepts from the calculus of variations. Let us begin with the case where the configuration manifold Q is a finite dimensional real vector space. (Since we use coordinate free methods, some of the ideas can be extended to the infinite dimensional case, using the appropriate functional analysis generalities.) Let Q^d denote the dual vector space. Denote the duality function

$$Q^d \times Q \to R$$

by

$$(q^d, q) \to \langle q^d, q \rangle .$$

Let T continue as a time interval manifold, say

$$0 \leqslant t < \infty .$$

Definition: A *mass law* is a linear map

$$m: Q \to Q^d \tag{14.2}$$

which is *symmetric* in the sense that

$$\langle m(q), q' \rangle = \langle m(q'), q \rangle$$

$$\text{for } q, q' \in Q. \tag{14.3}$$

A *force law* is a map

$$f: Q \times Q \times R \to Q^d. \tag{14.4}$$

A *mechanical system* (in Newton's sense) is a pair

$$(m, f)$$

of a mass law m and force law f. A curve $t \to q(t)$ is a *trajectory* of the mechanical system if it is a map

$$\mathbf{q}: T \to Q,$$

$$t \to q(t)$$

such that

$$m\left(\frac{d^2 q}{dt^2}\right) = f\left(q(t), \frac{dq}{dt}, t\right). \tag{14.5}$$

One can derive the energy relation

$$\mathrm{KE}(\mathbf{q}) = \frac{1}{2}\left\langle m\left(\frac{d\mathbf{q}}{dt}\right), \frac{dq}{dt}\right\rangle \equiv \textit{kinetic energy}.$$

$$\frac{\mathrm{d}}{\mathrm{dt}}(\mathrm{KE}(\mathbf{q})) = [\text{using (20.4) and (20.5)}],$$

$$f\left(q, \frac{dq}{dt}, t\right), \frac{dq}{dt}, \tag{14.6}$$

$$W[\mathbf{q}; (a,b)) = \int_a^b \left\langle f\left(q, \frac{dq}{dt}, t\right), \frac{dq}{dt}\right\rangle dt$$

is the *work* in the time interval $a \leqslant t \leqslant b$.

Now, we can free this material from the constraint that Q be a vector space by using the Ehresmann jet-space formalism. Identify

$$Q \times Q \times T$$

with $J^1(T, Q)$, the space of one-jets of maps $T \to Q$. Let x denote a point of $J^1(T, Q)$ identified (when Q is a vector space) with a triple:

$$(t, q, \dot{q}),$$

$$t \in T, q \in Q, \dot{q} \in Q.$$

Let

$$\pi: J^1(T, Q) \to Q \tag{14.7}$$

be the projection map:

$$\pi(x) = \pi(t, q, \dot{q}) = q. \tag{14.8}$$

Let E be the vector bundle over $J^1(T, Q)$, which results from lifting the tangent bundle $T(Q)$ to X via the projection map (14.7). Explicitly, a point of E is an ordered pair

$$(x, v) \tag{14.9}$$

where

$$x \in J^1(T, Q)$$

and

$$v \in Q_{\pi(x)}.$$

Let E^d be the *dual* vector bundle to E, i.e., a point of E^d is a pair (x, θ).

$$x \in J^1(T, Q), \quad \theta \in Q^d_{\pi(x)}.$$

The *force law* f should then be interpreted as a cross-section map

$$f: J^1(T, Q) \to E. \tag{14.10}$$

The mass law is a linear bundle map

$$m: E \to E^d. \tag{14.11}$$

Thus, the right-hand side of Eq. (14.5) is

$$f(\partial^1 \mathbf{q}(t)), \tag{14.12}$$

which is an element of the fiber of E^d above the point

$$\partial^1 \mathbf{q}(t)$$

of $J^1(T, Q)$.

In order to have Newton's equation (14.5) make sense on a general manifold Q (i.e., to put it into a fully "covariant" form) it is necessary to make sense of the left-hand side of Eq. (14.5). This is where the construction of Sec. 6 enters: We have shown that the vector bundle E can be identified with the two-jet space:

$$J^2(T, Q) = E. \tag{14.13}$$

The "acceleration" vector

$$\frac{d^2 q}{dt^2}(t)$$

should be identified [via Eq. (14.13)] with an element of E. We can write Eq. (14.5) as

$$m(\partial^2 \mathbf{q}(t)) = f(\partial^2 q(t)). \tag{14.14}$$

We can now translate Lagrange's equations into this framework. Write them in the following form:

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}^a}\right) - \frac{\partial L}{\partial q^a} = f_a(q, \dot{q}, t),$$

$$\dot{q}^a = \frac{dq^a}{dt} \quad (1 \leqslant a, b \leqslant n = \dim Q). \tag{14.15}$$

Do the derivation on the left-hand side of (14.15):

$$\frac{\partial^2 L}{\partial \dot{q}^a \partial \dot{q}^b} = \frac{d^2 q^b}{dt^2} + \frac{\partial^2 L}{\partial \dot{q}^a \partial q^b}\frac{dq^b}{dt} + \frac{\partial^2 L}{\partial t \partial \dot{q}^a} - \frac{\partial L}{\partial q^a} = f_a. \tag{14.16}$$

We can derive from Eq. (14.16) the definition of a mass and force law which leads to Eq. (14.14):

$$m(x)(v) = \left(\frac{\partial^2 L}{\partial \dot{q}^a \partial \dot{q}^b}(x) v^b\right)\frac{\partial}{\partial q^a}, \tag{14.17}$$

$$f(x) = \left(f_a + \frac{\partial L}{\partial q^a} - \frac{\partial^2 L}{\partial \dot{q}^a \partial q^b}\dot{q}^b + \frac{\partial^2 L}{\partial t \partial \dot{q}^a}\right) dq^a \tag{14.18}$$

for

$$x = (q, : \dot{q}, t) \in J^1(T, Q), \quad v \in Q_{\pi(x)}, \quad (x, v) \in E.$$

15. MECHANICAL SYSTEMS AS PFAFFIAN SYSTEMS

We can now formulate what a "general mechanical system" might mean from the point of view of the theory of Pfaffian systems. For simplicity, we deal only with systems without constraints.

Let X be a manifold of dimension $(2n + 1)$. Choose indices and the summation convention as follows: $1 \leqslant a, b \leqslant n$. Suppose the following data is given on X:

(a) a real-valued function $t \in \mathscr{F}(X)$,

(b) a completely integrable Pfaffian system $\mathscr{I} \subset \mathscr{D}^1(X)$ whose leaves are dimension $(n + 1)$, such that

$$v(t) \neq 0 \tag{15.1}$$

for each nonzero tangent vector to X such that

$$\mathscr{I}(v) = 0 \tag{15.2}$$

(In particular, $dt \neq 0$ at each point of X.),

(c) a Pfaffian system $\mathscr{C}$, determined by a cross-section map

$$x \to G^{n+1}(T(X)), \tag{15.3}$$

(d) as an $\mathscr{F}(X)$ module, $\mathscr{C}$ is contained in the submodule spanned by $\mathscr{I}$ and dt,

(e) a two-form Ω on X which lies in the Grassmann algebra ideal generated by $\mathscr{C}$. Symbolically,

$$\Omega \in \mathscr{D}^1(X) \wedge \mathscr{C}. \tag{15.4}$$

The curves

$$\mathbf{x}: \tau \to x(\tau)$$

in X which satisfy the conditions

$$\frac{d}{d\tau}(t(x(\tau))) \neq 0, \tag{15.5}$$

$$\mathscr{C}\left(\frac{dx}{d\tau}\right) = 0, \tag{15.6}$$

$$\frac{dx}{d\tau} \quad \Omega = 0, \tag{15.7}$$

are called the *trajectories* of the system. ($dx/d\tau$ denotes the tangent vector curve to $\mathbf{x}$.)

Condition (15.5) means that these trajectories can be

reparametrized (locally) by the function t, i.e., so that

$$t(x(\tau)) = \tau. \tag{15.8}$$

We shall deal with these trajectory curves in this form.

Let us make these assumptions explicit in terms of moving frames, i.e., local bases of the $\mathscr{F}(X)$ modules involved. Suppose then that

(ω^a) is an $\mathscr{F}(X)$ basis of $\mathscr{I}$.

Let (ω^a_b) be forms (which exist locally by complete integrability of $\mathscr{I}$) such that

$$d\omega^a = \omega^a_b \wedge \omega^b. \tag{15.9}$$

Condition (d) implies that there is (locally) a basis θ^a of $\mathscr{C}$ of the form

$$\theta^a = \omega^a - v^a\, dt, \tag{15.10}$$

where $v^a \in \mathscr{F}(X)$. Condition (15.4) implies that Ω can be written in the following form:

$$\Omega = \eta_a \wedge \theta^a, \tag{15.11}$$

where (η_a) are a set of one-forms. The trajectory curves $\tau \to x(\tau)$ then satisfy the following Pfaffian equations:

$$\eta_a\left(\frac{dx}{d\tau}\right) = 0$$
$$= \theta^a\left(\frac{dx}{d\tau}\right). \tag{15.12}$$

Let $\mathscr{P}$ be the Pfaffian system generated by η_a and θ^a. Equation (15.12) then says that $\tau \to x(\tau)$ is an *integral curve* of $\mathscr{P}$.

We can describe the trajectories in terms of second order, quasilinear differential equations in the following way.

Let (q^a) be a set of functions on X (or possibly an open subset of X) such that the (dq^a) generate the module $\mathscr{I}$. [In other words, the (q^a) are local coordinates of the leaf space $\mathscr{I} \setminus X$.] Then,

$$\omega^a = A^a_b dq^b, \tag{15.13}$$

$$\theta^a = \omega^a - v^a\, dt$$
$$= A^a_b dq^b - v^a\, dt. \tag{15.14}$$

Suppose that the η_i are of the following form:

$$\eta_a = B_{aj}\theta^b + C_{ab}dv^b + C_a st. \tag{15.15}$$

The integral curves $\sigma: T \to X$ of the Pfaffian system $\mathscr{P}$ [which satisfy Eq. (15.12)] then satisfy the following differential equations:

$$C_{ab}(\sigma(t))\frac{dv^b}{dt} + C_a = 0, \tag{15.16}$$

$$v^a(t) = A^a_b(\sigma(t))\frac{dq^b}{dt}, \tag{15.17}$$

$$q^a(t) = q^a(\sigma(t)). \tag{15.18}$$

Notice that this is a second order ordinary differential equation for the $t \to q(t)$, which is linear in the second derivatives.

16. FINAL REMARKS

In this paper I have described how mechanical systems of a finite number of degrees of freedom, but of a general type in terms of external forces, constraints, etc., could be described in terms of the differential geometric theories of Cartan and Ehresmann. The ground has been prepared for the application of the theory of *equivalence* of geometric structures, the application of the theory of *pseudogroups*, the study of global properties, and so on. In this way, I believe that we are now in a position to more fully realize possibilities of applying differential geometry to mechanics and electromagnetic theory that were suggested many years ago by Gabriel Kron[5] and the RAAG Memoirists.[6] The introduction of control-theoretic concepts[12–15] also provides a fertile field of potential application of the formalism.

[1] R. Abraham and J. E. Marsden, *Foundations of Mechanics* (Addison-Wesley, Reading, MA, 1979), 2nd ed.
[2] V. Arnold, *Methods of Classical Mechanics*, Springer Graduate Texts on Mathematics, No. 60 (Springer, New York, 1978).
[3] E. Cartan, *Leçons sur les Invariants Intégraux* (Herman, Paris, 1922).
[4] C. Ehresmann, "Les prolongements d'un variété différentielles, I–V," C. R. Acad. Sci. **233**, 598, 777, 1081 (1951); **234**, 1028, 1424 (1952).
[5] G. Kron, "Non-Riemannian geometry of rotating electrical machinery," J. Math. Mech. **13**, 103–194 (1934).
[6] K. Kondo, *The RAAG Memoirs* (Gakujutsu Bunken Fukyu-Kai, Tokyo, 1955).
[7] R. Hermann, *Differential Geometry and the Calculus of Variations* (Academic, New York, 1968).
[8] R. Hermann, *Geometry, Physics, and Systems* (Dekker, New York, 1973).
[9] E. Whittaker, *Analytical Dynamics* (Cambridge U. P., Cambridge, 1959).
[10] C. Caratheodory, *Calculus of Variations and Partial Differential Equations* (Holden-Day, San Francisco, 1965).
[11] J. Neimark and N. Fufaev, *Dynamics of Nonholonomic Systems* (American Mathematical Society, Providence, RI, 1972).
[12] G. Meyer and L. Cicolani, "Application of nonlinear system inverses to automatic flight control design-system concepts and flight evaluations," *AGARDograph on Theory and Applications of Optimal Control in Aerospace Systems*, edited by P. Kant (NATO, London, 1980).
[13] R. Su, "On the linear equivalents of nonlinear systems," Systems Control Lett. (to be published).
[14] L. R. Hunt and R. Su, "Multi-input nonlinear systems," SIAM J. Control Optimization (to be published).
[15] R. Hermann, "The theory of equivalence of Pfaffian systems and input systems under feedback," *Mathematical System Theory* (to be published).
[16] E. Cartan, *Les Systèmes Différentiels Extérieurs et leurs Applications Geométriques* (Herman, Paris, 1945).
[17] E. Cartan, "Sur l'Equivalence Absolue de Certains Systemes d'Equations Différentielles," *Oeuvres Completes* (Gauthier-Villars, Paris, 1953), pp. 1133–1141.
[18] E. Cartan, Les Systemes de Pfaff a Cinq Variables," *Oeuvres Completes* (Gauthier-Villars, Paris, 1953), pp. 927–1010.
[19] E. Cartan, "Sur l'Integration de Certains Systemes de Pfaff de Caractere Deux," *Oeuvres Completes* (Gauthier-Villars, Paris, 1953), pp. 483–502.
[20] E. Cartan, Sur Certaines Expressions Differentielles et le Probleme de Pfaff, Collected Works, Ann. Ec. Norm. **16**, 239–332 (1899).
[21] E. Goursat, *Leçons sur le Probleme de Pfaff* (Herman, Paris, 1922).
[22] W. L. Chow, "Systeme von linearen partiellen differentialgleichungen ersten Ordmung," Math. Ann. **117**, 98–105 (1939).
[23] R. Hermann, "Cartan connections and the equivalence problems for geometric structures," Contrib. Diff. Eq. **3**, 199–248 (1964).
[24] S. Sternberg, *Lectures on Differential Geometry* (Prentice-Hall, Englewood Cliffs, NJ, 1964).
[25] R. Hermann, "E. Cartan's geometric theory of partial differential equations," Adv. Math. **1**, 265–317 (1965).
[26] R. B. Gardner, "Invariants of Pfaffian systems," Trans. Am. Math. Soc. **126**, 514–533 (1967).
[27] H. Goldschmidt and S. Sternberg, "The Hamilton-Jacobi formalism in the calculus of variations," Ann. Inst. Fourier **23**, 203–267 (1973).
[28] F. Takens, "A global version of the inverse problem of the calculus of variation," J. Diff. Geom. (to be published).
[29] I. Anderson and T. Duchamp, "On the existence of global variational principles," Am. J. Math. **102**, 781–868 (1979).
[30] R. Hermann, "Some differential geometric aspects of the Lagrange variational problem," Ill. J. Math. **6**, 634–673 (1962).

The equivalence of two approaches to the Feynman integral

G. W. Johnson
Department of Mathematics, University of Nebraska, Lincoln, Nebraska 68588-0323

(Received 5 February 1982; accepted for publication 25 June 1982)

Two apparently quite different Banach algebras of functions have been introduced and studied recently in connection with the theory of the "Feynman integral." The functions in both spaces have been shown to be "Feynman integrable," but two different definitions of the "Feynman integral" were used. We show here that the two spaces are in fact isometrically isomorphic as Banach algebras where the correspondence is given by what is essentially an extension (or restriction) map. Further, the "Feynman integrals," in the two different senses, of corresponding functions are equal. The equivalence between these two theories is surprisingly easy to prove but has a number of consequences for both theories. In the last section of the paper we give a few simple but useful consequences and make some remarks about our experience so far in using the equivalence.

PACS numbers: 02.50.Sk, 03.65.Db

I. INTRODUCTION

We begin by giving information necessary for our discussion of the Banach algebra S. The space S was introduced and studied by Cameron and Storvick in Ref. 1 and was studied further in Refs. 2–6.

Fix $t>0$. For notational simplicity, we will confine our attention throughout to the interval $[0,t]$ and, with one exception, to a single space dimension. Actually the arguments work just as well for a general interval $[a,b]$ and for an arbitrary number of space dimensions.

$C[0,t]$ will denote the set of $\mathbb{R}$-valued, (i.e., real valued) continuous functions on $[0,t]$. $C_0[0,t]$ will denote Wiener space, that is, the set of functions x in $C[0,t]$ such that $x(0)=0$; m will denote Wiener measure on $C_0[0,t]$.

A subset A of $C_0[0,t]$ is said to be scale-invariant measurable provided ρA is Wiener measurable for every $\rho>0$. It is easy to see that the class $\mathcal{S}$ of scale-invariant measurable sets forms a σ-algebra. N in $\mathcal{S}$ is said to be scale-invariant null provided $m(\rho N)=0$ for every $\rho>0$. A property which holds except on a scale-invariant null set is said to hold scale-invariant almost everywhere (s-a.e). For a rather detailed discussion of scale-invariant measurability and its relation with other topics see Ref. 7.

Let F be a C-valued (i.e. complex-valued) function on $C_0[0,t]$ which is s-a.e. defined and scale-invariant measurable and such that the Wiener integral

$$J(\lambda)=\int_{C_0[0,t]} F(\lambda^{-1/2}x)dm(x)$$

exists as a finite number for all $\lambda>0$. If there exists a function $J^*(\lambda)$ analytic in $C^+:=\{\lambda \text{ in } C:\mathrm{Re}\,\lambda>0\}$ such that $J^*(\lambda)=J(\lambda)$ for all $\lambda>0$, then $J^*(\lambda)$ is defined to be the analytic Wiener integral of F over $C_0[0,t]$ with parameter λ, and, for λ in C^+, we write

$$\int_{C_0[0,t]}^{\text{an } w_\lambda} F(x)dm(x):=J^*(\lambda).$$

Let q be a real parameter ($q\neq 0$) and let F be a function whose analytic Wiener integral exists for λ in C^+. If the following limit exists, we call it the analytic Feynman integral of F over $C_0[0,t]$ with parameter q, and we write

$$\int_{C_0[0,t]}^{\text{an } f_q} F(x)dm(x):=\lim_{\lambda\to -iq}\int_{C_0[0,t]}^{\text{an } w_\lambda} F(x)dm(x)$$

where λ approaches $-iq$ through C^+.

Remark: Equality s-a.e. is an equivalence relation for functions on $C_0[0,t]$. It is the appropriate equivalence relation for this setting as was discussed briefly in Ref. 4.

The definition of the Banach algebra S involves the Paley–Wiener–Zygmund (P.W.Z.) integral,[8] a relatively simple type of stochastic integral which we now define.

Let $\{\phi_j\}$ be a complete orthonormal set of $\mathbb{R}$-valued functions of bounded variation on $[0,t]$. For v in $L_2:=L_2[0,t]$ and n a positive integer, let

$$v_n(s):=\sum_{j=1}^{n}\left(\int_0^t v(\tau)\phi_j(\tau)d\tau\right)\phi_j(s). \tag{1.1}$$

The P.W.Z. integral $\int_0^t v(s)\tilde{d}x(s)$ is defined by

$$\int_0^t v(s)\tilde{d}x(s):=\lim_{n\to\infty}\int_0^t v_n(s)dx(s)$$

for all x in $C_0[0,t]$ for which the limit exists. See Doob's book[9] for an alternate method of defining this stochastic integral. Some of the key properties of the P.W.Z. integral are conveniently summarized in Ref. 5 with some additional information given in Ref. 4. We will limit ourselves here to a brief mention of a few facts: (i) For each v in $L_2[0,t]$ the P.W.Z. integral exists for s-a.e. x in $C_0[0,t]$. (ii) $\int_0^t v(s)\tilde{d}x(s)$ is a Borel function of (v,x) on $L_2[0,t]\times C_0[0,t]$. (iii) If v is of bounded variation on $[0,t]$, the P.W.Z. integral $\int_0^t v(s)\tilde{d}x(s)$ is s-a.e. equal to the Riemann–Stieltjes integral $\int_0^t v(s)dx(s)$. (iv) The P.W.Z. integral has the expected linearity properties when they are interpreted properly. [See Ref. 5 for a more detailed statement of (iv).]

It is natural to think of the condition $x(0)=0$ as simply a normalizing condition. Given x in $C[0,t]$, we define

$$\int_0^t v(s)\tilde{d}x(s):=\int_0^t v(s)\tilde{d}\,[x(s)-x(0)]$$

provided the integral on the right exists.

We are now ready to define the space S. Let $M(L_2)$ be the collection of C-valued, countably additive measures on

 0022-2488/82/112090-07$02.50 © 1982 American Institute of Physics

$\mathscr{B}(L_2)$, the Borel class of $L_2[0,t]$. $M(L_2)$ is a Banach algebra under the total variation norm where convolution is taken as the multiplication. Let σ be in $M(L_2)$. Consider the function $\tilde{\sigma}$ defined for s-a.e. x in $C_0[0,t]$ by the formula

$$\tilde{\sigma}(x) := \int_{L_2} \exp\left\{i\int_0^t v(s)\tilde{d}x(s)\right\} d\sigma(v). \tag{1.2}$$

An element of S is an equivalence class $[\tilde{\sigma}]$ of functions which are s-a.e. equal to $\tilde{\sigma}$ for some σ in $M(L_2)$. One often uses somewhat loose terminology and refers to the elements of S as functions, however the distinction between $\tilde{\sigma}$, $[\tilde{\sigma}]$, and $F \in [\tilde{\sigma}]$ will be important to us in certain places. Cameron and Storvick show that the correspondence $\sigma \to [\tilde{\sigma}]$ is injective (Ref. 1; Theorem 2.1) and carries convolution into pointwise multiplication. Letting $||[\tilde{\sigma}]|| := ||\sigma||$ we have that S is a Banach algebra. The analytic Feynman integral exists for every $[\tilde{\sigma}]$ in S (Ref. 1; Theorem 5.1).

In light of (1.2), there is a natural way of regarding F in $[\tilde{\sigma}]$ as defined on $C[0,t]$: If x in $C_0[0,t]$ is such that $F(x)$ is defined, then for any c in $\mathbb{R}$, take $F(x+c) := F(x)$.

Next we give the information necessary for our discussion of the Banach algebra $\mathscr{F}(H)$ of "Fresnel integrable functions." The fundamental work on the space $\mathscr{F}(H)$ was done by Albeverio and Høegh-Krohn.[10–12] They do a particularly effective job of relating their approach to the "Feynman integral" to problems in quantum mechanics and quantum field theory. Their approach was influenced by the earlier seminal work of Ito[13,14] and DeWitt-Morette.[15,16] In turn much of the impetus for the recent beautiful work of Truman[17–19] is found in the work of Albeverio and Høegh-Krohn.

Let H be a separable Hilbert space over $\mathbb{R}$. Let $M(H)$ be the collection of C-valued, countably additive measures on $\mathscr{B}(H)$, the Borel class of H. $M(H)$ is a Banach algebra under the total variation norm where convolution is taken as the multiplication. Given μ in $M(H)$, $\hat{\mu}$ is defined for every h in H by the formula $\hat{\mu}(h) := \int_H \exp\{i(h,h_1)\}d\mu(h_1)$. By definition, $\mathscr{F}(H) := \{\hat{\mu} : \mu \text{ is in } M(H)\}$. The correspondence $\mu \to \hat{\mu}$ is injective and carries convolution into pointwise multiplication. Hence, letting $||\hat{\mu}|| := ||\mu||$, we have that $\mathscr{F}(H)$ is a Banach algebra. The Fresnel integral $\mathscr{F}(\hat{\mu})$ is defined for $\hat{\mu}$ in $\mathscr{F}(H)$ by the formula

$$\mathscr{F}(\hat{\mu}) := \int_H \exp\{-(i/2)||h||^2\} d\mu(h). \tag{1.3}$$

There is a particular Hilbert space H which covers the applications of the Albeverio and Høegh-Krohn theory to ordinary quantum mechanics (but not to quantum field theory). It is this space which will concern us throughout the rest of this paper.

Let $\mathscr{H}$ be the space of $\mathbb{R}$-valued functions γ on $[0,t]$ which are absolutely continuous and with derivative $D\gamma$ in $L_2[0,t]$. Let $H := \{\gamma \text{ in } \mathscr{H} : \gamma(t) = 0\}$. The following inner product makes H into a separable Hilbert space over $\mathbb{R}$:

$$(\gamma_1,\gamma_2)_H := \int_0^t (D\gamma_1)(s)(D\gamma_2)(s)ds. \tag{1.4}$$

Given μ in $M(H)$ and γ in H

$$\hat{\mu}(\gamma) = \int_H \exp\left(i\int_0^t (D\gamma)(s)(D\gamma_1)(s)ds\right) d\mu(\gamma_1). \tag{1.5}$$

Note that $\hat{\mu}$ makes sense on $\mathscr{H}$ and, in fact, if γ is in H and c is in $\mathbb{R}$, $\hat{\mu}(\gamma + c) = \hat{\mu}(\gamma)$.

2. THE EQUIVALENCE OF THE TWO APPROACHES TO THE FEYNMAN INTEGRAL

Lemma 1 and Propositions 1 and 2 below are surely known, but since we would like this paper to be simple and reasonably self-contained, we will sketch parts of the proofs.

Lemma 1: Let v be in $L_2[0,t]$ and let γ be in $\mathscr{H}$. Then the P.W.Z. integral $\int_0^t v(s)\tilde{d}\gamma(s)$ exists and we have

$$\int_0^t v(s)\tilde{d}\gamma(s) = \int_0^t v(s)(D\gamma)(s)ds. \tag{2.1}$$

Proof: Let v_n be given by (1.1). The Riemann–Stieltjes integral (RS) $\int_0^t v_n(s)d\gamma(s)$ exists. It is not difficult to argue that $\int_0^t v_n(s)d\gamma(s)$ also exists and equals the Riemann–Stieltjes integral where we are thinking of $\int_0^t v_n(s)d\gamma(s)$ as a Lebesgue integral with respect to the measure induced by the function of bounded variation γ. Since $D\gamma$ is in $L_2[0,t] \subset L_1[0,t]$, we also have the formula $\int_0^t v_n(s)d\gamma(s) = \int_0^t v_n(s)(D\gamma)(s)ds$. Now $||v_n - v||_2 \to 0$, and so $v_n \to v$ weakly, and we can write

$$(\mathrm{RS})\int_0^t v_n(s)d\gamma(s) = \int_0^t v_n(s)(D\gamma)(s)ds \to \int_0^t v(s)(D\gamma)(s)ds$$

as desired.

Proposition 1: The differentiation map D is an isometric isomorphism of H onto $L_2[0,t]$. The integration map

$$(Iv)(s) := \int_t^s v(\tau)d\tau \tag{2.2}$$

is the inverse of D.

Proof: Given v in $L_2[0,t]$, let Iv be given by (2.2). Then Iv is absolutely continuous, $(Iv)(t) = 0$ and $DIv = v$. Hence Iv is in H and D maps H onto $L_2[0,t]$. D is injective because of the requirement that $\gamma(t) = 0$ for γ in H. The fact that D preserves inner products is built into the definition (1.4) of the inner product in H. ■

Since D and I are isometric isomorphisms, they are Borel measurable, and so

$$\mathscr{D}\mu := \mu \circ D^{-1}, \tag{2.3}$$

and

$$\mathscr{I}\sigma := \sigma \circ I^{-1} \tag{2.4}$$

map $M(H)$ into $M(L_2)$ and $M(L_2)$ into $M(H)$, respectively.

Proposition 2: $\mathscr{D}$ is a Banach algebra isometric isomorphism of $M(H)$ onto $M(L_2)$. $\mathscr{I} = \mathscr{D}^{-1}$.

Proof: This proof is again simple, but we will carry out some parts of it.

(a) $\mathscr{D}$ is onto. Let σ be in $M(L_2)$. Then $\mathscr{I}\sigma$ is in $M(H)$ and $\mathscr{D}(\mathscr{I}\sigma) = \sigma$ since $\sigma \circ I^{-1} \circ D^{-1} = \sigma$.

(b) $\mathscr{D}(\mu_1 * \mu_2) = \mathscr{D}(\mu_1) * \mathscr{D}(\mu_2)$. Let B be in $\mathscr{B}(L_2)$. Using the Change of Variables Theorem to justify the third equality, we can write

$$[\mathscr{D}(\mu_1) * \mathscr{D}(\mu_2)](B)$$

$$= \int_{L_2} (\mathscr{D}\mu_1)(B - v)d(\mathscr{D}\mu_2)(v)$$

$$= \int_{L_2} \mu_1[D^{-1}(B-v)]d(\mu_2 \circ D^{-1})(v)$$

$$= \int_H \mu_1[D^{-1}B - D^{-1}D\gamma]d\mu_2(\gamma)$$

$$= \int_H \mu_1[D^{-1}B - \gamma]d\mu_2(\gamma)$$

$$= (\mu_1 * \mu_2)(D^{-1}B)$$

$$= [\mathscr{D}(\mu_1 * \mu_2)](B).$$

(c) $||\mathscr{D}\mu|| = ||\mu||$. Given a Hilbert space H_0, let $C_b(H_0)$ denote the space of C-valued, bounded, continuous functions on H_0. $C_b(H_0)$ is a Banach space under the supremum norm. Let $B_1(C_b(H_0))$ denote the unit ball of $C_b(H_0)$. It is clear that every element μ_0 of $M(H_0)$ defines, via integration, an element of the dual of $C_b(H_0)$. In fact this imbedding of $M(H_0)$ in the dual of $C_b(H_0)$ is a Banach space isometric isomorphism. Hence we can write

$$||\mathscr{D}\mu|| = \sup\left\{\left|\int_{L_2} f(v)d(\mathscr{D}\mu)(v)\right| : f \text{ is in } B_1(C_b(L_2))\right\}$$

$$= \sup\left\{\left|\int_{L_2} f(v)d(\mu \circ D^{-1})(v)\right| : f \text{ is in } B_1(C_b(L_2))\right\}$$

$$= \sup\left\{\left|\int_H f(D\gamma)d\mu(\gamma)\right| : f \text{ is in } B_1(C_b(L_2))\right\}$$

$$\leqslant ||\mu||,$$

since $f \circ D$ is in $B_1(C_b(H))$ for every f in $B_1(C_b(L_2))$.

To get the opposite inequality, let σ in $M(L_2)$ be such that $\mathscr{I}\sigma = \mu$ and then argue much as above that $||\mathscr{I}\sigma|| \leqslant ||\sigma||$; that is, $||\mu|| \leqslant ||\mathscr{D}\mu||$. ■

Let μ be in $M(H)$ and let $\sigma = \mathscr{D}\mu$. In addition to the functions $\hat{\mu}$ and $\tilde{\sigma}$, we may consider the ordinary Fourier transform $\hat{\sigma}$ of σ. $\hat{\sigma}$ is defined for every u in $L_2[0,t]$ by the formula $\hat{\sigma}(u) := \int_{L_2} \exp\{i\int_0^t v(s)u(s)ds\}d\sigma(v)$. The next result describes the relationship between $\hat{\mu}$, $\tilde{\sigma}$, and $\hat{\sigma}$.

Theorem 1: Let $\sigma = \mathscr{D}\mu$ where μ is in $M(H)$. Then

$$\hat{\mu} = \hat{\sigma} \circ D = \tilde{\sigma}|_H. \tag{2.5}$$

Proof: Let γ be in H. By Lemma 1 and the Change of Variable Theorem we can write

$$\tilde{\sigma}(\gamma) = \int_{L_2} \exp\left\{i\int_0^t v(s)\dot{\gamma}(s)\right\}d\sigma(v)$$

$$= \int_{L_2} \exp\left\{i\int_0^t v(s)(D\gamma)(s)ds\right\}d\sigma(v)$$

$$= \hat{\sigma}(D\gamma)$$

$$= \int_{L_2} \exp\left\{i\int_0^t v(s)(D\gamma)(s)ds\right\}d(\mu \circ D^{-1})(v)$$

$$= \int_H \exp\left\{i\int_0^t (D\gamma_1)(s)(D\gamma)(s)ds\right\}d\mu(\gamma_1)$$

$$= \hat{\mu}(\gamma).$$

The correspondence between the Banach algebras $\mathscr{F}(H)$ and S can now be easily established by using the facts assembled above and certain facts already in the literature.

Theorem 2: The map $\phi: S \to \mathscr{F}(H)$ defined by $\phi([\tilde{\sigma}]) = \tilde{\sigma}|_H$ identifies S and $\mathscr{F}(H)$ isometrically and isomorphically as Banach algebras. The inverse of ϕ is given by $\phi^{-1}(\hat{\mu}) = [\widetilde{\mathscr{D}\mu}]$.

Remark: The maps ϕ and ϕ^{-1} can alternately be described in a way that is less precise but perhaps easier to remember. Let $\tilde{\sigma}$ "in" S be given by (1.2), then $\phi(\tilde{\sigma}) = \tilde{\sigma}|_H$; or given $\hat{\mu}$ in $\mathscr{F}(H)$, $\phi^{-1}(\hat{\mu})$ is the unique element $\tilde{\sigma}$ "in" S such that $\tilde{\sigma}|_H = \hat{\mu}$.

Proof: Let $[\tilde{\sigma}]$ be in S. Let $\mu = \mathscr{I}\sigma$ with $\mathscr{I}$ given by (2.4). Cameron and Storvick have shown[1] that the map ϕ_1 sending $[\tilde{\sigma}]$ to σ is a Banach algebra isometric isomorphism of S onto $M(L_2)$. By Proposition 2, $\mathscr{I}$, sending σ to $\mathscr{I}\sigma = \mu$, is an isometric isomorphism of $M(L_2)$ onto $M(H)$. Further, the map ϕ_2 sending $\mathscr{I}\sigma = \mu$ to $\widehat{\mathscr{I}\sigma} = \hat{\mu}$ is an isometric isomorphism of $M(H)$ onto $\mathscr{F}(H)$ as is known and is discussed by Albeverio and Høegh-Krohn.[11] Hence $\phi := \phi_2 \circ \mathscr{I} \circ \phi_1$ is a Banach algebra isometric isomorphism of S onto $\mathscr{F}(H)$. Finally, by Theorem 1, the action of $\phi([\tilde{\sigma}]) = \hat{\mu} = \widehat{\mathscr{I}\sigma}$ on H agrees with the action of $\tilde{\sigma}$ on H.

We end this section by showing that there is a simple relationship between the Fresnel integral of $\hat{\mu}$ and the analytic Feynman integral of the corresponding $\tilde{\sigma}$.

Theorem 3: Let μ belong to $\mathscr{F}(H)$ and let $[\widetilde{\mathscr{D}\mu}]$ be the corresponding element of S. Then the Fresnel integral of $\hat{\mu}$ equals the analytic Feynman integral with parameter $q = 1$ of $\widetilde{\mathscr{D}\mu}$; that is

$$\int_{C_0[0,t]}^{\text{an } f_1} (\widetilde{\mathscr{D}\mu})(x)dm(x) = \mathscr{F}(\hat{\mu}). \tag{2.6}$$

In fact, for any F in $[\widetilde{\mathscr{D}\mu}]$,

$$\int_{C_0[0,t]}^{\text{an } f_1} F(x)dm(x) = \mathscr{F}(\hat{\mu}).$$

Proof: Cameron and Storvick (Ref. 1; Theorem 5.1) show that

$$\int_{C_0[0,t]}^{\text{an } f_1} (\widetilde{\mathscr{D}\mu})(x)dm(x) = \int_{L_2} \exp\{||v||_2^2|2i\}d(\mathscr{D}\mu)(v).$$

Albeverio and Høegh-Krohn (Ref. 11; p. 18) have the formula

$$\mathscr{F}(\hat{\mu}) = \int_H \exp\{-i||\gamma||_H^2|2\}d\mu(\gamma).$$

But by the Change of Variables Theorem

$$\int_{L_2} \exp\{||v||_2^2|2i\}d(\mathscr{D}\mu)(v)$$

$$= \int_{L_2} \exp\{-i||v||_2^2|2\}d(\mu \circ D^{-1})(v)$$

$$= \int_H \exp\{-i||D\gamma||_2^2|2\}d\mu(\gamma)$$

$$= \int_H \exp\{-i||\gamma||_H^2|2\}d\mu(\gamma).$$

3. SOME CONSEQUENCES OF THE EQUIVALENCE

In this section we give some simple consequences of the theory in Sec. 2 and make some general comments, based on our experience to date, on how to make use of the equivalence between S and $\mathscr{F}(H)$. We intend to pursue this topic further in later work.

There are of course certain results which Theorems 1, 2, and 3 allow us to immediately carry over from one theory to the other. For example, any statements about S and $\mathscr{F}(H)$ as

Banach algebras are obviously of this type. Certain other results can be obtained with a minimal amount of effort. For example, there are certain functions on Wiener space which are of physical interest and which have been shown to be in S[1,4,5] but which were not known to be in $\mathcal{F}(H)$. These are functions arising from certain time dependent[1,4] and quadratic potentials,[5] respectively. In fact, before the present work, it was believed by this author and others that the functions arising from quadratic potentials[5] were examples of functions in S whose "restrictions" to H were not in $\mathcal{F}(H)$. However, the theory in Sec. 2 and the results of Ref. 5 allow one to show quite easily that these functions are in $\mathcal{F}(H)$. We now formally state the results we have just been discussing.

Corollary 1: Let $\theta:[0,t]\times\mathbb{R}\to C$ be a function which for each s in $[0,t]$ is the Fourier–Stieltjes transform of a C-valued, countably additive Borel measure σ_s on $\mathbb{R}$; that is,

$$\theta(s,u)=\int_{\mathbf{R}}\exp\{iuv\}d\sigma_s(v). \tag{3.1}$$

We suppose that for each Borel subset B of $\mathbb{R}$, $\sigma_s(B)$ is a Borel measurable function of s and that $||\sigma_s||$ is in $L_1[0,t]$. Then the function

$$f(\gamma):=\exp\left\{\int_0^t\theta(s,\gamma(s)-\gamma(0))ds\right\} \tag{3.2}$$

is in $\mathcal{F}(H)$.

Remarks: (i) The multivariable version of this with $\mathbb{R}$ replaced by $\mathbb{R}^v$ is valid and no harder to treat. (ii) When the potential θ is independent of s, the functions (3.2) are well-known to be in $\mathcal{F}(H)$ and have been extensively treated in Refs. 10–12,19.

Corollary 2 is the only result in this paper which we will state in a multivariable setting. We do so, since, in this case, going from the one variable to the multivariable setting is not entirely routine. Here H^v will denote the space of functions $\vec{\gamma}$ on $[0,t]$ to $\mathbb{R}^v$ each of whose components belongs to H.

Corollary 2: Let

$$g(\vec{\gamma}):=\exp\left\{-\int_0^t(A(s)[\vec{\gamma}(s)-\vec{\gamma}(0)],[\vec{\gamma}(s)-\vec{\gamma}(0)])ds\right\}, \tag{3.3}$$

where $\{A(s)=(a_{ij}(s)):0\leqslant s\leqslant t\}$ is a commutative family of v by v real, symmetric, positive definite matrices such that the (necessarily positive) eigenvalues $\{p_1(s),\ldots,p_v(s)\}$ have square roots which are of bounded variation on $[0,t]$. Then g is in $\mathcal{F}(H^v)$.

Remarks: (i) It is easy to show (Ref. 5; Corollary 1) that the hypothesis on the eigenvalues is satisfied if the functions a_{ij} are continuous and of bounded variation on $[0,t]$. (ii) Since $\mathcal{F}(H^v)$ is a Banach algebra, one has immediately that products of functions as in Corollary 2, and the v-dimensional version of Corollary 1, are also in $\mathcal{F}(H^v)$.

We finish this discussion with a few comments on the proof of Corollary 2. Virtually the same comments can be made about the proof of Corollary 1. The function (3.3) can be thought of as acting on the v–dimensional space $C^v[0,t]$. Let us denote that function by G. It was shown in Ref. 5 that G "belongs" to the Banach algebra S; more precisely, it was shown that there exists σ in $M(L_2^v[0,t])$ such that G is in $[\tilde{\sigma}]$. Since the theory in Sec. 2 tells us that $\tilde{\sigma}|_{H^v}$ is in $\mathcal{F}(H^v)$, one might think that Corollary 2 follows immediately. The situation is a bit more complicated than that however since H^v is a negligible subset of $C^v[0,t]$, and G in $[\tilde{\sigma}]$ implies just that $G=\tilde{\sigma}$ s-a.e. on Wiener space. However, if one reads the proofs in Ref. 5, one sees that $G=\tilde{\sigma}$ on H^v and so Corollary 2 follows. Some minor modifications of the proofs in Ref. 5 are needed to see this, but the modifications are actually simplifications and are rather obvious and so won't be included here. [The essential idea is that in arguments in Ref. 5 that are made for m$\times$m-a.e. (x,w), replace x by fixed γ in H and argue for m-a.e. w.]

When Corollaries 1 and 2 are combined with various theorems of Albeverio and Høegh-Krohn[10,11] and Truman,[19] one obtains further corollaries. Specifically several theorems from Refs. 10, 11, 19 involve a hypothesis that a function involved is in $\mathcal{F}(H)$. All such results are now seen to apply to the functions (3.2) and (3.3) and their products. Certain other theorems from Refs. 10, 12, and 19 are done for special classes of functions in $\mathcal{F}(H)$. It seems likely that at least some of these results are valid for functions of the forms (3.2) and (3.3), but this remains to be investigated.

Theorem 2 of Truman's paper[19] explores the connection between his "Feynman" map acting on $\mathcal{F}(H)$ and the Wiener integral. We will see that the work in Sec. 2 allows us to extend and to better understand Truman's Theorem.

Truman shows that for $f=\hat{\mu}$ in $\mathcal{F}(H)$, the Feynman map $\mathcal{F}^s(\mathcal{I}ms\leqslant 0)$ is given by the formula

$$\mathcal{F}^s(f)=\int_H\exp\{-is||\gamma||_H^2|2\}d\mu(\gamma). \tag{3.4}$$

Note that $\mathcal{F}^1(f)$ is just the Fresnel integral (1.3), $\mathcal{F}(f)$, of f. Since the result of present interest only concerns the action of $\mathcal{F}^s$ on $\mathcal{F}(H)$, we can and will simplify our discussion by regarding (3.4) as the definition of $\mathcal{F}^s$.

Next we state Truman's result. By $C_0(0,t)$ Truman means the space of continuous functions on $[0,t]$ which vanish at t. E denotes the integral with respect to the Wiener measure on $C_0(0,t)$. As noted earlier the functions in S and $\mathcal{F}(H)$ may be thought of as acting on $C[0,t]$ and $\mathcal{H}$ respectively in a natural way, and the fact that the elements of $C_0(0,t)$ vanish at t rather than at 0 is actually of no consequence for our present purposes.

Theorem 4: (Ref. 19, p. 80) Let f be in $\mathcal{F}(H)$, then $\mathcal{F}^s(f)$ is a regular analytic function of s in $\mathcal{I}ms<0$, continuous in $\mathcal{I}ms\leqslant 0$. If f in $\mathcal{F}(H)$ is a continuous functional $f:C_0(0,t)\to C$, then

$$\mathcal{F}(f)=\lim_{\epsilon\to 0^+}\mathcal{F}^{1-i\epsilon}(f),\quad E(f)=\lim_{\epsilon\to 0}\mathcal{F}^{-i+\epsilon}(f). \tag{3.5}$$

Interpolation gives

$$|\mathcal{F}^{se^{i\alpha}}(f)|\leqslant||f||^{1-(2\alpha)/\pi}||f||_\infty^{(2\alpha)/\pi},$$
$$0\leqslant\alpha\leqslant\pi/2,\quad s>0, \tag{3.6}$$

where $||f||_\infty=\sup\{|f(\gamma)|:\gamma \text{ is in } C_0(0,t)\}$.

Remarks: (i) The first assertion is a straightforward consequence of (3.4). The second assertion follows immediately from the first as soon as one shows that $E(f)=\mathcal{F}^{-i}(f)$. The inequality (3.6) is easy for $\alpha=0$ and $\alpha=\pi/2$, and, for gen-

eral α, is then a consequence of the Hadamard Three Lines Theorem (Ref. 20; p. 33). (ii) Note that the second and third assertions of the theorem deal with f in $\mathcal{F}(H)$ for which it is *assumed* that there is an extension to a function which is defined on all of $C_0(0,t)$ which is continuous in the sup norm. However, the work in Sec. 2 assures us that for *every* $f=\hat{\mu}$ in $\mathcal{F}(H)$, there is associated a unique element $[\tilde{\sigma}]$ in S such that $\tilde{\sigma}|_H = f$. We will see below that if we take any F in $[\tilde{\sigma}]$ as the "extension" of f, the second and third assertions of Truman's Theorem go through.

Corollary 3: Let $f=\hat{\mu}$ be in $\mathcal{F}(H)$, then $\mathcal{F}^s(f)$ is a regular analytic function of s in $\mathcal{I}ms<0$, continuous in $\mathcal{I}ms\leqslant 0$. Let $\sigma=\mathcal{D}\mu$ so that $\tilde{\sigma}|_H = f$. Then for any F in $[\tilde{\sigma}]$ we have

$$\int_{C_0[0,t]} F(x)dm(x) = \mathcal{F}^{-i}(f), \tag{3.7}$$

and so

$$\mathcal{F}(f) = \lim_{\epsilon\to 0^+} \mathcal{F}^{1-i\epsilon}(f),$$

$$\int_{C_0[0,t]} F(x)dm(x) = \lim_{\epsilon\to 0} \mathcal{F}^{-i+\epsilon}(f). \tag{3.8}$$

Interpolation gives

$$|\mathcal{F}^{se^{i\alpha}}(f)| \leqslant ||f||^{1-(2\alpha)/\pi} ||F||_\infty^{(2\alpha)/\pi}, \quad 0\leqslant\alpha\leqslant\pi/2, \quad s>0. \tag{3.9}$$

Proof: Our first assertion is exactly the same as Truman's first assertion. The second follows immediately from the work in Sec. 2. We now show (3.7). (3.8) will then follow immediately.

$$\begin{aligned}
&\int_{C_0[0,t]} F(x)dm(x) \\
&\quad = \int_{C_0[0,t]} \tilde{\sigma}(x)dm(x) \\
&\quad = \int_{C_0[0,t]} \left[\int_{L_2} \exp\left\{i\int_0^t v(s)\tilde{d}x(s)\right\} d\sigma(v)\right] dm(x) \\
&\quad = \int_{L_2} \left[\int_{C_0[0,t]} \exp\left\{i\int_0^t v(s)\tilde{d}x(s)\right\} dm(x)\right] d\sigma(v).
\end{aligned}$$

Applying a basic Wiener integration formula [Ref. 5; (2.1)] to this last expression and then using the fact that $(2\pi)^{-1/2}\exp(-u^2/2)$ is its own Fourier transform, we obtain

$$\begin{aligned}
&\int_{C_0[0,t]} F(x)dm(x) \\
&\quad = \int_{L_2} \left[(2\pi)^{-1/2}\int_{\mathbb{R}} \exp\{i||v||_2 u\}\exp(-u^2/2)du\right] d\sigma(v) \\
&\quad = \int_{L_2} \exp(-||v||_2^2/2)d\sigma(v) \\
&\quad = \int_{L_2} \exp(-||v||_2^2/2)d(\mu\circ D^{-1})(v) \\
&\quad = \int_H \exp(-||D\gamma||_2^2/2)d\mu(\gamma) \\
&\quad = \int_H \exp(-||\gamma||_H^2/2)d\mu(\gamma) \\
&\quad = \mathcal{F}^{-i}(f),
\end{aligned}$$

where this last equality follows from (3.4).

The argument we have just made can be trivially extended to show that for $s>0$,

$$\mathcal{F}^{-is}(f) = \int_{C_0[0,t]} F(s^{-1/2}x)dm(x).$$

(3.9) now follows immediately for $\alpha=\pi/2$ and $\alpha=0$. The Hadamard Three Lines Theorem yields the result for general α just as in Truman's setting.

There are many further aspects of the relationship between $\mathcal{F}(H)$ and S which remain to be explored. Also it seems likely that the work of Albeverio and Høegh-Krohn which is directed toward quantum field theory ought to have a "Wiener space" counterpart. However, we will be satisfied in this paper with one additional application. Before doing that we give a few rather vague but perhaps helpful general impressions derived from our (still limited) experience in applying the results of Sec. 2.

(1) Once one realizes the nature of the correspondence between S and $\mathcal{F}(H)$, the formal outline of arguments often translates readily from one space to the other.

(2) In order to get information about specific functions f in $\mathcal{F}(H)$ out of information about the corresponding "function" in S, some additional argument is needed because H is a negligible subset of $C[0,t]$. It appears that such arguments will usually be rather simple. (This point was already discussed a little in connection with Corollaries 1 and 2.)

(3) When one is attempting to carry results from the Hilbert space setting to the Wiener space setting, it seems that some measure–theoretic subtleties may arise. This will be illustrated somewhat by our next application, but, in this case, a well known result about translation in Wiener space immediately takes care of the situation. In another result which we have just recently proved, the subtleties are not so readily overcome and, indeed, are intimately related to an apparently as yet unanswered question asked earlier in Ref. 7, p. 165. This result seems too involved to include since it is our hope to keep this paper simple and reasonably self–contained.

(4) Is the Hilbert space setting or the Wiener space setting more convenient to work in? Hilbert spaces of course have many pleasant properties, and one seems less likely to encounter delicate measure–theoretic questions in that setting. On the other hand a great deal is known about Wiener space, and there is a certain probabilistic intuition which one can bring to bear. In point of fact, certain things have been done in each setting which were not done in the other. Which setting is more convenient is probably, at least in part, a function of one's background and the particular problem at hand.

There is concern by some that the space of paths H is not appropriate for quantum mechanics. Feynman commented in his original paper[21] that the paths which should contribute most to his integral are the highly nondifferentiable paths familiar from Wiener space. Along related lines, Truman comments (Ref. 18, p. 1745) that the paths in H may be too smooth to provide a good model for quantum mechanics. Whether this should be considered a serious problem or not is not really clear to us, but the results relating $\mathcal{F}(H)$ and S say that, at least for a certain large class of

potentials, it essentially doesn't matter whether the underlying space of paths is taken to be H or Wiener space.

In their monograph (Ref. 11, pp. 19–20), Albeverio and Høegh-Krohn showed how the Fresnel integral is transformed under a translation. In a recent paper,[2] Cameron and Storvick have proved a similar result in the setting of the Banach algebra S. We will show how Cameron and Storvick's result can be obtained from Albeverio and Høegh-Krohn's result via the work in Sec. 2. We should mention that both results have their roots in the Cameron and Martin Translation Theorem for Wiener space.[22]

We will need two simple results about how certain measures are transformed by the map $\mathscr{D}: M(H) \to M(L_2)$. We need only very special cases of this for our present purposes, but we state rather general Propositions since they seem likely to be useful in the future. We omit the simple proofs.

Proposition 3: Let μ be in $M(H)$ and let $\sigma = \mathscr{D}\mu$ be the corresponding measure in $M(L_2)$.

(a) Let h be a C-valued Borel measurable function on H. Then h is in $L_1(\mu)$ if and only if $h \circ I$ is in $L_1(\sigma)$. Further, if h is in $L_1(\mu)$ and μ_h is defined by $d\mu_h(\gamma) := h(\gamma)d\mu(\gamma)$, then μ_h is in $M(H)$ and $\mathscr{D}(\mu_h) = \sigma_{h \circ I}$, where $d\sigma_{h \circ I}(v) := h(I(v))d\sigma(v)$.

(b) Let g be a C-valued Borel measurable function on $L_2[0,t]$. Then g is in $L_1(\sigma)$ if and only if $g \circ D$ is in $L_1(\mu)$. Further, if g is in $L_1(\sigma)$ and σ_g is defined by $d\sigma_g(v) := g(v)d\sigma(v)$, then σ_g is in $M(L_2)$ and $\mathscr{I}(\sigma_g) = \mu_{g \circ D}$, where $d\mu_{g \circ D}(\gamma) = g(D(\gamma))d\mu(\gamma)$.

Proposition 4: Let μ be in $M(H)$ and let $\sigma = \mathscr{D}\mu$ be the corresponding measure in $M(L_2)$.

(a) Let R_H be an injective map from H into H which carries Borel sets to Borel sets. Then $\mu \circ R_H$ is in $M(H)$ and $\mathscr{D}(\mu \circ R_H) = \sigma \circ R_L$, where $R_L = D \circ R_H \circ I$.

(b) Let R_L be an injective map from $L_2[0,t]$ into $L_2[0,t]$ which carries Borel sets to Borel sets. Then $\sigma \circ R_L$ is in $M(L_2)$ and $\mathscr{I}(\sigma \circ R_L) = \mu \circ R_H$, where $R_H = I \circ R_L \circ D$.

Given a in $\mathscr{H}$, let $T_a: C[0,t] \to C[0,t]$ be defined by $T_a x := x + a$. The restriction of T_a to $\mathscr{H}$ carries $\mathscr{H}$ into $\mathscr{H}$.

Theorem 5: (Albeverio and Høegh-Krohn). Let a be in $\mathscr{H}$ and let $f = \hat{\mu}$ be in $\mathscr{F}(H)$. Let $f_a := f \circ T_a$ and define $f^{(a)}: \mathscr{H} \to C$ by $f^{(a)}(\gamma) := \exp\{-i(a,\gamma)_H\} f(\gamma)$. Then $f_a = \hat{\mu}_1$ and $f^{(a)} = \hat{\mu}_2$ are in $\mathscr{F}(H)$, where μ_1 and μ_2 are defined by $d\mu_1(\gamma) := \exp\{i(a,\gamma)_H\} d\mu(\gamma)$ and $\mu_2 := \mu \circ T_a$. Further

$$\mathscr{F}(f_a) = \exp\{i||a||_H^2/2\}\mathscr{F}(f^{(a)}). \tag{3.10}$$

Corollary 4: (Cameron and Storvick). Let a be in $\mathscr{H}$ and let $F \in [\tilde{\sigma}]$ where $[\tilde{\sigma}]$ is in S. Let $F_a := F \circ T_a$ and define $F^{(a)}: C[0,t] \to C$ by $F^{(a)}(x) := \exp\{i\int_0^t (Da)(s)\tilde{d}x(s)\} F(x)$. Then $F_a \in [\tilde{\sigma}_1]$ and $F^{(a)} \in [\tilde{\sigma}_2]$, where σ_1 and σ_2 are defined by $d\sigma_1(v) := \exp\{i\int_0^t v(s)\tilde{d}a(s)\} d\sigma(v)$ and $\sigma_2 := \sigma \circ T_{Da}$. Further

$$\int_{C_0[0,t]}^{\text{an } f_1} F_a(x)dm(x) = \exp\{i||a||_H^2/2\} \int_{C_0[0,t]}^{\text{an } f_1} F^{(a)}(x)dm(x).$$

Proof: Let $\mu = \mathscr{I}\sigma$ and let $f = \hat{\mu}$. Let μ_1 and μ_2 be as in Theorem 5. Let $\tau_1 = \mathscr{D}\mu_1$ and $\tau_2 = \mathscr{D}\mu_2$. By Proposition 3 and Lemma 1,

$$\begin{aligned} d\tau_1(v) &= \exp\{i(a,Iv)_H\} d\sigma(v) \\ &= \exp\left\{i\int_0^t (Da)(s)(DIv)(s)ds\right\} d\sigma(v) \\ &= \exp\left\{i\int_0^t v(s)\tilde{d}a(s)\right\} d\sigma(v) = d\sigma_1(v). \end{aligned}$$

By Proposition 4, $\tau_2 = \sigma \circ D \circ T_a \circ I$. Thus for B in $\mathscr{B}(L_2)$, $\tau_2(B) = \sigma(DIB + Da) = (\sigma \circ T_{Da})(B) = \sigma_2(B)$. Thus $\sigma_1 = \mathscr{D}\mu_1$ and $\sigma_2 = \mathscr{D}\mu_2$.

By Theorem 5, $\mathscr{F}(\hat{\mu}_1) = \exp\{i||a||_H^2/2\} \mathscr{F}(\hat{\mu}_2)$. It follows immediately from Theorem 3 that

$$\int_{C_0[0,t]}^{\text{an } f_1} \tilde{\sigma}_1(x)dm(x) = \exp\{i||a||_H^2/2\} \int_{C_0[0,t]}^{\text{an } f_1} \tilde{\sigma}_2(x)dm(x).$$

Thus the theorem will be proved as soon as we show that $F_a \in [\tilde{\sigma}_1]$ and $F^{(a)} \in [\tilde{\sigma}_2]$. (It is in showing that $F_a \in [\tilde{\sigma}_1]$ that one needs to be a little careful with the measure theory.) We write $F \approx G$ when $F = G$ s-a.e.

$$\begin{aligned} F_a \in [\tilde{\sigma}_1]: \tilde{\sigma}_1(x) &\approx \int_{L_2} \exp\left\{i\int_0^t v(s)\tilde{d}x(s)\right\} d\sigma_1(v) \\ &\approx \int_{L_2} \exp\left\{i\int_0^t v(s)\tilde{d}\,[x(s) + a(s)]\right\} d\sigma(v) \\ &\approx \tilde{\sigma}(x + a). \end{aligned}$$

The proof that $F_a \in [\tilde{\sigma}_1]$ will be finished if we show that $\tilde{\sigma}(x + a) = F(x + a)$ for s-a.e. x in $C_0[0,t]$. Let $N := \{x \text{ in } C_0[0,t]: F(x) \neq \tilde{\sigma}(x)\}$. N is scale-invariant null since $F \in [\tilde{\sigma}]$. Now $\tilde{\sigma}(x + a) = F(x + a)$ or, equivalently, $\tilde{\sigma}[x + a - a(0)] = F[x + a - a(0)]$ except for x's such that x is in $N - a + a(0)$. So it suffices to show that $\rho N - \rho a + \rho a(0)$ is Wiener null where ρ is an arbitrary positive number. But $m(\rho N) = 0$ since N is scale-invariant null. Also $\rho a - \rho a(0)$ is in $\mathscr{H}$, and, it is well known from the translation theory in Wiener space, that translation by such elements preserves sets of measure 0.

$$\begin{aligned} F^{(a)} \in [\tilde{\sigma}_2]: \tilde{\sigma}_2(x) &\approx \int_{L_2} \exp\left\{i\int_0^t v(s)\tilde{d}x(s)\right\} d\sigma_2(v) \\ &\approx \int_{L_2} \exp\left\{i\int_0^t v(s)\tilde{d}x(s)\right\} d(\sigma \circ T_{Da})(v) \\ &\approx \int_{L_2} \exp\left\{i\int_0^t [v(s) - (Da)(s)]\tilde{d}x(s)\right\} d\sigma(v) \\ &\approx \exp\left\{-i\int_0^t (Da)(s)\tilde{d}x(s)\right\} F(x) \\ &\approx F^{(a)}(x). \end{aligned}$$

ACKNOWLEDGMENTS

This work was begun while the author was on leave at the Mathematisches Institut, Universität Erlangen–Nürnberg. The author gratefully acknowledges the support of the University of Nebraska as well as the support and hospitality of the Institut and especially the kind help of Professor Kölzow and Dr. Tischer.

[1] R. H. Cameron and D. A. Storvick, "Some Banach algebras of analytic Feynman integrable functionals," in *Analytic Functions, Kozubnik, 1979, Springer Lecture notes in Mathematics*, Vol. 798 (Springer, Berlin, 1980), pp. 18–67.

[2]R. H. Cameron and D. A. Storvick, "A new translation theorem for the analytic Feynman integral" Preprint, U. of Minnesota, Minneapolis, Minnesota.
[3]R. H. Cameron and D. A. Storvick, "Analytic Feynman integral solutions of an integral equation related to the Schroedinger equation," J. D'Analyse Math. **38**, 34–66 (1980).
[4]G. W. Johnson and D. L. Skoug, "Notes on the Feynman integral, I," Pac. J. Math. **93**, 313–324 (1981).
[5]G. W. Johnson and D. L. Skoug, "Notes on the Feynman integral, II," J. Func. Anal. **41**, 277–289 (1981).
[6]G. W. Johnson and D. L. Skoug, "Notes on the Feynman integral, III: The Schroedinger equation," to appear in the Pac. J. Math.
[7]G. W. Johnson and D. L. Skoug, "Scale-invariant measurability in Wiener space," Pac. J. Math. **83**, 157–176 (1979).
[8]R. E. A. C. Paley, N. Wiener, and A. Zygmund, "Notes on random functions," Math. Z. **37**, 647–688 (1933).
[9]J. L. Doob, *Stochastic Processes* (Wiley, New York, 1953).
[10]S. Albeverio and R. Høegh-Krohn, "Feynman path integrals and the corresponding method of stationary phase," in *Feynman Path Integrals* (Marseille, 1978), *Springer Lecture Notes in Physics*, Vol. **106** (Springer, 1978), pp. 3–57.
[11]S. Albeverio and R. Høegh-Krohn, "Mathematical Theory of Feynman Path Integrals," *Springer Lecture Notes in Mathematics*, Vol. **523** (Springer, Berlin, 1976).
[12]S. Albeverio and R. Høegh-Krohn, "Oscillatory integrals and the method of stationary phase in infinitely many dimensions, with applications to the classical limit of quantum mechanics, I," Invent. Math. **40**, 59–106 (1977).
[13]K. Ito, "Wiener integral and Feynman integral," *Proceedings of the 4th Berkeley Symposium Mathematical Statistics and Probability* (Univ. California Press, Berkeley, 1961), Vol. II, pp. 227–238.
[14]K. Ito, "Generalized uniform complex measures in the Hilbertian metric space with their application to the Feynman path integral," *Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability*, (Univ. California Press, Berkeley, 1961), Vol. II, pp. 145–161.
[15]C. M. DeWitt-Morette, "Feynman path integral definition without limiting procedure," Comm. Math. Phys. **28**, 47–67 (1972).
[16]C. M. DeWitt-Morette, "Feynman path integrals, I. Linear and affine techniques, II. The Feynman–Green function," Comm. Math. Phys. **37**, 63–81 (1974).
[17]A. Truman, "Some applications of vector space measures to non-relativistic quantum mechanics," Dublin, 1977, *Springer Lecture Notes in Mathematics*, Vol. **644** (Springer, Berlin, 1978), pp. 418–451.
[18]A. Truman, "The Feynman maps and the Wiener integral," J. Math. Phys. **19**, 1742–1750 (1978).
[19]A. Truman, "The polygonal path formulation of the Feynman path integral," in *Feynman Path Integrals* (Marseille, 1978), *Springer Lecture Notes in Physics*, Vol. **106** (Springer, Berlin, 1978), pp. 73–102.
[20]M. Reed and B. Simon, *Fourier Analysis, Self-Adjointness* (Academic, New York, 1975).
[21]R. P. Feynman, "Space–time approach to nonrelativistic quantum mechanics," Rev. Mod. Phys. **20**, 367–387 (1948).
[22]R. H. Cameron and W. T. Martin, "Transformations of Wiener integrals under translations," Ann. Math. **45**, 386–396 (1944).

Approximate methods for the solution of the Chandrasekhar *H*-equation

C. T. Kelley
Department of Mathematics, North Carolina State University, Raleigh, North Carolina 27650

(Received 23 April 1982; accepted for publication 25 June 1982)

We consider two methods of approximate solution to matrix valued analogs of the Chandrasekhar *H*-equation. We give conditions under which they converge. The first method is a generalization of approximation of the integral by a quadrature. The second is Newton's method.

PACS numbers: 02.60.Cb, 95.30.Jn

I. INTRODUCTION

Various analogs of the Chandrasekhar *H*–equation,[1]

$$H(\mu) = 1 + \frac{c}{2}\int_0^1 \frac{\mu}{\mu+\nu} H(\nu)d\psi\, H(\mu), \tag{1}$$

are useful in the solution of exit distribution problems in radiative transfer and neutron transport.[1-7] In this paper we consider two types of approximate methods for solution of these problems.

For example, it is known that Eq. (1) has, for $0<c<1$, two solutions, only one of which is of importance physically. This solution, which we denote by H, is analytical in c for $|c|<1$ and continuous for $|c|\leqslant 1$, $0\leqslant\mu\leqslant 1$.

If we approximate $d\nu$ by

$$d\psi_N(\nu) = \sum_{i=1}^{N} a_{N_i}\,\delta(\nu - x_{N_i})\,d\nu, \tag{2}$$

with

$$a_{N_i} > 0\,,$$
$$\int_0^1 d\psi_N = 1\,, \tag{3}$$

and let H_N be the physical solution to

$$H_N(\mu) = 1 + \frac{c}{2} H_N(\mu)\int_0^1 \frac{\mu}{\mu+\nu} H_N(\nu)d\psi_N(\nu)\,, \tag{4}$$

one would like to prove that H_N converges to H in some sense. If a_{N_i} and x_{N_i} are the weights and nodes for a quadrature scheme, Kelley[8] has given conditions under which H_N converges to H uniformly in μ for any fixed c, $|c|<1$. In a different situation, Masson[9] has shown convergence of a scheme of this type for $0<c<\frac{8}{9}$.

The methods of Ref. 8 require estimates of $H_N(\mu)$ that are independent of N. Such estimates cannot be made for $c=1$ and only partial success for $0<c<1$ has been made in matrix-valued problems when these methods are used.

In Sec. II of this paper we show that under certain hypothesis approximations of the form of Eq. (4) converge to the physical solutions of the matrix valued analogs of Eq. (1) that arise in multigroup neutron transport[2] for $|c|<1$.

We require some notation that allows us to write these matrix-valued *H*-equations in a compact way.[3,10] The *H*-equation for multigroup neutron transport[2] is a coupled system of equations for two $n\times n$ matrix-valued functions H_l and H_r

$$H_l(\mu) = I + \frac{c\mu}{2}\int_0^1 \frac{\Psi(\nu)H_r(\nu)}{\mu+\nu}\,d\nu\, H_l(\mu)\,,$$
$$H_r(\mu) = I + \frac{c\mu}{2} H_r(\mu)\int_0^1 \frac{H_l(\nu)\Psi(\nu)}{\mu+\nu}\,d\nu\,. \tag{5}$$

Here $\Psi(\nu)$ is a known $n\times n$ matrix with non-negative entries. We normalize Ψ by requiring $\| \int_0^1 \Psi(\nu)d\nu \|_{sp} = 1$, where $\|\ \|_{sp}$ denotes spectral radius. With this normalization H_r and H_l exist for all $|c|\leqslant 1$.

For convenience we let X denote the space of $2n\times 2n$ matrices of the form

$$A = \begin{pmatrix} A_l & 0 \\ 0 & \tilde{A}_r \end{pmatrix}, \tag{6}$$

where A_r and A_l are $n\times n$ matrices and $\sim$ denotes transpose. We define a norm on X by

$$\|A\|_X = \sum_{i,j=1}^{n} |a_{ij}^{(r)}| + |a_{ij}^{(l)}|\,, \tag{7}$$

where $a_{ij}^{(r)}$ and $a_{ij}^{(l)}$ are the entries of A_r and A_l.

For $A\in X$ given by Eq. (6) we define A^* by

$$A^* = \begin{pmatrix} A_r & 0 \\ 0 & \tilde{A}_l \end{pmatrix}, \tag{8}$$

We define H and D as X-valued functions of $\mu\in[0,1]$ by

$$H = \begin{pmatrix} H_l & 0 \\ 0 & \tilde{H}_r \end{pmatrix},$$
$$D = \begin{pmatrix} \Psi & 0 \\ 0 & \tilde{\Psi} \end{pmatrix}, \tag{9}$$

With this notation established we may write Eq. (5) as an equation for H^* as

$$H^*(\mu) = E + \frac{c}{2}\int_0^1 H(\nu)\frac{\mu}{\mu+\nu} D(\nu)d\nu\, H^*(\mu) \tag{10}$$
$$= E + (c/2)(LH)H^*\,.$$

Here E is the $2n\times 2n$ identity matrix and L denotes the integral operator.

The approximations we consider here are formed by replacing $D(\nu)d\nu$ by $dD_N(\nu)$ and letting H_N be the physical solution, if it exists, to

$$H_N^* = E + \frac{c}{2}(L_N H_N) H_N^*\,, \tag{11}$$

where, for X-valued F,

 0022-2488/82/112097-04$02.50 © 1982 American Institute of Physics

$$(L_N F)(\mu) = \int_0^1 \frac{\mu}{\mu+\nu} F(\nu) dD_N(\nu) . \tag{12}$$

The measure dD_N will be assumed to have the form

$$dD_N = \begin{pmatrix} d\Psi_N & 0 \\ 0 & d\tilde{\Psi}_N \end{pmatrix} . \tag{13}$$

We let $\|\cdot\|_{sp}$ denote spectral radius. Our assumptions on the sequence $d\Psi_N$ are

(H1) $\| \int_0^1 d\Psi_N \|_{sp} \leqslant 1$.

(H2) For every $n \times n$ matrix valued function f with non-negative entries

$$\int_0^1 f(\nu) d\Psi_N(\nu) \geqslant 0 .$$

(H3) For every continuous $n \times n$ matrix valued function f

$$\lim_{N\to\infty} \int_0^1 f(\nu) d\Psi_N(\nu) = \int_0^1 f(\nu)\Psi(\nu) d\nu .$$

In Sec. II we prove

Theorem (1.1): Assume the sequence $\{dD_N\}$ satisfies (H1)–(H3). Then H_N exists for $|c| \leqslant 1$ and

$$\lim_{N\to\infty} \max_{0<\mu<1} \| H(\mu) - H_N(\mu) \|_X = 0 \tag{14}$$

uniformly for c in compact subsets of $\{c \mid |c| < 1\}$.

A second approximation method we consider is Newton's method. In Sec. III we extend some results of Ref. 11 to this matrix-valued situation. Problems are encountered for $c = 1$ and we show how results of Reddien,[12] Griewank and Osborne,[13] and Decker and the author[14,15] may be applied to that situation. These results therefore apply both to multigroup neutron transport and to Rayleigh scattering of polarized light.[1] We also consider the chord method in this setting. The analysis in Sec. III applies as well to approximate equations like Eq. (11) if $\| \int_0^1 d\Psi_N \|_{sp} = 1$.

II. PROOF OF THEOREM (1.1)

Let K and K_N be $n \times n$ matrix valued functions given, for $x > 0$, by

$$K(x) = \int_0^1 e^{-|x|/\nu} \Psi(\nu) \frac{d\nu}{\nu} , \tag{15}$$

$$K_N(x) = \int_0^1 e^{-|x|/\nu} \frac{1}{\nu} d\Psi_N . \tag{16}$$

For matrices A, B we say $A \geqslant B$ if $a_{ij} \geqslant b_{ij}$ for all i and j. We let $|A|$ be the matrix with entries $|a_{ij}|$. Assumption (H2) implies that $K_N \geqslant 0$; $K \geqslant 0$ by our assumption that $\Psi \geqslant 0$.

We let $\mathscr{L}$ denote the Banach space of X-valued integrable functions on $(0, \infty)$ with norm

$$\|F\| = \int_0^\infty \|F(x)\|_X dx . \tag{17}$$

Let $T, T_N \in \mathscr{L}$ be given

$$T = \begin{pmatrix} K & 0 \\ 0 & \tilde{K} \end{pmatrix} , \tag{18}$$

$$T_N = \begin{pmatrix} K_N & 0 \\ 0 & \tilde{K}_N \end{pmatrix} . \tag{19}$$

Consider the integral equations in $\mathscr{L}$,

$$\Gamma(x) - \frac{c}{2} \int_0^\infty T(x-y)\Gamma(y) dy = \frac{c}{2} T(x) , \tag{20}$$

$$\Gamma_N(x) - \frac{c}{2} \int_0^\infty T_N(x-y)\Gamma_N(y) dy = \frac{c}{2} T_N(x) . \tag{21}$$

Our assumptions on Ψ and $d\Psi_N$ imply, as in Refs. 2 and 3, the following theorem:

Theorem (2.1): For $|c| < 1$, Eqs. (20) and (21) have unique solutions in $\mathscr{L}$. Moreover, for $|c| < 1$,

$$H^*(\mu) = E + \int_0^\infty e^{(-x/\mu)} \Gamma(x) dx , \tag{22}$$

$$H_N^*(\mu) = E + \int_0^\infty e^{(-x/\mu)} \Gamma_N(x) dx . \tag{23}$$

Theorem (2.1) implies that we need only show that Γ_N converges to in Γ to prove Theorem (1.1). This will be an immediate consequence of the following lemma.

Lemma (2.2): $\lim_{N\to\infty} \|T_N - T\|_{\mathscr{L}} = 0$.

Proof: By assumption (H3) $T_N(x)$ converges to $T(x)$ for each fixed $x \neq 0$. Also the definition of $\| \|_X$ implies that if $F \in \mathscr{L}$,

$$F = \begin{pmatrix} F_l & 0 \\ 0 & \tilde{F}_r \end{pmatrix} ,$$

then

$$\begin{aligned} \|F\|_{\mathscr{L}} &= \int_0^\infty \sum_{i,j=1}^n |F_{ij}^{(l)}(x)| + |F_{ij}^{(r)}(x)| dx \\ &= \sum_{i,j=1}^n \left(\int_0^\infty |F_{ij}^{(l)}(x)| dx + \int_0^\infty |F_{ij}^{(r)}(x)| dx \right) \\ &= \left\| \int_0^\infty |F(x)| dx \right\|_X . \end{aligned}$$

Hence, if we can show that

$$\lim_{N\to\infty} \int_0^\infty |T_N(x)| dx = \int_0^\infty |T(x)| dx \tag{24}$$

we will be done.[16] But, as $T, T_N \geqslant 0$ we have

$$\int_0^\infty |T_N(x)| dx = \int_0^1 dD_N(\nu) .$$

This implies Eq. (24) by (H3). The proof of the lemma is complete.

To complete the proof of the theorem we let $(c/2)\mathscr{T}$ and $(c/2)\mathscr{T}_N$ denote the integral operators in Eqs. (20) and (21), respectively. We have,[2] for $F \in L^1(X)$,

$$\left\| \left(\frac{c}{2}\mathscr{T} - \frac{c}{2}\mathscr{T}_N \right) F \right\|_{\mathscr{L}} \leqslant |c| \, \|T - T_N\|_{\mathscr{L}} \, \|F\|_{\mathscr{L}} . \tag{25}$$

Hence $\mathscr{T}_N$ converges to $\mathscr{T}$ in the operator norm on $\mathscr{L}$. As[2] $[I - (c/2)\mathscr{T}_N]^{-1}$ exists for all N, $[I - (c/2)\mathscr{T}_N]^{-1}$ must converge to $[I - (c/2)\mathscr{T}]^{-1}$ in norm. Let $|||\cdot|||$ denote the operator norm on $\mathscr{L}$.

Now choose $\epsilon > 0$, let N be sufficiently large so that for $k > N$

$$\left| \left| \left| \left(I - \frac{c}{2}\mathscr{T}_k \right)^{-1} - \left(I - \frac{c}{2}\mathscr{T} \right)^{-1} \right| \right| \right| < \frac{\epsilon}{2\|T\|_{\mathscr{L}}} , \tag{26}$$

$$\|T_k - T\|_{\mathscr{L}} < \frac{\epsilon}{2} \times \left[\left\| \left\| \left(I - \frac{c}{2}\mathscr{T} \right)^{-1} \right\| \right\| + \frac{\epsilon}{2\|T\|_{\mathscr{L}}} \right]. \tag{27}$$

We have, for $k > N$,

$$\begin{aligned} \|\Gamma - \Gamma_k\|_{\mathscr{L}} &\leqslant \| \left[\left(I - \frac{c}{2}\mathscr{T} \right)^{-1} - \left(I - \frac{c}{2}\mathscr{T}_k \right)^{-1} \right] T \|_{\mathscr{L}} \\ &+ \left\| \left(I - \frac{c}{2}\mathscr{T}_k \right)^{-1} (T_k - T) \right\|_{\mathscr{L}} \\ &< \frac{\epsilon}{2} + \left[\left\| \left\| \left(I - \frac{c}{2}\mathscr{T} \right)^{-1} \right\| \right\| \right. \\ &\left. + \frac{\epsilon}{2\|T\|} \right] \|T_k - T\|_L < \epsilon . \end{aligned}$$

This completes the proof of Theorem (1.1) as $\epsilon > 0$ was arbitrary.

III. NEWTON'S METHOD

The theorems and proofs in this chapter are given in the context of Eq. (15); however, they apply as well to the matrix-valued H-equation that arises in scattering of polarized light in the case $c = 1$.

We rewrite Eq. (10) as

$$\mathscr{F}_c(H(c)) = H(c) - E - (c/2)\, H(c) Q H(c) = 0 , \tag{28}$$

where, if $\mathscr{C}$ denotes the space of continuous X-valued functions on (0,1) with the uniform norm, the operator Q is given, for $F \in \mathscr{C}$, by

$$(\mathrm{QF})(\mu) = \int_0^1 [\mu/(\mu + \nu)]\, D(\nu) F^*(\nu) d\nu . \tag{29}$$

$\mathscr{F}_c$ is a continuous nonlinear map on the space $\mathscr{C}$. The Frechet derivative of $\mathscr{F}_c$ evaluated at $G \in \mathscr{C}$ is a linear map on $\mathscr{C}$ given, for $u \in \mathscr{C}$ by

$$\mathscr{F}_c'(G)u = u - (c/2)\, GQu - (c/2)\, uQG . \tag{30}$$

The classical Newton–Kantorovich theorem[17], stated in our context, is the following

Theorem (3.1): If $\mathscr{F}_c'(H(c))$ has a bounded inverse on $\mathscr{C}$ and if H_0 is sufficiently close to $H(c)$ the Newton iterates

$$H_n = H_{n-1} - \mathscr{F}_c'(H_{n-1})^{-1} \mathscr{F}_c(H_{n-1}) \tag{31}$$

converge to $H(c)$ and there is $K > 0$ such that

$$\|H(c) - H_n\| < K \|H(c) - H_{n-1}\|^2 . \tag{32}$$

We require the following lemma due to Mullikin and Victory[2]. The notation is from Ref. 18.

Lemma (3.2): $\mathscr{F}_c'(H(1))$ has a one-dimensional null space N and closed range X with $\mathscr{C} = N \oplus X$. N is spanned by $\Phi \in \mathscr{C}$ where $\Phi(\mu) = \mu P H(\mu)$ and

$$P = P^* = \begin{pmatrix} P & 0 \\ 0 & \tilde{P} \end{pmatrix}, \tag{33}$$

P is an $n \times n$ matrix with positive entries and $P^2 = P$. Finally, $\mathscr{F}_c'(H(1))$ has an inverse for $|c| \leqslant 1$, $c \neq 1$.

The following theorem implies that for $|c| < 1$, the Newton iterates converge to $H(c)$ if H_0 is sufficiently close to $H(c)$.

Theorem (3.3): For $|c| < 1$, $\mathscr{F}_c'(H(c))^{-1}$ exists on $\mathscr{C}$.

Proof: Define a bounded linear operator $\mathscr{A}(c)$ by

$$\mathscr{A}(c)F = FQH(c) + H(c)QF . \tag{34}$$

It is known[10] that if $H(c)$ is expanded in a power series in c about $c = 0$, then the coefficients are matrices with non-negative entries. Hence, as $D(\nu) \geqslant 0$,

$$\begin{aligned} |\mathscr{A}(c)F| &\leqslant |F|Q|H(c)| + |H(c)|Q|F| \\ &\leqslant |F|QH(|c|) + H(|c|)Q|F| \\ &\leqslant |F|QH(1) + H(1)Q|F| \\ &= \mathscr{A}(1)|F| . \end{aligned} \tag{35}$$

For $|c| < 1$ we have, by Lemma (3.2), that the series

$$\sum_{m=0}^{\infty} \left(\frac{c}{2} \right)^m (\mathscr{A}(1))^m = \mathscr{F}_c'(H(1))^{-1} \tag{36}$$

converges in the operator norm. By Eq. (35), the series

$$\sum_{m=0}^{\infty} \left(\frac{c}{2} \right)^m (\mathscr{A}(c))^m$$

converges as well and this is $\mathscr{F}_c'(H(c))^{-1}$. This completes the proof.

The case $c = 1$ is very different. Here Lemma (3.2) implies that the hypothesis of Theorem (3.1) do not hold. We will show that if the initial guess H_0 is carefully chosen, that the Newton iterates still converge to (H2) but at a slower rate than that given by Eq. (32).

We define an inner product $[\cdot,\cdot]$ on $\mathscr{C}$ for $F,G \in \mathscr{C}$, as follows:

$$F(\mu) = \begin{pmatrix} F_l(\mu) & 0 \\ 0 & \tilde{F}_r(\mu) \end{pmatrix},$$

$$G(\mu) = \begin{pmatrix} G_l(\mu) & 0 \\ 0 & \tilde{G}_r(\mu) \end{pmatrix},$$

let

$$[F,G] = \sum_{i,j=1}^{n} \int_0^1 (F_{ij}^{(r)}(\mu)\, G_{ij}^{(r)}(\mu) + F_{ij}^{(l)}(\mu) G_{ij}^{(l)}(\mu)) d\mu . \tag{37}$$

We now define a projection $\mathscr{P}_N$ from $\mathscr{C}$ to N by

$$\mathscr{P}_N F = \frac{[F,\Phi^*]}{[\Phi,\Phi^*]} \Phi , \tag{38}$$

where $\Phi^* \geqslant 0$ is orthogonal to the range of $\mathscr{F}(H)$ in the sense that $[\mathscr{F}'(H)F,\Phi^*] = 0$.[2] We wish to apply the following result due to Decker and the author.[14]

Theorem (3.4). Let $c = 1$. If

$$\mathscr{P}_N(\Phi Q \Phi) \neq 0 , \tag{39}$$

and H_0 is chosen sufficiently close to $H(1)$ subject to the additional condition

$$\|H_0 - \mathscr{P}_N H_0\| \leqslant \theta \|\mathscr{P}_N H_0\| , \tag{40}$$

with θ sufficiently small, then the Newton iterates converge to $H(1)$ and

$$\lim_{n \to \infty} \frac{\|H_n - H(1)\|}{\|H_{n-1} - H(1)\|} = \frac{1}{2} . \tag{41}$$

We now observe that Eq. (39) holds. To see this we simply note that $\Phi \geqslant 0$, $\Phi \neq 0$ $\Phi^* \geqslant 0$, $\Phi^* \neq 0$, if $\mu \neq 0$. As $D(\nu) \geqslant 0$ and not identically zero, $[\Phi Q \Phi, \Phi^*] > 0$. This implies Eq.

(39).

The chord method was considered for scalar H-equations in Ref. 11. The chord iterates are, for $n \geqslant 1$,

$$K_n = K_{n-1} - \mathscr{F}_c'^{-1}(K_0)\,\mathscr{F}_c(K_{n-1})\,. \tag{42}$$

The following theorem is a direct consequence of Theorem (3.3), general results on the chord method in Ref. 17, Eq. (39), and new work of Decker and the author.[15]

Theorem (3.5). For $|c| < 1$ the chord iterates K_n converge to $H(c)$ if K_0 is sufficiently close to $H(c)$. Moreover there is $0 < \sigma(c) < 1$ so that

$$\lim_{n\to\infty} \frac{\|K_n - H(c)\|_\mathscr{E}}{\|K_{n-1} - H(c)\|_\mathscr{E}} = \sigma(c)\,. \tag{43}$$

If $c = 1$, K_0 is sufficiently close to $H(c)$ and if

$$\|K_0 - \mathscr{P}_N K_0\| \leqslant \theta\, \|\mathscr{P}_N K_0\| \tag{44}$$

for some θ sufficiently small then the chord iterates converge to $H(c)$ and there is $\sigma > 0$ so that

$$\|K_n - H(c)\| \leqslant \sigma/n\,. \tag{45}$$

The reader should note that at $c = 1$ convergence of both Newton's method and the chord method is much slower than for $|c| < 1$. For the chord method convergence is so slow that the method should not be used.

ACKNOWLEDGMENTS

The author would like to thank Professor T. W. Mullikin of Purdue University for a helpful conversation on Theorem (3.3). This work was supported by NSF grant MCS–7902659A01.

[1]S. Chandrasekhar, *Radiative Transfer* (Dover, New York, 1960).
[2]T. W. Mullikin and H. D. Victory, "*N*-group neutron transport: criticality problem in slab geometry," J. Math. Anal. Appl. **58** 605–630 (1977).
[3]C. T. Kelley, "Convolution and *H*-equations for operator-valued functions with applications to neutron transport theory," J. Math. Phys. **81** 764–769 (1977).
[4]I. W. Busbridge, "On inhomogeneous stellar atmospheres," Astrophys. J. **133**, 198–209 (1961).
[5]T. W. Mullikin, "Mathematics of radiative transfer," in *Theoretical Principles in Astrophysics and Relativity*, edited by N. R. Lebovitz, W. M. Ried, and P. O. Vandervoort (University of Chicago Press, Chicago, 1978), Chap. 6.
[6]C. T. Kelley, "Multi-group neutron transport in inhomogeneous slabs," J. Int. Eqns. **3**, 261–275 (1981).
[7]C. T. Kelley, "Energy dependent radiative transfer in inhomogeneous slabs" (to appear in J. Integral Eqns.).
[8]C. T. Kelley, "Approximate methods for exit distribution problems in inhomogeneous slabs." *Proceedings of 7th International Conference on Transport Theory* (Lubbock, TX, 1981), Prog. in Nucl. Energy, **8**, 227–234 (1981).
[9]D. Masson, "Moment problem approximation to the Chandrasekhar *H*-equation," J. Math. Phys. **22**, 462–464 (1981).
[10]C. T. Kelley and T. W. Mullikin, "Solution by iteration of *H*-equations in multigroup neutron transport," J. Math. Phys. **19**, 500–501 (1978).
[11]C. T. Kelley, "Solution of the Chandrasekhar *H*-equation by Newton's method," J. Math. Phys. **21**, 1625–1628 (1980).
[12]G. W. Reddien, "Newton's Method and high order singularities," Comput. Math. Appl. **5**, 79–86 (1980).
[13]A. Griewank and M. R. Osborne, "Newton's method for singular problems when the dimension of the null space is > 1," SIAM J. Numer. Anal. **18**, 179–189 (1981).
[14]D. W. Decker and C. T. Kelley, "Newton's method at singular points I.," SIAM J. Numer. Anal. **17**, 66–70 (1980).
[15]D. W. Decker and C. T. Kelley, "Sublinear convergence of the chord method at singular points," (submitted for publication).
[16]W. R. Rudin, *Real and Complex Analysis* (McGraw-Hill, New York, 1966).
[17]L. V. Kantorovich and G. P. Akilov, *Functional Analysis in Normed Spaces* (Pergamon, New York, 1964).
[18]C. T. Kelley, "Analytic continuation of an operator-valued *H*-function with applications to neutron transport theory," J. Math. Phys. **19**, 494–499 (1978).

Lagrangians for spherically symmetric potentials

Marc Henneaux[a)]
Center for Theoretical Physics and Center for Studies in Statistical Mechanics, University of Texas, Austin, Texas 78712

L. C. Shepley
Center for Relativity, University of Texas, Austin, Texas 78712

(Received 25 November 1981; accepted for publication 12 March 1982)

Two Lagrangians are s-equivalent (s for "solution") if they yield equations of motion having the same set of solutions. We consider Lagrangians s-equivalent to $T - V$, where T is flat space kinetic energy and V is a spherically symmetric potential. We show that for n = dimension of space $\geqslant 3$, there are many s-equivalent Lagrangians which cannot be formed from $T - V$ by multiplication by a constant or addition of a total time derivative. In general these s-equivalent Lagrangians lead to inequivalent quantum theories in the sense that the energy spectra are different.

PACS numbers: 03.20. + i, 02.30.Wd, 46.10. + z

I. INTRODUCTION

In the last couple of years, there has been a revived interest in the inverse problem of the calculus of variations for Newtonian mechanics.[1] One of the main motivations for taking up that study is the well-known fact that the classical equations of motion of some dynamical models do not uniquely determine their Lagrangian. This results in an ambiguity for the Hamiltonian, with a corresponding ambiguity in both the quantum and the statistical theories of these systems (uniqueness aspect of the inverse problem). Another motivation is provided by the usefulness of Lagrangians in the search for exact or approximate solutions to the equations of motion, as well as in the study of their stability (existence aspect of the inverse problem).

Recently, one of us has shown how the inverse problem can be handled—and in fact, solved completely for each particular set of forces—by studying the restrictions on the coordinate-velocity Poisson brackets implied by the existence of a second-order variational principle.[2] (i) This method yields necessary and sufficient conditions for the existence and the uniqueness of a Lagrangian. These conditions can be dealt with and have a geometrical content. (ii) It answers the question: "Do the equations of motion determine the commutation relations?"[3] (iii) It gives a procedure to construct explicitly the Lagrangian(s) (if any). The purpose of the present paper is to illustrate that latter point in the important case of spherically symmetric potentials, with a special emphasis on the $1/r$ potential problem.

We prove that there always exist many s-equivalent Lagrangians for spherically symmetric potentials in more than two dimensions (the two-dimensional case has been treated in Ref. 2).

For example, consider (in n dimensions) the Lagrangian

$$L = T - V + \gamma J/r^2. \tag{1.1}$$

The kinetic energy is (we take the mass $m = 1$ throughout)

$$T = \frac{1}{2}\dot{q}^i\dot{q}^i \quad \left(\cdot = \frac{d}{dt}\right),$$

and the potential energy is a function only of radius r and time t:

$$V = V(r,t) \quad \text{where} \quad r = (q^iq^i)^{1/2}.$$

J is the magnitude of angular momentum:

$$\begin{aligned} J &= [r^2\dot{q}^i\dot{q}^i - (\dot{q}^iq^i)^2]^{1/2} \\ &= [\tfrac{1}{2}(q^s\dot{q}^t - q^t\dot{q}^s)(q^s\dot{q}^t - q^t\dot{q}^s)]^{1/2}. \end{aligned} \tag{1.2}$$

(Summation over repeated indices is understood.) The Lagrangian L yields equations equivalent to

$$\ddot{q}^i = -\frac{\partial V}{\partial q^i}. \tag{1.3}$$

However, only for dimension $n = 1$ (when $J = 0$) or $n = 2$ is the term $\gamma J/r^2$ a total derivative for the constant $\gamma \neq 0$.

We also show that in three dimensions ($n = 3$) there are as many equivalence classes of Lagrangians that lead to a variational principle equivalent to (1.3) as arbitrary functions of two variables. All the Lagrangians can be expressed in terms of one curvilinear integral that involves the Lagrange parentheses of the coordinates and the velocities. These parentheses are explicitly determined for any choice of the above mentioned arbitrary function.

We also discuss some of the properties of the Lagrangians associated with (1.3). In particular, it is pointed out that the introduction of a small anisotropy in the potential singles out the usual one ($L_0 = T - V$): When such an interaction is switched on, the equations derive from only one variational principle, which reduces in a continuous manner to the standard one in the limit of no anisotropic interaction. In our opinion, however, the fact that the standard Lagrangian plays a privileged role does not mean that the other Lagrangians might not be useful in the study of some classical questions.

We finally turn to the Kepler (or hydrogen atom) problem, for which we derive the quantum theory implied by the specific Lagrangian (1.1). It is found that the spectrum of the energy levels is different from the usual one. This result clearly shows that, although L and L_0 are equivalent from the classical viewpoint, they yield inequivalent quantum effects. Why the correct quantum theory should be based on L_0 rather than L may be related to the fact that L_0 plays a

[a)] Chercheur des Instituts Internationaux de Physique et de Chimie, founded by E. Solvay.

 0022-2488/82/112101-07$02.50 © 1982 American Institute of Physics

privileged role, but the relationship is not at all clear.

All this is done after a brief survey of the results of Ref. 2 (in Sec. II).

II. INVERSE PROBLEM OF THE CALCULUS OF VARIATIONS: BASIC EQUATIONS

Consider a system of second order differential equations in normal form,

$$\ddot{q}^i = f^i(q,\dot{q},t). \tag{2.1}$$

Upon introducing n new variables, it can be cast in the first order form

$$\dot{q}^i = u^i, \quad \dot{u}^i = f^i(q,u,t), \tag{2.2}$$

or, in more compact notation,

$$\dot{z}^\lambda = f^\lambda(z,t) \tag{2.3}$$

with $z^\lambda = (q^i,u^i)$, $f^\lambda = (u^i,f^i)$ $(\lambda = 1,\ldots,2n)$.

The following theorem was proved in Ref. 2:

Theorem: A necessary and sufficient condition for the existence of a second order Lagrangian $L(q,\dot{q},t)$ yielding Euler–Lagrange equations equivalent to (2.1) is that the linear, algebraic-differential equations for the 2-form $\sigma_{\lambda\mu}(z,t)$,

$$\text{(i) } \mathbf{d}\boldsymbol{\sigma} = 0, \tag{2.4}$$

$$\text{(ii) } \pounds_f\boldsymbol{\sigma} + \partial_t\boldsymbol{\sigma} = 0, \tag{2.5}$$

$$\text{(iii) } \boldsymbol{\sigma}[\mathbf{X}_i \wedge \mathbf{X}_j] = 0, \tag{2.6}$$

$$\text{(iv) } \det\boldsymbol{\sigma} \neq 0 \tag{2.7}$$

possess a solution. Here, the symbols $\mathbf{d}$, $\pounds_f$, and ∂_t are, respectively, exterior differentiation, Lie differentiation along the vector field f^λ tangent to the trajectories, and partial differentiation with respect to t (at constant z^λ). Also, the vector fields $\mathbf{X}_i$ are tangent to the u^i-coordinate lines:

$$\mathbf{X}_i = \frac{\partial}{\partial u^i}$$

("vertical vectors"). For later convenience we deliberately adopt here notations that do not fully reflect the well-known invariance of the formalism under time-dependent transformations $t' = t$, $z'^\lambda = z'^\lambda(z,t)$. We also note that the second-order problem is expressed in terms of given configuration variables q^i and not in the $2n$-dimensional z^μ-space. This explains why some of Eqs. (2.4)–(2.7) [Eq. (2.6) more precisely] are only invariant under point transformations, i.e., transformations of the form $q'^i = q'^i(q,t)$, $u'^i = u^j\partial q'^i/\partial q^j$ (no velocities in the transformation law for the coordinates). This is a major difference from the so-called "first-order" problem, for which the condition (2.6) is absent (see, for example, Refs. 2 and 4).

Equations (2.4)–(2.7) for $\boldsymbol{\sigma}$ are easy to handle. Taken separately, they all possess an infinity of solutions. However, in more than one dimension, the differential equation (2.5) is in general incompatible with the algebraic conditions (2.6)–(2.7), and there is no Lagrangian for (2.1). This is because (2.5) and (2.6) imply the algebraic conditions

$$\boldsymbol{\sigma}[(\pounds_f + \partial_t)^m(\mathbf{X}_i \wedge \mathbf{X}_j)] = 0 \quad \text{for} \quad m = 0,1,2,\cdots \tag{2.8}$$

on $\boldsymbol{\sigma}$. These constitute a system of linear, homogeneous equations for the $n(2n-1)$ components $\sigma_{\lambda\mu}$, the rank of which depends on the forces (through $\pounds_f$). In general, i.e., for sufficiently arbitrary forces, this rank is equal to $n(2n-1)$. Consequently, the only solution to (2.8) is $\boldsymbol{\sigma} = 0$, which clearly violates (2.7).

It is convenient to recast Eqs. (2.4)–(2.7) for $\boldsymbol{\sigma}$ into initial value equations at an arbitrarily chosen time, say $t = 0$.

$$\mathring{\boldsymbol{\sigma}}[(\pounds_f + \partial_t)^m(\mathbf{X}_i \wedge \mathbf{X}_j)]\,|_{t=0} = 0 \quad \text{for} \quad m = 0,1,2,\cdots, \tag{2.9}$$

$$\mathbf{d}\mathring{\boldsymbol{\sigma}} = 0, \tag{2.10}$$

$$\det\mathring{\boldsymbol{\sigma}} \neq 0, \tag{2.11}$$

and evolution equations

$$\pounds_f\boldsymbol{\sigma} + \partial_t\boldsymbol{\sigma} = 0, \quad \text{with} \quad \boldsymbol{\sigma}(t=0) = \mathring{\boldsymbol{\sigma}}. \tag{2.12}$$

These latter equations propagate the initial data without ambiguity in a manner that maintains the constraints (2.9)–(2.11) (at least locally), so that the number of independent solutions to (2.4)–(2.7) is equal to the number of independent solutions to the initial value problem.

The above theorem also sheds light on the uniqueness aspect of the inverse problem. We call two Lagrangians $L_{(1)}$ and $L_{(2)}$ trivially equivalent if

$$L_{(1)} = \alpha L_{(2)} + \frac{dg}{dt} \quad (\alpha = \text{const} \neq 0).$$

Each Lagrangian which is trivially equivalent to a given $L_{(1)}$ will yield exactly the same equations of motion (except for the overall factor α). In contrast, two Lagrangians are s-equivalent (s for "solution") if they yield equations of motion which have the same set of solutions. Thus L of (1.1) is s-equivalent to $L_0 = T - V$ but not trivially equivalent to it. We say a given $L_{(1)}$ is essentially unique if the only Lagrangians s-equivalent to $L_{(1)}$ are trivially equivalent to it.

The following result was proved in Ref. 2:

Theorem: There is a one-to-one correspondence between trivial equivalence classes of Lagrangians for (2.1) and equivalence classes of solutions $\sigma_{\lambda\mu}(z,t)$ of (2.4)–(2.7), where two solutions of (2.4)–(2.7) are "equivalent" if and only if $\sigma_{(1)\lambda\mu} = \alpha\sigma_{(2)\lambda\mu}$ $(\alpha = \text{const} \neq 0)$.

This theorem reduces the discussion of the uniqueness aspect of the inverse problem to the simpler study of the uniqueness of the solutions to the system (2.4)–(2.7)—or, equivalently, (2.9)–(2.11).

We note finally that the connection between the 2-forms $\boldsymbol{\sigma}$ and the Lagrangians is provided by

$$\mathbf{d}L = \pounds_f\mathbf{a} + \partial_t\mathbf{a}, \tag{2.13}$$

where the 1-forms $\mathbf{a}$ obey

$$\mathbf{da} = \boldsymbol{\sigma} \quad \text{and} \quad \mathbf{a}(\mathbf{X}_i) = 0. \tag{2.14}$$

Once the 2-forms $\boldsymbol{\sigma}$ are known, the Lagrangians can be obtained by elementary integrations along paths in the z^μ-space (see example below—we do not treat here global problems that might arise from the nontrivial topology of the z^μ-space, which is the tangent bundle to configuration space). Formulas (2.13)–(2.14) identify $\sigma_{\lambda\mu}$ as the coordinate-velocity Lagrange parentheses:

$$\boldsymbol{\sigma} = \begin{pmatrix} \mathbf{A} & -\mathbf{B} \\ \mathbf{B} & 0 \end{pmatrix}, \tag{2.15}$$

where the n by n matrices $\mathbf{A}$ and $\mathbf{B}$ are equal to

$$B_{ij} = \frac{\partial^2 L}{\partial \dot{q}^j \partial \dot{q}^i}, \quad A_{ij} = \frac{\partial^2 L}{\partial q^i \partial \dot{q}^j} - \frac{\partial^2 L}{\partial q^j \partial \dot{q}^i}. \tag{2.16}$$

Note that $\mathbf{A} = -\mathbf{A}^T$, $\mathbf{B} = \mathbf{B}^T$.

Results similar to those mentioned above have been derived in the recent preprint[5] (which only considers time-independent Lagrangians) and Ref. 6.

III. VELOCITY-FREE FORCES DERIVABLE FROM A POTENTIAL

Let us now specialize the previous general discussion to the case when the forces (2.1) are velocity-free and derive from a potential

$$f^i = -\frac{\partial V}{\partial q^i}, \quad V = V(q,t). \tag{3.1}$$

It is well known that, upon variation, the Lagrangian $T - V$ reproduces the equations of motion (2.1). Accordingly, the only nontrivial part of the inverse problem of the calculus of variations consists in its uniqueness aspect. We reformulate it below as a standard problem of linear algebra for $n \times n$ matrices, coupled to the differential problem (2.10).

As shown in Ref. 2, Eqs. (2.8) with $m = 1$ and $m = 2$ are, respectively, equivalent to $\mathbf{B} = \mathbf{B}^T$ and $\mathbf{A} = 0$. As a result, Eq. (2.5) reduces to

$$\partial_t \mathbf{B} + u^m \frac{\partial \mathbf{B}}{\partial q^m} + f^m \frac{\partial \mathbf{B}}{\partial u^m} = 0 \tag{3.2}$$

and states that the elements of the matrix $\mathbf{B}$ are all constants of the motion (compare with Ref. 7; note that the "Hessian matrix" $\mathbf{B}$ is equal to the unit matrix for the Lagrangian $T - V$).

On the other hand, Eqs. (2.8) with $m = 3$ read

$$[\mathbf{B}, \mathbf{D}^0] = 0, \tag{3.3}$$

where the symmetric matrix $\mathbf{D}^0$ is given by[2]

$$D^0{}_{ij} = \frac{\partial f^i}{\partial q^j} \tag{3.4}$$

(indices are lowered and raised with the Euclidean metric δ_{ij}). All time derivatives of (3.3) vanish, and, using (3.2), we get Eqs. (2.8) with $m > 3$. These equations are explicitly

$$[\mathbf{B}, \mathbf{D}^l] = 0, \quad l = 1,2,\cdots, \tag{3.5}$$

where the matrices D^l are recursively defined by

$$\mathbf{D}^{l+1} = \frac{\partial \mathbf{D}^l}{\partial t} + u^m \frac{\partial \mathbf{D}^l}{\partial q^m} + f^m \frac{\partial \mathbf{D}^l}{\partial u^m}. \tag{3.6}$$

Again, we equivalently recast the equations for $\mathbf{B}$ into initial value equations

$$[\mathring{\mathbf{B}}, \mathring{\mathbf{D}}^l] = 0, \quad l = 0,1,2,\cdots, \tag{3.7}$$

$$\det \mathring{\mathbf{B}} \neq 0, \tag{3.8}$$

$$\frac{\partial \mathring{B}_{ij}}{\partial q^k} - \frac{\partial \mathring{B}_{ik}}{\partial q^j} = 0, \tag{3.9}$$

$$\frac{\partial \mathring{B}_{ij}}{\partial \dot{q}^k} - \frac{\partial \mathring{B}_{ik}}{\partial \dot{q}^j} = 0, \tag{3.10}$$

and evolution equations (3.2). Equations (3.9) and (3.10) are just what become the closure conditions $\mathbf{d}\dot{\boldsymbol{\sigma}} = 0$ since $\mathbf{A} = 0$.

The above interesting way (3.7) of rewriting the algebraic conditions (2.9) is due to Sarlet (who has also shown how to extend it to the case of general forces).[6] It leaves us with well-known equations of basic algebra, which simply state that the matrix $\mathring{\mathbf{B}}$ commutes with all the elements of the Lie algebra $\mathscr{A}(\mathring{\mathbf{D}}^l)$ generated by the symmetric matrices $\mathring{\mathbf{D}}^l$ [i.e., the smallest Lie subalgebra of $\mathscr{G}\ell(n,R)$ that contains all $\mathbf{D}^l$.] Indeed, from the Jacobi identity for the commutators of matrices, one easily infers $[\mathring{\mathbf{B}}, [\mathring{\mathbf{D}}^l, \mathring{\mathbf{D}}^m]] = 0$ and then $[\mathring{\mathbf{B}}, \mathbf{D}] = 0$ for all $\mathbf{D} \in \mathscr{A}(\mathring{\mathbf{D}}^l)$. If the representation $\mathscr{A}(\mathring{\mathbf{D}}^l)$ is irreducible, the matrix $\mathring{\mathbf{B}}$ must be a multiple of the identity, $\mathring{\mathbf{B}} = \lambda \mathbf{I}$, where λ is a constant function of q^i, $\dot{q}^i$ by (3.9)–(3.10)[2] (we assume $n > 2$). The Lagrangian $T - V$ is consequently essentially unique (up to the trivial equivalence relation mentioned above).

If, however, the representation $\mathscr{A}(\mathring{\mathbf{D}}^l)$ is reducible, the general solution to the algebraic problem (3.7) is

$$\mathring{\mathbf{B}} = \sum_\alpha \lambda_\alpha \mathbf{M}_\alpha, \tag{3.11}$$

where the matrices $\mathbf{M}_\alpha (\alpha \geqslant 2)$ constitute a complete set of independent particular solutions to (3.7) and where the λ_α are arbitrary functions of q^i and $\dot{q}^i$. The closure relations (3.9) then impose differential restrictions on these functions (see below). Once the initial value problem is solved, the evolution equations (3.2) can be integrated and yield $\mathbf{B}$ at all times.

IV. ALGEBRAIC PART OF THE INITIAL VALUE PROBLEM FOR SPHERICALLY SYMMETRIC POTENTIALS

In this paragraph, Eqs. (3.7) are solved in the case of spherically symmetric potentials,

$$V = V(r,t) \quad \text{with} \quad r = [q^i q^i]^{1/2}.$$

From the equality

$$\mathring{D}_{ik} = \frac{\partial f^k}{\partial q^i} = -\left(\frac{V'}{r}\right)' \frac{q^k q^i}{r} - \frac{V'}{r} \delta_{ik} \tag{4.1}$$

(where $'$ is $\partial/\partial r$), the first equation (3.7) for the matrix $\mathring{\mathbf{B}}$ reads

$$\frac{1}{r}\left(\frac{V'}{r}\right)' [\mathring{\mathbf{B}}, \mathbf{q} \otimes \mathbf{q}] = 0, \tag{4.2}$$

where $(\mathbf{q} \otimes \mathbf{q})_{ij} = q^i q^j$.

When the dynamical system is a free particle or an isotropic harmonic oscillator (with positive or negative spring coefficient), this equation and the other equations of (3.7) impose no restriction on $\mathring{\mathbf{B}}$. The algebraic problem (3.7) is thus trivial and we will assume from now on that $V \neq br^2 + c$. In that case Eq. (4.2) is equivalent to

$$[\mathring{\mathbf{B}}, \mathbf{q} \otimes \mathbf{q}] = 0. \tag{4.3}$$

As one easily checks, the whole content of Eqs. (3.7) is that the matrix $\mathring{\mathbf{B}}$—which depends on both q^i and u^i—must commute with the four-dimensional algebra generated by $\mathbf{q} \otimes \mathbf{q}$ and $\mathbf{u} \otimes \mathbf{u}$.

In order to proceed further, we remove from R^{2n} all the "singular points" (q^i, u^i) where the n-dimensional vectors q^i and u^i are linearly dependent. This is a set of measure zero. If

the initial conditions do not belong to that set, the trajectory will never hit it by the conservation of angular momentum.

The general solution to the above problem is given in the remaining part of R^{2n} (i.e., in R^{2n} - {singular points}), by

$$\mathring{\mathbf{B}}(q,u) = \lambda(q,u)\mathbf{I} + \mathbf{P}(q,u)\mathbf{M}(q,u)\mathbf{P}(q,u), \tag{4.4}$$

where $\mathbf{M}(q,u)$ is an arbitrary symmetric matrix and where $\mathbf{P}(q,u)$ is the projection on the surface (of codimension 2) orthogonal to both q^i and u^i:

$$\mathbf{P}(q,u) = \mathbf{I} - J^{-2}[q^2\mathbf{u}\otimes\mathbf{u} + u^2\mathbf{q}\otimes\mathbf{q} - (u\cdot q)(\mathbf{q}\otimes\mathbf{u} + \mathbf{u}\otimes\mathbf{q})]. \tag{4.5}$$

Here, $u^2 = u^i u^i$, $q^2 = r^2 = q^i q^i$, and $u\cdot q = u^i q^i$. As in the beginning of this paper we set

$$J^2 = u^2 q^2 - (u\cdot q)^2.$$

The projection $\mathbf{P}$ possesses the following properties:

$$\mathbf{P}^2 = \mathbf{P}, \quad \mathbf{P} = \mathbf{P}^T, \tag{4.6}$$

$$\mathbf{P} = \mathbf{I} - J^{-2}\mathbf{S}^2 \tag{4.7}$$

and annihilates the vectors q^i and u^i:

$$P_{ij}q^j = P_{ij}u^j = 0 \quad (\text{rank } \mathbf{P} = n-2). \tag{4.8}$$

In (4.7), the antisymmetric matrix

$$\mathbf{S} = \mathbf{u}\otimes\mathbf{q} - \mathbf{q}\otimes\mathbf{u}$$

is the ordinary angular momentum. In three dimensions,

$$S_{ij} = \epsilon_{ijk}J_k, \quad J_i = \tfrac{1}{2}\epsilon_{ijk}S_{jk},$$

where J_k is the vector usually referred to as the angular momentum; note that

$$S_{ij}S_{ij} = -\text{tr}\mathbf{S}^2 = 2J^2.$$

The matrix $\mathring{\mathbf{B}}$ depends on $1 + \frac{1}{2}(n-2)(n-1)$ arbitrary functions, for only the projected part of $\mathbf{M}$ perpendicular to the plane (q^i,u^i) remains in (4.4). It is one of the remarkable features of $\mathring{\mathbf{B}}$ that it is independent of the specific form of the forces. This is, of course, a consequence of the force-independent structure of Eqs. (4.3) and will have important implications to which we shall return.

When $n = 2$, the general solution (4.4) depends on only one function since $\mathbf{P} = 0$. Consequently, the Lagrangian for the two-dimensional spherically symmetric problem is essentially unique.

When $n = 3$, the general solution to (4.3) involves two arbitrary functions. It can be rewritten as

$$\mathring{\mathbf{B}}(q,u) = \lambda(q,u)\mathbf{I} + \mu(q,u)\mathbf{P}(q,u), \tag{4.9}$$

where $\mu(q,u)$ is an arbitrary function.

For higher dimensions, the general solution to (4.3) depends on more arbitrary functions and is harder to handle. We shall thus restrict our attention to the case $n = 3$, which already presents interesting features.

V. DIFFERENTIAL PART OF THE INITIAL VALUE PROBLEM (THREE DIMENSIONS)

With $\mathring{\mathbf{B}}$ given by (4.9), the conditions (3.9) and (3.10) become first-order partial differential equations for λ and μ. When the indices i, j, k all take different values, these equations only involve the function μ. They read explicitly

$$T_i\mu = -u^i\mu, \tag{5.1}$$

$$R_i\mu = q^i\mu, \tag{5.2}$$

where the vector fields T_i and R_i are given by

$$T_i = S_{ij}\frac{\partial}{\partial q^j}, \quad R_i = S_{ij}\frac{\partial}{\partial u^j}. \tag{5.3}$$

The other equations in (3.9) and (3.10) are

$$\frac{\partial\lambda}{\partial q^i} = 0, \quad \frac{\partial\lambda}{\lambda u^i} = 0 \tag{5.4}$$

[use has been made of (5.1) and (5.2) to simplify their form]. Their general solution is evidently

$$\lambda = c, \tag{5.5}$$

where c is an arbitrary nonvanishing constant (so that $\det\mathring{\mathbf{B}} \neq 0$).

Let us now solve (5.1) and (5.2) for the unknown function $\mu(q,u)$. This system only contains four independent equations since the determinant of the odd-dimensional antisymmetric matrix $\mathbf{S}$ vanishes ($J^i S_{ik} = 0$; note also that $J^i u_i = 0 = J^i q_i$). It is a "complete" system, in the sense that no new equation is obtained if one takes the brackets of (5.1), (5.2) with (5.1), (5.2). This follows from the identities obeyed by the Lie brackets of the vector fields T_i and R_i,

$$[T_i,T_j] = u_jT_i - u_iT_j - 2S_{ij}u^l\frac{\partial}{\partial q^l}, \tag{5.6}$$

$$[T_i,R_j] = u_jR_i + q_iT_j - S_{ij}\left(u^l\frac{\partial}{\partial u^l} - q^l\frac{\partial}{\partial q^l}\right), \tag{5.7}$$

$$[R_i,R_j] = q_iR_j - q_jR_i + 2S_{ij}q^l\frac{\partial}{\partial u^l}, \tag{5.8}$$

and from the equations

$$u^l\frac{\partial\mu}{\partial q^l} = 0 = q^l\frac{\partial\mu}{\partial u^l} = 0,$$

$$u^l\frac{\partial\mu}{\partial u^l} = q^l\frac{\partial\mu}{\partial q^l} = -\mu,$$

which are algebraic consequences of (5.1) and (5.2). To prove that latter point, the following relations were found useful:

$$u_iS_{jk} - u_jS_{ik} = u_kS_{ji}, \tag{5.9}$$

$$q_iS_{jk} - q_jS_{ik} = q_kS_{ji}. \tag{5.10}$$

Let $\mu_1(q,u)$ and $\mu_2(q,u)$ be two solutions of (5.1) and (5.2). Then their quotient $\mu_1/\mu_2 = \nu$ is a solution to the equations

$$T_i\nu = 0, \quad R_i\nu = 0. \tag{5.11}$$

Accordingly, the general solution to (5.1) and (5.2) is given by

$$\mu = \mathring{\mu}\nu, \tag{5.12}$$

where $\mathring{\mu}$ is a particular solution and where ν is the general solution to Eq. (5.11). Since this latter system is complete, it possesses two independent solutions (2 = dimension of the space, 2×3, minus number of independent equations, 4).

It is easy to check that J^{-1} is a particular solution to (5.1) and (5.2) and that

$$\alpha_i = J_i/J \tag{5.13}$$

yields two independent solutions to the homogeneous system (5.11) (only two because $\Sigma\alpha_i^2 = 1$). It then follows that the general solution to (5.1) and (5.2) reads

$$\mu = g(\alpha_i)/J, \tag{5.14}$$

where $g(\alpha_i)$ is an arbitrary function of its arguments. This completes the resolution of the initial value problem.

VI. THE BRACKETS $[q^i,\dot{q}^j]$ AND THE LAGRANGIANS (THREE DIMENSIONS)

Since the matrix $\mathring{\mathbf{B}}$ only involves the components of the angular momentum, which are all conserved quantities, the general solution to (3.2) has no explicit time dependence and is simply given by

$$\mathbf{B} = \mathring{\mathbf{B}} = \mathbf{I} + [g(\alpha_i)/J]\mathbf{P}. \tag{6.1}$$

The nonvanishing constant c in (5.5) has been taken equal to 1. There is no loss of generality in so doing, since the Lagrangians are only determined up to an arbitrary multiplicative constant (plus an arbitrary total time derivative).

The matrix inverse to $\mathbf{B}$ (i.e., the matrix of the Poisson brackets $[q^i,\dot{q}^j]$, reads explicitly

$$\mathbf{B}^{-1} = \mathbf{I} - \frac{g(\alpha_i)}{J + g(\alpha_i)}\mathbf{P} \tag{6.2}$$

since $\mathbf{P}^2 = \mathbf{P}$. It is well defined everywhere in R^{2n}- {singular points} provided the function $g(\alpha_i)$ obeys $J + g(\alpha_i) \neq 0$. Note again that, although the forces do appear in the evolution equations (3.2), the Lagrange parentheses $(q_i,q_j) = 0$, $(q_i,\dot{q}_j) = B_{ij}$, and $(\dot{q}_i,\dot{q}_j) = 0$ are independent of their specific nature. Together with the fact that the projection $\mathbf{P}$ annihilates the forces ($P_{ij} f^j = 0$, the forces are radial), this has the following interesting consequence: Let $L[V]$ be one of the Lagrangians for the spherically symmetric potential $V(r,t)$ and $L_0[V]$ be the standard one $(T - V)$.

Theorem: The difference $\Delta L = L[V] - L_0[V]$ is independent of the spherically symmetric potential $V(r,t)$, i.e., $T - \tilde{V} + \Delta L$ is a good Lagrangian for any other spherically symmetric potential $\tilde{V}(r,t)$.

Proof: Although it is obvious that $\Delta\sigma$ defined by

$$\Delta\sigma = \begin{pmatrix} 0 & -(g/J)\mathbf{P} \\ (g/J)\mathbf{P} & 0 \end{pmatrix}$$

is independent of the form of $V(r)$, we must show that ΔL is. From (2.13) and (2.14) we find

$$\mathbf{d}(\Delta L) = \pounds_f(\Delta\mathbf{a}) \quad \text{with } \mathbf{d}\Delta\mathbf{a} = \Delta\sigma \text{ and } \mathbf{a}(\mathbf{X}_i) = 0, \tag{6.3}$$

where we have taken $\Delta\mathbf{a}$ to be explicitly independent of t. The components of $\pounds_f(\Delta\mathbf{a})$ are [remember $f_\mu = (u^i, -\partial V/\partial q^i)$]

$$[\pounds_f(\Delta\mathbf{a})]_\mu = \left(u^i \frac{\partial}{\partial q^j}(\Delta a_i), a_i\right),$$

where we have used the fact that

$$f^i \frac{\partial}{\partial u^j}(\Delta a_i) = f^j\left[-\frac{g}{J}P_{ij}\right] = 0.$$

Thus $\mathbf{d}(\Delta L) = \pounds_f(\Delta\mathbf{a})$ for any $V(r,t)$ and hence $T - V + \Delta L$ is a good Lagrangian regardless of the form of $V(r,t)$.

This theorem shows that, in order to get all the Lagrangians for the spherically symmetric potential (in three dimensions), one simply has to solve the equations

$$\mathbf{d}\Delta\mathbf{a} = \Delta\sigma, \quad \Delta\mathbf{a}(\mathbf{X}_i) = 0, \tag{6.4}$$

and

$$\mathbf{d}(\Delta L) = \pounds_f(\Delta\mathbf{a}) \tag{6.5}$$

with $f^\lambda = (u^i,0)$. The second equation is easy to solve once $\Delta\mathbf{a}$ is known:

$$\Delta L(z^\mu) = \int_{z_0^\mu}^{z^\mu} (\pounds_f \Delta\mathbf{a})_\nu dz^\nu.$$

Here, the integral is taken along any path joining an arbitrarily chosen fiducial point z_0^μ to the point z^μ. In more explicit terms, we have

$$\Delta L(z) = \int_{z_0}^{z} \left[u^j\left(\frac{\partial}{\partial q^j}\Delta a_i\right)dq^i + \Delta a_i du^i\right]. \tag{6.6}$$

We thus turn to the task of determining $\Delta\mathbf{a}$ from Eqs. (6.4). These equations read explicitly

$$\frac{\partial \Delta a_i}{\partial \dot{q}^j} = \frac{g(\alpha_k)}{J} P_{ij} \tag{6.7}$$

and

$$\frac{\partial \Delta a_i}{\partial q^j} - \frac{\partial \Delta a_j}{\partial q^i} = 0, \tag{6.8}$$

since $\Delta\mathbf{a} = (\Delta a_i,0)$.

The formal solution of (6.7) is

$$\Delta a_i = \int_{u_0}^{u} \frac{g}{J} P_{ij} du^j + C_i(q). \tag{6.9}$$

Because of Eqs. (3.9) and (3.10), (6.8) implies

$$C_i(q) = \frac{\partial \phi}{\partial q^i}$$

for some function $\phi(q)$. Finally, we find that

$$\Delta L = u^j \Delta a_j \tag{6.10}$$

directly satisfies (6.5) and therefore is the solution (6.6) up to the possible addition of a constant.

The conclusion of this whole analysis is that the general Lagrangian for nontrivial spherically symmetric potentials ($V \neq a + br^2$) involves, in three dimensions, an arbitrary function of two variables. It is given by $T - V + \Delta L$, where ΔL is obtained by one integration [(6.9), (6.10)]. It is to be remarked that this Lagrangian is also a good one when the forces are linear or in more than three dimensions, but it is not the most general one in those cases.

VII. INTRODUCTION OF AN ANISOTROPIC INTERACTION

It has been argued in Ref. 2 that systems admitting many inequivalent Lagrangians are rather peculiar. We shall illustrate that general assertion by introducing an anisotropic perturbation of the simplest type,

$$V(r,t) \to V(r,t) + \beta q^1 \tag{7.1}$$

and by showing that when the constant β does not vanish, there is only one trivial equivalence class of Lagrangians (the standard one).

The effect of the perturbation βq^1 is to change the first component of the force by the amount $-\beta$. Accordingly, the first algebraic equations for the matrix $\mathring{\mathbf{B}}$ read again

$$[\mathring{\mathbf{B}},\mathbf{q}\otimes\mathbf{q}]=0. \tag{7.2}$$

The matrix $\mathring{\mathbf{D}}$ is indeed unchanged and we assume $V\neq a+br^2$. (If $V=a+br^2$, one must also add an anharmonic interaction γr^4, for example, to single out the usual equivalence class of Lagrangians.[2]) The next algebraic equations are

$$[\mathring{\mathbf{B}},\mathbf{u}\otimes\mathbf{q}+\mathbf{q}\otimes\mathbf{u}]=0, \tag{7.3}$$

from which one successively infers, using the Jacobi identity, that

$$[\mathring{\mathbf{B}},\mathbf{S}]=0 \tag{7.4}$$

and

$$[\mathbf{B},\mathbf{u}\otimes\mathbf{u}]=0. \tag{7.5}$$

As in the spherically symmetric case, the matrix $\mathring{\mathbf{B}}$ must commute with the algebra generated by $\mathbf{q}\otimes\mathbf{q}$ and $\mathbf{u}\otimes\mathbf{u}$. This implies (4.9):

$$\mathring{\mathbf{B}}(q,u)=\lambda(q,u)\mathbf{I}+\mu(q,u)\mathbf{P}(q,u). \tag{7.6}$$

When $\beta\neq 0$, this is not the end of the algebraic problem, however. Equations (3.7) with $l\geqslant 2$ are equivalent to the additional condition

$$[\mathring{\mathbf{B}},\mathbf{\Lambda}]=0,$$

where the matrix elements of $\mathbf{\Lambda}$ are $\Lambda_{ij}=\beta(\delta_{i1}q^j+\delta_{j1}q^i)$. Since $\mathbf{P}$ and $\mathbf{\Lambda}$ do not commute except on a set of measure zero, the function μ in (7.6) must vanish. The matrix $\mathring{\mathbf{B}}$ reduces to

$$\mathring{\mathbf{B}}(q,u)=\lambda(q,u)\mathbf{I}, \tag{7.7}$$

from which one infers, using (3.9) and (3.10) that λ is constant.[2] It then follows that the Lagrangians for the potentials (7.1) are all given by the formula

$$\alpha(T-V)\quad(\text{with }\alpha=\text{const}\neq 0) \tag{7.8}$$

up to a total time derivative.

Consequently, it is only for the special case of potentials characterized by $\beta=0$ that an ambiguity in the choice of the Lagrangian arises. This indicates, at least in the case at hand, that the existence of many equivalent Lagrangians results more from mathematical simplifications than from physical considerations.

VIII. STUDY OF A PARTICULAR LAGRANGIAN

Although the Lagrangians $T-V+\Delta L\,[g(\alpha_i)]$ are, as we have just stressed, rather peculiar when $g(\alpha_i)\neq 0$, it is nevertheless interesting to study further some of their properties and, in particular, the kinds of quantum theory that they lead to. To that end, we shall treat in detail the case $g(\alpha_i)=\gamma=\text{const}>0$, for which the integration (6.9) is elementary. {The integrand of (6.9) is the derivative with respect to u^j of $(\gamma/J)\,[u^i-(u^sq^s/r^2)q^i]$.} By a suitable choice of the functions $C_i(q)$, one gets

$$\Delta a_i=\frac{\gamma}{J}\left(u^i-\frac{u^sq^s}{r^2}q^i\right). \tag{8.1}$$

This implies that L is

$$L=T-V+\gamma J/r^2, \tag{8.2}$$

in agreement with (1.1).

The transition to the Hamiltonian is made in the usual way. Since the additional term $\gamma J/r^2$ in (8.2) is homogeneous of the first degree in the velocities, the Hamiltonian, as a function of the coordinates and the velocities, is the same as for the standard Lagrangian:

$$H=\tfrac{1}{2}u^2+V(r) \tag{8.3}$$

(but the Poisson brackets $[q^i,u^j]$ are different). In order to express it in terms of the coordinates and the momenta, one needs to invert explicitly the Legendre definition

$$p^i=\frac{\partial L}{\partial u^i}=u^i+\frac{\gamma}{Jr^2}(u^ir^2-q^iu^sq^s). \tag{8.4}$$

The result is

$$u^i=p_i-\frac{\gamma}{\overline{J}r^2}(p_ir^2-q^ip_sq^s), \tag{8.5}$$

where $\overline{J}_k$ is the Hamiltonian angular momentum

$$\overline{J}_k=\epsilon_{kij}(p_iq^j), \tag{8.6}$$

$$\overline{J}=(\overline{J}^2)^{1/2}=[\,p_sp_sq^tq^t-(p_sq^s)^2\,]^{1/2}. \tag{8.7}$$

It is $\overline{J}_k$ that generates infinitesimal rotations by the Poisson bracket operation. It is related to J_k by the relation

$$\overline{J}_k=J_k(1+\gamma/J), \tag{8.8}$$

from which one deduces

$$\overline{J}=J+\gamma. \tag{8.9}$$

With (8.5), the Hamiltonian becomes

$$H=\tfrac{1}{2}p^2+V-\gamma\overline{J}/r^2+\gamma^2/2r^2. \tag{8.10}$$

To finish with the classical theory, we point out that we could have derived the coordinate-velocity form (8.3) of the Hamiltonian directly from the general equation

$$\sigma_{\lambda\mu}f^\mu=\frac{\partial H}{\partial z^\lambda}+\frac{\partial a_\lambda}{\partial t}, \tag{8.11}$$

which relates H, a_λ, and the forces for any system.[2] In our particular case $\partial a_\nu/\partial t$ vanishes, whereas only $\sigma_0=\left(\begin{smallmatrix}0 & -I\\ I & 0\end{smallmatrix}\right)$ contributes in (8.11), since the forces are radial and accordingly annihilated by the projection $\mathbf{P}$. The same conclusion holds even when the arbitrary function $g(\alpha_i)$ on which $\sigma_{\lambda\mu}$ depends is not constant: the Hamiltonian, as a function of the coordinates and the velocities, is given by (8.3) for all Lagrangians that yield the dynamical equations (1.2). We shall call (8.10) the unorthodox Hamiltonian.

IX. QUANTIZATION OF THE UNORTHODOX HAMILTONIAN FOR THE HYDROGEN ATOM

When the potential is proportional to $1/r$, the quantization of (8.10) is straightforward. First of all, there is no ordering problem if one writes H as (8.10), since $\overline{J}$ and r^2 commute. Second of all, the additional terms $-\gamma\overline{J}/r^2$ and $\gamma^2/2r^2$ do not destroy the spherical symmetry so that one can simultaneously diagonalize H, $\overline{J}^2$, and J_z. Of course, these two properties are not peculiar to the $1/r$-potential and hold for any $V(r)$.

As in the usual treatments, we write for the wavefunction ψ,

$$\psi_{lm}(r,\theta,\varphi)=\frac{u(r)}{r}Y_l^m(\theta,\varphi), \tag{9.1}$$

where the functions $Y_l^m(\theta,\varphi)$ are the standard spherical harmonics. Inserting (9.1) in the Schrödinger equation, we obtain the hydrogen-atom radial equation for $u(r)$ with $l(l+1)$ replaced by

$$l(l+1) - 2\gamma[l(l+1)]^{1/2} + \gamma^2 = \{[l(l+1)]^{1/2} - \gamma\}^2$$

in the centrifugal term (this is the effect of the new terms). Note that from (8.9), the numbers $\{[l(l+1)]^{1/2} - \gamma\}^2 = \tilde{l}(\tilde{l}+1)$ are the eigenvalues of the coordinate-velocity angular momentum J^2. Here there is no reason why $\tilde{l}$ should be an integer, or why two different $\tilde{l}$'s should differ by an integer, since γ is an arbitrary real number.

The spectrum of H is given by the Balmer formula with the principal quantum number replaced by $N + \tilde{l} + 1$. Since $\tilde{l}$ is not an integer, the degeneracy peculiar to the Coulomb potential is removed, which clearly indicates that H possesses a spectrum different from the usual one. This was to be anticipated, since the commutation relations $[q^i,u^j]$ are different from the standard case.

In view of the remarks made in Sec. VII, we should state that we feel that the unorthodox H and its quantum theory are not to be considered as physical. Much needs to be done to make completely unambiguous the passage from a classical system to its quantized analog. The importance of the present example is to underscore yet again that in a system having several Lagrangians which are equivalent from the classical point of view, criteria from without the system must be employed to determine the correct quantization procedure.

ACKNOWLEDGMENTS

One of us (M.H.) wishes to thank Professor I. Prigogine, Professor C. Teitelboim, and Professor J. A. Wheeler for much encouragement.

[1]See, for example, the reviews: S. M. Santilli, *Foundations of Classical Mechanics I* (Springer-Verlag, New York, 1978); and S. Hojman and L. C. Shepley, "Lagrangianos Equivalentes," Rev. Mex. Fis. **28**, 149–205 (1982).
[2]M. Henneaux, "Equations of Motion, Commutation Relations and Ambiguities in the Lagrangian Formalism," preprint (Université Libre de Bruxelles, June 1981) [to appear in Ann. Phys. (N. Y.), 1982].
[3]E. P. Wigner, "Do the Equations of Motion Determine the Quantum Mechanical Commutation Relations?," Phys. Rev. **77**, 711–712 (1950); S. Okubo, "Does equation of motion determine commutation relation?," Phys. Rev. D **22**, 919–923 (1980).
[4]S. Hojman and L. Urrutia, "On the inverse problem of the calculus of variations," J. Math. Phys. **22**, 1896–1903 (1981).
[5]R. L. Schafir, "The Variational Principle and Natural Transformations. I. Autonomous Dynamical Systems," preprint (King's College, London, 1981) (to be published in Math. Proc. Camb. Phil. Soc.).
[6]W. Sarlet, "The Helmholtz Conditions Revisited, a New Approach to the Inverse Problem of Lagrangian Dynamics," Preprint RUGsam 81-11 (University of Ghent, September 1981) [to appear in J. Phys. A, 1982].
[7]S. Hojman and H. Harleston, "Equivalent Lagrangians: Multidimensional case," J. Math. Phys. **22**, 1414–1419 (1981). See also D. Currie and E. J. Saletan, "q-Equivalent Particle Hamiltonians. I. The classical one-dimensional case," J. Math. Phys. **7**, 967–974 (1966) and M. Henneaux, "On a Theorem by Hojman and Harleston," Hadronic J. **4**, 2137–2143 (1981).

Structure of three-twistor particles[a)]

B. Lukács, Z. Perjés, and Á. Sebestyén
Central Research Institute for Physics, H-1525 Budapest 114, P. O. B. 49, Hungary

E. T. Newman and J. Porter
University of Pittsburgh, Pittsburgh, Pennsylvania 15260

(Received 10 December 1981; accepted for publication 4 June 1982)

The simplest physical system to have a nontrivial intrinsic structure in Minkowski space-time is a three-twistor particle. We investigate this structure and the two pictures of the particle as an extended object in space-time and as a point in unitary space. We consider the effect of twistor translations on the mass triangle defined by the partial center of mass points in space-time. Finally we consider the connections between twistor rotations and spin and we establish the spin deficiency formula.

PACS numbers: 03.30. + p, 11.30.Ly

1. INTRODUCTION

In Penrose's theory of twistors, zero-mass objects are assigned a fundamental role. The zero-mass particles are represented (classically) by a single twistor while the massive particles are represented by several twistors.[1–6] The basic idea of twistor particle theory is that the kinematic variables, e.g., momentum and angular momentum, associated with the massive particles can be expressed in terms of two or more twistors. (This description is via the so-called kinematic twistor). On the other hand, the internal structure of the particle does depend critically on the number of twistors used in the description. (Frequently in twistor theory[6–8] leptons are described by two-twistor systems while hadrons are described by three-twistor systems.) The linear transformations among the two or three (or more) twistors which preserve the kinematic twistor (or variables) are referred to as internal symmetry transformations,[7] (IST). The two-twistor particles have the simplest space-time description. They can be thought of as either a real center of mass world-line with an associated momentum and spin or as a complex center of mass world-line.

Three-twistor particles are the simplest systems which *do* possess an extended structure in complex Minkowski space. A three-twistor particle with twistors X^α, Y^α,Z^α can be thought of as being (in some sense) composed of the three pairs (X^α,Y^α), (Y^α,Z^α), and (Z^α,X^α) of twistors[5] and thus it would have a substructure of three two-twistor massive subsystems. These structures are, however, not disjoint since any pair of them has a twistor in common. Furthermore, the world-lines, masses, spins, etc., of the parts make up those of the entire system. A point to be emphasized is that the subsystems and their kinematic properties *are not* invariant under the IST.

The present work is an attempt to study the geometry of this twistor decomposition of three-twistor systems. The two main questions investigated are (a) what are the changes in the subsystems caused by the action of the IST, and (b) what is the relationship of the total system variables to those of the related subsystem variables. For pedagogical reasons we have chosen a purely classical approach, most of the results easily surviving quantization.

In Sec. 2 the general theory[6–8] of massive n-twistor particles is reviewed. At the center of the formalism lies the kinematic twistor in terms of which the momentum, angular momentum, and center of mass line of the particle are expressed. As mentioned before, the IST of the system leaves the kinematic twistor invariant.

In Sec. 3 we specialize to three-twistor systems and study the IST, which turns out to be the inhomogeneous SU(3) group [ISU(3)]. Three spaces play a critical role here: (1) twistor space T on which we choose three points $Z_i^\alpha = (X^\alpha,Y^\alpha,Z^\alpha)$ or, alternately, three copies of T, i.e., $T\times T\times T$; (2) complex Minkowski space which has the naturally chosen triple of points x^a,y^a,z^a, each being the intersections of the twistor pairs (Y^α,Z^α), (Z^α,X^α), and (X^α,Y^α); and (3) unitary space, a three-complex-dimensional affine space on which the elementary representation of the ISU(3) group acts as the isometries. It can be viewed as the homogeneous space ISU(3)/SU(3).

In Sec. 4 we investigate the translation subgroup of the ISU(3) group and show that a translation along a given axis in unitary space leaves the corresponding twistor unchanged while shifting the other two in complex space-time along their subsystem or *partial* center of mass line. We also find a unique correspondence between the time development of the system in Minkowski space and a special translation in unitary space.

In Sec. 5 we study the SU(3) subgroup of ISU(3) and show how from its generators we can find a unique "complex internal center of mass" world line in unitary space. This is in analogy with the use of the homogeneous Lorentz group generators to find the center of mass.[1] We further discuss the internal or unitary spin (which is analogous to the Pauli–Lubanski vector) and show its relationship to the space-time spin. Section 6 deals with the mass triangle defined by the partial center of mass points, while Sec. 7 presents the spin deficiency formula, i.e., the relationship between the total spin and the constituent spins. According to this formula the total spin is the sum of the massive subsystem spins minus the spins of the three twistors.

[a)] Work supported by a National Science Foundation Cooperation Grant.

 0022-2488/82/112108-08$02.50 © 1982 American Institute of Physics

In the concluding sections we use the Penrose blob notation[9] to facilitate lengthy algebraic computation. An introduction to the blob notation is given in an Appendix.

2. THE TWISTOR CONSTITUENTS

Consider a massive particle in Minkowski space-time as a system of $n \geqslant 2$ massless constituent twistors Z_i^α, $i = 1,...,n$. The particle has the kinematical twistor[1]

$$A^{\alpha\beta} = 2Z_i^{(\alpha} I^{\beta)\gamma} \overline{Z}^i_\gamma , \tag{2.1}$$

where $I^{\alpha\beta}$ is the infinity twistor which breaks the conformal invariance. The summation convention holds for Roman twistor "flavor" indices (flavor indices share the property of Greek twistor indices that they are raised and lowered by complex conjugation, $(\overline{Z_i^\alpha}) = (\overline{Z}^i_\alpha)$. Each term on the right of Eq. (2.1) for a fixed value of i is the kinematical twistor of one of the massless constituents.

The kinematical twistor is decomposed into the spinor parts

$$[A^{\alpha\beta}] = \begin{bmatrix} -2i\mu^{AB} & p^A_{B'} \\ p^{A'}_B & 0 \end{bmatrix} . \tag{2.2}$$

Here μ^{AB} is the total angular momentum spinor (symmetric in its indices) and the Hermitian spinor p^A_B, is the four-momentum. The *center-of-mass line* of the system consists of the points of real Minkowski space-time

$$x^{AA'}(\tau) \overset{def}{=} m^{-2}(\mu^{AB} p^{A'}_B + \bar{\mu}^{A'B'} p^A_{B'}) + \tau m^{-1} p^{AA'} , \tag{2.3}$$

where $m^2 \overset{def}{=} p^A_{B'} p^{B'}_A$ is the squared rest mass and the real parameter τ is the proper time. It does not appear to be possible to express the center of mass in terms of the kinematical quantities in a manifestly twistor-invariant form, due to the fact that the concept of the center of mass in not invariant with respect to translations. However, using the constituent twistors directly, one can define a *center of mass point twistor*[8]

$$R^{\alpha\beta} \overset{def}{=} 2m^{-2} Z_i^\alpha Z_k^\beta \bar{M}^{ik} . \tag{2.4}$$

The quantities

$$M_{ik} \overset{def}{=} Z_i^\alpha Z_k^\beta I_{\alpha\beta} \tag{2.5}$$

are called *mass amplitudes* and for $n > 2$ are the *partial mass amplitudes* of the two-component subsystem labeled by i and k, and are such that the mass squared of the system may be written as a sum of *partial mass squares*

$$m^2 = M_{ik} \bar{M}^{ik} . \tag{2.6}$$

The point twistor (2.4) decomposes according to

$$R^{\alpha\beta} = \begin{bmatrix} -\frac{1}{2} r_{RR'} r^{RR'} \epsilon^{AB} & ir^A_{B'} \\ -ir^B_{A'} & \epsilon_{A'B'} \end{bmatrix} , \tag{2.7}$$

where $r^{AA'}$, the *center of mass point*, is a point of the complex Minkowski space-time. The point $r^{AA'}$ lies on the *complex center of mass line* of the system. The complex center of mass line is the set of points[10]

$$z^{AA'} = m^{-2}(\mu^{AB} p^{A'}_B + \bar{\mu}^{A'B'} p^A_{B'}) + iS^{AA'} m^{-2} + \lambda p^{AA'} m^{-1}, \tag{2.8}$$

where λ is complex and $S_{AA'}$ denotes the Pauli–Lubanski spinor

$$S_{AA'} = i(\mu_{AB} p^B_{A'} - \bar{\mu}_{A'B'} p^{B'}_A) . \tag{2.9}$$

In the rest frame of the particle defined by the form of the four-momentum

$$[p^{AA'}] = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} ,$$

the Pauli–Lubanski spinor becomes proportional to the nonrelativistic spin. Thus the spin $j_{AA'}$ is given

$$j_{AA'} = m^{-1} S_{AA'} . \tag{2.10}$$

The Pauli–Lubanski spinor is the sole nonvanishing part of the spin twistor[5]

$$S^\beta_\alpha = \tfrac{1}{2}(\bar{A}_{\alpha\rho} A^{\rho\beta} + \tfrac{1}{2} m^2 \delta^\beta_\alpha) , \tag{2.11}$$

according to

$$S^\beta_\alpha = \begin{bmatrix} 0 & 0 \\ -S^{A'B} & 0 \end{bmatrix} . \tag{2.12}$$

The internal symmetry transformations of the kinematical twistor are[11]

$$\begin{aligned} \hat{Z}^\alpha_i &= U^k_i (Z^\alpha_k + \Lambda_{kl} I^{\alpha\beta} \overline{Z}^l_\beta) , \\ \overline{Z}^i_\alpha &= \overline{U}^i_k (\overline{Z}^k_\alpha + \overline{\Lambda}^{kl} I_{\alpha\beta} Z^\beta_l) , \end{aligned} \tag{2.13}$$

where $[U^i_k]$ is a $n \times n$ unitary matrix and $[\Lambda_{kl}]$ is skew. The transition $Z^\alpha_i \to \hat{Z}^\alpha_i$ amounts to selecting a new set of massless constituents for which the kinematical twistor of the system remains unchanged. Thus the kinematical variables discussed remain unchanged under internal symmetry transformations, with the exception of the center-of-mass twistor. The center-of-mass point is defined directly in terms of constituents and it has been shown by Hughston[8] that internal transformations (2.13) move the complex center-of-mass point over the entirety of the complex center-of-mass line.

The central dogma of twistor particle theory asserts that the state of the system at any instant of time is completely described by the values of the constituent twistors Z^α_i and by their complex conjugates. The space of n-twistors $T \times T \times ... \times T$ admits a naturally defined symplectic form $id\hat{Z}^\alpha_i d\overline{Z}^i_\alpha$. In the sense of Hamiltonian dynamics Z^α_i and $\overline{Z}^i_\alpha$ together play the role of canonically conjugate variables. Accordingly, any function of the form $f(Z^\alpha_i, \overline{Z}^i_\alpha)$ is called a *dynamical quantity*.[12]

We introduce the Poisson bracket of dynamical quantities $f(Z,\overline{Z})$ and $g(Z,\overline{Z})$

$$[f,g] = -i\left(\frac{\partial f}{\partial Z^\alpha_r}\frac{\partial g}{\partial \overline{Z}^r_\alpha} - \frac{\partial f}{\partial \overline{Z}^r_\alpha}\frac{\partial g}{\partial Z^\alpha_r}\right) . \tag{2.14}$$

The Poisson bracket is antisymmetric in f and g and is real when both f and g are real. From (2.14) we identify the general coordinates q^i_α and canonically conjugate momenta p^α_i as follows:

$$q^\alpha_i = Z^\alpha_i , \quad p^i_\alpha = i\overline{Z}^i_\alpha . \tag{2.15}$$

The twistor variables have the Poisson brackets

$$[Z^\alpha_i, Z^\beta_k] = 0, \ [Z^\alpha_i, \overline{Z}^k_\beta] = -i\delta^k_i \delta^\alpha_\beta . \tag{2.16}$$

A transformation generated by a dynamical quantity f is given

$$\delta Z_i^{\alpha} = [Z_i^{\alpha} f] , \qquad \delta \overline{Z}_i^{\alpha} = [\overline{Z}_{\alpha}^{i} f] . \tag{2.17}$$

The unitary transformations $\hat{Z}_i^{\alpha} = U_i^k Z_k^{\alpha}$ are generated by the functions

$$B_k^i = Z_k^{\alpha} \overline{Z}_{\alpha}^{i} , \tag{2.18}$$

forming a Hermitian matrix, and the transformations $\tilde{Z}_i^{\alpha} = Z_i^{\alpha} + \Lambda_{ik} I^{\alpha\beta} \overline{Z}_{\beta}^{k}$ are generated by the mass amplitudes

$$M_{ik} = Z_i^{\alpha} Z_k^{\beta} I_{\alpha\beta} \text{ and } \overline{M}^{ik} = \overline{Z}_{\alpha}^{i} \overline{Z}_{\beta}^{k} I^{\alpha\beta} . \tag{2.19}$$

The Λ transformations commute and are called *internal translations.*

As long as we are concerned with a massive particle in free motion, the decomposition into allowed twistor constituents is immaterial for the motion of the particle. This reflects the invariance of the kinematical twistor with respect to the internal transformations (2.13). The idea is, however, that the behavior of the particle in interactions should depend on the substructures present in a twistor decomposition.

The n-twistor particle where $n \geqslant 3$, possesses massive parts. Such a substructure consists of two or more null constituents. Clearly, a two-twistor particle has no massive subsystems. The simplest place to study massive subsystems is a three-twistor particle. In the next section we discuss some features unique to three-twistor particles.

3. STRUCTURE OF THREE-TWISTOR PARTICLES

The internal structure of a massive particle described by three-twistors can be examined in terms of the three two-twistor subsystems obtained by considering the three twistors pairwise. Each such two-twistor subsystem defines a massive particle in space-time with well defined (real and complex) center-of-mass line, spin, and center-of-mass point. These physical properties of the subsystems combine to yield the properties of the entire system in unexpected and interesting ways, given an ordered triple of twistors $(Z_1^{\alpha}, Z_2^{\alpha}, Z_3^{\alpha}) \in T$. (It is sometimes preferable to think of the triple as a point in $T \times T \times T$.) Any one of these, Z_1^{α}, has its kinematical twistor, $A_i^{\alpha\beta}$, and any pair of these, $(Z_i^{\alpha}, Z_j^{\alpha})$, $i<j$, has its associated kinematical twistor $A_{ij}^{\alpha\beta}$. While a single twistor describes a massless system in Minkowski space-time, a pair of twistors describes a massive system. The internal symmetry transformations change the kinematical twistors of the one- and two-twistor subsystems while the kinematical twistor of the entire system is unchanged and, in fact, this constrains the changes in $A_i^{\alpha\beta}$ and $A_{ij}^{\alpha\beta}$ since

$$A^{\alpha\beta} = \frac{1}{2} \sum_{\substack{i,j \\ i<j}} A_{ij}^{\alpha\beta} = \sum_i A_i^{\alpha\beta} = \sum_{i,j} A_{ij}^{\alpha\beta} - \sum_i A_i^{\alpha\beta} .$$

The internal states will be used in the description of interactions and the manner in which the various concepts are linked is of importance. To proceed further, the internal symmetry group must be examined more closely. The internal translations are generated by the partial mass amplitudes, which can be given equivalently by

$$d^i = \tfrac{1}{2} \epsilon^{ijk} M_{jk} \text{ and } \bar{d}_i = \tfrac{1}{2} \epsilon_{ijk} \overline{M}^{jk} . \tag{3.1}$$

The mass squared is easily given by $m^2 = 2 d^i \bar{d}_i$ and is positive. The unitary transformations are generated by B_k^i which satisfy

$$\bar{d}_i B_k^i d^k = 0 . \tag{3.2}$$

Writing B_k^i as a trace-free part A_k^i plus a trace results in

$$B_k^i = A_k^i + \delta_k^i B , \tag{3.3}$$

where $B = \frac{1}{3} B_r^r$. From (3.2), the trace may be written as

$$B = 2m^{-2} \bar{d}_i A_k^i d^k . \tag{3.4}$$

Thus the trace of the generators of the unitary transformations can be written in terms of the remaining 14 internal symmetry generators and without loss of generality the unitary transformations will be restricted to SU(3) in the remainder of the paper.

As in (3.1), introducing alternative translation parameters

$$t^i = \tfrac{1}{2} \epsilon^{ijk} \Lambda_{jk} , \tag{3.5}$$

the internal symmetry transformations take the form

$$\hat{Z}_i^{\alpha} = U_i^k (Z_k^{\alpha} - \epsilon_{klm} t^l I^{\alpha\beta} \overline{Z}_{\beta}^{m}) , \tag{3.6}$$

with U_i^k an element of SU(3) and t^l an element of $\mathbb{C}^3$. This transformation is represented by the pair $(\underline{U}, \underline{t})$ and the group product structure follows from the composition of two successive transformations. Thus $(\underline{U}, \underline{t})$ followed by $(\underline{U}', \underline{t}')$ gives after a short calculation

$$(\underline{U}', \underline{t}') \cdot (\underline{U}, \underline{t}) = (\underline{U}' \underline{U}, \ \underline{t} + \underline{U}^{+} \underline{t}') \tag{3.7}$$

where $\underline{U}^{+}$ is the Hermitian adjoint of $\underline{U}$. Equation (3.7) defines the 14-parameter group denoted by ISU(3) and called the inhomogeneous SU(3) group of internal symmetry transformations (IST). This group acts on $\mathbb{C}^3$ with coordinates z^i, $i = 1,2,3$ as a group of point transformations where $(\underline{U}, \underline{t})$ gives

$$\hat{Z}^i = U_k^i (z^k + t^k) . \tag{3.8}$$

That is to say, the ISU(3) is realized as the isometry group on $\mathbb{C}^3$ of the Hermitian line element $dz^i d\bar{z}_i$ and this gives $\mathbb{C}^3$ the structure[13] of a unitary space, U^3. An alternate point of view is to consider U^3 as the homogeneous space ISU(3)/SU(3). We will give later yet another method of obtaining U^3.

The transformations in (3.6) which act on T^3 constitute the twistor realization of ISU(3). The same elements of ISU(3) act on U^3 via (3.8). We need not bother to compute the generators of the ISU(3) group in the isometry representation since they are already available in the twistor realization [cf. Eqs. (2.18), (3.3), and (3.1)]

$$A_k^i = Z_k^{\alpha} \overline{Z}_{\alpha}^{i} - \tfrac{1}{3} \delta_k^i Z_r^{\alpha} \overline{Z}_{\alpha}^{r} , \qquad d^i = \tfrac{1}{2} \epsilon^{ijk} Z_j^{\alpha} I_{\alpha\beta} Z_k^{\beta} , \qquad \bar{d}_i = \tfrac{1}{2} \epsilon_{ijk} \overline{Z}_{\alpha}^{j} I^{\alpha\beta} \overline{Z}_{\beta}^{k} . \tag{3.9}$$

We identify the translation generators d^i and $\bar{d}_i$ as the components of the *complex internal momentum* of a particle in the unitary space. The generators A_k^i of SU(3) rotations constitute the *total internal angular momentum* of the particle.

Under a translation $\hat{z}^i = z^i + t^i$, the total angular momentum and B change as [cf. Eq. (3.7)]

$$\hat{A}^i_{\ k} = A^i_{\ k} + d^i \bar{l}_k + t^i \bar{d}_k - \tfrac{1}{3}(d^r \bar{l}_r + t^r \bar{d}_r)\delta^i_k \,, \quad (3.10)$$

$$\hat{B} = B + \tfrac{2}{3}(d^i \bar{l}_i + t^i \bar{d}_i) \,, \quad (3.11)$$

and the complex momentum remains invariant. The behavior of dynamical quantities with respect to SU(3) rotations is implicit in our tensor notation.

For future use we wish to spell out the meaning of the transformation (3.10); the $A^i_{\ k}$ are the generating functions of the isometries in unitary space with the origin as a fixed point while the $\hat{A}^i_{\ k}$ are the generating functions of isometries keeping the point $z^i = t^i$ fixed. In this manner $\hat{A}^i_{\ k}$ can be thought of as a tensor field on unitary space with $t^i = z^i$.

To summarize this section, there are three spaces which play fundamental roles here. The first is twistor space T on which we take three points $(Z^\alpha_1, Z^\alpha_2, Z^\alpha_3)$ to specify our massive system. (An alternative and sometimes necessary point of view would be to choose a twistor from each of three different copies of T). The IST, i.e., ISU(3), acts on these three points preserving the kinematic twistor. Since pairs of twistors define points of complex Minkowski space, the three twistors define three points in complex Minkowski space which are moved about by the IST. The third space is the unitary space U^3 having the isometry group ISU(3). The generators of ISU(3) define a vector field d^i and tensor field $A^i_{\ j}$ on the unitary space in a manner analogous to the way the Lorentz group defines the momentum and angular momentum fields on Minkowski space. In Sec. 5 we will show how $A^i_{\ j}$ and d^i define an internal center of mass line and internal spin in analogy to the way angular momentum and momentum determine a center of mass and spin.

4. INTERNAL TRANSLATIONS

We now explore the effect of internal translations on the structure of the three-twistor particle. Consider first a translation

$$\hat{z}^i = z^i + t^i \,,$$

with $t^i = (0,0,\Lambda)$, by the complex amount Λ in the z^3 direction of the unitary space. In the twistor realization, Eq. (3.7), the momentum parts of the constituent twistors $[\pi_{iA'}$, where $Z^\alpha_i = (\omega^A_i, \pi_{iA'})]$ remains unaffected,

$$\hat{\pi}_{iA'} = \pi_{iA'} \,, \quad (4.1)$$

and

$$\hat{\omega}^A_1 = \omega^A_1 + \Lambda \bar{\pi}^{2A} \,,$$
$$\hat{\omega}^A_2 = \omega^A_2 - \Lambda \bar{\pi}^{1A} \,, \quad (4.2)$$
$$\hat{\omega}^A_3 = \omega^A_3 \,.$$

We compare the change in the ω parts with the effect of a translation in Minkowski space-time

$$\tilde{x}^{AA'} = x^{AA'} + a^{AA'} \,. \quad (4.3)$$

This latter gives

$$\tilde{\omega}^A = \omega^A + i a^{AA'} \pi_{A'} \,, \quad \tilde{\pi}_{A'} = \pi_{A'} \,, \quad (4.4)$$

for any twistor. Consider, in particular, the pair Z^α_1 and Z^α_2. Choose

$$a^{AA'} = \lambda (\bar{\pi}^A_1 \pi^{1A'} + \bar{\pi}^A_2 \pi^{2A'}) \,, \quad (4.5)$$

where λ is complex. From Eqs. (4.4),

$$\tilde{\omega}^A_1 = \omega^A_1 + i\lambda \bar{\pi}^{2A} \pi^{A'}_2 \pi_{1A'} \,,$$
$$\tilde{\omega}^A_2 = \omega^A_2 + i\lambda \bar{\pi}^{1A} \pi^{A'}_1 \pi_{2A'} \,,$$

or using (2.5),

$$\tilde{\omega}^A_1 = \omega^A_1 + i\lambda M_{12} \bar{\pi}^A_2 \,,$$
$$\tilde{\omega}^A_2 = \omega^A_2 - i\lambda M_{12} \bar{\pi}^A_1 \,, \quad (4.6)$$

Choosing the parameter λ to be

$$\lambda = -i\Lambda / M_{12} \quad (4.7)$$

we have the following result:

The special internal translation with $(t^i) = (0,0,\Lambda)$ shifts the twistors Z^α_1 and Z^α_2 parallelly along the time-like center of mass line of the massive two-twistor subsystem they represent and leaves the twistor Z^α_3 invariant. A similar result is obtained for translations along the other two axes.

Consider next the space-time translation in the direction of the total four-momentum of the system

$$a^{AA'} = \tau(\bar{\pi}^{1A} \pi^{A'}_1 + \bar{\pi}^{2A} \pi^{A'}_2 + \bar{\pi}^{3A} \pi^{A'}_3) \,. \quad (4.8)$$

Is this possibly an internal translation? From Eqs. (4.4) we obtain

$$\tilde{\omega}^A_1 = \omega^A_1 + i\tau(M_{12} \bar{\pi}^{2A} + M_{13} \bar{\pi}^{3A}) \,,$$
$$\tilde{\omega}^A_2 = \omega^A_2 + i\tau(M_{21} \bar{\pi}^{1A} + M_{23} \bar{\pi}^{3A}) \,, \quad (4.9)$$
$$\tilde{\omega}^A_3 = \omega^A_3 + i\tau(M_{31} \bar{\pi}^{1A} + M_{32} \bar{\pi}^{2A}) \,,$$

or, in matrix form using (3.1),

$$\begin{pmatrix} \tilde{\omega}^A_1 \\ \tilde{\omega}^A_2 \\ \tilde{\omega}^A_3 \end{pmatrix} = \begin{pmatrix} \omega^A_1 \\ \omega^A_2 \\ \omega^A_3 \end{pmatrix} + i\tau \begin{pmatrix} 0 & d^3 & -d^2 \\ -d^3 & 0 & d^1 \\ d^2 & -d^1 & 0 \end{pmatrix} \begin{pmatrix} \bar{\pi}^{1A} \\ \bar{\pi}^{2A} \\ \bar{\pi}^{3A} \end{pmatrix} . \quad (4.10)$$

This defines an internal translation with

$$(t^i) = i\tau(d^1, d^2, d^3) \,. \quad (4.11)$$

What we have is a translation in the direction of the unitary momentum d^i by the amount τ. The significance of this result lies in the fact that it establishes a map from the time development of the system in space-time to the development in unitary space parallel to the unitary momentum.

To conclude this section we observe that translations of the form (4.11) exhaust the unitary translations which can be pictured equivalently as space-time translations. The reason for this is that space-time translations not along the centre-of-mass line of the system alter the angular momentum. However, the angular momentum is preserved by all internal transformations since these preserve the kinematic twistor.

5. SPIN AND ROTATION

In addition to its Minkowski space structure (momentum, mass, angular momentum, center of mass, etc.) a three-twistor particle has an associated unitary space structure,

namely a point (or origin) in unitary space and a complex "internal center of mass" world-line also in unitary space with a related internal spin tensor (which is the unitary space analog of the Pauli–Lubanski spin vector).

To see the point structure we note that three-twistor space has 24 real dimensions ($3\times4\times2$) while the kinematic twistor $A^{\alpha\beta}$ has ten real components (momentum and angular momentum) and thus the kinematic subspace defined by $A^{\alpha\beta}$ constant is 14-real-dimensional. The equivalence classes of points in this space (eight-dimensional) defined as those points connected by SU(3) transformations, i.e., $Z'^{\alpha}_{i} = U^{j}_{i} Z^{\alpha}_{j}$, $U^{j}_{i} \in$ SU(3), can be identified with the points of unitary space. The equivalence classes can be parametrized by points in $\mathbb{C}^3$ (six real dimensions) i.e., by the translations $Z'^{\alpha}_{i} = Z^{\alpha}_{i} + \Lambda_{ij} I^{\alpha\beta} \bar{Z}^{j}_{\beta}$, from some arbitrarily chosen "origin" Z^{α}_{i}.

(Note that by associating this arbitrarily chosen origin with the group identity element, the kinematic subspace can be considered as the ISU(3) group space. Note further that if we had considered originally the group IU(3), the U(1) part would have an action on the ISU(3) manifold which would not be the action of an ISU(3) element. Nevertheless, *locally* one could duplicate the U(1) action by an ISU(3) element. This explains from a group theoretical point of view the relationship (3.4) between the U(1) generator and the ISU(3) generator).

In order to understand and see the internal center of mass line and internal spin tensor we define

$$C^{i}_{j} = A^{i}_{j} + \tfrac{1}{2} B \delta^{i}_{j}, \tag{5.1}$$

and obtain from (3.10) and (3.11) the transformation law under translations $\hat{z}^{i} = z^{i} + t^{i}$

$$\hat{C}^{i}_{j} = C^{i}_{j} + d^{i} \bar{t}_{j} + t^{i} \bar{d}_{j}. \tag{5.2}$$

C^{i}_{j} can be decomposed into the four parts

$$C^{i}_{j} = \alpha d^{i} \bar{d}_{j} + \alpha^{i} \bar{d}_{j} + d^{i} \bar{\alpha}_{j} - \frac{2}{m^2} S^{i}_{j}, \tag{5.3}$$

where

$$\begin{aligned} &S^{i}_{j} d^{j} = S^{i}_{j} \bar{d}_{i} = S^{i}_{i} = 0, \; S^{i}_{k} = \overline{S}^{i}_{j}, \\ &\alpha^{i} \bar{d}_{i} = d^{i} \bar{\alpha}_{i} = 0, \\ &\alpha = \left(\frac{2}{m^2}\right)^2 C^{i}_{j} d^{j} \bar{d}_{i} = \frac{3}{m^2} B, \; \alpha^{i} = \frac{2}{m^2} C^{i}_{j} d^{j} - \alpha d^{i}, \\ &\bar{\alpha}_{j} = \frac{2}{m^2} C^{i}_{j} \bar{d}_{i} - \alpha \bar{d}_{j}. \end{aligned} \tag{5.4}$$

(Note that the Hermitian adjoint is defined by $\overline{S}^{i}_{k} = \overline{S}^{k}_{i}$).

If we now insert (5.3) into (5.2) with

$$t^{i} = -\alpha^{i} + i\tau d^{i}, \; \bar{t}_{i} = -\bar{\alpha}_{i} - i\bar{\tau} \bar{d}_{i}, \tag{5.5}$$

we obtain

$$\hat{C}^{i}_{j} = (\alpha + i(\tau - \bar{\tau})) d^{i} \bar{d}_{j} - \frac{2}{m^2} S^{i}_{j}. \tag{5.6}$$

Thus along the *internal center of mass line* defined by (5.5), $\hat{C}^{i}_{j}$ has only the first and last terms of its canonical decomposition. The Hermitian tensor S^{i}_{j} called the *internal spin-tensor* can be explicitly solved for and written $S^{i}_{j} = \epsilon^{iab} \epsilon_{jcd} (A^{c}_{a} + \tfrac{1}{2} B \delta^{c}_{a}) \bar{d}_{b} d^{d}$. From its derivation or by direct calculation it is seen that it is (essentially) the invariant part of C^{i}_{j} under translations, i.e.,

$$[S^{i}_{k}, d^{j}] = 0, \quad [S^{i}_{k}, \bar{d}_{j}] = 0, \tag{5.7}$$

and that it has a canonical decomposition

$$S^{i}_{j} = S(X^{i} \overline{X}_{j} - Y^{i} \overline{Y}_{j}), \tag{5.8}$$

with

$$\begin{aligned} &S^{i}_{j} X^{j} = S X^{i}, \; S^{i}_{j} Y^{j} = -S Y^{i}, \; S \geqslant 0, \\ &X^{i} \bar{d}_{i} = Y^{i} \bar{d}_{i} = 0, \; X^{i} \overline{X}_{i} = Y^{i} \overline{Y}_{i} = 1. \end{aligned}$$

The vector $S^{i} = (S)^{1/2} X^{i}$ contains all the information in S^{i}_{j}. From (5.4) and the first term in (5.6) we have

$$\hat{B} = B + i \frac{m^2}{3} (\tau - \bar{\tau}).$$

Up to the choice of origin $\hat{B}$ can be identified with the imaginary part of τ. There exists a real line which is imbedded in the complex line and parametrized by the real part of τ defined by $\hat{B} = 0$.

To reiterate the material of this section, we have shown that a point in three-twistor space selects an origin in unitary space. $A^{i}_{j}(Z)$ represent the eight generators of SU(3) rotations about this point while d^{i} are the generators of the three complex translations. The tensor field $\hat{A}^{i}_{j}(Z^{\alpha}_{i}, t^{i})$ on the unitary space represents the SU(3) generators about the point $z^{i} = t^{i}$. Assuming that the three twistors Z^{α}_{i} are held fixed (i.e., we have a given internal structure) then simply from the algebraic structure of A^{i}_{j} and $\hat{A}^{i}_{j}$ one is led to the complex line and internal spin-tensor. At the present we make no attempt at a physical interpretation of the "origin", the internal center of mass line, and internal spin tensor other than to say that they are to represent the internal structure of the three-twistor particle. A different choice of the three-twistors obtained from the IST would represent a different particle having a different origin, world-line, and spin tensor but with the same kinematic values.

For a fixed numerical value of t^{i}, $\hat{A}^{i}_{j}$ are SU(3) generators. This is true in particular on the center of mass world-line. However an examination of (5.5) shows that t^{i} is a function of the Z^{α}_{i} and $\overline{Z}^{i}_{\alpha}$ and the functional dependence of $\hat{A}^{i}_{j}$ on the Z^{α}'s is changed. From (5.7) we see that the S^{i}_{j} generate transformations which keep the mass line fixed and by direct calculation we have

$$[S^{i}_{j}, S^{k}_{l}] = (d^{k} \bar{d}_{j} - \delta^{k}_{j} d^{r} \bar{d}_{r}) S^{i}_{l} - (d^{i} \bar{d}_{l} - \delta^{i}_{l} d^{r} \bar{d}_{r}) S^{k}_{j} \tag{5.9}$$

which are the SU(2) Poisson brackets. When the transformations generated by S^{i}_{j} are referred to arbitrary three-twistor systems, then they do not belong to ISU(3) since S^{i}_{j} is not a linear combination of the A^{i}_{j}'s and d^{i}'s. We can consider, however, the restriction of these transformations to systems with a fixed momentum $d^{i} = D^{i}$. Then, S^{i}_{j} are ISU(3) generators. We thus have a parallel, in unitary space, to space-time spin as the generator of O(3) transformations.[14]

Theorem: The unitary spin S^{i}_{j} generates the SU(2) subgroup of ISU(3) leaving the mass amplitudes invariant. [cf. Eq. (5.7)].

The mass amplitudes determine the scalar products of the momentum parts π_{iA} of constituent twistors by M_{ik}

$= \pi_{iA'}\,\pi_k^{A'}$. Hence the effect of transformations generated by S^i_j is a rigid rotation of the momentum spinors $\pi_{iA'}$ (together with the frame defined by any pair of them) about the total four-momentum $p^A{}_{A'}$. For systems with a fixed value of the unitary momentum, $d^i = D^i$, we may choose coordinates $D^i = (m/\sqrt{2})\,\delta^i_3$. In this coordinate system, the unitary spin has the component form, with J_a real,

$$[S^i_k] = \frac{1}{2}m^2\begin{bmatrix} J_3 & J_1 + iJ_2 & 0 \\ J_1 - iJ_2 & -J_3 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Thus the Poisson brackets (5.9), restricted to systems with a fixed momentum may be written

$$[\,J_a, J_b\,] = i\epsilon_{abc}\,J_c, \quad a,b,c, = 1,2,3. \tag{5.10}$$

It has been known for some time[7] that the magnitudes of the space-time spin and of the unitary spin are equal

$$\tfrac{1}{2}m^2\,S^i_k\,S^k_i = -S_{AA'}S^{AA'} \equiv j^2 m^{-2}, \tag{5.11}$$

where we introduce the real parameter j which can take any real value for a classical system (and it will take the values 0, $\frac{1}{2}$, 1, $\frac{3}{2}$,$\cdots$ after quantization). This common magnitude is a Casimir invariant of both the Poincaré and of the inhomogeneous SU(3) groups, the second common Casimir invariant being the mass square

$$d^i\bar{d}_i = p_{AA'}\,p^{AA'} = m^2/2\,.$$

There is a further property that connects the space-time spin and (the negative of) unitary spin: The projections of the spin twistor (2.11) onto the constituent twistors are the negative of the components of the unitary spin

$$S^\alpha_\beta\,Z^\beta_i\,\bar{Z}^k_\alpha = -S^k_i\,. \tag{5.12}$$

This result (which can be proved by direct computation) generalizes a relation holding for two-twistor particle spins.[5] For a two-twistor particle, however, the unitary spin is replaced by the conformally invariant quadratic expressions

$$Z^\alpha_i\,\sigma^j_k\,\bar{Z}^k_\alpha, \quad i,k = i,2\,,$$

where $[\sigma^j_k]$ are the Pauli matrices.

6. THE MASS TRIANGLE

The subsystem of twistors Z^α_1 and Z^α_2 has the squared rest mass

$$m^2_{12} = 2M_{12}\,\bar{M}^{12} = 2d^3\bar{d}_3\,.$$

The center of mass point twistor (2.4) of the subsystem may be written

$$R^{\alpha\beta}_{12} = \frac{1}{M_{12}}(Z^\alpha_1\,Z^\beta_2 - Z^\alpha_2\,Z^\beta_1)\,. \tag{6.1}$$

Using Eq. (6.1) and the center of mass twistors of the remaining massive subsystems, we obtain the center of mass of the three-twistor particle as the linear combination

$$R^{\alpha\beta} = \frac{2}{m^2}(M_{12}\,\bar{M}^{12}R^{\alpha\beta}_{12} + M_{23}\,\bar{M}^{23}\,R^{\alpha\beta}_{23} + M_{31}\,\bar{M}^{31}\,R^{\alpha\beta}_{31})\,. \tag{6.2}$$

The nondiagonal spinor part of (6.2) which is linear in the position vectors [cf. Eq. (2.7)] is

$$r^{AA'} = m^{-2}(m^2_{12}\,r^{AA'}_{12} + m^2_{23}\,r^{AA'}_{23} + m^2_{31}\,r^{AA'}_{31})\,. \tag{6.3}$$

Thus the center-of-mass point $r^{AA'}$ of the particle is a weighted mean of the partial center of masses. Hence the four mass center points lie in a plane. The center of mass $r^{AA'}$ is in the barycenter of the triangle formed by the partial mass centers. Note, however, that the weights are mass squares rather than masses. We now compute the sides of the triangle.

The invariant distance of two complex points in Minkowski space-time

$$r^{AA'}_1 = x^{AA'}_1 + iy^{AA'}_1\,, \qquad r^{AA'}_2 = x^{AA'}_2 + iy^{AA'}_2\,, \tag{6.4}$$

can be expressed in terms of the point twistors $R^{\alpha\beta}_1$ and $R^{\alpha\beta}_2$ as[8]

$$R^{\alpha\beta}_1\,R_{2\alpha\beta} = -\tfrac{1}{2}R^{\alpha\beta}_1\,\epsilon_{\alpha\beta\gamma\delta}\,R^{\gamma\delta}_2 = (r^{AA'}_1 - r^{AA'}_2)(r^1_{AA'} - r^2_{AA'})\,. \tag{6.5}$$

The length of the imaginary part of vector $r^{AA'}$ is

$$R^{\alpha\beta}\bar{R}_{\alpha\beta} = 4y^{AA'}\,y_{AA'}\,. \tag{6.6}$$

The partial mass centers are null-separated in complex Minkowski space-time since any pair of them lie on a common twistor. Let us, however, consider the real mass triangle. From Eq. (6.5), the condition of null separation for the arbitrary subsystems a and b is

$$-R^{\alpha\beta}_{a}R_{b\alpha\beta} = (x_a - x_b)^2 - (y_a - y_b)^2 + 2i(x_a - x_b)\cdot(y_a - y_b) = 0\,. \tag{6.7}$$

Hence using Eq. (2.8), we can express the side c connecting the mass points a and b in terms of the spin vectors $S^{AA'} = m^2 y^{AA'}$ as

$$c^2 \overset{def}{=} (x_a - x_b)^2 - (m^{-2}S_a - m^{-2}S_b)^2. \tag{6.8}$$

It is quite surprising that the spin difference appears in a side length of the mass triangle. From Eq. (6.8) we further have that each side of the mass triangle is orthogonal to the difference of the spins at the endpoints of the side.

To complete the analysis of the mass triangle, we now ask how the lengths of the sides of the mass triangle depend directly on the constituent twistors. A direct substitution into Eq. (6.8) yields unwieldy results. Instead we consider the variants of Eq. (6.7):

$$\begin{aligned} -\bar{R}^{a}_{\alpha\beta}\,\bar{R}^{b\alpha\beta} &= (x_a - x_b)^2 - (y_a - y_b)^2 - 2i(x_a - x_b)\cdot(y_a - y_b)\,, \\ -\bar{R}^{\alpha\beta}_{a}\,\bar{R}^{b}_{\alpha\beta} &= (x_a - x_b)^2 - (y_a + y_b)^2 + 2i(x_a - x_b)\cdot(y_a + y_b)\,, \\ -\bar{R}^{a\alpha\beta}\,\bar{R}_{b\alpha\beta} &= (x_a - x_b)^2 - (y_a + y_b)^2 - 2i(x_a - x_b)\cdot(y_a + y_b)\,, \end{aligned} \tag{6.9}$$

Taking the sum, and adding the lengths of imaginary parts [Eq. (6.6)],

$$c^2 = (x - x)^2$$
$$= \tfrac{1}{4}\Big[R^{\alpha\beta}\,\bar{R}_{\alpha\beta} + R^{\alpha\beta}\bar{R}_{\alpha\beta} - ((\bar{R}^{\alpha\beta} + \bar{R}^{\alpha\beta})(\bar{R}_{\alpha\beta} + \bar{R}_{\alpha\beta})\Big]. \tag{6.10}$$

For calculational convenience and to illustrate its usefulness, Penrose's blob notation will be used to write the lengths in terms of the constituent twistors (cf. Appendix). Let us write

$$Z_1^\alpha = \quad , \quad Z_2^\alpha = \quad , \quad Z_3^\alpha = \quad . \tag{6.11}$$

From Eq. (2.8), and $\underline{a} = (13)$, $\underline{b} = (23)$,

$$R^{\alpha\beta}\,R_{\alpha B} = \tfrac{2}{\quad} \quad , \tag{6.12}$$

where the mass amplitudes are

$$\underset{a}{M} = \quad , \qquad \overset{b}{M} = \quad . \tag{6.13}$$

Proceeding with the evaluation of terms in Eq. (6.10) we obtain the expressions for the sides of the mass triangle:

$$a^2 = \tfrac{1}{2} \quad ,$$
$$b^2 = \tfrac{1}{2} \quad , \tag{6.14}$$
$$c^2 = \tfrac{1}{2} \quad .$$

Thus, typically, in tensor notation

$$c^2 = \frac{2}{(Z_1^\alpha Z_3^\beta I_{\alpha\beta})(\bar{Z}_\gamma^1 \bar{Z}_\delta^3 I^{\gamma\delta})(Z_2^\mu Z_3^\nu I_{\mu\nu})(\bar{Z}_\lambda^2 \bar{Z}_\rho{}^3 I^{\lambda\rho})} \times \{(Z_3^\kappa Z_1^{[\eta} I_{\kappa\eta}) Z_2^{\zeta]} \bar{Z}_{[\zeta}^2 (\bar{Z}_{\sigma]}^1 \bar{Z}_\tau^3 I^{\tau\sigma}) Z_3^\omega \bar{Z}_\omega^3 - (Z_3^\kappa Z_1^{[\eta} I_{\kappa\eta}) Z_2^{\zeta]} \bar{Z}_\zeta^3 \bar{Z}_\tau^3 I^{\tau\sigma} \bar{Z}_{[\sigma}^1 \bar{Z}_{\omega]}^2 Z_3^\omega\}$$

where antisymmetrization in indices in the brackets is understood.

7. THE SPIN DEFICIENCY FORMULA

In this section we prove the spin deficiency theorem. The content of this theorem is that joining three two-twistor particles by identifying their constituents pairwise gives a total spin which is the sum of the spins diminished by the spins of the twistor constituents.

$$S_k^i = S_{12k}^i + S_{23k}^i + S_{31k}^i - S_{1k}^i - S_{2k}^i - S_{3k}^i \,. \tag{7.1}$$

Equation (2.1) takes the form for a two-twistor particle in the blob notation: $\quad = \quad + \quad$. Inserting this in the spin twistor (2.1) written in terms of blobs as

$$2 \quad = \quad + \; m^2/2 \,, \tag{7.2}$$

and using the identity[5]

$$= \tfrac{1}{\quad} \{ \quad + \quad + \quad - \quad - \quad \}, \tag{7.3}$$

we obtain

$$2 \quad = (\quad - \quad)(\quad - \quad) + 2 \quad + 2 \quad . \tag{7.4}$$

We now compute the spin of a three-twistor particle similarly. Inserting the rest-mass square

$$\tfrac{1}{2}m^2 = \quad + \quad + \quad , \tag{7.5}$$

and the kinematical twistor

$$= \quad + \quad + \quad .$$

in Eq. (7.2), we have

$$2 \quad = (\quad + \quad + \quad) \times (\quad + \quad + \quad) + (\quad + \quad + \quad) \,.$$

Using a judicious amount of identities of the form (7.3) for various subsystems, we obtain for the spin of the three-twistor particle

$$2 \quad = (\quad - \quad - \quad) \quad + (\quad - \quad - \quad) \quad + (\quad - \quad - \quad) \quad \tag{7.7}$$
$$+ 2 \quad + 2 \quad + 2 \quad \quad 2 \quad + 2 \quad + 2 \quad .$$

When we compare this expression with the spin twistors of the massive subsystems [cf. Eq. (7.4)] and with the spin twistors

(7.8)

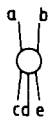
of the zero-mass constituents, we obtain the spin definiency formula in the blob notation

(7.9)

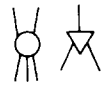
APPENDIX: THE BLOB NOTATION

The blob notation of abstract tensor systems was first introduced by Penrose.[9] Its advantages over the more conventional formalisms of Ricci, Levi–Civita, and Einstein are probably best understood in terms of the physiology of human perception.

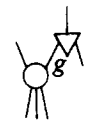
A tensor is drawn in the diagrammatic notation as a blob with arms and legs depicting the upper and lower indices

$T^{ab}_{cde} =$  $U^{f}_{gh} =$ 

The outer product of tensors T^{ab}_{cde} and U^{f}_{gh} is the juxtaposition of blobs

$T^{ab}_{cde}\, U^{f}_{gh} =$ 

To contract a pair of indices, one connects the corresponding arm and leg,

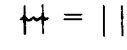
$T^{ag}_{cde}\, U^{f}_{gh} =$ 

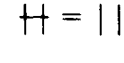
A Kronecker symbol δ^{a}_{b} is represented by a line segment

$\delta^{a}_{b} =$

The notation converts identities of the kind $U^{f}_{gr}\, \delta^{r}_{h} = U^{f}_{gh}$ into trivials partitions of some index line.

Symmetrization and skewing in like indices is denoted according to the scheme

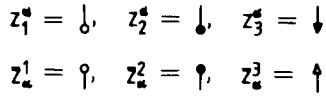




The dimension n of the tensor system is given by the loop

$n = \delta^{a}_{a} =$

In twistor theory one is interested in dimension $n = 4$. Taking twistor complex conjugates is an involution that has the effect of turning the symbols upside down. Thus the blobs of twistors Z^{α}_{1}, Z^{α}_{2}, Z^{α}_{3}, and of their complex conjugates are drawn



The skew unit twistor and the infinity twistor are denoted, respectively,

$\epsilon^{\alpha\beta\gamma\delta} =$  and $I^{\alpha\beta} =$

It is useful in computations to keep in mind some of their algebraic properties in the blob notation such as

[1]Our notation is in agreement with R. Penrose and M. A. H. MacCallum, Phys. Rep. 6C, 241 (1973).
[2]R. Penrose, Int. J. Theor. Phys. **1**, 61 (1968).
[3]K. P. Tod and Z. Perjés, Gen. Rel. Grav. 7, 903 (1976).
[4]K. P. Tod, Reps. Math. Phys. **11**, 339 (1977).
[5]Z. Perjés, Phys. Rev. D**11**, 2031 (1975).
[6]Z. Perjés, Phys. Rev. D**20**, 1857 (1979).
[7]The coincidence of internal and space-time Casimir invariants has been observed in Penrose, "Twistors and Particles; an Outline," in *Quantum Theory and the Structures of Time and Space*, edited by L. Castell, M. Drieschner, and C. F. von Weizsacker (Hanser, Munich, 1975).
[8]L. P. Hughston: *Twistors and Particles, Lecture Notes in Physics*, Vol. 97 (Springer, Berlin, 1979).
[9]R. Penrose, "Applications of Negative Dimensional Tensors," in *Combinatorial Mathematics*, edited by D. J. A. Welsh (Academic, New York, 1971).
[10]E. T. Newman and J. Winicour, J. Math. Phys. **15**, 1113 (1974).
[11]R. Penrose and G. A. J. Sparling, A Note on the *n*-Twistor Internal Symmetry Group, "in *Advances in Twistor Theory*, edited by L. P. Hughston and R. S. Ward, (Pitman, San Francisco, 1979).
[12]Ter Haar, *Elements of Hamiltonian Mechanics* (North-Holland, Amsterdam, 1968).
[13]H. S. M. Coxeter: Regular Complex Poltopes (Cambridge U. P., Cambridge, 1974).
[14]E. P. Wigner, Ann. Math. **40**, 149 (1939).

Trace identities in the inverse scattering transform method associated with matrix Schrödinger operators

L. Martínez Alonso and E. Olmedilla
Departamento de Métodos Matemáticos de la Física, Facultad de Ciencias Físicas, Universidad Complutense, Madrid-3, Spain

(Received 5 February 1982; accepted for publication 14 May 1982)

Trace identities arising in the scattering theory of one-dimensional matrix Schrödinger operators are deduced. They derive from the properties of an asymptotic expansion of the trace of the resolvent kernel in inverse powers of the spectral parameter. Applications of these trace identities for characterizing infinite families of conservation laws for nonlinear evolution equations are given.

PACS numbers: 03.65.Nk, 02.30.Tr

1. INTRODUCTION

One of the most important properties of the nonlinear equations which are integrable by means of the inverse scattering transform method is the existence of an infinite family of local conservation laws. A natural explanation of this fact, for the case of the Korteweg–de Vries equation, was provided by Faddeev and Zakharov[1] by using the trace identities arising in the scattering transform theory for the Schrödinger equation. Shortly thereafter trace identities for other spectral problems were applied to the analysis of important nonlinear wave equations as, for instance, the nonlinear Schrödinger equation[2] and the sine–Gordon equation.[3] The interest in trace identities is not only motivated by their connection with conservation laws; they are also a fundamental step in arriving at a description of completely integrable equations in terms of systems of action-angle variables. This latter application of trace identities is particularly relevant in the semiclassical quantization of completely integrable wave equations[4,5] as well as in the quantum inverse scattering transform method.[6]

In this paper we derive the trace identities associated with general matrix Schrödinger spectral problems. The inverse scattering transform theory for Hermitian matrix Schrödinger operators was analyzed by Wadati and Kamijo[7] who also indicated explicit examples of Lax pair equations. On the basis of this theory, Calogero and Degasperis[8] deduced a wide class of nonlinear evolution equations for which the scattering data evolve in a simple form. However, as already observed by Wadati and Kamijo, several of the more important evolution equations appear to be related to non-Hermitian matrix Schrödinger operators, for which a general inverse scattering transform theory has not yet been formulated.

The starting point of our derivation of trace identities is the asymptotic expansion of the trace of the resolvent operator. This expansion can be obtained in two different ways: firstly, by using the algebraic properties of the diagonal of the resolvent kernel in the context of symbolic calculus of differential operators,[9] and, secondly, by means of the analytic properties of the trace of the resolvent as a function of the spectral parameter. The first procedure provides an asymptotic expansion in which the coefficients appear as functionals with polynomial densities depending on the matrix elements of the potential and their derivatives. By the second procedure these coefficients can be expressed in terms of scattering data and this leads us to the trace identities. The paper is organized as follows.

Section 2 deals with the main properties of the Jost solutions and the resolvent kernel for matrix Schrödinger operators. Several of these properties, those included in Theorems 1 and 2, are listed without proof since they derive easily from methods which are similar to the ones used in the scalar case.[10,11] We do, however, provide a detailed proof of Proposition 1 of Sec. 2 since it leads to the characterization of bound states as zeros of the determinant of the Wronskian of two matrix Jost solutions. Section 3 is concerned with the derivation of the trace identities. In Part A a recursion relation is given which enables us to calculate explicitly the coefficients H_n of the asymptotic expansion for the trace of the resolvent operator. In Part B of Sec. 3 we use the analyticity properties of the trace of the resolvent in order to get the expressions of the coefficients H_n in terms of scattering data. Finally, Sec. 4 includes the application of the trace identities to obtain infinite sets of conservation laws for several relevant families of integrable nonlinear evolution equations.

2. MATRIX SCHRÖDINGER SPECTRAL PROBLEMS

A. Notation and basic properties

We will be concerned with the $N \times N$ matrix Schrödinger operator

$$L = -\partial_{xx} + V(x), \quad -\infty < x < \infty, \tag{2.1}$$

where $V(x)$ is a complex $N \times N$ matrix function, non-Hermitian in general, satisfying[12]

$$\int_{-\infty}^{\infty} (1 + x^2)|V(x)|\, dx < \infty. \tag{2.2}$$

Here and below we denote $|M| = \max_j \Sigma_k |M_{jk}|$ for a given matrix M. The operator L acts on vector functions with N components, but in studying its spectral properties it proves useful to consider the following two eigenvalue problems:

$$-\partial_{xx}F + V(x)F = k^2 F, \tag{2.3a}$$

$$-\partial_{xx}\overline{F} + \overline{F}V(x) = k^2 \overline{F}, \tag{2.3b}$$

where both F and $\overline{F}$ are assumed to be $N \times N$ matrix functions. We define the Jost solutions $F_{\pm}$ of (2.3a) and $\overline{F}_{\pm}$ of (2.3b) as those matrix functions verifying the integral equations

 0022-2488/82/112116-06$02.50 © 1982 American Institute of Physics

$$F_{\pm}(k,x) = e^{\pm ikx}\mathbb{1} - \int_x^{\pm\infty} \frac{\sin k(x-y)}{k} V(y)F_{\pm}(k,y)\,dy, \tag{2.4a}$$

$$\overline{F}_{\pm}(k,x) = e^{\pm ikx}\mathbb{1} - \int_x^{\pm\infty} \frac{\sin k(x-y)}{k} \overline{F}_{\pm}(k,y)V(y)\,dy, \tag{2.4b}$$

where $\mathbb{1}$ is the $N\times N$ identity matrix. These integral equations can be analyzed by means of the well-known method of successive approximations.[10,11] In order to describe the results which are derived from this method we introduce the two subsets

$$\mathbb{C}^+ = \{k\in\mathbb{C}:\mathrm{Im}\,k>0\}, \quad \overline{\mathbb{C}}^+ = \{k\in\mathbb{C}:\mathrm{Im}\,k\geqslant 0\}. \tag{2.5}$$

One proves the following theorem.

Theorem 1: The Jost solutions $F_{\pm}(k,x)$, $\overline{F}_{\pm}(k,x)$ exist for all $k\in\overline{\mathbb{C}}^+$ and, as functions of k, are analytic on $\mathbb{C}^+$ and continuous on $\overline{\mathbb{C}}^+$. Moreover, they satisfy the bounds[13]

$$|e^{\mp ikx}\widetilde{F}_{\pm}(k,x)| = 1 + O(1/|k|), \quad |k|\to\infty, \tag{2.6}$$

where $\widetilde{F}_{\pm}$ denotes either $F_{\pm}$ or $\overline{F}_{\pm}$.

Given two solutions F and $\overline{F}$ of (2.3a) and (2.3b), respectively, the Wronskian

$$[\overline{F};F] = \overline{F}(\partial_x F) - (\partial_x\overline{F})F \tag{2.7}$$

is independent of x. This property leads at once to the following relations, valid for all nonzero real k:

$$F_+(k,x) = F_-(-k,x)A(k) + F_-(k,x)B(k), \tag{2.8}$$

$$F_-(k,x) = F_+(-k,x)C(k) + F_+(k,x)D(k), \tag{2.9}$$

$$\overline{F}_+(k,x) = C(k)\overline{F}_-(-k,x) - D(-k)\overline{F}_-(k,x), \tag{2.10}$$

$$\overline{F}_-(k,x) = A(k)\overline{F}_+(-k,x) - B(-k)\overline{F}_+(k,x), \tag{2.11}$$

where A, B, C, D are the matrices defined by

$$A(k) = (1/2ik)[\overline{F}_-(k);F_+(k)],$$
$$C(k) = -(1/2ik)[\overline{F}_+(k);F_-(k)], \tag{2.12}$$
$$B(k) = -(1/2ik)[\overline{F}_-(-k);F_+(k)],$$

$$D(k) = (1/2ik)[\overline{F}_+(-k),F_-(k)]. \tag{2.13}$$

The compatibility of the relations (2.8)–(2.11) implies the constraints

$$C(-k)A(k) + D(k)B(k) = A(-k)C(k) + B(k)D(k) = \mathbb{1}, \tag{2.14a}$$

$$D(-k)A(k) + C(k)B(k) = B(-k)C(k) + A(k)D(k) = 0. \tag{2.14b}$$

The matrices $A(k)$ and $C(k)$ are specially important for our subsequent discussion on the trace identities. Their definition as given in (2.12) can be extended to $\mathrm{Im}\,k>0$, and the properties of the Jost solutions together with the analysis of the integral equations (2.4) imply

Theorem 2: The matrix functions $A(k)$ and $C(k)$ are analytic on $\mathbb{C}^+$ and continuous on $\overline{\mathbb{C}}^+ - \{0\}$. Moreover, they satisfy

$$|A(k)-\mathbb{1}| = O(1/|k|), \quad |C(k)-\mathbb{1}| = O(1/|k|), \quad |k|\to\infty. \tag{2.15}$$

As we shall see later, $\det A(k) = \det C(k)$, but at this point of our exposition it is convenient to content ourselves with the following partial result.

Proposition 1: The functions $\det A(k)$ and $\det C(k)$ have the same set of zeros in $\overline{\mathbb{C}}^+ - \{0\}$.

Proof: Suppose we have $k_0\in\overline{\mathbb{C}}^+ - \{0\}$ such that $\det A(k_0) = 0$. Then there will be a nonzero vector such that $A(k_0)a = 0$. According to (2.12) this implies that

$$\overline{F}_-(k_0,x)\partial_x\varphi(x) = (\partial_x\overline{F}_-(k_0,x))\varphi(x), \tag{2.16}$$

where $\varphi(x)\equiv F_+(k_0,x)a$. Due to the asymptotic behavior of $\overline{F}_-$ as $x\to-\infty$, the matrix $\overline{F}_-(k_0,x)$ will be invertible for all x in some interval $(-\infty,r)$. Therefore (2.16), considered as a first-order differential equation for $\varphi(x)$, has N linearly independent solutions defined on $(-\infty,r)$. But since the Wronskian $[\overline{F}_-(k_0);F_-(k_0)]$ vanishes, each column of $F_-(k_0,x)$ is a solution of (2.16). In addition, these columns are linearly independent for all x in some interval $(-\infty,r')$. Hence, for all x in the interval $(-\infty,\min(r,r'))$ the general solution of (2.16) is a linear combination of the columns of $F_-(k_0,x)$. As a consequence a nonzero vector b exists verifying

$$F_+(k_0,x)a = F_-(k_0,x)b. \tag{2.17}$$

Note that this equation must hold for all $x\in\mathbb{R}$ since $F_+(k_0,x)a$ and $F_-(k_0,x)b$ are solutions of $(L-k^2)\varphi = 0$, and then, if they coincide on an interval they coincide also on the whole line $-\infty<x<\infty$. Now, if we use the definition of $C(k)$ as given in (2.12) we have

$$C(k_0)b = -(1/2ik_0)[\overline{F}_+(k_0);F_+(k_0)]a = 0. \tag{2.18}$$

That is to say, $\det C(k_0) = 0$. In a similar fashion, one proves that each zero of $\det C(k)$ is also a zero of $\det A(k)$. Q.E.D.

Henceforth we will denote

$$Z = \{k\in\overline{\mathbb{C}}^+ - \{0\}:\det A(k) = \det C(k) = 0\}. \tag{2.19}$$

Obviously, from (2.17) it follows that each $k\in Z\cap\mathbb{C}^+$ determines an exponentially decreasing eigenfunction $F_+(k,x)a = F_-(k,x)b$ of the matrix Schrödinger operator L, and then k^2 is a proper eigenvalue of L. However, we notice that if the potential matrix $V(x)$ is non-Hermitian then the set Z may have elements with $\mathrm{Im}\,k = 0$ and these do not correspond to bound states of L.

From the point of view of scattering theory, the Jost solutions F_+ and F_- describe waves incident from the left and the right, respectively. In this way, taking into account the relations (2.8) and (2.9) we deduce that the transmission and reflection coefficients are given by

$$T(k) = C(k)^{-1}, \quad R(k) = D(k)C(k)^{-1}, \tag{2.20}$$

in the case of right incidence, and by

$$\hat{T}(k) = A(k)^{-1}, \quad \hat{R}(k) = B(k)A(k)^{-1}, \tag{2.21}$$

for the left-incidence case. Moreover, by Eqs. (2.14) one readily finds the following relations:

$$T(-k) = (1-\hat{R}(-k)\hat{R}(k))\hat{T}(k)^{-1}, \tag{2.22a}$$

$$R(k) = -\hat{T}(k)\hat{R}(-k)\hat{T}(-k)^{-1}. \tag{2.22b}$$

B. The resolvent kernel

The resolvent operator of L is an integral operator acting on N-component functions and its kernel admits the following representation in terms of the Jost solutions of (2.3a) and (2.3b):

$$R(k,x,y)=\begin{cases}(i/2k)F_+(k,x)A(k)^{-1}\overline{F}_-(k,y), & x\geqslant y,\\ (i/2k)F_-(k,x)C(k)^{-1}\overline{F}_+(k,y), & x\leqslant y.\end{cases} \tag{2.23}$$

In order to justify this fact let us prove

Lemma 1: Given $k\in\overline{\mathbb{C}}^+-(Z\cup\{0\})$ and $x_0\in\mathbb{R}$, then the matrix functions

$$F(x)=(1/2ik)(F_+(k,x)A(k)^{-1}\overline{F}_-(k,x_0)-F_-(k,x)C(k)^{-1}\overline{F}_+(k,x_0)) \tag{2.24a}$$

and

$$\overline{F}(x)=(i/2k)(F_+(k,x_0)A(k)^{-1}\overline{F}_-(k,x)-F_-(k,x_0)C(k)^{-1}\overline{F}_+(k,x)) \tag{2.24b}$$

are solutions of (2.3a) and (2.3b), respectively, verifying

$$F(x_0)=\overline{F}(x_0)=0,\quad (\partial_x F)(x_0)=(\partial_x\overline{F})(x_0)=\mathbb{1}. \tag{2.25}$$

Proof: Consider the function $F(x)$; clearly it is a solution of (2.3a) and it verifies the Wronskian relations

$$[\overline{F}_+(k);F]=\overline{F}_+(k,x_0), \tag{2.26a}$$

$$[\overline{F}_-(k);F]=\overline{F}_-(k,x_0). \tag{2.26b}$$

If we think of (2.26a) as a first-order differential equation for F it is easy to conclude that F must be of the form

$$F(x)=F_+(k,x)M+F_0(k,x), \tag{2.27}$$

where M is some constant matrix and F_0 stands for the solution of (2.3a) satisfying the boundary conditions (2.25). Now, if (2.27) is inserted into (2.26b) and the Wronskian $[\overline{F}_-;F]$ is evaluated at $x=x_0$, it follows at once that $M=0$. Therefore $F=F_0$ and the enunciated property for F is proved. The proof of the statement for $\overline{F}$ is completely similar. Q.E.D.

As a consequence of this lemma, it follows that $R(k,x,y)$ verifies the equations

$$-\partial_{xx}R(k,x,y)+(V(x)-k^2)R(k,x,y)=\delta(x-y), \tag{2.28}$$

$$-\partial_{xx}R(k,y,x)+R(k,y,x)(V(x)-k^2)=\delta(x-y). \tag{2.29}$$

Now, from the properties of the Jost solutions it can be seen that for $k\in\mathbb{C}^+-Z$ the function $R(k,x,y)$ decreases exponentially as (x,y) goes to infinity. Hence (2.28) and (2.29) mean that the integral operator determined by the kernel $R(k,x,y)$ is a bounded two-sided inverse operator of $(L-k^2)$ provided that $k\in\mathbb{C}^+-Z$. Furthermore, it is clear that, as a function of k, $R(k,x,y)$ is analytic on $\mathbb{C}^+-Z$ and continuous on $\overline{\mathbb{C}}^+-(Z\cup\{0\})$. All these properties imply at once that $R(k,x,y)$ is the kernel of the resolvent operator of L.

In a recent paper[14] Ragnisco has given an expression for the kernel of the resolvent operator which contains some trivial misprints: the resolvent kernel $R^{(-)}(x,y,k)$ should change sign, and it is analytically continuable in the upper (not lower) half k-plane.

3. DERIVATION OF THE TRACE IDENTITIES

A. The trace of the resolvent operator

As it has been shown by Gel'fand and Dikii,[9] the diagonal of the resolvent kernel of a differential operator is an interesting algebraic object. In the case of the matrix Schrödinger operator (2.1) we have that according to (2.23) the restriction of $R(k,x,y)$ to the diagonal $x=y$ is given by the matrix function

$$\begin{aligned}R(k,x)&=(i/2k)F_+(k,x)A(k)^{-1}\overline{F}_-(k,x)\\&=(i/2k)F_-(k,x)C(k)^{-1}\overline{F}_+(k,x).\end{aligned} \tag{3.1}$$

Observe that the equality between both expressions for $R(k,x)$ is a consequence of Lemma 1. Following the method of Gel'fand–Diki based on the symbolic calculus of differential operators one proves[9,15] the existence of an asymptotic expansion of $R(k,x)$ for $|k|\to\infty$;

$$R(k,x)=\frac{i}{2k}\sum_{n=0}^{\infty}R_n(x)k^{-2n},\quad R_0=\mathbb{1}, \tag{3.2}$$

where the coefficients $R_n(n\geqslant 1)$ are polynomials, without constant term, depending on the potential $V(x)$ and its derivatives. The explicit form of these coefficients can be calculated by taking into account the expression (3.1) for $R(k,x)$ as a product of solutions of (2.3a) and (2.3b). Indeed, one readily finds that (2.3a) and (2.3b) imply

$$L_V R(k,x)=k^2\partial_x R(k,x), \tag{3.3}$$

where L_V is the operator defined by

$$L_V R\equiv\frac{1}{4}\left(-\partial_{xxx}R+\partial_x\{V,R\}+\{V,\partial_x R\}-\left[V,\int_{+\infty}^{x}[V,R](x')\,dx'\right]\right), \tag{3.4}$$

and $\{\ ,\ \}$ denotes the anticommutator operation for matrices. Insertion of (3.2) into (3.3) leads to the recursion relation

$$\partial_x R_{n+1}=L_V R_n,\quad R_0=\mathbb{1}, \tag{3.5}$$

which allows us to calculate the coefficients R_n. For example, one obtains

$$\begin{aligned}&R_1=\tfrac{1}{2}V,\quad R_2=-\tfrac{1}{8}(V_{xx}-3V^2),\\&R_3=\tfrac{1}{32}(V_{xxx}-5(V^2)_{xx}+5V_x^2+10V^3).\end{aligned} \tag{3.6}$$

We define the trace of the resolvent operator of L as the following integral:

$$\int_{-\infty}^{\infty}\operatorname{tr}\left(R(k,x)-\frac{i}{2k}\mathbb{1}\right)dx, \tag{3.7}$$

where tr( ) denotes the trace operation of matrices. The analysis of the integral equations (2.4) shows that for $\operatorname{Im}k>0$ the Jost solutions verify the asymptotic behavior[16]

$$F_+(k,x)\underset{x\to-\infty}{\sim}A(k)e^{ikx},\quad F_-(k,x)\underset{x\to+\infty}{\sim}C(k)e^{-ikx}, \tag{3.8a}$$

$$\overline{F}_+(k,x)\underset{x\to-\infty}{\sim}C(k)e^{ikx},\quad \overline{F}_-(k,x)\underset{x\to+\infty}{\sim}A(k)e^{-ikx}. \tag{3.8b}$$

Then by (3.1) we have that the integral (3.7) converges for $k\in\mathbb{C}^+-Z$. Now, substitution of (3.2) into (3.7) yields the asymptotic expansion

$$\int_{-\infty}^{\infty}\operatorname{tr}\left(R(k,x)-\frac{i}{2k}\mathbb{1}\right)dx=\frac{i}{2k}\sum_{n=1}^{\infty}H_n[V]k^{-2n}, \tag{3.9}$$

where the coefficients $H_n[V]$ are the functionals depending on V given by

$$H_n[V] = \int_{-\infty}^{\infty} \mathrm{tr} R_n(x)\, dx. \tag{3.10}$$

The above mentioned properties of the coefficients R_n imply that the densities of the functionals H_n are polynomials depending on the matrix elements of the potential V and their derivatives with respect to x. For instance, the first few functionals are

$$H_1[V] = \frac{1}{2}\int_{-\infty}^{\infty} \mathrm{tr}\, V\, dx, \quad H_2[V] = \frac{3}{8}\int_{-\infty}^{\infty} \mathrm{tr}\, V^2\, dx,$$
$$H_3[V] = \frac{5}{32}\int_{-\infty}^{\infty} \mathrm{tr}(V_x^2 + 2V^3)\, dx. \tag{3.11}$$

B. Trace identities

We are going to obtain another expression for the asymptotic expansion (3.9). Our derivation is based on the following identity:

Proposition 2: For all $k\in\mathbb{C}^+ - Z$ it is verified that

$$\int_{-\infty}^{\infty} \mathrm{tr}\left(R(k,x) - \frac{i}{2k}\mathbb{1}\right) dx = -\frac{1}{2k}\frac{\partial_k \det A(k)}{\det A(k)} = -\frac{1}{2k}\frac{\partial_k \det C(k)}{\det C(k)}. \tag{3.12}$$

Proof: Given two solutions F and $\overline{F}$ of (2.3a) and (2.3b), respectively, we have the identity

$$\partial_x[\overline{F}(k);\partial_k F(k)] = -2k\overline{F}(k)F(k). \tag{3.13}$$

Thus, from (3.1) and (3.13) we get

$$\mathrm{tr}\, R(k,x) = -(i/4k^2)\partial_x \mathrm{tr}[A(k)^{-1}\overline{F}_-(k,x);\partial_k F_+(k,x)] = -(i/4k^2)\partial_x \mathrm{tr}[C(k)^{-1}\overline{F}_+(k,x);\partial_k F_-(k,x)]. \tag{3.14}$$

In this way, taking into account the asymptotic behavior (3.8) of the Jost solutions for $\mathrm{Im}\, k > 0$, one finds

$$\int_{-\infty}^{\infty} \mathrm{tr}\left(R(k,x) - \frac{i}{2k}\mathbb{1}\right) dx = -\frac{1}{2k}\mathrm{tr}(A(k)^{-1}\partial_k A(k)) = -\frac{1}{2k}\mathrm{tr}(C(k)^{-1}\partial_k C(k)). \tag{3.15}$$

Therefore, by using the matrix identity

$$\partial_k \det M(k) = \det M(k)\, \mathrm{tr}(M(k)^{-1}\partial_k M(k)), \tag{3.16}$$

the result follows. Q.E.D.

Now, we are ready to improve the statement given in Proposition 1.

Proposition 3: $\det A(k) = \det C(k)$ for all $k\in\overline{\mathbb{C}}^+ - \{0\}$.

Proof: From (3.12) we have that

$$\partial_k\left(\frac{\det A(k)}{\det C(k)}\right) = 0, \tag{3.17}$$

for all $k\in\mathbb{C}^+ - Z$. Because of the analyticity of $\det A(k)$ on the open set $\mathbb{C}^+$ the zeros of $\det A(k)$ on $\mathbb{C}^+$ are isolated points and then $\mathbb{C}^+ - Z$ is an open connected set. Hence, by (3.17) and due to the fact that

$$\lim_{|k|\to\infty} \frac{\det A(k)}{\det C(k)} = 1, \tag{3.18}$$

we deduce that $\det A(k)$ and $\det C(k)$ coincide on $\mathbb{C}^+ - Z$. Moreover, since both functions are continuous on $\overline{\mathbb{C}}^+ - \{0\}$, they coincide also on $\overline{\mathbb{C}}^+ - \{0\}$. Q.E.D.

Because $\det A(k)\to 1$ as $|k|\to\infty$ there will be some $r>0$ such that $\det A(k)$ is never 0 on the simple connected open set $\mathbb{C}_r^+ \equiv \{k\in\mathbb{C}^+ : |k| > r\}$. Consequently, $\det A(k)$ has an analytic logarithm on $\mathbb{C}_r^+$ and Eq. (3.12) may be rewritten as

$$\int_{-\infty}^{\infty} \mathrm{tr}\left(R(k,x) - \frac{i}{2k}\mathbb{1}\right) dx = -\frac{1}{2k}\partial_k \ln[\det A(k)], \quad k\in\mathbb{C}_r^+. \tag{3.19}$$

At this point we make the assumption that no zero of $\det A(k)$ lies on the real axis. This is satisfied automatically when the potential $V(x)$ is Hermitian, but it is not assured in the general case. From this assumption and the properties of $\det A(k)$ we deduce that Z is a finite set $\{k_l : l = 1,...,m\} \subset \mathbb{C}^+$. Moreover, if we denote by n_l the order of the zero k_l $(l = 1,...,m)$, it follows that the function

$$f(k) \equiv [\det A(k)] \prod_l \left(\frac{k+k_l}{k-k_l}\right)^{n_l} \tag{3.20}$$

is analytic on $\mathbb{C}^+$, continuous and never 0 on $\overline{\mathbb{C}}^+ - \{0\}$, and $f(k)\to 1$ as $|k|\to\infty$. Hence, $f(k)$ admits a logarithm branch which is analytic on $\mathbb{C}^+$, continuous on $\overline{\mathbb{C}}^+ - \{0\}$, and such that $\ln f(k)\to 0$ as $|k|\to\infty$. In this way, according to the Beckenbach's generalization of Cauchy's Integral theorem,[17] we have

$$\ln f(k) = \frac{1}{2\pi i}\oint_\gamma \frac{\ln f(z)}{z-k}\, dz, \quad \mathrm{Im}\, k > 0, \tag{3.21}$$

where γ is the contour consisting of a semicircle of radius r in the upper half-plane, plus the real axis from $-r$ to $+r$, except the origin which is avoided along a semicircle of radius ϵ. Now, because $\ln f(z)\to 0$ as $|z|\to\infty$ the large semicircle does not contribute to the integral (3.21) in the limit $r\to\infty$. On the other hand, it can be proved[18] that $|A(k)| \leqslant \mathrm{const}/|k|$ as $k\to 0$. Then $f(k)$ can become infinite when $k\to 0$ no faster than k^{-N}, and it implies that the contribution of the small semicircle to the integral (3.21) vanishes as $\epsilon\to 0$. Therefore it follows that

$$\ln f(k) = \frac{1}{2\pi i}\,\text{⨍}\int_{-\infty}^{\infty} \frac{\ln f(q)}{q-k}\, dq, \quad \mathrm{Im}\, k > 0. \tag{3.22}$$

Here ⨍ signifies that the Cauchy principal value of the integral must be taken at $q = 0$. Analogously, by integrating along a similar contour in the lower half-plane, we find

$$0 = \frac{1}{2\pi i}\,\text{⨍}\int_{-\infty}^{\infty} \frac{\ln f(-q)}{q-k}\, dq, \quad \mathrm{Im}\, k > 0. \tag{3.23}$$

Now, by adding (3.22) and (3.23),

$$\ln f(k) = \frac{1}{2\pi i}\,\text{⨍}\int_{-\infty}^{\infty} \frac{\ln(f(q)f(-q))}{q-k}\, dq, \quad \mathrm{Im}\, k > 0. \tag{3.24}$$

Then, by using (3.20), we deduce that for $k\in\mathbb{C}_r^+$,

$$\ln[\det A(k)] = -\sum_l \ln\left(\frac{k+k_l}{k-k_l}\right)^{n_l} + \frac{1}{2\pi i}\,\text{⨍}\int_{-\infty}^{\infty} \frac{\ln[\det(A(q)A(-q))]}{q-k}\, dq. \tag{3.25}$$

This expression leads to the asymptotic expansion

$$\ln[\det A(k)] = \frac{i}{2k}\sum_{n=0}^{\infty}\left\{\frac{4i}{2n+1}\sum_l n_l k_l^{2n+1} + \frac{1}{\pi}\int_{-\infty}^{\infty} q^{2n}\ln[\det(A(q)A(-q))]\,dq\right\}k^{-2n}. \quad (3.26)$$

In addition, from Proposition 3 and the relations (2.22) we have

$$\det(A(q)A(-q)) = [\det(\mathbb{1} - R(q)R(-q))]^{-1}. \quad (3.27)$$

Therefore, by substituting (3.9) and (3.26) into (3.19), we get that the coefficients of the asymptotic expansion of the trace of the resolvent operator may be expressed in terms of scattering data in the following form:

$$H_n[V] = 2i\sum_l n_l k_l^{2n-1} - \frac{2n-1}{2\pi} \times \int_{-\infty}^{\infty} q^{2n-2}\ln[\det(\mathbb{1} - R(q)R(-q))]\,dq, \quad n \geqslant 1. \quad (3.28)$$

These relations are the trace identities for matrix Schrödinger operators. Observe that the existence of the integrals appearing in (3.28) requires that $R(q)$ must decrease rapidly as $|q| \to \infty$. This asymptotic behavior for $R(q)$ is assured by assuming that the potential $V(x)$ is infinitely differentiable and along with its derivatives decreases rapidly at infinity.[19]

4. CONSERVATION LAWS FOR NONLINEAR EVOLUTION EQUATIONS

By means of the trace identities (3.30) we can derive in a unified way infinite families of conservation laws for several of the most important nonlinear evolution equations solvable by the inverse scattering transform method. To see this, let us consider the Calogero–Degasperis equations[8] for $N \times N$ matrix functions $V(x,t)$;

$$\partial_t V = 2\beta_0(\underline{L},t)V_x + \alpha_n(\underline{L})[\sigma_n, V] + \beta_n(\underline{L})\underline{G}\sigma_n, \quad (4.1)$$

where the notation conventions are as in Ref. 8. It was proved by Calogero and Degasperis that the evolution law of the scattering data of the matrix Schrödinger equation under the flows (4.1) is such that the eigenvalues remain invariant and the reflection coefficient for right incidence evolves in time according to

$$R(k,t) = \exp\left[4ik\int_0^t dt'\,\beta_0(-4k^2,t')\right] \times \exp[t(\alpha_n(-4k^2) + 2ik\beta_n(-4k^2))\sigma_n] \times R(k,0)\exp[t(-\alpha_n(-4k^2) + 2ik\beta_n(-4k^2))\sigma_n]. \quad (4.2)$$

Thus, we find that

$$R(k,t)R(-k,t) = \exp[t(\alpha_n(-4k^2) + 2ik\beta_n(-4k^2))\sigma_n] \times R(k,0)R(-k,0)\exp[-t(\alpha_n(-4k^2) + 2ik\beta_n(-4k^2))\sigma_n], \quad (4.3)$$

and then we deduce

$$\det(\mathbb{1} - R(k,t)R(-k,t)) = \det(\mathbb{1} - R(k,0)R(-k,0)). \quad (4.4)$$

Therefore, from the trace identities (3.28) it is obvious that the functionals $H_n[V]$ are conserved under the evolution equations (4.1). We emphasize that the equations (4.1), for which our result applies, contain the time-dependent term $2\beta_0(\underline{L},t)V_x$.

The class of evolution equations (4.1) includes,[20] among others, the KdV equation, the modified KdV equation, the nonlinear Schrödinger equation, and the sine-Gordon equation. In this way, from the functionals $H_n[V]$ we can get conservation laws for such equations by means of appropriate choices of the matrix function V. For example

(1) Modified Korteweg–de Vries equation:

$$q_t = q_{xxx} + 6q^2q_x, \quad V = \begin{pmatrix} -q^2 & q_x \\ -q_x & -q^2 \end{pmatrix}. \quad (4.5)$$

(2) Nonlinear Schrödinger equation:

$$i\psi_t = -\psi_{xx} - 2|\psi|^2\psi, \quad V = \begin{pmatrix} -|\psi|^2 & \psi_x \\ -\psi_x^* & -|\psi|^2 \end{pmatrix}. \quad (4.6)$$

(3) Sine-Gordon equation:

$$u_{tx} = \sin u, \quad V = \begin{pmatrix} -\frac{1}{4}u_x^2 & -\frac{1}{2}u_{xx} \\ \frac{1}{2}u_{xx} & -\frac{1}{4}u_x^2 \end{pmatrix}. \quad (4.7)$$

In order to understand the relationship among these sets of conservations laws and the ones already known for these equations, let us consider the generalized Zakharov–Shabat spectral problem

$$(\partial_x - Q + ik\sigma_3)\varphi = 0, \quad Q(x) = \begin{pmatrix} 0 & q(x) \\ r(x) & 0 \end{pmatrix}, \quad (4.8)$$

and its Jost solutions $\psi(k,x)$ and $\bar{\psi}(k,x)$, verifying the conditions

$$\begin{pmatrix} \bar{b}(k)e^{-ikx} \\ a(k)e^{ikx} \end{pmatrix} \underset{x\to-\infty}{\leftarrow} \psi(k,x) \underset{x\to+\infty}{\rightarrow} \begin{pmatrix} 0 \\ 1 \end{pmatrix}e^{ikx}, \quad (4.9)$$

$$\begin{pmatrix} \bar{a}(k)e^{-ikx} \\ -b(k)e^{ikx} \end{pmatrix} \underset{x\to-\infty}{\leftarrow} \bar{\psi}(k,x) \underset{x\to+\infty}{\rightarrow} \begin{pmatrix} 1 \\ 0 \end{pmatrix}e^{-ikx}. \quad (4.10)$$

Now, by defining $V = Q^2 + Q_x$ it follows that

$$-\partial_{xx} + V - k^2 = (-\partial_x - Q + ik\sigma_3)(\partial_x - Q + ik\sigma_3). \quad (4.11)$$

This implies that the Jost function $F_+(k,x)$ of the 2×2 matrix Schrödinger problem with potential $V = Q^2 + Q_x$ is given by

$$F_+(k,x) = (\bar{\psi}(-k,x)\ \ \psi(k,x)); \quad (4.12)$$

hence, the corresponding matrix $A(k)$ is

$$A(k) = \begin{pmatrix} \bar{a}(-k) & 0 \\ 0 & a(k) \end{pmatrix}. \quad (4.13)$$

Therefore, by (3.9), (3.19), and (4.13) it follows that

$$\frac{i}{2k}\sum_{n=1}^{\infty} H_n[Q^2 + Q_x]k^{-2n} = -\frac{1}{2k}\partial_k[\ln a(k) + \ln\bar{a}(-k)]. \quad (4.14)$$

On the other hand, it is known[21] that the conservation laws for the evolution equations solvable by the inverse scattering transform associated with (4.8) are the coefficients of two

asymptotic expansions for the logarithms of $a(k)$ and $\bar{a}(k)$ of the form

$$\ln a(k) = -\sum_{n=1}^{\infty} C_n[Q]k^{-n}, \quad \operatorname{Im} k > 0, \tag{4.15a}$$

$$\ln \bar{a}(k) = \sum_{n=1}^{\infty} C_n[Q]k^{-n}, \quad \operatorname{Im} k < 0. \tag{4.15b}$$

As a result we find

$$H_n[Q^2 + Q_x] = 2i(2n-1)C_{2n-1}[Q]. \tag{4.16}$$

That is to say, the conservation laws $H_n[Q^2 + Q_x]$ provided by the trace relations (3.28) are members of the known families of conservation laws for the evolution equations associated with the generalized Zakharov–Shabat spectral problems.

Another interesting nonlinear equation for which the functionals $H_n[V]$ determine an infinite set of conservation laws is the m-component nonlinear Schrödinger equation

$$i\partial_t \psi_\alpha = -\partial_{xx}\psi_\alpha - 2\sum_{\beta=1}^{m} |\psi_\beta|^2 \psi_\alpha, \quad \alpha = 1,\ldots,m. \tag{4.17}$$

As it has been observed by Wadati,[22] if we define the $(m+1)\times(m+1)$ matrices

$$J = \begin{pmatrix} 1 & 0\cdots 0 \\ 0 & \\ \vdots & -1_{m\times m} \\ 0 & \end{pmatrix}, \quad U(x) = \begin{pmatrix} 0 & \psi_1(x)\cdots\psi_m(x) \\ \psi_1^*(x) & \\ \vdots & 0_{m\times m} \\ \psi_m^*(x) & \end{pmatrix}, \tag{4.18}$$

then, under the evolution law associated with (4.17), the matrix Schrödinger operator

$$L = -\partial_{xx} + V(x), \quad V \equiv -U^2 + JU_x \tag{4.19}$$

evolves according to the Lax pair equation

$$\partial_t L = [B,L], \quad B \equiv 2iJ\partial_{xx} - 2iU\partial_x - iJV. \tag{4.20}$$

From (4.20) it is clear that the eigenvalues of L are constants of the motion. Furthermore, it is easy to realize that the reflection coefficient for right incidence associated with L evolves in time in the following form:

$$R(k,t) = \exp(-2ik^2tJ)R(k,0)\exp(2ik^2tJ). \tag{4.21}$$

This implies that $\det(1 - R(k,t)R(-k,t))$ is time independent. Therefore, by virtue of the trace identities (3.28), we have that the functionals $H_n[-U^2 + JU_x]$ are conservation laws for the evolution equation (4.17).

[1] L. D. Faddeev and V. E. Zakharov, Funkcional Anal. Priložen. **5**, 18 (1971).
[2] V. E. Zakharov and S. V. Manakov, Theor. Math. Phys. **19**, 332 (1974).
[3] L. D. Faddeev and L. A. Takhtajan, Theor. Math. Phys. **21**, 160 (1974).
[4] L. D. Faddeev and V. E. Korepin, Phys. Rep. **42**, 1 (1978).
[5] L. D. Faddeev, "A Hamiltonian Interpretation of the Inverse Scattering Method" in *Solitons*, edited by R. K. Bullough and P. J. Caudrey (Springer, Berlin, 1980).
[6] See the references quoted in the article by M. Fowler, "The Quantum Inverse Scattering Method and Applications to Spin Chains" in *Physics in One Dimension*, edited by J. Bernasconi and T. Schneider, *Springer Series in Solid-State Sciences*, Vol. 23 (Springer, Berlin, 1981).
[7] M. Wadati and T. Kamijo, Prog. Theor. Phys. **52**, 397 (1974).
[8] F. Calogero and A. Degasperis, Nuovo Cimento B **39**, 1 (1977).
[9] I. M. Gel'fand and L. A. Diki, Usp. Math. Nauk **30**, 67 (1975); Funkcional Anal. Priložen. **10**, 18 (1976); Funkcional Anal. Priložen. **10**, 13 (1976); Funkcional Anal. Priložen. **11**, 11 (1977); Funkcional Anal. Priložen. **12**, 8 (1978).
[10] L. D. Faddeev, Trudy Mat. Inst. Steklov **73**, 314 (1964).
[11] P. Deift and E. Trubowitz, Comm. Pure Appl. Math. **32**, 121 (1979).
[12] This condition is more suitable than the usual one, $(1+|x|)|V| \in L^1(\mathbf{R})$, in order to characterize the behavior of the scattering coefficients at $k = 0$. See Ref. 11.
[13] The bounds for F_+ and $\tilde{F}_+$ (respectively, F_- and $\tilde{F}_-$) are uniform with respect to x in the intervals (α,∞) [respectively, $(-\infty,\beta)$], where α,β are arbitrary finite numbers.
[14] O. Ragnisco, Lett. Nuovo Cimento **31**, 651 (1981).
[15] E. Olmedilla, L. Martínez Alonso, and F. Guil, Nuovo Cimento B **61**, 49 (1981).
[16] See the proof of Lemma 6 of Ref. 11. Similar arguments lead to (3.8a) and (3.8b).
[17] E. F. Beckenbach, Bull. Amer. Math. Soc. **49**, 615 (1943).
[18] See part V of Theorem 1 of Ref. 11. The statement remains true for the matrix case.
[19] See Part IV of Theorem 1 of Ref. 11.
[20] A. Degasperis, "Spectral Transform and Solvability of Nonlinear Evolution Equations" in *Nonlinear Problems in Theoretical Physics*, edited by A. F. Rañada, *Lecture Notes in Physics*, Vol. 98 (Springer, Berlin, 1979).
[21] A. C. Newell, "The Inverse Scattering Method" in *Solitons*, edited by R. K. Bullough and P. J. Caudrey (Springer, Berlin, 1980).
[22] M. Wadati, "Generalized Matrix Form of the Inverse Scattering Method" in *Solitons*, edited by R. K. Bullough and P. J. Caudrey (Springer, Berlin, 1980).

Langer's method for weakly bound states of the Helmholtz equation with symmetric profiles

George L. Johnston
Plasma Fusion Center, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

(Received 9 March 1982; accepted for publication 30 June 1982)

Use of the harmonic oscillator equation as the comparison equation in the application of Langer's method to bound states of the Helmholtz equation, $w'' + k_0^2 g(z) w(z) = 0$, with symmetric profiles $k_0^2 g(z)$, produces the WKB eigenvalue condition, which asserts the equality of the phase integral of the original equation between the turning points to $(n + 1/2)\pi$. In the case of weakly bound states, this condition gives eigenvalue estimates of low accuracy. Use of the Helmholtz equation with the symmetric Epstein profile, $G(x) = [\tilde{E} + U_0(\cosh \alpha x)^{-2}]$, as the comparison equation provides the basis for a convenient method to obtain eigenvalue estimates of substantially increased accuracy in the case of weakly bound states. In addition to the usual condition of equality of the phase integrals of the original and comparison equations between the turning points, the conditions $k_0^2 g(0) = G(0)$ and $k_0^2 g(\infty) = G(\infty)$ are imposed. An eigenvalue condition which is a simple generalization of the usual WKB eigenvalue condition is obtained. Its application to selected diverse examples of the Helmholtz equation indicates that it has a broad range of utility.

PACS numbers: 03.65.Sq, 03.65.Ge

I. INTRODUCTION

We consider the one-dimensional Helmholtz equation,

$$w'' + k_0^2 g(z) w(z) = 0, \tag{1}$$

in the interval $-\infty < z < \infty$. The profile $k_0^2 g(z)$ is characterized by the symmetric curve of Fig. 1. It is positive at $z = 0$. It decreases monotonically away from $z = 0$ and approaches a finite negative limit as $|z| \to \infty$. Thus it has two turning points. The profile $k_0^2 g(z)$ depends on a parameter, which we do not denote explicitly. The functional dependence of the profile on the parameter is arbitrary. The Schrödinger equation is thus considered here a special case of the Helmholtz equation. For a discrete set of values of the parameter, the eigenvalues, the equation has solutions, the eigenfunctions or bound states, which approach zero as $|z| \to \infty$. We consider weakly bound states. Accordingly, we have the condition $k_0^2 g(0) \gtrsim -k_0^2 g(\infty)$. We are interested in obtaining estimates of eigenvalues which are more accurate than those provided by the usual WKB eigenvalue condition. This condition, which can be obtained by Langer's method, using the Schrödinger equation for the harmonic oscillator as the comparison equation, gives eigenvalue estimates of low accuracy in the case of weakly bound states.

In Sec. II we derive Langer's transformation. In Sec. III we present the usual Langer's method treatment of the problem of bound states with two turning points, using the harmonic oscillator equation as the comparison equation. We examine in detail the reason why the resulting eigenvalue condition gives eigenvalue estimates of low accuracy in the case of weakly bound states. In Sec. IV we present a parallel development, using the symmetric Epstein equation, an example of the Helmholtz equation in which the profile is $G(x) = [\tilde{E} + U_0(\cosh \alpha x)^{-2}]$, as the comparison equation. Note that the parametric dependence of this equation is not completely determined, as it is in the case of the harmonic oscillator equation, by the usual condition of equality of the phase integrals of the original and comparison equations between the turning points. We argue that the optimal disposition of the additional parametric dependence is achieved by imposing the additional conditions $k_0^2 g(0) = G(0)$ and $k_0^2 g(\infty) = G(\infty)$. This choice leads to an eigenvalue condition which is a simple generalization of the usual WKB eigenvalue condition. It asserts the equality of the phase integral of the original equation between the turning points, not to $(n + 1/2)\pi$, as in the WKB eigenvalue condition, but to an algebraic function of $k_0^2 g(0)$ and $k_0^2 g(\infty)$. In Sec. V we examine the utility of our eigenvalue condition by applying it to selected diverse examples of the Helmholtz equation. In order to provide a rigorous test, we examine ground states. We find that our eigenvalue condition yields eigenvalue estimates of substantially increased accuracy, relative to WKB eigenvalue estimates, for the Helmholtz equation with a wide range of dependences of $k_0^2 g(z)$ on z and the eigenvalue.

II. LANGER'S TRANSFORMATION

Following Langer,[1] we first express $w(z)$ in the form

$$w(z) = u(z) v[x(z)], \tag{2}$$

where $u(z)$, $v(x)$, and $x(z)$ are functions which are to be determined; $v(x)$ is a new dependent variable and $x(z)$ is a new independent variable. Introducing (2) into (1), we obtain the differential equation

$$ux'^2 \frac{d^2 v}{dx^2} + (ux'' + 2u'x') \frac{dv}{dx} + (u'' + k_0^2 g u) v(x) = 0. \tag{3}$$

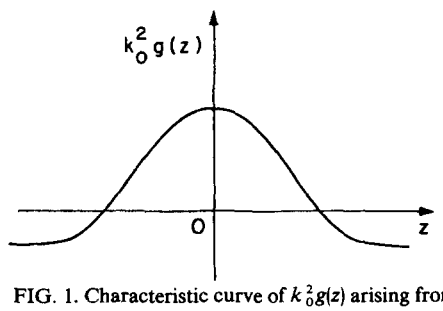


FIG. 1. Characteristic curve of $k_0^2 g(z)$ arising from symmetric profiles which correspond to weakly bound eigenstates.

 0022-2488/82/112122-05$02.50 © 1982 American Institute of Physics

Equating to zero the coefficient of dv/dx, we obtain

$$ux'' + 2u'x' = 0, \tag{4}$$

which, upon integration, yields

$$u(z) = Nx'^{-1/2}, \tag{5}$$

where N is an undetermined normalization constant. Dividing (3) by ux'^2, we obtain the equation

$$\frac{d^2v}{dx^2} + \left(\frac{k_0^2 g}{x'^2} + \frac{u''}{ux'^2}\right)v(x) = 0. \tag{6}$$

The term u''/ux'^2 in (6) presents a serious obstacle to further progress in obtaining convenient approximate solutions of (1). The lowest order approximation is obtained by neglecting this unwanted term. In the Appendix we investigate the conditions under which neglect of the unwanted term is justified.

We equate the quantity $k_0^2 g(z)/x'^2$ to $G(x)$, the profile function in the equation

$$\frac{d^2v}{dx^2} + G(x)v(x) = 0, \tag{7}$$

which is referred to as the comparison equation. It satisfies the following requirements. First, it must be similar to the original equation, (1). This is an imprecise requirement, but it is central to our program. In general, the comparison equation is required to have the same structure of turning points and singularities as the original equation. The symmetric Epstein equation clearly satisfies this requirement in the present case. The harmonic oscillator equation satisfies this requirement for strongly bound states. The divergence between the profiles of the original and comparison equations at large values of the independent variables has a negligible effect on the analysis. For weakly bound states, on the other hand, the divergence is significant. Accordingly, for such states the harmonic oscillator equation does not satisfy the requirement that it be similar to the original equation. The second requirement is that the comparison equation must be analytically solvable. The third requirement is that the phase integral of the comparison equation must be integrable in closed form. The second and third requirements are satisfied by both comparison equations.

The transformations of independent variables, which is expressed as the functional relation $x = x(z)$, is determined from the relation between $k_0^2 g(z)/x'^2$ and $G(x)$ which we have imposed. We write it as a differential equation:

$$G(x)x'^2 = k_0^2 g(z). \tag{8}$$

Extracting the square root of this equation and integrating, we obtain the functional relation $x = x(z)$ implicitly, as the equality of two definite integrals,

$$\int_{x_0}^{x} [G(\xi)]^{1/2} d\xi = \int_{z_0}^{z} [k_0^2 g(s)]^{1/2} ds, \tag{9}$$

integrating from corresponding reference points x_0 and z_0. For the problem at hand it is convenient to choose $x_0 = 0$ and $z_0 = 0$ as the reference points. The transformation of independent variables must provide that the turning points of the original and comparison equations correspond to each other. Thus we require that

$$\int_{x_-}^{x_+} [G(\xi)]^{1/2} d\xi = \int_{z_-}^{z_+} [k_0^2 g(s)]^{1/2} ds, \tag{10}$$

where $z_\pm$ and $x_\pm$ are the right- and left-hand turning points of the original and transformed equations, respectively.

III. HARMONIC OSCILLATOR COMPARISON EQUATION

The standard treatment of the problem of two turning points[2,3] is based on the use of the harmonic oscillator equation, in which

$$G(x) = \tilde{E} - U_0 x^2, \tag{11}$$

as the comparison equation. The eigenvalues of this equation are given by[4]

$$\tilde{E}_n = U_0^{1/2}(1 + 2n), \quad (n = 0,1,2,\cdots). \tag{12}$$

In this case, the condition stated in (10) becomes

$$\int_{z_-}^{z_+} [k_0^2 g(s)]^{1/2} ds = \int_{x_-}^{x_+} [U_0^{1/2}(1 + 2n) - U_0 x^2]^{1/2} dx. \tag{13}$$

The value of the integral on the right-hand side of (13) is $(n + 1/2)\pi$. Thus we obtain the eigenvalue condition

$$\int_{z_-}^{z_+} [k_0^2 g(s)]^{1/2} ds = (n + 1/2)\pi. \tag{14}$$

This is the same as WKB eigenvalue condition.[5] Note that the parameteric dependence of the comparison equation in this case is completely determined by the requirement that the phase integrals of the original and comparison equations between the turning points be equal to each other.

We can now see in detail the reason why the eigenvalue condition (14) gives eigenvalue estimates of low accuracy in the case of weakly bound states. The phase integrals of $k_0^2 g(z)$ and $G(x)$ contain substantial contributions from ranges of the independent variables in which the behavior of the two profiles deviate significantly from each other. In the case of strongly bound states, on the other hand, ranges of the independent variables in which the behavior of $k_0^2 g(z)$ and $G(x)$ deviate significantly from each other lie outside the range of integration.

IV. SYMMETRIC EPSTEIN COMPARISON EQUATION

A class of examples of the Helmholtz equation which are analytically solvable in terms of solutions of the hypergeometric equation was applied to ionospheric radio waves by Epstein[6] and to quantum mechanics by Eckart.[7] The symmetric Epstein equation has the profile

$$G(x) = [\tilde{E} + U_0(\cosh \alpha x)^{-2}]. \tag{15}$$

For weakly bound states, this equation is similar to the original equation. The harmonic oscillator equation is not. The eigenvalue condition for the symmetric Epstein equation is[8]

$$\tilde{E} = -\tfrac{1}{4}\alpha^2[-(1 + 2n) + (1 + 4U_0\alpha^{-2})^{1/2}]^2, \tag{16}$$

where n takes nonnegative integral values starting from zero. There is a finite number of levels, determined by the condition

$$2n < (1 + 4U_0^{-2})^{1/2} - 1. \tag{17}$$

The phase integral between the turning points of the

comparison equation is

$$\int_{x_-}^{x_+} [\tilde{E} + U_0(\cosh \alpha x)^{-2}]^{1/2} dx = \pi\alpha^{-1}[U_0^{1/2} - (-\tilde{E})^{1/2}]. \quad (18)$$

In contrast to the parametric dependence of the harmonic oscillator equation, that of the symmetric Epstein equation is not completely determined by the condition of equality of the phase integrals of the original and comparison equations between the turning points. Thus we must impose additional conditions in order to obtain an eigenvalue condition. The phase integral condition is unique. The choice of additional conditions is more arbitrary. In making it, we are guided by considerations of simplicity and plausibility. We choose to impose the conditions that U_0 and $\tilde{E}$ be equal to the corresponding quantities of the original equation, i.e.,

$$U_0 = k_0^2 g(0) - k_0^2 g(\infty) \equiv \Delta(k_0^2 g), \quad (19)$$

and

$$\tilde{E} = G(\infty) = k_0^2 g(\infty). \quad (20)$$

We eliminate the remaining quantity in (18), namely α, by means of (16), making use of (19) and (20) to express it in terms of $\Delta(k_0^2 g)$ and $k_0^2 g(\infty)$. Note that there is no simple and plausible way to choose α. Thus the conditions that we have imposed in order to obtain an eigenvalue condition are the appropriate ones. The following expressions for α as a function of U_0 and $\tilde{E}$ are obtained from (16) for $n = 0$ and $n \neq 0$, respectively:

$$\alpha_0 = (-\tilde{E}_0)^{-1/2}[U_0 - (\tilde{E}_0)], \quad (21)$$

$$\alpha_{n\neq 0} = \frac{2(1+2n)(-\tilde{E}_n)^{1/2}}{[(1+2n)^2 - 1]} \times \left[\left\{1 + \frac{[U_0 - (-\tilde{E}_n)][(1+2n)^2 - 1]}{(1+2n)^2(-\tilde{E}_n)}\right\}^{1/2} - 1\right]. \quad (22)$$

The eigenvalue condition is thus given by the relation

$$\int_{z_-}^{z_+} [k_0^2 g(s)]^{1/2} ds = \pi\{\alpha_n[\Delta(k_0^2 g), k_0^2 g(\infty)]\}^{-1} \times \{[\Delta(k_0^2 g)]^{1/2} - [-k_0^2 g(\infty)]^{1/2}\}. \quad (23)$$

This condition is a simple generalization of the WKB eigenvalue condition, (14). In the case of the ground state, $n = 0$, the eigenvalue condition has a particularly simple form, namely

$$\int_{z_-}^{z_+} [k_0^2 g(s)]^{1/2} ds = \pi\left\{\frac{[-k_0^2 g(\infty)]^{1/2}}{[\Delta(k_0^2 g)]^{1/2} + [-k_0^2 g(\infty)]^{1/2}}\right\}. \quad (24)$$

In the case of strongly bound ground states, $\{[-k_0^2 g(\infty)]/\Delta(k_0^2 g)\} \to 1$ and the quantity in curly brackets in (24) approaches 1/2, in agreement with the WKB eigenvalue condition. In the case of strongly bound higher states, the right-hand side of the eigenvalue condition (23) approaches $(n + 1/2)\pi$, in agreement with the WKB eigenvalue condition.

V. NUMERICAL EXAMPLES

In order to examine the utility of the eigenvalue condition which we have obtained, we apply it to selected diverse examples of the Helmholtz equation. In order to provide a rigorous test, we examine ground states. Since the scale lengths of higher states are smaller relative to that of the profile, we expect that the eigenvalue estimates for them will be more accurate. The results which we obtain indicate that our eigenvalue condition yields eigenvalue estimates which are of substantially increased accuracy, relative to WKB eigenvalue estimates, for the Helmholtz equation with a wide range of dependences of $k_0^2 g(z)$ on z and the eigenvalue.

The first equation which we consider is the Schrödinger equation with a Gaussian potential:

$$k_0^2 g(z) = E + V_0 \exp(-z^2). \quad (25)$$

The second equation is the Schrödinger equation with a potential whose magnitude is the square of a Lorentzian:

$$k_0^2 g(z) = E + V_0(1 + z^2)^{-2}. \quad (26)$$

These equations provide an indication of the applicability of our eigenvalue condition to equations in which the asymptotic behavior of $k_0^2 g(z)$ as $|z| \to \infty$ is different from that of $G(x)$ as $|x| \to \infty$. Since the asymptotic behavior of the difference $[G(x) - G(\infty)]$ is exponential, namely, as $x \to \infty$

$$[G(x) - \tilde{E}] \sim 4U_0 \exp(-2\alpha x), \quad (27)$$

the equations considered provide an appropriate indication. In the case of the Gaussian potential, the asymptotic behavior is

$$[k_0^2 g(z) - E] \sim V_0 \exp(-z^2). \quad (28)$$

In the case of the squared Lorentzian potential, the asymptotic behavior is algebraic:

$$[k_0^2 g(z) - E] \sim V_0 z^{-4}. \quad (29)$$

Note that parameters multiplying the independent variables in the potentials of these equations can be scaled away in the eigenvalue condition, so that the particular equations considered do not embody restrictions on their generality with respect to such parameters.

The other equations which we consider are examples of the Helmholtz equation in which

$$k_0^2 g(z) = E + V_0 \exp(s\sqrt{-E} - z^2), \quad (s = \pm 1), \quad (30a,b)$$

$$k_0^2 g(z) = E + V_0 \exp[-(1 - s\sqrt{-E})z^2], \quad (s = \pm 1), \quad (31a,b)$$

Note that the dependence of $k_0^2 g(z)$ on z of (30) and (31) is the same as that of (25). Considered together, these four equations provide an indication of the applicability of our eigenvalue condition to equations with a wide range of dependences of $k_0^2 g(z)$ on the eigenvalue. We consider $E = k_0^2 g(\infty)$ to be the effective eigenvalue. In (30), $[k_0^2 g(0) - k_0^2 g(\infty)]$ increases (decreases) with increasing $(-E)$ for $s = 1$ $(s = -1)$. In (31), the width of $[k_0^2 g(z) - k_0^2 g(\infty)]$ increases (decreases) with increasing $(-E)$ for $s = 1$ $(s = -1)$. In (30) and (31), we choose $[k_0^2 g(z) - k_0^2 g(\infty)]$ to be functions of $\sqrt{-E}$, instead of $-E$. Thereby we produce stronger varia-

tions of $[k_0^2 g(z) - k_0^2 g(\infty)]$ with respect to $-E$ for small values of $-E$, which will occur in the numerical examples, and more severe tests of our eigenvalue condition.

In order to provide a simple and uniform basis for evaluating the accuracy of the estimates provided by our eigenvalue condition, in each case we choose the value of V_0 for which the ground state eigenvalue estimate given by the WKB eigenvalue condition is zero. For the examples of $k_0^2 g(z)$ chosen, there is in each case a bound ground state corresponding to this value of V_0. Thus the relative error produced by the WKB eigenvalue condition in each case is equal to -1. The choice of a squared Lorentzian instead of a Lorentzian has been made in (26) in order to ensure the existence of the phase integral for $E = 0$.

The numerical integration of the phase integrals presents a particular difficulty. In the neighborhood of a turning point at $z = z_t$, the dominant behavior of the integrand is proportional to $(z - z_t)^{1/2}$. The first derivative of the integrand increases in magnitude without limit as $z \to z_t$. A convenient technique for the numerical evaluation of the phase integrals involves the introduction of a simple transformation of the variable of integration which removes the singularity in the derivative of the integrand. For the class of integrands considered here on the interval $0 \leqslant z \leqslant z_+$, a suitable transformation is

$$z(t) = z_+[\tfrac{1}{2}(3t - t^3)], \quad (0 \leqslant t \leqslant 1). \tag{32}$$

With the introduction of this transformation, the phase integrals between the turning points can be expressed in the form

$$I = 3z_+ \int_0^1 \{k_0^2 g[z(t)]\}^{1/2}(1 - t^2)dt. \tag{33}$$

In the neighborhood of $t = 1$, the dominant behavior of $(k_0^2 g)^{1/2}$ is given by

$$\{k_0^2 g[z(t)]\}^{1/2} \simeq [-\tfrac{3}{2} z_+ k_0^2 g'(z_+)]^{1/2}(1 - t). \tag{34}$$

Use of the transformation (32) results in a dramatic improvement in the accuracy of the numerical integration of the phase integrals.

The numerical integration of the differential equations to determine the exact eigenvalues is performed using a fourth-order Runge–Kutta approximation.

The numerical results are shown in Table I. For each differential equation, the value of V_0, the exact value of $-E$, $-E_e$, the approximate value of $-E$ obtained from our eigenvalue condition, $-E_a$, and the relative error of $-E$, $[\Delta(-E)/(-E)] = [(-E_a)/(-E_e) - 1]$, are presented in succeeding columns. Recall that the relative error produced by the WKB eigenvalue condition is -1 because the values of V_0 have been selected to give zero WKB ground state eigenvalue.

The relative error produced by our eigenvalue condition is very small in each case. The sign of the relative error is the same for all equations except the equation with the squared Lorentzian potential. Presumably, this behavior is a consequence of the fact that in all the other equations $[k_0^2 g(z) - k_0^2 g(\infty)]$ is a Gaussian in z. Considered together, the six equations suggest that our eigenvalue condition yields eigenvalue estimates of substantially increased accuracy, relative to WKB eigenvalue estimates, for the Helmholtz equation with a wide range of dependences of $k_0^2 g(z)$ on z and the eigenvalue.

ACKNOWLEDGMENTS

We are happy to acknowledge helpful discussions with A. Baños, Jr. This work was supported in part by the U.S. Office of Naval Research.

APPENDIX: NEGLIGIBILITY OF UNWANTED TERM

In Sec. II we neglected the unwanted term in (6), u''/ux'^2, in order to permit the development of the lowest order approximation to the solution of (1). Here we examine the unwanted term in detail and attempt to understand the conditions under which its neglect can be justified.

The explicit form of $u(z)$ is

$$u(z) = N\left\{\frac{G[x(z)]}{k_0^2 g(z)}\right\}^{1/4}. \tag{A1}$$

An explicit statement of the condition for the negligibility of the unwanted term is

$$\left|\frac{u''}{uk_0^2 g(z)}\right| \ll 1. \tag{A2}$$

The condition which is typically invoked to justify the satisfaction of the inequality (A2) is that the ratio of the scale length on which the solution of (1) varies to the scale length on which the profile varies is small compared to unity or is

TABLE I. Numerical results. Entries in (a) designate equations in the text. Corresponding to each equation, (b) gives the value of V_0, (c) the exact value of $-E$, (d) the approximate value of $-E$ given by our eigenvalue condition, and (e) the relative error, $[(-E_a)/(-E_e) - 1]$. For each equation, the value of V_0 is chosen so that the WKB eigenvalue condition gives a ground state at $E = 0$. Thus, the relative error produced by the WKB eigenvalue condition in each case is equal to -1.

(a) Equation	(b) V_0	(c) $-E_e$	(d) $-E_a$	(e) $[\Delta(-E)/(-E)]$
(25)	$(1/8)\pi$	0.080 100 3	0.081 536 4	0.017 928 0
(26)	1/4	0.028 924 1	0.028 559 0	−0.012 625 2
(30a)	$(1/8)\pi$	0.152 063	0.157 108	0.033 172 1
(30b)	$(1/8)\pi$	0.053 936 6	0.054 520 1	0.010 816 1
(31a)	$(1/8)\pi$	0.068 058 4	0.068 881 8	0.012 098 3
(31b)	$(1/8)\pi$	0.102 742	0.105 456	0.026 420 1

less than or approximately equal to unity, depending on the situation considered. We shall find that the ratio can be permitted to exceed unity in some cases.

The role played by this ratio in determining the negligibility of the unwanted term is particularly clear in the case of the WKB approximation, which can be obtained from Langer's transformation by using a comparison equation in which $G(x)$ is a constant. In that case $u(x)$ is the WKB swelling factor and

$$\frac{u''}{u[k_0^2 g(z)]} = \frac{5}{16}\frac{[k_0^2 g'(z)]^2}{[k_0^2 g(z)]^3} - \frac{1}{4}\frac{[k_0^2 g''(z)]}{[k_0^2 g(z)]^2}. \quad \text{(A3)}$$

If the scale length of $g(z)$ is κ_0^{-1}, so that $|g'(z)| \simeq \kappa_0 |g(z)|$ and $|g''(z)| \simeq \kappa_0^2 |g(z)|$, the condition (A2) implies that $\kappa_0^2 \ll k_0^2$. If $g(z)$ is of order unity, k_0^{-1} is the scale length of the solution of (1).

If $G(x)$ is not a constant, the situation is considerably more complicated and explicit results can be obtained only in particular circumstances. For example, in a study of the computation of the reflection and transmission coefficients for a family of symmetric profiles, using a comparison equation in which

$$G(x) = x^2 + \eta, \quad \text{(A4)}$$

where η is a constant, Baños[3] includes a detailed consideration of the unwanted term associated with an original equation which is a symmetric Epstein equation with profile

$$k_0^2 g(z) = k_0^2 \tanh^2(z/2\lambda), \quad \text{(A5)}$$

in which case $\eta = 0$. He develops a two-term expansion in powers of x^2 for the unwanted term:

$$u''/ux'^2 = \alpha - \beta x^2 + \cdots, \quad \text{(A6)}$$

with coefficients

$$\alpha = (1/16\pi)(\lambda_0/\lambda), \quad \beta = (3/128\pi^2)(\lambda_0/\lambda)^2, \quad \text{(A7)}$$

where $\lambda_0 = 2\pi/k_0$. Examining the coefficients of the expansion, he concludes that a necessary and sufficient criterion for the applicability of Langer's method in this case is $\lambda/\lambda_0 \gg 1$.

It is impossible to perform a similar analysis in the case considered here. We are dealing with an eigenvalue problem and we use a more complicated comparison equation. Furthermore, the condition $k_0^2 g(0) = 0$, which makes an explicit calculation possible in that case, is a very special one which is not available to us. We can, however, develop a fairly specific intuitive understanding of the situation by considering (A1). Note that u is proportional to the one-fourth power of the ratio of profiles and hence that u varies weakly with deviations of the ratio from its value of unity at $z = 0$ and in the limit as $|z| \to \infty$. For states which are sufficient weakly bound, the scale length of the state may exceed that of the profile. If the profile of the comparison equation is sufficiently similar to that of the original equation, the eigenvalue condition may nevertheless give an eigenvalue estimate of high accuracy. Unfortunately, it is a practical impossibility to ascertain this on an *a priori* basis.

[1] R. E. Langer, Phys. Rev. **51**, 669 (1937). For a very readable account of Langer's method, see M. V. Berry and K. E. Mount, Rep. Prog. Phys. **35**, 315 (1972).

[2] S. C. Miller, Jr., and R. H. Good, Jr., Phys. Rev. **91**, 174 (1953).

[3] A. Baños, Jr., J. Math. Phys. **14**, 963 (1973).

[4] L. D. Landau and E. M. Lifshitz, *Quantum Mechanics* (Addison-Wesley, Reading, MA, 1958), Sec. 21, p. 64.

[5] E. Merzbacher, *Quantum Mechanics* (Wiley, New York, 1970), 2nd ed., p. 122.

[6] P. S. Epstein, Proc. Natl. Acad. Sci. (USA) **16**, 627 (1930). For an excellent account of the three-parameter family of profiles first introduced by Epstein, see R. A. Phinney, Rev. Geophys. Space Sci. **8**, 517 (1970).

[7] C. Eckart, Phys. Rev. **35**, 1303 (1930).

[8] L. D. Landau and E. M. Lifshitz, *Quantum Mechanics* (Addison-Wesley, Reading, MA, 1958), Sec. 21, p. 69.

Formal solutions of inverse scattering problems. IV. Error estimates

Reese T. Prosser
Department of Mathematics, Dartmouth College, Hanover, New Hampshire 03755

(Received 17 February 1982; accepted for publication 30 April 1982)

The formal solutions of certain three-dimensional inverse scattering problems presented in papers I–III in this series [J. Math. Phys. **10**, 1819 (1969); **17**, 1175 (1976); **21**, 2648 (1980)] are employed here to obtain quantitative estimates on the error resulting from the use of the Born approximations in both direct and inverse potential scattering problems. These estimates are uniformly valid at all energies, and for all sufficiently weak potentials.

PACS numbers: 03.80. + r

1. INTRODUCTION

It is known that the first Born approximation for the direct problem of potential scattering also provides, when suitably interpreted, a first approximation for the associated inverse problem. We consider here the question of just how good an approximation this is for both the direct and the inverse problems of potential scattering in three dimensions. We give quantitative estimates of the errors involved in this approximation in terms of familiar norms on the data. These estimates are derived from results of our previous work.[1]

2. THE DIRECT PROBLEM

The scattering of a quantum mechanical wave function $\varphi(\mathbf{x}, \mathbf{k})$ from a fixed potential $V(\mathbf{x})$ in three dimensions is governed by the time-independent Schödinger equation

$$(\nabla^2 + k^2)\varphi(\mathbf{x}, \mathbf{k}) = V(\mathbf{x})\varphi(\mathbf{x}, \mathbf{k}). \tag{1}$$

The solution, which is to consist of an ingoing plane wave plus an outgoing scattered wave, may be expressed as

$$\varphi(\mathbf{x}, \mathbf{k}) = e^{i\mathbf{k}\cdot\mathbf{x}} - \int \frac{e^{i|\mathbf{k}|\,|\mathbf{x}-\mathbf{y}|}}{4\pi|\mathbf{x}-\mathbf{y}|} V(\mathbf{y})\varphi(\mathbf{y}, \mathbf{k})\, d\mathbf{y}. \tag{2}$$

As $|\mathbf{x}| \to \infty$ the behavior of $\varphi(\mathbf{x}, \mathbf{k})$ is given by

$$\varphi(\mathbf{x}, \mathbf{k}) \to e^{i\mathbf{k}\cdot\mathbf{x}} - \frac{e^{i|\mathbf{k}|\,|\mathbf{x}|}}{4\pi|\mathbf{x}|} T(\mathbf{k}', \mathbf{k}) + O\left(\frac{1}{|\mathbf{x}|^2}\right). \tag{3}$$

Here $\mathbf{k}' = (|\mathbf{k}|/|\mathbf{x}|)\mathbf{x}$, and $T(\mathbf{k}', \mathbf{k})$ is given by

$$T(\mathbf{k}', \mathbf{k}) = \int e^{-i\mathbf{k}\cdot\mathbf{y}} V(\mathbf{y})\varphi(\mathbf{y}, \mathbf{k})\, d\mathbf{y}. \tag{4}$$

In this way, the "on-shell" T-matrix $T(\mathbf{k}', \mathbf{k})$ contains the scattering data.

Substituting (2) into (4) and rearranging, we find

$$T(\mathbf{k}'\mathbf{k}) = V(\mathbf{k}' - \mathbf{k}) - \int V(\mathbf{k}' - \mathbf{k}'') \times \frac{1}{\mathbf{k}''^2 - \mathbf{k}^2 + i0} T(\mathbf{k}'', \mathbf{k})\, d\mathbf{k}'' \tag{5}$$

or, more formally,

$$T = V - V\Gamma T, \tag{6}$$

where $\Gamma T(\mathbf{k}', \mathbf{k})$ is the matrix $(\mathbf{k}'^2 - \mathbf{k}^2 + i0)^{-1} T(\mathbf{k}', \mathbf{k})$. Equation (6) may be solved by iteration, yielding the Born series for T:

$$T = V - V(\Gamma V) + V[\Gamma V(\Gamma V)] - \cdots. \tag{7}$$

Following Friedrichs,[2] we define an appropriate class of integral kernels and a suitable Friedrichs norm $\| \ \|_F$ for this class, such that

$$\|K\Gamma M\|_F \leqslant \|K\|_F \, \|M\|_F \tag{8}$$

and such that

$$\|V\|_F \leqslant a < 1. \tag{9}$$

Then the Born series for T converges in this norm,[3] and we have

$$\|T\|_F \leqslant a(1-a)^{-1}. \tag{10}$$

If we denote by T_n the sum of the first n terms of the Born series, i.e., T_n is the nth Born approximation, then

$$\|T - T_n\|_F \leqslant \sum_{k=n+1}^{\infty} \|V\|^k \leqslant \sum_{k=n+1}^{\infty} a^k = \frac{a^{n+1}}{1-a}. \tag{11}$$

In particular, for the first Born approximation T_1, we have

$$\|T - T_1\|_F \leqslant a^2/(1-a), \tag{12}$$

and each succeeding approximation is better by a factor of a.

Of course it is well known that the Born series converges geometrically. The trouble here is that estimates for the "on-shell" T-matrix in three dimensions require the use of a Friedrichs norm satisfying (8). In order to draw useful conclusions, we must relate this norm to norms of a more familiar form.

In Appendix A we show that for a particular choice of Friedrichs norm and any convolution kernel $F(\mathbf{k}' - \mathbf{k})$ we always have

$$10\,\|F\|_0 \leqslant \|F\|_F \leqslant 1000\,\|F\|_1, \tag{13}$$

where

$$\|F\|_1 = \sup_{\mathbf{k},j} \{(1 + |\mathbf{k}|)|D^j F(\mathbf{k})|\}, \tag{14}$$

with $D^j F$ any derivative of F of order j, $0 \leqslant j \leqslant 2$, and

$$\|F\|_0 = \sup_{\mathbf{k}} |F(\mathbf{k})|. \tag{15}$$

It follows that if

$$\|V\|_1 = \sup_{\mathbf{k}} \{1 + |\mathbf{k}|)|D^j V(\mathbf{k})|\} \leqslant (0.001)a < 0.001, \tag{16}$$

 0022-2488/82/112127-04$02.50 © 1982 American Institute of Physics

and if $B(\mathbf{k}'-\mathbf{k}) = T((\mathbf{k}'-\mathbf{k})/2, (\mathbf{k}-\mathbf{k}')/2)$ is the backscatter matrix, then $\|B\|_F \leqslant \|T\|_F$, so

$$\|B - B_n\|_0 \leqslant \sup_{\mathbf{k}} \{|B(\mathbf{k}) - B_n(\mathbf{k})|\} \leqslant (0.1)\frac{a^{n+1}}{(1-a)}. \quad (17)$$

We note that this estimate is *uniform* in $\mathbf{k}$.

3. THE INVERSE PROBLEM

We summarize here the results of Ref. 1. We may rewrite (6) as

$$T = V - V\Gamma T = VU, \quad (18)$$

where U is the ingoing Möller wave matrix,

$$U = I - \Gamma T. \quad (19)$$

This operator is known to be unitary when the Born series converges, i.e., when (9) holds.[2] Hence $UU^* = I$, where

$$U^* = I + \Gamma T^*, \quad (T^*(\mathbf{k}', \mathbf{k}) = \overline{T(k, k')}). \quad (20)$$

Applying (20) to (18), we find

$$T + T\Gamma T^* = V. \quad (21)$$

We now define the projection operator Θ, which takes kernels $K(\mathbf{k}'\mathbf{k})$ in our Friedrichs class into convolution kernels, by

$$\Theta K(\mathbf{k}'-\mathbf{k}) = K\left(\frac{\mathbf{k}'-\mathbf{k}}{2}, \frac{\mathbf{k}-\mathbf{k}'}{2}\right). \quad (22)$$

In particular,

$$\Theta V(\mathbf{k}'-\mathbf{k}) = V(\mathbf{k}'-\mathbf{k}) \quad (23)$$

and

$$\Theta T(\mathbf{k}', \mathbf{k}) = B(\mathbf{k}'-\mathbf{k}), \quad (24)$$

where $B(-\mathbf{k}) = T(-\mathbf{k}, \mathbf{k})$ contains the backscattering data. Inserting (23) and (24) into (21), we find

$$V = B + \Theta(T\Gamma T^*). \quad (25)$$

Inserting (25) in turn into (21) and solving for T, we obtain

$$T = B + (I - \Theta)(T\Gamma T^*). \quad (26)$$

We show in Ref. 1 that Eq. (26) *characterizes* the scattering kernels obtained from local potentials, and that this equation can be solved by iteration for T in terms of B, provided

$$\|B\|_F \leqslant b < \tfrac{1}{8}, \quad (27)$$

in which case we have

$$\|T\|_F \leqslant 2b < \tfrac{1}{4}. \quad (28)$$

If we now take for our first approximation

$$T_1(\mathbf{k}', \mathbf{k}) = 0 \quad (29)$$

and obtain our higher order approximations by interation of (27),

$$T_{n+1}(\mathbf{k}', \mathbf{k}) = B(\mathbf{k}'-\mathbf{k}) + (I - \Theta)(T_n\Gamma T_n^*)(\mathbf{k}', \mathbf{k}), \quad (30)$$

then we show in Ref. 1 that

$$\|T - T_n\|_F \leqslant \tfrac{1}{2}(8b)^n(1-8b)^{-1} \quad (31)$$

so that the convergence of these approximations T_n to T is also geometric. Finally, our nth approximation to V is obtained from (25):

$$V_n = B + \Theta T_n\Gamma T_n^*. \quad (32)$$

How good are these approximations? We show in Ref. 1 that if (27) holds, then, for all n,

$$\|T_n\|_F \leqslant 2b < \tfrac{1}{4}, \quad (33)$$

and so

$$\|V_n\|_F \leqslant b + 4b^2 < \tfrac{3}{16}. \quad (34)$$

Moreover, it follows from (31) and (32) that

$$\|V - V_n\|_F < \tfrac{1}{4}(8b)^{n+1}(1-8b)^{-1}. \quad (35)$$

In particular, $V_1(\mathbf{k}'-\mathbf{k}) = B(\mathbf{k}'-\mathbf{k})$ is the first Born approximation for the inverse problem, and (34) says

$$\|V - V_1\|_F \leqslant 16b^2(1-8b)^{-1}. \quad (36)$$

Each succeeding approximation is better by a factor of $8b$. Thus V_2 is not much of an improvement over V_1 unless $8b < \frac{1}{2}$, in which case V_2 is better at least by a factor of 2.

In terms of more familiar norms, if we assume

$$\|B\|_1 = \sup_{\mathbf{k},j} \{1 + |\mathbf{k}|)|D^jK(\mathbf{k})|\} \leqslant (0.001)b, \quad (37)$$

then we conclude from (13) that (35) and (36) hold with $\|\ \|_F$ replaced by $10\|\ \|_0$. Thus,

$$\|V - V_n\|_0 = \sup_{\mathbf{k}} \{|V(\mathbf{k}) - V_n(\mathbf{k})|\}$$
$$\leqslant (0.025)(8b)^{n+1}(1-8b)^{-1} \quad (38)$$

which gives a uniform estimate for the error in the nth approximation to $V(\mathbf{k})$ which is uniform in $\mathbf{k}$.

It is unfortunate that the conditions (16) and (37) are so stringent as to preclude any but a theoretical interest. Presumably better conditions can be given with more *a priori* knowledge; e.g., with knowledge that the potential is radial. In retrospect, it is perhaps not surprising that such stringent conditions are required to include so broad a class of three-dimensional potentials. Nevertheless, it should be emphasized again that the results described here hold for all sufficiently weak potentials, and for *all* sufficiently weak scattering data.

ACKNOWLEDGMENT

I am indebted to Harry Moses for suggesting this problem.

APPENDIX A

In order to derive Eq. (13) we introduce here a particular Friedrichs norm suggested by Schwartz.[4]

Let $A(s, t)$ be a (possibly vector valued) function of two real variables (s, t). We define

$$MA(s, t) = (1 + |s|)^{1/4}|A(s, t)|,$$
$$M'A(s, t) = (1 + |t|)^{1/4}|A(s, t)|, \quad (A1)$$

$$NA(s, t) = \sup_{|h|<1} (2h)^{-1/2}|A(s+h, t) - A(s-h, t)|,$$

$$N'A(s, t) = \sup_{|h|<1} (2h)^{-1/2}|A(s, t+h) - A(s, t-h)|,$$

and

$$|A| = \sup_{s,t} |A(s, t)|. \quad (A2)$$

Then we set

$$\|A\|_S = \max\{|MM'A|, |MN'A|, |M'NA|, |NN'A|\}. \quad \text{(A3)}$$

Lemma 1: Suppose $\|A\|_S < \infty$ and set

$$\Gamma A = \frac{A(s,t)}{s-t+i0}. \quad \text{(A4)}$$

Then ΓA is a bounded integral operator on $L^2(\mathbb{R})$, and

$$\|\Gamma A f\|_2 \leqslant C\|A\|_S \|f\|_2. \quad \text{(A5)}$$

Lemma 2: Suppose $\|A\|_S < \infty$, $\|B\|_S < \infty$, and set

$$B\Gamma A(s,t) = \int_{-\infty}^{+\infty} \frac{B(s,\sigma)A(\sigma,t)\,d\sigma}{\sigma - t + i0}. \quad \text{(A6)}$$

Then

$$\|B\Gamma A\|_S \leqslant (10+2\pi)\|B\|_S \|A\|_S. \quad \text{(A7)}$$

Proof: Lemmas 1 and 2 are both proved by Schwartz (Ref. 4, p. 626) with undetermined constants in the estimates (A5) and (A7). We consider here only Lemma 2 in order to show that with our normalizations in (A3), we can set the const = 17 in (A7).

Following Schwartz, we set

$$f(s,\sigma,t) = B(s,\sigma)A(\sigma,t) \quad \text{(A8)}$$

and

$$g(s,t) = P\int_{-\infty}^{+\infty} \frac{f(s,\sigma,t)\,d\sigma}{\sigma - t}, \quad \text{(A9)}$$

where P denotes the Cauchy principal part. Then, since $P\int_{-\infty}^{+\infty}[1/(\sigma-t)]\,d\sigma = 0$, we have

$$\begin{aligned}
|g(s,t)| &\leqslant \int_{|\sigma-t|>1} \frac{|f(s,\sigma,t)-f(s,t,t)|}{|\sigma-t|}\,d\sigma \\
&\quad + \int_{|\sigma-t|>1} \frac{|\sigma-t|^{-1/2}}{|\sigma-t|}\,d\sigma \\
&\leqslant \|A\|\,\|B\| \int_{|\sigma-t|>1} \frac{|\sigma-t|^{-1/2}}{|\sigma-t|}\,d\sigma \\
&\quad + 2\|A\|\,\|B\| \int_{|\sigma-t|<1} \left|\frac{\sigma-t}{\sigma-t}\right|^{1/2} d\sigma \\
&\leqslant \|A\|_S \|B\|_S \{4+8\} = 12\|A\|_S\|B\|_S. \quad \text{(A10)}
\end{aligned}$$

Moreover, since $P\int_{-\infty}^{+\infty}[1/(\sigma^2-h^2)]\,d\sigma = 0$, we also have

$$\begin{aligned}
&|g(s,t+h)-g(s,t-h)| \\
&\leqslant 2h\left|P\int_{-\infty}^{+\infty} \frac{f(s,\sigma,t)}{(\sigma-t)^2-h^2}\,d\sigma\right| \\
&= 2h\left|P\int_{-\infty}^{+\infty} \frac{f(s,\sigma+t,t)}{\sigma^2-h^2}\,d\sigma\right| \\
&\leqslant 2h\int_0^{\infty} \frac{|f(s,\sigma+t,t)-f(s,h+t,t)|}{|\sigma^2-h^2|}\,d\sigma \\
&\quad + 2h\int_0^{\infty} \frac{|f,t-\sigma,t)-f(s,t-h,t)|}{|\sigma^2-h^2|}\,d\sigma \\
&\leqslant 4h\|A\|\,\|B\| \int_0^{\infty} \frac{|\sigma-h|^{1/2}}{|\sigma^2-h^2|}\,d\sigma \\
&= (2h)^{1/2}\|A\|\,\|B\|\,2^{3/2}\int_0^{\infty} \frac{|\sigma-1|^{1/2}}{|\sigma^2-1|}\,d\sigma \\
&= (2h)^{1/2}\|A\|\,\|B\|\,6. \quad \text{(A11)}
\end{aligned}$$

Now, since (cf. Ref. 4 p. 615)

$$\int_{-\infty}^{+\infty} \frac{f(s,\sigma,t)}{\sigma-t+i0}\,d\sigma = P\int_{-\infty}^{+\infty} \frac{f(s,\sigma,t)}{\sigma-t}\,d\sigma - i\pi f(s,t,t), \quad \text{(A12)}$$

we conclude that $\|B\Gamma A\|_S \leqslant (10+2\pi)\|A\|_S\|B\|_S$, as required.

Now suppose $F(\mathbf{k})$ is a function of the 3-vector $\mathbf{k}\in\mathbb{R}^3$, and set

$$\begin{aligned}
\|F\|_0 &= \sup_{\mathbf{k}} |F(\mathbf{k})|, \\
\|F\|_1 &= \sup_{\mathbf{k},j} \{(1+|\mathbf{k}|)|D^jF(\mathbf{k})|\}, \quad \text{(A13)}
\end{aligned}$$

where D_j is any derivative of F of order j, $0\leqslant j\leqslant 2$. If we regard $F(\mathbf{k}'-\mathbf{k})$ as a convolution kernel, then we may define the Schwartz norm $\|F\|_S$ for F as follows:

Set $r = |\mathbf{k}|$, $r' = |\mathbf{k}'|$, $\mathbf{k} = r\boldsymbol{\omega}$, $\mathbf{k}' = r'\boldsymbol{\omega}'$, and

$$\begin{aligned}
|G(r,r')| = \max\Big\{&\sup_{\omega} \frac{1}{4\pi}\int_{\Omega} |F(r\boldsymbol{\omega}-r'\boldsymbol{\omega}')|\,d\boldsymbol{\omega}', \\
&\sup_{\omega'} \frac{1}{4\pi}\int_{\Omega} |F(r\boldsymbol{\omega}-r'\boldsymbol{\omega}')|\,d\boldsymbol{\omega}\Big\}. \quad \text{(A14)}
\end{aligned}$$

Then $|G(r,r')|$ is the Holmgren norm of the operator $G(r,r')$ acting on $L_2(\Omega)$, where Ω is the unit sphere in $\mathbb{R}^3$, and so $|G(r,r')|$ bounds the operator norm. Now if we set $s = r^2$, $t = r'^2$, and

$$A(s,t) = \begin{cases} (st)^{1/4}F(s^{1/2}\boldsymbol{\omega} - t^{1/2}\boldsymbol{\omega}') & s,t>0, \\ 0 & \text{otherwise}, \end{cases} \quad \text{(A15)}$$

then we can define $\|A\|_S$ by (A3), and define $\|F\|_S$ by

$$\|F\|_S = \|A\|_S. \quad \text{(A16)}$$

Lemma 3: Suppose $\|F\|_1 < \infty$. Then we have

$$\|F\|_0 \leqslant \|F\|_S \leqslant 16\pi\|F\|_1. \quad \text{(A17)}$$

Proof: This Lemma is also proved by Schwartz (Ref. 4, p. 633), again with an undetermined constant in the estimate (A17). We show here that we can take const = 16π.

We start by observing that if $r<r'$, then $|r\boldsymbol{\omega}-r'\boldsymbol{\omega}'| \geqslant 1/2r'|\boldsymbol{\omega}-\boldsymbol{\omega}'|$ which is easily seen by drawing the appropriate picture. Hence, if $|F(\mathbf{k})|\leqslant C(1+|\mathbf{k}|)^{-1}$, then

$$|F(r\boldsymbol{\omega}-r'\boldsymbol{\omega}')| \leqslant C(1+\mu|\boldsymbol{\omega}-\boldsymbol{\omega}'|)^{-1}, \quad \text{(A18)}$$

where

$$\mu = \tfrac{1}{2}\max(r,r'). \quad \text{(A19)}$$

Hence we have, with φ = angle between $\boldsymbol{\omega}$ and $\boldsymbol{\omega}'$,

$$\begin{aligned}
\frac{1}{C}\int_{\Omega} F(r\boldsymbol{\omega}-r'\boldsymbol{\omega}')|d\boldsymbol{\omega} &\leqslant 2\pi\int_0^{\pi} \frac{\sin\varphi\,d\varphi}{(1+\mu\sin(\varphi/2))} \\
&= 4\pi\int_0^{\pi/2} \frac{\sin 2\beta\,d\beta}{(1+\mu\sin\beta)} \\
&= 8\pi\int_0^1 \frac{z\,dz}{(1+\mu z)} \\
&= \frac{8\pi}{\mu^2}\int_1^{1+\mu} \frac{z-1}{z}\,dx \\
&= (8\pi/\mu^2)|\mu - \ln(1+\mu)| \\
&\leqslant 4\pi/(1+\mu)^2. \quad \text{(A20)}
\end{aligned}$$

Hence we have [cf. (A14)]

$$|G(r, r')| \leq \frac{4\pi C}{(1+\mu)^2} \leq \frac{4\pi}{(1+\mu)^2} \|F\|_1. \tag{A21}$$

The same is true, of course, if F is replaced by $D^j F$, where D^j is any derivative of order j, $0 \leq j \leq 2$. Hence, with $A(s, t)$ given by (A15) we have

$$(1 + |s|^{1/2})(1 + |t|^{1/2})|A(s, t)| \leq 16\pi \|F\|_1. \tag{A22}$$

Similarly,

$$\begin{aligned} &|A(s+h, t) - A(s-h, t)| \\ &\quad \leq \left| \frac{\partial F}{\partial r}(s^{1/2}\omega - t^{1/2}\omega') \right| \left| \frac{(s+h)^{1/2} - (s-h)^{1/2}}{(2h)^{1/2}} \right| \\ &\quad \leq \frac{4\pi \|F\|_1}{(1+\mu)^2}, \end{aligned} \tag{A23}$$

and

$$\begin{aligned} &|A(s+h, t+h) - A(s+h, t-h) - A(s-h), t+h) \\ &\quad + A(s-h, t-h)| \\ &\quad \leq \frac{\partial^2 F(s^{1/2}\omega - t^{1/2}\omega')}{\partial r \, \partial r'} \frac{(s+h)^{1/2} - (s-h)^{1/2}}{(2h)^{1/2}} \\ &\quad \times \frac{(t+h)^{1/2} - (t-h)^{1/2}}{(2h)^{1/2}} \leq \frac{4\pi \|F\|_1}{(1+\mu)^2}. \end{aligned} \tag{A24}$$

Combining all these estimates, we obtain the upper bound in (A17).

For the lower bound we simply observe that from (A3) we have

$$\max\{|MN'A|, |NN'A|\} \leq \|A\|_S. \tag{A25}$$

Hence, if we set $t = 0$ in (39), we find

$$\sup_{|h|<1} (2h)^{-1/2}(1 + |s|)^{1/4}|A(s, h)| \leq \|A\|_S, \tag{A26}$$

while if we set $s = t = 0$, we find

$$\left.\sup_{\substack{|h|<1 \\ |h'|<1}} (2h)^{-1/2}(2h')^{-1/2}|A(h, h')| \leq \|A\|_S\right\}. \tag{A27}$$

It follows that

$$\begin{aligned} &\sup (st)^{-1/4}|A(s, t)| \\ &\quad = \sup G(s^{1/2}, t^{1/2}) \leq \|A\|_S. \end{aligned} \tag{A28}$$

Now from (A14) we find

$$\begin{aligned} &G(s^{1/2}, 0) \\ &\quad = \max\left\{\sup_\omega |F(s^{1/2}\omega)|, \frac{1}{4\pi}\int_\Omega |F(s^{1/2}\omega| \, d\omega\right\} \\ &\quad = \sup_\omega |F(s^{1/2}\omega)|. \end{aligned} \tag{A29}$$

Hence by putting $t = 0$ in (A28) we conclude

$$\sup_{\mathbf{k}} |F(\mathbf{k})| = \sup_{\omega, r} |F(r, \omega)| \leq \|A\|_S \tag{A30}$$

which gives the lower bound in (A17).

If we now define the Friedrichs norm $\|A\|_F$ for A by

$$\|A\|_F = (10 + 2\pi)\|A\|_S, \tag{A31}$$

then we find, according to Lemma 2, that

$$\|B\Gamma A\|_F \leq \|B\|_F \|A\|_F \tag{A32}$$

and, according to Lemma 3, that

$$10\|F\|_0 \leq \|F\|_F \leq 1000\|F\|_1. \tag{A33}$$

APPENDIX B

As a simple example, we consider explicitly the case where the backscatter data are given by

$$B(\mathbf{k}) = \epsilon/(1 + \mathbf{k}^2) \tag{B1}$$

with ϵ a small parameter to be determined. One sees easily from (37) that

$$\|B\|_1 = 2\epsilon. \tag{B2}$$

Hence if we choose ϵ so that

$$32\epsilon \leq 0.001, \tag{B3}$$

then B satisfies (37) with $b = \frac{1}{16}$, and our analysis applies.

The first Born approximation V_1 to V is now

$$V_1(\mathbf{k}) = B(\mathbf{k}) = \epsilon/(1 + \mathbf{k}^2). \tag{B4}$$

In configuration space this becomes

$$V_1(\mathbf{x}) = \epsilon \, e^{-|x|}/4\pi|\mathbf{x}|. \tag{B5}$$

The error in this approximation is given by (38), which reduces to

$$|V(\mathbf{k}) - V_1(\mathbf{k})| \leq (0.025)(1/2)^2(1/2)^{-1} = 0.0125, \tag{B6}$$

valid for all k.

[1] R. T. Prosser, "Formal solutions of inverse scatteriang problems, I," J. Math. Phys. **10**, 1819–1822 (1969); II, *ibid.* **17**, 1775–1779 (1976); III, *ibid.* **21**, 2648–2653 (1980).

[2] K. O. Friedrichs, Perturbations of Continuous Spectra, Comm. Pure Appl. Math. **1**, 361–405 (1948).

[3] This condition, of course, rules out potentials admitting bound states (Ref. 2).

[4] J. Schwartz, "Some non-self-adjoint operators. I," Comm. Pure Appl. Math. **13**, 609–639 (1960); II, *ibid.* **14**, 619–626 (1961).

Perturbation theory of inelastic resonances

S. Bosanac[a)]
R. Bošković Institute, 41001 Zagreb, Croatia, Yugoslavia

(Received 8 May 1981; accepted for publication 21 August 1981)

Perturbation theory of inelastic resonances is developed. From the fact that the resonances appear as the roots of the Jost function, we show that perturbation coefficients are obtained without use of the complete set. The theory is generalized to the case when the unperturbed Hamiltonian is p-fold degenerate. The near degenerate case is also discussed, and the radius of convergence for the perturbation series is estimated. We also treat the perturbation theory of residues both in the nondegenerate and degenerate case.

PACS numbers: 03.80 + r, 03.65.Nk

1. INTRODUCTION

The theory of resonances has been an intriguing subject of research ever since the first papers by Breit and Wigner.[1,2] Results show that this problem can be treated in two ways, depending on what kind of information one wants to obtain. In the earlier line of research resonances were defined in terms of the complex eigenenergies of the Hamiltonian, while in later work the resonances were defined in terms of the complex angular momentum.[3–5] The two approaches are complementary but the Regge theory, as the latter is usually referred to, offers more insight if we are concerned with the calculation of cross sections,[6,7] or with the classification of resonances.[8] In many cases, however, the complex energy formalism is used as an alternative, such as in the study of decay properties of well-defined states,[9] in the analysis of the resonance cross sections where only a few partial waves are involved, and lately in atom-surface resonances.[10–12] For historic reasons, when we talk in this article about the resonances we will understand the complex energy formalism.

In order to understand resonances several techniques were developed: a partitioning approach,[13,14] the formal treatment,[15–19] and lately, the coordinate rotation method.[20,21] Of all the methods, perturbation theory is the most attractive since it gives the simplest understanding of how resonances are formed. However, there are obvious limitations to such an approach, as clearly pointed out by Fonda *et al.*[22] Nevertheless, there is still room for improvement in the perturbation theory and one is offered here. The perturbation theory presented in this article is an extension of the one recently developed for the inelastic Regge poles.[23]

Before giving the outline of the theory, let us make a few general comments about the shortcomings of the perturbation approach that has been in use so far. If we set aside the R-matrix method as not being perturbation theory in the true sense, we have only the one based on the Green's function formalism.[24] It offers a simple and lucid description of resonances. However, its formal nature obscures some difficulties when it is used as a basis for developing a perturbation theory of resonances: (a) the wave functions are obtained from the iteration of Fredholm type integral equations, a procedure which in many cases is not convergent,[25] (b) it requires definition of a complete set of functions, e.g., for determining the "level shifts," (c) it is not clear how to treat the degenerate resonances, i.e., the case when the unperturbed Hamiltonian is degenerate, of (d) how to treat resonances which are true resonances in the unperturbed Hamiltonian. Furthermore, the theory does not offer a simple way for calculating the peturbation expansion of the residues of the S-matrix.

In this article we develop a perturbation theory based on the fact that the resonance poles appear as the roots of a determinant of the Jost function.[26,27] The Jost function will be determined by an integral equation of Volterra type for which the iteration procedure produces a series which is absolutely covergent.[28] We also note that in the Jost function formalism there is no need for separation of channels into closed and open ones, such as in the approach of Feshbach.[14] Therefore we have complete symmetry between the resonances which are bound states and the true resonances in the unperturbed Hamiltonian. Furthermore, at no point do we use a complete set for determining the perturbation coefficients. The advantage of this was shown in the simple case of a single channel problem.[29] This means that the level shift is given only in terms of the solution for the particular unperturbed state which is subjected to the perturbation.

Another difference from the previous theory is the choice of variable which is expanded in a perturbation series. In our approach we expand the wave number, while usually the expansion is done for the energy variable. Although a minor point, in some cases we obtain better results as shown in a single channel case.[29]

We also show in this article how to treat the more general case in which there is a p-fold degeneracy in the unperturbed Hamiltonian. An estimate is given for the radius of convergence in both the nondegenerate and degenerate cases. It is shown that if we treat the nondegenerate problem as the degenerate one, with a suitable definition of the perturbation, the radius of convergence of the perturbation series is considerably enlarged.

In this article we will only treat the Feshbach-type resonances, i.e., the resonances which are bound states of the unperturbed Hamiltonian. However, the theory is of general validity and can be applied to the resonances which are also true resonances in the unperturbed Hamiltonian. In such cases, though, one should make an analytic continuation of the potential in the complex coordinate plane.[29]

In Secs. 4 and 8 we derive a perturbation theory for the

a)This work was supported in part by the grant NSF F6F006-Y.

 0022-2488/82/112131-09$02.50 © 1982 American Institute of Physics

residues both for the nondegenerate and degenerate case, respectively. We obtain results which are in agreement with the formal theory of resonances, i.e., the modulus of residues in nondegenerate theory is proportional to the partial width of resonances. In the degenerate case we find the same agreement with an additional factor measuring the order of degeneracy. In addition to this result we also find the phase of residues, which in the formal theory is essentially not determined.

In this article we will only consider the s-state resonances. The theory for nonzero angular momentum resonances can be obtained by simply modifying the potential matrix by the centrifugal term.

2. PERTURBATION THEORY FOR NONDEGENERATE POLES

Let the set of multichannel equations describing inelastic collision be

$$\psi'' = (V - K^2)\psi, \tag{2.1}$$

where V is the $n \times n$ matrix and K^2 is the diagonal matrix of channel energies. Let us also assume that the first few channel energies are negative, in which case they describe states that are not observed asymptotically. Our main interest, as was pointed out in the Introduction, is to calculate their contribution to the S matrix, in particular when their configuration allows formation of Feshbach-type resonances. In the model which will be considered, we will assume that the interaction between channels is weak, in which case the potential matrix V allows separation,

$$V = V_0 + \epsilon V', \tag{2.2}$$

where V' contains all the off-diagonal elements of V. V_0 is therefore diagonal.

The set of equations (2.1) has a solution in the form of a Volterra type integral equation

$$\psi = \psi_0 + \frac{\epsilon}{2i} K^{-1} \int_0^r G(r,r') V'(r') \psi(r') dr', \tag{2.3}$$

for which it can be proved that the iteration series converges absolutely.[28] The kernal is defined as

$$G(r,r') = f_0^+(r') f_0^-(r) - f_0^+(r) f_0^-(r'), \tag{2.4}$$

where the functions ψ_0 and $f_0^{\pm}$ are the regular and irregular solutions of the uncoupled equations (2.1), respectively. They are, therefore, the diagonal matrices.

For large r the irregular solutions are defined as

$$f^{\pm}(r) \sim e^{\mp ikr}, \quad r \to \infty \tag{2.5}$$

while the regular solution is given in a form of a linear combination

$$\psi \sim f_0^-(r) J^+ + f_0^+(r) J, \tag{2.6}$$

where J and J^+ are the Jost functions (in fact the $n \times n$ matrices, but we will refer to them as functions), and their ratio $J^+ J^{-1}$ gives the S matrix. From the understanding that the poles of the S matrix have physical interpretation of bound states and resonances, these observables are given from the solution of the equation

$$F = \mathrm{Det}(J) = 0. \tag{2.7}$$

It can be shown that the Jost function J is given by[27]

$$J = J_0 - \frac{\epsilon}{2i} K^{-1} \int_0^\infty dr\, f_0^-(r) V'(r) \psi(r) dr, \tag{2.8}$$

where the unperturbed Jost function J_0 is diagonal and corresponds to the solution of (2.1) with $\epsilon = 0$. By iterating (2.3) the resulting series for (2.8) is absolutely convergent, with the same restrictions as for (2.3). Therefore, the Jost function J is given as a power series in

$$J = J_0 + \epsilon J_1 + \frac{\epsilon^2}{2} J_2 + \cdots, \tag{2.9}$$

where

$$J_1 = -\frac{1}{2i} K^{-1} \int_0^\infty dr\, f_0^- V' \psi_0 \tag{2.10}$$

and

$$J_2 = \tfrac{1}{2} K^{-1} \int_o^\infty dr\, f_0^- V' K^{-1} \int_o^r G(r,r') V'(r') \psi_0(r') dr', \tag{2.11}$$

with the kernel defined in (2.4). In our treatment, the higher order corrections to J will not be required.

The derivation of the basic equations for obtaining the poles of the S matrix have been so far only formal and no conditions on the potential were given. As pointed out by Newton,[30] the function F, given by (2.7), is analytic near the roots which correspond to the resonances, but the Jost function (2.8) may not be defined in the same vicinity if more than one channel is closed. The reason is that, in general, ψ is exponentially increasing function for large r and and if the matrix elements of V' are not sufficiently rapidly going to zero in the same limit, the integral in (2.8) is infinite. Although for an absolute convergence of (2.3) it is sufficient that the first moments of V are finite, the same is not true for (2.8). Therefore we will define a cutoff in the potential matrix $V(r)$ at some large $r = R$. Beyond that point $V(r)$ is identically zero. For such a potential (2.8) is absolutely convergent, but the roots are now R-dependent. In order to obtain the R-independent roots we will in the final expressions for the perturbation coefficients of the roots take the limit $R \to \infty$, if such a limit exists. As will be shown, this limit exists and we will associate such roots with those corresponding to the true noncutoff potential $V(r)$, for which we do not require more than that the first moments are finite. For convenience, in the derivation of the roots we will omit any reference to R, but having in mind that they are defined in such manner.

We will now solve the equation for the poles of the S matrix (2.7) with the assumption (2.9), provided that no two roots of equation

$$F_0 = \mathrm{Det}(J_0) = 0 \tag{2.12}$$

for different channels are equal (the assumption of nondegeneracy). However, before doing that let us comment on F. The F is multidimensional in K but by replacing each element of K by

$$K_i = (k^2 - E_i)^{1/2}, \tag{2.13}$$

where E_i is the threshold energy of the ith channel, we look for the roots of (2.7) in the variable k. The set of all Riemann surfaces for each K_i is replaced by a single surface in the

variable k, with as many branch points as there are channels. Therefore, we should be careful how to define the physical sheet of k in order to get physically meaningful results.

Let us now assume that one zero of Eq. (2.7) is κ. It is obviously a function of ϵ, and for a small perturbation we can develop κ in a power series,

$$\kappa = k_0 + \epsilon k_1 + (\epsilon^2/2)\, k_2 + \cdots, \tag{2.14}$$

where $k_0 = \kappa(\epsilon = 0)$ and

$$k_1 = \frac{d\kappa}{d\epsilon}(\epsilon = 0), \quad k_2 = \frac{d^2\kappa}{d\epsilon^2}(\epsilon = 0). \tag{2.15}$$

The assumption on (2.14) is that the roots of F_0 are not close to the branch points in the k plane so that when the perturbation is included the radius of convergence in ϵ is smaller than 1. Otherwise, the series (2.14) is not convergent. We will also discuss other limitations on (2.14) in the subsequent sections.

It has been shown that the coefficients of expansion (2.15) are related to the function F by[23]

$$k_1 = -\mathring{F}/F', \quad k_2 = -(1/F')[\mathring{\mathring{F}} + 2k_1\mathring{F}' + F_1^2 F''], \tag{2.16}$$

where we use the notation $\mathring{F} = \partial F/\partial\epsilon$ and $F' = \partial F/\partial k$. Therefore in order to obtain the coefficients in the expansion (2.14) we require partial derivatives of F in both ϵ and k variables, and we must take their limit $k \to \kappa(\epsilon)$ and then take the limit $\epsilon \to 0$. However, in order to avoid calculation of k_2 through the second derivatives of F, we can evaluate $d\kappa/d\epsilon$ for finite ϵ but to terms of order $O(\epsilon^2)$, i.e.,

$$d\kappa/d\epsilon = k_1 + \epsilon k_2 + O(\epsilon^2), \tag{2.17}$$

in which case we get k_2 without using (2.16). Such a procedure considerably simplifies calculation of k_2 in the degenerate case.

The positive imaginary roots of F_0 in K are of particular importance to our analysis. When the channels are uncoupled they represent the bound states but with the coupling they become decaying states, i.e., Feshbach-type resonances. Other roots of F_0, in particular the complex ones, will not be treated in this article. Let us therefore assume that k_0 is the bound state root of the pth element of uncoupled Jost function, i.e., $j_p(k_0) = 0$. In such a case F' to zeroth order in ϵ_1 is given by[31]

$$F' = j_1 j_2 \cdots j_{p-1} j'_p j_{p+1} \cdots j_n \equiv j'_p P(j), \tag{2.18}$$

where j_i are the diagonal elements of J_0, and P designates the product of all j_i except the pth. However, to obtain (2.17) we need F' to $O(\epsilon^2)$. We therefore take the limit of F' for $k \to \kappa(\epsilon)$ and expand F' in powers of ϵ. For small, but nonzero ϵ, we can write

$$\begin{aligned} j_p &\sim j'_p(\kappa - k_0) \sim \epsilon k_1 j'_p, \\ j_i &\sim j_{i0} + j'_{i0}(\kappa - k_0) \sim j_{i0} + \epsilon k_1 j'_{i0}, \quad i \neq p \end{aligned} \tag{2.19}$$

where we have used (2.14). The derivative of j_p is taken for $\epsilon \to 0$. We find

$$F' = j'_p P(j) + \epsilon k_1 j'_p P(j) \sum_{i=1\neq p}^{n} \frac{j'_i}{j_i} + O(\epsilon)^2. \tag{2.20}$$

Let us now find $\mathring{F}$ to $O(\epsilon)^2$. By taking into account that the diagonal terms of J_1 are zero, we obtain

$$\mathring{F} = \epsilon P(j)\left[(J_2)_{pp} - 2\sum_{i=1\neq p}^{n} \frac{1}{j_i}(J_1)_{pi}(J_1)_{ip}\right] + O(\epsilon)^2, \tag{2.21}$$

hence k_1 as given by (2.17), is

$$k_1 = 0 \tag{2.22}$$

and k_2 is

$$k_2 = \frac{1}{j'_p}\left[2\sum_{i=1\neq p}^{n} \frac{1}{j_i}(J_1)_{pi}(J_1)_{ip} - (J_2)_{pp}\right] \tag{2.23}$$

as result analogous to the Regge poles.[23] The matrix elements of J_1 and J_2 are given by (2.10) and (2.11), respectively.

Higher order coefficients are obtained in a similar manner.

3. DISCUSSION OF k_2

As we have shown, for a perturbation of the form (2.2), the first order correction in the expansion of the poles of the S matrix in ϵ is exactly zero. The leading term is k_2, given by (2.23). However, such a form of k_2 can be reduced to a simpler form, more tractable for discussion. If we take into account (2.10) and (2.11) and use (2.6) we obtain, after some algebra,

$$\begin{aligned} k_2 = -\frac{i}{2k}\sum_{l=1\neq p}^{n} \frac{1}{K_l j_l}\Bigg[&\int_0^\infty \psi_p V_{pl}\psi_l\, dr \int_r^\infty f_l^- V_{lp}\psi_p\, dr' \\ &+ \int_0^\infty dr\, \psi_p V_{pl} f_l^- \int_0^r dr' \psi_l V_{lp}\psi_p\Bigg], \end{aligned} \tag{3.1}$$

where we have dropped all the zeros referring to the unperturbed solutions. The function ψ_p in (3.1) is the unperturbed bound state.

We have also used the relationship for the normalization constant of the bound state ψ_p, given by[17]

$$ij_p^+ \frac{\partial j_p}{\partial K_p} = \int_0^\infty \psi_p^2 dr = N^2. \tag{3.2}$$

Since j' is derivative with respect to k, we have

$$ij'_p j_p^+ k_p = kN^2, \tag{3.3}$$

where k is real. Therefore ψ_p in (3.1) is a normalized function.

Some of the channels are closed and some are open. We will now show that k_2 is finite despite the fact that in (3.1) there are matrix elements connecting the pth channel with the other closed channels. Let us assume, for simplicity, that the first s channels are closed (it follows that $p \leqslant s$) in which case the solution f_i^-, $i \leqslant s$, for large r

$$f_l^-(r) \sim e^{-K_l r}, \quad r \to \infty. \tag{3.4}$$

However, ψ_i is

$$\psi_i \sim e^{K_i r} j_i, \tag{3.5}$$

therefore the integrand in (3.1) containing ψ_i is

$$\begin{aligned} e^{-K_p r} V_{pi}\Bigg(&e^{K_i r}\int_r^\infty e^{-K_i r' - K_p r'} V_{ip}\, dr' \\ &+ e^{-K_i r}\int_0^r e^{K_i r' - K_p r'} V_{ip}\, dr'\Bigg) \end{aligned} \tag{3.6}$$

and goes to zero for $r \to \infty$ at least as $\exp(-2K_p r)$. There-

fore, the coefficient k_2 is finite.

Let us now calculate the portion of the sum in (3.1) corresponding to the closed channels only. j_l in such a case is real and K_l is positive imaginary, therefore

$$k_2(\text{cl.ch.}) = -\frac{1}{2k}\sum_{l=1\neq p}^{s}\frac{1}{K_l j_l}\left[\int_0^\infty \psi_p V_{pl}\psi_l dr\int_r^\infty f_l^- V_{lp}\psi_p dr' + \int_0^\infty dr\,\psi_p V_{pl} f_l^- \int_0^r dr'\psi_l V_{lp}\psi_p\right] \tag{3.7}$$

is real, meaning that the interaction of ψ_p with the closed channels cannot produce a decaying state, but this term only contributes to the level shift. The imaginary part of k_2 is entirely due to the interaction of ψ_p with the open channels, as one expects on physical grounds.

The imaginary part of k_2 is therefore

$$\mathrm{Im}(k_2) = \frac{1}{2i}(k_2 - k_2^*) = -\frac{1}{4k}\sum_{l=\text{open}}\frac{1}{k_l j_l j_l^\dagger}\left[\int_0^\infty \psi_p V_{pl}\psi_l dr\right]^2, \tag{3.8}$$

which is negative since j_l^+ is the complex conjugate of j_l. The remaining real part of k_2 is obtained by subtracting (3.7) and (3.8) from (3.1).

The resonance energy is now

$$k_R^2 = (k_0 + \tfrac{1}{2}k_2)^2 = k_0^2 + k_0[\mathrm{Re}(k_2) + i\mathrm{Im}(k_2)] + \tfrac{1}{4}k_2^2, \tag{3.9}$$

from where we obtain the level shift and its width,

$$\Delta = k_0\mathrm{Re}(k_2), \quad \Gamma = k_0\mathrm{Im}(k_2). \tag{3.10}$$

By noticing that Re (k_2) is calculated from (3.1), as described earlier, we notice that the level shift is given with no reference to the complete set of unperturbed solutions of Hamiltonian, as it is given, for example, in Feshbach theory. The level shift is entirely determined by the interaction of the bound state ψ_p, which is being perturbed, and all other channels.

4. PERTURBATION EXPANSION OF RESIDUES FOR NONDEGENERATE POLES

The diagonal S-matrix elements for the open channels were shown by Newton[30] to be

$$S_{ii} = F(-K_i)/F, \tag{4.1}$$

where F was defined in (2.7) and $F(-K_i)$ means that the ith element of K in F is replaced by its negative value. The off-diagonal elements are then given by

$$S_{ij}^2 = S_{ii}S_{jj} - F(-K_i, -K_j)/F, \tag{4.2}$$

where $F(-K_i, -K_j)$ means that both variables K_i and K_j have changed their sign.

By definition the residue of S_{ii} is

$$\beta_i = \lim_{k\to\kappa}(k-\kappa)S_{ii} \tag{4.3}$$

and it can be easily shown that the off-diagonal residues are given by

$$\lim_{k\to\kappa}(k-\kappa)S_{ij} = (\beta_i\beta_j)^{1/2}. \tag{4.4}$$

Therefore only the diagonal residues are independent. The pole $\kappa(\epsilon)$ in (4.3) is from the closed channels.

The residues are also a function of ϵ and in the weak coupling limit we can write

$$\beta(\epsilon) = \beta_0 + \epsilon\beta_1 + (\epsilon^2/2)\beta_2 + \cdots, \tag{4.5}$$

where the subscript refering to the ith channel was omitted for convenience. However, it should be born in mind that the expansion (4.5) applies to every open channel.

In order to obtain the leading term in the expansion (4.5), let us first take the limit (4.3)

$$\beta = \lim_{k\to\kappa}\frac{F(-K)}{F'} \tag{4.6}$$

and calculate the leading terms of $F(-K)$ and F' in ϵ. The index of K in (4.6) was dropped with the understanding that the minus sign applies to only one open channel.

As it was shown in (2.18), F' is of the order ϵ^0 and given by

$$F' = j_p' P(j). \tag{4.7}$$

On the other hand, $F(-K)$ is zero in the limit $\epsilon\to 0$ since $F_0(-K)$ is a product of the elements of unperturbed Jost function and one of them in the closed channels is zero. Therefore

$$\beta_0 = 0. \tag{4.8}$$

To obtain the leading term of $F(-K)$ in ϵ, we should notice that j_p is given in the form

$$j_p(\kappa(\epsilon)) \sim j_p + j_p'(\kappa - k_0) + \tfrac{1}{2}j_p''(\kappa - k_0)^2 \sim (\epsilon^2/2)j_p' k_2 + O(\epsilon^3), \tag{4.9}$$

where we have taken into account that k_0 is a zero of j_p and that $k_1 = 0$, as shown in Sec. II.

The matrix $J(-K)$, as a function of ϵ, contains terms of the order ϵ^0 only on the diagonal, except the pth element, which is of the order ϵ^2 and given by

$$J_{pp}(-K) \sim (\epsilon^2/2)[k_2 j_p' + (J_2)_{pp}], \tag{4.10}$$

where (4.9) was taken into account. The terms of the order ϵ are all the off-diagonal elements of $J(-K)$. The determinant of $J(-K)$ can now be evaluated as an expansion in ϵ. Since the pth diagonal element of $J(-K)$ is of the order ϵ^2, the factor ϵ which multiplies the pth column and the pth row can be taken out leaving a determinant of the matrix with the elements of order ϵ^0 along the diagonal and in the pth row and column. Therefore we get

$$F(-K_0) = \epsilon^2 P(j)\left[\tfrac{1}{2}k_2 j_p' + \tfrac{1}{2}(J_2(-K_0))_{pp} - \sum_{l=1\neq p,0}^{n}\frac{(J_1)_{pl}(J_1)_{lp}}{j_l} - \frac{(J_1(-K_0))_{p0}(J_1(-K_0))_{0p}}{j_0(-K_0)}\right]\frac{j_0(-K_0)}{j_0}, \tag{4.11}$$

where k_2 is given by (2.23). By inserting k_2 into (4.11) we get

$$F(-K_0)=\epsilon^2\frac{P(j)j_0(-K_0)}{2j_0}\left[(J_2)_{pp}-(J_2(-K_0))_{pp}+2\frac{(J_1)_{p0}(J_1)_{0p}}{j_0}-2\frac{(J_1(-K_0))_{p0}(J_1(-K_0))_{0p}}{j_0(-K_0)}\right]. \tag{4.12}$$

Noticing that for the open channels $j(-k)=j^+(k)=j^*(k)$ and that

$$(J_1(-K_0))_{p0}=(J_1)_{p0},\quad J_1(-K_0)_{0p}=(J_1^*)_{0p}, \tag{4.13}$$

where we have taken into account that ψ is symmetric with respect to the change $k\to -k$, and also that

$$(J_2)_{pp}=(J_2(-K_0))_{pp}, \tag{4.14}$$

we obtain for $F(-K)$

$$F(-K_0)=-\epsilon^2\frac{P(j)}{4j_0^2K_0K_p}\frac{1}{j_p^+}\left[\int_0^\infty dr\psi_0V'_{0p}\psi_p\right]^2. \tag{4.15}$$

Therefore, the leading term in the expansion (4.5) is of the order ϵ^2 giving for β^2

$$\beta_2=-\frac{i}{2kk_0j_0^2}\left[\int_0^\infty dr\,\psi_0V'_{0p}\psi_p\right]^2, \tag{4.16}$$

where ψ_p is normalized.

Higher order coefficients are obtained by calculating terms of the order ϵ^2 in $F(-K)$ and of the order ϵ in F'.

5. PERTURBATION THEORY FOR DEGENERATE POLES

In Sec. 2 we described a perturbation theory for the poles of S matrix with the assumption that only one element of the Jost function in the uncoupled equations has the root k_0. In this section we will take a general case where p elements of the uncoupled Jost function have the same root k_0, out of the total of n elements j_i.

For a clearer presentation let us introduce a new notation. We will designate by $g(k)$ the p elements of uncoupled Jost function, which have the root k_0, and we designate $h(k)$ those which are nonzero for k_0. The dimension of g is p and the dimension of h is s, giving the total $p+s=n$. We will designate by A the first-order perturbation to the Jost function J_0 and the second order by B, therefore we have

$$J=\begin{vmatrix}J_{11}, & J_{12}\\ J_{21}, & J_{22}\end{vmatrix}=\begin{vmatrix}g, & 0\\ 0, & h\end{vmatrix}+\epsilon\begin{vmatrix}A_{11}, & A_{12}\\ A_{21}, & A_{22}\end{vmatrix}+\frac{\epsilon^2}{2}\begin{vmatrix}B_{11}, & B_{12}\\ B_{21}, & B_{22}\end{vmatrix}+\cdots, \tag{5.1}$$

where we have assumed, as a matter of convenience, that g forms an upper left submatrix $p\times p$. Let us also assume that A has nonzero terms on the diagonal, in contrast with Sec. II. The reason for this will be discussed later.

The determinant of J is designated as F, and its zeros

$$F=0$$

are the poles of the S matrix. Let one of the roots of (5.2) be $\kappa(\epsilon)$. It can be expanded in the power series in

$$\kappa(\epsilon)=k_0+\epsilon k_1+(\epsilon^2/2)k_2+\cdots \tag{5.3}$$

and for small ϵ we can write for J_{11}

$$J_{11}\sim\epsilon(k_1g'+A_{11})+(\epsilon^2/2)(g'k_2+k_1^2g''+2k_1A'_{11}+B_{11})+O(\epsilon^3), \tag{5.4}$$

where the terms on the right-hand side are evaluated for $k=k_0$. The last result shows that the matrix elements of J are all of the order ϵ except the diagonal ones corresponding to h.

The coefficient k_1 in (5.3) is

$$k_1=-\mathring{F}/F' \tag{5.5}$$

and if we use the representation

$$F=\exp[\mathrm{Tr}(\ln(J))] \tag{5.6}$$

we can write for (5.5)

$$k_1=-\frac{\mathrm{Tr}[J^{-1}\mathring{J}]}{\mathrm{Tr}[J^{-1}J^1]},\quad \epsilon\to 0. \tag{5.7}$$

The inverse of J is singular in the limit $\epsilon\to 0$ because if we take into account (5.4), only the submatrix J_{22} is nonzero. However, for a finite but small ϵ, we can write approximately

$$J^{-1}\sim\begin{vmatrix}\epsilon(k_1g'+A_{11}), & \epsilon A_{12}\\ \epsilon A_{21}, & h\end{vmatrix}^{-1}\sim\begin{vmatrix}(1/\epsilon)(k_1g'+A_{11})^{-1}, & O(\epsilon^0)\\ O(\epsilon^0), & h^{-1}\end{vmatrix}. \tag{5.8}$$

Hence the traces in (5.7) are

$$\mathrm{Tr}(J^{-1}\mathring{J})\sim\frac{1}{\epsilon}\mathrm{Tr}[(k_1g'+A_{11})^{-1}A_{11}]=\frac{p}{\epsilon}-\frac{k_1}{\epsilon}\times\mathrm{Tr}[(g'k_1+A_{11})^{-1}g'],$$
$$\mathrm{Tr}(J^{-1}J^1)\sim(1/\epsilon)\mathrm{Tr}[k_1g'+A_{11})^{-1}g'] \tag{5.9}$$

and the coefficient k_1 is

$$k_1=k_1-\frac{p}{\mathrm{Tr}[(g'k_1+A_{11})^{-1}g']}. \tag{5.10}$$

Provided g' is a nonzero matrix, the last equation is an implicit equation for k_1 and has a solution if and only if

$$\mathrm{Det}(g'k_1+A_{11})=0 \tag{5.11}$$

giving p solutions for k_1. We will assume that they are all nonequal. Therefore, if p unperturbed zeros of J are degenerate, the first-order correction to the roots destroy the degeneracy, the splitting being given by the solutions of (5.11). In the next section we will show that in the case when $A=J_1$ all the roots k_1 are real.

The second-order coefficient k_2 is more difficult to obtain and before we calculate it we must perform certain transformations on the Jost function J. Let us define for k_1, as calculated from (5.11), a matrix U which diagonalizes $k_1I+g'^{-1}A_{11}$,

$$U^{-1}(k_1+g'^{-1}A_{11})U=\lambda. \tag{5.12}$$

As will be shown in the next section, if we put $A=J_1$, the

matrix U is not unitary since $g'^{-1}A_{11}$ is not symmetric. However, by a simple transformation of the form $Ng'^{-1}A_{11}N^{-1}$, where N is a diagonal matrix with the nonzero elements, we obtain a symmetric matrix for which we know that the eigenvalues are real. Therefore, we can write $U = N^{-1}T$, where T is unitary, hence the eigenvalues λ are all real. It turns out that the elements of N are the normalization constants for the degenerate states.

We now show that one eigenvalue of λ_l is zero, all the others being nonzero. Taking the determinant of (5.12) the left-hand side gives (5.11), while the right hand side is a product of the eigenvalues. Since the k_1 are all different, by our assumption, Eq. (5.11) has only first-order zeros, therefore the product of the eigenvalues λ_l is zero in the first order. Hence only one of the λ_l is zero. We can now define the index l of λ_l. It indicates that the l th eigenvalue is zero and for the convenience we will assume that it corresponds to the l th root of (5.11). If this is not the case we can always arrange the sequence of K_1 in order to get this correspondence. In our discussion we will drop the index l having in mind, when necessary, that all the derivations are done for a particular k_1.

Let us now transform the matrix $g'^{-1}J$ by

$$Y = \begin{vmatrix} U^{-1}, & 0 \\ 0 & I \end{vmatrix} \begin{vmatrix} g'^{-1}, & 0 \\ 0, & I \end{vmatrix} J \begin{vmatrix} U, & 0 \\ 0, & I \end{vmatrix} = \begin{vmatrix} U^{-1}g'^{-1}J_{11}U, & U^{-1}g'^{-1}J_{12} \\ J_{21}U, & J_{22} \end{vmatrix}, \tag{5.13}$$

in which case when $k = \kappa(\epsilon)$ the matrix Y_{11} is

$$Y_{11} \sim \epsilon\lambda + \epsilon^2/2U^{-1}(k_2 + g'^{-1}k_1^2 g'' + 2k_1 g'^{-1}A'_{11} + g'^{-1}B_{11})U \equiv \epsilon\lambda + (\epsilon^2/2)\gamma. \tag{5.14}$$

The second-order coefficient k_2 is now obtained as in (2.17), where Y now replaces J. We get

$$\frac{d\kappa(\epsilon)}{d\epsilon} = -\frac{\mathrm{Tr}[Y^{-1}\mathring{Y}]}{\mathrm{Tr}[Y^{-1}Y']} = k_1 + \epsilon k_2 + O(\epsilon^2), \tag{5.15}$$

where Y^{-1} is singular in the limit $\epsilon \to 0$. By taking into account that the lth eigenvalue λ is zero, and that Y_{11} is of the form (5.14), we obtain Y^{-1} as a power series in

$$(Y^{-1})_{ij} = \frac{c}{\epsilon^2}\alpha_{ij} + \frac{c}{\epsilon}\beta_{ij} + O(\epsilon^0) = \frac{M_{ji}}{\mathrm{Det}(Y)}(-)^{i+j}, \tag{5.16}$$

where M_{ji} are the minors corresponding to the element Y_{ji}. The constant C is suitably chosen in order that the elements α_{ij} and β_{ij} are formally simple. The constant C will cancel in (5.15).

It can be shown that α is zero everywhere except in the position (1,1) where we can take it to be

$$\alpha_{11} = 1. \tag{5.17}$$

The nonzero elements of β, when the minors M_{ji} in (5.16) are calculated and the normalization (5.17) is assumed, are

$$\beta_{ii} = \frac{1}{\lambda_i}\left[\tfrac{1}{2}\gamma_{ii} - \sum_{r=1}^{s}\frac{a_{e,r+p}a_{r+p,e}}{h_r}\right], \quad i \leqslant p, i \neq e \tag{5.18a}$$

$$\beta_{ll} = \sum_{r=1\neq l}^{p}\frac{\gamma_{rr}}{2\lambda_r} + \sum_{r=1}^{s}\frac{1}{h_r}\left[a_{r+p,r+p} - \sum_{q=1,\neq l}^{p}\frac{a_{q,r+p}a_{r+p,q}}{\lambda_q}\right], \tag{5.18b}$$

$$\beta_{il} = -\frac{1}{\lambda_i}\left[\tfrac{1}{2}\gamma_{il} - \sum_{r=1}^{s}\frac{a_{l,rp}a_{r+p,l}}{h_r}\right], \quad i \leqslant p \tag{5.18c}$$

$$\beta_{li} = -\frac{1}{h_{i-p}}a_{li}, \quad \beta_{ie} = -\frac{1}{h_{i-p}}a_{il}, \quad i > p. \tag{5.18d}$$

In Eqs. (5.18a)–(5.18d) we have used notation

$$a_{li} = (U^{-1}g'^{-1}A_{12})_{li}, \quad a_{il} = (A_{12}U)_{il}$$
$$a_{ij} = (U^{-1}g'^{-1}A_{11}U)_{ij}, \quad j \leqslant p \tag{5.19}$$

where $i > p$. The trace of $Y^{-1}\mathring{Y}$ can now be calculated and it is given, to the order ϵ_1^0 by

$$\mathrm{Tr}[Y^{-1}\mathring{Y}] = \frac{1}{\epsilon^2}a_{ll} + \frac{1}{\epsilon}\left[\beta_{ll}a_{ll} + (U^{-1}g'^{-1}(k_1A'_{11} + B_{11})U)_{ll} + \sum_{i=1\neq l}^{p}\beta_{ii}a_{ii} + \sum_{i=1}^{s}(\beta_{p+i,l}a_{l,p+i} + \beta_{l,p+1,i}a_{p+i,l})\right] + O(\mathring{\epsilon}). \tag{5.20}$$

Similarly, we can calculate the trace of $Y^{-1}Y'$,

$$\mathrm{Tr}[Y^{-1}Y'] = \frac{1}{\epsilon^2} + \frac{1}{\epsilon}\left[\beta_{ll} + (U^{-1}g'^{-1}(k_1 g'' + A'_{11})U)_{ll} + \sum_{i=1\neq l}^{p}\beta_{ii}\right] + O(\mathring{\epsilon}). \tag{5.21}$$

Their ratio gives (5.15), from which we obtain k_1 and k_2. We have

$$\frac{\mathrm{Tr}[Y^{-1}\mathring{T}]}{\mathrm{Tr}[Y^{-1}Y']} = -k_1 - \epsilon k_2 = a_{ll}\left[1 + \frac{\epsilon}{a_{ll}} \times\left[\frac{p-1}{2}\gamma_{ll} - (p+1)\sum_{r=1}^{s}\frac{a_{l,r+p}a_{r+p,l}}{h_r} + (U^{-1}g'^{-1}(B_{11} + 2k_1A'_{11} + k_1^2 g'')U)_{ll}\right]\right] \tag{5.22}$$

from which we obtain

$$k_1^{(l)} = -a_{ll}, \tag{5.23}$$

which is just repeating our previous result that the l th eigenvalue is zero. The coefficient k_2 is

$$k_2 = -k_1^2(U^{-1}g'^{-1}g''U)_{ll} - 2k_1(U^{-1}g'^{-1}A'_{11}U)_{ll} - (U^{-1}g'^{-1}B_{11}U)_{ll} + 2\sum_{r=1}^{s}\frac{a_{l,r+p}a_{r+p,l}}{h_r}. \tag{5.24}$$

Since the coefficient k_2 is given for a particular k_1, we have p values of k_2.

6. DISCUSSION OF DEGENERATE PERTURBATION SERIES

As it was shown in the last section, when there is a p-

fold degeneracy in the unperturbed system the first-order term in the expansion (5.3) splits the degeneracy. The magnitude of the split is calculated from the equation

$$\mathrm{Det}[g'k_1 + A_{11}] = 0 \tag{6.1}$$

and if A is taken to be J_1, defined by (2.11), we can prove that the k_1 are all real. By taking g' out of the bracket in (6.1), we get

$$(g'^{-1}A_{11})_{pq} = -\frac{1}{2kN_p^2}\int_0^\infty \psi_p V'_{pq}\psi_q dr, \tag{6.2}$$

where the relationship (3.3) was used.

In general the matrix (6.2) is not symmetric but it can be made symmetric if it is transformed by $Ng'^{-1}A_{11}N^{-1}$, where N is a diagonal matrix of the normalization constants of ψ. Since the determinant in (6.1) is invariant to such a transformation we obtain an equivalent equation

$$\mathrm{Det}[k_1 + Ng'^{-1}A_{11}N^{-1}] = 0. \tag{6.3}$$

Therefore, Eq. (6.1) is equivalent to solving the eigenvalue problem for k_1, with the symmetric and real matrix, hence the roots are all real. This is also the proof that the transformation matrix U from (5.12) is not unitary but that the eigenvalues λ are real since U can always be replaced by $U = N^{-1}T$, where T is unitary. In fact, the eigenvalues λ, for the l th root k_1, are given by

$$\lambda_i = k_1^{(l)} - k_1^{(i)}. \tag{6.4}$$

The second-order coefficient k_2 in (5.24) resembles the one nondegenerate case in (2.24), except for the terms with k_1 and the transformation matrix U. This is to be expected since k_1 is equal to zero for the nondegenerate case, and the matrix U is a mixing parameter between degenerate levels. In order to simplify the discussion of k_2 let us define a matrix k_2, for each k_1, with the property

$$k_2 = (U^{-1}\mathbf{k}_2 U)_{ll}. \tag{6.5}$$

Let us also replace $g \to j$, $A \to J_1$, and $B \to J_2$. The matrix $\mathbf{k}_2$ can then be represented as a sum of two terms; one in which the sum over the intermediate states in J_2 extends over the degenerate levels only and the other which contains all the other summation indices, i.e.,

$$\mathbf{k}_2(\mathrm{I}) = -k_1^2\frac{j''}{j'} + \frac{1}{i}k_1 j'^{-1}\frac{\partial}{\partial k}\left(K^{-1}\int_0^\infty dr f^- V'\psi\right) - \frac{i}{2k}N^{-1}\int_0^\infty dr\psi V' K^{-1}\int_0^r K(r,r')V'(r')\psi dr' N \tag{6.6}$$

and

$$(\mathbf{k}_2(\mathrm{II}))_{pq} = -\frac{i}{2k}N_p^{-1}N_q\sum_{S=\text{nond.}}\frac{1}{K_s j_s}\times\left[\int_0^\infty \psi_p V_{ps}\psi_s dr\int_r^\infty f_s^- V_{sq}\psi_q dr' + \int_0^\infty dr\,\psi_p V_{ps}\, f_s^-\int_0^r dr'\psi_s V_{sq}\psi_q\right]. \tag{6.7}$$

In the last expression N is the normalization matrix with the elements defined by (3.2), and the functions ψ corresponding to the bound states are all normalized.

The matrix $\mathbf{k}_2(1)$ is real which can be proved by noticing that the wave vectors k_p are imaginary and that the Jost functions j are real. Therefore, this part of $\mathbf{k}_2$ will only contribute to the real part of k_2. The imaginary part of k_2 comes from $\mathbf{k}_2(11)$ and can be obtained as in (3.8),

$$\mathrm{Im}(\mathbf{k}_2(11))_{pq} = -\frac{1}{4k}\sum_{s=\text{open}}\frac{N_p^{-1}N_q}{K_s j_s j_s^+}\times\int_0^\infty dr\,\psi_p V_{ps}\psi_s\int_0^\infty dr'\psi_s V_{sq}\psi_q, \tag{6.8}$$

thus giving for the imaginary part of k_2

$$\mathrm{Im}(k_2) = -\frac{1}{4k}\sum_{s=\text{open}}\frac{1}{K_s j_s j_s^+}\times\left[\sum_{p=\text{deg}} T_{pl}\int_0^\infty dr\,\psi_p V_{ps}\psi_s\right]^2, \tag{6.9}$$

which is in close analogy with (3.8) for the nondegenerate case. However, the real part of k_2 contains two terms which are not present in the nondegenerate case.

The p-fold pole k_0 thus produces, when the coupling is introduced, p complex poles. The poles are clustered around k_0, with the spacing between them being $k_1^{(i)} - k_1^{(j)}$. This implies that the diagonal S-matrix elements near such poles do not have the simple form

$$S_{ii} \sim \frac{\beta_i}{k - \kappa(\epsilon)} \tag{6.10}$$

but are given by

$$S_{ii} \sim \sum_{l=1}^{p}\frac{\beta_i^{(l)}}{k - \kappa_l(\epsilon)}. \tag{6.11}$$

It is also clear that the usual form for the Breit–Wigner resonance is destroyed and replaced by a more complicated structure, in fact the complicated structure given by

$$|S_{ii}|^2 \sim \sum_{l,j=1}^{p}\frac{\beta_i^{(l)}\beta_i^{*(j)}}{(k-\kappa_l)(k-\kappa_j^*)}. \tag{6.12}$$

The off diagonal elements of the S-matrix are

$$S_{ij} = \left(\sum_{l,m}\frac{\beta_i^{(l)}\beta_j^{-(m)}}{(k-\kappa_l)(k-\kappa_m)}\right)^{1/2}. \tag{6.13}$$

In Eqs. (6.12) and (6.13) we have neglected the background term, which is present in the exact Breit–Wigner formula.

7. NEARLY DEGENERATE CASE

Let us assume that in the uncoupled Jost function J_0 only one j_l, say j_1, is zero for $k = k_0$. Let us also assume that $p - 1$ other j_l are approximately zero, meaning

$$j_l(k) \sim j_l(k_0) + j_l'(k_0)(k - k_0), \quad 1 < l \leqslant p \tag{7.1}$$

where $|j_1(k_0)| \sim 0$. The question is whether in such a case one should use the nondegenerate perturbation theory of Sec. 2 or one can define $j_l(k_0)$ as a perturbation and in the zeroth order treat the problem as a p-fold degenerate one. In the notation of Sec. 5 we define in the latter case

$$g_l = j_l(k) - j_l(k_0), \quad A_{11} = (J_1)_{ll} + j_l(k_0) \tag{7.2}$$

and the problem is formally replaced by a degenerate perturbation scheme.

To answer this question we will analyze the two-channel case, where the whole problem will unfold in simple analytic terms. The problem which we solve, analogous to (2.7), is

$$\mathrm{Det}\begin{vmatrix} g(x), & \epsilon a(x) \\ \epsilon a(x), & h(x) \end{vmatrix} = 0. \tag{7.3}$$

Let g have the root $x = x_0$. Near the true solution x of (7.3) we can write

$$\mathrm{Det}\begin{vmatrix} g'(x-x_0), & \epsilon a + \epsilon a'(x-x_0) \\ , & h + h'(x-x_0) \end{vmatrix} = 0, \tag{7.4}$$

where h is small. Equation (7.4) is quadratic in x and the solution is given by

$$x - x_0 = \frac{hg' - 2\epsilon^2 aa' - [(hg' - 2\epsilon^2 aa')^2 + 4\epsilon^2 a^2(g'h' - \epsilon^2 a'^2)]^{1/2}}{2(g'h' - \epsilon^2 a'^2)}. \tag{7.5}$$

The expansion of $x(\epsilon)$ in the powers of ϵ is now equivalent to (2.14) and the radius of convergence can be explicitly obtained from (7.5). In fact there are two radii for ϵ; one coming from the zero of the denominator and the other from the zero of the function under the square root. The denominator gives for the radius

$$\epsilon_1 = |a'/(g'h')^{1/2}|, \tag{7.6}$$

while the square root gives

$$\epsilon^2 = \left| \frac{ka'}{2[ag'(ah' - a'h)]^{1/2}} \right| \tag{7.7}$$

and the radius of convergence for the series is

$$\epsilon = \min(\epsilon_1, \epsilon_2). \tag{7.8}$$

For $|h| \to 0$ the radius ϵ_1 is constant; however, ϵ_2 goes to zero. Therefore the perturbation expansion based on (7.4) has zero radius of convergency, i.e., the perturbation series (2.14) is not convergent.

Let us define h as a perturbation, in which case we have for (7.3)

$$\mathrm{Det}\begin{vmatrix} g'(x-x_0) & , \ \epsilon a + \epsilon a'(x-x_0) \\ & , \ \epsilon h + h'(x-x_0) \end{vmatrix} \tag{7.9}$$

and for $\epsilon \to 0$ the problem reduces to the degenerate case. From the solution of (7.9) we also find two radii of convergence; one given by (7.6) and the other

$$\epsilon_2 = \left| \frac{h^2 g'^2 + 4a^2 g'h'}{4hg'^2 a} \right|. \tag{7.10}$$

In the limit $h \to 0$ the radius ϵ_2 is infinite, therefore the radius of convergence for the series (5.3) is determined by (7.6) and it is independent of h. This fact may be of advantage for treating some cases which are strictly nondegenerate as the degenerate ones.

Although we analyzed the simple 2 channel problem, we see that in general it is better to reformulate the nearly degenerate problem as being exactly degenerate, using the transformation (7.2).

8. DEGENERATE PERTURBATION THEORY FOR RESIDUES

In Sec. 4 we showed how to develop the perturbation expansion for the residues in the case when the relevant poles are not degenerate in the unperturbed Hamiltonian. Here we will assume, as in Sec. 5, that p channels in the uncoupled equations are degenerate in the pole k_0. The notation will be that of Sec. 5.

The residue in the oth open channel is

$$\beta_0 = \frac{1}{2\pi i} \oint_C \frac{F(-k_0)}{F}\, dk, \tag{8.1}$$

where the integration contour C encircles the relevant pole. We will assume that the perturbation ϵ destroys the degeneracy, therefore all the poles from the cluster of p poles are simple. Let us also, for convenience, drop all the reference to the oth channel and write $F(-k_0)$ simply as $F(-K)$.

By expanding β into a series

$$\beta = \beta_0 + \epsilon\beta_1 + (\epsilon^2/2)\beta_2 + \cdots, \tag{8.2}$$

we find the residue in terms of the solutions of the unperturbed Hamiltonian. The coefficients from (8.2) are therefore

$$\beta_m = \frac{1}{2\pi i} \oint_C \frac{d^m}{d\epsilon^m} \frac{F(-K_0)}{F}\, dk, \tag{8.3}$$

where the integration is performed prior to taking the limit $\epsilon \to 0$. This is important since if the limit $\epsilon \to 0$ is taken first, all the poles from the cluster become k_0 and the contour integration C is then undefined.

The coefficient β_1 is thus obtained as

$$\begin{aligned} \beta_1 &= \frac{1}{2\pi i} \oint \left[\frac{\mathring{F}(-K)}{F} - \frac{\mathring{F}F(-K)}{F^2} \right] dk \\ &= \lim_{k \to \kappa(\epsilon)} [k - \kappa(\epsilon)] \frac{\mathring{F}(-K)F - F(-K)\mathring{F}}{F^2}, \end{aligned} \tag{8.4}$$

where $\kappa(\epsilon)$ is given by (5.3) and the limit $\epsilon \to 0$ is implicitly assumed. By taking the limit in (8.4) we obtain

$$\beta_1 = (1/F'^2)[\mathring{F}(-K)F - F(-K)\mathring{F}]. \tag{8.5}$$

The determinants $F = \mathrm{Det}(J)$ and $F(-K)$ are invariant to the transformation (5.13), therefore the limit $\epsilon \to 0$ of (8.5) can be proved to be zero, as in the nondegenerate case.

The second order contribution β_2 is obtained from (8.3) and we find

$$\begin{aligned} \beta_2 = \frac{1}{F'} &\left[\overset{\circ\circ}{F}(-K) - \overset{\circ\circ}{F} \frac{F'(-K)}{F'} \right] \\ &+ 2\lim_{k \to \kappa} \frac{d}{dk} \left[(k-\kappa)^2 \left(\frac{\mathring{F}^2 F(-K)}{F^3} - \frac{\mathring{F}\mathring{F}(-K)}{F^2} \right) \right], \end{aligned} \tag{8.6}$$

where the limit $\epsilon \to 0$ is also assumed. After taking the derivative with respect to k and then the limit $k \to \kappa$, we get for β_2

$$\begin{aligned} \beta_2 = \frac{1}{F'} &\left(\overset{\circ\circ}{F}(-K) - \overset{\circ\circ}{F} \frac{F'(-K)}{F'} \right) \\ &+ \frac{1}{F'} \left(\frac{F}{F'} \right)^2 \left[F''(-K) - F'' \frac{F'(-K)}{F'} \right] \\ &- \frac{2}{F'} \frac{F}{F'} \left[\mathring{F}'(-K) - \mathring{F}' \frac{F'(-K)}{F'} \right]. \end{aligned} \tag{8.7}$$

The problem is now to find the derivatives of F in the limit $\epsilon \to 0$. By taking the limit in F' one finds that the dominant power is ϵ^{p-1}, therefore we look for the terms in the bracket which are of the same order of magnitude. Let us discuss calculation of $\overset{\circ\circ}{F}$. The other derivatives are calcuated in a similar manner. By recalling the rule for the derivative of F,[31] we findthat the dominant power in $\overset{=}{F}$ is ϵ^{p-2} if we take derivatives of the l th column and any other from the degenerate subset. In this case we have

$$\overset{\circ\circ}{F} = 2\epsilon^{p-2} h_1 \cdots h_s \lambda_1 \cdots \lambda_p \times \sum_{i=1\neq l}^{p} \frac{1}{\lambda_i} \operatorname{Det} \begin{vmatrix} a_{ii}, & a_{li} \\ a_{il}, & a_{ll} \end{vmatrix} + O(\epsilon^{p-1}), \tag{8.8}$$

where the product of λ_i excludes λ_l. The derivative $\overset{\circ\circ}{F}(-k_0)$ differs from $\overset{\circ\circ}{F}$ in replacing h_0 by $h_0(-k_0) = h_0^*$.

We also find

$$F'(-k_0)/F' = h_0^*/h_0, \tag{8.9}$$

therefore the terms of the order ϵ^{p-2}, in the first bracket in (8.8), cancel, hence

$$\overset{\circ\circ}{F}(-K) - \overset{\circ\circ}{F}(F'(-K)/F') = O(\epsilon^{p-1}). \tag{8.10}$$

The next highest order terms in F come from two sources: (a) taking the derivative of the l th column and any column from the set s, (b) from (8.8) but including the terms of the order ϵ^{p-1}. However, when such derivatives are calculated, we note that, when the subtraction (8.10) is done, all the terms which do not include derivatives of the oth column cancel. Therefore we have

$$\overset{\circ\circ}{F} \sim 2\epsilon^{p-1} h_1 \cdots h_s \lambda_1 \cdots \lambda_p \left[\frac{1}{h_0} \operatorname{Det} \begin{vmatrix} a_{ll}, & a_{lo} \\ a_{ol}, & a_{oo} \end{vmatrix} + \sum_{i=1\neq l}^{p} \frac{1}{\lambda_i h_0} \operatorname{Det} \begin{vmatrix} a_{ii}, & 0, & a_{io} \\ 0, & a_{ll}, & a_{lo} \\ a_{oi}, & a_{ol}, & a_{oo} + k_1 h_0' \end{vmatrix} \right] + O(\epsilon^p). \tag{8.11}$$

Similarly we obtain derivatives $\overset{\circ}{F}'$ and F''. Having now calculated the derivatives, it can be shown that

$$\overset{=}{F} - 2a_{ll}\overset{\circ}{F}' + a_{ll}^2 F'' \sim -2p \frac{a_{ol}a_{lo}}{h_0} h_1 \cdots h_s \lambda_1 \cdots \lambda_p \tag{8.12}$$

and the coefficient β_2 is

$$\beta_2 = (2p/h_0^2)(h_0^* a_{ol} a_{lo} - h_0 a_{ol}^* a_{lo}^*), \tag{8.13}$$

or if A is replaced by J_1

$$\beta_2 = -\frac{p}{2j_0^2} \frac{i}{kK_0} \left[\sum_{s=1}^{p} \int_0^\infty \psi_0 V_{os}' \psi_s T_{sl} \right]^2, \tag{8.14}$$

which is our desired leading term in the residue. Since there are p poles in the cluster we have p residues of type (8.14). They differ from one another in the unitary matrix T.

9. CONCLUSION

In this article we have developed a perturbation theory for Feshbach-type resonances, although this restriction is not essential. We have also treated the problem with the p-fold degeneracy in the unperturbed Hamiltonian, which is of importance in atom-surface scattering, where such a case often arises. We have also shown, on a simple case of two channels, that when the channels are nearly degenerate in the unperturbed Hamiltonian, the radius of convergence of the perturbation series is greatly reduced and the series may not converge. In other words, the nondegenerage perturbation theory cannot be smoothly continued to the degenerate one. This fact is of importance when the resonance cross section is a function of a continuous parameter, e.g., azimuthal angle in the atom-surface scattering, but for some discrete values several bound states in the closed channels of the unperturbed Hamiltonian become degenerate.

ACKNOWLEDGMENT

The author wishes to acknowledge a very fruitful discussion with Professor D. Micha, in particular the part concerning the degeneracy problem in atom-surface scattering.

[1] G. Breit and F. L. Yost, Phys. Rev. **48**, 203 (1935).
[2] G. Breit and E. P. Wigner, Phys. Rev. **49**, 519 (1936).
[3] T. Regge, Nuovo Cimento **14**, 951 (1959).
[4] R. G. Newton, *The Complex j-Plane* (Benjamin, New York, 1964).
[5] S. Bosanac, Mol. Phys. **35**, 1057 (1978).
[6] H. M. Nussenzveig, J. Math. Phys. **10**, 82, 125 (1969).
[7] S. Bosanac, Phys. Rev. A **19**, 125 (1979).
[8] This point is of particular importance in high energy physics and references to related literature can be found in P. D. B. Collins; *Regge Theory and High Energy Physics* (Cambridge U. P., Cambridge, 1977).
[9] M. L. Goldberger and K. M. Watson, *Collision Theory* (Wiley, New York, 1964).
[10] G. Wolken, J. Chem. Phys. **58**, 3047 (1973).
[11] H. Chow and E. D. Thompson, Surf. Sci. **54**, 269 (1976).
[12] S. Bosanac and D. Micha, J. Chem. Phys. (to be published).
[13] G. Breit, Phys. Rev. **58**, 506 (1940).
[14] H. Feshbach, Ann. Phys. (NY) **5**, 357 (1958).
[15] P. L. Kapur and R. Peierls, Proc. R. Soc. (London) Ser. A **166**, 277 (1938).
[16] A. J. F. Siegert, Phys. Rev. **56**, 750 (1939).
[17] E. P. Wigner, Phys. Rev. **70**, 15, 606 (1946).
[18] A. N. Lane and R. G. Thomas, Rev. Mod. Phys. **30**, 257 (1958).
[19] L. Fonda and R. G. Newton, Ann. Phys. (NY) **10**, 490 (1960).
[20] E. Brandes and P. Froelich, Phys. Rev. A **16**, 2207 (1977).
[21] Int. J. Quantum Chem **XIV** (1978), entirely devoted to this subject.
[22] L. Fonda and R. G.Newton, Nuovo Cimento **14**, 1027 (1959).
[23] S. Bosanac, J. Math. Phys. **21**, 1881 (1980).
[24] The formal theory is reviewed by R. G. Newton, *Scattering Theory of Waves and Particles* (McGraw-Hill, New York, 1966), Sec. 16.5.
[25] S. Weinberg, Phys. Rev. **131**, 440 (1963).
[26] R. Jost, Helv. Phys. Acta **20**, 256 (1947).
[27] R. G. Newton, Ann. Phys. (NY) **4**, 29 (1958).
[28] V. deAlfaro and T. Regge, *Potential Scattering* (North-Holland, Amsterdam, 1965).
[29] S. Bosanac, Fizika (Zagreb) **12**, 77 (1980).
[30] R. G. Newton, J. Math. Phys. **2**, 188 (1961).
[31] In the calculation of F and F' we use the rule for the derivative of the determinant: $F' = \Sigma_i \operatorname{Det}(J)'_i$, where the index outside the bracket designates the derivative of the ith column.

Direct and inverse scattering by a sphere of variable index of refraction[a)]

C. Eftimiu
McDonnell Douglas Research Laboratories, St. Louis, Missouri 63166

(Received 25 February 1982; accepted for publication 1 July 1982)

It is shown that in cases with spherical symmetry, a Liouville transformation leads from the wave equation to a Schrödinger-like equation with energy-independent potential. The direct problem can be solved by iteration and the inverse problem by the Marchenko formalism. An exactly solvable example is given.

PACS numbers: 03.80. + r, 03.65.Nk, 02.30. + g

1. INTRODUCTION

A plane (scalar) wave of frequency $\omega = ck$ propagates (say) in the positive direction of the z axis of a system of coordinates and interacts with an object of finite dimensions, characterized by an index of refraction $n(\mathbf{r}) \geqslant 1$. As a result of the interaction, a scattered wave ψ is generated so that the total field at a point $\mathbf{r}$ is of the form

$$\psi(k,\mathbf{r}) = e^{ikz} + \psi_s(k,\mathbf{r}) \tag{1.1}$$

and satisfies everywhere the equation

$$\Delta\psi + k^2 n^2 \psi = 0. \tag{1.2}$$

At large distances from the scatterer, ψ_s is an outgoing spherical wave:

$$\psi_s \underset{r\to\infty}{\sim} A(k,\theta,\phi)\frac{e^{ikr}}{r}, \tag{1.3}$$

where $A(k,\theta,\phi)$, a function of the spherical angles θ and ϕ but not of the radial coordinate $r = |\mathbf{r}|$, is the scattering amplitude whose modulus squared yields the scattering cross section, a quantity directly accessible to measurement.

Simply stated, the direct scattering problem is to determine the scattering amplitude A, when the function n (identically equal to unity outside the domain occupied by the obstacle) is known. The inverse scattering problem consists of determining the function n (and thereby also the shape of the scatterer) from information regarding the scattering amplitude A.

Both problems are important, but their solutions are unevenly developed. While the direct problem has been extensively studied, and is well understood, a comprehensive solution of the inverse problem is yet to be developed.

To date, the most complete results regarding the inverse scattering problem have been obtained for quantum-mechanical scattering by a potential V, when the wave function is a solution of the Schrödinger equation:

$$\Delta\psi + k^2\psi = V\psi, \tag{1.4}$$

primarily in one dimension.[1,2]

At first sight, it would seem that whatever results exist concerning Eq. (1.4) carry over in the study of Eq. (1.2), since the latter can be cast in the form of a Schrödinger equation simply by writing

$$V = -k^2[n^2(\mathbf{r}) - 1]. \tag{1.5}$$

Moreover, in so doing, the problems might appear to be even simpler because, in contrast with the situation commonly arising in quantum mechanics, Eq. (1.2) has no bound-state solutions [i.e., square integrable solutions which, at least for finite-range potentials like (1.5), can correspond only to non-positive values of k^2]. This physically obvious fact has a simple mathematical proof. Indeed, by rewriting Eq. (1.2) as

$$k^2 n^2 \psi^2 = (\text{grad}\,\psi)^2 - \text{div}(\psi\,\text{grad}\,\psi), \tag{1.6}$$

and assuming that both ψ and its gradient are square integrable, it follows that

$$k^2 \int n^2\psi^2\, d^3\mathbf{r} = \int (\text{grad}\,\psi)^2\, d^3\mathbf{r}, \tag{1.7}$$

which clearly requires $k^2 > 0$. (The possibility that $k^2 = 0$ implies that grad $\psi \equiv 0$, and also, since ψ is assumed square integrable, that $\psi \equiv 0$).

However, a closer look at the formalism developed to solve the inverse problem in quantum mechanics reveals that the situation concerning Eq. (1.2) is not a simple transposition of results. The main obstacle is the fact that, in contrast with the assumption made in quantum mechanics, the potential (1.5) is k dependent, in a manner that invalidates most conclusions regarding the analytic properties of the solutions of Eq. (1.4) in quantum mechanics, and that are essential in the solution of the inverse problem.

Nevertheless, it will be shown below that, at least in the one-dimensional spherically symmetric case, the quantum mechanical developments, both for the direct and inverse problems, can still be carried over to the scattering by a sphere of variable index of refraction as described by Eq. (1.2), although not quite in the simplistic manner suggested by the relation (1.5). The key ingredient in the procedure is a Liouville-type transformation leading from Eq. (1.2) to a Schrödinger-like equation with a k-independent potential, as described in Sec. 2. The direct scattering problem is examined in Sec. 3, and the inverse problem is considered in Sec. 4. Throughout the paper, the procedure is illustrated by a specific example. The portions of the text referring to this example are marked with a vertical line.

2. LIOUVILLE TRANSFORMATION

If it is known that n depends only on the radial variable r, Eq. (1.2) separates in spherical coordinates and for the radial dependence of the solution one obtains the reduced equation

[a)] This work was conducted under the McDonnell Douglas Independent Research and Development program.

 0022-2488/82/112140-07$02.50 © 1982 American Institute of Physics

$$\left[\frac{d^2}{dr^2} - \frac{l(l+1)}{r^2} + k^2 n^2(r)\right] u_l(r) = 0 . \tag{2.1}$$

The so-called physical solution of this equation is a solution regular in the origin $r = 0$, and which asymptotically behaves like

$$u_l(r) \underset{r\to\infty}{\sim} e^{i\delta_l} \sin\left(kr - \frac{l\pi}{2} + \delta_l\right), \tag{2.2}$$

where $\delta_l(k)$ are the scattering phase shifts. Correspondingly, the scattering amplitude is

$$A(k,\theta) = \frac{1}{k} \sum_{l=0}^{\infty} (2l+1) \cdot e^{i\delta_l} \cdot \sin\delta_l \cdot P_l(\cos\theta) . \tag{2.3}$$

Rather than investigating the solutions of Eq. (2.1) directly, one subjects it first to the simultaneous change of variable and function (the Liouville transformation[3]):

$$r \to \rho : \rho(r) = \int_0^r n(s)\, ds , \tag{2.4}$$

$$u \to \psi : \psi_l(\rho) = (n(r))^{1/2} u_l(r) . \tag{2.5}$$

Note that

$$\rho(0) = 0 \tag{2.6}$$

and that, if r_0 is the radius of the sphere outside which $n \equiv 1$,

$$\rho(r) = \rho_0 + r - r_0 \quad \text{for } r \geqslant r_0 , \tag{2.7}$$

where

$$\rho_0 = \int_0^{r_0} n(s)\, ds . \tag{2.8}$$

Also,

$$\rho(r) \underset{r\to\infty}{\sim} n(0) \cdot r . \tag{2.9}$$

The transformation (2.4) can of course be inverted:

$$r(\rho) = \int_0^{\rho} \frac{ds}{\nu(s)} , \tag{2.10}$$

where

$$\nu(\rho) = n(r(\rho)) . \tag{2.11}$$

Applying the transformation (2.4)–(2.5) to the Eq. (2.1), one obtains

$$\left[\frac{d^2}{d\rho^2} - \frac{l(l+1)}{R^2(\rho)} + k^2\right] \psi_l(\rho) = V(\rho)\psi_l(\rho) , \tag{2.12}$$

where

$$R(\rho) = \nu(\rho) \cdot r(\rho) \tag{2.13}$$

and

$$V(\rho) = [\nu(\rho)]^{-1/2} \cdot \frac{d^2}{d\rho^2} [\nu(\rho)]^{1/2} \tag{2.14}$$

for the validity of which it is required that $\nu(\rho)$ be twice differentiable.

It is important to note that the potential $V(\rho)$ vanishes for $\rho > \rho_0$. It will also be assumed to be bounded everywhere.

The Schrödinger-like Eq. (2.12) with the k-independent potential (2.14) constitutes the starting point of all considerations to follow. The only difference between Eq. (2.12) and the usual Schrödinger equation consists in the form of the centrifugal potential which is given in terms of the distance (2.13) rather than simply ρ. Note, however, that

$$R(\rho) = \rho + r_0 - \rho_0 \quad \text{for } \rho \geqslant \rho_0 \tag{2.15}$$

and that

$$R(\rho) \underset{r\to\infty}{\sim} \rho \tag{2.16}$$

if, as we assume, $\nu(\rho)$ and its first derivative are everywhere continuous.

To illustrate the general formalism, the special case of a sphere of radius r_0 and an index of refraction

$$n(r) = \begin{cases} \left[1 - \gamma^2 \left(\frac{r - r_0}{r_0}\right)^2\right]^{-1}, & r \leqslant r_0 \\ 1, & r \geqslant r_0 \end{cases} \tag{2.17}$$

will be treated throughout the paper. In (2.17), γ^2 is an arbitrary parameter except for the restriction

$$0 < \gamma^2 < 1 . \tag{2.18}$$

In this case, (2.4) yields

$$\rho(r) = \begin{cases} \rho_0 + \frac{r_0}{\gamma} \tanh^{-1}\gamma \frac{r - r_0}{r_0}, & 0 \leqslant r \leqslant r_0, \\ r + \rho_0 - r_0 & r_0 \leqslant r, \end{cases} \tag{2.19}$$

with

$$\rho_0 = \rho(r_0) = (r_0/\gamma) \tanh^{-1}\gamma . \tag{2.20}$$

Given (2.19), the inverse transformation is

$$r(\rho) = \begin{cases} r_0 + \frac{r_0}{\gamma} \tanh \frac{\gamma}{r_0}(\rho - \rho_0) & (0 \leqslant \rho \leqslant \rho_0), \\ \rho + r_0 - \rho_0 & (\rho_0 \leqslant \rho) \end{cases} \tag{2.21}$$

and

$$\nu(\rho) = \begin{cases} \cosh^2 \frac{\gamma}{r_0}(\rho - \rho_0) & (0 \leqslant \rho \leqslant \rho_0) \\ 1 & (\rho_0 \leqslant \rho) . \end{cases} , \tag{2.22}$$

Finally, in this case, from (2.14)

$$V(\rho) = \begin{cases} \gamma^2/r_0^2 & (0 \leqslant \rho < \rho_0) , \\ 0 & (\rho_0 < \rho) . \end{cases} \tag{2.23}$$

3. THE DIRECT PROBLEM

In the direct problem the objective is to determine the phase shifts $\delta_l(k)(l = 0,1,2, \cdots)$, and hence the scattering amplitude (2.3), when the scatterer, i.e., the function $n(r)$, is known. As indicated, one way to proceed is to find the physical solution of Eq. (2.12) and examine its asymptotic form to identify the phase shifts. A related solution the so-called regular solution, will be sought initially. This solution of Eq. (2.12) is defined through its behavior in the origin, which is prescribed to be

$$\phi_l(\rho) \underset{\rho\to 0}{\sim} \frac{\rho^{l+1}}{(2l+1)!!} . \tag{3.1}$$

To construct this solution, one customarily writes for ϕ_l an integral equation, the solution of which satisfies both Eq. (2.12) and the condition (3.1). Equation (2.12) is transcribed in the form

$$\left[\frac{d^2}{d\rho^2} - \frac{l(l+1)}{\rho^2} + k^2\right] \phi_l(\rho) = V_l(\rho)\, \phi_l(\rho) , \tag{3.2}$$

where

$$V_l(\rho) = V(\rho) - l(l+1)\left[\frac{1}{\rho^2} - \frac{1}{R^2(\rho)}\right], \tag{3.3}$$

and the Green's function

$$G_l(k,\rho,\rho') = \begin{cases} g_l(k,\rho,\rho'), & \rho' < \rho, \\ 0, & \rho' > \rho \end{cases} \tag{3.4}$$

is introduced, where

$$\begin{aligned} g_l(k,\rho,\rho') = {} & \tfrac{1}{2} i(-1)^l k\rho\rho' \\ & \times [h_l^{(1)}(k\rho)\cdot h_l^{(1)}(-k\rho') \\ & - h_l^{(1)}(k\rho')\cdot h_l^{(1)}(-k\rho)] . \end{aligned} \tag{3.5}$$

The integral equation for the regular solution is thus

$$\begin{aligned} \phi_l(k,\rho) = {} & \rho k^{-l} j_l(k\rho) \\ & + \int_0^\rho g_l(k,\rho,\rho')\cdot V_l(\rho')\,\phi_l(k,\rho')d\rho' , \end{aligned} \tag{3.6}$$

in which the free term is a solution of Eq. (3.2) without the potential V_l, and which has the behavior (3.1).

Equation (3.6), however, need not be solved for all ρ. Indeed, if one introduces the notation

$$\phi_l(k,\rho) = \begin{cases} \phi_l^I(k,\rho) & (0 \leqslant \rho \leqslant \rho_0), \\ \phi_l^{II}(k,\rho) & (\rho_0 \leqslant \rho), \end{cases} \tag{3.7}$$

and remembers that $V(\rho) = 0$ for $\rho > \rho_0$, one can write for $\phi_l^{II}(k,\rho)$ the equation

$$\left(\frac{d^2}{d\rho^2} - \frac{l(l+1)}{(\rho - \rho_0 + r_0)^2} + k^2\right)\phi_l^{II}(k,\rho) = 0 \quad (\rho \geqslant \rho_0), \tag{3.8}$$

the most general solution of which is

$$\phi_l^{II}(k,\rho) = A_l\, r j_l(kr) + B_l\, r y_l(kr) , \tag{3.9}$$

where r stands for $\rho - \rho_0 + r_0$, the expression of $r(\rho)$ for $\rho \geqslant \rho_0$. The constants A_l and B_l are to be found by matching ϕ_l^{II} (and its first derivative) with ϕ_l^I (and its first derivative) at $\rho = \rho_0$, after Eq. (3.6) has been solved for $\rho \leqslant \rho_0$. The expression (3.9) also yields the asymptotic form of the regular solution:

$$\begin{aligned} \phi_l(k,\rho) \underset{\rho\to\infty}{\sim} {} & \frac{1}{k}\left[A_l \sin\left(kr - \frac{l\pi}{2}\right)\right. \\ & \left. - B_l \cos\left(kr - \frac{l\pi}{2}\right)\right] \\ = {} & C_l \sin\left(kr - \frac{l\pi}{2} + \delta_l\right), \end{aligned} \tag{3.10}$$

where

$$\tan\delta_l = -B_l/A_l \tag{3.11}$$

and

$$C_l = (1/k)\,[A_l^2 + B_l^2]^{1/2} . \tag{3.12}$$

The existence of a unique solution $\phi_l^I(k,\rho)$ of Eq. (3.6) for $\rho \leqslant \rho_0$ follows from standard arguments,[4] and requires that the potential $V_l(\rho)$ satisfy the condition

$$\int_0^{\rho_0} d\rho\cdot\rho\,|V_l(\rho)| < \infty . \tag{3.13}$$

This condition is easily seen to be fulfilled in the present case, since $V(\rho)$ is, by assumption, bounded and

$$\frac{1}{\rho^2} - \frac{1}{R^2(\rho)} \underset{\rho\to 0}{\sim} \frac{n'(0)}{n(0)}\frac{1}{\rho} . \tag{3.14}$$

The fact that the same notation for phase shifts has been used in (3.10) as in (2.2) and (2.3) requires an explanation. The phase shifts δ_l defined by (3.11) are *not* the phase shifts of the quantum mechanical scattering problem. Indeed, the latter are defined through the asymptotic form of the physical solution for this problem

$$\psi_l(k,\rho) \underset{\rho\to\infty}{\sim} e^{i\bar\delta_l}\cdot\sin(k\rho - \tfrac{1}{2}l\pi + \bar\delta_l) . \tag{3.15}$$

Since $\psi_l(\rho)$ is also regular, and a regular solution is essentially unique, it follows that

$$\psi_l(k,\rho) = C_l^{-1} e^{i\bar\delta_l}\phi_l(k,\rho) \tag{3.16}$$

and

$$\bar\delta_l(k) = \delta_l(k) - k(\rho_0 - r_0) . \tag{3.17}$$

Up to this point, neither δ_l [as defined in (3.10)–(3.11)] nor $\bar\delta_l$ [as defined in (3.16)–(3.17)] have been given any physical meaning. The quantum mechanical problem described by Eq. (3.2) is, in this context, only an auxiliary mathematical tool: no experiments can be carried out in the space in which ρ is the radial distance. However, the regular and physical solutions of the real problem, described by Eq. (2.1), can be found easily by inverting the Liouville transformation (2.4) and (2.5). Indeed, the regular solution $v_l(k,r)$ of Eq. (2.1), which is defined through the behavior in the origin $r = 0$,

$$v_l(k,r) \underset{r\to 0}{\sim} \frac{r^{l+1}}{(2l+1)!!} , \tag{3.18}$$

is

$$v_l(k,r) = [n(r)]^{-1/2}[n(0)]^{-(l-1/2)}\,\phi_l(k,\rho(r)) , \tag{3.19}$$

while the physical solution of Eq. (2.1) is

$$u_l(k,r) = C_l^{-1}\,[n(0)]^{l+1/2}\,e^{i\delta_l}\,v_l(k,\rho(r)) \tag{3.20}$$

and has the asymptotic form (2.2). This relation identifies δ_l as the real phase shifts of the problem and justifies the use of the same notation in (2.2) and (3.10). The relation (3.17) still warrants additional comments, which will evolve from the ensuing discussion of the Jost solution. For Eq. (2.12), this solution is defined through the boundary condition

$$f_l(k,\rho) \underset{\rho\to\infty}{\sim} e^{i(k\rho + l\pi/2)} \quad (k \neq 0) . \tag{3.21}$$

The integral equation incorporating this condition is

$$f_l(k,\rho) = w_l(u\rho) - \int_\rho^\infty g_l(k,\rho,\rho')\cdot V_l(\rho')f_l(k,\rho')\cdot d\rho' , \tag{3.22}$$

where

$$w_l(k\rho) = ie^{i\pi l}\,k\rho\cdot h_l^{(1)}(k\rho) , \tag{3.23}$$

and is a solution of Eq. (2.12) without the potential term, satisfying (3.21).

Again, Eq. (3.22) need not be solved for all ρ. If one writes

$$f_l(k,\rho)=\begin{cases} f_l^I(k,\rho) & (0<\rho\leqslant\rho_0),\\ f_l^{II}(k,\rho) & (\rho_0\leqslant\rho),\end{cases} \tag{3.24}$$

one sees that

$$f_l^{II}(k,\rho)=ie^{i\pi l}\cdot kr\cdot h_l^{(1)}(kr)\cdot e^{ik(\rho_0-r_0)} \tag{3.25}$$

with $r=\rho-\rho_0+r_0$, the expression of $r(\rho)$ for $\rho\geqslant\rho_0$. To find $f_l(k,\rho)$, one needs only to solve Eq. (3.22) for $f_l^I(k,\rho)$,

$$f_l^I(k,\rho)=f_l^{II}(k,\rho_0)+w_l(k\rho)-w_l(k\rho_0)-\int_\rho^{\rho_0} g_l(k,\rho,\rho')\cdot V_l(\rho')\cdot f_l^I(k,\rho')\,d\rho' \ . \tag{3.26}$$

The existence of a unique solution of Eq. (3.26) again follows if

$$\int_0^{\rho_0}\rho'|V_l(\rho')|e^{(|\mathrm{Im}\,k|-\mathrm{Im}\,k)\rho'}\,d\rho'<\infty\ , \tag{3.27}$$

a condition which is indeed fulfilled if $\mathrm{Im}\,k\geqslant 0$ since (3.13) holds. It should also be noted that (3.13) holds not only for the first moment of the potential, but *a fortiori* for any higher moment. One can thus conclude that $f_l(k,\rho)$ exists for all $\rho>0$ and all $k\neq 0$ in the upper k half-plane, where it is also analytic. As $|k|\to\infty$ in this region, $f_l(k,\rho)$ behaves like

$$f_l(k,\rho)\underset{|k|\to\infty}{\sim} e^{i(k\rho+l\pi/2)}\quad(\rho>0)\,. \tag{3.28}$$

The Jost solution $\mathscr{f}_l(k,r)$ for the Eq. (2.1), defined through the asymptotic condition

$$\mathscr{f}_l(k,r)\underset{r\to\infty}{\sim} e^{i(kr-l\pi/2)}\,, \tag{3.29}$$

can be easily constructed:

$$\mathscr{f}_l(k,r)=[n(r)]^{-1/2}f_l(k,\rho(r))\,e^{-ik(\rho_0-r_0)}\,. \tag{3.30}$$

However, as $|k|\to\infty$ in $\mathrm{Im}\,k>0$ with $r>0$,

$$\mathscr{f}_l(k,r)\underset{|k|\to\infty}{\sim}[n(r)]^{-1/2}\,i^l\,e^{ik[\rho(r)-\rho_0+r_0]}\,. \tag{3.31}$$

It follows thus that, since

$$\rho_0-r_0=\int_0^{r_0}[n(s)-1]\,ds>0\,, \tag{3.32}$$

the Jost solution for the Eq. (2.1) does not have a domain of analyticity in the k plane independent of r. As will become clear in the following section, this circumstance is basically the reason for the failure of the direct application of the quantum mechanical inverse scattering formalism to Eq. (2.1).

Finally, the Jost function for the Eq. (2.12), defined as the Wronskian

$$F_l(k)=(-k)^l\,W\{f_l(k,\rho),\phi_l(k,\rho)\} \tag{3.33}$$

can be evaluated by substituting for f_l and ϕ_l their asymptotic ($k\neq 0$, $\rho\to\infty$) expressions:

$$F_l(k)=|F_l(k)|e^{-i\delta_l(k)} \tag{3.34}$$

with

$$|F_l(k)|=k^{l+1}\,C_l\,. \tag{3.35}$$

The integral representation

$$F_l(k)=1+(-k)^l\int_0^\infty w_l(k\rho)V_l(\rho)\,\phi_l(k,\rho)\,d\rho \tag{3.36}$$

yields the bound

$$|F_l(k)-1|<C\int_0^\infty\frac{\rho}{1+|k|\rho}\,|V_l(\rho)|e^{(|\mathrm{Im}\,k|-\mathrm{Im}\,k)\rho}\,d\rho\,, \tag{3.37}$$

which in turn shows, since

$$|V_l(\rho)|\underset{\rho\to\infty}{\sim}\mathscr{O}(\rho^{-3})\,, \tag{3.38}$$

that

$$\lim_{|k|\to\infty}F_l(k)=1\quad(\mathrm{Im}\,k\geqslant 0)\,, \tag{3.39}$$

i.e. (modulo 2π),

$$\lim_{k\to\infty}\bar{\delta}_l(k)=0\quad(k\ \text{real})\,. \tag{3.40}$$

This is a physically natural circumstance in quantum mechanics: as the energy of the incident beam increases, the effect of the force represented by the (energy-independent) potential weakens, and in the limit $k\to\infty$, the presence of such a force is totally ignored, and the incident beam is no longer scattered. The relation (3.17) shows, however, that this is no longer the case for the scattering described by Eq. (2.1) or (1.2). Indeed, as $k\to\infty$, $\delta_l(k)$ increases (linearly). This is clearly due to the fact that, for the problem described by Eq. (1.2), as k increases, the force represented by the potential (1.5) also increases, and the incident wave never escapes the effects of the force.

In any event, it is worth emphasizing that one can still obtain solutions to Eq. (2.1) by iterations, provided these are conducted on equations like (3.6) or (3.22) for which they are guaranteed to converge.

For $l=0$, the special case given by (2.17) is exactly solvable. In this case,

$$\phi_0(k,\rho)=\begin{cases}\alpha^{-1}\sin\alpha\rho & (0\leqslant\rho\leqslant\rho_0),\\ C_0\sin(kr+\delta_0) & (\rho_0\leqslant\rho),\end{cases} \tag{3.41}$$

where

$$\alpha^2=k^2-\gamma^2/r_0^2 \tag{3.42}$$

and

$$C_0=[k^{-2}\cos^2\alpha\rho_0+\alpha^{-2}\sin^2\alpha\rho_0]^{1/2}\,, \tag{3.43}$$

$$\tan\delta_0=\frac{k\tan\alpha\rho_0-\alpha\tan kr_0}{k\tan\alpha\rho_0\tan kr_0+\alpha}\,. \tag{3.44}$$

From the last expression, one can check directly that

$$\delta_0(k)\underset{k\to\infty}{\sim}k(\rho_0-r_0)\,. \tag{3.45}$$

The physical solution is

$$\psi_0(k,\rho)=C_0^{-1}\,e^{i\bar{\delta}_0}\,\phi_0(k,\rho) \tag{3.46}$$

with C_0 given by (3.43) and $\bar{\delta}_0$ by (3.17) and (3.44):

$$\bar{\delta}_0=\tan^{-1}((k/\alpha)\tan\alpha\rho_0)-k\rho_0\,. \tag{3.47}$$

Again, one can check directly that $\bar{\delta}_0$ vanishes in the limit $k\to\infty$, like k^{-1}.

The Jost solution is

$$f_0(k,\rho) = \begin{cases} e^{ik\rho_0}\,[\cos\alpha(\rho-\rho_0) + (ik/\alpha)\sin\alpha(\rho-\rho_0)] & (0\leqslant\rho\leqslant\rho_0)\,, \\ e^{ik\rho}, & (\rho_0\leqslant\rho) \end{cases} \tag{3.48}$$

and the Jost function is

$$F_0(k) = e^{ik\rho_0}\,(\cos\alpha\rho_0 - (ik/\alpha)\sin\alpha\rho_0)\,. \tag{3.49}$$

The corresponding exact solutions of the equation (2.1) for $l = 0$ can be found from (3.19), (3.20), and (3.30), by using (2.19).

For $l \neq 0$, the problem is no longer exactly solvable but, as mentioned above, it still can be solved by iteration.

4. THE INVERSE PROBLEM

In this problem it is assumed that it is known that the scatterer has the shape of a sphere with a radially distributed index of refraction, but neither its radius r_0 nor the function $n(r)$ is given. What is assumed given is some information regarding the scattering amplitude $A(k,\theta)$, and this information need not be total: one can assume that $A(k,\theta)$ is known only at one particular value of k for all values of θ, which implies that all phase shifts $\delta_l(k)(l = 0,1,\cdots)$ are known for a fixed k, or one can assume that one phase shift $\delta_l(k)$ is known for all values of k. Only the latter assumption will be discussed in this paper.

The assumption that one $\delta_l(k)$ is known refers to one of the phase shifts defined in (2.2). In the quantum mechanical formalism, the corresponding phase shift $\bar{\delta}_l(k)$, as shown in the previous section, can be calculated from the formula

$$\bar{\delta}_l(k) = \delta_l(k) - k\lim_{s\to\infty}\,[s^{-1}\delta_l(s)]\,. \tag{4.1}$$

Since excellent comprehensive descriptions[1,2] of the inverse scattering problem in Quantum Mechanics exist, only a summary of the salient points will be given here. Moreover, the approach proposed by Karlsson[5] will be adopted, since it emphasizes the analytic properties of Jost's solution in the complex k plane.

One begins by using the differential equation satisfied by the functions $k\rho j_l(k\rho)$ and $w_l(k\rho)$ to show that

$$\int_\rho^\infty w_l(k\rho')\,j_l(k'\rho')d\rho' = \frac{w_l(k\rho)\dfrac{d}{d\rho}\,[k'\rho j_l(k'\rho)] - k'\rho j_l(k'\rho)\dfrac{d}{d\rho}\,w_l(k\rho)}{k'^2 - (k+i\epsilon)^2}\,, \tag{4.2}$$

and then considers the integral

$$\mathcal{F}_l(k,\rho) = \frac{(-1)^l}{\pi i}\int_C dk'\,f_l(k',\rho) \times\int_\rho^\infty d\rho' w_l(k\rho')k'\rho' j_l(k'\rho')\,, \tag{4.3}$$

where the closed contour of integration, C, consists of the real axis and a semicircle of indefinitely large radius in the upper half-plane. The integral (4.3) is evaluated in two different ways.

First, because the integrand is an analytic function in the upper half-plane, except for a pole at $k + i\epsilon$, the theorem of residues can be employed. Since at his pole the numerator in (4.2) becomes a Wronskian, one obtains immediately the result

$$\mathcal{F}_l(k,\rho) = f_l(k,\rho)\,. \tag{4.4}$$

Second, the contributions from the real axis and semicircular portions of C are separated, viz.

$$\mathcal{F}_l(k,\rho) = \mathcal{F}_l^{(1)}(k,\rho) + \mathcal{F}_l^{(2)}(k,\rho)\,, \tag{4.5}$$

and these contributions are evaluated separately. Because

$$k\rho j_l(k\rho) = (i/2)\,[w_l(-k\rho) - (-1)^l w_l(k\rho)]\,, \tag{4.6}$$

the contribution of the real axis portion can be written as

$$\mathcal{F}_l^{(1)}(k,\rho) = \frac{(-1)^l}{2\pi}\int_{-\infty}^{\infty} dk' \times[f_l(-k',\rho) - (-1)^l f_l(k',\rho)] \times\int_\rho^\infty d\rho'\,w_l(k\rho')w_l(k'\rho')\,. \tag{4.7}$$

Using the relation between the Jost solution and the physical solution,[1]

$$\psi_l(k,\rho) = (i/2)\,[f_l(-k,\rho) - (-1)^l e^{2i\bar{\delta}_l(k)} f_l(k,\rho)]\,, \tag{4.8}$$

one obtains

$$\mathcal{F}_l^{(1)}(k,\rho) = \frac{1}{2\pi}\int_{-\infty}^{\infty} dk'\,[e^{2i\bar{\delta}_l(k')} - 1]\,f_l(k',\rho) \times\int_\rho^\infty d\rho'\,w_l(k\rho')w_l(k'\rho') + \frac{(-1)^l}{\pi i}\int_{-\infty}^{\infty} dk'\psi_l(k',\rho) \times\int_\rho^\infty d\rho' w_l(k\rho')w_l(k'\rho')\,. \tag{4.9}$$

Because the integrand of the term containing the physical solution is an entire function of k' (there are no bound states in this problem), this integral equals its negative evaluated along the semicircle portion of C. Thus it can be combined with $\mathcal{F}_l^{(2)}$ and evaluated by using the asymptotic forms of the solutions for large k in Im $k \geqslant 0$. The combined result is $w_l(k\rho)$. Thus, the integral equation

$$f_l(k,\rho) = w_l(k\rho) + \frac{1}{2\pi} \times\int_{-\infty}^{\infty} dk'\,[e^{2i\delta_l(k')} - 1]\,f_l(k',\rho) \times\int_\rho^\infty d\rho' w_l(k\rho')w_l(k'\rho')\,, \tag{4.10}$$

is obtained which, in principle, solves the inverse scattering quantum mechanical problem: given $\bar{\delta}_l(k)$, one solves Eq. (4.10) for the Jost solution $f_l(k,\rho)$ which, when substituted in

Eq. (3.2), allows immediate identification of the potential $V_l(\rho)$.

However, Eq. (4.10) has the unpleasant feature of being singular, a circumstance which, when possible, is preferably avoided. In the present case, it can be avoided by recasting the solution according to the so-called Marchenko formulation. By introducing the function

$$A_l(\rho,\sigma)=\frac{(-1)^l}{\pi i}\int_{-\infty}^{\infty}[f_l(k,\rho)-w_l(k,\rho)]k\sigma j_l(k\sigma)\,dk \tag{4.11}$$

and taking into account the bounds

$$|f_l(k,\rho)-w_l(k,\rho)|<C\left(\frac{1+|k|\rho}{|k|\rho}\right)^l e^{-\rho\,\mathrm{Im}\,k}, \tag{4.12}$$

$$|k\sigma j_l(k\sigma)|<e^{\sigma\,\mathrm{Im}\,k},$$

one recognizes that

$$A_l(\rho,\sigma)=0 \quad \text{for } \rho>\sigma. \tag{4.13}$$

The integral representation

$$f_l(k,\rho)=w_l(k\rho)+\int_{\rho}^{\infty}A_l(\rho,\sigma)w_l(k\sigma)\,d\sigma \tag{4.14}$$

is obtained by taking the Hankel transform of Eq. (4.11). Substitution of this representation in Eq. (4.10) yields the Marchenko equation

$$A_l(\rho,\sigma)=B_l(\rho,\sigma)+\int_{\rho}^{\infty}d\tau\,A_l(\rho,\tau)B_l(\tau,\sigma), \tag{4.15}$$

where

$$B_l(\rho,\sigma)=\frac{1}{2\pi}\int_{-\infty}^{\infty}dk\,[e^{2i\bar{\delta}_l(k)}-1]w_l(k\rho)w_l(k\sigma). \tag{4.16}$$

The last step in the solution consists of establishing a connection between $A_l(\rho,\sigma)$ and the potential $V_l(\rho)$. This step can be done directly by substituting the integral representation (4.14) in the integral equation (3.22) satisfied by the Jost solution. For $l=0$, one obtains the integral equation

$$A_0(\rho,\sigma)=\frac{1}{2}\int_{(\rho+\sigma)/2}^{\infty}V_0(\tau)\,d\tau+\int_{(\rho+\sigma)/2}^{\infty}d\tau \times\int_0^{(\sigma-\rho)/2}V_0(\sigma-\omega)A_0(\sigma-\omega,\sigma+\omega)\,d\omega\ (\sigma\geqslant\rho), \tag{4.17}$$

which yields the simple relation

$$V_0(\rho)=-2\frac{d}{d\rho}A_0(\rho,\rho). \tag{4.18}$$

For $l\neq 0$, the equation corresponding to Eq. (4.17) still can be found, but it has a considerably more complicated expression (for $l=0$ the Hankel transform reduces to a Fourier transform). Nevertheless, one still can show that, for $\rho=\sigma$, this relation simplifies drastically and yields, for any l, the result

$$V_l(\rho)=-2\frac{d}{d\rho}A_l(\rho,\rho). \tag{4.19}$$

A simple derivation of this result was given by Blažek.[6]

In the Marchenko formulation, solution of the quantum mechanical inverse problem proceeds as follows: given $\bar{\delta}_l(k)$, one first evaluates the integral (4.16) and then solves the Volterra-type Marchenko equation (4.15). Finally, the potential $V_l(\rho)$ is calculated from (4.19).

Solving the quantum mechanical problem does not yet solve the problem under consideration; one still must find $n(r)$ from a given $V_l(\rho)$. For $l=0$ and $V_l(\rho)=V(\rho)$, $\nu(\rho)$ can be found by solving the equation

$$\frac{d^2}{d\rho^2}\mu(\rho)-V(\rho)\mu(\rho)=0, \quad \mu(\rho)=[\nu(\rho)]^{1/2}, \tag{4.20}$$

which is, in principle, manageable. However, for arbitrary l, viewed as an equation for $\nu(\rho)$, Eq. (3.3) has a formidable aspect. Not only is it integrodifferential, it is also strongly nonlinear. Nothing is known at the present time about whether, given $V_l(\rho)$, Eq. (3.3) has a solution, whether it is unique, or, as it should be independent of l. But even if the answers to such questions were favorable in all respects, one would still have to construct the solution. To this end, the customary approach is by iteration, and one could think of at least two such procedures. One procedure would start from the origin $\rho=0$, assuming that in the first approximation

$$R^{(1)}(\rho)=\rho \quad \text{for all } \rho\geqslant 0. \tag{4.21}$$

Equation (3.3) then becomes

$$\left[\frac{d^2}{d\rho^2}-V_l(\rho)\right][\nu^{(1)}(\rho)]^{1/2}=0 \tag{4.22}$$

and, with the condition $\nu^{(1)}(\rho)\to 1$ as $\rho\to\infty$, yields $\nu(\rho)$ in the first approximation. Hence, via Eq. (2.13), one obtains the second approximation for $R(\rho)$:

$$R^{(2)}(\rho)=\nu^{(1)}(\rho)\int_0^{\rho}\frac{ds}{\nu^{(1)}(s)} \tag{4.23}$$

and the process is repeated.

Another procedure would start from large values of ρ, for which it is known that $V(\rho)=0$. To a first approximation one would take then

$$V^{(1)}(\rho)=0 \quad \text{for all } \rho\geqslant 0. \tag{4.24}$$

With given $V_l(\rho)$, Eq. (3.3) would yield, thus, a first approximation $R^{(1)}(\rho)$. Substitution of this approximation in the relation [derived immediately below, Eq. (4.27)]

$$\nu^{(2)}(\rho)=\exp\left\{-\int_{\rho}^{\infty}\frac{R^{(1)\prime}(s)-1}{R(s)}\,ds\right\} \tag{4.25}$$

leads, via Eq. (2.14), to the next approximation for $V(\rho)$, and the process is repeated. Naturally, one would need to show that such iterations are convergent.

The construction problem could be significantly simplified if the potential $V_l(\rho)$ were to be determined for two values of l, rather than just one. In this case, writing Eq. (3.3) for each l, and subtracting, results in an immediate determination of the function $R(\rho)$. Knowledge of $R(\rho)$ directly implies knowledge of $\nu(\rho)$. Indeed, differentiating (2.13) and taking into account (2.11) yields the equation

$$\frac{\nu'}{\nu}=\frac{R'-1}{R} \quad \left[\lim_{\rho\to\infty}\nu(\rho)=1\right], \tag{4.26}$$

the solution of which is

$$\nu(\rho) = \exp\left\{ -\int_{\rho}^{\infty} \frac{R'(s) - 1}{R(s)}\, ds \right\}. \tag{4.27}$$

Suppose, for example, that $\bar{\delta}_l(k)$ is given for $l = 0$ by (3.47), with unspecified physical meaning of the parameters kr_0, $k\sigma_0$, and γ^2. Then, the integral (4.16), which for $l = 0$ reads

$$B_0(\rho,\sigma) = \frac{1}{2\pi} \int_{-\infty}^{\infty} dk\, [e^{2i\bar{\delta}_0(k)} - 1]\, e^{ik(\rho+\sigma)}, \tag{4.28}$$

needs to be calculated first. This calculation can be done by expanding the integrand in powers of γ^2 and finding B_0 in the form of a series:

$$B_0(\rho,\sigma;\gamma^2) = B_0(\rho,s;0) + \gamma^2 \times \left[\frac{\partial}{\partial\gamma^2} B_0(\rho,\sigma;\gamma^2)\right]_{\gamma^2=0} + \cdots. \tag{4.29}$$

One obtains successively

$$B_0(\rho,\sigma;0) = 0 \tag{4.30}$$

and

$$\left[\frac{\partial}{\partial\gamma^2} B_0(\rho,\sigma;\gamma^2)\right]_{\gamma^2=0} = \frac{i}{\pi}\int_{-\infty}^{\infty} \frac{dk}{4(kr_0)^2} (\sin 2k\rho_0 - 2k\rho_0) e^{ik(\rho+\sigma)} = \frac{1}{2r_0^2}\left(\rho_0 - \frac{\rho+\sigma}{2}\right)\theta\left(\rho_0 - \frac{\rho+\sigma}{2}\right), \tag{4.31}$$

where $\theta(x)$ is the Heaviside step function. In fact, the entire series can be generated and its sum found:

$$B_0(\rho,\sigma;\gamma^2) = \frac{\gamma^2}{r_0^2}\left(\rho_0 - \frac{\rho+\sigma}{2}\right) \sum_{n=0}^{\infty} \frac{(-1)^n}{n!(n+2)!} \times \left[\frac{\gamma}{r_0}\left(\rho_0 - \frac{\rho+\sigma}{2}\right)\right]^{2n} \theta\left(\rho_0 - \frac{\rho+\sigma}{2}\right) = \left[4\left(\rho_0 - \frac{\rho+\sigma}{2}\right)\right]^{-1} \times J_2\left(\frac{2\gamma}{r_0}\left[\rho_0 - \frac{\rho+\sigma}{2}\right]\right) \times \theta\left(\rho_0 - \frac{\rho+\sigma}{2}\right). \tag{4.32}$$

The solution to the Marchenko equation (4.15) now can be found, again by seeking it in the form of a power series in γ^2:

$$A_0(\rho,\sigma) = \frac{1}{2}\sum_{n=0}^{\infty} \left(\frac{\gamma^2}{r_0^2}\right)^{n+1} \frac{1}{n!(n+1)!} \left(\frac{\sigma-\rho}{2}\right)^n \times \left(\rho_0 - \frac{\rho+\sigma}{2}\right)^{n+1} \theta\left(\rho_0 - \frac{\rho+\sigma}{2}\right) = \frac{\gamma}{2r_0}\left[\frac{\rho_0 - (\rho+\sigma)/2}{(\sigma-\rho)/2}\right]^{1/2} \times I_1\left(\frac{2\gamma}{r_0}\left[\frac{\sigma-\rho}{2}\left(\rho_0 - \frac{\rho+\sigma}{2}\right)\right]^{1/2}\right) \times \theta\left(\rho_0 - \frac{\rho+\sigma}{2}\right). \tag{4.33}$$

Finally, from (4.18) one gets

$$V_0(\rho) = V(\rho) = (\gamma^2/r_0^2)\,\theta(\rho_0 - \rho), \tag{4.34}$$

which identifies ρ_0 as the value of ρ beyond which the potential vanishes.

From Eq. (4.20), for $\rho > \rho_0$, with the condition that $\lim_{\rho\to\infty} \mu(\rho) = 1$, one obtains $\mu(\rho) \equiv 1$. For $0 \leqslant \rho \leqslant \rho_0$, the solution of (4.20) can be found easily and matched at $\rho = \rho_0$ (together with its derivative) with the solution found for $\rho > \rho_0$. The final answer is Eq. (2.22), as expected. Equation (2.10) determines $r(\rho)$, including $r_0 = r(\rho_0)$ which now is identified as the radius of the sphere. Inverting $r(\rho)$, one finally finds $n(r) = \nu(\rho(r))$, Eq. (2.17).

[1] R. G. Newton, *Scattering Theory of Waves and Particles* (McGraw-Hill, New York, 1966).

[2] K. Chadan and P. C. Sabatier, *Inverse Problems in Quantum Scattering Theory* (Springer-Verlag, New York, 1977).

[3] R. Courant and D. Hilbert, *Methods of Mathematical Physics* (Interscience, New York, 1963), Vol. I, 63.3. The use of the Liouville transformation in the inverse scattering context appears to have been first proposed by I. Kay and H. Moses, J. Appl. Phys. **27**, 1503 (1956). Two more recent references are: P. Sabatier, J. Math. Phys. **19**, 2410 (1978) and S. Coen, J. Math. Phys. **22**, 1127 (1981).

[4] R. G. Newton, J. Math. Phys. **1**, 319 (1960).

[5] B. R. Karlsson, Phys. Rev. D **10**, 1985 (1974); also in *Applied Inverse Problems*, edited by P. C. Sabatier (Springer-Verlag, New York, 1978), p. 226.

[6] M. Blažek, Commun. Math. Phys. **3**, 282 (1966).

On the nonexistence of static solutions of the Einstein–Weyl equations in general relativity

Charalampos A. Kolassis
Institut Henri Poincaré, Laboratoire de Physique Théorique, Equipe de Recherche Associée au C.N.R.S. NO. 533, 11, rue Pierre et Marie Curie, 75231 Paris Cedex 05, France

(Received 22 April 1981; accepted for publication 7 August 1981)

It is proved that the combined gravitational-neutrino equations in general relativity admit no nontrivial solutions in a static space-time provided that the energy density of the neutrino field is nonnull for all observers.

PACS numbers: 04.20.Cv

I. INTRODUCTION

In general relativity and in classical field theory a neutrino field interacting with a gravitational field is represented by a two-spinor ξ^A which satisfies the generalized Weyl equation

$$\sigma^\mu{}_{A\dot{X}}\xi^A{}_{;\mu}=0, \tag{1.1}$$

and the Einstein gravitational equations

$$R_{\mu\nu}=-T_{\mu\nu}, \tag{1.2}$$

where $\sigma^\mu{}_{A\dot{X}}$ are the generalized Pauli matrices and $T_{\mu\nu}$ is the neutrino energy momentum tensor.

A fruitful manner to treat the above equations is to complete ξ^A with a spinor χ^A in such a way that (ξ^A,χ^A) form a spinor frame. This spinor frame gives rise to a null tetrad $(l^\mu,k^\mu,m^\mu,\bar{m}^\mu)$ as follows:

$$l^\mu=\sigma^\mu{}_{A\dot{X}}\xi^A\bar{\xi}^{\dot{X}}, \tag{1.3a}$$

$$k^\mu=\sigma^\mu{}_{A\dot{X}}\chi^A\bar{\chi}^{\dot{X}}, \tag{1.3b}$$

$$m^\mu=\sigma^\mu{}_{A\dot{X}}\xi^A\bar{\chi}^{\dot{X}}, \tag{1.3c}$$

$$\bar{m}^\mu=\sigma^\mu{}_{A\dot{X}}\chi^A\bar{\xi}^{\dot{X}}. \tag{1.3d}$$

The vector l^μ is interpreted as the neutrino flux vector.

For a given neutrino field ξ^A the null tetrad $(l^\mu,k^\mu,m^\mu,\bar{m}^\mu)$ is not uniquely determined; we can perform a "null rotation about l^μ" expressed by the formulas

$$l^\mu=l'^\mu, \tag{1.4a}$$

$$k^\mu=k'^\mu+\Psi m'^\mu+\bar{\Psi}\bar{m}'^\mu+\Psi\bar{\Psi}l'^\mu, \tag{1.4b}$$

$$m^\mu=m'^\mu+\bar{\Psi}l'^\mu. \tag{1.4c}$$

In terms of the spin coefficients associated with the null tetrad[1] the Weyl equation takes the equivalent form[2]:

$$\rho=\epsilon, \tag{1.5a}$$

$$\beta=\tau. \tag{1.5b}$$

Furthermore, if we will suppose that the neutrino field is of class E_1, we can prove[2] the existence of a null tetrad with respect to which the neutrino energy-momentum tensor assumes the form

$$T_{\mu\nu}=-\tfrac{1}{4}\{\Lambda l_\mu l_\nu+2\omega[g_{\mu\nu}-2(l_\mu k_\nu+l_\nu k_\mu)]+2i(\bar{\sigma}m_\mu m_\nu-\sigma\bar{m}_\mu\bar{m}_\nu)\}, \tag{1.6}$$

with the following restrictions on the spin coefficients:

$$k=0, \tag{1.7}$$

$$\alpha-2\bar{\tau}=0, \tag{1.8}$$

$$4\omega^2-\sigma\bar{\sigma}\geqslant 0. \tag{1.9}$$

The quantities Λ and ω are defined by

$$\Lambda=2i(\bar{\gamma}-\gamma), \tag{1.10}$$

$$\omega=(i/2)(\rho-\bar{\rho}). \tag{1.11}$$

If in particular,

$$4\omega^2-\sigma\bar{\sigma}=0,$$

then we can perform the null rotation (1.4a)–(1.4c) with the function Ψ restricted by the equation

$$2\omega\Psi-i\bar{\sigma}\bar{\Psi}=0, \tag{1.12}$$

and give to the real part of Ψ any convenient value.

The components of the Einstein field equations with respect to the null tetrad can be written[3]

$$\Phi_{00}=0, \tag{1.13a}$$

$$\Phi_{01}=0, \tag{1.13b}$$

$$\Phi_{02}=(i/4)\sigma, \tag{1.13c}$$

$$\Phi_{11}=\tfrac{1}{4}\omega, \tag{1.13d}$$

$$\Phi_{12}=0, \tag{1.13e}$$

$$\Phi_{22}=-\tfrac{1}{8}\Lambda. \tag{1.13f}$$

II. THE NEUTRINO FIELD IN A STATIC SPACE-TIME

Static space-time is characterized by the existence of a timelike hypersurface orthogonal Killing vector field n^μ, that is

$$n^\mu n_\mu>0, \tag{2.1}$$

$$\mathcal{L}_n g_{\mu\nu}=0, \tag{2.2}$$

$$n_\lambda n_{\mu;\nu}+n_\nu n_{\lambda;\mu}+n_\mu n_{\nu;\lambda}=0. \tag{2.3}$$

From these equations it follows that n^μ is an eigenvector of the Ricci tensor $R_{\mu\nu}$ and consequently by virtue of the Einstein field equations (1.2) we must have

$$T_{\mu\nu}n^\mu=\phi n_\nu. \tag{2.4}$$

As in this paper we consider a static space-time generated by a neutrino field of class E_1 we can prove[4] that from (2.2) it

 0022-2488/82/112147-04$02.50 © 1982 American Institute of Physics

follows that

$$\mathcal{L}_n l^\mu = 0, \tag{2.5a}$$

$$\mathcal{L}_n k^\mu = 0, \tag{2.5b}$$

$$\mathcal{L}_n m^\mu = -ism^\mu, \tag{2.5c}$$

where s is a real constant.

To exploit the Eqs. (2.1), (2.3), (2.4), and (2.5a)–(2.5c) we need to expand n^μ in terms of the null tetrad

$$n^\mu = al^\mu + bk^\mu - cm^\mu - \bar{c}\bar{m}^\mu, \tag{2.6}$$

where a, b are real, and c complex.

Now the Eqs. (2.1), (2.4), (2.5a)–(2.5c) and (2.3) can be written equivalently

$$ab - c\bar{c} > 0, \tag{2.7}$$

$$2(\omega - 2\phi)a - b\Lambda = 0, \tag{2.8a}$$

$$(\omega - 2\phi)b = 0, \tag{2.8b}$$

$$(\omega + 2\phi)c - i\bar{c}\bar{\sigma} = 0, \tag{2.8c}$$

$$b_{,\mu} = (l_{\nu;\mu} - l_{\mu;\nu})n^\nu, \tag{2.9a}$$

$$a_{,\mu} = (k_{\nu;\mu} - k_{\mu;\nu})n^\nu, \tag{2.9b}$$

$$\bar{c}_{,\mu} = (m_{\nu;\mu} - m_{\mu;\nu})n^\nu - ism_\mu, \tag{2.9c}$$

$$c\Delta b + b\bar{\delta}a - a\bar{\delta}b + an^\mu\bar{m}^\nu l_{\mu;\nu} - bn^\mu\bar{m}^\nu k_{\mu;\nu} - cn^\nu k^\mu l_{\nu;\mu} = 0, \tag{2.10a}$$

$$b\bar{\delta}\bar{c} - \bar{c}\bar{\delta}b + c\delta b - bn^\mu\bar{m}^\nu m_{\mu;\nu} - cn^\nu m^\mu l_{\nu;\mu} + \bar{c}n^\mu\bar{m}^\nu l_{\mu;\nu} = 0, \tag{2.10b}$$

$$c\delta a - a\delta c - \bar{c}\bar{\delta}a + an^\nu m^\mu\bar{m}_{\nu;\mu} - cn^\nu m^\mu k_{\nu;\mu} + \bar{c}n^\mu\bar{m}^\nu k_{\mu;\nu} = 0, \tag{2.10c}$$

where Δ, δ, and $\bar{\delta}$ are the intrinsic derivatives associated with the null tetrad.[5]

From (2.7) it is clear that we must have $a \neq 0$ and $b \neq 0$ and therefore the system (2.8a)–(2.8c) is equivalent to the system

$$2\phi = \omega, \tag{2.11a}$$

$$\gamma = \bar{\gamma}, \tag{2.11b}$$

$$2\omega c - i\bar{\sigma}\bar{c} = 0. \tag{2.11c}$$

Also, from (1.9), (2.11b) and the form (1.6) of the energy-momentum tensor, it is clear that if $\rho - \bar{\rho} = 0$ then the neutrino field reduces to a ghost field. As it is known[6] these fields are allowed by static space-time configurations, but in the present paper they are not considered. So, we assume

$$\rho - \bar{\rho} \neq 0. \tag{2.12}$$

In the next section we give the proof of the following theorem which appears as a generalization of a theorem proved by Wainwright.[7]

Theorem: There are no nontrivial solutions of the Einstein–Weyl equations in a static space-time for a neutrino field of class E_1.

For convenience, in the remainder of the paper the restrictions (1.5a), (1.5b), (1.7), (1.8), (2.7), (2.11b), (2.12), and the Einstein equations (2.13a)–(2.13f) will be used without explicit reference. For the Ricci and Bianchi identities we will adopt a special reference notation, e.g., by (2.R) [respectively, (2.B)] we mean the second Ricci identity (respectively, the second Bianchi identity) in the listing given by Pirani[8] or by Flaherty.[9]

III. PROOF OF THE THEOREM

First, from (2.11c) we observe that if $4\omega^2 - \sigma\bar{\sigma} > 0$ the quantity c vanishes. But this can also be achieved if $4\omega^2 - \sigma\bar{\sigma} = 0$. In fact, in this case, we can perform the null rotation (1.4a)–(1.4c) with the function Ψ restricted by (1.12). In the new null tetrad $(l'^\mu, k'^\mu, m'^\mu, \bar{m}'^\mu)$ the Killing vector n^μ can be expanded as follows:

$$n^\mu = a'l'^\mu + b'k'^\mu - c'm'^\mu - \bar{c}'\bar{m}'^\mu,$$

where the components a', b', c' are related to a, b, c, by

$$a' = a + b\Psi\bar{\Psi} - c\bar{\Psi} - \bar{c}\Psi, \tag{3.1a}$$

$$b' = b, \tag{3.1b}$$

$$c' = c - b\Psi. \tag{3.1c}$$

Now if $c \neq 0$, we choose

$$\mathrm{Re}\Psi = (1/b)\mathrm{Re}c,$$

and then, from (1.12), (2.11c) and (3.1c) it follows that $c' = 0$. Therefore, in all cases we can set

$$c = 0, \tag{3.2}$$

and assume that the null tetrad is uniquely determined. Introducing (3.2) into Eq. (2.9c) we obtain

$$\bar{\pi} + \tau = 0, \tag{3.3}$$

$$a\sigma = b\bar{\lambda}, \tag{3.4}$$

$$a\bar{\rho} - b\bar{\mu} = is. \tag{3.5}$$

Also, by virtue of (3.2) Eqs. (2.10a)-(2.10c) become

$$b\bar{\delta}a - a\bar{\delta}b + 2abk^\mu\bar{m}^\nu l_{\mu;\nu} = 0,$$

$$n^\mu\bar{m}^\nu m_{\mu;\nu} = 0.$$

These two equations with the help of (2.9a)–(2.9b), (3.3) and the formulas given in the appendix yield

$$\nu = 0, \tag{3.6}$$

$$a\rho = b\bar{\mu}. \tag{3.7}$$

For the calculations which follow it will be helpful if we notice that by virtue of (3.4) and (3.7) the quantities $\lambda\sigma$ and $\rho\mu$ are real. Also, instead of (3.5) we consider in our proof the equation

$$a(\rho - \bar{\rho}) = -is, \tag{3.8}$$

which follows from (3.5) and (3.7).

Equations (2.9a) and (2.9b) with the help of (3.2), (3.3), and (3.6) can be written in the form

$$Da = 2\gamma b, \tag{3.9a}$$

$$\Delta a = -2\gamma a, \tag{3.9b}$$

$$\delta a = -4\tau a, \tag{3.9c}$$

$$Db = (\rho + \bar{\rho})b, \tag{3.10a}$$

$$\Delta b = -(\rho + \bar{\rho})a, \tag{3.10b}$$

$$\delta b = 2\tau b. \tag{3.10c}$$

For convenience, in addition to the conventions adopted at the end of Sec. II, we agree that in the sequel Eqs. (3.3),

(3.4), (3.6)–(3.8), (3.9a)–(3.9c), and (3.10a)–(3.10c) will also be used without explicit reference.

To obtain further restrictions on the spin coefficients and the components of the Weyl conformal tensor we are forced to consider the integrability conditions of the Eqs. (3.4), (3.7), and (3.8) in conjunction with the Ricci and Bianchi identities. So, we will proceed by first examining integrability conditions of the Eq. (3.8), namely, its D, δ, and Δ derivatives.

By D differentiation of (3.8) and the help of (1.R) we obtain

$$\gamma+\mu+\bar{\mu}=0. \tag{3.11}$$

From (3.R), (4.R), and (5.R) we derive

$$\delta(\rho-\bar{\rho})=4\tau(\rho-\bar{\rho})-2\Psi_1, \tag{3.12}$$

and so, the δ derivative of (3.8) yields

$$\Psi_1=0. \tag{3.13}$$

Also, from (6.R) we derive

$$\Delta(\rho-\bar{\rho})=2\gamma(\rho-\bar{\rho})+\bar{\Psi}_2-\Psi_2, \tag{3.14}$$

and so, the Δ derivative of (3.8) yields

$$\Psi_2=\bar{\Psi}_2. \tag{3.15}$$

By taking into account (3.11) and (3.15) we obtain from (12.R)

$$\delta\bar{\tau}=(2\mu-\bar{\mu})\rho-\lambda\sigma+\tau\bar{\tau}-\Psi_2+\Phi_{11}, \tag{3.16}$$

and hence, (8.R) and (17.R) can be written, respectively,

$$D\mu=(\bar{\mu}-3\mu)\rho+2\lambda\sigma+\tau\bar{\tau}-\Phi_{11}+2\Psi_2, \tag{3.17}$$

$$\Delta\rho=-\bar{\rho}\mu-3\rho\bar{\mu}-2\lambda\sigma-\tau\bar{\tau}+\Phi_{11}-2\Psi_2. \tag{3.18}$$

Substituting (3.18) into (6.R) and taking into account (3.11) we obtain

$$D\gamma=4\rho\mu-2\lambda\sigma-2\tau\bar{\tau}+2\Phi_{11}-\Psi_2. \tag{3.19}$$

By D differentiation of (3.11) and the help of (3.15), (3.17), and (3.19) we derive

$$3\Psi_2=(\rho-\bar{\rho})(\mu-\bar{\mu})-2\lambda\sigma. \tag{3.20}$$

Let us now consider the integrability conditions of Eq. (3.4). At first, by D differentiation of (3.4) and the help of (2.R) we obtain

$$D\bar{\lambda}=(\rho-3\bar{\rho})\bar{\lambda}+(a/b)\Psi_0, \tag{3.21}$$

and then, this equation with (7.R), (16.R), and (3.11) gives

$$\Delta\sigma=-2(\mu+\bar{\mu})\sigma-(a/b)\Psi_0. \tag{3.22}$$

With the help of (3.22), (3.11), and the Ricci identity (10.R) which by virtue of (3.11) can be written

$$\Delta\lambda=(\mu+\bar{\mu})\lambda-\Psi_4, \tag{3.23}$$

the Δ differentiation of (3.4) yields

$$(a/b)\Psi_0=(b/a)\bar{\Psi}_4. \tag{3.24}$$

the imaginary parts of (1.R) and (3.17) can be written, respectively,

$$D(\rho-\bar{\rho})=2(\rho+\bar{\rho})(\rho-\bar{\rho}),$$

$$D(\mu-\bar{\mu})=-(\rho+\bar{\rho})(\mu-\bar{\mu}).$$

By virtue of these two equations and Eqs. (2.R) and (3.21) the D derivative of (3.20) becomes

$$3D\Psi_2=3(\rho+\bar{\rho})\Psi_2-2(\lambda\Psi_0+\bar{\lambda}\bar{\Psi}_0). \tag{3.25}$$

The substitution of (3.25) into (3.B) with the help of (3.13), (3.15) and (9.B) gives

$$\lambda\Psi_0=-3\bar{\rho}\Psi_2-2(2\bar{\rho}+\rho)\Phi_{11}+\bar{\sigma}\Phi_{02}. \tag{3.26}$$

From (3.R), (5.R), (11.R), and the help of (3.13) we obtain

$$\bar{\delta}\sigma=4\sigma\bar{\tau}. \tag{3.27}$$

By virtue of this equation the $\bar{\delta}$ derivative of (3.4) yields

$$\delta\lambda=-2\lambda\tau. \tag{3.28}$$

Acting on ρ with the commutator of the D and δ-derivative operators[5] and using (1.R)–(5.R), (3.13), and (3.27) we find

$$\lambda\delta\sigma=4\lambda\sigma\tau-(a/b)\Psi_0\bar{\tau}. \tag{3.29}$$

As by virtue of (3.13) Eq. (3.12) becomes

$$\delta(\rho-\bar{\rho})=4(\rho-\bar{\rho})\tau, \tag{3.30}$$

it follows that

$$\delta(\mu-\bar{\mu})=-2(\mu-\bar{\mu})\tau. \tag{3.31}$$

Now, by virtue of (3.28)–(3.31) the δ derivative of (3.20) can be written

$$3\delta\Psi_2=6\Psi_2\tau+2(a/b)\Psi_0\bar{\tau}. \tag{3.32}$$

From (9.R), (15.R), and (18.R) we derive

$$\Delta\tau=0, \tag{3.33}$$

$$\Psi_3=0, \tag{3.34}$$

and then by substitution of (3.32) into (4.B) and taking into account (3.13), (3.30), (3.34), and (10.B) we obtain

$$2(a/b)\Psi_0\bar{\tau}=3(\Psi_2-2\Phi_{11})\tau+3\Phi_{02}\bar{\tau}. \tag{3.35}$$

By D differentiation of (3.7) and the help of (3.11), (3.17), (3.20), and (1.R) we find

$$3\Phi_{11}=\rho\mu+\rho\bar{\mu}+\bar{\rho}\mu-\lambda\sigma+3\tau\bar{\tau}. \tag{3.36}$$

By virtue of (3.14), (3.15), (3.22)–(3.24), (3.26), and (14.R) the Δ derivative of (3.20) can be written in the form

$$\Delta\Psi_2=-(\mu+\bar{\mu})(3\Psi_2+4\Phi_{11}). \tag{3.37}$$

Let us write down the imaginary part of Eq. (3.26):

$$\lambda\Psi_0-\bar{\lambda}\bar{\Psi}_0=(\rho-\bar{\rho})(3\Psi_2+2\Phi_{11})+2\bar{\sigma}\Phi_{02}.$$

Then, taking the Δ derivative of this equation and using (2.R), (2.B), (3.13)–(3.15), (3.11), (3.22)–(3.24), (3.26), and (3.37) we obtain

$$(i/4)(\lambda\Psi_0+\bar{\lambda}\bar{\Psi}_0)=2(\mu+\bar{\mu})\bar{\sigma}\Phi_{02}+(\bar{\rho}\mu-\rho\bar{\mu})(3\Psi_2+10\Phi_{11}). \tag{3.38}$$

But as from (3.20) and (3.26) it follows that

$$(i/4)(\lambda\Psi_0+\bar{\lambda}\bar{\Psi}_0)=2(\mu+\bar{\mu})\bar{\sigma}\Phi_{02}+2(\rho\bar{\mu}-\bar{\rho}\mu)\Phi_{11}+(3/16)(\rho^2-\bar{\rho}^2),$$

finally Eq. (3.38) yields

$$(\rho+\bar{\rho})(\Psi_2+4\Phi_{11}+(1/16)(b/a))=0. \tag{3.39}$$

Here, we are forced to consider two cases, i.e., $\rho+\bar{\rho}=0$ (nonexpanding case) or $\rho+\bar{\rho}\neq 0$ (expanding case). We will show that both cases contradict the assumption (2.12).

A. Nonexpanding case

We consider the case where

$$\rho + \bar{\rho} = 0. \tag{3.40}$$

Then, from (1.R) follows

$$\rho^2 + \sigma\bar{\sigma} = 0, \tag{3.41}$$

$$D\rho = 0. \tag{3.42}$$

Equation (3.41) can be written equivalently in the form

$$\rho\mu = \lambda\sigma. \tag{3.43}$$

By virtue of (3.40) and (3.43) Eqs. (3.20), (3.36), and (3.26) become

$$3\Psi_2 = 2\rho\mu, \tag{3.44}$$

$$3\Phi_{11} = -2\rho\mu + 3\tau\bar{\tau}, \tag{3.45}$$

$$\lambda\Psi_0 = \rho(3\Psi_2 + \Phi_{11}). \tag{3.46}$$

Also, from (3.R)–(5.R), and the help of (3.13) and (3.40) we derive

$$\rho\tau = \sigma\bar{\tau}. \tag{3.47}$$

By taking into account (3.40), (3.42), (2.R), (3.R), and (3.13) the D derivative of (3.47) yields

$$\Psi_0\bar{\tau} = 0. \tag{3.48}$$

Now, if

$$\tau \neq 0, \tag{3.49}$$

then,

$$\Psi_0 = 0,$$

and so, from (3.44)–(3.46) it follows that

$$3\tau\bar{\tau} = -4(a/b)\rho\bar{\rho},$$

which is in contradiction to (2.7) and (3.49).

On the other hand, if

$$\tau = 0, \tag{3.50}$$

then, from (3.16), (3.40), (3.43)–(3.45) we derive

$$\rho\mu = 0.$$

But by virtue of (3.7) this equation leads to

$$\rho = 0.$$

which contradicts (2.12).

B. Expanding case

We consider the case where

$$\rho + \bar{\rho} \neq 0. \tag{3.51}$$

Hence, from (3.39) it follows that

$$\Psi_2 + 4\Phi_{11} + \tfrac{1}{16}(b/a) = 0. \tag{3.52}$$

By δ differentiation of this equation and the help of (3.30), (3.32), and (3.35) we obtain

$$(2\Phi_{11} + \tfrac{3}{16}(b/a))\tau + \Phi_{02}\bar{\tau} = 0. \tag{3.53}$$

By Δ differentiation of (3.53) and the help of (3.11), (3.14), (3.15), (3.22), and (3.33) we obtain

$$\tfrac{3}{4}(\rho + \bar{\rho})\tau = -i(a/b)\Psi_0\bar{\tau}. \tag{3.54}$$

From (3.35), (3.52)–(3.54) we derive

$$\tau[\tfrac{1}{2}(\rho + \bar{\rho}) - 4i(2\Phi_{11} + (1/16)(b/a))] = 0,$$

and then, by virtue of (3.51)

$$\tau = 0. \tag{3.55}$$

The substitution of (3.55) into (3.16) leads to

$$\rho\bar{\mu} = 2\rho\mu - \lambda\sigma + \Phi_{11} - \Psi_2. \tag{3.56}$$

But, as by virtue of (3.15) the right-hand side of (3.56) is real, we have

$$\rho\bar{\mu} = \bar{\rho}\mu,$$

or, by virtue of (3.7)

$$\rho^2 = \bar{\rho}^2.$$

But, clearly, this last equation is in contradiction to (2.12) and (3.51).

ACKNOWLEDGMENTS

I would like to thank Professor A. Papapetrou and Dr. G. Le Denmat for valuable discussions.

APPENDIX

For convenience, we list the expansions of $l_{\mu;\nu}$, $k_{\mu;\nu}$, and $m_{\mu;\nu}$ on the null tetrad.

$$\begin{aligned} l_{\mu;\nu} = {} & (\epsilon + \bar{\epsilon})l_\mu k_\nu - k\bar{m}_\mu k_\nu - \bar{k}m_\mu k_\nu + (\gamma + \bar{\gamma})l_\mu l_\nu - \tau\bar{m}_\mu l_\nu - \bar{\tau}m_\mu l_\nu \\ & - (\bar{\alpha} + \beta)l_\mu\bar{m}_\nu - (\alpha + \bar{\beta})l_\mu m_\nu + \rho\bar{m}_\mu m_\nu + \bar{\rho}m_\mu\bar{m}_\nu + \sigma\bar{m}_\mu\bar{m}_\nu + \bar{\sigma}m_\mu m_\nu, \\ k_{\mu;\nu} = {} & -(\epsilon + \bar{\epsilon})k_\mu k_\nu + \pi m_\mu k_\nu + \bar{\pi}\bar{m}_\mu k_\nu - (\gamma + \bar{\gamma})k_\mu l_\nu + \nu m_\mu l_\nu + \bar{\nu}\bar{m}_\mu l_\nu \\ & + (\bar{\alpha} + \beta)k_\mu\bar{m}_\nu + (\alpha + \bar{\beta})k_\mu m_\nu - \lambda m_\mu m_\nu - \bar{\lambda}\bar{m}_\mu\bar{m}_\nu - \mu m_\mu\bar{m}_\nu - \bar{\mu}\bar{m}_\mu m_\nu, \\ m_{\mu;\nu} = {} & -kk_\mu k_\nu + \bar{\pi}l_\mu k_\nu - (\bar{\epsilon} - \epsilon)m_\mu k_\nu - \tau k_\mu l_\nu + \bar{\nu}l_\mu l_\nu - (\bar{\gamma} - \gamma)m_\mu l_\nu \\ & + \sigma k_\mu\bar{m}_\nu - \bar{\lambda}l_\mu\bar{m}_\nu + (\bar{\alpha} - \beta)m_\mu\bar{m}_\nu + \rho k_\mu m_\nu - \bar{\mu}l_\mu m_\nu + (\bar{\beta} - \alpha)m_\mu m_\nu. \end{aligned}$$

[1]Readers unfamiliar with the spin coefficient formalism are reffered to the original paper by E. Newman and R. Penrose, J. Math. Phys. **3**, 566 (1962).

[2]J. Wainwright, J. Math. Phys. **12**, 828 (1971); In the present paper, the scalar ϕ associated with the neutrino field [see Eqs. (2.2) and (2.4) in this reference] is chosen so that $\phi = 1$.

[3]For the definition of the components Φ_{ij} of the Ricci tensor with respect to a null tetrad see Ref. 1, p. 570, Eq. (4.3b).

[4]C. A. Kolassis, "On the effect of space-time isometries on the neutrino field," J. Math. Phys. **23**, 1630 (1982). Here, Eq. (2.11b) follows from Eq. (2.1) and Eq. (2.4). Then, a pure radiative neutrino field is excluded. Furthermore, Eq. (2.5b) holds whenever the null tetrad is chosen so as to satisfy Eq. (3.2).

[5]For the definition of the intrinsic derivatives D, Δ, δ, $\bar{\delta}$ and for their commutators acting on scalars see Ref. 1, p. 567, Eqs. (2.12) and p. 570, Eqs. (4.4).

[6]T. M. Davis and J. R. Ray, J. Math. Phys. **16**, 80 (1975).

[7]J. Wainwright, Nuovo Cimento B **22**, 131 (1974).

[8]F. A. E. Pirani, in *Lectures on General Relativity*, 1964 *Brandeis Summer Institute*, edited by S. Deser and K. Ford (Prentice-Hall, Englewood Cliffs, 1965), Vol. 1, pp. 350–351.

[9]E. Flaherty, "Hermitian and Kalherian Geometry in Relativity," in *Lecture Notes in Physics*, edited by J. Ehlers *et al.* (Springer, New York, 1976), pp. 128–132.

Geometrical perturbation theory: action-principle surface terms in homogeneous cosmology

Robert H. Gowdy
Department of Physics, Virginia Commonwealth University, Richmond, Virginia 23284

(Received 4 November 1981; accepted for publication 14 May 1982)

The coordinate-independent formulation of spacetime perturbation theory is applied to the analysis of surface terms which appear in variations of the Einstein Action of general relativity. Restricted variations of the type used to study the dynamics of homogeneous universes require a corrected action functional. A general expression for the correction is found.

PACS numbers: 04.20.Cv, 02.40. – k, 98.80.Dr

1. INTRODUCTION

An action principle can provide a wonderfully simple way to pass from the general formulation of a field theory to a restricted version in which only a few degrees of freedom are allowed. However, there are problems with this approach—it can yield an incorrect action principle.

If the full field equations extremize the action I when all smooth variations of the field are allowed, then an action principle for a restricted theory can be found by evaluating I on a restricted class of fields. For example, evaluating the Einstein action functional on spacetimes with Bianchi type IX homogeneous space sections yields a valid action principle for Misner's mixmaster universes.[1] However, it was soon pointed out that this approach fails for all of the Bianchi types with commutation coefficients which are not trace-free.[2]

The difficulty arises because the action principle that one starts with yields the correct field equations only under a restricted family of variations. The usual restriction is that the field and its derivatives vanish outside of a compact domain so that surface integrals may be discarded from the variations of the action. More general variations give rise to "natural boundary conditions" which cannot always be satisfied. Homogeneous cosmology contemplates a restriction to fields which are the *same* everywhere on a space section and must therefore satisfy the natural boundary conditions *everywhere*. These conditions cannot be obtained from the restricted variation and must be imposed separately. Furthermore, they may impose unacceptable restrictions on the class of solutions. Thus, there is no reason to expect the Einstein Action Principle to work for homogeneous cosmologies with noncompact space sections.[3]

A brief description of perturbation geometry is given in Sec. 2 of this paper. However, it will be necessary to refer back to the previous two papers of this series for a full introduction to the subject.[4,5] Section 3 discusses the first order variation problem and its implications for homogeneous cosmology. Previous discussions of this issue have been cast in 3 + 1-Hamiltonian canonical language. Here a spacetime-geometrical Lagrangian approach is used as a way to discover the geometrical roots of the problem. The corrected action which partially rescues the restricted action principal for Class B models is introduced in Sec. 4. An alternative way to deal with these models as "inherently tilted cosmologies" is suggested in Sec. 5.

2. SPACETIME PERTURBATION THEORY: THE GEOMETRICAL APPROACH

The basic idea of Geometrical Perturbation Theory is to extend the use of the covariant derivative. Because of the Equivalence Principal and its close cousin the Principal of Minimal Coupling, all of the derivatives in the Lagrangian of a meaningful field theory are covariant derivatives. Thus it makes sense to enlarge the spacetime geometry by introducing perturbation parameters as additional dimensions and to take covariant derivatives in perturbation directions.

The enlarged geometry is called a $4 + n$ deformation. It consists of a manifold with four spacetime coordinates x^α and n perturbation parameters $x^{(A)}$. Think of it as a stack of spacetimes each labelled by a set of perturbation parameters. The spacetime metric tensor is represented by a second rank contravariant tensor field γ^{ab}, which is tangent to the spacetimes in the deformation. A choice of perturbation gauge corresponds to the introduction of the projection tensor field $H^a{}_b$, which projects onto the spacetime tangent spaces. Projection in the perturbation directions is accomplished by the identification gauge tensor

$$\iota^a{}_b := \delta^a{}_b - H^a{}_b. \tag{2.1}$$

As usual, derivatives are denoted by indexes preceded by commas for directional derivatives of components, semicolons for spacetime covariant derivatives, and vertical bars for deformation derivatives which have been twice projected onto the spacetime surfaces. A deformation-covariant derivative is denoted by a dot. The full definition of the deformation-covariant derivative can be found in the two previous papers of this series. The properties which one needs to use this type of derivative are

(1) it agrees with the spacetime derivative for vector and tensor fields which are tangent to spacetime,

(2) the derivative of the gauge tensor is just the second fundamental form of spacetime according to

$$H^a{}_{s\cdot b} = h^a{}_{bs}, \tag{2.2}$$

(3) the second fundamental form has the nonzero components

$$h^{\alpha\beta}{}_{(S)} = \tfrac{1}{2}\gamma^{\alpha\beta}{}_{,(S)} \tag{2.3}$$

in a coordinate system aligned with the spacetimes and the gauge-choice,

(4) the deformation curvature tensor is defined so that covariant derivatives can be interchanged by means of the

 0022-2488/82/112151-04$02.50 © 1982 American Institute of Physics

usual Ricci identity,

(5) the deformation curvature tensor obeys Bianchi identities which have the same form as in normal Riemanian geometry, and

(6) in a coordinate system which is aligned with spacetime and with the gauge tensor, only the following components of H, the connection Γ, and Riemann are nonzero:

$$H^{\alpha}{}_{\beta} = \delta^{\alpha}{}_{\beta}\,, \tag{2.4}$$

$$\Gamma^{\alpha}{}_{\beta(C)} = \Gamma^{\alpha}{}_{(C)\beta} = -h^{\alpha}{}_{\beta(C)}\,, \tag{2.5}$$

$$\Gamma^{\alpha}{}_{\beta\gamma} = {}^{g}\Gamma^{\alpha}{}_{\beta\gamma}$$
$$= \text{spacetime connection coefficients},$$

$$R^{\alpha}{}_{\beta\gamma\delta} = {}^{g}R^{\alpha}{}_{\beta\gamma\delta}$$
$$= \text{spacetime curvature components},$$

$$R^{\kappa}{}_{(C)\alpha\beta} = h^{\kappa}{}_{\beta(C);\alpha} - h^{\kappa}{}_{\alpha(C);\beta}\,, \tag{2.6}$$

$$R^{\kappa}{}_{(C)\alpha(B)} = h^{\kappa}{}_{\alpha(C),B} - h^{\kappa}{}_{\rho(C)}\,h^{\rho}{}_{\alpha(B)}\,, \tag{2.7}$$

$$R^{\kappa}{}_{\gamma(A)(B)} = h^{\kappa}{}_{\rho(B)}h^{\rho}{}_{\gamma(A)} - h^{\kappa}{}_{\rho(A)}h^{\rho}{}_{\gamma(B)}\,, \tag{2.8}$$

$$R^{\kappa\gamma}{}_{\alpha(B)} = h^{\kappa}{}_{\alpha(B)}{}^{;\gamma} - h^{\kappa}{}_{\alpha(B)}{}^{;\gamma}\,. \tag{2.9}$$

The previous paper[5] of this series explained the use of action principles in deformation geometry. One begins with a deformation M which can include any number of perturbation parameters and embeds it in a larger deformation M^*. The action functional defines a function on M^*. The value $I(P)$ of this function at a point P of M^* is found by integrating the Lagrangian density over the spacetime which passes through P. The field equations on M are then found by insisting that the function I should have an extremum on M for every set of fields on M^*.

The extended manifold M^* requires an additional coordinate $x^{(*)}$. The deformation M appears as a surface of constant $x^{(*)}$ in M^*.

3. FIRST ORDER VARIATIONS AND HOMOGENEOUS COSMOLOGY

The Einstein Action at a point P of a deformation is given by the integral

$$I(P) = \int_{\Sigma P} {}^{4}\sigma R \tag{3.1}$$

over the spacetime Σ_P through the point P. In the language of deformation geometry, the Einstein Action Principle says that the point P lies on an Einstein spacetime if and only if all of the derivatives of I vanish at P in every deformation which includes the spacetime through P.

The previous paper[5] showed that the Einstein Action Principle

$$I_{,d} = 0$$

yields the condition

$$\int_{\partial\Sigma} \sigma_a G^{a}{}_{r}\,\iota^{r}{}_{d} + \int_{\Sigma} {}^{4}\sigma G^{a}{}_{r} h^{r}{}_{ad} = 0\,. \tag{3.2}$$

For simplicity, I will assume that the deformation is compact so that each spacetime in it is also compact. Thus, each spacetime Σ in the deformation is only a compact subset of a fully extended spacetime and will always have a boundary $\partial\Sigma$. In a closed universe one can arrange for the boundary to consist of just two disjoint space sections. In an open universe, there will always be a timelike boundary hypersurface. In certain special cases such as spacetimes with black holes it may be useful to have lightlike boundary surfaces along the event horizons.

When the spacetimes which are allowed in the deformation are unrestricted, the second fundamental form h is arbitrary and one recovers the Einstein equations from the deformations with $h = 0$ on the boundary. However, if the spacetimes are restricted to a particular isometry class, then the second fundamental form h will also be restricted and one cannot get rid of the surface integral.

This deformation geometry approach reveals a startling fact about these much-maligned surface terms—they are just the deformation components of the deformation Einstein Tensor. An obvious option is to set them equal to zero everywhere as gauge conditions on the deformation. That option will be discussed in more detail elsewhere. However, in the present case, the usual construction of a homogeneous cosmological model requires the construction of a rigidly defined deformation gauge which does not leave enough freedom to eliminate the surface terms.

Consider a spacetime which is described by the nonholonomic basis vector fields e_{μ} and the metric components $g^{\alpha\beta}$. Denote the commutation coefficients of the spacelike basis vectors by $c_{\mu\nu}{}^{\rho}$, where

$$c_{\mu\nu}{}^{\rho}\,e_{\rho} = [e_{\mu}, e_{\nu}]$$

and require the only nonvanishing coefficients to be the spacelike ones. Further require these coefficients to be the constants which characterize generators of one of the Bianchi Lie Groups. Take the timelike basis vector field to be just $e_0 = \partial/\partial t$, where t is the time coordinate. Finally, require the metric components to be functions of t alone. These are the spacetimes which are considered by homogeneous cosmology. Fix the perturbation gauge by declaring the time coordinate t and the Bianchi group coordinates to be the aligned spacetime coordinates and by insisting that the nonholonomic metric components are simple functions of the aligned perturbation parameters. It is also customary to require the time coordinate to be hypersurface orthogonal so that the shift vector is zero.

Contract Eq.(2.9) to obtain the expression

$$G^{a}{}_{r}\,\iota^{r}{}_{d} = h_{d}{}^{\cdot a} - h^{ab}{}_{d\cdot b}\,. \tag{3.3}$$

The first term of this expression is just the derivative of a one-form which has no spacetime components. The deformation connection coefficients all vanish for such a one-form so that the derivative consists of ordinary directional derivatives along the basis directions. One can split it up into spacelike and timelike parts (and possibly a perturbation part too). The spacelike derivatives vanish by virtue of Eq. (2.3) and the restriction to metric components which depend on t alone. Any perturbation components are projected away by the metric tensor when the index is raised. Thus, the surface term in the action principle becomes just

$$\int \sigma_a R^{a}{}_{d} = \int \sigma_0 \gamma^{00} \frac{dh_d}{dt} - \int \sigma_{\alpha} h^{\alpha\beta}{}_{d;\beta}\,.$$

The time derivative term can be integrated along the

timelike boundary hypersurfaces and can be removed from the spacelike hypersurface boundaries by restricting the deformation there. It is the second term which causes the trouble. Evaluating this term and neglecting all derivatives one easily finds that the surface term is just

$$\int \sigma_a \, R^{a}{}_{(*)} = \text{time derivatives} + \int \sigma_\alpha \, h^{\alpha\beta}{}_{(*)} \, c_{\beta\sigma}{}^{\sigma} . \tag{3.4}$$

Evidently, when the spacelike one-form with components

$$c_b := c_{bs}{}^{s}$$

is not zero, varying the Einstein Action with respect to the constant metric components will not yield the Einstein equations unless one imposes the gauge condition:

$$c_b \, h^{b}{}_{sd} = 0. \tag{3.5}$$

Without this gauge condition, one is using the wrong action functional for homogeneous variations of the Class B Bianchi-type cosmologies (types II, III, IV, V, $\mathrm{VI}_{h \neq -1}$, and $\mathrm{VII}_{h \neq 0}$).

4. CORRECTED ACTION FOR HOMOGENEOUS LAGRAGANGIAN COSMOLOGY

If the usual Einstein Action is the wrong choice, is there a correct choice for these models? The simplest idea is to add the spacetime integral of a divergence to the action and hope that the variation of this divergence will itself be a divergence which cancels the troublesome term in the variational surface integral. The obvious divergence to add is the divergence of the troublesome vector field c^r. This vector field is well defined on a deformation which consists of Class B Bianchi-type spacetimes. On such a deformation, we consider the actions

$$I'(P) = \int_{\Sigma P} {}^{4}\sigma(R + kc^{r}{}_{,r}) , \tag{4.1}$$

where k is a constant to be determined.

Now perform the deformation-covariant derivative of the added term:

$$\int (\sigma c^{r}{}_{,r})_{,d} = \int \sigma(c^{r}{}_{,rd} - h_d c^{r}{}_{,r}) .$$

After some straightforward manipulations, the proposed term is found to have the deformation-covariant derivative

$$\int (\sigma c^{r}{}_{,r})_{,d} = 2 \int \sigma_r \, c_a \bar{h}^{ra}{}_{d} ,$$

where $\bar{h}$ is the trace-reversed second fundamental form. Because the calculation uses the unfamiliar deformation-geometry formalism, I have included it in the Appendix.

Evidently the proposed correction term fails to eliminate a term proportional to the trace of the second fundamental form. But all is not lost. We have not completely fixed the gauge choice for these deformations. It is still possible to make time-coordinate transformations of the form

$$t = T(t, x^{(*)}).$$

These transformations change the spacetime volume element. With these transformations, one can guarantee that the spacetime volume element is preserved by Lie dragging in perturbation directions. The corresponding gauge condition is just

$$h_d = 0.$$

With the gauge now completely fixed, the correct action functional for this restricted class of spacetimes is

$$I'(P) = \int_{\Sigma P} {}^{4}\sigma(R - \tfrac{1}{2} c^{r}{}_{,r}) . \tag{4.2}$$

This action principle should give rise to a completely regular system of Lagrangian dynamical equations which are consistent with Einstein's equations. The remaining difficulty is that the second fundamental form is not arbitrary for the allowed deformations. The isometry group requires h to have constant components—that is no problem since the same isometry group requires the Einstein tensor to have constant components. However, the time gauge-condition $\mathrm{Tr}\, h = 0$ means that extremizing the action gives only the tracefree part of Einstein's equations.

Fortunately the Einstein equations are not all independent. Once the tracefree part of the Einstein tensor (minus the stress-energy tensor if matter is present) is required to be zero, the contracted Bianchi identities (and stress-energy conservation laws) require the trace to be a constant for all spacetime. The tracefree field equations in the form

$$G^{ab} - \tfrac{1}{4} G\gamma^{ab} = 0$$

can be combined with the initial value equation

$$G^{00} = 0$$

to ensure that G vanishes.

By using Eq. (4.2) one can reduce a Class B cosmological model to a system of dynamical equations which are obtained from an action principle. These equations must be augmented by the initial value equation, which is often called the Hamiltonian constraint. If the Hamiltonian constraint is used to eliminate a variable (a conformal factor or perhaps the lapse function for example) from the action, then a completely reduced dynamical system should result.

5. THE NATURAL EINSTEIN GAUGE IN HOMOGENEOUS COSMOLOGY

Instead of modifying the action, one might consider modifying the restrictions which are imposed on the spacetime metric in order to admit the elegant natural gauge conditions given by Eq. (3.5). The only difficulty with those conditions is that the usual definition of universal time in these cosmological models implies the conditions

$$h^{0s}{}_{d} = 0 . \tag{5.1}$$

Combining these conditions with Eq. (3.5) restricts the geometries which the deformation can connect.

One possible way out of the problem is to give up the usual $g_{0i} = 0$ restriction and allow *tilted* time surfaces. Equation (3.5) can then be used as the four gauge conditions which follow from the coordinate conditions

$$c_r \, \gamma^{rs} = \delta^{s}_{3} . \tag{5.2}$$

The difficulty then comes when one attempts to incorporate these particular coordinate conditions into a Hamiltonian

formalism. The usual coordinate conditions restrict the lapse and shift parts of the metric, functions which play natural roles as Lagrange multipliers for the initial value equations. Equation (5.2), however, restricts parts of the space metric which usually appear as dynamical fields.

6. DISCUSSION

The natural boundary conditions of the Einstein Action Principal are known to be the source of the difficulty with Hamiltonian formulations of Class B Bianchi model cosmologies.[3] Here, the difficulty has been displayed in the simpler context of spacetime-covariant Lagrangian dynamics. Thus, it is not a difficulty of the Hamiltonian canonical formalism alone.

The general form of the natural boundary term is given by Eq. (3.3). The part of it which causes trouble in a Class B model is shown in Eq. (3.4). Both of these equations reveal something unusual about these terms: Either one could be set to zero everywhere as a *gauge condition* on the metric variation. The Einstein Action Principal appears to be selecting a natural gauge condition rather than a natural boundary condition. Thus, a possible interpretation of the difficulty with Class B models is that their natural "Einstein gauges" conflict with the gauges which simplify Hamiltonian Canonical Formalisms. For example, Eq. (5.2) indicates that a Class B model wants to have a hypersurface-orthogonal geodesic *space* coordinate.

APPENDIX: COMPUTATION OF SURFACE COUNTERTERM

First, move the derivative across the integral by regarding it as a Lie derivative on the deformation and then treat it as a deformation-covariant derivative on the integrand.

$$\left(\int \sigma c^{r}{}_{\cdot r}\right)_{\cdot d}$$
$$= \int \sigma[c^{r}{}_{\cdot r\cdot s}\,\iota^{s}{}_{d} - h_{d}\,c^{r}{}_{\cdot r}]$$
$$= \int \sigma[c^{r}{}_{\cdot sr}\,\iota^{s}{}_{d} + c^{a}R^{r}{}_{asr}\,\iota^{s}{}_{d} - h_{d}\,c^{r}{}_{\cdot r}]$$
$$= \int [(c^{r}{}_{\cdot s}\,\iota^{s}{}_{d})_{\cdot r} - c^{r}{}_{\cdot s}\,\iota^{s}{}_{d\cdot r} - c^{a}R_{as}\,\iota^{s}{}_{d} - h_{d}\,c^{r}{}_{\cdot d}] .$$

Now evaluate each of the above terms in turn:

$$c^{r}{}_{\cdot s}\,\iota^{s}{}_{d} = \gamma^{rj}c_{j\cdot s}\,\iota^{s}{}_{d} = -\gamma^{rj}c_{k}\,\Gamma^{k}{}_{js}\iota^{s}{}_{d}$$
$$= -\gamma^{rj}c_{k}(-h^{k}{}_{sd}) = c^{a}h^{r}{}_{ad} ,$$
$$c^{r}{}_{\cdot s}\iota^{s}{}_{d\cdot r} = -c^{a}{}_{\cdot r}\,h^{r}{}_{ad}$$
$$-c^{a}R_{as}\iota^{s}{}_{d} = -c_{a}(h_{d}{}^{\cdot a} - h^{a}{}_{cd}{}^{\cdot c})$$
$$= c_{a}\,h^{a}{}_{rd}{}^{\cdot r} - c_{r}h_{d}{}^{\cdot r} .$$

Collect these partial results, recognize a total divergence, and obtain the desired surface integral.

$$\left(\int \sigma c^{r}{}_{\cdot r}\right)_{\cdot d}$$
$$= \int \sigma[(c^{a}h^{r}{}_{ad})_{\cdot r} + c^{a}{}_{\cdot r}\,h^{r}{}_{ad} + c_{a}h^{a}{}_{rd}{}^{\cdot r}$$
$$- c_{r}h_{d}{}^{\cdot r} - h_{d}c^{r}{}_{\cdot r}]$$
$$= \int \sigma[(c^{a}h^{r}{}_{ad})_{\cdot r} + (c^{a}h^{r}{}_{ad})_{\cdot r} - (c^{r}h_{d})_{\cdot r}]$$
$$= 2\int \sigma_{r}c_{a}(h^{ra}{}_{d} - \tfrac{1}{2}\gamma^{ra}h_{d}) .$$

[1] M. P. Ryan, *Hamiltonian Cosmolgy* (Springer, Heidelberg, 1972).
[2] M. MacCallum and A. Taub, Commun. Math. Phys. **25**, 173 (1972); M. P. Ryan, J. Math. Phys. **15**, 812–815 (1974).
[3] J. Isenberg and J. Nester, "Canonical Gravity" in *General Relativity and Gravitation*, edited by A. Held (Plenum, New York, 1980), see pp. 84–89; T. Regge and C. Teitelboim, Ann. Phys. (N.Y.) **88**, 286 (1974). A way to save the Hamiltonian formalism for these cosmologies by introducing force terms has been presented by Jantzen: R. Jantzen, "Spatially Homogeneous Cosmologies, Background and Dynamics", U. Cal.–Berkeley Preprint and Commun. Math. Phys. **64**, 211 (1979).
[4] R. H. Gowdy, J. Math. Phys. **19**, 2294–2304 (1978).
[5] R. H. Gowdy, J. Math. Phys. **22**, 988–994 (1981).

A class of solutions of the generalized Lund–Regge model

Dipankar Ray
Physics Department, Queen Mary College, Mile End Road, London E1 4NS, England [a)]

(Received 13 February 1980; accepted for publication 16 June 1981)

A class of solutions are obtained for the generalized Lund–Regge model of Corones [J. Math. Phys. **19**, 2431 (1978)] and its Euclidean counterpart. As a consequence, a new solution is noted for the original Lund–Regge model.

PACS numbers: 04.20.Tb

I. INTRODUCTION

In a recent paper, Corones[1] has studied the following form of coupled equations for a generalized Lund–Regge model:

$$\Theta_{11} - \Theta_{22} - 4g(\Theta) + h(\Theta)(\lambda_1^2 - \lambda_2^2) = 0,$$
$$\lambda_{11} - \lambda_{22} = 2p(\Theta)(\lambda_1\Theta_1 - \lambda_2\Theta_2), \tag{1.1}$$

where $\Theta_1 = \partial\Theta/\partial x^1$ and so on.

Θ and λ are unspecified field variables. $g(\Theta)$, $h(\Theta)$, and $p(\Theta)$ are unspecified functions of Θ. A transformation $x^2 \to ix^1$ transforms (1.1) to

$$\Theta_{11} + \Theta_{22} - 4g(\Theta) + h(\Theta)(\lambda_1^2 + \lambda_2^2) = 0,$$
$$\lambda_{11} + \lambda_{22} = p(\Theta)(\lambda_1\Theta_1 + \lambda_2\Theta_2). \tag{1.2}$$

Thus if (1.1) is a set of coupled equations in the two-dimensional space–time continuum, where x^1 is timelike and x^2 is spacelike, then (1.2) is its Euclidean counterpart, where both x^1 and x^2 are spacelike.

Equations of the form (1.1) and (1.2) occur in a number of physical problems, although the definition of the field variables Θ and λ as well as the functions $g(\Theta)$, $h(\Theta)$, and $p(\Theta)$ vary from one problem to another. Particular examples of (1.1) occur in the study of relativistic strings in a uniform external field,[2] and the nonlinear σ model in $1+1$ dimensions. Particular examples of (1.2) occur in the study of two-dimensional Heisenberg ferromagnets[3] and in the Ginzburg–Pitaevski approach to superfluids.[4] Also Corones[1] has shown that an infinite number of equations of the form (1.1), including the two above, admit soliton solutions.

The present note provides a unified approach towards obtaining some particular solutions of (1.1) and (1.2) for unspecified $g(\Theta)$, $h(\Theta)$, and $p(\Theta)$ subject to the restriction that

$$g(\Theta) \not\equiv 0, \quad h(\Theta) \not\equiv 0, \quad p(\Theta) \not\equiv 0. \tag{1.3}$$

To obtain these solutions, we assume that

$$\Theta = \Theta(z),$$

where $z = k(x^{1^2} + \epsilon x^{2^2}) + lx^1 + mx^2$,

k, l, and m are constants,

and $$\epsilon = -1 \quad \text{for Eqs. (1.1)}$$
$$= +1 \quad \text{for Eqs. (1.2).} \tag{1.4}$$

It is obvious that Eqs. (1.1) are Lorentz-invariant and Eqs. (1.2) are invariant under arbitrary translations and rotation in two-dimensional Euclidean space.

Therefore, for Eqs. (1.1), without loss of generality, one can express z in one of the following forms.

(i) $z = x^1$ with $\Theta_z \neq 0$,
(ii) $z = x^2$ with $\Theta_z \neq 0$,
(iii) $z = x^1 + x^2$ with $\Theta_z \neq 0$,
(iv) $z = x^{1^2} - x^{2^2}$ with $\Theta_z \neq 0$,
(v) $\Theta = \text{const}$, where the form of z is immaterial.

In a similar manner for (1.2), z can be expressed in one of the following ways:

(i) $z = x^1$, $\Theta_z \neq 0$,
(ii) $z = x^{1^2} + x^{2^2}$,
(iii) $\Theta = \text{const}$, where z is immaterial.

Before we study these cases one by one we have to establish two lemmas.

II. TWO LEMMAS

Lemmas: If χ is a function of any two independent variables ξ and η satisfying

$$\chi_{\xi\xi} \mp \chi_{\eta\eta} = U(\xi)\chi_\xi, \tag{2.1a}$$
$$\chi_\xi^2 \mp \chi_\eta^2 = V(\xi), \tag{2.1b}$$
$$U(\xi) \not\equiv 0, \quad V(\xi) \not\equiv 0,$$

then

$$\chi_\eta = \text{const}. \tag{2.2}$$

We shall prove this for the negative sign in (2.1). The other is similar. From (2.1b),

$$\chi_\xi\chi_{\xi\eta} - \chi_\eta\chi_{\eta\eta} = 0. \tag{2.3}$$

From (2.1a) and (2.3)

$$\chi_\eta\chi_{\xi\xi} - \chi_\xi\chi_{\xi\eta} = U(\xi)\chi_\xi\chi_\eta. \tag{2.4}$$

If $\chi_\eta \equiv 0$, the (2.2) is obviously satisfied. On the other hand, if $\chi_\eta \not\equiv 0$, one can divide (2.4) by χ_η^2 and integrate to obtain

$$\chi_\xi = \chi_\eta \exp\left[\int U(\xi)\, d\xi\right] W(\eta),$$

where $W(\eta)$ is some function of η.

From this equation and (2.1b) one can solve χ_ξ and χ_η in terms of $U(\xi)$, $V(\xi)$, and $W(\eta)$. The condition that

[a)]Present Address: Center of Advanced Study in Applied Mathematics, University College of Science, 92, Acharyya Prafulla Chandra Road, Calcutta-700 009, India.

 0022-2488/82/112155-04$02.50 © 1982 American Institute of Physics

$\partial\chi_\xi/\partial\eta = \partial\chi_\eta/\partial\xi$ then gives (2.2). Therefore, (2.2) holds in either case.

III. SOLUTIONS OF (1.1)

Case (i) $z = x^1, \Theta_z \neq 0$

Here

$$\Theta = \Theta(x^1). \tag{3.1}$$

Equation (1.1) is now of the form (2.1) with the negative sign, where λ is taken as χ, x^1 as ξ, and χ^2 as a. Therefore, using (3.1) and the lemma, we obtain from (1.1b)

$$\lambda = K^1 x^2 + A^1 \int \exp\left[2\int p(\Theta)\, d\Theta\right] dx^1.$$

Formula (1.1a) then gives

$$\int \frac{d\Theta}{(B^1 + 8\int g(\Theta)\, d\Theta - 2\int h(\Theta)\{A^{1^2}\exp[4\int p(\Theta)\, d\Theta] - K^{1^2}\}\, d\Theta)^{1/2}} = x^1, \tag{3.2}$$

where A^1, B^1, and K^1 are constants of integration.

Case (ii) $z = x^2$ with $\Theta_z \neq 0$

Here

$$\Theta = \Theta(x^2). \tag{3.3}$$

Similarly, here

$$\lambda = K''x^1 + A''\int \exp\left[2\int p(\Theta)\, d\Theta\right] dx^2,$$

$$\int \frac{d\Theta}{(B'' - 8\int g(\Theta)\, d\Theta - 2\int h(\Theta)\{A''^2\exp[4\int p(\Theta)\, d\Theta] - K''^2\}\, d\Theta)^{1/2}} = x^2. \tag{3.4}$$

Case (iii) $z = x^1 + x^2$, $\Theta \neq$ const

Define

$$\alpha = x^1 + x^2,$$
$$\beta = x^1 - x^2. \tag{3.5}$$

Therefore,

$$\Theta = \Theta(\alpha), \tag{3.6}$$

Formula (1.1) reduces to

$$-g(\Theta) + h(\Theta)\lambda_\alpha\lambda_\beta = 0, \tag{3.7a}$$

$$\lambda_{\alpha\beta} = p(\Theta)\lambda_\beta\Theta_\alpha. \tag{3.7b}$$

Equation (3.7b) can be readily integrated to give

$$\lambda = \psi + \phi\exp\left[\int p(\Theta)\, d\Theta\right], \tag{3.8}$$

where $\psi = \psi(\alpha)$, $\phi = \phi(\beta)$.

Formula (3.7a) then gives

$$\phi_\beta\exp\left[\int p(\Theta)\, d\Theta\right]\left\{\psi_\alpha + \phi\exp\left[\int p(\Theta)\, d\Theta\right]p(\Theta)\Theta_\alpha\right\} = g(\Theta)/h(\Theta). \tag{3.9}$$

From (1.3) and (3.9),

$$\phi_\beta \not\equiv 0. \tag{3.10}$$

Therefore, (3.9) can be rewritten as

$$\frac{\psi_\alpha\exp[-\int p(\Theta)q\, d\Theta]}{p(\Theta)\Theta_\alpha} + \phi = \frac{g(\Theta)\exp[-2\int p(\Theta)\, d\Theta]}{h(\Theta)p(\Theta)\Theta_\alpha}\cdot\frac{1}{\phi_\beta}. \tag{3.11}$$

Since ψ, $g(\Theta)$, $p(\Theta)$, and $h(\Theta)$ are functions of α and ϕ is a function of β, differentiating (3.11) with respect to α and β, we have

$$\left[\frac{g(\Theta)\exp[-2\int p(\Theta)\, d\Theta]}{h(\Theta)p(\Theta)\Theta_\alpha}\right]_\alpha\left(\frac{1}{\phi_\beta}\right)_\beta = 0.$$

However, if $(1/\phi_\beta)_\beta = 0$, i.e., $1/\phi_\beta = \text{const}$, then the right-hand side of (3.11) is a function of α only and hence so is the left-hand side which is possible only if ϕ is a constant which violates (3.10). Therefore,

$$\frac{g(\Theta)\exp[-2\int p(\Theta)\, d\Theta]}{h(\Theta)p(\Theta)\Theta_\alpha} = \text{const}. \tag{3.12}$$

Then the right-hand side of (3.11) is a function of β and so must be the left-hand side; therefore,

$$\frac{\psi_\alpha\exp[-\int p(\Theta)\, d\Theta]}{p(\Theta)\Theta_\alpha} = \text{const}. \tag{3.13}$$

Using (3.8), (3.11), (3.12), and (3.13), the complete solutions of (3.7), which are the complete solutions of (1.1) for Θ as a function of $x^1 + x^2$, this can be written as

$$B\int \frac{h(\Theta)p(\Theta)\exp[2\int p(\Theta)\, d\Theta]}{g(\Theta)}\, d\Theta = x^1 + x^2,$$
$$\psi = A\int \exp\left[\int p(\Theta)\, d\Theta\right]p(\Theta)\, d\Theta,$$
$$\phi = -A \pm \sqrt{2B(x^1 - x^2) + C}, \tag{3.14}$$
$$\lambda = \psi(x^1 + x^2) + \phi(x^1 - x^2)\exp\left[\int p(\Theta)\, d\Theta\right],$$

where A, B, and C are constants of integration.

Case (iv) $z = x'^2 - x^{2^2}$, $\Theta \neq$ **const**

Define

$$y = \ln(x'^2 - x^{2^2}), \tag{3.15}$$

$$s = \ln \frac{x^1 + x^2}{x^1 - x^2}.$$

Therefore,

$$\Theta = \Theta(y). \tag{3.16}$$

Formula (1.1) can then be rewritten as

$$\lambda_{yy} - \lambda_{ss} = 2p(\Theta)\Theta_y \lambda_y \ , \tag{3.17a}$$

$$\lambda_y^2 - \lambda_s^2 = \frac{e^y g(\Theta) - \Theta_{yy}}{h(\Theta)} \ . \tag{3.17b}$$

Therefore, using (3.16) and Sec. II,

$$\lambda_s = \text{const}. \tag{3.18}$$

Using (3.18), one can integrate (3.17a) to give

$$\lambda = M\exp\left(2\int p\, d\Theta\right) + N \ln \frac{x^1 + x^2}{x^1 - x^2}. \tag{3.19a}$$

(3.17b) then becomes

$$\left[M^2 \exp\left(4\int p\, d\Theta\right) - N^2\right] h(\Theta) = e^y g(\Theta) - \Theta_{yy}, \tag{3.19b}$$

where

$$\Theta = \Theta(y) \tag{3.19c}$$

and

$$y = \ln(x'^2 - x^{2^2}). \tag{3.19d}$$

If p, h, and g are known functions of Θ, then (3.19b) is an ordinary second order differential equation for Θ as a function of y which can be solved numerically by computer. λ and y are then given respectively by (3.19a) and (3.19d).

Case (v) $\Theta =$ **const**

Here (1.1) gives

$$\lambda_1^2 - \lambda_2^2 = \text{const},$$

$$\lambda_{11} - \lambda_{22} = 0.$$

The solution can be readily seen as

$$\lambda = A'''x^1 + B'''x^2, \tag{3.20}$$

where A''' and B''' are constants satisfying

$$A'''^2 - B'''^2 = 4g(\Theta)/h(\Theta)$$

and Θ is a constant.

IV. SOLUTIONS OF (1.2)

Case (i) $z = x^1$, $\Theta_z \neq 0$

This case is similar to case (i) of Sec. III. Proceeding in a similar way, we get the solution as

$$\lambda = K_0 x^2 + A_0 \int \exp\left(2\int p(\Theta)\, d\Theta\right) dx^1,$$

$$\int \frac{d\Theta}{(B_0 + 8\int g(\Theta)\, d\Theta - 2\int h(\Theta)\{A_0^2 \exp[4\int p(\Theta)\, d\Theta] + K^2\}\, d\Theta)^{1/2}} = x^1. \tag{4.1}$$

where A_0, B_0, and K_0 are constants of integration.

Case (ii) $z = x'^2 + x^{2^2}$, $\Theta_z \neq 0$

This case is similar to case (iv) of Sec. III. Proceeding in a similar way, we get

$$\lambda = M_0 \exp\left(2\int p\, d\Theta\right) + 2N_0 \tan^{-1} \frac{x^2}{x^1}, \tag{4.2a}$$

$$\left[M_0^2 \exp\left(4\int p\, d\Theta\right) + N_0^2\right] h(\Theta) = e^{y_0} g(\Theta) + \Theta_{y_0 y_0}, \tag{4.2b}$$

where

$$\Theta = \Theta(y_0), \tag{4.2c}$$

$$y_0 = \ln(x'^2 + x^{2^2}) \ , \tag{4.2d}$$

M_0 and N_0 are constants of integration.

As before, (4.2b) is an ordinary second order differential equation for Θ as a function of y_0 if p, h, and g are known functions of Θ. This equation can be solved numerically. λ and y_0 are given respectively by (4.2a) and (4.2d).

Case (iii) $\Theta =$ **const**

This case is similar to case (v) of Sec. III.

The solutions are

$$\lambda = A_1 x^1 + B_1 x^2, \tag{4.3a}$$

where A_1 and B_1 are constants satisfying

$$A_1^2 + B_1^2 = 4g(\Theta)/h(\Theta) \tag{4.3b}$$

and Θ is a constant.

V. SOLUTIONS FOR ORIGINAL LUND–REGGE MODEL

The original Lund–Regge model[2] is a special case of (1.1), where

$$g(\Theta) = -(c \sin\Theta\cos\Theta)/4,$$

$$h(\Theta) = \cos\Theta/\sin^3\Theta, \tag{5.1}$$

$$p(\Theta) = 1/\sin\Theta\cos\Theta.$$

In this case all the solutions for $\Theta = \Theta(x^1)$, $\Theta = \Theta(x^2)$, $\Theta = \Theta(x^1 - x^2)$, and $\Theta =$ const already appear in the literature.[5] However, in the case of $\Theta = \Theta(x'^2 - x^{2^2})$, i.e., in case (iv) of Sec. III, we observe some new solutions for the original Lund–Regge model.[2]

For the original Lund–Regge model the equations for case (iv) in Sec. III, i.e., (3.19), reduce to

$$\lambda = M \tan^2\Theta + N \ln \frac{x^1 + x^2}{x^1 - x^2}, \tag{5.2a}$$

$$(M^2\tan^2\Theta - N^2)\frac{\cos\Theta}{\sin^3\Theta} = e^y g(\Theta) - \Theta_{yy}\,, \tag{5.2b}$$

$$\Theta = \Theta(y), \tag{5.2c}$$

$$y = \ln(x^{1^2} - x^{2^2}) \tag{5.2d}$$

Equation (5.1b) can be solved numerically as noted before and λ and y are given by (5.1a) and (5.1d). The special case of (5.1) where $M = 0$ and the special case of (5.1) where $N = 0$ already appear in the literature.[5]

VI. CONCLUSION

In conclusion, all the solutions of (1.1) subject to (1.4) are given by (3.2), (3.4) (3.14), (3.19), and (3.20). In the particular case of the original Lund–Regge model this indicates a new solution (5.2). Solutions of (1.2) subject to (1.4) are given by (4.1) and (4.2).

[1]J. Corones, J. Math. Phys. **19**, 2431 (1978).
[2]F. Lund and T. Regge, Phys. Rev. D **14**, 1524 (1976).
[3]S. Trimper, Phys. Lett. A **70**, 114 (1979).
[4]V. L. Ginsburg and V. L. Pitaveski, Zh. Eksp. Theor. Fiz. **34**, 1240 (1958) [Sov. Phys. JETP **7**, 858 (1958)].
[5]D. Ray, Phys. Rev. D **18**, 3879 (1978).

Orthogonal transitivity, invertibility and null geodesic separability in type D vacuum solutions of Einstein's field equations with cosmological constant[a)]

S. R. Czapor and R. G. McLenaghan
Department of Applied Mathematics, University of Waterloo, Waterloo, Ontario, Canada

(Received 5 January 1982; accepted for publication 18 June 1982)

It is shown without explicit integration that all Petrov type-D vacuum solutions with cosmological constant admit at least a two-parameter, abelian, orthogonally transitive group of local isometries. In the case when the orbits are non-null the group is invertible and a symmetric null tetrad is shown to exist in which it is manifest that the principal null congruences defined by the type-D Weyl tensor are indistinguishable. General forms for the metrics are given for both the case of non-null and null group orbits. It is also demonstrated that there always exists a system of coordinates for these solutions in which the Hamilton–Jacobi equation for the null geodesics is solvable by separation of variables, a fact that explains the existence of a conformal Killing tensor therein. A partial invariant characterization of Kinnersley's type-D vacuum solutions is given from which it follows that all his metrics except those of Case III admit a (0,2) Yano–Killing tensor and hence a full (0,2) Killing tensor.

PACS numbers: 04.20.Jb, 02.30.Jr

1. INTRODUCTION

Kinnersley[1] observed that every member of his exhaustive set of solutions of Einstein's vacuum field equations for Petrov type D admits at least a two-parameter abelian isometry group. More direct proofs of this property have been given by Hughston and Sommers[2] and by Weir and Kerr.[3] The former authors construct the Killing vectors from a second-rank Killing spinor whose existence in the solutions had previously been demonstrated by Walker and Penrose.[4] These workers and others[5–7] employed the Killing spinor to construct a second-rank conformal Killing tensor and, in all but Kinnersley Case III metrics, a full Killing tensor in order to explain the existence in the Kerr (–Newman) solution of a fourth constant of the motion for the (charged-) particle orbits which had been brought to light earlier by the work of Carter[8,9] as the separation constant in his separable solution of the corresponding Hamilton–Jacobi equation. Carter's result has been extended to most of Kinnersley's metrics by Matravers,[10] who shows that for all metrics in Cases II and IV the Hamilton–Jacobi equation is solvable by separation of variables while for the Case III.A (the *C*-metric) it is solvable in this way only for the null geodesics, the separation constant in this case giving rise to a quadratic first integral and a corresponding conformal Killing tensor. (These separability properties for all but the Case III metrics may be inferred indirectly from the existence of Killing tensors.[11,12])

The results just described have been given a unified treatment in a recent paper by Debever and McLenaghan[13] (referred to in the sequel as DM) for the case of the Petrov type-D electrovac solutions with cosmological constant for an aligned nonsingular Maxwell field satisfying the generalized Goldberg–Sachs theorem.[14,15] These authors showed, without explicit integration, (i) that all such solutions admit at least a two-parameter, abelian, orthogonally transitive group of local isometries; (ii) that in the case of non-null orbits, when the group is invertible, there exists a symmetric null tetrad in which it is manifest that the principal null congruences of the type-D Weyl tensor are indistinguishable; (iii) that there exist solutions (which seem to have been overlooked by other workers) with null group orbits that are orthogonally transitive but not invertible; (iv) that the Hamilton–Jacobi equation for the null geodesics is always solvable by separation of variables in these solutions and that this fact explains the existence of a conformal Killing tensor therein.

The purpose of the present paper is to extend these results to the vacuum solutions with cosmological constant. It may seem that this result is merely a special case for vanishing Maxwell field of the electrovac result just described. However, while the results are identical the proof is considerably more delicate in the vacuum case. The apparent reason for this is that some structure is lost when the electromagnetic field is set to zero. In addition to the results just described we give a partial classification of Kinnersley's metrics in terms of a set of invariants defined in terms of the Newman-Penrose[16] (NP) spin coefficients for a null tetrad whose real null vectors are principal null vectors of the type-D Weyl tensor. We also express in terms of these invariants the conditions that must be satisfied in order that the rank-two Killing spinor of these solutions define a second-rank Yano–Killing tensor[17] and therefrom a second-rank Killing tensor. It is then easily seen that these conditions are satisfied by all the Kinnersley metrics except those of Case III.[18]

The proof of our results follows the same lines as that given in DM and divides into two cases according to whether both the null congruences defined by the Weyl tensor are either expanding or nonexpanding or whether one is expanding and the other is not. In the first case it is shown that the solutions admit a Riemannian–Maxwellian invertible structure and hence by a theorem of Debever, McLenaghan, and Tariq[19] (quoted as DMT in the sequel, earlier versions of which were given by Debever[20,21]) possess an invertible, abelian two parameter isometry group. The second case requires

[a)] This work was supported in part by a grant from the Natural Sciences and Engineering Research Council of Canada.

 0022-2488/82/112159-09$02.50 © 1982 American Institute of Physics

a separate treatment. Parts of the proof which are identical with those given in DM are omitted.

Our calculations are performed using the NP formalism and the equivalent complex vectorial formalism of Cahen, Debever, and Defrise[22,23] which are related in DMT. The notation is identical to that of DM.

2. HYPOTHESES AND STATEMENT OF RESULTS

We consider space-times which are solutions of Einstein's vacuum field equations with cosmological constant

$$R_{ij} = \lambda g_{ij} \tag{2.1}$$

such that the Weyl tensor C_{ijkl} is Petrov type D. Thus there exist real null vector fields l and n such that at each point

$$l^i l^j C_{hij[k}\, l_{m]} = n^i n^j C_{hij[k}\, n_{m]} = 0. \tag{2.2}$$

This class of solutions will be denoted by $\mathfrak{D}_0$. The main results of this paper are as follows.

Theorem 1: Every solution in $\mathfrak{D}_0$ admits at least a two-parameter orthogonally transitive abelian isometry group. If the group orbits are non-null, the group is invertible, and there exist coordinates (u, v, w, x) in which the metric has the form

$$\begin{aligned} ds^2 = &-e(Ldu + Mdv)^2 + eR^2dw^2 \\ &+ e(Ndu + Pdv)^2 - eT^2dx^2, \end{aligned} \tag{2.3a}$$

where $e = 1$ (spacelike orbits) or $e = -1$ (timelike orbits), L, M, R are real-valued functions independent of u and v, and N, P, T are functions independent of u and v satisfying

$$\bar{N} = -eN, \quad \bar{P} = -eP, \quad \bar{T} = eT. \tag{2.3b}$$

If the group orbits are null, the group is not invertible, and there exist coordinates such that the metric has the form

$$\begin{aligned} ds^2 = &\, 2Rdw(Ldu + Mdv) \\ &- (Ndu + Pdv)^2 - T^2dx^2, \end{aligned} \tag{2.4}$$

where L, M, N, P, R, T are real-valued functions independent of u and v.

Invertibility is clear in the case of the metric (2.3) since the transformation

$$(u, v, w, x) \to (-u, -v, w, x) \tag{2.5}$$

is an isometry. The symmetric null tetrad for the metric (2.3) is given by[24]

$$\begin{aligned} \theta^1 &:= n_i dx^i = (1/\sqrt{2})(Ldu + Mdv + Rdw), \\ \theta^2 &:= l_i dx^i = (e/\sqrt{2})(-Ldu - Mdv + Rdw), \\ \theta^3 &:= -\bar{m}_i dx^i = (1/\sqrt{2})(Ndu + Pdv + Tdx), \\ \theta^4 &:= -m_i dx^i = (e/\sqrt{2})(-Ndu - Pdv + Tdx). \end{aligned} \tag{2.6}$$

The transformation (2.5) induces the following transformation of the above tetrad:

$$\theta^1 \to e\theta^2, \ \theta^2 \to e\theta^1, \ \theta^3 \to e\theta^4, \ \theta^4 \to e\theta^3. \tag{2.7}$$

This is the involution L of the Riemannian–Maxwellian invertible structure which exists in the solutions of $\mathfrak{D}_0$ with non-null group orbits.

An example[25] of a solution in $\mathfrak{D}_0$ with null orbits is given by Kinnersley's Case II.E metric with $m = b = 0$.

Theorem 2: For every solution in $\mathfrak{D}_0$ there exist coordinates (u, v, w, x) for which the Hamilton–Jacobi equation for the null geodesics

$$g^{ij}\frac{\partial S}{\partial x^i}\frac{\partial S}{\partial x^j} = 0 \tag{2.8}$$

is solvable by separation of variables; that is, it has a complete integral of the form

$$S = \alpha u + \beta v + S_1(w, \alpha, \beta, \gamma, \delta) + S_2(x, \alpha, \beta, \gamma, \delta), \tag{2.9}$$

where $\alpha, \beta, \gamma, \delta$ are constant. In these coordinates the metrics have the following forms:

Non-null orbits:

$$\begin{aligned} ds^2 = e^{2\psi}[&-e(Ldu + Mdv)^2 + eR^2dw^2 \\ &+ e(Ndu + Pdv)^2 - eT^2dx^2], \end{aligned} \tag{2.10}$$

Null orbits:

$$\begin{aligned} ds^2 = e^{2\psi}[&2Rdw(Ldu + Mdv) \\ &- (Ndu + Pdv)^2 - T^2dx^2], \end{aligned} \tag{2.11}$$

where the metric functions satisfy

$$\psi_u = \psi_v = 0, \quad \left(\psi_u = \frac{\partial \psi}{\partial u}\right), \tag{2.12}$$

$$R_x = T_w = 0, \tag{2.13}$$

$$(L/Z)_w = (M/Z)_w = (N/Z)_x = (P/Z)_x = 0, \tag{2.14}$$

where $Z = LP - MN$, as well as the conditions of Theorem 1.

By substituting S of the form (2.9) into Eq. (2.8), we obtain, via (2.10) and (2.11) respectively:

Non-null orbits:

$$\begin{aligned} &Z^{-2}(\alpha P - \beta N)^2 - R^{-2}(S_1'(w))^2 \\ &\quad = Z^{-2}(\beta L - \alpha M)^2 - T^{-2}(S_2'(x))^2, \end{aligned} \tag{2.15}$$

Null orbits:

$$\begin{aligned} &R^{-1}S_1'(w)Z^{-1}(\alpha P - \beta N) \\ &\quad = \tfrac{1}{2}[Z^{-2}(\beta L - \alpha M)^2 + T^{-2}(S_2'(x))^2], \end{aligned} \tag{2.16}$$

which both separate into ordinary differential equations in S_1 and S_2 in view of (2.13) and (2.14).

The proof of Theorem 2 is identical to that given in DM. Separability of the Hamilton–Jacobi equation

$$g^{ij}\frac{\partial S}{\partial x^i}\frac{\partial S}{\partial x^j} = m^2 \tag{2.17}$$

for the non-null geodesics requires that the conformal factor be of the form

$$e^{2\psi} = f(w) + g(x), \tag{2.18}$$

which is equivalent to ψ satisfying the differential equation

$$\psi_{wx} + 2\psi_w\psi_x = 0. \tag{2.19}$$

To determine when this condition is satisfied requires a knowledge of the explicit form of the function ψ that can only be determined by solving the remaining vacuum field equations for the metrics (2.10) and (2.11). Such an analysis, which will be presented elsewhere, shows that (2.19) is satisfied for all but the Case III Kinnersley metrics.[26]

It follows directly from Theorem 2 and Eqs. (2.15) and (2.16) that we have the following:

Corollary: Every solution in $\mathfrak{D}_0$ admits a second rank

conformal Killing tensor B^{ij} defined by:

Non-null orbits:

$$B^{ij}p_ip_j = \tfrac{1}{2}[R^{-2}p_w^2 + T^{-2}p_x^2 - Z^{-2}(Pp_u - Np_v)^2 - Z^{-2}(Lp_v - Mp_u)^2], \tag{2.20}$$

Null orbits:

$$B^{ij}p_ip_j = \tfrac{1}{2}[R^{-1}p_wZ^{-1}(Pp_u - Np_v) + \tfrac{1}{2}T^{-2}p_x^2 + \tfrac{1}{2}Z^{-2}(Lp_v - Mp_u)^2], \tag{2.21}$$

where $p_i = \partial S/\partial x^i$.

The tensors defined above give explicit forms for the conformal Killing tensors whose existence in the $\mathfrak{D}_0$ solutions was demonstrated by Walker and Penrose[27] and others.[28] We recall that these authors construct the tensor in question by first showing the existence for every solution in $\mathfrak{D}_0$ of a second-rank (conformal) Killing spinor[29] X_{AB} satisfying the twistor equation

$$\nabla_{\dot{A}(A}X_{BC)} = 0. \tag{2.22}$$

In the $\mathfrak{D}_0$ solutions X_{AB} has the form

$$X_{AB} = \phi o_{(A}\iota_{B)}, \tag{2.23}$$

where o_A and ι_B are principal spinors of the type-D Weyl tensor corresponding, respectively, to the principal null vectors l and n defined by Eq. (2.2). The complex function ϕ in Eq. (2.23) satisfies

$$d\phi = \tfrac{1}{2}\phi\theta, \tag{2.24}$$

where θ is a complex one-form defined by

$$\theta = -2(\rho\theta^1 - \mu\theta^2 + \tau\theta^3 - \pi\theta^4), \tag{2.25}$$

and ρ, μ, τ, and π denote NP spin coefficients associated to o_A and ι_B. The conformal Killing tensor B_{ij} is expressible in terms of the Killing spinor as

$$B_{ij} := X_{AB}\bar{X}_{\dot{A}\dot{B}} = \tfrac{1}{2}\phi\bar{\phi}(l_{(i}n_{j)} + m_{(i}\bar{m}_{j)}), \tag{2.26}$$

and satisfies the equations

$$B_{(ij;k)} - \tfrac{1}{3}g_{(ij}B_{k)}{}^{l}{}_{;l} = 0. \tag{2.27}$$

It follows from Eq. (2.24) and DM Eq. (7.7), which holds for both the bracketed metrics on the right-hand sides of Eqs. (2.10) and (2.11), that

$$d(\phi\bar{\phi}) = \tfrac{1}{2}\phi\bar{\phi}(\theta + \bar{\theta}) = 0, \tag{2.28}$$

which in turn implies that $\phi\bar{\phi}$ is constant. It is then easily verified using Eq. (2.6) that, if $\phi\bar{\phi} = \tfrac{1}{2}e$, the right-hand side of Eq. (2.26) yields a value of B_{ij} and hence of B^{ij} which is identical to that given by Eq. (2.20). A similar argument applies to the tensor defined by Eq. (2.21).

Penrose and Floyd[30] have shown that a second tensor which is skew symmetric may be constructed from X_{AB} namely

$$f_{ij} := X_{AB}\epsilon_{\dot{A}\dot{B}} + \epsilon_{AB}\bar{X}_{\dot{A}\dot{B}}. \tag{2.29}$$

It can be shown that f_{ij} satisfies the Yano–Killing equation[31]

$$f_{i(j;k)} = 0 \tag{2.30}$$

if and only if X_{AB} satisfies the *additional* skew Hermitian condition[32]

$$\nabla_{\dot{A}B}X^{B}{}_{A} + \nabla_{\dot{A}B}\bar{X}^{B}{}_{\dot{A}} = 0. \tag{2.31}$$

It will be shown in Secs. 3 and 6 that all Kinnersley's metrics except those in Case III satisfy this condition and hence admit such a tensor, sometimes called a Yano–Killing tensor. It follows from Eq. (2.30) that the symmetric tensor[33]

$$K_{ij} := f_{ik}f_j{}^k \tag{2.32}$$

satisfies the Killing tensor equation

$$K_{(ij;k)} = 0. \tag{2.33}$$

The trace-free part of K_{ij} is just the tensor B_{ij}. Thus by the remark following Eq. (2.31) the Kinnersley metrics in all but Case III admit the full Killing tensor. It should be noted that as in the case of B_{ij} the tensor K_{ij} may be obtained from the separable solution of the Hamilton–Jacobi equation (2.17) when the condition (2.19) holds. It is also worth mentioning that the complex Killing vector of Hughston and Sommers for the $\mathfrak{D}_0$ metrics may be expressed as

$$\xi_i := H_i{}^j{}_{;j}, \tag{2.34}$$

where

$$H_{ij} := X_{AB}\epsilon_{\dot{A}\dot{B}}. \tag{2.35}$$

It follows from previous remarks that ξ_i is complex, hence defining two real Killing vectors, only for the Case III metrics. A method of construction of a second Killing vector for the metrics in the remaining cases has been given by Sommers.[34]

3. BASIC EQUATIONS FOR THE CLASS $\mathfrak{D}_0$

Since the Weyl tensor C_{ijkl} is Petrov type-D we may choose a local null tetrad $(l, n, m, \bar{m})$ whose real null vectors l and n are principal null vectors of C_{ijkl}. When the field equations are also taken into account it follows that the only nonzero NP components of the curvature are $\Psi := \Psi_2$ and Λ. The tetrad is then fixed up to the transformation

$$l' = e^a l, \quad n' = e^{-a}n, \quad m' = e^{ib}m, \tag{3.1}$$

where a and b are real-valued functions. Under (3.1) the NP spin coefficients transform as follows:

$$\kappa' = e^{2a+ib}\kappa, \quad \tau' = e^{ib}\tau, \quad \sigma' = e^{a+2ib}\sigma, \quad \rho' = e^a\rho, \tag{3.2}$$

$$\pi' = e^{-ib}\pi, \quad \nu' = e^{-2a-ib}\nu, \quad \mu' = e^{-a}\mu, \quad \lambda' = e^{-a-2ib}\lambda, \tag{3.3}$$

$$\epsilon' = e^a[\epsilon + \tfrac{1}{2}Dp], \quad \gamma' = e^{-a}[\gamma + \tfrac{1}{2}\Delta p], \tag{3.4}$$

$$\beta' = e^{ib}[\beta + \tfrac{1}{2}\delta p], \quad \alpha' = e^{-ib}[\alpha + \tfrac{1}{2}\bar{\delta}p], \tag{3.5}$$

where $p = a + ib$.

If we define

$$I_1 := \pi\bar{\pi} - \tau\bar{\tau}, \tag{3.6a}$$

$$I_2 := \bar{\mu}\rho - \mu\bar{\rho}, \tag{3.6b}$$

$$I_3 := \rho\bar{\pi} - \bar{\rho}\tau, \tag{3.6c}$$

$$I_4 := \mu\bar{\pi} - \bar{\mu}\tau, \tag{3.6d}$$

$$I_5 := \rho\bar{\pi} + \bar{\rho}\tau, \tag{3.6e}$$

it is easily verified that these quantities transform under (3.1) as follows:

$$I_1' = I_1, \tag{3.7a}$$

$$I_2' = I_2, \tag{3.7b}$$

$$I_3' = e^{a+ib} I_3, \tag{3.7c}$$

$$I_4' = e^{-a+ib} I_4, \tag{3.7d}$$

$$I_5' = e^{a+ib} I_5. \tag{3.7e}$$

We shall use these transformation laws along with those given in Eqs. (3.2) and (3.3) to give in Sec. 6 an invariant characterization of Kinnersley's cases.

On account of the choice of tetrad, Bianchi's identities take the form

$$\kappa\Psi = \sigma\Psi = \nu\Psi = \lambda\Psi = 0, \tag{3.8}$$

$$D\Psi = 3\rho\Psi, \quad \Delta\Psi = -3\mu\Psi, \tag{3.9}$$

$$\delta\Psi = 3\tau\Psi, \quad \bar{\delta}\Psi = -3\pi\Psi. \tag{3.10}$$

Since $\Psi \neq 0$ it follows that

$$\kappa = \sigma = \nu = \lambda = 0, \tag{3.11}$$

which expresses the well-known fact that *both* the null congruences associated to the type-D Weyl tensor are *geodesic* and *shear-free*. The remaining Eqs. (3.9) and (3.10) may be written compactly as

$$d\Psi = -\tfrac{3}{2}\Psi\theta, \tag{3.12}$$

where θ is defined by Eq. (2.25). It follows that the integrability conditions for Bianchi's identities may be expressed as

$$d\theta = 0. \tag{3.13}$$

This equation is also the integrability condition for the Eq. (2.24) satisfied by the function ϕ appearing in the Killing spinor of the form (2.23). The rest of the Eq. (2.22) satisfied by X_{AB} can be written as

$$\kappa\phi = \sigma\phi = \nu\phi = \lambda\phi = 0, \tag{3.14}$$

which is identically satisfied in this case on account of Eq. (3.11). It thus follows that the existence of a Killing spinor of the form (2.23) imposes the *same* restrictions on the spacetime geometry as the vacuum Bianchi identities. This relation becomes transparent when one notes (following Walker and Penrose[35]) that $\phi = \Psi^{-1/3}$ is a solution of Eq. (2.24) if and only if Ψ is a solution of Eq. (3.12).

The additional condition (2.31) that must be satisfied in order that X_{AB} define a Yano–Killing tensor, when combined with Eq. (2.24) implies the further restrictions

$$\rho\phi + \bar{\rho}\bar{\phi} = 0, \tag{3.15}$$

$$\mu\phi + \bar{\mu}\bar{\phi} = 0, \tag{3.16}$$

$$\tau\phi + \bar{\pi}\bar{\phi} = 0. \tag{3.17}$$

The consistency conditions for these equations, expressed in terms of the invariants (3.6), are[36]

$$I_1 = I_2 = I_3 = I_4 = 0, \tag{3.18}$$

of which only *three* are algebraically independent in the generic situation. Later in this section it will be shown that $I_1 = I_2 = 0$ for *all* solutions in $\mathfrak{D}_0$, while in general I_3 and I_4 are nonzero. In Sec. 6 it will be seen that $I_3 = I_4 = 0$ for all but the Case III Kinnersley metrics.

The solutions in $\mathfrak{D}_0$ also admit a "test" self-dual Maxwell field

$$\overset{+}{F} := BZ^2, \tag{3.19}$$

where $Z^2 := \theta^1 \wedge \theta^2 - \theta^3 \wedge \theta^4$. It can be shown that

$$dZ^2 = \theta \wedge Z^2, \tag{3.20}$$

where θ is defined by Eq. (2.25). In view of this $\overset{+}{F}$ satisfies the source-free Maxwell equations if and only if the function B is a solution of the equation

$$dB = -B\theta, \tag{3.21}$$

the integrability condition of which is also given by Eq. (3.13). Thus the Eq. (3.21) has a solution if and only if the Eq. (3.12) has one; more explicitly (following Hughston and Sommers[37]) $B = \Psi^{2/3}$ is a solution of (3.21) if and only if Ψ is a solution of (3.12). We note that the two-form $\overset{+}{F}$ defined by Eq. (3.19) is that of the Riemannian–Maxwellian invertible structure possessed by the solutions of $\mathfrak{D}_0$ with non-null group orbits[38]

We next write the NP field equations for the tetrad already chosen taking into account the Eq. (3.11), obtaining

$$D\rho = \rho^2 + (\epsilon + \bar{\epsilon})\rho, \tag{3.22}$$

$$D\tau = (\tau + \bar{\pi})\rho + (\epsilon - \bar{\epsilon})\tau, \tag{3.23}$$

$$D\alpha - \bar{\delta}\epsilon = (\rho - 2\epsilon + \bar{\epsilon})\alpha - \bar{\beta}\epsilon + (\epsilon + \rho)\pi, \tag{3.24}$$

$$D\beta - \delta\epsilon = (\bar{\rho} - \bar{\epsilon})\beta - (\bar{\alpha} - \bar{\pi})\epsilon, \tag{3.25}$$

$$D\gamma - \Delta\epsilon = (\tau + \bar{\pi})\alpha + (\bar{\tau} + \pi)\beta - (\epsilon + \bar{\epsilon})\gamma - (\gamma + \bar{\gamma})\epsilon + \tau\pi + \Psi - \Lambda, \tag{3.26}$$

$$\bar{\delta}\pi = -\pi^2 - (\alpha - \bar{\beta})\pi, \tag{3.27}$$

$$D\mu - \delta\pi = \mu\bar{\rho} + \pi\bar{\pi} - (\epsilon + \bar{\epsilon})\mu - (\bar{\alpha} - \beta)\pi + \Psi + 2\Lambda, \tag{3.28}$$

$$\Delta\pi = -(\bar{\tau} + \pi)\mu - (\gamma - \bar{\gamma})\pi, \tag{3.29}$$

$$\delta\rho = (\bar{\alpha} + \beta)\rho + (\rho - \bar{\rho})\tau, \tag{3.30}$$

$$\delta\alpha - \bar{\delta}\beta = \mu\rho + \alpha\bar{\alpha} + \beta\bar{\beta} - 2\alpha\beta + (\rho - \bar{\rho})\gamma + (\mu - \bar{\mu})\epsilon - \Psi + \Lambda, \tag{3.31}$$

$$\bar{\delta}\mu = -(\mu - \bar{\mu})\pi - (\alpha + \bar{\beta})\mu, \tag{3.32}$$

$$\Delta\mu = -\mu^2 - (\gamma + \bar{\gamma})\mu, \tag{3.33}$$

$$\delta\gamma - \Delta\beta = (\tau - \bar{\alpha} - \beta)\gamma + \mu\tau - (\gamma - \bar{\gamma} - \mu)\beta, \tag{3.34}$$

$$\delta\tau = \tau^2 - (\bar{\alpha} - \beta)\tau, \tag{3.35}$$

$$\Delta\rho - \bar{\delta}\tau = -\bar{\mu}\rho - \tau\bar{\tau} - (\alpha - \bar{\beta})\tau + (\gamma + \bar{\gamma})\rho - \Psi - 2\Lambda, \tag{3.36}$$

$$\Delta\alpha - \bar{\delta}\gamma = (\bar{\gamma} - \bar{\mu})\alpha + (\bar{\beta} - \bar{\tau})\gamma. \tag{3.37}$$

By expanding the left-hand side of Eq. (3.13) using (2.25), we obtain the following equations which are independent of the field equations:

$$\Delta\rho + D\mu = (\gamma + \bar{\gamma})\rho - (\epsilon + \bar{\epsilon})\mu + (\pi\bar{\pi} - \tau\bar{\tau}), \tag{3.38}$$

$$\delta\mu + \Delta\tau = (\gamma - \bar{\gamma})\tau - (\bar{\alpha} + \beta)\mu, \tag{3.39}$$

$$\bar{\delta}\rho + D\pi = (\alpha + \bar{\beta})\rho - (\epsilon - \bar{\epsilon})\pi, \tag{3.40}$$

and in addition,

$$\bar{\delta}\tau + \delta\pi = (\bar{\alpha} - \beta)\pi + (\alpha - \bar{\beta})\tau + (\bar{\mu}\rho - \mu\bar{\rho}), \tag{3.41}$$

which follows from (3.28), (3.36), and (3.38).

It should be noted at this point that in the *electrovac* case, in contrast to the present case, the integrability condi-

tions for the Bianchi identities are not identical to those for Maxwell's equations; by applying the NP commutators to the quantity Ψ one obtains three additional equations which are independent of the field equations and the integrability conditions for Maxwell's equations [namely, DM equations (4.18a), (4.18c), and (4.18d)]. One of these extra equations immediately yields the *key equations*

$$I_1 = \pi\bar{\pi} - \tau\bar{\tau} = 0, \tag{3.42}$$

$$I_2 = \bar{\mu}\rho - \mu\bar{\rho} = 0. \tag{3.43}$$

The remainder of this section will be devoted to proving that (3.42) and (3.43) also hold in the present case. This is done by an examination of cases involving the quantities μ, ρ, τ, and π. It happens that many of these cases are impossible. For example, if $\rho = \tau = 0$ then Eq. (3.36) implies that $\Psi = -2\Lambda = \text{const}$. If we then assumed either $\mu \neq 0$ or $\pi \neq 0$ we would reach a contradiction via (3.9) or (3.10), respectively. In what follows we mention only the admissible cases.

A. Case I(a). $\pi\ \rho \neq 0, \quad \mu \neq 0$

Equation (3.23) immediately implies that $\tau \neq 0$. By means of a tetrad transformation of the form (3.1), we may set (on dropping the primes on the transformed quantities)

$$\rho = \pi. \tag{3.44}$$

It then follows from (3.23), (3.27), and (3.40) that

$$\alpha = \epsilon. \tag{3.45}$$

Comparison of (3.29) and (3.38) yields

$$D\mu = 2\gamma\rho - (\alpha + \bar{\alpha})\mu + (\rho + \bar{\tau})\mu + (\rho\bar{\rho} - \tau\bar{\tau}), \tag{3.46}$$

while the difference of (3.30) and (3.41) gives

$$\bar{\delta}\tau = -2\beta\rho + (\alpha - \bar{\beta})\tau - (\rho - \bar{\rho})\tau + (\bar{\mu}\rho - \mu\bar{\rho}). \tag{3.47}$$

Finally, one obtains from (3.29), (3.36), and (3.47) the equation

$$\Psi + 2\Lambda = 2\rho(\gamma - \beta) + (\mu - \tau)(\rho - \bar{\rho} + \bar{\tau}). \tag{3.48}$$

To obtain further conditions, it is necessary to examine some of the commutators of ρ. The $[\Delta, D]$ commutator yields

$$\Delta[\rho^2 + (\alpha + \bar{\alpha})\rho] - D[-(\rho + \bar{\tau})\mu - (\gamma - \bar{\gamma})\rho]$$
$$= (\gamma + \bar{\gamma})D\rho + (\alpha + \bar{\alpha})\Delta\rho - (\bar{\rho} + \tau)\bar{\delta}\rho - (\rho + \bar{\tau})\delta\rho,$$

or

$$\rho(\Delta\alpha + D\gamma) - \rho(D\bar{\gamma} - \Delta\bar{\alpha}) + 2\rho\Delta\rho + (\rho + \bar{\tau})D\mu + \mu D\bar{\tau}$$
$$+ (\mu - 2\bar{\gamma})D\rho + (\bar{\rho} + \tau)\bar{\delta}\rho + (\rho + \bar{\tau})\delta\rho = 0.$$

On account of Eq. (3.26) this becomes

$$2\rho\Delta\alpha = \rho[2\alpha(\gamma + \bar{\gamma}) - 2\beta(\rho + \bar{\tau}) - 2\gamma\bar{\tau} - (\Psi - \bar{\Psi})$$
$$- \mu\bar{\rho} - (\rho - \bar{\rho})\tau] + \bar{\tau}[2\alpha\mu + (\bar{\rho} + \bar{\tau})(\tau - \mu)]. \tag{3.49}$$

Similarly, by applying the $[\bar{\delta}, D]$ and $[\bar{\delta}, \delta]$ commutators and using Eqs. (3.24) and (3.31) respectively we obtain

$$D\alpha = \alpha(\rho - \alpha + \bar{\alpha}) + \tfrac{1}{2}\rho^2, \tag{3.50}$$

$$\bar{\delta}\alpha = \alpha(\alpha + \bar{\beta} - \rho) - \tfrac{1}{2}\rho^2, \tag{3.51}$$

and

$$2\rho\delta\alpha = \rho[2\alpha(\bar{\alpha} - \beta) + 2\gamma(\rho - \bar{\rho}) - (\Psi - \bar{\Psi}) + \mu\rho$$
$$- 2\beta\bar{\rho} + \bar{\tau}(\mu - \tau)] + \bar{\rho}[2\alpha\tau + (\tau - \mu)(\bar{\rho} + \bar{\tau})]. \tag{3.52}$$

Since $\rho \neq 0$, Eq. (3.50) implies that $\alpha \neq 0$. Therefore we can apply the $[\Delta, D]$ commutator to the quantity α to obtain

$$\Delta[\alpha(\rho - \alpha + \bar{\alpha}) + \tfrac{1}{2}\rho^2] - D[\alpha(\gamma + \bar{\gamma}) - \beta(\rho + \bar{\tau}) - \tfrac{1}{2}(\Psi$$
$$- \bar{\Psi}) - \tfrac{1}{2}\mu\bar{\rho} - \gamma\bar{\tau} - \tfrac{1}{2}\tau(\rho - \bar{\rho}) + \tfrac{1}{2}\bar{\tau}(2\alpha\mu + (\tau - \mu)(\bar{\rho} + \bar{\tau}))/\rho]$$
$$= (\gamma + \bar{\gamma})D\alpha + (\alpha + \bar{\alpha})\Delta\alpha - (\bar{\rho} + \tau)\bar{\delta}\alpha - (\rho + \bar{\tau})\delta\alpha.$$

Upon performing the required differentiations, substituting for the NP derivatives using the equations obtained thus far, substituting for the quantity Ψ using (3.48), and simplifying, we obtain

$$\bar{\Psi}(2\alpha - \rho + \bar{\rho} - \bar{\tau}) = 0.$$

Since $\Psi \neq 0$, we then have

$$\tau = 2\bar{\alpha} + \rho - \bar{\rho}. \tag{3.53}$$

This in turn yields

$$\Delta\tau = 2\Delta\bar{\alpha} + \Delta\rho - \Delta\bar{\rho}, \tag{3.54}$$

and so by (3.39)

$$\delta\mu = (\gamma - \bar{\gamma})\tau - (\bar{\alpha} + \beta)\mu - 2\Delta\bar{\alpha} - \Delta\rho + \Delta\bar{\rho}. \tag{3.55}$$

Now, by examining three further commutators, we can solve for the remaining unknown NP derivatives (namely, $\Delta\beta$, $\Delta\gamma$, and $\delta\beta$). First,

$$[\delta, \Delta]\rho = \delta[-(\rho + \bar{\tau})\mu - (\gamma - \bar{\gamma})\rho] - \Delta[(\bar{\alpha} + \beta)\rho$$
$$+ (\rho - \bar{\rho})\tau]$$
$$= (\tau - \bar{\alpha} - \beta)\Delta\rho + (\mu - \gamma + \bar{\gamma})\delta\rho,$$

which on account of (3.34), (3.37), and (3.39) yields

$$2\rho\Delta\beta = -\bar{\tau}\delta\mu - \mu(\delta\bar{\tau} + 2\delta\rho) + \bar{\rho}\Delta\tau + \tau(\Delta\bar{\rho}$$
$$- 2\Delta\rho) + \rho[2\bar{\alpha}\mu - 2(\gamma - \bar{\gamma})(\tau - \beta) - \mu\tau]. \tag{3.56}$$

Since $\mu \neq 0$, we may apply $[\Delta, D]$ to μ to obtain, via Eq. (3.26),

$$2\rho\Delta\gamma = -(2\gamma + \mu + \bar{\rho})\Delta\rho + [2(\alpha + \bar{\alpha}) - \bar{\tau} - \rho]\Delta\mu$$
$$+ (\tau - \mu)\Delta\bar{\tau} - 2\mu D\mu - \rho\Delta\bar{\rho} + \bar{\tau}\Delta\tau$$
$$- (\bar{\rho} + \tau)\bar{\delta}\mu - (\rho + \bar{\tau})\delta\mu - \mu[(\bar{\rho} + \tau)(\alpha + \bar{\beta})$$
$$+ (\rho + \bar{\tau})(\bar{\alpha} + \beta) - 2(\alpha + \bar{\alpha})(\gamma + \bar{\gamma})$$
$$+ (\tau\rho + \bar{\tau}\bar{\rho}) + (\Psi + \bar{\Psi}) - 2\Lambda]. \tag{3.57}$$

Finally, by applying $[\bar{\delta}, \delta]$ to τ we find

$$2\rho\delta\beta = -2\tau\bar{\delta}\tau + (\bar{\mu} - 2\beta - \tau)\delta\rho + (\tau - \mu)\delta\bar{\rho}$$
$$+ [2(\alpha - \bar{\beta}) - \rho + \bar{\rho}]\delta\tau - \bar{\rho}\delta\mu + \rho\delta\bar{\mu}$$
$$- (\mu - \bar{\mu})D\tau - (\rho - \bar{\rho})\Delta\tau + \tau[\mu\rho + \bar{\mu}\bar{\rho}$$
$$+ 2\alpha\bar{\alpha} + 2\beta\bar{\beta} - 2(\alpha\beta + \bar{\alpha}\bar{\beta}) + (\rho - \bar{\rho})(\gamma - \bar{\gamma})$$
$$+ (\mu - \bar{\mu})(\alpha - \bar{\alpha}) - (\Psi + \bar{\Psi}) + 2\Lambda]. \tag{3.58}$$

The last step in the procedure (which requires a good deal of algebra[39]) is to apply the commutator $[\delta, \Delta]$ to the quantity α. Following the same approach as that used to obtain (3.53), we find the relation

$$(\bar{\Psi}/\rho)[\rho\bar{\rho} - \tau\bar{\tau} + \mu\bar{\rho} - \bar{\mu}\rho] = 0,$$

which implies that

$$I_1 = \pi\bar{\pi} - \tau\bar{\tau} = \rho\bar{\rho} - \tau\bar{\tau} = 0$$

and

$$I_2 = \bar{\mu}\rho - \mu\bar{\rho} = 0.$$

B. Case I(b). $\pi\,\rho\neq 0,\quad \mu = 0$

As before, we have $\tau\neq 0$ by (3.23). We remark that the case $\tau\pi\mu\neq 0,\ \rho = 0$ is also admissible, but represents the same situation as the present case (since by interchanging the tetrad vectors l and n we pass from one to the other).

Clearly Eq. (3.43) holds trivially. Since $\pi\,\rho\tau\neq 0$, we may choose the tetrad of the previous case and proceed to obtain Eq. (3.53). Taking $\bar{\delta}$ of both sides of this equation yields

$$(\rho+\bar{\rho})(\pi\bar{\pi}-\tau\bar{\tau}) = 0.$$

If $(\rho+\bar{\rho}) = 0$, we have

$$\pi\bar{\pi}-\tau\bar{\tau} = 0$$

from Eq. (3.38). Thus

$$I_1 = I_2 = 0.$$

C. Case II(a). $\pi = \tau = 0,\quad \mu\,\rho\neq 0$

Eq. (3.43) follows from Eq. (3.41).

D. Case II(b). $\rho = \mu = 0,\quad \pi\tau\neq 0$

Eq. (3.42) follows from Eq. (3.38).

E. Case III. $\rho = \mu = \tau = \pi = 0$

The key equations hold trivially in this case. This completes the proof of the validity of the Eqs. (3.42) and (3.43).

4. PROOF OF THEOREM 1 FOR NON-NULL ORBITS

In order to prove Theorem 1 in the case of non-null orbits, we shall use the following theorem implicit in the work of Debever, McLenaghan, and Tariq,[40] which is essentially a restatement of their main theorem:

Theorem 3: If any space-time V_4 admits a local null tetrad $(l, n, m, \bar{m})$ in which the NP spin coefficients and curvature components satisfy the following relations:

$$\mu = -e\rho,\ \pi = -e\tau,\ \nu = -e\kappa,\ \lambda = -e\sigma,\quad \gamma = -e\epsilon,\ \beta = -e\alpha, \tag{4.1}$$

$$A := 2(\bar{\alpha}-e\alpha)+(e\kappa-\bar{\kappa})-(\tau-e\bar{\tau}) = 0,$$

$$iE := 2(\epsilon-\bar{\epsilon})-(\rho-\bar{\rho})+e(\sigma-\bar{\sigma}) = 0, \tag{4.2}$$

$$\Phi_{22} = \Phi_{00},\ \Phi_{02} = \Phi_{20},\ \Phi_{21} = \Phi_{01}, \tag{4.3}$$

$$\Psi_4 = \Psi_0,\ \Psi_3 = \Psi_1, \tag{4.4}$$

$$\mathscr{D}\rho = \mathscr{D}\sigma = \mathscr{D}\kappa = \mathscr{D}\tau = \mathscr{D}\alpha = \mathscr{D}\epsilon = 0,$$

$$\mathscr{L}\rho = \mathscr{L}\sigma = \mathscr{L}\kappa = \mathscr{L}\tau = \mathscr{L}\alpha = \mathscr{L}\epsilon = 0, \tag{4.5}$$

where

$$e^2 = 1,\ \mathscr{D} := D - e\Delta \text{ and } \mathscr{L} := \delta - e\bar{\delta},$$

then V_4 possesses locally an invertible 2-parameter abelian isometry group and there exist coordinates (u, v, w, x) such that the metric has the form (2.3).

We now note that by choosing a null tetrad as in Sec. 3, and in view of (2.25), we have (4.3), (4.4), and parts of (4.1) and (4.5) already satisfied. Thus we need only establish that (4.2) and the remainders of (4.1) and (4.5) may be satisfied using the tetrad freedom given by (3.1). It bears mention that this task is somewhat simpler in the electrovac case because of DM Eqs. (4.18c, 4.18d) and (4.21b–4.21e).

A. Case I. $\pi\,\rho\mu\neq 0$

On account of Eqs. (3.42) and (3.43) we may use the available tetrad freedom to set

$$\mu = -e\rho, \tag{4.6}$$

$$\pi = -e\tau. \tag{4.7}$$

This fixes the tetrad up to the transformation

$$l' = l,\quad n' = n,\quad m' = -m.$$

Now Eqs. (3.22), (3.33), and (3.38) imply that

$$\rho[(\gamma+\bar{\gamma})+e(\epsilon+\bar{\epsilon})] = 0,$$

while Eqs. (3.27), (3.35), and (3.41) yield

$$\tau[(\alpha-\bar{\beta})-e(\bar{\alpha}-\beta)] = 0,$$

so that (since $\rho\tau\neq 0$) we have

$$(\gamma+\bar{\gamma})+e(\epsilon+\bar{\epsilon}) = 0, \tag{4.8}$$

$$(\bar{\alpha}-\beta)-e(\alpha-\bar{\beta}) = 0. \tag{4.9}$$

The equations (3.29), (3.30), and (3.39) imply

$$\bar{\rho}\tau - e\rho\bar{\tau} = 2[(\bar{\alpha}+\beta)\rho+e(\gamma-\bar{\gamma})\tau], \tag{4.10}$$

and similarly (3.23), (3.32), and (3.40) imply

$$\rho\bar{\tau} - e\bar{\rho}\tau = 2[(\alpha+\bar{\beta})\rho+e(\epsilon-\bar{\epsilon})\tau]. \tag{4.11}$$

Then (4.10) and (4.11) together yield

$$\rho[(\bar{\alpha}+\beta)+e(\alpha+\bar{\beta})]+\tau[(\epsilon-\bar{\epsilon})+e(\gamma-\bar{\gamma})] = 0. \tag{4.12}$$

Taking the conjugate of (4.12) and multiplying by e yields

$$\bar{\rho}[(\bar{\alpha}+\beta)+e(\alpha+\bar{\beta})]-e\bar{\tau}[(\epsilon-\bar{\epsilon})+e(\gamma-\bar{\gamma})] = 0. \tag{4.13}$$

The Eqs. (4.12) and (4.13) form a system whose determinant may be zero or nonzero.

1. Sub-case (a). $I_3 = -(\bar{\rho}\tau+e\rho\bar{\tau}) = 0$ (4.14)

Differentiation of (4.14) with respect to D and Δ yields

$$(\bar{\rho}^2\tau+e\rho^2\bar{\tau})+(\epsilon-\bar{\epsilon})(\bar{\rho}\tau-e\rho\bar{\tau}) = 0 \tag{4.15}$$

and

$$(\bar{\rho}^2\tau+e\rho^2\bar{\tau})-e(\gamma-\bar{\gamma})(\bar{\rho}\tau-e\rho\bar{\tau}) = 0, \tag{4.16}$$

respectively. Subtracting the second of these relations from the first yields

$$2\bar{\rho}\tau[(\epsilon-\bar{\epsilon})+e(\gamma-\bar{\gamma})] = 0,$$

which implies (since $\bar{\rho}\tau - e\rho\bar{\tau} = 2\bar{\rho}\tau\neq 0$)

$$(\epsilon-\bar{\epsilon})+e(\gamma-\bar{\gamma}) = 0. \tag{4.17}$$

This, with (4.8), then gives

$$\gamma = -e\epsilon, \tag{4.18}$$

and with (4.12) implies

$$(\bar{\alpha}+\beta)+e(\alpha+\bar{\beta}) = 0. \tag{4.19}$$

Eqs. (4.19) and (4.9) then yield

$$\beta = -e\alpha. \tag{4.20}$$

When Eqs. (4.18) and (4.20) are taken into account Eq. (4.10) takes the form

$$A\rho - iE\tau = 0. \tag{4.21}$$

Since $\bar{A} = -eA$ we also have

$$A\bar{\rho} - iE(e\bar{\tau}) = 0. \tag{4.22}$$

The determinant of the system (4.21) and (4.22) is $I_5 = \bar{\rho}\tau - e\rho\bar{\tau}$, which is nonzero by (4.14). Hence

$$A = iE = 0. \tag{4.23}$$

We have now satisfied all of the conditions (4.1) and (4.2) of Theorem 3.

Attention will now be turned to the differential conditions (4.5). It follows from the field equations that

$$\mathscr{D}\rho = \mathscr{D}\tau = \mathscr{L}\rho = \mathscr{L}\tau = 0. \tag{4.24}$$

Therefore from (4.23) we have

$$\begin{aligned}\mathscr{D}(\epsilon - \bar{\epsilon}) = \mathscr{D}(\bar{\alpha} - e\alpha) = \mathscr{L}(\epsilon - \bar{\epsilon}) \\ = \mathscr{L}(\bar{\alpha} - e\alpha) = 0.\end{aligned} \tag{4.25}$$

Now by applying the commutator

$$[\Delta, D] = -e(\epsilon + \bar{\epsilon})\mathscr{D} - (\bar{\tau} - e\tau)\mathscr{L}$$

to the quantity ρ we obtain

$$\mathscr{D}(\epsilon + \bar{\epsilon}) = 0.$$

Similarly, by applying the commutator

$$[\bar{\delta}, \delta] = e(\rho - \bar{\rho})\mathscr{D} + (\alpha + e\bar{\alpha})\mathscr{L}$$

to τ we obtain

$$\mathscr{L}(\bar{\alpha} + e\alpha) = 0.$$

Combining these relations with (4.25) yields

$$\mathscr{D}\epsilon = \mathscr{L}\alpha = 0. \tag{4.26}$$

From Eqs. (3.24), (3.25), (3.34), and (3.36) we obtain

$$e\mathscr{D}\alpha + \mathscr{L}\epsilon = 0, \tag{4.27}$$

and hence (by taking the conjugate and multiplying be e)

$$\mathscr{D}\bar{\alpha} - \mathscr{L}\bar{\epsilon} = 0. \tag{4.28}$$

The sum of (4.27) and (4.28) yields, in view of (4.25),

$$\mathscr{D}(\bar{\alpha} + e\alpha) = 0,$$

and so

$$\mathscr{L}\epsilon = \mathscr{D}\alpha = 0. \tag{4.29}$$

Thus we have established all of the conditions of Theorem 3.

2. Sub-case (b). $I_3 = -(\bar{\rho}\tau + e\rho\bar{\tau}) \neq 0$ (4.30)

Since the determinant is nonzero, the only solution to the Eqs. (4.13) and (4.12) is

$$(\alpha + \bar{\beta}) + e(\bar{\alpha} + \beta) = 0, \tag{4.31}$$

$$(\gamma - \bar{\gamma}) + e(\epsilon - \bar{\epsilon}) = 0. \tag{4.32}$$

These equations together with (4.8) and (4.9) imply

$$\beta = -e\alpha, \tag{4.33}$$

and

$$\gamma = -e\epsilon. \tag{4.34}$$

Now, if $I_5 = \bar{\rho}\tau - e\rho\bar{\tau} \neq 0$ we may proceed as in the previous case since the proof of the conditions (4.5) is independent of the condition (4.14). Hence we assume that

$$I_5 = \bar{\rho}\tau - e\rho\bar{\tau} = 0. \tag{4.35}$$

Differentiating (4.35) by D yields

$$2(\epsilon - \bar{\epsilon})\tau + e(e\tau - \bar{\tau})\rho = 0, \tag{4.36a}$$

which by (4.35) implies

$$2(\epsilon - \bar{\epsilon}) + (\rho - \bar{\rho}) = 0. \tag{4.36b}$$

Thus we have

$$iE = -2(\rho - \bar{\rho}). \tag{4.37}$$

Differentiation of (4.35) by δ yields

$$(\rho - \bar{\rho})\tau - 2(e\alpha - \bar{\alpha})\rho = 0 \tag{4.38a}$$

which implies

$$(\tau - e\bar{\tau}) - 2(e\alpha - \bar{\alpha}) = 0. \tag{4.38b}$$

Thus

$$A = 2(e\bar{\tau} - \tau). \tag{4.39}$$

The field equations again imply

$$\mathscr{L}\rho = \mathscr{L}\tau = \mathscr{D}\rho = \mathscr{D}\tau = 0; \tag{4.40}$$

then by applying the commutator

$$[\mathscr{L}, \mathscr{D}] = iE[(\tau/\rho)(D + e\Delta) - (\delta + e\bar{\delta})]$$

to the quantities ρ and τ we obtain

$$iE\tau[2(\epsilon + \bar{\epsilon}) + (\rho + \bar{\rho})] = 0 \tag{4.41a}$$

and

$$-iE\tau[(\tau + e\bar{\tau}) - 2(\bar{\alpha} + e\alpha)] = 0. \tag{4.41b}$$

Suppose that $iE \neq 0$. Then along with the Eq. (4.38), the above relations imply

$$2\epsilon + \rho = 2\alpha - e\tau = 0. \tag{4.42}$$

Combining this with Eqs. (3.22), (3.26), and (3.33) we find

$$-e\tau\bar{\tau} + \rho\bar{\rho} + e(\Psi - \Lambda) = 0,$$

which implies

$$\Psi = \bar{\Psi}.$$

Hence by (3.9) we would have

$$\rho = \bar{\rho},$$

and thus $iE = 0$, which is a contradiction. We conclude that $iE = 0$ [and so by Eq. (4.21), $A = 0$ as well], and note that by Eqs. (4.36b) and (4.38b) we also have

$$\rho - \bar{\rho} = \epsilon - \bar{\epsilon} = \bar{\tau} - e\tau = \bar{\alpha} - e\alpha = 0$$

in this case. The remainder of the proof is now identical to that of the previous case.

3. Case II(a). $\pi = \tau = 0, \quad \mu\rho \neq 0$

The proof in this case is identical to that of the electrovac case, where Eqs. (3.39) and (3.40) replace DM Eqs. (4.21e) and (4.21b), respectively.

4. Case II(b). $\pi\tau \neq 0, \quad \mu = \rho = 0$

Again the proof is the same as for the electrovac case, with Eqs. (3.39) and (3.40) replacing DM Eqs. (4.21d) and (4.21c), respectively.

5. Case III. $\mu = \rho = \pi = \tau = 0$

As none of the DM Eqs. (4.18) or (4.21) apply in this case, the proof is once more identical to that of DM. We note

that by Eq. (3.28)

$$\Psi = -2\Lambda,$$

so that this case is nontrivial only when $\Lambda \neq 0$.

5. PROOF OF THEOREM 1 FOR NULL ORBITS

We now wish to prove Theorem 1 in the case of null group orbits, i.e., $\rho \neq 0, \mu = 0$, or vice versa. Since one finds that the case $\rho \neq 0, \mu = \pi = 0$, is impossible, we consider the case $\pi\rho \neq 0, \mu = 0$ (cf. Case I(b) of Sec. 3).

It suffices to demonstrate that the proof for the electrovac case holds in the vacuum case as well, despite the absence of DM Eqs. (4.21b and 4.21c). This amounts to establishing that the relation

$$(\rho - \bar{\rho}) - 2(\epsilon - \bar{\epsilon}) = 0 \tag{5.1}$$

holds for our choice of tetrad.

In view of (3.42), we may use the transformation (3.1) to set

$$\pi = \tau. \tag{5.2}$$

This is preserved by further transformations of the form

$$l' = e^a l, \quad n' = e^{-a} n, \quad m' = m. \tag{5.3}$$

Now Eqs. (3.29) and (3.39) imply

$$\gamma = \bar{\gamma} \tag{5.4}$$

and Eqs. (3.27), (3.28), (3.35), (3.36), and (3.38) together imply that

$$\alpha + \bar{\alpha} = \beta + \bar{\beta}. \tag{5.5}$$

On account of (5.4) we may choose the function a in (5.3) to set

$$\gamma = 0, \tag{5.6}$$

which is preserved by further transformations (5.3) such that

$$\Delta a = 0. \tag{5.7}$$

We now wish to make such a transformation to set

$$\tau + \bar{\tau} = 2(\alpha + \bar{\beta}). \tag{5.8}$$

This requires that, in addition to (5.7), the function α satisfy

$$\delta a = \tfrac{1}{2}(\tau + \bar{\tau}) - (\bar{\alpha} + \beta), \tag{5.9a}$$

$$\bar{\delta} a = \tfrac{1}{2}(\tau + \bar{\tau}) - (\alpha + \bar{\beta}). \tag{5.9b}$$

This is indeed possible since the commutation relations

$$[\delta, \Delta] a = (\tau - \bar{\alpha} - \beta)\Delta a$$

and

$$[\bar{\delta}, \delta] a = (\bar{\rho} - \rho)\Delta a - (\bar{\alpha} - \beta)\bar{\delta} a + (\alpha - \bar{\beta})\delta a$$

hold. We note that (5.6) and (5.8) are preserved by further transformations satisfying

$$\Delta a = \delta a = 0, \tag{5.10}$$

and also that (5.5) and (5.8) together imply

$$\alpha = \beta. \tag{5.11}$$

Hence Eq. (5.8) reads

$$\tau + \bar{\tau} = 2(\alpha + \bar{\alpha}). \tag{5.12}$$

If we now apply the commutator $[\Delta, D]$ to ρ, we obtain

$$\Delta(\epsilon + \bar{\epsilon}) = -(\tau/\rho)(\tau + \bar{\tau})[(\rho - \bar{\rho}) - 2(\epsilon - \bar{\epsilon})],$$

which implies that

$$(\tau + \bar{\tau})[(\tau/\rho) + (\bar{\tau}/\bar{\rho})][(\rho - \bar{\rho}) - 2(\epsilon - \bar{\epsilon})] = 0. \tag{5.13}$$

By Eqs. (3.27), (3.36), and (3.38) we also have

$$\Psi + 2\Lambda + 4\alpha\tau = 0. \tag{5.14}$$

If we assume

$$\tau + \bar{\tau} = 0, \tag{5.15}$$

then by Eqs. (5.12) and (5.14) we have

$$\Psi = \bar{\Psi}. \tag{5.16}$$

Equation (3.9) then implies that

$$\rho = \bar{\rho}; \tag{5.17}$$

taking D of both sides of Eq. (5.15) yields

$$\epsilon = \bar{\epsilon}, \tag{5.18}$$

which establishes Eq. (5.1).

If, on the other hand, we suppose that $\tau + \bar{\tau} \neq 0$, and

$$I_5 = \rho\bar{\tau} + \bar{\rho}\tau = 0, \tag{5.19}$$

then by taking D of both sides of this relation we obtain

$$(\rho - \bar{\rho}) + 2(\epsilon - \bar{\epsilon}) = 0. \tag{5.20}$$

By Eqs. (3.10), (3.27), and (3.35) we have

$$(\delta + \bar{\delta})\Psi = (\delta + \bar{\delta})\tau = 0; \tag{5.21a}$$

TABLE I. Invariant characterization of Kinnersley's metrics.

Quantity							Kinnersley
ρ	τ	μ	I_3	I_5	Case	Examples	case
$\neq 0$	0	–	–	–	II(a)	NUT metrics	I
$\neq 0$	$\neq 0$	$\neq 0$ / 0	0	–	I(a) / null orbits	Kerr metric	II
$\neq 0$	$\neq 0$	$\neq 0$ / 0	$\neq 0$	0	I(b)(ii) / null orbits	C metric	III.A
$\neq 0$	$\neq 0$	$\neq 0$ / 0	$\neq 0$	$\neq 0$	I(b)(i) / null orbits	C-NUT metric	III.B
0	$\neq 0$	0	–	–	II(b)	B metric	IV
0	0	0	–	–	III	Robinson[41] Bertotti[42]	–

hence by Eq. (5.14) we also have

$$(\delta + \bar{\delta})\alpha = 0. \tag{5.21b}$$

Now, by applying the commutator $[\bar{\delta},\delta]$ to ρ, and noting Eq. (5.21b), we obtain

$$\tau(\alpha\rho + \bar{\alpha}\bar{\rho}) = 0.$$

In view of Eqs. (5.12) and (5.19) this becomes

$$(\tau - 2\bar{\alpha})(\alpha + \bar{\alpha}) = 0. \tag{5.22}$$

Since

$$\tau + \bar{\tau} = 2(\alpha + \bar{\alpha}) \neq 0,$$

we have

$$\tau - 2\bar{\alpha} = 0. \tag{5.23}$$

Immediately Eq. (5.14) implies

$$\Psi = \bar{\Psi},$$

which, along with Eqs. (3.9) and (5.20), once again establishes (5.1). The remainder of the proof is now identical to that given in DM.

6. CORRESPONDENCE WITH KINNERSLEY'S RESULTS

Because of the transformation laws (3.2), (3.3), and (3.7) the NP spin coefficients ρ, μ, τ, and π, and the quantities I_3, I_4, and I_5 defined by Eqs. (3.6) may be used to obtain a partial invariant characterization of Kinnersley's metrics. We first note that since $I_1 = 0$ for all solutions in $\mathfrak{D}_0$ it follows that $\tau = 0$ if and only if $\pi = 0$ [both of these equations being invariant under the general tetrad transformation (3.1)]. In terms of Kinnersley's quantities τ^0 and π^0 it follows from his Eqs. (3.5), (3.6c), and (3.25) that $\tau^0 = \pi^0 = 0$ iff $\tau = 0$ (or $\pi = 0$). Hence Kinnersley's Case I metrics are characterized by $\rho \neq 0$ and $\tau = 0$ [it follows from Eqs. (3.9) and (3.28) that $\mu \neq 0$ in this case], while the Case IV metrics are characterized by $\rho = \mu = 0$, and $\tau \neq 0$. For the Cases II and III one has $\rho\tau \neq 0$; in order to distinguish these cases one employs the quantity I_3 which in view of Kinnersley's Eqs. (3.1), (3.5), (3.6c), and (3.25) may be expressed as

$$I_3 = 2\bar{\rho}\pi^0. \tag{6.1}$$

Thus Kinnersley's Cases II and III are characterized, respectively, by the conditions $I_3 = 0$ and $I_3 \neq 0$. Similarly, the Case III.A and Case III.B metrics may be distinguished according to whether I_5 vanishes or not. The above results are summarized in Table I. The reader will observe that in the table we have distinguished in Cases II and III between $\mu = 0$ and $\mu \neq 0$. This distinction corresponds to the orbits of the two-parameter abelian group of isometries being null or non-null respectively.[43] In the case $\rho\mu \neq 0$ (non-null orbits) a symmetric tetrad can be introduced in which Eqs. (4.6) and (4.7) hold implying

$$I_3 = -eI_4. \tag{6.2}$$

It follows that $I_3 = 0$ iff $I_4 = 0$. This implies that the invariant I_4 also vanishes for the Kinnersley Case II metrics, a fact which is not entirely obvious from Kinnersley's equations.

ACKNOWLEDGMENTS

We would like to express our appreciation to Professor R. Debever and Dr. N. Tariq for many enlightening discussions. One of us (S.C.) would like to thank the Natural Sciences and Engineering Research Council of Canada for the award of a Postgraduate Scholarship.

[1] W. Kinnersley, J. Math. Phys. **10**, 1195 (1969).
[2] L. P. Hughston and P. Sommers, Commun. Math. Phys. **33**, 129 (1973).
[3] G. J. Weir and R. P. Kerr, Proc. R. Soc. London Ser. A **355**, 31 (1977).
[4] M. Walker and R. Penrose, Commun. Math. Phys. **18**, 265 (1970).
[5] L. P. Hughston, R. Penrose, P. Sommers, and M. Walker, Commun. Math. Phys. **27**, 303 (1972).
[6] L. P. Hughston and P. Sommers, Commun. Math. Phys. **32**, 147 (1973).
[7] M. Demianski and M. Francaviglia, J. Phys. A: Math. Gen. **14**, 173 (1981).
[8] B. Carter, Phys. Rev. **174**, 1559 (1968).
[9] B. Carter, contribution to *Black Holes*, edited by B. DeWitt and C. DeWitt (Gordon and Breach, New York, 1973), p. 118.
[10] D. Matravers, Util. Math. **2**, 55 (1972).
[11] N. M. J. Woodhouse, Commun. Math. Phys. **44**, 9 (1975).
[12] Ref. 7.
[13] R. Debever and R. G. McLenaghan, J. Math. Phys. **22**, 1711 (1981).
[14] I. Robinson and A. Schild, J. Math. Phys. **4**, 484 (1962).
[15] W. Kundt and A. H. Thompson, C. R. Acad. Sci. (Paris) **254**, 4257 (1962).
[16] E. Newman and R. Penrose, J. Math. Phys. **3**, 566 (1962).
[17] See, for example, R. G. McLenaghan, and Ph. Spindel, Phys. Rev. D **20**, 409 (1979) and the references therein.
[18] See also M. Demianski and M. Francaviglia, Int. J. Theor. Phys. **19**, 675 (1980).
[19] R. Debever, R. G. McLenaghan, and N. Tariq, Gen. Rel. Grav. **10**, 853 (1979).
[20] R. Debever, in *Proceedings of the Conference on Differential Geometrical Methods in Mathematical Physics, Lecture Notes in Mathematics No. 570* (Springer, Berlin, 1977), p. 518.
[21] R. Debever, Bull. Cl. Sci. Acad. R. Belg. **LXII**, 662 (1976).
[22] R. Debever, Cah. Phys. **168–169**, 303 (1964).
[23] M. Cahen, R. Debever, and L. Defrise, J. Math. Mech. **16**, 761 (1967).
[24] Ref. 19.
[25] An electrovac generalization of this solution is given in Ref. 13, Eq. (2.26). Another electrovac solution with null group orbits and nonvanishing cosmological constant has been given by R. Debever and N. Kamran, Bull. Cl. Sci. Acad. R. Belg. **LXVI**, 585 (1980), Eqs. (4.3) and (7.15)–(7.17).
[26] See also Refs. 11 and 7 where this separability property is established indirectly from the existence of a Killing tensor for these metrics.
[27] Ref. 4.
[28] Refs. 5 and 6.
[29] We employ the conventions and notations of Penrose for two-component spinors. See R. Penrose, in *Battelle Rencontres 1967 Lectures in Mathematics and Physics*, edited by C. M. DeWitt and J. A. Wheeler (Benjamin, New York, 1968), p. 121.
[30] R. Penrose, Ann. New York Acad. Sci. **224**, 125 (1973).
[31] Ref. 17.
[32] B. Carter and R. G. McLenaghan, Phys. Rev. D **19**, 1093 (1979).
[33] Ref. 30.
[34] P. Sommers, "Killing tensors and type {2, 2} spacetimes" (Ph.D. thesis, University of Texas at Austin, 1973).
[35] Ref. 4.
[36] See also C. D. Collinson, Int. J. Theor. Phys. **15**, 311 (1976).
[37] Ref. 6.
[38] Refs. 19 and 21.
[39] This and other calculations in this section have been checked by computer using the ALTRAN language.
[40] Ref. 19.
[41] I. Robinson, Bull. Acad. Pol. Sci. Ser. Sci. Math. Astr. Phys. **7**, 351 (1959).
[42] B. Bertotti, Phys. Rev. **116**, 1331 (1959).
[43] See Theorem 1 Sec. 2 Eqs. (2.4) and (2.3).

NUT 4-momenta are forever[a)]

Abhay Ashtekar[b)]
Physics Department, Syracuse University, Syracuse, New York 13210 and
Département de Physique, Université de Clermont-Fd., 63170 Aubière, France

Amitabha Sen[c)]
Department of Physics and Astronomy, University of Maryland, College Park, Maryland 20742

(Received 8 April 1982; accepted for publication 21 May 1982)

The asymptotic structure of the gravitational field at null infinity is re-examined by allowing certain potentials to develop "wire singularities", keeping the physical fields smooth. This relaxation of the regularity conditions leads to the introduction of the Newman–Unti–Tamburino (NUT) 4-momentum which is the "magnetic" or the "dual" counterpart of the Bondi–Sachs 4-momentum. It is shown that, unlike the Bondi–Sachs 4-momentum, the NUT 4-vector is *absolutely* conserved even in the presence of gravitational radiation. Thus, while the gravitational field resembles the nonabelian Yang–Mills fields in its "electric" properties, it is analogous to the abelian Maxwell field in its "magnetic" properties. It is pointed out that gravitational fields with nonvanishing NUT 4-momenta may have a substantial role in quantum gravity even though they are not physically significant in classical general relativity.

PACS numbers: 04.50. + h

I. INTRODUCTION

In spite of the fact that they have not been detected in the laboratory, magnetic monopoles have drawn a great deal of attention ever since Dirac appealed to their possible existence to account for the observed quantization of the electric charge. In recent years, interest in monopoles has continued to grow in the broader setting of nonabelian gauge theories. There are several factors which motivate these investigations. For example, semiclassical analyses indicate that the classical solutions representing monopole configurations may play an important role in the more complete quantum theory. Similarly, there have been speculations on the existence and the role of monopoles in the early stages of the evolution of the universe. But perhaps the strongest reason behind this sustained interest is simply that monopole configurations arise very naturally in the classical description and that there is no theoretical reason to rule them out as being unphysical. One is therefore led to ask if classical theories dealing with other interactions admit analogous configurations and, if so, what properties they have. A natural candidate for such a theory is general relativity.

That general relativity does admit solutions with "magnetic" masses has been clear since the early sixties when the NUT solution (to Einstein's vacuum equation) was discovered[1] and analyzed.[2] This solution provided a good intuitive picture of the structure associated with the presence of magnetic masses. However, it did not immediately lead to general theorems encompassing arbitrary gravitational configurations with magnetic properties. Thus, for example, one did not even have a precise notion of what is to be meant by the term "magnetic mass" in the general context. To improve this situation, Ramaswamy and Sen[3] have recently introduced a framework to describe such configurations in the absence of gravitational radiation. This framework led to a precise definition of magnetic or "dual' energy-momentum as well as to certain theorems concerning the properties of the associated gravitational fields. However, this framework is also limited: Because it restricts itself to situations without gravitational radiation, it cannot answer certain physically interesting questions. For example, in the abelian Maxwell theory, both the electric and the magnetic charges are absolutely conserved while in the nonabelian Yang–Mills theories, both charges are radiated away across $\mathscr{I}^+$; they satisfy balance equations rather than absolute conservation laws. What is the situation in general relativity? One's first guess would be that general relativity mimics the nonabelian case and allows both the usual energy-momentum and the dual energy-momentum to radiate away across $\mathscr{I}^+$. As far as the usual (or "electric") energy-momentum is concerned, this guess is quite correct: The Bondi–Sachs[4] analysis provides the required balance equations. In the case of the dual (or magnetic) quantities, however, the situation is more complicated. For, in the stationary case, the presence of the dual energy-momentum *implies* the existence of wire singularities in certain potentials for Weyl curvature, so that the situation is analogous to the abelian case rather than the nonabelian one where one can and does have everywhere regular configurations carrying magnetic charges. One is therefore led to ask: Can the dual energy-momentum be radiated away? The available framework cannot answer this question since it deals only with those situations in which the Bondi news vanishes identically.

The purpose of this paper is to extend the framework to cases in which the Bondi news fails to vanish. This will enable us to answer the question raised above. We shall find that the magnetic quantities *cannot* be radiated away even when gravitational waves carry away energy-momentum across $\mathscr{I}^+$: NUT 4-momenta are forever! Thus, although

[a)] Research supported in part by a grant from the Research and Equipment Fund of Syracuse University.
[b)] Alfred P. Sloan Research Fellow. Supported in part by NSF Contract No. PHY80-08155.
[c)] Fellow of the Center for Theoretical Physics. Supported in part by NSF Contract No. PHY79-06940.

 0022-2488/82/112168-11$02.50 © 1982 American Institute of Physics

the gravitational field is analogous to the nonabelian Yang–Mills fields in its electric aspects, it mimics the abelian Maxwell field in its magnetic aspects. We shall see that this flip comes about because the gravitational field determines the space-time geometry.

Although there is a well-defined sense in which our framework reduces to that in Ref. 3 in the case when the Bondi news vanishes, our overall approach and viewpoint is substantially different from that of Ref. 3.[5]

The present viewpoint is motivated by the following considerations. Space-times with magnetic masses are not likely to be of interest as models of gravitational fields of macroscopic objects in classical relativity because one expects[3] their metrics to exhibit either wire singularities or causality violations. Hence, the primary interest in these objects comes from quantum gravity. Indeed, there is a scenario in which such configurations fit in nicely at the quantum level. This is provided by the so-called asymptotic quantization framework[6,7] in which one constructs the asymptotic quantum states on abstractly defined null infinity, $\mathscr{I}^{\pm}$, and then attempts to obtain the \$-operator to map the states on $\mathscr{I}^{-}$ to those on $\mathscr{I}^{+}$. For concreteness, let us adopt a path integral approach. Then, to obtain the transition amplitudes, one would fix two configurations of radiative modes of the gravitational field, one on the abstractly defined $\mathscr{I}^{-}$ and another on the abstractly defined $\mathscr{I}^{+}$, and permit all possible 4-geometries which can interpolate between the two. (Thus, $\mathscr{I}^{+}\cup\mathscr{I}^{-}$ is to be considered as an abstract light cone, not attached to any specific space-time; each permissible path exhibits it as the null infinity of specific space-time.) Now, the 4-geometries which interpolate between $\mathscr{I}^{-}$ and $\mathscr{I}^{+}$ do not have to be C^{∞} everywhere; it will suffice that they be, in some sense, paths with finite action.[8] Indeed, if one adopts the Euclidean (or, rather, Riemannian) path integral approach, one is led to allow 4-geometries which are even complex or singular in the Lorentzian domain[9]; the only requirements are that the intrinsic structure on $\mathscr{I}^{+}\cup\mathscr{I}^{-}$ induced by the path be regular and the Euclidean action of the path be finite. By contrast, the 4-geometries which carry magnetic masses exhibit only mild pathologies: the curvature tensors are regular everywhere and it is only that the (covariant) metric fails to be invertible along the so-called wire singularities. We shall therefore adopt the attitude that such paths should be included in the integral and analyze the way in which the asymptotic kinematic framework needs to be enlarged to incorporate them.

With this motivation, we proceed as follows. Fix a 3-manifold, $\mathscr{I}$, topologically $S^2\times R$, which is equipped with the usual[10] (smooth) universal structure of null infinity. The radiative modes of the gravitational field can be represented by certain equivalence classes, $\{D\}$, of derivative operators on this $\mathscr{I}$. (Ultimately, $\mathscr{I}$ will serve either as $\mathscr{I}^{+}$ or as $\mathscr{I}^{-}$, $\mathscr{I}^{+}\cup\mathscr{I}^{-}$ being the required abstract light cone.) The physically observable fields, the Bondi news, N_{ab}, and the radiative part of the Weyl tensor, ${}^{*}K^{ab}$, can be constructed from the curvature of $\{D\}$.[7,11] The idea now is to allow the connections $\{D\}$ to develop "wire singularities" in such a way that the fields N_{ab} and ${}^{*}K^{ab}$ remain smooth. When this is done, one can introduce a new[5] quantity associated with the translation subgroup of the Bondi–Metzner–Sachs[4] (BMS) group. This is the NUT 4-momentum, the magnetic analog of the Bondi–Sachs energy-momentum. The NUT 4-vector vanishes if $\{D\}$ is smooth; its origin is similar to that of the magnetic charge in Maxwell's theory. Thus, the primary difference between the present approach and that of Ref. 3 is the following. In Ref. 3, Ramaswamy and Sen restricted themselves to space-times which are C^{∞} but permitted null infinity, $\mathscr{I}$, to acquire various topologies. In the present analysis, by contrast, we restrict the topology of $\mathscr{I}$ to $S^2\times R$ but allow connections on $\mathscr{I}$ to develop wire singularities. This change in strategy seems to be essential in order to introduce the notion of the NUT 4-momentum in the presence of gravitational radiation. The new strategy also fits in more naturally in the asymptotic quantization scheme.

In Sec. II we introduce the basic framework. Section III is the heart of the paper: It contains the definition of the NUT 4-momentum and the result that it is absolutely conserved. Section IV discusses the issue of the enlargement of the space of classical vacua due to the presence of the NUT 4-momentum. Details of the effect of this enlargement on the asymptotic quantum states will be discussed elsewhere.

II. THE FRAMEWORK

Fix a C^{∞} 3-manifold $\mathscr{I}$, topologically $S^2\times R$, equipped with a collection of pairs (q_{ab}, n^a) of C^{∞} fields satisfying the following conditions: (i) The vector field n^a is complete and its manifold of orbits is diffeomorphic to S^2; (ii) $q_{ab}\,V^b = 0$ iff V^b is proportional to n^b; (iii) $\mathscr{L}_n\, q_{ab} = 0$; and (iv) pairs (q_{ab}, n^a) and $(\bar{q}_{ab}, \bar{n}^a)$ are both in the collection iff there exists a C^{∞} function ω on $\mathscr{I}$ such that $\bar{q}_{ab} = \omega^2 q_{ab}$ and $\bar{n}^a = \omega^{-1} n^a$. [Note that (ii) and (iii) imply that ω of (iv) must satisfy $\mathscr{L}_n\,\omega = 0$.] The integral curves of n^a will be referred to as the *generators* of $\mathscr{I}$ and the space of these generators will be denoted by $\mathscr{S}$. The 3-manifold $\mathscr{I}$ will serve as the "kinematic arena" for what follows; fields of physical interest will be simply "painted" on $\mathscr{I}$. Note that, although $\mathscr{I}$ is equipped with the usual[10] structure of null infinity—q_{ab} is the degenerate metric and n^a plays the role of the null normal—it is to be thought of as a 3-manifold in its own right without reference to any background space-time.

The dynamical variables representing gravitational degrees of freedom can be introduced in terms of suitable connections on $\mathscr{I}$. Let us first recall this procedure from Ref. 7. Fix a "conformal frame", i.e., a pair (q_{ab}, n^a) on $\mathscr{I}$. Denote by $\mathscr{C}$, the space of C^{∞} torsion-free derivative operators D on $\mathscr{I}$, satisfying

$$D_a\, q_{bc} = 0 \quad \text{and} \quad D_a\, n^b = 0. \tag{II.1}$$

These derivative operators lead to certain tensor fields. First, one can show that there exists[12] a unique, second-rank, symmetric tensor field ρ_{ab} on $\mathscr{I}$, satisfying

$$\rho_{ab}\, n^b = 0, \quad \rho_{ab}\, q^{ab} = R^0, \quad \text{and} \quad D_{[a}\,\rho_{b]c} = 0, \tag{II.2}$$

where D is any element of $\mathscr{C}$; q^{ab} is any tensor field satisfying $q^{ab}q_{am}\,q_{bn} = q_{mn}$ on $\mathscr{I}$; and R^0 is the pull-back to $\mathscr{I}$ of the scalar curvature of the metric induced on $\mathscr{S}$ by q_{ab}. If the induced metric happens to be the metric of a unit 2-sphere,

the conformal frame (q_{ab}, n^a) is referred to as the *Bondi (conformal) frame*. Equation (II.2) implies that, in a Bondi frame, $\rho_{ab} = q_{ab}$. The tensor field ρ_{ab} is "kinematic" since it is independent of the choice of the connection used in its definition. A connection-dependent field is provided by the curvature tensor. It can be shown[7] that the Riemann tensor[13] $R_{abc}{}^d$ of any connection D on $\mathscr{I}$ has the form[14]

$$R_{abc}{}^d = q_{c[a} S_{b]}{}^d + S_{c[a} \delta_{b]}{}^d , \tag{II.3}$$

where the D-dependent tensor field $S_a{}^b$ satisfies $S_a{}^m q_{mb} =: S_{ab} = S_{(ab)}$. Set

$$N_{ab} = S_{ab} - \rho_{ab} \tag{II.4}$$

and

$$*K^{ab} = 2\,\epsilon^{amn} D_m S_n{}^b . \tag{II.5}$$

Then N_{ab} is the Bondi news tensor and $*K^{ab}$ is the "radiative part" of the asymptotic Weyl curvature.[11] In the gravitational radiation theory, one can regard D and $S_a{}^b$ as "potentials" and N_{ab} and $*K^{ab}$ as "physical fields". (For details, see Refs. 12 and 7.)

What is the available gauge freedom? One can show that any two elements D and D' in $\mathscr{C}$ are related by

$$(D_a - D'_a) K_b = \Sigma_{ab} n^c K_c , \quad \forall\, K_b \tag{II.6}$$

by some C^∞ Σ_{ab}, independent of K_b, satisfying $\Sigma_{[ab]} = 0$ and $\Sigma_{ab} n^b = 0$. The connections D and D' *are gauge related* iff

$$\Sigma_{ab} = f q_{ab} \tag{II.7}$$

for some C^∞ function f on $\mathscr{I}$. It is easy to show that if D and D' are gauge related, Eqs. (II.6) and (II.7) imply

$$S'^b_a = S_a{}^b - 2D_a f n^b , \tag{II.8}$$

$$N'_{ab} = N_{ab} \quad \text{and} \quad *K'^{ab} = *K^{ab} ; \tag{II.9}$$

physical fields N_{ab} and $*K^{ab}$ are insensitive to gauge transformations. Thus, it is the equivalence classes $\{D\}$ of gauge related connections, rather than the individual elements of $\mathscr{C}$, that represent the true radiative degrees of freedom of the gravitational field. In fact, under conformal rescalings, $(q_{ab}, n^a) \to (\bar{q}_{ab}, \bar{n}^a) = (\omega^2 q_{ab}, \omega^{-1} n^a)$, potentials D and $S_a{}^b$ do not have unambiguous transformation properties. Equivalence classes, on the other hand, transform in a well-defined manner:

$$\{\bar{D}_a\} K_b = \{D_a\} K_b - 2\omega^{-1} K_{(a} D_{b)} \omega + \omega^{-1} (\omega^m K_m) q_{ab} \tag{II.10}$$

and

$$\{\bar{S}_a{}^b\} = \omega^{-2} \{S_a{}^b\} - 2\omega^{-3} D_a \omega^b + 4\omega^{-4} \omega^b D_a \omega - \omega^{-4} \delta^b{}_a \omega^m D_m \omega , \tag{II.11}$$

where ω^b is any vector field on $\mathscr{I}$ such that $q_{ab}\omega^b = D_a \omega$ and $\{S_a{}^b\}$ is the equivalence class of tensors obtained from $\{D\}$. (It is the ambiguity in ω^b that prevents us from specifying the transformation properties of individual D and $S_a{}^b$.)[7] The fields, of course, transform without ambiguities:

$$\bar{N}_{ab} = N_{ab} \quad \text{and} \quad *\bar{K}^{ab} = \omega^{-5} *K^{ab} . \tag{II.12}$$

We now wish to enlarge this framework by permitting the potentials $(D, S_a{}^b)$ to develop "wire singularities" which leave the C^∞ character of the fields N_{ab} and $*K^{ab}$ unaffected. Since $N_{ab} = S_{ab} - \rho_{ab}$ and since ρ_{ab} is C^∞ (being kinematical), $S_{ab} = S_a{}^m q_{mb}$ must be C^∞. Thus the "singular part" of $S_a{}^b$ must be of the type $A_a n^b$. Furthermore, since $*K^{ab} = 2\epsilon^{amn} D_m S_n{}^b$ is to be C^∞, $D_{[a} A_{b]}$ must also be C^∞. Can we satisfy these constraints and still be left with $(D, S_a{}^b)$ with "wire singularities"? Our experience with magnetic monopoles in Maxwell's theory suggests that this may be possible. We therefore proceed as follows. Fix a C^∞ connection D^0 in $\mathscr{C}$ whose curvature is trivial, i.e., $N^0_{ab} = 0$ and $*K^0{}_{ab} = 0$. Then, any connection D on $\mathscr{I}$ satisfying Eq. (II.1) is completely determined by a tensor field Σ_{ab}: $D_a K_b = D^0_a K_b + \Sigma_{ab} n^c K_c$. Hence, the equivalence class $\{D\}$ to which D belongs is determined by $\gamma_{ab} := \Sigma_{ab} - \frac{1}{2} q_{ab} \Sigma_{mn} q^{mn}$, the trace-free part of Σ_{ab}. If D is C^∞, so is γ_{ab}. We can now enlarge the space $\mathscr{C}$ of connections. Denote by $\overline{\mathscr{C}}$ the space of torsion free connections D on $\mathscr{I}$ satisfying Eq. (II.1) and the following conditions: (i) the physical fields N_{ab} and $*K^{ab}$ constructed from D are C^∞ everywhere on $\mathscr{I}$; (ii) Σ_{ab} obtained from D is C^∞ everywhere on $\mathscr{I}$ except possibly on a finite number of generators; and (iii) $N_{ab}\, \gamma_{cd}\, g^{ac}$ is bounded on $\mathscr{I}$. [It is easy to verify that $\mathscr{L}_n \gamma_{ab} = -\frac{1}{2} N_{ab}$. Hence, by condition (i), the irregularities of γ_{ab}, or of $\{D\}$, must lie along whole generators of $\mathscr{I}$. This justifies the name "wire singularities."] We shall say that D and D' in $\overline{\mathscr{C}}$ are gauge related if they satisfy $(D'_a - D_a) K_b = f q_{ab} n^c K_c$ for some function f which is C^∞ except possibly on a finite number of generators.[15] Denote, as before, the equivalence classes of gauge related connections by $\{D\}$. We shall say that $\{D\}$ is *regular* if it contains an element D of $\overline{\mathscr{C}}$ which is everywhere C^∞. Otherwise $\{D\}$ will be said to be *nutty*. Note that $\{D\}$ is regular if and only if the corresponding γ_{ab} is a C^∞ tensor field on $\mathscr{I}$. It will turn out that $\{D\}$ is nutty if and only if the NUT 4-momentum of $\{D\}$ fails to vanish.

Remarks: (i) In the above construction, we chose a connection D^0 with trivial curvature (i.e., a "classical vacuum"). Note, however, that the notion of regular and nutty configurations does *not* depend upon the specific choice of D. (ii) Note that the notion of a C^∞ connection is not gauge invariant in $\overline{\mathscr{C}}$: one can make "singular gauge transformations". The notion of regularity of equivalence classes, on the other hand, *is* gauge invariant. (iii) Given an element D of $\overline{\mathscr{C}}$, one can compute the tensor field $S_a{}^b$ and, given a C^∞ connection D^0 with trivial curvature, the tensor fields Σ_{ab} and γ_{ab}. The singularity structure of γ_{ab} is gauge invariant since γ_{ab} itself remains unchanged under the gauge transformations. The situation is different for $S_a{}^b$, and Σ_{ab}: By making singular gauge transformations, one can move around the wire singularities in these fields. Indeed, in the case of $S_a{}^b$, the situation is completely analogous to that in electrodynamics. The "singular" part of $S_a{}^b$ is contained in $A_a := S_a{}^b l_b$, where l_b is any covector field satisfying $l_a n^a = 1$ on $\mathscr{I}$ and, under a gauge transformation, we have $A_a \to A_a - 2D_a f$. Hence, as in the case of a magnetic monopole, although one cannot simply gauge away the wire singularity in $S_a{}^b$, one *can* move it around.

To conclude this section, we give some examples of nut-

ty configurations $\{D\}$ on $\mathscr{I}$. Fix a Bondi conformal frame (q_{ab}, n^b) and a derivative operator D^0 with trivial curvature. Then, there exists[7] a 4-parameter family of covector fields l_a on $\mathscr{I}$, satisfying $n^a l_a = 1$ and $D_{[a} l_{b]} = 0$ such that $D_a{}^0 l_b = 0$. Choose one of these and set $l_a = D_a u$ and label the 2-sphere cross sections, $u =$ constant, of $\mathscr{I}$ by coordinates θ and ϕ satisfying $\mathscr{L}_n \theta = 0$ and $\mathscr{L}_n \phi = 0$. Thus, u, θ, ϕ provide a chart for $\mathscr{I}$ (modulo the usual problems at the poles $\theta = 0$ and $\theta = \pi$). Set $\sqrt{2}\, m^a = (\partial/\partial\theta)^a + (i/\sin\theta)(\partial/\partial\phi)^a$. We are now ready to define the required nutty connections. Define Σ_{ab} by

$$\Sigma_{ab} = g m_a m_b + \bar{g} \bar{m}_a \bar{m}_b \,, \tag{II.13}$$

with

$$g = il(1 - \cos\theta)(1 + \cos\theta)^{-1} \,,$$

and consider the derivative operator D defined by

$$(D_a - D^0{}_a) K_b = \Sigma_{ab}\, n^c K_c \tag{II.14}$$

for all covector fields K_c. (Note that Σ_{ab} satisfies the required properties: $\Sigma_{ab} = \Sigma_{(ab)}$ and $\Sigma_{ab} n^b = 0$. In addition, it is tracefree: $\Sigma_{ab}\, q^{ab} = 0$. Hence, $\Sigma_{ab} \equiv \gamma_{ab}$.) Since $\Sigma_{ab} \equiv \gamma_{ab}$ has a wire singularity along $\theta = \pi$, D fails to belong to $\mathscr{C}$. To see if it belongs to $\overline{\mathscr{C}}$, let us compute its curvature tensor. Since $\mathscr{L}_n \gamma_{ab} = 0$, the news tensor N_{ab} of D vanishes identically. Hence, $S_{ab} = \rho_{ab} = q_{ab}$. This implies that $S_a{}^b$ is of the form

$$S_a{}^b = \delta_a{}^b + A_a n^b \tag{II.15a}$$

for some covector field A_a. Equation (II.3) and the definition of D now imply[16]

$$A_a = -2l_a - 4l(1 - \cos\theta)\, D_a \phi \,. \tag{II.15b}$$

Finally, Eq. (II.5) yields

$$^*K^{ab} = 8l n^a n^b \,. \tag{II.16}$$

Thus, although D has a wire singularity at $\theta = \pi$, it is C^∞ everywhere else, and $^*K^{ab}$ and N_{ab} are C^∞ everywhere on $\mathscr{I}$. Finally, since $N_{ab} = 0$, $\gamma_{ab} N_{cd}\, q^{bc}$ is trivially bounded on $\mathscr{I}$. Thus, D belongs to $\mathscr{C}$ and $\{D\}$ is nutty. Examples of nutty configurations $\{D'\}$ with nonvanishing news are easy to construct. Set

$$\gamma'_{ab} = \gamma_{ab} + \lambda_{ab} \,, \tag{II.17}$$

where λ_{ab} is a C^∞, symmetric, tracefree tensor field on $\mathscr{I}$, transverse to n^a the closure of whose support excludes the generator $\theta = \pi$ of $\mathscr{I}$. Then, if $\mathscr{L}_n \lambda_{ab}$ fails to vanish, $\{D'\}$ would be a nutty configuration with nonzero news.

Remark: Although the connection D above was introduced intrinsically on $\mathscr{I}$, without any reference to an interior space-time, it is in fact the connection induced[12,7] on the null boundary I in the Penrose completion of the NUT space-time (without the Misner[2] identification.) The conformally rescaled NUT metric is given by[3]

$$\begin{aligned} g_{ab} dx^a dx^b &= -U[x^2\, du^2 - 2U^{-1}\, du\, dx \\ &\quad + 2A\,(x^2\, du - U^{-1}\, dx)\, d\phi] \\ &\quad + (1 + l^2x^2)\, d\theta^2 + [(1 + l^2x^2)\sin^2\theta - UA^2x^2]\, d\phi^2 \,, \end{aligned} \tag{II.18}$$

where

$$U = 1 - \frac{2(mx + l^2x^2)}{1 + l^2x^2} \,, \quad A = 2l(1 - \cos\theta)$$

and m and l are constants. The null infinity, I, is the surface $x = 0$. The coordinate u runs from $-\infty$ to ∞ (no periodicity!) and θ and ϕ take on their usual 2-sphere values. On I, we have

$$\begin{aligned} g_{ab}\, dx^a dx^b|_I &= +2\, du\, dx - 2A\, dx\, d\phi + d\theta^2 + \sin^2\theta\, d\phi^2 \,. \end{aligned}$$

Thus, the wire singularity at $\theta = \pi$ persists even at I. Set $l_a = \nabla_a u$ and, as before, $\sqrt{2}\, m^a = (\partial/\partial\theta)^a + (i/\sin\theta)(\partial/\partial\phi)^a$. Then, a direct computation yields

$$\sigma|_I \equiv m^a m^b \nabla_a l_b|_I = -il(1 - \cos\theta)(1 + \cos\theta)^{-1} \,. \tag{II.19}$$

Hence, it follows[7,11] that the connection D induced on I by ∇ is precisely the same as that of Eq. (II.13) (with D^0, the connection defined by $D^0_a l_b = 0$). The pull-back to $\mathscr{I}$ of the (4-dimensional) tensor field $(R^a{}_b - \frac{1}{6} R\delta^a{}_b)$, constructed from g_{ab}, is precisely the $S_a{}^b$ of Eq. (II.15) and the restriction to $\mathscr{I}$ of $^*K^{ambn} \nabla_m x \nabla_n x$ is precisely the $^*K^{ab}$ of Eq. (II.16). (Here R_{ab} is the Ricci tensor of g_{ab} and $^*K^{abcd}$ is related to the dual $^*C^{abcd}$ of the Weyl tensor via $^*K^{abcd} = x^{-1}\, {}^*C^{abcd}$.)

III. THE NUT 4-MOMENTUM

Fix a radiative mode $\{D\}$ on $\mathscr{I}$. The NUT 4-momentum of $\{D\}$ evaluated at any cross section C of $\mathscr{I}$ is to be a 4-vector which lies in the space dual to that of BMS translations. Thus, given a cross section C and a BMS translation ξ^a on $\mathscr{I}$, one is to obtain a number, $^*Q_{(\xi)}[C]$, the component of the NUT 4-momentum along the translation ξ^a. From general considerations, one expects $^*Q_{(\xi)}[C]$, to have the following properties: (i) $^*Q_{(\xi)}[C]$, should depend linearly on the BMS translation ξ^a; (ii) it should depend only on $\{D\}$, C, and ξ^a and not on any auxiliary structure such as a conformal frame or a choice of gauge; (iii) it should, in some sense, be the magnetic analog of $Q_{(\xi)}[C]$, the component of the Bondi 4-momentum at C, along the translation ξ^a; and (iv) it should vanish if $\{D\}$ is regular, i.e., it should be a measure of the "strength of the wire singularities" in $\{D\}$.

In the first part of this section, we shall introduce a definition of $^*Q_{(\xi)}[C]$ satisfying these four properties and, in the second part, show that $^*Q_{(\xi)}[C]$ cannot be radiated away.

A. Definition of $^*Q_{(\xi)}[C]$

Our definition may be motivated as follows. For simplicity, let us fix a Bondi conformal frame $(\bar{q}_{ab}, \bar{n}^a)$ on $\mathscr{I}$. (Thus, $\bar{q}_{ab}$ is a 2-sphere metric.) Then $\bar{n}^a$ itself is a BMS (time $-$) translation. Let us therefore begin by setting $\xi^a = \bar{n}^a$ and look for $Q_{(\bar{n})}[C]$. Since $\bar{A}_a = \bar{S}_a{}^b \bar{l}_b$ resembles the vector potential in Maxwell's theory, it is natural to set, in view of the definition of the magnetic monopoles in Maxwell's theory,

$$^*Q_{(\bar{n})}[C] = \frac{1}{8\pi} \overline{\int_C} \bar{D}_{[a} \bar{A}_{b]}\, ds^{ab} \,, \tag{III.1}$$

where to perform the integral $\overline{\int}_C$, one first surrounds the

singularities (of $\bar{D}$) on C by closed loops, integrates $\bar{D}_{[a}\bar{A}_{b]}$ on the part of C outside these loops where A_b is C^∞, and then takes the limit as the loops shrink to zero. (Recall: $\bar{l}_a$ is any covector field satisfying $\bar{l}_a \bar{n}^a = 1$.) If $\bar{D}$ (in $\{\bar{D}\}$, whose curvature tensor determines $S_a{}^b$) is a C^∞ connection, ${}^*Q_{(\bar{n})}[C]$ vanishes, as required by condition (iv) above. We can therefore look simply for a suitable extension of Eq. (III.1) to all BMS translations ξ^a. To guess the correct extension, we first re-express ${}^*Q_{(\bar{n})}[C]$ using a new conformal frame $(q_{ab}, n^a) := (\omega^{-2}\bar{q}_{ab}, \omega\bar{n}^a)$ (which is not necessarily a Bondi frame). Using Eqs. (II.10) and (II.11), one obtains

$$
\begin{aligned}
{}^*Q_{(\bar{n})}[C] &= {}^*Q_{(\omega^{-1}n)}[C] \\
&= \frac{1}{8\pi}\overline{\int_C} ds^{ab}\left[\omega^{-1} D_{[a}A_{b]} + (D_m\omega^{-1}) S_b{}^m l_a\right. \\
&\quad \left. + 2q^{mn}(D_bD_n\omega^{-1}) D_a l_m\right], \qquad \text{(III.2)}
\end{aligned}
$$

where $A_b = S_b{}^m l_m$ with $l_m n^m = 1$, and q^{mn} is any inverse of q_{mn} (i.e., any tensor field satisfying $q^{mn}q_{ma}q_{nb} = q_{ab}$). Using Eq. (II.5) and setting $\omega^{-1} = \alpha$, we can reexpress Eq. (III.2) as follows:

$$
\begin{aligned}
{}^*Q_{(\alpha n)}[C] = \frac{1}{8\pi}\overline{\int_C} ds^{ab}\Big(&\frac{\alpha}{4}\,{}^*K^{pm} l_m \epsilon_{pab} \\
&+ (\alpha D_a l_m + l_a D_m \alpha) S_b{}^m \\
&+ 2q^{mn}(D_bD_n\alpha) D_a l_m\Big). \qquad \text{(III.3)}
\end{aligned}
$$

This is our definition of the NUT 4-momentum: In any conformal frame (q_{ab}, n^a) on $\mathscr{I}$, the right side of Eq. (III.3) gives the component of the NUT 4-momentum along the BMS translation $\xi^a = \alpha n^a$, evaluated at the cross section C of $\mathscr{I}$.

The above discussion serves only to motivate our definition of ${}^*Q_{(\alpha n)}[C]$. We shall now show explicitly that the expression is well defined and has the required properties.

Lemma 1.1: The integral in Eq. (III.3) exists and is finite for all BMS translations αn^a and all connections D in $\bar{\mathscr{C}}$.

Proof: Fields α, ϵ_{abc}, q^{ab}, and l_a are smooth, being kinematical, while ${}^*K^{ab}$ and $S_a{}^m q_{mb}$ are smooth by definition of $\bar{\mathscr{C}}$. Hence, in the integrand of Eq. (III.3), only the following terms fail to be manifestly C^∞:

$$
\begin{aligned}
F_{ab} &:= \alpha(D_a l_m) S_b{}^m + 2q^{mn}(D_bD_n\alpha)(D_a l_m) \\
&= (D_a l_m) q^{mn}(\alpha S_{bn} + 2D_bD_n\,\alpha) \\
&= (D_a l_m) q^{mn}(\alpha N_{bn} + \rho_{bn} + 2D_bD_n\,\alpha) \\
&= (D_a l_m) q^{mn}(\alpha N_{bn} + \mu q_{bn})
\end{aligned}
$$

for some scalar field μ where, in the last step, we have used the fact that α is a BMS translation.[17] Now, consider any C^∞ connection D^0 with trivial curvature. Then, $(D_a - D^0_a) l_m = \Sigma_{am} l_p n^p = \Sigma_{am}$. Hence,

$$
\begin{aligned}
F_{ab} &= (\Sigma_{am} + D^0_a l_m) q^{mn}(\alpha N_{bn} + \mu q_{bn}) \\
&= (\gamma_{am} + \tfrac{1}{2}\Sigma q_{am} + D^0_a l_m) q^{mn}(\alpha N_{bn} + \mu q_{bn}) \\
&= \alpha\gamma_{am} q^{mn} N_{bn} + \tfrac{1}{2}\Sigma N_{ab} + \mu\Sigma_{ab} + \alpha(D^0_a l_m) \\
&\quad \times q^{mn}(\alpha N_{bn} + \mu q_{bn}),
\end{aligned}
$$

where $\Sigma = \Sigma_{ab} q^{ab}$. (In the last step, we have used the fact that $N_{ab} q^{ab} = 0$.) By definition of $\bar{\mathscr{C}}$, $\gamma_{am} q^{mn} N_{bn}$ is bounded on $\mathscr{I}$. Hence, we have ${}^*Q_{(\alpha n)}[C] = \overline{\int} ds^{ab} B_{ab} + \overline{\int} ds^{ab}(\mu\Sigma_{ab} + \frac{1}{2}\Sigma N_{ab})$, where B_{ab} is a bounded 2-form on C. The first of these two integrals is clearly finite; $\overline{\int} ds^a B_{ab} = \int ds^{ab} B_{ab}$. The second integral vanishes because Σ_{ab} and N_{ab} are symmetric. Hence the result.

Remarks: (i) Why did we use the integral $\overline{\int}$ in Eqs. (III.1)–(III.3) rather than just the usual integral? The reason is that even when $\{D\}$ belongs to $\bar{\mathscr{C}}$ (so that ${}^*K^{ab}$ and N_{ab} are smooth and $\gamma_{ab} N_{cd} q^{ac}$ remains bounded), as we saw in the proof of Lemma 1.1, the integrand contains a singular term, Σ_{ab}. Hence, even though Σ_{ab} is symmetric wherever it is nonsingular, $\int_C(\mu\Sigma_{ab} + \frac{1}{2}\Sigma N_{ab})\, ds^{ab}$ is not well defined *a priori*. However, since the singularities of Σ_{ab} on C lie at a finite number of isolated points, we can first surround these points by small loops, l_i containing 2-surface areas S_i, compute

$$
\int_{(C - \cup_i S_i)} \Sigma_{ab}\, ds^{ab}, \qquad \text{(III.4)}
$$

and then take the limit as S_i shrinks to zero. Since at each stage in the limit the integral in Eq. (III.4) vanishes, $\overline{\int_C} \Sigma_{ab}\, dS^{ab}$ is well defined and its value is zero. Thus, the use of $\overline{\int}$ simply serves to regularize the integral in the obvious way. (ii) Note that the integral in Eq. (III.3) fails to be well defined if αn^a is a BMS supertranslation rather than a BMS translation: In the proof of Lemma 1.1, we used the condition $2D_aD_b\,\alpha + \rho_{ab} = \mu q_{ab}$, which is satisfied precisely when αn^a is a BMS translation. Hence, it appears that NUT supermomenta may not be well defined within the type of framework that is being pursued here. This situation is quite different from that in the case of the Bondi–Sachs quantities, where the obvious extension of the expression of the 4-momentum does lead to a meaningful definition of supermomentum.[12,18] The difference arises because in the case of the Bondi–Sachs quantities one deals only with smooth fields on $\mathscr{I}$ so that the analog of Lemma 1.1 is automatically satisfied for all supertranslations. We shall return to this issue in Sec. IV.

Lemma 1.2: The integral in Eq. (III.3) is gauge invariant, i.e., remains unchanged if D, $S_a{}^b$, and ${}^*K^{ab}$ is replaced by D', $S'_a{}^b$, and ${}^*K'^{ab}$ of Eqs. (II.6)–(II.9).

Proof: Since $(D'_a - D_a) K_b = f q_{ab}\, n^c K_c$ for all covector fields K_c and $S'_a{}^b = S_a{}^b - 2(D_a f) n^b$ and ${}^*K'^{ab} = {}^*K^{ab}$, the difference, Δ, between the integrals involving primed and unprimed quantities is given by

$$
\begin{aligned}
8\pi\Delta = &\overline{\int_C} ds^{ab}\left[(\alpha D'_a l_m + l_a D'_m\,\alpha) S'_b{}^m\right. \\
&- (\alpha D_a l_m + l_a D_m\,\alpha) S_b{}^m \\
&\left. + 2q^{mn}(D'_bD'_n\,\alpha) D'_a l_m - 2q^{mn}(D_bD_n\,\alpha) D_a l_m\right] \\
= &\overline{\int_C} ds^{ab}\left[\alpha f q_{am}(S_b{}^m - 2n^m D_b f)\right. \\
&- 2\alpha(D_a l_m) n^m D_b f - 2l_a(D_m\,\alpha) n^m D_b f \\
&\left. + 2q^{mn}(D_bD_n\,\alpha) f q_{am}\right].
\end{aligned}
$$

Using equations $n^m q_{am} = 0$, $\mathscr{L}_n\,\alpha = 0$, and $n^m D_a l_m = D_a(n^m l_m) = 0$, one obtains

$$
\Delta = \frac{1}{8\pi}\overline{\int_C} ds^{ab}(\alpha f S_{ab} + 2f D_aD_b\,\alpha).
$$

Since $S_{[ab]}=0$ and $D_{[a}D_{b]}\alpha=0$ we have [19] $\Delta=0$.

Lemma 1.3: The integral in Eq. (III.3) is conformally invariant, i.e., remains unchanged under transformations (II.10)–(II.12).

Proof: Consider the conformal frame[20] $(\bar{q}_{ab}, \bar{n}^a) \equiv(\omega^2 q_{ab}, \omega^{-1}n^a)$. Let $\xi^a=\bar{\alpha}\,\bar{n}^a$ be a BMS translation vector field. Then, using Eq. (III.3), we have

$$
\begin{aligned}
{}^*Q_{(\bar{\alpha}\bar{n})}[C] = \frac{1}{8\pi}\overline{\int_C} ds^{ab} & \times\left[\frac{\bar{\alpha}}{4}\,{}^*\bar{K}^{\,pm}\bar{l}_m\,\bar{\epsilon}_{pab} + (\bar{\alpha}\bar{D}_a\bar{l}_m + \bar{l}_a\bar{D}_m\,\bar{\alpha})\,S_b{}^m\right. \\
& \left. + 2\bar{q}^{mn}(\bar{D}_b\bar{D}_n\,\bar{\alpha})\,\bar{D}_a\bar{l}_m\right]. \qquad \text{(III.5)}
\end{aligned}
$$

To express the integral in terms of quantities defined in the unbarred conformal frame (q_{ab}, n^a), we set $l_a=\omega^{-1}\bar{l}_a$ (so that $n\cdot l=1$), $\alpha=\omega^{-1}\bar{\alpha}$ (so that $\xi^a=\alpha n^a$), and use Eqs. (II.10)–(II.12) to obtain

$$
\frac{\bar{\alpha}}{4}\,{}^*K^{\,pm}\bar{l}_m\,\bar{\epsilon}_{pab}=\frac{\alpha}{4}\,{}^*K^{\,pm}l_m\,\epsilon_{pab}\,, \qquad \text{(III.6)}
$$

$$
\begin{aligned}
& q^{mn}(\bar{D}_a\bar{l}_m)(\bar{\alpha}\bar{S}_{bn}+2\bar{D}_b\,\bar{D}_n\,\bar{\alpha}) \\
& \quad = q^{mn}(D_a l_m)(\alpha S_{bn}+2D_bD_n\,\alpha) \\
& \qquad + 2\omega^{-1}(D_a l_b)\,\omega^p D_p\,\alpha \\
& \qquad - 2\omega^{-1}l_a\,\omega^n D_bD_n\,\alpha + 2\omega^{-2}(\omega^p D_p\,\omega) \\
& \qquad \times(D_a l_b) - \omega^{-1}\alpha l_a(D_n\,\omega)\,S_b{}^n \\
& \qquad - 2\omega^{-2}l_a(D_b\,\omega)\,\omega^p D_p\,\alpha - \omega^{-3}(\omega^p D_p\alpha) \\
& \qquad \times l_a\,D_b\,\omega + \Delta_{ab}\,, \qquad \text{(III.7)}
\end{aligned}
$$

where Δ_{ab} is a symmetric tensor field, and

$$
\begin{aligned}
& \bar{l}_a(\bar{D}_m\,\bar{\alpha})\,\bar{S}_b{}^m \\
& \quad = l_a(D_m\,\alpha)\,S_b{}^m + (l_aD_m\,\alpha)(-2\omega^{-1}D_b\,\omega^m \\
& \qquad + 4\omega^{-2}\,\omega^m D_b\,\omega) \\
& \qquad - l_a(D_b\,\alpha)(\omega^{-2}\,\omega^p D_p\,\omega) + \alpha l_a(D_m\,\omega)(4\omega^{-3} \\
& \qquad \times\omega^m D_b\,\omega - 2\omega^{-2}D_b\,\omega^m) \\
& \qquad + \omega^{-1}\,\alpha(D_m\,\omega)l_a\,S_b{}^m - \alpha l_a(D_b\,\omega)(\omega^{-3}\,\omega^p\,D_p\,\omega)\,. \qquad \text{(III.8)}
\end{aligned}
$$

Here, as before, ω^a is any vector field satisfying $q_{ab}\,\omega^a = D_b\,\omega$. Since the sum of the left sides of Eqs. (III.6)–(III.8) constitutes the integrand in (III.5), we obtain

$$
\begin{aligned}
{}^*Q_{(\bar{\alpha}\bar{n})}[C] = \frac{1}{8\pi}\overline{\int_C} ds^{ab} & \left[\frac{\alpha}{4}\,{}^*K^{\,bm}\,l_m\,\epsilon_{pab}\right. \\
& + (\alpha D_a l_m + l_a D_m\alpha)S_b{}^m \\
& + 2q^{mn}(D_bD_n\,\alpha)\,D_a l_m \\
& + 2D_a(l_b\omega^{-1}\,\omega^p D_p\,\alpha) \\
& \left. + D_a(l_b\,\omega^{-2}\,\alpha\omega^p\,D_p\,\omega) + \Delta_{ab}\right].
\end{aligned}
$$

Since $\Delta_{ab}=\Delta_{(ab)}$, the last term does not contribute to the integral. Since $l_p(2\omega^{-1}\,\omega^b D_b\,\alpha+\omega^{-2}\,\alpha\omega^b D_b\,\omega)$ is a smooth vector field, the integral over C of its curl vanishes. Hence,

$$
{}^*Q_{(\bar{\alpha}\bar{n})}[C] = {}^*Q_{(\alpha n)}[C]\,.
$$

Lemma 1.4: The integral in Eq. (III.3) is independent of the choice of the covector field l_a (satisfying $l_a n^a=1$) made in its evaluation.

Proof: Let l_a and l'_a be two covector fields on $\mathscr{I}$ satisfying $l\cdot n=l'\cdot n=1$. Set $l_a-l'_a=v_a$. Then $v\cdot n=0$. Since the integral in Eq. (III.3) is linear in l_a, the difference Δ in its value obtained by using l_a and l'_a is given by

$$
\begin{aligned}
\Delta &= \frac{1}{8\pi}\overline{\int_C} ds^{ab}\left[\frac{\alpha}{4}\,{}^*K^{\,bm}v_m\,\epsilon_{pab}\right. \\
&\quad \left. + (\alpha D_a v_m + v_a D_m\,\alpha)\,S_b{}^m + 2q^{mn}(D_bD_n\,\alpha)(D_a v_m)\right] \\
&= \frac{1}{8\pi}\overline{\int_C} ds^{ab}\,[D_a(\alpha S_b{}^m v_m) - (D_a\,\alpha)\,S_b{}^m v_m \\
&\quad + (D_m\,\alpha)\,S_b{}^m v_a \\
&\quad + 2q^{mn}(D_bD_n\,\alpha)\,D_a v_m\,] \\
&= \frac{1}{8\pi}\overline{\int_C} ds^{ab}\,D_a(\alpha S_{bn}v_m\,q^{mn}) + 2D_b(q^{mn}(D_n\,\alpha)(D_a v_m)) \\
&= \frac{1}{8\pi}\overline{\int_C} ds^{ab}\,D_a\,[\alpha S_{bn}v_m\,q^{mn} - 2q^{mn}(D_n\,\alpha)(D_b v_m)] \\
&= 0,
\end{aligned}
$$

since the integrand is the curl of a C^∞ vector field.

Remark: One can use the freedom available in the choice of the conformal frame (q_{ab}, n^a) and the covector field l_a to simplify the expression of the NUT 4-momentum. Let (q_{ab}, n^a) be a Bondi frame. Introduce a function u on $\mathscr{I}$ such that $n^a D_a u=1$ and set $l_a=D_a\,u$. Then, $D_{[a}l_{b]}=0$. Using the fact that $D_aD_b\,\alpha$ is proportional to q_{ab} in a Bondi frame, Eq. (III.3) now simplifies to

$$
\begin{aligned}
& {}^*Q_{(\alpha n)}[C] \\
& \quad = \frac{1}{8\pi}\overline{\int_C} ds^{ab} \\
& \qquad \times\left[\frac{\alpha}{4}\,{}^*K^{\,pm}l_m\,\epsilon_{pab} + (\alpha D_a l_m + l_a D_m\,\alpha)\,S_b{}^m\right]. \qquad \text{(III.9)}
\end{aligned}
$$

Finally, if the function u is so chosen that $u=$ constant on C, a further simplification occurs:

$$
{}^*Q_{(\alpha n)}[C] = \frac{1}{8\pi}\overline{\int_C} ds^{ab}\,\frac{\alpha}{4}\,({}^*K^{\,bm}\,l_m\,\epsilon_{pab} + (D_a l_m)\,S_b{}^m) \qquad \text{(III.10a)}
$$

$$
= \frac{1}{8\pi}\overline{\int_C} ds^{ab}\,\alpha D_{[a}(S_{b]}{}^m\,l_m)\,. \qquad \text{(III.10b)}
$$

Lemma 1.5: The integral in Eq. (III.3) vanishes if $\{D\}$ is regular.

Proof: Denote by D a C^∞ connection in $\{D\}$. Since the integral in question is gauge invariant (Lemma 1.2), we can use this D and its $S_a{}^b$ in the integrand. We have

$$
\begin{aligned}
& 8\pi\,{}^*Q_{(\alpha n)}[C] \\
& \quad = \overline{\int_C} ds^{ab}\,[\alpha D_a(S_b{}^m\,l_m) + (D_m\,\alpha)\,S_b{}^m\,l_a \\
& \qquad + 2q^{mn}(D_b\,D_n\,\alpha)\,D_a l_m\,]
\end{aligned}
$$

$$= \overline{\int_C} ds^{ab} \left[D_a(\alpha S_b{}^m l_m) - (D_a \alpha) S_b{}^m l_m \right.$$
$$+ (D_m \alpha) S_b{}^m l_a$$
$$\left. + 2q^{mn}(D_b D_n \alpha)(D_a l_m) \right]$$
$$= \overline{\int_C} ds^{ab} \left[D_a(\alpha S_b{}^m l_m) - 2q^{mn}(D_n \alpha)(D_b l_m) \right]$$
$$= 0,$$

since the integrand is the curl of a C^∞ 1-form.

Collecting results from Lemmas (1.1)–(1.5) we obtain

Theorem 1: For each choice of cross section C of $\mathscr{I}$, the NUT 4-momentum of any $\{D\}$ (with D in $\overline{\mathscr{C}}$) is a well-defined, gauge and conformally invariant, linear mapping from the space of BMS translations to the reals. If $\{D\}$ is regular, its NUT 4-momentum vanishes identically.

Remarks: (i) Note that the information concerning the NUT 4-vector is contained entirely in the radiative mode $\{D\}$: We used only those fields on $\mathscr{I}$ which can be constructed directly from $\{D\}$. This is in striking contrast with the definition of the Bondi–Sachs 4-vector, where one uses also the longitudinal modes (via the field K^{ab} or, in the Newman–Penrose notation, Re $\psi_2{}^0$) which cannot be recovered from $\{D\}$ alone. This difference is somewhat puzzling. Note however that this feature is not a peculiarity of general relativity: The situation is completely analogous in both abelian and nonabelian gauge theories, where the magnetic charges can be recovered from the radiative data on $\mathscr{I}$ (the pull-back of the vector potential) although the expression of the electric charge involves longitudinal modes (Re ϕ_1^0, in the Newman–Penrose notation). (ii) Consider the case when $N_{ab} = 0$. Introduce a smooth function u on $\mathscr{I}$ such that $n^a D_a u = 1$ and $u =$ constant on C. Then, by choosing $l_a = D_a u$, one obtains

$$*Q_{(\alpha n)}[C] = \frac{1}{8\pi} \overline{\int_C} ds^{ab}$$
$$\times \left[\frac{\alpha}{4} *K^{pm} l_m \epsilon_{pab} \right.$$
$$\left. + (D_a l_m) q^{mn}(2D_b D_n \alpha + \alpha \rho_{bn}) \right]$$
$$= \frac{1}{32\pi} \overline{\int_C} ds^{ab} \alpha *K^{pm} l_m \epsilon_{pab} . \qquad \text{(III.11)}$$

Next, since $N_{ab} = 0$, it follows[12,3] that there exists a C^∞ function μ on $\mathscr{I}$ such that $\mathscr{L}_n \mu = 0$ and $*K^{pm} = \mu n^p n^m$. Hence, we have

$$*Q_{(\alpha n)}[C] = \frac{1}{32\pi} \overline{\int_C} ds^{ab} (\alpha \mu \epsilon_{ab}) . \qquad \text{(III.12)}$$

Finally, since the integrand is a 2-form orthogonal to and Lie-derived by n^a, we can replace C by the 2-sphere $\mathscr{S}$ of generators of $\mathscr{I}$:

$$*Q_{(\alpha n)} = \frac{1}{32\pi} \int_{\mathscr{S}} \hat{\alpha}\hat{\mu}\hat{\epsilon}_{ab} \, ds^{ab} . \qquad \text{(III.13)}$$

This is precisely the expression introduced by Ramaswamy and Sen [Eq. (25b) in Ref. 3]. In presence of news, on the other hand, one must return to Eq. (III.3) or, with a special choice of a conformal factor and l_a, to Eq. (III.10): Not only can one not project down to $\mathscr{S}$, but even Eq. (III.11) fails to coincide with Eq. (III.3). The additional "correction term" involving news in Eq. (III.10) is analogous to the correction term needed in the case of the Bondi–Sachs 4-momentum to incorporate the presence of news. Had we omitted it, the NUT 4-vector would have failed to vanish in the case when $\{D\}$ is regular. More importantly, it would have failed to satisfy the absolute conservation law.

Perhaps the simplest way to see that the NUT 4-vector is the magnetic analog of the Bondi–Sachs 4-vector even in presence of news is via the Newman–Penrose framework.[21] A straightforward translation[11] of notation yields, from Eq. (III.10),

$$*Q_{(\alpha n)}[C] = \frac{1}{4\pi} \overline{\int_C} \alpha \, \mathrm{Im}(\psi_2{}^0 + \sigma \dot{\bar{\sigma}}) \sin\theta \, d\theta \, d\phi , \qquad \text{(III.14)}$$

where Im $A = (1/2i)(A - \bar{A})$ is the imaginary part of A. The expression for the Bondi–Sachs 4-momentum is

$$Q_{(\alpha n)}[C] = \frac{1}{4\pi} \int_C \alpha \, \mathrm{Re}(\psi_2{}^0 + \sigma \dot{\bar{\sigma}}) \sin\theta \, d\theta \, d\phi . \qquad \text{(III.15)}$$

As is well known, duality rotations in space-time correspond to a switch of real and imaginary parts in the Newman–Penrose notation. For example, in gauge theories, if we set $\Phi_1^0 = F_{ab}(l^a n^b + m^a \bar{m}^b)$ on $\mathscr{I}$, the magnetic charge $*Q[C]$ is given by

$$*Q[C] = \frac{1}{4\pi} \int_C \mathrm{Im} \, \Phi_1^0 \sin\theta \, d\theta \, d\phi ,$$

while the electric charge $Q[C]$ has the expression

$$Q[C] = \frac{1}{4\pi} \int_C \mathrm{Re} \, \Phi_1^0 \sin\theta \, d\theta \, d\phi$$

at the cross section C of $\mathscr{I}$. (Here, we have suppressed the internal indices for simplicity.)

Remark: In the Newman–Penrose framework, one has the identity $\mathrm{Im}(\psi_2^0 + \sigma \dot{\bar{\sigma}}) = \mathrm{Im} \, \bar{\delta}^2 \sigma$. Hence, we have

$$*Q_{(\alpha n)}[C] = \frac{1}{4\pi} \overline{\int_C} (\mathrm{Im} \, \bar{\delta}^2 \sigma) \alpha \sin\theta \, d\theta \, d\phi$$
$$= \frac{1}{4\pi} \overline{\int_C} (\mathrm{Im} \, \bar{\delta}^2 \, \alpha\sigma) \sin\theta \, d\theta \, d\phi , \qquad \text{(III.16)}$$

since $\bar{\delta}^2 \alpha = 0$ when αn^a is a BMS translation. It is therefore clear that the NUT 4-momentum vanishes identically when the shear, σ, is regular. Similarly, the reason behind our assumption that $N_{ab} \gamma_{cd} q^{bc}$ be bounded is also clear: This boundedness is necessary and sufficient for Im $\bar{\delta}^2 \sigma$ to be bounded (if the "fields" $*K^{ab}$ and N_{ab} are regular.) Thus, the basic ideas of the present framework become simple and transparent in the Newman–Penrose formalism. The reason why we have not worked with this formalism in the main body of the paper is twofold. First, it is both aesthetically pleasing and convenient to have available the expressions [such as Eq. (III.3)] for basic quantities without reference to

specific conformal frames and choices of l_a. Second, since the Newman–Penrose framework fixes, right in the beginning, a slicing of $\mathscr{I}$ by 2-spheres (u = constant) to which expressions such as Eq. (III.14) are tied, it is ill-suited to analyze the question of whether or not there is a flux of NUT 4-momentum through an *arbitrary* patch of $\mathscr{I}$. As the example of the linkage formula[22] for the ("electric") supermomentum shows, it *can* happen that the flux of a quantity through the region bounded by *any* two u = const cuts of $\mathscr{I}$ can vanish in *every* Bondi frame even though the flux through a "generic" patch of $\mathscr{I}$ fails to vanish. Thus, the central result on absolute conservation of the NUT 4-momentum would have been difficult to prove in the Newman–Penrose notation. As is often the case, the "intrinsic" framework seems to be better suited for proving the basic theorems while the calculations in specific examples are easier in the Newman–Penrose formalism.

B. ${}^*Q_{(\alpha n)}[C]$ is absolutely conserved

Consider a region Δ of $\mathscr{I}$ bounded by any two cross sections C_1 and C_2. Then, it follows by Stokes' theorem that

$$Q_{(\alpha n)}[C_1] - Q_{(\alpha n)}[C_2] = \overline{\int_\Delta} D_{[a}F_{bc]}\, dS^{abc}, \qquad \text{(III.17)}$$

where

$$8\pi F_{bc} = \frac{\alpha}{4} {}^*K^{pm} l_m\, \epsilon_{pbc} + (\alpha D_b l_m + l_b D_m \alpha) S_c{}^m + 2q^{mn}(D_c D_n \alpha) D_b l_m . \qquad \text{(III.18)}$$

Here, the integral $\overline{\int}$ is performed, as before, by first surrounding the singularities of D by boundaries, then integrating on the region where D is regular, and finally taking the limit in which boundaries pinch to the wire singularities (see Fig. 1). Taking the curl of each of the three terms in the expression of F_{bc}, we obtain

$$\epsilon^{abc} D_a(\tfrac{1}{4}\alpha\, {}^*K^{pm} l_m\, \epsilon_{pbc}) = \tfrac{1}{2} {}^*K^{am}(\alpha D_a l_m + l_m D_a \alpha),$$

$$\epsilon^{abc} D_a(\alpha D_b l_m + l_b D_m \alpha) S_c{}^m = -\tfrac{1}{2} {}^*K^{am}(\alpha D_a l_m + l_a D_m \alpha) + \epsilon^{abc}[S_b{}^m (D_a l_m) D_c \alpha + (D_c l_a)(D_m \alpha) S_b{}^m + l_a (D_c D_m \alpha) S_b{}^m],$$

and

$$\epsilon^{abc} D_a(2q^{mn}(D_c D_n \alpha) D_b l_m) = -\epsilon^{abc}[S_b{}^m (D_a l_m) D_c \alpha + (D_c l_a)(D_m \alpha) S_b{}^m + l_a (D_c D_m \alpha) S_b{}^m].$$

Thus, $\epsilon^{abc} D_a F_{bc} = 0$, whence $D_{[a}F_{bc]} = 0$ in the region where D is C^∞. Therefore, it follows that

$$\overline{\int_\Delta} D_{[a}F_{bc]}\, dS^{abc} = 0 .$$

Thus, we have

Theorem 2: Fix a radiative mode $\{D\}$ and any two cross sections C_1 and C_2 of $\mathscr{I}$. Then, $Q_{(\alpha n)}[C_1] = Q_{(\alpha n)}[C_2]$ for any BMS translation αn^a.

Remark: The absolute conservation of the NUT 4-momentum is surprising in the light of the fact that in the nonabelian gauge theories, magnetic charges can be radiated through $\mathscr{I}$. How does this difference arise in the two cases?

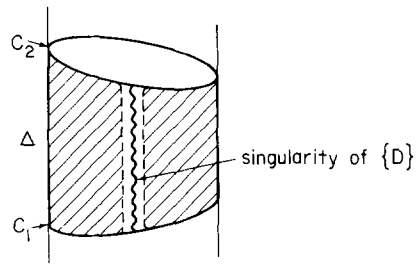


FIG. 1. Consider, for simplicity, the case in which the singularity of $\{D\}$ lies along a single generator. Applying Stokes' theorem to the shaded part P of Δ, we obtain, $\int_{\partial p} F_{ab}\, dS^{ab} = \int_p D_{[a}F_{bc]}\, dS^{abc}$, where F_{ab} is defined in Eq. (III.18). Since F_{ab} is bounded even on the singularity of $\{D\}$ (see the proof of Lemma 1.1), the contribution to the integral over ∂p from the tube surrounding the singularity vanishes in the limit as the tube shrinks to converge on the singularity. Therefore, in the limit, one obtains Eq. (III.17).

Let us begin by recalling the situation in the Yang–Mills theory in Minkowski space. Fix a cross section C of the null infinity of Minkowski space. The Yang–Mills magnetic charge is given by

$${}^*Q^\alpha{}_\beta[C] = \int_C \underset{\sim}{F}{}^\alpha_{ab\,\beta}\, dS^{ab} \int_C \underset{\sim}{F}{}^\alpha_{ab\,\beta}\, dS^{ab},$$

where $\underset{\sim}{F}_{ab}$ is the pull-back to $\mathscr{I}$ of the Yang–Mills field F_{ab}. (Here, and in what follows, internal indices are represented by Greek letters.) On $\mathscr{I}$, $\underset{\sim}{F}_{ab}$ satisfies

$$\underset{\sim}{D}_{[a}\underset{\sim}{F}{}^\alpha_{bc]\,\beta} \equiv \underset{\sim}{D}_{[a}\underset{\sim}{F}{}^\alpha_{bc]\,\beta} - A^\alpha_{[a|\lambda|}\, F^\lambda_{bc]\,\beta} + A^\lambda_{[a|\beta|}\, F^\alpha_{bc]\,\lambda} = 0,$$

where $\underset{\sim}{D}$ is the gauge covariant operator. Hence, the flux of ${}^*Q^\alpha{}_\beta$ through a region Δ of $\mathscr{I}$, bounded by cross sections C_1 and C_2, is given by

$${}^*Q^\alpha{}_\beta[C_1] - {}^*Q^\alpha{}_\beta[C_2] = \int_\Delta \underset{\sim}{D}_{[a}\underset{\sim}{F}_{bc]}\, dS^{abc}$$

which is, in general, nonzero. Thus, it is because Stokes' theorem involves derivative operators D while the Yang–Mills equation involves the gauge covariant operator $\underset{\sim}{D}$ that ${}^*Q^\alpha{}_\beta$ can be radiated away. In general relativity, because of the geometrical nature of the gravitational field, the analog of the gauge covariant operator is just a torsion-free affine connection; the F_{bc} which features in Eq. (III.17) satisfies $D_{[a}F_{bc]} = 0$. Thus, it is because the derivative operator which features in the field equations can also be used directly in Stokes' theorem, that ${}^*Q_{(\alpha n)}$ is absolutely conserved. To summarize, the difference arises from the fact that there is no distinction between the space-time and the internal indices in general relativity because the gravitational field describes the geometry of space-time itself, rather than of a bundle constructed on space-time.

IV. CLASSICAL VACUA

The flux of the Bondi–Sachs 4-momentum through any patch of $\mathscr{I}$ vanishes identically if the news tensor N_{ab} of $\{D\}$ is zero. Hence, while dealing with the radiative modes of the gravitational field, it is natural to regard the equivalence classes $\{D\}$ with vanishing news as the "classical ground states", or, "classical vacua".[23] In the Yang–Mills theory as well as in several exactly soluble models in 1 + 1 dimensional space-times, the analogous classical field configurations play a significant role in the quantum domain. It turns out

that the situation is similar in the gravitational case. In this section, we classify the classical vacua in detail and indicate the role played by this "classical vacuum degeneracy" in the quantum theory. The details of quantum considerations will appear elsewhere.

In Ref. 7, a partial classification was carried out by restricting attention to smooth configurations $\{D\}$ with vanishing N_{ab} as well as ${}^*K^{ab}$. We now consider the general case. Fix a Bondi frame (q_{ab}, n^a) and two configurations $\{D\}$ and $\{D'\}$, which have vanishing news tensors *and the same NUT* 4-momentum. Let γ_{ab} characterize the difference between $\{D\}$ and $\{D'\}$. Then, we have

Theorem 3: There exist functions α and β satisfying $\mathcal{L}_n \alpha = 0$ and $\mathcal{L}_n \beta = 0$ such that $\gamma_{ab} \doteq D_a D_b \alpha + q^{np} n^m \epsilon_{mn(a} D_{b)} D_p \beta$, where $\doteq$ stands for the equality of tracefree parts. α and β are unique up to an addition of BMS translations.

Proof: Let D and D' be any connections in $\{D\}$ and $\{D'\}$, respectively. Since the news tensors vanish, curvature tensors $S_a{}^b$ and $S'_a{}^b$ of D and D' are given by $S_a{}^b = \delta_a{}^b + A_a n^b$ and $S'_a{}^b = \delta_a{}^b + A'_a n^b$ for some vectors A_a and A'_a satisfying

$$ {}^*K^{ab} = 2\epsilon^{amn}(D_m A_n) n^b \quad \text{and} \quad {}^*K'^{ab} = 2\epsilon^{amn}(D'_m A'_n) n^b . \tag{IV.1} $$

Since the news vanishes, we have[3,12]

$$ {}^*K^{ab} = \mu n^a n^b \quad \text{and} \quad {}^*K'^{ab} = \mu' n^a n^b \tag{IV.2} $$

for some C^∞ functions μ and μ' satisfying $\mathcal{L}_n \mu = 0$ and $\mathcal{L}_n \mu' = 0$. Therefore, it follows that $4D_{[m} A_{n]} = \mu \epsilon_{amn} n^a$ and $4D_{[m} A'_{n]} = \mu' \epsilon_{amn} n^a$. Using the available gauge freedom in the choice of individual connections in any given equivalence class, we can always find D in $\{D\}$ and D' in $\{D'\}$ such that $A_a n^a = 0$ and $A'_a n^a = 0$ (except possibly on isolated generators.) From now on, we shall assume that such a choice is made. Then, since $n^a D_{[a} A_{b]} = 0$ and $n^a D_{[a} A'_{b]} = 0$, it follows that $\mathcal{L}_n A_a = 0$ and $\mathcal{L}_n A'_a = 0$, set

$$ v_a = A_a - A'_a . \tag{IV.3} $$

Then,

$$ v_a n^a = 0, \quad \mathcal{L}_n v_a = 0, \text{ and } 4D_{[a} v_{b]} = (\mu - \mu')\epsilon_{abc} n^c . \tag{IV.4} $$

Thus, v_a is the pull-back to $\mathscr{I}$ of a 1-form $\hat{v}_a$ on the 2-sphere $\mathscr{S}$ of generators. Hence, we can project the last of Eq. (IV.4) on $\mathscr{S}$[24]:

$$ 4\hat{D}_{[a} \hat{v}_{b]} = (\hat{\mu} - \hat{\mu}') \hat{\epsilon}_{ab} , \tag{IV.5} $$

where $\hat{D}$ and $\hat{\epsilon}_{ab}$ are the derivative operator and the alternating tensor on $\mathscr{S}$, compatible with the 2-sphere metric $\hat{q}_{ab}$ thereon, obtained by projection of q_{ab} on $\mathscr{I}$, and $\hat{\mu}$ and $\hat{\mu}'$ are the projections to $\mathscr{S}$ of μ and μ'. Since $\{D\}$ and $\{D'\}$ have the *same* NUT 4-momentum, it follows from Eq. (III.13) that

$$ \int_{\mathscr{S}} \hat{\alpha} (\hat{\mu} - \hat{\mu}') \hat{\epsilon}_{ab} \, dS^{ab} = 0 \tag{IV.6} $$

for all BMS translation αn^a. Hence, it follows that $(\hat{\mu} - \hat{\mu}')$ cannot be in the kernel of the operator $\hat{D}^m \hat{D}_m$ (or, $\hat{D}^m \hat{D}_m + 2$). This enables us to solve Eq. (IV.5). The *general* solution is

$$ \hat{v}_a = \hat{\epsilon}_{am} \hat{D}^m \hat{h} + \hat{D}_a \hat{K} , \tag{IV.7a} $$

where $\hat{h}$ satisfies

$$ 2\hat{D}^a \hat{D}_a \hat{h} = (\hat{\mu} - \hat{\mu}') . \tag{IV.7b} $$

[Note that the freedom to add a constant to $\hat{h}$ in Eq. (IV.7b) does not affect the solution (IV.7a).] Here, $\hat{K}$ is an arbitrary function[25] (possibly singular at a finite number of points) on $\mathscr{S}$.

Let us now consider Σ_{ab} which relates D and D',

$$ (D_a - D'_a) K_b = \Sigma_{ab} n^c K_c \tag{IV.8} $$

for all K_b. Using the expression (II.3) of Riemann tensors of D and D' and the definition (IV.3) of v_a, it follows that Σ_{ab} must satisfy

$$ D_{[a} \Sigma_{b]c} = q_{c[a} v_{b]} . \tag{IV.9} $$

Since $N_{ab} = 0$ and $N'_{ab} = 0$ this equation can also be projected down to $\mathscr{S}$:

$$ \begin{aligned} \hat{D}_{[a} \hat{\Sigma}_{b]c} &= \hat{q}_{c[a} \hat{v}_{b]} \\ &= \hat{q}_{c[a} \hat{\epsilon}_{b]m} \hat{D}^m \hat{h} + \hat{q}_{c[a} \hat{D}_{b]} \hat{K} . \end{aligned} \tag{IV.10} $$

A particular solution to this equation is

$$ \hat{\Sigma}_{ab} = \hat{D}_a \hat{D}_b \hat{K} + \hat{\epsilon}_{m(a} \hat{D}_{b)} \hat{D}^m \hat{\beta} , \tag{IV.11a} $$

where $\hat{\beta}$ satisfies

$$ \tfrac{1}{2} \hat{D}^a \hat{D}_a \hat{\beta} + \hat{\beta} = -\hat{h} . \tag{IV.11b} $$

The last equation admits a solution because Eqs. (IV.7b) and (IV.6) imply that $\hat{h}$ cannot be in the kernel of the operator $(\hat{D}^a \hat{D}_a + 2)$. Finally, the general solution to the homogeneous equation, $\hat{D}_{[a} \hat{\Sigma}_{b]c} = 0$, is known[7] to be $\Sigma_{ab} = \hat{D}_a \hat{D}_b \hat{g} + \hat{g} q_{ab}$. Hence, the general solution to (IV.10) is

$$ \hat{\Sigma}_{ab} = \hat{D}_a \hat{D}_b \hat{\alpha} + \hat{\epsilon}_{m(a} \hat{D}_{b)} \hat{D}^m \hat{\beta} + \hat{g} \hat{q}_{ab} , $$

where we have set[25] $\hat{\alpha} = \hat{K} + \hat{g}$. Since Σ_{ab} on $\mathscr{I}$ is the pull-back of $\hat{\Sigma}_{ab}$, the tracefree part, $\gamma_{ab} = \Sigma_{ab} - \frac{1}{2} q_{ab} q^{mn} \Sigma_{mn}$, of Σ_{ab} is given by

$$ \gamma_{ab} \doteq D_a D_b \alpha + q^{np} n^m \epsilon_{mn(a} D_{b)} D_p \beta . \tag{IV.12} $$

Finally, we consider the issue of uniqueness of α and β. Note first, that given Eq. (IV.12), one can work backwards and show that $\hat{\mu}$ and $\hat{\mu}'$ must be related by

$$ (\hat{D}^a \hat{D}_a + 2)(\hat{D}^m \hat{D}_m) \hat{\beta} = \hat{\mu}' - \hat{\mu} . \tag{IV.13} $$

[See Eqs. (IV.11b) and (IV.7b).] Hence, if $(\bar{\alpha}$ and $\bar{\beta})$ are also permissible potentials for the given γ_{ab} [via Eq. (IV.12)], $\bar{\beta}$ must also satisfy Eq. (IV.13). Hence, $\hat{\Delta} = \hat{\beta} - \bar{\beta}$ satisfies

$$ (\hat{D}^a \hat{D}_a + 2)(\hat{D}^m \hat{D}_m) \hat{\Delta} = 0 . $$

This is possible if and only if Δn^a is a BMS translation, i.e., if and only if $D_a D_b \Delta$ is proportional to q_{ab} on $\mathscr{I}$. Equation (IV.12) for (α, β) and (α', β') now implies that $\delta = \alpha - \alpha'$ must satisfy $D_a D_b \delta \doteq 0$, i.e., δn^a must also be a BMS translation. Thus α and β are each unique up to the addition of a BMS translation.

Remarks: (i) Let $\{D^0\}$ denote a regular configuration with vanishing N_{ab} and ${}^*K^{ab}$. Clearly, the NUT 4-momentum of $\{D^0\}$ vanishes identically. Let $\{D\}$ be any classical vacuum with vanishing NUT 4-momentum. Let us use $\{D^0\}$ for $\{D'\}$ in Theorem 3. Since $A^0_a = 0$ and $\mu^0 = 0$, it follows

that A_a must satisfy $4\hat{D}_{[a}\hat{A}_{b]} = \hat{\mu}\,\hat{\epsilon}_{ab}$. Since the right side is C^∞ and $\int \hat{\mu}\,\hat{\alpha}\,\hat{\epsilon}_{ab}\,dS^{ab} = 0$ for all BMS translations $\hat{\alpha}$, it follows that (possibly after a gauge transformation) A_a is C^∞. Hence, the functions α and β of Eq. (IV.12) are also C^∞ modulo gauge whence γ_{ab} is C^∞ and $\{D\}$ is regular. Thus, every classical vacuum with vanishing NUT 4-momentum is regular. Let us now consider a general configuration $\{D\}$ for which the news vanishes in the past (or future) of some cross section. If the NUT 4-momentum of $\{D\}$ vanishes, then the restriction of $\{D\}$ to the past of the given cross section is a *regular* classical vacuum. That is, the tensor field γ_{ab}, relating $\{D\}$ to $\{D^0\}$, is C^∞ in the past of this cross section. Since the news tensor N_{ab} of $\{D\}$ satisfies $N_{ab} = -2\mathcal{L}_n\,\gamma_{ab}$ and since N_{ab} is, by assumption, C^∞, γ_{ab} is C^∞ everywhere on $\mathcal{I}$, i.e., $\{D\}$ is regular. Thus, configurations $\{D\}$ whose news vanishes in the distant past (or future) on $\mathcal{I}$ are nutty if and only if their NUT 4-momentum fails to vanish. (ii) In the Newman–Penrose notation,[11] a classical vacuum is a configuration σ with $\dot{\sigma} = 0$ (which is C^∞ everywhere except for a finite number of generators and has the property that Im $\bar{\delta}^2\sigma$ is bounded on $\mathcal{I}$.)

Denote by V the space of classical vacua. Since the news tensor, N_{ab}, of *any* connection $\{D\}$ is given by $N_{ab} = -2\mathcal{L}_n\,\gamma_{ab}$, where γ_{ab} is the tensor field relating $\{D\}$ to any $\{D^0\}$ with vanishing news and ${}^*K^{ab}$, it follows that V is an affine space. We can use Theorem 3 to explore the structure of V. Using the notion of the NUT 4-momentum, we can first divide V into subspaces, each subspace $V({}^*p_a)$, containing the vacua with a given, fixed NUT 4-momentum ${}^*p_a$. Theorem 3 tells us that each $V({}^*p_a)$ is an *affine* subspace of V and that any two elements $\{D\}$ and $\{D'\}$ of $V({}^*p_a)$ are related to each other by two elements, α and β, of the quotient ST/T of the BMS supertranslation group by its translation subgroup. [Note that the value ${}^*p_a$ of the NUT 4-momentum does not affect the "size" of the corresponding subspace $V({}^*p_a)$.] Thus, there exist a 2-fold degeneracy: The presence of the NUT charge leads to a 4-parameter family of sectors within V and each sector contains as many vacuum configurations as there are elements of $(ST/T)\otimes(ST/T)$. In quantum theory the two sources of degeneracy play distinct roles. The first, the NUT charge, leads via superselection to inequivalent representations of the CCR, i.e., to "charged sectors". Within each sector, there is further degeneracy which manifests itself via infrared problems. Thus, within each (NUT $-$) charge sector there are infrared subsectors which emerge due to fluctuations of classical configurations connecting two distinct classical vacua[6] with the same NUT 4-momentum.

V. DISCUSSION

We saw in Sec. III that, when $N_{ab} = 0$, the present definition of the NUT 4-momentum reduces to that in Ref. 3. In this subsection, we shall show further that there is a sense in which the Ramaswamy–Sen construction itself can be recovered from the present framework.

The remarks in Sec. II and Eq. (III.1) suggests that a description free of "wire singularities" in A_a can be obtained by constructing a smooth U(1) bundle B over $\mathcal{I}$ with a smooth connection one-form A_α as in electrodynamics. (Here Greek indices refer to B.) Fix such a bundle with a connection. Then, since A_α defines horizontal subspaces of B, one can pull-back both contra- and covariant tensor fields on $\mathcal{I}$ onto B. The resulting tensor fields on B will necessarily be smooth. This construction serves to clarify the meaning of ${}^*Q_{(\bar{n})}[C]$ in Eq. (III.1) as measuring the "twist" in the U(1) bundle but is otherwise devoid of physical significance. For example, the connection A_α does not play the role of the connection in the physics of gravitational radiation.

Now consider the case when $N_{ab} = 0$. Then, we claim, the bundle B can be reduced to a U(1) bundle $\hat{B}$ over the space $\mathcal{S}$ of generators of $\mathcal{I}$. Furthermore, there is a unique connection $A_{\hat{\alpha}}$ on $\hat{B}$ induced by the connection A_α on B. To establish these results, let n^α be the *unique* lift to B of n^a. In other words, $n^\alpha A_\alpha = 0$ and $\pi^*(n^\alpha) = n^a$, where $\pi: B \to \mathcal{I}$ is the bundle projection. Let $\hat{B}$ be the manifold of orbits of n^α. Since the news vanishes, $n^\alpha A_\alpha = 0 \Rightarrow \mathcal{L}_n A_\alpha = 0$. Hence there exists a unique one-form $A_{\hat{\alpha}}$ on $\hat{B}$ whose pull-back on B is A_α. Next we show that n^α maps whole fibers of B into whole fibers, i.e., $\mathcal{L}_n v^\alpha = 0$, where v^α is the vertical vector field. In general $\mathcal{L}_n v^\alpha = k v^\alpha$ for some function k. [Let m_a, $\bar{m}_a$, l_a be linearly independent covectors on $\mathcal{I}$ such that $\mathcal{L}_n m_a = \mathcal{L}_n \bar{m}_a = \mathcal{L}_n l_a = 0$ and let $m_\alpha = \pi_*(m_a)$, etc. If n'^α is *any* lift of n^a then $\mathcal{L}_{n'} m_\alpha = 0$, etc., since $\mathcal{L}_{n'} m_\alpha = \pi_*(\mathcal{L}_n m_a)$, etc. Now v^α is uniquely determined up to a scale factor by $v^\alpha m_\alpha = v^\alpha \bar{m}_\alpha = v^\alpha l_\alpha = 0$. Then clearly $(\mathcal{L}_{n'} v^\alpha) m_\alpha = 0$, etc., whence $\mathcal{L}_{n'} v^\alpha = k' v^\alpha$.] We show that $k = 0$. Since $A_\alpha v^\alpha = 1$, we have $0 = \mathcal{L}_n A_\alpha v^\alpha = A_\alpha \mathcal{L}_n v^\alpha = k A_\alpha v^\alpha = k$. It is now easy to see that $\hat{B}$ is indeed a U(1) bundle over $\mathcal{S}$ and that $A_{\hat{\alpha}}$ is a connection on $\hat{B}$.

This reduced bundle $\hat{B}$ is naturally isomorphic to the null boundary, $\mathcal{I}_{R-S}$, constructed in Ref. 3. [There is a family of structure preserving isomorphisms between $\hat{B}$ and $\mathcal{I}_{R-S}$. If the isomorphism on the base space $\mathcal{S}$ of B and $\mathcal{I}_{R-S}$ is fixed, then the isomorphism is unique up to local U(1) transformations.] Note, however, the conceptual difference between the two. $\mathcal{I}_{R-S}$ is to be regarded as a boundary of a physical space-time while $\hat{B}$ is a construction which incorporates the gauge transformations on $\mathcal{I}$ and is not directly related to space-time.

[1] E. T. Newman, L. Tamburino, and T. Unti, J. Math. Phys. **4**, 915 (1963).
[2] C. W. Misner, J. Math. Phys. **4**, 924 (1963).
[3] S. Ramaswamy and A. Sen, J. Math. Phys. **22**, 2612 (1981).
[4] H. Bondi, M. G. J. Van der Burg, and A. W. K. Metzner, Proc. R. Soc. (London) Ser. A **269**, 21 (1962); R. K. Sachs, *ibid.* **270**, 103 (1962).
[5] Note also that our expression for the NUT 4-momentum contains terms involving Bondi news and is therefore not the "obvious" generalization of Eq. (21) of Ref. 3. Also, we shall find that, in presence of news, supermomenta cannot be introduced in a satisfactory manner.
[6] A. Ashtekar, Phys. Rev. Lett. **46**, 573 (1981); in *Quantum Gravity 2*, edited by C. J. Isham, R. Penrose, and D. W. Sciama (Oxford U. P., Oxford, 1975).
[7] For details, see A. Ashtekar, J. Math. Phys. **22**, 2885 (1981).
[8] See, e.g., A. Magnon-Ashtekar, J. Math. Phys. **22**, 2012 (1981).
[9] See, e.g., S. W. Hawking, D. Page, and C. N. Pope, Nucl. Phys. B **170**, 283 (1980).
[10] B. G. Schmidt, M. Walker, and P. D. Sommers, Gen. Relativ. Gravit. **6**, 489 (1975).

[11]In the Newman–Penrose notation, $N_{ab}m^a m^b = -2\dot{\sigma}$ and the five independent components of ${}^*K^{ab}$ correspond to ψ_4^0, ψ_3^0 and Im ψ_2^0. The two degrees of freedom in $\{D\}$ are coded in σ. For details, see C. N. Kozomeh and E. T. Newman, "A Note on Asymptotically Flat Spaces II" (preprint). See also Sec. III of Ref. 3 or Appendix A of Ref. 7.

[12]R. Geroch, in *Asymptotic Structure of Space-time*, edited by P. Esposito and L. Witten (Plenum, New York, 1977).

[13]Our conventions are the following. Curvature tensor is defined by $D_{[a}D_{b]}k_c = \frac{1}{2}\mathscr{R}_{abc}{}^d k_d$. q_{ab} has signature $(0++)$. ϵ^{abc} is defined by $\epsilon^{abc}\epsilon^{mnr}q_{bn}q_{am} = 2n^c n^r$ and ϵ_{abc} by $\epsilon^{abc}\epsilon_{abc} = 3!$

[14]Note that there is an error of a factor of 2 in the expression of $\mathscr{R}_{abc}{}^d$ in Ref. 7. Equation (II.3) of the present paper corrects this error.

[15]Why do we not require that f be C^∞? This requirement would lead us to treat $\{D\}$ and $\{D'\}$ as being distinct if $(D'_a - D_a)k_b = fq_{ab}n^c k_c$ and if f has wire singularities, and add spurious degrees of freedom to the radiative modes $\{D\}$. (Note, that, even if f has wire singularities, $\{D'\}\in\overline{\mathscr{C}}$ iff $\{D\}\in\overline{\mathscr{C}}$. Thus, if $\{D\}$ is admissible, so is $\{D'\}$.)

[16]Note that by a gauge transformation of Eq. (II.8) with $f = -u$, A_a can be transformed to $-4l(1-\cos\theta)D_a\phi$, which is the familiar expression of the vector potential of a magnetic monopole in Maxwell's theory.

[17]Note that, since all derivative operators satisfying Eq. (II.1) have the same action on covectors satisfying $k_a n^a = 0$, $D_a D_b\,\alpha$ is a C^∞ tensor field even when $\{D\}$ is nutty. Hence, in particular, μ is C^∞.

[18]R. K. Sachs, Phys. Rev. **128**, 2851 (1962); J. Winicour, J. Math. Phys. **9**, 861 (1968); A. Ashtekar and M. Streubel, Proc. R. Soc. (London) Ser. A **376**, 585 (1981); R. Geroch and J. Winicour, J. Math. Phys. **22**, 803 (1981). Note, however, that these references contain two distinct definitions of supermomentum.

[19]Possible wire singularities in f do not cause complications in the proof because we have used the regularized integral $\overline{\int}$.

[20]Note that ω is *any* C^∞, nowhere vanishing function on $\mathscr{I}$, satisfying $\mathscr{L}_n\omega = 0$; unlike in the argument in the beginning of this section, ω is unrelated to α.

[21]E. T. Newman and R. Penrose, Proc. R. Soc. (London) Ser. A **305**, 175 (1968). Since the duality transformation uses the 4-dimensional alternating tensor ϵ_{abcd}, it is somewhat cumbersome to see its effect on quantities defined intrinsically on $\mathscr{I}$. Since the required manipulations are well known in the Newman–Penrose framework, we have chosen this method here.

[22]J. Winicour, J. Math. Phys. **9**, 861 (1968); R. Geroch and J. Winicour, *ibid.* **22**, 803 (1981).

[23]Note that this terminology is different from that in Ref. 7, where a configuration $\{D\}$ was called a classical vacuum if it has a vanishing N_{ab} *and* ${}^*K^{ab}$, i.e., if its curvature is trivial. From radiative energy considerations, the present terminology appears to be more natural.

[24]Since all connections D in $\overline{\mathscr{C}}$ agree on covectors K_a satisfying $K_a n^a = 0$, $\hat{D}$ on $\mathscr{I}$ is a C^∞ connection. Hence, solutions to the differential equations that follow are C^∞ modulo gauge.

[25]Arbitrariness in the choice of $\hat{K}$ corresponds to the freedom in the choice of connections D and D' in $\{D\}$ and $\{D'\}$. If we make a gauge transformation $D_a w_b \to \bar{D}_a w_b + fq_{ab}n^c w_c$ such that $A_a n^a = \bar{A}_a n^a = 0$, then $\hat{v}_a \to \hat{v}_a + 2D_a f$. One can use this gauge freedom to ensure that the function α of Eq. (IV.12) is C^∞.

Theory of nonlinear Langevin equation with quadratic noise

K. Wódkiewicz
Max-Planck Institute for Quantum Optics, D-8046 Garching, Federal Republic of Germany and Institute of Theoretical Physics, Warsaw University, Warsaw 00-681, Poland[a]

(Received 13 April 1982; accepted for publication 28 May 1982)

A generalized Langevin equation with a multiplicative linear and quadratic Ornstein–Uhlenbeck stochastic noise is investigated. An exact generalized Fokker–Planck equation for this Langevin equation is obtained. This generalized Fokker–Planck equation has the form of an integro–differential equation. An exact form of the integral kernel is given in terms of an operator-valued matrix-continued fraction. In the limit of small fluctuations and short coherence time of the noise a standard Fokker–Planck equation with drift and diffusion is derived in a systematic way.

PACS numbers: 05.40. + j

I. INTRODUCTION

Langevin equations with an external driving noise have been used in physics for many years in order to describe the statistical or the dynamical properties of microscopic systems characterized by a small number of variables. In the last years a lot of effort has been spent in the investigation of nonlinear open systems subjected to an external driving noise. Examples of such dynamical systems can be found in many branches of physics.[1] If an atom or a molecule is driven by a partially coherent external source of light the simplest stochastic equation which describes the atomic degrees of freedom has a multiplicative noise due to the external fluctuations of the driving electric field.[2–5] But it is well known that the main dynamical properties of the atomic system, for instance Stark broadening,[6,7] ionization rates,[8,9] or photon counting[10] of the fluorescence light, are related to the fluctuations of the field intensity, i.e., to the square of the fluctuating electric field amplitude. This means that beside the multiplicative linear noise we have to add a quadratic multiplicative noise to our dynamical Langevin equation. For obvious physical reasons the power spectrum of the superimposed external noise has in most cases a finite bandwidth, i.e., it cannot be treated as a standard Gaussian white noise. These physical applications and the property of the driving noise are the main motivation for the investigation in this paper of the following generalized Langevin equation:

$$\frac{\partial}{\partial t}\rho(t) = F(\rho) + ix(t)G(\rho) + ix^2(t)H(\rho), \tag{1.1}$$

where F, G, and H are arbitrary functions of ρ and $x(t)$ is an Ornstein–Uhlenbeck stochastic process. The nonlinear terms G and H are in most cases the result of an adiabatic elimination of some of the degrees of freedom of the atomic system.[11–14] Stochastic equations of this type have been also obtained in electronics[15,16] and chemical reactions.[17]

In this paper we shall derive an exact generalized Fokker–Planck equation for the density probability of the dynamical variable ρ. This generalized Fokker–Planck equation has the form of an integro-differential equation with an infinite number of derivatives. In Secs. II and III we derive this exact equation. We show that the form of this equation is closely related to the so-called vector tridiagonal recurrence relation.[18] A natural way of investigating such a tridiagonal recurrence relation is by use of the theory of matrix-continued fractions. Our generalized Fokker–Planck equation can be solved formally in terms of such an operator-valued matrix-continued fraction. We derive the general form of this operator-continued fraction for the probability distribution associated with the stochastic equation (1.1). The derivation and the form of this exact generalized Fokker–Planck equation for the stochastic equation (1.1) with the Ornstein–Uhlenbeck noise is new. This equation is the main result of this paper. In Sec. IV we introduce a systematic approximate method based on a power series expansion of the generalized Fokker–Planck equation. In the limit of small fluctuations and short coherence time of the noise we obtain a standard Fokker–Planck equation, i.e., an equation involving only a drift and diffusion term. We derive in this case the form of these two important terms as functions of arbitrary F, G, and H. With this form the stationary solution of this standard Fokker–Planck equation is obtained. In fact we show that a standard Fokker–Planck equation, i.e., an equation containing only first- and second-order derivatives, is obtained only in the lowest order of the parameter describing the strength of the noise. Higher-order contributions bring higher derivatives to the proper equation for the probability distribution. However, in many physical applications these additional contributions can be approximated by a standard Fokker–Planck operator. We comment in Sec. IV on a recent method proposed for an approximate derivation of a Fokker–Planck equation related to the Langevin equation (1.1).[19] Finally, in the last section V we discuss a generalization of our approach to the calculation of a transition probability needed when correlations of ρ are investigated.

II. LIOUVILLE EQUATION

It is well known that we can transform the nonlinear Langevin equation (1.1) into the following linear Liouville-type equation[20]:

$$\frac{\partial}{\partial t}\varphi(t) = (M_0 + ix(t)M_1 + ix^2(t)M_2)\varphi(t), \tag{2.1}$$

where the stochastic density $\varphi(t)$ is defined as follows:

$$\varphi(t,\xi) = \delta(\xi - \rho(t)), \tag{2.2}$$

and the linear differential operators M_0, M_1, and M_2 are given by the following expressions:

[a] Permanent address.

 0022-2488/82/112179-06$02.50 © 1982 American Institute of Physics

$$M_0 = -\frac{\partial}{\partial \xi} F(\xi), \quad M_1 = -\frac{\partial}{\partial \xi} G(\xi),$$
$$M_2 = -\frac{\partial}{\partial \xi} H(\xi). \tag{2.3}$$

From the Liouville equation, by a stochastic average over all possible realizations of the Gaussian random process $x(t)$ we can calculate the onefold Fokker–Planck distribution:

$$p(t,\xi) = \langle \varphi(t,\xi) \rangle, \tag{2.4}$$

and accordingly we can obtain all the moments of ρ: $\langle \rho^n(t) \rangle = \int d\xi\, \xi^n p(t,\xi)$. Because of the nonwhite character of the noise $x(t)$ the onefold probability distribution $p(t,\xi)$ is not sufficient to compute correlations of the dynamical variable $\rho(t)$ like, for example, $\langle \rho(t)\rho(t') \rangle$ Nevertheless, we shall call this probability distribution a Fokker–Planck distribution and will come back to the correlation function problem in the last paragraph of this work.

It is well known that if $H = 0$ (i.e., $M_2 = 0$) and for a Gaussian white noise $\langle x_w(t) x_w(t') \rangle = 2D\delta(t - t')$ the stochastic average of the Liouville equation given by Eq. (2.1) can be performed exactly and as a result we obtain the Stratonovich–Fokker–Planck equation[21–24]:

$$\frac{\partial}{\partial t} p(t,\xi) = (M_0 - DM_1^2)\, p(t,\xi). \tag{2.5}$$

This important result obtained for a white noise $x_w(t)$ but for two arbitrary noncommuting operators M_0 and M_1 will be the key to the investigation of the proper Fokker–Planck equation associated with the Langevin equation (1.1) with the Ornstein–Uhlenbeck process $x(t)$. The Ornstein–Uhlenbeck process itself can be described by a Langevin equation with a white noise driving term:

$$\frac{d}{dt} x = -\frac{1}{\tau_c} x + x_w, \tag{2.6}$$

where in order to obtain the correlation function

$$\langle x(t) x(t') \rangle = \frac{\Gamma}{\tau_c} e^{-|t - t'|/\tau_c} \tag{2.7}$$

we have to assume that the diffusion constant D of the white noise $x_w(t)$ is given by $D = \Gamma/\tau_c^2$. Because the power spectrum of the correlation function has a Lorentzian shape with bandwidth $1/\tau_c$ in many physical applications τ_c plays an important role of a coherence time of the external noise $x(t)$. In the limit of $\tau_c \to 0$ the Ornstein–Uhlenbeck process tends to the white noise case with the diffusion constant given by Γ.

III. GENERALIZED FOKKER–PLANCK EQUATION

In this section we shall derive an exact equation for $p(t,\xi)$ given by the definition (2.4) for our Langevin equation (1.1). We shall call this equation a generalized Fokker–Planck equation though not all the statistical properties of $\rho(t)$ can be obtained from $p(t,\xi)$.

Following the methods of generating functions used in many branches of mathematical physics[20,25] we introduce the following function:

$$\phi(t,\lambda) = e^{i\lambda x(t)} \varphi(t), \tag{3.1}$$

where for notational convenience we have dropped the ξ-dependence. This function plays the role of a generating function for the Fokker–Planck distribution in the sense that its stochastic expectation value evaluated at the point $\lambda = 0$ gives exactly $p(t,\xi)$:

$$p(t,\xi) = \langle \phi(t,0) \rangle. \tag{3.2}$$

By simple time differentiation of (3.1) and with the help of Eq. (2.6) we obtain the following stochastic differential equation for $\phi(t,\lambda)$:

$$\frac{\partial}{\partial t}\phi = \left(-\frac{1}{\tau_c}\lambda \frac{\partial}{\partial \lambda} + M_0 + M_1 \frac{\partial}{\partial \lambda} - iM_2 \frac{\partial^2}{\partial \lambda^2} + ix_w\lambda \right)\phi. \tag{3.3}$$

Note that all multiplications by the noise $x(t)$ have been converted into $\partial/\partial\lambda$ derivatives and that the only noise in the stochastic equation (3.3) is given by the white noise $x_w(t)$. This means that we can average exactly this equation using the result (2.5) with $D = \Gamma/\tau_c^2$. We obtain in this way an exact differential equation satisfied by $\langle \phi(t) \rangle$:

$$\frac{\partial}{\partial t}\langle \phi \rangle = \left(-\frac{1}{\tau_c}\lambda \frac{\partial}{\partial \lambda} + M_0 + M_1 \frac{\partial}{\partial \lambda} - iM_2 \frac{\partial^2}{\partial \lambda^2} - \frac{\Gamma}{\tau_c^2}\lambda^2 \right)\langle \phi \rangle. \tag{3.4}$$

Following from the definitions (2.7) and (3.1) the initial condition for the expectation value of ϕ is equal to $\langle \phi(0,\lambda) \rangle = \exp(-\lambda^2\Gamma/2\tau_c)\,\delta(\xi - \rho(0))$, where we have assumed that $\rho(0)$ is a statistically independent quantity. In order to have a differential equation with a λ-independent initial condition we introduce an additional function g defined as follows:

$$g(t,\lambda) = e^{\lambda^2\Gamma/2\tau_c} \langle \phi(t,\lambda) \rangle. \tag{3.5}$$

This function satisfies the following equation:

$$\frac{\partial}{\partial t} g = \left(\left(-\frac{1}{\tau_c} + i2\frac{\Gamma}{\tau_c} M_2 \right)\lambda \frac{\partial}{\partial \lambda} + M_0 + i\frac{\Gamma}{\tau_c} M_2 - \frac{\Gamma}{\tau_c} M_1 \lambda + M_1 \frac{\partial}{\partial \lambda} - i\frac{\Gamma^2}{\tau_c^2} M_2 \lambda^2 - iM_2 \frac{\partial^2}{\partial \lambda^2} \right) g, \tag{3.6a}$$

with the initial condition

$$g(t,\lambda)|_{\lambda = 0} = \delta(\xi - \rho(0)) = P_0. \tag{3.6b}$$

Because we need the value of the generating function only for $\lambda = 0$ we expand g in a power series in λ:

$$g(t,\lambda) = \sum_{n=0}^{\infty} \frac{a_n(t)\lambda^n}{n!}, \tag{3.7}$$

where, following from the definitions (3.2) and (3.5),

$$p(t,\xi) = a_0(t). \tag{3.8}$$

It is clear that the differential equation (3.6) couples all the coefficients a_n in (3.7). We derive the following recurrence relation for a_n from Eq. (3.6):

$$\frac{\partial}{\partial t}a_n = \left(M_0 + n\left(-\frac{1}{\tau_c} + i3\frac{\Gamma}{\tau_c}M_2\right)\right)a_n - i\frac{\Gamma}{\tau_c}M_1 a_{n-1} + M_1 a_{n+1} - i\frac{\Gamma^2}{\tau_c^2}n(n-1)M_2 a_{n-2} - iM_2 a_{n+2}, \quad n = 0,1,2,\ldots . \tag{3.9}$$

Performing the Laplace transform $X_n(z) = \int_0^\infty dt\, e^{-zt}a_n(t)$, $t \geqslant 0$, we can transform the relation (3.9) into the following time-independent recurrence relation:

$$(z + A_n)X_n = B_n X_{n-1} + C_{n+1}X_{n+1} + D_n X_{n-2} + E_{n+2}X_{n+2} + \delta_{n0}P_0, \tag{3.10}$$

where the operator-valued coefficients are defined as follows:

$$A_n = n/\tau_c - i(\Gamma/\tau_c)(2n+1)M_2 - M_0,$$
$$B_n = -(\Gamma/\tau_c)nM_1, \quad C_{n+1} = M_1,$$
$$D_n = -i(\Gamma^2/\tau_c^2)n(n-1)M_2, \quad E_{n+2} = -iM_2, \tag{3.11}$$

and the initial condition as pointed out in Eq. (3.6b) is given by $a_n(0) = \delta_{n0}P_0$. Note that this recurrence relation has a tridiagonal form only if $M_1 = 0, M_2 \neq 0$ and $M_1 \neq 0, M_2 = 0$, i.e., only in the presence of a linear or a quadratic noise. In these two cases that we can solve exactly the tridiagonal recurrence relation for $X_0(z)$ in terms of an operator-continued fraction.[18] For $H = 0$, i.e., $M_2 = 0$, we obtain

$$X_0(z) = \frac{1}{z - M_0 + \tilde{K}_1(z)}P_0, \tag{3.12a}$$

with

$$\tilde{K}_1(z) = M_1 \frac{\Gamma/\tau_c}{z - M_0 + M_1 \dfrac{\Gamma/\tau_c}{z + 1/\tau_c - M_0 + \cdots}M_1}M_1. \tag{3.12b}$$

For $G = 0$, i.e., $M_1 = 0$, we obtain

$$X_0(z) = \frac{1}{z - i(\Gamma/\tau_c)M_2 - M_0 + \tilde{K}_2(z)}P_0, \tag{3.13a}$$

with

$$\tilde{K}_2(z) = 2M_2 \frac{\Gamma^2/\tau_c^2}{z - M_0 + \dfrac{2}{\tau_c} - 5i\dfrac{\Gamma}{\tau_c}M_2 + 20M_2 \dfrac{\Gamma^2/\tau_c^2}{z - M_0 + 4/\tau_0 - 9i(\Gamma/\tau_c)M_2 + \cdots}M_2}M_2. \tag{3.13b}$$

In these two cases we can write the following generalized Fokker–Planck equation after performing the inverse Laplace transform of Eqs. (3.12a) and (3.13a). We obtain for the two cases the following equations:

$$\frac{\partial}{\partial t}p = M_0 p - \int_0^t ds\, K_1(t-s)\,p(s) \quad \text{for } H = 0 \tag{3.14a}$$

and

$$\frac{\partial}{\partial t}p = \left(M_0 + \frac{i\Gamma}{\tau_c}M_2\right)p - \int_0^t ds\, K_2(t-s)\,p(s) \quad \text{for } G = 0, \tag{3.14b}$$

where $\int_0^\infty dt\, K_{1,2}(t)e^{-zt} = \tilde{K}_{1,2}(z)$ for $t \geqslant 0$. Note that the formal form of Eq. (3.14a) with (3.12b) has already been derived in statistical mechanics[10] and used recently for many problems in quantum optics except for the difference that F and G were linear maps of a vector ρ.[16] The derivation of the continued fraction (3.12b) for application in quantum optics has been obtained in a completely different way.[26]

In our case of Fokker–Planck equations (3.14a) and (3.14b) $K_{1,2}$ are differential operators with an infinite number of $\partial/\partial\xi$ derivatives acting on the Fokker–Planck probability distribution $p(t,\xi)$. This fact, that such a generalized Fokker–Planck equation has an infinite number of derivatives, is not a surprise at all. It is well known that for white noise higher-order derivatives should appear in the proper equation for the probability density $p(t,\xi)$. Recently it has also been shown that we can attribute an interpretation to each step in the continued fraction (3.13b).[27] If the external noise $x(t)$ is approximated by a finite sequence of independent random telegraph signals it has been shown that a finite truncation of the continued fraction (3.13b) leads to an exact solution of the Langevin equation (1.1) driven by an external random telegraph noise.[27] The continued fraction for the case of only a quadratic noise, i.e., with $G = 0$ has not been obtained in the literature before. Note that $\tilde{K}_2(z)$ has no limit for $\tau_c \to 0$. This reflects the singular character of quadratic white noise in a stochastic differential equation (1.1). For linear noise only, we check that

$$\tilde{K}_1(z) \underset{\tau_c \to 0}{\to} \Gamma M_1^2$$

as predicted by Eq. (2.5).

For both $G \neq 0$ and $H \neq 0$ the recurrence relation (3.9) is no longer tridiagonal. This reflects the various statistical correlations of the linear noise with its quadratic contribution. Because of these correlations the general solution of the stochastic equation (1.1) cannot be obtained by a simple superposition of the linear and quadratic case. Nevertheless, it has been shown[18] that the recurrence relation of the type given by Eq. (3.10) can be transformed into a tridiagonal vector recurrence relation of the following form:

$$(z - \mathbb{Q}_n)\mathbb{X}_n = \mathbb{P}_{n+1}\mathbb{X}_{n+1} + \mathbb{R}_n\mathbb{X}_{n-1} + \delta_{n0}\mathbb{P}_0, \tag{3.15}$$

where we have introduced vectors

$$\mathbb{X}_n = \begin{pmatrix} X_n \\ X_{n+1} \end{pmatrix} \quad \text{and} \quad \mathbb{P}_0 = \begin{pmatrix} P_0 \\ 0 \end{pmatrix} \tag{3.16}$$

and the following operator-valued matrices:

$$\mathbb{Q}_n = -\begin{Vmatrix} A_{2n} & C_{2n+1} \\ B_{2n+1} & A_{2n+1} \end{Vmatrix}, \quad \mathbb{P}_n = \begin{Vmatrix} E_{2n} & 0 \\ C_{2n} & E_{2n+1} \end{Vmatrix}, \quad \mathbb{R}_n = \begin{Vmatrix} D_{2n} & B_{2n} \\ 0 & D_{2n+1} \end{Vmatrix}. \tag{3.17}$$

Now following the tridiagonal recurrence relation solution we can write the exact solution of our vector tridiagonal recurrence relation (3.15):

$$\mathbb{X}_0(z) = \frac{1}{z - \mathbb{Q}_0 - \widetilde{\mathbb{K}}(z)}\,\mathbb{P}_0, \tag{3.18a}$$

where

$$\widetilde{\mathbb{K}}(z) = \mathbb{P}_1 \frac{1}{z - \mathbb{Q}_1 - \mathbb{P}_2 \dfrac{}{z - \mathbb{Q}_2 - \mathbb{P}_3 \cdots}\mathbb{R}_2}\,\mathbb{R}_1. \tag{3.18b}$$

We can convert this operator-continued fraction into the following integro-differential equation:

$$\frac{\partial}{\partial t}\,\mathfrak{a}_0(t,\xi) = \mathbb{Q}_0\mathfrak{a}_0(t,\xi) + \int_0^t ds\,\mathbb{K}(t-s)\mathfrak{a}_0(s,\xi), \tag{3.19a}$$

where

$$\mathfrak{a}_0(t,\xi) = \begin{pmatrix} p(t,\xi) \\ a_1(t,\xi) \end{pmatrix} \tag{3.19b}$$

and $\mathbb{K}(t)$ is the inverse Laplace transform of the operator-continued fraction (3.18b). Because of the vector character of the tridiagonal recurrence relation our Fokker–Planck equation (3.19) also has a vector form with the auxiliary function $a_1(t,\xi)$. That the additional term $a_1(t,\xi)$ should appear in Eq. (3.19) is clear from the form of the recurrence relation (3.9). If $H = 0$ a_0 couples to a_1, a_1 couples to a_0 and a_2, and so on. In this case a_1 can be eliminated from Eq. (3.19), leading to the continued fraction (3.12). If $G = 0$ a_0 couples to a_2, a_2 couples to a_0 and a_4, and no a_1 coefficient is involved. For both $H \neq 0$ and $G \neq 0$ a_1 mixes the two stochastic processes showing that they cannot be superimposed independently. The appearance of the a_1 coefficient in Eq. (3.19) indicates the correlations between the $x(t)$ and $x^2(t)$ contributions of the noise in the solution. By a Laplace transform we can solve again our Eq. (3.19) exactly:

$$X_0(z) = \frac{1}{z - M_0 - i(\Gamma/\tau_c)M_2 - \widetilde{\mathfrak{R}}(z)}\,P_0, \tag{3.20a}$$

where the operator $\widetilde{\mathfrak{R}}(z)$ is given by

$$\widetilde{\mathfrak{R}}(z) = -\mathbb{K}_{11}(z) - (\mathbb{K}_{12} - M_1)\frac{1}{z - M_0 + 1/\tau_c - i3(\Gamma/\tau_c)M_2} \times \left(\mathbb{K}_{21} + \frac{\Gamma}{\tau_c}M_1\right) \tag{3.20b}$$

with the $\mathbb{K}_{ij}(z)$ denoting the matrix elements of the operator-valued matrix $\mathbb{K}(z)$ given by Eq. (3.18b). Again we can transform this Laplace solution into an exact integro-differential equation for the probability distribution:

$$\frac{\partial}{\partial t}\,p(t,\xi) = \left(M_0 + \frac{i\Gamma}{\tau_c}M_1\right)p(t,\xi) + \int_0^t ds\,\mathfrak{R}(t-s)\,p(s,\xi), \tag{3.21}$$

where $\int_0^\infty dt\,e^{-zt}\mathfrak{R}(t) = \widetilde{\mathfrak{R}}(z)$. This formal but exact equation is the main result of this paper. We have obtained an exact generalized Fokker–Planck equation for the probability distribution associated with the Langevin equation (1.1).

IV. STANDARD FOKKER–PLANCK EQUATION

In this paragraph we shall derive an approximate standard-form Fokker–Planck equation from our main equation (3.21). By a standard-form Fokker–Planck equation we mean a second-order differential equation with a drift and diffusion term:

$$\frac{\partial p(t,\xi)}{\partial t} = -\frac{\partial}{\partial \xi}\left(\mathscr{A}(\xi) + \mathscr{B}(\xi)\frac{\partial}{\partial \xi}\right)p(t,\xi). \tag{4.1}$$

Such a form enables us to study, for example, the steady-state properties of the distribution probability $p(t,\xi)$ which has the following well-known form:

$$p_{st}(\xi) = N\exp\left(-\int^{\xi}\frac{\mathscr{A}(\xi')}{\mathscr{B}(\xi')}\,d\xi'\right), \tag{4.2}$$

where N is a normalization constant.

From our exact equation for $p(t,\xi)$ we now derive a differential equation (4.1) which can be written in the following form:

$$\frac{\partial}{\partial t}\,p(t,\xi) = \hat{\Gamma}p(t,\xi), \tag{4.3}$$

where $\hat{\Gamma}$ is the Fokker–Planck operator. We have in our theory two independent physical parameters τ_c and Γ/τ_c. In the limit of small fluctuations, i.e., if $\Gamma/\tau_c < 1$ and for time $t > \tau_c$ we will look for the Fokker–Planck operator in the following power series form:

$$\hat{\Gamma} = \sum_{n=0}^{\infty}\left(\frac{\Gamma}{\tau_c}\right)^n\hat{\Gamma}^{(n)}, \tag{4.4}$$

where the condition $t > \tau_c$ allows us to forget about memory effects so that $\hat{\Gamma}^{(n)}$ are differential operators rather than integro–differential ones. This method can be implemented in a systematical way following closely the method of adiabatic elimination of additional degrees of freedom from the exact equation (3.19a) or, what is equivalent, by a systematic power series expansion of the operator $\widetilde{\mathfrak{R}}(z)$ in Eq. (3.21) for $t > \tau_c$ and $\Gamma/\tau_c < 1$. From Eq. (3.21) it is clear that the only contribution of zero order in Γ/τ_c is given by the standard drift term:

$$\hat{\Gamma}^{(0)} = M_0. \tag{4.5}$$

The contribution of the order Γ/τ_c can be obtained from Eq. (3.15) performing a proper elimination of $\mathbb{Q}_n$ up to the order linear in Γ/τ_c. It is clear that in this case only $\mathbb{Q}_0$ and $\mathbb{Q}_1$ will contribute and this is equivalent to a truncation of the continued fraction to the first fraction. Performing this elimination we obtain from Eqs. (3.15) and (3.19)

$$\frac{\partial}{\partial t}\,\mathfrak{a}_0(t,\xi) = \mathbb{A}_0\mathfrak{a}_0(t,\xi) + \mathbb{P}_1\int_0^t ds\,e^{\mathbb{A}_1^{(0)}(t-s)}\mathbb{R}_1 e^{\mathbb{A}_0(s-t)}\mathfrak{a}_0(t,\xi), \tag{4.6}$$

where the subscript (0) in $\mathbb{A}_1^{(0)}$ indicates that only terms of the order $(\Gamma/\tau_c)^0$ should be retained. Because $\mathbb{A}_1^{(0)} = \mathbb{A}_0 - (2/\tau_c)I$, we calculate that for $t > \tau_c$

$$\int_0^t ds\, e^{\mathbb{A}_0(t-s)}\mathbb{R}_1 e^{\mathbb{A}_0(s-t)-(2/\tau_c)(t-s)}$$

$$=\frac{1-e^{-t(\mathbb{L}+2/\tau_c)}}{2/\tau_c+\mathbb{L}}\mathbb{R}_1\simeq\frac{1}{2/\tau_c+\mathbb{L}}\mathbb{R}_1, \tag{4.7a}$$

where $\mathbb{L}\cdot=[\cdot,\mathbb{A}_0]$ is a vector Liouville operator acting in Eq. (4.7a) on $\mathbb{R}_1$. This approximation leads to

$$\hat{\Gamma}^{(1)}=iM_2-M_1\frac{1}{1/\tau_c+L}M_1, \tag{4.7b}$$

where the scalar Liouville operator $L\cdot=[\cdot,A_0]$ now acts on the operator M_1. Up to the order Γ/τ_c the Fokker–Planck equation takes the following form:

$$\frac{\partial}{\partial t}p=\left(M_0+\frac{i\Gamma}{\tau_c}M_2-\frac{\Gamma}{\tau_c}M_1\frac{1}{1/\tau_c+L}M_1\right)p. \tag{4.8}$$

Note that the quadratic noise has contributed only to the drift term and the diffusion part has only contributions due to the linear noise. The form of this diffusion part has been obtained first in Ref. 20 and then discussed in detail in Refs. 4 and 8. What usually is done in most of the physical applications is really a new approximation which follows by expending the inverse Liouville operator as a Neumann series:

$$\frac{1}{1/\tau_c+L}=\sum_{n=0}^{\infty}(-1)^n\tau_c^{n+1}L^n \tag{4.9}$$

and than neglecting all terms proportional to a given power of τ_c. This approximation is consistent with $t>\tau_c$ and has an additional requirement that τ_c is small. One can show[28] that the Fokker–Planck operator from Eq. (4.7b) can be written in the following form:

$$\frac{\partial}{\partial t}p(t,\xi)=\left(-\frac{\partial}{\partial\xi}F-i\frac{\Gamma}{\tau_c}\frac{\partial}{\partial\xi}H-\frac{\partial}{\partial\xi}G\frac{\partial}{\partial\xi}\mathfrak{U}\right)p(t,\xi), \tag{4.10}$$

where the function $\mathfrak{U}(\xi)$ is given by the following series:

$$\mathfrak{U}(\xi)=\Gamma\sum_{n=0}^{\infty}(-1)^n\tau_c^n u_n(\xi), \tag{4.11a}$$

where u_n are solutions of the following recurrence relation:

$$u_n=F\frac{\partial}{\partial\xi}u_{n-1}-u_{n-1}\frac{\partial}{\partial\xi}F,\quad n=1,\cdots\quad\text{with}\quad u_0=G. \tag{4.11b}$$

This leads to the following formula for the drift and diffusion functions in Eq. (4.1):

$$\mathscr{A}(\xi)=F(\xi)+\frac{i\Gamma}{\tau_c}H(\xi)+G(\xi)\mathfrak{U}'(\xi), \tag{4.12a}$$

$$\mathscr{B}(\xi)=G(\xi)\mathfrak{U}(\xi). \tag{4.12b}$$

After some lengthy algebra we obtain the following contribution to the Fokker–Planck operator up to order $(\Gamma/\tau_c)^2$:

$$\begin{aligned}\hat{\Gamma}^{(2)}=&2M_2\frac{1}{2/\tau_c+L}M_2+M_1\frac{1}{1/\tau_c+L}\\&\times(M_1-i3M_2)\frac{1}{1/\tau_c+L}\\&\times M_1-i2M_1\frac{1}{1/\tau_c+L}\\&\times M_1\frac{1}{2/\tau_c+L}M_2-i2M_2\frac{1}{2/\tau_c+L}\\&\times M_1\frac{1}{1/\tau_c+L}M_1+2M_1\frac{1}{1/\tau_c+L}\\&M_1\frac{1}{2/\tau_c+L}M_1\frac{1}{1/\tau_c+L}M_1.\end{aligned} \tag{4.13}$$

Note that in this order we now have a contribution to the diffusion term coming from the quadratic noise itself and cross contributions coming from "contractions" of $x(t)$ with $x^2(t)$. The operator $\hat{\Gamma}^{(2)}$ contains terms proportional to $\partial^4/\partial\xi^4$ and in this sense $\hat{\Gamma}^{(2)}$ is giving contributions beyond the usual Fokker–Planck description. We can truncate $\hat{\Gamma}^{(2)}$ to an equation of second order but following the arguments of Ref. 29 we shall keep these terms, since our equation

$$\frac{\partial}{\partial t}p=\left(\hat{\Gamma}^{(0)}+\frac{\Gamma}{\tau_c}\hat{\Gamma}^{(1)}+\frac{\Gamma^2}{\tau_c^2}\hat{\Gamma}^{(2)}\right)p \tag{4.14a}$$

can be written in the following form:

$$\frac{\partial}{\partial t}p=-\frac{\partial}{\partial\xi}J, \tag{4.14b}$$

i.e., the evaluation of the stationary probability distribution from the condition $J=0$ with proper boundary conditions is possible in the standard way. At this point we should comment on a recent derivation of the Fokker–Planck equation with a quadratic noise given in Ref. 19. Apart from the important difference that in Ref. 19 H is chosen in such a way that M_1 commutes with M_2, the main approach of that paper has been based on a nonlinear transformation of the noise $x(t)+x^2(t)$ into a Gaussian white noise with an effective diffusion constant. It has been shown next that such a transformation leads to a Fokker–Planck equation with an effective diffusion part and a shifted drift term.[19] It is easy to check that $\hat{\Gamma}^{(0)}$ and $\hat{\Gamma}^{(1)}$ [Eqs. (4.5) and (4.7b)] and the first contribution to $\hat{\Gamma}^{(2)}$ in Eq. (4.13) specialized for H proportional to G, i.e., for the case of $[M_1,M_2]=0$ and with the Liouville operators expanded to the lowest order in τ_c, lead precisely to the result of Ref. 19. The nonlinear transformation of the noise fails to predict the mixed terms given by the other operators in Eq. (4.13). These terms come from correlations between the linear and the quadratic part of the noise in the nonlinear Langevin equation (1.1).

V. TRANSITION PROBABILITY

In order to compute correlation functions of $\rho(t)$ we need to generalize our results of the previous sections. We introduce instead of $p(t,\xi)$ the following functionally dependent probability transition[30]:

$$p[t,\xi\,|\mathscr{K}]=\left\langle\varphi(t,\xi)\exp\left(i\int_0^t d\tau\,\mathscr{K}(\tau)\rho(\tau)\right)\right\rangle, \tag{5.1}$$

where the functional dependence on an abritrary smooth function $\not{\kappa}(t)$ has been shown explicitly. We see immediately from Eq. (5.1) that

$$\int d\xi\, p[t,\xi\,|\not{\kappa}] = \left\langle \exp\left(i\int_0^t d\tau\, \not{\kappa}(\tau)\rho(\tau)\right)\right\rangle \tag{5.2}$$

is the generating functional for the dynamical variable $\rho(t)$. From the definition (5.1) we can compute arbitrary correlations of $\rho(t)$ by a functional differentiation of the generalized transition probability (5.1):

$$\langle \rho(t_1)\cdots\rho(t_n)\rangle = \int d\xi\, \frac{\delta^n}{\delta(-i\not{\kappa}(t_1))\cdots\delta(-i\not{\kappa}(t_2))} p[t,\xi\,|\not{\kappa}]|_{\not{\kappa}=0}. \tag{5.3}$$

From the definition 5.1 we can derive the proper Liouville equation for

$$\varphi\,[t,\xi\,|\not{\kappa}] = \varphi\,(t,\xi)\exp\left(i\int_0^t d\tau\, \not{\kappa}(\tau)\rho(\tau)\right). \tag{5.4}$$

After simple algebra we conclude that $\varphi[t,\xi\,|\not{\kappa}]$ satisfies the same equation as that $\varphi(t)$ from Sec. I with the only difference that[30]

$$M_0 = -\frac{\partial}{\partial\xi} F(\xi) + i\xi\not{\kappa}(t). \tag{5.5}$$

This means that we can carry out the same procedure as in the previous sections with the difference that the recurrence relation (3.9) now has time-dependent coefficients. From the mathematical point of view the solution for the transition probability is not much more complicated that the construction of the dynamical probability density from Eq. (3.9). In fact, recently an exact dynamical solution of this problem has been obtained for the case when[31]

$$F(\rho) = f_1(t)\rho, \quad G(\rho) = f_2(t)\rho, \quad H(\rho) = f_3(t)\rho \tag{5.6}$$

with $f_1(t)$, $f_2(t)$, and $f_3(t)$ arbitrary time-dependent functions. The external term $i\xi\not{\kappa}(t)$ in the drift term, in the Fokker–Planck equation, can be treated in the same way as the external current effects are calculated in quantum electrodynamics or field theory.

ACKNOWLEDGMENTS

The author would like to thank Professor H. Walther for his invitation and hospitality extended to him at the Max-Planck Institute for Quantum Optics where this work has been done and Dr. J. Cresser for helpful comments about the final form of this manuscript.

[1] H. Haken, *Synergetics-An Introduction* (Springer, Berlin, 1978).
[2] J. H. Eberly, Phys. Rev. Lett. **37**, 1387 (1976).
[3] K. Wódkiewicz, Phys. Rev. A **19**, 1686 (1979).
[4] K. Wódkiewicz, in *Spectral Line Shapes*, edited by B. Wende (de Gruyter, Berlin, 1981), p. 955.
[5] A. Schenzle and H. Brand, Phys. Rev. A **20**, 1628 (1979).
[6] A. T. Georges and P. Lambropoulos, Phys. Rev. A **20**, 991 (1979).
[7] P. Zoller, Phys. Rev. A **20**, 1019 (1979).
[8] J. J. Yeh and J. H. Eberly, Phys. Rev. A **24**, 888 (1981).
[9] P. Zoller and P. Lambropoulos, J. Phys. B **13**, 69 (1980).
[10] R. Barakat and J. Blake, Phys. Rep. **60C**, 225 (1980).
[11] H. Haken, Rev. Mod. Phys. **47**, 65 (1975).
[12] H. Haken, *Handbook der Physik* (Springer, Berlin, 1970), Vol. XXV/2c.
[13] W. H. Louisell, *Quantum Statistical Properties of Radiation* (Wiley, New York, 1973).
[14] R. Balescu, *Equilibrium and Nonequilibrium Statistical Mechanics* (Wiley, New York, 1975).
[15] S. Kai, T. Kai, M. Takata, and M. Hirakawa, J. Phys. Soc. Jpn. **47**, 1379 (1979).
[16] T. Kawakubo, A. Yanagita, and S. Kabashima, J. Phys. Soc. Jpn. **50**, 1451 (1981).
[17] F. Schlögl, Ann. Phys. (NY) **45**, 155 (1967).
[18] H. Risken and H. D. Vollmer, Z. Phys. B **39**, 339 (1980).
[19] M. San Miguel and J. M. Sancho, Z. Phys. B **43**, 361 (1981).
[20] N. G. Van Kampen, Phys. Rep. **24C**, 171 (1976).
[21] R. F. Fox, J. Math. Phys. **13**, 1196 (1972).
[22] R. F. Fox, Phys. Rep. **48C**, 179 (1978).
[23] K. Wódkiewicz, J. Math. Phys. **20**, 45 (1979).
[24] R. L. Stratonovich, *Topics in the theory of random noise* (Gordon and Breach, New York, 1963), Vol. 1.
[25] E. W. Montroll in *Lectures in Theoretical Physics*, University of Colorado, Boulder, Colorado, edited by A. O. Barut and W. E. Britten (Gordon and Breach, New York, 1967).
[26] S. N. Dixit, P. Zoller, and P. Lambropoulos, Phys. Rev. A **21**, 1289 (1980).
[27] K. Wódkiewicz, Z. Phys. B **42**, 95 (1981).
[28] M. Kuś and K. Wódkiewicz, Phys. Lett. A **90**, 23 (1982).
[29] N. G. Van Kampen, Adv. Chem. Phys. **15**, 65 (1969).
[30] K. Wódkiewicz, Phys. Lett. A **84**, 56 (1981).
[31] K. Wódkiewicz and M. S. Zubairy, Phys. Rev. A (submitted for publication).

The occupational statistics for indistinguishable trimers on a $3\times N$ lattice space[a)]

J. L. Hock, P. E. Licato, and R. B. McQuistan
Department of Physics and Laboratory for Surface Studies, University of Wisconsin—Milwaukee, Milwaukee, Wisconsin 53201

(Received 18 March 1982; accepted for publication 29 April 1982)

A 20 term recursion relation is derived that describes exactly the occupational degeneracy for trimers distributed on a $3\times N$ lattice space. The associated generating functions, expectation, normalization, dispersion, and continuous representation are also developed and discussed.

PACS numbers: 05.50. + q, 02.30. + g

I. INTRODUCTION

The purpose of the present paper is to examine the occupation statistics that arise when trimer particles (which occupy three linearly contiguous lattice sites) are distributed on a rectangular $3\times N$ lattice space. This research represents a generalization of previously published[1-4] results concerning the occupational degeneracy of dimers distributed on $1\times N$ and $2\times N$ lattice spaces.

Specifically, we seek first to determine $A[q,N]$, the multiplicity of independent arrangements that arise when q indistinguishable trimers are distributed on a $3\times N$ rectangular lattice space (see Fig. 1). Computer-generated numerical values for $A[q,N]$ are presented in Table I for several values of q and N.

We will use a set of theoretic arguments to obtain an exact recursion relationship satisfied by $A[q,N]$. This recursion relationship takes on a particularly simple, Fibonnacci-like, form when the lattice space is completely filled.

On the basis of this recursion we will develop the appropriate generating functions which we will utilize to determine the expectation, normalization, dispersion, and continuous representation for $A[q,N]$.

II. RECURSION RELATIONSHIP FOR $A[q,N]$

The diagrams shown in Fig. 2 serve to define the 12 lattice spaces, $A, B,\ldots,L$. Then, for example, $G[q,N]$ represents the set of all the possible arrangements of q trimer particles on a G_N-space, N columns long. By examining the state of occupation of the compartments in the Nth column, we can decompose the degeneracy of a particular lattice space into subsets, each of which represents the degeneracy for other kinds of lattice spaces depicted in Fig. 2. For example, consider the A_N-space.

The possible states of occupation of the Nth column of an A_N-space can be seen in Fig. 3. There are nine unique states of occupation of the Nth column. We observe that if the top (or bottom) compartment in the Nth column is occupied and the other two compartments are vacant a B_{N-1}-space is created. All other arrangements on the remaining compartments involve $q-1$ trimer particles on a B_{N-1}-space. Hence $2B[q-1,N-1]$ represents the set of all arrangements on an A_N-space when the Nth column is occupied as shown in Fig. 3. The factor of two arises because the trimer particle can be placed in either the top or bottom compartment to create a B_{N-1}-space. By considering the seven possible subsets shown in Fig. 3 we may write for the decomposition of $A[q,N]$:

$$\begin{aligned}A[q,N] = {} & A[q,N-1] + 2B[q-1,N-1] \\ & + C[q-1,N-1] + A[q-1,N-1] \\ & + 2D[q-2,N-1] + E[q-2,N-1] \\ & + A[q-3,N-3]. \end{aligned} \tag{1}$$

A similar examination of the state of occupation of the other lattice spaces yields

$$\begin{aligned}B[q,N] = {} & F[q,N-1] + G[q-1,N-1] \\ & + H[q-1,N-1] + I[q-2,N-2], \end{aligned} \tag{2}$$

$$\begin{aligned}C[q,N] = {} & J[q,N-1] + 2K[q-1,N-1] \\ & + L[q-2,N-2], \end{aligned} \tag{3}$$

$$D[q,N] = I[q,N-1] + F[q-1,N-2], \tag{4}$$

$$E[q,N] = L[q,N-1] + J[q-1,N-2], \tag{5}$$

$$\begin{aligned}F[q,n] = {} & A[q,N-1] + B[q-1,N-1] \\ & + C[q-1,N-1] + D[q-2,N-1], \end{aligned} \tag{6}$$

$$G[q,N] = J[q,N-1] + K[q-1,N-1], \tag{7}$$

$$H[q,N] = F[q,N-1] + G[q-1,N-1], \tag{8}$$

$$I[q,N] = A[q,N-1] + B[q-1,N-1], \tag{9}$$

$$\begin{aligned}J[q,N] = {} & A[q,N-1] + 2B[q-1,N-1] \\ & + E[q-2,N-1], \end{aligned} \tag{10}$$

$$K[q,N] = F[q,N-1] + H[q-1,N-1], \tag{11}$$

$$L[q,N] = A[q,N-1] + C[q-1,N-1]. \tag{12}$$

Thus, we see in Eqs. (1)–(12) that there is closure in the sense that each of the 12 lattice spaces can be decomposed into subsets which are also members of the 12 lattice spaces.

These 12 equations (1)–(12) can be solved for $A[q,N]$ in terms of the A's only yielding the following recursion rela-

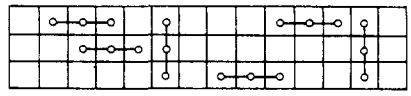


FIG. 1. One of the 1,247,559 independent arrangements that can arise when six trimers are distributed on a 3×14 lattice space.

[a)] This work was supported in part by the Air Force Office of Scientific Research under Grant AFOSR 81-0192.

 0022-2488/82/112185-05$02.50 © 1982 American Institute of Physics

TABLE I. Computer-generated numerical values for $A[q,N]$ for several values of q and N.

N \ q	0	1	2	3	4	5	6	7	8	9	10	Sum of rows
1	1											1
1	1	1										2
2	1	2	1									4
3	1	6	6	2								15
4	1	10	24	18	3							56
5	1	14	55	76	32	4						182
6	1	18	102	238	204	60	6					619
7	1	22	165	535	793	510	123	9				2 158
8	1	26	244	1058	2 278	2 420	1 174	212	13			7 426
9	1	30	339	1861	5 385	8 343	6 675	2 496	360	19		25 509
10	1	34	450	3008	11 093	23 214	27 256	17 056	5130	634	28	87 904

tion:

$$\begin{aligned} A[q,N] = {} & A[q,N-1] + 2A[q-1,N-1] \\ & + A[q-1,N-2] - A[q-2,N-2] \\ & + A[q-1,N-3] + 3A[q-2,N-3] \\ & + 3A[q-3,N-3] + A[q-2,N-4] \\ & - 5A[q-4,N-4] - A[q-3,N-5] \\ & - 6A[q-4,N-5] + 2A[q-5,N-5] \\ & - A[q-4,N-6] - 3A[q-5,N-6] \\ & - 3A[q-6,N-6] + 4A[q-7,N-7] \\ & - 2A[q-7,N-8] - A[q-8,N-8] \\ & + A[q-9,N-9] - A[q-10,N-10]. \end{aligned} \tag{13}$$

With appropriate initial conditions Eq. (13) describes exactly the occupational degeneracy for an A_N-space, i.e., for a $3\times N$ lattice space.

We observe that the analogous recursion relationships for the 11 other lattice spaces, $B_N,\ldots,L_N$, all have exactly the same form as Eq. (13). Thus, Eq. (13) is a characteristic of a $3\times N$ lattice space populated with trimers regardless of the configuration of the end compartments. It is the initial conditions only that are the source of the differences among $A[q,N]$, $B[q,N],\ldots,L[q,N]$. If the initial conditions for two lattice spaces are the same, it follows that the occupational degeneracy for them is identical. Thus, of the 12 spaces shown in Fig. 2, only six of them are unique, and we see that

$$B[q,N] = C[q,N], \tag{14}$$

$$D[q,N] = E[q,N], \tag{15}$$

$$F[q,N] = J[q,N], \tag{16}$$

$$G[q,N] = H[q,N] = K[q,N], \tag{17}$$

$$I[q,N] = L[q,N]. \tag{18}$$

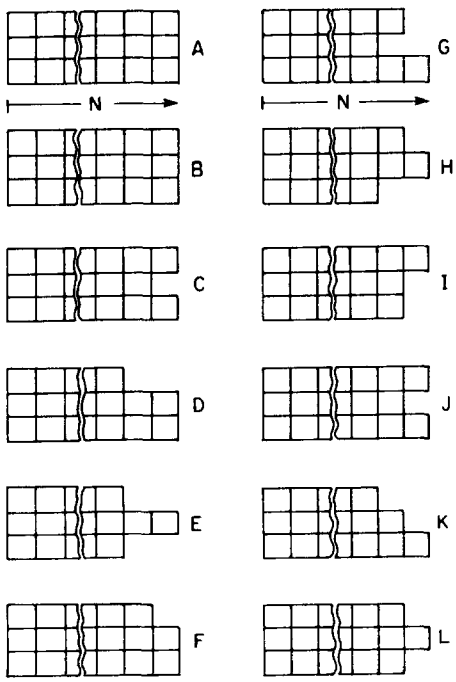


FIG. 2. These 12 drawings serve to define the 12 lattice spaces used to develop a recursion relationship satisfied by $A[q,N]$.

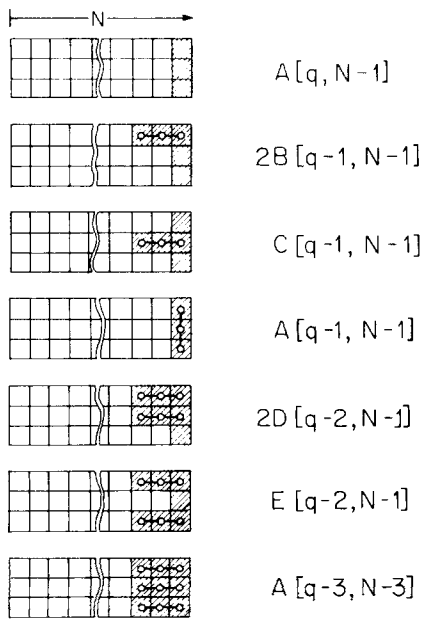


FIG. 3. These seven drawings show the decomposition of $A[q,N]$ into subsets involving one or more of the lattice spaces shown in Fig. 2.

Recognition of these relationships greatly simplifies the calculation necessary to obtain Eq. (13) from Eqs. (1)–(12).

We note here that if $q > N$, no arrangements are possible. However, if $q = n$, i.e., if the space is completely filled, Eq. (13) reduces to

$$\begin{aligned} A^*[q,N] \equiv & A[N] \\ = & 2A[N-1] - A[N-2] \\ & + 3A[N-3] - 5A[N-4] \\ & + 2A[N-5] - 3A[N-6] \\ & + 4A[N-7] - A[N-8] \\ & + A[N-9] - A[N-10]. \end{aligned} \tag{19}$$

Equation (19), however, is not in its simplest form. By appropriately reindexing, adding and subtracting multiples of

$$A[N] = A[N-1] + A[N-3], \tag{20}$$

we can obtain Eq. (19) (see the saturated diagonal in Table I). If we assume that[5]

$$A[N] = cr^N, \tag{21}$$

where c is a constant, Eq. (20) yields

$$r^3 - r^2 - 1 = 0.$$

Upon imposing the initial conditions,

$$A[0] = A[1] = A[2] = 1, \tag{22}$$

(see Table I) we obtain

$$A[N] = c_1 r_1^N + c_2 r_2^N + c_3 r_3^N, \tag{23}$$

where

$$r_1 = \tfrac{1}{3} + D_+ = 1.465571233, \tag{24}$$

$$\begin{aligned} r_2 &= \frac{1}{3} - \frac{1}{2}D_+ + \frac{i\sqrt{3}}{2}D_- \\ &= -0.232785616 + i[0.792551993], \end{aligned} \tag{25}$$

$$\begin{aligned} r_3 &= \frac{1}{3} - \frac{1}{2}D_+ - \frac{i\sqrt{3}}{2}D_- \\ &= r_2^* = -0.232785616 - i[0.792551993], \end{aligned} \tag{26}$$

where

$$\begin{aligned} D_\pm = & \left[\frac{1}{2}\left(\frac{29}{27}\right)\right]^{1/3}\left\{\left[1+\left(1-\frac{2^2}{29^2}\right)^{1/2}\right]^{1/3}\right. \\ & \left. \pm\left[1-\left(1-\frac{2^2}{29^2}\right)^{1/2}\right]^{1/3}\right\}, \end{aligned} \tag{27}$$

and where the initial conditions yield

$$c_1 = 0.611491991,$$

$$c_2 = 0.194254004 - i[0.122549691],$$

$$c_3 = c_2^*.$$

As $N \to \infty$, the real root dominates so that Eq. (23) becomes

$$A[N] = c_1 r_1^N. \tag{28}$$

At $N = 10$ Eq. (28) yields $A[10] = 27.96$. This compares with the actual value of 28.

III. GENERATING FUNCTIONS

We form the polynomials

$$f_N(x) \equiv \sum_{q=0}^{N} A[q,N]x^q. \tag{29}$$

Utilizing Eq. (13), Eq. (29) yields

$$\begin{aligned} f_{N+10}(x) = & [1+2x]f_{N+9}(x) + x[1+x]f_{N+8}(x) \\ & + x[1+3x+3x^2]f_{N+7}(x) \\ & + x^2[1-5x^2]f_{N+6}(x) \\ & - x^3[1+6x-2x^2]f_{N+5}(x) \\ & - x^4[1+3x+3x^2]f_{N+4}(x) \\ & + 4x^7 f_{N+3}(x) - x^7[2+x]f_{N+2}(x) \\ & + x^9 f_{N+1}(x) - x^{10}f_N(x). \end{aligned} \tag{30}$$

With the appropriate initial conditions gleaned from Table I, Eq. (30) will generate the numerical values for $A[q,N]$ contained in Table I.

To determine $G(x,y)$, the bivariant generating function, defined by

$$G(x,y) \equiv \sum_{N=0}^{\infty} f_N(x)\, y^N, \tag{31}$$

we impose the initial conditions on $f_N(x)$ and obtain

$$\begin{aligned} G(x,y) = & H(x,y)[1 - y(1+2x) \\ & - y^2x(1-x) - y^3x(1+3x \\ & + 3x^2) - y^4x(1-5x^2) \\ & + y^5x^3(1+6x-2x^2) \\ & + y^6(1+3x+3x^2) - 4y^7x^7 \\ & + y^8x^7(2+x) - y^9x^9 + y^{10}x^{10}]^{-1} \end{aligned} \tag{32}$$

where $H(x,y)$ is a polynomial of ninth degree in x and y.

IV. NORMALIZATION

In this section we seek an expression for the calculation of Δ_N, the normalization of the statistics appropriate to Eq. (13), for large values of N. Δ_N is defined by

$$\Delta_N \equiv \sum_{q=0}^{N} A[q,N]. \tag{33}$$

Utilizing Eq. (13) we obtain

$$\begin{aligned} \Delta_{N+10} = & 3\Delta_{N+9} + 7\Delta_{N+7} - 4\Delta_{N+6} \\ & - 5\Delta_{N+5} - 7\Delta_{N+4} + 4\Delta_{N+3} \\ & - 3\Delta_{N+2} + \Delta_{N+1} - \Delta_N \end{aligned} \tag{34}$$

[see Eq. (30) with $x = 1$]. The initial conditions on Δ_N can be seen from Table I and yield a generating function for the normalization:

$$\begin{aligned} G[1,y] &= G(y) = \sum_{N=0}^{\infty} \Delta_N y^N \\ &= \frac{1 - y - 2y^2 - 4y^3 - y^4 - y^5 + y^6 - y^7}{1 - 3y - 7y^3 + 4y^4 + 5y^5 + 7y^6 - 4y^7 + 3y^8 - y^9 + y^{10}} \\ &= \sum_{j=1}^{10} \frac{k_j}{1 - R_j y}, \end{aligned} \tag{35}$$

where the R_j's are the reciprocals of the roots of the denominator, i.e., the R_j's are the roots of

$$y^{10} - 3y^9 - 7y^7 + 4y^6 + 5y^5 - 7y^4 - 4y^3 + 3y^2 - y + 1 = 0. \tag{36}$$

Thus

$$G(y) = \sum_{j=1}^{10} k_j \sum_{N=0}^{\infty} (R_j y)^N = \sum_{N=0}^{\infty} \left\{ \sum_{j=0}^{10} k_j R_j^N \right\} y^N \tag{37}$$

so that, in comparison with Eqs. (29) and (31), we have

$$\Delta_N = \sum_{j=1}^{10} k_j R_j^N. \tag{38}$$

To determine the k_j's we let $y \to R_j^{-1}$, then of the terms in Eq. (35) only the jth term is important. Thus

$$\lim_{y \to R_j^{-1}} \left\{ G(y) - \frac{k_j}{1 - R_j y} \right\} = 0. \tag{39}$$

Applying L'Hospital's rule, we obtain

$$k_j = \frac{1 - R_j - 2R_j^2 - 4R_j^3 + R_j^4 - R_j^5 + R_j^6 - R_j^7}{R_j[3 + 21R_j^2 - 16R_j^3 - 25R_j^4 - 42R_j^5 + 28R_j^6 - 24R_j^7 + 9R_j^8 - 10R_j^9]}. \tag{40}$$

As $N \to \infty$ only one of the terms of the sum in Eq. (38) is significant, the term associated with the dominant root of Eq. (36). In this case R_1, the dominant root, is $R_1 = 3.44139062$, so that $k_1 = 0.374354638$ and thus

$$\lim_{N \to \infty} \Delta_N = 0.374354638\ [3.44139062]^N. \tag{41}$$

For $N = 10$ Eq. (41) yields 87,221.4 while from Table I we see that $\Delta_{10} = 87904$, an error of 0.78%.

V. THE EXPECTATION, DISPERSION, AND CONTINUOUS REPRESENTATION OF $A[q,N]$

In this section we begin by calculating $\langle \theta \rangle_N$ the expectation value of the coverage:

$$\langle \theta \rangle_N \equiv \frac{\langle q \rangle_N}{N}, \tag{42}$$

where

$$\langle q \rangle_N \equiv \sum_{q=0}^{N} qA\,[q,N] \Big/ \sum_{q=0}^{N} A\,[q,N] \tag{43}$$

or

$$\langle q \rangle_N \Delta_N = \sum_{q=0}^{N} qA\,[q,N]. \tag{44}$$

Utilizing Eq. (13) and assuming for large values of N that

$$\langle \theta \rangle_N = \langle \theta \rangle_{N-1} = \cdots = \langle \theta \rangle_\infty, \tag{45}$$

we obtain

$$\begin{aligned} \langle \theta \rangle_\infty = &\{2R_1^9 - R_1^8 + 16R_1^7 - 18R_1^6 - 17R_1^5 - 37R_1^4 \\ &+ 28R_1^3 - 22R_1^2 + 9R_1 - 10\} \Big/ \\ &\{3R_1^9 + 21R_1^7 - 16R_1^6 \\ &- 25R_1^5 - 42R_1^4 + 28R_1^3 \\ &- 24R_1^2 + 9R_1 - 10\} \\ &= 0.589731995. \end{aligned} \tag{46}$$

Thus, the maximum number of arrangements occurs when the lattice space is approximately 59% filled. Figure 4 shows $A\,[q,12]$ as a function of q. In this particular case, the maximum occurs at $q = 7$ or $\langle \theta \rangle_{12} \approx 0.58$.

$\langle \theta^2 \rangle_N$, the dispersion in θ, defined by

$$\langle \theta^2 \rangle_N = \frac{\langle q^2 \rangle_N}{N^2} = \frac{1}{N^2 \Delta_N} \sum_{q=0}^{N} q^2 A\,[q,N], \tag{47}$$

is obtained from Eq. (13) [under the assumption in Eq. (45)]. Then

$$\sigma_N \equiv [\langle \theta^2 \rangle_N - \langle \theta \rangle_N^2]^{1/2} = 0.380110326[N]^{-1/2}, \tag{48}$$

so that for large values of N, $A\,[q,N]$ can be represented as a

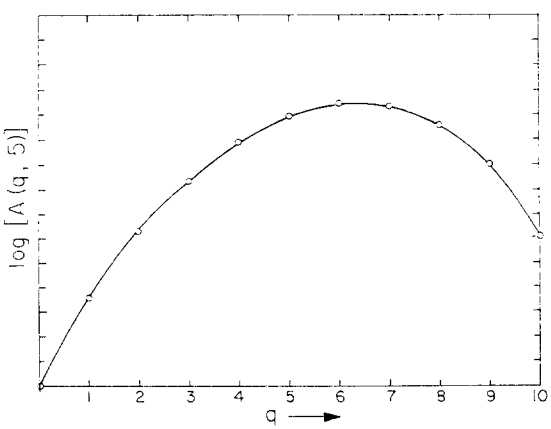


FIG. 4. ln $A\,[q,12]$ as a function of q. The distribution has a maximum at $q = 7$ or $\theta_{max} = 0.58$.

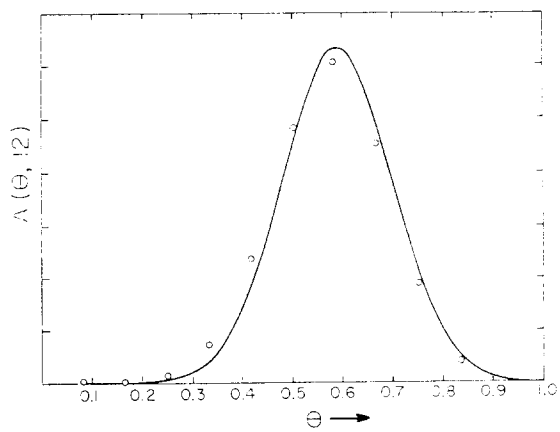


FIG. 5. $A\,[\theta,12]$ as computed exactly from the recursion relation, Eq. (13) (dots), and $\lim_{N \to \infty} A\,[\theta,N]$ as determined by Eq. (49) are plotted as functions of θ.

Gaussian distribution

$$A\,[\theta,N\,] = A_{\max}\exp\left\{\frac{-\,[\theta-\langle\theta\rangle_{\infty}\,]^2}{2\sigma_N^2}\right\}, \tag{49}$$

where

$$A_{\max} = 0.392901436[3.444139062]^N\,[N\,]^{1/2} \tag{50}$$

as determined by the normalization.

Figure 5 shows a comparison of $A\,[\theta,N\,]$ as calculated according to Eq. (13) [see Table I] for $N = 12$ with $A\,[\theta,N\,]$ as given by Eq. (49).

VI. CONCLUSIONS

We have considered the statistics arising when indistinguishable trimers are distributed on a $3\times N$ lattice space. Specifically, we have determined a recursion relationship that yields exactly the occupational degeneracy for this situation.

On the basis of this recursion relationship we have calculated the associated generating functions, normalization, expectation, and dispersion of the distribution. A continuous representation of the degeneracy is also presented.

ACKNOWLEDGMENTS

The authors wish to express their appreciation to Professor D. Snider for many helpful discussions. We also wish to thank Mr. R. Ristic for drawing the figures.

[1] D. Lichtman and R. B. McQuistan, J. Math. Phys. **8**, 2441 (1967).
[2] R. B. McQuistan, Nuovo Cimento B **58**, 86 (1968).
[3] R. B. McQuistan and S. J. Lichtman, J. Math. Phys. **11**, 3095 (1970).
[4] R. B. McQuistan, J. Math. Phys. **13**, 242 (1972).
[5] A. J. Pares and K. H. Simmons, Am. J. Phys. **48**, 159 (1980).

On the convergence of reflectionless approximants to confining potentials

I. Sabba Stefanescu
Institut für Theoretische Kernphysik der Universität Karlsruhe, 75 Karlsruhe 1, West Germany

(Received 5 June 1981; accepted for publication 1 August 1981)

We discuss the problem of the convergence of the reflectionless approximants $V_N(x)$, proposed by Quigg, Rosner, and Thacker, to the true confining potential $V(x)$, both for "wrong" $[u'(0) = 0]$ and "correct" $[u(0) = 0]$ boundary conditions. We show that, for any function $\phi(x)$ with continuous derivative and compact support contained in $(0, \infty)$, $\int_0^\infty \phi(x)V_N(x)\,dx \to \int_0^\infty \phi(x)V(x)\,dx$ as $N \to \infty$, if $V(x)$ is appropriately restricted. Namely, for "wrong" boundary conditions, $V(x)$ should be such that the difference $\sigma(E)$ between its spectral measure and the free one obeys $|\sigma(E)| < \text{const} \times E^{-\epsilon}$, $\epsilon > 0$; for "correct" boundary conditions, the spectral measure of $V(x)$ should be sufficiently close to a semiclassical estimate and, in this approximation, $V(x) < Cx^{2-\epsilon}$.

PACS numbers: 12.40.Qq

I. INTRODUCTION

In a recent series of papers,[1-5] Quigg, Rosner, Schonfeld, Thacker, and collaborators proposed, investigated, and used effectively a method to construct approximants to the confining quark-antiquark potential for heavy quarks, from knowledge of a finite number of S wave energy levels E_n and of the values $\Psi_n(0)$ of the corresponding wavefunctions at the origin. The latter are obtained from the leptonic decay width of the state by means of the Weisskopf–van Royen formula.[6] Using the reduced wavefunctions $u_n(x) = x\Psi_n(x)$, their approximant consists of a one-dimensional symmetric reflectionless potential which has prescribed odd parity bound states and values of $u_n'(0)$. Their method contains an arbitrary parameter—called Λ in the following—determining the onset of the continuum spectrum of the approximant.

More precisely, given Λ and the positions of a finite number of both even- and odd-parity energy levels, one can determine a unique symmetric reflectionless potential which supports them, by means of a linear set of equations, described in Refs. 1 and 4. The positions of the *a priori* unknown even-parity levels are fixed in such a potential from knowledge of the odd-parity ones and the derivative at the origin of the corresponding wavefunctions. This computation amounts to the determination of the zeros of a certain rational function, whose poles, residua, and asymptotic behavior are known (see Sec. IV). It is a well-posed mathematical problem[7,8] to find out whether the sequence of approximants $V_N(x)$ constructed in this way from an increasing number of energy levels converges, in some sense, to the true potential $V(x)$.

Since the relation of the spectral function to the potential is most easily described by means of the Gel'fand–Levitan equations (Refs. 9 and 10), a natural way to proceed is to check whether the kernels of the integral equations (for the half-line) associated to the soliton potentials $V_N(x)$ and to the true confining potentials $V(x)$ converge in the uniform norm towards each other, uniformly for all x, $0 \leqslant x \leqslant x_0$, x_0 finite, and whether the free terms also converge uniformly towards each other as $N \to \infty$. This will ensure that the potentials tend to each other, in some sense, as $N \to \infty$.

However, the Gel'fand–Levitan equations in the usual form are established for confining potentials only in the case of "wrong" boundary conditions at the origin,[10] i.e., $u_n'(0) = 0$. Therefore, an easier problem appears to be that in which the soliton approximants are constructed starting from the values of the even-parity levels and from the magnitude of the corresponding wavefunction at the origin, rather than from the data on the odd-parity levels.

In this (unphysical) setting, we show that, if the difference $\sigma(E)$ between the spectral measure of the confining potential and the free measure goes to zero quicker than $1/E^\epsilon$, for some $\epsilon > 0$, then the sequence of approximants $V_N(x)$ does converge to the true potential $V(x)$ in the sense that, for any function $\phi(x)$ with continuous derivative and compact support contained in $(0, \infty)$,

$$\lim_{N\to\infty} \int_0^\infty \phi(x)V_N(x)\,dx = \int_0^\infty \phi(x)V(x)\,dx. \tag{1.1}$$

The convergence problem with physical boundary conditions is then studied following Ref. 7, by reducing it to the previous case of "wrong" boundary conditions. However, the study of the difference of the kernels becomes more complicated, and the author is able to prove a statement like (1.1) only in a restricted class of potentials whose spectral function stays uniformly sufficiently close to its semiclassical estimate (see Sec. IV) and which, roughly speaking, increase less fast than x^2.

The paper is organized as follows:

In Sec. II, we discuss the approximation problem in general terms and state sufficient conditions for convergence to be achieved. The problem will then be to check that these criteria are satisfied by the soliton approximants.

The reflectionless approximants are defined naturally on the whole line $-\infty < x < \infty$, whereas the Gel'fand–Levitan equations require the spectral function of $V_N(x)$, regarded as a radially symmetrical potential. So, the next problem is the construction of the Jost functions and spectral measures for a symmetric superposition of solitons. The results are simple and are derived in Secs. III and IV.

In Sec. V, we prove the convergence of the approximants constructed from "wrong" boundary conditions, in the sense of Eq. (1.1).

Sections VI and VII contain a proof of the convergence of reflectionless approximants constructed from physical

 0022-2488/82/112190-15$02.50 © 1982 American Institute of Physics

boundary conditions, for a restricted class of potentials. Conclusions are presented in Sec. VIII.

II. CRITERIA FOR CONVERGENCE

We consider the S-wave Schrödinger equation for the reduced wavefunction $u(x) = x\Psi(x)$ and a confining potential $V(x)$ [a rescaling of x has been performed, so that $(2\mu)^{1/2}/\hbar = 1$, μ = reduced mass of the quark],

$$u''(x) + (E - V(x))u(x) = 0, \tag{2.1}$$

with two kinds of boundary conditions: $u'(0) = 0$ and $u(0) = 0$ [and $u(x) \in L^2(0, \infty)$]. The eigenvalues and eigenfunctions corresponding to these conditions are denoted by E_n^e, $u_{n,e}(x)$ and by E_n^o, $u_{n,o}(x)$, respectively (with indices derived from "even" and "odd"). The eigenfunctions $u_{n,e(o)}(x)$ are normalized so that

$$\int_0^\infty (u_{n,e(o)})^2(x)\,dx = 1. \tag{2.2}$$

The spectral functions associated with these problems are defined as

$$\rho_e(E) = \sum_{E_i^e < E} (u_{i,e})^2(0), \tag{2.3}$$

$$\rho_o(E) = \sum_{E_i^o < E} (u'_{i,o})^2(0). \tag{2.4}$$

We assume $V(x)$ in (2.1) is a continuous function on $0 \leqslant x < \infty$, with a continuous derivative. It is known[9,10] that it is uniquely determined by the spectral function $\rho_e(E)$, via the Gel'fand–Levitan equation

$$F(r,u) + K(r,u) + \int_0^r ds\, K(r,s)F(s,u) = 0, \tag{2.5}$$

where

$$F(r,u) = \lim_{N\to\infty} \int_{-\infty}^{N} d\sigma_e(E)\cos(uE^{1/2})\cos(rE^{1/2}), \tag{2.6}$$

$$\sigma_e(E) = \begin{cases} \rho_e(E) & \text{for } E < 0, \\ \rho_e(E) - (2/\pi)E^{1/2} & \text{for } E > 0, \end{cases} \tag{2.7}$$

and

$$V(r) = 2\frac{d}{dr}K(r,r). \tag{2.8}$$

If $V(x)$ is bounded from below, which we assume, we are free to suppose that $\rho_e(E) = 0$ for $E < 0$. We define

$$\Omega(t) = \lim_{N\to\infty} \int_{-\infty}^{N} d\sigma_e(E)\cos(tE^{1/2})\,dE. \tag{2.9}$$

Clearly,

$$F(r,u) = \tfrac{1}{2}(\Omega(r+u) + \Omega(r-u)). \tag{2.10}$$

It can be shown[10] that, if $V(x)$ has a continuous derivative on $0 \leqslant x < \infty$ then $\Omega(t)$ has two derivatives as a function of t, for all t. Further,[9] $K(r,s)$, the solution of (2.5), has two continuous derivatives with respect to r and s for all r, s in the triangle $0 \leqslant s \leqslant r$.

It is also true that,[11] as $E \to \infty$,

$$\sigma_e(E) \to 0. \tag{2.11}$$

For each fixed r, Eq. (2.5) is a Fredholm equation of the second kind for $K(r,s)$. It is shown in Ref. 10 that is has a unique solution for all $r > 0$, if $\rho_e(E)$ is the spectral function of a continuous potential. The class of spectral functions $\rho_e(E)$ that are generated as $V(x)$ varies over the possible continuous potentials is described in Ref. 10.

It is natural to regard the left-hand side of Eq. (2.1) as a mapping into itself of the Banach space $\mathbb{C}(0, r)$ of continuous functions on $[0, r]$, endowed with the uniform norm

$$\|f\|_r = \sup_{0\leqslant x\leqslant r} |f(x)|. \tag{2.12}$$

It can be shown (Ref. 12, p. 101) that the norm of the integral operator F^0 with kernel $F(r,u)$, mapping $\mathbb{C}(0, r)$ into itself is

$$\|F^0\|_r = \max_{0\leqslant u\leqslant r} \int_0^r ds\, |F(s,u)| \tag{2.13}$$

provided $F(s,u)$ is a continuous function of both s and u (as is guaranteed for our class of potentials).

We wish to investigate the effect of the replacement of the exact $\rho_e(E)$ in Eq. (2.6) by an approximant $\rho_e^{(N)}(E)$, constructed from (approximate) knowledge of the first N energy levels and the corresponding values of $u_{n,e}(0)$. As N increases, $\rho_e^{(N)}(E)$ is supposed to get increasingly close to the exact $\rho_e(E)$ and our purpose is to examine whether the potentials $V_N(x)$, corresponding to $\rho_e^{(N)}(E)$ by the Gel'fand–Levitan equations, approach the original $V(x)$. For confining potentials, the approximation of $\rho_e(E)$ can occur in two ways:

(i) the functions $\rho_e^{(N)}(E)$ are constructed from the exact values of $u_{n,e}^2(0)$, E_n^e for $n \leqslant N$, and a quite arbitrary function for $E > E_{N,e}$, chosen only in such a way as to make $\rho_e^{(N)}(E)$ an admissible spectral measure for a potential (with continuous derivative).

(ii) the values $E_n^{e(N)}$, $(u_{n,e}^{(N)})^2(0)$ occurring in $\rho_e^{(N)}(E)$ for the energy levels with $n \leqslant N$ are different from those of the true potential, but approach them as $N \to \infty$.

Clearly, we expect the study of the first type of approximants to be easier.

From Eqs. (2.10) and (2.13) we see that, if the functions $\Omega_N(t)$ constructed from the approximate measures converge to the function $\Omega(t)$ of the exact potential, uniformly on some finite interval $0 \leqslant t \leqslant 2r_0$, the approximate integral operators F_N^0 approach the exact one in norm, uniformly for all r, $0 \leqslant r \leqslant r_0$.

Indeed,

$$\begin{aligned} \|F^0 - F_N^0\|_r &= \max_{0<t<r} \int_0^r |F(s,t) - F_N(s,t)|\,ds \\ &\leqslant \tfrac{1}{2}\Big(\max_{0<t<r} \int_0^r |\Omega(s+t) - \Omega_N(s+t)|\,ds \\ &\quad + \max_{0<t<r} \int_0^r |\Omega(s-t) - \Omega_N(s-t)|\,ds\Big) \end{aligned} \tag{2.14}$$

and, for any ϵ, we can choose N so that, for all $|x| \leqslant 2r_0$, $|\Omega(x) - \Omega_N(x)| \leqslant \epsilon$ [notice that $\Omega(x)$ is even in x]. Clearly, the free terms in the Gel'fand–Levitan equation also tend uniformly to the true one, inside any triangle $0 \leqslant s \leqslant r \leqslant r_0$.

It follows then that the solutions of the Gel'fand–Levi-

tan equations tends to the true solution in the uniform norm in the triangle $0\leqslant s\leqslant r\leqslant r_0$. To this end, we refer first to Appendix A, where we show that the quantity $\|(1-F^0)^{-1}\|_r$ is a continuous function of r, for $r>0$. Let then $d(r_0)=\max_{0<r<r_0}\|(1-F^0)^{-1}\|_r$; clearly, $d(r_0)<\infty$. Given $\epsilon<1/r_0$, we can choose then N so that $\|F-F_N\|_r\leqslant\epsilon/d(r_0)$ [and so $\|F^0-F_N^0\|_r\leqslant\epsilon r_0/d(r_0)$], for all $0\leqslant r\leqslant r_0$. We can then write, at fixed r

$$
\begin{aligned}
&\|K(r,t)-K_N(r,t)\|_r\\
&\quad=\|(1-F^0)^{-1}F-(1-F_N^0)^{-1}F_N\|_r\\
&\quad\leqslant\|(1-F^0)^{-1}\|_r\|F-F_N\|_r+\|(1-F^0)^{-1}-(1+(1-F^0)^{-1}(F^0-F_N^0))^{-1}(1-F^0)^{-1}\|_r\times\|F_N\|_r\\
&\quad\leqslant\epsilon+\|(1-F^0)^{-1}-(1-(1-F^0)^{-1}(F^0-F_N^0)+((1-F^0)^{-1}(F^0-F_N^0))^2-\cdots)(1-F^0)^{-1}\|_r\|F_N\|_r\\
&\quad\leqslant\epsilon+\|(1-F^0)^{-1}\|_r^2\|F^0-F_N^0\|_r\|F_N\|_r/(1+\|(1-F^0)^{-1}(F^0-F_N^0)\|_r)\\
&\quad\leqslant\epsilon+(d(r_0)\epsilon r_0\|F_N\|_r/(1-\epsilon r_0)).
\end{aligned}
\tag{2.15}
$$

In (2.15), in the second step, we have expanded $(1+(1-F^0)^{-1}(F^0-F_N^0))^{-1}$ in a Neumann series. From (2.5) we conclude that, in particular, $K(r,r)$ and $K_N(r,r)$ get arbitrarily close to each other as N increases, uniformly for $0\leqslant r\leqslant r_0$. We point out that use of a norm different from the uniform one would not have allowed the conclusion that the values of $K_N(r,s)$ along the diagonal $r=s$ approach $K(r,r)$ uniformly for $0\leqslant r\leqslant r_0$, and this is essential in order to state anything about the potentials.

It is now enough to choose a function $\phi(r)$ to have compact support contained in $(0, r_0)$, and with a continuous derivative. Then

$$
\begin{aligned}
&\left|\int_0^\infty\phi(r)\frac{d}{dr}(K(r,r)-K_N(r,r))\,dr\right|\\
&\quad=\left|\int_0^\infty\phi'(r)(K(r,r)-K_N(r,r))\,dr\right|\\
&\quad\leqslant\sup_{0\leqslant r\leqslant r_0}|K(r,r)-K_N(r,r)|\times r_0\sup_{0\leqslant r\leqslant r_0}|\phi'(r)|.
\end{aligned}
\tag{2.16}
$$

The last line can be made arbitrarily small for N sufficiently large, as is shown in (2.15).

So, if $\Omega_N(t)$ converges to $\Omega(t)$ uniformly for $0\leqslant t\leqslant t_0$, the corresponding $V_N(x)$ approximates $V(x)$ increasingly well in the sense of (1.1). In particular, if the approximant $\rho_e^{(N)}(E)$ reproduces the first N energy levels of $V(x)$ exactly,

$$
\begin{aligned}
|\Omega(t)-\Omega_N(t)|\leqslant&\left|\Omega(t)-\int_0^\Lambda\cos(tE^{1/2})\,d\sigma_e(E)\right|\\
&+\left|\lim_{M\to\infty}\int_\Lambda^M\cos(tE^{1/2})\,d\sigma_e^{(N)}(E)\right|,\quad E_N<\Lambda<E_{N+1}.
\end{aligned}
\tag{2.17}
$$

We can show that the first term tends uniformly to zero, as $\Lambda\to\infty$ for $0\leqslant x\leqslant x_0$, in our class of potentials (see Appendix B). The approximant $\Omega_N(t)$ is defined by means of an *a priori* unknown function $\sigma_e^{(N)}(E)$, for $E>\Lambda$. This function can in particular be chosen to be differentiable at all $E>\Lambda$; this would correspond to an approximation of $V(x)$ by a nonconfining potential. Assuming this to be the case, we gain from (2.17) a sufficient condition for $\sigma_e^{(N)}(E)$ in order that convergence be fulfilled:

$$
\lim_{\Lambda\to\infty}\left|\int_\Lambda^\infty\cos(tE^{1/2})\,d\sigma_e^{(N)}(E)\right|=0
\tag{2.18}
$$

uniformly in t, $0\leqslant t\leqslant t_0$.

A first obvious question is whether it is trivial to find $\sigma_e^{(N)}(E)$ satisfying (2.18) or not. In the opinion of the author, it is easy to do so, although the solution of the associated Gel'fand–Levitan equation might be complicated. Let, for instance ($\epsilon>0$)

$$
\sigma_e^{(N)}(E)=(\Lambda/E)^\epsilon\sigma_e(\Lambda),\quad E>\Lambda,
\tag{2.19}
$$

where $\sigma_e(\Lambda)$ is estimated from knowledge of the first N bound states ($E_{N+1}>\Lambda>E_N$). Then

$$
\left|\int_\Lambda^\infty d\sigma_e^{(N)}(E)\cos(tE^{1/2})\right|\leqslant\sigma_e(\Lambda)
\tag{2.20}
$$

$\sigma_e(\Lambda)$ vanishes, as $\Lambda\to\infty$, by virtue of (2.11).

A second question arises in connection with the reflectionless approximants of Refs. 1–5 as mentioned in the Introduction, if we construct them starting from prescribed even-parity energy levels and corresponding values of $u_{n,e}(0)$, do they converge to the true potential? To answer this question, one must actually determine $\sigma_e^{(N)}(E)$ for this construction (this is done in Secs. III and IV) and check whether (2.18) is fulfilled (this is done in Sec. V).

We now turn to the problem of approximation starting from the odd-parity energy levels and the values of $(u_{n,o})'(0)$. A first complication is that the kernel of the Gel'fand–Levitan equation can no longer be written in the form (2.6) but rather as[9]:

$$
\begin{aligned}
F(r,u)&=\frac{\partial^2}{\partial r\partial u}\int_{-\infty}^\infty\frac{(1-\cos(rE^{1/2}))(1-\cos(uE^{1/2}))}{E^2}\,d\sigma_o(E)\\
&\equiv\frac{\partial^2}{\partial r\partial u}\tilde F(r,u),
\end{aligned}
\tag{2.21}
$$

with

$$
\sigma_o(E)=\begin{cases}\rho_o(E), & E<0,\\ \rho_o(E)-(2/3\pi)E^{3/2}, & E>0\end{cases}.
\tag{2.22}
$$

It can be shown[13] that, as $E\to\infty$,

$$
\sigma_o(E)=O(E).
\tag{2.23}
$$

The previous arguments concerning the approximation of F can in principle be repeated; however, one must now prove the uniform convergence of the derivatives of the approximants $\tilde F_N(r,u)$ to $\tilde F(r,u)$ and this appears to be somewhat difficult.[14] So, following Ref. 7, we try to reduce the problem to the one with wrong boundary conditions. Namely, given the spectral function $\rho_o^{(N)}(E)$ of the approximant constructed from the correct boundary conditions, it is possible to find

the spectral function corresponding to "wrong" boundary conditions $\tilde{\rho}_e^{(N)}(E)$ (see Sec. IV).

For the latter approximant, one can apply completely the reasoning of this section. However, condition (2.6) is not sufficient for convergence: the difference between the exact spectral function $\rho_e(E)$ and the approximate one $\tilde{\rho}_e^{(N)}(E)$ extends down to $E=0$. Indeed, it is not in general true that the estimates of the positions of the energy levels with wrong boundary conditions, $E_{n,e}^{(N)}$, obtained from $\tilde{\rho}_e^{(N)}(E)$ coincide with the true ones.

So, to prove the uniform convergence of $\Omega_N(t)$, it is sufficient to prove, apart from (2.18), that, uniformly in t,

$$\lim_{N\to\infty}\left|\sum_{i=1}^{N}(\gamma_i\cos(tE_{i,e}^{1/2})-\gamma_i^{(N)}\cos(tE_{i,e}^{N\,1/2}))\right|=0, \quad (2.24)$$

where $\gamma_i=(u_{i,e}^2)(0)$, $\gamma_i^{(N)}$ is its estimate from the approximant $\tilde{\rho}_e^{(N)}(E)$.

We shall perform this explicitly for the reflectionless approximants to a special class of potentials in Secs. VI and VII.

In the following, we shall only refer to Eqs. (2.18) and (2.24) as sufficient conditions for convergence. In the next section, we present a derivation of the Jost functions for a symmetric superposition of solitons, as a preliminary step towards finding their spectral measure.

III. THE JOST FUNCTIONS OF THE SOLITON POTENTIALS

We consider the Schrödinger equation for an approximant $V_N^0(x)=V_N(x)-\Lambda$ of the type considered in Refs. 1–5, shifted in energy by an amount Λ, so that it tends to zero as $x\to\infty$,

$$u''(x)+(E^0-V_N^0(x))u(x)=0. \quad (3.1)$$

We introduce two regular solutions of (3.1), $\Phi(E,x)$, $\Psi(E,x)$, by the requirements

$$\begin{aligned}&\Phi(E^0,0)=0, \quad \Phi'(E^0,0)=1,\\ &\Psi(E^0,0)=-1, \quad \Psi'(E^0,0)=0.\end{aligned} \quad (3.2)$$

We denote

$$k=(E^0)^{1/2}, \quad (3.3)$$

where the root has a cut along the positive real axis and is equal to $+i$ on the negative real axis. The Jost solutions $f(\pm k,x)$ of (3.1) are introduced by the requirement

$$f(\pm k,x)\sim e^{\pm ikx}, \quad x\to\infty \quad (3.4)$$

—there always exist such solutions for potentials which approach zero at infinity, as the $V_N^0(x)$ do. We define two Jost functions $F_o(k)$, $F_e(k)$, corresponding to $\Phi(k^2,x)$, $\Psi(k^2,x)$, for $k\in\mathbb{R}$,

$$\Phi(k^2,x)=(i/2k)[F_o(-k)f(k,x)-F_o(k)f(-k,x)], \quad (3.5a)$$

$$\Psi(k^2,x)=(i/2k)[F_e(-k)f(k,x)-F_e(k)f(-k,x)]. \quad (3.5b)$$

It can be shown[15] that $F_o(k)$ and $F_e(k)$ can be extended to holomorphic functions of k in the upper half-plane[16] and that they satisfy

$$F_o(k)\sim 1, \quad F_e(k)\sim -ik, \quad |k|\to\infty. \quad (3.6)$$

The difference in the asymptotic behavior of $F_1(k)$, $F_2(k)$ can be traced back to the fact that, for large k,

$$\Phi(k^2,x)\sim\sin kx/k, \quad \Psi(k^2,x)\sim-\cos kx. \quad (3.7)$$

Further, the $F_{o,e}(k)$ have the reality property with respect to $\mathrm{Re}\,k=0$:

$$F_{o,e}(-k^*)=F_{o,e}^*(k). \quad (3.8)$$

We next introduce briefly the terminology of one-dimensional scattering.[15] One defines two independent solutions of (3.1):

$$f_+(k,x)\sim e^{ikx}, \quad x\to+\infty, \quad (3.9a)$$

$$f_-(k,x)\sim e^{-ikx}, \quad x\to-\infty. \quad (3.9b)$$

Notice, $f_+(k,x)=f(k,x)$ of the "half-line" (or centrally symmetrical three-dimensional) problem. Since the equation is even in k, $f_+(-k,x)$, $f_-(-k,x)$ are also solutions of (1.1) and one introduces coefficients $c_{11}(k)$, $c_{12}(k)$ (see Ref. 15, p. 299) by means of ($k\in\mathbb{R}$)

$$f_-(k,x)=c_{11}(k)f_+(k,x)+c_{12}(k)f_+(-k,x). \quad (3.10)$$

The transmission and reflection coefficients are

$$T(k)=1/c_{12}(k), \quad R(k)=\frac{c_{11}(k)}{c_{12}(k)} \quad (3.11)$$

and a soliton potential is defined to be such that $c_{11}(k)\equiv 0$. It is true that (Ref. 15, p. 300)

$$|c_{12}(k)|^2=1+|c_{11}(k)|^2 \quad (3.12)$$

so that $|c_{12}(k)|=1$, for $k\in\mathbb{R}$, in our case. Further, it can be shown that $c_{12}(k)$ is holomorphic in the upper half k plane and $c_{12}(k)\to 1$, as $k\to\infty$. The equations by which the potential $V_N^0(x)$ is determined form the condition $c_{11}(k)\equiv 0$ and the residua and positions of its first N bound states are given, e.g., in Ref. 4 [Eqs. (2.8) and (2.9)]. If one further requires that

$$V_N^0(x)=V_N^0(-x) \quad (3.13)$$

the residua are completely determined from the positions of the bound states [see Eqs. (3.6) and (3.12) of Ref. 4] and $V_N^0(x)$ is completely fixed from knowledge of the latter ones.

The symmetry (3.13) implies a relation between the functions $f_+(k,x)$, $f_-(k,x)$. It is namely true that

$$f_-(k,x)=f_+(k,-x). \quad (3.14)$$

Indeed, if $f_+(k,x)$ is a solution of (3.1), $f_+(k,-x)$ is also a solution by (3.13), and it behaves like $e^{ik(-x)}$ as $(-x)\to\infty$. Consequently, it behaves like (3.9b) and it is equal to $f_-(k,x)$. As a consequence of (3.14), the solutions $\Phi(k^2,x)$, $\Psi(k^2,x)$ in (3.2) must be multiples of the combinations $f^+(k,x)$ and $f^-(k,x)$ defined below:

$$\Phi(k^2,x)=f^-(k,x)/g_o(k)\equiv\tfrac{1}{2}(f_+(k,x)-f_-(k,x))/g_o(k), \quad (3.15a)$$

$$\Psi(k^2,x)=f^+(k,x)/g_e(k)\equiv\tfrac{1}{2}(f_+(k,x)+f_-(k,x))/g_e(k). \quad (3.15b)$$

We can now relate the Jost functions $F_{o,e}(k)$ of (3.5) to the coefficient $c_{12}(k)$ of Eq. (3.10). Indeed, we get from (3.5) and

(3.10) that $[W(f_1,f_2)\equiv f_1\partial f_2/\partial x - f_2\partial f_1/\partial x]$

$$F_o(k) = W(\Phi(k,x), f_+(k,x)) = (1/g_o(k))W(f^-(k,x), f_+(k,x)) = -ik\frac{c_{12}(k)}{g_o(k)}, \tag{3.16a}$$

$$F_o(-k) = W(\Phi(k,x), f_+(-k,x)) = -ik/g_o(k), \tag{3.16b}$$

and similarly

$$F_e(k) = W(\Psi(k,x), f_+(k,x)) = ikc_{12}(k)/g_e(k), \tag{3.17a}$$

$$F_e(-k) = W(\Psi(k,x), f_+(-k,x)) = -ik/g_e(k), \tag{3.17b}$$

where we have used the fact that

$$c_{11}(k) \equiv -(1/2ik)W(f_-(k,x), f_+(-k,x)) \equiv 0, \tag{3.18a}$$

$$c_{12}(k) \equiv (1/2ik)W(f_-(k,x), f_+(k,x)). \tag{3.18b}$$

It follows from (3.16) and the properties of $c_{12}(k)$ and $F_o(k)$ that $g_o(k)$ is a holomorphic function in the upper half-plane, which has zeros at the positions of the even-parity energy levels and which behaves asymptotically like $-ik$. Similarly, from (3.17) we deduce that $g_e(k)$ is holomorphic in the upper half-plane, has zeros at the positions of the odd-parity levels and tends asymptotically to -1. From Eqs. (3.8), (3.16), and (3.17), we get information concerning the phase of $g_o(k)$ and $g_e(k)$ on the real axis, namely that

$$g_o(k)/g_o^*(k) = -c_{12}(k), \quad k\in\mathbb{R}, \tag{3.19a}$$

$$g_e(k)/g_e^*(k) = c_{12}(k), \quad k\in\mathbb{R}. \tag{3.19b}$$

Now, knowledge of the zeros, of the asymptotic behavior, and of the phase of $g_{e,o}(k)$ fixes them completely. We shall consider for simplicity only the case of $g_e(k)$ and of an even number of levels (in the one-dimensional potential) $N_0 = 2N$. Then, it follows from (3.12) that

$$c_{12}(k) = \prod_{i=1}^{N}\frac{(ik+\lambda_i)(ik+\mu_i)}{(ik-\lambda_i)(ik-\mu_i)}, \tag{3.20}$$

where, $i\lambda_i$, $i\mu_i$ are the positions of the even- and odd-parity levels, respectively. Consequently, we can write

$$g_e(k) = \prod_{i=1}^{N}(ik+\mu_i)g_e^1(k), \tag{3.21}$$

where $g_e^1(k)$ is a function holomorphic in the upper half-plane, with no zeros there and behaving at infinity like $(-1)/(ik)^N$. We deduce further, from (3.19b) that

$$g_e^1(k)/g_e^1(k)^* = (-1)^N\prod_{i=1}^{N}(ik+\lambda_i)/\prod_{i=1}^{N}(ik-\lambda_i). \tag{3.22}$$

It is now easy to see that the following choice for $g_e^1(k)$ solves the problem:

$$g_e^1(k) = -1/\prod_{i=1}^{N}(ik-\lambda_i). \tag{3.23}$$

We conclude that

$$g_e(k) = -\prod_{i=1}^{N}(ik+\mu_i)/\prod_{i=1}^{N}(ik-\lambda_i). \tag{3.24a}$$

Similarly, it turns out that

$$g_o(k) = -ik\prod_{i=1}^{N}(ik+\lambda_i)/\prod_{i=1}^{N}(ik-\mu_i). \tag{3.24b}$$

As a consequence, we have found the Jost functions $F_{e,o}(k)$ of our problem, by means of (3.16a), (3.17a), (3.24a), and (3.24b):

$$F_o(k) = \prod_{i=1}^{N}(ik+\mu_i)/\prod_{i=1}^{N}(ik-\lambda_i), \tag{3.25a}$$

$$F_e(k) = \left(\prod_{i=1}^{N}(ik+\lambda_i)/\prod_{i=1}^{N}(ik-\mu_i)\right)\times(-ik). \tag{3.25b}$$

Similar expressions can be obtained for an odd number of levels, but we shall see in the next section that only the case of an even number of levels is likely to be interesting in practice.

IV. THE SPECTRAL MEASURES OF THE SOLITON POTENTIALS

We are now in a position to construct the spectral functions for the soliton potentials, by means of the Jost functions obtained above. We recall to this end that, e.g., the spectral function $\rho_e(E)$ of the problem with boundary condition $u'(0) = 0$, Eq. (2.3), is the integral over the discontinuity in the E plane of that function $R_e(E)$ which is such that the combination

$$\tilde{f}(E,x) = \Phi(E,x) + R_e(E)\Psi(E,x) \tag{4.1}$$

has square integrable modulus over $(0, \infty)$ (see, e.g., Ref. 17, p. 207). Clearly, these latter functions are multiples (with k dependent coefficients) of the Jost solutions of Eq. (3.1). Consequently, to get $R_e(E)$, we just solve Eqs. (3.4), (3.5) for $f(k,x)$; one gets

$$f(k,x) = \frac{\Phi(k,x)F_e(k) - \Psi(k,x)F_o(k)}{(i/2k)(F_e(k)F_o(-k) - F_e(-k)F_o(k))}. \tag{4.2}$$

We conclude that

$$R_o(E^o) = -F_e(k)/F_o(k), \tag{4.3}$$

$$R_e(E^o) = -F_o(k)/F_e(k). \tag{4.4}$$

Using the expressions (3.25) for the Jost functions and the variable

$$E = E^o + \Lambda \tag{4.5}$$

we get

$$\frac{d\rho_o^{(N)}}{dE} = \frac{1}{\pi}\left(\prod_{i=1}^{N}(E-E_i^c)/\prod_{i=1}^{N}(E-E_i^o)\right)(E-\Lambda)^{1/2}, \quad E>\Lambda, \tag{4.6a}$$

$$\frac{d\rho_e^{(N)}}{dE} = \frac{1}{\pi}\left(\prod_{i=1}^{N}(E-E_i^o)/\prod_{i=1}^{N}(E-E_i^c)\right)(E-\Lambda)^{-1/2}, \quad E>\Lambda, \tag{4.6b}$$

where $E_i^c = \Lambda - \lambda_i^2$, $E_i^o = \Lambda - \mu_i^2$ are the energies of the even- and odd-parity levels.

For $E<\Lambda$, the spectral function coincides by definition with the one of the true confining potential.

We recall that in Eqs. (4.6), the positions of the even-parity levels in $d\rho_o^{(N)}/dE$ (and odd-parity levels in $d\rho_e^{(N)}/dE$) are unknown *a priori*, but must be determined as functions of Λ, of the positions of the odd (even-, respectively)-parity lev-

els and their residua. It is therefore necessary to express $d\rho_{e,o}^{(N)}/dE$ in Eq. (3.6) directly in terms of these "observable" quantities. We consider first the case of $d\rho_e^{(N)}/dE$ (the "wrong" boundary conditions).

We follow then Ref. 4 and introduce the function

$$\bar{\Phi}(x,k)=\begin{cases} T(k)f_-(k,x)e^{ikx}, & \text{Im } k>0, \\ f_+^*(k^*,x)e^{ikx}, & \text{Im } k<0, \end{cases} \tag{4.7}$$

where $f_-(k,x), f_+(k,x)$ are the solutions of Eq. (3.1), defined in (3.9) and $T(k)$ is given by (3.11). It is shown in Refs. 1 and 4 that $\bar{\Phi}(x,k)$ is, for reflectionless potentials, a rational function in the whole k plane, for any x, and that $\bar{\Phi}(x,k)\to 1$ as $|k|\to\infty$. Further, $\bar{\Phi}(x,k)$ has poles at the points $i\lambda_p$, $i\mu_p$, i.e., at the positions of the even- and odd-parity levels (shifted by Λ), with residua equal to, e.g., $if_+(i\lambda_p, x)\exp(-\lambda_p x)\times[\int_{-\infty}^{\infty} f_+^2(i\lambda_p,x)dx]^{-1}$ (cf. Ref. 15, p. 302). If one introduces normalized wavefunctions $u_{p,e(o)}(x)$, $\int_0^\infty u_{p,e(o)}^2(x)dx=1$, the residuum becomes equal to $iu_p(x)\alpha_p(x)/2$, where $\alpha_p(x)=c_p\exp(-\lambda_p x)$ is the asymptotic behavior of $u_p(x)$. One can also check that

$$\bar{\Phi}(x,\ -i\lambda_p(\mu_p))_{,}=u_{p,e(o)}(x)/\alpha_{p,e(o)}(x). \tag{4.8}$$

As a consequence, the function of $E=E^o+\Lambda$,

$$\mathcal{F}_e(E)=\bar{\Phi}(0,k)\bar{\Phi}(0,-k) \tag{4.9}$$

has poles only at the positions of the even-parity levels, with residua equal to $-\lambda_p u_{p,e}^2(0)=-(\Lambda-E_p^e)^{1/2}u_{p,e}^2(0)$; has zeros at the positions of the odd-parity levels; and approaches 1 at infinity. So, the following identity holds:

$$\prod_{i=1}^{N}\frac{E-E_i^o}{E-E_i^e}=1-\sum_{k=1}^{N}\frac{u_{k,e}^2(0)(\Lambda-E_k^e)^{1/2}}{E-E_k^e}. \tag{4.10}$$

Notice, the right-hand side of Eq. (4.10) contains only "measurable" quantities (for the "wrong" boundary conditions). The positions of the odd-parity levels are completely determined by (4.9) as the zeros of a polynomial of degree N in E. Notice that, in order that the odd-parity levels lie below Λ, the inequality

$$\sum_{k=1}^{N}\frac{u_{k,e}^2(0)}{(\Lambda-E_k^e)^{1/2}}\leqslant 1 \tag{4.11}$$

must be fulfilled. The case of N_0 odd appears when inequality is attained in (4.11) and is, therefore, rather special. It can, however, be treated as a particular case of the situation with N_0 even.[4] In the following, we shall work only with the right-hand side of (4.10), so that we shall include those situations when (4.11) is fulfilled. No inequality analogous to (4.11) occurs if the correct boundary conditions are used.

We conclude therefore, from (4.10),

$$\frac{d\rho_e^{(N)}}{dE}=\frac{1}{\pi}\frac{1}{(E-\Lambda)^{1/2}}\left(1-\sum_{k=1}^{N}\frac{u_{k,e}^2(0)(\Lambda-E_k^e)^{1/2}}{E-E_k^e}\right). \tag{4.12}$$

We now turn to the calculation of $d\rho_o^{(N)}/dE$, Eq. (2.4). We follow again Ref. 4 and introduce

$$\mathcal{F}_0(0,E)=G(0,k)G(0,-k), \tag{4.13}$$

where $G(x,k)$ is defined as in Eq. (3.8) of Ref. 4:

$$\begin{aligned} G(x,k)&\equiv\frac{\partial\bar{\Phi}}{\partial x}(x,k)-ik\bar{\Phi}(x,k) \\ &=-ik+\frac{1}{2}\sum_{k=1}^{N_0}\alpha_k(x)u_k(x)+\frac{i}{2}\sum_{p=1}^{N}\frac{\alpha_{p,o}(x)u'_{p,o}(x)}{k-i\mu_p} \\ &\quad+\frac{i}{2}\sum_{p=1}^{N}\frac{\alpha_{p,e}(x)u'_{p,o}(x)}{k-i\lambda_p}. \end{aligned} \tag{4.14}$$

The function $\mathcal{F}_0(0,E)$ is meromorphic in the E plane, has poles at the odd-parity levels, with residua equal to $-(u'_{p,o}(0))^2(\Lambda-E_p^0)^{1/2}$, and vanishes at $E=\Lambda$ (see Ref. 4), for an even number $N_0=2N$ of energy levels. Writing a dispersion relation for $\mathcal{F}_0(0,E)/(E-\Lambda)$, we obtain Eq. (4.22) of Ref. 4, in the present notation:

$$\prod_{p=1}^{N}\frac{E-E_p^e}{E-E_p^o}=1+\sum_{p=1}^{N}\frac{(u'_{p,o}(0))^2}{(\Lambda-E_p^o)^{1/2}(E-E_p^o)}. \tag{4.15}$$

Equation (4.15) determines the even-parity levels completely, if the odd-parity ones and the corresponding $u'_{p,o}(0)$ are known. This time, we get n values of E_p^e for any choice of Λ, in contrast to the previous case (it is namely clear from (4.15) that no root E_p^e will be larger than Λ). The case of N_0 odd is of no practical interest since the subtraction constant of $G(x,k)$ is *a priori* unknown, and would appear as a further parameter. We conclude from Eq. (4.6a) ($E\geqslant\Lambda$),

$$\frac{d\rho_o^{(N)}}{dE}=\frac{1}{\pi}(E-\Lambda)^{1/2}\left(1+\sum_{p=1}^{N}\frac{(u'_{p,o}(0))^2}{(\Lambda-E_p^o)^{1/2}(E-E_p^o)}\right). \tag{4.16}$$

As mentioned in Sec. II, in order to verify that the potentials corresponding to (4.16) approximate the original confining potential increasingly well, as $N\to\infty$, we shall not use the Gel'fand-Levitan equation (2.21), but rather compute the spectral function of the same potential with "wrong" boundary conditions. This is already given by (4.12), however, in terms of the unknown quantities $u_{k,e}^2(0)$, E_k^e, $k=1,\ldots,N$. In order to express it in terms of $(u'_{k,o}(0))^2$, E_k^o we return to Eqs. (4.3), (4.4); we get, using (4.15),

$$\begin{aligned} R_e^{(N)}(E)&=\frac{1}{R_o^{(N)}(E)}=\frac{1}{(E-\Lambda)^{1/2}}\prod_{i=1}^{N}\frac{E-E_i^o}{E-E_i^e} \\ &=\frac{1}{(E-\Lambda)^{1/2}} \\ &\quad\times\left[1\Big/\left(1+\sum_{p=1}^{N}\frac{(u'_{p,o}(0))^2}{(\Lambda-E_p^o)^{1/2}(E-E_p^o)}\right)\right]. \end{aligned} \tag{4.17}$$

The function $R_e^{(N)}(E)$ has poles at the positions $E_{i,e}^{\Lambda}$, determined by the zeros of the denominator of (4.17). Their residua $\gamma_{i,e}^{\Lambda}$, occurring in Eq. (2.24) are

$$\gamma_{i,e}^{\Lambda}=\left(\frac{dR_o^{(N)}}{dE}\bigg|_{E=E_{i,e}^{\Lambda}}\right)^{-1}. \tag{4.18}$$

They depend on the value of Λ. The function $R_e^{(N)}(E)$ has further a cut for $E\geqslant\Lambda$ with a discontinuity just equal to (four times) the reciprocal of the discontinuity of $R_o^{(N)}(E)$ (the lat-

ter function is purely imaginary for $E \geqslant \Lambda$),

$$\frac{d\rho_e^{(N)}}{dE} = \frac{1}{\pi}\frac{1}{(E-\Lambda)^{1/2}} \times\left(1\Big/\left(1+\sum_{p=1}^{N}\frac{(u'_{p,o}(0))^2}{(\Lambda-E_p^o)^{1/2}(E-E_p^o)}\right)\right). \tag{4.19}$$

By means of Eqs. (4.12), (4.18), and (4.19) we construct in the next sections the kernels of the Gel'fand–Levitan equation determining the potential $V_N(x)$ and compare them with the kernels of the exact Gel'fand–Levitan equation for $V(x)$. We shall namely verify whether the convergence criteria, Eqs. (2.18) and (2.24), are fulfilled.

V. CONVERGENCE OF THE APPROXIMANTS FOR "WRONG" BOUNDARY CONDITIONS

In this section we show that, if the potential $V(x)$ is such that there exist constants C, ϵ, so that

$$|\sigma_e(E)| < CE^{-\epsilon}, \tag{5.1}$$

then the reflectionless approximants constructed with the information coming from the wrong boundary conditions converge to the true potential, in the sense of Eq. (1.1). According to Sec. II, it is enough to check that the convergence criterion (2.18) is fulfilled, i.e., as $\Lambda\to\infty$,

$$|\Omega_e(\Lambda,t)| \equiv \left|\int_\Lambda^\infty \cos(tE^{1/2})\, d\sigma_e^{(N)}(E)\right| \to 0 \tag{5.2}$$

uniformly with respect to t, $0\leqslant t\leqslant t_0$.

To this end, we use explicitly Eq. (4.12) and write

$$\Omega_e(\Lambda,t) = \int_\Lambda^\infty \cos(tE^{1/2})\left[\frac{1}{\pi}\frac{1}{(E-\Lambda)^{1/2}}\left(1-\int_0^\Lambda \frac{d\rho_e(\mu)(\Lambda-\mu)^{1/2}}{E-\mu}\right)-\frac{1}{\pi}E^{-1/2}\right]dE, \tag{5.3}$$

where we have replaced

$$\frac{d\rho_e}{dE} = \sum_{i=1}^{\infty} u_{i,e}^2(0)\delta(E-E_i^e). \tag{5.4}$$

Writing

$$d\rho_e(\mu) = d\sigma_e(\mu) + d\left((2/\pi)\mu^{1/2}\right) \tag{5.5}$$

one can check explicitly that

$$\frac{1}{\pi}\frac{1}{(E-\Lambda)^{1/2}}\left(1-\int_0^\Lambda \frac{d\left((2/\pi)\mu^{1/2}\right)(\Lambda-\mu)^{1/2}}{E-\mu}\right)-\frac{1}{\pi}E^{-1/2}\equiv 0, \tag{5.6}$$

so that we are left with

$$\Omega_e(\Lambda,t) \equiv \frac{1}{\pi}\int_\Lambda^\infty \frac{\cos(tE^{1/2})}{(E-\Lambda)^{1/2}}\left(\int_0^\Lambda \frac{d\sigma_e(\mu)(\Lambda-\mu)^{1/2}}{E-\mu}\right)dE. \tag{5.7}$$

We introduce the variables

$$E=\Lambda e, \quad \mu=\Lambda x, \quad \tau = t\sqrt{\Lambda} \tag{5.8}$$

and write[18]:

$$\begin{aligned}\Omega_e(\Lambda,t) &= \frac{1}{\pi}\int_1^\infty \frac{\cos(\tau\sqrt{e})}{(e-1)^{1/2}}\int_0^\Lambda \frac{d\sigma(\mu)(1-x)^{1/2}}{e-x}\,de \\ &= \frac{1}{\pi}\int_1^\infty \frac{\cos(\tau\sqrt{e})-\cos\tau}{(e-1)^{1/2}}de\int_0^\Lambda \frac{d\sigma(\mu)(1-x)^{1/2}}{e-x} + \frac{\cos\tau}{\pi}\int_1^\infty \frac{de}{(e-1)^{1/2}}\int_0^\Lambda \frac{d\sigma(\mu)(1-x)^{1/2}}{e-x} \\ &\equiv I'_\Lambda + I''_\Lambda.\end{aligned} \tag{5.9}$$

The integral I''_Λ can be evaluated explicitly:

$$I''_\Lambda = (2/\pi)\sigma(\Lambda)\cos\tau. \tag{5.10}$$

Clearly, as $\Lambda\to\infty$, $I''_\Lambda\to 0$ uniformly with respect to t. Integrating by parts in I'_Λ, we get

$$I'_\Lambda = \frac{1}{2\pi}\int_1^\infty \frac{\cos(\tau\sqrt{e})-\cos\tau}{(e-1)^{1/2}}\left|\frac{1}{\Lambda}\int_0^\Lambda \frac{\sigma(\mu)(e+x-2)}{(e-x)^2(1-x)^{1/2}}\,d\mu\right|de. \tag{5.11}$$

Clearly, for $t=0$, $I'_\Lambda = 0$. We part I'_Λ into two:

$$\begin{aligned}I'_\Lambda = I_{\Lambda 1} + I_{\Lambda 2} \equiv &\frac{1}{2\pi}\int_1^\infty \frac{\cos(\tau\sqrt{e})-\cos\tau}{e-1}\left|\frac{1}{\Lambda}(e-1)^{1/2}\int_0^\Lambda \frac{\sigma(\mu)\,d\mu}{(e-x)(1-x)^{1/2}}\right|de \\ &+\frac{1}{2\pi}\int_1^\infty \frac{\cos(\tau\sqrt{e})-\cos\tau}{e-1}\left|\frac{2}{\Lambda}(e-1)^{1/2}\int_0^\Lambda \frac{\sigma(\mu)(1-x)^{1/2}}{(e-x)^2}\,d\mu\right|de.\end{aligned} \tag{5.12}$$

Consider now $I_{\Lambda 1}$:

$$|I_{\Lambda 1}| < \frac{1}{2\pi}\int_1^\infty \frac{|\cos(\tau\sqrt{e}) - \cos\tau|}{e-1}(e-1)^{1/2}\int_0^1 \frac{|\sigma(\Lambda x)|\,dx}{(e-x)(1-x)^{1/2}}\,de$$
$$= \frac{1}{2\pi\Lambda^\epsilon}\int_1^\infty \frac{|\cos(\tau\sqrt{e}) - \cos\tau|}{e-1}(e-1)^{1/2}\int_0^1 \frac{\Lambda^\epsilon x^\epsilon|\sigma(\Lambda x)|\,dx}{x^\epsilon(e-x)(1-x)^{1/2}}\,de$$
$$< \frac{C}{2\pi\Lambda^\epsilon}\int_1^\infty \frac{|\cos(\tau\sqrt{e}) - \cos\tau|}{e-1}\frac{1}{\sqrt{e}}\,de, \tag{5.13}$$

where we have used the fact that $|\sigma(\Lambda x)|$ is bounded by $C_1/\Lambda^\epsilon x^\epsilon$, and that the second integral behaves like $C_2/(e-1)^{1/2}$ as $e\to 1$. The constant C appears in the majorization

$$C_1(e-1)^{1/2}\int_0^1 \frac{dx}{x^\epsilon(e-x)(1-x)^{1/2}} < C/\sqrt{e}, \quad e>1. \tag{5.14}$$

Further,

$$|I_{\Lambda 1}| < \frac{C}{\pi\Lambda^\epsilon}\int_4^\infty \frac{de}{(e-1)\sqrt{e}} + \frac{C}{2\pi\Lambda^\epsilon}\int_1^4 \frac{|\cos(\tau\sqrt{e}) - \cos\tau|}{(e-1)\sqrt{e}}\,de \equiv I'_{\Lambda 1} + I''_{\Lambda 1}. \tag{5.15}$$

Clearly, both integrals are convergent, for finite Λ. The first one is independent of t and vanishes as $\Lambda\to\infty$. Consider then $I''_{\Lambda 1}$ with $\kappa = \sqrt{e}, \kappa = \kappa - 1$:

$$I''_{\Lambda 1} = \frac{1}{2\pi\Lambda^\epsilon}\int_1^2 \frac{|\cos\tau\kappa - \cos\tau|}{\kappa^2 - 1}\,d\kappa$$
$$= \frac{1}{2\pi\Lambda^\epsilon}\int_0^1 \frac{1}{\kappa+2}\,\frac{|(\cos(\tau\kappa)-1)\cos\tau + \sin\tau\sin\tau\kappa|}{\kappa}$$
$$\leqslant \frac{1}{2\pi\Lambda^\epsilon}\int_0^1 \frac{|\cos\tau|}{\kappa+2}\,\frac{|\cos\tau\kappa - 1|}{\kappa}\,d\kappa + \frac{1}{2\pi\Lambda^\epsilon}\int_0^1 \frac{|\sin\tau||\sin\tau\kappa|}{\kappa(\kappa+2)}\,d\kappa$$
$$\equiv \mathcal{J}_1(\Lambda,t)|\cos\tau| + \mathcal{J}_2(\Lambda,t)|\sin\tau|. \tag{5.16}$$

Let now $\tau\kappa = u$, in both $\mathcal{J}_1, \mathcal{J}_2$. Then

$$\mathcal{J}_1(\Lambda,t) = \frac{\tau}{2\pi\Lambda^\epsilon}\int_0^\tau \frac{1-\cos u}{u(u+2\tau)}\,du \leqslant \frac{1}{4\pi\Lambda^\epsilon}\int_0^1 \frac{1-\cos u}{u}\,du + \frac{\tau\theta(\tau-1)}{2\pi\Lambda^\epsilon}\int_1^\tau \frac{du}{u(u+2\tau)}$$
$$\leqslant \begin{cases} \text{const}/\Lambda^\epsilon & \text{for } \tau\leqslant 1, \\ \text{const}/\Lambda^\epsilon + \text{const}\times \ln\tau/\Lambda^\epsilon & \text{for } \tau\geqslant 1. \end{cases} \tag{5.17}$$

But $\tau\leqslant t_0\sqrt{\Lambda}$, where t_0 is a number larger than t, and for $\tau\geqslant 1$, $|\ln\tau|\leqslant\ln(t_0\sqrt{\Lambda})$. So, for all $t\leqslant t_0$,

$$\mathcal{J}_1(t,\Lambda)\leqslant \text{const}\times\max[\ln(\Lambda^{1/2}t_0)/\Lambda^\epsilon,\ 1/\Lambda^\epsilon]. \tag{5.18}$$

Therefore, $\mathcal{J}_1(t,\Lambda)$ approaches zero uniformly, as $\Lambda\to\infty$, for all $t\in[0,t_0]$. Clearly, in the same way,

$$\mathcal{J}_2(t,\Lambda) = \frac{\tau}{2\pi\Lambda^\epsilon}\int_0^\tau \frac{|\sin u|}{u(u+2\tau)}\,du \leqslant \frac{1}{4\pi\Lambda^\epsilon}\int_0^1 \frac{|\sin u|}{u} + \frac{\tau\theta(\tau-1)}{2\pi\Lambda^\epsilon}\int_1^\tau \frac{du}{u(u+2\tau)} \tag{5.19}$$

and the majorization in (5.17) can be applied with a suitable constant. For $I_{\Lambda 2}$ in (5.12), we write

$$|I_{\Lambda 2}| \leqslant \frac{1}{\pi\Lambda^\epsilon}\int_1^\infty \frac{|\cos(\tau\sqrt{e}) - \cos\tau|}{e-1}\left[(e-1)^{1/2}\int_0^1 \frac{|\sigma(\Lambda x)\Lambda^\epsilon x^\epsilon|(1-x)^{1/2}\,dx}{x^\epsilon(e-x)^2}\right]de$$
$$\leqslant \frac{\text{const}}{\Lambda^\epsilon}\int_1^\infty \frac{|\cos(\tau\sqrt{e}) - \cos\tau|}{e-1}\left[(e-1)^{1/2}\int_0^1 \frac{dx}{x^\epsilon(e-x)(1-x)^{1/2}}\right]de$$
$$\leqslant \frac{\text{const}}{\Lambda^\epsilon}\int_1^\infty \frac{|\cos(\tau\sqrt{e}) - \cos\tau|}{e-1}\frac{1}{\sqrt{e}}\,de, \tag{5.20}$$

and this is treated as in $I_{\Lambda 1}$.

We conclude that, as announced, there exists a constant C_0, so that, for $0\leqslant t\leqslant t_0$,

$$\Omega_e(\Lambda,t)\leqslant C_0\max[1/\Lambda^\epsilon,\ \ln(\Lambda^{1/2}t_0)/\Lambda^\epsilon]. \tag{5.21}$$

Clearly, $|\Omega_e(\Lambda,t)|\to 0$, as $\Lambda\to\infty$, uniformly with respect to t, $0\leqslant t\leqslant t_0$, so that, in the class of potentials delimited by (5.1), the convergence condition (2.18) is fulfilled. This concludes our proof.

VI. CONVERGENCE OF THE APPROXIMANTS FOR CORRECT BOUNDARY CONDITIONS

In this and the next section, we assume that the spectral function $\rho_o(E)$ of the confining potential $V(x)$ satisfies the following conditions, for E large enough:

$$\text{(i)}\quad \left|\frac{(u'_{n,o})^2(0)}{E_{n+1}-E_n} - \frac{1}{\pi}\sqrt{E_n}\right| < \frac{C}{E_n^\epsilon}, \quad n>N_1 \tag{6.1}$$

for some constants $C, \epsilon > 0$; this means that the semiclassical estimate for $u'_{n,o}$ [see Ref. 19, Eq. (4.21)] is increasingly accurate, as $E \to \infty$. (We drop from now on the index o on E_n^o.)

(ii) The energy levels E_n admit of an approximation, for $n > N_1$:

$$|E_n - a_0 n^\alpha - a_1 n^{\alpha-\delta_1} - \cdots| < \frac{\text{const}}{n^{1-\alpha+\epsilon'}}, \tag{6.2}$$

$$0<\alpha<1, \quad 0<\delta_1<1, \quad \epsilon'>0.$$

Equation (6.2) is satisfied, in a semiclassical estimate, by potentials that rise less fast than $x^{2-\epsilon}$, $\epsilon > 0$ [see Ref. 19, Eq. (4.33)]. There is no loss of generality to assume $N_1 = 0$ in (6.1) and (6.2) since one can take sufficiently large constants on the right-hand side.

We further assume that

(iii) the cutoff Λ_N, $E_N < \Lambda_N < E_{N+1}$, is chosen in such a way that, for all N,

$$C_1 < (\Lambda_N - E_N)/(E_{N+1} - E_N) < C_2 \tag{6.3}$$

for some constants C_1, C_2.

To this end, we notice we can write

$$\rho(E) = \rho_1(E) + \rho_2(E), \tag{6.4a}$$

$$\rho_1(E) = \sum_{E_i<E} \frac{1}{\pi} E_i^{1/2}(E_{i+1} - E_i), \tag{6.4b}$$

$$|\rho_2(E)| \equiv \left| \sum_{E_i<E} \left((u'_{i,o})^2(0) - \frac{1}{\pi} E_i^{1/2}(E_{i+1} - E_i) \right) \right|$$
$$\leqslant \sum_{E_i<E} \frac{C}{E_i^\epsilon}(E_{i+1} - E_i) \leqslant \text{const} \times \int_0^E \frac{d\mu}{\mu^\epsilon}$$
$$= \text{const} \times E^{1-\epsilon}, \tag{6.4c}$$

and

$$\sigma(E) = \rho(E) - \frac{2}{3\pi} E^{3/2}$$
$$= \rho_1(E) - \frac{1}{\pi} \int_0^E E^{1/2}\, dE + \rho_2(E)$$
$$= \sigma_1(E) + \rho_2(E) \equiv \sigma_1(E) + \sigma_2(E). \tag{6.5}$$

In Eq. (6.5), $\sigma_1(E)$ can be viewed as the difference between a Riemann sum for the integral $(1/\pi)\int_0^E E'^{1/2}\, dE'$ and the integral itself. The function $\rho_2(E)$ (of locally bounded variation) has a total variation $\int_0^E |d\rho_2(\mu)|$ up to the value E which increases like $E^{1-\epsilon}$. On the other hand, $\sigma_1(E)$ cannot be estimated usefully from its total variation; the latter is simply the sum $\text{var}\,\rho_1 + \text{var}((2/3\pi)E^{3/2})$ which is $O(E^{3/2})$. This in turn violates the estimate (2.23) of $\sigma(E)$.[20] However, more careful estimates of $\sigma_1(E)$ are possible because of its simple form. The decomposition (6.5) will be much used in the following.

Let us also remark that the main reason why the author does not proceed as in the previous section, and supplementary assumptions like (i) and (ii) were introduced, is not related to the awkward form (4.19) of the spectral measure, but to the appearance of $(\Lambda - E_i^o)^{1/2}$ in the denominator of (4.15). This forbids one to integrate by parts as in the previous section, unless a more careful estimate for $\sigma(\Lambda) - \sigma(\mu)$ for μ sufficiently close to Λ is available (see below). It is certainly possible that weaker assumptions than (i) and (ii) can justify the estimates which we are going to derive.

Before proceeding, we estimate $\sigma_1(E)$ in more detail, in a manner that will also be used in other calculations. We write for $E_N < E < E_{N+1}$,

$$|\sigma_1(E)|$$
$$= \left| \sum_{n=1}^N \left(\frac{1}{\pi} E_n^{1/2}(E_{n+1} - E_n) - \frac{1}{\pi} \int_{E_1}^{E_{N+1}} E^{1/2}\, dE \right) \right.$$
$$\left. + \frac{1}{\pi} \int_E^{E_{N+1}} E^{1/2}\, dE - \frac{1}{\pi} \int_0^{E_1} E^{1/2}\, dE \right|$$
$$\leqslant \sum_{n=1}^N \frac{1}{2\pi} \frac{1}{E_n^{1/2}} (E_{n+1} - E_n)^2$$
$$+ \frac{1}{\pi} E_{N+1}^{1/2}(E_{N+1} - E_N)$$
$$\times \frac{E_{N+1} - E}{E_{N+1} - E_N} + \frac{2}{3\pi} E_1^{3/2}, \tag{6.6}$$

where we have applied mean value theorems twice to estimate the first term. On the other hand, from (6.2) we get that, for large n,

$$E_{n+1} - E_n \leqslant C/n^{1-\alpha} + O(1/n^{1-\alpha+\epsilon'}), \quad \epsilon' > 0. \tag{6.7}$$

Using (6.7), we estimate the first sum in (6.6):

$$\sum_{n=1}^N \frac{1}{2\pi} \frac{1}{E_n^{1/2}} (E_{n+1} - E_n)^2 \leqslant C \sum_{n=1}^N \frac{1}{n^{\alpha/2}} ((n+1)^\alpha - n^\alpha)^2$$
$$\leqslant C \int_1^N x^{3\alpha/2-2}\, dx \leqslant C_1 E_N^{3/2-1/\alpha} \quad (\alpha > 2/3) \tag{6.8}$$

if $\alpha = \frac{2}{3}$, we get a logarithmic increase; if $\alpha < 2/3$, the sum converges and is majorized by a constant). From (6.7) one also sees that

$$(1/\pi) E_{N+1}^{1/2}(E_{N+1} - E_N) < \text{const} \times E_N^{3/2-1/\alpha}. \tag{6.9}$$

We conclude that

$$|\sigma_1(E)| < C E_N^{[3/2-1/\alpha]_+}, \tag{6.10}$$

where the bracket is zero if $\frac{3}{2} - 1/\alpha < 0$, and means a logarithmic increase if $\alpha = \frac{2}{3}$. The estimate (6.10) agrees with (2.23) for all $\alpha \leqslant 2$, which is more than is allowed by assumption (ii).

In the same manner, we derive in Appendix C the following estimates:

(a) Let $E_N > \mu > k\Lambda_N$, $0 < k < 1$ or $E_{N+1} < \mu < k_1 \Lambda_N$, $k > 1$, Λ_N obeying (6.3); then, for some constants C'_1, C'_2 (independent of N, but depending on k, k_1), $\tilde{\epsilon} = 1 - [3/2 - 1/\alpha]_+$,

$$|\sigma_1(\Lambda_N) - \sigma_1(\mu)|$$
$$< C'_1 \Lambda_N^{-\tilde{\epsilon}} |\Lambda_N - \mu| + C'_2 (E_{N+1} - E_N) E_N^{1/2}; \tag{6.11a}$$

if $E_N < \mu < E_{N+1}$, then (trivially),

$$|\sigma_1(\Lambda_N) - \sigma_1(\mu)| \leqslant \text{const} \times E_N^{1/2} |\Lambda_N - \mu| \tag{6.11b}$$

(b) Let

$$\sigma_\Lambda^{(k)}(\mu) = \int_{k\Lambda_N}^{\mu} \frac{d\sigma_1(\nu)}{(\Lambda - \nu)^{1/2}}, \quad k\Lambda_N < \mu < \Lambda_N, \ 0 < k < 1. \tag{6.12}$$

Then,

$$|\sigma_\Lambda^{(k)}(\mu)| < \text{const}(k) \Lambda_N^{1/2-\tilde{\epsilon}} + \text{const}(k) E_N^{1/2}(E_{N+1} - E_N)^{1/2}. \tag{6.13}$$

If $0<\alpha<1$, we can write (6.13) as

$$|\sigma_\Lambda^{(k)}(\mu)|<\text{const}\times\Lambda_N^{1/2-\delta},\quad \delta>0. \tag{6.14}$$

We can now verify, for large $\Lambda\equiv\Lambda_N$,[21] whether (2.18) is fulfilled. We write

$$\Omega_0(\Lambda,t)\equiv\int_\Lambda^\infty \cos(t\sqrt{E})\left(\frac{d\rho_e^{(N)}}{dE}-\frac{1}{\pi}\frac{1}{\sqrt{E}}\right)dE, \tag{6.15}$$

where we have replaced $d\sigma_e^{(N)}(E)$ by the difference between (4.19) and the free measure for "wrong" boundary conditions. Using (4.19) and writing as in Sec. V,

$$\begin{aligned}\sum_{p=1}^{N}&\frac{(u'_{p,o})^2(0)}{(\Lambda-E_p)^{1/2}(E-E_p)}\\ &\equiv\int_0^\Lambda\frac{d\rho_o(\mu)}{(\Lambda-\mu)^{1/2}(E-\mu)}\\ &=\frac{1}{\pi}\int_0^\Lambda\frac{\mu^{1/2}\,d\mu}{(\Lambda-\mu)^{1/2}(E-\mu)}+\int_0^\Lambda\frac{d\sigma_o(\mu)}{(\Lambda-\mu)^{1/2}(E-\mu)}\\ &=\frac{E^{1/2}}{(E-\Lambda)^{1/2}}+\int_0^\Lambda\frac{d\sigma_o(\mu)}{(\Lambda-\mu)^{1/2}(E-\mu)}-1,\end{aligned} \tag{6.16}$$

we get

$$\Omega_0(\Lambda,t)=\int_\Lambda^\infty\frac{\cos(t\sqrt{E})}{\sqrt{E}}\,\frac{\int_0^\Lambda d\sigma_o(\mu)/((\Lambda-\mu)^{1/2}(E-\mu))}{1+\int_0^\Lambda d\rho_o(\mu)/((\Lambda-\mu)^{1/2}(E-\mu))}\,dE. \tag{6.17}$$

The denominator in (6.17) is strictly larger than $\sqrt{E}$, for all $E\geqslant\Lambda$. We shall prove then that the majorization of $|\Omega_0(\Lambda,t)|$,

$$|\Omega_0(\Lambda,t)|\leqslant\tilde{\Omega}_N\equiv\int_\Lambda^\infty\frac{dE}{\sqrt{E}}\left|\int_0^\Lambda\frac{d\sigma_o(\mu)}{(\Lambda-\mu)^{1/2}(E-\mu)}\right| \tag{6.18}$$

converges to zero, as $\Lambda\to\infty$, which will establish (2.18). To this end, we consider more carefully the integral $N(\Lambda,E)$:

$$\begin{aligned}N(\Lambda,E)&\equiv\int_0^\Lambda\frac{d\sigma_o(\mu)}{(\Lambda-\mu)^{1/2}(E-\mu)}\\ &=\int_0^{k\Lambda}\frac{d\sigma_o(\mu)}{(\Lambda-\mu)^{1/2}(E-\mu)}+\int_{k\Lambda}^\Lambda\frac{d\sigma_o(\mu)}{(\Lambda-\mu)^{1/2}(E-\mu)}\\ &\equiv N_1(\Lambda,E)+N_2(\Lambda,E),\quad 0<k<1.\end{aligned} \tag{6.19}$$

To estimate $N_1(\Lambda,E)$, we integrate by parts and use (6.10) and (6.4c); $\bar{\epsilon}=\min[\epsilon,\tilde{\epsilon}]$:

$$\begin{aligned}|N_1(\Lambda,E)|&<\frac{|\sigma_o(k\Lambda)|}{(\Lambda(1-k))^{1/2}(E-k\Lambda)}+\frac{1}{2}\int_0^{k\Lambda}\frac{|\sigma(\mu)|\,d\mu}{(\Lambda-\mu)^{3/2}(E-\mu)}+\int_0^{k\Lambda}\frac{|\sigma(\mu)|\,d\mu}{(\Lambda-\mu)^{1/2}(E-\mu)^2}\\ &\leqslant\text{const}\times\Lambda^{-1/2-\bar{\epsilon}}\frac{1}{E/\Lambda-k}+\text{const}\times\Lambda^{-1/2-\bar{\epsilon}}\int_0^k\frac{x^{1-\bar{\epsilon}}\,dx}{(1-x)^{3/2}(E/\Lambda-x)}\\ &\quad+\text{const}\times\Lambda^{-1/2-\bar{\epsilon}}\int_0^k\frac{x^{1-\bar{\epsilon}}}{(1-x)^{1/2}(E/\Lambda-x)^2}\equiv\tilde{N}_1(\Lambda,E).\end{aligned} \tag{6.20}$$

We can now calculate $\int_\Lambda^\infty dE\,\tilde{N}_1(\Lambda,E)E^{-1/2}$; changing the integration variable to $e=E/\Lambda$, it is straightforward to verify that the result vanishes like $\Lambda^{-\bar{\epsilon}}$ as $\Lambda\to\infty$.

For $N_2(\Lambda,E)$, we use the decomposition (6.5) for $\sigma(E)$ and then (6.13) for $\sigma_1(E)$ after integration by parts:

$$\begin{aligned}|N_2(\Lambda,E)|&<\int_{k\Lambda}^\Lambda\frac{|d\sigma_2(\mu)|}{(\Lambda-\mu)^{1/2}(E-\mu)}+\left|\int_{k\Lambda}^{E_N}\frac{d\sigma_1(\mu)}{(\Lambda-\mu)^{1/2}(E-\mu)}\right|+\left|\int_{E_N}^\Lambda\frac{d\sigma_1(\mu)}{(\Lambda-\mu)^{1/2}(E-\mu)}\right|\\ &\equiv(N_{21}+N_{22}+N_{23})(\Lambda,E).\end{aligned} \tag{6.21}$$

It is true then that, in turn,

$$N_{21}(\Lambda,E)<\text{const}\times\Lambda^{-1/2-\epsilon}\int_k^1\frac{dx}{x^\epsilon(1-x)^{1/2}(E/\Lambda-x)}+\text{const}\,\frac{(\Lambda-E_N)^{1/2}}{E-E_N}, \tag{6.22a}$$

$$N_{22}(\Lambda,E)<\text{const}\times\Lambda^{1/2-\bar{\epsilon}}\frac{1}{E-E_N}+\text{const}\,\frac{E_N^{1/2}(E_{N+1}-E_N)^{1/2}}{E-E_N}, \tag{6.22b}$$

$$N_{23}(\Lambda,E)<\Lambda^{1/2}\int_{E_N}^\Lambda\frac{d\mu}{(\Lambda-\mu)^{1/2}(E-\mu)}. \tag{6.22c}$$

It is now easy to estimate the integrals $\int_\Lambda^\infty dE\,\tilde{N}_{2i}(\Lambda,E)E^{-1/2}$ with $\tilde{N}_{2i}(\Lambda,E)$, $i=1,2,3$, the right-hand sides of (6.22). The integral over $\tilde{N}_{21}(\Lambda,E)$ vanishes like $\Lambda^{-\epsilon}$ (the second term even quicker). The integral over the first term in $\tilde{N}_{22}$ vanishes like $\ln\Lambda/\Lambda^{\bar{\epsilon}}$. This vanishing is independent of the value of α in Eq. (6.2) (provided it is strictly less than 2). However, the second term in (6.22b) and $\tilde{N}_{23}(\Lambda,E)$ yield contributions proportional to $(E_{N+1}-E_N)^{1/2}\ln\Lambda$ and $(E_{N+1}-E_N)^{1/2}$ respectively, which do not vanish as $N\to\infty$ unless $\alpha<1$.

This completes the proof of the convergence to zero of $\tilde{\Omega}_N$, Eq. (6.18), and thus of the condition (2.18). In the next section, we investigate the remaining condition (2.24).

VII. CONVERGENCE OF THE APPROXIMANTS FOR CORRECT BOUNDARY CONDITIONS.II.

To estimate (2.24), we divide it into three parts:

$$\left|\sum_{i=1}^{N}\gamma_i\cos(t\sqrt{E_i^e})-\sum_{i=1}^{N}\gamma_i^{\Lambda}\cos(t\sqrt{E_i^{\Lambda}})\right|<\left|\sum_{i=1}^{N}(\gamma_i-\gamma_i^{\Lambda})\cos(t\sqrt{E_i}^{\Lambda})\right|$$
$$+\left|\sum_{i=1}^{N}\gamma_i(\cos(t\sqrt{E_i^e})-\cos(t\sqrt{E_i^{\Lambda}}))\right|<\sum_{i=1}^{N}\left|1/\left(\frac{dR_o}{dE}\right)_{E_i^e}-1/\left(\frac{dR_o}{dE}\right)_{E_i^{\Lambda}}\right|$$
$$+\sum_{i=1}^{N}\left|1/\left(\frac{dR_o}{dE}\right)_{E_i^{\Lambda}}-1/\left(\frac{dR_o^{(N)}}{dE}\right)_{E_i^{\Lambda}}\right|+\frac{t}{2}\sum_{i=1}^{N}|E_i^e-E_i^{\Lambda}|/\left(\left(\frac{dR_o}{dE}\right)_{\bar{E}_i}\sqrt{\bar{E}_i}\right)\equiv S_1^{\Lambda}+S_2^{\Lambda}+\frac{t}{2}S_3^{\Lambda}. \tag{7.1}$$

In (7.1), we have used the definition of the residua $\gamma_i=(dR_o/dE)^{-1}_{E_i^e}$, $\gamma_i^{\Lambda}=(dR_o^{(N)}/dE)^{-1}|_{E_i^{\Lambda}}$ [cf. Eq. (4.18)], and E_i^e, E_i^{Λ} are the even-parity energy levels, the latter as estimated from the positions of the zeros of the approximant $R_o^{(N)}$ [Eq. (4.15), (4.17)]. A mean value theorem has been applied for the last term; $\tilde{E}_i$ lies between E_i and E_i^{Λ}. The t dependence has been removed from S_i^{Λ}, so that we try to prove uniform convergence of (2.24) in $0\leqslant t\leqslant t_0$ by showing that S_1^{Λ}, S_2^{Λ}, and S_3^{Λ} vanish, as $\Lambda\to\infty$.

The following majorizations are straightforward:

$$S_1^{\Lambda}\leqslant\sum_{i=1}^{N}\left.\frac{d^2R_o}{dE^2}\right|_{\bar{E}_i}|E_i^e-E_i^{\Lambda}|/\left[\min_{E_i\leqslant E\leqslant E_{i+1}}\left(\frac{dR_o}{dE}\right)\right]^2,\quad \bar{E}_i\in(E_i^e,\ E_i^{\Lambda}), \tag{7.2a}$$

$$S_2^{\Lambda}\leqslant\sum_{i=1}^{N}\left|\left(\frac{dF_{\Lambda}}{dE}\right)_{E_i^{\Lambda}}\right|/\min_{E_i\leqslant E\leqslant E_{i+1}}\left(\frac{dR_o}{dE}\ \frac{dR_o^{(N)}}{dE}\right), \tag{7.2b}$$

$$S_3^{\Lambda}\leqslant\sum_{i=1}^{N}|E_i^e-E_i^{\Lambda}|/E_i^{1/2}\left|\min_{E_i\leqslant E\leqslant E_{i+1}}\left(\frac{dR_o}{dE}\right)\right|, \tag{7.2c}$$

where F_{Λ} is the difference between R_o and $R_o^{(N)}$. It is possible to express F_{Λ} by means of an unsubtracted dispersion relation, in terms of $\sigma_o(E)$. Indeed, in the class of functions obeying (i), $R_o\sim\sqrt{E}+O(E^{-\bar{\epsilon}})$, $\bar{\epsilon}>0$, for $E\to\infty$ and $R_o^{(N)}\sim(-E)^{1/2}+O(E^{-1/2})$. Using (6.16), one gets

$$F_{\Lambda}(E)=\int_{\Lambda}^{\infty}\frac{d\sigma_o(E')}{E'-E}-\frac{1}{\pi}\int_{\Lambda}^{\infty}\frac{(E'-\Lambda)^{1/2}}{E'-E}\int_0^{\Lambda}\frac{d\sigma_o(\mu)}{(\Lambda-\mu)^{1/2}(E'-\mu)}dE'. \tag{7.3}$$

With the help of F_{Λ}, we can estimate $E_i^e-E_i^{\Lambda}$ in (7.2a)–(7.2c): at $E=E_i^{\Lambda}$, $R_o(E)=F_{\Lambda}(E)$; so, from a mean value theorem,

$$|E_i^e-E_i^{\Lambda}|<|F_{\Lambda}(E_i^{\Lambda})|/\min_{E_i\leqslant E\leqslant E_{i+1}}\frac{dR_o}{dE}. \tag{7.4}$$

Consequently, the task ahead of us is to obtain convenient upper bounds for $F_{\Lambda}(E)$, dF_{Λ}/dE, $d^2R_o/dE^2|_{\bar{E}_i}$ and lower bounds for dR_o/dE, $dR_o^{(N)}/dE$ in an interval $E_i\leqslant E\leqslant E_{i+1}$.

The bounds for $F_{\Lambda}(E)$ and dF_{Λ}/dE are obtained from (7.3) by integration by parts and use of (6.11), (6.20), and (6.22). It is shown in Appendix D that a constant exists, so that

$$|F_{\Lambda}(E)|\leqslant\text{const}\times\Lambda^{1/2}\frac{E_{N+1}-E_N}{\Lambda-E},\quad E\leqslant\Lambda. \tag{7.5}$$

Similarly,

$$\frac{dF_{\Lambda}}{dE}(E)\leqslant\text{const}\times\Lambda^{1/2}\frac{E_{N+1}-E_N}{(\Lambda-E)^2}. \tag{7.6}$$

Estimates for dR_o/dE are obtained by differentiating the subtracted dispersion relation for R_o:

$$\frac{dR_o}{dE}=\sum_{i=1}^{\infty}\frac{(u_{i,o}')^2(0)}{(E_i-E)^2}. \tag{7.7}$$

Clearly ($M\neq N$)

$$\min_{E_M<E<E_{M+1}}\frac{dR_o}{dE}\geqslant\sum_{i\leqslant M}\frac{(u_{i,o}')^2(0)}{(E_i-E_{M+1})^2}+\sum_{i\geqslant M+1}\frac{(u_{i,o}')^2(0)}{(E_i-E_M)^2}=R_{1,0}+R_{2,0}. \tag{7.8}$$

Using (6.1), we get, e.g., for $R_{1,0}$,

$$R_{1,0}(E)\geqslant\sum_{i\leqslant M}\frac{E_i^{1/2}(E_{i+1}-E_i)}{(E_i-E_{M+1})^2}-\sum_{i\leqslant M}\frac{E_{i+1}-E_i}{(E_i-E_{M+1})^2}\frac{C}{E_i^{\epsilon/2}}=R_{11}-R_{12}. \tag{7.9}$$

But, from (6.2), $E_M\leqslant E\leqslant E_{M+1}$,

$$R_{11}(E)\geqslant\text{const}\sum_{i\leqslant M-1}\frac{E_{i+1}^{1/2}(E_{i+1}-E_i)}{(E_{i+1}-E_{M+1})^2}+\frac{E_M^{1/2}}{E_{M+1}-E_M}$$
$$>\text{const}\int_0^{E_M}\frac{\mu^{1/2}\,d\mu}{(\mu-E_{M+1})^2}+\frac{E_M^{1/2}}{E_{M+1}-E_M}\geqslant C\frac{\sqrt{E_M}}{E_{M+1}-E_M}. \tag{7.10}$$

By majorizing $R_{12}(E)$ by an integral, one checks that

$$R_{12}(E)\leqslant\text{const}\frac{1}{E_{M+1}-E_M},\quad E_M\leqslant E\leqslant E_{M+1} \tag{7.11}$$

which shows that, for M sufficiently large, a constant C_0 exists so that

$$R_{1,0}(E)\geqslant C_0(\sqrt{E_M})/(E_{M+1}-E_M),\quad E_M\leqslant E\leqslant E_{M+1}. \tag{7.12}$$

Taking into account $R_{2,0}$, Eq. (7.12) will just alter the value of C_0. Differentiating the subtracted dispersion relation for $R_o^{(N)}(E)$ and using the fact that its imaginary part along the cut is positive, whereas the spectral function for $E<\Lambda$ coin-

cides with that of $R_o(E)$, one gets

$$\frac{dR_o^{(N)}(E)}{dE} \geqslant C\frac{\sqrt{E_M}}{E_{M+1}-E_M}, \quad E_M \leqslant E \leqslant E_{M+1}. \tag{7.13}$$

By means of (7.5)–(7.13), we are now in a position to estimate the last two terms in (7.2). Indeed, consider (7.2b) first:

$$S_2^{\Lambda} \leqslant \text{const} \times E_N^{1/2}(E_{N+1}-E_N)\sum_{n=1}^{N}\frac{(E_{n+1}-E_n)^2}{(\Lambda-E_n)^2E_n}. \tag{7.14}$$

This sum is parted into two: one of them, $S_2^{\Lambda,1}$, limited to kN, for some $0<k<1$, and a second one, $S_2^{\Lambda,2}$, from kN to N. Clearly,

$$S_2^{\Lambda,1} \leqslant \text{const}\frac{E_{N+1}-E_N}{(E_N)^{3/2}}\sum_{n=1}^{kN}\frac{(E_{n+1}-E_n)^2}{E_n}. \tag{7.15}$$

The last sum behaves, for large N, like $\Sigma_{n=1}^{kN} n^{\alpha-2}$, which is convergent (for $\alpha<1$).[22] So, $S_2^{\Lambda,1}\to 0$, as $\Lambda\to\infty$. The second sum can be evaluated using (6.2):

$$\begin{aligned} S_2^{\Lambda,2} &< E_N^{1/2}(E_{N+1}-E_N)\sum_{n=kN}^{N}\frac{(E_{n+1}-E_n)^2}{E_n(\Lambda-E_n)^2} \\ &< E_N^{1/2}(E_{N+1}-E_N)\times\text{const}\times\sum_{n=k}^{N}\frac{n^{\alpha-2}}{((N+1)^{\alpha}-n^{\alpha})^2}. \end{aligned} \tag{7.16}$$

For k appropriately chosen (sufficiently close to 1 if α is small), the last sum can be majorized like

$$\begin{aligned} S_2^{\Lambda,2} &< \text{const}\times\Big((N^{\alpha-2}/((N+1)^{\alpha}-N^{\alpha})^2 \\ &\quad + \int_{kN}^{N}\frac{x^{\alpha-2}}{((N+1)^{\alpha}-x^{\alpha})^2}\,dx\Big)E_N^{1/2}(E_{N+1}-E_N). \end{aligned} \tag{7.17}$$

Estimating the integrals with care at the upper limit, one gets that

$$S_2^{\Lambda,2} < \text{const}\times(E_{N+1}-E_N)/E_N^{1/2} \tag{7.18}$$

which vanishes as $\Lambda\to\infty$.

The sum S_3^{Λ} is majorized following the same steps as for S_2^{Λ}. One gets

$$\begin{aligned} S_3^{\Lambda} &< \text{const}\times E_N^{1/2}(E_{N+1}-E_N)\sum_{n=1}^{N}\frac{(E_{n+1}-E_n)^2}{E_n^{3/2}(\Lambda-E_n)} \\ &< \text{const}\times(E_{N+1}-E_N)\frac{\ln N}{N} \end{aligned} \tag{7.19}$$

and the last sum vanishes, as $N\to\infty$, for all $\alpha<2$.

We are now left with S_1^{Λ}. It is easy to obtain a bound on d^2R_o/dE^2 using (6.1), $E_M \leqslant \bar{E}_M \leqslant E_{M+1}$, $M\neq N$:

$$\begin{aligned} \left|\frac{d^2R_o}{dE^2}\right|_{\bar{E}_M} &< \text{const}\times E_M^{1/2}(E_{M+1}-E_M) \\ &\quad\times\left(\frac{1}{(E_{M+1}-\bar{E}_M)^3}+\frac{1}{(\bar{E}_M-E_M)^3}\right) \\ &\quad + \bar{E}_M\left(\frac{1}{(E_{M+1}-\bar{E}_M)^2}+\frac{1}{(\bar{E}_M-E_M)^2}\right). \end{aligned} \tag{7.20}$$

The majorization (7.20) is obtained by writing d^2R/dE^2 as a sum over positive terms (coming from $E_i \leqslant E_M$) and negative ones (from $E_i \geqslant E_{M+1}$) and simply finding upper bounds for the absolute values of the two sums, using (6.1). Equation (7.20) is still not useful, since we do not yet know how close $\bar{E}_M$ can be to E_M or E_{M+1}. Since $\bar{E}_M$ lies between E_M^e and E_M^{Λ} the question is how close to E_M or E_{M+1} the latter can lie. We can obtain a coarse estimate of this, sufficient for our purposes, as follows: we estimate the unique solutions of the equations

$$R(E)=0 \text{ and } R(E)=\tilde{F}_{\Lambda}(E), \quad E_M<E<E_{M+1} \tag{7.21}$$

[where $\tilde{F}_{\Lambda}(E)$ is the majorization of F_{Λ} on the right-hand side of (7.5)] by writing, e.g., for the former $[R(E_M^e)=0]$,

$$\begin{aligned} &\frac{E_M^e}{E_M}\frac{(u'_{M,o})^2(0)}{E_M-E_M^e}+\frac{E_M^e}{E_{M+1}}\frac{(u'_{M+1,o})^2(0)}{E_{M+1}-E_M^e} \\ &\quad = -\sum_{i=1}^{M-1}\frac{E_M^e(u'_{i,o})^2(0)}{E_i(E_i-E_M^e)} \\ &\quad\quad -\sum_{i=M+2}^{\infty}\frac{E_M^e(u'_{i,o})^2(0)}{E_i(E_i-E_M^e)}-\bar{C}\equiv -\tilde{R}(E_M^e) \end{aligned} \tag{7.22}$$

($\bar{C}$ is a subtraction constant) and finding upper bounds for the quantities on the right. We multiply (7.22) by $(E_{M+1}-E_M^e)E_M/((u'_{M,o})^2(0)E_M^e)$ and get a bound on the ratio $(E_{M+1}-E_M^e)/(E_M^e-E_M)$:

$$\begin{aligned} &\frac{E_{M+1}-E_M^e}{E_M^e-E_M} \\ &\quad < \max_{E_M<E<E_{M+1}}|\tilde{R}(E)|\frac{E_{M+1}-E_M}{(u'_{M,o})^2(0)}\frac{E_M}{E_M^e}. \end{aligned} \tag{7.23}$$

The quantity $\max_{E_M<E<E_{M+1}}|\tilde{R}(E)|$ can be majorized in a straightforward way[23] by $\text{const}\times E_M^{1/2}\ln E_M$, if one uses (6.1). This shows that

$$(E_{M+1}-E_M^e)/(E_M^e-E_M) < \text{const}\times\ln E_M. \tag{7.24}$$

Clearly, by multiplying (7.22) by $(E_M-E_M^e)E_{M+1}/((u'_{M+1,o})^2(0)E_M^e)$, we obtain an inequality similar to (7.24) for the ratio $(E_M^e-E_M)/(E_{M+1}-E_M^e)$. We need estimates like (7.24) also for the solutions $E_M^{\Lambda,1}$ of the second equation (7.21) (it is true that $|E_M^{\Lambda,1}-E_M^e| > |E_M^{\Lambda}-E_M^e|$). However, from (7.5) one sees that the quantity $F_{\Lambda}(E)/\sqrt{E}$, which is relevant for estimates similar to (7.23) is bounded by a constant for all $E<E_N$. Thus, the term coming from $R(E)$ will be dominant, and $E_M^{\Lambda,1}$ satisfies the same inequality (7.24) as E_M^e. It follows that $\bar{E}_M$ also obeys it. We are now in a position to show that (7.2a) tends to zero, as $\Lambda\to\infty$. Indeed, replacing $|E_i^e-E_i^{\Lambda}|$ by its upper bound (7.4), using (7.20), with the denominators replaced by (7.24), and the lower bound (7.12), one obtains

$$S_1^{\Lambda} < \text{const}\times E_N^{1/2}(E_{N+1}-E_N)\sum_{i=1}^{N}\frac{E_{i+1}-E_i}{E_i(E_{N+1}-E_i)}\ln^2 E_i. \tag{7.25}$$

We then use (6.2) and proceed as for S_2^{Λ}, S_3^{Λ}: a first sum $S_1^{\Lambda,1}$, extending to kN, $0<k<1$, can be majorized by

$$S_1^{\Lambda,1} < \text{const}\frac{E_{N+1}-E_N}{\sqrt{E_N}}\sum_{n=1}^{kN}\frac{\ln^2 n}{n}. \tag{7.26}$$

As $N\to\infty$, $S_1^{\Lambda,1}$ vanishes, since the sum in (7.26) diverges only logarithmically, whereas the factor in front goes to zero like a power. The remaining sum $S_1^{\Lambda,2}$ can be evaluated by an

integral which, after scaling, reads

$$S_1^{\Lambda,2} < \text{const}\,\frac{E_{N+1}-E_N}{\sqrt{E_N}}\,(\ln E_N)^2\int_k^{1-1/(N+1)}\frac{dx}{1-x^\alpha}\,. \tag{7.27}$$

Again, as $N\to\infty$, $S_1^{\Lambda,2}$ vanishes, since the integral diverges only logarithmically at the upper end. This concludes the proof.

Let us notice that all majorizations in this section are valid for $0<\alpha<2$.

VIII. CONCLUSIONS

In the previous sections, we have proved the convergence of the reflectionless approximants proposed in Refs. 1–5, constructed from the first N levels (positions and residua) of a confining potential, to the potential itself, as $N\to\infty$, if the latter belongs to a restricted class (5.1) or (6.1) and (6.2). It might be worth pointing out a difference between previous work on this problem[1,4] and the present one: we do not try to prove the convergence of the approximants starting from a given sequence of even- and odd-parity levels, but rather from the spectral function of the radially symmetrical potential, for correct or "wrong" boundary conditions. This is, in the opinion of the author, of more physical relevance (since the latter are the real "input" data) and, presumably, also mathematically more easily tractable. The problem of proving convergence starting from a given sequence of even- and odd-parity levels and, further, of characterizing such possible sequences is, to the knowledge of the author, unsolved in general (the latter part is solved in Ref. 10 for a restricted class of potentials ($\Omega_{1/4}$) of potentials).

The results that we obtained seem to be close to what is desired for the case of "wrong" boundary conditions. For the correct boundary conditions, the restriction $0<\alpha<1$ in Eq. (6.2) might appear unsatisfactory, particularly since numerical study[1,4] shows good convergence for approximants of the harmonic oscillator potential. However, we repeat, the problem of this paper is different from the one investigated in Ref. 1 where the even- and odd-parity levels are taken as prescribed. For the present author, the main difficulty in proving convergence for $\alpha>1$ lies in the presence of the root $(\Lambda-E_i)^{1/2}$ in the denominator of the spectral function (4.15) of the approximant. To prove that the difference between those parts of the kernels corresponding to $E>\Lambda$ has a vanishingly small norm as $\Lambda\to\infty$, the author found that a majorization like (6.14) was essential. The latter holds only if $0<\alpha<1$. It is possible that the reduction to the case of wrong boundary conditions can be short circuited for the class of potentials with spectral functions delimited by (6.1) and (6.2), if one starts from first principles, as in Ref. 10.

Most of this paper was concerned with a proof that reflectionless approximants converge to the true potential. One of the difficulties was the construction of the spectral measure on the half-line for such potentials (Secs. III and IV). One could therefore wonder whether these superpositions of solitons are likely to be in some sense preferred approximants (i.e., whether a "better" convergence can be achieved using them, rather than other potentials). The present study fails unfortunately to find any special quality that might justify such a preference. The author believes that there exists a large class of potentials [see Eq. (2.19)] for which convergence in the sense of (1.1) can be proved, at least for the "wrong" boundary conditions. It may certainly be true that the reflectionless potentials achieve more than this and converge, e.g., even pointwise to $V(x)$. Moreover, the very elegant algebra accompanying these soliton approximants,[4] as well as the simple Jost functions that belong to them [Eq. (3.25)] lead one to the belief that they do have a special significance for this kind of problem, although one that is unclear at present.

ACKNOWLEDGMENTS

The author wishes to thank H. Grosse and F. Lambert for useful discussions about the problems of this paper.

APPENDIX A: THE CONTINUITY OF $\|(1-F^0)^{-1}\|_r$

According to Ref. 9, $(1-F_r^0)^{-1}$ exists and is bounded for all r. It is then easy to prove that $\|(1-F_r^0)^{-1}\|_r$ is a continuous function of r, $0<r<r_0$ by following Lemma 1.3.1 of Ref. 10. We change variables $s=rs'$, $u=ru'$ in the kernel $F(s,u)$ of F^0, so that we can regard the latter as an operator from $\mathbb{C}(0,1)$ to $\mathbb{C}(0,1)$. The norm (2.13) of F^0 stays unchanged (the kernel is modified to $rF(rs',ru')$) and one can check that $\|F_r^0-F_{r_1}^0\|$ can be made arbitrarily small, for r_1 close to r. We can then evaluate $\|\Delta\|\equiv\|(1-F_r^0)^{-1}-(1-F_{r_1}^0)^{-1}\|$, the norm being understood in $\mathbb{C}(0,1)$ according to (2.12). We write

$$\Delta=(1+(1-F_r^0)^{-1}(F_r^0-F_{r_1}^0))^{-1}\,(1-F_r^0)^{-1}-(1-F_r^0)^{-1} \tag{A1}$$

and expand the first brackets in a Neumann series for r close enough to r_1, so that

$$\epsilon_1=\epsilon\|(1-F_r^0)^{-1}\|<1,\quad \epsilon\equiv\|F_r^0-F_{r_1}^0\|. \tag{A2}$$

We get

$$\|\Delta\|<(\epsilon_1/(1-\epsilon_1))\|(1-F_r^0)^{-1}\|. \tag{A3}$$

Since $1-F_r^0$ has a bounded inverse, we see from (A3) that we can make $\|\Delta\|$ arbitrarily small for r_1 close enough to r. But,

$$\|\Delta\|>|\|(1-F_r^0)^{-1}\|-\|(1-F_{r_1})^{-1}\||\| \tag{A4}$$

and

$$\|(1-F_r^0)^{-1}\|=\|(1-F^0)^{-1}\|_r \tag{A5}$$

(by the definition of the norm). Equations (A4) and (A3) prove our assertion.

APPENDIX B: THE UNIFORM CONVERGENCE OF $\Omega^{(N)}(t)$

We wish to show that, if $\sigma_e(E)$ is the difference between the spectral measure of a confining potential with a continuous derivative and the free measure for "wrong" boundary conditions, then the limit

$$\lim_{N\to\infty}\int_0^N\cos(t\sqrt{E})\,d\sigma_e(E)=\Omega(t) \tag{B1}$$

is uniform with respect to t, $0\leqslant t\leqslant t_0$.

We shall use to this end the following statements of

Ref. 10.

(i) The functions

$$\Delta_N(t)\equiv\frac{1}{\pi}\int_0^N\cos(t\sqrt{E})\frac{dE}{\sqrt{E}}-\int_0^N\Psi_1(t,E)\,d\rho_e(E) \tag{B2}$$

converge uniformly to zero as $N\to\infty$, $0\leqslant t\leqslant t_0$, $[\Psi_1(t,E)\equiv-\Psi(t,E)$ of Eq. (3.2)].

(ii) The following representation holds (inverse of the Povzner–Levitan representation[15]):

$$\cos(t\sqrt{E})=\Psi_1(t,E)+\int_0^t H(t,x)\Psi_1(x,E)\,dx, \tag{B3}$$

where $H(t,x)$ has two continuous derivatives with respect to t and x.

(iii) $H(0,0)=0, H(t,0)$ has two derivatives with respect to t for $t\geqslant 0$.

(iv) $\Omega(t)=H(t,0)$. (B4)

With this, following Ref. 10, one checks that, for large N,

$$\Omega^{(N)}(t)\equiv\int_0^N\cos(t\sqrt{E})\,d\sigma_e(E)$$
$$=\frac{2}{\pi}\int_0^t dx\,H(t,x)\frac{\sin x\sqrt{N}}{x}+o(1), \tag{B5}$$

where $o(1)$ is uniform with respect to $t, 0\leqslant t\leqslant t_0$.

Then,

$$|\Omega(t)-\Omega^{(N)}(t)|$$
$$\leqslant\left|H(t,0)-\frac{2}{\pi}\int_0^t H(t,x)\frac{\sin x\sqrt{N}}{x}\,dx\right|+o(1)$$
$$\equiv T(t)+o(1), \tag{B6}$$

$$T(t)\leqslant\left|H(t,0)\left(1-\frac{2}{\pi}\int_0^t\frac{\sin x\sqrt{N}}{x}\,dx\right)\right|$$
$$+\frac{2}{\pi}\left|\int_0^t(H(t,x)-H(t,0))\frac{\sin x\sqrt{N}}{x}\,dx\right|$$
$$=T_1(t)+T_2(t). \tag{B7}$$

Consider $T_1(t)$; it follows from (iii) that, given ϵ, we can choose $\bar{t}_0(\epsilon)$, so that, for $0\leqslant t\leqslant\bar{t}_0$, $H(t,0)<\epsilon$. For $t\leqslant t_0$,

$$T_1(t)\leqslant\epsilon\left(1+\frac{2}{\pi}\left|\int_0^t\frac{\sin x\sqrt{N}}{x}\,dx\right|\right)\leqslant C_1\epsilon. \tag{B8}$$

For $t\geqslant t_0$,

$$T_1(t)\leqslant\sup_{\bar{t}_0\leqslant t\leqslant t_0}|H(t,0)|\left|\frac{2}{\pi}\int_{t\sqrt{N}}^\infty\frac{\sin u}{u}\,du\right|. \tag{B9}$$

If $\alpha(\epsilon)$ is defined such that, for $\alpha_1>\alpha$,

$$\left|\frac{2}{\pi}\int_{\alpha_1}^\infty\frac{\sin u}{u}\,du\right|<\epsilon, \tag{B10}$$

it is clear it is enough to choose $N>\alpha^2(\epsilon)/\bar{t}_0^2(\epsilon)$, to make (B9) smaller than $\epsilon\sup_{0<t<t_0}|H(t,0)|$, for all $\bar{t}_0\leqslant t\leqslant t_0$. This makes however $T_1(t)$ uniformly smaller than const$\times\epsilon$, for $0\leqslant t\leqslant t_0$.

We now turn to $T_2(t)$. We denote $\tilde{G}(t,x)=(H(t,x)-H(t,0))/x$. The function $\tilde{G}(t,x)$ is a continuous function of t,x on the compact set $0\leqslant x\leqslant t\leqslant t_0$ and therefore uniformly continuous. Let $C=\sup_{0<x<t<t_0}|\tilde{G}(t,x)|$. Further, given ϵ, there exists a set of $N(\epsilon)$ points t_i, $0<t_1<t_2<\cdots<t_0\equiv t_{N(\epsilon)}$, with the property that, for any x, t, $t_i\leqslant t\leqslant t_{i+1}$, $0\leqslant x\leqslant t$, both

$$|\tilde{G}(t,x)-\tilde{G}(t_i,x)|\leqslant\epsilon,\quad|t_{i+1}-t_i|\leqslant\epsilon. \tag{B11}$$

Also, all the integrals $I_j\equiv\int_0^{t_j}\tilde{G}(t_j,x)\sin xN^{1/2}\,dx, j=1,2,\ldots,N$, tend to zero as $N\to\infty$, by the Riemann–Lebesgue lemma; consequently, given ϵ, we can find $\bar{N}$ so that all of them are smaller than ϵ, for $N>\bar{N}$. But, for $t_k\leqslant t\leqslant t_{k+1}$,

$$T_{2N}(t)=\int_0^t\tilde{G}(t,x)\sin xN^{1/2}\,dx$$
$$=\int_0^{t_k}(\tilde{G}(t,x)-\tilde{G}(t_k,x))\sin xN^{1/2}\,dx$$
$$+\int_0^{t_k}\tilde{G}(t_k,x)\sin xN^{1/2}\,dx+\int_{t_k}^t\tilde{G}(t,x)\sin xN^{1/2}\,dx, \tag{B12}$$

and it follows from the above that

$$|T_{2N}(t)|<\epsilon t_0+\epsilon+\epsilon C \tag{B13}$$

for N sufficiently large. This concludes our proof. This reasoning is mapped after Ref. 24, Sec. 698.

Notice, for the boundary condition $u'(0)+hu(0)=0$, $h\neq 0$, one gets a supplementary term in $T_1(t)$:

$$T_1'(t)\equiv h\int_t^\infty\frac{\sin x\sqrt{N}}{x}\,dx \tag{B14}$$

which does not tend to zero uniformly in t, as $N\to\infty$.

APPENDIX C: COMPLEMENT TO SECTION IV

We prove in this Appendix the estimates (6.11) and (6.13).

We start with (6.11); let E_M be the energy level closest to μ from above; then $(0<k<1, k_1>1)$,

$$|\sigma_1(\Lambda_N)-\sigma_1(\mu)|\leqslant|\sigma_1(\Lambda_N)-\sigma_1(E_{N+1}-0)|$$
$$+|\sigma_1(E_{N+1}-0)-\sigma_1(E_M-0)|$$
$$+|\sigma_1(E_M-0)-\sigma_1(\mu)|. \tag{C1}$$

The first and third term are easily majorized by

$$|\sigma_1(\Lambda_N)-\sigma_1(E_{N+1}-0)|\leqslant\text{const}\times E_N^{1/2}(E_{N+1}-E_N), \tag{C2a}$$

$$|\sigma_1(E_M)-\sigma_1(\mu)|<\text{const}\times E_M^{1/2}(E_{M+1}-E_M)$$
$$<\text{const}\times E_N^{1/2}(E_{N+1}-E_N), \tag{C2b}$$

where use has been made of (6.2) and $\mu>k\Lambda_N$ (or $\mu<k_1\Lambda_N$). The second term in (C1) is estimated using Eq. (6.2):

$$|\sigma_1(E_{N+1}-0)-\sigma_1(E_M-0)|\leqslant\sum_{n=M}^N\frac{1}{2\sqrt{E_n}}(E_{n+1}-E_n)^2$$
$$<\text{const}\times(\Lambda_N^{[3/2-1/\alpha]_+}-\mu^{[3/2-1/\alpha]_+})$$
$$<\text{const}\times\Lambda_N^{[3/2-1/\alpha]_+-1}(\Lambda_N-\mu). \tag{C3}$$

So, (6.11) is valid for all $\alpha<2$, $\tilde{\epsilon}=1-[\frac{3}{2}-1/\alpha]_+$. Consider next $\sigma_\Lambda^{(k)}(\mu)$, Eq. (6.12). We integrate by parts and use Eq.

(6.3) and (6.11), if $\mu < E_N$:

$$|\sigma_\Lambda^{(k)}(\mu)| < \left| \frac{\sigma_1(k\Lambda) - \sigma_1(\Lambda)}{(\Lambda(1-k))^{1/2}} \right| + \left| \frac{\sigma_1(\mu) - \sigma_1(\Lambda)}{(\mu - \Lambda)^{1/2}} \right| + \frac{1}{\pi} \int_{k\Lambda}^{\mu} \frac{|\sigma_1(\nu) - \sigma_1(\Lambda)|}{(\Lambda - \nu)^{3/2}} d\nu$$
$$< \text{const} \times \Lambda^{-\tilde{\epsilon}} + \text{const} \times \Lambda^{-\tilde{\epsilon}} (\Lambda - \mu)^{1/2} + \text{const} \times E_N^{1/2} (E_{N+1} - E_N)^{1/2} + \text{const} \times \Lambda^{-\tilde{\epsilon}} \int_{k\Lambda}^{\mu} \frac{d\nu}{(\Lambda - \nu)^{1/2}} + \text{const} \times E_N^{1/2} (E_{N+1} - E_N)^{1/2}$$
$$< \text{const} \times (\Lambda^{1/2 - \tilde{\epsilon}} + E_N^{1/2} (E_{N+1} - E_N)^{1/2}) . \quad \text{(C4)}$$

If $\Lambda > \mu > E_N$, only the last term is relevant.

APPENDIX D: COMPLEMENT TO SECTION VII

In this Appendix, we justify the majorization (7.5). We write

$$F_\Lambda(E) = F_{\Lambda 1}(E) - F_{\Lambda 2}(E), \quad \text{(D1)}$$

$$F_{\Lambda 1}(E) = \int_\Lambda^\infty \frac{d\sigma(\mu)}{\mu - E}, \quad \text{(D2)}$$

$$F_{\Lambda 2}(E) = \int_\Lambda^\infty \frac{dE'(E' - \Lambda)^{1/2}}{E' - E} \int_0^\Lambda \frac{d\sigma(\mu)}{(\Lambda - \mu)^{1/2}(E' - \mu)}. \quad \text{(D3)}$$

We consider first $F_{\Lambda 1}(E)$ and divide $\sigma(\mu)$ into two as in (6.5). The part $F_{\Lambda 1}^{(2)}$ corresponding to $\sigma_2(\mu)$ can be majorized by an integral over $|d\sigma_2(\mu)|$ as in (6.21), and so

$$|F_{\Lambda 1}^{(2)}(E)| \leqslant \Lambda^{-\epsilon} \ln \frac{E_{N+1} - E}{\Lambda}. \quad \text{(D4)}$$

The part generated by $\sigma_1(\mu)$ is again divided into two: for $\Lambda \leqslant \mu \leqslant k_1\Lambda$ and $k_1\Lambda \leqslant \mu \leqslant \infty$, $k_1 > 1$. The second part is easily majorized by $\text{const} \times \Lambda^{-\tilde{\epsilon}}$, after integration by parts. For the first one, we write

$$F_{\Lambda 1,1}^{(1)} = \frac{\sigma_1(\mu) - \sigma_1(\Lambda)}{\mu - E} \Big|_\Lambda^{k_1\Lambda} + \int_\Lambda^{k_1\Lambda} \frac{\sigma_1(\mu) - \sigma_1(\Lambda)}{(\mu - E)^2} d\mu \quad \text{(D5)}$$

and use (6.11) to majorize the numerator of the integrand. (The first term disappears like $\text{const} \times \Lambda^{-\tilde{\epsilon}}$.)

$$\left| \int_\Lambda^{k_1\Lambda} \frac{\sigma(\mu) - \sigma(\Lambda)}{(\mu - E)^2} d\mu \right| < \int_\Lambda^{k_1\Lambda} \frac{\Lambda^{-\tilde{\epsilon}}(\mu - \Lambda)}{(\mu - E)^2} d\mu + \int_{E_{N+1}}^{k_1\Lambda} \frac{\Lambda^{1/2}(E_{N+1} - E_N)}{(\mu - E)^2} d\mu + \int_\Lambda^{E_{N+1}} \frac{\Lambda^{1/2}(\mu - \Lambda)}{(\mu - E)^2} d\mu. \quad \text{(D6)}$$

It is easy to check that the dominant contributions come from the second and third term, so that

$$F_{\Lambda 1}^{(1)} < \text{const} \times E_N^{1/2} \frac{E_{N+1} - \Lambda}{E_{N+1} - E} + \text{const} \times E_N^{1/2} \ln \frac{E_{N+1} - E}{\Lambda - E}. \quad \text{(D7)}$$

For $F_{\Lambda 2}(E)$, we use the majorizations (6.20) and (6.22). The contribution of $\tilde{N}_1$ is negligible; using $(E' - \Lambda)^{1/2}/(E' - E_N) < 1/(E' - \Lambda)^{1/2}$, for $E' > \Lambda$, one sees that the part corresponding to $\tilde{N}_{21}$ is majorized by $(\Lambda - E_N)^{1/2}/(\Lambda - E)^{1/2}$. From $\tilde{N}_{22}$, one obtains a bound like $C(E_{N+1} - E_N)^{1/2} \Lambda^{1/2}/(\Lambda - E)^{1/2}$, whereas $\tilde{N}_{23}$ contributes $\ln(1 + (\Lambda - E_N)^{1/2}/(\Lambda - E)^{1/2})$. Summing up, one concludes that (7.5) is correct.

[1] H. B. Thacker, C. Quigg, and J. L. Rosner, Phys. Rev. D **18**, 274 (1978).
[2] H. B. Thacker, C. Quigg, and J. L. Rosner, Phys. Rev. D **18**, 287 (1978).
[3] C. Quigg, J. L. Rosner, and H. B. Thacker, Fermilab-Pub-79/52-THY.
[4] J. L. Schonfeld, W. Kwong, J. L. Rosner, C. Quigg, and H. B. Thacker, Ann. Phys. (N. Y.) **128**, 1 (1980).
[5] C. Quigg and J. L. Rosner, Fermilab-Pub-81/13-THY; P. Moxhay, J. L. Rosner, and C. Quigg, Fermilab-Pub-81/14-THY.
[6] R. P. van Royen and V. F. Weisskopf, Nuovo Cimento A **50**, 617 (1967).
[7] H. Grosse and A. Martin, Nucl. Phys. B **148**, 413 (1979).
[8] H. Grosse and A. Martin, Phys. Rep. **60**, 343 (1980).
[9] I. M. Gel'fand and B. M. Levitan, Am. Math. Soc. Transl. Ser. **2**, 1 (1955).
[10] G. M. Gasymov and B. M. Levitan, Russian Math. Surveys **19**, 1 (1964).
[11] B. M. Levitan and I. S. Sargsyan, Russian Math. Surveys **15**, 1 (1960).
[12] K. Jörgens, *Lineare Integraloperatoren* (Teubner, Stuttgart, 1970).
[13] B. M. Levitan, Izv. Akad. Nauk SSSR, Ser. Mat. **17**, 331 (1953).
[14] Taking the derivatives inside the integral in (2.21) cannot be easily justified for confining potentials: the resulting integrals are not absolutely convergent. Reference 10 justifies this for the case $u'(0) = 0$.
[15] K. Chadan and P. C. Sabatier, *Inverse Problems in Quantum Scattering Theory* (Springer, Berlin, 1977).
[16] Reference 15 treats only the case of $F_o(k)$, but it is easy to generalize it to $F_c(k)$.
[17] R. D. Richtmyer, *Principles of Advanced Mathematical Physics* (Springer, Berlin, 1978).
[18] We drop the subscript e on σ_e in the rest of this section.
[19] C. Quigg and J. L. Rosner, Phys. Rep. **56**, 167 (1979).
[20] Compare Ref. 14 concerning the Gel'fand–Levitan equation with correct boundary conditions.
[21] For simplicity, we drop from now on the subscript N in Λ_N. It is understood that $\Lambda \to \infty$ through a sequence of values obeying (6.3).
[22] This is not necessary for the argument.
[23] One follows the procedure used for majorizing d^2R/dE^2; i.e., one parts $\tilde{R}(E)$ into negative and positive contributions and places upper bounds on the absolute values of both.
[24] G. M. Fichtenholz, *Differential und Integralrechnung* (Deutscher, Berlin, 1963).

On the finiteness of the number of discrete eigenvalues in neutron transport theory[a)]

Dan G. Cacuci
Oak Ridge National Laboratory, Oak Ridge, Tennessee 37830

(Received 22 March 1982; accepted for publication 11 June 1982)

This work examines the finiteness of the number of discrete eigenvalues occurring in neutron slowing down and transport in an infinite medium consisting of a single nuclide scattering elastically, isotropically, and with energy-independent cross sections. This finiteness is related to convergence conditions for the Legendre moments of the scattering function. It is shown that the convergence criterion obtained by Case for one-velocity neutron transport is not satisfied by hydrogen at large lethargies. A new convergence criterion, weaker than that of Case, is derived and shown to be satisfied by all nuclides, including hydrogen, at all lethargies. Although derived within the more general framework of energy-dependent transport, the new convergence criterion also holds for the Legendre moments of the scattering function in one-velocity transport.

PACS numbers: 28.20.Lh, 03.65.Nk

I. INTRODUCTION

In the treatment of problems in neutron transport with anisotropic scattering, it is customary to seek solutions by expanding the scattering function in Legendre polynomials. This leads to an eigenvalue problem involving a three-term difference equation. Such equations are typical in the theory of orthogonal polynomials.

The close connection between the theory of orthogonal polynomials and the solutions of the one-velocity linear neutron-transport equation with anisotropic scattering was noted by Inonu.[1] On the other hand, Case[2] showed that the methods of scattering theory are a powerful tool for investigating the properties of orthogonal polynomials. Developing this interplay, Case[2] showed that the discrete eigenvalues arising in one-velocity transport with anisotropic scattering to *all* orders in terms of Legendre polynomials (henceforth referred to as "infinite order anisotropy") are real or purely imaginary, simple, $\geqslant 1$ in absolute magnitude, finite in number, and occur in $\pm$ pairs. (These properties are well known[3] for one-velocity transport with *finite*-order anisotropic scattering.) Among these properties, the finiteness of the number of discrete eigenvalues is mathematically the most difficult[2] to prove, and occurs only if the Legendre moments of the scattering function satisfy certain convergence criteria.

It appears that only Inonu[1] and Case[2] deal explicitly with such convergence criteria—for monoenergetic transport. Inonu[1] considers that the discrete eigenvalues are finite in number provided that the Legendre moments f_n of the scattering function satisfy the convergence criterion $\Sigma_{n=0}^{\infty}(2n+1)|f_n| < \infty$. Case[2] arrived at the convergence criterion $\Sigma_{n=0}^{\infty} n^2|f_n| < \infty$, a criterion he considered to be "about as weak as possible," as a requirement for the finiteness of the number of discrete eigenvalues.

On the other hand, it is known[4,5] that by means of a Laplace transform in lethargy, the Boltzmann equation describing neutron slowing-down and transport under certain conditions (i.e., infinite medium, plane geometry, azimuthally independent scattering, and cross sections that are constant in energy) is transformed into an equation of the same form as the one-velocity equation with infinite-order anisotropy. The only difference[4,5] from the standard[1-3] one-velocity form is that the Legendre coefficients of the (Laplace-transformed) scattering function depend upon an additional parameter, i.e., the transform variable.

Due to the close similarity to the one-velocity equation, it is natural to expect that this Laplace-transformed equation would possess a finite number of discrete eigenvalues only if the Legendre moments of the scattering function satisfy a convergence condition of the same form as derived by Case[2] for monoenergetic transport, i.e., $\Sigma_{n=0}^{\infty} n^2|f_n| < \infty$. As is shown in Sec. II, though, this condition is *not* satisfied by hydrogen in a region that corresponds to the region of large lethargies. This failure to satisfy Case's convergence criterion could imply the occurrence of infinitely many discrete eigenvalues in this region. This provided the motivation for reexamining the question of the finiteness of the number of discrete eigenvalues, and the consequent convergence conditions for the Legendre moments of the scattering function.

Section II succintly reviews the mathematical treatment of the space- and lethargy-dependent Boltzmann equation that describes the constant-cross-section problem. In Sec. IIIA, methods parallel to those used by Case[2] for one-velocity transport are used to arrive at a three-term difference eigenvalue equation. This constitutes the starting point for the new procedure, presented in Sec. IIIB, that leads to a new convergence criterion for the Legendre moments of the scattering function. Finally, Sec. IV summarizes and concludes this work.

II. MOTIVATION: NEUTRON SLOWING DOWN AND TRANSPORT IN THE CONSTANT-CROSS-SECTION PROBLEM

Consider the time-independent homogeneous Boltzmann equation for the case of neutron transport under the following conditions: infinite medium, plane geometry, azimuthally independent scattering, and cross sections that are constant in energy. The explicit form is

[a)] Research sponsored by the U. S. Department of Energy under contract W-7405-eng-26 with Union Carbide.

 0022-2488/82/112205-10$02.50 © 1982 American Institute of Physics

$$\left(\mu\frac{\partial}{\partial x}+1\right)\psi(x,\mu,u)$$
$$=c\int_{\Omega'}\int_{u'}f(\mathbf{\Omega}\cdot\mathbf{\Omega}',u-u')\psi(x,\mu',u')du'd\Omega', \tag{1}$$

where $\psi(x,\mu,u)$ is the angular flux density of the neutrons per unit solid angle and lethargy. The scattering function for isotropic scattering in the center of mass system by a single element of mass A is

$$f(\mathbf{\Omega}\cdot\mathbf{\Omega}',u)-u')=\frac{\exp[-(u-u')]}{8\pi}\frac{(A+1)^2}{A}$$
$$\times\delta\left\{\mathbf{\Omega}\cdot\mathbf{\Omega}'-\frac{A+1}{2}\exp[-(u-u')/2]\right.$$
$$\left.-\frac{A-1}{2}\exp[(u-u')/2]\right\},$$
$$\text{for } (u-u')\in[0,\epsilon] \text{ with } \epsilon=2\ln\frac{A+1}{A-1},$$
$$=0 \quad \text{for } (u-u')>\epsilon. \tag{2}$$

All quantities in Eqs. (1) and (2) are in the standard notation of transport theory and have their usual meaning.[1–4] In this work, the physical problem described by Eqs. (1) and (2) is referred to as the constant-cross-section problem.

Expanding the scattering function in spherical harmonics, applying a one-sided Laplace transform in lethargy and using the convolution theorem, Eq. (1) becomes

$$\left(\mu\frac{\partial}{\partial x}+1\right)\psi(x,\mu,\eta)$$
$$=c\sum_{n=0}^{\infty}\frac{2n+1}{2}g_n(\eta)P_n(\mu)\int_{-1}^{1}P_n(\mu')\psi(x,\mu',\eta)\,d\mu', \tag{3}$$

where

$$\psi(x,\mu,\eta)\equiv\mathscr{L}[\psi(x,\mu,u)], \tag{4}$$

$$g_n(\eta)\equiv\mathscr{L}[f_n(u)]=\int_0^{\epsilon}\exp(-\eta u)f_n(u)du, \tag{5}$$

and $f_n(u)$ is the lethargy-dependent nth Legendre moment of the scattering function.

Equation (3) has the same form[1,2] as the one-velocity transport equation with anisotropic scattering to all orders in terms of Legendre polynomials. (For brevity, this will henceforth be referred to as "infinite order anisotropy".) The only difference from the standard one-velocity form is that the scattering coefficients now depend on an additional parameter, the transform variable η.

Seeking a solution to Eq. (3) of the form

$$\psi(x,\mu,\eta)=\phi(\nu,\mu,\eta)\exp(-x/\nu) \tag{6}$$

and defining

$$\phi_n(\nu,\eta)\equiv\int_{-1}^{1}\phi(\nu,\mu,\eta)P_n(\mu)\,d\mu \tag{7}$$

leads to the following eigenvalue equation for $\phi(\nu,\mu,\eta)$:

$$(\nu-\mu)\phi(\nu,\mu,\eta)=\tfrac{1}{2}c\nu M(\mu,\nu,\eta), \tag{8}$$

where

$$M(\mu,\nu,\eta)=\sum_{n=0}^{\infty}(2n+1)g_n(\eta)P_n(\mu)\phi_n(\nu,\eta). \tag{9}$$

At this stage, the normalization of $\phi(\nu,\mu,\eta)$ is arbitrary. Without loss of generality, it is possible to take

$$\phi_0(\nu,\eta)\equiv\int_{-1}^{1}\phi(\nu,\mu,\eta)\,d\mu=1, \tag{10}$$

and to define

$$\phi_{-1}(\nu,\eta)=0. \tag{11}$$

Using the procedure developed[3] for the one-velocity case, it can be shown[4,5] that (i) the interval $\nu\in(-1,1)$ constitutes the continuous part of the spectrum of the eigenvalue equation [i.e., Eq. (8)], and the continuum eigenfunctions corresponding to these values of ν are

$$\phi(\nu,\mu,\eta)=\frac{c\nu}{2}P\frac{M(\mu,\nu,\eta)}{\nu-\mu}+\lambda(\nu,\eta)\delta(\nu-\mu), \tag{12}$$

where P denotes Cauchy principal value, and

$$\lambda(\nu,\eta)=\frac{\Lambda^{+}(\nu,\eta)+\Lambda^{-}(\nu,\eta)}{2}. \tag{13}$$

In Eq. (13), $\Lambda^{+}(\nu,\eta)$ and $\Lambda^{-}(\nu,\eta)$ are the boundary values of the dispersion function[1–5]

$$\Lambda(\nu,\eta)=1-\frac{c\nu}{2}\int_{-1}^{1}\frac{M(\mu,\nu,\eta)}{\nu-\mu}\,d\mu \tag{14}$$

as the cut $(-1,1)$ is approached from above and below, respectively. (ii) for all values of ν outside the interval $(-1,1)$, the solutions of Eq. (8) are the discrete eigenfunctions

$$\phi(\nu_j,\mu,\eta)=\frac{1}{2}c\nu_j\frac{M(\mu,\nu_j,\eta)}{\nu_j-\mu}, \tag{15}$$

where the discrete eigenvalues ν_j are the roots of the equation

$$\Lambda(\nu_j,\eta)=0. \tag{16}$$

The solutions (e.g., Green's function, spatial moments of the Green's function) to typical problems described by equations of the form of Eq. (3) involve summations over the number of discrete eigenvalues ν_j. But even in monoenergetic transport, it is not obvious that the number of ν_j remains necessarily finite in the case of infinite-order anisotropy. As Case has shown,[2] it is the investigation of conditions for this finiteness that ultimately leads to convergence requirements for the Legendre moments of the scattering function.

To date, it appears that only the works of Inonu[1] and Case[2] on monoenergetic transport make explicit use of convergence conditions for the Legendre moments of the scattering function. Thus, Inonu[1] asserts that the number of discrete eigenvalues remains finite provided that

$$\sum_{n=0}^{\infty}(2n+1)|f_n|<\infty, \tag{17}$$

where f_n is the nth Legendre moment of the scattering function $f(\mathbf{\Omega}\cdot\mathbf{\Omega}')$ in monoenergetic transport. Note, though, that Inonu did not rigorously investigate the behavior of the dispersion function $\Lambda(\nu)$ on the cut $\nu\in(-1,1)$ and, in particular, at the points $\nu=\pm1$. Using methods of scattering theory applied to orthogonal polynomials, Case[2] showed that the

behavior of $\Lambda(\nu)$ on the cut $\nu\in(-1,1)$ requires special attention. Precisely as a consequence of this behavior, Case's method of proof led to

$$\sum_{n=0}^{\infty} n^2 |f_n| < \infty \tag{18}$$

as a requirement for the finiteness of the number of discrete eigenvalues; Case[2] states that this requirement is "about as weak as possible." Obviously, this requirement is stronger than Inonu's [cf. Eq. (17)].

Since the functions $g_n(\eta)$ play in our case [viz. Eq. (3)] the same role as played by the Legendre moments f_n in monoenergetic transport, Case's results, i.e., Eq. (18), implies that the number of discrete eigenvalues ν_j would be finite if

$$\sum_{n=0}^{\infty} n^2 |g_n(\eta)| < \infty. \tag{19}$$

As shown in Appendix A, the above condition is *not* satisfied by hydrogen for values of η such that $\mathrm{Re}(\eta) < \frac{1}{4}$. In particular, the series $\Sigma_{n=0}^{\infty} n^2 |g_n(0)|$ is divergent when $A = 1$ (i.e., hydrogen). In view of asymptotic relationships[6] between a function and its Laplace transform, the region $\eta \to 0$ corresponds to the region $u \to \infty$. Thus, the failure of Eq. (19) in the region of small η could imply, according to Case's arguments, the occurrence of infinitely many discrete eigenvalues at large lethargies in hydrogen.

On the other hand, it is known[4,5,7-12] that at large lethargies the phase space distribution of neutrons is described by the "age-theory" solution. In particular, it has been shown[4,12] that even in hydrogen age-theory is the correct limiting form for the neutron flux density at large lethargies. Thus, there is no obvious physical reason to expect the occurrence of infinitely many discrete eigenvalues in this region. This motivated a reexamination of the question regarding the finiteness of the number of discrete eigenvalues, and the consequent convergence conditions for the Legendre moments of the scattering function.

III. FINITENESS OF THE NUMBER OF DISCRETE EIGENVALUES, AND THE RESULTING CONVERGENCE CRITERION FOR THE LEGENDRE MOMENTS OF THE SCATTERING FUNCTION

A. Preliminaries

Multiplying Eq. (8) by $P_k(\mu)$, integrating over μ from -1 to $+1$, and using the orthogonality and recursion properties of the Legendre polynomials leads to the following three-term recursion relation for the polynomials $\phi_n(\nu,\eta)$:

$$(2n+1)\gamma_n(\eta)\nu\phi_n(\nu,\eta) = (n+1)\phi_{n+1}(\nu,\eta) + n\phi_{n-1}(\nu,\eta), \tag{20}$$

where

$$\gamma_n(\eta) = 1 - cg_n(\eta). \tag{21}$$

Note that the polynomials $\phi_n(\nu,\eta)$ are uniquely determined by Eqs. (20), (10), and (11).

As Case[2] noted, an equivalent form of the eigenvalue problem defined by Eq. (8) is to find those ν for which Eq. (20) has bounded solutions. Those solutions which are merely bounded correspond to the continuum $\nu\in(-1,1)$. Those ν_j for which there are square-summable solutions are the discrete eigenvalues.

Consider, along with Case,[2] the more symmetrical polynomials $\Phi(\nu,n,\eta)$ defined by

$$\Phi(\nu,n,\eta) \equiv \left[\frac{2n+1}{2}\gamma_n\right]^{1/2} \phi_n(\nu,\eta). \tag{22}$$

Substituting Eq. (22) in Eq. (20) gives

$$A(n+1,\eta)\Phi(\nu,n+1,\eta) + A(n,\eta)\Phi(\nu,n-1,\eta) = \nu\Phi(\nu,n,\eta), \tag{23}$$

where

$$A(n,\eta) \equiv n[(2n+1)(2n-1)\gamma_n\gamma_{n-1}]^{-1/2}. \tag{24}$$

Using the asymptotic expressions derived in Appendix B, it can be shown that

$$\lim_{n\to\infty} A(n,\eta) = \tfrac{1}{2} \tag{25}$$

and the above convergence is such that

$$\lim_{n\to\infty} n \ln n[A(n,\eta) - A(\infty,\eta)] = 0. \tag{26}$$

Then for large n, the solutions of Eq. (23) tend to solutions of

$$\tfrac{1}{2}\{\tilde{\Phi}(\nu,n+1,\eta) - \tilde{\Phi}(\nu,n-1,\nu)\} = \nu\tilde{\Phi}(\nu,n,\eta). \tag{27}$$

Two linearly independent solutions of Eq. (27) are

$$\tilde{\Phi}(\nu,n,\eta) = z^{\pm n}, \tag{28}$$

where

$$\nu = \tfrac{1}{2}(z + z^{-1}). \tag{29}$$

Note that Eq. (29) defines a conformal mapping such that the entire ν plane, cut along $[-1,1]$, is mapped onto the open unit disk $K: |z| < 1$, while the unit circle $\Gamma: |z| = 1$ is the image of the cut $\nu\in[-1,1]$ traversed twice. Thus,[2] for $z = e^{i\theta}$, (θ = real), all solutions of Eq. (27) are bounded as $n \to \infty$. It follows that all solutions of Eq. (23) are bounded for

$$\nu = \cos\theta, \quad -1 \leqslant \cos\theta \leqslant 1. \tag{30}$$

In particular, solutions $\Phi(\nu,n,\eta)$ subject to the "initial conditions" obtained by using Eqs. (10) and (11), i.e.,

$$\Phi(\nu,-1,\eta) = 0, \qquad \Phi(\nu,0,\eta) = (\gamma_0/2)^{1/2}, \tag{31}$$

are bounded. Thus, with these initial conditions, the spectrum of Eq. (23) [or, alternately, of Eq. (20)] has a continuous part for $-1 \leqslant \nu < 1$, confirming a fact already known in the literature.[1-5]

Consider, along with Case,[2] two auxiliary solutions $\Phi^{\pm}(\nu,n,\eta)$ of Eq. (23) defined by the boundary conditions

$$\lim_{n\to\infty} |\Phi^{\pm}(\nu,n,\eta) - z^{\pm n}| = 0 \quad \text{for} \quad \begin{matrix} |z| \leqslant 1 \\ |z| \geqslant 1 \end{matrix}. \tag{32}$$

Equation (32) can now be used to define the functions $f_{\pm}(z)$ (i.e., the analog[2] of the Jost function) as

$$f_{\pm}(z) = A(0,\eta)\Phi^{\pm}(\nu,-1,\eta). \tag{33}$$

Then square-summable solutions of Eq. (23) subject to the

initial conditions given by Eq. (31) exist for those z_i, with $|z_i| \leqslant 1$, that are the roots of

$$f_+(z_i) = 0. \tag{34}$$

Thus, the discrete eigenvalues ν_j, which in the usual[1,3-5] treatment of the transport equation are the zeros of $\Lambda(\nu)$ [cf. Eq. (16)], are related to the zeros of $f_+(z)$ within the unit circle $|z| \leqslant 1$. Indeed, Case[2] has shown that $f_+(z)$ and $\Lambda(\nu)$ are related; his proof for the case of monoenergetic transport can be taken over without change to show that, in the present case,

$$\Lambda(\nu,\eta) = \left[\frac{2\pi\gamma_0(\eta)}{z^{-2}-1}\right]^{1/2} f_+(z). \tag{35}$$

Thus,if $f_+(z)$ has a finite number of zeros z_i within the unit circle $|z| \leqslant 1$, then the number of discrete eigenvalues ν_j is also finite.

The investigation of the conditions under which $f_+(z)$ has a finite number of zeros proceeds as follows. The starting point is the same as Case's,[2] i.e., the equation

$$\nu\Psi^+(\nu,n,\eta) - a^0(n+1)\Psi^+(\nu,n+1,\eta) - a^0(n)\Psi^+(\nu,n-1,\nu) = \nu c g_n(\eta)\Psi^+(\nu,n,\eta), \tag{36}$$

where

$$\Psi^+(\nu,n,\eta) \equiv \Phi^+(\nu,n,\eta)/[\gamma_n(\eta)]^{1/2}, \tag{37}$$

$$a^0(n) \equiv n/(4n^2-1)^{1/2}, \tag{38}$$

and Ψ^+ satisfies the condition

$$\lim_{n\to\infty} |\Psi^+(\nu,n,\eta) - z^n| = 0, \quad \text{for } |z| \leqslant 1. \tag{39}$$

Equation (36) is solved by introducing a Green's function $G(\nu,n;m)$ defined as

$$G(\nu,n;m) = \begin{cases} 0 \text{ for } n \geqslant m \\ \dfrac{\sqrt{2}}{f_0^+(z)}\left[\Psi_0^+(m)\Psi_0(n) - \Psi_0(m)\Psi_0^+(n)\right], \quad n \leqslant m, \end{cases} \tag{40}$$

where Ψ_0 and Ψ_0^+ are the "regular" and " + " solutions of Eq. (36) with $c = 0$, i.e.,

$$\Psi_0(n) = \left(\frac{2n+1}{2}\right)^{1/2} P_n(\nu) \tag{41}$$

and

$$\Psi_0^+(n) = (2n+1)^{1/2} f_0^+(z) Q_n(\nu). \tag{42}$$

In Eqs. (41) and (42), P_n are the Legendre polynomials, Q_n are the Legendre functions of the second kind, and

$$f_0^+(z) = \left(\frac{z^{-2}-1}{2\pi}\right)^{1/2}. \tag{43}$$

The solution Ψ^+ can be written as

$$\Psi^+(\nu,n,\eta) = \Psi_0^+(\nu,n) + \sum_{m=n+1}^{\infty} G(\nu,n;m)\nu c g_m(\eta)\Psi^+(\nu,m,\eta). \tag{44}$$

Solving Eq. (44) by iteration gives

$$\Psi^+(\nu,n,\eta) = \sum_{i=0}^{\infty} \Psi^{+(i)}(\nu,n,\eta), \tag{45}$$

where

$$\Psi^{+(0)}(\nu,n) = \Psi_0^+(\nu,n), \tag{46}$$

and

$$\Psi^{+(i)}(\nu,n,\eta) = \sum_{m=n+1}^{\infty} G(\nu,n;m)\nu c g_m \Psi^{+(i-1)}(\nu,m,\eta). \tag{47}$$

Using Eqs. (23), (24), (29), and (37), it is observed that Eq. (33) can be rewritten as

$$f_+(z) = \frac{z+z^{-1}}{2}\gamma_0^{1/2}\Psi^+(\nu,0,\eta) - \frac{1}{(3\gamma_0)^{1/2}}\Psi^+(\nu,1,\eta), \tag{48}$$

where Ψ^+ is given by Eq. (45).

The stage is now set to proceed with the main derivations, whose objective is to determine conditions under which the function Ψ^+ is such that $f_+(z)$ given by Eq. (48) has a finite number of zeros z_i in the domain $|z| \leqslant 1$. At this stage, our method of proof departs from that used by Case,[2] and is presented in Sec. IIIB. For comparison purposes, though, it is useful to summarily recall *Case's procedure*,[2] which consists of the following three steps:

(i) Show that the series in Eq. (45) converges uniformly provided that $\Sigma_0^\infty m|f_m| < \infty$ [where the f_m are the monoenergetic transport equivalents of our $g_m(\eta)$] for $|z| < 1$.

(ii) Show that the series of derivatives of Eq. (45) converges, provided that $\Sigma_0^\infty m^{3/2}|f_m| < \infty$, for $|z| < 1$, so that analyticity of $\Psi^1(\nu,n)$ within the unit circle results.

(iii) Show that Ψ^+ given by Eq. (45) is continuous on the unit circle $|z| = 1$, provided that $\Sigma_{m=0}^\infty m^2|f_m| < \infty$. Note that it is this step that gives rise to the convergence criterion mentioned in Eq. (18).

B. The convergence criterion for the Legendre moments of the scattering function

This section presents the procedure leading to the necessary and sufficient convergence criterion that has to be satisfied by the Legendre moments $g_n(\eta)$ in order that $f(z)$ have a finite number of zeros z_i within $|z| \leqslant 1$. This procedure relies on the following sequence of steps:

(A) Determine the necessary and sufficient conditions for $\Psi^{+(i)}(\nu,n,\eta)$ to be analytic on the domain K:$|z| < 1$, with boundary Γ:$|z| = 1$, and to be continuous on $\bar{K}$.

(B) Use Weierstrass's theorem[13] on uniformly convergent series of analytic functions. This theorem will be used with $\Psi^{+(i)}$ and $\Psi^+(\nu,n,\eta)$.

(C) Use the following theorem[13]:

Theorem: If $h(z) \equiv 0$ is analytic on a domain K, there can be no limit point of zeros of $h(z)$ in K. This theorem will be used with $h(z) \equiv f_+(z)$.

(D) Show that $f_+(z)$ satisfying the above theorem does not have an accumulation point of zeros on the boundary Γ.

It then follows that $f_+(z)$ has a finite number of zeros z_i within $|z| \leqslant 1$. Consequently, the related discrete eigenvalues ν_j are finite in number.

The analytic propertis of $\Psi^{+(i)}(\nu,n,\eta)$ on the domain K:$|z| < 1$ are determined as follows. When $\nu = \frac{1}{2}(z+z^{-1})$ is *not* on the segment $[-1.1]$, and n is large, the polynomial $Q_n(\nu)$ satisfies the inequality[14]

$$|Q_n(\nu)| < \frac{|z|^{n+1}}{[1-|z|^2]^{1/2}}\left(\frac{\pi}{n}\right)^{1/2}, \quad |z|<1, \tag{49}$$

and the polynomial $P_n(\nu)$ satisfies the inequality

$$|P_n(\nu)| < \frac{|z|^{-n}}{(2n+1)^{1/2}}\left(\frac{4}{\pi}\right)^{1/2} \times \frac{1}{[|1-z|\,|1+z|\,(1-|z|^2)]^{1/2}}, |z|<1. \tag{50}$$

(The last inequality is obtained from Eq. (7.1.9) of Ref. 15.)

Using Eqs. (43) and (49) in Eq. (42) gives

$$\begin{aligned}|\Psi_0^+(\nu,n)| &< (2n+1)^{1/2}\frac{1}{|z|}\left(\frac{1-z^2}{2\pi}\right)^{1/2}\frac{|z|^{n+1}}{(1-|z|^2)^{1/2}}\left(\frac{\pi}{n}\right)^{1/2} \\ &\leqslant C_0|z|^n\left[\frac{1-z^2}{1-|z|^2}\right]^{1/2},\end{aligned} \tag{51}$$

where C_0 is a positive constant.

Taking absolute values in Eq. (47) gives

$$|\Psi^{+(i)}(\nu,n)| \leqslant \sum_{m=n+1}^{\infty} |G(\nu,n;m)|\,|cg_m|\,|\Psi^{+(i-1)}(\nu,m)| \tag{52}$$

and, in particular,

$$|\Psi^{+(1)}(\nu,n)| \leqslant \sum_{m=n+1}^{\infty} |G(\nu,n;m)|\,|cg_m|\,|\Psi_0^+(\nu,m)|. \tag{53}$$

Note that henceforth, as in Eqs. (52) and (53), the explicit dependence on η is supressed for notational simplicity.

Substituting Eqs. (41) and (42) in Eq. (40) gives

$$G(\nu,n;m) = \begin{cases} 0; & n \geqslant m, \\ [(2n+1)(2m+1)]^{1/2}\,[Q_m(\nu)P_n(\nu) - P_m(\nu)Q_n(\nu)]; & n<m \end{cases}. \tag{54}$$

Taking absolute values, and using the inequalities (49) and (50) in the above expression leads to

$$\begin{aligned}|G(\nu,n;m)| &< [(2n+1)(2m+1)]^{1/2}\left|\frac{z^2+1}{2z}\right| \\ &\quad \times \{|P_m|\,|Q_n| + |P_n|\,|Q_m|\} \\ &\leqslant C_1|z|^{n-m}\frac{1}{(1-|z|^2)(|1-z^2|)^{1/2}},\end{aligned} \tag{55}$$

where C_1 is a positive constant. Using Eqs. (51) and (55) in Eq. (53) gives

$$|\Psi^{+(1)}(\nu,n)| < C_0\left(\frac{1-z^2}{1-|z|^2}\right)^{1/2}\frac{C_1|z|^n}{(1-|z|^2)(|1-z^2|)^{1/2}} \times \sum_{m=n+1}^{\infty}|cg_m|. \tag{56}$$

Setting $i=2$ in Eq. (52), and using Eq. (56) gives

$$\begin{aligned}|\Psi^{+(2)}(\nu,n)| &< C_0\left(\frac{1-z^2}{1-|z|^2}\right)^{1/2}\left[\frac{C_1}{(1-|z|^2)\,|1-z^2|^{1/2}}\right]^2 \\ &\quad \times |z|^n \sum_{m=n+1}^{\infty}|cg_m| \sum_{m'=m+1}^{\infty}|cg_{m'}| \\ &= C_0\left(\frac{1-z^2}{1-|z|^2}\right)^{1/2}\left[\frac{C_1}{(1-|z|^2)\,|1-z^2|^{1/2}}\right]^2 \\ &\quad \times \frac{|z|^n}{2!}\left(\sum_{m=n+1}^{\infty}|cg_m|\right)^2.\end{aligned} \tag{57}$$

Continuation of this process leads to

$$|\Psi^{+(i)}(\nu,n)| < C_0\left(\frac{1-z^2}{1-|z|^2}\right)^{1/2}\left[\frac{C_1}{(1-|z|^2)\,|1-z^2|^{1/2}}\right]^i \times \frac{|z|^n}{i!}\left(\sum_{m=n+1}^{\infty}|cg_m|\right)^i. \tag{58}$$

In view of Eq. (58), the sequence $\{\Psi^{+(i)}(\nu,n\}$ is uniformly bounded on every compact subset of K, provided that

$$\sum_{m=1}^{\infty}|cg_m| < \infty. \tag{59}$$

Thus, each term in the sum appearing in Eq. (47) is analytic within K. It then follows that $\Psi^{+(i)}(\nu,n)$ is analytic on K, provided the convergence condition (59) is satisfied.

The behavior of $\Psi^{+(i)}(\nu,n)$ on the boundary Γ: $|z|=1$ is examined next. The boundary Γ is characterized by

$$z = e^{i\theta}, \text{ or } \nu = \tfrac{1}{2}(z+z^{-1}) = \cos\theta. \tag{60}$$

Since $Q_n(\nu)$ has logarithmic singularities at $\nu = \pm 1$, the relationship[14]

$$Q_n(\nu) = P_n(\nu)Q_0(\nu) - W_{n-1}(\nu), \tag{61}$$

where

$$W_{n-1}(\nu) = \sum_{k=1}^{n}\frac{1}{k}P_{k-1}(\nu)\,P_{n-k}(\nu), \tag{62}$$

with

$$W_{-1}(\nu) \equiv 0, \tag{63}$$

is used to recast Eq. (54) in the form

$$G(\nu,n;m) = \begin{cases}[(2n+1)(2m+1)]^{1/2}\,[P_m(\nu)\,W_{n-1}(\nu) \\ \qquad - P_n(\nu)\,W_{m-1}(\nu)] \quad \text{for } m>n, \\ 0, \quad \text{for } m \leqslant n.\end{cases} \tag{64}$$

Using the inequality[15]

$$|P_n(\cos\theta)| \leqslant 1, \quad 0 \leqslant \theta \leqslant \pi, \tag{65}$$

and taking absolute values in Eq. (62) gives

$$|W_{n-1}(\cos\theta)| \leqslant \sum_{k=1}^{n}\frac{1}{k}|P_{k-1}(\cos\theta)|\,|P_{n-k}(\cos\theta)| \leqslant \sum_{k=1}^{n}\frac{1}{k}. \tag{66}$$

Taking logarithms in the inequality[16]

$$\exp(x) > \left(1 + \frac{x}{y}\right)^y > \exp\left(\frac{xy}{x+y}\right), \quad x > 0,\ y > 0,$$

setting $x = 1$ and $y = k$, and summing over k from 1 to n leads to

$$\ln(n+1) \leqslant \sum_{k=1}^{n} \frac{1}{k} < 1 + \ln n. \tag{67}$$

Taking absolute values in Eq. (64), using Eqs. (65)–(67), and noting that $m > n$, leads to the bound

$$|G(\cos\theta, n; m)| < C_2' m \ln m, \tag{68}$$

where C_2' is a positive constant.

At the points $\theta = 0$ and $\theta = \pi$ (i.e., at $\nu = \pm 1$), the values $P_n(1) = 1$ and $P_n(-1) = (-1)^n$ can be used in conjunction with Eqs. (62) and (64) to obtain

$$|G(\pm 1, n; m)| = [(2n+1)(2m+1)]^{1/2} \left| \sum_{k=n}^{m} \frac{1}{k} \right|. \tag{69}$$

Since both m and n can become large, but always $m \geqslant n$, Eq. (69) is bounded by

$$|G(\pm 1, n; m)| < C_2'' m \ln m. \tag{70}$$

It follows from Eqs. (68) and (70) that

$$|G(\cos\theta, n; m)| < C_3 m \ln m, \tag{71}$$

where $C_3 \equiv \max(C_2', C_2'')$ is a positive constant. The m-dependence of the right-hand side of inequality (71) appears to be as weak as possible, because of the behavior of $G(\pm 1, n; m)$ indicated by Eq. (69).

Using the inequality[17]

$$(\sin\theta)^{1/2} |Q_n(\cos\theta)| < (\pi/2n)^{1/2}, \quad 0 < \theta < \pi, \tag{72}$$

in Eq. (42) gives

$$|\Psi_0^+(e^{i\theta}, n)| = (2n+1)^{1/2} \left| \left(\frac{e^{-2i\theta} - 1}{2\pi} \right)^{1/2} \right| |Q_n(\cos\theta)| < C_4, \quad 0 \leqslant \theta \leqslant \pi, \tag{73}$$

where C_4 is a positive constant. Note that this inequality holds even at $\theta = 0, \pi$, since $[f_0^+(e^{i\theta}) Q_n(\cos\theta)]$ vanishes at these points.

From Eqs. (53), (71), and (73), it follows that

$$|\Psi^{+(1)}(\cos\theta, n)| < C_3 C_4 \sum_{m=n+1}^{\infty} m \ln m |c g_m|. \tag{74}$$

The same procedure is used to show that

$$|\Psi^{+(2)}(\cos\theta, n)| < C_4 \frac{C_3^2}{2!} \left(\sum_{m=n+1}^{\infty} m \ln m |c g_m| \right)^2 \tag{75}$$

and, in general,

$$|\Psi^{+(i)}(\cos\theta, n)| < C_4 \frac{C_3^i}{i!} \left(\sum_{m=n+1}^{\infty} m \ln m |c g_m| \right)^i \equiv M_i. \tag{76}$$

Therefore, the function $\Psi^{+(i)}$ are continuous on Γ, provided that

$$\sum_{m=1}^{\infty} m \ln m |g_m| < \infty. \tag{77}$$

Note that if the convergence criterion (77) is satisfied, the series $\Sigma_{i=1}^{\infty} M_i$ is also convergent since

$$\lim_{i\to\infty} \frac{M_{i+1}}{M_i} = \lim_{i\to\infty} \frac{C_3}{i+1} \sum_{m=n+1}^{\infty} m \ln m |c g_m| = 0. \tag{78}$$

In view of Eq. (78), Weierstrass's M test[13] ensures that the series

$$\sum_{i=1}^{\infty} \Psi^{+(i)}(\cos\theta, n) \tag{79}$$

is uniformly convergent on Γ.

It has thus been demonstrated that $\Psi^{+(i)}(\nu, n, \eta)$ is analytic on K and continuous on $\bar{K}$ provided that the convergence criterion (77) is satisfied. This completes step (A), as outlined at the beginning of this section, and also satisfies the prerequisites for using Weierstrass's theorem stated in step (B). From this theorem, it follows that

$$\Psi^+(\nu, n) = \sum_{i=0}^{\infty} \Psi^{+(i)}(\nu, n)$$

is analytic on K. (80)

Consequently,

$$f_+(z) = \frac{z + z^{-1}}{2} (\gamma_0)^{1/2} \Psi^+(z, 0) - \frac{1}{(3\gamma_0)^{1/2}} \Psi^+(z, 1) \tag{48}$$

is analytic on K, except perhaps for a simple pole at $z = 0$.

In view of the theorem stated in step (C), $f_+(z)$ has no accumulation point of zeros in K. Conceivably, there could be an accumulation point of zeros on the boundary Γ. However, this would require $f_+(z)$ to have an essential singularity on Γ, implying that Ψ^+ has an essential singularity on Γ. But Ψ^+ cannot have an essential singularity on Γ since, in view of Eqs. (76) and (45),

$$|\Psi^+(\cos\theta, n)| < C_4 \exp\left(C_3 \sum_{m=n+1}^{\infty} m \ln m |c g_m| \right),$$

i.e., $\Psi^+(\cos\theta, n)$ is bounded in view of Eq. (77). This contradiction completes the proof that $f_+(z)$ can have at most a finite number of zeros z_i within $|z| \leqslant 1$ [Step (D) has thus been accomplished]. It therefore follows that the related discrete eigenvalues ν_i are finite in number, provided that the convergence criterion (77) is satisfied.

IV. SUMMARY AND CONCLUSIONS

The finiteness of the number of discrete eigenvalues occurring in neutron slowing down and transport in the "constant-cross-section problem" has been examined. By taking a Laplace transform in lethargy, the Boltzmann equation governing this problem has been transformed into an equation of the same form as the one-velocity transport equation with "infinite-order" anisotropy. The only difference from the standard one-velocity form is that the Legendre moments of the scattering function depend on the transform variable.

Case[2] showed that, in monoenergetic transport with infinite-order anisotropy, the discrete eigenvalues are finite in number provided that the Legendre moments f_n of the scattering function satisfy the convergence criterion $\Sigma_{n=0}^{\infty} n^2 |f_n| < \infty$. However, it is shown in this work that the equivalent in the constant-cross-section problem of Case's

convergence criterion is *not* satisfied by hydrogen in a region that corresponds asymptotically to the region of large lethargy; this is where the age-theory solution is the limiting form taken on by the space- and lethargy-dependent neutron collision density. Since there seemed to be no physical reason to expect the occurrence of infinitely many discrete eigenvalues in this region, the problem of convergence conditions for the Legendre moments of the scattering function has been reexamined.

The preliminaries for this reexamination paralleled the methods of scattering theory applied to orthogonal polynomials as used by Case[2] in monoenergetic transport. However, a new procedure has been devised to derive the convergence criterion for the Legendre moments of the scattering function. The underlying derivations showed that Case's convergence criterion is not necessary, although it is sufficient, for the discrete eigenvalues to be finite in number. On the other hand, it appears that the convergence criterion, used by Inonu[1] is insufficient; this insufficiency arises from the apparent failure of this criterion to ensure continuity properties on the cut $[-1,1]$ in the ν eigenvalue space. The new convergence criterion derived in this work, i.e,

$$\sum_{m=1}^{\infty} m \ln m |g_m(\eta)| < \infty,$$

seems to be both necessary and sufficient for the discrete eigenvalues to be finite in number. Unlike Case's convergence criterion, the new convergence criterion *is* satisfied by hydrogen at small values of η that correspond to the region described by age-theory.

The derivations leading to the new convergence criterion have been carried out within the more general framework of energy-dependent slowing-down and transport. It is apparent, though, that this convergence criterion also holds for the Legendre moments of the scattering function in one-velocity transport.

APPENDIX A: THE SERIES $\Sigma_{n=0}^{\infty} \eta^{\beta}|g_n(\eta)|$ AND $\Sigma_{n=1}^{\infty} n \ln n|g_n(\eta)|$

1. The series $\Sigma_{n=0}^{\infty} \eta^{\beta}|g_n(\eta)|$, $\beta =$ constant

When $A > 1$, D'Alembert's criterion can be used in conjunction with Eqs. (B17) and (B22) of Appendix B to show that

$$\lim_{n\to\infty} \frac{(n+1)^{\beta}}{n^{\beta}} \left| \frac{g_{n+1}(\eta)}{g_n(\eta)} \right| = 0, \text{ for } \beta = \text{const}, \tag{A1}$$

for any fixed (large or small) value of η. Hence, the series $\Sigma_{n=0}^{\infty} n^{\beta}|g_n(\eta)|$ is absolutely convergent for any $A > 1$ and any constant β.

When $A = 1$, i.e., hydrogen, the series under consideration becomes

$$\sum_{n=0}^{\infty} n^{\beta}|g_n(\eta)| = \sum_{m=0}^{\infty} v_m(\eta), \tag{A2}$$

where

$$v_m(\eta) = (2m)^{\beta}|g_{2m}(\eta)| + (2m+1)^{\beta}|g_{2m+1}(\eta)|. \tag{A3}$$

For large m, the expressions given by Eqs. (B29) and (B31) of Appendix B are now used in conjunction with Eq. (A3) to obtain

$$\lim_{m\to\infty} v_m(\eta) = 0, \text{ if } \beta < 2\,\mathrm{Re}(\eta) + \tfrac{5}{2}, \tag{A4}$$

and

$$\frac{v_m}{v_{m+1}} = 1 + (2\eta + \tfrac{5}{2} - \beta)\frac{1}{m} + O(m^{-2}). \tag{A5}$$

It follows from Eq. (A5) and Gauss's criterion for convergence that the series

$\Sigma_{n=0}^{\infty} n_{\beta}|g_n(\eta)|$ is absolutely convergent if and only if
$\beta < 2\,\mathrm{Re}(\eta) = \tfrac{3}{2}$. (A6)

This result implies that Case's requirement,[2] i.e, $\Sigma_{n=0}^{\infty} n^2|g_n(\eta)| < \infty$, is *not* satisfied if $\mathrm{Re}(\eta) \leqslant \frac{1}{4}$. In particular, at $\eta = 0$, the sereis $\Sigma_{n=0}^{\infty} n^2|g_n(0)|$ is divergent in view of (A6). This fact can be confirmed directly by using the particularly simple form taken on by $g_n(0)$, i.e.,

$$g_n(0) = 2\int_0^1 xP_n(x)\,dx = \begin{cases} 1 \text{ for } n = 0, \\ \frac{2}{3} \text{ for } n = 1, \\ 0 \text{ for } n = \text{odd},\ n > 1, \\ \dfrac{(-1)^{n/2}\, n!}{2^{n-1}(n-1)(n+2)[n/2)!]^2}, \text{ for } n = \text{even}. \end{cases} \tag{A7}$$

Using Eq. (A7), it is easy to prove that the series

$$\sum_{n=0}^{\infty} n^2|g_n(0)| = \frac{2}{3} + \sum_{m=0}^{\infty} (2m)^2|g_{2m}(0)| \text{ is divergent,} \tag{A8}$$

since

$$\frac{(2m)^2|g_{2m}(0)|}{[2(m+1)]^2|g_{2(m+1)}(0)|} = 1 + \frac{1}{2m} + O(m^{-2}).$$

2. The series $\Sigma_{n=1}^{\infty} n \ln n|g_n(\eta)|$

When $A > 1$, D'Alembert's criterion can be used in conjunction with Eqs. (B17) and (B22) to show that the series $\Sigma_{n=1}^{\infty} n \ln n|g_n(\eta)|$ is absolutely convergent for any fixed value of η.

For hydrogen, i.e., $A = 1$, the series can be written as

$$\sum_{n=1}^{\infty} n \ln n|g_n(\eta)| = \sum_{m=1}^{\infty} u_m, \tag{A9}$$

where

$$u_m = 2m|g_{2m}(\eta)| \ln(2m) + (2m+1)|g_{2m+1}(\eta)| \ln(2m+1). \tag{A10}$$

For large m, Eqs. (B29) and (B31) can be used in conjunction with Eq. (A9) to obtain

$$\frac{u_m}{u_{m+1}} = 1 + \frac{1}{m}(2\eta + \tfrac{3}{2}) + O\left(\frac{1}{m \ln m}\right). \tag{A11}$$

Thus, according to Gauss's criterion, the series

$\Sigma_{n=1}^{\infty} n \ln n|g_n(\eta)|$ is absolutely convergent if
$\mathrm{Re}(\eta) > -\frac{1}{4}$. (A12)

APPENDIX B: ASYMPTOTIC EXPRESSIONS FOR THE FUNCTIONS $g_n(\eta)$

Recall that the functions $g_n(\eta)$ are defined as the Laplace transforms of the coefficients $f_n(u)$ of the Legendre expansion of the scattering kernel $f(\bar{\Omega}\cdot\bar{\Omega}',u)$, i.e.,

$$g_n(\eta)\equiv\mathcal{L}[f_n(u)].$$

Two cases have to be distinguished, depending on the nuclear mass A of the scatterer:

(1) When $A=1$ (i.e., hydrogen), the quantity

$$\epsilon\equiv\ln\frac{A+1}{A-1} \tag{B1}$$

becomes infinite and therefore

$$g_n(\eta)=\int_0^\infty e^{-(\eta+1)u}P_n(e^{-u/2})\,du. \tag{B2}$$

(2) For any $A>1$, the expression of $g_n(\eta)$ is

$$g_n(\eta)=\alpha\int_0^\epsilon e^{-(\eta+1)u}\times P_n\left[\frac{A+1}{2}\exp\left(-\frac{u}{2}\right)-\frac{A-1}{2}\exp\left(\frac{u}{2}\right)\right]du, \tag{B3}$$

where

$$\alpha\equiv(A+1)^2/4A. \tag{B4}$$

The functions $g_n(\eta)$ have received considerable attention in the earlier literature, and important results concerning their properties have been obtained by Placzek,[7] Marshak,[8] Wick,[9] Bethe *et al.*,[10] and Holte.[11] The purpose of this Appendix is to present several new results regarding the asymptotic expressions taken on by these functions as one or more of the variables n,η, and A become large or small. These new asymptotic expressions are particularly useful for investigating convergence conditions for infinite sums involving the functions $g_n(\eta)$.

1. The case $A>1$

As Eq. (B3) indicates, the functions $g_n(\eta)$ are in this case entire functions of η. It does not seem possible, though, to obtain a closed-form analytic expression for $g_n(\eta)$ by directly carrying out the integration in Eq. (B3). Nevertheless, it is possible to obtain asymptotic expressions that can in turn be used to examine convergence properties of infinite sums involving the functions $g_n(\eta)$. These functions are expected to take on distinct asymptotic expressions, according to whether η is large or small. Recall that the region where η is large (small) corresponds to the region of small (large) lethargies u.

a. Asymptotic form of $g_n(\eta)$ for large η

When η is large, an asymptotic expression for $g_n(\eta)$ that is valid for *any* values of $n\geqslant 0$ and $A>1$ can be obtained by applying Watson's Lemma[6] to the Legendre moments

$$f_n(u)=\alpha e^{-u}P_n\left(\frac{A+1}{2}e^{-u/2}-\frac{A-1}{2}e^{u/2}\right). \tag{B5}$$

This lemma can be used because $|f_n(u)|<K_1\exp(K_2u)$, where K_1 and K_2 are some positive constants. According to Watson's lemma, the asymptotic expression of $g_n(\eta)$ for large η is

$$g_n(\eta)\sim\alpha\sum_{r=0}^{\infty}C_r(n,A)\eta^{-r-1}, \tag{B6}$$

where the coefficients $C_r(n,A)$ arise in the MacLaurin expansion

$$f_n(u)=\alpha\sum_{r=0}^{\infty}C_r(n,A)\frac{u^r}{r!}. \tag{B7}$$

The straightforward algebra required to determine the coefficients $C_r(n,A)$ has been performed in Ref. 18. As a function of n, the expression of $C_r(n,A)$ becomes increasingly cumbersome as n increases. Thus, although Eq. (B6) is valid for any $n\geqslant 0$, it is cumbersome to use for examining convergence properties of infinite sums involving the functions $g_n(\eta)$, since for this purpose the behavior of $g_n(\eta)$ for large n becomes of primary interest. When both η and n become large, the asymptotic expression for $g_n(\eta)$ becomes considerably simpler than Eq. (B6), and can be obtained as follows.

Equation (B3) is recast in the form

$$g_n(\eta)=2\pi\int_{-1}^{1}d\mu\,P_n(\mu)\,g(\mu,n), \tag{B8}$$

where

$$g(\mu,n)=\frac{\alpha}{\pi}(A^2-1+\mu^2)^{-1/2}\times\left[\frac{\mu+(\mu^2+A^2-1)^{1/2}}{A+1}\right]^{2(\eta+1)}. \tag{B9}$$

As already discussed, the region of large η corresponds to the region of small values of u around $u=0$. Thus, when η is large, the angle

$$\theta\equiv\arccos\mu=\arccos\left[\frac{A+1}{2}\exp(-u/2)-\frac{A-1}{2}\exp(u/2)\right] \tag{B10}$$

is a very small angle (i.e., $\theta\approx 0$). When, in addition to θ being small, n becomes large, $P_n(\cos\theta)$ takes on the form[15]

$$\left(\frac{\sin\theta}{\theta}\right)^{1/2}P_n(\cos\theta)=J_0[(n+\tfrac{1}{2})\theta]-\frac{\theta}{24(n+\frac{1}{2})}J_1[(n+\tfrac{1}{2})\theta]+\sigma, \tag{B11}$$

where

$$|\sigma|<0.03\,\theta^4\quad\text{if } 0<\theta\leqslant\pi/(2n)$$
$$|\sigma|<0.25\,\theta^{5/2}n^{-3/2}\quad\text{if } \pi/(2n)<\theta\leqslant\pi/2. \tag{B12}$$

Expanding Eq. (B9) for small θ yields

$$g(\cos\theta,\eta)=\frac{\alpha}{\pi A}\exp[-\eta_1\theta^2+\eta_2\theta^4-\eta_3\theta^6+O(\theta^8)], \tag{B13}$$

where

$$\eta_1=\frac{1}{A}\left(\eta+1-\frac{1}{2A}\right), \tag{B14}$$

$$\eta_2 = \frac{1}{12A}\left(1 - \frac{3}{A^2}\right)\left(\eta + 1 - \frac{1}{A}\cdot\frac{2 - 3/A^2}{1 - 3/A^2}\right), \tag{B15}$$

$$\eta_3 = \frac{1}{360A}\left(1 - \frac{30}{A^2} - \frac{45}{A^4}\right) \times\left(\eta + 1 - \frac{4}{A^2}\frac{2 - 15/A^2 + 15/A^4}{1 - 30/A^2 + 45/A^4}\right). \tag{B16}$$

Equations (B11) and (B13) are now used in Eq. (B8), and the prescribed integration over μ is transformed into an equivalent integration over θ. Retaining consistently terms up to $O(\theta^4) = O(n^{-4})$, and using the relationship

$$\int_0^\infty \theta^{n+\nu/2} e^{-\alpha\theta} J_\nu(2\beta\sqrt{\theta})\, d\theta = n!\beta^\nu e^{-\beta^2/\alpha} L_n^\nu(\beta^2/\alpha),$$

(where L_n^ν are the Laguerre Polynomials) gives

$$g_n(\eta) = \frac{\alpha}{A\eta_1}\left\{1 - \frac{1}{12\eta_1}\left[\frac{5}{4} - \frac{(n+\frac{1}{2})^2}{4}\right]\right\} \times\exp\left[-\frac{(n+\frac{1}{2})^2}{4\eta_1}\right] + R, \tag{B17}$$

where

$$R \leqslant \frac{1}{\eta_1^3}\left\{2\left(0.03 + \frac{1}{1440} + \eta_2\right) + \frac{1}{576}\left[2 - \frac{(n+\frac{1}{2})^2}{4}\right]\right\}. \tag{B18}$$

Note that by imposing the additional restrictions $\eta \gg A$ and $An^2 \gg 4\eta$, Eq. (B17) reduces to the asymptotic expression obtained earlier by Wick,[9] i.e.,

$$g_n(\eta) \sim \frac{\alpha}{\eta}\exp\left[-\frac{A}{4\eta}(n+\tfrac{1}{2})^2\right]$$

b. Asymptotic form of $g_n(\eta)$ for small η

It has been shown by Price[19] that whenever the quantity x defined as

$$x = \frac{2}{A}(\eta + 1) \tag{B19}$$

has values up to unity, $g_n(\eta)$ can be represented as

$$g_n(x) = \left(\frac{A+1}{A}\right)^2\left(\frac{\pi}{2x}\right)^{1/2} I_{n+1/2}(x) + O(1/A^2), \quad |x| < 1, \tag{B20}$$

where $I_{n+1/2}$ is the modified Bessel function of the first kind. For large values of n, the representation

$$I_{n+1/2}(x) = \sum_{k=0}^{\infty} \frac{\left(\frac{x}{2}\right)^{2k+n+(1/2)}}{k!\Gamma(k+n+3/2)} \tag{B21}$$

can be used in Eq. (B20) to obtain

$$g_n(x) = \left(\frac{A+1}{A}\right)^2 e^{-x}\frac{x^n}{(2n+1)!!} + O(1/n). \tag{B22}$$

2. The case $A = 1$ (hydrogen)

With the change of variables $x = e^{-u/2}$, Eq. (B2) becomes

$$g_n(\eta) = 2\int_0^1 x^\lambda P_n(x)dx, \tag{B23}$$

where

$$\lambda = 2\eta + 1. \tag{B24}$$

The integration in Eq. (B23) can be carried out exactly, by employing the formulas[20]

$$\int_0^1 x^\lambda P_{2m}(x)dx = \frac{(-1)^m\Gamma(m - \lambda/2)\Gamma(\frac{1}{2} + \lambda/2)}{2\Gamma(-\lambda/2)(m + \frac{3}{2} + \lambda/2)}, \quad \text{for } \mathrm{Re}\,\lambda > -1$$

and

$$\int_0^1 x^\lambda P_{2m+1(x)dx} = \frac{(-1)^m\Gamma(m + \frac{1}{2} - \lambda/2)\Gamma(1 + \lambda/2)}{2\Gamma(\frac{1}{2} - \lambda/2)\Gamma(m + 2 + \lambda/2)}, \quad \text{for } \mathrm{Re}\,\lambda > -2.$$

This gives

$$g_{2m}(\eta) = \frac{(-1)^m\Gamma(m - \lambda/2)\Gamma(\frac{1}{2} + \lambda/2)}{\Gamma(-\lambda/2)\Gamma(m + \frac{3}{2} + \lambda/2)}, \quad \text{for } \mathrm{Re}\,\lambda > -1 \tag{B25}$$

and

$$g_{2m+1}(\eta) = \frac{(-1)^m\Gamma(m + \frac{1}{2} - \lambda/2)\Phi(1 + \lambda/2)}{\Gamma(\frac{1}{2} - \lambda/2)\Gamma(m + 2 + \lambda/2)}, \quad \text{for } \mathrm{Re}\,\lambda > -2. \tag{B26}$$

When m becomes large, asymptotic expressions for $g_{2m}(\eta)$ and $g_{2m+1}(\eta)$ can be obtained by using the asymptotic representation[20]

$$\ln\Gamma(z) = z\ln z - z - \tfrac{1}{2}\ln z + \ln(2\pi)^{1/2} + O(z^{-2}), \quad \text{for large } z, \tag{B27}$$

and the fact that

$$\ln(m + a) = \ln m + \frac{a}{m} + O(m^{-2}), \quad m \text{ large}. \tag{B28}$$

Thus, using Eqs. (B27) and (B28) in Eq. (B25) gives

$$|g_{2m}(\eta)| \sim Km^{-(2\eta + 5/2)}, \tag{B29}$$

where

$$K = \frac{\Gamma(\frac{1}{2} + \lambda/2)}{\Gamma(-\lambda/2)}. \tag{B30}$$

Similarly, using Eqs. (B27) and (B28) in Eq. (B26) leads to

$$|g_{2m+1}(\eta)| \sim K_1 m^{-(2\eta + 5/2)}, \tag{B31}$$

where

$$K_1 = \frac{\Gamma(1 + \lambda/2)}{\Gamma(\frac{1}{2} - \lambda/2)}. \tag{B32}$$

ACKNOWLEDGMENTS

Stimulating discussions with M. C. G. Hall of Oak Ridge National Laboratory are gratefully acknowledged. Special thanks are due to Ruth Fahey for her expert typing of this paper.

[1]E. Inonu, J. Math. Phys. **11**, 568 (1970).
[2]K. M. Case, J. Math. Phys. **15**, 974 (1974).
[3]K. M. Case and P. F. Zweifel, *Linear Transport Theory* (Addison-Wesley, Reading, Mass., 1967).
[4]J. J. McInerney, Nucl. Sci. Eng. **22**, 215 (1965).
[5]D. G. Cacuci and H. Goldstein, J. Math. Phys. **18**, 2436 (1977).
[6]D. V. Widder, *The Laplace Transform* (Princeton U. P., Princeton, N.J., 1972). See also I. N. Sneddon, *The Use of Integral Transforms* (McGraw-Hill, New York, 1972).
[7]G. Placzek, Manhattan Project Report A-25 (unpublished).
[8]R. E. Marshak, Rev. Mod. Phys. **19**, 185 (1947).
[9]G. C. Wick, Phys. Rev. **75**, 738 (1949).
[10]H. A. Bethe, L. Tonks, and H. Hurwitz, Jr., Phys. REv. **80**, 11 (1950).
[11]G. Holte, Ark. Math., Astr., Fys., Band **35** a, 1 (1948), and Ark. Fys., **2** (1950).
[12]H. J. Amster and K. C. Chan, Nucl. Sci. Eng. **61**, 388 (1976).
[13]A. I. Markushevich, *Theory of Functions of a Complex Variable, Vol. 1* (Prentice-Hall, Englewood Cliffs, N.J., 1965).
[14]E. W. Hobson, *The Theory of Spherical and Ellipsoidal Harmonics* (Cambridge U. P., London, 1931), p. 61.
[15]G. Szego, *Orthogonal Polynomials*, 4th ed. (Am. Math. Soc., Providence, RI, 1975), Vol. XXIII.
[16]M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions* (U. S. Government Printing Office, Washington, D.C., 1964), Natl. Bur. Stand. Appl. Math. Series, Vol. 55, p. 70.
[17]Ref. 14, p. 309.
[18]D. G. Cacuci, "Analytical Calculations of Neutron Slowing Down and Transport in the Constant-Cross-Section Problem," ORNL/TM-6378, Oak Ridge National Laboratory, Oak Ridge, TN, April 1978.
[19]J. P. Price, Atomic Energy Research Establishment, T/R 523, April 1950.
[20]I. S. Gradshtein and I. M. Ryzhik, *Tables of Integrals, Series and Products* (Academic, New York, 1973), 7th printing.