Unification of gauge and gravitation theories

R. Tabensky *

Departamento de Física, Facultad de Ciencias Físicas y Matemáticas, Universidad de Chile, Santiago, Chile
(Received 15 April 1975)

A new approach to local Gauge theories is presented. The theory is developed in a curved space–time, and therefore gravitation is not neglected. Besides the Yang–Mills A_μ^a vector bosons associated with a symmetry group, scalar bosons $g_{\alpha\beta}$ appear just as naturally. The Lagrangian describing the interaction of these fields is the Ricci scalar for an extended Riemannian geometry.

1. INTRODUCTION

The only strictly nonphenomenological theory physicists have today is the Einstein theory of gravitation. No potentials nor experimental parameters have to be introduced. Einstein field equations $R_{\mu\nu}=0$ are the simplest one can write for a pseudo-Riemanian geometry.

In this paper an extended pseudo-Riemanian (ERG) structure based on any semisimple Lie algebra is introduced. This structure is as flexible as ordinary tensor analysis. The corresponding Einstein's equations $R_{ab}=0$ turn out to be a Yang—Mills theory fully coupled to gravity. The theory is coordinate and gauge invariant. Besides the usual Yang—Mills vector bosons A_μ^α a number of independent scalar bosons make their appearance naturally.

The most economic way of presenting the ERG is by using modern differential geometry. This language shall be used in this introduction to give a short algebraic account, all topological questions set aside, of the theory of gravitation. The main reason to do so is that almost all formulas remain valid for ERG an thus notation is set straight.

Let M be a four-dimensional differentiable manifold[1] called space—time and $C^\infty(M)$ the set of all infinitely differentiable real functions on M. The basic object in this theory are vector fields denoted by X, Y, Z. X is a mapping of $C^\infty(M)$ into itself with the following properties:

$$X(\alpha f+\beta g)=\alpha Xf+\beta Xg, \qquad \alpha,\beta \text{ reals}, \tag{1.1}$$

$$Xfg=(Xf)g+f(Xg), \qquad f,g\in C^\infty(M). \tag{1.2}$$

Let $D(M)$ be the set of all vector fields. $D(M)$ is turned into a module over the ring $C^\infty(M)$ by defining fX and $X+Y$ as follows[1]:

$$fX: g\to f(Xg)$$
$$X+Y: g\to Xg+Yg \quad , \quad f,g\in C^\infty(M) \tag{1.3}$$

The Lie product

$$[X,Y]=XY-YX, \tag{1.4}$$

where XY stands for a composition of mappings, is also a vector field. The Lie product satisfies the Jacobi identity $[X,[Y,Z]]+[Y,[Z,X]]+[Z,[X,Y]]=0$. A general solution of Eqs. (1.1) and (1.2) may be found in any coordinate system of M.[1] Any X takes the form $(Xf)(x^\mu)=\xi^\mu(x)\partial f(x)/\partial x^\mu$, where $\xi^\mu\in C^\infty(M)$. That is,

$$X=\xi^\mu\partial_\mu. \tag{1.5}$$

The ξ^μ are the components of X in the partial derivative basis ∂_μ.

An affine connection is a rule ∇ which assigns to each $X\in D(M)$ a linear mapping $\nabla_x: D(M)\to D(M)$ satisfying the following two conditions:

$$\nabla_{fX+gY}=f\nabla_X+g\nabla_Y, \qquad f,g\in C^\infty(M), \tag{1.6}$$
$$\nabla_X(fY)=f\nabla_X(Y)+(Xf)Y.$$

These properties define $\nabla_X Y$ for all X and Y if $\nabla_X Y$ is specified for a basis in $D(M)$, that is, if $\Gamma^\lambda_{\mu\nu}\in C^\infty(M)$, defined by

$$\nabla_{\partial_\mu}(\partial_\nu)=\Gamma^\lambda_{\mu\nu}\partial_\lambda \tag{1.7}$$

are given. $\Gamma^\lambda_{\mu\nu}$ are the Christoffel symbols for a pseudo-Riemanian geometry.

The linear space $D(M)$ is provided with a nondegenerate metric $\langle X,Y\rangle\in C^\infty(M)$ with the following properties[2]:

$$\langle X,Y\rangle=\langle Y,X\rangle, \tag{1.8}$$
$$\langle fX+gY,Z\rangle=f\langle X,Z\rangle+g\langle Y,Z\rangle,$$

$$Z\langle X,Y\rangle=\langle\nabla_Z X,Y\rangle+\langle X,\nabla_Z Y\rangle. \tag{1.9}$$

Equation (1.9) relates the affine and metric structures; it is better known by the equivalent statement that the covariant derivative of the metric $g_{\mu\nu}=\langle\partial_\mu,\partial_\nu\rangle$ vanishes.

The following two expressions are fundamental:

$$T(X,Y)=\nabla_X Y-\nabla_Y X-[X,Y], \tag{1.10}$$

$$R(X,Y)=\nabla_X\nabla_Y-\nabla_Y\nabla_X-\nabla_{[X,Y]}. \tag{1.11}$$

$T(X,Y)$ is a vector field with the following linear property: $T(fX+gY,Z)=fT(X,Y)+gT(Y,Z)$ for any $f,g\in C^\infty(M)$. The same property holds for variable Z since $T(X,Z)=-T(Z,X)$. The vector field $R(X,Y)Z$ has the same linear property for the variables $X, Y,$ and Z. Equations (1.10) and (1.11) are related to the usual torsion and Riemann tensors in component form by the following expressions:

$$T(\partial_\mu,\partial_\nu)=T^\lambda_{\mu\nu}\partial_\lambda$$

and (1.12)

$$R(\partial_\mu,\partial_\nu)\partial_\lambda=R^\rho_{\lambda\mu\nu}\partial_\rho.$$

Both $T^\lambda_{\mu\nu}$ and $R^\rho_{\lambda\mu\nu}$ are expressed entirely in terms of $\Gamma^\lambda_{\mu\nu}$. A pseudo-Riemanian space is defined by $T(X,Y)=0$ or $T^\lambda_{\mu\nu}=0$. In this case the affine connection is entirely determined by the nondegenerate metric through the following identity[1]:

$$2\langle X, \nabla_Z Y\rangle = Z\langle X, Y\rangle + Y\langle X, Z\rangle - X\langle Y, Z\rangle + \langle Z, [X, Y]\rangle + \langle Y, [X, Z]\rangle - \langle X, [Y, Z]\rangle \tag{1.13}$$

derived from (1.9) and $T(X, Y) = 0$ by cyclic permutation. When Eq. (1.13) is written for the basis ∂_μ, the usual Christoffel symbols

$$\Gamma^\lambda_{\mu\nu} = \tfrac{1}{2} g^{\lambda\rho}(g_{\rho\mu,\nu} + g_{\rho\nu,\mu} - g_{\mu\nu,\rho}) \tag{1.14}$$

are obtained. $g^{\lambda\rho}$ stands for the inverse matrix of $g_{\lambda\rho}$.

It is now possible to write the Einstein invariant action

$$S = \int d^4x \sqrt{g}\, R, \tag{1.15}$$

where $g = \det g_{\mu\nu}$ and $R = g^{\lambda\nu} R^\rho_{\lambda\rho\nu}$ is the Ricci scalar. The fields to be varied are seen to be $g_{\mu\nu}$ from (1.14). Upon variation the field equations $R^\rho_{\mu\rho\nu} = R_{\mu\nu} = 0$ are obtained.

The mathematical apparatus just described may be thought of as a way of generating fields and field euqations in a natural way.

In the next section the notion of vector fields is generalized in such a way as to retain most of the apparatus just described, in particular, the ability to write the Einstein action (1.15) invariant under coordinate and gauge transformations and thus obtain a Yang—Mills field theory fully coupled to gravity.[3]

2. EXTENDED RIEMANNIAN GEOMETRY

In this section the notion of a vector field is generalized on the basis of any semisimple Lie algebra[4,5] associated to a group of symmetry. Everything else will be an exact parallel of ordinary Riemannian geometry.

Let L be a semisimple, finite-dimensional Lie algebra over the reals. $L^\infty(M)$ denotes the set of infinitely differentiable functions $M \to L$, where M is the space—time. Differentiability may be defined componentwise. $L^\infty(M)$ plays here the role of $C^\infty(M)$ in the definition of vector fields given in the Introduction. $L^\infty(M)$ is turned into a Lie algebra over the ring $C^\infty(M)$ by the following definitions of $\theta + \phi$, $f\theta$, and $[\theta, \phi]$ for $\theta, \phi \in L^\infty(M)$ and $f \in C^\infty(M)$:

$$\theta + \phi : x^\mu \to \theta(x^\mu) + \phi(x^\mu),$$
$$f\theta : x^\mu \to f(x^\mu)\theta(x^\mu), \tag{2.1}$$

and

$$[\theta, \phi]x^\mu \to [\theta(x^\mu), \phi(x^\mu)].$$

The last bracket denotes the Lie product in L. x^μ stands for points (or coordinates) in M.

An extended vector field (EVF) X is defined to be a simultaneous mapping of the form

$$X : L^\infty(M) \to L^\infty(M),$$
$$X : C^\infty(M) \to C^\infty(M) \tag{2.2}$$

with the following derivation type properties:

$$X(\theta + \phi) = X\theta + X\phi, \qquad \theta, \phi \in L^\infty(M), \tag{2.3}$$
$$X(f\theta) = fX\theta + (Xf)\theta, \qquad f, g \in C^\infty(M), \tag{2.4}$$
$$X[\theta, \phi] = [X\theta, \phi] + [\theta, X\phi], \tag{2.5}$$
$$X(\alpha f + \beta g) = \alpha Xf + \beta Xg, \qquad \alpha, \beta \text{ reals}, \tag{2.6}$$
$$X(fg) = (Xf)g + fXg. \tag{2.7}$$

Let $ED(M)$ be the set of all vector fields. $ED(M)$ may in turn be converted into a Lie algebra over the ring $C^\infty(M)$ by canonical definitions of $X + Y$, fX and $[X, Y] = XY - YX$ for $X, Y \in ED(M)$.

Let l be the dimension of L. It shall now be shown that $ED(M)$ is of dimension $l + 4$, by exhibiting a natural basis for $ED(M)$.

Notice that the action of X on differentiable functions may be expressed as $X = \xi^\mu \partial_\mu$ since (2.6) and (2.7) are the same as (1.1) and (1.2). Any element $\theta \in L^\infty(M)$ may be written as

$$\theta = f^\alpha \theta_\alpha, \quad \alpha = 1 \ldots l, \quad f^\alpha \in C^\infty(M), \tag{2.8}$$

where θ_α is a basis of $L^\infty(M)$. From Eq. (2.3) and (2.4) it follows that

$$X\theta = (Xf^\alpha)\theta_\alpha + f^\alpha X^\beta_\alpha \theta_\beta, \quad X^\beta_\alpha \in C^\infty(M). \tag{2.9}$$

In order to make life simple, the bases θ_α are chosen in such a way as to have constant structure constants, that is, $X(C^\gamma_{\alpha\beta}) = 0$, where the $C^\gamma_{\alpha\beta}$ are defined by

$$[\theta_\alpha, \theta_\beta] = C^\gamma_{\alpha\beta}\theta_\gamma, \quad \alpha, \beta, \gamma = 1 \cdots l. \tag{2.10}$$

It is always possible to have $X(C^\gamma_{\alpha\beta}) = 0$ by choosing θ_α to be l independent constant mappings since $X(C^\gamma_{\alpha\beta}) = \xi^\mu \partial_\mu C^\gamma_{\alpha\beta}$ for some ξ^μ. The main reason for choosing $C^\gamma_{\alpha\beta}$ constant is that the mapping $X : \theta \to X(f^\alpha)\theta_\alpha$ satisfies all axioms (2.3)—(2.7) and thus the linear mapping $X : \theta \to f^\alpha X^\beta_\alpha \theta_\beta$ is a derivation of $L^\infty(M)$, that is, X satisfies (2.5), as simple verification will show. It is well known that a derivation of a semisimple Lie algebra is necessarily inner,[1] which amounts to the fact that for each X^β_α there is a $\phi \in L^\infty(M)$ such that $X^\beta_\alpha \theta_\beta = [\phi, \theta_\alpha]$. Any extended vector field X takes the form given by

$$X\theta = \xi^\mu \partial_\mu f^\alpha \theta_\alpha + [\phi, \theta] \tag{2.11}$$

when $C^\gamma_{\alpha\beta,\mu} = 0$. This is a base-dependent expression for X, and therefore the study of its transformation properties is most important. A direct computation based on Eq. (2.11) shows that under the transformation $\bar{\theta}_\alpha = L^\beta_\alpha \theta_\beta [L^\beta_\alpha \in C^\infty(M)]$, ξ^μ and ϕ transform accordingly to

$$\bar{\xi}^\mu = \xi^\mu \tag{2.12}$$

and

$$\bar{\phi} = \phi + g_0^{\alpha\delta} C^\gamma_{\alpha\epsilon} L^{-1\beta}_\gamma \xi^\mu L^\epsilon_{\beta,\mu}\, \theta_\delta,$$

where

$$g_{0\alpha\beta} = C^\gamma_{\alpha\delta} C^\delta_{\beta\gamma}. \tag{2.13}$$

$L^{-1\alpha}_\beta$ and $g_0^{\alpha\beta}$ are the inverse matrices of L^α_β and $g_{0\alpha\beta}$. $g_0^{\alpha\beta}$ exists because L is semisimple. From Eq. (2.12) it is seen that the difference $\phi_1 - \phi_2$, $\phi_1, \phi_2 \in L^\infty(M)$, for a fixed four-vector ξ^μ, is an invariant. This implies that, for any four $A_\mu \in L^\infty(M)$, $\Omega = \phi - \xi^\mu A_\mu$ is an invariant under base transformation. When (2.11) is written in terms of the invariant Ω,

$$X\theta = \xi^\mu(\partial_\mu \theta + [A_\mu, \theta]) + [\Omega, \theta]$$

is obtained. ∂_μ is a base dependent mapping that carries the instruction to differentiate componentwise. In order to simplify notation let $\mathcal{L}_\Omega$ stand for $\mathcal{L}_\Omega(\theta)=[\Omega,\theta]$, commonly called the Lie derivative with respect to Ω. X then takes the form

$$X=\xi^\mu(\partial_\mu+\mathcal{L}_{A_\mu})+\mathcal{L}_\Omega. \tag{2.14}$$

The transformation law for A_μ is easily calculated from Eq. (2.12), the following result,

$$\bar{A}_\mu=A_\mu+g_0^{\varepsilon\delta}C^\gamma_{\varepsilon\alpha}L^{-1\beta}_\gamma L^\alpha_{\beta,\mu}\theta_\delta, \tag{2.15}$$

is obtained for the base transformation $\bar{\theta}_\alpha=L^\beta_\alpha\theta_\beta$. A_μ is arbitrary but cannot be set equal to zero for all bases because of the transformation law (2.15). Notice that $\partial_\mu+\mathcal{L}_{A_\mu}$ is an invariant element of $ED(M)$ under base changes in $L^\infty(M)$. The freedom A_μ shall be used to simplify field equations.

From Eq. (2.14) a natural basis for $ED(M)$ may be defined by

$$D_\mu=\partial_\mu+\mathcal{L}_{A_\mu}, \tag{2.16}$$

$$D_\alpha=L_{\theta_\alpha}. \tag{2.17}$$

D_μ is the usual "covariant derivative" of local gauge theories once represented in a vector space. Any extended vector field X may be expressed in terms of D_μ and D_α by

$$X=\xi^a D_a,\quad a=(\mu,\alpha). \tag{2.18}$$

Indices $a, b, c, \cdots$ range over the space–time indices $\mu,\nu,\lambda,\cdots$ and the Lie algebra indices $\alpha,\beta,\gamma,\cdots$. Notice that $\mathcal{L}_{\Omega^\alpha\theta_\alpha}=\Omega^\alpha\mathcal{L}_{\theta_\alpha}$. Equation (2.18) may be used to map $C^\infty(M)$ into $C^\infty(M)$ if

$$D_\mu f=f_{,\mu},\quad f\in C^\infty(M),$$
$$\text{and} \tag{2.19}$$
$$D_\alpha f=0$$

are assumed.

Lie products for the basis D_a may be calculated with the help of $[\mathcal{L}_\theta,\mathcal{L}_\phi]=\mathcal{L}_{[\theta,\phi]}$ to get

$$\begin{aligned}&[D_\mu,D_\nu]=F^\alpha_{\mu\nu}D_\alpha,\\&[D_\mu,D_\alpha]=A^\beta_\mu C^\gamma_{\beta\alpha}D_\gamma,\\&[D_\alpha,D_\beta]=C^\gamma_{\alpha\beta}D_\gamma,\end{aligned} \tag{2.20}$$

where the following definitions have been used:

$$A_\mu=A^\alpha_\mu\theta_\alpha$$
$$\text{and} \tag{2.21}$$
$$F^\alpha_{\mu\nu}=A^\alpha_{\nu,\mu}-A^\alpha_{\mu,\nu}+A^\gamma_\mu A^\delta_\nu C^\alpha_{\gamma\delta}.$$

It shall now be shown that the usual gauge transformations of the Yang–Mills fields A^α_μ correspond to base transformation in $ED(M)$ of a certain kind.

Let us consider the transformations $\bar{\theta}_\alpha=L^\beta_\alpha\theta_\beta$ that leave the structure constants $C^\gamma_{\alpha\beta}$ invariant. Due to semi-simplicity of L these transformations are in one-to-one correspondence with $L^\infty(M)$, that is, for each L^β_α there is a $\phi\in\mathcal{L}^\infty(M)$ such that

$$e^\phi\theta_\alpha\bar{e}^\phi=L^\beta_\alpha\theta_\beta,\quad L^\beta_\alpha\in C^\infty(M), \tag{2.22}$$

where $e^\phi\theta_\alpha\bar{e}^\phi=\theta_\alpha+[\phi,\theta_\alpha]+\frac{1}{2}[\phi,[\phi,\theta_\alpha]]+\cdots$, and is therefore an element of $\mathcal{L}^\infty(M)$ when convergent. (Topological difficulties are ignored.) $\mathcal{L}^\beta_\alpha$, the adjoint representation matrices, obey integrability conditions, as a consequence of Eq. (2.22), which are derived below. By differentiating (2.22) the following is obtained:

$$[\bar{e}^\phi e^\phi_{,\mu},\theta_\alpha]=L^\beta_{\alpha,\mu}\bar{e}^\phi\theta_\beta e^\phi,$$

and therefore

$$H^\beta_\mu C^\gamma_{\beta\alpha}=L^\beta_{\alpha,\mu}L^{-1\gamma}_\beta, \tag{2.23}$$

where

$$\bar{e}^\phi e^\phi_{,\mu}=H^\beta_\mu\theta_\beta, \tag{2.24}$$

which when differentiated once again gives the integrability condition

$$H^\gamma_{\mu,\nu}-H^\gamma_{\nu,\mu}-H^\alpha_\mu H^\beta_\nu C^\gamma_{\alpha\beta}=0. \tag{2.25}$$

The transformation $\bar{\theta}_\alpha=L^\beta_\alpha\theta_\beta$ induces the transformation $\bar{D}_\alpha=L^\beta_\alpha D_\beta$ and $\bar{D}_\mu=D_\mu$ in $ED(M)$ which will be called gauge transformations. A simple computation based on Eqs. (2.20), (2.23), and (2.25) shows that under a gauge transformation A^α_μ transforms as follows:

$$\bar{A}^\alpha_\mu=L^{-1\alpha}_\beta A^\beta_\mu+H^\alpha_\mu \tag{2.26}$$

and that

$$\bar{F}^\alpha_{\mu\nu}=L^{-1\alpha}_\beta F^\beta_{\mu\nu}=\bar{A}^\alpha_{\nu,\mu}-\bar{A}^\alpha_{\mu,\nu}+\bar{A}^\gamma_\mu\bar{A}^\delta_\nu C^\alpha_{\gamma\delta}. \tag{2.27}$$

Remember that by definition $\bar{C}^\alpha_{\gamma\delta}=C^\alpha_{\gamma\delta}$. Equations (2.26) and (2.27) is the reason for the name gauge transformations.

3. EXTENDED EINSTEIN FIELD EQUATIONS

Fields are introduced through the concepts of affine connection ∇_X and metric $\langle X,Y\rangle$ just as in the Introduction. ∇_X is the mapping $\nabla_X: ED(M)\to ED(M)$ with properties (1.6). The metric $\langle X,Y\rangle\in C^\infty(M)$ is nondegenerate and obeys (1.8) and (1.9). The extended Christoffel symbols are Γ^c_{ab} defined by

$$\nabla_{D_a}(D_b)=\Gamma^c_{ab}D_c. \tag{3.1}$$

The metric tensor g_{ab} is by definition

$$g_{ab}=\langle D_a,D_b\rangle. \tag{3.2}$$

The structure fields C^c_{ab} are given by

$$[D_a,D_b]=C^c_{ab}D_c. \tag{3.3}$$

The torsion and Riemann tensors are defined by (1.10) and (1.11), and satisfy the same linear properties discussed in the Introduction. The component form for the torsion and Riemann tensors are the following:

$$T(D_a,D_b)=T^c_{ab}D_c \tag{3.4}$$

and

$$R(D_a,D_b)D_c=R^d_{cab}D_d, \tag{3.5}$$

which in terms of Γ^c_{ab} and C^c_{ab} read

$$T^c_{ab}=\Gamma^c_{ab}-\Gamma^c_{ba}-C^c_{ab} \tag{3.6}$$

and

$$R^d_{cab} = D_a(\Gamma^d_{bc}) - D_b(\Gamma^d_{ac}) + \Gamma^e_{bc}\Gamma^d_{ae} - \Gamma^e_{ac}\Gamma^d_{bd} - C^e_{ab}\Gamma^d_{ec}. \tag{3.7}$$

The connection between Γ^c_{ab} and g_{ab}, analogous to Eq. (1.14), is now

$$\Gamma^c_{ab} = \tfrac{1}{2}g^{cd}[D_a g_{db} + D_b g_{da} - D_d g_{ab}] + \tfrac{1}{2}g^{cd}[g_{ae}C^e_{db} + g_{be}C^e_{da}] + \tfrac{1}{2}C^c_{ab}, \tag{3.8}$$

where $T^c_{ab} = 0$ and formula (1.13) has been used.

The Einstein action for EVF is just

$$S = \int d^4x\sqrt{g}R, \tag{3.9}$$

where $g = \det g_{ab}$ and $R = g^{cb}R^a_{cab}$. No more invariants linear in R^a_{cbd} can be built from the Riemann tensor because it satisfies the same symmetry properties (and Bianchi identities) as the ordinary Riemann tensor. S is a coordinate invariant because $\sqrt{g}$ transforms under coordinate changes just like its counterpart in general relativity. By construction R is invariant under any base change in $ED(M)$, but $\sqrt{g}$ is invariant only under those transformations with determinant ± 1. Fortunately, the semisimplicity of L implies that $\det L^\beta_\alpha = 1$, where the L^β_α are the local adjoint representation matrices of L defined in Eq. (2.22). That is, S is invariant under gauge transformations. S appears to be functional of the fields g_{ab} and A^α_μ, but this is not quite so because A^α_μ are arbitrary. Said differently, under the base transformation $\bar{D}_\mu = D_\mu + E^\alpha_\mu D_\alpha$, $\bar{D}_\alpha = D_\alpha$, which leaves the Lagrangian invariant, may be used to set $A^\alpha_\mu = 0$, or alternatively $g_{\mu\alpha} = 0$, because $g_{\mu\alpha}$ transforms according to

$$\bar{g}_{\mu\alpha} = \langle \bar{D}_\mu, \bar{D}_\alpha \rangle = g_{\mu\alpha} + E^\beta_\mu g_{\beta\alpha}.$$

The second condition $g_{\mu\alpha} = 0$ is preferable because it is coordinate and gauge invariant. On the other hand, the condition $A^\alpha_\mu = 0$ is better in deriving the field equations by variation of S. The variation shall not be carried out explicitly since it is the same as in general relativity. The result is the field equation

$$R_{ab} = 0, \tag{3.10}$$

where R_{ab} is the Ricci tensor R^c_{acb}.

The field equation (3.10) shall be written out explicitly when $g_{\mu\alpha} = 0$, but before we exhibit the Lagrangian in terms of the fields $g_{\mu\nu}$, $g_{\alpha\beta}$, and A^α_μ under the same assumption $g_{\mu\alpha} = 0$. It should be clear that this last condition involves no loss of generality.

Just like in general relativity the Lagrangian $\sqrt{g}R$ of Eq. (3.9) involves second derivatives in the fields but these are exact divergences that may be ignored. The elimination of these terms is rather involved if use is not made of the following variation identities[6]:

$$\delta\Gamma^d_{cb} = \tfrac{1}{2}g^{da}[(\delta g_{ab})_{;c} + (\delta g_{ac})_{;b} - (\delta g_{bc})_{;a}]$$

and

$$\delta R_{cb} = (\delta\Gamma^a_{bc})_{;a} - (\delta\Gamma^a_{ac})_{;b}. \tag{3.11}$$

The following fully covariant derivatives have been used:

$$\xi_{a;b} = \xi_{a,b} - \Gamma^c_{ba}\xi_c$$

and

$$\xi^a_{\ ;b} = \xi^a_{\ ,b} + \Gamma^a_{bc}\xi^c, \tag{3.12}$$

where ordinary differentiation with respect to group indices vanish. All other tensors derivatives behave like derivatives of products of covariant vector fields, that is, like $(\xi^a\eta_b\gamma_c)_{;d}$. After elimination of divergences the Lagrangian (3.9) takes the form

$$\mathcal{L} = \sqrt{g}[\Gamma^a_{bd}\Gamma^d_{ac}g^{bc} - \Gamma^d_{bc}\Gamma^a_{ad}g^{bc}]. \tag{3.13}$$

The Christoffel symbols are given in the Appendix. With the help of them the Lagrangian (3.13) takes the following explicit form:

$$\mathcal{L} = \sqrt{g_Y}\mathcal{L}_E - \tfrac{1}{4}\sqrt{g}F_{\alpha\mu\nu}F^{\alpha\mu\nu} - \tfrac{1}{4}\sqrt{g}g^{\mu\nu}g^{\alpha\beta}g^{\gamma\delta}g_{\alpha\gamma|\mu}g_{\beta\delta|\nu} - \tfrac{1}{4}\sqrt{g}[g_{\gamma\gamma'}g^{\alpha\alpha'}g^{\beta\beta'}C^\gamma_{\alpha\beta}C^{\gamma'}_{\alpha'\beta'} + 2g^{\beta\alpha}C^\gamma_{\beta\alpha}C^\beta_{\gamma\delta}] + \frac{1}{\sqrt{g_Y}}\frac{\partial\sqrt{g_Y}}{\partial x^\mu}\frac{\partial\sqrt{g}}{\partial x^\nu}g^{\mu\nu} - \frac{\partial\sqrt{g_Y}}{\partial x^\lambda}\sqrt{g_E}\{\Gamma^\lambda_{\mu\nu}\}g^{\mu\nu}, \tag{3.14}$$

where the following definitions have been used:

$$g_E = \det g_{\mu\nu}, \qquad g_Y = \det g_{\alpha\beta}, \tag{3.15}$$

and

$$g_{\alpha\beta|\mu} = g_{\alpha\beta,\mu} - A^\gamma_\mu(C^\delta_{\gamma\beta}g_{\delta\alpha} + C^\delta_{\gamma\alpha}g_{\delta\beta}). \tag{3.16}$$

The last expression is a covariant derivative for gauge and coordinate transformations, and corresponds exactly to the representation of D_μ in the linear space defined by $g_{\alpha\beta}$. (3.16) is the rule of minimal coupling for $g_{\alpha\beta}$ commonly used in flat space gauge theories. Indices are raised or lowered with $g_{\mu\nu}$ or $g_{\alpha\beta}$ according to their nature. $\mathcal{L}_E$ stands for the Einstein Lagrangian.

A word should be said on the relation between fully covariant derivatives given by Eqs. (3.12) and the coordinate-gauge invariant derivative (3.16). First of all, the full covariance of the derivatives (3.12) can be seen from the identity

$$\nabla_{D_a}(\xi^b D_b) = (\xi^c_{\ ,a} + \Gamma^c_{ab}\xi^b)D_c \tag{3.17}$$

derived from rules (1.6) and (3.1). By full covariance we mean covariance under any base change in $ED(M)$. The Ricci scalar R is fully invariant. On the other hand, the really physically relevant transformations are those induced by coordinate changes and base changes in the Lie algebra L which correspond to degrees of indeterminism introduced in the theory right from the beginning. If transformations are restricted to coordinate-gauge transformations, then group indices and space–time indices do not mix, that is, the ξ^α transform among themselves without mixing with ξ^μ and vice versa. Under this restricted group, quantities like

$$\xi^\gamma_{\ |\mu} \equiv \xi^\gamma_{\ ,\mu} + \Gamma^\gamma_{\mu\beta}\xi^\beta \tag{3.18}$$

or

$$\xi^\mu_{\ |\gamma} = \xi^\mu_{\ ,\gamma} + \Gamma^\mu_{\gamma\lambda}\xi^\lambda \tag{3.19}$$

where $\xi^\mu_{\ ,\gamma} = 0$, behave covariantly. Notice that the relation between (3.12), (3.18), and (3.19) is that the summation over index "c" of Eq. (3.12) is performed only over

the group index β in Eq. (3.18) and over a space–time index in Eq. (3.19). This is what is meant by a "bar" covariant derivative in this paper. The field equations (3.10) when spelled out in terms of $g_{\mu\nu}$, A^{α}_{μ}, and $g_{\alpha\beta}$ read

$$0 = R_{\mu\nu} = R^{E}_{\mu\nu} - \left[\frac{1}{\sqrt{g_Y}}\frac{\partial\sqrt{g_Y}}{\partial x^{\mu}}\right]_{|\nu} - \tfrac{1}{2}F^{\alpha}_{\lambda\nu}F_{\alpha}{}^{\lambda}{}_{\mu}$$
$$- \tfrac{1}{4}g^{\beta\gamma}g^{\alpha\delta}g_{\gamma\alpha|\mu}g_{\delta\beta|\nu},$$
$$0 = R_{\alpha\mu} = \tfrac{1}{2}\left(F^{\nu}_{\alpha\mu|\nu} + F^{\nu}_{\alpha\mu}\frac{1}{\sqrt{g_Y}}\frac{\partial\sqrt{g_Y}}{\partial x^{\nu}} + \tfrac{1}{2}F^{\beta}_{\nu\mu}g_{\alpha\beta|}{}^{\nu}\right.$$
$$\left. - g^{\gamma\delta}g_{\beta\gamma|\mu}C^{\beta}{}_{\delta\alpha}\right),$$

and

$$0 = R_{\alpha\beta} = \tfrac{1}{2}\left(-\frac{1}{\sqrt{g_Y}}(\sqrt{g_Y}g_{\alpha\beta|}{}^{\nu})_{|\nu} + \tfrac{1}{2}F_{\alpha\mu\nu}F^{\mu\nu}_{\beta} + \tfrac{1}{2}C^{\delta\gamma}_{\alpha}C_{\beta\delta\gamma}\right.$$
$$\left. - C^{\delta\gamma}_{\alpha}C_{\delta\gamma\beta}\right).$$

4. CONCLUSION

The mathematical structure constructed in this paper, very similar to Einstein's theory of gravitation, renders an unambiguous way of associating besides spin-1 bosons of the Yang–Mills type, spin-0 bosons with definite transformation properties, with any semisimple group of symmetry. The theory is fully coupled to spin-2 gravitons, and therefore this all happens in a curved space–time. There is no problem in extending the formalism so as to include nonsemisimple groups of the type $U(1)\times$ (semisimple) and thus account for photons in the usual way. Hopefully elementary bosons could be fit into this scheme. Fermions seem to have no natural place in this formalism. The spontaneous symmetry breaking for the Lagrangian (3.14) is currently under study.

Note added in proof: After this work was completed the author was advised of a paper by R. Kerner, Ann. Inst. H. Poincaré **9**, 143 (1968), which contains related ideas.

ACKNOWLEDGMENTS

My thanks to R. K. Sacks for having taught me the power and beauty of geometry and to P. Cordero for having introduced me to gauge theories. My thanks to R. Benguria and D. Villarroel for many stimulating discussions. I am most grateful for the happy patience of Sonia Encalada in typing this manuscript.

APPENDIX

Christoffel symbols defined by equation (3.8) are given with all indices lowered by $g_{\mu\nu}$ or $g_{\alpha\beta}$ according to their nature:

$$\Gamma_{\lambda\mu\nu} = \{\lambda, \mu\nu\}, \qquad \Gamma_{\alpha\mu\nu} = \tfrac{1}{2}F_{\alpha\mu\nu},$$
$$\Gamma_{\lambda\mu\alpha} = \Gamma_{\lambda\alpha\mu} = \tfrac{1}{2}F_{\alpha\lambda\mu}, \qquad \Gamma_{\alpha\beta\nu} = \tfrac{1}{2}g_{\alpha\beta|\nu},$$
$$\Gamma_{\lambda\alpha\beta} = -\tfrac{1}{2}g_{\alpha\beta|\lambda}, \qquad \Gamma_{\alpha\nu\beta} = \tfrac{1}{2}[g_{\alpha\beta,\nu} + A^{\gamma}_{\nu}(C_{\alpha\gamma\beta} - C_{\beta\gamma\alpha}]$$
$$\Gamma_{\alpha\beta\gamma} = \tfrac{1}{2}(C_{\alpha\beta\gamma} + C_{\beta\alpha\gamma} + C_{\gamma\alpha\beta}),$$

where $\{\lambda, \mu\nu\}$ are the Christoffel symbols of the second kind built out of $g_{\mu\nu}$ only. Bar covariant derivatives are defined by Eq. (3.18).

*Present address is Departmento de Física, Universidade de Brasilia, Basilia 70.000.

[1]Singurdur Helgason, *Differential Geometry and Symmetric Spaces* (Academic, New York, 1962).
[2]Robert Hermann, *Differential Geometry and the Calculus of Variations* (Academic, New York, 1968).
[3]C. N. Yang, Phys. Rev. Lett. **33**, 445 (1974).
[4]S. Weinberg, Phys. Rev. D **7**, 1068 (1973).
[5]H. Fritzsch and P. Minkowski, Phys. Rev. Lett. B **53**, 373 (1974).
[6]A. H. Taub, "Relativistic Fluid Dynamics. I," Ciclo Bressanone 7-16, June 1970.

Generalized Källén–Pauli equation*†

Cvavb Chandra Raju

Department of Physics, University of Udaipur, Udaipur 313001, India
(Received 27 March 1975; revised manuscript received 29 May 1975)

The Källén–Pauli (KP) equation in the 2 V particle model was solved. In addition to satisfying all the requirements of the integral equation, this solution is found to be reducing to that of the ordinary KP equation. Unlike earlier authors, we found that there is a resonance in the $V\theta$ sector also. The solution given here shows that the deductive method used in the case of the ordinary KP equation does not hold good in the present case. The uniqueness of the solution is yet to be proved.

PACS numbers: 11.10.

I. INTRODUCTION

An equation similar to the ordinary Källén–Pauli equation can be obtained in the case of the Lee model with two V particles. This generalized KP equation was solved by many authors.[1–3] The solution obtained in Ref. 1 was shown by the author of Ref. 2 to be inadequate. Following the procedure of Ref. 2, the authors of Ref. 3 obtained a solution and applied it to the study of overlapping resonances in three-particle final states.

We have noted that the solutions of the GKP found in Refs. 1, 2, and 3 do not reduce to the solution of the ordinary KP equation when either of the interaction constants in the model is set equal to zero. This certainly implies the solution obtained in Ref. 3 is wrong.

In the present work we have applied a method used in Ref. 4. The deductive method applied in Ref. 5 and others[6–9] did not yield the correct solution.

In Ref. 3 the constant B [see Eq. (18)] is not handled properly. Hence we gave the steps leading to it elaborately. The solution found here correctly reduces to that of the ordinary KP equation. For completeness we have also given here the $N\theta$ sector as well. Because of the very nature of the method, the uniqueness cannot be decided until and unless the unitarity is proved. We have not yet been able to prove the unitarity in the general case here.

In Secs. II and III we give the model and the $N\theta$ sector. In Secs. IV, V, and VI the GKP equation and its solution are given. In Secs. VII and VIII the $V\theta$ scattering amplitude and the $N\theta\theta$ production amplitude are evaluated. In IX we examine the question of uniqueness by means of the unitarity at the resonance energy in the $N\theta$ sector.

II. THE MODEL

In this model there are two V particles. Because of the introduction of a second V particle, there will be a two-particle resonance in the $N\theta$ sector. Both the V particles are subjected to the elementary interactions $V_1 \rightleftarrows N\theta$ and $V_2 \rightleftarrows N\theta$. Here we choose the parameters such that only one of the V particles is stable. The model is described by the following Hamiltonian:

$$H=\sum_{i=1}^{2} m_i V_i^{+}V_i + m_N N^{+}N+\sum_{\kappa}\omega_\kappa a_\kappa^{+}a_\kappa + \sum_{i=1}^{2}\sum_{\kappa}\frac{u(\omega_\kappa)}{(2\omega_\kappa\Omega)^{1/2}} g_i V_i^{+} N a_\kappa + \sum_{i=1}^{2}\sum_{\kappa}\frac{u(\omega_\kappa)}{(2\omega_\kappa\Omega)^{1/2}} g_i V_i N^{+} a_\kappa^{+}, \tag{1}$$

where $\omega_\kappa^2=\mu^2+\kappa^2$, μ being the mass of the θ particle. In (1) m_1 and m_2 are the bare masses of V_1 and V_2. Without loss of generality we can set the mass of the N-particle m_N equal to zero. The coupling constants g_1 and g_2 are real. It is apparent from the Hamiltonian that the V and N particles are taken to be static and the θ particle obeys a relativistic energy-momentum relation. The cutoff function $u(\omega)$ is so chosen as to make all the relevant integrals finite, and also so as not to allow any ghost states. The equal time commutators obeyed by the field operators are

$$[a_\kappa, a_{\kappa'}^{+}]=\delta(\kappa-\kappa'),$$
$$[V_1, V_1^{+}]=[V_2, V_2^{+}]=[N, N^{+}]=1. \tag{2}$$

All other commutators between the field operators vanish. Moreover, if one wishes, for the V's and N particle, anticommutators can be taken.

III. $N\theta$ SECTOR

The lowest sector is spanned by the physical $|\mathbf{V}\rangle$ particle state and the $N\theta$ scattering states. We choose the stable $|\mathbf{V}\rangle$ particle state to have a mass equal to zero. This restriction does not give rise to any special effects. The physical $|\mathbf{V}\rangle$ state can be written as a linear combination of the bare states of the sector, and so we get

$$|\mathbf{V}\rangle=\alpha_1 V_1^{+}|0\rangle+\alpha_2 V_2^{+}|0\rangle+\sum_{\kappa}\phi_1(\omega_\kappa)N^{+}a^{+}|0\rangle. \tag{3}$$

From the Schrödinger equation

$$H|\mathbf{V}\rangle=E|\mathbf{V}\rangle \tag{4}$$

we find that

$$\phi_1(\omega)=-(\alpha_1 g_1+\alpha_2 g_2)u(\omega)/(2\omega\Omega)^{1/2}(\omega-E) \tag{5}$$

and we define

$$H_1(E)=(E-m_1)(E-m_2)+[g_1^2(E-m_2)+g_2^2(E-m_1)]\times\frac{1}{4\pi^2}\int_{\mu}^{\infty}\frac{\kappa u^2(\omega)\,d\omega}{(\omega-E)}. \tag{6}$$

Equation (6) is the definition of $H_1(E)$. In the present case we take the eigenvalue E to be *zero*. From Eq. (4) we obtain $H_1(0)=0$, and this condition reduces (6) to

$$\left(\frac{g_1^2}{m_1}+\frac{g_2^2}{m_2}\right)^{-1}=I_1(0), \tag{7}$$

where

$$I_1(E)=\frac{1}{4\pi^2}\int_\mu^\infty \frac{\kappa u^2(\omega)}{\omega-E}\,d\omega. \tag{8}$$

The normalization of the physical $|\mathbf{V}\rangle$ state restricts that

$$(\alpha_1 g_1+\alpha_2 g_2)^{-2}=\frac{(g_1^2/m_1^2+g_2^2/m_2^2)}{(g_1^2/m_1+g_2^2/m_2)^2}+\frac{1}{4\pi^2}\int_\mu^\infty\left|\frac{\kappa u^2(\omega)}{\omega^2}\right|d\omega, \tag{9}$$

as can be easily verified.

We also define a function $G(\omega)$ by the relation,

$$\begin{aligned}G(\omega+i\epsilon)&=\frac{H_1(\omega+i\epsilon)}{g_1^2(\omega-m_2)+g_2^2(\omega-m_1)}\\&=\frac{(\omega-m_1)(\omega-m_2)}{(g_1^2+g_2^2)(\omega-\omega_C)}+\frac{1}{4\pi^2}\int_\mu^\infty\frac{\kappa' u^2(\omega')}{\omega'-\omega-i\epsilon}\,d\omega',\end{aligned} \tag{10}$$

where, of course, we need that

$$G(0)=0.$$

In (10)

$$\omega_C=(g_1^2m_2+g_2^2m_1)/(g_1^2+g_2^2). \tag{11}$$

From Eq. (11) one notices immediately that

$$\omega_C=m_1 \quad \text{when } g_1^2=0$$

and (12)

$$\omega_C=m_2 \quad \text{when } g_2^2=0.$$

Subtracting $G(0)$ from $G(\omega)$ of (10) yields

$$G(\omega+i\epsilon)=\omega h(\omega+i\epsilon), \tag{13}$$

where

$$\begin{aligned}h(\omega+i\epsilon)&=\frac{\omega_C\omega-(m_1+m_2)\omega_C+m_1m_2}{(g_1^2+g_2^2)\omega_C(\omega-\omega_C)}\\&\quad+\frac{1}{4\pi^2}\int_\mu^\infty\frac{\kappa' u^2(\omega')}{\omega'(\omega'-\omega-i\epsilon)}\,d\omega'.\end{aligned} \tag{14}$$

Using (14), we can define a renormalized charge. We take from (14) that

$$h(0)=1/g^2. \tag{15}$$

We can easily show that right-hand side of Eq. (15) is identical to Eq. (9). Thus we find that

$$g^{-2}=(\alpha_1 g_1+\alpha_2 g_2)^{-2}. \tag{16}$$

For later convenience we rewrite (14) as $h(\omega+i\epsilon)+1/g^2-h(0)$, and this gives

$$h(\omega+i\epsilon)=\frac{1}{g^2}\left(1-\frac{g^2B\omega}{\omega_C-\omega}+\frac{g^2\omega}{4\pi^2}\int_\mu^\infty\frac{\kappa' u^2(\omega')}{\omega'^2(\omega'-\omega-i\epsilon)}\,d\omega'\right), \tag{17}$$

where in (17) we have

$$B=\frac{\omega_C^2-(m_1+m_2)\omega_C+m_1m_2}{(g_1^2+g_2^2)\omega_C^2}. \tag{18}$$

It may be noted from (18) that $B=0$ whenever either $g_1^2=0$ or $g_2^2=0$. In other words, the pole term in (17) vanishes when either of the interaction constants is switched off. In what follows we always take ω_C less than μ or very near μ and ω_C cannot be taken equal to zero, since then B will be infinite. This is because of our choice of the physical mass of the $|\mathbf{V}\rangle$ particle to be zero. Hereafter, whenever we say that B is zero, we mean either of the interaction constants in (1) is set equal to zero. In Ref. 3, it is pointed out that $B>0$. (see Appendix B there). If this restriction is imposed, we cannot, in fact, reduce our results to the one-particle case. Their restriction on B is unnecessary.

Whenever we refer to the functions $h(\omega)$ and $G(\omega)$, we mean Eqs. (17) and (13), respectively. From Eq. (17) one immediately notes that whenever B is zero the function in Eq. (17) will be identical to a similar function in the ordinary KP equation but for a numerical multiplicative factor.

We immediately find from Eqs. (13) and (17) that

$$\mathrm{Im}[G(\omega+i\epsilon)]=\theta(\omega)(4\pi)^{-1}\kappa u^2(\omega) \tag{19}$$

and

$$\mathrm{Im}[h(\omega+i\epsilon)]=\theta(\omega)(4\pi\omega)^{-1}\kappa u^2(\omega) \tag{20}$$

The $N\theta$ scattering matrix in the $2V$-particle case can be found out from earlier works. However, to acquaint the reader with our notation, we give here a brief derivation of the amplitude. For what follows it will be convenient to set

$$f(\omega)=u(\omega)/(2\omega\Omega)^{1/2}. \tag{21}$$

In the continuum we have

$$\begin{aligned}|N\theta^{\mathrm{in}}_{\kappa_0}\rangle&=a^+_{\kappa_0}N^+|0\rangle+\beta_1(\kappa_0)V_1^+|0\rangle+\beta_2(\kappa_0)V_2^+|0\rangle\\&\quad+\int\alpha(\kappa_0,\kappa')a^+_{\kappa'}N^+|0\rangle\,d^3\kappa',\end{aligned} \tag{22}$$

while the Schrödinger equation

$$H|N\theta^{\mathrm{in}}_{\kappa_0}\rangle=\omega_0|N\theta^{\mathrm{in}}_{\kappa_0}\rangle$$

gives

$$\beta_1(\kappa_0)=g_1f(\omega_0)(\omega_0-m_2)[H_1(\omega_0+i\epsilon)]^{-1}, \tag{23}$$

$$\beta_2(\kappa_0)=g_2f(\omega_0)(\omega_0-m_1)[H_1(\omega_0+i\epsilon)]^{-1}, \tag{24}$$

and

$$\alpha(\kappa,\kappa_0)=\frac{f(\omega_\kappa)f(\omega_0)}{(\omega_0-\omega+i\epsilon)}\,[G(\omega_0+i\epsilon)]^{-1}. \tag{25}$$

The $N\theta$ scattering amplitude is defined by

$$\begin{aligned}\langle N\theta^{\mathrm{out}}_{\kappa}|N\theta^{\mathrm{in}}_{\kappa_0}\rangle&=S_{N\theta}\\&=\delta(\kappa-\kappa_0)+2\pi i\delta(\omega-\omega_0)T_1(\omega_0).\end{aligned} \tag{26}$$

The T matrix, in (26), is given by

$$T_1(\omega_0)=-f^2(\omega_0)[G(\omega_0+i\epsilon)]^{-1}. \tag{27}$$

From Eq. (27) we notice that there is a resonance in the $N\theta$ sector for values of ω_0 around ω_C.

IV. $V\theta$ SECTOR

For the scattering of a θ particle on a V particle we want an eigenstate of the total Hamiltonian of the form

$$|V\theta^{\mathrm{in}}_{\kappa_0}\rangle=a^+_{\kappa_0}|\mathbf{V}\rangle+|\chi^+\rangle \tag{28}$$

with

$$H|V\theta^{\mathrm{in}}_{\kappa_0}\rangle=\omega_0|V\theta^{\mathrm{in}}_{\kappa_0}\rangle. \tag{29}$$

$|\mathbf{V}\rangle$ denotes the stationary eigenstate of the total Hamiltonian describing a physical V particle of mass zero and

$$|\chi^+\rangle = \int \psi_1(\kappa', \kappa_0)|V_1\theta_{\kappa'}\rangle\, d^3\kappa' + \int \psi_2(\kappa', \kappa_0)|V_2\theta_{\kappa'}\rangle\, d^3\kappa' + \int\int \psi_3(\kappa', \kappa'', \kappa_0)|N\theta_{\kappa'}\theta_{\kappa''}\rangle\, d^3\kappa'\, d^3\kappa''. \tag{30}$$

with outgoing waves only in ψ_1, ψ_2, and ψ_3.

The Schrödinger eigenvalue equation gives

$$(\omega_0 - \omega - m_1)\psi_1(\kappa, \kappa_0) = \frac{-gg_1 f(\omega_0)f(\omega)}{\omega} + 2g_1 \int f(\omega')\psi_3(\kappa', \kappa, \kappa_0)\, d^3\kappa', \tag{31}$$

$$(\omega_0 - \omega - m_2)\psi_2(\kappa, \kappa_0) = \frac{-gg_2 f(\omega_0)f(\omega)}{\omega} + 2g_2 \int f(\omega')\psi_3(\kappa', \kappa, \kappa_0)\, d^3\kappa', \tag{32}$$

and

$$2(\omega_0 - \omega - \omega')\psi_3(\kappa, \kappa', \kappa_0) = g_1[f(\omega)\psi_1(\kappa', \kappa_0) + f(\omega')\psi_1(\kappa, \kappa_0)] + g_2[f(\omega)\psi_2(\kappa', \kappa_0) + f(\omega')\psi_2(\kappa, \kappa_0)]. \tag{33}$$

By comparing (29) and (30) we find that

$$\psi_2(\kappa, \kappa_0) = \frac{g_2(\omega_0 - \omega - m_1)}{g_1(\omega_0 - \omega - m_2)}\, \psi_1(\kappa, \kappa_0). \tag{34}$$

After a little algebra we obtain from (34) and (31)

$$H_1(\omega_0 - \omega)\frac{\psi_1(\kappa, \kappa_0)}{(\omega_0 - \omega - m_2)} = \frac{-gg_1 f(\omega_0)f(\omega)}{\omega} - \int \frac{f(\omega')f(\omega)}{\omega' - \omega_0 + \omega - i\epsilon} \times \frac{\psi_1(\kappa', \kappa_0)}{(\omega_0 - \omega - m_2)}\,[g_1^2(\omega_0 - \omega' - m_2) + g_2^2(\omega_0 - \omega' - m_1)]\, d^3\kappa'. \tag{35}$$

Let

$$\psi_1(\kappa, \kappa_0) = \frac{gg_1 f(\omega_0)f(\omega)(\omega_0 - \omega - m_2)\phi(\kappa', \kappa_0)}{g_1^2(\omega_0 - \omega - m_2) + g_2^2(\omega_0 - \omega - m_1)}. \tag{36}$$

Inserting (36) into (35), we obtain

$$G(\omega_0 - \omega)\phi(\kappa, \kappa_0) = -\frac{1}{\omega} - \int \frac{f^2(\omega')\phi(\kappa', \kappa_0)}{\omega' - \omega_0 + \omega - i\epsilon}\, d^3\kappa'. \tag{37}$$

Doing the angular integrals, we find

$$G(\omega_0 - \omega)\phi(\omega, \omega_0) = -\frac{1}{\omega} - \frac{1}{\pi}\int_\mu^\infty \frac{\mathrm{Im}[G(\omega')]}{(\omega' - \omega_0 + \omega - i\epsilon)}\, \phi(\omega', \omega_0)\, d\omega', \tag{38}$$

where in arriving at (38) we have assumed that $\phi(\kappa, \kappa_0)$ is independent of the vectors. Let

$$\phi(\omega, \omega_0) = -M(\omega, \omega_0)/\omega(\omega_0 - \omega). \tag{39}$$

We find for (38),

$$h(\omega_0 - \omega)M(\omega, \omega_0) = 1 + \frac{\omega}{\pi}\int_\mu^\infty \frac{\mathrm{Im}[h(\omega')]}{(\omega' - \omega_0 - i\epsilon)}\, \frac{M(\omega', \omega_0)}{(\omega' - \omega_0 + \omega - i\epsilon)}\, d\omega'. \tag{40}$$

Equation (40) is what we call the generalized KP equation. One immediately notices that (40) reduces to the usual KP equation when $B = 0$. The function $M(\omega, \omega_0)$ vanishes for $\omega = \omega_0 - \omega_C$. This is because the function $M(\omega, \omega_0)$ is proportional to $h^{-1}(\omega_0 - \omega)$. [See Eq. (40).] The function $h^{-1}(\omega_0 - \omega)$ has a zero at $\omega = \omega_0 - \omega_C$. [See Eq. (47).] Obviously $\phi(\omega, \omega_0)$ also will be zero for $\omega = \omega_0 - \omega_C$. This is an extra condition that the GKP equation has to satisfy.

V. SOLUTION OF THE INTEGRAL EQUATION

Before solving the integral equation, we define certain functions which we need later.

We take

$$G_1(\omega) = (\omega_C - \omega)G(\omega + i\epsilon) \tag{41}$$

and

$$h_1(\omega) = (\omega_C - \omega)h(\omega + i\epsilon). \tag{42}$$

From (42) one immediately notes that $h_1(\omega)$ does not have a pole at $\omega_C = \omega$ unlike $h(\omega)$. Also we find that

$$\frac{1}{G_1(\omega)} = \frac{g^2}{\omega_C \omega} + \frac{1}{\pi}\int_\mu^\infty \mathrm{Im}\frac{1}{G_1(\omega')}\, \frac{d\omega'}{\omega' - \omega - i\epsilon}. \tag{43}$$

Since $G_1^{-1}(\omega)$ has a pole at $\omega = 0$ and

$$\frac{d}{d\omega}[G_1(\omega)]_{\omega=0} = \frac{\omega_C}{g^2}, \tag{44}$$

the relation in (43) can be verified by doing the integral as a contour integral. Multiplying Eq. (43) by ω, one obtains

$$\frac{1}{h_1(\omega)} = \frac{g^2}{\omega_C} + \frac{\omega}{\pi}\int_\mu^\infty \mathrm{Im}\frac{1}{G_1(\omega')}\, \frac{d\omega'}{\omega' - \omega - i\epsilon}. \tag{45}$$

In (45), if we write $(\omega_0 - \omega)$ for ω, we obtain

$$\frac{1}{h_1(\omega_0 - \omega)} = \frac{g^2}{\omega_C} + \frac{(\omega_0 - \omega)}{\pi}\int_\mu^\infty \mathrm{Im}\frac{1}{G_1(\omega')}\, \frac{d\omega'}{\omega' - \omega_0 + \omega - i\epsilon}. \tag{46}$$

From the relation (46) we easily find that

$$\frac{1}{h(\omega_0 - \omega)} = (\omega_C - \omega_0 + \omega) \times \left(\frac{g^2}{\omega_C} + \frac{(\omega_0 - \omega)}{\pi}\int_\mu^\infty \mathrm{Im}\frac{1}{G_1(\omega')}\, \frac{d\omega'}{\omega' - \omega_0 + \omega - i\epsilon}\right). \tag{47}$$

In order to solve the integral equation (40), one might be tempted to use (47) and follow the procedure of Ref. 4. But this procedure cannot give the correct solution for the simple reason the relation (47) does not reduce at all to a similar relation of the one V-particle equation whenever B is made zero. In other words although the function $h(\omega)$ reduces to a similar function of one V particle whenever B is zero, the function $h^{-1}(\omega)$ does not reduce to the inverse of a similar function of one V particle whenever B is zero (see Appendix C). Therefore, if we solve the integral equation (40) with the help of Eq. (47), the solution so obtained will not reduce to the one V-particle solution whenever B is made zero. Since the integral equation itself reduces to the ordinary KP equation whenever B is zero, we certainly require that the solution also must be reducible to the solution of the ordinary KP equation, whenever B is zero.

In order to find a solution which reduces to the ordinary KP solution whenever B is zero, we need another representation for the inverse of $h(\omega)$. To this end we define a function $\zeta(\omega)$, where

$$\mathcal{G}(\omega) = h(\omega) + \frac{B\omega}{(\omega_C - \omega)}$$
$$= \frac{1}{g^2}\left(1 + \frac{g^2\omega}{4\pi^2}\int_\mu^\infty \frac{\kappa' u^2(\omega')\,d\omega'}{\omega'^2(\omega' - \omega - i\epsilon)}\right). \quad (48)$$

The function given in (48) differs from $h(\omega)$ of Eq. (17) in that it does not have the pole term. Moreover, whenever B is zero, $g^2\mathcal{G}(\omega)$ will be indentical to the function $G_P(\omega)$ used in Refs. 4 and 5 and others if g^2 is taken as the renormalized coupling constant. But g^2 in (48) is as defined in Eq. (16). We define another function $K(\omega)$, where

$$K(\omega) = \omega\mathcal{G}(\omega). \quad (49)$$

From (48) and (49) we note that

$$\mathcal{G}(0) = h(0) = 1/g^2, \quad (50)$$

$$K(0) = 0, \quad (51)$$

and

$$\mathrm{Im}\,\mathcal{G}(\omega + i\epsilon) = (4\pi\omega)^{-1}\kappa u^2(\omega) = \mathrm{Im}[h(\omega + i\epsilon)]. \quad (52)$$

We also find easily that,

$$\frac{1}{\mathcal{G}(\omega_0 - \omega)} = g^2 + \frac{(\omega_0 - \omega)}{\pi}\int_\mu^\infty \mathrm{Im}\frac{1}{K(\omega')}\frac{d\omega'}{\omega' - \omega_0 + \omega - i\epsilon}. \quad (53)$$

This relation follows directly from Ref. 4. From Eq. (48), we immediately find

$$\mathcal{G}(\omega)/h(\omega) = 1 + B\omega/h_1(\omega). \quad (54)$$

In passing we also note that $h(\omega)$ and $\mathcal{G}(\omega)$ both tend to two different but related constants as $\omega \to \infty$. In what follows we do not need these representations, anyway.

In order to solve Eq. (38), we can define, say,

$$M(\omega, \omega_0) = [\mathcal{G}(\omega_0 - \omega)/h(\omega_0 - \omega)]N(\omega, \omega_0)$$

and insert it into Eq. (40) and solve for $N(\omega, \omega_0)$ and from the latter we can find $M(\omega, \omega_0)$. But here we follow an equivalent and simpler procedure. To this end we write (40) using (48), as

$$\mathcal{G}(\omega_0 - \omega)M(\omega, \omega_0)$$
$$= \frac{(\omega_0 - \omega)BM(\omega, \omega_0)}{(\omega_C - \omega_0 + \omega)} + 1$$
$$+ \frac{\omega}{\pi}\int_\mu^\infty \frac{\mathrm{Im}[h(\omega')]}{(\omega' - \omega_0 - i\epsilon)}\frac{M(\omega', \omega_0)\,d\omega'}{(\omega' - \omega_0 + \omega - i\epsilon)} \quad (55)$$

and solve Eq. (55) by following a procedure similar to that of Ref. 4. Dividing through by $\mathcal{G}(\omega_0 - \omega)$, we obtain

$$M(\omega, \omega_0) = \frac{(\omega_0 - \omega)BM(\omega, \omega_0)}{(\omega_C - \omega_0 + \omega)\mathcal{G}(\omega_0 - \omega)} + \frac{1}{\mathcal{G}(\omega_0 - \omega)}$$
$$\times\left(1 + \frac{\omega}{\pi}\int \frac{\mathrm{Im}[h(\omega')]}{(\omega' - \omega_0 - i\epsilon)}\frac{M(\omega', \omega_0)\,d\omega'}{(\omega' - \omega_0 + \omega - i\epsilon)}\right). \quad (56)$$

Following Ref. 4 closely, the quantity inside the braces divided by $\mathcal{G}(\omega_0 - \omega)$ can be written at once as

$$C_1(\omega_0) + \frac{(\omega_0 - \omega)}{\pi}\int_\mu^\infty \mathrm{Im}\frac{1}{K(\omega')}\frac{d\omega' X(\omega', \omega_0)}{(\omega' - \omega_0 + \omega - i\epsilon)}, \quad (57)$$

where

$$X(\omega', \omega_0) = 1 + \mathcal{G}^*(\omega')M(\omega', \omega_0)$$
$$+ \frac{1}{\pi}\int_\mu^\infty \frac{\mathrm{Im}[h(\omega'')]}{\omega'' - \omega_0 - i\epsilon}M(\omega'', \omega_0)\,d\omega''$$
$$- \frac{1}{\pi}\int_\mu^\infty \frac{\mathrm{Im}[h(\omega'')]}{\omega'' - \omega' - i\epsilon}M(\omega'', \omega_0)\,d\omega''. \quad (58)$$

While arriving at the second term in (58), we made use of the relation (52). In (57) $C_1(\omega_0)$ is a constant function of ω_0 and ω_C and is given by

$$C_1(\omega_0) = g^2 + \frac{\omega_0 g^2}{\pi}\int_\mu^\infty \frac{\mathrm{Im}[h(\omega')]}{\omega'(\omega' - \omega_0)}M(\omega', \omega_0)\,d\omega'. \quad (59)$$

We have to make a judicious choice for $X(\omega', \omega_0)$ from among the functions that have occured in the theory. The possible candidates for the dimensionless function $X(\omega', \omega_0)$ are

$$C_2(\omega_0)h^{-1}(\omega_0 - \omega'), \quad C_2(\omega_0)\frac{\mathcal{G}(\omega')}{h(\omega')}h^{-1}(\omega_0 - \omega'),$$
$$C_2(\omega_0)\mathcal{G}^{-1}(\omega_0 - \omega'), \quad \text{and} \quad C_2(\omega_0)\frac{\mathcal{G}(\omega')}{h(\omega')}\mathcal{G}^{-1}(\omega_0 - \omega') \quad (60)$$

since each one of them reduces to a similar choice made in Ref. 4 apart from satisfying the requirements here. For each choice we have to examine the unitarity, and then decide one from among the choices in (60). But the unitarity here is not as simple as in Ref. 10. We have not been able to prove unitarity in this model. Nevertheless, for ω_C slightly greater than μ and less than 2μ we have been able to show that the unitarity holds good only for the first two choices in (60). The results for the second choice in (60) are given in Appendix A. Here we consider only the first choice.

So we take for (57)

$$C_1 + C_2\frac{(\omega_0 - \omega)}{\pi}\int_\mu^\infty \mathrm{Im}\frac{1}{K(\omega')}\frac{d\omega'}{(\omega' - \omega_0 + \omega - i\epsilon)}\frac{1}{h(\omega_0 - \omega')}. \quad (61)$$

Inserting Eq. (61) into (56), we obtain

$$M(\omega, \omega_0) = \frac{(\omega_0 - \omega)BM(\omega, \omega_0)}{(\omega_C - \omega_0 + \omega)\mathcal{G}(\omega_0 - \omega)} + C_1 - C_2(\omega_0 - \omega)I(\omega_0 - \omega), \quad (62)$$

where

$$I(\omega_0 - \omega) = -\frac{1}{\pi}\int_\mu^\infty \mathrm{Im}\frac{1}{K(\omega')}\frac{1}{h(\omega_0 - \omega')}\frac{d\omega'}{(\omega' - \omega_0 + \omega - i\epsilon)}. \quad (63)$$

Rearranging (62) and using (48), we immediately obtain the solution of our integral equation (40) as

$$M(\omega, \omega_0) = \frac{\mathcal{G}(\omega_0 - \omega)}{h(\omega_0 - \omega)}[C_1 - (\omega_0 - \omega)C_2 I(\omega_0 - \omega)]. \quad (64)$$

The solution given in (64) satisfies all the conditions discussed below Eq. (40). The factor before the parenthesis simply cancels out when $B = 0$.

VI. EVALUATION OF THE CONSTANTS

The solution given in (64) will be complete if we can determine the constants C_1 and C_2. We follow a different procedure to that of Ref. 4 in evaluating these con-

stants. In order to determine the two constants C_1 and C_2, we need two equations connecting them. To this end we evaluate $M(0)$ and $M(\omega_0)$ from (40) and then from (64). We have from (40) and (64).

$$M(0)=\frac{1}{h(\omega_0)}=\frac{\mathcal{G}(\omega_0)}{h(\omega_0)}(C_1-\omega_0 C_2 A_0), \tag{65}$$

where

$$A_0=\frac{1}{\pi}\int_\mu^\infty \mathrm{Im}\frac{1}{K(\omega')}\,\frac{1}{G(\omega_0-\omega')}\,d\omega'. \tag{66}$$

From Eq. (65) we obtain

$$C_1\mathcal{G}_0-C_2K_0A_0=1, \tag{67}$$

where in (67) and hereafter we write for simplicity

$$\mathcal{G}(\omega_0)=\mathcal{G}_0,\quad K(\omega_0)=K_0,\quad \text{and}\quad h(\omega_0)=h_0. \tag{68}$$

Inserting (64) into (40) and then taking $\omega=\omega_0$, we find

$$\begin{aligned}\frac{C_1}{g^2}=1&+C_1\frac{\omega_0}{\pi}\int_\mu^\infty \frac{\mathrm{Im}[h(\omega')]}{\omega'(\omega'-\omega_0-i\epsilon)}\,\frac{\mathcal{G}(\omega_0-\omega')}{h(\omega_0-\omega')}\,d\omega'\\&-C_2\frac{\omega_0}{\pi}\int_\mu^\infty \frac{\mathrm{Im}[h(\omega')]\mathcal{G}(\omega_0-\omega')}{\omega'(\omega'-\omega_0-i\epsilon)h(\omega_0-\omega')}\\&\times(\omega_0-\omega')I(\omega_0-\omega')\,d\omega'.\end{aligned} \tag{69}$$

The integrals in (69) can be easily computed by contour integration. The same method applies in Ref. 4 also (see Appendix D). But here in (69) the function $h(\omega')$ has a pole at ω_C. As in dispersion theory we exclude ω_C by a cross cut from the infinite circle. If this is not done, both the constants C_1 and C_2 attain an indeterminate form for $\omega_0=\omega_C$ for any value of ω_C. With this in mind, we evaluate the integrals as contour integrals. The very first integral in (69) is given by

$$\begin{aligned}&C_1\omega_0\sum_{\text{poles of } z^{-1}(z-\omega_0)^{-1}}\mathrm{Res}\frac{h(z)}{z(z-\omega_0)}\,\frac{\mathcal{G}(\omega_0-z)}{h(\omega_0-z)}\\&=C_1\omega_0\left(\frac{h_0}{\omega_0}-\frac{\mathcal{G}_0}{h_0g^2\omega_0}\right).\end{aligned}$$

The second integral in (69), by using (63) and interchanging the orders of integration, can be written as

$$\begin{aligned}&-\frac{C_2\omega_0}{\pi}\int_\mu^\infty \mathrm{Im}\left(\frac{1}{K(\omega'')}\right)\frac{d\omega''}{h(\omega_0-\omega'')}\\&\times\frac{1}{\pi}\int_\mu^\infty \frac{\mathrm{Im}[h(\omega')]\mathcal{G}(\omega_0-\omega')}{\omega'(\omega'-\omega_0+\omega'')h(\omega_0-\omega')}\,d\omega';\end{aligned} \tag{71}$$

and the very last integral in (71) can be done by residue method and we have

$$\begin{aligned}&\frac{1}{\pi}\int_\mu^\infty \frac{\mathrm{Im}[h(\omega')]\,d\omega'}{(\omega'-i\epsilon)(\omega'-\omega_0+\omega''-i\epsilon)}\,\frac{\mathcal{G}(\omega_0-\omega')}{h(\omega_0-\omega')}\\&=\sum_{\text{poles of } z^{-1}(z-\omega_0+\omega'')^{-1}}\mathrm{Res}\frac{h(z)\mathcal{G}(\omega_0-z)}{z(z-\omega_0+\omega'')h(\omega_0-z)}\\&=\frac{1}{g^2}\frac{\mathcal{G}_0}{h_0}\frac{1}{(\omega''-\omega_0)}+\frac{h(\omega_0-\omega'')}{(\omega_0-\omega'')}\,\frac{\mathcal{G}(\omega'')}{h(\omega'')}.\end{aligned} \tag{72}$$

Inserting (72) into (71) and using (66) and (49), we obtain for (71)

$$\frac{C_2K_0A_0}{g^2h_0}+C_2\frac{\omega_0}{\pi}\int_\mu^\infty \mathrm{Im}\frac{1}{K(\omega')}\,\frac{1}{\omega'-\omega_0-i\epsilon}\,\frac{\mathcal{G}(\omega')}{h(\omega')}. \tag{73}$$

The very last integral can be again obtained by residue method. Thus

$$\frac{1}{\pi}\int_\mu^\infty \mathrm{Im}\frac{1}{K(\omega')}\,\frac{\mathcal{G}(\omega')}{h(\omega')}\,\frac{d\omega'}{(\omega'-\omega_0-i\epsilon)}=\frac{1}{K_0}\,\frac{\mathcal{G}_0}{h_0}-\frac{g^2}{\omega_0}. \tag{74}$$

Collecting all the results, we obtain for (69)

$$0=1+C_1\left(h_0-\frac{\mathcal{G}_0}{h_0g^2}-\frac{1}{g^2}\right)+C_2\left(\frac{K_0A_0}{g^2h_0}+\frac{1}{h_0}-g^2\right). \tag{75}$$

Now with the help of (75) and (67) it is a simple matter to determine the constants C_1 and C_2. We find,

$$\begin{aligned}C_1=&\left(K_0A_0\mathcal{G}_0-\frac{K_0A_0\mathcal{G}_0}{h_0g^2}+g^2\mathcal{G}_0-\frac{\mathcal{G}_0}{h_0}\right)\\&\times\left[\mathcal{G}_0\left(\frac{K_0A_0}{g^2}-K_0A_0h_0+\mathcal{G}_0g^2-\frac{\mathcal{G}_0}{h_0}\right)\right]^{-1},\end{aligned} \tag{76}$$

and

$$C_2=\left(\mathcal{G}_0+h_0+\frac{\mathcal{G}_0}{h_0g^2}-\frac{1}{g^2}\right)\left(\frac{K_0A_0}{g^2}-K_0A_0h_0+\mathcal{G}_0g^2-\frac{\mathcal{G}_0}{h_0}\right)^{-1} \tag{77}$$

The expression given in (76) could be simplified still further. But we leave it like that so that one can easily see how it reduces to a similar expression of the ordinary KP equation. The constants C_1 and C_2 also reduce to D_1 and D_2 of the ordinary KP equation whenever B is zero (compare with Appendix C).

VII. Vθ S MATRIX

The $V\theta$ scattering matrix can be computed. We use the definitions

$$S_{V\theta}=\langle V\theta_\kappa^{\text{out}}\,|\,V\theta_{\kappa_0}^{\text{in}}\rangle=\delta(\kappa-\kappa_0)+2\pi\delta(\omega-\omega_0)T(\omega_0); \tag{78}$$

one easily finds from Ref. 4 that

$$T(\omega_0)=f^2(\omega_0)M(\omega_0,\omega_0)/\omega_0. \tag{79}$$

From Eqs. (64) and (76) we obtain

$$\begin{aligned}T(\omega_0)&=f^2(\omega_0)C_1/\omega_0\\&=f^2(\omega_0)\left(K_0A_0\mathcal{G}_0-\frac{K_0A_0\mathcal{G}_0}{h_0g^2}+g^2\mathcal{G}_0-\frac{\mathcal{G}_0}{h_0}\right)\\&\times\left[K_0\left(\frac{K_0A_0}{g^2}-K_0A_0h_0+g^2\mathcal{G}_0-\frac{\mathcal{G}_0}{h_0}\right)\right]^{-1},\end{aligned} \tag{80}$$

where $f(\omega_0)$ is as defined in Eq. (21).

To show that K_0 is also present in the $N\theta$ scattering amplitude $S_{N\theta}$, we write Eq. (27), using Eqs. (49) and (54), and obtain

$$T_1(\omega_0)=-[f^2(\omega_0)/K_0][1+B\omega_0/h_1(\omega_0)]. \tag{81}$$

VIII. THE PRODUCTION AMPLITUDE

The production amplitude $V\theta_0\to N\theta_1\theta_2$ can be easily computed from $\psi_3(\omega_1,\omega_2,\omega_0)$ of Eq. (33). We find

$$\begin{aligned}\psi_3(\omega_1,\omega_2,\omega_0)=&\frac{-gf(\omega_0)f(\omega_1)f(\omega_2)}{2(\omega_0-\omega_1-\omega_2+i\epsilon)}\\&\times\left(\frac{M(\omega_1,\omega_0)}{\omega_1(\omega_0-\omega_1)}+\frac{M(\omega_2,\omega_0)}{\omega_2(\omega_0-\omega_2)}\right).\end{aligned} \tag{82}$$

The production amplitude P defined by

$$S_{\text{Prod}}=\langle N\theta_1\theta_2^{\text{out}}\,|\,V\theta_0^{\text{in}}\rangle=2\pi i\delta(\omega_1+\omega_2-\omega_0)P(\omega_1,\omega_0)$$

is given by the term containing the $\delta(\omega_1+\omega_2-\omega_0)$ in $\psi_3(\omega_1,\omega_2,\omega_0)$ of (82). Substituting $\omega_2=\omega_0-\omega_1$, we find

$$P = \frac{gf(\omega_0)f(\omega_0-\omega_1)f(\omega_1)}{\omega_1(\omega_0-\omega_1)}[M(\omega_1)+M(\omega_0-\omega_1)] \tag{83}$$

$$= \frac{gf(\omega_0)f(\omega_0-\omega_1)f(\omega_1)C_2}{G(\omega_0-\omega_1)G(\omega_1)} + \frac{gf(\omega_0)f(\omega_0-\omega_1)f(\omega_1)}{\omega_1 G(\omega_0-\omega_1)} \times(1+h_0C_1-C_2g^2). \tag{84}$$

Equation (84) is obtained by using the results of Appendix B. In Eq. (84) the constants C_1 and C_2 are as given in Eqs. (76) and (77). Whenever B is zero, the second term in (84) simply vanishes and the result again reduces to that of one particle case. The simplicity of this result should be noted.

IX. UNIQUENESS

In order to decide the uniqueness of the solution, one has to prove the unitarity. We have not been able to prove the unitarity. But still we can do one thing. We let ω_C take values slightly greater than μ but less than 2μ. That is, let $\omega_C = \mu+\epsilon$, where ϵ is a small positive constant but not equal to zero and is less than μ. We now take $\omega_0 = \omega_C$. Under these circumstances there will be no production amplitude and the unitarity relation in the $V\theta$ sector reduces to the elastic unitarity relation. But for $\omega_0 = \omega_C$, we find from (76) and (77)

$$C_1 = 0 \quad \text{and} \quad C_2 = -1/K_0A_0 \tag{85}$$

and the unitarity is trivially satisfied.

For any other choice of $X(\omega', \omega_0)$ in (60) other than the second choice, the unitarity fails for $\omega_0 = \omega_C$. That is the reason we have ruled out all other choices in (60).

Still the second choice as discussed in Appendix A also satisfies the unitarity for $\omega_0 = \omega_C$ and C_1 and C_2 have the same values as in (85). Therefore, to decide the uniqueness, one has to prove the unitarity in the general case.

X. DISCUSSION

Our solution satisfies all the requirements of the integral equation and is the only solution (known to the author) which reduces to the single V-particle case whenever either of the interaction constants is switched off.

As in the $N\theta$ sector, there is a resonance in the $V\theta$ sector also. Until and unless the unitarity is proved, it is dangerous to apply the model to a study of other mechanisms, such as Pierel's mechanism.

ACKNOWLEDGMENTS

The author is very thankful to Dr. P. K. Srivastava for suggesting this problem. He is also thankful to Dr. C. L. Mehta for his keen interest in the work.

APPENDIX A

If we take

$$X(\omega', \omega_0) = C_2' \frac{\mathcal{G}(\omega')}{h(\omega')} \frac{1}{h(\omega_0-\omega')} \tag{A1}$$

and insert into (62), we obtain

$$M'(\omega, \omega_0) = \frac{\mathcal{G}(\omega_0-\omega)}{h(\omega_0-\omega)} \times\{C_1' - C_2'(\omega_0-\omega)I'(\omega_0-\omega)], \tag{A2}$$

where

$$I'(\omega_0-\omega) = -\frac{1}{\pi}\int_\mu^\infty \mathrm{Im}\frac{1}{K(\omega')} \frac{1}{h(\omega_0-\omega')} \times \frac{\mathcal{G}(\omega')\,d\omega'}{h(\omega')(\omega'-\omega_0+\omega-i\epsilon)}. \tag{A3}$$

Following a similar procedure as in the text, we find

$$C_1' = \left(K_0A_0\mathcal{G}_0 - \frac{K_0A_0\mathcal{G}_0}{h_0g^2} - \frac{\mathcal{G}_0^2}{h_0^2} + g^2\mathcal{G}_0\right) \times\left[\mathcal{G}_0\left(\frac{K_0A_0}{g^2} - K_0A_0h_0 + g^2\mathcal{G}_0 - \frac{\mathcal{G}_0^2}{h_0^2}\right)\right]^{-1}. \tag{A4}$$

and

$$C_2'' = \left(\mathcal{G}_0 + h_0 - \frac{\mathcal{G}_0}{h_0g^2} - \frac{1}{g^2}\right) \times\left(\frac{K_0A_0}{g^2} - K_0A_0h_0 + g^2\mathcal{G}_0 - \frac{\mathcal{G}_0^2}{h_0^2}\right)^{-1}. \tag{A5}$$

One immediately verifies that $C_1 = 0$ for $\omega_0 = \omega_C$ and $C_2 = -1/K_0A_0$.

APPENDIX B

Using Eq. (64) in (83), we find

$$P = [gf(\omega_0)f(\omega_1)f(\omega_0-\omega_1)/G(\omega_0-\omega_1)G(\omega_1)] \times\{C_1[h(\omega_1)\mathcal{G}(\omega_0-\omega_1) + \mathcal{G}(\omega_1)h(\omega_0-\omega_1)] - C_2[h(\omega_1)K(\omega_0-\omega_1)I(\omega_0-\omega_1) + h(\omega_0-\omega_1)K(\omega_1)I(\omega_1)]. \tag{B1}$$

In order to simplify this result, we have to find an expression for

$$-C_2[h(\omega_1)K(\omega_0-\omega_1)I(\omega_0-\omega_1) + h(\omega_0-\omega_1)K(\omega_1)I(\omega_1)].$$

To achieve this, we insert (64) directly into (40) to obtain

$$h(\omega_0-\omega)M(\omega,\omega_0) = 1 + \frac{C_1\omega}{\pi}\int_\mu^\infty \frac{\mathrm{Im}[h(\omega')]\,d\omega'}{(\omega'-\omega_0-i\epsilon)(\omega'-\omega_0+\omega-i\epsilon)} \frac{\mathcal{G}(\omega_0-\omega')}{h(\omega_0-\omega')} - C_2\frac{\omega}{\pi}\int_\mu^\infty \frac{\mathrm{Im}[h(\omega')]}{(\omega'-\omega_0-i\epsilon)} \frac{(\omega_0-\omega')I(\omega_0-\omega')}{(\omega'-\omega_0+\omega-i\epsilon)} \times\frac{\mathcal{G}(\omega_0-\omega')}{h(\omega_0-\omega')}\,d\omega'. \tag{B2}$$

The integrals in (2) can be again obtained by residue method as explained in the text. This gives

$$h(\omega_0-\omega)M(\omega,\omega_0) = 1 + C_1\,h_0 - \left(\frac{h(\omega_0-\omega)\mathcal{G}(\omega)}{h(\omega)}\right) + \frac{C_2}{h(\omega)} + C_2\frac{K(\omega)h(\omega_0-\omega)I(\omega)}{h(\omega)} - C_2g^2. \tag{B3}$$

Inserting (64) on the right-hand side of Eq. (3) and simplifying, we obtain

$$\mathcal{G}(\omega_0-\omega)[C_1 - C_2(\omega_0-\omega)I(\omega_0-\omega)]$$

$$= 1 + C_1\left(h_0 - \frac{h(\omega_0-\omega)\mathcal{G}(\omega)}{h(\omega)}\right) + C_2\frac{K(\omega)h(\omega_0-\omega)I(\omega)}{h(\omega)}$$

$$- C_2 g^2 + \frac{C_2}{h(\omega)}\,. \tag{B4}$$

Simplifying and rearranging Eq. (4), we find

$$- C_2[K(\omega)h(\omega_0-\omega)I(\omega) + h(\omega)K(\omega_0-\omega)I(\omega_0-\omega)]$$
$$= - C_1[h(\omega)\mathcal{G}(\omega_0-\omega) + h(\omega_0-\omega)\mathcal{G}(\omega)]$$
$$+ h(\omega)(1 + C_1 h_0 - C_2 g^2) + C_2. \tag{B5}$$

When Eq. (5) is inserted into (1), we obtain Eq. (84), which is the desired result.

APPENDIX C

When B is zero, the integral equation (40) reduces to

$$\mathcal{G}(\omega_0-\omega)M_1(\omega,\omega_0)$$
$$= 1 + \frac{\omega}{\pi}\int_\mu^\infty \frac{\mathrm{Im}[\mathcal{G}(\omega')]}{(\omega'-\omega_0-i\epsilon)}\,\frac{M_1(\omega',\omega_0)\,d\omega}{(\omega'-\omega_0+\omega-i\epsilon)}\,, \tag{C1}$$

which is the ordinary KP equation in terms of our function $\mathcal{G}(\omega_0-\omega)$. The solution of (C1) is given by

$$M_1(\omega,\omega_0) = D_1 - D_2(\omega_0-\omega)J(\omega_0-\omega), \tag{C2}$$

where

$$J(\omega_0-\omega) = -\frac{1}{\pi}\int_\mu^\infty \frac{d\omega'}{\mathcal{G}(\omega_0-\omega')}\,\frac{1}{\omega'-\omega_0+\omega-i\epsilon}\,\mathrm{Im}\frac{1}{K(\omega')}$$

$$D_1 = \frac{g^2 + K_0A_1}{\mathcal{G}_0(g^2 - K_0A_1)} \tag{C3}$$

and

$$D_2 = \frac{2}{(g^2 - K_0A_1)} \tag{C4}$$

with

$$A_1 = \frac{1}{\pi}\int_\mu^\infty \frac{1}{K(\omega_0-\omega')}\,\mathrm{Im}\frac{1}{K(\omega')}\,d\omega'. \tag{C5}$$

Our function $M(\omega)$ must reduce to $M_1(\omega)$ above.

APPENDIX D

Inserting Pagnamenta's solution into his integral equation (Ref. 4), we obtain

$$G(\omega_0-\omega)\phi_1(\omega,\omega_0)$$
$$= 1 + C_1\frac{\omega}{\pi}\int_\mu^\infty \frac{\mathrm{Im}[G(\omega')]\,d\omega'}{(\omega'-\omega_0-i\epsilon)(\omega'-\omega_0+\omega-i\epsilon)}$$
$$- C_2\frac{\omega}{\pi}\int_\mu^\infty \frac{\mathrm{Im}[G(\omega')]}{(\omega'-\omega_0-i\epsilon)}\,\frac{I(\omega_0-\omega')(\omega_0-\omega')}{(\omega'-\omega_0+\omega-i\epsilon)}\,d\omega',$$

where all the symbols in the above expressionare as in Ref. 4. For the first integral we obtain

$$\frac{C_1\omega}{\pi}\int_\mu^\infty \frac{\mathrm{Im}[G(\omega')]\,d\omega'}{(\omega'-\omega_0)(\omega'-\omega_0+\omega)}$$
$$= -C_1\frac{\omega}{\pi}\int_\mu^\infty \frac{\mathrm{Im}[G(\omega')]}{(\omega'-\omega_0+\omega-i\epsilon)}\,\frac{G(\omega_0-\omega')}{H(\omega_0-\omega')}\,d\omega'$$
$$= C_1\omega\left(\sum_{\text{poles}}\mathrm{Res}\,\frac{G(z)G(\omega_0-z)}{(z-\omega_0+\omega)H(\omega_0-z)}\right)$$
$$= C_1\omega\left(\frac{G(\omega_0)}{\omega} - \frac{G(\omega_0-\omega)}{\omega}\right);$$

for the second integral we obtain

$$- C_2\frac{\omega}{\pi}\int_\mu^\infty \mathrm{Im}\,\frac{1}{H(\omega'')}\,\frac{d\omega''}{G(\omega_0-\omega'')}$$
$$\times\frac{1}{\pi}\int_\mu^\infty \frac{\mathrm{Im}[G(\omega')]}{(\omega'-\omega_0+\omega''-i\epsilon)}\,\frac{d\omega'}{(\omega'-\omega_0+\omega-i\epsilon)}\,.$$

The very last integral by residue method gives

$$\left(\frac{G(\omega_0-\omega)}{\omega''-\omega} + \frac{G(\omega_0-\omega'')}{\omega-\omega''}\right),$$

and, inserting this into the above relation, we find

$$C_2G(\omega_0-\omega)\omega I(\omega)$$
$$+ C_2\frac{\omega}{\pi}\int_\mu^\infty \mathrm{Im}\frac{1}{H(\omega')}\,\frac{1}{\omega'-\omega-i\epsilon}\,d\omega'.$$

Doing the above integral again by the residue method, we find

$$C_2G(\omega_0-\omega)\omega I(\omega) + C_2\omega\left(\frac{1}{H(\omega)} - \frac{1}{\omega}\right).$$

Collecting all the terms, we obtain

$$G(\omega_0-\omega)\phi_1(\omega,\omega_0) = 1 + C_1G(\omega_0) - C_1G(\omega_0-\omega)$$
$$+ C_2G(\omega_0-\omega)\omega I(\omega) + C_2\omega\left(\frac{1}{H(\omega)} - \frac{1}{\omega}\right),$$

which is the relation (31) of Ref. 4. We have used the above method in our integrals.

*Part of a thesis to be submitted to the Indian Institute of Technology (Delhi) for Ph. D.
†This work is supported by an I. I. T. Fellowship.

[1]P. K. Srivastava, Phys. Rev. **131**, 461 (1963).
[2]T. M. Luke, Phys. Rev. **141**, 1373 (1966).
[3]F. S. C. Cheung and C. M. Sommerfield, Phys. Rev. **152**, 1401 (1966).
[4]A. Pagnamenta, J. Math. Phys. **6**, 955 (1965).
[5]A. Pagnamenta, J. Math. Phys. **7**, 356 (1966).
[6]C. Sommerfield, J. Math. Phys. **6**, 1170 (1965).
[7]E. M. Kazes, J. Math. Phys. **6**, 1172 (1965).
[8]L. M. Scarfone, J. Math. Phys. **9**, 246 (1968).
[9]T. D. Lee and G. C. Wick, Nucl. Phys. B **10**, 1 (1969).
[10]R. D. Amado, Phys. Rev. **122**, 697 (1961).

On analytic nonlocal potentials. IV. A closed contour dispersion relation

Te Hai Yao

Mathematics Department, Bedford College, Regent's Park, London NW1, England
(Received 29 April 1975)

We give a class of analytic nonlocal potentials which give rise to a closed contour s-wave dispersion relation for sufficiently small coupling constant.

I. INTRODUCTION AND THE RESULT

In Refs. 1 and 2 dispersion relations are proved for classes of analytic non-local potentials. However, in these relations the left-hand contours are not closed. Here we introduce a further class of analytic nonlocal potentials which give rise to a closed contour s-wave dispersion relation for sufficiently small coupling constant.

The class of potentials $V(\mathbf{x},\mathbf{x}')$ we shall consider is defined by

$$V(\mathbf{x},\mathbf{x}')=V(x,x',\cos\nu),$$

$$x=|\mathbf{x}|>0,\quad x'=|\mathbf{x}'|>0,\quad 1\geq\cos\nu\geq-1,$$

where ν is the angle between $\mathbf{x}$ and $\mathbf{x}'$, with

$$V(x,x',\cos\nu)=g\,\frac{\exp(-\gamma x)}{x}\,\frac{\exp(-\gamma x')}{x'}\times\int_0^\infty\int_0^\infty d\beta d\beta'\exp(-\beta x)\exp(-\beta' x')\times\sigma(\beta,\beta',\cos\nu),$$

where g is real, $\sigma(\beta,\beta',\cos\nu)$ is continuous in β, β', and $\cos\nu$ in $\infty>\beta\geq 0$, $\infty>\beta'\geq 0$, $1\geq\cos\nu\geq-1$, is holomorphic in β and β' in the region $U(\epsilon)\times U(\epsilon)$, where $U(\epsilon)$ consists of all points at distances less than some arbitrary $\epsilon>0$ from the interval $[0,\infty)$ for all $\cos\nu$ in the above range, and satisfies further:

(i) For $\infty>\beta\geq 0$, $\infty>\beta'\geq 0$, $1\geq\cos\nu\geq-1$, $\sigma(\beta,\beta',\cos\nu)$ is real and $\sigma(\beta,\beta',\cos\nu)=\sigma(\beta',\beta,\cos\nu)$.

(ii) For $(\beta,\beta')\in U(\epsilon)\times U(\epsilon)$ and $1\geq\cos\nu\geq-1$, we have

$$|\sigma(\beta,\beta',\cos\nu)|\leq\frac{\text{const}}{|\beta+\beta_0|^{(1+\lambda)}\,|\beta'+\beta_0|^{(1+\lambda)}}$$

for some $\beta_0>\epsilon>0$ and $\lambda>0$.

(iii)

$$\sigma(\beta,\beta',\cos\nu)\underset{\beta\to 0}{=}O(\beta^\eta)$$

for $\beta\in U(\epsilon)$ and for some $\eta>\frac{1}{2}$, uniformly in $\beta'\in U(\epsilon)$ and $1\geq\cos\nu\geq-1$.

An example of such a spectral function is given by

$$\sigma(\beta,\beta',\cos\nu)=a\beta^\eta\exp[-\mu(\beta+\beta')^2]\beta'^\eta$$

with $a>0$, $\eta>\frac{1}{2}$, $\mu>0$.

For each potential belonging to this class with sufficiently small g, we have, for the s-wave scattering amplitude $T(k)$, the following dispersion relation:

$$T(E)=D+\frac{(E-E_s)}{\pi}\times\left[\int_{-\infty}^{-\gamma}+\int_0^\infty\right]dE'\,\frac{\mathrm{Im}[f(E')]}{E'-E_s}\cdot\frac{1}{E'-E},\qquad(\mathrm{I}.1)$$

where E and E_s are not on $(-\infty,-\gamma]$ and $[0,\infty)$, and D and E_s are constants.

II. PROOF OF THE RESULT

For potentials belonging to the above class, the s-wave partial potential $V_0(x,x')$, defined by

$$V_0(x,x')=2\pi(xx')\int_{-1}^{+1}d\cos\nu\;V(x,x',\cos\nu)$$

has the representation

$$V_0(x,x')=g\int_\gamma^\infty\int_\gamma^\infty d\beta d\beta'\exp(-\beta x)\exp(-\beta' x')\,s(\beta,\beta'),$$

where $s(\beta,\beta')$ satisfies:

(i) For $\infty>\beta\geq\gamma$, $\infty>\beta'\geq\gamma$, $s(\beta,\beta')$ is real and $s(\beta,\beta')=s(\beta',\beta)$.

(ii) $s(\beta,\beta')$ is holomorphic in $W(\epsilon_0)\times W(\epsilon_0)$, where $W(\epsilon_0)$ is the region consisting of all points at distances less than some arbitrary ϵ_0, satisfying $\gamma>\epsilon_0>0$ and $\epsilon>\epsilon_0>0$, from the interval $[\gamma,\infty)$, and for (β,β') in this region, we have

$$|s(\beta,\beta')|\leq\frac{\text{const}}{|\beta|^{(1+\lambda)}\,|\beta'|^{(1+\lambda)}}.$$

(iii)

$$S(\beta,\beta')\underset{\beta=\gamma}{=}O((\beta-\gamma)^\eta)$$

for $\beta\in W(\epsilon_0)$, uniformly in $\beta'\in W(\epsilon_0)$.

The idea in the proof of (1.1) is the use of the formula[3] (*note*: in Ref. 3 we have put $g=1$)

$$S(g;k)=\frac{1+\int_\gamma^\infty d\beta[\xi^-(g;k;\beta)/(k^2+\beta^2)]}{1+\int_\gamma^\infty d\beta[\xi^+(g;k;\beta)/(k^2+\beta^2)]},$$

valid for k in the complex plane cut from $i\gamma$ to $i\infty$ and from $-i\gamma$ to $-i\infty$, where $S(g;k)$ is the s-wave S matrix and $\xi^\pm(g;k;\beta)$ are solutions of the following integral equations:

$$\xi^\pm(g;k;\beta)=g\,\xi^{\pm(0)}(k;\beta)+g\int_\gamma^\infty d\beta' J(k;\beta,\beta')\xi^\pm(g;k;\beta')$$

with

$$\xi^{\pm(0)}(k;\beta)=\int_\gamma^\infty d\beta'\,\frac{s(\beta,\beta')}{\beta'\mp ik},$$

$$J(k;\beta,\beta')=\frac{1}{k^2+\beta^2}\int_\gamma^\infty d\beta''\,\frac{s(\beta,\beta'')}{\beta'+\beta''}.$$

These solutions are given by

$$\xi^{\pm}(g;k;\beta)=g\,\xi^{\pm(0)}(k;\beta)$$
$$+g^2\int_\gamma^\infty d\beta'\,\frac{D(g;k;\beta,\beta')}{D(g;k)}\,\xi^{\pm(0)}(k;\beta'),$$

with

$$\delta_0(k)=1,\quad \delta_1(k)=0,$$

$$\delta_n(k)=\frac{(-1)^n}{n!}\begin{vmatrix} 0 & n-1 & 0 & \cdots & 0 & 0\\ \tau_2(k) & 0 & n-2 & & 0 & 0\\ \tau_3(k) & \tau_2(k) & 0 & & 0 & 0\\ \vdots & & & & & \vdots\\ \tau_n(k) & \tau_{n-1}(k) & \tau_{n-2}(k) & \cdots & \tau_2(k) & 0 \end{vmatrix},\quad n\geq 2$$

$$\delta_0(k;\beta,\beta')=J(k;\beta,\beta'),\quad \delta_1(k;\beta,\beta')=J^2(k;\beta,\beta'),$$

$$\delta_n(k;\beta,\beta')=\frac{(-1)^n}{n!}\begin{vmatrix} J(k;\beta,\beta') & n & 0 & \cdots & 0 & 0\\ J^2(k;\beta,\beta') & 0 & n-1 & & 0 & 0\\ J^3(k;\beta,\beta') & \tau_2(k) & 0 & & 0 & 0\\ \vdots & & & & & \\ J^{n+1}(k;\beta,\beta') & \tau_n(k) & \tau_{n-1}(k) & \cdots & \tau_2(k) & 0 \end{vmatrix},\quad n\geq 2,$$

where

$$\tau_n(k)=\mathrm{Tr}[J^n(k)],$$

$J(k)$ being the Hilbert–Schmidt operator in $L^2(\gamma,\infty)$ with kernel $J(k;\beta,\beta')$.

We now consider values of k with $\mathrm{Re}k>0$, $\mathrm{Im}k>\gamma$. Then $J^n(k;\beta,\beta')$, $n=1,2,\cdots$, admit holomorphic extensions in (β,β') into $W(\epsilon_0)\times W(\epsilon_0)$ and can be written, for $n>1$, as

$$J^n(k;\beta,\beta')=\frac{1}{k^2+\beta^2}\int_C\cdots\int_C d\beta_1'\cdots d\beta_{n-1}'$$
$$\times\prod_{i=1}^{n-1}\frac{1}{k^2+\beta_i'^2}\int_\gamma^\infty\cdots\int_\gamma^\infty d\rho_1\cdots d\rho_n$$
$$\times\frac{s(\beta,\rho_1)}{\beta_1'+\rho_1}\cdot\prod_{i=1}^{n-1}\frac{s(\beta_i',\rho_{i+1})}{\beta_{i+1}'+\rho_{i+1}}$$

with $\beta_n'=\beta'$, where C is the contour in the complex plane consisting of the line segment from γ to $\gamma+i\epsilon_0/2$ and the half-line from $\gamma+i\epsilon_0/2$ to $\infty+i\epsilon_0/2$ (Fig. 1). Hence $\tau_n(k)$ has the representation, for k in the same region,

where

$$D(g;k)=\exp[-\tau_1(k)]\sum_{n=0}^\infty g^n\delta_n(k),$$

$$D(g;k;\beta,\beta')=\exp[-\tau_1(k)]\sum_{n=0}^\infty g^n\,\delta_n(k;\beta,\beta'),$$

$$\tau_n(k)=\int_C\cdots\int_C d\beta\,d\beta_1'\cdots d\beta_{n-1}'\,\frac{1}{k^2+\beta^2}$$
$$\times\prod_{i=1}^{n-1}\frac{1}{k^2+\beta_i'^2}\int_\gamma^\infty\cdots\int_\gamma^\infty d\rho_1\cdots d\rho_n$$

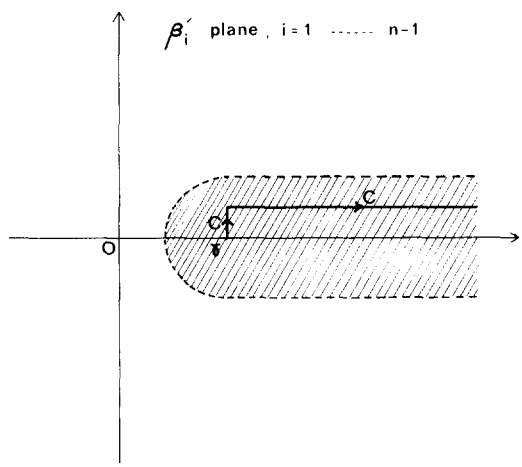


FIG. 1. The shaded region is $W(\epsilon_0)$.

$$\times \frac{s(\beta,\rho_1)}{\beta_1'+\rho_1}\cdot\left(\prod_{i=1}^{n-2}\frac{s(\beta_i',\rho_{i+1})}{\beta_{i+1}'+\rho_{i+1}}\right)\quad\frac{s(\beta_{n-1}',\rho_n)}{\beta+\rho_n}.$$

Consequently, $\tau_n(k)$ and $\delta_n(k)$ can be continued holomorphically across the cut $[i\gamma,i\infty)$ to the region $-\epsilon_0/2<\mathrm{Re}k<0$, $\mathrm{Im}k>\gamma$, on a new Riemann sheet Γ_1, in particular to the region $-\epsilon_0/4<\mathrm{Re}k<0$, $\mathrm{Im}k>\gamma$, on Γ_1.

For $\mathrm{Re}k>0$, $\mathrm{Im}k>\gamma$, the kernel $J(k;\beta,\beta')$ is a L^2 kernal along the contour C. If we let $\tilde{J}(k)$ be the Hilbert–Schmidt operator with kernel $J(k;\beta,\beta')$ in the space $L^2(C)$ of square integrable functions on the contour C, then the operator $\tilde{J}(k)$ can be continued holomorphically across the cut $[i\gamma,i\infty)$ to the region $-\epsilon_0/2<\mathrm{Re}k<0$, $\mathrm{Im}k>\gamma$, on the new Riemann sheet Γ_1. If we let $V_+(\epsilon_0)$ to be the union of the set $\mathrm{Re}k\geq 0$, $\mathrm{Im}k>\gamma$ on the original Riemann sheet Γ_0 and the region $-\epsilon_0/4<\mathrm{Re}k<0$, $\mathrm{Im}k>\gamma$ on the Riemann sheet Γ_1, then we have, for $k\in V_+(\epsilon_0)$,

$$\tau_n(k)=\mathrm{Tr}[\tilde{J}^n(k)]$$

and[4]

$$|\delta_n(k)|\leq \exp(n/2)\|\tilde{J}(k)\|^n/n^{n/2},\quad n\geq 2.$$

We also have

$$\|\tilde{J}(k)\|\leq \text{const}$$

for $k\in V_+(\epsilon_0,\epsilon_1)$, the part of $V_+(\epsilon_0)$ at a distance greater than some arbitrary $\epsilon_1>0$ from the point $i\gamma$. Hence $D(g;k)$ is holomorphic in k in $V_+(\epsilon_0,\epsilon_1)$ and hence in $V_+(\epsilon_0)$. Further, we have, given $1>\theta>0$, the relation

$$1+\theta>|D(g;k)|>1-\theta,$$

$$k\in V_+(\epsilon_0),\ \mathrm{Re}k\geq 0,$$

for g sufficiently small, since we can show that $\|\tilde{J}(k)\|\leq$ const for $k\in V_+(\epsilon_0)$, $\mathrm{Re}k\geq 0$, using property (iii) of $s(\beta,\beta')$.

Similarly, we can continue $D(g;k)$ holomorphically from $\mathrm{Re}k<0$, $\mathrm{Im}k>\gamma$ on the original Riemann sheet Γ_0 across the cut $[i\gamma,i\infty)$ to the region $\epsilon_0/2>\mathrm{Re}k>0$, $\mathrm{Im}k>\gamma$ on a new Riemann sheet Γ_2, leading to a function holomorphic in k in the part $V_-(\epsilon_0)$ on the Riemann surface consisting of the set $\mathrm{Re}k<0$, $\mathrm{Im}k>\gamma$ on the original Riemann sheet Γ_0, and the set $\epsilon_0/4>\mathrm{Re}k\geq 0$, $\mathrm{Im}k>\gamma$ on the Riemann sheet Γ_2; further we have, given $1>\theta>0$, the relation

$$1+\theta>|D(g;k)|>1-\theta,$$

$$k\in V_-(\epsilon_0),\ \mathrm{Re}k\geq 0$$

for g sufficiently small.

We now consider the integrals

$$I^{\pm}(g;k)\equiv\int_\gamma^\infty d\beta\,\frac{\xi^{\pm}(g;k;\beta)}{k^2+\beta^2}.$$

We have

$$I^{\pm}(g;k)=g\int_\gamma^\infty d\beta\,\frac{1}{k^2+\beta^2}\int_\gamma^\infty d\beta'\times\frac{s(\beta,\beta')}{\beta'\mp ik}+\frac{g^2}{D(g;k)}\int_\gamma^\infty d\beta\,\frac{1}{k^2+\beta^2}$$

$$\times\int_\gamma^\infty d\beta' D(g;k;\beta,\beta')\times\int_\gamma^\infty d\beta''\,\frac{s(\beta',\beta'')}{\beta''\mp ik}.$$

For $\mathrm{Re}k>0$, $\mathrm{Im}k>\gamma$, on Γ_0, the integrals

$$I_1^{\pm}(g;k)\equiv g\int_\gamma^\infty d\beta\,\frac{1}{k^2+\beta^2}\int_\gamma^\infty d\beta'\frac{s(\beta,\beta')}{\beta'\mp ik}$$

can be written as

$$I_1^{\pm}(g;k)=g\int_C d\beta\,\frac{1}{k^2+\beta^2}\int_C d\beta'\,\frac{s(\beta,\beta')}{\beta'\mp ik}$$

and hence can be continued, as functions of k, holomorphically from this region across the cut $[i\gamma,i\infty)$ to the region $-\epsilon_0/2<\mathrm{Re}k<0$, $\mathrm{Im}k>\gamma$, on Γ_1, leading to functions holomorphic in k in $V_+(\epsilon_0)$. Further, using property (iii) of $s(\beta,\beta')$, we obtain

$$|I_1^{\pm}(g;k)|\leq |g|\cdot\text{const},\quad k\in V_+(\epsilon_0),\ \mathrm{Re}k\geq 0.$$

Again, for $\mathrm{Re}k>0$, $\mathrm{Im}k>\gamma$, on Γ_0, we have

$$I_2^{\pm}(g;k)\equiv g^2\int_\gamma^\infty d\beta\,\frac{1}{k^2+\beta^2}\int_\gamma^\infty d\beta'\times D(g;k;\beta,\beta')\int_\gamma^\infty d\beta''\,\frac{s(\beta',\beta'')}{\beta''\mp ik}$$

$$=g^2\sum_{n=0}^\infty g^n\int_\gamma^\infty\int_\gamma^\infty\int_\gamma^\infty d\beta\,d\beta'\,d\beta''\times\frac{1}{k^2+\beta^2}\,\delta_n(k;\beta,\beta')\,\frac{s(\beta',\beta'')}{\beta''\mp ik}$$

$$=g^2\sum_{n=0}^\infty g^n\int_C\int_C\int_C d\beta\,d\beta'\,d\beta''\times\frac{1}{k^2+\beta^2}\,\delta_n(k;\beta,\beta')\,\frac{s(\beta',\beta'')}{\beta''\mp ik}.$$

The integrals inside the summation in the last line are holomorphic in $V_+(\epsilon_0)$.

For β and β' on the contour C, and $k\in V_+(\epsilon_0)$, we have[5]

$$|\delta_0(k;\beta,\beta')|\leq \text{const}/\{|\beta|^{(1+\lambda)}[(\mathrm{Re}\beta')^2+\gamma^2]^{1/2}\},$$

$$\text{also}\quad\leq\text{const}\cdot(\beta-\gamma)^\eta,\quad|\beta-i\gamma|<\zeta,\ \text{for some }\zeta;$$

$$|\delta_1(k;\beta,\beta')|\leq \text{const}/\{|\beta|^{(1+\lambda)}[(\mathrm{Re}\beta')^2+\gamma^2]^{1/2}\},$$

$$\text{also}\quad\leq\text{const}\cdot(\beta-\gamma)^\eta,\quad|\beta-i\gamma|<\zeta;$$

$$|\delta_n(k;\beta,\beta')|\leq\frac{\text{const}}{|\beta|^{(1+\lambda)}\sqrt{(\mathrm{Re}\,\beta')^2+\gamma^2}}\cdot\frac{e^{(n-1)/2}\|\tilde{J}(k)\|^{n-1}}{n^{(n-2)/2}},\quad n\geq 2,$$

$$\text{also}\quad\leq\text{const}\cdot(\beta-\gamma)^\eta e^{(n-1)/2}\|\tilde{J}(k)\|^{n-1}/n^{(n-2)/2},\quad|\beta-i\gamma|<\zeta,\ n\geq 2;$$

and

$$|I_2^{\pm}(g;k)|\leq g^2\left\{\text{const}+|g|\cdot\text{const}\right.$$

$$+\sum_{n=2}^{\infty} |g|^n \frac{e^{(n-1)/2}}{n^{(n-2)/2}} \cdot \text{const}\Bigg\}$$

for $k \in V_+(\epsilon_0, \epsilon_1)$, using $\|\tilde{J}(k)\| \leq \text{const}$ there. Hence $I_2^{\pm}(g;k)$ are holomorphic in k in $V_+(\epsilon_0, \epsilon_1)$, and hence in $V_+(\epsilon_0)$. Hence the integrals $I_2^{\pm}(g;k)$ have been continued, as functions of k, homomorphically from the region $\mathrm{Re}k > 0$, $\mathrm{Im}k > \gamma$, on Γ_0, across the cut $[i\gamma, i\infty)$ to the region $-\epsilon_0/2 < \mathrm{Re}k < 0$, $\mathrm{Im}k > \gamma$, on Γ_1, leading to functions holomorphic in k in $V_+(\epsilon_0)$. Further, using $\|\tilde{J}(k)\| \leq \text{const}$ for $k \in V_+(\epsilon_0)$, $\mathrm{Re}k \geq 0$, we have

$$|I_2^{\pm}(g;k)| \leq g^2 \cdot \text{const}$$

for g sufficiently small. Hence the integrals $I^{\pm}(g;k)$ have been continued, as functions in k, holomorphically across the cut $[i\gamma, i\infty)$ from the region $\mathrm{Re}k > 0$, $\mathrm{Im}k > \gamma$, on Γ_0, to the region $-\epsilon_0/2 < \mathrm{Re}k < 0$, $\mathrm{Im}k > \gamma$, on Γ_1, leading to functions holomorphic in k in $V_+(\epsilon_0)$. Further, we have, for $k \in V_+(\epsilon_0)$, $\mathrm{Re}k \geq 0$:

$$|I^{\pm}(g;k)| \leq |g| \cdot \text{const}$$

for g sufficiently small.

Similarly, these integrals $I^{\pm}(g;k)$ can be continued from the region $\mathrm{Re}k < 0$, $\mathrm{Im}k > \gamma$, on Γ_0, holomorphically to the region $\epsilon_0/2 > \mathrm{Re}k > 0$, $\mathrm{Im}k > \gamma$, on Γ_2, leading to functions holomorphic in k in $V_-(\epsilon_0)$, and further we have, for $k \in V_-(\epsilon_0)$, $\mathrm{Re}k \leq 0$:

$$|I^{\pm}(g;k)| \leq |g| \cdot \text{const}$$

for g sufficiently small.

Consequently, $S(g;k)$ is holomorphic in k in the union of $V_+(\epsilon_0)$ and $V_-(\epsilon_0)$, and since it is holomorphic in the sheet Γ_0 cut from $i\gamma$ to $i\infty$ and from $-i\gamma$ to $-i\infty$, it is holomorphic in the union V of these sets. Further, using property (iii) of $s(\beta, \beta')$, we can show that, given $1 > \theta > 0$, we have

$$1 + \theta > |D(g;k)| > 1 - \theta$$

and

$$|I^{\pm}(g;k)| \leq |g| \cdot \text{const}$$

for k in the sheet Γ_0 cut from $i\gamma$ to $i\infty$ and from $-i\gamma$ to $-i\infty$, for g sufficiently small. Consequently, we have shown that $S(g;k)$ is holomorphic in k in V, and that for k in the union of the sets $\{k \in V_+(\epsilon_0),\ \mathrm{Re}k \geq 0\}$, $\{k \in V_-(\epsilon_0),\ \mathrm{Re}k \leq 0\}$, and $\{k \in \Gamma_0$, cut from $i\gamma$ to $i\infty$ and from $-i\gamma$ to $-i\infty\}$, we have

$$1 + \theta > |S(g;k)| > 1 - \theta$$

given θ satisfying $1 > \theta > 0$, for g sufficiently small.

Hence using[2]

$$T(-k^*) = T(k)^*$$

for k in the sheet Γ_0 cut from $i\gamma$ to $i\infty$ and from $-i\gamma$ to $-i\infty$, we arrive at the subtracted dispersion relation (1.1).

ACKNOWLEDGMENT

The author wishes to acknowledge the financial support of the Science Research Council.

*Present address: Institute for Advanced Studies, School of Theoretical Physics, 10 Burlington Road, Dublin 4, Ireland.

[1]T. H. Yao, J. Math. Phys. **14**, 1141 (1973).
[2]T. H. Yao, J. Math. Phys. **15**, 1211 (1974).
[3]T. H. Yao, J. Math. Phys. **15**, 1232 (1974), Eq. (2.32)
[4]F. Smithies, *Integral Equations* (Cambridge U. P., London, 1958), Sec. 6.5, Eq. (3).
[5]F. Smithies, Ref. 4, Sec. 6.5, Eqs. (4) and (10).

On conserved quantities in kinematic dynamo theory

I. Lerche*

Center for Space Research, Department of Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139
(Received 16 June 1975)

Using a Lagrangian approach to the magnetic induction equation in an infinite medium, we demonstrate that there exist seven conserved quantities which, by analogy with classical mechanics, we label as "energy," "momentum," and "angular momentum." For prescribed fluid motions we spell out the detailed conservation equations. For a fluid motion which is turbulent we also give the average conserved quantities. In a pragmatic sense it is expected that these conservation laws will be of use in attempts to obtain numerically accurate solutions to the turbulent kinematic dynamo equations. Since the magnetic induction equation is not self-adjoint, numerical attempts to date have to impose some extraneous ad hoc "criteria of goodness" at any given level of numerical truncation. The conserved quantities given here provide an internal check of the accuracy of any numerical calculation without the necessity for arbitrarily imposed external criteria of accuracy. As such they should be a powerful tool in rapidly increasing the accuracy of numerical solutions to the kinematic dynamo equations. We also point out that the conserved quantities can be used to indicate the possibility of kinematic dynamo activity *ahead* of any detailed calculations.

I. INTRODUCTION

In recent years the basic theory of kinematic dynamo action in turbulent media has progressed extremely rapidly (see the review by Gubbins[1] for a partial list of contributors to the field) since the pioneering work of Parker (1955).[2] It has become apparent that simple kinematic dynamo models can account for the sustained presence of magnetic fields in astrophysical bodies and, occasionally, even for the observed spatial and temporal variations of, say, the classical "butterfly" pattern of sunspot evolution—at least in gross morphological outline.

One of the main technical difficulties remaining in attempts to improve our understanding of the dynamo phenomenon is that only for very simple models of bulk fluid velocity, bulk velocity shear, and velocity turbulence can the kinematic dynamo equations be solved analytically. And when numerical calculations are used on more complex problems, the fact that the kinematic dynamo equations are not self-adjoint normally implies that a small numerical error can rapidly grow to an unacceptable level—leaving the results of the computation open to question (but see Gubbins[1]). For (relatively) simple situations the numerical calculations of Stix,[3] and those of Rose and Levy,[4] which use a variational principle (Lerche[5]), would appear to have overcome to some extent the numerical rounding error problem. But in more complex situations there is still a considerable amount of room for improvement. The question we investigate in this paper is related to these problems. It would be very useful if we could obtain some conserved quantities from the kinematic dynamo equations for arbitrary fluid motions. Such quantities could then be used in two ways: first as checks at any given stage of a numerical calculation; second as necessary, but not sufficient, conditions for kinematic dynamo activity.

The first application is fairly obvious. If we can find conserved quantities, then at each stage of a numerical calculation one can check to see if they are conserved. If not, then a numerical iteration can be used to improve the accuracy of the calculation at each stage and comparison again made with the conserved quantities until some desired degree of accuracy is attained. This is an improvement over the variational principle[5] already in use because, as we have remarked already, the kinematic dynamo equations are not self-adjoint and thus neither is the variational method. Accordingly one has to impose some arbitrary criterion of accuracy on the variational method. This leads to "criteria of goodness" which are *not* contained in the kinematic dynamo equations but are dependent on the individual and are, therefore, imposed in an ad hoc manner (i.e., there is *no* Rayleigh–Ritz statement available which guarantees that one is getting closer to the correct answer as more terms are included in the variational method). It would, therefore, be extremely useful to have available conserved quantities which arise from the kinematic dynamo equations. These would give criteria of goodness which do not depend on external imposition.

The second use, namely as necessary conditions for kinematic dynamo action, would also appear to be of some interest. What we would seek here is the analog of Cowling's theorem which states that steady axisymmetric motion cannot support a steady axisymmetric magnetic field. And Cowling's argument depends only on symmetry statements. Thus conserved quantities must be related to the symmetry of a given situation and must also give information on the *necessity* of certain symmetry breaking motions for kinematic dynamo action. Such conserved quantities would then be of use in detailing classes of motions which *could not* be responsible for dynamo action. And they could also be used for detailing classes of motions which *may* give rise to kinematic dynamo activity.

In this paper we will set up seven conserved quantities (which follow from the magnetic induction equation in an infinite medium) which we call "energy," "momentum," and "angular momentum" by analogy to the classically conserved quantities of mechanics.

We shall then perform a statistical averaging to obtain the average conserved quantities in the kinematic dynamo equations; and we shall then illustrate how they

behave under simple situations. Essentially then we shall be constructing "virial theorems" for the magnetic induction equation and the kinematic dynamo equations.

II. LAGRANGIANS, HAMILTONIANS, AND CONSERVED QUANTITIES

A. General remarks

For an infinite medium of constant electrical resistivity, η, possessing a given velocity field $\mathbf{v}(\mathbf{x}, t)$, the equation describing the evolution of the magnetic field $\mathbf{B}(\mathbf{x}, t)$ is

$$\frac{\partial B_i}{\partial t} = \eta \nabla^2 B_i + \epsilon_{ijk}\epsilon_{klm} \frac{\partial}{\partial x_j} (v_l B_m). \tag{1}$$

From Eq. (1) it is well known[5] that it is possible to construct the Lagrangian

$$\begin{aligned} L &= \int d^3\mathbf{x}\, dt \left(B_i^+ \frac{\partial B_i}{\partial t} + \eta \frac{\partial B_i^+}{\partial x_j} \frac{\partial B_i}{\partial x_j} + \epsilon_{ijk}\epsilon_{klm} v_l B_m \frac{\partial B_i^+}{\partial x_j} \right) \\ &\equiv \int d^3\mathbf{x}\, dt\, \mathcal{L}, \end{aligned} \tag{2}$$

such that extremal variation of the Lagrangian with respect to B_i^+ yields Eq. (1), while extremal variation of the Lagrangian with respect to B_i yields the equation adjoint to Eq. (1):

$$\frac{\partial B_i^+}{\partial t} + \eta \nabla^2 B_i^+ - \epsilon_{mjk}\epsilon_{kli} v_l \frac{\partial B_m^+}{\partial x_j} = 0. \tag{3}$$

Thus $\mathbf{B}^+$ is the adjoint vector field to $\mathbf{B}$. For later purposes it is opportune here to write Eqs. (1) and (3) in the conventional Euler–Lagrange form using the Lagrangian density $\mathcal{L}$ which, in principle, is a functional of B_i, B_i^+, $\partial B_i/\partial x_j$, $\partial B_i^+/\partial x_j$, $\partial B_i/\partial t$, $\partial B_i^+/\partial t$, and of $\mathbf{v}(\mathbf{x}, t)$. Thus

$$\frac{\partial \mathcal{L}}{\partial B_i^+} - \frac{\partial}{\partial x_j} \left(\frac{\partial \mathcal{L}}{\partial (\partial B_i^+/\partial x_j)} \right) - \frac{\partial}{\partial t} \left(\frac{\partial \mathcal{L}}{\partial \dot{B}_i^+} \right) = 0, \tag{4}$$

and

$$\frac{\partial \mathcal{L}}{\partial B_i} - \frac{\partial}{\partial x_j} \left(\frac{\partial \mathcal{L}}{\partial (\partial B_i/\partial x_j)} \right) - \frac{\partial}{\partial t} \left(\frac{\partial \mathcal{L}}{\partial \dot{B}_i} \right) = 0, \tag{5}$$

where $\dot{B}_i = \partial B_i/\partial t$, $\dot{B}_i^+ = \partial B_i^+/\partial t$. From the form of the Lagrangian density (2) we also have a canonical momentum density of

$$\Pi_i \equiv \frac{\partial \mathcal{L}}{\partial \dot{B}_i} = B_i^+, \tag{6}$$

so that we can construct a Hamiltonian density

$$\mathcal{H} = \Pi_i \dot{B}_i - \mathcal{L} = -\eta \frac{\partial B_i}{\partial x_j} \frac{\partial B_i^+}{\partial x_j} - \epsilon_{ijk}\epsilon_{klm} v_l B_m \frac{\partial B_i^+}{\partial x_j}, \tag{7}$$

and a Hamiltonian H through

$$H = \int d^3\mathbf{x}\, dt\, \mathcal{H}. \tag{8}$$

B. Construction of conserved quantities

The construction of conserved quantities is based on symmetry arguments. The magnetic induction equation and its adjoint, together with the velocity field $\mathbf{v}(x, t)$, should not depend on how the spatial coordinates are chosen nor where the origin of time is measured from.

(i) *"Energy" construction*: If, in the Lagrangian (2), t is replaced by $t' = t + \tau$, where τ is arbitrary but infinitesimal no difference should ensue in the Lagrangian. But since the regime of temporal integration in Eq. (2) is over all values of time, this implies that the Lagrangian *density* should remain unchanged under such a shift in the origin of time measurements. Now, if we had a solution to Eqs. (1) and (3) (subject to appropriate boundary and/or initial value conditions), then the Lagrangian density would be representable as an explicit function of space and time:

$$\mathcal{L} = \mathcal{L}(\mathbf{x}, t). \tag{9}$$

Then changing the origin of temporal measurements will change the Lagrangian density:

$$\mathcal{L}(\mathbf{x}, t + \tau) = \mathcal{L}(\mathbf{x}, t) + \tau \frac{\partial \mathcal{L}}{\partial t}(\mathbf{x}, t) + O(\tau^2) \tag{10}$$

But, from Eq. (9) since $\mathbf{B}$, $\mathbf{B}^+$, and $\mathbf{v}$ are also functions of $\mathbf{x}$ and t, it follows that

$$\begin{aligned} \tau \frac{\partial \mathcal{L}}{\partial t} &= \tau \frac{\partial \mathcal{L}}{\partial B_j} \frac{\partial B_j}{\partial t} + \tau \frac{\partial \mathcal{L}}{\partial \dot{B}_j} \frac{\partial \dot{B}_j}{\partial t} \\ &\quad + \tau \frac{\partial \mathcal{L}}{\partial B_j^+} \frac{\partial B_j^+}{\partial t} + \tau \frac{\partial \mathcal{L}}{\partial \dot{B}_j^+} \frac{\partial \dot{B}_j^+}{\partial t} \\ &\quad + \tau \left(\frac{\partial \mathcal{L}}{\partial (\partial B_j/\partial x_\alpha)} \right) \frac{\partial \dot{B}_j}{\partial x_\alpha} + \tau \left(\frac{\partial \mathcal{L}}{\partial (\partial B_j^+/\partial x_\alpha)} \right) \\ &\quad \times \frac{\partial \dot{B}_j^+}{\partial x_\alpha} + \tau \frac{\partial \mathcal{L}}{\partial v_j} \frac{\partial v_j}{\partial t}. \end{aligned} \tag{11}$$

We can now use (4) and (5) to eliminate $\partial \mathcal{L}/\partial B_i$ and $\partial \mathcal{L}/\partial B_i^+$ in Eq. (11). Upon so doing we obtain

$$\begin{aligned} &\tau \frac{\partial}{\partial t} \left(\dot{B}_j \frac{\partial \mathcal{L}}{\partial \dot{B}_j} + \dot{B}_j^+ \frac{\partial \mathcal{L}}{\partial \dot{B}_j^+} - \mathcal{L} \right) \\ &= -\tau \frac{\partial v_j}{\partial t} \frac{\partial \mathcal{L}}{\partial v_j} + \tau \frac{\partial}{\partial x_\alpha} \left(\dot{B}_j^+ \frac{\partial \mathcal{L}}{\partial (\partial B_j^+/\partial x_\alpha)} \right. \\ &\quad \left. + \dot{B}_j \frac{\partial \mathcal{L}}{\partial (\partial B_j/\partial x_\alpha)} \right) \\ &\equiv -\tau \frac{\partial v_j}{\partial t} \frac{\partial \mathcal{L}}{\partial v_j} + \tau \nabla \cdot \mathbf{Q}. \end{aligned} \tag{12}$$

But Eq. (12) must hold true for all infinitesimal values of τ. Thus

$$\frac{\partial \mathcal{H}}{\partial t} = \nabla \cdot Q - \frac{\partial v_j}{\partial t} \frac{\partial \mathcal{L}}{\partial v_j}. \tag{13}$$

We now integrate Eq. (13) over the whole of coordinate space to obtain

$$\begin{aligned} \frac{\partial}{\partial t} \int \mathcal{H}\, d^3\mathbf{x} &= -\int d^3\mathbf{x} \frac{\partial v_j}{\partial t} \frac{\partial \mathcal{L}}{\partial v_j} \\ &\equiv -\epsilon_{ilk}\epsilon_{kjm} \int d^3\mathbf{x} B_m \frac{\partial B_i^+}{\partial x_l} \frac{\partial v_j}{\partial t}. \end{aligned} \tag{14}$$

The surface term $\int \mathbf{Q} \cdot d\mathbf{S}$ gives no contribution to Eq. (14) in an infinite medium of constant resistivity.

Equation (14) is equivalent to the power theorem in classical mechanics. It represents the rate at which the spatially averaged Hamiltonian density changes with time due to the acceleration $\partial \mathbf{v}/\partial t$.

(ii) *"Momentum" conservation*: In a manner similar to that for "energy" conservation [Eq. (13)] we can change the origin of the spatial coordinates from $\mathbf{x}$ to $\mathbf{x}+\boldsymbol{\xi}$, where $\boldsymbol{\xi}$ is infinitesimal to obtain "momentum" conservation. Proceeding by analogous steps, we obtain the equation

$$\frac{\partial}{\partial t}\left(\frac{\partial \mathcal{L}}{\partial \dot{B}_j^*}\frac{\partial B_j^*}{\partial x_i}+\frac{\partial \mathcal{L}}{\partial \dot{B}_j}\frac{\partial B_j}{\partial x_i}\right)=-\frac{\partial \mathcal{L}}{\partial v_j}\frac{\partial v_j}{\partial x_i}+\frac{\partial T_{\alpha i}}{\partial x_\alpha}, \tag{15}$$

where

$$T_{\alpha i}=\mathcal{L}\delta_{\alpha i}-\frac{\partial B_j^*}{\partial x_i}\frac{\partial \mathcal{L}}{\partial(\partial B_j^*/\partial x_\alpha)}-\frac{\partial B_j}{\partial x_i}\frac{\partial \mathcal{L}}{\partial(\partial B_j/\partial x_\alpha)}.$$

When Eq. (15) is integrated over the whole of coordinate space, and where once again the surface term $\int T_{\alpha i}dS_\alpha$ is ignored, we obtain the conserved "momentum" statement

$$\frac{\partial}{\partial t}\int d^3\mathbf{x}\left(\frac{\partial \mathcal{L}}{\partial \dot{B}_j^*}\frac{\partial B_j^*}{\partial x_i}+\frac{\partial \mathcal{L}}{\partial \dot{B}_j}\frac{\partial B_j}{\partial x_i}\right)=-\int d^3\mathbf{x}\frac{\partial \mathcal{L}}{\partial v_j}\frac{\partial v_j}{\partial x_i}. \tag{16}$$

(iii) *"Angular momentum" conservation*: We have already seen that the physics of the magnetic induction equation (1) must be invariant with respect to changes in the origins of the temporal and spatial coordinates systems. It must also be invariant with respect to *rotation* of the inertial spatial coordinate system to another inertial spatial coordinate system. Thus in the Lagrangian density if we replace x_i by $x_i'=x_i+\epsilon R_{ij}x_j$ with $R_{ij}=-R_{ji}$ (the rotation matrix) we must again obtain a conserved quantity—the analog of "angular momentum" in classical mechanics. By proceeding in a manner similar to that for "energy" conservation it can be shown after some elementary, but tedious, algegra that the relevant statement is

$$\frac{\partial}{\partial t}\int d^3\mathbf{x}B_j^*\left(x_\beta\frac{\partial}{\partial x_i}-x_i\frac{\partial}{\partial x_\beta}\right)B_j$$

$$=-\int\frac{\partial \mathcal{L}}{\partial v_j}\left(x_\beta\frac{\partial}{\partial x_i}-x_i\frac{\partial}{\partial x_\beta}\right)v_j d^3\mathbf{x} \tag{17}$$

$$+\int v_l B_m\epsilon_{klm}\left(\epsilon_{j\beta k}\frac{\partial}{\partial x_i}-\epsilon_{jik}\frac{\partial}{\partial x_\beta}\right)B_j^* d^3\mathbf{x},$$

where we have already integrated over the whole of coordinate space, and have again set the appropriate surface term to zero. If we multiply Eq. (17) by $\epsilon_{\alpha\beta i}$, we obtain

$$\frac{\partial}{\partial t}\int d^3\mathbf{x}L_\alpha=-\epsilon_{\alpha\beta i}\int d^3\mathbf{x}\frac{\partial \mathcal{L}}{\partial v_j}\left(x_\beta\frac{\partial}{\partial x_i}-x_i\frac{\partial}{\partial x_\beta}\right)v_j$$

$$-2\int d^3\mathbf{x}\left[B_\alpha^*\epsilon_{jlm}-B_j^*\epsilon_{\alpha lm}\right]\frac{\partial}{\partial x_j}(v_l B_m), \tag{18}$$

where the angular momentum density $\mathbf{L}$ is given by

$$L_\alpha=\epsilon_{\alpha\beta i}B_j^*\left(x_\beta\frac{\partial}{\partial x_i}-x_i\frac{\partial}{\partial x_\beta}\right)B_j. \tag{19}$$

Equation (19) expresses the rate of change of the "angular momentum" of the field in terms of the applied "torque."

Written out explicitly in terms of $\mathbf{v}, \mathbf{B}$ and $\mathbf{B}^*$ the seven equations (14), (16), and (18) are

$$\frac{\partial}{\partial t}\int d^3\mathbf{x}\left(\eta\frac{\partial B_i}{\partial x_j}\frac{\partial B_i^*}{\partial x_j}+\epsilon_{ijk}\epsilon_{klm}v_l B_m\frac{\partial B_i^*}{\partial x_j}\right)$$

$$=\epsilon_{ijk}\epsilon_{kjm}\int d^3\mathbf{x}B_m\frac{\partial B_i^*}{\partial x_l}\frac{\partial v_j}{\partial t}, \tag{20a}$$

$$\frac{\partial}{\partial t}\int d^3\mathbf{x}B_j^*\frac{\partial B_j}{\partial x_i}$$

$$=-\epsilon_{\alpha lk}\epsilon_{kjm}\int d^3\mathbf{x}B_m\frac{\partial B_\alpha^*}{\partial x_l}\frac{\partial v_j}{\partial x_i}, \tag{20b}$$

and

$$\frac{\partial}{\partial t}\int d^3\mathbf{x}\epsilon_{\alpha\beta i}B_j^*\left(x_\beta\frac{\partial}{\partial x_i}-x_i\frac{\partial}{\partial x_\beta}\right)B_j$$

$$=-\epsilon_{plk}\epsilon_{kjm}\epsilon_{\alpha\beta i}\int d^3\mathbf{x}B_m\frac{\partial B_p^*}{\partial x_l}\left(x_\beta\frac{\partial}{\partial x_i}-x_i\frac{\partial}{\partial x_\beta}\right)v_j$$

$$-2\int d^3\mathbf{x}\left[B_\alpha^*\epsilon_{jlm}-B_j^*\epsilon_{\alpha lm}\right]\frac{\partial}{\partial x_j}(v_l B_m). \tag{20c}$$

III. STATISTICAL AVERAGING OF CONSERVED QUANTITIES

A. Statistically exact conserved quantities

Now in dealing with situations in which the velocity field is turbulent, but possesses an ensemble average value, it is usual to note that an ensemble average of the individual members of a turbulent system can be performed. Thus we write

$$v_i=V_i+\delta v_i,\quad B_i=b_i+\delta B_i,\quad B_i^*=b_i^*+\delta B_i^*$$

where $V_i=\langle v_i\rangle$, $b_i=\langle B_i\rangle$, $b_i^*=\langle B_i^*\rangle$ and angular brackets denote an ensemble averaging process. Then upon ensemble averaging Eqs. (1) and (3) we obtain

$$\frac{\partial b_i}{\partial t}-\eta\nabla^2 b_i=\epsilon_{ijk}\epsilon_{klm}\frac{\partial}{\partial x_j}\left[V_l b_m+\langle\delta v_l\delta B_m\rangle\right], \tag{21a}$$

$$\frac{\partial \delta B_i}{\partial t}-\eta\nabla^2\delta B_i=\epsilon_{ijk}\epsilon_{klm}\frac{\partial}{\partial x_j}$$

$$\times\left(V_l\delta B_m+\delta v_l b_m+\delta v_l\delta b_m-\langle\delta v_l\delta b_m\rangle\right), \tag{21b}$$

$$\frac{\partial b_i^*}{\partial t}+\eta\nabla^2 b_i^*=\epsilon_{mjk}\epsilon_{kli}\left(V_l\frac{\partial b_m^*}{\partial x_j}+\left\langle\delta v_l\frac{\partial\delta B_m^*}{\partial x_j}\right\rangle\right), \tag{21c}$$

$$\frac{\partial\delta B_i^*}{\partial t}+\eta\nabla^2\delta B_i^*=\epsilon_{mjk}\epsilon_{kli}\left(V_l\frac{\partial\delta B_m^*}{\partial x_j}+\delta v_l\frac{\partial b_m^*}{\partial x_j}\right.$$

$$\left.+\delta v_l\frac{\partial\delta b_m^*}{\partial x_j}-\left\langle\delta v_l\frac{\partial\delta b_m^*}{\partial x_j}\right\rangle\right). \tag{21d}$$

Further, if we now ensemble average the conservation equations (20), we obtain

$$\frac{\partial}{\partial t}\int d^3\mathbf{x}\left[\eta\left\langle\frac{\partial b_i}{\partial x_j}\frac{\partial b_i^+}{\partial x_j}+\left\langle\frac{\partial\delta B_i}{\partial x_j}\frac{\partial\delta B_i^+}{\partial x_j}\right\rangle\right\rangle+\epsilon_{ijk}\epsilon_{klm}\left(V_l b_m\frac{\partial b_i^+}{\partial x_j}\right.\right.$$
$$\left.+\langle\delta v_l\delta B_m\rangle\frac{\partial b_i^+}{\partial x_j}+b_m\left\langle\delta v_l\frac{\partial\delta B_i^+}{\partial x_j}\right\rangle+V_l\left\langle\delta B_m\frac{\partial\delta B_i^+}{\partial x_j}\right\rangle\right)\Bigg]$$
$$=\epsilon_{ilk}\epsilon_{kjm}\int d^3\mathbf{x}\left(b_m\frac{\partial b_i^+}{\partial x_l}\frac{\partial V_j}{\partial t}+\frac{\partial V_j}{\partial t}\left\langle\delta B_m\frac{\partial\delta B_i^+}{\partial x_l}\right\rangle\right.$$
$$+b_m\left\langle\frac{\partial\delta v_j}{\partial t}\frac{\partial\delta B_i^+}{\partial x_l}\right\rangle+\left\langle\delta B_m\frac{\partial\delta v_j}{\partial t}\right\rangle\frac{\partial b_i^+}{\partial x_l}$$
$$\left.+\left\langle\frac{\partial\delta v_j}{\partial t}\frac{\partial\delta B_i^+}{\partial x_l}\delta B_m\right\rangle\right), \tag{22a}$$

$$\frac{\partial}{\partial t}\int d^3\mathbf{x}\left(b_j^+\frac{\partial b_\alpha^+}{\partial x_l}+\left\langle\delta B_j^+\frac{\partial\delta B_j}{\partial x_i}\right\rangle\right)$$
$$=-\epsilon_{\alpha lk}\epsilon_{kjm}\int d^3\mathbf{x}\left(b_m\frac{\partial b_\alpha^+}{\partial x_l}\frac{\partial V_j}{\partial x_i}\right.$$
$$+\frac{\partial V_j}{\partial x_i}\left\langle\delta B_m\frac{\partial\delta B_\alpha^+}{\partial x_l}\right\rangle+b_m\left\langle\frac{\partial\delta B_\alpha^+}{\partial x_l}\frac{\partial\delta v_j}{\partial x_i}\right\rangle$$
$$\left.+\frac{\partial b_2^+}{\partial x_l}\left\langle\delta B_m\frac{\partial\delta v_j}{\partial x_i}\right\rangle+\left\langle\frac{\partial\delta B_\alpha^+}{\partial x_l}\delta B_m\frac{\partial\delta v_j}{\partial x_i}\right\rangle\right) \tag{22b}$$

and

$$\frac{\partial}{\partial t}\int\epsilon_{\alpha\beta i}d^3\mathbf{x}\,(b_j^+\Omega_{\beta i}b_j+\langle\delta B_j^+\Omega_{\beta i}\delta B_j\rangle)$$
$$=-\epsilon_{plk}\epsilon_{kjm}\epsilon_{\alpha\beta i}\int d^3\mathbf{x}\left(b_m\frac{\partial b_p^+}{\partial x_l}\Omega_{\beta i}V_j+\left\langle\delta B_m\frac{\partial\delta B_p^+}{\partial x_l}\right\rangle\Omega_{\beta i}V_j\right.$$
$$+b_m\left\langle\frac{\partial\delta B_p^+}{\partial x_l}\Omega_{\beta i}\delta v_j\right\rangle+\frac{\partial b_p^+}{\partial x_l}\langle\delta B_m\Omega_{\beta i}\delta v_j\rangle$$
$$\left.+\left\langle\delta B_m\frac{\partial\delta B_p^+}{\partial x_l}\Omega_{\beta i}\delta v_j\right\rangle\right)-2\int d^3\mathbf{x}\left([b_\alpha^+\epsilon_{jlm}-b_j^+\epsilon_{\alpha lm}]\right.$$
$$\times\frac{\partial}{\partial x_j}\left(V_l b_m+\langle\delta v_l\delta B_m\rangle\right)+\langle\delta B_\alpha^+\epsilon_{jlm}-\delta B_j^+\epsilon_{\alpha lm}\rangle$$
$$\left.\times\frac{\partial}{\partial x_j}\left(\delta v_l b_m+V_l\delta B_m\right)\right)$$
$$+\left\langle(\delta B_\alpha^+\epsilon_{jlm}-\delta B_j^+\epsilon_{\alpha lm})\frac{\partial}{\partial x_j}(\delta v_l\delta B_m)\right\rangle, \tag{22c}$$

where $\Omega_{\beta i}=x_\beta\partial/\partial x_i-x_i\partial/\partial x_\beta$. As well as the ensemble average conserved quantities given by Eqs. (22), for each member of the ensemble we have individually conserved quantities obtained by subtracting Eqs. (22a), (22b), and (22c) from Eqs. (20a), (20b), and (20c), respectively.

Without further information on the form of the large-scale, ensemble averaged velocity field $\mathbf{V}(\mathbf{x},t)$ and without more information on the form of the velocity turbulence, it is difficult to proceed further with the general development of the conserved quantities. However, there does exist one particular class of situations which are often investigated in kinematic dynamo theory for their mathematical simplicity. This class occurs when correlations of quantities involving more than two random elements at a time are set to zero by *fiat*. This approximation is known under a variety of names; quasilinear theory, first order smoothing theory, the random Born approximation, etc. Its general realm of validity has not been established,[6] but its invalidity in several special cases has been established.[6–9] And, in fairness, we should also point out that its validity in two special cases has also been established recently.[7] We shall consider, then, the behavior of the conserved quantities under the quasilinear approximation.

B. Quasilinear approximation to conserved quantities

In the ensemble averaged equations ignore all correlations involving more than two random quantities. Under this approximation Eqs. (21a) and (21c) remain the same, while Eqs. (21b) and (21b) give respectively

$$\frac{\partial\delta B_i}{\partial t}-\eta\nabla^2\delta B_i=\epsilon_{ijk}\epsilon_{klm}\frac{\partial}{\partial x_j}(V_l\delta B_m+b_m\delta v_l), \tag{23a}$$

$$\frac{\partial\delta B_i^+}{\partial t}+\eta\nabla^2\delta B_i^+=\epsilon_{mjk}\epsilon_{kli}\left(V_l\frac{\partial\delta B_m^+}{\partial x_j}+\delta v_l\frac{\partial b_m^+}{\partial x_j}\right). \tag{23b}$$

Equations (22) then reduce to

$$\frac{\partial}{\partial t}\int d^3\mathbf{x}\left[\eta\left(\frac{\partial b_i}{\partial x_j}\frac{\partial b_i^+}{\partial x_j}+\left\langle\frac{\partial\delta B_i}{\partial x_j}\frac{\partial\delta B_i^+}{\partial x_j}\right\rangle\right)+\epsilon_{ijk}\epsilon_{klm}\left(V_l b_m\frac{\partial b_i^+}{\partial x_j}\right.\right.$$
$$\left.+\frac{\partial b_i^+}{\partial x_j}\langle\delta v_l\delta B_m\rangle+b_m\left\langle\delta v_l\frac{\partial\delta B_i^+}{\partial x_j}\right\rangle+V_l\left\langle\delta B_m\frac{\partial\delta B_i^+}{\partial x_j}\right\rangle\right)\Bigg]$$
$$=\epsilon_{ilk}\epsilon_{kjm}\int d^3\mathbf{x}\left(b_m\frac{\partial b_i^+}{\partial x_l}\frac{\partial V_j}{\partial t}+\frac{\partial V_j}{\partial t}\left\langle\delta B_m\frac{\partial\delta B_i^+}{\partial x_l}\right\rangle\right.$$
$$\left.+b_m\left\langle\frac{\partial\delta v_j}{\partial t}\frac{\partial\delta B_i^+}{\partial x_l}\right\rangle+\frac{\partial b_i^+}{\partial x_l}\left\langle\delta B_m\frac{\partial\delta v_j}{\partial t}\right\rangle\right). \tag{24a}$$

$$\frac{\partial}{\partial t}\int d^3\mathbf{x}\left(b_j^+\frac{\partial b_j}{\partial x_i}+\left\langle\delta B_j^+\frac{\partial\delta B_j}{\partial x_i}\right\rangle\right)$$
$$=-\epsilon_{\alpha lk}\epsilon_{kjm}\int d^3\mathbf{x}\left(b_m\frac{\partial b_\alpha^+}{\partial x_l}\frac{\partial V_j}{\partial x_i}+\frac{\partial V_j}{\partial x_i}\left\langle\delta B_m\frac{\partial\delta B_\alpha^+}{\partial x_l}\right\rangle\right.$$
$$\left.+b_m\left\langle\frac{\partial\delta B_\alpha^+}{\partial_l}\frac{\partial\delta v_j}{\partial x_i}\right\rangle+\frac{\partial b_\alpha^+}{\partial x_l}\left\langle\delta B_m\frac{\partial\delta v_j}{\partial x_i}\right\rangle\right), \tag{24b}$$

$$\frac{\partial}{\partial t}\int d^3\mathbf{x}\,\epsilon_{\alpha\beta i}\,(b_j^+\Omega_{\beta i}b_j+\langle\delta B_j^+\Omega_{\beta i}\delta B_j\rangle)$$
$$=-\epsilon_{plk}\epsilon_{kjm}\epsilon_{\alpha\beta i}\int d^3\mathbf{x}\left(b_m\frac{\partial b_p^+}{\partial x_l}\Omega_{\beta i}V_j\right.$$
$$+\left\langle\delta B_m\frac{\partial\delta B_p^+}{\partial x_l}\right\rangle\Omega_{\beta i}V_j+b_m\frac{\partial\delta B_p^+}{\partial x_l}\Omega_{\beta i}\delta v_j$$
$$\left.+\frac{\partial b_p^+}{\partial x_l}\langle\delta B_m\Omega_{\beta i}\delta v_j\rangle\right)-2\int d^3\mathbf{x}\left((b_\alpha^+\epsilon_{jlm}-b_j^+\epsilon_{\alpha lm})\right.$$

$$\times \frac{\partial}{\partial x_j}\,(V_l b_m + \langle \delta v_l \delta B_m \rangle) + \Big\langle (\delta B^+_\alpha \epsilon_{jlm} - \delta B^+_j \epsilon_{\alpha lm})$$

$$\times \frac{\partial}{\partial x_j}\,(\delta v_l b_m + V_l \delta B_m) \Big\rangle \Big\rangle . \tag{24c}$$

IV. A SIMPLE ILLUSTRATION

A. General remarks

Suppose that the velocity pattern is static, $\partial v_i/\partial t = 0$. Then energy conservation gives

$$\int d^3\mathbf{x} \left(\frac{\partial B_i}{\partial x_j} \frac{\partial B^+_i}{\partial x_j} \eta + \epsilon_{ijk}\epsilon_{klm} v_l B_m \frac{\partial B^+_i}{\partial x_j} \right) = E = \text{const.} \tag{25}$$

Thus, if there exists a most rapidly growing mode for B_i, Eq. (25) then implies that it must be exactly compensated by a decaying mode of B^+_i with the same time scale in order that E be constant. Reversing the argument, if one can show that all modes of B^+_i are decaying, then a guarantee exists that at least one mode of B_i is growing and dynamo action ensues. To put this point mathematically, consider again Eqs. (1) and (3). Write

$$B_i(\mathbf{x}, t) = \sum_\sigma B_i(\mathbf{x}, \sigma) \exp(\sigma t),$$

$$B^+_i(\mathbf{x}, t) = \sum_\sigma B^+_i(\mathbf{x}, \sigma) \exp(\sigma^+ t),$$

to obtain

$$\sigma B_i - \eta \nabla^2 B_i + \epsilon_{ijk}\epsilon_{klm} \frac{\partial}{\partial x_j}(v_l B_m) = 0, \tag{26a}$$

$$\sigma^+ B^+_m + \eta \nabla^2 B^+_m - \epsilon_{ijk}\epsilon_{klm} v_l \frac{\partial B^+_i}{\partial x_j} = 0. \tag{26b}$$

Now multiply Eq. (26a) by $B^+_i(\sigma^+)$, Eq. (26b) by $B_m(\sigma)$, add the results, and integrate over coordinate space to obtain

$$(\sigma + \sigma^+) \int d^3\mathbf{x}\, B_i(\mathbf{x}, \sigma) B^+_i(\mathbf{x}, \sigma^+) = 0, \tag{27}$$

where, once again, the surface terms have been set to zero.[8] From (27) we see that for $\sigma \neq -\sigma^+$, $\int d^3\mathbf{x} B_i(\mathbf{x}, \sigma) \times B^+_i(\mathbf{x}, \sigma^+)$ is zero (i.e., the normal modes of B^+_i are orthonormal to the normal modes of B_i at different eigen values), while for $\sigma = -\sigma^+$ we choose (in the standard manner) the orthonormality condition

$$\int d^3\mathbf{x}\, B_i(\mathbf{x}, \sigma) B^+_i(\mathbf{x}, \sigma^+) = \delta_{\sigma+\sigma^+}. \tag{28}$$

Hence, $\mathrm{Re}(\sigma) = -\mathrm{Re}(\sigma^+)$ and $\mathrm{Im}(\sigma) = -\mathrm{Im}(\sigma^+)$ for those modes satisfying Eq. (28). Thus, if $\mathrm{Re}\sigma^+ < 0$, implying a decaying (in time) adjoint field B^+_i, then $\mathrm{Re}\sigma > 0$, implying a growing (in time) dynamo field B_i.

Now this result (28) should also be contained in the Hamiltonian statement (25). To show that it is, consider Eq. (26b). Multiply it by $B_m(\sigma, \mathbf{x})$ and integrate the result over all of coordinate space to obtain

$$\sigma^+ \int B_m(\mathbf{x}, \sigma) B^+_m(\mathbf{x}, \sigma^+) d^3\mathbf{x}$$

$$= \int d^3\mathbf{x} \left(\eta \frac{\partial B_i}{\partial x_j}(\mathbf{x}, \sigma) \frac{\partial B^+_i}{\partial x_j}(\mathbf{x}, \sigma^+) \right.$$

$$\left. + \epsilon_{ijk}\epsilon_{klm} v_l B_m(\mathbf{x}, \sigma) \frac{\partial B^+_i}{\partial x_j}(\mathbf{x}, \sigma^+) \right). \tag{29}$$

With the temporal dependence included, Eq. (29) becomes

$$\int B_m(\mathbf{x}, t) \frac{\partial B^+_m}{\partial t}(\mathbf{x}, t)\, d^3\mathbf{x} = \int \mathcal{H}\, d^3\mathbf{x}. \tag{30}$$

But, according to the energy statement (25) we have

$$\frac{d}{dt} \int B_m(\mathbf{x}, t) \frac{\partial B^+_m}{\partial t}(\mathbf{x}, t)\, d^3\mathbf{x} = \frac{d}{dt} \int \mathcal{H}\, d^3\mathbf{x} = 0. \tag{31}$$

Writing $B_m(\mathbf{x}, t)$ *and* $B^+_m(\mathbf{x}, t)$ in terms of their temporal normal mode dependences, the conservation of the Hamiltonian yields, from Eq. (31),

$$\sigma^+(\sigma + \sigma^+) \int B_m(\mathbf{x}, \sigma) B^+_m(\mathbf{x}, \sigma^+) d^3\mathbf{x} = 0. \tag{32}$$

In other words, conservation of the Hamiltonian implies, and is implied by, the orthonormality statement (28) for static velocity patterns.

Thus, if it can be shown for a given velocity pattern that *any* one mode of the adjoint field B^+_i is growing (or decaying) then, under Eq. (28), there must exist a decaying (or growing) normal mode for the magnetic field B_i.

And, depending on the static velocity pattern, it can happen that the adjoint equation for B^+_i is considerably simpler to solve than the equation for B_i. It is in this sense that we regard the conserved quantities energy, momentum, and angular momentum as being criteria capable of indicating the possibility for kinematic dynamo activity ahead of detailed calculations.

It is, of course, not necessary to use modes with an exponential time dependence [and a space dependence determined by, say, Eqs. (26a) and (26b), which are spatially complete in $0 \leq |\mathbf{x}| \leq \infty$]. We could equally well use a set of modes which are spatially complete in coordinate space [say, $\exp(i\mathbf{k}\cdot\mathbf{x})$] and then obtain the appropriate time dependence. The argument given above would still obtain, of course (i.e., we can expand plane-waves $\exp(i\mathbf{k}\cdot\mathbf{x})$ in the form of any other set, spatially complete in $0 \leq |\mathbf{x}| \leq \infty$).[8]

B. A particular velocity pattern

In order to illustrate this approach, we consider the velocity pattern

$$v_i = a_{ij} x_j \tag{33}$$

with $a_{ij} = \text{const}$ and $a_{ii} = 0$ so that the velocity pattern (33) represents incompressible flow, $\partial v_i/\partial x_i = 0$. As we shall see directly, this particular pattern leads to a decaying temporal dependence for the modes of **B**. Thus it does not give dynamo action. However, it is useful as an illustration of the general technique.

Then with

$$B^+_i(\mathbf{x}, t) = \int B^+_i(\mathbf{k}, t) \exp(i\mathbf{k}\cdot\mathbf{x}) d^3\mathbf{k}, \tag{34}$$

Eq. (3) gives

$$\frac{\partial B^+_i}{\partial t} - \eta k^2 B^+_i + \epsilon_{mjk}\epsilon_{kli} a_{l\alpha} \frac{\partial}{\partial k_\alpha}(k_j B^+_m) = 0, \tag{35}$$

which can be written as

$$\frac{\partial B_i^+}{\partial t} - a_{m\alpha}k_m \frac{\partial B_i^+}{\partial k_\alpha} + a_{mi}B_m^+ - \eta k^2 B_i^+ = - a_{m\alpha}k_i \frac{\partial B_m^+}{\partial k_\alpha}. \tag{36}$$

In the interests of simplicity, we shall consider only the situation where $a_{m\alpha} = a\epsilon_{m\alpha\gamma}\lambda_\gamma/\lambda$ with a and λ constants. Change variables in Eq. (36) with

$$\tau = at, \quad \kappa_i = k_i \eta^{1/2}/a. \tag{37}$$

Equation (36) then reduces to

$$\frac{\partial B_i^+}{\partial \tau} - \kappa^2 B_i^+ + \epsilon_{mi\gamma}\lambda_\gamma \lambda^{-1} B_m^+ = - \kappa_i \epsilon_{m\alpha\gamma}\lambda_\gamma \lambda^{-1} \frac{\partial B_m^+}{\partial \kappa_\alpha}, \tag{38}$$

which is readily solved. First write

$$B_\parallel^+ = B_i^+ \lambda_i/\lambda, \quad U_\alpha = \epsilon_{m\alpha\gamma}\lambda_\gamma \lambda^{-1} B_m^+, \tag{39}$$

so that

$$\frac{\partial B_\parallel^+}{\partial \tau} - \kappa^2 B_\parallel^+ = - \lambda^{-1}(\boldsymbol{\lambda} \cdot \boldsymbol{\kappa}) \frac{\partial U_\alpha}{\partial \kappa_\alpha}, \tag{40a}$$

and

$$\frac{\partial U_\beta}{\partial \tau} - \kappa^2 U_\beta + \epsilon_{i\beta\gamma}\lambda_\gamma \lambda^{-1} U_i = - \epsilon_{i\beta\gamma}\kappa_i \lambda_\gamma \lambda^{-1} \frac{\partial U_\alpha}{\partial \kappa_\alpha}. \tag{40b}$$

Now rotate the κ coordinates until λ points along the Z axis. Write $\kappa_x = \kappa_\perp \cos\theta$, $\kappa_y = \kappa_\perp \sin\theta$, $\kappa_\parallel = \kappa_z$, and set

$$U_x \pm iU_y = \tfrac{1}{2} \exp[\tau(\kappa_\parallel^2 + \kappa_\perp^2)](s \pm r) \exp(\pm i\theta), \tag{41}$$

when Eqs. (40b) reduce to

$$\frac{\partial s}{\partial \tau} - \frac{\partial s}{\partial \theta} = 0, \tag{42a}$$

$$\frac{\partial r}{\partial \tau} - (1 + \cos 2\theta)\frac{\partial r}{\partial \theta} - r \sin 2\theta = - i\kappa_\perp \left(2\kappa_\perp \tau s + s \cos(2\theta)\kappa_\perp^{-1} + \frac{\partial s}{\partial \theta} \sin(2\theta)\kappa_\perp^{-1}\right). \tag{42b}$$

With $\xi = \tan\theta - 2\tau$, Eqs. (42a) and (42b) can be solved readily to give

$$s = s(\theta + \tau) \tag{43a}$$

and

$$r = G(\xi) \exp[H(\xi, \tau)] - i\kappa_\perp \exp[-H(\xi, \tau)] \times \int^\tau d\tau' M(\xi, \tau') \exp[H(\xi, \tau')], \tag{43b}$$

where

$$H(\xi, \tau) = \tfrac{1}{2} \ln[1 + (\xi + 2\tau)^2],$$

$$M(\xi, \tau) = \left(2\kappa_\perp \tau s(\theta + \tau) + s(\theta + \tau)\cos(2\theta)\kappa_\perp^{-1} + \kappa^{-1} \sin(2\theta) \frac{\partial s(\theta + \tau)}{\partial \theta}\right)_{\theta = \tan^{-1}(\xi)}$$

and where $G(\xi)$ and $s(\theta + \tau)$ are arbitrary, but periodic, functions of their arguments.

Note, then, that both U_x and U_y are growing functions of time for arbitrary, but real, $\mathbf{k}$ so that, from conservation of the Hamiltonian (25), we expect that all the normal modes of the magnetic field will be decaying in time.

To show directly that this is indeed the case, consider the magnetic induction equation (1) with $v_i = a\epsilon_{ijk}x_j\lambda_k/\lambda$. It is obvious, of course, that this defeats the general purpose of using the conserved Hamiltonian and the behavior of the adjoint field $\mathbf{B}^+$ to infer the behavior of $\mathbf{B}$. However, it is of use to illustrate the point that the temporal dependence of $\mathbf{B}$ does follow from knowledge of the behavior of $\mathbf{B}^+$ together with the conserved Hamiltonian statement. Upon Fourier transforming Eq. (1), we obtain

$$\frac{\partial B_i}{\partial t} + \eta k^2 B_i - a\epsilon_{ijk}\epsilon_{klm}\epsilon_{l\alpha\beta}\lambda_\beta \lambda^{-1} k_j \frac{\partial B_m}{\partial k_\alpha} = 0. \tag{44}$$

Again write $\tau = at$, $\kappa_i = k_i \eta^{1/2}/a$, and rotate the $\mathbf{k}$ coordinates until λ is pointing in the Z direction. Then Eq. (44) gives

$$\frac{\partial B_i}{\partial \tau} + \kappa^2 B_i + \left(\kappa_x \frac{\partial}{\partial \kappa_y} - \kappa_y \frac{\partial}{\partial \kappa_x}\right) B_i + \epsilon_{i\alpha\beta}\lambda_\beta \lambda^{-1} B_\alpha = 0. \tag{45}$$

Then with $\kappa_x = \kappa_\perp \cos\theta$, $\kappa_y = \kappa_\perp \sin\theta$, $\kappa_z = \kappa_\parallel$, Eq. (44) reduces to

$$\frac{\partial B_x}{\partial \tau} + \kappa^2 B_x - \frac{\partial B_x}{\partial \theta} + B_y = 0, \tag{44a}$$

$$\frac{\partial B_y}{\partial \tau} + \kappa^2 B_y - \frac{\partial B_y}{\partial \theta} - B_x = 0, \tag{46b}$$

$$\frac{\partial B_z}{\partial \tau} + \kappa^2 B_z - \frac{\partial B_z}{\partial \theta} = 0. \tag{46c}$$

With $\mathbf{B} = \mathbf{b} \exp(-\kappa^2 \tau)$ and $\xi = \theta + \tau$ Eqs. (46) reduce to

$$\frac{\partial b_x}{\partial \tau} + b_y = 0, \quad \frac{\partial b_y}{\partial \tau} - b_x = 0, \quad \frac{\partial b_z}{\partial \tau} = 0$$

with solution

$$b_z = b_z(\xi), \quad b_x = a_1(\xi) \exp(i\tau) + a_2(\xi) \exp(-i\tau),$$

$$b_y = i[a_1(\xi) \exp(i\tau) - a_2(\xi) \exp(-i\tau)],$$

where $b_z(\xi)$, $a_1(\xi)$ and $a_2(\xi)$ are arbitrary, but periodic, functions of their arguments. Thus the temporal behavior of $\mathbf{B}$ is decaying in time, with a decay rate ηk^2 which, apart from a periodic oscillation, is exactly the opposite temporal dependence of the normal modes of $\mathbf{B}^+$. Thus $\sigma = -\sigma^+$ by direct construction. Hence a static velocity pattern

$$v_i = a\epsilon_{ijk}x_j\lambda_k/\lambda$$

gives rise to damped, but oscillatory, magnetic fields in an infinite medium; and this fact can be demonstrated either by direct solution of the magnetic induction equation (1) or by solution of the adjoint field equation together with use of the conserved Hamiltonian.

While we recognize that the particular situation chosen is one in which we can solve either the induction equation or its adjoint equation with equal facility, the point we wish to make is that we can infer the temporal behavior of the normal modes of the magnetic field $\mathbf{B}$ by first solving for the adjoint field's normal modes and then using the conserved Hamiltonian statement, without having to solve the induction equation at all.

V. DISCUSSION AND CONCLUSION

In this paper we have obtained the conserved quantities "energy," "momentum," and "angular momentum" that follow from the magnetic induction equation, and its adjoint, in an infinite medium of constant resistivity. We also obtained the average conserved quantities in the case that the fluid velocity of the medium is turbulent, and we then gave the approximate form of the conserved quantities under the quasilinear assumption to the turbulent flow.

There were, and are, two main purposes for deriving the conserved quantities. The first, as we have already mentioned in the Introduction, is the fact that in any realistic fluid flow the kinematic dynamo equations are most often solved numerically using a variational principle.[5] The non-self-adjointness of the equations then means that the variational principle is also not self-adjoint. This in turn requires some extraneous criteria of goodness to be imposed on the numerical solutions at any given level of truncation. And the difficulties that then arise are due to the fact that the criteria of goodness used are not independent of the human element. Thus imposition of different criteria of goodness by different investigators means that any particular numerical solution is likely to be inaccurate to a degree which is difficult to define. The conserved energy, momentum, and angular momentum alleviate this problem to the extent that they provide internal criteria of goodness which arise directly from the induction equation instead of having to be externally imposed. Thus in any particular numerical calculation one can check to see if the numerics is preserving the energy, momentum, and angular momentum, and the numerical step size can be adjusted until they are preserved. This is clearly an improvement over ad hoc criteria of accuracy.

The second main purpose was to obtain criteria that may be used to indicate whether a particular fluid motion is at all capable of giving rise to kinematic dynamo activity. The point here is that it is notoriously difficult to obtain solutions to the kinematic dynamo equations even for simple fluid velocity patterns. Thus, before one commits an inordinate amount of time and effort in an attempt to solve the kinematic dynamo equations under some particular velocity field, it would be useful if one had criteria that could at least give an indication that there is a possibility for dynamo action ahead of any detailed calculations. The conserved energy, momentum, and angular momentum are clearly useful in this respect. For example, we considered the case of a static velocity pattern, $\partial v_i/\partial t = 0$. Then energy conservation gave

$$\int d^3\mathbf{x} \left(\eta \frac{\partial B_i}{\partial x_j} \frac{\partial B_i^+}{\partial x_j} + \epsilon_{ijk}\epsilon_{klm} v_l B_m \frac{\partial B_i^+}{\partial x_j} \right) = E = \text{const.}$$

We then showed that conservation of the Hamiltonian together with solutions to the adjoint field equation implied the temporal dependence of the magnetic field, and we verified this by solving directly the magnetic induction equation.

The point to be made here is that, depending on the velocity pattern, it can happen that the adjoint equation for $\mathbf{B}^+$ is considerably simpler to solve than the equation for $\mathbf{B}$. Then by inference, and without knowing directly the solution to the induction equation, one can categorically assert whether the induction equation admits of a dynamo mode or not.

It is in this sort of sense that we regard the conserved quantities, energy, momentum, and angular momentum as being criteria capable of indicating the possibility for kinematic dynamo activity, ahead of detailed calculations. We shall report on several such investigations at a later date.

ACKNOWLEDGMENT

The work reported here was performed during the first half year of 1975 while I was a visitor at the Center for Space Research Massachusetts Institute of Technology, Cambridge, Massachusetts. I am grateful to the Associate Director of the Center, Dr. H.S. Bridge, for the hospitality accorded me during my visit. The work was supported, in part, by the United States Air Force, Air Force Systems Command under contract F-19628-75-C-0101. It was also supported, in part, by M.I.T.

*Permanent address: Department of Physics, University of Chicago, Chicago, Ill. 60637.

[1] D. Gubbins, Rev. Geophys. & Sp. Phys. **12** (1974).
[2] E.N. Parker, Astrophys. J. **122**, 293 (1955).
[3] M. Stix, Astron. and Astrophys. **24**, 275 (1973).
[4] W.K. Rose and E.H. Levy, Astrophys. J. **193**, 419.(1974)
[5] I. Lerche, Astrophys. J. **176**, 225 (1972).
[6] U. Frisch, *Wave Propagation in Random Media* in *Probabilistic Methods in Applied Mathematics, Vol. I*, edited by A.T. Bharucha-Reid (Academic, New York, 1968).
[7] I. Lerche and E.N. Parker, J. Math. Phys. **14**, 1949 (1973).
[8] R.H. Kraichnan, J. Math. Phys. **2**, 124 (1962).
[9] P.H. Roberts and A.M. Soward, J. Math. Phys. **16**, 609 (1975); see, however, I. Lerche and E.N. Parker, J. Math. Phys. **16**, 1838 (1975).

Properties of massless relativistic fields under the conformal group

G. Post

Institut für Theoretische Physik, RWTH Aachen, F.R. Germany
(Received 1 April 1975)

Using the ladder representations of SU(2,2), we derive explicit transformation laws for massless free fields with arbitrary helicities under global conformal transformations.

1. INTRODUCTION

The group SU(2, 2) is known to act on Minkowski space—time as an extension of the group of Poincaré transformations and dilations. It is, precisely, a fourfold covering group of the conformal group $C(M^4)$ of a compactification M_c^4 of ordinary Minkowski space M^4; see Ref. 1. This paper deals with the well-known role of the conformal group as a symmetry of zero-mass wave equations.[2–4] We consider some global aspects of this symmetry. The starting point is the work of Mack and Todorov,[5] who showed that the irreducible Wigner representations of the Poincaré group, characterized by mass zero and the helicities $\lambda = 0, \pm 1/2, \pm 1, \ldots$, can be extended to representations of the Lie algebra of the conformal group by generators of unitary transformations. This series of representations is called the ladder series of SU(2, 2).

After a review (Sec. 2) of the action of SU(2, 2) on Minkowski space and its (incomplete) decomposition into the subgroups of Lorentz transformations, dilations, translations, and special conformal transformations, we discuss in Secs. 3 and 4 the results of Ref. 5 from the global point of view: The ladder representation of the SU(2, 2) Lie algebra is exponentiated to a unitary representation of the matrix group itself. The connection between the irreducible parts of the ladder series and the Wigner representations is investigated in detail. For instance, the helicity determines the global behavior; only odd helicities belong to representations of the conformal group itself, even helicities to $SO_0(4, 2)$, and half-integer helicities to SU(2, 2). For the special conformal transformations, we calculate the ladder representations explicitly in terms of integral transforms on the zero mass shell. Finally, we use the results of Secs. 3 and 4 to extend the work of Weinberg[6,7] on zero-mass field equations to include the transformation properties of these fields under finite conformal transformations.

2. DECOMPOSITION OF SU (2,2) AND THE CONFORMAL GROUP C (M^4)

In this section, we collect the notational conventions and the basic formulas about the 15-parameter group SU(2, 2) and the conformal group $C(M^4)$ of compactified Minkowski space M_c^4.[1] We discuss the factorization of these groups in terms of Lorentz group ($SL(2, \mathbb{C})$), dilations, translations, and special conformal transformations.

Notations

The Minkowski scalar product is

$$x \cdot y = x^T g y = x_0 y_0 - \mathbf{x}\mathbf{y}, \quad x, y \in \mathbb{R}^4 \ (=: M^4). \tag{2.1}$$

The basis of Pauli matrices

$$\{\sigma_0, \boldsymbol{\sigma}\} = \{1, \sigma_1, \sigma_2, \sigma_3\} = \left\{ \begin{pmatrix} 1 & \\ & 1 \end{pmatrix}, \begin{pmatrix} & 1 \\ 1 & \end{pmatrix}, \begin{pmatrix} & -i \\ i & \end{pmatrix}, \begin{pmatrix} 1 & \\ & -1 \end{pmatrix} \right\} \tag{2.2}$$

establishes a linear isomorphism of $\mathbb{R}^4$ with the set $H(2\times 2)$ of Hermitian 2×2 matrices, resp. of $\mathbb{C}^4$ with the set $\mathbb{C}(2\times 2)$ of all complex 2×2 matrices:

$$x = (x_0, \mathbf{x}) \mapsto \underset{\sim}{x} := x_0 \cdot 1 + \mathbf{x}\boldsymbol{\sigma}, \tag{2.3}$$

inverse map: $B \mapsto \underline{B}$; $(\underline{B})_\mu = \frac{1}{2}\mathrm{tr}(B\sigma_\mu)$.

The covering homomorphism Λ: $SL(2, \mathbb{C}) \to$ Lorentz group $L_+^\uparrow$ is defined by the following action of $SL(2, \mathbb{C})$ on $H(2\times 2)$:

$$\begin{aligned} &SL(2, \mathbb{C}) \times H(2\times 2) \to H(2\times 2): (A, B) \mapsto ABA^* \\ &\quad \Leftrightarrow \underline{ABA^*} = \Lambda(A)\underline{B} \quad [\Lambda(A) \in L_+^\uparrow]. \end{aligned} \tag{2.4}$$

The space inversion $\hat{}: (x_0, \mathbf{x}) \mapsto (x_0, -\mathbf{x})$ in $\mathbb{C}^4$ induces an inversion of 2×2 matrices, denoted by a hat:

$$\begin{aligned} &B = (b_0 \cdot 1 + \mathbf{b}\boldsymbol{\sigma}) \mapsto \hat{B} = (b_0 \cdot 1 - \mathbf{b}\boldsymbol{\sigma}) = (\sigma_2 B \sigma_2)^T, \\ &\begin{pmatrix} a & \hat{} & b \\ c & & d \end{pmatrix} = \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}. \end{aligned} \tag{2.5}$$

We note the rules $[A, B \in \mathbb{C}(2\times 2)]$:

$$\begin{aligned} &\widehat{A+B} = \hat{A} + \hat{B}, && \hat{A}A = A\hat{A} = (\underline{A} \cdot \underline{A}) \cdot 1 = \det A \cdot 1, \\ &\hat{A}\hat{B} = \hat{B}\hat{A}, && \det A \neq 0 \Rightarrow A^{-1} = \hat{A}\,(\det A)^{-1}, \\ &\hat{\hat{A}} = A, && \underline{A} \cdot \underline{B} = \tfrac{1}{2}\mathrm{tr} A\hat{B}, \\ &\mathrm{tr}\hat{A} = \mathrm{tr}A, && \det(1 + A) = 1 + \mathrm{tr}A + \det A. \end{aligned} \tag{2.6}$$

The group SU(2, 2) is, in standard matrix realization,

$$\left.\begin{aligned} U(2, 2) &= \{G \in GL(4, \mathbb{C}) \mid G^* I G = I\} \\ SU(2, 2) &= \{G \in U(2, 2) \mid \det G = 1\} \end{aligned}\right\},$$

$$I = \begin{pmatrix} 1 & \\ & -1 \end{pmatrix} = \begin{pmatrix} \sigma_0 & \\ & -\sigma_0 \end{pmatrix}. \tag{2.7}$$

We shall use 2×2 block notation for 4×4 matrices.

Another realization of SU(2, 2)[8] is more convenient to our purposes: We perform a similarity transformation

$$G \mapsto UGU^{-1} \quad \text{with} \quad U = \frac{1}{\sqrt{2}} \begin{pmatrix} l & -l \\ il & il \end{pmatrix} \quad \text{(unitary)}. \tag{2.8}$$

The formula reads in block notation

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix} \mapsto \frac{1}{2} \begin{pmatrix} A-B-C+D & -i(A+B-C-D) \\ i(A-B+C-D) & A+B+C+D \end{pmatrix} \equiv \begin{pmatrix} R & S \\ T & Q \end{pmatrix}. \tag{2.9}$$

The invariant metric transforms into

$$J = UIU^{+} = \begin{pmatrix} & -il \\ il & \end{pmatrix} \tag{2.10}$$

Subgroups of SU (2,2)

From now on, [S] U(2, 2) will be considered in J realization:

$$[\mathrm{S}]\,\mathrm{U}(2,2) = \{G \in \mathrm{GL}(4,\mathbb{C}) \,|\, G^{+}JG = J,\ [\det G = 1]\}. \tag{2.11}$$

As will be shown at the end of this section, the action of SU(2, 2) on the (minkowski) space $H(2\times 2)$ is as follows:

$$\left(G = \begin{pmatrix} R & S \\ T & Q \end{pmatrix},\ B = \underset{\sim}{b}\right) \mapsto (RB+S)(TB+Q)^{-1} \equiv \subset G(\underset{\sim}{b}). \tag{2.12}$$

$G \mapsto \subset G$ is the covering homomorphism of SU(2, 2) onto the conformal group $\mathrm{C}(M^4)$.

With (2. 11) we obtain the matrix representation of the Lie algebra [s] u(2, 2) of [S] U(2, 2):

$$[\mathrm{s}]\,\mathrm{u}(2,2) = \left\{ \begin{pmatrix} a & b \\ c & -a^{+} \end{pmatrix} \;\middle|\; b, c \in H(2\times 2);\ a \in \mathbb{C}(2\times 2); \right\} \quad [\mathrm{tr}\, a \in \mathbb{R}]\ . \tag{2.13}$$

We are now ready to comment on some subgroups of [S] U(2, 2):

$$\begin{aligned} G &:= \left\{ \begin{pmatrix} R & S \\ T & Q \end{pmatrix} \in U(2,2) \;\middle|\; S = T = 0 \right\} \\ &= \left\{ l(A) := \begin{pmatrix} A & \\ & (A^{+})^{-1} \end{pmatrix} \;\middle|\; A \in \mathrm{GL}(2,\mathbb{C}) \right\} \\ E &:= G \cap \mathrm{SU}(2,2) = \{ l(A) \,|\, \det A \in \mathbb{R} \} \\ L &:= \{ l(A) \,|\, \det A = 1 \} \\ D &:= \{ l(A) \,|\, A = \exp t,\ t \in \mathbb{R} \},\quad d(t) := l(\exp(t \cdot 1)) \\ U &:= \{ l(A) \,|\, A = \exp(it),\ t \in \mathbb{R} \}\ [\text{the center of U(2, 2)}] \\ Z &:= U \cap \mathrm{SU}(2,2) = \{ i, -1, -i, 1 \}\ [\text{center of SU(2, 2)}] \end{aligned} \tag{2.14}$$

These subgroups of U(2, 2) act on Minkowski space by means of Eq. (2. 12):

$$\subset l(A)\underset{\sim}{x} = A\underset{\sim}{x}A^{+} \tag{2.15}$$

The groups U and Z act trivially; the group L is isomorphic to $\mathrm{SL}(2,\mathbb{C})$ and acts as the covering (2. 4) of the Lorentz group: $\subset l(A) = \Lambda(A)$. Finally, the group D is the dilation group of Minkowski space: $\subset d(t)\underset{\sim}{x} = \exp(2t)\underset{\sim}{x}$.

The group E is the product LDZ; it is the union of two components ($\det A \gtrless 0$), and each $l(A) \in E$ has the unique decomposition

$$l(A) = l(A')\, d(t)\epsilon$$

with

$$\begin{aligned} Z \ni \epsilon &= \left\{ \begin{matrix} 1 : \det A > 0 \\ i : \det A < 0 \end{matrix} \right\} = (\mathrm{sgn}\, \det A)^{1/2}, \\ D \ni d(t),&\quad e^{t} = |\det A|^{1/2}, \\ L \ni l(A'),&\quad A' = A \cdot \exp(-t) \cdot \epsilon^{-1} \in \mathrm{SL}(2,\mathbb{C}). \end{aligned} \tag{2.16}$$

We introduce two four-parameter Abelian subgroups of SU(2, 2):

$$\begin{aligned} T &:= \left\{ t(B) = \begin{pmatrix} 1 & B \\ 0 & 1 \end{pmatrix} = \exp \begin{pmatrix} 0 & B \\ 0 & 0 \end{pmatrix} \;\middle|\; B \in H(2\times 2) \right\}, \\ \mathrm{S} &:= \left\{ s(B) = \begin{pmatrix} 1 & 0 \\ B & 1 \end{pmatrix} = \exp \begin{pmatrix} 0 & 0 \\ B & 0 \end{pmatrix} \;\middle|\; B \in H(2\times 2) \right\}. \end{aligned} \tag{2.17}$$

The group T induces Minkowski space translations: $\subset t(\underset{\sim}{b})\underset{\sim}{x} = \underset{\sim}{x} + \underset{\sim}{b}$. S acts as the "special conformal group" on Minkowski space: With the help of the rules (2. 6), (2. 12) we obtain ($x \in M^4$)

$$\subset s(B): x \mapsto [x + (x \cdot x)y]/\omega(x,y) \tag{2.18}$$

with $y = \hat{\underset{\sim}{B}}$ and $\omega(x,y) := 1 + 2(x \cdot y) + (x \cdot x)(y \cdot y)$.

Finally, we consider the two subgroups of "triangular" matrices:

$$\begin{aligned} P &:= \left\{ \begin{pmatrix} R & S \\ T & Q \end{pmatrix} \in \mathrm{SU}(2,2) \;\middle|\; T = 0 \right\}, \\ K &:= \left\{ \begin{pmatrix} R & S \\ T & Q \end{pmatrix} \in \mathrm{SU}(2,2) \;\middle|\; S = 0 \right\}, \end{aligned} \tag{2.19}$$

Some properties of these groups:

a. P is the semidirect product $P = T \circledS E$; each element of P has the form

$$\begin{pmatrix} A & B \\ 0 & (A^{+})^{-1} \end{pmatrix} = t(B(A^{+})^{-1}) \cdot l(A) \tag{2.20}$$

b. $P = \{G \in \mathrm{SU}(2,2) \,|\, (\subset G)(\underset{\sim}{x}) \text{ exists } \forall\ \underset{\sim}{x} \in H(2\times 2)\}$, so that P is the maximal subgroup of SU(2, 2) leaving invariant the uncompactified M^4. The image of P under $\subset$ is the Poincaré group including dilations.

a′. K is isomorphic to P; $K = S \circledS E$ (semidirect product); each element of K has the form

$$\begin{pmatrix} A & 0 \\ B & (A^{+})^{-1} \end{pmatrix} = s(BA^{-1}) \cdot l(A) \quad \text{(unique decomposition)}. \tag{2.21}$$

b′. $K = \{G \in \mathrm{SU}(2,2) \,|\, (\subset G)(0) = 0\}$, so that K is the maximal subgroup of SU(2, 2) which leaves the origin $O \in M^4_c$ fixed, i.e., K is the little group of O. $\subset K$ consists of Lorentz-, dilation-, and special-conformal group.

Factorization of SU (2,2)

We add some remarks on the "standard factorization"[9] of SU(2, 2). We wish to obtain a decomposition of any

$G \in SU(2,2)$ into a product of a "special conformal transformation" ($\in S$) and a "dilation-Poincaré-transformation" ($\in P$).

The ansatz

$$G = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \overset{!}{=} \begin{pmatrix} 1 & 0 \\ K & 1 \end{pmatrix} \begin{pmatrix} A' & B' \\ 0 & A' \end{pmatrix} = \begin{pmatrix} A' & B' \\ KA' & KB'+A' \end{pmatrix} \tag{2.22}$$

shows that the decomposition exists iff $A = A'$ is invertible. In this case, it is also unique: $A' = A$, $B' = B$, $K = CA^{-1}$. Collecting the formulas (2.22), (2.20) and (2.16), we have the following:

Result: The matrix $G = \left(\begin{smallmatrix} A & B \\ C & D \end{smallmatrix}\right) \in SU(2,2)$ can be decomposed into the product

$$G = s(K)\,t(H)\,l(A) = s(K)t(H)l(A')d(t)\epsilon \in STLDZ \tag{2.23}$$

iff $\det A \neq 0$. The factorization is unique, $K = CA^{-1}$ and $H = BA^*$ are Hermitian matrices, and A', t, ϵ are given by Eqs. (2.16).

With the homomorphism $\subset$, this decomposition carries over to the conformal group; the factor Z disappears. Thus, only "almost" every conformal transformation has a factorization into special conformal transformation, translation, Lorentz transformation, and dilation (in this order, and unique).

Here is an example for an SU(2,2) matrix which cannot be decomposed, because its submatrix A vanishes:

$$Z_1 = \begin{pmatrix} 0 & -l \\ l & 0 \end{pmatrix}, \quad (\subset Z_1)x = -\frac{\hat{x}}{(x \cdot x)} \; (x \in M^4). \tag{2.24}$$

The set $N = \{\det A = 0\}$ of "non-parametrizable" matrices has the codimension 1 in the SU(2,2) manifold and separates the two domains $\det A \gtrless 0$. N does not form a group, nor does $SU(2,2) \setminus N$. We get from the definition (2.19): $n \in N \Rightarrow np \in N \; \forall \, p \in P$. Thus N is a union of cosets $n \cdot P$. Indeed, the standard factorization is just the attempt to parametrize the coset space $SU(2,2)/P = \{G \cdot P\}$ by the special conformal subgroup S ($\cong \mathbb{R}^4$), each $s \in S$ representing its coset $s \cdot P$. But the coset space is compact and is thus not completely covered by a coordinate system $\mathbb{R}^4$.

By analogy, the coset space $SU(2,2)/K$ is not covered by the translation group T. In fact, $SU(2,2)/K$ is exactly the compactified Minkowski space M_c^4, and T is the ordinary part of it. The homomorphism $\subset : SU(2,2) \to C(M^4)$ results from the natural action of SU(2,2) on its coset space by K. In particular, the formula (2.12) is read off from the following ansatz, concerning those cosets which contain a representative element of T:

$$\begin{pmatrix} R & S \\ T & Q \end{pmatrix} \cdot \begin{pmatrix} l & B \\ 0 & l \end{pmatrix} = \begin{pmatrix} l & \subset G(B) \\ 0 & l \end{pmatrix} \cdot k \quad \text{with some } k \in K.$$

3. THE LADDER REPRESENTATIONS OF U(2,2)

We shall recall the work of Mack and Todorov[5] on the ladder representations of U(2,2), considering the group itself instead of the Lie algebra.

Notations and definition

In the space $\mathbb{C}^2$ of two complex variables (z_1, z_2), $z_j = u_j + iv_j$, we have the differential operators

$$\partial_j = \tfrac{1}{2}(\partial_{u_j} - i\partial_{v_j}), \quad \bar{\partial}_j = \tfrac{1}{2}(\partial_{u_j} + i\partial_{v_j}).$$

We introduce

$$z = \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} \quad \bar{z} = \begin{pmatrix} \bar{z}_1 \\ \bar{z}_2 \end{pmatrix}, \quad \partial = \begin{pmatrix} \partial_1 \\ \partial_2 \end{pmatrix}, \quad \bar{\partial} = \begin{pmatrix} \bar{\partial}_1 \\ \bar{\partial}_2 \end{pmatrix}.$$

The corresponding rows will be denoted by $z^T = (z_1, z_2)$, etc. We also define the four-column $\varphi = \left(\begin{smallmatrix} \xi \\ \bar{\eta} \end{smallmatrix}\right)$ and the four-row $\tilde{\varphi} = (-\bar{\eta}^T, z^T)$. The ladder representation of U(2,2) can now be defined on the space $L^2(\mathbb{C}^2)$ of square integrable functions on $\mathbb{C}^2$, with respect to Lebesgue measure. U(2,2) is considered, for the moment, in a matrix realization relative to the metric $J' = \left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$. Then the ladder representation[5] assigns to each matrix h' of the Lie algebra u(2,2) the following differentiation–multiplication operator on $L^2(\mathbb{C}^2)$: $\mathcal{L}'(h') := \tilde{\varphi} h' \varphi$. $\mathcal{L}'$ is a skew-self-adjoint representation of u(2,2), as follows from the construction of φ and $\tilde{\varphi}$ discussed in Ref. 5. See also Ref. 10 for general u(p,q) ladder representations.

We prefer the metric J [(2.10)], which is unitarily equivalent to J':

$$J = VJ'V^+ \text{ with } V = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 & (1-i)l \\ (1+i)l & 0 \end{pmatrix}.$$

So $h = Vh'V^{-1}$ is our Lie algebra element in the J metric, and we define $\mathcal{L}(h) := \mathcal{L}'(h') = \tilde{\varphi} V^{-1} h V \varphi$. Explicit calculation with Eq. (2.13) gives us the following formula, which we shall take as the definition of the u(2,2) ladder representation:

$$\mathcal{L}\left(h = \begin{pmatrix} a & b \\ c & -a^* \end{pmatrix}\right) = z^T a \partial + \bar{\partial}^T a^* \bar{z} + iz^T b \bar{z} + i\bar{\partial}^T c \partial. \tag{3.1}$$

Exponentiation of $\mathcal{L}$

First, we prove that $\mathcal{L}$ can be exponentiated to a unitary (reducible) representation of the matrix group U(2,2). It has been shown,[10] by an application of a theorem of Nelson, that the ladder representations of any u(p,q) algebra can be integrated to unitary representations of the universal covering groups $\mathrm{U}(p,q)$.

Since $\overline{\mathrm{U}}(p,q) = \overline{\mathrm{SU}}(p,q) \times \mathbb{R}$, they generate in particular representations of $\overline{\mathrm{SU}}(p,q)$. Now the covering $\pi : \overline{\mathrm{SU}}(p,q) \to \mathrm{SU}(p,q)$ has been investigated by Schaaf,[11] and it maps essentially a real line onto a circle.

Thus, in order to prove that $\mathcal{L}$ is a representation of SU(2,2) itself, it suffices to show $\exp[2\pi \mathcal{L}(\Gamma)] = 1$, where Γ is the generator of the above-mentioned circle subgroup. In standard metric (2.7) this subgroup is[11]

$$\exp(\Gamma t) = G(t) = \begin{pmatrix} \begin{pmatrix} \exp(it) & 0 \\ 0 & 1 \end{pmatrix} & 0 \\ 0 & \begin{pmatrix} \exp(-it) & 0 \\ 0 & 1 \end{pmatrix} \end{pmatrix}.$$

We obtain in J metric

$$\Gamma = \begin{pmatrix} 0 & \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \\ \begin{pmatrix} -1 & 0 \\ 0 & 0 \end{pmatrix} & 0 \end{pmatrix} \quad \text{[by Eq. (2.9)]},$$

$$\mathcal{L}(\Gamma) = i(u_1^2 - \tfrac{1}{4}\partial_{u1}^2 + v_1^2 - \tfrac{1}{4}\partial_{v1}^2) \quad \text{[by Eq. (3.1)]}.$$

Since this is essentially the Hamilton operator of a two-dimensional isotropic harmonic oscillator, we can easily find a complete orthogonal set $\{f_{nmk} | n, m, k, = 0, 1, 2, \cdots\}$ of $L^2(\mathbb{C}^2)$ so that $\mathcal{L}(\Gamma) f_{nmk} = i(n+m+1) f_{nmk}$. This implies $\exp[2\pi \mathcal{L}(\Gamma)] = 1$ and thus finishes the proof that $\mathcal{L}$ generates a representation of SU(2, 2). The extension of this property to U(2, 2) follows from the explicit calculations below. □

Next, we start from Eq. (3.1) to obtain the explicit form of the exponentiated ladder representation (also called $\mathcal{L}$) on several subgroups of U(2, 2).

G: The group $G \subset$ U(2, 2) [Eqs. (2.14)] is the exponential of the Lie algebra $\{\left(\begin{smallmatrix} a & 0 \\ 0 & -a^+ \end{smallmatrix}\right)\}$. We give the result

$$[\mathcal{L}\, l(A) f]z = \det A\, f(A^T z),$$
$$A \in \mathrm{GL}(2, \mathbb{C}),\ z \in \mathbb{C}^2,\ f \in L^2. \tag{3.2}$$

This is obviously a unitary representation of G, and the calculation of the corresponding Lie algebra representation (using one-parameter subgroups) produces exactly (3.1), $b = c = 0$.

T: For the translation group T [Eqs. (2.17)], the definition (3.1) yields immediately the following unitary representation:

$$[\mathcal{L}\, t(B) f]z = \exp(i z^T B \bar{z}) f(z) \tag{3.3}$$

S: In order to treat the group S [Eqs. (2.17)], we introduce the unitary automorphism $\mathcal{J}$ of $L^2(\mathbb{C}^2)$, which is induced by the Fourier transform in $\mathbb{C}^2$:

$$(\mathcal{J} f)w = (1/2\pi)^2 \int dz\, f(z) \exp[i(z, w)], \quad (z, w) := \mathrm{Re} z^T \bar{w},$$
$$(\mathcal{J}^{-1} f)z = (1/2\pi)^2 \int dw\, f(w) \exp[-i(z, w)], \quad z, w \in \mathbb{C}^2, \tag{3.4}$$

The following operator identity holds[12]:

$$\mathcal{J}^{-1}(\bar{\partial}^T B \partial)\mathcal{J} = z^T(-\tfrac{1}{4}B)\bar{z} \quad [B \in H(2\times 2)]. \tag{3.5}$$

We can now combine Eqs. (3.1), (3.3), (3.5) and obtain

$$s(B) = \begin{pmatrix} 1 & 0 \\ B & 1 \end{pmatrix} = \exp\begin{pmatrix} 0 & 0 \\ B & 0 \end{pmatrix} \mapsto \mathcal{L}\, s(B) = \mathcal{J}[\mathcal{L}\, t(-\tfrac{1}{4}B)]\mathcal{J}^{-1}. \tag{3.6}$$

This unitary representation of the subgroup S can be written as an integral transform, if the submatrix B is nonsingular, that is, $\underline{B}$ spacelike or timelike (see Ref. 12):

$$[\mathcal{L}\, s(B) f]z = (1/2\pi)^2 \int dw\, f(w) L_B(z - w)$$
$$\text{with } L_B(z) := -4(\det B)^{-1} \exp(i z^T B^{-1} \bar{z}). \tag{3.7}$$

Finally, we add some remarks.

a. The Fourier transformation $\mathcal{J}$ itself is the representation $\mathcal{L}$ of an element of SU(2, 2): $\mathcal{J} = \mathcal{L}(Z_2)$ with [see Eq. (2.24)]

$$Z_2 = Z_1 \cdot d(\log 2) = \begin{pmatrix} 0 & -\frac{1}{2} \\ 2 & 0 \end{pmatrix} = t(-\tfrac{1}{2}) s(2) t(-\tfrac{1}{2}). \tag{3.8}$$

A proof is obtained by applying (3.3) and (3.7) to (3.8).

b. The following relations are easily verified:

$$\mathcal{J}(\mathcal{L}\, l(A))\mathcal{J}^{-1} = \mathcal{L}\, l((A^+)^{-1}),$$
$$\mathcal{J}(\mathcal{L}\, t(B))\mathcal{J}^{-1} = \mathcal{L}\, s(-4B). \tag{3.9}$$

Decomposition of the (reducible) representation $\mathcal{L}$

The space $L^2(\mathbb{C}^2)$ is the direct sum of orthogonal subspaces H_n:

$$L^2(\mathbb{C}^2) = \bigoplus_{n \in \mathbb{Z}} H_n \quad (n = 0, \pm 1, \pm 2 \ldots)$$
$$H_n := \{f \in L^2(\mathbb{C}^2) \mid f(\exp(it)z) = \exp(int) f(z) \quad t \in \mathbb{R},\ z \in \mathbb{C}^2\}. \tag{3.10}$$

The H_n are invariant under the representation $\mathcal{L}$, since $f(z) \mapsto f(\exp(it)z)$ is easily seen to commute with the operators $\mathcal{L}\, l(A)$, $\mathcal{L}\, t(B)$, and $\mathcal{J}$, which generate the whole set $\mathcal{L}[\mathrm{U}(2, 2)]$. We denote the restriction of $\mathcal{L}$ to H_n by $\mathcal{L}_n$ and call the collection of $\mathcal{L}_n$ the ladder series. The irreducibility of the representations $\mathcal{L}_n$ of U(2, 2) is shown in the next section.

We classify the $\mathcal{L}_n$ according to their behavior with respect to the center $Z = \{i, i^2, i^3, i^4\}$ of SU(2, 2). Equations (3.2), (3.10) imply $\mathcal{L}(i) f(z) = -f(iz)$,

$$f \in H_n \Rightarrow \mathcal{L}_n(i) f(z) = -i^n f(z). \tag{3.11}$$

The group $\mathrm{SO}_0(4.2)$ occurs in a sequence of homomorphisms

$$\mathrm{SU}(2, 2) \to \mathrm{SO}_0(4.2) \to \mathrm{C}(M^4)$$
$$\text{with centers } \{\pm 1, \pm i\} \to \{1, -1\} \to \{1\}.$$

Therefore, (3.11) leads to the classification:

a. For $n \in 2\mathbb{Z} + 1$,

$\mathcal{L}_n$ is a faithful representation of SU(2, 2).

b. For $n \in 4\mathbb{Z}$, (3.12)

$\mathcal{L}_n$ induces a faithful rep. of $\mathrm{SO}_0(4, 2)$.

c. For $n \in 4\mathbb{Z} + 2$,

$\mathcal{L}_n$ induces a faithful rep. of $\mathrm{C}(M^4)$.

4. EXTENSION OF THE $m = 0$ WIGNER REPRESENTATIONS TO SU(2,2)

Following the lines of Mack and Todorov,[5] we obtain the realization of $\mathcal{L}_n$ as a continuation of the irreducible unitary representation u_λ of the inhomogeneous SL(2, $\mathbb{C}$), belonging to the mass $m = 0$ and the helicity $\lambda = -n/2$. The integral transforms representing special conformal transformations are calculated.

Notations, definitions

$$\Omega := \{(p_0, \mathbf{p}) \in \mathbb{R}^4 \mid p_0 > 0,\ p \cdot p = 0\} \tag{4.1}$$

is the zero mass shell (forward light cone), an orbit of SL(2, $\mathbb{C}$) [Eq. (2.4)].

Definition: E': $\mathbb{C}^2 \to \mathbb{C}(2\times 2)$: $w \mapsto 2ww^+ = 2\begin{pmatrix} w_1\bar{w}_1 & w_1\bar{w}_2 \\ w_2\bar{w}_1 & w_2\bar{w}_2 \end{pmatrix}$. (4.2)

E' maps $\mathbb{C}^2 \setminus \{0\}$ onto the set of rank-one Hermitian 2×2 matrices with positive trace, that is to say, $\underline{E' w_\mu}$

$= w^+\sigma_\mu w$ is a light vector $\in \Omega$. By Eq. (2.4), $A \in \mathrm{SL}(2,\mathbb{C})$ $\Rightarrow \underline{E'(Aw)} = \Lambda(A)\underline{E'w}$.

E' induces a bijection of the set of orbits $\exp(i\mathbb{R})\cdot w$ ($w \in \mathbb{C}^2$, $w \neq 0$) of the group $\mathrm{U}(1) = \{\exp(it)\cdot l_2 \mid t \in \mathbb{R}\}$, onto Ω, since $E'w_1 = E'w_2 \Leftrightarrow w_2 = \exp(i\beta)w_1$ ($0 \leq \beta < 2\pi$). In the ladder representation, $A \in \mathrm{SL}(2,\mathbb{C})$ acts on $\mathbb{C}^2$ as follows: $z \mapsto (A^T)^{-1}z$. By Eqs. (2.5)–(2.6),

$$A \in \mathrm{GL}(2,\mathbb{C}) \Rightarrow (\det A)(A^T)^{-1} = S^{-1}AS$$
$$\text{with unitary } S = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}. \tag{4.3}$$

For this reason we define a modification of E':

Definition: $Ez = E'(Sz)$ $(z \in \mathbb{C}^2)$,

$$Ez = \widehat{[E'z]}^T = 2\widehat{\bar{z}\,z^T}, \quad \underline{Ez} = (z^T\sigma_0\bar{z}, -z^T\boldsymbol{\sigma}\bar{z}),$$
$$A \in \mathrm{GL}(2,\mathbb{C}) \Rightarrow E((A^T)^{-1}z) = |\det A|^{-2}A(Ez)A^+ \tag{4.4}$$

We introduce parameters for Ω and $\mathbb{C}^2$ in the following manner: We select a fixed point $k \in \Omega$ and a family of transformations

$$\{f_p \in \mathrm{SL}(2,\mathbb{C}) \mid p \in \Omega,\ \Lambda f_p(k) = p\}. \tag{4.5}$$

Explicit choices:

$$k = (\tfrac{1}{2}, 0, 0, \tfrac{1}{2}),$$
$$\underset{\sim}{k} = \tfrac{1}{2}(\sigma_0 + \sigma_3) = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} = E'w_k \text{ with } w_k := \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 0 \end{pmatrix}, \tag{4.6}$$

$$f_p = d_{\hat{p}}\cdot b_{p_0} \quad \text{if } p = (p_0, \mathbf{p}) = p_0\cdot(1,\hat{p}) \quad (|\hat{p}| = 1).$$

b_{p_0} is a boost which carries k into $(p_0, 0, 0, p_0)$:

$$b_{p_0} = \mathrm{diag}((2p_0)^{1/2},\ (2p_0)^{-1/2}) \in \mathrm{SL}(2,\mathbb{C}). \tag{4.7a}$$

$d_{\hat{p}}$ is a rotation [$\in \mathrm{SU}(2)$] carrying the 3 axis into the direction $\hat{p}$. If $0 < \vartheta < \pi$, $0 \leq \varphi < 2\pi$ are the polar coordinates of $\hat{p}$ [$\leftrightarrow (1, \vartheta, \varphi)$], two possible choices for $d_{\hat{p}}$ can be given:

$$d_1(\hat{p}) = \exp(-\tfrac{1}{2}i\sigma_3\cdot\varphi)\exp(-\tfrac{1}{2}i\sigma_2\cdot\vartheta),$$
$$d_2(\hat{p}) = \exp(-\tfrac{1}{2}i\sigma_3\cdot\varphi)\exp(-\tfrac{1}{2}i\sigma_2\cdot\vartheta)\exp[\tfrac{1}{2}i\sigma_3(-\varphi)]. \tag{4.7b}$$

We define

$$w_p := f_p w_k, \quad E'w_p = \underset{\sim}{p},$$
$$z_p := S^{-1}w_p \quad Ez_p = \underset{\sim}{p}. \tag{4.8}$$

$\{w_p\}$ is a representative family of the space of orbits $\mathbb{C}^2_0/\mathrm{U}(1)$. With an additional parameter $0 \leq \beta < 2\pi$, we have a coordinate system of (almost) the whole of $\mathbb{C}^2$:

$$w = \exp(i\beta)w_p \leftrightarrow (\beta, p) \in [0, 2\pi)\times\Omega, \quad \underline{E'w} = p$$
$$z = S^{-1}w = \exp(i\beta)z_p, \quad \underline{Ez} = p. \tag{4.9}$$

Explicitly, if $d_1(\hat{p})$ is chosen

$$\begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = \exp(i\beta)\,p_0^{1/2}\cdot\begin{pmatrix} \exp(i\varphi/2) & \sin(\vartheta/2) \\ -\exp(-i\varphi/2) & \cos(\vartheta/2) \end{pmatrix}.$$

Finally, we recall[13] the Wigner mass-zero representations u_λ of the inhomogeneous $\mathrm{SL}(2,\mathbb{C})$ [$\mathrm{ISL}(2,\mathbb{C}) = T \circledS L$, see (2.14), (2.17)], belonging to helicity $\lambda \in \mathbb{Z}/2$: The representation space is $L^2(\Omega; d^3\mathbf{p}/2p_0)$, and, up to equivalence, u_λ is described as follows [$F \in L^2(\Omega)$]:

$$\text{Translation } t(B)\text{: } u_\lambda t(B)F(p) = \exp(i\underline{Bp})F(p),$$
$$A \in \mathrm{SL}(2,\mathbb{C})\text{:} \qquad u_\lambda l(A)F(p) = [Q(A,p)]^{2\lambda}F(\Lambda(A)^{-1}p). \tag{4.10}$$

Here, Q has values in the complex unit circle, satisfies the relations

$$Q(AA', p) = Q(A,p)Q(A', \Lambda(A)^{-1}p), \tag{4.11}$$

and can be implemented by a unitary representation Δ of a little group:

$$k \in \Omega\text{: } G_k = \{A \in \mathrm{SL}(2,\mathbb{C}) \mid \Lambda(A)k = k\} \text{ (little group of } k) \tag{4.12}$$

$Q(A,p) = \Delta(f_p^{-1}Af_{\Lambda(A)^{-1}p})$ with a family $\{f_p\}$, Eq. (4.5).

Explicitly, with the choices (4.6) of k and f_p, we get

$$G_k = \left\{ g_{z\varphi} = \begin{pmatrix} 1 & z \\ 0 & 1 \end{pmatrix}\cdot\exp(-\tfrac{1}{2}i\sigma_3\cdot\varphi) \,\middle|\, z \in \mathbb{C},\ 0 \leq \varphi < 4\pi \right\},$$
$$\Delta(g_{z\varphi}) = \exp(-i\varphi/2). \tag{4.13}$$

Remark: If $\mathrm{ISL}(2,\mathbb{C})$ is enlarged by the dilation group D [see Eq. (2.14)], then the most general extension of (4.10) is[12]

$$u_{\lambda\delta}\,d(t)F(p) = \exp[2t(1+i\delta)]\,F(\exp(2t)p) \quad (\delta \in \mathbb{R}) \tag{4.14}$$

The new continuous parameter δ arises from the extension of the representation Δ of the little group.

Reduction of $\mathcal{L}_n$ to ISL(2,ℂ) (→ POINCARÉ GROUP)

With the parameters (4.9), the integral in $\mathbb{C}^2$ can be written

$$\int dz_1 dz_2 = \tfrac{1}{4}\int_0^{2\pi} d\beta \int \frac{d^3\mathbf{p}}{2p_0} \qquad (z = \exp(i\beta)\cdot z_p). \tag{4.15}$$

This leads to the following characterization of the spaces H_n [definition (3.10)]:

$$f \in H_n \Leftrightarrow \exists\, F \in L^2(\Omega;\, d^3\mathbf{p}/2p_0)\text{: } f(z) = F(p)\exp(in\beta)$$
$$[F(p) = f(z_p)]. \tag{4.16}$$

(4.16) establishes an isomorphism $f \mapsto F$: $H_n \to L^2(\Omega)$. In particular,

$$f \in H_n,\ g \in H_m \Rightarrow \int \bar{f}\,g\,dz = (\pi/2)\int \bar{F}\,G(d^3\mathbf{p}/2p_0)\cdot\delta_{nm}. \tag{4.17}$$

We will now calculate the ladder representations $\mathcal{L}_n$ of the subgroups T, L, and D, acting as unitary operators on $L^2(\Omega)$.

a. The group T of translations: $\mathcal{L}_n t(B)f(z) = \exp(i\,z^TB\,\bar{z})$ $(f \in H_n)$,

$$z = \exp(i\beta)z_p \Rightarrow z^TB\bar{z} = \mathrm{tr}(B\bar{z}z^T) = \tfrac{1}{2}\mathrm{tr}(B\,\widehat{Ez}) = \underline{B}\cdot p,$$
$$\text{with (4.16)} \Rightarrow \mathcal{L}_n\, t(B)F(p) = \exp(i\,\underline{Bp})F(p). \tag{4.18}$$

b. The group L [$\cong \mathrm{SL}(2,\mathbb{C})$]:

$$\angle_n\, l(A) f(z) = f(A^T z) \quad [A \in \mathrm{SL}(2,\mathbb{C})]$$

$$\Rightarrow \angle_n\, l(A) F(p) = \angle_n\, l(A) f(z_p) = f(A^T z_p).$$

Because of (4.4) and (4.9), there is a phase factor $\tilde{Q}(A,p)$ ($|\tilde{Q}| = 1$) such that

$$A^T z_p = [\tilde{Q}(A,p)]^{-1}\, z_{\Lambda(A)^{-1}p} \tag{4.19}$$

$$\Rightarrow \angle_n\, l(A) F(p) = [\tilde{Q}(A,p)]^{-n}\, F(\Lambda(A)^{-1}p). \tag{4.20}$$

c. The group D of dilations:

$$\angle_n\, d(t) f(z) = \exp(2t) f(e^t z) \text{ by (2.14), (3.2)}$$

$$\Rightarrow \angle_n\, d(t) F(p) = \exp(2t) F(e^{2t} p). \tag{4.21}$$

We shall now prove: (4.18) and (4.20) are the same as the representation (4.10) of ISL$(2,\mathbb{C})$, with $n = -2\lambda$.

We only have to show $Q = \tilde{Q}$. For convenience, we shall switch to the family w_p (4.8). Then (4.19) reads [see definition (4.3)]:

$$A^{-1} w_p = \tilde{Q}(A,p)^{-1} w_{\Lambda(A)^{-1}p}$$

$$\Leftrightarrow \tilde{Q}(A,p) w_k = (f_p^{-1} A f_{\Lambda(A)^{-1}p}) w_k, \quad \text{since } w_p = f_p w_k.$$

Therefore, $\tilde{Q}(A,p)$ is just the phase produced by the little group G_k [(4.13)] when acting on the spinor $w_k = (1/\sqrt{2})\binom{1}{0}$. It is easily seen that it coincides with Δ, defined in (4.13). Thus $Q = \tilde{Q}$, by Eq. (4.12). ∎

The proof is independent of the special choice (4.7a–b) of f_p; another set $\{f_p\}$ simply redefines the representative family z_p and the isomorphism $f \mapsto F$. In Ref. 5, (4.7a–b) with $d_1(\hat{p})$ is chosen (implicitly).

Result: The ladder representation $\angle_n$ of SU(2,2) is an extension of the (already irreducible) mass-zero representation with helicity $\lambda = -n/2$ of ISL$(2,\mathbb{C})$. We have [Eq. (3.12)] a representation of $C(M^4)$ if $\lambda = \pm 1, \pm 3, \cdots$, of $\mathrm{SO}_0(4,2)$ if $\lambda = 0, \pm 2, \pm 4, \cdots$, or only of SU(2,2) if λ is half-integer. This classification was discussed by Castell.[14]

As to dilations, we note that $\angle_n$ (4.21) corresponds to the choice $\delta = 0$ in (4.14), that is, the trivial extension of the little group representation. This is in fact the only choice compatible with the representation of the special conformal generators.[15]

Special conformal transformations on $L^2(\Omega)$

We shall first deal with the representation of the inversion Z_1 [Eq. (2.26)], which is essentially the Fourier transform on $L^2(\mathbb{C}^2)$. By (3.8) we find

$$\angle\, Z_1 f(w) = (1/2\pi)^2\, 4 \int dz \exp[2i \mathrm{Re}(z^T \bar{w})] f(z). \tag{4.22}$$

The real bilinear form $\mathrm{Re}(z^T B \bar{w})$ [$z, w \in \mathbb{C}^2$, $B \in H(2\times 2)$] will be evaluated in the coordinates (4.9): (4.23)

$$z = z_p \exp(i\alpha), \quad z_p{}^T B \bar{z}_q =: |z_p{}^T B \bar{z}_q| \cdot \exp[-i\varphi(B,p,q)],$$

$$w = z_q \exp(i\beta), \quad |z_p{}^T B \bar{z}_q|^2 = \tfrac{1}{4}\,\mathrm{tr}(\underset{\sim}{p} \hat{B} \underset{\sim}{q} \hat{B})$$

[derived with (2.6), (4.8)]

$$\Rightarrow \mathrm{Re}(z^T B \bar{w}) = \tfrac{1}{2}[\mathrm{tr}(\underset{\sim}{p} \hat{B} \underset{\sim}{q} \hat{B})]^{1/2} \cos[\alpha - \beta - \varphi(B,p,q)]. \tag{4.24}$$

We can now calculate $\angle_n Z_1$ on $L^2(\Omega)$: Let $f \in H_n$, i.e., $f(z) = F(p)\exp(in\alpha)$. Then

$$\angle_n Z_1 F(q) = \angle_n Z_1 f(z_q)$$

$$= \left(\frac{1}{2\pi}\right)^2 \int F(p) \frac{d^3\mathbf{p}}{2p_0} \int_0^{2\pi} d\alpha \exp(in\alpha)$$

$$\times \exp[2i\, \mathrm{Re}(z^T \bar{z}_q)].$$

By Eq. (4.24) with $B = 2$: $2\,\mathrm{Re}(z^T \bar{z}_q) = r_{pq} \cos(\alpha - \varphi_{pq})$, where $r_{pq} := [2(p_0 q_0 + \mathbf{pq})]^{1/2}$; and φ_{pq} is the polar angle of $z_p{}^+ z_q$. We use the following relations for Bessel functions[16]:

$$\frac{1}{2\pi} \int_0^{2\pi} d\alpha \exp\{i[n\alpha + r\cos(\alpha - \varphi)]\} = i^n \exp(in\varphi) J_n(r) \tag{4.25}$$

($r > 0$, $\varphi \in \mathbb{R}$, $n \in \mathbb{Z}$)

$$\Rightarrow \angle_n Z_1 F(q) = \frac{1}{2\pi}\, i^n \int \frac{d^3\mathbf{p}}{2p_0} F(p)$$

$$\times J_n \sqrt{2(p_0 q_0 + \mathbf{pq})}\; (\exp(i\varphi_{pq}))^n. \tag{4.26}$$

The factor $\exp(i\varphi_{pq})$ depends on the choice of the f_p in (4.5). If we take (4.7a–b) with $d_2(\hat{p})$ (contrary to Mack–Todorov), it can be expressed in terms of the elements of the Hermitian matrices

$$\underset{\sim}{p} = \begin{pmatrix} u & v \\ s & t \end{pmatrix}, \quad \underset{\sim}{q} = \begin{pmatrix} u' & v' \\ s' & t' \end{pmatrix}: \quad \exp(i\varphi_{pq}) = \frac{uu' + v\bar{v}'}{|uu' + v\bar{v}'|}. \tag{4.27}$$

The mass-zero integral transforms given by Castell in Ref. 17 are exactly the same as (4.26)–(4.27), with $n = -2\lambda$. The integral kernel R discussed by Kastrup and Mayer[3,4] is obtained in the special case $\lambda = 0$, if $\angle_0 Z_1$ is combined with the space inversion $(x_0, \mathbf{x}) \mapsto (x_0, -\mathbf{x})$ [in $L^2(\Omega)$: $F(p_0, \mathbf{p}) \mapsto F(p_0, -\mathbf{p})$]. R represents the inversion $x \mapsto -x/x \cdot x$ in M^4, which is not an element of the connected group $C(M^4)$.

Second, we shall treat the group S acting on $L^2(\Omega)$, but only those special conformal transformations $s(B)$ which have a regular matrix B (that is, $\underset{\sim}{B} \cdot \underset{\sim}{B} \neq 0$), because of formula (3.7). Our task consists in restricting (3.7) to H_n and calculating the corresponding integral transform in $L^2(\Omega)$. The calculation of $\angle_n s(B)$ is very similar to that of $\angle_n Z_1$, and so we omit the details:

Since B and B^{-1} are Hermitian, we get from (3.7)

$$L_B(z - w) = -4(\det B)^{-1}$$

$$\times \exp\{i[z^T B^{-1} \bar{z} + w^T B^{-1} \bar{w} - 2\,\mathrm{Re}(z^T B^{-1} \bar{w})]\}.$$

With the use of the parameters below (4.23) for z and w and of Eqs. (4.24)–(4.25) for the expression $-2\,\mathrm{Re}(z^T B^{-1} \bar{w})$, the following result for $\angle_n s(B)$ is obtained:

$$\angle_n s(B) F(q) = \int (d^3\mathbf{p}/2p_0) F(p)\, L_n^B(p,q),$$

$$L_n^B(p,q)=\frac{1}{2\pi}\cdot i^n\exp\left[i\left(\frac{(p+q)\cdot\underline{B}}{\underline{B}\cdot\underline{B}}\right)\right]\times J_n\left(\frac{\sqrt{\mathrm{tr}(\underline{p}B\underline{q}B)}}{\underline{B}\cdot\underline{B}}\right)(\exp(i\alpha))^{-n}. \tag{4.28}$$

$\exp(i\alpha)$ is the phase of $-z_p{}^T B^{-1}\bar{z}_q$, depending again on the choice of $\{f_p\}$. We recall that $s(B)$ acts on M_c^4 as the conformal transformation

$$x\mapsto[x+(x\cdot x)\hat{\underline{B}}]/\omega(\hat{\underline{B}},x).$$

Thus, for all "nonlightlike" special conformal transformations, the series of ladder representations $\mathcal{L}_n$ is given by the integral kernels (4.28) in momentum space.

5. CONFORMAL TRANSFORMATIONS OF MASSLESS FREE FIELDS

The classical and quantized free fields on Minkowski space which correspond to the mass zero representations of ISL(2, $\mathbb{C}$) are considered. Their transformation under the special conformal group S is investigated.

Conformal transformations of classical fields

The conformal transformation properties of the irreducible Minkowski space spinor fields will be reviewed first.[18]

We define the space $\mathcal{S}_{rs}$ of classical spinor fields as the set of all (generalized) functions ϕ on M^4 with values $\phi(x)$ in the $(2s+1)(2r+1)$-dimensional representation space of the irreducible representation D_{rs} of SL(2, $\mathbb{C}$). With respect to our Λ [Eq. (2.4)], D_{rs} is defined as

$$D_{rs}:=D_r\otimes\bar{D}_s,\quad r,s=0,\tfrac{1}{2},1,\tfrac{3}{2},\cdots,$$

$$\bar{D}_s(A)=D_s((A^+)^{-1}),\quad A\in\mathrm{SL}(2,\mathbb{C}),$$

$$D_r(\tfrac{1}{2}\sigma_k)=-iD_r(-\tfrac{1}{2}i\sigma_k),\quad \tfrac{1}{2}\sigma_k\in\mathrm{sl}(2,\mathbb{C}),\ -\tfrac{1}{2}i\sigma_k\in\mathrm{su}(2).$$

The last equation says that D_r of the Lie algebra sl(2, $\mathbb{C}$) is obtained by antianalytic extension from the well-known representations D_r of su(2).

The usual Poincaré transformation behavior of spinor fields is given by the following (highly reducible) representation $\mathcal{R}_{rs}$ of ISL(2, $\mathbb{C}$) on $\mathcal{S}_{rs}$:

$$\mathcal{R}_{rs}(t(B)l(A))\phi(x)=D_{rs}(A)\,[\phi((A)^{-1}x-\underline{B})] \tag{5.1}$$

$\mathcal{R}_{rs}$ is defined by "inducing" from the representation D_{rs} of the little group SL(2, $\mathbb{C}$) of $O\in M^4$. We want to extend $\mathcal{R}_{rs}$ to the conformal group.[18] We start from the little group of O with respect to SU(2, 2), which is the group K of Eqs. (2.21)–(2.23). Which are the possible extensions of D_{rs} from $\mathrm{SL}(2,\mathbb{C})\tilde{=}L$ to K? The group law forces us to represent $S\subset K$ by 1, and the dilation $d(t)$ by a complex number $\exp(2zt)$ ($z\in\mathbb{C}$), see Ref. 18. The central element i can be represented by i^m, where $m=0,1,2,3$ is restricted because of $D_{rs}(-1)=(-1)^{2(r+s)}$:

$$r+s\quad\text{integer}\Rightarrow m=0,2,$$

$$r+s\ \text{half-integer}\Rightarrow m=1,3.$$

Thus, we have the following class of extensions of D_{rs} [with Eq. (2.18)]:

$$D_{rs\,\epsilon m}(s(B)l(A')d(t)\epsilon)=D_{rs}(A')\exp(2zt)\epsilon^m. \tag{5.2}$$

By the use of translations, we induce the corresponding extensions of $\mathcal{R}_{rs}$, that is, we define for $G\subset\mathrm{SU}(2,2)$ (see Sec. 2):

$$(\mathcal{R}_{rs\epsilon m}G)\phi(x)=D_{rs\epsilon m}(t(-\underline{x})Gt((\complement G)^{-1}\underline{x}))\,\phi((\complement G)^{-1}x)). \tag{5.3}$$

This formula is not defined if $(\complement G)^{-1}x\notin M^4$.

For special conformal transformations $s(B)$, (5.3) becomes after decomposing the little group element $t(-\underline{x})s(B)t((1-B\underline{x})^{-1})$ as in Eq. (5.2):

$$\mathcal{R}_{rs\epsilon m}s(B)\phi(x)=|\omega|^z\epsilon^m D_{rs}\left(\frac{1-\underline{x}B}{|\omega|^{1/2}}\right)\phi\left(\frac{x-(x\cdot x)\hat{\underline{B}}}{\omega}\right),$$

$$\omega=\omega(-\hat{\underline{B}},x)=\det(1-\underline{x}B)\quad[\text{see Eq. (2.20)}],$$

$$\epsilon=\begin{cases}1:\omega>0\\ i:\omega<0\end{cases}\equiv(\mathrm{sgn}\omega)^{1/2}. \tag{5.4}$$

The massless classical fields

On the one hand, we have the ladder representations $\mathcal{L}_n$ of the conformal group SU(2, 2), on the other hand the conformal transformation properties (5.2)–(5.4) of spinor fields. Can we single out spaces of conformal invariant free fields in $\mathcal{S}_{rs}$, i.e., can $\mathcal{L}_n$ be imbedded into some of the representations $\mathcal{R}_{rs\epsilon m}$?

First, we recall how Weinberg[6] solved the analogous problem for the Poincaré group ISL(2, $\mathbb{C}$): Are there subspaces of $\mathcal{S}_{rs}$ which obey "free massless field equations", or, more specifically, which belong to the irreducible representations u_λ of Eqs. (4.10)? Is u_λ "contained" in $\mathcal{R}_{rs}$ for given values (λ,r,s); i.e., does there exist an injection $L^2(\Omega)\to\mathcal{S}_{rs}$ such that $F\mapsto\phi\Rightarrow u_\lambda F\mapsto\mathcal{R}_{rs}\phi$?

We try an integral kernel ansatz for the linear map $F\mapsto\phi$. The translation behavior of Eqs. (4.10), (5.1) restricts it to the conventional form

$$\phi(x)=\left(\frac{1}{2\pi}\right)^{3/2}\int\frac{d^3\mathbf{p}}{2p_0}\exp(-ip\cdot x)u(p)F(p). \tag{5.5}$$

The $(2r+1)(2s+1)$-component kernel function u must obey the following law, which is equivalent to the correct SL(2, $\mathbb{C}$) behavior of $F\mapsto\phi$:

$$u(p)=D_{rs}(A)[\bar{u}_\lambda l(A)u(p)]\quad[A\in\mathrm{SL}(2,\mathbb{C})]. \tag{5.6}$$

The complex conjugate representation $\bar{u}_\lambda$ of u_λ is defined by conjugating the unitary multiplicators in (4.10). $\bar{u}_\lambda$ is understood to act on all components of $u(p)$. (5.6) follows easily from the unitarity of u_λ and the definition of the scalar product in $L^2(\Omega)$.

Since u_λ was induced from a little group G_k with a family f_p, the function u is completely determined by its value at the point k:

$$u(p)=D_{rs}(f_p)u(k) \tag{5.7}$$

and the problem (5.6) is reduced to an eigenvalue problem for G_k:

$$u(k)=D_{rs}(g_k)(\bar{\Delta}g_k)^{2\lambda}u(k)\quad(g_k\in G_k). \tag{5.8}$$

The solution $u(k)$ exists and is unique up to normalization, *iff* $s-r=\lambda$. This was shown by Weinberg; we denote

the resulting function (5.7) by $u_{rs}(p)$. With the usual basis of eigenvectors of $D_{rs}(\frac{1}{2}\sigma_3)$ for the representation space of D_{rs}: $\{e_m \otimes e_n \mid -r \leq m \leq r;\ -s \leq n \leq s\}$ and with $k = (\frac{1}{2}, 0, 0, \frac{1}{2})$, we obtain $u_{rs}(k) = e_{-r} \otimes e_s$. Using the family f_p of Eqs. (4.7) with $d_1(\hat{p})$, we get in the above basis

$$(u_{rs})_{mn}(p) = (2p_0)^{r+s} \exp[-i\varphi(m+n)]\, d^{(r)}_{m,-r}(\vartheta)\, d^{(s)}_{n,s}(\vartheta) \qquad [\mathbf{p} \hat{=} (p_0, \vartheta, \varphi)]. \tag{5.9}$$

Result: The mass-zero representation u_λ is contained (only) in the spaces of fields $\mathcal{S}_{rs}$ with $s - r = \lambda$ [Eqs. (5.5), (5.9)]

We treat now the same problem for SU(2, 2): Can the ladder representation $\mathcal{L}_n$ be imbedded into a given field representation $\mathcal{R}_{rs\varepsilon m}$? As discussed above, the restriction to ISL(2, $\mathbb{C}_-$) alone fixes the mapping $F \mapsto \phi$ and yields the condition $n = -2\lambda = 2(r-s)$:

$$\phi^{(+)}(x) = \left(\frac{1}{2\pi}\right)^{3/2} \int \frac{d^3\mathbf{p}}{2p_0}\, u_{rs}(p) \exp(-ip\cdot x) F(p). \tag{5.10}$$

There are further constraints derived from other subgroups of SU(2, 2): The group D of dilations fixes the number z in (5.2). z must be chosen so that the analog of (5.8) for the extended little group is still valid; we state the result[12]: $z = -(1 + r + s)$.

The central element i fixes the number m: From (5.2) and (3.11) we get $m = n + 2 = 2 - 2\lambda = 2(1 + r - s)$. Thus, the set of conditions is

$$\begin{aligned} &n = 2(r-s) \qquad (n = -2\lambda), \\ &z = -(1 + r + s), \\ &m = 2(1 + r - s), \end{aligned} \tag{5.11}$$

The previous discussion suffices to determine the allowed $\mathcal{R}_{rs\varepsilon m} =: \mathcal{R}_{rs}$.

We shall now prove that $\mathcal{L}_n$ is indeed contained in the representations $\mathcal{R}_{rs}$ [Eq. (5.12) below] if $n = 2(r-s)$. Since this property is satisfied by construction of (5.12) for the subgroups ISL(2, $\mathbb{C}$), D, and the center, and since these generate SU(2, 2) together with the group S, it is sufficient to give the proof for the special conformal transformations. The infinitesimal conformal transformations belonging to (5.12) are given in Ref. 18, and for these generators it is shown in Ref. 5 (in a quantized version) that they are obtained by the transformation (5.10) from the special conformal generators of the ladder representation $\mathcal{L}_n$. Since the group S is Abelian ($\tilde{=} \mathbb{R}^4$), the compatibility is also satisfied for the finite group elements, and the proof is finished.

Result: The map (5.10): $L^2(\Omega) \to \mathcal{S}_{rs}$ imbeds the ladder representation $\mathcal{L}_n$ of SU(2, 2) into all the Minkowski space spinor field representations $\mathcal{R}_{rs}$ with $2(r-s) = n$, extended to SU(2, 2) by the laws

$$\begin{aligned} &\mathcal{R}_{rs}\, t(\underline{b})\phi(x) = \phi(x-b) \\ &\mathcal{R}_{rs}\, l(A)\, \phi(x) = D_{rs}\, \phi(\Lambda(A)^{-1}x) \quad [A \in \mathrm{SL}(2, \mathbb{C})], \\ &\mathcal{R}_{rs}\, d(t)\, \phi(x) = \exp[2t(-1-r-s)] \\ &\qquad\qquad \phi(\exp(-2t)x) \text{ (dilations)}, \end{aligned} \tag{5.12}$$

$$\mathcal{R}_{rs}\, s(B)\phi(x) = |\omega|^{-1-r-s}\ (\mathrm{sgn}\omega)^{1+r-s} \cdot D_{rs}\left(\frac{1 - \underline{x}B}{|\omega|^{1/2}}\right) \times \phi\left(\frac{x - (x\cdot x)\hat{\underline{B}}}{\omega}\right)$$

[special conformal tr, $\omega = \omega(-\underline{\hat{B}}, x)$, $(-1)^{1/2} \equiv i$].

The transformations (5.12) thus leave invariant the momentum spectrum Ω. Though the conformal transformations act as integral transforms in momentum space, they are transformations of the local "differential geometric" type in Minkowski space.

The fields defined by (5.10) belong to positive frequencies only. But it is also possible to imbed negative frequency parts into fields transforming with (5.12): We define the series of starred ladder representations of U(2, 2): $\mathcal{L}^*_n = \overline{\mathcal{L}_{-n}}$. The bar means conjugation of the factors and integral kernels in (4.18), (4.20), (4.21), (4.26), (4.28). Calculation gives

$$\begin{aligned} &\mathcal{L}^*_n\, l(A) = \mathcal{L}_n\, l(A) \qquad [A \in \mathrm{GL}(2, \mathbb{C}),\ \det A \in \mathbb{R}], \\ &\mathcal{L}^*_n\, t(B) = \mathcal{L}_n\, t(-B), \\ &\mathcal{L}^*_n\, s(B) = \mathcal{L}_n\, s(-B). \end{aligned} \tag{5.13}$$

With a function $G(p) \in L^2(\Omega)$ transforming with $\mathcal{L}^*_n$, we can define the negative frequency field

$$\phi^{(-)}(x) = \left(\frac{1}{2\pi}\right)^{3/2} \int \frac{d^3\mathbf{p}}{2p_0}\, u_{rs}(p) \exp(ip\cdot x) G(p), \tag{5.14}$$

which also transforms with $\mathcal{R}_{rs}$ like the positive frequency part (5.10), if the same condition $n = 2(r-s)$ is satisfied. This is easily seen for the Poincaré group and the dilations, and will be shown for the group S: In an abbreviated notation, with the $L^2(\Omega)$ scalar product, we have

$$\begin{aligned} &\phi^{(+)}(x) = \langle \bar{u}_{rs},\ \mathcal{L}_n t(-\underline{x})F\rangle, \qquad F \text{ transforming with } \mathcal{L}_n, \\ &\phi^{(-)}(x) = \langle \bar{u}_{rs}, \mathcal{L}_n t(\underline{x})G\rangle, \qquad G \text{ transforming with } \mathcal{L}^*_n, \end{aligned}$$

$$\begin{aligned} \mathcal{R}_{rs}\, s(B)\phi^{(+)}x &= \langle \bar{u}_{rs},\ \mathcal{L}_n[t(-\underline{x})s(B)]F\rangle \\ &= \langle \bar{u}_{rs},\ \mathcal{L}_n\{t(-\underline{x})s(B)t([\complement s(B)]^{-1}\underline{x})\} \\ &\quad \mathcal{L}_n t(-[\complement s(B)]^{-1}\underline{x})F\rangle \\ &\equiv D(B, x)\phi^{(+)}([\complement s(B)]^{-1}x) \quad [\text{see (5.12)}], \end{aligned}$$

$$\begin{aligned} \mathcal{R}^{(-)}_{rs}\, s(B)\phi^{(-)}x &:= \langle \bar{u}_{rs},\ \mathcal{L}_n[t(\underline{x})s(-B)]G\rangle \\ &= \langle \bar{u}_{rs},\ \mathcal{L}_n\{t(\underline{x})s(-B)t([\complement\, s(-B)]^{-1}(-\underline{x}))\} \\ &\quad \times \mathcal{L}_n t(-[\complement\, s(-B)]^{-1}(-\underline{x}))G\rangle \\ &= D(-B, -x)\, \phi^{(-)}([\complement\, s(B)]^{-1}x) = \mathcal{R}_{rs}\, s(B)\phi^{(-)}x, \end{aligned}$$

since $-[\complement\, s(-B)]^{-1}(-x) = [\complement\, s(B)]^{-1}x$, and $D(-B, -x) = D(B, x)$ [from (5.12)].

Result: The representation $\mathcal{L}_n$ can be imbedded into $\mathcal{R}_{rs}$ with positive frequencies, and $\mathcal{L}^*_n$ into $\mathcal{R}_{rs}$ with negative frequencies, if $n = 2(r-s)$, with the help of definitions (5.10) and (5.14).

Massless quantized fields

Finally, we consider quantized free massless fields of the type of Weinberg's $2(2r+1)$-component fields.[6]

We start with a set of creation and annihilation operators with canonical (anti)commutators:

$$\{a_\lambda^+(p);\ a_\lambda(p)\,|\,p\in\Omega;\ \lambda\in\mathbf{Z}/2\},$$
$$[a_\lambda(p),\ a^+_{\lambda'}(p')]_\pm=\delta_{\lambda\lambda'}(2p_0)\delta^3(\mathbf{p}-\mathbf{p}'). \tag{5.15}$$

In particular, we write for $F\in L^2(\Omega)$, transforming with $\mathcal{L}_n$ $(n=-2\lambda)$:

$$F=\int\frac{d^3\mathbf{p}}{2p_0}F(p)\,a_\lambda^+(p)\,|0\rangle=(\text{formally})\ \langle\overline{F},a_\lambda^*\rangle\,|0\rangle \tag{5.16}$$

with a conformal invariant vacuum $|0\rangle$. We denote by U the representation of SU(2, 2) in the Fock space belonging to the series of ladder representations, that is, generated by the $a_\lambda^+(p)$ and $a_\lambda(p)$ out of $|0\rangle$.

Then the following SU(2, 2) transformation behavior of the operators can be inferred from (5.16) ($\mathcal{L}_n$ is unitary):

$$U_g\,a_\lambda(p)\,U_g^{-1}=[\mathcal{L}_n(g^{-1})a_\lambda](p),\quad g\in\mathrm{SU}(2,2),$$
$$U_g\,a_\lambda^+(p)U_g^{-1}=[\overline{\mathcal{L}_n}(g^{-1})a_\lambda^+](p)=[\mathcal{L}^*_{-n}(g^{-1})a_\lambda^+](p). \tag{5.17}$$

For the special conformal group, these are nonlocal transformations. It is seen from the *Result* below Eq. (5.14) that we can construct two operator fields $[n=-2\lambda=2(r-s)]$:

$$\varphi^{(+)}(x)=\left(\frac{1}{2\pi}\right)^{3/2}\int\frac{d^3\mathbf{p}}{2p_0}\,u_{rs}(p)\exp(-ip\cdot x)\,a_\lambda(p)$$

(positive frequencies)

$$\chi^{(-)}(x)=\left(\frac{1}{2\pi}\right)^{3/2}\int\frac{d^3\mathbf{p}}{2p_0}\,u_{sr}(p)\exp(ip\cdot x)\,a_\lambda^+(p)$$

(negative frequencies)

with the transformation properties

$$U_g\,\varphi^{(+)}(x)\,U_g^{-1}=[\mathcal{R}_{rs}(g^{-1})\,\varphi^{(+)}](x),\quad g\in\mathrm{SU}(2,2),$$
$$U_g\,\chi^{(-)}(x)\,U_g^{-1}=[\mathcal{R}_{sr}(g^{-1})\chi^{(-)}](x).$$

In order to obtain fields with all frequencies, we must add to these two fields $\varphi^{(+)}$ and $\chi^{(-)}$ the fields of opposite helicity $-\lambda$:

$$\varphi^{(-)}(x)=\left(\frac{1}{2\pi}\right)^{3/2}\int\frac{d^3\mathbf{p}}{2p_0}\,u_{rs}(p)\exp(ip\cdot x)\,a^+_{-\lambda}(p)$$

(transforms with $\mathcal{R}_{rs}$),

$$\chi^{(+)}(x)=\left(\frac{1}{2\pi}\right)^{3/2}\int\frac{d^3\mathbf{p}}{2p_0}\,u_{sr}(p)\exp(-ip\cdot x)\,a_{-\lambda}(p)$$

(transforms with $\mathcal{R}_{sr}$).

Then the combined field

$$\begin{pmatrix}\varphi\\ \chi\end{pmatrix}=\begin{pmatrix}\xi_L\varphi^{(+)}+\eta_R\varphi^{(-)}\\ \xi_R\chi^{(+)}+\eta_L\chi^{(-)}\end{pmatrix},\quad \xi_{L,R},\eta_{L,R}\in\mathbb{C}$$

contains the helicities λ and $-\lambda$, both positive and negative frequencies, and transforms reducibly according to the sum $\mathcal{R}_{rs}\oplus\mathcal{R}_{sr}$ under SU(2, 2). In the case $\lambda>0$ and $(r,s)=(-\lambda,0)$ we obtain Weinberg's $2(2r+1)$-component fields. They are thus not only Poincaré- but also conformal-invariant.

The difficulties with differently transforming positive and negative frequency parts (Swieca—Völkel[19]) do not arise in our case of three (odd) space dimensions.

ACKNOWLEDGMENT

The author would like to thank Professor H. A. Kastrup for helpful discussions and suggestions.

[1] T. Go, H.A. Kastrup, and D. Mayer, Rep. Math. Phys. (to be published).
[2] L. Gross, J. Math. Phys. **5**, 687 (1964).
[3] H.A. Kastrup, Phys. Rev. **140**, B183 (1965).
[4] H.A. Kastrup and D.H. Mayer, J. Math. Phys. **11**, 1041 (1970).
[5] G. Mack and I.T. Todorov, J. Math. Phys. **10**, 2078 (1969).
[6] S. Weinberg, Phys. Rev. **134**, B882 (1964).
[7] S. Weinberg, Phys. Rev. **133**, B1318 (1964).
[8] W. Rühl, Commun. Math. Phys. **27**, 53 (1972).
[9] M. Hortaçsu, R. Seiler, and B. Schroer, Phys. Rev. D **5**, 2519 (1972).
[10] R. L. Anderson, J. Fischer, and R. Raczka, Proc. Roy. Soc. Lond. A **302**, 491 (1968).
[11] M. Schaaf, Rep. Math. Phys. **4**, 275 (1973).
[12] G. Post, Diplomarbeit TH Aachen, 1974.
[13] A.S. Wightman, "L'invariance dans la mécanique quantique relativiste" in *Dispersion Relations*, edited by DeWitt and Omnes (Wiley, New York, 1960).
[14] L. Castell, Nucl. Phys. B **13**, 231 (1969).
[15] A.H. Clark, Jr., thesis, Princeton University, 1968.
[16] N.J. Vilenkin, *Fonctions spéciales et théorie de la représentation des groupes* (Dunod, Paris, 1960), pp. 207 ff.
[17] L. Castell, J. Math. Phys. **11**, 2999 (1970).
[18] G. Mack and A. Salam, Ann. Phys. **53**, 174 (1969).
[19] J.A. Swieca and A.H. Völkel, Commun. Math. Phys. **29**, 319 (1973); B. Schroer and J.A. Swieca, Phys. Rev. D **10**, 480 (1973).

The generalized Wiener–Feynman path integrals

Amar Maheshwari

Department of Physics, Himachal Pradesh University, Simla-171001, India
(Received 12 June 1975)

The generalized Wiener–Feynman path integrals are defined by the primitive mappings of the canonical Gaussian measure on a Hilbert space of real square integrable functions. The expressions of the covariance of the pro (pseudo) measures are found to be form covariant. The measures known in the literature by names as Wiener–Feynman and Uhlenbeck–Ornstein appear as special cases of no particular remark in our general definition. The connection of the primitive mapping with the general class of linear Cameron–Martin transformations is established.

I. INTRODUCTION

A definition of Feynman path integrals as a image of the canonical Gaussian pseudomeasure on a Hilbert space of real square integrable functions defined on the time interval $T=[0, T]$ by means of primitive mapping has been given recently.[1,2] This definition leads to new techniques for computing path integrals. The primitive mappings given in Ref. 2 map the canonical pro (pseudo) measure to Wiener—Feynman type of pro (pseudo) measures on the vector spaces Φ_-, Φ_+ and Φ of continuous functions defined on the time interval $T=[0, T]$. It is well known that Gaussian measures other than Wiener—Feynman based on Markovian probability distributions can be introduced for applications to physical problems in quantum statistics[3] and quantum mechanics.[4] The aim of this paper is to define the most general Gaussian pro (pseudo) measures by appropriate primitive mappings of the canonical Gaussian pro (pseudo) measure on Hilbert space of real square integrable functions. The measures known in the literature by names as Wiener—Feynman, Uhlenbeck—Ornstein appear as special cases of no particular remark in our general definition. The remarkable result of this paper is that in the global definition of pro (pseudo) measures by the Fourier transform on the dual space all Gaussian measures are form covariant. Therefore the integration with the general Gaussian pro (pseudo) measures, which is needed for generating a WKB type of evaluation of Feynman path integrals,[4] can be done with the same ease as with the Wiener—Feynman measure using the new techniques.

II. NOTATIONS AND DEFINITIONS

For the sake of completeness we recall here from Ref. 2 some pertinent definitions which we shall need in this paper. The space Φ_- is the space of paths x defined on $(0, T]$ such that $x(t)\to 0$ when $t\to 0$; Φ_+ is the space of paths x defined on $[0, T)$ such that $x(t)\to 0$ when $t\to T$; and $\Phi\subset\Phi_-$ is the space of paths x defined on $(0, T)$ such that $x(t)\to 0$ when $t\to 0$ and when $t\to T$. The topology on vector spaces Φ_-, Φ_+, Φ is the norm topology induced by the uniform norm $\|x\|=\mathrm{Sup}\,|x(t)|$ for all t in the range of x. The integral with pseudomeasures ω_- of a function F on the space Φ_- is a complex number written symbolically

$$K=\int_{\Phi_-} F(x)\,d\omega_-(x). \tag{1}$$

Its generalization to integrals with pseudomeasures ω_+ and ω on Φ_+ and Φ related to ω_- on Φ_- is obvious. The pseudomeasures are defined by their Fourier transforms as functions on the dual space $/\!\!/$ of $\Phi_\mp$, which is the space of bounded measures μ on T:

$$\mathcal{F}\omega_\mp=\exp\left(-\frac{i}{2}W_\mp\right), \tag{2}$$

where variance $W_\mp$ is the quadratic form on $/\!\!/$ defined by

$$W_\mp=\int_T d\mu(r)\int_T d\mu(s)\,G_\mp(r,s) \tag{3}$$

in terms of the covariance functions $G_\mp(r, s)$. The canonical Gaussian pseudomeasure on a Hilbert space H of real square integrable functions on T is the Gaussian of variance

$$I(f)=(f,f)_H=\int_0^T f(t)f(t)\,dt. \tag{4}$$

We shall need the transformation properties of a Gaussian pseudomeasure under linear mappings. Let $\mathbf{X}$ and $\mathbf{Y}$ be two Hausdorff, topological vector spaces locally convex; let $\mathbf{X}'$ and $\mathbf{Y}'$ be their topological duals. Let P be a linear continuous mapping from $\mathbf{X}$ into $\mathbf{Y}$ and $\tilde{P}$ be the transposed mapping from $\mathbf{X}'$ to $\mathbf{Y}'$ defined by

$$\langle\tilde{P}y', x\rangle=\langle y', Px\rangle. \tag{5}$$

Let ω be a Gaussian pseudomeasure on $\mathbf{X}$ of variance W. The image of W under P is a Gaussian pseudomeasure ω_p on $\mathbf{Y}$, whose Fourier transform is

$$\mathcal{F}\omega_p=\exp\left(-\frac{i}{2}W_p\right)\ \text{with}\ W_p=W\circ\tilde{P}. \tag{6}$$

This transformation relation together with the equation

$$\int_{\mathbf{Y}} F(y)\,d\omega_p(y)=\int_{\mathbf{X}} F\circ P(x)\,d\omega(x) \tag{7}$$

form the basis of our formalism.

III. PRIMITIVE MAPPINGS AND THE GENERAL PSEUDOMEASURE

Let $D(\sigma)$ be the solution of the differential equation

$$\frac{d^2D(\sigma)}{d\sigma^2}+p(\sigma)D(\sigma)=0, \tag{8}$$

with the boundary conditions $D(T)=1$ and $D'(T)=0$. We denote by $\mathcal{D}(\sigma)$ the solution of the same differential equation with the boundary conditions $\mathcal{D}(0)=1$ and $\mathcal{D}'(0)=0$. Let P_-, P_+, and P be the primitive mappings defined by

$$P_-: /\!\!/\to\Phi_-\ \text{by}\ P_-f(t)=x(t)=D(t)\int_0^t\frac{f(\sigma)\,d\sigma}{D(\sigma)}, \tag{9}$$

$$P_+ : \mathcal{H} \to \Phi_+ \text{ by } P_+ f(t) = x(t) = \mathcal{D}(t) \int_t^T \frac{f(\sigma)\,d\sigma}{\mathcal{D}(\sigma)}, \tag{10}$$

$$P : \mathcal{H} \to \Phi \text{ by } Pf(t) = x(t),$$

$$x(t) = D(t) \int_0^t \frac{f(\sigma)\,d\sigma}{D(\sigma)} - D(t) \left(\int_0^t \frac{d\xi}{D^2(\xi)} \right) \times \left(\int_0^t \frac{d\xi}{D^2(\xi)} \right)^{-1} \int_0^T \frac{f(\sigma)\,d\sigma}{D(\sigma)}. \tag{11}$$

The transposed primitive mappings are defined from Eq. (5) as

$$(\tilde{P}_\mp \mu, f) = \langle \mu, P_\mp f \rangle \tag{12}$$

and are easily found to be

$$\tilde{P}_- \mu(\sigma) = \frac{1}{D(\sigma)} \int_T \theta(t-\sigma) D(t)\,d\mu(t), \tag{13}$$

$$\tilde{P}_+ \mu(\sigma) = \frac{1}{\mathcal{D}(\sigma)} \int_T \theta(\sigma - t)\ \mathcal{D}(t)\,d\mu(t), \tag{14}$$

$$\tilde{P}\mu(\sigma) = \frac{1}{D(\sigma)} \left[\int_T \theta(r-\sigma) D(r)\,d\mu(r) - \left(\int_0^T \frac{d\xi}{D^2(\xi)} \right)^{-1} \int_T d\mu(r) D(r) \int_0^r \frac{d\xi}{D^2(\xi)} \right]. \tag{15}$$

The transformation relation Eq. (6) together with the definition of canonical Gaussian pseudomeasure gives

$$W_\mp = I(\tilde{P}_\mp \mu(\sigma)) = \int_0^T d\sigma\ \tilde{P}_\mp \mu(\sigma) \cdot \tilde{P}_\mp \mu(\sigma). \tag{16}$$

Substituting Eqs. (13), (14), and (15) into Eq. (16), we find

$$W_\mp(\mu) = \int_T d\mu(r) \int_T d\mu(s)\, G_\mp(r, s),$$

$$G_-(r, s) = D(r) D(s) \left(\theta(s-r) \int_0^r \frac{d\sigma}{D^2(\sigma)} + \theta(r-s) \int_0^s \frac{d\sigma}{D^2(\sigma)} \right) \tag{17}$$

$$G_+(r, s) = \mathcal{D}(r)\, \mathcal{D}(s) \left(\theta(s-r) \int_s^T \frac{d\sigma}{\mathcal{D}^2(\sigma)} + \theta(r-s) \int_r^T \frac{d\sigma}{\mathcal{D}^2(\sigma)} \right), \tag{18}$$

$$G(r, s) = D(r) D(s) \left(\int_0^T \frac{d\sigma}{D^2(\sigma)} \right)^{-1} \left[\theta(s-r) \left(\int_0^r \frac{d\sigma}{D^2(\sigma)} \right) \times \left(\int_s^T \frac{d\sigma}{D^2(\sigma)} \right) + \theta(r-s) \left(\int_0^s \frac{d\sigma}{D^2(\sigma)} \right) \left(\int_r^T \frac{d\sigma}{D^2(\sigma)} \right) \right]. \tag{19}$$

To illustrate the computational steps leading to Eqs. (17), (18), and (19) from Eq. (16), we calculate the variance $W(\mu)$ on Φ:

$$W \circ \tilde{P}\mu = \int_0^T d\sigma \left\{ \frac{1}{D(\sigma)} \left[\int_T \theta(r-\sigma) D(r)\,d\mu(r) - \left(\int_0^T \frac{d\xi}{D^2(\xi)} \right)^{-1} \int_T d\mu(r) D(r) \int_0^r \frac{d\xi}{D^2(\xi)} \right] \right\}^2$$

$$= \int_T d\mu(r) D(r) \int_T d\mu(s) D(s) \left[\int_0^T \frac{\theta(r-\sigma)\theta(s-\sigma)\,d\sigma}{D^2(\sigma)} + \left(\int_0^r \frac{d\xi}{D^2(\xi)} \right) \left(\int_0^s \frac{d\xi}{D^2(\xi)} \right) \left(\int_0^T \frac{d\xi}{D^2(\xi)} \right)^{-1} - 2 \left(\int_0^T \frac{d\xi}{D^2(\xi)} \right)^{-1} \left(\int_0^s \frac{d\xi}{D^2(\xi)} \right) \left(\int_0^T \frac{d\sigma\,\theta(r-\sigma)}{D^2(\sigma)} \right) \right]$$

$$= \int_T d\mu(r) \int_T d\mu(s) D(r) D(s) \left(\int_0^T \frac{d\sigma}{D^2(\sigma)} \right)^{-1} \times \left[\theta(s-r) \left(\int_0^r \frac{d\sigma}{D^2(\sigma)} \right) \left(\int_s^T \frac{d\sigma}{D^2(\sigma)} \right) + \theta(r-s) \left(\int_0^s \frac{d\sigma}{D^2(\sigma)} \right) \left(\int_r^T \frac{d\sigma}{D^2(\sigma)} \right) \right].$$

From our general results we next calculate the particular cases of Wiener–Feynman and Uhlenbeck–Ornstein pro (pseudo) measures.

Wiener-Feynman

This measure is defined by $D(\sigma) = \mathcal{D}(\sigma) = 1$. The primitive mappings for this case are obtained by substituting $D(\sigma) = 1$ in Eqs. (9), (10), and (11):

$$P_- : \mathcal{H} \to \Phi_- \text{ by } P_- f(t) = x(t) = \int_0^t f(\sigma)\,d\sigma \tag{20a}$$

$$P_+ : \mathcal{H} \to \Phi_+ \text{ by } P_+ f(t) = x(t) = \int_t^T f(\sigma)\,d\sigma \tag{20b}$$

$$P : \mathcal{H} \to \Phi \text{ by } Pf(t) = x(t) = \int_0^t f(\sigma)\,d\sigma - \frac{t}{T} \int_0^T f(\sigma)\,d\sigma. \tag{20c}$$

The covariance $G_\mp(r, s)$ follows from Eqs. (17), (18), and (19) and are as follows:

$$G_-(r, s) = \theta(s-r) r + \theta(r-s) s, \tag{21a}$$

$$G_+(r, s) = \theta(s-r)(T-s) + \theta(r-s)(T-r), \tag{21b}$$

$$G(r, s) = \theta(s-r) \frac{r(T-s)}{T} + \theta(r-s) \frac{s(T-r)}{T}. \tag{21c}$$

The mappings, Eq. (20), and the covariances $G_\mp(r, s)$ are the same as those given in Ref. 2. The Wiener promeasure is equivalent to the probability distribution

$$P_{\mathrm{WF}}(x_{k+1} | x_k) = \frac{1}{[2\pi(t_{k+1} - t_k)]^{1/2}} \exp\left(- \frac{(x_{k+1} - x_k)^2}{2(t_{k+1} - t_k)} \right). \tag{22}$$

Uhlenbeck-Ornstein

This measure is defined by $p(t) = \Omega^2$ $D(\sigma) = \cos\Omega(T-\sigma)$ and $\mathcal{D}(\sigma) = \cos\Omega\sigma$. We find

$$P_- : \mathcal{H} \to \Phi_- \text{ by } P_- f(t) = x(t) = \cos\Omega(T-t) \int_0^t \frac{f(\sigma)\,d\sigma}{\cos\Omega(T-\sigma)}. \tag{23a}$$

$$P_+ : \mathcal{H} \to \Phi_+ \text{ by } P_+ f(t) = x(t) = \cos\Omega t \int_t^T \frac{f(\sigma)\,d\sigma}{\cos\Omega\sigma}, \tag{23b}$$

$$P: \mathcal{H} \to \Phi \quad \text{by } Pf(t) = x(t)$$

$$= \cos\Omega(T-t) \int_0^t \frac{f(\sigma)\,d\sigma}{\cos\Omega(T-\sigma)} - \frac{\sin\Omega t}{\sin\Omega T} \times \int_0^T \frac{f(\sigma)\,d\sigma}{\cos\Omega(T-\sigma)}\,. \tag{23c}$$

The covariances $G_{\mp}(r, s)$ in this case are given by the following expressions:

$$G_-(r,s) = \frac{1}{\Omega\cos\Omega T}\,[\theta(s-r)\sin\Omega r\,\cos\Omega(T-s) + \theta(r-s)\sin\Omega s\,\cos\Omega(T-r)], \tag{29a}$$

$$G_+(r,s) = \frac{1}{\Omega\cos\Omega T}\,[\theta(s-r)\cos\Omega r\,\sin\Omega(T-s) + \theta(r-s)\cos\Omega s\,\sin\Omega(T-r)], \tag{29b}$$

$$G(r,s) = \frac{1}{\Omega\sin\Omega T}\,[\theta(s-r)\sin\Omega r\,\sin\Omega(T-s) + \theta(r-s)\sin\Omega s\,\sin\Omega(T-r)]. \tag{29c}$$

The expression (23b) is the same as given in Refs. 2 and 3. The Uhlenbeck–Ornstein promeasure is equivalent to the distribution

$$P_{UO}(x_{k+1}|x_k) = \left(\frac{\Omega}{\pi\{1-\exp[-2\Omega(t_{k+1}-t_k)]\}}\right)^{1/2} \cdot \times \exp\left(-\frac{\Omega\{x_{k+1}-x_k\exp[-\Omega(t_{k+1}-t_k)]\}^2}{\{1-\exp[-2\Omega(t_{k+1}-t_k)]\}}\right). \tag{25}$$

IV. CONNECTION WITH THE CAMERON-MARTIN TRANSFORMATION

We decompose the primitive mapping P_-, defined in Eq. (9), as

$$P_- = P_{CM} \circ P_{-WF}, \tag{26}$$

where P_{-WF} is the primitive mapping, Eq. (20a), from the Hilbert space $\mathcal{H}$ to Φ_- corresponding to the Wiener–Feynman pro (pseudo) measure. This defines a linear mapping P_{CM} of Φ_- to Φ_- (Ref. 5);

$$P_{CM}x(t) = y(t) = x(t) + D(t)\int_0^t \frac{D'(\sigma)}{D^2(\sigma)}\,x(\sigma)\,d\sigma. \tag{27}$$

It has been established in Ref. 6 that P_{-WF} leads to a promeasure corresponding to Wiener integral and in Refs. 1, 2 that P_{-WF} leads to a pseudomeasure corresponding to the Feynman integral. That is

$$\int_{\Phi_-} F(x)\,d\omega_{-F} = \int_{\Phi_-} F(x)\,\exp\left(\frac{i}{2}\int_0^T \dot{x}^2(t)\,dt\right)\mathcal{D}(x) \tag{28}$$

where

$$\mathcal{F}\omega_{-F}(\mu) = \exp\left(-\frac{i}{2}\int_T d\mu(r)\int_T d\mu(s)\,\inf(r,s)\right).$$

If $F = f \circ P_{CM}$, we can write the left-hand side of Eq. (28) from the transformation relation given in Eq. (7) as a function space integral of f with the pseudomeasure of covariance $G_-(r, s)$ given in Eq. (17):

$$\int_{\Phi_-} F(x)\,d\omega_{-F}(x) = \int_{\Phi_-} f(x)\,d\omega_-(x). \tag{29}$$

The right-hand side of Eq. (28) can be written as a path integral with Feynman measure using the Cameron–Martin linear transformation of the Wiener–Feynman measure. This is given in the form required by Gelfand and Yaglom.[7] We find

$$\int_{\Phi_-} F(x)\exp\left(\frac{i}{2}\int_0^T \dot{x}^2(t)\,dt\right)\mathcal{D}(x) = [D(0)]^{1/2}\int_{\Phi_-} f(x)\exp\left[\frac{i}{2}\left(\int_0^T \dot{x}^2(t)\,dt + \int_0^T \frac{D''(\sigma)}{D(\sigma)}\,x^2(\sigma)\,d\sigma\right)\right]\mathcal{D}(x). \tag{30}$$

Equating the right-hand sides of Eqs. (29) and (30), and using Eq. (8), we get

$$\int_{\Phi_-} f(x)\,d\omega_-(x) = [D(0)]^{1/2}\int_{\Phi_-} f(x) \times \exp\left(\frac{i}{2}\int_0^T dt[\dot{x}^2(t) - p(t)x^2(t)]\right)\mathcal{D}(x). \tag{31}$$

We have thus established the connection between our definition of the pro (pseudo) measures and the corresponding Feynman type of integral which is to be calculated by the "time-slicing" procedure. The advantage of our definition of Feynman type of integrals with pseudomeasures and techniques is that integral of cylindrical functions can be readily calculated by mapping $P_u: \Phi_{\mp} \to \mathbb{R}^u$. It gives many new results, Refs. 2 and 8, including the diagram technique. It has been shown in Ref. 2 that the covariance $G(r, s)$ of the pseudomeasure on Φ is the Feynman Greens function. $G(r, s)$ given in Eq. (19) is the Greens function of the differential operator $d^2/dt^2 - p(t)$.

The method of primitive mappings for defining pseudomeasures on spaces Φ_-, Φ developed in this paper is particularly suited to the study of systems whose configuration space is a Riemann multiply connected manifold.[9] The results of Koval'chik[10] for various types of Wiener integrals can be easily obtained using our Eqs. (7) and (17).

ACKNOWLEDGMENT

This work is an outcome of a discussion I had with Professor Cecile DeWitt-Morette at Karpacz, Poland. I am thankful to her.

[1]C. DeWitt-Morette, Commun. Math. Phys. **28**, 47 (1972).
[2]C. DeWitt-Morette, Commun. Math. Phys. **37**, 63 (1974).
[3]A. Maheshwari, J. Stat. Phys. **12**, 11 (1975).
[4]M. C. Gutzwiller, J. Math. Phys. **12**, 343 (1971); V. P. Maslov, Theor. Math. Phys. **2**, 21 (1970).
[5]R. H. Cameron, and W. T. Martin, Trans. Am. Math. Soc. **58**, 184 (1945).
[6]N. Bourbaki, *Elements de mathematiques* (Paris, Hermann, 1969), Vol. VI, Chap. IX.
[7]I. M. Gelfand and A. M. Yaglom, J. Math. Phys. **1**, 48 (1960).
[8]A. Maheshwari, J. Phys. A 8, 1019 (1975); Acta Universitatis Wratislaviensis, Proceedings of the XIIth Winter School of Theoretical Physics in Karpacz, Poland (1975).
[9]C. DeWitt-Morette and A. Maheshwari, Acta Universitatis Wratislaviensis, Proceedings of the XIIth Winter School of Theoretical Physics in Karpacz, Poland (1975).
[10]I. M. Koval'chik, Russ. Math. Surveys **18**, 97 (1963).

A theorem on stress–energy tensors

Eric A. Lord

Department of Mathematics, Indian Institute of Science, Bangalore 560012, India
(Received 6 March 1975)

The equality of the symmetrized Noether stress–energy tensor (Belinfante's tensor) and the canonical stress–energy tensor (functional derivative of the Lagrangian density with respect to the metric) is established by methods based on the formalism of tetrads and Ricci rotation coefficients. The result holds for any Lagrangian which contains no derivatives of the fields higher than first order.

The equality of Belinfante's symmetrized stress–energy (SE) tensor and the canonical SE tensor (functional derivative of the Lagrangian density with respect to the metric) was demonstrated for integral spin fields by Rosenfeld.[1] For fields with half-integral spin it is not immediately clear how the canonical SE tensor should be defined. Goedecke[2] has shown that, subject to a certain prescription for carrying out the variation of the metric, Rosenfeld's equality holds for the Dirac field and for the coupled Maxwell and Dirac fields. Goedecke conjectured that a general proof of the equality for any field should be possible. The proof presented here is based on the tetrad formalism,[3–5] in which the Noether SE tensor and the Noether spin tensor are defined as functional derivatives of the Lagrangian density with respect to the tetrad components and the Ricci rotation coefficients.

Let $L(\phi, \partial_\mu\phi)$ be the Lagrangian of a set of fields ϕ in a Cartesian coordinate system in Minkowski spacetime. We generalize it to a Lagrangian density $\mathfrak{L}$ in a curvilinear coordinate system in the following way.[3] Introduce a tetrad h_μ^α and convert all the coordinate based indices $(\mu, \nu, \cdots)$ on ϕ to tetrad based indices $(\alpha, \beta, \cdots)$, by contractions with h_μ^α or its inverse h_α^μ. Introduce a set of Ricci rotation coefficients $\lambda^{\alpha\beta}{}_\mu$ for the purpose of constructing a derivative of ϕ that is covariant for coordinate transformations and spacetime dependent Lorentz rotations of the tetrad. Then

$$\mathfrak{L}(\phi, \partial_\mu\phi, h_\mu^\alpha, \lambda^{\alpha\beta}{}_\mu) = hL(\phi, \phi_\alpha) \tag{1}$$

is the required generalization, invariant under tetrad rotations and a scalar density of weight 1 under coordinate transformations. We have used the notation

$$h = |h_\mu^\alpha| = (-g)^{1/2} \tag{2}$$

and

$$\phi_\alpha = h_\alpha^\mu(\partial_\mu\phi + \tfrac{1}{2}\lambda^{\alpha\beta}{}_\mu G_{\alpha\beta}\phi). \tag{3}$$

The quantity in brackets in (3) is the *covariant derivative* ϕ_μ and the $G_{\alpha\beta}$ are the constant matrices which generate the Lorentz rotations in the ϕ-representation.

Two tensor densities t_α^μ and $s^\mu{}_{\alpha\beta}$ are defined as the functional derivatives of $\mathfrak{L}$ with respect to h_μ^α and $\lambda^{\alpha\beta}{}_\mu$. That is, for the infinitesimal variations

$$\delta h_\mu^\alpha = \zeta_\mu^\alpha, \quad \delta\lambda^{\alpha\beta}{}_\mu = \zeta^{\alpha\beta}{}_\mu \tag{4}$$

we have

$$\delta\mathfrak{L} \sim t_\alpha^\mu\zeta_\mu^\alpha + \tfrac{1}{2}s^\mu{}_{\alpha\beta}\zeta^{\alpha\beta}{}_\mu$$

(where $\sim$ denotes that all *divergences* have been omitted from an equation).

Suppose the original tetrad was $h_\mu^\alpha = \delta_\mu^\alpha$ before the variation (and hence $g_{\mu\nu} = \eta_{\mu\nu}$, i.e., the coordinate system was Cartesian), and suppose also the Ricci rotation coefficients were zero. Then, when we work only to first order in the infinitesimal quantities, the distinction between tetrad based indices and coordinate based indices is not relevant. All vector indices are raised and lowered with the Minkowski metric.

Substitute for the variations in (5) those brought about by an infinitesimal coordinate transformation $x^\mu \to x^\mu + \xi^\mu$:

$$\zeta_\mu^\alpha = -\partial_\mu\xi^\alpha, \quad \zeta^{\alpha\beta}{}_\mu = 0 \tag{6}$$

(to first order). Integrate (5) over a region of four-space on the boundary of which ξ^α vanishes and apply Gauss's theorem. We obtain the identity

$$\partial_\mu t_\alpha^\mu = 0 \tag{7}$$

in the Cartesian system. Similarly, substitute the variations brought about by an infinitesimal tetrad rotation with parameters $\lambda_{\alpha\beta} = -\lambda_{\beta\alpha}$:

$$\zeta_{\alpha\beta} = \lambda_{\alpha\beta}, \quad \zeta^{\alpha\beta}{}_\mu = \partial_\mu\zeta^{\alpha\beta}, \tag{8}$$

and we obtain the identity

$$\partial_\mu s^\mu{}_{\alpha\beta} + 2t_{[\alpha\beta]} = 0. \tag{9}$$

The rotation coefficients and the tetrad have been treated here as independent fields in the variation. However, since we have chosen the rotation coefficients to be zero in the initial reference system and since we know how they transform,[3] we can easily show that they can be constructed from the tetrad. The infinitesimal form is

$$\zeta_{\alpha\beta\mu} = \partial_\mu\zeta_{[\alpha\beta]} + \partial_\alpha\zeta_{(\beta\mu)} - \partial_\beta\zeta_{(\alpha\mu)}. \tag{10}$$

Substituting this in (5) gives

$$\delta\mathfrak{L} \sim (t^{(\alpha\beta)} + \partial_\rho s^{(\alpha\beta)\rho})\zeta_{(\alpha\beta)} - (t^{[\alpha\beta]} + \tfrac{1}{2}\partial_\rho s^{\rho\alpha\beta})\zeta_{[\alpha\beta]}. \tag{11}$$

The second term vanishes on account of (9) and the first term can be reexpressed, using (9), as

$$\delta\mathfrak{L} \sim \theta^{\alpha\beta}\zeta_{(\alpha\beta)}, \tag{12}$$

where the symmetric tensor $\theta^{\alpha\beta}$ is

$$\theta^{\alpha\beta} = t^{\alpha\beta} + \partial_\rho(\tfrac{1}{2}s^{\rho\alpha\beta} - s^{(\alpha\beta)\rho}). \tag{13}$$

For *integral spin fields*, an alternative (and more usual) way of generalizing $L(\phi, \partial_\mu\phi)$ to curvilinear

coordinates is by means of the metric and the Christoffel symbols, without the introduction of a tetrad:

$$\mathfrak{L}(\phi, \partial_\mu\phi, g_{\mu\nu}, \partial_\rho g_{\mu\nu}). \tag{14}$$

Because (1) and (14) are equal in the initial (Cartesian) system, they are equal in any system, because they have the same transformation properties for coordinate changes and tetrad changes. The canonical SE tensor density is defined by arbitrary variation of the metric $\delta g_{\mu\nu} = \xi_{\mu\nu}$:

$$\delta\mathfrak{L} \sim \tfrac{1}{2} T^{\mu\nu}\xi_{\mu\nu} \tag{15}$$

so that

$$T_{\mu\nu} = 2\left(\partial_\rho\left(\frac{\partial\mathfrak{L}}{\partial\partial_\rho g^{\mu\nu}}\right) - \frac{\partial\mathfrak{L}}{\partial g^{\mu\nu}}\right). \tag{16}$$

For *half-integer* spin this tensor is undefined, because a spinor index is *essentially* related to tetrad rotations, not to coordinate transformations. The Lagrangian density of a half-integer spin field *necessarily* contains the tetrad components. However, if we are interested only in the reference systems that differ infinitesimally from gartesian ones, the canonical SE tensor (16) can be defined for half-integer spin *provided* we destroy the independence of infinitesimal tetrad rotation and infinitesimal coordinate transformations. The simplest way of doing this is to impose the restriction

$$\zeta_{[\alpha\beta]} = 0 \tag{17}$$

on the tetrad variations. This condition is implicit in Goedecke's treatment of the Dirac field, though he does not explicitly introduce the tetrad concept.

Now, because of the orthonormality of the tetrad

$$\eta_{\alpha\beta} h^\alpha_\mu h^\beta_\nu = g_{\mu\nu} \tag{18}$$

we have

$$\xi_{\mu\nu} = 2\zeta_{(\mu\nu)}, \tag{19}$$

and (15) is

$$\delta\mathfrak{L} \sim T^{\mu\nu}\zeta_{(\mu\nu)}. \tag{20}$$

Comparison of (12) and (20) gives immediately, in the Cartesian system (by an integration and application of Gauss's theorem),

$$T_{\mu\nu} = \theta_{\mu\nu}. \tag{21}$$

[Incidentally, the identity $\partial_\mu T^{\mu\nu} = 0$ follows from substituting in (15) the variation $\xi_{\mu\nu} = -\partial_\mu\xi_\nu - \partial_\nu\xi_\mu$, integrating, and applying Gauss's theorem.]

Thus the equality of the Belinfante SE tensor and the canonical SE tensor is established if it can be shown that $t_{\mu\nu}$ is actually the Noether SE tensor, i.e., we have to show that

$$t^\mu_\nu = \delta^\mu_\nu\mathfrak{L} - \pi^\mu\partial_\nu\phi, \quad s^\mu{}_{\alpha\beta} = \pi^\mu G_{\alpha\beta}\phi, \tag{22}$$

where

$$\pi^\mu = \partial\mathfrak{L}/\partial\partial_\mu\phi. \tag{23}$$

These follow from the form (1) of $\mathfrak{L}$. Note that, in (1), $L(\phi, \phi_\alpha)$ is constructed only from ϕ, ϕ_α and the Minkowski metric; the tetrad components do not occur explicitly, but only in the structure of ϕ_α. For variation of the tetrad and rotation coefficients,

$$\begin{aligned}\delta\mathfrak{L} &= \delta hL + h\delta L = \delta hL + h(\partial L/\partial\phi_\alpha)\delta\phi_\alpha \\ &= \zeta^\mu_\mu\mathfrak{L} - (\partial\mathfrak{L}/\partial\phi_\alpha)(-\zeta^\mu_\alpha\phi_\mu + \tfrac{1}{2}\zeta^{\alpha\beta}{}_\mu G_{\alpha\beta}\phi) \\ &= (\delta^\mu_\nu\mathfrak{L} - \pi^\mu\phi_\nu)\zeta^\nu_\mu + \tfrac{1}{2}(\pi^\mu G_{\alpha\beta}\phi)\zeta^{\alpha\beta}{}_\mu.\end{aligned}$$

Comparing this expression with (5) identifies the tensors t^μ_ν and $s^\mu{}_{\alpha\beta}$, which coincide with (22) in the Cartesian system.

[1] L. Rosenfeld, Mem. Acad. Roy. Belg. **18**, No. 6 (1940).
[2] G.H. Goedecke, J. Math. Phys. **15**, 792 (1974).
[3] T.W.B. Kibble, J. Math. Phys. **2**, 212 (1961).
[4] D.W. Sciama, Proc. Camb. Phil. Soc. **54**, 72 (1958).
[5] D.W. Sciama, *Recent Developments in General Relativity* (Pergamon, London, and PWN, Warsaw, 1962).

Killing vectors in empty space algebraically special metrics. II

A. Held

Institut für Theoretische Physik, Universität Bern, 3012 Bern, Switzerland
(Received 14 May 1975)

Empty space algebraically special metrics possessing an expanding degenerate principal null vector and Killing vectors are investigated. Attention is centered on that class of Killing vector (called nonpreferred) which is necessarily spacelike in the asymptotic region. A detailed analysis of the relationship between the Petrov–Penrose classification and these Killing vectors is carried out.

1. INTRODUCTION

Despite the fifty odd years of work since the appearance of the Einstein equations, the number of known solutions, and in particular those which may be considered as having physical significance, remains limited. The traditional approach leads in all but the simplest cases to insoluble nonlinear differential equations and then more or less grinds to a halt. Since the outlook for further progress in this direction is not very encouraging, one should perhaps seek other roads to information. In particular, one can ask what geometric and physical information may be extracted from a metric without knowing its exact form, i. e., without having the solution to its associated minimal differential equation. [1]

To this end the author has been developing a technique[2,3] applicable to vacuum algebraically special metrics with expanding degenerate principal null direction (metrics of this sort will henceforth be referred to as EDPN metrics) which allows calculations to be done without appeal to coordinates and enables an investigation to be carried past the barrier of insoluble differential equations.

As a first application, this technique has been used to investigate Killing vectors in vacuum EDPN metrics. It was found that these vectors are of two distinct types, called preferred and nonpreferred (cf. below). An analysis of those metrics containing a preferred Killing vector has previously been published. [4] The investigation is completed with this paper which focuses attention on those metrics containing at least one nonpreferred Killing vector.

Some of the results presented have been obtained using an algebraic approach by others (notably Petrov[5] and Collinson and French[6]); however, it is felt that the additional insight obtained through the approach presented here justifies their rederivation. The coordinate free approach shows that the algebraically special metrics are rich in geometric structure, and some of the proofs presented are based on appeal to these geometric structures.

The paper is divided into sections as follows. Section 2 contains a brief resume of the technique and of some of the more pertinent results obtained in Ref. 4. Sections 3 and 4 deal with general properties of EDPN metrics and introduce (for when needed) a specific choice of coordinate system and gauge. They also contain additional information on the operators of the formalism. A detailed analysis of how many Killing vectors of each type may be present in the various subclassifications of algebraically special metrics (type N, $\{3,1\}$, etc.,) is carried out in Sec. 5, and finally a brief discussion of the results is given.

2. REVIEW OF THE TECHNIQUE

It deals with an algebra of objects and differential operators all of which are of good spin and boost weight in the sense of Ref. 7. That is, given a standard null tetrad l_a, n_a, m_a, and $\bar{m}_a$ (the bar represents complex conjugation) such that

$$l^a n_a = -m^a \bar{m}_a = 1 \tag{2.1}$$

and all other scalar products vanish, an object η is said to be of weight (p,q) if, under the tetrad gauge transformation,

$$l^a \to \lambda l^a, \quad n^a \to \lambda^{-1} n_a \tag{2.2a}$$

$$m^a \to \exp(i\theta) m^a, \tag{2.2b}$$

$$\eta \to \lambda^t \exp(is\theta)\eta, \tag{2.3}$$

where

$t = (p+q)/2$ is the boost weight,

$s = (p-q)/2$ is the spin weight.

The fundamental objects of the formalism are those elements of the Newman–Penrose (N–P) spin-coefficient algebra[8] which possess good weight. Following the practice initiated in Ref. 7, some of these will be designated differently than is customary. Those spin coefficients differing from the notation of Ref. 8 are (the original notation appears in brackets)

$$\rho'(-\mu), \quad \kappa'(-\nu), \quad \sigma'(-\lambda), \quad \tau'(-\pi).$$

The differential operators[2,3] are essentially those those of Ref. 7, modified to take advantage of the simplifying properties of algebraically special empty space metrics. To define the new operators, introduce the auxiliary vectors

$$\tilde{\alpha}_a = n^b \nabla_a l_b - \left[\tau\bar{\tau}\left(\frac{1}{\rho}+\frac{1}{\bar{\rho}}\right)+\frac{1}{2}\left(\frac{\Psi_2}{\rho}+\frac{\bar{\Psi}_2}{\bar{\rho}}\right)\right] l_a + \frac{\rho}{\bar{\rho}}\bar{\tau} m_a + \frac{\rho}{\rho}\tau\bar{m}_a, \tag{2.4a}$$

$$\tilde{\beta}_a = \bar{m}^b \nabla_a m_b + \left[\Omega^\circ \tau\bar{\tau} + \frac{1}{2}\left(\frac{\Psi_2}{\rho}-\frac{\bar{\Psi}_2}{\bar{\rho}}\right)\right] l_a - \frac{\rho}{\bar{\rho}}\bar{\tau} m_a + \frac{\bar{\rho}}{\rho}\tau\bar{m}_a, \tag{2.4b}$$

where

$$\Omega^\circ = \frac{1}{\bar{\rho}} - \frac{1}{\rho} \quad (\rho \neq 0 \text{ is assumed throughout}).$$

The derivative operators þ, $\tilde{\text{þ}}'$, $\tilde{\delta}$, $\tilde{\delta}'$ are then defined by

$$\nabla_a - t\tilde{\alpha}_a + s\tilde{\beta}_a = n_a\text{þ} + l_a(\tilde{\text{þ}}' + \bar{\tau}\tilde{\delta} + \tau\tilde{\delta}') - \bar{\rho}m_a\tilde{\delta}' - \rho\bar{m}_a\tilde{\delta}). \tag{2.5}$$

These operators have weights $\text{þ}(1,1)$, $\tilde{\text{þ}}'(-1,-1)$, $\tilde{\delta}(0,-2)$ and $\tilde{\delta}'(-2,0)$, and also possess the properties that

$$[\text{þ},\tilde{\text{þ}}']\eta^\circ = [\text{þ},\tilde{\delta}]\eta^\circ = [\text{þ},\tilde{\delta}']\eta^\circ = 0 \tag{2.6}$$

if

$$\text{þ}\eta^\circ = 0. \tag{2.7}$$

(A degree sign $\circ$ will be used to mark any quantity annihilated by þ.) In addition

$$\tilde{\bar{\text{þ}}}' = \bar{\tilde{\text{þ}}}', \quad \tilde{\bar{\delta}} = \bar{\tilde{\delta}}'. \tag{2.8}$$

For a more detailed treatment of the operators, including symmetry properties and the technique of integration without coordinates, the reader is referred to Refs. 2 and 7.

In Ref. 3 the equations were written out and partially integrated for the standard choice of tetrad, i.e., one chosen so that

$$\kappa = \tau = \tau' = \sigma = \sigma' = \Psi_0 = \Psi_1 = 0. \tag{2.9}$$

The results were

$$\rho' = \bar{\rho}\rho'^\circ - \tfrac{1}{2}(\rho^2 + \rho\bar{\rho})\Psi_2^\circ, \tag{2.10}$$

$$\kappa' = \kappa'^\circ - \rho\Psi_3^\circ - \tfrac{1}{2}\rho^2\tilde{\delta}'\Psi_2^\circ - \tfrac{1}{2}\rho^3\Psi_2^\circ\tilde{\delta}'\Omega^\circ, \tag{2.11}$$

$$\Psi_2 = \rho^3\Psi_2^\circ, \tag{2.12}$$

$$\Psi_3 = \rho^2\Psi_3^\circ + \rho^3\tilde{\delta}'\Psi_2^\circ + \tfrac{3}{2}\rho^4\Psi_2^\circ\tilde{\delta}'\Omega^\circ, \tag{2.13}$$

$$\Psi_4 = \rho\Psi_4^\circ + \rho^2\tilde{\delta}'\Psi_3^\circ + \rho^3(\Psi_3^\circ\tilde{\delta}'\Omega^\circ + \tfrac{1}{2}\tilde{\delta}'\tilde{\delta}'\Psi_2^\circ) + \rho^4(\tfrac{3}{2}\tilde{\delta}'\Psi_2^\circ\tilde{\delta}'\Omega^\circ + \tfrac{1}{2}\Psi_2^\circ\tilde{\delta}'\tilde{\delta}'\Omega^\circ) + \tfrac{3}{2}\rho^5\Psi_2^\circ(\tilde{\delta}'\Omega^\circ)^2. \tag{2.14}$$

ρ, ρ'°, κ'°, Ψ_2°, Ψ_3°, and Ψ_4° are undetermined functions which are subject to the relations

$$\tilde{\delta}'\rho'^\circ = -(\Psi_3^\circ + \Omega^\circ\kappa'^\circ), \tag{2.15}$$

$$\tilde{\text{þ}}'\rho'^\circ = \tilde{\delta}\kappa'^\circ, \tag{2.16}$$

$$\tilde{\delta}'\kappa'^\circ = -\Psi_4^\circ, \tag{2.17}$$

$$\tilde{\delta}\Psi_4^\circ = \tilde{\text{þ}}'\Psi_3^\circ, \tag{2.18}$$

$$\tilde{\delta}\Psi_3^\circ = \tilde{\text{þ}}'\Psi_2^\circ, \tag{2.19}$$

$$\tilde{\delta}\Psi_2^\circ = 0, \tag{2.20}$$

$$\tilde{\delta}\tilde{\delta}'\Omega^\circ = 2\Omega^\circ\bar{\rho}'^\circ + \Psi_2^\circ - \bar{\Psi}_2^\circ, \tag{2.21}$$

$$\tilde{\text{þ}}'\Omega^\circ = \bar{\rho}'^\circ - \rho'^\circ. \tag{2.22}$$

The derivative operators acting on ρ are

$$\text{þ}\rho = \rho^2, \quad \tilde{\text{þ}}'\rho = \rho^2\bar{\rho}'^\circ - \tfrac{1}{2}\rho^3\Psi_2^\circ - \tfrac{1}{2}\bar{\rho}\rho^2\bar{\Psi}_2^\circ,$$
$$\bar{\delta}\rho = 0, \quad \tilde{\delta}'\rho = \rho^2\tilde{\delta}'\Omega^\circ, \tag{2.23}$$

and the commutators are

$$[\text{þ},\tilde{\text{þ}}'] = -\tfrac{1}{2}(\rho^2\Psi_2^\circ + \bar{\rho}^2\bar{\Psi}_2^\circ)\text{þ}, \tag{2.24}$$

$$[\text{þ},\tilde{\delta}] = 0, \tag{2.25}$$

$$[\text{þ},\tilde{\delta}'] = 0, \tag{2.26}$$

$$[\tilde{\delta},\tilde{\delta}'] = \left(\frac{\bar{\rho}'^\circ}{\bar{\rho}} - \frac{\rho'^\circ}{\rho} + \tfrac{1}{2}(2+\rho\Omega^\circ)\Psi_2^\circ - \tfrac{1}{2}(2-\bar{\rho}\Omega^\circ)\bar{\Psi}_2^\circ\right)\text{þ} + \Omega^\circ\tilde{\text{þ}}' + p\rho'^\circ - q\bar{\rho}'^\circ, \tag{2.27}$$

$$[\tilde{\text{þ}}',\tilde{\delta}] = \left(-\frac{\bar{\kappa}'^\circ}{\bar{\rho}} + \bar{\Psi}_3^\circ + \tfrac{1}{2}\bar{\rho}\tilde{\delta}\bar{\Psi}_2^\circ - \tfrac{1}{2}\bar{\rho}^2\bar{\Psi}_2^\circ\tilde{\delta}\Omega^\circ\right)\text{þ} + q\bar{\kappa}'^\circ, \tag{2.28}$$

$$[\tilde{\text{þ}}',\tilde{\delta}'] = \left(-\frac{\kappa'^\circ}{\rho} + \Psi_3^\circ + \tfrac{1}{2}\rho\tilde{\delta}'\Psi_2^\circ + \tfrac{1}{2}\rho^2\Psi_2^\circ\tilde{\delta}'\Omega^\circ\right)\text{þ} + p\kappa'^\circ. \tag{2.29}$$

Not as in the N–P formalism,[8] these commutators are not identities but are additional restrictions which must be satisfied before a solution is valid.

The following results concerning Killing vectors in vacuum EDPN metrics were obtained in Ref. 4.

If ξ^a is a Killing vector and

$$\xi^a = \xi_0 n^a + \xi_2 l^a - \bar{\xi}_1 m^a - \xi_1\bar{m}^a, \tag{2.30}$$

then, in the standard choice of tetrad, expressions for ξ_0, ξ_1, ξ_3 are given by

$$\xi_0 = \xi_0^\circ, \tag{2.31a}$$

$$\xi_1 = \xi_1^\circ/\rho, \tag{2.31b}$$

$$\xi_2 = \frac{1}{2}\left(\frac{1}{\rho} + \frac{1}{\bar{\rho}}\right)\tilde{\text{þ}}'\xi_0^\circ + \xi_1^\circ\tilde{\delta}'\Omega^\circ - \bar{\rho}'^\circ\xi_0^\circ - \tfrac{1}{2}\Omega^\circ\tilde{\text{þ}}'\xi_0^\circ + \tfrac{1}{2}(\rho\Psi_2^\circ + \bar{\rho}\bar{\Psi}_2^\circ)\xi_0^\circ, \tag{2.31c}$$

where ξ_0° and ξ_1° are undetermined functions which satisfy the relations

$$\tilde{\delta}\xi_0^\circ = \Omega^\circ\xi_1^\circ, \tag{2.32}$$

$$\tilde{\text{þ}}'\xi_1^\circ = \tilde{\delta}\xi_1^\circ = 0, \tag{2.33}$$

$$\tilde{\delta}\bar{\xi}_1^\circ + \tilde{\delta}'\xi_1^\circ = -2\tilde{\text{þ}}'\xi_0^\circ, \tag{2.34}$$

$$\xi_0^\circ\tilde{\text{þ}}'\rho'^\circ - \xi_1^\circ\tilde{\delta}'\rho'^\circ - \xi_1^\circ\tilde{\delta}\rho'^\circ + 2\rho'^\circ\tilde{\text{þ}}'\xi_0^\circ = 0, \tag{2.35}$$

$$\xi_0^\circ\tilde{\text{þ}}'\Psi_2^\circ - \xi_1^\circ\tilde{\delta}'\Psi_2^\circ + 3\Psi_2^\circ\tilde{\text{þ}}'\xi_0^\circ = 0, \tag{2.36}$$

$$\xi_0^\circ\tilde{\text{þ}}'\Omega^\circ - \xi_1^\circ\tilde{\delta}'\Omega^\circ - \bar{\xi}_1^\circ\tilde{\delta}\Omega^\circ + 3\Omega^\circ\tilde{\text{þ}}'\xi_0^\circ = 0. \tag{2.37}$$

This last equation was not explicitly displayed in Ref. 4, but is inherent in the condition that ξ_2 be real.

In general, if η is a (p,q) quantity and ξ^a is a Killing vector, then

$$\xi_0^\circ\tilde{\text{þ}}'\eta - \xi_1^\circ\tilde{\delta}'\eta - \bar{\xi}_1^\circ\tilde{\delta}\eta + t\eta\tilde{\text{þ}}'\xi_0^\circ - \tfrac{1}{2}s(\tilde{\delta}'\xi_1^\circ - \tilde{\delta}\bar{\xi}_1^\circ) = 0. \tag{2.38}$$

Equations (2.35)–(2.37) differ from the general case (2.38) in that they are part of the necessary and sufficient conditions that ξ^a defined by (2.30) be a Killing vector. The normalization of ξ^a is given by

$$\tfrac{1}{2}\xi_a\xi^a = -\frac{\xi_1^\circ\bar{\xi}_1^\circ}{\rho\bar{\rho}} + \xi_0^\circ\left[\frac{1}{2}\left(\frac{1}{\rho} + \frac{1}{\bar{\rho}}\right)\tilde{\text{þ}}'\xi_0^\circ + \xi_1^\circ\tilde{\delta}'\Omega^\circ - \bar{\rho}'^\circ\xi_0^\circ - \tfrac{1}{2}\Omega^\circ\tilde{\text{þ}}'\xi_0^\circ\right] + \tfrac{1}{2}(\rho\Psi_2^\circ + \bar{\rho}\bar{\Psi}_2^\circ)(\xi_0^\circ)^2. \tag{2.39}$$

From this expression it is clear in the asymptotic region ($\rho \to 0$) unless $\xi_1^\circ = 0$, the Killing vector ξ^a will be spacelike. Therefore, EDPN metrics which are asymptotically stationary must contain a Killing vector such that $\xi_1^\circ = 0$. A Killing vector possessing this property is referred to as "preferred."[4] The existence of a preferred Killing vector restricts the space to an extent that all such metrics may be found. This program was carried out in Ref. 4. The results obtained there

which will be used frequently in this paper are as follows:

There can be at most one preferred Killing vector. If one exists, then

$$\kappa'^{\circ} = \rho'^{\circ} - \bar{\rho}'^{\circ} = 0 \tag{2.40}$$

and there exists a gauge in which $\tilde{\alpha}_a = 0$. [This last property follows directly from (2.40).[3]]

Using Eq. (2.34) and the commutation relations, it is a simple matter to obtain the following useful results:

$$\tilde{\text{þ}}'\tilde{\text{þ}}'\xi_0^{\circ} = -\kappa'^{\circ}\xi_1^{\circ} - \bar{\kappa}'^{\circ}\bar{\xi}_1^{\circ}, \tag{2.41}$$

$$\tilde{\text{ð}}\tilde{\text{þ}}'\xi_0^{\circ} = -\bar{\kappa}'\xi_0^{\circ} + (\bar{\rho}'^{\circ} - \rho'^{\circ})\xi_1^{\circ}, \tag{2.42}$$

$$\tilde{\text{ð}}'\tilde{\text{þ}}'\xi_0^{\circ} = -\kappa'^{\circ}\xi_0^{\circ} + (\rho'^{\circ} - \bar{\rho}'^{\circ})\bar{\xi}_1^{\circ}. \tag{2.43}$$

3. A GAUGE AND A COORDINATE SYSTEM

While a good many of the results to be presented will be obtained without appeal to a gauge or a coordinate system, the author has been unable to obtain the desired goal of eliminating them entirely. It is therefore desirable to develop a gauge and a coordinate system which harmonize as closely as possible with the technique of investigation.

As is well known, in the standard choice of tetrad the vector m_a is proportional to a gradient and so may be written as

$$-\bar{P}m_a = \bar{\zeta}_{,a}. \tag{3.1}$$

A natural choice for two of the coordinate functions is ζ and $\bar{\zeta}$. Equation (3.1) defines ζ to within a transformation of the form

$$\zeta' = \zeta'(\zeta), \tag{3.2}$$

which in turn induces the transformation

$$P \to P' = \frac{\partial \zeta}{\partial \zeta'} P. \tag{3.3}$$

The factor P obeys the Blöchlinger equations[3]

$$\begin{aligned} &\text{þ}P = \rho P \to P = \rho P^{\circ}, \\ &\tilde{\text{þ}}'P^{\circ} = \tilde{\text{ð}}'P^{\circ} = 0. \end{aligned} \tag{3.4}$$

With the choice of ζ and $\bar{\zeta}$ as x^3 and x^4 respectively, the tetrad vectors l^a and n^a are confined to a subspace of the tangent space spanned by $\partial/\partial x^1$ and $\partial/\partial x^2$.

Since the operators þ and $\tilde{\text{þ}}'$, are involutive under commutation, there exists a gauge such that[3]

$$l^a\tilde{\alpha}_a = n^a\tilde{\alpha}_a = l^a\tilde{\beta}_a = n^a\tilde{\beta}_a = 0. \tag{3.5}$$

Let this choice be considered the standard choice. This still leaves the gauge freedom

$$\lambda \to \lambda(\zeta, \bar{\zeta}), \quad \theta \to \theta(\zeta, \bar{\zeta}). \tag{3.6}$$

The coordinate x^2 will be chosen to be an affine parameter along the integral curves of the degenerate principal null direction. Several choices of x^1 are viable, for all of which

$$l^a x^1_{,a} = 0, \tag{3.7}$$

so that if Eq. (2.7) holds, then

$$\eta^{\circ} = \eta^{\circ}(x^1, \zeta, \bar{\zeta}). \tag{3.8}$$

If, in addition, $\tilde{\text{þ}}'\eta^{\circ} = 0$, then

$$\eta^{\circ} = \eta^{\circ}(\zeta, \bar{\zeta}). \tag{3.9}$$

With these choices, the operators $\tilde{\text{ð}}$, $\tilde{\text{ð}}'$ [operating on a degree marked (p, q) object] may be displayed in the convenient form

$$\tilde{\text{ð}}\eta^{\circ} = \left(\frac{m^1}{\rho}\frac{\partial}{\partial x^1} - qm^a\tilde{\alpha}_a + \text{ð}_2\right)\eta^{\circ}, \tag{3.10a}$$

$$\tilde{\text{ð}}'\eta^{\circ} = \left(\frac{\bar{m}^1}{\rho}\frac{\partial}{\partial x^1} - p\bar{m}^a\tilde{\alpha}_a + \bar{\text{ð}}_2\right)\eta^{\circ}, \tag{3.10b}$$

where ð_2 is defined by

$$\text{ð}_2\eta = P^{\circ}\bar{P}^{\circ - s}\frac{\partial}{\partial \zeta}\bar{P}^{\circ s}\eta \tag{3.11}$$

and is to within a factor the "edth" of Newman and Penrose,[9,10] defined on a two-surface with metric

$$ds = \frac{d\zeta\, d\bar{\zeta}}{\bar{P}^{\circ}P^{\circ}}. \tag{3.12}$$

Although the original operator was defined for $P^{\circ} = \frac{1}{2}(1 + \zeta\bar{\zeta})$, it is clearly the same animal. It is of interest to note that this two-dimensional edth admits of a general coordinate free definition by means of the following commutative diagram.

$$\begin{array}{ccc} T_{\mu\ldots\nu} & \xrightarrow{m^{\sigma}\nabla_{\sigma}(\bar{m}^{\sigma}\nabla_{\sigma})} & S_{\mu\ldots\nu} \\ \Pi(m,\bar{m}) \downarrow & & \downarrow \Pi(m,\bar{m}) \\ T & \xrightarrow{\text{ð}_2(\bar{\text{ð}}_2)} & S \end{array}$$

m^{σ} and $\bar{m}^{\sigma}$ are vector defined on the two space with normalization

$$m^{\sigma}\bar{m}_{\sigma} \pm 1 = m^{\sigma}m_{\sigma} = \bar{m}^{\sigma}\bar{m}_{\sigma} = 0, \tag{3.13}$$

and $\Pi(m, \bar{m})$ is any operator which contracts the tensors $T_{\mu\ldots\nu}$ and $S_{\mu\ldots\nu}$ with the dyad vectors m^{σ} and $\bar{m}^{\sigma}$ an appropriate number of times to form the spin weighted scalars T and S respectively. (If the metric has a positive definite signature, then m^{σ} is complex). Unfortunately, no such elementary description of the properties of the four-dimensional operators þ, $\text{þ}'$, ð, $\text{ð}'$ exists.[7]

As a simple application of the above, consider the product $\bar{\xi}_1^{\circ}P^{\circ}$, where $\bar{\xi}_1^{\circ}$ is the scalar defined by (2.30) with ξ^a a nonpreferred Killing vector. $\bar{\xi}_1^{\circ}P^{\circ}$ is a (0,0) object which obeys

$$\text{þ}(\bar{\xi}_1^{\circ}P^{\circ}) = \tilde{\text{þ}}'(\bar{\xi}_1^{\circ}P^{\circ}) = \tilde{\text{ð}}'(\bar{\xi}_1^{\circ}P^{\circ}) = 0, \tag{3.14a}$$

so that

$$\bar{\xi}_1^{\circ} = \bar{\xi}(\zeta)/P^{\circ}, \tag{3.14b}$$

where $\bar{\xi}(\zeta)$ is an unknown function. If desired, the $\bar{\xi}(\zeta)$ associated with any one Killing vector may be absorbed into P° by using the coordinate freedom (3.2) with

$$\zeta'(\zeta) = \int \frac{d\zeta}{\xi(\zeta)}. \tag{3.15}$$

Combining this with Eqs. (2.30) and (3.1), it follows that the coordinate ζ may be chosen such that one nonpreferred Killing vector may always be put in the form

$$\xi^1\frac{\partial}{\partial x^1} + \xi^2\frac{\partial}{\partial x^2} + \bar{\rho}\frac{\partial}{\partial \zeta} + \rho\frac{\partial}{\partial \bar{\zeta}}. \tag{3.16}$$

For metrics of the Robinson–Trautman type[11]

($\Omega^\circ = 0$), the operator $\tilde{\delta}$ may be reduced further. Since $\rho'^\circ = \bar{\rho}'^\circ$ [Eq. (2.22)], the commutator $[\tilde{\delta}, \tilde{\delta}']\eta^\circ$ vanishes if η° is a (p,p) quantity and there is a gauge[3] in which

$$m^a \tilde{\alpha}_a = \bar{m}^a \tilde{\alpha}_a = 0. \tag{3.17}$$

But $\tilde{\alpha}_a$ transforms as

$$\tilde{\alpha}_a \to \tilde{\alpha}_a + \lambda + \lambda - {}^1\lambda_{,a} \tag{3.18}$$

so that in any other gauge, $m^a\tilde{\alpha}_a$ is of the form $\lambda^{*-1}m^a\lambda^*_{,a}$, where λ^* relates the gauge under consideration to one determined by (3.17). In addition, since the tetrad vector l^a is twist free, the coordinate function x^1 may be chosen such that

$$x^1_{,a} \sim l_a. \tag{3.19}$$

If this be done, then m^a takes the simple form $\bar{\rho}P^\circ\delta^a_\xi$ and the expressions (3.10) with $P^\circ = \bar{P}^\circ$ reduce to

$$\tilde{\delta}\eta^\circ = P^{\circ\,1-s}\lambda^{*q}\frac{\partial}{\partial\xi}(P^{\circ s}\lambda^{*-q}\eta^\circ), \tag{3.20a}$$

$$\tilde{\delta}'\eta^\circ = P^{\circ\,1+s}\lambda^{*p}\frac{\partial}{\partial\bar{\xi}}(P^{\circ\,-s}\lambda^{*-p}\eta^\circ). \tag{3.20b}$$

4. SOME GENERAL RESULTS

The existence of a single nonpreferred Killing vector ξ^a in a vacuum EDPN metric introduces the additional functions ξ°_0 and ξ°_1 and the associated equations (2.32)–(2.37) into the problem, but as far as the author has been able to determine, there are no associated physical restrictions. This statement no longer holds for those EDPN metrics with more than one Killing vector, and the balance of this paper will be concerned with these additional restrictions.

As a first step in this program, one calculates the commutator of two Killing vectors. If the tetrad vectors are tied to the intrinsic geometry of the space it is possible to derive a simple expression for the commutator of two Killing vectors (cf. Appendix, Ref. 4). This result is applied here as follows. Assume that there are r linearly independent Killing vectors $\underset{(\alpha)}{\xi^a}$ ($\alpha, \beta, \gamma = 1, \ldots, r$) and that the commutation relations are given by

$$[\underset{(\alpha)}{\xi}, \underset{(\beta)}{\xi}]^a = \underset{\alpha\beta}{c}{}^{\gamma}\, \underset{(\gamma)}{\xi^a} \tag{4.1}$$

where $\underset{\alpha\beta}{c}{}^{\gamma}$'s are the structure constants of the associated Lie group. Inserting the expression (2.31) for the $\underset{(\alpha)}{\xi^a}$'s into the commutation relations of Ref. 4 and using the integration techniques of Ref. 2, one arrives at the relations

$$\underset{\alpha\beta}{c}{}^{\gamma}\, \underset{(\gamma)}{\xi^\circ_0} = \underset{[(\beta)}{\xi^\circ_0}\tilde{\mathfrak{p}}'\, \underset{(\alpha)]}{\xi^\circ_0} + \underset{[(\alpha)}{\bar{\xi}^\circ_1}\, \underset{(\beta)]}{\xi^\circ_1}\,\Omega^\circ, \tag{4.2a}$$

$$\underset{\alpha\beta}{c}{}^{\gamma}\, \underset{(\gamma)}{\xi^\circ_1} = \underset{[(\alpha)}{\xi^\circ_1}\tilde{\delta}'\, \underset{(\beta)]}{\xi^\circ_1}, \tag{4.2b}$$

$$\Omega^\circ\kappa'^\circ = 0, \tag{4.2c}$$

where

$$\underset{[(\beta)}{\xi^\circ_0}\tilde{\mathfrak{p}}'\, \underset{(\alpha)]}{\xi^\circ_0} = \underset{(\beta)}{\xi^\circ_0}\tilde{\mathfrak{p}}'\, \underset{(\alpha)}{\xi^\circ_0} - \underset{(\alpha)}{\xi^\circ_0}\tilde{\mathfrak{p}}'\, \underset{(\beta)}{\xi^\circ_0}, \quad \text{etc.}$$

Equation (4.2c) involves considerable labor and includes the use of the general expression (2.38) applied to both ρ'° and $\tilde{\delta}'\Omega^\circ$.

Equations (2.17) and (2.14) show that if $\kappa'^\circ = 0$, the metric is nonradiative, while $\Omega^\circ = 0$ means that the metric is Robinson–Trautman. Equation (4.1c) thus gives the general result:

If a vacuum EDPN metric has two or more Killing vectors, it must be either nonradiative or a Robinson–Trautman metric.

Petrov[5] has shown that a type $\{2,1,1\}$, D, or $\{3,1\}$ vacuum metric can have at most 4 Killing vectors. This result is now rederived and extended to the following:

A vacuum EDPN metric can admit at most four Killing vectors, and when this is the case, one of them must be preferred.

The proof consists of showing that there exists a homomorphism of the Lie algebra of Killing vectors onto a two-dimensional subspace of the tangent space and that the kernel of this homomorphism is a preferred Killing vector. Since the maximal number of Killing vectors which can act on a space of two dimensions is three and there is at most one preferred Killing vector, the result follows trivially once the homomorphism is established.

Equation (4.2b) when written out using the coordinate system of Sec. 3 and with $\underset{(\alpha)}{\xi^\circ_1}$ expressed as in (3.14) appear as

$$\underset{\alpha\beta}{c}{}^{\gamma}\, \underset{(\gamma)}{\xi}(\bar{\xi}) = \underset{(\alpha)}{\xi}(\bar{\xi})\frac{\partial}{\partial\bar{\xi}}\underset{(\beta)}{\xi}(\bar{\xi}) - \underset{(\beta)}{\xi}(\bar{\xi})\frac{\partial}{\partial\bar{\xi}}\underset{(\alpha)}{\xi}(\bar{\xi}). \tag{4.3}$$

This relation clearly induces a homomorphism of the Lie algebra of vectors $\underset{(\alpha)}{\xi^a}\partial/\partial x^a$ onto the two-dimensional subspace of the tangent space spanned by the vectors $\partial/\partial x^3$ ($\partial/\partial\xi$) and $\partial/\partial x^4$ ($\partial/\partial\bar{\xi}$). The kernel of the homomorphism consists of Killing vectors $\underset{(\alpha)}{\xi^a}$ for which $\underset{(\alpha)}{\xi^\circ_1} = 0$, and such a Killing vector is by definition a preferred Killing vector. This is the required homomorphism and so the result is proved.

A second homomorphism also exists for Robinson–Trautman metrics. Since ξ°_0 is real, using the expressions (3.20) for $\tilde{\delta}$ and $\tilde{\delta}'$ in Eq. (2.32) and its complex conjugate leads to

$$\frac{\partial}{\partial\bar{\xi}}\lambda^{*-1}\underset{(\alpha)}{\xi^\circ_0} = \frac{\partial}{\partial\bar{\xi}}\lambda^{*-1}\underset{(\alpha)}{\xi^\circ_0} = 0, \tag{4.4}$$

so that

$$\underset{(\alpha)}{\xi^\circ_0} = \underset{(\alpha)}{H(u)}\lambda^*(\xi, \bar{\xi}, u), \tag{4.5}$$

where $\underset{(\alpha)}{H(u)}$ is an undetermined function of u ($=x^1$).

Using (4.5) in the commutator expression (4.2a), we have

$$\underset{\alpha\beta}{c}{}^{\gamma}\, \underset{(\gamma)}{H(u)} = \frac{n^1}{\lambda^*}\left(\underset{(\alpha)}{H(u)}\frac{\partial}{\partial u}\underset{(\beta)}{H(u)} - \underset{(\beta)}{H(u)}\frac{\partial}{\partial u}\underset{(\alpha)}{H(u)}\right). \tag{4.6}$$

If the left-hand side is nonvanishing, then n^1/λ^* must be a function of u only and can be set equal to 1 using the freedom inherent in (3.19). This establishes a homomorphism of the Lie algebra of Killing vectors onto the one-dimensional subspace of the tangent space spanned by the vector $\partial/\partial u$.

If the lhs of (4.6) vanishes, then the functions $H(u)$ are constant multiples of each other. Both possibilities show that the Killing vectors may be chosen so that $H(u)=0$, $\alpha>1$. Therefore, for all but at most one
$_{(\alpha)}$
Killing vector

$$l_a \underset{(\alpha)}{\xi^a} = \underset{(\alpha)}{\xi^\circ_0} = 0. \tag{4.7}$$

This allows one to establish the following result:

If an R–T metric has three Killing vectors, then it must be nonradiative.

As in the preceeding paragraph, the Killing vectors may be chosen so that $\underset{(2)}{\xi^\circ_0} = \underset{(3)}{\xi^\circ_0} = 0$. Equation (2.42) then yields

$$\underset{(2)}{\xi^\circ_1}\kappa'^\circ + \overline{\underset{(2)}{\xi^\circ_1}}\,\overline{\kappa}'^\circ = 0, \tag{4.8a}$$

$$\underset{(3)}{\xi^\circ_1}\kappa'^\circ + \overline{\underset{(3)}{\xi^\circ_1}}\,\overline{\kappa}'^\circ = 0, \tag{4.8b}$$

so that either $\kappa'^\circ = 0$ or the determinant of coefficients vanishes. If the latter is true, then

$$\underset{(2)}{\xi^\circ_1} = c\,\underset{(3)}{\xi^\circ_1}, \quad \overline{\underset{(2)}{\xi^\circ_1}} = c\,\overline{\underset{(3)}{\xi^\circ_1}} \tag{4.9}$$

with c real. Operating with þ, $\tilde{\text{þ}}'$, $\tilde{\text{ð}}$, $\tilde{\text{ð}}'$, on (4.9) yields

$$\text{þ}c = \tilde{\text{þ}}'c = \tilde{\text{ð}}c = \tilde{\text{ð}}'c = 0. \tag{4.10}$$

Since c is a (0, 0) quantity, it must be a constant. But this cannot be, as the vectors $\underset{(2)}{\xi^a}$ and $\underset{(3)}{\xi^a}$ are linearly independent. Therefore, $\kappa'^\circ = 0$ and as before it follows that the metric is nonradiative.

This establishes the rule that a radiating vacuum EDPN metric can have at most two Killing vectors, and if it does, then it is a Robinson–Trautman metric.

5. THE SPECIFIC CASES

If a vacuum metric is restricted to be of a specific Petrov–Penrose type, then the number and type of possible Killing vectors is further limited. This problem was first investigated by Petrov,[5] who did not restrict himself to EDPN metrics. In this section the problem is reexamined, but with the additional restriction that the metrics be EDPN.

Type N: $\psi_2^\circ = \psi_3^\circ = 0$

One sees from Eq. (2.17) that if $\kappa'^\circ = 0$, $\Psi_4^\circ = 0$ and the space is flat, so that a type N space will not support a preferred Killing vector. There remains then the question of whether or not it can contain two nonpreferred Killing vectors. Collinson[12] has shown that if the degenerate principal null vector is twisting ($\Omega^\circ \neq 0$), there can be at most one Killing vector. This result follows trivially from Eq. (4.2c), i.e., if there are two Killing vectors then

$$\Omega^\circ\kappa'^\circ = 0.$$

But if the principal null vector is twisting, $\Omega^\circ \neq 0$, so that $\kappa'^\circ = 0$ and the space is flat.

If the principal null direction is nontwisting, then $\Omega^\circ = 0$ and Eq. (4.2c) is automatically satisfied so that it is possible for the metric to have more that one Killing vector. In Sec. 4 it was shown that if there exist three Killing vectors, then $\kappa'^\circ = 0$. But this cannot obtain for a type N metric; therefore, it can have at most two Killing vectors. Collinson and French[6] give two examples of such metrics.

Type $(3,1)$: $\psi_2^\circ = 0$

Assume that a type $\{3,1\}$ metric has r linearly independent Killing vectors $\underset{(\alpha)}{\xi^a}$ ($\alpha = 1, r$, $r>1$). Assume also that none of these are preferred. Then by Eq. (4.2c) there are two cases.

Case 1: $\Omega^\circ \neq 0$, $\kappa'^\circ = \psi_4^\circ = 0$

Using the field equations (2.18), (2.19), the expression (2.38) for the Lie derivatives of Ψ_3° are

$$\underset{(1)}{\xi^\circ_1}\tilde{\text{ð}}'\Psi_3^\circ + (\tilde{\text{ð}}'\underset{(1)}{\xi^\circ_1} - 2\tilde{\text{þ}}'\underset{(1)}{\xi^\circ_0})\Psi_3^\circ = 0, \tag{5.1a}$$

$$\underset{(2)}{\xi^\circ_1}\tilde{\text{ð}}'\Psi_3^\circ + (\tilde{\text{ð}}'\underset{(2)}{\xi^\circ_1} - 2\tilde{\text{þ}}'\underset{(2)}{\xi^\circ_0})\Psi_3^\circ = 0. \tag{5.1b}$$

If $\tilde{\text{ð}}'\Psi_3^\circ = 0$, then direct calculation of the commutator $[\tilde{\text{ð}}, \tilde{\text{ð}}']\Psi_3^\circ$ shows that $\Psi_3^\circ = 0$ as well. But this cannot be, therefore for consistancy the determinant of coefficients of (5.1) must vanish, i.e.,

$$\begin{vmatrix} \underset{(1)}{\xi^\circ_1} & (\tilde{\text{ð}}'\underset{(1)}{\xi^\circ_1} - 2\tilde{\text{þ}}'\underset{(1)}{\xi^\circ_0}) \\ \underset{(2)}{\xi^\circ_1} & (\tilde{\text{ð}}'\underset{(2)}{\xi^\circ_1} - 2\tilde{\text{þ}}'\underset{(2)}{\xi^\circ_0}) \end{vmatrix} = 0. \tag{5.2}$$

Operating on (5.2) with $\tilde{\text{þ}}'$, and adding the result to its complex conjugate yields

$$3(\overline{\rho}'^\circ - \rho'^\circ)\begin{vmatrix} \underset{(1)}{\xi^\circ_1} & \overline{\underset{(1)}{\xi^\circ_1}} \\ \underset{(2)}{\xi^\circ_1} & \overline{\underset{(2)}{\xi^\circ_1}} \end{vmatrix} - \begin{vmatrix} (\tilde{\text{ð}}'\underset{(1)}{\xi^\circ_1} + \tilde{\text{ð}}\,\overline{\underset{(1)}{\xi^\circ_1}}) & \tilde{\text{þ}}'\underset{(1)}{\xi^\circ_0} \\ (\tilde{\text{ð}}'\underset{(2)}{\xi^\circ_1} + \tilde{\text{ð}}\,\overline{\underset{(2)}{\xi^\circ_1}}) & \tilde{\text{þ}}'\underset{(2)}{\xi^\circ_0} \end{vmatrix} = 0. \tag{5.3}$$

The second determinant vanishes by virtue of Eq. (2.34) and the first determinant vanishes only if $\underset{(1)}{\xi^\circ_1} = c\,\underset{(2)}{\xi^\circ_1}$, c real. As shown at the end of Sec. 4, this means that c is a real constant. Therefore, either the Killing vectors may be chosen so that one of them is preferred or the determinant is nonvanishing and

$$\rho'^\circ = \overline{\rho}'^\circ. \tag{5.4}$$

Given Eq. (5.4), it is a trivial matter to show that there is a preferred Killing vector. Since $\kappa'^\circ = 0$ as well, there exists a (1, 1) object θ_0° such that[3]

$$\text{þ}\theta_0^\circ = \text{þ}'\theta_0^\circ = \tilde{\text{ð}}\theta_0^\circ = \tilde{\text{ð}}'\theta_0^\circ = 0. \tag{5.5}$$

Define a vector θ^a by

$$\theta^a = \theta_0^\circ n^a - \rho'^\circ\theta_0^\circ l^a. \tag{5.6}$$

Equations (2.31)–(2.34) are satisfied by this vector, and conditions (2.35)–(2.37) are identically fulfilled. Therefore, the vector defined by (5.6) is a preferred Killing vector.

The case $\rho'^\circ = \overline{\rho}'^\circ$ has produced a preferred Killing vector, but it has not eliminated the possibility of two or more nonpreferred. To put a limit on these, return to the matrix Eq. (5.2) in the form

$$\underset{(1)}{\xi^\circ_1} = c\,\underset{(2)}{\xi^\circ_1}, \tag{5.7a}$$

$$\tilde{\text{ð}}'\underset{(1)}{\xi^\circ_1} - 2\tilde{\text{þ}}'\underset{(1)}{\xi^\circ_0} = c(\tilde{\text{ð}}'\underset{(2)}{\xi^\circ_1} - 2\tilde{\text{þ}}'\underset{(2)}{\xi^\circ_0}). \tag{5.7b}$$

Operating on (5.7a) with þ, $\tilde{\text{þ}}'$, $\tilde{\delta}$ and on (5.7b) with $\tilde{\delta}'$ reproduces Eq. (5.5), and c must be a constant. If c is real, then $\underset{(1)}{\xi^a}$ and $\underset{(2)}{\xi^a}$ essentially differ by a preferred Killing vector, and so we assume that c is complex. In that case, $\underset{(1)}{\xi^a}$ and $\underset{(2)}{\xi^a}$ may be chosen so that

$$\underset{(1)}{\xi_1^\circ} = i\,\underset{(2)}{\xi_1^\circ} \tag{5.8}$$

Writing out equation (2.34) for $\underset{(1)}{\xi^a}$ and $\underset{(2)}{\xi^a}$,

$$\tilde{\delta}'\,\underset{(1)}{\xi_1^\circ} + \tilde{\delta}\,\underset{(1)}{\bar{\xi}_1^\circ} = -2\tilde{\text{þ}}'\,\underset{(1)}{\xi_0^\circ}, \tag{5.9a}$$

$$\tilde{\delta}'\,\underset{(2)}{\xi_1^\circ} + \tilde{\delta}\,\underset{(2)}{\bar{\xi}_1^\circ} = -2\tilde{\text{þ}}'\,\underset{(2)}{\xi_0^\circ}. \tag{5.9b}$$

Multiplying (5.9a) by i and using (5.8) in (5.9b), we have

$$\tilde{\delta}'\,\underset{(1)}{\xi_1^\circ} = -\tilde{\text{þ}}'\,\underset{(1)}{\xi_0^\circ} + i\tilde{\text{þ}}'\,\underset{(2)}{\xi_0^\circ}. \tag{5.10}$$

Operating on this equation with $\tilde{\delta}$, and using commutator (2.27)],

$$\rho'^\circ\,\underset{(1)}{\xi_1^\circ} = 0. \tag{5.11}$$

But this is a contradiction, for $\rho'^\circ = 0 \Rightarrow \Psi_3^\circ = 0$. Therefore:

A type $\{3,1\}$ EDPN metric with $\Omega^\circ \neq 0$ can have at most two Killing vectors, and if it has the maximum number, then one must be preferred.

An example of such a metric has been given by Robinson.[13]

Case 2: $\Omega^\circ = 0$ (Robinson-Trautman)

There are no restrictions on the existence of two nonpreferred Killing vectors, and Collinson and French[6] have given an example. If one looks at the possibility of three Killing vectors, one finds that such a metric can exist provided that one of the Killing vectors is preferred.

If there exist three Killing vectors, then $\kappa'^\circ = 0$ (cf. Sec. 4) and a preferred Killing vector may be constructed as above. The general expression for a type $\{3,1\}$ Robinson–Trautman metric with three Killing vectors also appears in Collinson and French[6] and one of the Killing vectors is indeed preferred.

Types {2,1,1} and *D*

A type $\{2,1,1\}$ or type D metric is free to have two Killing vectors with no restrictions beyond those imposed by the results of Sec. 4. However, if one asks that the metric have three Killing vectors, one finds that not only must the metric be type D, but that there also exists a fourth Killing vector. The proof of this proceeds as follows.

Let a space be type $\{2,1,1\}$ ($\Psi_2^\circ \neq 0$) and let it contain three Killing vectors $\underset{(\alpha)}{\xi^a}$ ($\alpha = 1,3$). The first step is to show that

$$\kappa'^\circ = \rho'^\circ - \bar{\rho}'^\circ = 0. \tag{5.12}$$

The first part ($\kappa'^\circ = 0$) has been proved in Sec. 4, and if for any $\underset{(\alpha)}{\xi^a}$, $\tilde{\text{þ}}'\,\underset{(\alpha)}{\xi_0^\circ} = 0$, the second part follows from Eq. (2.42). Therefore, it may be assumed that $\tilde{\text{þ}}'\,\underset{(\alpha)}{\xi_0^\circ} \neq 0$ for any $\underset{(\alpha)}{\xi^a}$. Writing the Lie derivative of Ψ_2° using Eq. (2.36),

$$3\tilde{\text{þ}}'\,\underset{(\alpha)}{\xi_0^\circ}\Psi_2^\circ + \underset{(\alpha)}{\xi_0^\circ}\tilde{\text{þ}}'\Psi_2^\circ - \underset{(\alpha)}{\xi_1^\circ}\tilde{\delta}'\Psi_2^\circ = 0 \quad (\alpha = 1,3). \tag{5.13}$$

Since $\Psi_2^\circ \neq 0$, for consistancy the determinant of the matrix of coefficients of Ψ_2°, $\tilde{\delta}'\Psi_2^\circ$, $\tilde{\text{þ}}'\Psi_2^\circ$ must vanish. Operating on this determinant with $\tilde{\delta}'$, adding the result to its complex conjugate, and using (2.34), we have

$$(\bar{\rho}'^\circ - \rho'^\circ)\left|\underset{(\alpha)}{\xi_0^\circ}\ \underset{(\alpha)}{\bar{\xi}_1^\circ}\ \underset{(\alpha)}{\xi_0^\circ}\right| + \Omega^\circ\left|\tilde{\text{þ}}'\,\underset{(\alpha)}{\xi_0^\circ}\ \underset{(\alpha)}{\xi_1^\circ}\right| = 0, \tag{5.14}$$

where the straight brackets represent the square determinant formed from $\alpha = 1,3$.

Operating on (5.14) with $\tilde{\text{þ}}'$, one obtains

$$2(\bar{\rho}'^\circ - \rho'^\circ)\left|\tilde{\text{þ}}'\,\underset{(\alpha)}{\xi_0^\circ}\ \underset{(\alpha)}{\bar{\xi}_1^\circ}\ \underset{(\alpha)}{\xi_1^\circ}\right| = 0 \tag{5.15}$$

so that either $\bar{\rho}'^\circ = \rho'^\circ$ as required, or

$$\left|\tilde{\text{þ}}'\,\underset{(\alpha)}{\xi_0^\circ}\ \underset{(\alpha)}{\bar{\xi}_1^\circ}\ \underset{(\alpha)}{\xi_1^\circ}\right| = 0. \tag{5.16}$$

If the latter, then using (5.16) in Eq. (5.14), together with the fact that the determinant $\left|\underset{(\alpha)}{\xi_0^\circ}\ \underset{(\alpha)}{\bar{\xi}_1^\circ}\ \underset{(\alpha)}{\xi_1^\circ}\right|$ is nonvanishing yields the required $\bar{\rho}'^\circ = \rho'^\circ$.

From Eqs. (2.41)–(2.43) it follows that the (0,0) objects $\tilde{\text{þ}}'\,\underset{(\alpha)}{\xi_0^\circ}$ are all constants. Therefore, the Killing vectors may be chosen so that $\tilde{\text{þ}}'\,\underset{(2)}{\xi_0^\circ} = \tilde{\text{þ}}'\,\underset{(3)}{\xi_0^\circ} = 0$. Equation (2.36) gives

$$\underset{(2)}{\xi_0^\circ}\tilde{\text{þ}}'\Psi_2^\circ + \underset{(2)}{\xi_1^\circ}\tilde{\delta}'\Psi_2^\circ = 0, \tag{5.17a}$$

$$\underset{(3)}{\xi_0^\circ}\tilde{\text{þ}}'\Psi_2^\circ + \underset{(3)}{\xi_1^\circ}\tilde{\delta}'\Psi_2^\circ = 0, \tag{5.17b}$$

so that either

$$\tilde{\text{þ}}'\Psi_2^\circ = \tilde{\delta}'\Psi_2^\circ = 0 \tag{5.18}$$

or

$$\underset{(2)}{\xi_0^\circ} = c\,\underset{(3)}{\xi_0^\circ}, \quad \underset{(2)}{\xi_1^\circ} = c\,\underset{(3)}{\xi_1^\circ}. \tag{5.19}$$

with c a real constant.

This second possibility is excluded by the linear independence of $\underset{(2)}{\xi^a}$ and $\underset{(3)}{\xi^a}$, and so (5.18) is proved

From Eq. (2.36), with the use of (5.18) one sees that $\tilde{\text{þ}}'\,\underset{(1)}{\xi_0^\circ} = 0$ as well. Since $\underset{(\alpha)}{\xi_1^\circ} = 0$ for at most one Killing vector, one can now choose the Killing vectors such that $\underset{(1)}{\xi_1^\circ}$, $\underset{(2)}{\xi_1^\circ} \neq 0$. Equation (2.35) yields

$$\underset{(1)}{\xi_1^\circ}\tilde{\delta}'\rho'^\circ + \underset{(1)}{\bar{\xi}_1^\circ}\tilde{\delta}\rho'^\circ = 0, \tag{5.20a}$$

$$\underset{(2)}{\xi_1^\circ}\tilde{\delta}'\rho'^\circ + \underset{(2)}{\bar{\xi}_1^\circ}\tilde{\delta}\rho'^\circ = 0. \tag{5.20b}$$

Since $\underset{(1)}{\xi^a}$ and $\underset{(2)}{\xi^a}$ are linearly independent nonpreferred Killing vectors, it follows that

$$\tilde{\delta}'\rho'^\circ = 0. \tag{5.21}$$

This in turn implies [(2.15)]

$$\Psi_3^\circ = 0. \tag{5.22}$$

Similarly from Eq. (2.37), $\tilde{\delta}'\Omega^\circ = 0$, and finally from (2.15), $\Psi_4^\circ = 0$.

It has now been established that the existence of three Killing vectors requires

$$\Psi_4^\circ = \Psi_3^\circ = \tilde{\delta}'\psi_2^\circ = \tilde{\delta}'\Omega^\circ = 0. \tag{5.23}$$

A glance at Eqs. (2.13) and (2.14) suffices to show that (3.23) is a sufficient condition that the metric is type D and the first part of the proof is complete.

One could in fact demonstrate the existence of a fourth Killing vector by a lengthy calculation; however, it is easier to cheat and look in Kinnersley's catalog of type D metrics,[14] where one notices that type D metrics have either two or four Killing vectors. This completes the proof.

CONCLUSION

An extensive analysis of vacuum EDPN metrics containing Killing vectors has been carried out. It was shown that the number of possible Killing vectors is highly dependent upon the Petrov–Penrose classification. It was also shown that with three Killing vectors a vacuum EDPN metric is necessarily nonradiative and that if it has two Killing vectors and is radiative then it must be a Robinson–Trautman metric. This leads to the result that a radiating vacuum EDPN metric can only have cylindrical symmetry if it is Robinson–Trautman.

Also worthy of comment is the fact that there is slightly more freedom in a type $\{3,1\}$ metric than in a $\{2,1,1\}$ metric in that that the former may contain three Killing vectors and the latter but two.

The philosophy behind the analysis was to avoid the use of coordinates to as great an extent as possible. While it is unlikely that solutions to essential physical problems lie in the as yet undiscovered vacuum EDPN metrics, it is felt that the coordinate free approach used is worthy of further investigation in the hope that its philosophy, though probably not its mechanics, may be applied to the analysis of algebraically general metrics. A thorough understanding of the algebraically special solutions must necessarily be the first step in such a program.

[1]Minimal differential equations (MDE) refers to what is left of the Einstein equations after all the imposed geometric restrictions have been built in and all the coordinate and gauge freedom has been used in their simplification. An example of an MDE is the Robinson–Trautman equation.[11]

[2]A. Held, Comm. Math. Phys. **37**, 311 (1974).
[3]A. Held, Comm. Math. Phys. (to be published).
[4]A. Held, General Relativity Gravitation (to be published).
[5]A. Z. Petrov *Einstein Spaces* (Pergamon, London, 1969).
[6]C. D. Collinson and D. C. French, J. Math. Phys. **8**, 701 (1967).
[7]R. Geroch, A. Held, and R. Penrose, J. Math. Phys. **14**, 874 (1973).
[8]E. T. Newman and R. Penrose, J. Math. Phys. **3**, 566 (1962).
[9]E. T. Newman and R. Penrose, J. Math. Phys. **7**, 863 (1966).
[10]J. N. Goldberg *et al.*, J. Math. Phys. **8**, 2155 (1967).
[11]I. Robinson and A. Trautman, Proc. Roy. Soc. (London) A **265**, 463 (1962).
[12]C. D. Collinson, J. Phys. A **2**, 621 (1969).
[13]I. Robinson, General Relativity Gravitation **6**, (to be published).
[14]W. Kinnersley, J. Math. Phys. **10**, 1195 (1969).

Interpretation of Kato's invariance principle in scattering theory

John D. Dollard

Department of Mathematics, University of Texas at Austin, Austin, Texas 78712
(Received 14 April 1975)

A simple proof is given that Kato's invariance principle holds for a class of generalized piecewise linear (GPL) functions, under the sole assumption of existence of the Møller wave operators. The invariance principle for GPL functions is viewed as an expression of our freedom to change the scale of time and shift the zero point of energy. It is remarked that whenever scattering theory can be done, it can be done with bounded Hamiltonians. The motion of a particle is studied when its Hamiltonian is replaced by a GPL function of this Hamiltonian.

INTRODUCTION

Let H_0 and H denote self-adjoint operators on a Hilbert space $\mathcal{H}$. (In this paper, H_0 and H will be thought of as representing respectively the free and interacting Hamiltonians for a quantum-mechanical system. We take $\hbar = 1$.) Let $W_\pm(H, H_0)$ denote the Møller wave operators (if any) determined by H and H_0. That is,

$$W_\pm(H, H_0) = s\text{-}\lim_{t \to \pm\infty} \exp(iHt) \exp(-iH_0 t) \tag{1}$$

provided the indicated limits exist. [In some scattering problems, it is necessary to define $W_\pm(H, H_0)$ as the limits of $\exp(iHt)\exp(-iH_0 t)P$, where P is a certain projection operator which commutes with $\exp(-iH_0 t)$. Equation (1) and all other equations in this paper are written for the case in which P is the identity operator. Our discussion applies with easy modifications to cases in which P is not the identity operator.] Kato's invariance principle[1,2] states that if ϕ is a special kind of function then under certain circumstances it is possible to replace H and H_0 in Eq. (1) by operators $\phi(H)$ and $\phi(H_0)$, and obtain the *same* Møller wave operators. That is, the limits

$$W_\pm(\phi(H), \phi(H_0)) = s\text{-}\lim_{t \to \pm\infty} \exp[i\phi(H)t] \exp[-i\phi(H_0)t] \tag{2}$$

exist, and

$$W_\pm(\phi(H), \phi(H_0)) = W_\pm(H, H_0). \tag{3}$$

The type of function ϕ allowed in this invariance principle has been somewhat generalized since Kato's discovery, (see, for example, Ref. 3). However, the sort of condition imposed is well illustrated by Kato's original requirements: We will say that ϕ is a *Kato function* if ϕ is a real-valued function on $(-\infty, \infty)$ with the properties that the interval $(-\infty, \infty)$ can be divided up into a finite number of subintervals such that on each open subinterval ϕ is differentiable with ϕ' continuous, locally of bounded variation, and positive.

Kato first proved the above invariance principle (and more) for the case when ϕ is a Kato function and $H - H_0$ is a trace-class operator. The principle has since been proved for similar classes of functions ϕ under much weaker assumptions on H and H_0 (Refs. 2–7, and references therein). The purpose of this paper is to give an interpretation of the invariance principle and to prove it under the sole assumption that the limit in Eq. (1) exists, but for a different class of functions ϕ. As will be seen, our class of functions is uniformly dense (and more) in the set of Kato functions.

I. INTERPRETATION OF THE INVARIANCE PRINCIPLE

We assume that the limit in Eq. (1) exists, and think of the operators $W_\pm(H, H_0)$ as the Møller wave operators for a scattering process. We consider two kinds of change in our description of the scattering theory:

(A) *Change of the zero point from which energies are measured*: This is performed by substituting $H_0 + b$ and $H + b$ for H_0 and H, respectively, where b is a real number.

(B) *Change of the time scale*: This is performed by substituting ct for t, where c is a positive real number.

Changes of type (A) are physically trivial. Changes of type (B) modify the statements we make about finite times but not those about infinite times, because $ct \to \pm\infty$ if and only if $t \to \pm\infty$. Thus the Møller wave operators in the new description must be identical with those in the old. Indeed we have

$$\exp[i(H + b)ct] \exp[-i(H_0 + b)ct] = \exp(iHct) \exp(-iH_0 ct), \tag{4}$$

and when $t \to \pm\infty$ also $ct \to \pm\infty$, so that the right-hand side of Eq. (4) converges to the operator $W_\pm(H, H_0)$ of Eq. (1).

If we do not wish to emphasize the origin of the changes discussed above in our scattering theory, we can simply say that they arose as follows: In the expression $\exp(iHt)\exp(-iH_0 t)$, substitute $\phi(H_0)$ and $\phi(H)$ for H_0 and H respectively, where ϕ is the linear function on $(-\infty, \infty)$ defined by

$$\phi(\lambda) = c\lambda + cb. \tag{5}$$

Thus our statements about changing the description of the theory lead to a simple instance of the invariance principle: The Møller wave operators do not change if H_0 and H are replaced by $\phi(H_0)$ and $\phi(H)$, where ϕ is a real-valued linear function on $(-\infty, \infty)$, with positive derivative. For future reference we note that if ϕ is written as

$$\phi(\lambda) = c\lambda + d, \tag{6}$$

then this ϕ corresponds to the process of replacing t by ct and H_0 and H by $H_0 + d/c$ and $H + d/c$, respectively.

We now generalize these remarks. We first wish to study wavefunctions ψ that have kinetic energy restricted to some interval $(E_1, E_2]$. We will say that ψ has kinetic energy in $(E_1, E_2]$ if and only if ψ belongs to the subspace of $\mathcal{H}$ corresponding to the portion of the spectrum of H_0 in the interval $(E_1, E_2]$. The projection on this subspace is denoted by P_{E_1,E_2}, so that the subspace itself is $P_{E_1,E_2}\mathcal{H}$. [P_{E_1,E_2} can be written as a function of H_0 using the usual functional calculus for self-adjoint operators: If χ_{E_1,E_2} is the characteristic function of the interval $(E_1, E_2]$, then P_{E_1,E_2} is the same as $\chi_{E_1,E_2}(H_0)$.] If ψ is in the subspace $P_{E_1,E_2}\mathcal{H}$, then $\exp(-iH_0t)\psi$ is also in it, whatever the value of t. For this reason, when considering evolution of wavefunctions according to the free propagator $\exp(-iH_0t)$, we may, if desired, isolate and study separately the wavefunctions which have kinetic energy in $(E_1, E_2]$. When studying these wavefunctions, we are at liberty to make a shift of energy zero-point and change of time-scale. This would change the propagator from $\exp(-iH_0t)$ to $\exp[-i\phi(H_0)t]$, where ϕ is a certain linear function, as described above. (Note: We do *not* intend to change the class of wavefunctions studied. The instructions are: For $\psi \in P_{E_1,E_2}\mathcal{H}$, replace $\exp(-iH_0t)\psi$ by $\exp[-i\phi(H_0)t]\psi$.) We can extend this discussion as follows: Imagine splitting the real line up into a (possibly infinite) number of disjoint intervals. Each such interval determines a subspace of $\mathcal{H}$ (namely, wavefunctions with kinetic energy in that interval) and we could study evolution under the free propagator by isolating the subspaces and studying the evolution separately in each subspace. Because the evolutions in the various subspaces do not "interfere" with each other, there is nothing to prevent our making a different choice of time-scale and zero-point energy in each subspace (whimsical though this might seem). The net effect would be to replace H_0 in the propagator by various different linear functions of H_0 on different subspaces of $\mathcal{H}$. More generally, we could contemplate decomposing the real line into sets B_n other than intervals, and carrying out the above procedure. $\mathcal{H}$ would be split up into subspaces, each subspace consisting of wavefunctions with kinetic energy in one of the sets B_n, and on these various different subspaces we would be replacing H_0 in the propagator by various different linear functions of H_0. We will now see that such a procedure can be summarized by saying that the net effect is this: *On the entire Hilbert space* $\mathcal{H}$, we have replaced H_0 in the propagator by the operator $\phi(H_0)$, where ϕ is one of the "generalized piecewise linear" functions described below. For technical reasons, we ask that the sets B_n into which we decompose the real line are Borel sets.

Definition: ϕ is a *generalized piecewise linear (GPL) function* if and only if ϕ is a real-valued function on $(-\infty, \infty)$ with the representation

$$\phi(\lambda) = \sum_{n=0}^{\infty} (c_n\lambda + d_n)\chi_{B_n}(\lambda), \tag{7}$$

where c_n and d_n are real numbers, with $c_n > 0$ for all n, the B_n are pairwise disjoint Borel sets whose union is $(-\infty, \infty)$, and χ_{B_n} is the characteristic function of B_n:

$$\chi_{B_n}(\lambda) = \begin{cases} 1 & \text{if } \lambda \in B_n \\ 0 & \text{otherwise} \end{cases}. \tag{8}$$

Clearly, if ϕ has the representation (7), then ϕ is a linear function on each of the sets B_n, but the linear functions on different sets are different. If each B_n is an interval, then ϕ is piecewise linear in the usual sense. If we form $\phi(H_0)$ according to the functional calculus for operators, then $\chi_{B_n}(\lambda)$ becomes the projection $\chi_{B_n}(H_0)$ onto the subspace of $\mathcal{H}$ consisting of wavefunctions with kinetic energy lying in the set B_n. On this portion of the Hilbert space, $\phi(H_0)$ takes the value $c_nH_0 + d_n$; the propagator $\exp[-i\phi(H_0)t]$ appears as $\exp[-i(c_nH_0 + d_n)t]$, corresponding to the shift of zero-point energy which takes H_0 to $H_0 + d_n/c_n$ and the change of time-scale which takes t to c_nt. Thus if ϕ is a GPL function, replacement of H_0 by $\phi(H_0)$ in the free propagator has the effect we have described above, and can be considered as representing a different adjustment of time-scale and energy zero point in different subspaces of $\mathcal{H}$. [As a technical point relevant in the comparison of our results to Kato's results, we note that if ϕ is a GPL function then the absolutely continuous subspaces for H_0 and $\phi(H_0)$ are identical. The proof is elementary.]

We can carry out an entirely similar discussion for the full Hamiltonian H. We say that a wavefunction ψ has *total* energy in the set B_n if and only if ψ lies in the subspace of $\mathcal{H}$ corresponding to the portion of the spectrum of H lying in the set B_n. The projection on this subspace is $\chi_{B_n}(H)$, where B_n is as above. If ϕ is a GPL function, the effect of replacing $\exp(-iHt)$ by $\exp[-i\phi(H)t]$ would be to replace H in the propagator by various different linear functions of H on various subspaces of $\mathcal{H}$, corresponding again to a change of time-scale and energy zero point. An important point is that the subspaces this time would not be the same as those for the operator H_0, because $\chi_{B_n}(H)$ projects on the set of wavefunctions with *total* energy in B_n whereas $\chi_{B_n}(H_0)$ projects on the set of wavefunctions with *kinetic* energy in B_n.

The processes of passing in the propagators from H_0 to $\phi(H_0)$ and from H to $\phi(H)$ have this in common: In each case one is dealing with an energy operator, and selectively makes time-scale and energy zero-point adjustments in the portions of the theory concerning values of the energy operator in various sets. In both cases, when dealing with energies in B_n, we have changed the time-scale by c_n and shifted the energy zero point by d_n/c_n. In this sense, we have done the "same" thing to both energy operators. Our point now is this: If the limit $W_\pm(H, H_0)$ of Eq. (1) exists, and ϕ is a GPL function, then

$$\text{s-}\lim_{t \to \pm\infty} \exp[i\phi(H)t]\exp[-i\phi(H_0)t] = W_\pm(H, H_0). \tag{9}$$

We give a formal proof in the next section. However, the physical reason for Eq. (9) is easy to state: Equation (9) is a consequence of the conservation of energy. To understand this, imagine taking a wavefunction ψ with kinetic energy in the set B_n. Then $\exp[-i\phi(H_0)t]\psi$ is the same as $\exp[-i(c_nH_0 + d_n)t]\psi$. Abbreviating $W_\pm(H, H_0)$ to $W_\pm$, we have the following: Because the

limits in Eq. (1) are assumed to exist, the wavefunction $\exp[-i(c_nH_0+d_n)t]\psi$ will for $t\to\pm\infty$ asymptotically agree with $\exp[-i(c_nH+d_n)t]W_\pm\psi$. But by conservation of energy, $W_\pm\psi$ has *total* energy in the set B_n, so that $\exp[-i(c_nH+d_n)t]W_\pm\psi$ is the same as $\exp[-i\phi(H)t]W_\pm\psi$. Thus $\exp[-i\phi(H_0)t]\psi$ will asymptotically agree with $\exp[-i\phi(H)t]W_\pm\psi$, and in a standard way this implies that Eq. (9) holds at least when both sides of (9) act on a wavefunction with kinetic energy in the set B_n. Combining these results for all the sets B_n, we find that Eq. (9) holds on the entire Hilbert space. The invariance principle (9) is an expression of our freedom to adjust the scale of time and the zero point of energy. Although we used different subspaces when dealing with H_0 and H, our treatment of both was inspired by the same thought, and in the asymptotic limit our changes in these operators compensate to yield the original Møller wave operators.

II. INVARIANCE FOR GENERALIZED PIECEWISE LINEAR FUNCTIONS

Theorem 1: Suppose that the limits $W_\pm(H,H_0)$ indicated in Eq. (1) exist. Let ϕ be a GPL function. Then

$$\text{s-}\lim_{t\to\pm\infty}\exp[i\phi(H)t]\exp[-i\phi(H_0)t]=W_\pm(H,H_0). \tag{10}$$

Proof: We abbreviate $W_\pm(H,H_0)$ to $W_\pm$. Because $W_\pm$ satisfies the intertwining relations

$$\exp(iHt)\,W_\pm=W_\pm\exp(iH_0t), \tag{11}$$

we have

$$W_\pm-\exp(iHt)\exp(-iH_0t)=\exp(iHt)(W_\pm-1)\exp(-iH_0t). \tag{12}$$

Because the left-hand side of Eq. (12) converges to zero as $t\to\pm\infty$ and $\exp(iHt)$ is unitary, we therefore have

$$\text{s-}\lim_{t\to\pm\infty}(W_\pm-1)\exp(-iH_0t)=0. \tag{13}$$

Now Eq. (11) immediately implies that

$$\exp[i\phi(H)t]W_\pm=W_\pm\exp[i\phi(H_0)t], \tag{14}$$

and a calculation analogous to that of Eq. (12) shows that $\exp[i\phi(H)t)\exp[-i\phi(H_0)t]$ will converge to $W_\pm$ as $t\to\pm\infty$ if and only if

$$\text{s-}\lim_{t\to\pm\infty}(W_\pm-1)\exp[-i\phi(H_0)t]=0. \tag{15}$$

We now deduce Eq. (15) from Eq. (13). Let B_n be the Borel sets in the representation (7) for ϕ, and let P_n be the corresponding projection operators defined using H_0:

$$P_n=\chi_{B_n}(H_0) \tag{16}$$

Because the union of the sets B_n is $(-\infty,\infty)$, we have

$$\sum_{n=0}^{\infty}P_n=I, \tag{17}$$

where I denotes the identity operator on $\mathcal{H}$. For this reason, an elementary calculation shows that Eq. (15) will hold if and only if we have for each n the condition

$$\text{s-}\lim_{t\to\pm\infty}(W_\pm-1)\exp[-i\phi(H_0)t]P_n=0. \tag{18}$$

But we have

$$\exp[-i\phi(H_0)t]P_n=\exp[-i(c_nH_0+d_n)t]P_n\quad(c_n>0), \tag{19}$$

so that (18) is an obvious consequence of (13). This completes the proof.

Theorem 1 is interesting for several reasons:

(i) It holds under the sole assumption of existence of the Møller wave operators $W_\pm(H,H_0)$.

(ii) The character of the functions ϕ involved is somewhat different from that of the Kato functions mentioned earlier. A GPL function can be nondifferentiable in an essential way (i.e., so that differentiability almost everywhere cannot be achieved by correction on a set of measure zero), and need not be increasing on any interval. By remark (i), in any case in which the invariance principle has been proved for a certain type of function η, it can be generalized to composite functions of the type $\phi\circ\eta$, where ϕ is a GPL function.

(iii) If ψ is any Kato function and $\epsilon>0$, then there is a GPL function ϕ such that simultaneously ϕ approximates ψ within ϵ and the derivative ϕ' approximates ψ' within ϵ on the entire real line. Given this fact and the invariance principle proved for GPL functions in Theorem 1, it is perhaps not so surprising that an invariance principle can often be shown to hold for Kato functions, although it should be emphasized that mathematically the step from invariance for GPL functions to invariance for Kato functions is entirely nontrivial. The meaning of the invariance principle for Kato functions can be regarded as a generalization of its meaning for GPL functions, in which the adjustment of zero-point energy and time-scale, instead of being carried out in discrete steps on a countable number of sets, is carried out continuously along the spectrum of the energy operator. By analogy with the GPL case, if ψ is a Kato function, then writing

$$\psi(\lambda)=\psi'(\lambda)\lambda+d(\lambda), \tag{20}$$

we can think of the transition from H_0 to $\psi(H_0)$ as meaning that at the point λ of the spectrum of H_0, H_0 is being replaced by $H_0+d(\lambda)/\psi'(\lambda)$ and t is being replaced by $\psi'(\lambda)t$. Similar remarks apply to the transition from H to $\psi(H)$.

(iv) Since it is easy to think of bounded GPL functions, it follows from Theorem 1 that whenever scattering theory can be done, it can be done with bounded "Hamiltonians" $\phi(H_0)$ and $\phi(H)$. The author does not know whether or not a technical advantage can be gained from this remark. In any case, the remark provides a footnote to the usual statement that in most interesting quantum-mechanical problems one must deal with unbounded energy operators. We now briefly study a class of bounded GPL functions (selected mainly because of their conceptual simplicity) which correspond to a sort of repeated cutoff of the Hamiltonian.

Definition: For each $\mu>0$, we define ϕ_μ to be the GPL function such that

(i) $\phi_\mu(\lambda)=\lambda$ for $\lambda\in(-\mu,\mu]$

and

(ii) ϕ_μ is periodic with period 2μ.

The representation of ϕ_μ corresponding to (7) is

$$\phi_\mu(\lambda) = \sum_{n=-\infty}^{\infty} (\lambda - 2n\mu)\chi_{I_n}(\lambda), \tag{21}$$

where I_n is the interval $((2n-1)\mu, (2n+1)\mu]$. ϕ_μ corresponds to *no* adjustment of time-scale, only a repeated adjustment of energy zero point. Using these functions, it is easy to prove the following.

Theorem II: Let H_0 and H be self-adjoint operators on a Hilbert space $\mathcal{H}$. Then the operators $W_\pm(H, H_0)$ of Eq. (1) exist if and only if the limits

$$\text{s-}\lim_{t\to\pm\infty} \exp[i\phi_\mu(H)t]\exp[-i\phi_\mu(H_0)t] = W_\pm(\phi_\mu(H), \phi_\mu(H_0)) \tag{22}$$

exist for all $\mu > 0$ and are all the same. $W_\pm(H, H_0)$ is then equal to their common value.

Proof: Although the proof presents no difficulties, we give it for completeness: The "only if" part we have already proved. We now assume that all the limits in (22) exist and are equal, denoting their common value by $W_\pm$. Then for each $\mu > 0$ we have

$$\exp[i\phi_\mu(H)t]W_\pm = W_\pm \exp[i\phi_\mu(H_0)t] \tag{23}$$

and

$$\text{s-}\lim_{t\to\pm\infty} (W_\pm - 1)\exp[-i\phi_\mu(H_0)t] = 0. \tag{24}$$

Let P_μ denote the projection on the subspace of $\mathcal{H}$ corresponding to the portion of the spectrum of H_0 lying in the interval $(-\mu, \mu]$, and let Q_μ be the analogous projection for H and the interval $(-\mu, \mu]$. Then

$$\exp[i\phi_\mu(H_0)t]P_\mu = \exp(iH_0t)P_\mu \tag{25}$$

and

$$Q_\mu \exp[i\phi_\mu(H)t] = Q_\mu \exp(iHt). \tag{26}$$

Thus multiplying Eq. (23) on the left by Q_μ and on the right by P_μ, we have

$$Q_\mu \exp(iHt) W_\pm P_\mu = Q_\mu W_\pm \exp(iH_0t)P_\mu. \tag{27}$$

Now because P_μ and Q_μ approach the identity as $\mu \to \infty$, it easily follows that

$$\exp(iHt) W_\pm = W_\pm \exp(iH_0t), \tag{28}$$

so that to prove $\exp(iHt)\exp(-iH_0t)$ converges to $W_\pm$ we need only verify that Eq. (13) holds. To prove that Eq. (13) holds, it is enough to show that

$$\text{s-}\lim_{t\to\pm\infty} (W_\pm - 1)\exp(-iH_0t)P_\mu = 0 \quad \text{for all } \mu > 0. \tag{29}$$

But (29) is obvious because of (25) and (24). This completes the proof. Clearly, the proof actually depends only on the existence and equality of the limits in (22) for a sequence of subscripts μ tending to infinity.

Theorem II provides a partial converse of Theorem I. We remark that the general converse is false, as can be seen by fixing $\mu > 0$ and taking $H = H_0 + 2\mu$. Then $\phi_\mu(H)$ is the same as $\phi_\mu(H_0)$, so the limit in Eq. (22) exists for our one fixed μ. But of course the limit in Eq. (1) does not exist.

III. SCATTERING THEORY WITH HAMILTONIANS $\phi(H_0)$ AND $\phi(H)$

Let ϕ be a GPL function. In this section we study the consequences of accepting $\phi(H_0)$ and $\phi(H)$ literally as new Hamiltonians describing a physical scattering process. We will try to distinguish which of the following possibilities is correct:

Possibility I: $\phi(H_0)$ and $\phi(H)$ yield a scattering theory which describes physical happenings quite different from those in the scattering theory of H_0 and H. Nevertheless, the Møller wave operators for the two theories are the same.

Possibility II: The physical happenings in the scattering theory of $\phi(H_0)$ and $\phi(H)$ are essentially the same as those in the scattering theory of H_0 and H, and that is why the Møller wave operators for the two theories are the same.

Our interpretation of the change from H_0 and H to $\phi(H_0)$ and $\phi(H)$ is that it is physically trivial (as far as scattering theoretic statements go), at least on the subspaces of $\mathcal{H}$ on which $\phi(H_0)$ or $\phi(H)$ appear as linear functions of H_0 or H. This, of course, inclines us to possibility II, and it is this one which we shall defend. In order to discuss "physical happenings," we need a specific example of a scattering theory, which we take to be the quantum-mechanical potential scattering theory for a nonrelativistic spinless particle. We thus take for $\mathcal{H}$ the space $L^2(\mathbb{R}^3)$ of complex-valued square-integrable functions on three-dimensional Euclidean space, and we take for H_0 the (natural self-adjoint extension of the) operator $-\Delta/2m$ where Δ is the Laplacean and m the mass of the particle. H has the usual form

$$H = H_0 + V \tag{30}$$

where V is the potential energy operator for the particle. The "physical happening" to be discussed is the asymptotic motion of the particle as judged from the development in time of its position probability density (ppd).

We summarize some known facts about the ppd of a free particle in our theory.[9] Such a particle has a wavefunction ψ_t of the form

$$\psi_t = \exp(-iH_0t)f \tag{31}$$

with $f \in L^2(\mathbb{R}^3)$. The ppd determined by ψ_t is $|\psi_t(\mathbf{x})|^2$. We can make the following two statements (see Ref. 9 for more details):

(a) The ppd determined by ψ_t can asymptotically for large $|t|$ be replaced by $|m/t|^3 |\tilde{f}(m\mathbf{x}/t)|^2$, where $\tilde{f}$ denotes the Fourier transform of f. The function $|m/t|^3 \times |\tilde{f}(m\mathbf{x}/t)|^2$ is the same as the ppd for a classical free particle starting from the origin of coordinates at time $t = 0$, and having momentum probability density given by $|\tilde{f}(\mathbf{k})|^2$ (i.e., it has the same momentum probability density as the quantum mechanical particle under consideration).

(b) The probability $P(f, C)$ that the particle with wavefunction $\psi_t = \exp(-iH_0t)f$ will asymptotically for $t \to +\infty$

lie in a cone C with apex at the origin of coordinates is given by

$$P(f, C) = \int_C |\tilde{f}(\mathbf{k})|^2 d\mathbf{k}, \tag{32}$$

so that $P(f, C)$ is the probability that the momentum of the particle lies in C. (There is a similar statement for the case $t \to -\infty$.)

Statements (a) and (b) reinforce our belief in the naturalness of the quantum-mechanical description. Statement (a) is considerably stronger than statement (b), and implies statement (b) directly. In order to defend the idea that the physical happenings described by $\phi(H_0)$ are not much different from those described by H_0 (for large times), we will now show that if H_0 is replaced by $\phi_\mu(H_0)$, where ϕ_μ is as in Eq. (21), then (a) and (hence) (b) continue to hold, while for a general GPL function (and even for Kato functions η such that the derivative η' is polynomially bounded at infinity) we can still salvage (b). The reason we get better results for the functions ϕ_μ is that ϕ_μ represents no change in time-scale in the various subspaces of $\mathcal{H}$ corresponding to the intervals I_n of Eq. (21). Statement (a) depends explicitly on the parameter t, and in order to obtain such a result for an arbitrary $f \in \mathcal{H}$, one needs the same time-scale on the entire Hilbert space. [We could in fact prove that statement (a) continues to hold using any (possibly unbounded) GPL function ϕ such that, in Eq. (7), $c_n = 1$ for all n and the sets B_n are intervals. Again, this would represent no change in time-scale. The proof for such GPL functions closely resembles that given below for ϕ_μ. A simple generalization of statement (a) could be obtained by making the *same* change of time-scale on the *entire* Hilbert space.]

Proof of statement (a) with H_0 replaced by $\phi_\mu(H_0)$

Definition: f is a μ-function if and only if (i) f belongs to Schwartz' space $\mathcal{S}$ (Ref. 10) consisting of complex functions on $\mathbb{R}^3$ which together with all their derivatives vanish rapidly at large values of $|\mathbf{x}|$, and (ii) $\tilde{f}(\mathbf{k})$ vanishes in some shell around each of the spheres S_n in momentum space defined by

$$S_n = \left\{ \mathbf{k} : \frac{k^2}{2m} = (2n-1)\mu \right\}$$

for $n = 1, 2, \cdots$.

The point of the above definition is that if f is a μ-function then $\tilde{f}(\mathbf{k})$ vanishes near all the singularities of $\phi_\mu(k^2/2m)$, with the result that $\exp[-i\phi_\mu(H_0)t]f$ is also a μ-function, as is easily seen from the equation

$$\{\exp[-i\phi_\mu(H_0)t]f\}\tilde{}(\mathbf{k}) = \exp[-i\phi_\mu(k^2/2m)t]\tilde{f}(\mathbf{k}). \tag{34}$$

It is also easy to show that for any fixed μ, the μ-functions are dense in $L^2(\mathbb{R}^3)$. We now define for $n = 0, 1, 2, \cdots$ the functions

$$\chi_n(\mathbf{k}) = \begin{cases} 1 & \text{if } (2n-1)\mu < k^2/2m \leq (2n+1)\mu, \\ 0 & \text{otherwise} \end{cases} \tag{35}$$

so that χ_0 is the characteristic function of the portion of momentum space lying inside the sphere S_1, and each other χ_n is the characteristic function of the portion lying between the two spheres S_n and S_{n+1}. Then

$$\sum_{n=0}^{\infty} \chi_n(\mathbf{k}) = 1. \tag{36}$$

If f is a μ-function, we define f_n by

$$f_n(\mathbf{x}) = \frac{1}{(2\pi)^{3/2}} \int \exp(i\mathbf{k}\cdot\mathbf{x})\chi_n(\mathbf{k})\tilde{f}(\mathbf{k})\, d\mathbf{k}. \tag{37}$$

Note that

$$\tilde{f}_n(\mathbf{k}) = \chi_n(\mathbf{k})\tilde{f}(\mathbf{k}) \tag{38}$$

and

$$f(\mathbf{x}) = \sum_{n=0}^{\infty} f_n(\mathbf{x}). \tag{39}$$

Equation (39) holds both in the sense of pointwise convergence and in the strong sense. The f_n are clearly pairwise orthogonal, by Plancherel's theorem. We have

$$\begin{aligned} &\{\exp[-i\phi(H_0)t]f_n\}(\mathbf{x}) \\ &= \frac{1}{(2\pi)^{3/2}} \int \exp(i\mathbf{k}\cdot\mathbf{x}) \exp[-i\phi_\mu(k^2/2m)t]\chi_n(\mathbf{k})\tilde{f}(\mathbf{k})\, d\mathbf{k}) \\ &= \frac{1}{(2\pi)^{3/2}} \int \exp(i\mathbf{k}\cdot\mathbf{x}) \exp[-i(k^2/2m - 2n\mu)t]\chi_n(\mathbf{k})\tilde{f}(\mathbf{k})\, d\mathbf{k} \\ &= \{\exp[-i(H_0 - 2n\mu)t]f_n\}(\mathbf{x}). \end{aligned} \tag{40}$$

Thus

$$\begin{aligned} \exp[-i\phi(H_0)t]f &= \sum_{n=0}^{\infty} \exp[-i\phi_\mu(H_0)t]f_n \\ &= \sum_{n=0}^{\infty} \exp[-i(H_0 - 2n\mu)t]f_n. \end{aligned} \tag{41}$$

Now according to Ref. 9 we have

$$\exp(-iH_0t)f_n = C_tQ_tf_n \tag{42}$$

where the action of Q_t and c_t on $g \in L^2(\mathbb{R}^3)$ is given by

$$(Q_tg)(\mathbf{x}) = \exp(imx^2/2t)g(\mathbf{x}) \tag{43}$$

and

$$(C_tg)(\mathbf{x}) = \left(\frac{m}{it}\right)^{3/2} \exp(imx^2/2t)\tilde{g}\left(\frac{m\mathbf{x}}{t}\right). \tag{44}$$

Also, for any function $g \in L^2(\mathbb{R}^3)$, we have

$$\lim_{t\to\pm\infty} \|\exp(-iH_0t)g - C_tg\|^2 = 0. \tag{45}$$

An elementary estimate using the orthogonality of the f_n now shows that for the μ-function f we have

$$\lim_{t\to\pm\infty} \|\exp[-i\phi_\mu(H_0)t]f - \sum_{n=0}^{\infty} \exp(2n\mu it)C_tf_n\| = 0. \tag{46}$$

Now *a priori* the sum $\sum_{n=0}^{\infty} \exp(2n\mu it)c_tf_n$ is known to converge only in the strong sense. But because $(C_tf_n)(\mathbf{x})$ is proportional to $\tilde{f}_n(m\mathbf{x}/t)$, the terms in the sum are non-zero on different disjoint sets, so that the sum always reduces to one term at any point $\mathbf{x}$, and hence converges pointwise. Because of Eq. (46), the ppd determined by $\exp[-i\phi_\mu(H_0)t]f$ can asymptotically be replaced by the

ppd determined by $\sum_{n=0}^{\infty} \exp(2n\mu it) C_t f_n$. But this latter ppd is just

$$\left|\sum_{n=0}^{\infty} \exp(2n\mu it)\{C_t f_n\}(\mathbf{x})\right|^2 = \sum_{n=0}^{\infty} \left|\exp(2n\mu it)\{C_t f_n\}(\mathbf{x})\right|^2$$
$$= \sum_{n=0}^{\infty} \left|\{C_t f_n\}(\mathbf{x})\right|^2 = \left|\frac{m}{t}\right|^3 \sum_{n=0}^{\infty} \left|\tilde{f}_n\left(\frac{m\mathbf{x}}{t}\right)\right|^2$$
$$= \left|\frac{m}{t}\right|^3 \left|\tilde{f}\left(\frac{m\mathbf{x}}{t}\right)\right|^2,$$

so that we have proved statement (a) for the case when f is a μ-function. It is now not difficult to extend the result to any f in $L^2(\mathbb{R}^3)$, using the fact that the μ-functions are dense in $L^2(\mathbb{R}^3)$.

We thus see that the ppd for the wavefunction $\exp[-i\phi_\mu(H_0)t]f$ behaves asymptotically just like the ppd for the solution $\exp(-iH_0t)f$ of the free Schrödinger equation. If $W_\pm$ exists, then we can say something about wavefunctions propagating under the influence of the Hamiltonian $\phi_\mu(H)$ also. Namely, we have

$$\text{s-}\lim_{t\to\pm\infty} (\exp[-i\phi_\mu(H)t]W_\pm f - \exp[-i\phi_\mu(H_0)t]f) = 0 \quad (48)$$

and

$$\text{s-}\lim_{t\to\pm\infty} (\exp(-iHt)W_\pm f - \exp(-iH_0t)f) = 0. \quad (49)$$

On comparing these equations and using the fact that the ppd determined by $\exp[-i\phi_\mu(H_0)t]f$ and $\exp(-iH_0t)f$ are asymptotically the same, we see that the ppd's determined by $\exp[-i\phi_\mu(H)t]W_\pm f$ and $\exp(-iHt)W_\pm f$ are asymptotically the same. Thus asymptotically, at least, the motion of free particles and scattered interacting particles does not change much when H_0 and H are replaced by $\phi_\mu(H_0)$ and $\phi_\mu(H)$.

As remarked above, we can still salvage the scattering-into-cones formula (32) when we replace H_0 by $\phi(H_0)$, where ϕ is any GPL function. The proof is straightforward but tedious, and we do not give it. Further, using a slight modification of the proof of the theorem of Jauch, Lavine and Newton,[11] it is not difficult to show that if η is a Kato function whose derivative η' is polynomially bounded at infinity, then the formula (32) is still valid when H_0 is replaced by $\eta(H_0)$. These results offer a (somewhat weaker) corroboration of our thesis that the asymptotics of the theory are not substantially affected by replacing H_0 and H by (appropriate) functions of these operators.

[1] T. Kato, Pacific J. Math. 15, 171 (1965).
[2] T. Kato, *Perturbation Theory for Linear Operators* (Springer, New York, 1966).
[3] C. Chandler and A. G. Gibson, "Invariance Principle for Scattering with Long-Range (and other) Potentials" (to be published).
[4] J. A. Donaldson, A. G. Gibson, and R. Hersh, J. Funct. Anal. 14, 131 (1973).
[5] V. B. Matveev, Topics Math. Phys. 5, 77 (1972).
[6] V. B. Matveev, Theor. Math. Phys. 8, 663 (1971).
[7] L. A. Sakhnovich, Funct. Anal. Appl. 5, 49 (1971).
[8] J. D. Dollard, J. Math. Phys. 5, 729 (1964).
[9] J. D. Dollard, Commun. Math. Phys. 12, 193 (1969).
[10] L. Schwartz, *Théorie des distributions* (Hermann, Paris, 1951), Vol. II.
[11] J. M. Jauch, R. Lavine, and R. G. Newton, Helv. Phys. Acta 45, 325 (1972).

Weyl conform tensor of $\delta=2$ Tomimatsu–Sato spinning mass gravitational field

John E. Economou and Frederick J. Ernst

Department of Physics, Illinois Institute of Technology, Chicago, Illinois 60616
(Received 10 April 1975)

Extremely simple expressions are presented for the hitherto uncalculated invariants associated with the Weyl conform tensor of the $\delta=2$ Tomimatsu–Sato solution of Einstein's field equations.

The determination of the Weyl conform tensor can constitute a formidable calculational problem if one considers a spacetime such as one of those discovered[1] by Tomimatsu and Sato. In principle one can employ the formulas published[2] by Bardeen or Chandrasekhar and Friedman, as Tomimatsu and Sato themselves attempted to do.[3] The complexity of the problem is attested to by the fact that these authors settled for a study of a single term in one component of the Weyl tensor. We are happy to report that one can by an alternative procedure evaluate the *complete* Weyl tensor, and the result is far simpler than one might have anticipated.

I. WEYL TENSOR INVARIANTS

In the notation developed[4] by Hauser the invariant Riemann operator is defined by

$$\mathbb{R}=\tfrac{1}{2}\underline{e}^{\alpha}\wedge d^{2}\underline{e}_{\alpha}, \tag{1}$$

where the $\underline{e}_{\alpha}$ ($\alpha=1,2,3,4$) are basic tangent vectors. If e^{β} ($\beta=1,2,3,4$) designate those linear functionals (differential forms) such that

$$\underline{e}_{\alpha}e^{\beta}=\delta_{\alpha}^{\beta}\ , \tag{2}$$

then one can easily establish that

$$\mathbb{R}=\tfrac{1}{4}e^{\mu}e^{\nu}R_{\mu\nu}{}^{\alpha\beta}\underline{e}_{\alpha}\wedge\underline{e}_{\beta}. \tag{3}$$

(Note that differential forms are written to the *right* of vectors when they act as linear functionals upon those vectors, but are written to the *left* of vectors when a dyadic product is intended.)

Consider the duality operator

$$\mathbb{D}=\tfrac{1}{4}e^{\mu}e^{\nu}\epsilon_{\mu\nu\alpha\beta}\underline{e}^{\alpha}\wedge\underline{e}^{\beta}, \tag{4}$$

which has two threefold degenerate eigenvalues $\pm i$. From either of the three-dimensional subspaces select the *unique* eigenvector whose real part is equal to $-\rho^{-1}\underline{a}_{\phi}\wedge\underline{a}_{T}$, where $\underline{a}_{\phi}$ is the axial Killing vector, $\underline{a}_{T}$ is the temporal Killing vector, and ρ is the norm of $\underline{a}_{\phi}\wedge\underline{a}_{T}$. Designate this selected eigenvector of $\mathbb{D}$ by $\underline{B}_{0}$. By construction, the inner product of $\underline{B}_{0}$ with itself is given by $\underline{B}_{0}\rceil\underline{B}_{0}=-2$.

The eigenvector $\underline{B}_{0}$ of $\mathbb{D}$ may be augmented by two other eigenvectors $\underline{B}_{+}$ and $\underline{B}_{-}$ corresponding to the *same* eigenvalue as $\underline{B}_{0}$, chosen so that $\underline{B}_{0}\rceil\underline{B}_{+}=\underline{B}_{0}\rceil\underline{B}_{-}=\underline{B}_{+}\rceil\underline{B}_{+}=\underline{B}_{-}\rceil\underline{B}_{-}=0$ and $\underline{B}_{+}\rceil\underline{B}_{-}=1$. The fields $\underline{B}_{+}$ and $\underline{B}_{-}$ are determined up to a null rotation about $\underline{B}_{0}$ (and interchange $\underline{B}_{+}\leftrightarrow\underline{B}_{-}$). Therefore, the mere existence of two commuting Killing vectors singles out an almost completely specified basis for the bivector space. The freedom to perform null rotations about $\underline{B}_{0}$, under which $\underline{B}_{+}$ and $\underline{B}_{-}$ are multiplied by reciprocal fields and $\underline{B}_{0}$ remains unchanged, corresponds to the fact that one may perform Lorentz transformations along the ϕ direction and rotations about the ϕ direction.

If one decomposes the Riemann operator $\mathbb{R}$ into a part which commutes with the duality operator $\mathbb{D}$, and a part which anticommutes with it, the former part is essentially the Weyl conform part of the Riemann operator, while the latter is essentially the Ricci part. (The qualification may be deleted when the curvature scalar vanishes.) In general five complex fields are required to describe the Weyl conform part, but two of these vanish automatically in the case of a stationary axially symmetric field, when $\underline{B}_{+}$, $\underline{B}_{0}$, $\underline{B}_{-}$ are chosen in the manner described above. These five complex fields may be defined as follows:

$$C_{2}=B_{+}\lceil\mathbb{R}\rceil\underline{B}_{+}, \tag{5a}$$

$$C_{1}=-\tfrac{1}{2}B_{0}\lceil\mathbb{R}\rceil\underline{B}_{+}\ , \tag{5b}$$

$$C_{0}=\tfrac{1}{8}B_{0}\lceil\mathbb{R}\rceil\underline{B}_{0}+\tfrac{1}{2}B_{-}\lceil\mathbb{R}\rceil B_{+}\ , \tag{5c}$$

$$C_{-1}=-\tfrac{1}{2}B_{0}\lceil\mathbb{R}\rceil\underline{B}_{-}\ , \tag{5d}$$

$$C_{-2}=B_{-}\lceil\mathbb{R}\rceil\underline{B}_{-}\ . \tag{5e}$$

In Ref. 5 were presented the more complete decompositions of the 2-forms

$$\mathbb{R}\rceil\underline{B}_{+}=dv+vu, \tag{6a}$$

$$\mathbb{R}\rceil\underline{B}_{0}=du-2wv, \tag{6b}$$

$$\mathbb{R}\rceil\underline{B}_{-}=dw-wu\ , \tag{6c}$$

which can be seen to be compatible with the above definitions of the fields C_{s} ($s=2,\ldots,-2$).

Under a null rotation about $\underline{B}_{0}$ the field C_{s} transforms as a spin-weight s quantity. Hence C_{0} and $C_{2}C_{-2}$ are invariant under such null rotations. Consequently, in the case of a stationary axially symmetric spacetime, we may form from the Weyl tensor and the Killing vectors two complex scalar fields I_{1} and I_{2} such that if $\underline{B}_{+}$, $\underline{B}_{0}$, $\underline{B}_{-}$ are chosen in the way described earlier, then

$$C_{0}=I_{1}, \tag{7a}$$

$$C_{2}C_{-2}-9C_{0}^{2}=I_{2}\ . \tag{7b}$$

The fields C_{1} and C_{-1} vanish identically, so that they contain no additional information. In the case of a Petrov type D spacetime the invariant I_{2} vanishes. In the case of the Tomimatsu–Sato solutions neither I_{1} nor I_{2} vanishes.

II. TOMIMATSU-SATO $\delta = 2$ SOLUTION

Except for the Kerr metric, only three asymptotically flat stationary axially symmetric solutions of Einstein's equations have been discovered. These three were discovered by Tomimatsu and Sato (T—S), who employed the complex potential formalism described in Ref. 5.

The simplest T—S solution corresponds to the complex potential

$$\xi = N/D,$$

where in terms of symmetrical x—y coordinates.

$$N = p^2(x^4 - 1) + q^2(y^4 - 1) - 2ipq\,xy(x^2 - y^2), \tag{8a}$$

$$D = 2px(x^2 - 1) - 2iqy(1 - y^2). \tag{8b}$$

The constants p and q satisfy $p^2 + q^2 = 1$. From the complex potential

$$\mathcal{E} = (\xi - 1)/(\xi + 1) = (N - D)/(N + D) \tag{9}$$

the entire spacetime metric can be constructed by now well-known methods. The result can be expressed in the form

$$ds^2 = f^{-1}\left[P^2\left(\frac{dx^2}{x^2 - 1} + \frac{dy^2}{1 - y^2}\right) + \rho^2 d\phi^2\right] - f(dT - \omega\, d\phi)^2 , \tag{10}$$

where

$$f = (|N|^2 - |D|^2)/|N + D|^2, \tag{11}$$

$$P^2 = p^4(x^2 - y^2)^3/(|N|^2 - |D|^2) , \tag{12}$$

$$\rho^2 = (x^2 - 1)(1 - y^2). \tag{13}$$

III. CALCULATIONAL PROCEDURE

While in principle one should be able to evaluate I_1 and I_2 without specifying the bivector basis any more completely, our calculations were in fact based upon the specific null tetrad introduced in Ref. 5; namely,

$$\underline{k} = 2^{-1/2}[f^{1/2}\rho^{-1}(\underline{a}_\phi + \omega \underline{a}_T) + f^{-1/2}\underline{a}_T], \tag{14a}$$

$$\underline{m} = 2^{-1/2}[f^{1/2}\rho^{-1}(\underline{a}_\phi + \omega \underline{a}_T) - f^{-1/2}\underline{a}_T], \tag{14b}$$

$$\underline{t} = 2^{-1/2}f^{1/2}P[(x^2 - 1)^{1/2}\underline{a}_x + i(1 - y^2)^{1/2}\underline{a}_y] . \tag{14c}$$

The nonvanishing components of the Weyl tensor are then given by

$$C_2 = -\tfrac{1}{8}[2\eth\eth\mathcal{E} + f^{-1}(\eth\mathcal{E})^2], \tag{15a}$$

$$C_0 = -\tfrac{1}{8}[-2\eth^*\eth\mathcal{E} + f^{-1}(\eth^*\mathcal{E})(\eth\mathcal{E})] , \tag{15b}$$

$$C_{-2} = -\tfrac{1}{8}[2\eth^*\eth^*\mathcal{E} + f^{-1}(\eth^*\mathcal{E})^2], \tag{15c}$$

but the evaluation of C_0 is facilitated by employing the vacuum field equations to obtain the alternate equation

$$C_0 = v_k w_m - v_m w_k , \tag{16}$$

where the spin coefficients are given by

$$v_m = -2^{-3/2}f^{1/2}\rho^{-1}\eth\rho , \tag{17a}$$

$$w_k = -2^{-3/2}f^{1/2}\rho^{-1}\eth^*\rho , \tag{17b}$$

$$v_k = -2^{-3/2}[f^{1/2}\rho^{-1}\eth\rho - f^{-1/2}\eth\mathcal{E}] , \tag{17c}$$

$$w_m = -2^{-3/2}[f^{1/2}\rho^{-1}\eth^*\rho - f^{-1/2}\eth^*\mathcal{E}] . \tag{17d}$$

An alternative method of evaluating C_2C_{-2} involves using the Bianchi identities $d\mathbb{R} = 0$. In particular, we employed the identities[6]

$$3v_m C_0 - 2^{-1/2}f^{1/2}\eth C_0 = w_m C_2, \tag{18a}$$

$$3w_k C_0 - 2^{-1/2}f^{1/2}\eth^* C_0 = v_k C_{-2}, \tag{18b}$$

in order to take advantage of the earlier determination of C_0.

The simplicity of results for the $\delta = 2$ T—S solution; namely,

$$I_1 = p^4 Z^{-3} T , \tag{19a}$$

$$I_2 = 36p^8 Z^{-5}\rho^2 \tag{19b}$$

belies the amount of work which went into the derivation of these results. Here the fields T and Z are given by

$$T = 2 - px(x^2 - 3) - iqy(3 - y^2), \tag{20a}$$

$$Z = (N + D)/(x^2 - y^2). \tag{20b}$$

IV. SINGULARITIES OF THE WEYL TENSOR

According to Eq. (19b) the T—S solution is Petrov type D at all points for which $\rho = 0$, i.e., on the symmetry axis and upon the surfaces $x = \pm 1$. Elsewhere it is algebraically general. The Weyl tensor invariants become infinite only where $Z = 0$. It is in fact easy to show that Z vanishes nowhere off the symmetry plane $y = 0$, while on the symmetry plane it vanishes only at two ring singularities identified by Tomimatsu and Sato.

One implication of the present work which will require further study concerns the "directional singularities" at points where $x^2 = 1$ and $y^2 = 1$. Since the field T is perfectly well behaved at such points, the nature of the singularity is revealed by considering the behavior of Z as $x^2 \to 1$ and $y^2 \to 1$. While the value attained by Z in the limit depends upon the value of the ratio $\alpha = (x^2 - 1)/(1 - y^2)$ attained in the limit, there is no way to choose the limiting value of α so that $Z \to 0$. We, therefore, disagree with the statement of Tomimatsu and Sato[7] that it is possible to take the limit $x^2 \to 1$ and $y^2 \to 1$ in such a way that the Weyl tensor becomes infinite. Consequently, we feel that it would be appropriate to ask whether "points" such as $x = 1$, $y = 1$, are in fact *surfaces*, the nature of which has been obscured by the symmetrical x—y coordinates. We hope to be able to say more concerning this possibility in the future.

ACKNOWLEDGMENT

It is impossible to ennumerate all the subtle ways in which we have profited from our association with Professor Isidore Hauser.

[1] A. Tomimatsu and H. Sato, Phys. Rev. Lett. **29**, 1344 (1972).

[2] J. Bardeen, Astrophys. J. 162, 71 (1970); S. Chandrasekhar and J. Friedman, Astrophys. J. 175, 379 (1972).

[3] A. Tomimatsu and H. Sato, Progr. Theor. Phys. **50**, 95 (1973).

[4] This notation has been described in a number of IIT unpublished reports, and has been used sparingly in articles which have appeared recently in the *Journal of Mathematical Physics*.

[5] F. J. Ernst, J. Math. Phys. **15**, 1409 (1974).

[6] These equations are equivalent to the sixth and the third Bianchi identities listed in Eq. (4.5) of the well-known article by E. Newman and R. Penrose, J. Math. Phys. **3**, 566 (1962). Our Eq. (16) corresponds to their Eq. (4.2f).

[7] We are thinking especially of the comments in Ref. 3.

Black holes in a magnetic universe

Frederick J. Ernst

Department of Physics, Illinois Institute of Technology, Chicago, Illinois 60616
(Received 21 April 1975)

We present a general procedure for transforming asymptotically flat axially symmetric solutions of the Einstein-Maxwell equations into solutions resembling Melvin's magnetic universe. Specific applications yield metrics associated with black holes in a magnetic universe. It is hoped that these solutions will be of interest to astrophysicists studying gravitational collapse in the presence of strong magnetic fields.

I. INTRODUCTION

Seven years ago we showed[1] that stationary axially symmetric solutions of the coupled Einstein-Maxwell field equations can be generated from a pair of complex potentials $\mathcal{E}$ and Φ satisfying the field equations

$$(\mathrm{Re}\mathcal{E} + |\Phi|^2)\nabla^2\mathcal{E} = (\nabla\mathcal{E} + 2\Phi * \nabla\Phi)\circ\nabla\mathcal{E}, \tag{1.1}$$

$$(\mathrm{Re}\mathcal{E} + |\Phi|^2)\nabla^2\Phi = (\nabla\mathcal{E} + 2\Phi * \nabla\Phi)\circ\nabla\Phi. \tag{1.2}$$

Subsequently an elegant description of an invariance group K of these equations was provided by Kinnersley.[2] Starting with any particular solution of the Einstein–Maxwell equations one can now generate a large number of solutions by applying elements of the group K. Certain specific elements of K are associated with transformations developed independently by Harrison,[3] Ehlers,[4] and Ernst.[5] Other elements involve duality rotations or gauge transformations.

In general the procedure described above produces solutions which at this time seem to possess little physical relevance. For example, the solution may lack asymptotic flatness or some other desirable feature. If physics journals are to be spared the proliferation of articles presenting "new" solutions of no discernible physical relevance generated automatically by using the now well-known group K, then it is important that relativists acknowledge some criterion for publication. Because some pathological solutions have acquired a degree of distinction (e.g., the NUT solution), the appropriate place to draw the line is not particularly obvious.

While the solutions we shall present in this paper are not asymptotically flat, they possess an important redeeming feature; namely, they have nonsingular event horizons. Since these are the first exact solutions of the Einstein–Maxwell field equations corresponding to black holes in external magnetic fields, it is likely that they will be of interest to astrophysicists. Our metrics do not become asymptotically flat, but rather they resemble Melvin's magnetic universe.[6] Only in the case of a weak magnetic field can one describe the situation as that of a black hole in an approximately uniform field, but to our knowledge there is no evidence that nature favors anything resembling uniform magnetic fields in the vicinity of black holes, so one should temper one's prejudice. In fact, it is conceivable, although we have not attempted to prove it, that our approach constitutes the only way to add an external magnetic field to a black hole solution without destroying the nonsingular nature of the event horizon.

While the potential astrophysical applications of our solutions would, in our opinion, be sufficient to justify publication of these solutions, an additional justification can be found in the fact that the derivation of these solutions involves an unconventional identification of the $\mathcal{E}$ and Φ potentials, which is obtained by interchanging the roles of the timelike and spacelike Killing vector fields. Our previous prejudice in favor of associating $\mathrm{Re}\mathcal{E}$ with the norm of the timelike Killing vector was based upon the observation that the expression for $\mathcal{E}$ was generally simpler if you did this rather than associate $\mathrm{Re}\mathcal{E}$ with the norm of the spacelike Killing vector. Our present work demonstrates that it is sometimes advantageous to sacrifice this simplicity.

II. ADDITION OF A MAGNETIC FIELD

Throughout this paper stationary axially symmetric line elements will be expressed in the form

$$ds^2 = f^{-1}[-2P^{-2}d\zeta\, d\zeta^* + \rho^2 dT^2] - f(d\phi - \omega\, dT)^2, \tag{2.1}$$

where $f \neq 0$. The complex gravitational potential $\mathcal{E}$ and the complex electromagnetic potential Φ will be introduced as in Ref. 5, but with ϕ and T or the 3 and 4 indices interchanged. All the Einstein–Maxwell equations, including Eqs. (1.1) and (1.2), may be carried over from Ref. 5 without modification.

Our prescription for introducing a magnetic field involves employing a Harrison-type transformation; namely,

$$\mathcal{E}' = \Lambda^{-1}\mathcal{E}, \tag{2.2}$$

$$\Phi' = \Lambda^{-1}(\Phi - \tfrac{1}{2}B_0\mathcal{E}), \tag{2.3}$$

where

$$\Lambda = 1 + B_0\Phi - \tfrac{1}{4}B_0^2\mathcal{E}. \tag{2.4}$$

Under this transformation the fields f and ω are transformed into fields f' and ω' given by

$$f' = \mathrm{Re}\mathcal{E}' + |\Phi'|^2 = |\Lambda|^{-2}f, \tag{2.5}$$

$$\nabla\omega' = |\Lambda|^2\nabla\omega + \rho f^{-1}(\Lambda^*\nabla\Lambda - \Lambda\nabla\Lambda^*), \tag{2.6}$$

while the fields ρ and P are unmodified.

We shall illustrate the procedure by considering Minkowski space. The line element

$$ds^2 = [dz^2 + d\rho^2 - dT^2] + \rho^2 d\phi^2 \tag{2.7}$$

is of the general form (2.1), where $f = -\rho^2$, $\omega = 0$, $P = \rho^{-1}$, and $d\zeta = (2)^{-1/2}(dz + i\,d\rho)$. In this case the potentials are given by

$$\Phi = 0, \quad \mathcal{E} = -\rho^2. \tag{2.8}$$

Equation (2.4) gives

$$\Lambda = 1 + \tfrac{1}{4}B_0{}^2\rho^2, \tag{2.9}$$

while Eq. (2.3) yields the transformed electromagnetic potential

$$\Phi' = \tfrac{1}{2}\Lambda^{-1}B_0\rho^2, \tag{2.10}$$

and Eqs. (2.5) and (2.6) provide the transformed line element

$$ds^2 = \Lambda^2[dz^2 + d\rho^2 - dT^2] + \Lambda^{-2}\rho^2 d\phi^2. \tag{2.11}$$

From Eq. (2.10) the Cartan components of the magnetic field are easily shown to be

$$H_z = \Lambda^{-2}B_0, \quad H_\rho = H_\phi = 0. \tag{2.12}$$

This solution, which is called Melvin's magnetic universe, has been known for many years, and was studied extensively by Melvin and Thorne.[6]

III. SCHWARZSCHILD BLACK HOLE

The application of our procedure to any static vacuum line element is quite simple, the result being a static solution of the Einstein—Maxwell equations. By writing the Schwarzschild line element in the form

$$ds^2 = \left(\frac{dr^2}{1-2m/r} + r^2 d\theta^2 - (1-2m/r)dT^2\right) + r^2\sin^2\theta d\phi^2, \tag{3.1}$$

and comparing this with the general expression (2.1), we may identify the fields

$$f = -r^2\sin^2\theta, \quad \omega = 0, \quad \rho = (r^2 - 2mr)^{1/2}\sin\theta,$$
$$P = (r^2\sin\theta)^{-1}, \quad d\zeta = (2)^{-1/2}\left(\frac{dr}{(r^2-2mr)^{1/2}} + i\,d\theta\right). \tag{3.2}$$

Applying the procedure described in the previous section, we generate the line element of a Schwarzschild black hole in an external magnetic field; namely,

$$ds^2 = \Lambda^2\left(\frac{dr^2}{1-2m/r} + r^2 d\theta^2 - (1-2m/r)dT^2\right) + \Lambda^{-2}r^2\sin^2\theta d\phi^2. \tag{3.3}$$

where

$$\Lambda = 1 + \tfrac{1}{4}B_0{}^2 r^2\sin^2\theta. \tag{3.4}$$

In this case the Cartan components of the magnetic field are given by

$$H_r = \Lambda^{-2}B_0\cos\theta, \tag{3.5}$$

$$H_\theta = -\Lambda^{-2}B_0(1-2m/r)^{1/2}\sin\theta, \tag{3.6}$$

where the angular component vanishes upon the event horizon.

When $m = 0$ the metric (3.3) reduces to Melvin's magnetic universe, while for $m \neq 0$ there is an event horizon at $r = 2m$. From the form of Eq. (3.3) it is evident that standard Kruskal coordinates may be introduced in order to extend the solution across the nonsingular event horizon. The only singularity of the Weyl tensor occurs at $r = 0$, just as in the case of the usual Schwarzschild metric.

Everywhere on the axis the magnetic field has the value B_0. When $|B_0 m| \ll 1$ there is a region $2m \ll r \ll B_0^{-1}$ outside the event horizon where the space is approximately flat, and where the magnetic field is approximately uniform. In the case of a strong magnetic field, where $|B_0 m|$ is of order unity, there is no region outside the black hole where the space may be described as approximately flat or the magnetic field as approximately uniform. In fact, the stronger the field is, the more it is concentrated near the poles $\theta = 0$ and $\theta = \pi$.

This concentration of the field near the poles as $B_0 \to \infty$ explains why the Hajicek inequality[7]

$$I = \int_0^\pi \left(\frac{B^2}{A^2} + R^2E^2 + R^2H^2\right)A\,d\theta < 2, \tag{3.7}$$

can continue to be satisfied as $B_0 \to \infty$. For our solution we have

$$A = \Lambda^{-2}\sin\theta, \quad B = 0, \quad R = 2m\Lambda, \quad E = 0, \quad H = \Lambda^{-2}B_0\cos\theta, \tag{3.8}$$

where $\Lambda = 1 + B_0{}^2m^2\sin^2\theta$. The integration over the event horizon yields

$$I = \frac{(1+4B_0{}^2m^2)(3+2B_0{}^2m^2)}{6(1+B_0{}^2m^2)^2} - \frac{\tanh^{-1}[B_0m/(1+B_0{}^2m^2)^{1/2}]}{2B_0m(1+B_0{}^2m^2)^{5/2}}. \tag{3.9}$$

The second term is negative definite, while the first term is a monotonically increasing function of $|B_0 m|$, and is always less than 4/3. Thus, our exact solution easily satisfies the Hajicek limit with $I < 4/3$.

IV. REISSNER-NORDSTROM BLACK HOLE

The application of our procedure to the Reissner—Nordstrom black hole is not quite so simple. Since $\mathbf{E} \times \mathbf{H}$ serves as a source for the twist potential, the transformed metric is stationary rather than static. In this case we may identify

$$\Phi = -ie\cos\theta, \quad \mathcal{E} = -r^2\sin^2\theta - e^2\cos^2\theta. \tag{4.1}$$

Consequently, the transformed metric assumes the form

$$ds^2 = |\Lambda|^2\left(\frac{dr^2}{1-2m/r+e^2/r^2} + r^2 d\theta^2 - (1-2m/r+e^2/r^2)dT^2\right) + |\Lambda|^{-2}r^2\sin^2\theta(d\phi - \omega'\,dT)^2, \tag{4.2}$$

where

$$\Lambda = 1 + \tfrac{1}{4}B_0{}^2(r^2\sin^2\theta + e^2\cos^2\theta) - iB_0e\cos\theta. \tag{4.3}$$

Integration of Eq. (2.6) yields the following expression for ω':

$$\omega' = -2B_0er^{-1} + B_0{}^3er + \tfrac{1}{2}B_0{}^3e^3r^{-1} - \tfrac{1}{2}B_0{}^3er^{-1}(r^2 - 2mr + e^2)\sin^2\theta + \text{const.} \tag{4.4}$$

Finally, the Cartan components of the electric and magnetic fields may be evaluated from the electromagnetic potential Φ'. The result is expressible in the following form:

$$H_r + iE_r = \Lambda^{-2}\{i(e/r^2)[1 - \tfrac{1}{4}B_0{}^2(r^2\sin^2\theta + e^2\cos^2\theta)] + B_0(1 - \tfrac{1}{2}iB_0e\cos\theta)(1 - e^2/r^2)\cos\theta\}, \tag{4.5}$$

$$H_\theta + iE_\theta = -B_0\Lambda^{-2}(1 - \tfrac{1}{2}iB_0 e\cos\theta)(1 - 2m/r + e^2/r^2)^{1/2}\sin\theta. \tag{4.6}$$

V. KERR-NEWMAN BLACK HOLE

In principle nothing prevents one from writing out the metric of a Kerr—Newman black hole in an external magnetic field, although the evaluation of ω' and the Cartan components of the electric and magnetic fields is very tedious. The appropriate complex potentials in this case assume the following forms:

$$\Phi = e\frac{a - ir\cos\theta}{r + ia\cos\theta}, \tag{5.1}$$

$$\mathcal{E} = -\left(r^2 + a^2 - a\frac{2ma + i(2mr - e^2)\cos\theta}{r + ia\cos\theta}\right)\sin^2\theta - (4ma + ie^2\cos\theta)\frac{a - ir\cos\theta}{r + ia\cos\theta}. \tag{5.2}$$

If preliminary studies indicate that astrophysics would be well served by having the explicit solution for a Kerr—Newman black hole in an external magnetic field, the rest of the calculation can be facilitated by electronic symbol manipulation.[8]

ACKNOWLEDGMENT

We should like to thank Dr. R. Wald, whose lecture concerning approximate solutions corresponding to black holes in magnetic fields provided the original motivation for our study, and with whom we have had a number of fruitful discussions.

[1]F. J. Ernst, Phys. Rev. **168**, 1415 (1968).
[2]W. Kinnersley, J. Math. Phys. **14**, 651 (1973).
[3]B. K. Harrison, J. Math. Phys. **9**, 1744 (1968).
[4]J. Ehlers, *Les théories relativistes de la gravitation* (CNRS, Paris, 1959).
[5]F. J. Ernst, J. Math. Phys. **15**, 1409 (1974).
[6]M. A. Melvin, Phys. Rev. **139**, B225 (1965); also see K. S. Thorne, Phys. Rev. **139**, B244 (1965).
[7]P. Hajicek, J. Math. Phys. **15**, 1554 (1974).
[8]*Note added in proof*: This calculation has since been performed by F. J. Ernst and W. J. Wild, and an article concerning the results is scheduled to appear in J. Math. Phys.

The spectrum of the Liouville–von Neumann operator

Herbert Spohn

Fachbereich Physik der Universität München, München, Germany
(Received 21 April 1975)

We relate the pure point spectrum, the singularly continuous, and the absolutely continuous part of the spectrum of the Liouville–von Neumann operator $[H,\cdot]$ to the respective parts of the spectrum of the Hamiltonian operator H. As a consequence of this result we obtain a theorem about the weak* limit of the time evolution $W(t)$ of a normal state W for $t \to \infty$.

1. INTRODUCTION

In quantum statistical mechanics, the time evolution of a normal state (density matrix) W is governed by the Liouville–von Neumann equation which reads formally

$$i\frac{d}{dt}W=[H,W]. \tag{1}$$

Often, the spectrum of the Hamiltonian operator H is known to some extent, and one would like to convert this information to information about the spectrum of the Liouville–von Neumann operator $[H,\cdot]$. In this note, we intend to solve this problem. Hereby, a recent result by Prugovečki and Tip[1] is extended to the continuous spectrum. (In a more general context, Prugovečki and Tip related the pure point spectrum of the respective operators, a question which, for bounded Hamiltonian operators, has also been studied by Moyal.[2]) We will apply our result to the problem of the existence of

$$\text{w*-}\lim_{t\to\infty} W(t). \tag{2}$$

[$W(t)$ denotes the time evolution of the normal state W and w*-lim the weak* limit in the trace class.] This supplements the thorough investigation in Ref. 1.

2. NOTATIONS AND DEFINITIONS

Let $\mathcal{H}$ be a (separable or nonseparable) Hilbert space and $U(t)=\exp(iHt)$ be a strongly continuous one-parameter unitary group on $\mathcal{H}$ with generator H. Let $D(H)\subset\mathcal{H}$ be the domain of H. H is the *Hamiltonian operator* of the quantum mechanical system considered. Let $\mathcal{H}_{pp}$, $\mathcal{H}_{sc}$, and $\mathcal{H}_{ac}$ be the closed subspaces of discontinuity, of singular and of absolute continuity of $\mathcal{H}$ with respect to H[3,4] and P_{pp}, P_{sc}, P_{ac} be the (pairwise orthogonal) projection operators with ranges $\mathcal{H}_{pp}$, $\mathcal{H}_{sc}$, and $\mathcal{H}_{ac}$, respectively. We have $P_{pp}+P_{sc}+P_{ac}=1$.

We denote by $B_2(\mathcal{H})$ the Hilbert space of Hilbert–Schmidt operators on $\mathcal{H}$ [with the scalar product $\langle A\,|\,B\rangle=\mathrm{tr}(A^*B)$]. Then

$$B_2(\mathcal{H}) \ni A \mapsto U(t)AU(t)=\mathfrak{U}(t)A \tag{3}$$

defines a strongly continuous one-parameter unitary group on $B_2(\mathcal{H})$.[2] Let $\mathfrak{H}$ with domain $D(\mathfrak{H})\subset B_2(\mathcal{H})$ be the generator of $\mathfrak{U}(t)$: $\mathfrak{U}(t)=\exp(i\mathfrak{H}t)$. $\mathfrak{H}$ is called the *Liouville–von Neumann operator* corresponding to H. For bounded H we have $\mathfrak{H}A=[H,A]=HA-AH$ for all $A\in B_2(\mathcal{H})$. If H is unbounded, this does not hold (cf. Ref. 1), not even on $D(\mathfrak{H})$. However, from the representation of $\mathfrak{H}$ as a multiplication operator in the proof below, we see that $D=\{A\in B_2(\mathcal{H})\,|\,AH\in B_2(\mathcal{H})$ and $HA\in B_2(\mathcal{H})\}\subset D(\mathfrak{H})$ is a core for $\mathfrak{H}$ and that $\mathfrak{H}$ defined on D is essentially self-adjoint. (By a theorem due to Nelson [Ref. 5, Theorem 1.4] this can also be concluded directly, not using the spectral representation of H.) With this limitation in mind, we will keep to the suggestive notation $\mathfrak{H}=[H,\cdot]$. Let $\mathfrak{P}_{pp}$, $\mathfrak{P}_{sc}$, and $\mathfrak{P}_{ac}$ be the projection operators with ranges $B_2(\mathcal{H})_{pp}$, $B_2(\mathcal{H})_{sc}$, and $B_2(\mathcal{H})_{ac}$, respectively.

Finally, we denote by $\sigma(A)$, $\sigma_{pp}(A)$, $\sigma_{sc}(A)$, and $\sigma_{ac}(A)$ the spectrum, the pure point spectrum, the singularly continuous spectrum and the absolutely continuous spectrum of the self-adjoint operator A.

3. THE SPECTRUM

Theorem 1: Presuppositions as in Sec. 2. Then the spectrum of $[H,\cdot]$ and the spectrum of H are related in the following way:

$$\sigma([H,\cdot])=\{x-y\,|\,x,y\in\sigma(H)\},$$

$$\sigma_{pp}([H,\cdot])=\{x-y\,|\,x,y\in\sigma_{pp}(H)\},$$

$$\sigma_{sc}([H,\cdot])=\{x-y\,|\,x\in\overline{\sigma_{pp}(H)}\text{ and }y\in\sigma_{sc}(H),\ x\in\sigma_{sc}(H)\text{ and }y\in\overline{\sigma_{pp}(H)},\ x,y\in\sigma_{sc}(H)\},$$

$$\sigma_{ac}([H,\cdot])=\{x-y\,|\,\text{either } x\in\sigma_{ac}(H)\text{ or }y\in\sigma_{ac}(H)\text{ with } x,y\in\sigma(H)\},$$

$$\mathfrak{P}_{pp}B_2(\mathcal{H})=P_{pp}B_2(\mathcal{H})P_{pp},$$

$$\mathfrak{P}_{sc}B_2(\mathcal{H})=P_{pp}B_2(\mathcal{H})P_{sc}\oplus P_{sc}B_2(\mathcal{H})(P_{pp}+P_{sc}),$$

$$\mathfrak{P}_{ac}B_2(\mathcal{H})=B_2(\mathcal{H})P_{ac}\oplus P_{ac}B_2(\mathcal{H})(P_{pp}+P_{sc}).$$

Proof: (1) By the spectral theorem, there exists a unitary mapping

$$V:\mathcal{H}\to\bigoplus_{k\in I}L^2(R,d\mu_k) \tag{4}$$

such that

$$VD(H)=\left\{\psi\in\bigoplus_{k\in I}L^2(R,d\mu_k)\,\middle|\,\{x\psi_k(x)\}_{k\in I}\in\bigoplus_{k\in I}L^2(R,d\mu_k)\right\}$$

and for all $\psi\in VD(H)$

$$(VHV^{-1}\psi)_k(x)=x\psi_k(x), \tag{5}$$

$k\in I$, $x\in R$. Thus for all $\psi\in\oplus_{k\in I}L^2(R,d\mu_k)$

$$(VU(t)V^{-1}\psi)_k(x)=\exp(ixt)\psi_k(x). \tag{6}$$

Let P_k, $k\in I$, be the projection operator with range $L^2(R,d\mu_k)$. The image $\mathfrak{U}'(t)$ of $\mathfrak{U}(t)$ under V is given by

$$\mathfrak{U}'(t):B_2\left(\bigoplus_{k\in I}L^2(R,d\mu_k)\right)\ni A\mapsto VU(t)V^{-1}A(VU(t)V^{-1})^*. \tag{7}$$

Since $[P_k, VU(t)V^{-1}]=0$ for all $t\in R$ and all $k\in I$, we have $P_iVU(t)V^{-1}A(VU(t)V^{-1})^*P_j = VU(t)V^{-1}P_iAP_j(VU(t)V^{-1})^*$ Thus the closed subspaces $P_iB_2(\oplus_{k\in I} L^2(R, d\mu_k))P_j$ reduce $\mathfrak{U}'(t)$ for all $i,j\in I$.

By Ref. 3, Theorem VI. 23, there exists a unitary mapping

$$\mathfrak{B}_{ij}: L^2(R^2, d\mu_i\times\mu_j) \to P_iB_2\left(\bigoplus_{k\in I} L^2(R, d\mu_k)\right)P_j \tag{8}$$

such that the action of the Hilbert–Schmidt operator $\mathfrak{B}_{ij}f$, $f\in L^2(R^2, d\mu_i\times\mu_j)$, is defined by an integral operator:

$$(\mathfrak{B}_{ij}f)\psi_j(x) = \int_R f(x,y)\psi_j(y)\,d\mu_j(y)\in L^2(R, d\mu_i). \tag{9}$$

Since

$$\begin{aligned}&[VU(t)V^{-1}(\mathfrak{B}_{ij}f)(VU(t)V^{-1})^*\psi_j](x)\\ &\quad=\exp(ixt)(\mathfrak{B}_{ij}f)[\exp(-ixt)\psi_j(x)]\\ &\quad=\int_R \exp(ixt)f(x,y)\exp(-iyt)\psi_j(y)\,d\mu_j(y),\end{aligned} \tag{10}$$

we obtain

$$\mathfrak{B}_{ij}^{-1}\mathfrak{U}'(t)\mathfrak{B}_{ij}: f(x,y)\to\exp[i(x-y)t]f(x,y). \tag{11}$$

By Stone's theorem,

$$\begin{aligned}&D(\mathfrak{B}_{ij}^{-1}([VHV^{-1}, \cdot]\restriction P_iB_2P_j)\mathfrak{B}_{ij})\\ &\quad=\{f\in L^2(R^2, d\mu_i\times\mu_j)\mid \int_{R^2}|x-y|^2\\ &\qquad\times|f(x,y)|^2\,d\mu_i(x)\times\mu_j(y)<\infty\}\end{aligned}$$

and for all $f\in D(\mathfrak{B}_{ij}^{-1}([VHV^{-1}, \cdot]\restriction P_iB_2P_j)\mathfrak{B}_{ij})$

$$\mathfrak{B}_{ij}^{-1}([VHV^{-1}, \cdot]\restriction P_iB_2P_j)\mathfrak{B}_{ij}: f(x,y)\mapsto(x-y)f(x,y). \tag{12}$$

Thus $[H, \cdot]$ is represented as a direct sum of multiplication operators.

(2) Next, for all $i,j\in I$, we have to study the relation between the spectrum of

$$L_{ij}: L^2(R^2, d\mu_i\times\mu_j)\ni f(x,y)\mapsto(x-y)f(x,y) \tag{13}$$

on

$$D(L_{ij})=\{f\mid\int_{R^2}|x-y|^2\mid f(x,y)|^2\times d\mu_i(x)\times\mu_j(y)<\infty\}$$

and the spectrum of

$$H_j: L^2(R, d\mu_j)\ni\psi(x)\mapsto x\psi(x) \tag{14}$$

on

$$D(H_j)=\{\psi\mid\int_R x^2|\psi(x)|^2\,d\mu_j(x)<\infty\}.$$

The spectral measure $E_{ij}(\cdot)$ of L_{ij} is (Ref. 4, **X**, Example 1.9)

$$\|E_{ij}(\Delta)f\|^2=\int_{R^2}\mathrm{ch}\{x-y\in\Delta\}\,|f(x,y)|^2\,d\mu_i(x)\times\mu_j(y), \tag{15}$$

where $\mathrm{ch}\{\cdot\}$ is the characteristic function of the set $\{\cdot\}$.

(a) Let $\{x-y=z\}\cap\sigma(H_i)\times\sigma(H_j)=\emptyset$. Since $\sigma(H_i)\times\sigma(H_j)$ is closed, its distance d to the straight line $\{x-y=z\}$ is positive. This implies $|(x-y)-z|^{-1}\le d^{-1}$ for all $\{x,y\}\in\sigma(H_i)\times\sigma(H_j)$ and therefore z lies in the resolvent set of L_{ij}. Let $\{x-y=z\}\cap\sigma(H_i)\times\sigma(H_j)\neq\emptyset$ and $\{\omega_i,\omega_j\}$ be a point of the intersection. Since $\omega_i\in\sigma(H_i)$ and $\omega_j\in\sigma(H_j)$, there exist vectors $\psi_n\in L^2(R, d\mu_i)$ and vectors $\varphi_n\in L^2(R, d\mu_j)$ such that $\|\psi_n\|=1=\|\varphi_n\|$ and $\|(x-\omega_i)^{-1}\psi_n\|\to\infty$, $\|(y-\omega_j)^{-1}\varphi_n\|\to\infty$ as $n\to\infty$. We have $z=\omega_i-\omega_j$ and

$$2(x-\omega_i)^2+2(y-\omega_j)^2\ge[(x-\omega_i)-(y-\omega_j)]^2.$$

This implies $\|[(x-y)-z]^{-1}\psi_n\varphi_n\|\to\infty$ as $n\to\infty$ and therefore $z\in\sigma(L_{ij})$. Thus we have shown that

$$\sigma(L_{ij})=\{x-y\mid x\in\sigma(H_i),\ y\in\sigma(H_j)\}.$$

(b) $L_{ij}\restriction L^2(R^2, d\mu_i^{pp}\times\mu_j^{pp})$ has a pure point spectrum.

(c) $L_{ij}\restriction L^2(R^2, d\mu_i^{c}\times\mu_j)$ has a continuous spectrum, since the continuity of μ_i^c implies

$$\|E_{ij}(\{z\})f\|^2=\int_R\left(\int_R\mathrm{ch}\{x-y=z\}\,|f(x,y)|^2\,d\mu_i^c(x)\right)\times d\mu_j(y)=0 \tag{16}$$

for all $f\in L^2(R^2, d\mu_i^c\times\mu_j)$.

(d) $L_{ij}\restriction L^2(R^2, d\mu_i^{sc}\times\mu_j^{pp})$ has a singularly continuous spectrum. There exists a measurable set $\Delta\subset R$ with $\lambda(\Delta)=0$ and $\mu_i^{sc}(\Delta)=\mu_i^{sc}(R)$. ($\lambda$ is the one-dimensional Lebesgue measure.) Let $\Omega\subset R$ be the projection under 45° of the set $\Delta\times\{y\in R\mid\mu_j^{pp}(\{y\})\neq0\}$ onto the x axis. Then $\lambda(\Omega)=0$ and, by (15), $\|E_{ij}(\Omega)f\|^2=\|f\|^2$ for all $f\in L^2(R^2, d\mu_i^{sc}\times\mu_j^{pp})$.

(e) $L_{ij}\restriction L^2(R^2, d\mu_i^{sc}\times\mu_j^{sc})$ has a singularly continuous spectrum. Let Δ_1, $\Delta_2\subset R$ be measurable and Δ_ρ be the projection under 45° of $\Delta_1\times\Delta_2$ onto the x axis. Approximating $\Delta_1\times\Delta_2$ from outside by a union of open intervals, one obtains $\lambda(\Delta_\rho)\le\lambda(\Delta_1)+\lambda(\Delta_2)$. There exist measurable sets Δ, $\Delta'\subset R$ with $\lambda(\Delta)=0=\lambda(\Delta')$ and $\mu_i^{sc}(\Delta)=\mu_i^{sc}(R)$, $\mu_j^{sc}(\Delta')=\mu_j^{sc}(R)$. Let $\Omega'\subset R$ be the projection under 45° of $\Delta\times\Delta'$ onto the x axis. Then $\lambda(\Omega')\le\lambda(\Delta)+\lambda(\Delta')=0$ and, by (15), $\|E_{ij}(\Omega')f\|^2=\|f\|^2$ for all $f\in L^2(R^2, d\mu_i^{sc}\times\mu_j^{sc})$.

(f) $L_{ij}\restriction L^2(R^2, d\mu_i^{ac}\times\mu_j)$ has an absolutely continuous spectrum. Let $\lambda(\Delta)=0$. By the absolute continuity of μ_i^{ac} we obtain

$$\|E_{ij}(\Delta)f\|^2=\int_R\left(\int_R\mathrm{ch}\{x-y\in\Delta\}\,|f(x,y)|^2\,d\mu_i^{ac}(x)\right)\times d\mu_j(y)=0 \tag{17}$$

for all $f\in L^2(R^2, d\mu_i\times\mu_j)$.

(3) Using the invariance of the spectrum under unitary transformations, we have

$$\begin{aligned}&\sigma_{pp}([H, \cdot])=\bigcup_{i,j\in I}\sigma_{pp}(L_{ij}),\\ &\sigma_{pp}(H)=\bigcup_{j\in I}\sigma_{pp}(H_j),\\ &\sigma_l([H, \cdot])=\overline{\bigcup_{i,j\in I}\sigma_l(L_{ij})},\\ &\sigma_l(H)=\overline{\bigcup_{j\in I}\sigma_l(H_j)},\end{aligned} \tag{18}$$

where $l\in\{\phi, \mathrm{sc}, \mathrm{ac}\}$. Furthermore,

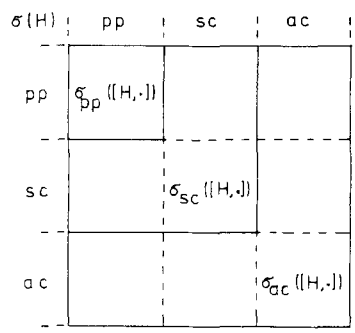
$$\mathfrak{P}_m=\sum_{i,j\in I}\mathfrak{P}_m^{ij} \tag{19}$$

holds in the strong limit, where $m\in\{\mathrm{pp}, \mathrm{sc}, \mathrm{ac}\}$ and $\mathfrak{P}_m^{ij}$ is the image under $\mathfrak{B}_{ij}$ and V^{-1} of the projection operator

which has as range the m-subspace of $L^2(R^2, d\mu_i \times \mu_j)$ corresponding to L_{ij}. Inserting the results of (2) (a) up to (2) (f) in (18) and (19) we obtain the assertions of the theorem. ∎

Corollary 2: $[H, \cdot]$ has a pure point (singularly continuous, absolutely continuous) spectrum if and only if H has a pure point (singularly continuous, absolutely continuous) spectrum.

The results of Theorem 1 can easily be summarized in the following figure:



(20)

The proof of Theorem 1 shows that the spectrum of $[H, \cdot]$ can be obtained by a simple graphical construction: one has to project under 45° the set $\sigma(H) \times \sigma(H)$ onto the x axis. In an obvious way, one even obtains the degeneracy of the spectrum of $[H, \cdot]$ (cf. Fig. 1).

We should note that the spectral measure equivalence classes of $[H, \cdot]$ can be expressed in terms of those of H, but it does not seem to be worthwhile to write down this rather lengthy expression. However, the spectral family $\mathfrak{E}(\chi)$ of $[H, \cdot]$ can easily be stated in terms of the spectral family $E(x)$ of H as

$$\|\mathfrak{E}(\chi)A\|_2^2 = \int_{R^2} \mathrm{ch}\{x - y < \chi\} \, \mathrm{tr}[A^* E(x) A E(y)] \tag{21}$$

for all $A \in B_2(\mathcal{H})$. A similar expression has been derived by Prugovečki.[6]

4. WEAK LIMIT OF NORMAL STATES

Let $B_\infty(\mathcal{H})$ denote the C^*-algebra of all compact operators on $\mathcal{H}$ and $B_1(\mathcal{H})$ the Banach space of all trace class operators on $\mathcal{H}$. Then $B_\infty(\mathcal{H})^* = B_1(\mathcal{H})$ and the weak* topology on $B_1(\mathcal{H})$ is the weakest topology on $B_1(\mathcal{H})$ in which all linear functionals $\mathrm{tr}A \cdot$, $A \in B_\infty(\mathcal{H})$, are continuous.[7] If $\lambda_k \in \sigma_{pp}(H)$, $k \in J$, then let Q_k be the projection operator which has as range the eigenspace belonging to λ_k. By Theorem 1, the projection operator $\mathfrak{P}_0$ which has as range the eigenspace belonging to the zero eigenvalue of $[H, \cdot]$ is given by $\mathfrak{P}_0 A = \sum_{k \in J} Q_k A Q_k$ for all $A \in B_2(\mathcal{H})$. In addition to Ref. 1 we now have the following:

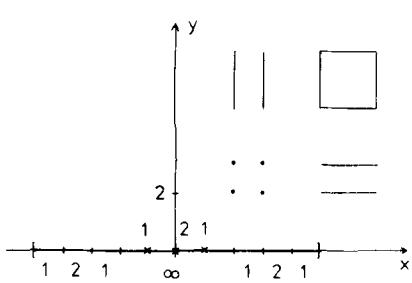


FIG. 1. We assume that the spectrum of H is simple and that $\sigma_{pp}(H) = \{2, 3\}$, $\sigma_{ac}(H) = [5, 7]$. The set $\sigma(H) \times \sigma(H)$ is drawn on the right. Projecting under 45° this set onto the x axis, we obtain the spectrum of $[H, \cdot]$ with $\sigma_{pp}([H, \cdot]) = \{-1, 0, 1\}$ and $\sigma_{ac}([H, \cdot]) = [-5, 5]$. Above the x axis we marked the degeneracy of the point spectrum of $[H, \cdot]$ and below the x axis the degeneracy of the absolutely continuous spectrum of $[H, \cdot]$.

Theorem 3: With notations and assumptions as in Sec. 2, let $W \in B_1(\mathcal{H})$.

(i) If $\sigma_{sc}(H) = \emptyset$, then

$$\mathrm{w}^*\text{-}\lim_{t \to \infty} U(t) W U(t)^* \tag{22}$$

exists and is equal to $\mathfrak{P}_0 W$ if and only if $Q_i W Q_j = 0$ for all $i \neq j \in J$.

(ii) If $\mathcal{H}$ is separable, then there exists a measurable set $\Delta \subset R_+$ of density zero[8] such that

$$\mathrm{w}^* - \lim_{t \to \infty,\, t \notin \Delta} U(t) W U(t)^* \tag{23}$$

exists and is equal to $\mathfrak{P}_0 W$ if and only if $Q_i W Q_j = 0$ for all $i \neq j \in J$.

Proof: If $W \in \mathfrak{P}_0 B_2(\mathcal{H})$, the assertions are obvious. Thus we can assume $W \in (1 - \mathfrak{P}_0) B_2(\mathcal{H})$.

"$\Leftarrow$" ad (i): By presupposition and by Theorem 1, $W \in \mathfrak{P}_{ac} B_2(\mathcal{H})$. A theorem of ergodic theory[9] implies

$$\lim_{t \to \infty} \mathrm{tr}[A U(t) W U(t)^*] = 0 \tag{24}$$

for all $A \in B_2(\mathcal{H})$. Since $B_2(\mathcal{H})$ lies dense in $B_\infty(\mathcal{H})$, there exists for every $A \in B_\infty(\mathcal{H})$ a sequence $A_n \in B_2(\mathcal{H})$, $n \in N$, such that $\|A_n - A\| \to 0$ as $n \to \infty$. Since

$$|\mathrm{tr}[(A_n - A) U(t) W U(t)^*]| \leq \|W\|_1 \, \|A_n - A\|, \tag{25}$$

$\mathrm{tr}[A_n U(t) W U(t)^*]$ converges uniformly to $\mathrm{tr}[A U(t) W U(t)^*]$ which proves the assertion.

ad (ii): By presupposition and by Theorem 1, $W \in (\mathfrak{P}_{sc} + \mathfrak{P}_{ac}) B_2(\mathcal{H})$. A theorem of ergodic theory[10] implies

$$\lim_{t \to \infty} \frac{1}{t} \int_0^t |\mathrm{tr}[A U(s) W U(s)^*]| \, ds = 0 \tag{26}$$

for all $A \in B_2(\mathcal{H})$. Since with $\mathcal{H}$ $B_2(\mathcal{H})$ is also separable, there exists a set $\Delta \subset R_+$ of density zero (where Δ may be taken as a set of intervals such that there are only a finite number of intervals of Δ in any arbitrarily chosen finite interval)[11] such that

$$\lim_{t \to \infty,\, t \notin \Delta} \mathrm{tr}[A U(t) W U(t)^*] = 0 \tag{27}$$

for all $A \in B_2(\mathcal{H})$. (If $\mathcal{H}$ is nonseparable, in general, one cannot choose Δ to be independent of A.) The same approximation as above proves the assertion.

"$\Rightarrow$": If $Q_i W Q_j \neq 0$ for some $i \neq j \in J$, then, by Theorem 1, there exists an $A \in B_\infty(\mathcal{H})$ such that $\mathrm{tr}[A U(t) W U(t)^*] = \exp[i(\lambda_i - \lambda_j)t] \mathrm{tr} A W$ with $\mathrm{tr} A W \neq 0$. Thus the limits (22) and (23) do not exist. ∎

Physically, a state $W \in \mathfrak{P}_{ac} B_2(\mathcal{H})$ is a scattering state

which intuitively means that, in some sense, it "flies off" to infinity. Theorem 3 makes precise that this flying off takes place in the weak* topology and not, as one would assume on first sight, in the weak topology (in which the above mentioned limit, in general, does not exist).

ACKNOWLEDGMENT

The author would like to thank Dr. W. Ochs for a critical reading of the manuscript.

[1]E. Prugovečki and A. Tip, J. Math. Phys. **15**, 275 (1974).
[2]J.E. Moyal, J. Math. Phys. **10**, 506 (1969).
[3]M. Reed and B. Simon, *Methods of Modern Mathematical Physics I: Functional Analysis* (Academic, New York, 1972).
[4]T. Kato, *Perturbation Theory for Linear Operators* (Springer-Verlag, Berlin, 1966).
[5]B. Simon, "Topics in Functional Analysis," *Mathematics of Contemporary Physics*, edited by R.F. Streater (Academic, London, 1972).
[6]Prugovečki, J. Math. Phys. **13**, 969 (1972).
[7]Sakai, *C^*-Algebras* and *W^*-Algebras* (Springer-Verlag, Berlin, 1972).
[8]$\Delta \subset R_+$ is of density zero, if $\lim_{t\to\infty}(1/t)\lambda(\Delta \cap [0,t[) = 0$.
[9]V.I. Arnold and A. Avez, *Ergodic Problems of Classical Mechanics* (Benjamin, New York, 1968).
[10]E. Hopf, *Ergodentheorie* (Springer-Verlag, Berlin, 1973, reprint 1970). The proofs of Refs. 9 and 10 extend to the more general case considered here.
[11]B.O. Koopman and J. von Neumann, Proc. Natl. Acad. Sci. USA **18**, 255 (1932).

On superpropagator for nonpolynomial Lagrangians with internal symmetry

A. K. Kapoor

Department of Physics, Indian Institute of Technology, Kanpur-208016, India
(Received 7 April 1975)

Using the exponential shift lemma, a method to evaluate the superpropagator for scalar functions of a multiplet of fields is developed. As an application we obtain the generating function $\langle T(\mathrm{Tr}\exp[\lambda\Phi(x)]\ \mathrm{Tr}\exp[\lambda'\Phi(y)])\rangle_0$ for vacuum expectation values $\langle T(\mathrm{Tr}\Phi^N(x)\mathrm{Tr}\Phi^N(y))\rangle_0$, when Φ is a 3×3 matrix, which is sufficient to get the results of Ashmore and Delbourgo on the matrix superpropagator for the chiral symmetry case.

1. INTRODUCTION

Nonpolynomial Lagrangians involving functions of matrix fields are of considerable interest because of their application to nonlinear chiral Lagrangians and gravity modified theories.[1] To perform calculations in these theories, the vacuum expectation values (VEV) of time ordered products of scalar and matrix functions of fields are needed. Transform methods were used by Delbourgo[2] to evaluate superpropagators of scalar and matrix functions of fields with isospin. Later, using an integral representation for determinant of a matrix raised to an arbitrary power, Ashmore and Delbourgo[3] succeeded in evaluating $\langle T(\mathrm{Tr}\Phi^N(x)\mathrm{Tr}\Phi^N(y))\rangle_0$, Φ being $\nu\times\nu$ matrix field. This was used by them to calculate matrix superpropagator. In this paper we develop a method, based on exponential shift lemma,[4] to obtain superpropagators for invariant functions of matrix fields. As an application, we calculate $\langle T(\mathrm{Tr}\exp[\lambda\Phi(x)]\times\mathrm{Tr}\exp[\lambda'\Phi(y)])\rangle_0$ and recover the results of Ashmore and Delbourgo.

2. INTEGRAL REPRESENTATION FOR SUPERPROPAGATOR

The exponential shift lemma[4] for a two point function states that

$$T[:F(\phi_k(x)):\,:F'(\phi_k(y)):]$$
$$=(1/\pi^k)\prod_k\int d^2u_k\exp(-\sum_k|u_k|^2)$$
$$\times F(\phi_k(x)+c_ku_k)F'(\phi_k+c_k'u_k^*), \tag{1}$$

where the c's obey

$$c_kc_k'=\Delta(x-y)\quad\text{(no summation over } k)$$

and are otherwise arbitrary. We will assume all c_k to be equal and drop the index k.

Let us now consider the case of a $\nu\times\nu$ Hermitian matrix field $\Phi_{\alpha\beta}$ satisfying,[5]

$$\langle T(\Phi_{\alpha\beta}(x)\Phi_{\alpha'\beta'}(y))\rangle_0=\delta_{\alpha\beta'}\delta_{\alpha'\beta}\Delta(x-y) \tag{2}$$

and consider the time ordered function

$$T[:F(\Phi_{\alpha\beta}(x)):\,:F'(\Phi_{\alpha\beta}(y)):], \tag{3}$$

where F and F' are assumed to be invariant under transformations $\Phi\to X\Phi X^{-1}$. Equation (1) now reads

$$T[:F(\Phi_{\alpha\beta}(x)):\,:F'(\Phi_{\alpha\beta}(y)):]$$
$$=\int(\Pi(1/\pi)\,du_{\alpha\beta})\exp(-\sum_{\alpha,\beta}|u_{\alpha\beta}|^2)$$
$$\times F(\Phi_{\alpha\beta}(x)+cu_{\alpha\beta})F'(\Phi_{\alpha\beta}(y)+c'u_{\alpha\beta}^*). \tag{4}$$

The vacuum expectation value of (3) is obtained by setting $\Phi=0$ in the right-hand side of (4):

$$\langle T[:F(\Phi_{\alpha\beta}(x)):\,:F'(\Phi_{\alpha\beta}(y)):]\rangle_0$$
$$=(1/\pi^{\nu^2})\int dU\exp[-\mathrm{Tr}(U^\dagger U)]F(cU)F'(c'U^*), \tag{5}$$

where the matrix U is defined by

$$(U)_{\alpha\beta}=u_{\alpha\beta}.$$

Let X be the matrix which diagonalizes a matrix U which has distinct eigenvalues, so that we have

$$U=XZX^{-1}, \tag{6}$$

where Z is a diagonal matrix. As the matrices not having distinct eigenvalues form a set of measure zero, changing variables of integration from U to X and Z, we obtain for (5)[6]

$$\int dx\,|J(X)|\prod_\alpha dz_\alpha\prod_{\alpha<\beta}|z_\alpha-z_\beta|^4$$
$$\times\exp\{-\mathrm{Tr}[Z^+X^+XZ(X^+X)^{-1}]\}F(cZ)F(cZ^*), \tag{7}$$

where we have explicitly written the dependence of the Jacobian on the elements z_α of the diagonal matrix Z. Following Ginibre,[7] integrations over X can be performed to obtain

$$N\int\prod_\alpha dz_\alpha\prod_{\alpha<\beta}|z_\alpha-z_\beta|^2\exp(-\sum_\alpha|z_\alpha|^2)F(cZ)F'(c'Z^*), \tag{8}$$

where N is a constant. Integrations over Z's can be performed using techniques developed in statistical mechanics.[8] Expression (8) is an integral representation for time ordered product (5). This integral can be completely solved if F and F' are known. We demonstrate it for the case of chiral superpropagator.

3. CHIRAL SUPERPROPAGATOR

We now use Eq. (8) to evaluate

$$A(\lambda\lambda'\Delta)=\langle T(\mathrm{Tr}:\exp[\lambda\Phi(x)]:\,\mathrm{Tr}:\exp[\lambda'\Phi(y)]:)\rangle_0, \tag{9}$$

where Φ is a 3×3 matrix. Using (8), we get

$$A = N\int dz_1\,dz_2\,dz_3$$

$$\times \prod_{\alpha<\beta} |z_\alpha - z_\beta|^2 \exp(-|z_1|^2 - |z_2|^2 - |z_3|^2)$$

$$\times \sum_\alpha \exp(\lambda c z_\alpha) \sum_\beta \exp(\lambda c' z_\beta^*). \tag{10}$$

$\prod_{\alpha<\beta}|z_\alpha - z_\beta|^2$ in the above integral is the square of absolute value of Vandermonde determinant of z's, i.e.,

$$\prod_{\alpha<\beta}|z_\alpha - z_\beta|^2 = \begin{vmatrix} 1 & 1 & 1 \\ z_1 & z_2 & z_3 \\ z_1^2 & z_2^2 & z_3^2 \end{vmatrix}^2$$

$$= \begin{vmatrix} 3 & \sum z_\alpha^* & \sum z_\alpha^{*2} \\ \sum z_\alpha & \sum z_\alpha^* z_\alpha & \sum z_\alpha^{*2} z_\alpha \\ \sum z_\alpha^2 & \sum z_\alpha^* z_\alpha^2 & \sum z_\alpha^{*2} z_\alpha^2 \end{vmatrix}. \tag{11}$$

Since the integrand in (10) is symmetric in z's, the first column of (11) may be replaced by 3, $3z_1$, $3z_1^2$ inside the integral. z_1 can then be eliminated from other two columns by subtracting suitable multiples of first column. This process can be repeated once again with z_2 and z_3 to arrive at the result

$$A = 6N\int dz_1\,dz_2\,dz_3 \exp(-|z_1|^2 - |z_2|^2 - |z_3|^2)$$

$$\times \sum_{\alpha,\beta} \exp(\lambda c z_\alpha) \exp(\lambda' c' z_\beta^*) \begin{vmatrix} 1 & z_2^* & z_3^{*2} \\ z_1 & z_2^* z_2 & z_3^{*2} z_3 \\ z_1^2 & z_2^* z_2^2 & z_3^{*2} z_3^2 \end{vmatrix}$$

$$\equiv N' \sum_{\alpha,\beta} A_{\alpha\beta}. \tag{12}$$

Each of the integrals $A_{\alpha\beta}$ can be evaluated by integrating each column separately and expanding the last two exponentials in powers of z, z^*, respectively, and making use of the relation

$$\int \exp(-|z|^2) z^m z^{*n} d^2z = \pi\,\delta_{mn}\Gamma(m+1). \tag{13}$$

We then obtain the result

$$A(\zeta) = 9 + 3(1 + \zeta + \zeta^2/6)(e^\zeta - 1), \tag{14}$$

where we have fixed N' using $A|_{\zeta=0} = 9$ and $\zeta = \lambda\lambda'\Delta$.

Equation (14) agrees with the result for $A(\zeta)$ [$= 3a(\zeta)$ in notation of Ref. 3] obtained by Ashmore and Delbourgo. Also from their work it is immediate that matrix superpropagator

$$\langle [\exp(\lambda\gamma_5\Phi(x))]_{\alpha\beta} [\exp(\lambda'\gamma_5\Phi(y))]_{\alpha'\beta'} \rangle_0$$

can be calculated using (14). Thus we completely recover the results of Ashmore and Delbourgo for the case when all components of $\Phi_{\alpha\beta}$ are independent. The method described here has the advantage of being simple and directly applicable to a large class of functions.

In the above formulation we have assumed that all components of Φ are independent and it is therefore not applicable, in the present form, when this condition is not satisfied. This excludes the important case of gravity. We hope to extend the formalism to include the possibility that Φ may be symmetric[9] enabling us to calculate the superpropagator for gravity.

ACKNOWLEDGMENT

I would like to thank Mr. M.S. Sri Ram for many useful discussions.

[1]For a review of nonpolynomial Lagrangian theories, see *Proceedings of the Coral Gables Conference on Fundamental Interactions at High Energies* 1971, edited by A. Salam (Gordon and Breach, New York, 1971).
[2]R. Delbourgo, J. Math. Phys. **13**, 464 (1972).
[3]J. Ashmore and R. Delbourgo, J. Math. Phys. **14**, 176, 569 (1973).
[4]See, for example, R. Delbourgo, A. Salam, and J. Strathdee, Phys. Rev. **187**, 1999 (1969).
[5]We have assumed that all components of Φ are independent.
[6]For evaluation of Jacobians of matrix transformations see W. L. Deemer and I. Olkin, Biometrika **38**, 345 (1951).
[7]J. Ginibre, J. Math. Phys. **6**, 440 (1965).
[8]See, for example, M. L. Mehta, *Random Matrices* (Academic, New York, 1967).
[9]In this case we will get integrals similar to that in Eq. (5), but now U is restricted to be complex symmetric.

Some examples of symmetrical perturbation problems in several complex variables

O. Gherman

*International Centre for Theoretical Physics, Trieste, Italy**

L. Méder

Physics Department, University of Craiova, Romania
(Received 19 March 1975)

We consider two examples of symmetrical perturbation problems in several complex variables and show that special singularities appear. We then discuss the spectral properties of the perturbed operators in connection with the involved symmetries.

Perturbation problems in several complex variables may involve unexpected singularities which disappear when all parameters except one are kept constant.[1] Our aim in this paper is to show how a simple symmetry pattern could generate a certain class of such singularities.

A TWO-DIMENSIONAL EXAMPLE

Rellich[1] considers a two-variable example of supplementary singularities of the perturbed eigenvalues, introduced by more than one complex variable. The eigenvalues of the matrix function

$$T(\kappa_1, \kappa_2) = \begin{pmatrix} \kappa_1 & \kappa_2 \\ \kappa_2 & -\kappa_1 \end{pmatrix}, \tag{1}$$

given by the branches of the function

$$\lambda(\kappa_1, \kappa_2) = (\kappa_1^2 + \kappa_2^2)^{1/2}, \tag{2}$$

are not differentiable at $(\kappa_1, \kappa_2) = (0, 0)$, although for any fixed value of κ_1 or κ_2 (even for $\kappa_1 = 0$ or $\kappa_2 = 0$) the one-variable partial functions generated by them are analytic in every neighborhood of the origin.

By writing (1) in the form

$$T(\kappa_1, \kappa_2) = \kappa_1 \sigma_z + \kappa_2 \sigma_x, \tag{3}$$

where the σ's denote the usual Pauli matrices, its covariance with respect to two-dimensional rotations suggests that it be generalized as follows.

Let us replace (1) by

$$T(x) = f_S(x^2)\,\mathbb{1} + \{f_V(x^2)\,\mathbb{1} + i\sigma_y f_A(x^2)\}\,(x_1\sigma_z + x_2\sigma_x). \tag{4}$$

where x means $(x_1, x_2) \in \mathbb{C}^2$ and x^2 means $x_1^2 + x_2^2$. The f's are assumed to be analytic in some neighborhood U of the origin of the x^2 plane. The eigenvalues will be given by the branches of the function $\lambda : \mathbb{C}^2 \to \mathbb{C}$ defined by

$$\lambda(x) = f_S(x^2) + (\{[f_V(x^2)]^2 + [f_A(x^2)]^2\} \cdot x^2)^{1/2}. \tag{5}$$

The square root in (5) preserves the same singularity as the square root in (2) if $\{f_V^2 + f_A^2\}$ does not vanish anywhere in U. The singularity at $x^2 = 0$ is (is not) removed if $\{f_V^2 + f_A^2\}$ has a zero of odd (even) order at $x^2 = 0$. Any other odd (even) order zero of $\{f_V^2 + f_A^2\}$ does (does not) introduce a supplementary branching point (in the x^2 plane) where the two eigenvalues of 4 coincide. If $f_V = 0$ or $f_A = 0$ (but not both), then the eigenvalues are given by the branches of the function defined by the formula

$$\lambda(x) = f_S(x^2) + f_1(x^2) \cdot (x^2)^{1/2}, \tag{5'}$$

where f_1 denotes f_A or f_V according to which of them is nonzero and the singularities show the same pattern as in the Rellich example. [If $f_1(x^2) = 0$ for some $x^2 \neq 0$, then the two branches of (5') cross incidentally.]

In order to understand the significance of the singular behavior of the eigenvalues of $T(x)$ at $x = 0$, let us examine its eigenprojections (and eigennilpotents) as functions of the vector variable x.

The eigenprojections are given by the function

$$P(x) = \frac{1}{2}\left(\mathbb{1} + \frac{[f_V(x^2)\mathbb{1} + i\sigma_y f_A(x^2)]\,(x_1\sigma_z + x_2\sigma_x)}{(\{[f_V(x^2)]^2 + [f_A(x^2)]^2\} \cdot x^2)^{1/2}}\right); \tag{6}$$

they are connected with the proper eigenvalues by the requirement of taking the same branches of the square root function.

We start by discussing the simplest case (essentially the same as that of Rellich's example) of when f_V and f_A do not bring any complication (i.e., $\{f_V^2 + f_A^2\}(0) \neq 0$ and thus there is a neighborhood $U' \subset U$ of $x^2 = 0$, where the only singularity of the eigenvalues, viewed as functions of x^2, is at the origin). In this case

$$P(x) = \tfrac{1}{2}[\mathbb{1} + \hat{f}(x^2) \cdot \hat{x}],$$

where

$$\hat{x} = (x_1\sigma_z + x_2\sigma_x)/(x^2)^{1/2},$$

and $\hat{f}(x^2)$ is a nonsingular matrix for $x^2 \in V'$ {because $[\hat{f}(x^2)]^2 = -\mathbb{1}$}. It is obvious that $\hat{x}$ [and thus $P(x)$] is defined only if $x^2 \neq 0$, i.e., if x is a nonisotropic vector in $\mathbb{C}^2$ (with the natural scalar product defined by the quadratic form $x \mapsto x^2$). This focuses our attention on $T(x)$ in a neighborhood not only of $x = 0$ but also of an arbitrary nonzero isotropic x. The two eigenvalues of $T(x)$ collapse together at any such x and they must be singular there since they branch (in x^2) at $x^2 = 0$. As x approaches a (nonzero) isotropic x_0 on a path passing through no other isotropic point, the two branches of $\hat{x}$ are unbounded. That means that the two branches of $P(x)$ cannot be continued to x_0, i.e., $T(x_0)$ may not be decomposed into a direct sum as a continuation of the decomposition of $T(x)$ along the path. It is easy to check that $D(x_0) \stackrel{\text{def}}{=} T(x_0) - f_S(0)\,\mathbb{1}$ [where $f_S(0)$ is the common

value of the two eigenvalues of $T(x_0)$] is nilpotent (and generally nonzero):

$$D(x_0)=[f_V(0)x_1+f_A(0)x_2]\sigma_z+[f_V(0)x_2-f_A(0)x_1]\sigma_x,$$

$$[D(x_0)]^2=\{[f_V(0)]^2+[f_A(0)]^2\}(x_1^2+x_2^2)=0. \tag{7}$$

Let us now focus our attention on $x_0=0$. Passing through $x_0=0$ along any complex "curve" with a definite "tangent" at $x_0=0$ is equivalent to passing along a "straight line" directed along this "tangent." It is obvious that the branches of $\hat{x}$ are constant and thus the two branches of $P(x)$ are continuous (even analytical) along a "straight line" $x=\kappa x_1$, $\kappa\in\mathbb{C}$ if $x_1^2\neq 0$ (and thus $x^2\neq 0$), and they have finite limits at $\kappa=0$, allowing them to be continued at $\kappa=0$. However, these limits depend on $\hat{x}$ and thus on the "direction" of the considered "line." This means that $T(0)$ may be decomposed into a direct sum in infinitely many ways, but none of these can be considered the "true" one. Moreover, for a "straight line" $x=\kappa x_1$, $\kappa\in\mathbb{C}$ with $x_1^2=0$, x^2 is 0 for any $\kappa\in\mathbb{C}$ and thus $T(x)$ has a unique eigenprojection ($=\mathbb{I}$) and has an eigen-nilpotent (7) which vanishes at $\kappa=0$ [and, if $f_A(0)\mathbb{I}+i\sigma_y f_V(0)\neq 0$, only there]. All these last conclusions are related to elementary considerations on $T(0)=f_S(0)\mathbb{I}$. We can infer, even at this elementary stage, that the nasty singularities of the eigenvalues in Rellich's example reflect in fact an over-all singular behavior of the whole spectral problem of the "perturbed" $T(x)$.

Concerning the role of the zeros of $\{f_V^2+f_A^2\}$, it may be checked that a zero at $x^2=0$ brings no essentially new structures: Its only effect is upon pole and branching properties of $\lambda(x)$ and $P(x)$. On the other hand, a zero of $\{f_V^2+f_A^2\}$ at $x^2\neq 0$ makes the eigenvalues and eigenprojections have a branching point (in the x^2 plane) connected[2] with the occurrence of a nonvanishing eigen-nilpotent $D(x)=T(x)-f_S(x^2)\cdot\mathbb{I}$ along the corresponding $x^2=\text{const}$ pseudosphere and with the unboundedness of the eigenprojections around this pseudosphere.

AN $O(3,C)$ EXAMPLE[3]

We start with a six-dimensional representation D of a certain covering group $\mathcal{A}$ of $O(3,C)$ of the following form: $SL(2,C)$ elements above proper rotations are mapped into $D^{(3/2,0)}\oplus D^{(1/2,0)}$ matrices[4] and the group elements above space inversion are represented trivially (by the corresponding 6×6 unit matrix).

We think of (this covering group of) $O(3,C)$ as of a transformation group in $\mathbb{C}^3$ [this is possible by means of the canonical covering homomorphism $R:\mathcal{A}\to O(3,C)$] and we shall be concerned with a perturbation problem for an operator-valued function $T:\mathbb{C}^3\to\mathcal{L}(\mathbb{C}^6)$ which is "covariant" in the sense that

$$D(\alpha)T(\mathbf{x})D(\alpha^{-1})=T(R(\alpha)\mathbf{x}).$$

To be definite, we write down the most general expression of such a $T(\mathbf{x})$ in the standard "angular momentum" basis[4] of $D^{(1/2,0)}\oplus D^{(3/2,0)}$: If $\mathbf{J}_{3/2}$, $\mathbf{J}_{1/2}$, $\mathbf{A}_{3/2,1/2}$, and $\mathbf{A}_{1/2,3/2}$ and defined by

$$\mathbf{J}_{3/2}\cdot\mathbf{x}=\frac{1}{2}\begin{pmatrix}3x_3 & \sqrt{3}\,x_- & 0 & 0\\ \sqrt{3}\,x_+ & x_3 & 2x_- & 0\\ 0 & 2x_+ & -x_3 & \sqrt{3}\,x_-\\ 0 & 0 & \sqrt{3}\,x_+ & -3x_3\end{pmatrix}\oplus\mathbb{0}_{2\times 2}, \tag{9a}$$

$$\mathbf{J}_{1/2}\cdot\mathbf{x}=\mathbb{0}_{4\times 4}\oplus\tfrac{1}{2}\begin{pmatrix}x_3 & x_-\\ x_+ & -x_3\end{pmatrix}, \tag{9b}$$

$$\mathbf{A}_{3/2,1/2}\cdot\mathbf{x}=\frac{1}{2}\left(\begin{array}{c|cc}\multirow{4}{*}{$\mathbb{0}_{4\times 4}$} & -\sqrt{3}\,x_- & 0\\ & 2x_3 & -x_-\\ & x_+ & 2x_3\\ & 0 & \sqrt{3}\,x_+\\ \hline \mathbb{0}_{2\times 4} & \multicolumn{2}{c}{\mathbb{0}_{2\times 2}}\end{array}\right) \tag{9c}$$

$$\mathbf{A}_{1/2,3/2}\cdot\mathbf{x}=(\mathbf{A}_{3/2,1/2}\circ\mathbf{x})^{\dagger}, \tag{9d}$$

where $x_\pm$ means $x_1\pm ix_2$, then we take[5]

$$\begin{aligned}T(\mathbf{x})=&A(\mathbf{x}^2)P_{3/2}+a(\mathbf{x}^2)P_{1/2}\\ &+2[B(\mathbf{x}^2)\circ(\mathbf{J}_{3/2}\circ\mathbf{x})+b(\mathbf{x}^2)(\mathbf{J}_{1/2}\circ\mathbf{x})\\ &\quad+\beta(\mathbf{x}^2)\circ(\mathbf{A}_{3/2,1/2}\circ\mathbf{x})+\tilde{\beta}(\mathbf{x}^2)\circ(\mathbf{A}_{1/2,3/2}\circ\mathbf{x})]\\ &+4[C(\mathbf{x}^2)\cdot(\mathbf{A}_{3/2,1/2}\cdot\mathbf{x})^2+\gamma(\mathbf{x}^2)\cdot(\mathbf{A}_{3/2,1/2}\cdot\mathbf{x})\cdot(\mathbf{J}_{1/2}\circ\mathbf{x})\\ &\quad+\tilde{\gamma}(\mathbf{x}^2)\circ(\mathbf{J}_{1/2}\circ\mathbf{x})\circ(\mathbf{A}_{1/2,3/2}\circ\mathbf{x})]\\ &+8D(\mathbf{x}^2)\circ(\mathbf{A}_{3/2,1/2}\cdot\mathbf{x})\cdot(\mathbf{A}_{1/2,3/2}\cdot\mathbf{x})(\mathbf{J}_{3/2}\cdot\mathbf{x}),\end{aligned} \tag{10}$$

where the "invariant coefficients"[6] $A, a, B, b, \beta, \tilde{\beta}, C, \gamma, \tilde{\gamma}, D$ are supposed to be analytic functions of $\mathbf{x}^2$ in some neighborhood of the origin of the $\mathbf{x}^2$ plane. The covariance (8) of T implies that, for any $\mathbf{x}\in\mathbb{C}^3$ and any $\alpha\in\mathcal{A}$, $T(\mathbf{x})$ and $T(R(\alpha)\mathbf{x})$ have the same spectral properties: Their spectrum is the same and their eigenprojections (and possible eigennilpotents) are related by a similarity transformation [by means of $D(\alpha)$]; this last statement shows that their eigenspaces are isomorphic, this isomorphism being given again by $D(\alpha)$. Consequently, spectral properties are determined by *orbits*. As there are essentially three different kinds of orbits [the orbit of $\mathbf{x}=0$, the isotropic cone (without $\mathbf{x}=0$) and nonisotropic pseudospheres], three different situations are to be discussed:

(1) $\mathbf{x}=0$. This case is trivial: $T(0)=A(0)P_{3/2}+a(0)P_{1/2}$ from (10), and this formula completes the spectral decomposition.

(2) All nonzero isotropic vectors form one orbit of $O(3,C)$, so it is enough to study $T(\mathbf{x})$, e.g. for $\mathbf{x}^{(0)}=(\frac{1}{2},i/2,0)$.

(3) Similarly, for each pseudosphere of nonisotropic vectors a representative of the form $\mathbf{x}^{(\kappa)}=(0,0,\kappa)$ may be taken.

The matrix of $T(\mathbf{x}^{(0)})$ is

$$T(\mathbf{x}^{(0)}) = \left(\begin{array}{cccc|cc} A(0) & \sqrt{3}\,B(0) & -\sqrt{3}\,C(0) & -3D(0) & -\sqrt{3}\,\beta(0) & -\sqrt{3}\,\gamma(0) \\ 0 & A(0) & 2B(0) & -\sqrt{3}\,C(0) & 0 & -\beta(0) \\ 0 & 0 & A(0) & \sqrt{3}\,B(0) & 0 & 0 \\ 0 & 0 & 0 & A(0) & 0 & 0 \\ \hline 0 & 0 & \tilde{\beta}(0) & \sqrt{3}\,\tilde{\gamma}(0) & a(0) & b(0) \\ 0 & 0 & 0 & \sqrt{3}\,\tilde{\beta}(0) & 0 & a(0) \end{array}\right). \tag{11}$$

Its spectral decomposition is dependent on the relation between $A(0)$ and $a(0)$. If $A(0)=a(0)$, it has only the eigenvalue $\lambda(\mathbf{x}^{(0)})=A(0)=a(0)$ and a (generally nondecomposable) eigennilpotent $D(\mathbf{x}^0)=T(\mathbf{x}^{(0)})-\lambda(\mathbf{x}^{(0)})\cdot \mathbb{1}_{6\times 6}$. However, if $A(0)\neq a(0)$, then there are two eigenvalues $A(0)$ and $a(0)$ and $T(\mathbf{x}^{(0)})$ can be decomposed as

$$T(\mathbf{x}^{(0)}) = [P_{A(0)}(\mathbf{x}^{(0)}) + D_{A(0)}(\mathbf{x}^{(0)})] + [P_{a(0)}(\mathbf{x}^{(0)}) + D_{a(0)}(\mathbf{x}^{(0)})], \tag{12}$$

where the eigenprojections are

$$P_{A(0)}(\mathbf{x}^{(0)}) = \left(\begin{array}{cccc|cc} 1 & 0 & \dfrac{\sqrt{3}\,\beta(0)\tilde{\beta}(0)}{[A(0)-a(0)]^2} & \dfrac{3[\beta(0)\tilde{\gamma}(0)+\tilde{\beta}(0)\gamma(0)]}{[A(0)-a(0)]^2} - \dfrac{6[B(0)-b(0)]\beta(0)\tilde{\beta}(0)}{[A(0)-a(0)]^3} & -\dfrac{\sqrt{3}\,\beta(0)}{A(0)-a(0)} & \dfrac{-\sqrt{3}\,\gamma(0)}{A(0)-a(0)} + \dfrac{\sqrt{3}\,\beta(0)[B(0)-b(0)]}{[A(0)-a(0)]^2} \\ 0 & 1 & 0 & \dfrac{\sqrt{3}\,\beta(0)\tilde{\beta}(0)}{[A(0)-a(0)]^2} & 0 & -\dfrac{\beta(0)}{A(0)-a(0)} \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ \hline 0 & 0 & \dfrac{\tilde{\beta}(0)}{A(0)-a(0)} & \dfrac{\sqrt{3}\,\tilde{\gamma}(0)}{A(0)-a(0)} - \dfrac{\sqrt{3}\,\tilde{\beta}(0)[B(0)-b(0)]}{[A(0)-a(0)]^2} & 0 & 0 \\ 0 & 0 & 0 & \dfrac{\sqrt{3}\,\tilde{\beta}(0)}{A(0)-a(0)} & 0 & 0 \end{array}\right), \tag{13a}$$

$$P_{a(0)}(\mathbf{x}^{(0)}) = \mathbb{1}_{6\times 6} - P_{A(0)}(\mathbf{x}^{(0)}), \tag{13b}$$

and the eigennilpotents are

$$D_{A(0)}(\mathbf{x}^{(0)}) = \left(\begin{array}{cccc|cc} 0 & \sqrt{3}\,B(0) & -\sqrt{3}\,C(0) - \dfrac{\sqrt{3}\,\beta(0)\tilde{\beta}(0)}{A(0)-a(0)} & -3D(0) - \dfrac{3[\beta(0)\tilde{\gamma}(0)+\tilde{\beta}(0)\gamma(0)]}{A(0)-a(0)} + \dfrac{[6B(0)-3b(0)]\beta(0)\tilde{\beta}(0)}{[A(0)-a(0)]^2} & 0 & \dfrac{-\sqrt{3}\,B(0)\beta(0)}{A(0)-a(0)} \\ 0 & 0 & 2B(0) & -\sqrt{3}\,C(0) - \dfrac{\sqrt{3}\,\beta(0)\tilde{\beta}(0)}{A(0)-a(0)} & 0 & 0 \\ 0 & 0 & 0 & \sqrt{3}\,B(0) & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & \dfrac{\sqrt{3}\,B(0)\tilde{\beta}(0)}{A(0)-a(0)} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right), \tag{14a}$$

$$D_{a(0)}(\mathbf{x}^{(0)}) = \left(\begin{array}{ccc|c|c} 0 & 0 & 0 & -\dfrac{3b(0)\beta(0)\tilde{\beta}(0)}{[A(0)-2(0)]^2} & 0 & -\dfrac{\sqrt{3}\,b(0)\beta(0)}{A(0)-a(0)} \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & -\dfrac{\sqrt{3}\,B(0)\tilde{\beta}(0)}{A(0)-a(0)} & 0 & b(0) \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right) \tag{14b}$$

The above results become slightly more transparent if the standard basis

$$\{|\tfrac{3}{2},\tfrac{3}{2}\rangle,\ |\tfrac{3}{2},\tfrac{1}{2}\rangle,\ |\tfrac{3}{2},-\tfrac{1}{2}\rangle,\ |\tfrac{3}{2},-\tfrac{3}{2}\rangle;\ |\tfrac{1}{2},\tfrac{1}{2}\rangle;\ |\tfrac{1}{2},-\tfrac{1}{2}\rangle\}$$

of $D^{(3/2,0)}\oplus D^{(1/2,0)}$ is rearranged so that the sequence of the "J_z" quantum numbers[7] become monotonically increasing or decreasing: Then $T(\mathbf{x}^{(0)})$ and all the summands in (12) become triangular matrices.

In the case of nonisotropic orbits, the spectral problem of $T(\mathbf{x}^{(\kappa)})$ is an ordinary (one-parameter) perturbation problem, with an analytical κ dependence of the eigenvalues and eigenprojections. However, a better insight into its "kinematics"[8] may be achieved by the same permutation of the basis vectors as above. It is easy to see that this makes $T(\mathbf{x}^{(\kappa)})$ decompose into the direct sum

$$T(\mathbf{x}^{(\kappa)}) = T_{3/2}(\kappa)\oplus T_{1/2}(\kappa)\oplus T_{-1/2}(\kappa)\oplus T_{-3/2}(\kappa) \tag{15}$$

(the indices refer to "J_z" quantum numbers) where $T_{\pm 3/2}(\kappa)$ are one-dimensional,

$$T_{\pm 3/2}(\kappa) = A(\kappa^2) \pm 3\kappa B(\kappa^2), \tag{16}$$

and $T_{\pm 1/2}(\kappa)$ are two-dimensional,

$$T_{\pm 1/2}(\kappa) = \begin{pmatrix} A(\kappa^2)\pm\kappa B(\kappa^2)+\kappa^2 C(\kappa^2)\pm\kappa^3 D(\kappa^2) & 2\kappa[\beta(\kappa^2)\pm\kappa\gamma(\kappa^2)] \\ 2\kappa[\tilde{\beta}(\kappa^2)\pm\kappa\tilde{\gamma}(\kappa^2)] & a(\kappa^2)\pm\kappa b(\kappa^2) \end{pmatrix}. \tag{17}$$

The spectral problem of $T(\mathbf{x}^{(\kappa)})$ reduces practially to that of $T_{\pm 1/2}(\kappa)$.

For any given κ the eigenvalues of $T(\mathbf{x}^{(\kappa)})$ are paired by the permutation induced by $\kappa \rightleftarrows -\kappa$. This is a consequence of $T(-\mathbf{x})$ being similar to $T(\mathbf{x})$: In particular, for any rotation bringing $(0,0,1)$ into $(0,0,-1)$ and any $\boldsymbol{\alpha}_0 \in \mathcal{A}$ which is mapped by R into this rotation

$$T(\mathbf{x}^{(-\kappa)}) = T(-\mathbf{x}^{(\kappa)}) = D(\boldsymbol{\alpha}_0)\,T(\mathbf{x}^{(\kappa)})\,D(\boldsymbol{\alpha}_0^{-1}). \tag{18}$$

The "kinematical" nature of the decomposition (15) may be understood by realizing that:

(i) the stability group of $\mathbf{x}^{(\kappa)}$ contains the group generated by $\frac{1}{2}\sigma_z$ and $\frac{1}{2}\sigma_z$,

(ii) the representation of this group, contained in the representation $D^{(3/2,0)}\oplus D^{(1/2,0)}$ of $SL(2,C)$ is deco. posed into (multiples of) inequivalent irreducible representations, labelled exactly by the "J_z" quantum numbers.

The decomposition (12) has no similar "kimematical" origin because the representation of the stability group of $\mathbf{x}_0$ [containing products of (trivially represented) inversion elements with members of the group of the matrices of the form $\left(\begin{smallmatrix}1 & \xi\\ 0 & 1\end{smallmatrix}\right)$, $\xi\in\mathbb{C}$] contained in the representation $D^{(3/2,0)}$ and $D^{(1/2,0)}$ of $SL(2,C)$:

(i) is reducible but not completely reducible, and

(ii) all representations induced by D on minimal invariant subspaces and on "minimal"[9] factor spaces are trivial [i.e., they are given by $\left(\begin{smallmatrix}1 & \xi\\ 0 & 1\end{smallmatrix}\right)\to$ the identity operator].

The next step is again to see how the spectral structure of $T(\mathbf{x})$ behaves when $\mathbf{x}^2 \to 0$, i.e., when $\mathbf{x}$ approaches 0 or a nonzero isotropic vector. The behavior should be suspected to be singular in both cases as

(i) both orbits with $\mathbf{x}^2 = 0$ differ from nonisotropic orbits *qualitatively* and

(ii) in our two-dimensional example [where $T(x)$ given by (4) was, as it is easy to check, covariant in the sense of (8) with respect to the identical representation of the Abelian group generated by $\frac{1}{2}\sigma_y$ and $\frac{1}{2}\sigma_y$, covering by the canonical spinor homomorphism the group $SO(2,C)$ of proper orthogonal transformations of $\mathbb{C}^2$] $T(x)$ was generally singular near any point lying on any isotropic orbit.

Let $\mathbf{x}$ approach any nonzero $\mathbf{x}_0$ along a path which intersects the isotropic cone in no other point except $\mathbf{x}_0$. Then, along this path $T(\mathbf{x})$ may be represented, for any $\mathbf{x}\neq\mathbf{x}_0$, as

$$T(\mathbf{x}) = D(\boldsymbol{\alpha}(\mathbf{x}))\,T(\tilde{\mathbf{x}})\,D(\boldsymbol{\alpha}^{-1}(\mathbf{x})), \tag{19}$$

where $\tilde{\mathbf{x}} = (0,0,\lambda)$ is in the same orbit as $\mathbf{x}$ and $\boldsymbol{\alpha}(\mathbf{x})$ is chosen so that $R(\boldsymbol{\alpha}(\mathbf{x}))\tilde{\mathbf{x}} = \mathbf{x}$. Now, a possible choice for $\boldsymbol{\alpha}(\mathbf{x})$ is

$$\boldsymbol{\alpha}_0(\mathbf{x}) = \begin{pmatrix} 1/\sqrt{2} & -x_3/\sqrt{2}\,\lambda \\ (\lambda - x_3)/\sqrt{2}\,x_- & (\lambda + x_3)/\sqrt{2}\,\lambda \end{pmatrix}, \tag{20}$$

and any other is obtained therefrom by multiplication on the right by an arbitrary diagonal $SL(2,C)$ matrix, i.e., with

$$\boldsymbol{\alpha}_\zeta = \begin{pmatrix} \zeta & 0 \\ 0 & \zeta^{-1} \end{pmatrix}, \tag{21}$$

with an arbitrary $\zeta \in \mathbb{C}$. This ambiguity does not affect $T(\mathbf{x})$ as $R(\boldsymbol{\alpha}_\zeta)\tilde{\mathbf{x}} = \tilde{\mathbf{x}}$ and D in (8) is a representation. The eigenprojections (and eigennilpotents) of $T(\mathbf{x})$ are functions of $T(\mathbf{x})$[10,11]; thus they are related to those of $T(\tilde{\mathbf{x}})$ by the same similarity transformation [by means of $D(\boldsymbol{\alpha}(x))$ in (9)] which relates $T(\mathbf{x})$ to $T(\tilde{\mathbf{x}})$.

The matrix of $D(\boldsymbol{\alpha}_0(\mathbf{x}))$ for an $\mathbf{x}$ in the standard basis is a direct sum of $\boldsymbol{\alpha}_0(\mathbf{x})$ itself and of $D^{(3/2,0)}(\boldsymbol{\alpha}_0(\mathbf{x}))$; the standard matrix of the latter is[12]

$$D^{(3/2,0)}\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} = \begin{pmatrix} \alpha^3 & \sqrt{3}\,\alpha^2\beta & \sqrt{3}\,\alpha\beta^2 & \beta^3 \\ \sqrt{3}\,\alpha^2\gamma & \alpha(\alpha\delta + 2\beta\gamma) & \beta(\beta\gamma + 2\alpha\delta) & \sqrt{3}\,\beta^2\delta \\ \sqrt{3}\,\alpha\gamma^2 & \gamma(\beta\gamma + 2\alpha\delta) & \delta(\alpha\delta + 2\beta\gamma) & \sqrt{3}\,\beta\delta^2 \\ \gamma^3 & \sqrt{3}\,\gamma^2\delta & \sqrt{3}\,\gamma\delta^2 & \delta^3 \end{pmatrix}, \tag{22}$$

with coefficients which are polynomials in those of its arguments. This means that if $\mathbf{x} \to \mathbf{x}_0$, $D(\boldsymbol{\alpha}_0(\mathbf{x}))$ behaves singularly (because of the negative powers of λ occurring in it). Thus the eigenprojections of $T(\mathbf{x})$ may also be expected to be singular when $\mathbf{x} \to \mathbf{x}_0$.

We showed previously in (15) that the decomposition of unity associated with $T(\mathbf{x}^{(\lambda)}) = T(\tilde{\mathbf{x}})$ is generally a refinement of that of J_z. By standard arguments this implies that the decomposition of unity for $T(\mathbf{x})$ (with $\mathbf{x}^2 \neq 0$) is a refinement of that of $\mathbf{J}\cdot\mathbf{x}$. But the eigenprojections of $\mathbf{J}\cdot\mathbf{x}$ may be obtained from those of J_z by the similarity transformation mediated by $D(\boldsymbol{\alpha}_0)$ and an explicit computation shows that they *are* singular (as their matrix elements contain negative powers of λ). This means that (at least some of) the eigenprojections of $T(\mathbf{x})$ must have (at least) these singularities in λ. (We only discuss here the behavior of these singularities on a path reaching $\mathbf{x}_0$ and not in a whole neighborhood of $\mathbf{x}_0$.)

The pairs of eigenvalues of $T(\mathbf{x})$ may be regarded as "branches" of functions which are analytical in $(\mathbf{x}^2)^{1/2}$. When $\mathbf{x} \to \mathbf{x}_0$, these "branches" collapse together. Thus we see that the analytical features of the spectral behavior of $T(\mathbf{x})$ in a neighborhood of any (nonzero) isotropic vector are similar to those met in our two-dimensional example. With inessential modification of the arguments, our conclusions concerning the spectral problem around the origin can also be taken over from the previous example. The singular behavior of the origin is related to its "having no direction," i.e.,

(i) to its being stable under any rotation [while $T(\mathbf{x})$ is nontrivially affected by rotations of $\mathbf{x}$] and

(ii) to its property that any of its neighborhoods contains both isotropic and nonisotropic vectors.

SUPPLEMENTARY REMARKS AND CONJECTURES

We are now able to draw some general conclusions concerning the behavior of the spectral properties of a symmetrically perturbed operator near "singular" orbits.

Let G be a Lie group and $\mathbb{E}$ a finite-dimensional complex vector space; let $R: G \to \mathcal{L}(\mathbb{E})$ be a group homomorphism, allowing G to be regarded as a (left) transformation group in $\mathbb{E}$. Let $\mathcal{D} \in \mathbb{E}$ be a domain symmetrical under G-action and containing the origin of $\mathbb{E}$. Let $\mathbb{H}$ be another vector space with a linear representation $D: G \to \mathcal{L}(\mathbb{H})$ of G acting in it. The representation D generates another representation $\tilde{D}: G \to \mathcal{L}(\mathcal{L}(\mathbb{H}))$ by the usual rule: for any $g \in G$ and $B \in \mathcal{L}(\mathbb{H})$ set

$$\tilde{D}(g)B = D(g)AD(g^{-1}). \tag{23}$$

Let, further, $T: \mathcal{D} \to \mathcal{L}(\mathbb{H})$ be a map, equivariant with respect to R and $\mathcal{D}$, i.e., let the diagram

$$\begin{array}{ccc} \mathcal{D} & \xrightarrow{T} & \mathcal{L}(\mathbb{H}) \\ {\scriptstyle R(g)}\downarrow & & \downarrow{\scriptstyle \tilde{D}(g)} \\ \mathcal{D} & \xrightarrow{T} & \mathcal{L}(\mathbb{H}) \end{array}$$

be commutative. This means a covariance scheme which generalizes (8): For any $g \in G$ and $x \in \mathcal{D}$

$$T(R(g)x) = D(g)\,T(x)\,D(g^{-1}). \tag{24}$$

It is obvious that such a scheme must impose severe restrictions on the spectral properties of each $T(x)$ ($x \in \mathcal{D}$). We do not intend now to go into an exhaustive study of its implications. Nevertheless, without any new or deep ideas some immediate conclusions may be formulated. The set of eigenvalues of $T(x)$ is G-invariant, i.e., it is constant along G-orbits in $\mathcal{D}$. Thus the eigenvalue set may be viewed as a function of these orbits. Up to an isomorphism of the eigenspaces, the spectral decomposition of $T(x)$ may also be viewed as a function of the same orbits. [This isomorphism is given, obviously, for an x and x' in the same G-orbit, by any $\tilde{D}(g)$ where g with $R(g)x = x'$ may be any member of a right coset of the stability group of x.]

Our previously stated examples and others involving orthogonal groups make us conjecture that the "kinematical" (i.e., symmetry-generated) singularities might be understood in terms of the topology of the orbit space $\mathcal{D}/R(G)$. Such singularities of the spectral structure of a symmetrically perturbed operator seem to occur along "singular" orbits. [We call an orbit whose stability group is $K \subset G$ "singular" if it has no neighborhood containing orbits whose stability group (in G) are not isomorphic to K.] In particular, the origin of $\mathbb{E}$ ("no perturbation") should always be a singular point of the spectral behavior of a symmetrically perturbed operator.

In addition, our first example shows that an en-

larging of the symmetry group may rule out some "non-kinematical" singularities. [There, a restriction of the form $f_V = 0$ or $f_A = 0$ could be interpreted as replacing covariance merely with respect to (the spinor covering group of) $SO(2, C)$ by covariance with respect to (the spinor covering group of) the full $O(2, C)$.]

ACKNOWLEDGMENTS

One of the authors (OG) wishes to thank Professor Abdus Salam, the International Atomic Energy Agency, and UNESCO for hospitality at the International Centre for Theoretical Physics, Trieste. Both authors wish to thank their colleagues from Craiova, L. Saliu, G. Steinbrecher, and L. Tătaru, for illuminating discussions.

*Permanent address: Physics Department, University of Craiova, Romania.

[1]F. Rellich, Math. Ann. **113**, 600 (1937).

[2]Excepting the case when $f_V(x^2) = f_A(x^2) = 0$ and $T(x)$ reduces (along the corresponding $x^2 = \text{const}$ pseudosphere) to $f_S(x^2) \cdot \mathbb{1}$.

[3]This may be considered as arising from a generalization and a continuation to (nonphysical) complex field intensities of the spectral problem of the Zeeman effect in a magnetic field of any direction and intensity for a spin 1 coupled with a spin $\frac{1}{2}$.

[4]H. Bacry, *Leçons sur la théorie des groupes et les symétries des particules élémentaries* (Gordon and Breach, Paris, London, and New York,1967), Chap. 6, Sec. 10, pp. 243—44, Chap. 7, pp. 275—95.

[5]$P_{3/2} = \mathbb{1}_{4\times4} \oplus \mathbb{0}_{2\times2}$ and $P_{1/2} = \mathbb{0}_{4\times4} \oplus \mathbb{1}_{2\times2}$ in the notations of (9a, b).

[6]Any other form involves another (equivalent) set of linearly independent spinor covariants and appropriate linear combination of the invariants.

[7]We denote $\mathbf{J}_{3/2} + \mathbf{J}_{1/2}$ by $\bar{J}$ and $\mathbf{J} \cdot \mathbf{x}^{(0)}$ by $\mathbf{J}_z$.

[8]We shall use the terms "kinematics" and "kinematical" throughout for features arising only from the symmetry scheme in (8).

[9]We say that M_1/M_2 is "minimal" if (M_1 and M_2 being invariant subspaces with $M_1 \supset M_2$, $M_1 \neq M_2$) there is no invariant subspace M' with $M_1 \supset M' \supset M_2$, $M' \neq M_1$, $M' \neq M_2$.

[10]They may be represented by suitable integrals involving the resolvent of $T(x)$.

[11]T. Kato, *Perturbation Theory for Linear Operators* (Springer, New York, 1966), Chap. I, Sec. 6, p. 55.

[12]N. J. Vilenkin, *Special Functions and the Theory of Group Representations* (American Mathematical Society, Providence, R.I., 1968), Chap. III, Sec. 3, p. 116.

Determination of nearest neighbor degeneracy on a one-dimensional lattice

C. C. Yan

Instituto de Física, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil
(Received 27 June 1975)

A general method is proposed for the determination of the degeneracy of nearest neighbor pairs on a one-dimensional lattice. Explicit results are given for the special cases involving two and three kinds of molecules.

I. INTRODUCTION

In statistical mechanical models involving nearest neighbor interactions, the most general form of the interaction energy on a one-dimensional lattice can be written as follows:

$$E(\{n_i\},\{n_{ij}\})=\sum_i V_i n_i+\tfrac{1}{2}\sum_{i,j}(1+\delta_{ij})V_{ij}n_{ij}, \qquad (1.1)$$

where $i,j=1,2\cdots k$, indicate different kinds of molecules and run through these values independently in the summations; n_i is the number of ith kind of molecules; n_{ij} is the number of nearest neighbor pairs between ith and jth kinds; V_i and V_{ij} are the energies for each ith kind of molecule and for each (ij) kind of pair. Suppose that the ith kind of molecules are rigid rods of length $(l_i-1)a$ with a being the unit length of the linear lattice, then each ith kind of molecule will occupy l_i lattice sites. Hence, assuming the total lattice sites to be L, we have

$$\sum_i l_i n_i=L. \qquad (1.2)$$

If the total number of molecules is N, we have further

$$\sum_i n_i=N. \qquad (1.3)$$

The partition function corresponding to a definite fixed set of $\{n_i\}$ is given by

$$Z(\{n_i\})=\sum_{\{n_{ij}\}}\exp[-\beta E(\{n_i\},\{n_{ij}\})] \qquad (1.4)$$

with $\beta=1/kT$. The partition function corresponding to a fixed (N,L) is

$$Z(N,L)=\sum_{\{n_i\}}Z(\{n_i\}), \qquad (1.4')$$

where the summations of $\{n_{ij}\}$ and $\{n_i\}$ are to be carried out in accordance with the given restrictions such as (1.2) and (1.3).

In the case with only two kinds of molecules, the problem can be reduced to several special cases. If we take $V_1=-V_2=H$, $V_{11}=V_{22}=-V_{12}=J$, we have the well-known Ising model[1]; if we take $V_1=U$, $V_2=0$, $V_{11}=V$, $V_{22}=V_{12}=0$, we have the lattice gas model.[2] These models are essentially equivalent, and various methods[3] can be applied to solve the problem without having to determine degeneracy of nearest neighbor pairs. However, in general, knowledge of the degeneracy associated with each type of nearest neighbor pair will be required in order to calculate partition function and thermodynamic quantities.

In a recent article, McQuistan[4] has determined the degeneracy of nearest neighbor pairs for the special case involving two kinds of molecules with $l_1=1$ and $l_2=2$ (one kind of molecule is the vacant site and the other kind of molecule is the dumbbell) generalizing his previous results for the case of $l_1=l_2=1$.

In the following, we shall show that for a system with two kinds of molecules, the degeneracy of nearest neighbor pairs can be readily determined for arbitrary l_1 and l_2. We shall generalize the results to a system involving three kinds of molecules. A further generalization to the situation involving more kinds of molecules can be similarly done with increasing labor.

II. A SYSTEM WITH TWO KINDS OF MOLECULES

We place n_1 molecules in a row thus creating maximum number of n_{11} pairs which is

$$(n_{11})_{\max}=n_1-1. \qquad (2.1)$$

Now if we put one molecule of the second kind at one end of this row, we create one (1,2) pair without affecting the (1,1) pairs. However, if we put one molecule of the second kind at any intervals of the row, we destroy one (1,1) pair and create two (1,2) pairs. Now if we place one molecule of the second kind next to any molecule of the second kind already in the row, we create one (2,2) pair without affecting the (1,1) pairs or (1,2) pairs. It is clear that the determination of the degeneracy will depend on the conditions at the ends. Denoting the end conditions by $[i,j]$, meaning that the left end is occupied by the ith kind of molecule and the right end by the jth kind of molecule, we have the following four cases.

(1) [1,1]: Let x be the number of intervals of (1,1) pairs selected to place one molecule of the second kind at each of those x intervals. Then we have

$$2x=n_{12}, \qquad (2.2)$$

$$n_1-1-x=n_{11}=(n_{11})_{\max}-x. \qquad (2.3)$$

Now there are (n_2-x) numbers of molecules of the second kind left unused. Since each time we place one of those unused molecules next to any of the x molecules already placed in the row we increase one (2,2) pair without altering n_{11} and n_{12}, we conclude that

$$n_2-x=n_{22}. \qquad (2.4)$$

The number of ways to select x intervals from a total of (n_1-1) available intervals is given by

$$\binom{n_1-1}{x} \tag{2.5}$$

To place (n_2-x) indistinguishable molecules into the distinguishable x positions is a Bose–Einstein type of distribution, and the number of ways is given by

$$\binom{n_2-1}{x-1}. \tag{2.6}$$

Thus the degeneracy for this case of given x with the end condition $[1,1]$ is

$$A_{11}=\binom{n_1-1}{x}\binom{n_2-1}{x-1}. \tag{2.7}$$

Here and in the following, it is to be understood that $\binom{p}{q}=0$ if $q>p$ or $p,\ q<0$.

(2) $[1,2]$: In this case, we first place one molecule of the second kind at the right end of the row, then select x intervals to place x molecules of the second kind. We thus have

$$2x+1=n_{12}, \tag{2.8}$$

$$n_1-1-x=n_{11}, \tag{2.9}$$

$$n_2-1-x=n_{22}. \tag{2.10}$$

After selecting x intervals from (n_1-1) total intervals, we now have to place (n_2-1-x) molecules into $(x+1)$ positions. Hence the result

$$A_{12}=\binom{n_1-1}{x}\binom{n_2-1}{x}. \tag{2.11}$$

(3) $[2,1]$: By symmetry this gives the same result as (2.11); hence

$$A_{21}=A_{12}=\binom{n_1-1}{x}\binom{n_2-1}{x}. \tag{2.12}$$

(4) $[2,2]$: By the same reasoning, we have the following:

$$2x+2=n_{12}, \tag{2.13}$$

$$n_1-1-x=n_{11}, \tag{2.14}$$

$$n_2-2-x=n_{22}, \tag{2.15}$$

$$A_{22}=\binom{n_1-1}{x}\binom{n_2-1}{x+1}. \tag{2.16}$$

If condition (1.3) is relaxed, we have from (1.2)

$$n_1=(L-l_2n_2)/l_1; \tag{2.17}$$

thus the degeneracy corresponding to odd n_{12} is obtained from a combination of (2.8), (2.11), (2.12), and (2.17) as

$$A(n_{12})=2A_{12}=2\binom{n_2-1}{(n_{12}-1)/2}\binom{(L-l_2n_2)/l_1-1}{(n_{12}-1)/2} \quad (n_{12}\text{ odd}). \tag{2.18}$$

Similarly for even n_{12}, we obtain from (2.2), (2.7), (2.13), (2.16), and (2.17)

$$A(n_{12})=\binom{n_2-1}{(n_{12}/2)-1}\binom{(L-l_2n_2)/l_1-1}{n_{12}/2}+\binom{n_2-1}{n_{12}/2}\binom{(L-l_2n_2)/l_1-1}{(n_{12}/2)-1} \quad (n_{12}\text{ even}). \tag{2.19}$$

To obtain the degeneracy in terms of n_{22} or n_{11}, we simply add up all the A_{ij} and express the x in terms of n_{22} or n_{11}. Thus

$$\begin{aligned}A(n_{22})&=\binom{n_2-1}{n_2-n_{22}-1}\binom{(L-l_2n_2)/l_1-1}{n_2-n_{22}}+2\binom{n_2-1}{n_2-n_{22}-1}\binom{(L-l_2n_2)/l_1-1}{n_2-n_{22}-1}\\&\quad+\binom{n_2-1}{n_2-n_{22}-1}\binom{(L-l_2n_2)/l_1-1}{n_2-n_{22}-2}\\&=\binom{n_2-1}{n_{22}}\binom{(L-l_2n_2)/l_1+1}{n_2-n_{22}},\end{aligned} \tag{2.20}$$

$$A(n_{11})=\binom{n_2+1}{(L-l_2n_2)/l_1-n_{11}}\binom{(L-l_2n_2)/l_1-1}{n_{11}}. \tag{2.21}$$

The partition function is given by

$$\begin{aligned}Z(\{n_i\})&=\exp(-\beta\varphi_1)\sum_x\exp(-\beta\varphi_3x)\\&\quad\times\left[\binom{n_1-1}{x}\binom{n_2-1}{x-1}\exp(-\beta\varphi_2)+2\binom{n_1-1}{x}\binom{n_2-1}{x}+\binom{n_1-1}{x}\binom{n_2-1}{x+1}\exp(\beta\varphi_2)\right],\end{aligned} \tag{2.22}$$

where

$$\begin{aligned}\varphi_1&=\sum_{i=1}^{2}V_in_i+\sum_{i=1}^{2}(n_i-1)V_{ii},\\\varphi_2&=V_{22}-V_{12},\\\varphi_3&=2V_{12}-V_{11}-V_{22}.\end{aligned} \tag{2.23}$$

III. A SYSTEM WITH THREE KINDS OF MOLECULES

In the previous section we have started from a situation with $n_{11}=n_1-1$, $n_{12}=0$, $n_{22}=0$ and ended up with four different possible arrangements. Now we shall start from each of these different arrangements to generate possible arrangements with introduction of the third kind of molecules. The principle is to create n_{13}, n_{23}, n_{33}, from n_{11}, and n_{22} without changing n_{12}. Since this process will depend on the end conditions, we shall again divide the discussion into the following cases.

(A1) $[1,1]\to[1,1]$: A notation of $[i,j]\to[l,k]$ means that in the original arrangement the left end is occupied by the ith kind while the right end is occupied by the jth kind, and, with the introduction of new kinds of molecules, the end conditions change to $[l,k]$.

Let y be the number of the third kind of molecules placed into y intervals of n_{11}, and z be the number of the third kind of molecules placed into z intervals of n_{22}; then in this case we have

$$\begin{aligned} &n_{12}=2x, \quad n_{11}=n_1-1-x-y,\\ &n_{13}=2y, \quad n_{22}=n_2-x-z,\\ &n_{23}=2z, \quad n_{33}=n_3-y-z. \end{aligned} \tag{3.1}$$

Since after establishing $[1,1]$ with two kinds of molecules, we have $\binom{n_1-1-x}{y}$ ways of choosing y intervals from (n_1-1-x) intervals of $(1,1)$ pairs; $\binom{n_2-x}{z}$ ways of choosing z intervals from (n_2-x) intervals of $(2,2)$ pairs and $\binom{n_3-1}{y+z-1}$ ways of distributing (n_3-y-z) unused and indistinguishable third kind of molecules into $(y+z)$ intervals. We have all together the following degeneracy:

$$A_1=\binom{n_3-1}{y+z-1}\binom{n_2-x}{z}\binom{n_1-1-x}{y}\binom{n_2-1}{x-1}\binom{n_1-1}{x}. \tag{3.2}$$

(A2) $[1,1]\to[1,3]$: Applying the same method, we have for this case

$$\begin{aligned} &n_{12}=2x, \quad n_{11}=n_1-1-x-y,\\ &n_{13}=2y+1, \quad n_{22}=n_2-x-z\\ &n_{23}=2z, \quad n_{33}=n_3-1-y-z, \end{aligned} \tag{3.3}$$

$$A_2=\binom{n_3-1}{y+z}\binom{n_2-x}{z}\binom{n_1-1-x}{y}\binom{n_2-1}{x-1}\binom{n_1-1}{x}. \tag{3.4}$$

(A3) $[1,1]\to[3,1]$: The results are the same as (A2).

(A4) $[1,1]\to[3,3]$:

$$\begin{aligned} &n_{12}=2x, \quad n_{11}=n_1-1-x-y,\\ &n_{13}=2y+2, \quad n_{22}=n_2-x-z,\\ &n_{23}=2z, \quad n_{33}=n_3-2-y-z, \end{aligned} \tag{3.5}$$

$$A_4=\binom{n_3-1}{y+z+1}\binom{n_2-x}{z}\binom{n_1-1-x}{y}\binom{n_2-1}{x-1}\binom{n_1-1}{x}. \tag{3.6}$$

(B1) $[1,2]\to[1,2]$:

$$\begin{aligned} &n_{12}=2x+1, \quad n_{11}=n_1-1-x-y,\\ &n_{13}=2y, \quad n_{22}=n_2-1-x-z,\\ &n_{23}=2z, \quad n_{33}=n_3-y-z, \end{aligned} \tag{3.7}$$

$$B_1=\binom{n_3-1}{y+z-1}\binom{n_2-1-x}{z}\binom{n_1-1-x}{y}\binom{n_2-1}{x}\binom{n_1-1}{x}. \tag{3.8}$$

(B2) $[1,2]\to[1,3]$:

$$\begin{aligned} &n_{12}=2x+1, \quad n_{11}=n_1-1-x-y,\\ &n_{13}=2y, \quad n_{22}=n_2-1-x-z,\\ &n_{23}=2z+1, \quad n_{33}=n_3-1-y-z, \end{aligned} \tag{3.9}$$

$$B_2=\binom{n_3-1}{y+z}\binom{n_2-1-x}{z}\binom{n_1-1-x}{y}\binom{n_2-1}{x}\binom{n_1-1}{x}. \tag{3.10}$$

(B3) $[1,2]\to[3,2]$:

$$\begin{aligned} &n_{12}=2x+1, \quad n_{11}=n_1-1-x-y,\\ &n_{13}=2y+1, \quad n_{22}=n_2-1-x-z,\\ &n_{23}=2z, \quad n_{33}=n_3-1-y-z, \end{aligned} \tag{3.11}$$

$$B_3=B_2.$$

(B4) $[1,2]\to[3,3]$:

$$\begin{aligned} &n_{12}=2x+1, \quad n_{11}=n_1-1-x-y,\\ &n_{13}=2y+1, \quad n_{22}=n_2-1-x-z,\\ &n_{23}=2z+1, \quad n_{33}=n_3-2-y-z, \end{aligned} \tag{3.12}$$

$$B_4=\binom{n_3-1}{y+z+1}\binom{n_2-1-x}{z}\binom{n_1-1-x}{y}\binom{n_2-1}{x}\binom{n_1-1}{x}.$$

(C1) $[2,1]\to[2,1]$: The results are the same as (B1).

(C2) $[2,1]\to[2,3]$: This case gives the same results as as (B3).

(C3) $[2,1]\to[3,1]$: Same as (B2).

(C4) $[2,1]\to[3,3]$: Same as (B4).

(D1) $[2,2]\to[2,2]$:

$$\begin{aligned} &n_{12}=2x+2, \quad n_{11}=n_1-1-x-y,\\ &n_{13}=2y, \quad n_{22}=n_2-2-x-z,\\ &n_{23}=2z, \quad n_{33}=n_3-y-z, \end{aligned} \tag{3.14}$$

$$D_1=\binom{n_3-1}{y+z-1}\binom{n_2-2-x}{z}\binom{n_1-1-x}{y}\binom{n_2-1}{x+1}\binom{n_1-1}{x}. \tag{3.15}$$

(D2) $[2,2]\to[2,3]$:

$$\begin{aligned} &n_{12}=2x+2, \quad n_{11}=n_1-1-x-y,\\ &n_{13}=2y, \quad n_{22}=n_2-2-x-z,\\ &n_{23}=2z+1, \quad n_{33}=n_3-1-y-z, \end{aligned} \tag{3.16}$$

$$D_2=\binom{n_3-1}{y+z}\binom{n_2-2-x}{z}\binom{n_1-1-x}{y}\binom{n_2-1}{x+1}\binom{n_1-1}{x}. \tag{3.17}$$

(D3) $[2,2]\to[3,2]$: Same as (D2).

(D4) $[2,2]\to[3,3]$:

$$n_{12}=2x+2, \quad n_{11}=n_1-1-x-y,$$

$$n_{13}=2y, \qquad n_{22}=n_2-2-x-z, \tag{3.18}$$
$$n_{23}=2z+2, \quad n_{33}=n_3-2-y-z,$$

$$D_4=\binom{n_3-1}{y+z+1}\binom{n_2-2-x}{z}\binom{n_1-1-x}{y}\binom{n_2-1}{x+1}\binom{n_1-1}{x}. \tag{3.19}$$

If we wish to express the above results in terms of any three of n_{ij} instead of x, y, z, all what we have to do is to express x, y, z in terms of the desired set of $\{n_{ij}\}$ in each case and add up all the nonconflicting cases. If in one case n_{ij} is required to be odd and in the other case the same n_{ij} is required to be even, then it is an example considered to be conflicting. If we write

$$n_1=(L-l_2n_2-l_3n_3)/l_1, \tag{3.20}$$

the restriction due to molecular sizes can be taken care of. Finally the partition function can be written as

$$Z(\{n_i\})=\exp[-\beta(C_0-V_{11})]\sum_{x,y,z}\exp[-\beta(C_1x+C_2y+C_3z)]$$
$$\times\{A_1+2A_2\exp(-\beta C_5)+A_4\exp(-2\beta C_5)$$
$$+2B_1\exp(-\beta C_4)+2B_2\exp[-\beta(C_4+C_5)]$$
$$+2B_2\exp[-\beta(C_4+C_6)]$$
$$+2B_4\exp[-\beta(C_4+C_5+C_6)]+D_1\exp(-2\beta C_4)$$
$$+2D_2\exp[-\beta(2C_4+C_6)]+D_4\exp[-2\beta(C_4+C_6)]\}, \tag{3.21}$$

where

$$C_0=\sum_{i=1}^{3}(V_i+V_{ii})n_i, \quad C_1=2V_{12}-V_{11}-V_{22},$$
$$C_2=2V_{13}-V_{11}-V_{33}, \quad C_3=2V_{23}-V_{22}-V_{33},$$
$$C_4=V_{12}-V_{22}, \quad C_5=V_{13}-V_{33}, \quad C_6=V_{23}-V_{33}.$$

[1]E. Ising, Z. Physik **31**, 253 (1925).
[2]T.D. Lee and C.N. Yang, Phys. Rev. **87**, 410 (1952).
[3]G.F. Newell and E.W. Montroll, Rev. Mod. Phys. **25**, 253 (1953).
[4]R.B. McQuistan, J. Math. Phys. **15**, 1192 (1974); J. Math. Phys. **13**, 1317 (1972).

A solution of the Korteweg–de Vries equation in a half-space bounded by a wall

H. E. Moses

Air Force Cambridge Research Laboratories, Hanscom Air Force Base, Massachusetts 01731
(Received 14 July 1975)

We give a solution of the Korteweg–de Vries equation in the half-space $0 < r < \infty$ with the boundary condition $V(0) = 0$. The boundary condition may be interpreted as the requirement that the plane which bounds the half-space be a rigid wall. Aside from possible physical interest, this solution, which is obtained from one of the potentials for the radial Schrödinger equation which do not scatter, appears to indicate that the radial Schrödinger equation and the corresponding Gel'fand–Levitan equation play a role in the case of the half-space bounded by a wall similar to that of the one-dimensional Schrödinger equation $(-\infty < x < \infty)$ and its corresponding Gel'fand–Levitan equation in the more usual full space treatment of the KdV equation. A possible interpretation of the solution presented in this paper is that it corresponds to the reflection of a wave by a wall, in which the incident wave is singular and the reflected wave is nonsingular but highly dispersive.

1. THE KORTEWEG-DeVRIES EQUATION. THE RADIAL GELFAND-LEVITAN EQUATION. POTENTIALS WITH ZERO SCATTERING PHASE AND AMPLITUDE

The Korteweg—De Vries equation for $V(x,t)$, namely,

$$V_t - 6VV_x + V_{xxx} = 0, \tag{1}$$

has become a subject of intense study, since it was shown (Ref. 1) that general classes of solutions can be obtained in the full space $-\infty < x < \infty$ by mapping the problem to the inverse problem for the Schrödinger equation in one dimension. A rather complete bibliography is given in Ref. 2.

One of the important classes of solutions are obtained from potentials for which the reflection coefficient is zero. Such potentials, derived in Ref. 3, lead to closed form solutions of (1) which correspond to the motion and interaction of highly stable modes called solitons.

It is natural to see whether one can find analogous solutions in the half-space $0 \leq r < \infty$. (We shall use r as the space variable instead of x when working in the half-space, since r represents the radial variable in the associated radial Schrödinger equation.) In the case of the radial Schrödinger equation the continuous spectrum eigenfunctions $\psi(r|k)$ satisfy

$$-\frac{d^2}{dr^2}\psi(r|k) + V(r)\psi(r|k) = k^2\psi(r|k), \tag{2}$$

where we require

$$\psi(0|k) = 0, \quad \frac{d}{dr}\psi(0|k) = \left(\frac{2}{\pi}\right)^{1/2} k \tag{3}$$

as boundary conditions. It can then be shown

$$\lim_{r\to\infty}[\psi(r|k) - (2/\pi)^{1/2}|f(k)|\sin(kr+\delta)] = 0, \tag{4}$$

where $f(k)$ is the Jost function and $\delta = \delta(k)$ is the phase shift. The eigenfunctions for the point eigenvalues E_i satisfy the equation

$$-\frac{d^2}{dr^2}\psi_i(r) + V(r)\psi_i(r) = E_i\psi_i(r). \tag{5}$$

Unlike the situation in full space $(-\infty < x < \infty)$ where the point eigenvalues E_i must be negative, in the present case E_i may be negative, zero, or positive. For $E_i < 0$, $E_i = 0$, $E_i > 0$ we must choose as boundary conditions $\psi_i(0) = 0$. It is convenient to chose the second boundary condition, namely the derivative at the origin in the following way:

$$\begin{aligned} \frac{d}{dr}\psi_i(0) &= (-E_i)^{1/2} \quad \text{for } E_i < 0, \\ &= (E_i)^{1/2} \quad \text{for } E_i > 0, \\ &= 1 \quad \text{for } E_i = 0. \end{aligned} \tag{6}$$

The normalizations of the eigenfunctions $\psi_i(r)$ are given by

$$\int_0^\infty [\psi_i(r)]^2\,dr = C_i. \tag{7}$$

Since $\psi_i(r)$ is real, the C_i are positive.

In the direct problem, the potential $V(r)$ is given, together with the boundary conditions at $r = 0$ for $\psi(r|k)$ and $\psi_i(r)$. It is required to find the Jost functions $f(k)$ $(= |f(k)|\exp(i\delta))$ from the eigenfunctions of the continuous spectrum using the asymptotic form (4). Furthermore, it is required to find the point eigenvalues E_i and the normalizations C_i.

In the inverse problem, one gives $|f(k)|$, E_i, C_i and calculates the potential $V(r)$ which would reproduce these quantities. For brevity we shall not give the Gel'fand—Levitan algorithm here but refer instead to Ref. 4 which also refers to earlier papers of others.

The analog of reflectionless potentials of Ref. 3 are those for which $|f(k)| \equiv 1$. The simplest of these potentials are given in Ref. 4. For $E_i > 0$ and $E_i = 0$ it is possible to find potentials for which $\delta(k) \equiv 1$ also so that $f(k) \equiv 1$. The potential for which there is only one point eigenvalue $E_1 = 0$ is the simplest of these and is given by

$$V(r) = 6r(r^3 - 2\lambda)/(r^3 + \lambda)^2, \quad \lambda = 3C_1. \tag{8}$$

[It should be mentioned that this potential for phaseless scattering was obtained earlier in Ref. 5 using more special methods.]

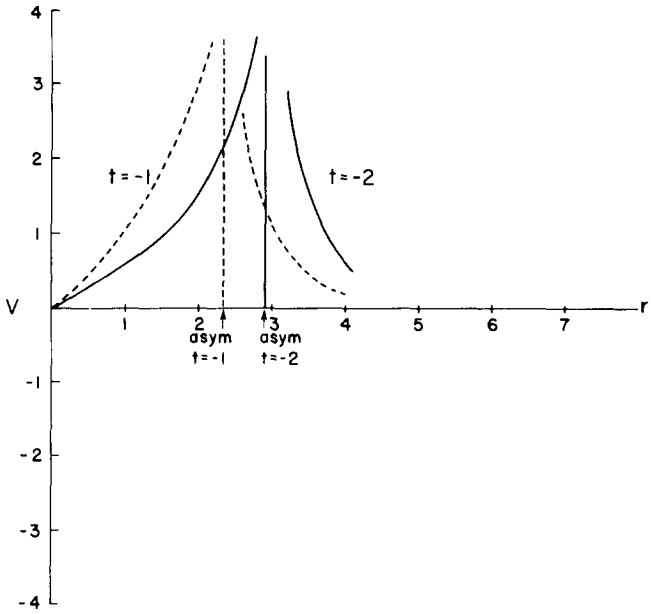


FIG. 1. Velocity as a function of distance from wall for $t<0$.

2. THE SOLUTION OF THE KORTEWEG-De VRIES EQUATION. INTERPRETATION

In Eq. (8) let us now consider λ as a function of the time t where t is a parameter. On substituting into the K-dV equation (1) we obtain a solution if

$$\frac{d}{dt}\lambda(t)=12$$

or

$$\lambda(t)=12t \tag{9}$$

on choosing the origin of time appropriately.

Thus we have as a solution of the KdV equation:

$$V(r,t)=6r(r^3-24t)/(r^2+12t)^2. \tag{10}$$

This solution is very different from those obtained using the Gel'fand–Levitan equation over the whole real axis.

For $t>0$ and $r\geq 0$, V is continuous. Moreover, $V(0,t)=0$. Figure 2 shows the highly dispersive wavelike motion which this solution exhibits, since V is zero at $r=0$, we may interpret $r=0$ as being a wall, if V is interpreted as being a velocity. From Fig. 2 we see that the "wave" has a minimum value at $r_{\min}=1.205\ t^{1/3}$ and that this minimum traveals with the velocity $v_{\min}=0.402\ t^{-2/3}$. V goes from its minimum value which is negative to its maximum which is positive through a stagnation point $r_s=2.88\ t^{1/3}$ which travels with the velocity $v_s=0.961\ t^{-2/3}$. The position of the maximum $r_{\max}$ is given by $r_{\max}=4.35\ t^{1/3}$ and its velocity is $v_{\max}=1.45\ t^{-2/3}$. Finally $V(r_{\min},t)=-0.851\ t^{-2/3}$, $V(r_{\max},t)=0.171\ t^{-2/3}$. From these quantities one can infer the shape of the "wave" and the character of its motion. For brevity, we shall not discuss this matter further.

Equation (10) also represents a solution with a wall for $t<0$. However, now V is positive for all $r>0$. Furthermore, V approaches infinite values on either side of the asymptote $r_{\text{asy}}=12^{1/3}(-t)^{1/3}=2.289\ (-t)^{1/3}$ which travels with the velocity $v_{\text{asy}}=-0.763\ (-t)^{1/3}$. Figure 1 gives V for two values of t when t is negative. One sees that the solution for negative t represents a singular "wave" moving toward the wall.

If one combines the motions for $t<0$ and $t>0$, one obtains an analog of a wave reflected by a wall in which the incident wave is singular and the reflected wave is continuous and highly dispersive. The situation for $t=0$ requires comment. In this case the asymptote for the motion $t<0$ coincides with the wall, and one has an exceptionally singular situation in that the velocity is infinite, not zero, at the wall. One can either take the attitude that at $t=0$ one has a kind of "impulse" on the wall at the instant of reflection or one can take the more general attitude that the solution for $t<0$ cannot be continued into the domain $t>0$. One simply has two different solutions in the two time domains. However, we do not wish to belabor the point, since our principal objective has been to show that one can obtain solutions in a half-space with a wall.

3. DISCUSSION

The fact that a solution of the inverse scattering problem for the radial Schrödinger equation yields a solution for the half-space with a wall suggests a connection between the inverse scattering problem for the radial equation and the KdV equation in the half-space generally. We are now investigating the possibility of such a connection, though the situation is much more complicated than the relation between the KdV equation and the inverse problem in full space. First of all, the point eigenvalues are functions of time, instead of being constant as in the full space case. Then too, the discrete and continuous spectrum are strongly coupled in time, again unlike the situation in full space. As a consequence, it appears that the solutions to the inverse

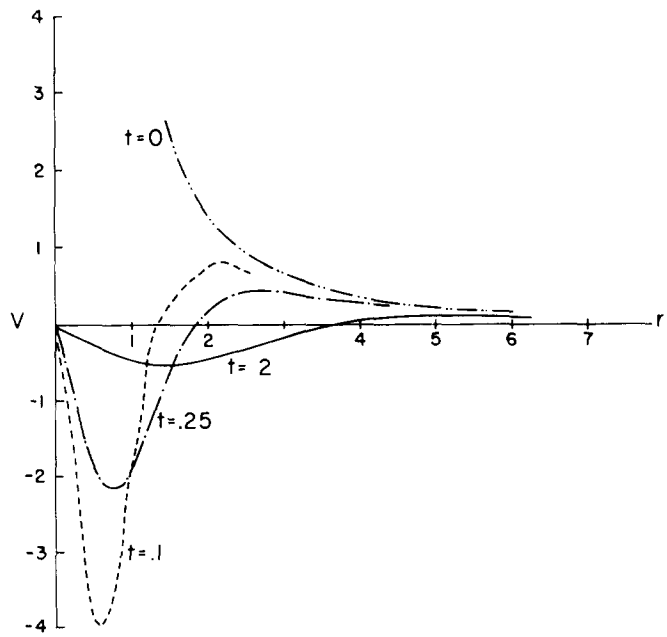


FIG. 2. Velocity as a function of distance from wall for $t>0$.

problem for which the continuous spectrum gives no contribution (i.e., for which $|f(k)| = 1$) but for which $E_i \neq 0$ do not give solutions to the KdV equation in the half-space with a wall.

In Ref. 6 solutions of the nonlinear Schrödinger equation in the half-space are discussed, though no explicit examples are given as in the present paper. The half-space solutions are obtained from the inverse problem in the full space by imposing suitable symmetry conditions. The authors find that they cannot relate the inverse problem in the half-space to their nonlinear equation in the half-space. We believe that the analog of our solution is not among those which can be obtained by their method. Moreover, the relation between our approach and that of Ref. 6 requires exploration of some difficulty, but this exploration is certainly worthwhile.

[1]C. S. Gardner, J. M. Greene, M. D. Kruskal, and R. M. Miura, Phys. Rev. Lett. **19**, 1095 (1967).
[2]C. S. Gardner, J. M. Greene, M. D. Kruskal, and R. M. Miura, Comm. Pure Appl. Math. **27**, 97 (1974).
[3]I. Kay and H. E. Moses. J. Appl. Phys. **27**, 1503 (1956).
[4]H. E. Moses and S. F. Tuan, Nuovo Cimento **13**, 197 (1959).
[5]V. Bargmann, Rev. Mod. Phys. **21**, 488 (1949).
[6]M. J. Ablowitz and H. Segur, J. Math. Phys. **16**, 1054 (1975).

Multigroup neutron transport. I. Full range*

R. L. Bowden and S. Sancaktar

Department of Physics, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061

P. F. Zweifel†

Department of Physics, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061 and The Rockefeller University, New York, New York 10021
(Received 25 July 1974; revised manuscript received 23 January 1975)

A functional analytic approach to the N-group, isotropic scattering, particle transport problem is presented. A full-range eigenfunction expansion is found in a particularly compact way, and the stage is set for the determination of the half-range expansion, which is discussed in a companion paper. The method is an extension of the work of Larsen and Habetler for the one-group case.

I. INTRODUCTION

Ever since Case[1] successfully applied the method of singular eigenfunction expansion to the one-speed, one-dimensional neutron transport equation, there have been efforts to generalize the treatment to the multigroup case. The basic equation considered for N groups is the following:

$$\mu \frac{\partial}{\partial x} \psi(x,\mu) + \Sigma\psi(x,\mu) = \int_{-1}^{1} C(\mu,s)\ \psi(x,s)\, ds. \qquad (1)$$

Here ψ is a column vector whose elements ψ_i represent the neutron angular density in the ith group, $1 \leq i \leq N$, and Σ is a diagonal matrix whose ith element σ_i is the neutron cross section in the ith group, ordered such that $\sigma_1 > \sigma_2 > \cdots > \sigma_n = 1$. The elements C_{ij} of the transfer matrix C represent the neutron scattering from group j to group i.

The somewhat simplified isotropic case (in which C is independent of μ and s) is treated in detail by Siewert and Zweifel[2] for a rather special situation physically relevant to radiative transfer, namely the determinant of every minor matrix of C of rank >1 was assumed to vanish. The more general case, in which the determinant of C was assumed not to vanish, has turned out to be rather difficult. Part of the problem is notational (a difficulty also encountered in Ref. 2), because the continuous spectrum is highly degenerate and because adjoint solutions must be introduced to calculate expansion coefficients. This leads to a bookkeeping task of no small magnitude. However, more fundamental is the difficulty that a satisfactory proof of half-range completeness of the singular eigenfunctions hinges on the signum of the so-called partial indices of the matrix Riemann–Hilbert problem, and these indices turn out to be difficult to pin down. A review of the partial index problem, and references to some attempts to deal with the problem have been given by Burniston *et al.*[3]

The notational problem alluded to above was solved in 1968 by Yoshimura and Katsuragi,[4] who also proved the relevant full-range completeness and orthogonality theorems. The more general case, in which C is a function of angles, was discussed by Silvenionnen and Zweifel,[5] where the further notational difficulties are handled, and sufficient conditions for full-range completeness presented. Half-range completeness can be deduced, in fact, if C is a symmetric matrix, as one obtains for example in thermal neutron problems with Maxwellian weights.

In 1973 a new development occurred in transport theory, namely the publication of a paper by Larsen and Habetler[6] in which the full- and half-range formulas originally obtained in Ref. 1 by heuristic arguments, were derived rigorously through functional analytic techniques. (A later paper by Larsen[7] extended these results to the anisotropic case, still, however, in one group theory.) These papers served not only to mollify the mathematicians who objected to Case and his disciples' cavalier treatment of the continuous spectrum, but also gave the hope of simplifying and generalizing the original results. For example, a paper[8] based on the results of Ref. 6 has extended the original expansion theorems of Case to an enormously larger class of functions. Basically, Ref. 6 made it possible to deal with linear operators in a Banach space, where previously one had to consider rather involved singular integral equations. Further, as we see by comparing the present paper with Ref. 4, the Larsen–Habetler technique is simpler and clearer than the standard technique of obtaining adjoint solutions, using Schmidt orthogonalization procedure, and calculating a large number of normalization integrals.

We propose to use the Larsen–Habetler technique to present an explicit solution to the half-range problem for a subcritical medium. (The extension to the more general case has not yet been found.) The solution will be given in terms of certain matrices X and Y which factor the dispersion matrix Λ.

It is well known, for example, in one-speed theory[1] that the matrix X which solves the Riemann-Hilbert problem also factors the Λ matrix [in Case's notation, $(1-c)(V_0^2 - Z^2)X(Z)X(-Z) = \Lambda(Z)$]. This "identity" as Case refers to it, in fact provides the basic connection between two apparently dissimilar methods for solving the same problem, i.e., the Case "eigenfunction expansion," leading to a Riemann–Hilbert problem, and the Wiener–Hopf method.

In the multigroup case, it turned out, perhaps because two matrices X and Y are involved in the factorization, that the Wiener–Hopf approach was more practical. The existence of the Wiener–Hopf factorization has been proved by Mullikin,[9] and although explicit solutions have not been found except in some special cases[10] they can be determined as the solutions of a certain nonlinear, non-singular integral equation. Our approach is reminiscent of that used by Siewert, Burniston, and Kriese[11] for the two group problem, their work being based on earlier work of Siewert and Ishigura.[12] These authors introduce matrices H and H^* which are, in fact, closely related to our matrices X^{-1} and Y^{-1}. (Along these lines, one should also note a paper by Burniston, Mullikin and Siewert.[13]) Perhaps even closer to our approach is that of Pahor and Suhadolc,[14] who uses a full range expansion to deduce the half-range formulas, again for subcritical media. These results are, however, based on heuristic arguments similar to Case's original work.

In the present paper, we will apply the Larsen–Habtler technique to the full-range, N group problem, partially to demonstrate that the functional analysis can be carried through to this case, and to set the stage for the half-range analysis which will be published separately. Throughout both papers, we shall restrict our attention to the case that C is a constant matrix, and det$C \neq 0$. We work in the function space U defined by

$$U=\{\psi \,|\, \mu\psi_i \text{ is differentiable in } x\in(-\infty,\infty) \text{ and is Hölder continuous in } \mu\in[-1,1]\}.$$

The idea of the Larsen–Habetler technique is to rewrite Eq. (1) as

$$\frac{\partial}{\partial x}\psi + K^{-1}\psi = 0,$$

where the reduced transport operator K^{-1}, which acts only on μ, is given by (for fixed x),

$$(K^{-1}\psi)(x,\mu) = (1/\mu)[\Sigma\psi(x,\mu) - C\int_{-1}^{1}\psi(x,s)\,ds], \quad \mu\neq 0. \tag{2}$$

The bounded inverse K of K^{-1} and the resolvent $(zI-K)^{-1}$ are then constructed. The identity

$$\psi(\mu) = (1/2\pi i)\oint_\tau (zI-K)^{-1}\psi(\mu)\,dz \tag{3}$$

is then used to obtain the full-range expansion. In Eq. (3), and otherwise as convenient, the dependence of ψ on x will be supressed.

In this paper, we study the properties of K and $(zI-K)^{-1}$ in Sec. II and derive the full-range expansion in Sec. III. We outline the solution of a typical transport problem in the Appendix.

II. THE RESOLVENT OPERATOR $(zI-K)^{-1}$

In this section we construct the bounded operator K and study the properties of the resolvent operator $(zI-K)^{-1}$. In addition, the following notation will be used in the subsequent analysis:

$$B \equiv (\Sigma - 2C)C^{-1}\Sigma,$$

$$D(z,\mu) \equiv (zI-\mu\Sigma^{-1})^{-1},$$

and

$$g_n = \int_{-1}^{1} s^n g(s)\,ds,$$

where g may be an element of U or an $N\times N$ matrix.

The operator K^{-1} in Eq. (2) can be easily inverted to yield K:

$$K\eta = \Sigma^{-1}\mu\eta + B^{-1}\eta_1. \tag{4}$$

Note that the diagonal terms of K are one-speed operators, while the off-diagonal terms are of rank one. If it is assumed that det $(\Sigma - 2C)\neq 0$, then K will be a bounded operator of U. This determinant condition is related to the critical condition for an infinite medium[15] and would reduce in the one-speed limit to the condition $c\neq 1$, which we note is also required by the analysis of Ref. 6. (Note the definition of the matrix C; the factor $1/2$ which multiplies the scalar c in the one-speed equation has been absorbed into C.)

We next determine the resolvent operator $(zI-K)^{-1}$ from Eq. (4) by a straightforward calculation and write it in the form

$$(zI-K)^{-1}\,\psi(\mu) = D(z,\mu)[\psi(\mu) + \Lambda^{-1}(z)\,(D\psi)_1(z)],$$

where the dispersion matrix Λ is defined by

$$\Lambda(z) \equiv B - D_1(z). \tag{5}$$

The function $(zI-K)^{-1}\psi$ is analytic in z except for a branch cut along $[-1,+1]$ due to the branch cuts in the element of Λ and poles at the zeros of the function

$$\Omega(z) \equiv \det\Lambda(z).$$

A simple calculation shows that the zeros of Ω are eigenvalues of K.

The above definition of Λ differs from that, for example, of Ref. 4 by an immaterial multiplicative factor which does not affect the condition $\Omega(z)=0$. In fact, Λ

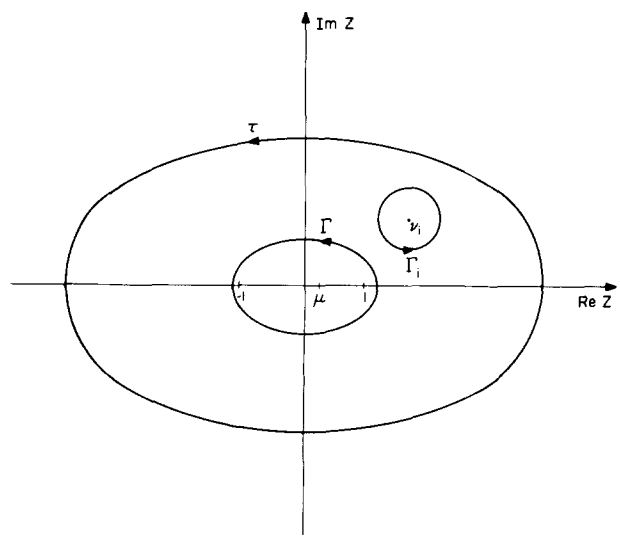


FIG. 1. The contour τ, surrounding the spectrum of K is deformed into the contour Γ surrounding the continuum plus circles Γ_i about each eigenvalue.

above is related to Λ usually encountered in the literature by $\Lambda_{\text{usual}} = C\Sigma^{-1}\Lambda\Sigma^{-1}$. The elements of $\Lambda(z)$ are even real functions of z. Thus the eigenvalues of K occur in pairs; if ν_i is an eigenvalue, so is $-\nu_i$. We shall label, for convenience, $-\nu_i = \nu_{i+n}$, where n is the number of pairs of eigenvalues. For simplicity, we will assume that each ν_i is a simple zero of Ω and lies outside the branch cut $[-1, 1]$. Although these restrictions are not necessary for the validity of the subsequent analysis, they simplify the paperwork. These restrictions can be relaxed and with some modifications the analysis would follow in a similar manner.

III. THE FULL-RANGE EXPANSION

We can use the identity in Eq. (3) to write for $|\mu| \leq 1$,

$$\psi(\mu) = (1/2\pi i) \oint_\tau D(z, \mu)[\psi(\mu) + \Lambda^{-1}(z)(D\psi)_1(z)]\,dz,$$

where τ encircles the spectrum of K which consists of the point spectrum given by the points $\{\nu_i\}$, $i = 1, \ldots, 2n$, and the rest of the spectrum (the union of the continuous and residual spectrums) given by the interval $[-1, 1]$.[16] We now deform the contour τ, as shown in Fig. 1, into small circles Γ_i centered about each eigenvalue ν_i, plus a contour Γ about the branch cut $[-1, 1]$ to obtain,

$$\psi(\mu) = \sum_{i=1}^{2n} \psi_{\nu_i}(\mu) + \psi_\Gamma(\mu),$$

where

$$\psi_{\nu_i}(\mu) \equiv (1/2\pi i) \oint_{\Gamma_i} D(z, \mu)[\psi(\mu) + \Lambda^{-1}(z)(D\psi)_1(z)]\,dz, \quad (6)$$

and

$$\psi_\Gamma(\mu) \equiv (1/2\pi i) \oint_\Gamma D(z, \mu)[\psi(\mu) + \Lambda^{-1}(z)(D\psi)_1(z)]\,dz.$$

Because of the simple pole of the elements of $\Lambda^{-1}(z)$ at $z = \nu_i$, $i = 1, \ldots, 2n$, we can use the residue theorem to write

$$\psi_{\nu_i}(\mu) = [\Omega'(\nu_i)]^{-1} D(\nu_i, \mu)\Lambda_c(\nu_i)(D\psi)_1(\nu_i), \quad (7)$$

where $\Omega'(\nu_i) = [d\Omega/dz]_{z=\nu_i}$ and $\Lambda_c(z)$ is the transpose of the cofactor matrix of $\Lambda(z)$. We note that ψ_{ν_i} is an eigenvector of K since from Eqs. (4), (5), and (7) we have

$$[\nu_i I - K]\psi_{\nu_i} = [\Omega'(\nu_i)]^{-1}[I - B^{-1}D_1(\nu_i)] \times \Lambda_c(\nu_i)(D\psi)_1(\nu_i) = 0. \quad (8)$$

Equation (7) can be expressed as the product of an eigenfunction of K, ψ_{ν_i}, times an expansion coefficient α_i as follows. Note that each column of $\Lambda_c(\nu_i)$ is proportional to the eigenvector of $\Lambda(\nu_i)$, i.e.,

$$[\Lambda_c(\nu_i)]_{lm} = \alpha_m \beta_l(\nu_i), \quad (9a)$$

where

$$\Lambda(\nu_i)\beta(\nu_i) = 0, \quad (9b)$$

and α_m, $m = 1, \ldots, N$, are constants depending on the elements of $\Lambda_c(\nu_i)$. Then we can write

$$\psi_{\nu_i}(\mu) = a_i \phi_{\nu_i}(\mu),$$

where

$$\phi_{\nu_i}(\mu) = D(\nu_i, \mu)\beta(\nu_i) \quad (10)$$

is the discrete eigenvector obtained in Ref. 4, and

$$a_i = \sum_{m=1}^{N} \alpha_m [(D\psi)_1(\nu_i)]_m [\Omega'(\nu_i)]^{-1} \quad (11)$$

is the expansion coefficient.

We now turn to the continuum term, ψ_Γ of the eigenfunction expansion. Define

$$M(z) \equiv \Lambda^{-1}(z)(D\psi)_1(z). \quad (12)$$

Then we can write

$$\psi_\Gamma(\mu) = (1/2\pi i) \oint_\Gamma D(z, \mu)[\psi(\mu) + M(z)]\,dz. \quad (13)$$

An application of the Cauchy integral theorem reduces the first term on the rhs of Eq. (13) to $\psi(\mu)$. To evaluate the remaining term, we deform the contour Γ as shown in Fig. 2. The contribution from Γ_{-1} and Γ_{+1} can be shown to vanish (see, for example, Ref. 6). The contribution from Γ' and Γ_μ give the first and second terms on the rhs of the following equation:

$$[(1/2\pi i) \oint D(z, \mu) M(z)\,dz]_i = (1/2\pi i) \Big\{ \int_{-1}^{1} D(\nu, \mu) \times [M^-(\nu) - M^+(\nu)]\,d\nu \Big\}_i + \tfrac{1}{2}[M_i^+(\mu/\sigma_i) + M_i^-(\mu/\sigma_i)], \quad i = 1, \ldots, N, \quad (14)$$

where we use the notation that $M^+(\nu)$ and $M^-(\nu)$ are the boundary values of $M(\nu \pm i\epsilon)$ as $\epsilon \to 0$ respectively. The integral on the rhs of Eq. (14) is understood to be the Cauchy principal value. If we further denote

$$[\Delta(\nu, \mu)]_{ml} = \delta_{ml}\delta(\sigma_m \nu - \mu),$$

and

$$[\psi_\Gamma(\mu)]_i = \begin{cases} \psi_i(\sigma_i \mu) & |\sigma_i \mu| \leq 1, \\ 0 & \text{otherwise}, \end{cases}$$

where δ_{ml} and $\delta(\sigma_m \nu - \mu)$ are the kronecker delta function and dirac delta distribution, respectively, we can

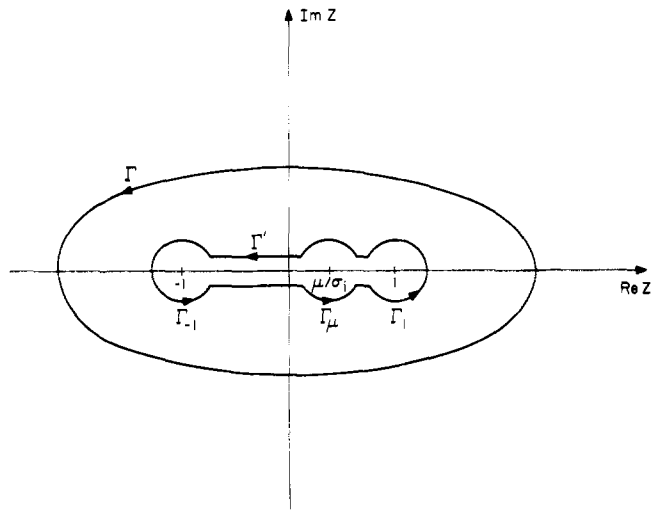


FIG. 2. The contour Γ about the continuum is squeezed down to the contour $\Gamma' \cup \Gamma_{-1} \cup \Gamma_{+1} \cup \Gamma_\mu$

write

$$\psi_\Gamma(\mu) = (1/2\pi i)\int_{-1}^{1} [D(\nu,\mu)\{M^-(\nu) - M^+(\nu)\} + \pi i \Sigma \Delta(\nu,\mu)(2\psi_\Sigma(\nu) + \{M^-(\nu) + M^+(\nu)\})]\,d\nu.$$

Noting from Eqs. (5) and (12) that

$$\psi_\Sigma(\mu) = (1/2\pi i\nu)\Sigma^{-1}[\Lambda\ (\nu) + \Lambda\ (\nu)] \times [M^-(\nu) - M^+(\nu)] - \tfrac{1}{2}[M^+(\nu) + M^-(\nu)],$$

we can write

$$\psi_\Gamma(\mu) = \int_{-1}^{1} \Phi(\nu,\mu) A(\nu)\,d\nu, \tag{15}$$

where

$$\Phi(\nu,\mu) \equiv \nu D(\nu,\mu) + \tfrac{1}{2}\Delta(\nu,\mu)\Sigma^{-1}[\Lambda^+(\nu) + \Lambda^-(\nu)], \tag{16}$$

and

$$A(\nu) \equiv (1/2\pi i\nu)[M^+(\nu) - M^-(\nu)]. \tag{17}$$

(Note that $\psi_\Gamma(\mu)$ is Holder continuous on $[-1,1]$.)

It has become customary to express formulas like Eq. (15) in terms of so-called continum eigenvectors multiplied by an expansion coefficient integrated on ν. [In fact, the columns of $\Phi(\nu,\mu)$ are exactly the continuum eigenvectors of Ref. 4.] Although the mathematical justification of this custom is not exact, it is instructive to note that the columns of the matrix Φ are indeed generalized eigenvectors of the operator K. By this we mean that

$$K\psi_\Gamma(\mu) = \int_{-1}^{1} \nu\Phi(\nu,\mu)\,\Lambda(\nu)\,d\nu, \quad |\nu| \leq 1. \tag{18}$$

To prove Eq. (18), we utilize the identity

$$(zI - K)^{-1}K\psi(\mu) = z(zI - K)^{-1}\ \psi(\mu) - \psi(\mu). \tag{19}$$

Then the analysis leading to Eq. (18) is carried out for Kf rather than f, using the rhs of Eq. (19). An almost identical calculation leads to Eq. (18).

IV. CONCLUSION

As a result of the analysis of the previous sections, we have established the following eigenfunction expansion for ψ, which we state as a theorem.

Theorem 1: Let $\psi \in U$. Then for a fixed x, ψ can be expanded as

$$\psi(x,\mu) = \sum_{i=1}^{2n} a_i(x)\,\phi_{\nu_i}(\mu) + \int_{-1}^{1} \Phi(\nu,\mu)A(\nu)\,d\nu, \tag{20}$$

where ϕ_{ν_i}, a_i, ϕ and A are given by Eqs. (10), (11), (16), and (17) respectively. Furthermore, the ϕ_{ν_i} are eigenvectors of K and $\Phi(\nu,\mu)$ is a generalized eigenvector in the sense of Eq. (19).

In an accompanying paper, the so-called half-range problem in which an eigenfunction expansion of ψ corresponding to the part of the spectrum containing elements z with re $z \geq 0$ will be considered. The ground work for that paper has been laid by the present paper.

Also, the authors expect an extension of the domain of Eq. (1) to a larger class of function spaces, namely $X_p = \{\psi \mid \psi_i$ differentiable in $x \in (-\infty, \infty)$ and $\mu\psi_i \in L_p[-1,1]$ in μ, $p > 1$, in a similar manner to Ref. 8 which extended the results of Ref. 6. The restriction of the present paper to the smaller space U was made so that equations like (14) could be written down.

Finally, in the authors' opinion, the major benefit of the above analysis, other than its mathematical rigor, is the fact that the cumbersome calculations of the so-called degenerate continuum eigenfunctions in various regions of the branch cut (see Ref. 4) are avoided. Also, the analysis is suitable for more generalized cases, namely the anisotropic multigroup transport equation.

ACKNOWLEDGMENTS

The authors would like to express their thanks to Dr. W. Greenberg and Dr. E. Larsen for their helpful discussions. Also, one of the authors is grateful to the Instituto Matematica Applicata, Universita Degli Studi di Firenza and to Professor Aldu Belleni-Morante for their kind hospitality during the period that this manuscript was rewritten.

APPENDIX A: REMARKS ON SOLUTIONS OF THE MULTIGROUP TRANSPORT EQUATION

Larsen and Habetler[6] have used the Case eigenfunction expansion to solve the transport equation by asserting that any solution which was "sufficiently smooth" could be expanded, and its expansion coefficients related to those of the source. The trouble with this approach is that it requires, *ex post facto*, a verification that the solution is indeed "sufficiently smooth." Here we follow a different approach, more in accord with Case's original work, of constructing a solution which satisfies the boundary conditions, and then relying on a uniqueness theorem to guarantee that there is no other solution.

Since, in the context of a full-range expansion, only infinite medium problems are really relevant, we consider the infinite medium Green's function, that is we seek solutions of the homogeneous equation

$$\frac{\partial}{\partial x}\psi + K^{-1}\psi = 0, \tag{A1}$$

subject to the conditions

$$\lim_{|x|\to\infty} \psi(x,\mu) = 0, \tag{A2}$$

$$\lim_{\substack{\epsilon\to 0 \\ \epsilon>0}} \{\psi(\epsilon,\mu) - \psi(-\epsilon,\mu)\} = \frac{\varphi(\mu)}{\mu} \equiv Q_o(\mu), \tag{A3}$$

where Q_o is assumed Hölder continuous. (Q is the source strength vector.)

Let us expand $Q_o(\mu)$ according to Theorem 1, rewritten for convenience in the following form:

$$\varphi_0(\mu) = \sum_{i=1}^{n} a_i\phi_{\nu_i}(\mu) + \sum_{k=1}^{n} a_k\phi_{-\nu_k}(\mu) + \int_0^1 \Phi(\nu,\mu)A(\nu)\,d\nu + \int_{-1}^{0} \phi(\nu,\mu)A(\nu)\,d\nu. \tag{A4}$$

Then we make the following

Assertion: The solution of Eqs. (A1), (A2), and (A3)

is given by

$$\psi(x,\mu)=\sum_{i=1}^{n} a_i \exp(-x/\nu_i)\,\phi_{\nu_i}(\mu)+\int_0^1 \Phi(\nu,\mu)$$

$$\times\exp(-x/\nu)A(\nu)\,d\nu,\quad x>0, \tag{A5a}$$

$$=-\sum_{k=1}^{n} a_k \exp(x/\nu_k)\,\phi_{-\nu_k}(\mu)-\int_{-1}^{0}\Phi(\nu,\mu)$$

$$\times\exp(-x/\nu)A(\nu)\,d\nu,\quad x<0. \tag{A5b}$$

Justification of Assertion: The assertion can be proved rigorously by developing a full functional calculus for K, as has been done[8] for the one-speed case. This could easily be carried out, but requires some detailed estimates which have not yet been made. A simpler procedure involves rewriting the transport equation in the form

$$\left(K\frac{\partial\psi}{\partial x}\right)(x,\mu)+\psi(x,\mu)=0. \tag{A6}$$

This is justified so long as $\mu\,\partial\psi/\partial x$ is Hölder continuous in μ for every x. (Even this is not really required since K is a bounded operator which could, by continuity, be extended to a complete function space, say $L_p[-1,1]$.) Then from Eqs. (8) and (19) we observe that $\psi(x,\mu)$ obeys the transport equation, and the boundary conditions (A2) and (A3) and is, hence, the unique solution.[17]

For the one speed problem, uniqueness is well known.[18] For the energy dependent transport equation, a number of uniqueness theorems have been shown,[18,19] and the multigroup case considered here can, with a little effort, be shown to be a special case of some of these treated there. However, we note that uniqueness for the one-dimensional case is more or less trivial; we sketch the proof because the argument involves the calculation of the norm of an integral operator and the same calculation is involved in a different way in the construction of the half-range eigenfunction expansion (accompanying paper).

It is well known [see Ref. 7, Sec. (3.6)] that the one speed transport equation can be written as an integral equation by introducing the Green's function. In exactly the same way, the multigroup transport equation with source Q (which may be a distributed source or a plane source as in the Green's function problem) is equivalent to an equation for the density ρ:

$$\rho(x)=\int_{-\infty}^{\infty} G(|x-x'|)\,\rho(x')\,dx'+\bar{Q}(x), \tag{A6a}$$

where

$$\bar{Q}(x)=\int_{-\infty}^{\infty}dx'\int_{-1}^{1}d\mu\,\frac{\exp(-\Sigma|x-x'|/|\mu|)}{|\mu|}CQ_0(x',\mu), \tag{A6b}$$

with

$$G(|x-x'|)=E(|x-x'|)\delta_{ij}, \tag{A7a}$$

and

$$E_{ij}\equiv E_1(\sigma_i|x-x'|)\delta_{ij}. \tag{A7b}$$

E_1 being defined in Ref. 17, Appendix E.

Now, a Fredholm equation like (A-6) is known to possess a unique solution if the norm of its kernel G is less than unity. In the next appendix we compute this norm in an L_1 space and conclude that our solution given by Eqs. (A5) is indeed unique if the infinite medium under consideration is subcritical. This is precisely the condition at which one arrives in one-speed theory, i.e., $c<1$.

APPENDIX B: CALCULATION OF $\|G\|$

We work with the Banach space

$$L=\bigoplus_{1}^{N} L_1,$$

with norm

$$\|f\|=\sum_{i=1}^{N}\int_{-\infty}^{\infty}|f_i|\,dx.$$

(We note the solution obtained in Appendix A is an element of L.) The operator norm is

$$\|G\|=\sup_{f\in L}\|Gf\|/\|f\|.$$

It is trivial to show that the integral operator G, with kernel $G(|x-x'|)$ is a map of L into itself. In particular, if A is a matrix of constants then

$$\|A\|=\sup_k\sum_{j=1}^{N}|A_{jk}|\equiv\|A\|_M.$$

Now, if K is a matrix of operators, then each matrix element K_{ij} has an L_1 norm which is denoted by $\|K_{ij}\|_1$. By writing the operator $G=EC=(E\Sigma)(\Sigma^{-1}C)$ we conclude that (E and Σ are diagonal)

$$\|G\|\le\|E\Sigma\|\cdot\|\Sigma^{-1}C\|$$

$$\le\sup_i\|E_{ii}\Sigma_{ii}\|\cdot\|\Sigma^{-1}C\|_M.$$

To compute $\|E\Sigma\|$, we may use Kato's criterion[20] which, for a difference kernel reduce to

$$\|E_{ii}\Sigma_{ii}\|\le\sup_x\int_{-\infty}^{\infty}\sigma_i E_1(\sigma_i|x-x'|)\,dx'=2.$$

Thus,

$$\|G\|\le 2\|\Sigma^{-1}C\|_M,$$

and the solution obtained in Appendix A will be unique if

$$2\|\Sigma^{-1}C\|_M\le 1.$$

This result is exactly one of those derived by Bowden[15] as the subcriticality condition for an infinite medium.

We shall find that this norm result plays a crucial role in the half-space problem discussed in the accompanying paper. We point out that for half-space problems, the range of integration is $(0,\infty)$ rather than $(-\infty,\infty)$. However, the norm calculated is identical [evaluated with the limits $(0,\infty)$].

*Based in part on a doctoral dissertation submitted by one of the authors (S.S.) to the Graduate School, VPI and SU. Research supported by the National Science Foundation under Grant No. GK-35903.

†John Simon Guggenheim Memorial Foundation Fellow.

[1]K.M. Case, Ann. Phys. (N.Y.) **9**, 1 (1960).
[2]C.E. Siewert and P.F. Zweifel, Ann. Phys. (N.Y.) **36**, 61 (1966); J. Math. Phys. **7**, 2092 (1966).
[3]E.E. Burniston, *et al.*, Nucl. Sci. Eng. **45**, 331 (1971).
[4]T. Yoshimura and S. Katsuragi, Nucl. Sci. Eng. **33**, 297 (1968).
[5]P. Silvennoinen and P.F. Zweifel, J. Math. Phys. **13**, 1114 (1973).
[6]E.W. Larsen and G.J. Habetler, Commun. Pure Appl. Math. **26**, 525 (1973).
[7]E.W. Larsen, Commun. Pure Appl. Math. **27**, 523 (1974).
[8]E.W. Larsen, S. Sancaktar, and P.F. Zweifel, J. Math. Phys. **16**, 1117 (1975).
[9]T.W. Mullikin, Transp. Theory Stat. Phys. **3**, 215 (1973). (1973).
[10]E.W. Larsen and P.F. Zweifel *Explicit Wiener–Hopf Factorizations* (to be published).
[11]C.E. Siewert, E.E. Burniston, and J.T. Kriese, J. Nucl. Energy **26**, 469 (1972).
[12]C.E. Siewert and Y. Ishiguro, J. Nucl. Energy **26**, 251 (1972).
[13]E.E. Burniston, T.W. Mullikin, and C.E. Siewert, J. Math. Phys. **13**, 1961 (1972).
[14]S. Pahor and A. Suhadolc, Transp. Theory Stat. Phys. **2**, 335 (1972).
[15]R.L. Bowden, Transp. Theory Stat. Phys. **4**, 25 (1975).
[16]L.C. Baird and P.F. Zweifel, Nuovo Cimento B **23**, 402 (1974).
[17]Mathematical purists (only) will object to this "justification" on several grounds, one being the necessity of a delicate proof that

$$K\frac{\partial\psi}{\partial x}(x,\mu)=\frac{\partial}{\partial x}(K\psi)(x,\mu).$$

The complete functional calculus, as in Ref. 8, could be developed to satisfy this question, but the question then arises as to how much effort "purity" is worth.
[18]K.M. Case and P.F. Zweifel, *Linear Transport Theory* (Addison-Wesley, Reading, Mass., 1967), p. 43.
[19]K.M. Case and P.F. Zweifel, J. Math. Phys. **4**, 1370 (1963).
[20]T. Kato, *Perturbation Theory for Linear Operators* (Springer, Berlin, 1966), p. 143, example 2.4.

Multigroup neutron transport. II. Half range*

R. L. Bowden and S. Sancaktar

Department of Physics, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061

P. F. Zweifel†

Department of Physics, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061 and The Rockefeller University, New York, New York 10021

(Received 25 July 1974; revised manuscript received 23 January 1975)

This paper accompanies a preceding one in which a functional analytic method was used to obtain the full-range expansion in multigoup neutron transport. In the present paper the analysis is extended to obtain the half-range expansion. The method used is an extension of the work of Larsen and Habetler for the one-group case. The results are given in terms of certain matrices which are solutions of coupled integral equations and which factor the dispersion matrix.

I. INTRODUCTION

In an accompanying paper,[1] hereafter referred as I, the full range eigenfunction expansion for the solution of a multigroup neutron transport equation is obtained from application of the resolvent operator technique of Larsen and Habetler.[2] We refer the reader to the introduction of I, in which we point out some of the advantages of the Larsen—Habetler technique over the usual Case orthogonal singular eigenmodes approach.[3] These advantages include mathematical rigor and simplified notation. But the major advantage is the subject of the present paper, namely that the technique can be applied just as easily to obtain the so-called half range eigenfunction expansion. Except for some special cases,[4,5] existence of the half range expansion, let alone explicit formulas, has never been demonstrated, due to a technicality in determining so called "partial indices."[6] (The only exception is a paper by Pahor and Suhadolc[7] whose ideas have some similarity to ours, but whose arguments are still based on the original Case approach. Like them, we are restricted by the condition $\|\Sigma^{-1}C\| < \frac{1}{2}$, which has been shown to be a condition that the infinite medium be subcritical.[8])

We should point out another major advantage of the present technique, namely its suitability to generalization. For example, the anisotropic case could be treated almost as easily as the isotropic case considered here. It is necessary in the analysis to write the transport equation in the form of an integral equation which, while tedious in high order anisotropy, is nonetheless feasible. Also, the condition on $\|\Sigma^{-1}C\|$ could be relaxed if the sufficient condition of Gohberg and Krein[9] for the convergence of Neumann series solutions to certain integral equations could be improved. (Our analysis is heavily based on a factorization theorem of T. W. Mullikin[10] which is, in turn, based on the work of Ref. 9.)

In the present paper we use the Larsen—Habetler technique and the results of the Mullikin theorem to obtain an explicit representation of the half-range expansion; the partial index question never arises. An outline of our presentation is as follows:

In Sec. III we find a projection operator E such that $(zI-K)^{-1}E\psi$ is analytic for $\mathrm{Re}z<0$, where ψ belongs to a certain function space and $(zI-K)^{-1}$ is the resolvent operator considered in I. It turns out that this operator contains certain noncanonical matrices X and Y which factor the dispersion matrix. Then, in Sec. III we obtain explicitly the half range eigenfunction expansion which can, in turn, be used to solve half-space transport problems in the usual way. The X and Y matrices are given in terms of the solutions of nonlinear, nonsingular integral equations, and they may be obtained numerically. This, in turn, will make it possible to obtain numerical solutions from the results of the present paper.

Appendix A is devoted to the conversion of the Mullikin results to the form which can be used in the present paper. In Appendix B a uniqueness theorem is proved for the matrices X and Y. Appendix C contains a solution to a half-space transport problem.

II. THE PROJECTION OPERATOR E

We consider the function space U' of N-dimensional column vectors ψ defined as

$$U' = \{\psi \mid \mu\psi_i \text{ is differentiable in } x \in (-\infty, \infty) \text{ and Hölder continuous in } \mu \in [0,1]\}.$$

The operator K defined by Eq. 4 of I is a bounded operator on the function space U defined in I. We closely follow the procedure of Ref. 2 and look for a projection operator E: $U' \to U$ with the two properties:

$$\text{(I)}\quad E\psi(\mu) = \psi(\mu), \quad 0 < \mu \leq 1, \tag{1}$$

and

(II) $(zI-K)^{-1}E\psi(\mu)$ is analytic in z for $\mathrm{Re}z<0$,

where we use the notation defined in I to write for $f \in U$,

$$(zI-K)^{-1}f(\mu) = D(z,\mu)\left[f(\mu) + \Lambda^{-1}(z)\int_{-1}^{1} sD(z,s)f(s)\,ds\right]. \tag{2}$$

(As in I, the x-dependence will be suppressed.) The idea of introducing the operator E is that the identity

$$\psi(\mu) = E\psi(\mu) = (1/2\pi i)\oint_{\tau}(zI-K)^{-1}E\psi(\mu)\,dz, \quad 0 < \mu \leq 1, \tag{3}$$

where τ is a contour encircling the spectrum of the operator K, will reduce to the half range expansion of ψ for $0 < \mu \leq 1$ because of property (II) above.

Let us now consider $\psi \in U'$ and write

$$(zI-K)^{-1}E\psi(\mu) = D(z,\mu)[E\psi(\mu)+G(z)], \tag{4}$$

where we have defined the column vector $G(z)$ by

$$G(z) = \Lambda^{-1}(z)\int_{-1}^{1} s\,D(z,s)E\psi(s)\,ds. \tag{5}$$

In order for $(zI-K)^{-1}E\psi(\mu)$ to be analytic in z for $\mathrm{Re}z < 0$, we require that:

(a) $G^+(\mu) = G^-(\mu) = G(\mu), \quad -1 < \mu < 0$

(b) $[G(\mu)]_i = -[E\psi(\sigma_i\mu)]_i, \quad -1 < \sigma_i\mu < 0, \quad i = 1,\ldots,N,$

and

(c) $G(-\nu_i) < \infty, \quad \mathrm{Re}\nu_i > 0, \quad i = 1,\ldots,n,$

where we recall that $\det\Lambda(\pm\nu_i) = \Omega(\pm\nu_i) = 0, \quad i = 1,\ldots,n.$

At this point, we introduce matrices $X(z)$ and $Y(z)$ which factor the dispersion matrix $\Lambda(z)$ according to

$$\Lambda(z) = Y(-z)X(z), \tag{6}$$

and which satisfy the following properties:

(i) $X(z)$ and $Y(z)$ are analytic in the complex z-plane cut along $[0,1]$,

(ii) $\det X(\nu_i) = \det Y(\nu_i) = 0$ and $\det X(-\nu_i) \neq 0$ and $\det Y$ $\det Y(-\nu_i) \neq 0$ for $\mathrm{Re}\nu_i > 0, \quad i = 1,2,\ldots,n,$ and

(iii) $\lim_{|z|\to\infty} X(z) = \text{constant}$, and $\lim_{|z|\to\infty} Y(z) = \text{constant}$.

The existence of the $X(z)$ and $Y(z)$ matrices with the properties listed above can be shown from the results of Mullikin.[10] This is done in Appendix A.

Let us find another representation of $G(z)$ by defining the operator E and the column vector function F so that the column vector function

$$Q(z) = \int_{-1}^{1} sD(z,s)E\psi(s)\,ds - Y(-z)\int_0^1 (1/z-s)F(s)\,ds, \tag{7}$$

is identically zero. Since $Q(z)$ vanishes as $|z| \to \infty$ and is analytic except perhaps for a cut along the interval $[-1,1]$, we need only to require that $Q(z)$ be continuous across that interval. Thus, using the Plemelj formulas and property (I) above, we find that $Q(z)$ is continuous across $[-1,1]$ if

$$F(\mu) = \mu Y^{-1}(-\mu)\Sigma^2\psi_\Sigma(\mu), \quad 0 \le \mu \le 1 \tag{8}$$

and

$$[E\psi]_i(\sigma_i\mu) = -[X^{-1}(\mu)\int_0^1 s(\mu-s)^{-1}\,Y^{-1}(-s)\Sigma^2\psi_\Sigma(s)\,ds]_i, \quad -1 \le \sigma_i\mu < 0, \tag{9}$$

where we have used Eq. (8) in Eq. (9) and we have defined the column matrix $\psi_\Sigma(\mu)$ such that

$$[\psi_\Sigma(\mu)]_i = \begin{cases} \psi_i(\sigma_i\mu), & 0 \le \sigma_i\mu \le 1, \\ 0, & \text{otherwise.} \end{cases}$$

Combining Eqs. (5), (7), (8), and (9), we determine $G(z)$ to be given by

$$G(z) = X^{-1}(z)\int_0^1 s(z-s)^{-1}Y^{-1}(-s)\Sigma^2\psi_\Sigma(s)\,ds. \tag{10}$$

Note that $G(z)$ given by Eq. (10) satisfies requirements (a), (b) and (c) listed above. Therefore our desired projection operator E is defined by Eqs. (1) and (9).

Note Eq. (9) that $E\psi(\mu)$ is Hölder continuous for $\mu < 0$ (recalling that X is analytic). Thus we verify that E maps into U, in particular, we are justified in applying the full range expansion of I to $E\psi$.

III. HALF RANGE EXPANSION

We now can combine Eqs. (3), (4) and (10) to obtain the half range expansion in the form

$$\psi(\mu) = \sum_{i=1}^{n} \psi_{\Gamma_i} + \psi_\Gamma,$$

where

$$\psi_{\Gamma_i}(\mu) = (1/2\pi i)\oint_{\Gamma_i} D(z,s)[E\psi(\mu)+G(z)]\,dz,$$

and

$$\psi_\Gamma(\mu) = (1/2\pi i)\oint_\Gamma D(z,s)[E\psi(\mu)+G(z)]\,dz.$$

Each contour Γ_i consists of a small circle centered about ν_i, and the contour Γ encircles the interval $(0,1)$. Applying the residue theorem, we can calculate ψ_{Γ_i} to be

$$\psi_{\Gamma_i}(\mu) = [\Omega'(\nu_i)]^{-1}D(\nu_i,\mu)[\det Y(\nu_i)] \times X_c(\nu_i)\int_0^1 s(\nu_i-s)^{-1}Y^{-1}(-s)\Sigma^2\psi_\Sigma(s)\,ds, \tag{11}$$

where $X_c(\nu_i)$ is the transpose of the cofactor of $X(\nu_i)$.

Noting that each column of $X_c(\nu_i)$ is proportional to the eigenvector $\beta(\nu_i)$ of $\Lambda(\nu_i)$, as defined by Eq. (I-9) we can write

$$[\det Y(\nu_i)][X_c(\nu_i)]_{im} = \alpha'_m\beta_i(\nu_i),$$

where α'_m is a constant. With the above definitions, we can express Eq. (11) as

$$\psi_{\Gamma_i}(\mu) = a'_i\phi_{\nu_i}(\mu), \tag{12}$$

where

$$a'_i = [\Omega'(\nu_i)]^{-1}\sum_{m=1}^{N}\alpha^1_m[\int_0^1 s(\nu_i-s)^{-1}Y^{-1}(-s)\Sigma^2\psi_\Sigma(s)\,ds]_m, \tag{13}$$

and $\phi_{\nu_i}(\mu)$ is the discrete eigenvector given in Eq. (I-10).

To find ψ_Γ, we apply the same integration technique to the contour Γ as was done in I, and we get

$$\psi_\Gamma(\mu) = \int_0^1 \Phi(\nu,\mu)A'(\nu)\,d\nu, \tag{14}$$

where

$$A'(\nu) = \frac{G^+(\nu)-G^-(\nu)}{2\pi i\nu}, \quad 0 \le \nu \le 1, \tag{15}$$

and $\Phi(\nu,\mu)$ is given by Eq. (I-16). Finally, we combine Eqs. (12) and (14) to write the half range expansion for $\psi \in U'$, $0 \le \mu \le 1$ as

$$\psi(\mu) = \sum_{i=1}^{N} a'_i\phi_{\nu_i}(\mu) + \int_0^1 \Phi(\nu,\mu)A'(\nu)\,d\nu,$$

where ϕ_{ν_i}, $\Phi(\nu,\mu)$, a'_i and $A'(\nu)$ are given by Eqs. (I-10), (I-16), (13), and (15), respectively.

The expansion coefficients are given in terms of the matrices $X(z)$ and $Y(z)$. By rearranging the results of Mullikin,[10] we show in Appendix A that $X(z)$ and $Y(z)$ satisfy the functional equations

$$X(z)=C^{-1}\Sigma+\frac{z}{2\pi i}\int_0^1\frac{Y^{-1}(-s)}{s(s-z)}[\Lambda^+(s)-\Lambda^-(s)]\,ds, \quad (16)$$

and

$$Y(z)=\Sigma-\frac{z}{2\pi i}\int_0^1\frac{\Lambda^+(s)-\Lambda^-(s)}{s(s-z)}X^{-1}(-s)\,ds. \quad (17)$$

If z is restricted to the interval $[-1,0]$, these functional equations reduce to nonlinear, nonsingular coupled matrix integral equations for $Y(-\mu)$ and $X(-\mu)$, $0\leq\mu\leq 1$. Using the results of Mullikin,[10] we show in Appendix A that a solution of this last set of equations exists if $\|\Sigma^{-1}C\|<\frac{1}{2}$. This is, of course, a sufficient condition and not a necessary condition. Furthermore, we show that any pair of matrices $X(z)$ and $Y(z)$ which satisfy the factorization to within a trivial factor, is unique (Mullikin has shown existence of at least one solution). Thus, one may proceed with confidence to evaluate X and Y numerically from Eqs. (16) and (17).

It is important to understand the difference between the present method and the "Case-type" approach. In the latter, one obtains the half range eigenfunction expansion as the solution of a singular integral equation on the line $[0,1]$, and requires a matrix L^{-1} analytic in the complex plane cut along $[0,1]$ such that $(L^-)^{-1}L^+=(\Lambda^-)^{-1}\Lambda^+$. The existence of such a matrix has been proved[11] but its behavior at infinity (the so-called "partial indices") is crucial to the proof of completeness, i.e., the proof that the singular integral equation for the expansion coefficient possesses a unique solution. Unfortunately, in many cases the calculation of the partial indices is impossible. In the approach taken here here, where we deal with the complex plane as a whole, the factorization as described by Eq. (6) is relevant, and the behavior at infinity is known. The poles of X^{-1} and Y^{-1} at the discrete eigenvalues presented no problem as was seen.

The major advantage of the approach used here is that numerical methods may be applied to Eqs. (16) and (17) and thus the expansion coefficients may be numerically evaluated. Using the canonical approach, it seems that at best existence can be proved, since no explicit solutions of the canonical problems are known. (See, however, Ref. 7.)

ACKNOWLEDGMENTS

The authors would like to express their appreciation to Professor T.W. Mullikin for communicating his results prior to publication. One of the authors (P.F.Z.) thanks Professor K.M. Case for valuable discussions, and is grateful to Istituto Matematica Applicata, Univer Universita degli Studi di Firenze and to Professor Aldo Belleni-Morante for their kind hospitality during the period that this manuscript was rewritten.

APPENDIX A: SOME PROPERTIES OF X AND Y MATRICES

T.W. Mullikin[10] has shown that if K_x is an $N\times N$ matrix operator

$$K_x f(x)=\int_0^\infty k(x,y)f(y)\,dy,$$

on vector functions f with norm

$$\|f\|=\sum_{i=1}^N\int_0^\infty|f_i(x)|\,dx,$$

and if $\|K_x\|<1$, then there exists a Weiner–Hopf factorization

$$[I-\hat{k}(z)]H_r(z)H_l(-z)=I \text{ for } \mathrm{Im}z=0,$$

where $H_r(z)$ and $H_l(z)$ are matrices, analytic for $\mathrm{Im}z>0$, and continuous and non-singular in $\mathrm{Im}z\geq 0$. Here, $\hat{k}$ is the Fourier transform of k,

$$\hat{k}(z)=\int_{-\infty}^\infty k(x)\exp(-izx)\,dx.$$

Mullikin also showed that H_r and H_l satisfy

$$H_r^{-1}(z)=I+(1/2\pi i)\int_{-\infty}^\infty H_l(t)\hat{k}(-t)(t+z)^{-1}\,dt, \quad \mathrm{Im}z>0, \quad (A1)$$

and

$$H_l^{-1}(z)=I+(1/2\pi i)\int_{-\infty}^\infty\hat{k}(t)H_r(t)(t+z)^{-1}dt, \quad \mathrm{Im}z>0. \quad (A2)$$

We can analytically extend the matrices $H_r(z)$ and $H_l(z)$ to the lower half of the z plane by defining

$$H^*(z)=\begin{cases}H_r(z), & \mathrm{Im}z\geq 0,\\ [I-\hat{k}(z)]H_l^{-1}(-z), & \mathrm{Im}z<0,\end{cases} \quad (A3)$$

and

$$H(z)=\begin{cases}H_l(z) & \mathrm{Im}z\geq 0\\ H_r^{-1}(-z)[I-\hat{k}(z)]^{-1} & \mathrm{Im}z<0.\end{cases} \quad (A4)$$

Now, $H_r(z)$ is analytic for $\mathrm{Im}z>0$ and $[I-\hat{k}(z)]^{-1}[H_l(-z)]^{-1}$ is analytic for $\mathrm{Im}z<0$ except for a branch cut along $[-i,-i\infty)$, due to $[I-k(z)]^{-1}$ and poles at the zeros of $\det[I-\hat{k}(z)]$. Since $H_r(z)=[I-\hat{k}(z)]\,[H_l(-z)]^{-1}$, $\mathrm{Im}z=0$, $H^*(z)$ is analytic everywhere in the complex plane except for the cut along $[-i,-i\infty)$, and poles at the zeros of $\det[I-\hat{k}(z)]$ in the lower half-plane.

Similar arguments follow for the matrix $H(z)$. Using Eqs. (A3) and (A4), one may easily show by direct substitution that

$$[I-\hat{k}(z)]H^*(z)H(-z)=I, \quad (A5)$$

is valid for all z.

To link Mullikin's results to the X and Y matrices used in this paper, we note that the multigroup transport equation with source Q in half-space problems may be reduced to an equation for the density ρ (cf. Eq. I-A7),

$$\rho(x)=\int_0^\infty k(|x-x'|)\rho(x')\,dx'+\bar{Q}(x)$$

where

$$k(|x-x'|)=E_1(\Sigma|x-x'|)C,$$

with

$$[E_1(\Sigma|x-x'|)]_{ij}=\delta_{ij}\int_0^1(1/\mu)\exp[-\sigma_i|x-x'|/\mu]\,d\mu.$$

Let us call K_x the operator with kernel k. Using this particular k, we calculate $\hat{k}$ and find that it is related to the dispersion matrix Λ by the relationship

$$\Lambda(z)=\Sigma[I-\hat{k}(-i/z)]C^{-1}\Sigma. \quad (A6)$$

From Eqs. (A5) and (A6) we obtain

$$\Lambda(z)=\Sigma[H(i/z)]^{-1}[H^*(-i/z)]^{-1}C^{-1}\Sigma. \quad (A7)$$

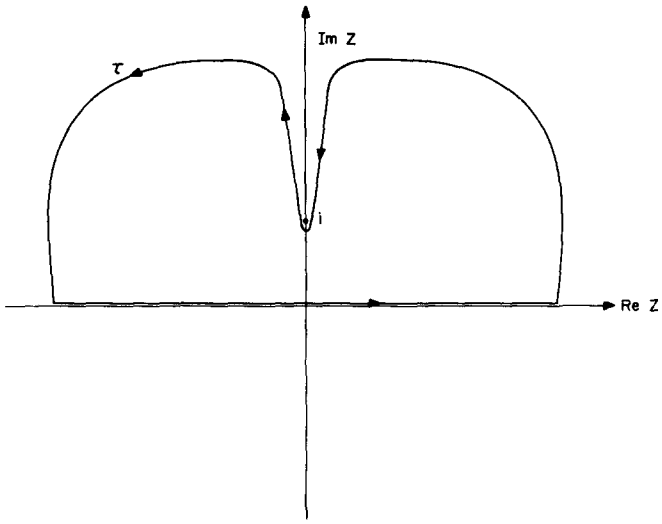


FIG. 1. The contour τ used for relating $\hat{k}(\omega)$ and $\Lambda(z)$.

If we now define

$$X(z)=[H^*(-i/z)]^{-1}C^{-1}\Sigma, \tag{A8}$$

and

$$Y(z)=\Sigma[H(-i/z)]^{-1}, \tag{A9}$$

we get Eq. (6). We recall from I that $\|k\|<1$ if $\|\Sigma^{-1}C\|<\frac{1}{2}$.

To determine Eqs. (16) and (17) for the X and Y matrices, we consider the contour τ given in Fig. 1 and note that the integrands in Eqs. (A1) and (A2) are analytic inside τ and have a branch cut $[i,i\infty)$ due to $\hat{k}$. Since $\hat{k}$ vanishes as $|z|\to\infty$, we can write the integral in Eq. (A1) as

$$\lim_{R\to\infty}\frac{1}{2\pi i}\int_{-R}^{R}H_1(t)\hat{k}(-t)\frac{dt}{t+z}$$
$$=\lim_{R\to\infty}\left[\int_{iR}^{i}H_1(\omega)\frac{\hat{k}^+(-\omega)\,d\omega}{\omega+z}\right.$$
$$\left.+\int_{i}^{iR}H_1(\omega)\frac{\hat{k}^-(-\omega)}{\omega+z}\,d\omega\right]. \tag{A10}$$

Using Eq. (A10), we can write Eq. (A1) as

$$H_r^{-1}(-i/z)$$
$$=I+\frac{z}{z\pi i}\int_0^1\frac{H_1(i/s)}{s(s+z)}[\hat{k}^+(i/s)-\hat{k}^-(i/s)]\,ds. \tag{A11}$$

We calculate $k^+(i/s)-k^-(i/s)$, $0<s\leq 1$ from Eq. (A6) and substitute into Eq. (A11) to get

$$H_r^{-1}(-i/z)=I+\frac{z}{2\pi i}\int_0^1\frac{ds\,H_1(i/s)}{s(s+z)}\Sigma^{-1}[\Lambda^+(s)-\Lambda^-(s)]\Sigma^{-1}C. \tag{A12}$$

Identifying $X(z)$ and $Y(z)$ from Eqs. (A8) and (A9), we get Eq. (16). A similar analysis on Eq. (A2) yields Eq. (17).

APPENDIX B: UNIQUENESS OF THE X AND Y MATRICES

The factorization of Λ, Eq. (16) [along with conditions (i)–(iii)] is unique up to right multiplication of Y by a constant invertible matrix R and left multiplication of X by R^{-1}. [From Eq. (9) we observe that such a transformation does not affect E, and hence leaves the solution of the transport equation unchanged.) In addition, solutions of the nonlinear equations, (16) and (17) always factor Λ. Furthermore, these equations "normalize" X and Y. To be quite specific, we state these results as lemmas.

Lemma 1: Any pair of matrices X and Y which satisfy Eqs. (16) and (17) provide a factorization Eq. (6) of Λ.

Remark: Only the converse of this has been proved by Mullikin.

Proof: Eqs. (16) and (17) may be combined to give

$$[Y(-z)-\Sigma][X(z)-C^{-1}\Sigma]=\frac{-z^2}{(2\pi i)^2}\int_0^1\int_0^1\frac{[\Lambda^+(s)-\Lambda^-(s)]}{s(s+z)}$$
$$\times X^{-1}(-s)\,\frac{Y^{-1}(t)[\Lambda^+(t)-\Lambda^-(t)]}{t(t-z)}\,ds\,dt. \tag{B1}$$

If the right hand side of (B1) is expanded by a partial fraction decomposition and common terms are cancelled, we obtain

$$Y(-z)X(z)=\Sigma C^{-1}\Sigma+(z/z\pi i)$$
$$\times\int_{-1}^{1}[\Lambda^+(s)-\Lambda^-(s)/s(s-z)]\,ds=\Lambda(z), \tag{B2}$$

proving the lemma. (The definition of Λ, Eq. (I-5), has been utilized.)

Lemma 2: Let $X(z)$ and $Y(z)$ satisfy Eqs. (16) and (17) plus the conditions (i)–(iii) following Eq. (6). Let $X'(z)$ and $Y'(z)$ satisfy the same equations and the same conditions. Then

$$X(z)=X'(z) \quad \text{and} \quad Y(z)=Y'(z)$$

Remark: In a sense, this result is unimportant, since we already know that a factorization can be computed from (16) and (17) and, unique or not, it will provide a solution to the transport equation. However, it is interesting to note the constraints on the solutions do not require more than verifying conditions (i)–(iii) mentioned above.

Proof: The matrices

$$D_1(z)=[Y'(z)]^{-1}Y(z) \text{ and } D_2(z)=X(z)[X'(z)]^{-1}$$

are analytic everywhere in the complex plane except perhaps for a cut along (0,1) and poles at $\{+\nu_i\}$, $i=1,\cdots,n$. Also, because X, Y and X', Y' both satisfy (10) and (17),

$$\lim_{|z|\to\infty}D_1(z)=\lim_{|z|\to\infty}D_2(z)=I. \tag{B3}$$

Now calculate

$$D_1(-z)D_2(z)=[Y'(-z)]^{-1}Y(-z)X(z)[X'(z)]^{-1}=I, \tag{B4}$$

where *Lemma* 1 has been invoked. Similarly,

$$D_1(z)D_2(-z)=I. \tag{B5}$$

These equations are valid for all z. Since $D_2(z)$ is analytic in the left half plane, it follows from Eq. (B4), that $D_1(-z)$ is also analytic in the left half plane, i.e, that $D_1(z)$ is analytic in the right half plane. Similarly, from (B5) we conclude that D_2 is analytic in the right half plane. Thus, D_1 and D_2 are analytic everywhere and approach I at infinity. Hence,

$$D_1(z)=D_2(z)=I$$

or

$$[Y'(z)]^{-1}Y(z)=I \text{ and } X(z)[X'(z)]^{-1}=I.$$

Similarly, by redefining D_1 and D_2, we can show

$$Y(z)[Y'(z)]^{-1}=I \text{ and } [X'(z)]^{-1}X(z)=I.$$

So

$$X(z)=X'(z) \quad \text{and} \quad Y(z)=Y'(z).$$

APPENDIX C: SOLUTION OF HALF-SPACE TRANSPORT PROBLEMS

The solution of half space problems is similar to the infinite medium case considered in I. Consider the Albedo problem. That is, we seek solutions of the transport equation in the source free half space subject to

$$\psi(x,\mu)\to 0 \quad \text{as} \quad |x|\to\infty, \tag{C1}$$

and

$$\psi(x,\mu)=\psi_0(\mu), \quad \mu>0, \tag{C2}$$

where ψ_0 represents the incident distribution. As in I, we expand ψ_0, in a "half-range expansion"

$$\psi_0(\mu)=\sum_{i=1}^{n} a_i'\phi_{\nu_i}(\mu)+\int_0^1 \Phi(\nu,\mu)A'(\nu)\,d\nu. \tag{C3}$$

Then

$$\psi(x,\mu)=\sum_{i=1}^{n} a_i' \exp(-x/\nu_i)\Phi_{\nu_i}(\mu)+ + \int_0^1 \Phi(\nu,\mu)\exp(-x/\nu)A'(\nu)\,d\nu. \tag{C4}$$

Equation (C4) is the (unique) solution because, as in I, it satisfied the transport equation and obeys the boundary conditions (C1) and (C2).

*Based in part on a doctoral dissertation submitted by one of the authors (S.S.) to the Graduate School, VPI and SU. Research supported by the National Science Foundation under Grant No. GK-35903.

†John Simon Guggenheim Memorial Foundation Fellow.

[1]R. L. Bowden, Selim Sancaktar and P.F. Zweifel, J. Math. Phys. **17**, 76 (1976), preceding paper. Equations in this paper are referred to as (I-n).
[2]E.W. Larsen and G.J. Habetler, Commun. Pure Appl. Math. **26**, 525 (1973).
[3]T. Yoshimura and S. Katsuragi, Nucl. Sci. Eng. **33**, 297 (1968).
[4]C. E. Siewert and P.F. Zweifel. Ann. Phys. (N.Y.) **36**, 61 (1966).
[5]C. E. Siewert and P.F. Zweifel, J. Math. Phys. **7**, 2092 (1966).
[6]E. E. Burniston, *et al.*, Nucl. Sci. Eng. **45**, 331 (1971).
[7]S. Pahor and A. Suhadolc, Transp. Theory Stat. Phys. **2**, 335 (1972).
[8]R. L. Bowden, Transp. Theory Stat. Phys. **4**, 25 (1975).
[9]I. C. Gohberg and M.G. Krein, Usp. Mat. Nauk **13**, 3 (1958) [AMS Transl. **14**, 217 (1960)].
[10]T.W. Mullikin, Transp. Theory Stat. Phys. **3**, 215 (1973).
[11]G. F. Mandzavidre and B.V. Huedelidre, Dokl. Akad. Nauk SSSR **123**, 791 (1958).

An upper bound on the energy gap in the $(\lambda\phi^4+\sigma\phi^2)_2$ model

Abel Klein

Department of Mathematics, University of California, Irvine, California 92664

Lawrence J. Landau

Institut de Physique Théorique, Université de Louvain, B-1348 Louvain-La-Neuve, Belgium
(Received 23 April 1975)

Let $\Delta E(\lambda,\sigma,m_0)$ be the energy gap in the infinite volume quantum field theory with bare mass m_0 and interaction $\lambda\phi^4+\sigma\phi^2$ in two space–time dimensions, obtained with either full-Dirichlet or half-Dirichlet boundary conditions. Then $\Delta E(\lambda,\sigma,m_0)/m_0<\exp[\sigma/(m_0^2+3\lambda/\Pi)]$ for all $\lambda>0$ and σ real. In particular $\Delta E(\lambda,0,m_0)<m_0$ for all $\lambda>0$ and $\Delta E(\lambda,\sigma,m_0)/m_0\to 0$ as $\sigma\to-\infty$. For half-Dirichlet boundary conditions one also has $\Delta E(\lambda,\sigma,m_0)/m_0<(1+2\sigma/m_0^2)^{1/2}$ for $\sigma\geq 0,\lambda>0$. In each pure phase of a $(\lambda\phi^4)_2$ theory, let m_{phys} be the energy gap and $\langle\phi\rangle$ the expectation of the field. Then $m_{\text{phys}}/m_0<\exp[2\Pi\langle\phi\rangle^2/(1+\Pi m_0^2/3\lambda)]$.

1. INTRODUCTION

Let $\Delta E(\lambda,\sigma,m_0)$ be the energy gap between the ground state and first excited state for the infinite volume field theory in two space–time dimensions with interaction $\lambda\phi^4+\sigma\phi^2$ and bare mass m_0. According to the Goldstone picture (see, e.g., Ref. 1a) one considers the "potential" part of the Lagrangian

$$V(\phi)=\lambda\phi^4+\sigma\phi^2+\tfrac{1}{2}m_0^2\phi^2.$$

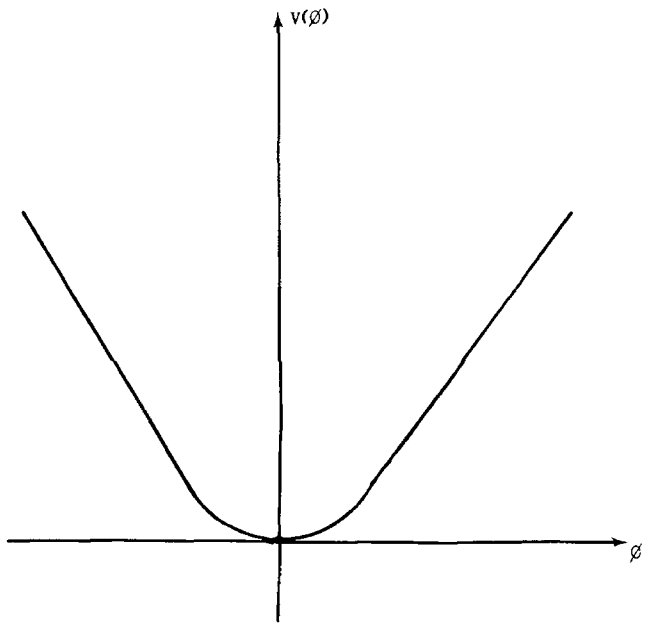
For $\sigma>-m_0^2/2$, $V(\phi)$ has the form shown in Fig. 1a, whereas for $\sigma<-m_0^2/2$ it has the form in Fig. 1b.

The simplest version of the Goldstone picture would then say that for $\sigma>-m_0^2/2$ there is a unique ground state, while for $\sigma<-m_0^2/2$ there is a doubly degenerate ground state. This picture is certainly too simple since it ignores Wick ordering.[2,3] In fact Figs. 1a and 1b are in some sense equivalent since the theory with parameters (λ,σ,m_0) is equivalent to the theory $(\lambda,0,m_\sigma)$ if[2,3]

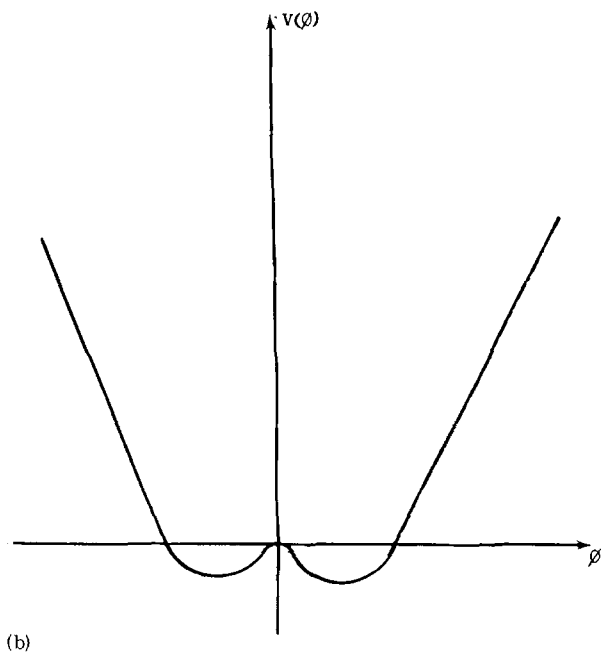
$$\sigma=\frac{m_\sigma^2-m_0^2}{2}+\frac{3\lambda}{2\Pi}\ln\frac{m_\sigma^2}{m_0^2} \tag{1}$$

However, one does expect the following to occur (e.g., Ref. 4): Holding the bare mass m_0 and the coupling constant λ fixed, ΔE should be strictly positive for large σ, and ΔE should monotonically decrease until at a critical value $\sigma=\sigma_c$ the gap ΔE becomes zero and the ground state degenerate. For values of $\sigma<\sigma_c$ the ground state remains degenerate and so $\Delta E=0$ (see Fig. 2).

Guerra, Rosen, and Simon[2] have shown the monotonicity of ΔE with σ and Glimm, Jaffe, and Spencer[1] have proved that for large positive σ (fixed m_0 and λ) the ground state is unique and $\Delta E>0$. One would like to show $\sigma_c>-\infty$, i.e., there exists a finite σ_c such that $\Delta E=0$ for $\sigma\leq\sigma_c$. (Dobrushin and Minlos[5] have announced a result of this type, but no details have appeared.) There have been no results for large coupling ($\sigma\to-\infty$); in particular, no estimates on how fast ΔE falls off as



(a)



(b)

FIG. 1.

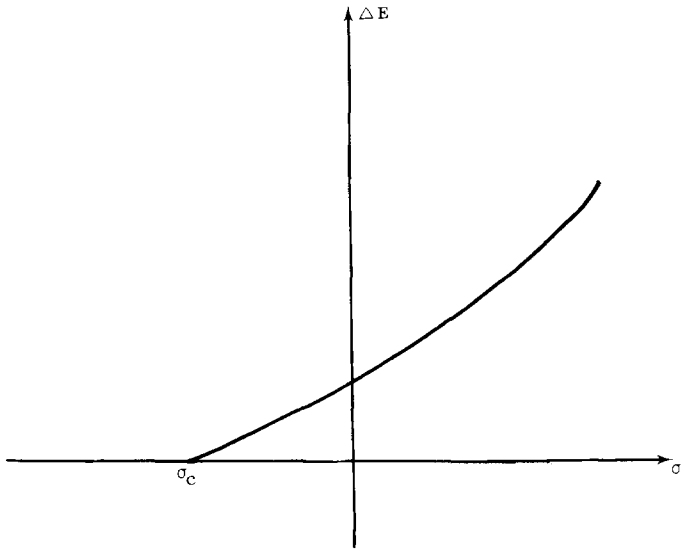


FIG. 2. Expected dependence of ΔE on σ for fixed bare mass m_0 and coupling constant λ.

σ decreases or even if it goes to zero as $\sigma \to -\infty$.

In this paper we obtain a bound on ΔE valid for all $\lambda > 0$, σ real and $m_0 > 0$, and show that ΔE decreases exponentially to zero as $\sigma \to -\infty$. The precise result is

Theorem 1: Let $\Delta E(\lambda, \sigma, m_0)$ be the energy gap for the infinite volume limit $(\lambda\phi^4 + \sigma\phi^2)_2$ theory with bare mass m_0, obtained with either half- or full-Dirichlet boundary conditions. Then

(i) $\dfrac{\Delta E(\lambda, \sigma, m_0)}{m_0} < \exp \dfrac{\sigma}{m_0^2 + 3\lambda/\Pi}$ for all $\lambda > 0$ and σ real.

In particular, $\Delta E(\lambda, 0, m_0) < m_0$ for all $\lambda > 0$ and $\Delta E(\lambda, \sigma, m_0)/m_0 \to 0$ as $\sigma \to -\infty$.

(ii) In the case of half-Dirichlet boundary conditions one also has

$$\frac{\Delta E(\lambda, \sigma, m_0)}{m_0} < \sqrt{1 + 2\sigma/m_0^2} \quad \text{for all } \lambda > 0 \text{ and } \sigma \geq 0. \qquad \square$$

This theorem gives the following picture for the upper bound on ΔE (Fig. 3). This supports the conventional picture (compare Figs. 2 and 3).

In comparing our result $\Delta E < m_0$ for $(\lambda\phi^4)_2$ with those results previously obtained, we note that Guerra, Rosen, and Simon[2] have proved that $\Delta E \leq m_0$ for sufficiently small λ (using perturbation theory). They have also shown that for fixed λ, ΔE is monotone increasing in the bare mass m_0. It follows by a dimensional argument that $\Delta E/\lambda^{1/2}$ is monotone decreasing in λ (which does not imply $\Delta E \leq m_0$). If on the other hand, if one knew that $\Delta E/m_0$ is monotone increasing in m_0, it would follow that ΔE is monotone decreasing in λ. Another related result, due to Glimm and Jaffe[6] is the estimate[7]

$$\frac{d}{d\sigma} \Delta E(\lambda, \sigma, m_0)^2 \leq 2 \quad \text{for } \sigma > \sigma_c.$$

If in addition one assumes the existence of a critical point such that $\Delta E(\lambda, 0, m_c) = 0$, it follows that

$$\Delta E(\lambda, 0, m_0) \leq \left(m_0^2 - m_c^2 + \frac{3\lambda}{\Pi} \ln \frac{m_0^2}{m_c^2}\right)^{1/2}$$

(which does not imply $\Delta E \leq m_0$).

We prove Theorem 1 in Sec. 4. The key ingredients are the existence[8] of the infinite volume limit for $\langle : \phi(x)^2 : \rangle$ together with a lower bound on $\langle : \phi(x)^2 : \rangle$ (Sec. 2). Here and throughout the paper we denote by $\langle \rangle_\Lambda$ a finite volume full- or half-Dirichlet expectation, and $\langle \rangle = \lim_{\Lambda \to R^2} \langle \rangle_\Lambda$. We will deal exclusively with full and half-Dirichlet states even if not always explicitly mentioned.

Heuristically the estimate $\Delta E \leq m_0$ for $\lambda\phi^4$ follows from the fact that $\langle : \phi(x)^2 : \rangle \geq 0$ (Ref. 9) and that the infinite volume limit two point function $S(x-y)$ has the form

$$\int_0^\infty d\rho(m) S_m(x-y)$$

where $S_m(x-y)$ is the two point function of a free field with mass m. From the canonical commutation relations[11] $\int_0^\infty d\rho(m) = 1$ and so

$$0 \leq \langle : \phi(0)^2 : \rangle = \lim_{x \to 0} [S(x) - S_{m_0}(x)]$$
$$= \lim_{x \to 0} \int d\rho(m) [S_m(x) - S_{m_0}(x)]$$

implies that there must be a contribution to $S(x)$ coming from masses $m \leq m_0$ [since $S_m(x) < S_{m_0}(x)$ if $m > m_0$]. We make this argument rigorous in Secs. 4 and 6.

Notice that we are using a small distance limiting procedure $\lim_{|x-y| \to 0} \langle : \phi(x)\phi(y) : \rangle$ rather than the more usual large distance behavior $\lim_{|x-y| \to \infty} \langle \phi(x)\phi(y) \rangle$ to obtain information on the energy gap.

The technical complication which arises with this approach is the identification of $\langle : \phi(x)^2 : \rangle$ with the limit $\lim_{y \to x} \langle : \phi(x)\phi(y) : \rangle$. One may formulate the problem as follows: Let ϕ_g denote the field ϕ with a momentum cutoff. Is it true that

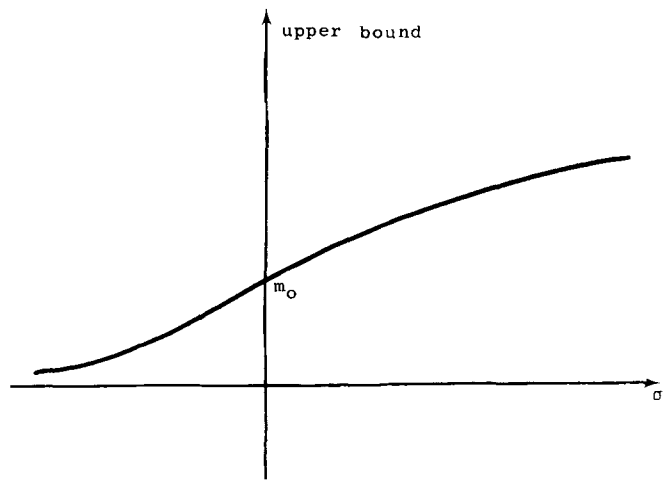


FIG. 3. Upper bound on ΔE for half-Dirichlet boundary conditions.

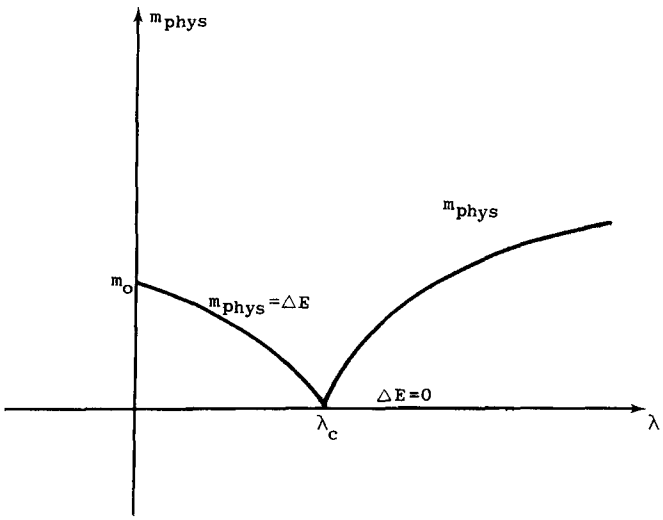


FIG. 4. Expected dependence of m_{phys} on λ for fixed bare mass m_0.

$$\lim_{\varepsilon\to\delta}\lim_{\Lambda\nearrow R^2}\langle:\phi_\varepsilon^2:\rangle_\Lambda=\lim_{\Lambda\nearrow R^2}\lim_{\varepsilon\to\delta}\langle:\phi_\varepsilon^2:\rangle_\Lambda\ ?$$

It is interesting that we will not need to show the equality; it will suffice to show

$$\lim_{\varepsilon\to\delta}\lim_{\Lambda\nearrow R^2}\langle:\phi_\varepsilon^2:\rangle_\Lambda\geq\lim_{\Lambda\nearrow R^2}\lim_{\varepsilon\to\delta}\langle:\phi_\varepsilon^2:\rangle_\Lambda$$

(see Sec. 3).

We use the integration by parts technique to obtain some information relating the mass gap m_{phys} and expectation $\langle\phi\rangle$ in each *pure* theory (corresponding to a decomposition into non-degenerate ground states) of $(\lambda\phi^4)_2$. According to the conventional picture ΔE decreases to zero as λ increases to λ_c (holding m_0 fixed and $\sigma=0$). For $\lambda>\lambda_c$, $\Delta E=0$. If, however, we consider the physical theory corresponding to each pure theory it is expected that $\Delta E\neq 0$. If m_{phys} denotes the energy gap between the ground state and first excited state in each "pure phase" then $m_{\text{phys}}=\Delta E$ if $\lambda<\lambda_c$. If $\lambda>\lambda_c$, $\Delta E=0$ but $m_{\text{phys}}>0$ (see Fig. 4).[1a]

If the $(\lambda\phi^4)_2$ theory has the expected particle interpretation, m_{phys} would be the smallest particle mass. Nothing is known about m_{phys} for $\lambda>\lambda_c$. We obtain an inequality relating m_{phys} and $\langle\phi\rangle$. More precisely, we consider first the interaction $\lambda\phi^4-\mu\phi$ with $\mu>0$. Then it is known that the ground state is unique.[12] Let ΔE_μ be the energy gap and $\langle\phi\rangle_\mu$ the expectation of ϕ.

Definition:

$$m_{\text{phys}}=\lim_{\mu\downarrow 0}\Delta E_\mu,\quad \langle\phi\rangle_+=\lim_{\mu\downarrow 0}\langle\phi\rangle_\mu.$$

Remark 1: These limits exist since both ΔE_μ and $\langle\phi\rangle_\mu$ are nonnegative and monotone decreasing as $\mu\downarrow 0$. $\langle\phi\rangle_\mu$ decreases by the Griffiths' inequalities and ΔE_μ decreases by the GHS inequalities (see Ref. 13, Theorem IX. 9). The following theorem is proven in Sec. 5.

Theorem 2:

$$\frac{\Delta E_\mu}{m_0}<\exp\left(\frac{2\Pi\langle\phi\rangle_\mu{}^2}{1+\Pi m_0^2/3\lambda}\right)$$

and, consequently,

$$\frac{m_{\text{phys}}}{m_0}<\exp\left(\frac{2\Pi\langle\phi\rangle_+{}^2}{1+\Pi m_0^2/3\lambda}\right).$$ □

Remark 2: Note that in contrast to the result $\Delta E/m_0<1$ (for $\mu=0$, $\sigma=0$) we see that it is possible for m_{phys}/m_0 to be >1 provided $\langle\phi\rangle_+$ is sufficiently large.

2. APPLICATIONS OF THE INTEGRATION BY PARTS FORMULA

Consider a theory with interaction $P(\phi)$ in a finite volume Λ in the lattice approximation.[2] An expectation $\langle\rangle_{\Lambda,\delta}$ is obtained by integrating with respect to the measure

$$\eta\exp\left(-\sum_k\delta^2:P(q_k):\right)\exp\left(-\frac{1}{2}\sum_{kl}q_k(S^{-1})_{kl}q_l\right)$$

where η is a normalization constant, δ is the lattice spacing, and S_{nm} is the covariance $\langle q_n q_m\rangle_{\Lambda,\delta}$ for the case $P=0$. The variable $q_k=\phi_\delta(k\delta)$ is the lattice field at the site $k\delta$ and the sums over k and l are finite sums such that $k\delta,\ l\delta\in\Lambda$.

The integration by parts formula[1a]

$$\int q_i F(q)\exp[-\tfrac{1}{2}\textstyle\sum q_k(S^{-1})_{kl}q_l]\,dq$$
$$=\sum_j S_{ij}\int\frac{\partial F}{\partial q_j}\exp[-\tfrac{1}{2}\textstyle\sum q_k(S^{-1})_{kl}q_l]\,dq$$

yields

$$\langle q_i\rangle_{\Lambda,\delta}=\langle-\sum_j\delta^2 S_{ij}:P'(q_j):\rangle_{\Lambda,\delta}.\tag{2}$$

Applying the integration by parts formula twice gives

$$\langle q_iq_j-S_{ij}\rangle_{\Lambda,\delta}+\sum_k\delta^2S_{ik}S_{jk}\langle:P''(q_k):\rangle_{\Lambda,\delta}$$
$$=\langle(\sum_k\delta^2S_{ik}:P'(q_k):)(\sum_l\delta^2S_{jl}:P'(q_l):)\rangle\rangle_{\Lambda,\delta}.\tag{3}$$

Equations (2) and (3) may be applied to several situations. Consider a $(\lambda\phi^4+\sigma\phi^2-\mu\phi)_2$ infinite volume theory. The infinite volume limit of[14] $\langle:\phi^2:(f)\rangle_\Lambda$ exists (Ref. 9 or Sec. 7) and so by translation invariance $\langle:\phi(x)^2:\rangle$ is a finite number $\gamma=\langle:\phi(0)^2:\rangle$. We prove a lower bound on $\langle:\phi(x)^2:\rangle$ which extends the positivity of $\langle:\phi(x)^2:\rangle$ for $(\lambda\phi^4)_2$ of Klein—Landau.[9]

Lemma 1: In a $(\lambda\phi^4+\sigma\phi^2-\mu\phi)_2$ infinite volume limit theory with either half- or full-Dirichlet boundary conditions

$$\langle:\phi(x)^2:\rangle\geq-\frac{1}{2\Pi}\frac{\sigma}{m_0^2+3\lambda/\Pi}.$$ □

Proof[15]: Putting $i=j$ in Eq. (3) gives

$$\langle q_i^2-S_{ii}\rangle_{\Lambda,\delta}+\sum_k\delta^2S_{ik}^2\langle:P''(q_k):\rangle_{\Lambda,\delta}$$
$$=\langle(\sum_k\delta^2S_{ik}:P'(q_k):)^2\rangle_{\Lambda,\delta}\geq 0.$$

With $P(\phi)=\lambda\phi^4+\sigma\phi^2-\mu\phi$,

$$\langle q_i^2-S_{ii}\rangle_{\Lambda,\delta}+12\lambda\sum_k\delta^2S_{ik}{}^2\langle:q_k{}^2:\rangle_{\Lambda,\delta}+2\sigma\sum_k\delta^2S_{ik}^2\geq 0.$$

Taking the limit $\delta\to 0$ and then $\Lambda\nearrow R^2$ gives (see Ref. 9)

$$\langle :\phi(x)^2: \rangle \geq \frac{-2\sigma \int d^2y S_{m_0}(x-y)^2}{1+12\lambda \int d^2y S_{m_0}(x-y)^2}$$

where translation invariance has been used.

As

$$\int d^2y S_{m_0}(x-y)^2 = \frac{1}{(2\Pi)^2}\int \frac{d^2p}{(p^2+m_0^2)^2} = \frac{1}{4\Pi m_0^2}$$

the lemma is proved. ∎

Consider now the interaction $P(\phi) = \lambda\phi^4 - \mu\phi$.

Lemma 2: Let g be a real smooth function of compact support ($g \in \mathcal{D}$), $\tilde{g}(p) = \int d^2x g(x) \exp(ip\cdot x)$. In a $(\lambda\phi^4 - \mu\phi)_2$ infinite volume limit theory with either half- or full-Dirichlet boundary conditions

$$\langle :\phi(g)^2: \rangle + \frac{3\lambda}{\Pi^2}\int d^2p \frac{|\tilde{g}(p)|^2}{(p^2+m_0^2)^2}\langle :\phi(0)^2: \rangle \geq \tilde{g}(0)^2 \langle \phi(0)\rangle^2. \quad \square$$

Proof: Formula (2) gives

$$\sum_n \delta^2 g(n\delta)\langle q_n\rangle_{\Lambda,\delta} = -\sum_n \delta^2 g(n\delta)\sum_m \delta^2 S_{nm}\langle 4\lambda : q_m^3 : -\mu\rangle_{\Lambda,\delta}$$

and formula (3) gives

$$\sum_n \delta^2 \sum_m \delta^2 g(n\delta)g(m\delta)\langle q_n q_m - S_{nm}\rangle_{\Lambda,\delta}$$

$$+ 12\lambda \sum_n \delta^2 \sum_m \delta^2 \sum_k \delta^2 g(n\delta)g(m\delta)S_{nk}S_{mk}\langle :q_k^2:\rangle_{\Lambda,\delta}$$

$$= \langle \{\sum_n \delta^2 \sum_m \delta^2 g(n\delta)S_{nm}[4\lambda : q_m^3 : -\mu]\}^2\rangle_{\Lambda,\delta}.$$

Using $\langle F^2\rangle \geq \langle F\rangle^2$ and taking the limits $\delta \to 0$, $\Lambda \nearrow R^2$ gives

$$\langle :\phi(g)^2: \rangle + 12\lambda \int d^2x d^2y d^2z g(x)g(y)$$

$$\times S(x-z)S(y-z)\langle :\phi(0)^2: \rangle \geq [\int d^2x g(x)]^2 \langle \phi(0)\rangle^2.$$

Performing the integrations completes the proof. ∎

As a final application of the integration by parts formula we obtain a bound on $\langle :\phi(g)^2: \rangle$.

Lemma 3: Let $0 \leq g \in \mathcal{D}$. In a $(\lambda\phi^4 + \sigma\phi^2 - \mu\phi)_2$ infinite volume limit theory with either half- or full-Dirichlet boundary conditions

$$|\langle :\phi(g)^2: \rangle| \leq \tilde{g}(0)^2 \left[\left(1+\frac{3\lambda}{\Pi m_0^2}\right)\langle :\phi(0)^2: \rangle + \frac{\sigma}{2\Pi m_0^2}\right]$$

$$+ \frac{1}{2\Pi^2}\int \frac{d^2p\,|\tilde{g}(p)|^2}{(p^2+m_0^2)^2}\,|6\lambda\langle :\phi(0)^2: \rangle + \sigma|.$$

In particular, $|\langle :\phi(g)^2: \rangle|$ remains bounded as $g \to \delta$. $\square$

Proof: Formula (3) gives

$$\langle q_i q_j - S_{ij}\rangle_{\Lambda,\delta} + \sum_k \delta^2 S_{ik}S_{jk}[12\lambda\langle :q_k^2:\rangle_{\Lambda,\delta} + 2\sigma] = F_{ij} \quad (4)$$

where

$$F_{ij} = \langle (\sum_k \delta^2 S_{ik}[4\lambda : q_k^3 : + 2\sigma q_k - \mu])(\sum_l \delta^2 S_{jl}[4\lambda : q_l^3 :$$

$$+ 2\sigma q_l - \mu])\rangle_{\Lambda,\delta}.$$

According to the Schwarz equality $|F_{ij}| \leq F_{ii}^{1/2}F_{jj}^{1/2}$ which implies

$$|\sum_i \delta^2 \sum_j \delta^2 g(i\delta)g(j\delta)F_{ij}| \leq \sum_j \delta^2 g(j\delta)\sum_i \delta^2 g(i\delta)F_{ii}. \quad (5)$$

Now Eq. (4) with $i=j$ leads to

$$\sum_i \delta^2 g(i\delta)F_{ii} = \sum_i \delta^2 g(i\delta)\langle q_i^2 - S_{ii}\rangle_{\Lambda,\delta}$$

$$+ \sum_i \delta^2 \sum_k \delta^2 g(i\delta)S_{ik}^2[12\lambda\langle :q_k^2:\rangle_{\Lambda,\delta} + 2\sigma]. \quad (6)$$

Multiply Eq. (4) by $\delta^2 g(i\delta)\delta^2 g(j\delta)$ and sum on i and j. Using Eq. (6) and inequality (5), and taking limits $\delta \to 0$, $\Lambda \nearrow R^2$ gives

$$|\langle :\phi(g)^2: \rangle + \frac{1}{\Pi^2}\int \frac{d^2p\,|\tilde{g}(p)|^2}{(p^2+m_0^2)^2}[3\lambda\langle :\phi(0)^2: \rangle + \sigma/2]|$$

$$\leq \int d^2x g(x)^2 \left(\langle :\phi(0)^2: \rangle + \frac{1}{\Pi m_0^2}[3\lambda\langle :\phi(0)^2: \rangle + \sigma/2]\right)$$

which immediately yields the lemma. ∎

The bound of Lemma 3 leads to the following theorem, which is proved in Sec. 6.

Theorem 3: Let $0 \leq g \in \mathcal{D}$, $\int d^2x g(x) = 1$, $g_n(x) = n^2 g(nx)$. The estimate of Lemma 3, $|\langle :\phi(g_n)^2: \rangle| \leq c$ for all n, implies that the infinite volume covariance $S(x-y)$ has the form

$$S(x-y) = \int_0^\infty d\rho(m) S_m(x-y) + \langle \phi(0)\rangle^2 + d$$

where $S_m(x-y)$ is the covariance of a free field of mass m, and

$$\int_0^\infty d\rho(m) = 1, \quad \int_0^\infty d\rho(m)\,|\ln m| < \infty.$$

Here $\langle \phi(0)\rangle$ is the vacuum expectation of ϕ ($=0$ if $\mu = 0$) and d comes from translation invariant states other than the vacuum ($=0$ in case of a nondegenerate ground state).

It follows that $\lim_{n\to\infty}\langle :\phi(g_n)^2: \rangle$ exists and equals

$$\frac{1}{2\Pi}\int_0^\infty d\rho(m)\ln\frac{m_0}{m} + \langle \phi(0)\rangle^2 + d. \quad \square$$

Remark 3: The fact that no δ-functions or derivatives of δ-functions appear in $S(x-y)$ could also be obtained by bounds showing existence of time-zero fields. That $\int_0^1 d\rho(m)\ln(1/m) < \infty$ follows from the fact that S is a tempered distribution [but $\int_1^\infty d\rho(m)\ln m < \infty$ requires Lemma 3], that $\int_0^\infty d\rho(m) = 1$ also follows from the canonical commutation relations. Also, if one works with time-zero fields an upper bound on $\langle :\phi(g)^2: \rangle$, $0 \leq g \in \mathcal{D}(R^1)$, follows directly from the Griffiths' inequalities and ϕ-bounds for $:\phi(g,0)^2:$ (the lower bound also follows directly from the Griffiths' inequalities, e.g., Lemma 4). Theorem 3 thus also holds for any even $P(\phi)_2$ model.

Remark 4: The estimate of Lemma 3 also leads immediately to the conclusion that ϕ is not in the Borchers' class of a free field[16] since if $\phi = \sum C_m :\phi^m:$ the singularity in the two point function of ϕ is too great to be compensated by S_{m_0}.

3. $\lim_{g\to\delta} \langle :\phi(g)^2: \rangle \geq \langle :\phi(0)^2: \rangle$

Lemma 4: Let $0 \leq g \in \mathcal{D}$, $\int d^2x g(x) = 1$, $g_n(x) = n^2 g(nx)$. In a $(\lambda\phi^4 + \sigma\phi^2 - \mu\phi)_2$ infinite volume limit theory with

either half- or full-Dirichlet boundary conditions

$$\lim_{n\to\infty}\langle :\phi(g_n)^2:\rangle \geq \langle :\phi(0)^2:\rangle. \qquad \square$$

Proof: The limit exists by the preceding theorem. Let $:\phi(g_n)^2:(f) = \int :\phi(g_n)(x)^2: f(x)\,d^2x$, where $\phi(g_n)(x)$ is the translate of $\phi(g_n)$ by the amount x, and $0 \leq f \in \mathcal{D}$. By the Griffiths' inequalities[2]

$$\langle :\phi(g_n)^2:(f)\rangle_\Lambda \leq \langle :\phi(g_n)^2:(f)\rangle = \langle :\phi(g_n)^2:\rangle \int d^2x f(x)$$

for all Λ (containing the support of $g*f$). Thus

$$\langle :\phi^2:(f)\rangle_\Lambda = \lim_{n\to\infty}\langle :\phi(g_n)^2:(f)\rangle_\Lambda \leq \lim_{n\to\infty}\langle :\phi(g_n)^2:\rangle \int d^2x f(x).$$

Since this holds for all Λ, we may take the limit as $\Lambda \nearrow R^2$, giving

$$\langle :\phi(0)^2:\rangle \int d^2x f(x) \leq \lim_{n\to\infty}\langle :\phi(g_n)^2:\rangle \int d^2x f(x). \qquad \blacksquare$$

4. THE UPPER BOUND ON THE ENERGY GAP for $(\lambda\phi^4 + \sigma\phi^2)_2$

Lemma 5: Let $S(x) = \int_{m_1}^\infty d\rho(m) S_m(x)$ where $S_m(x)$ denotes the covariance of a free field with mass m, $m_1 \geq 0$, $\int_{m_1}^\infty d\rho(m) = 1$. Let $0 \leq g \in \mathcal{D}$, $\int d^2x g(x) = 1$, $g_n(x) = n^2 g(nx)$. Define $S(g) = \int d^2x d^2y g(x) S(x-y) g(y)$. If $\lim_{n\to\infty} S(g_n) - S_{m_0}(g_n) = \gamma$, then $m_1/m_0 \leq \exp(-2\Pi\gamma)$ with equality holding if and only if $\rho = \delta(m - m_1)$. $\square$

Proof: Assume $m_1 \neq 0$ (otherwise the lemma is trivially true). Since $\int d\rho(m) = 1$, $S(g_n) \leq S_{m_1}(g_n)$, with equality only if $\rho = \delta(m - m_1)$. Thus

$$\gamma \leq \lim_{n\to\infty} S_{m_1}(g_n) - S_{m_0}(g_n)$$

$$= \lim_{n\to\infty} \int \frac{d^2p}{(2\Pi)^2}\left(\frac{1}{p^2+m_1^2} - \frac{1}{p^2+m_0^2}\right)|\tilde{g}_n(p)|^2$$

$$= \int \frac{d^2p}{(2\Pi)^2}\left(\frac{1}{p^2+m_1^2} - \frac{1}{p^2+m_0^2}\right) = \frac{1}{2\Pi}\ln\frac{m_0}{m_1}. \qquad \blacksquare$$

Proof of Theorem 1: For the $(\lambda\phi^4 + \sigma\phi^2)_2$ theory, the vacuum expectation $\langle\phi(0)\rangle = 0$ by $\phi \to -\phi$ symmetry. If the vacuum is not the only translation invariant state $\Delta E = 0$ and so Theorem 1 is trivially true. Thus we may assume $d = 0$ in the representation of S in Theorem 3 (Sec. 2).

Part (i) of Theorem 1 is now an immediate consequence of Lemmas 1, 4, and 5. [The measure ρ cannot be of the form $\delta(m - m_1)$ since then the field would be a free field[17] and this is known not to be the case.[16]] To prove Part (ii) of Theorem 1 we use the equivalence $(\lambda, \sigma, m_0) \sim (\lambda, 0, m_\sigma)$ given by Eq. (1) which is valid for half-Dirichlet states.[2] It follows that

$$\Delta E(\lambda, \sigma, m_0) = \Delta E(\lambda, 0, m_\sigma) < m_\sigma \qquad (7)$$

by Theorem 1(*i*). Now if $\sigma \geq 0$, then $m_\sigma \geq m_0$ and $\ln m_\sigma/m_0 > 0$. Thus $(m_\sigma^2 - m_0^2)/2 \leq \sigma$, i.e., $m_\sigma \leq \sqrt{m_0^2 + 2\sigma}$, which with Eq. (7) completes the proof of the theorem. $\blacksquare$

Remark 5: Since $x - 1 \geq \ln x$, Eq. (1) implies

$$\sigma \geq \left(\frac{m_0^2}{2} + \frac{3\lambda}{2\Pi}\right)\ln\frac{m_\sigma^2}{m_0^2}$$

which, together with Eq. (7), immediately implies the bound of Theorem 1(i). Thus for half-Dirichlet states, the exponential fall-off of ΔE as $\sigma \to -\infty$ is a direct consequence of the bound $\Delta E(\lambda, \sigma = 0, m_0) \leq m_0$ valid for all $m_0 > 0$.

5. THE RELATION BETWEEN m_{phys} and $\langle\phi\rangle$

It is known[12] that the ground state is unique for the $\lambda\phi^4 - \mu\phi$ theory if $\mu \neq 0$. Thus, $d = 0$ in the representation of S given in Theorem 3 (Sec. 2) and

$$S(x) = \int_{m_1}^\infty d\rho(m) S_m(x) + \langle\phi(0)\rangle_\mu^2. \qquad (8)$$

We may assume $m_1 \neq 0$ (otherwise Theorem 2 is trivially true). From Lemma 2

$$\lim_{g\to\delta}\langle :\phi(g)^2:\rangle_\mu + \frac{3\lambda}{\Pi m_0^2}\langle :\phi(0)^2:\rangle_\mu \geq \langle\phi(0)\rangle_\mu^2.$$

Thus from Lemma 4

$$\left(1 + \frac{3\lambda}{\Pi m_0^2}\right)\lim_{g\to\delta}\langle :\phi(g)^2:\rangle_\mu \geq \langle\phi(0)\rangle_\mu^2.$$

Using the representation (8) one has

$$\lim_{g\to\delta}\int_{m_1}^\infty d\rho(m)[S_m(g) - S_{m_0}(g)] \geq -\frac{\langle\phi(0)\rangle_\mu^2}{1 + \Pi m_0^2/3\lambda}.$$

The application of Lemma 5 now completes the proof of Theorem 2.

6. PROOF OF THEOREM 3

The most general relativistic two point function is given by the Källen—Lehmann representation (e.g., Ref. 18)

$$W(s, \mathbf{x}) = \int_0^\infty d\rho(m) \int_{-\infty}^\infty \frac{dp}{4\Pi\sqrt{p^2+m^2}} \exp[i(\sqrt{p^2+m^2}\,s - p\mathbf{x})] + c$$

where c is the contribution from translation invariant states and ρ is a regular Borel measure such that there exists an integer N satisfying $\int d\rho(m)/(1+m^2)^{N+1} < \infty$. Because of the singularity at $m = 0$, it is necessary in two space—time dimensions that $\int_0^1 d\rho(m) \ln 1/m < \infty$ which guarantees that W is a tempered distribution. [Let $f \in \mathcal{S}(R^2)$ be chosen so that $\tilde{f}(p_1, p_2) \geq 0$ and $\tilde{f}(p_1, p_2) \geq 1$ for $p_1^2 + p_2^2 \leq 3$. Then

$$\frac{1}{2\Pi}\int_0^1 d\rho(m) \ln\frac{1}{m} \leq \int d\rho(m) \int_{-\infty}^\infty \frac{dp}{4\Pi\sqrt{p^2+m^2}} \tilde{f}(\sqrt{p^2+m^2}, p)$$

$$= W(f) < \infty.$$

Consider

$$W_m(s, \mathbf{x}) = \int_{-\infty}^\infty \frac{dp}{4\Pi\sqrt{p^2+m^2}} \exp[i(\sqrt{p^2+m^2}\,s - p\mathbf{x})].$$

Analytically continuing this to $s = it$, $t > 0$ we have

$$\hat{S}_m(t, \mathbf{x}) = \int_{-\infty}^\infty \frac{dp}{4\Pi} \exp\frac{(-\sqrt{p^2+m^2}\,t - ip\mathbf{x})}{\sqrt{p^2+m^2}} \qquad (t > 0)$$

which is analytic in $\mathbf{x}$, t for $t > 0$. Notice that

$$(-\Delta + m^2)\hat{S}_m = 0 \quad \text{for } t > 0, \quad \Delta = \frac{\partial^2}{\partial t^2} + \frac{\partial^2}{\partial \mathbf{x}^2}.$$

Thus

$$\hat{S}_m^{(N)} \equiv \frac{(1+\Delta)^N}{(1+m^2)^N}\hat{S}_m$$

is equal to $\hat{S}_m$ for $t > 0$.

Therefore

$$S^{(N)}(t, \mathbf{x}) \equiv \int_0^\infty \frac{d\rho(m)}{(1+m^2)^N}(1+\Delta)^N \hat{S}_m(|t|, \mathbf{x}) + c$$
$$= \int_0^\infty \frac{d\rho(m)}{(1+m^2)^N}\int \frac{d^2p}{(2\Pi)^2}\frac{(1-p^2)^N}{p+m^2}\exp(ip\cdot x) + c$$

is a tempered distribution[19] which equals the analytic continuation of W to $s = it$, $t > 0$. It is Euclidean invariant and analytic away from $x = 0$ ("coinciding arguments"). Here $x = (t, \mathbf{x})$.

Now any tempered distribution which agrees with the analytic continuation of W away from $x = 0$ must differ from $S^{(N)}$ by a finite number of derivatives of δ-functions. By Euclidean invariance the derivatives must have the form $\sum_{j=0}^{J} a_j(-\Delta)^j\delta(x)$.

We have thus shown that the general form of the Fourier transform of the Euclidean covariance is

$$S(p) = \int_0^\infty \frac{d\rho(m)}{(1+m^2)^N}\frac{1}{(2\Pi)^2}\frac{(1-p^2)^N}{p^2+m^2} + c\delta(p) + \sum_{j=0}^{J} a_j(p^2)^j.$$

Changing the value of N in this representation for S corresponds to a change in the parameters a_j. We may thus suppose that S has been brought to a standard form where $N = 0$ if $\int_0^\infty [d\rho(m)/1+m^2] < \infty$, and otherwise N is chosen so that $\int_0^\infty [d\rho(m)/(1+m^2)^{N+1}] < \infty$ but $\int_0^\infty [d\rho(m)/(1+m^2)^N] = \infty$. We now consider the restrictions on S which follow from the estimate of Lemma 3:

$$|S(g_n) - S_{m_0}(g_n)| \leq c \quad \text{for all } n$$

where

$$S(g) = \int S(p)|\tilde{g}(p)|^2 d^2p \quad \text{and} \quad \tilde{g}_n(p) = \tilde{g}(p/n), \quad \tilde{g}(0) = 1.$$

It is easy to see that the only way the singularities of S and S_{m_0} can compensate as $n \to \infty$ is for $N = 0$ and $a_j = 0$ for all j. For an explicit proof, consider first the free covariance

$$S_{m_0}(g_n) = \int \frac{d^2p}{(2\Pi)^2}\frac{|\tilde{g}(p)|^2}{p^2+m_0^2/n^2} \underset{n\to\infty}{\sim} \ln n.$$

As for the behavior of $S(g)$, the δ-functions contribution behaves like

$$\sum_{j=0}^{J} n^{2(j+1)} a_j \int d^2p (p^2)^j |\tilde{g}(p)|^2.$$

The remaining contribution is (assuming $N > 0$)

$$\int \frac{d\rho(m)}{(1+m^2)^N}\int \frac{d^2p}{(2\Pi)^2}\frac{(1-p^2)^N}{p^2+m^2}\left|\tilde{g}\left(\frac{p}{n}\right)\right|^2$$
$$= \sum_{j=0}^{N}(-1)^j n^{2j} f_j(n)$$

where

$$f_j(n) = \binom{N}{j}\int \frac{d\rho(m)}{(1+m^2)^N}\int \frac{d^2p}{(2\Pi)^2}\frac{p^{2j}}{p^2+m^2/n^2}|\tilde{g}(p)|^2.$$

As $n \to \infty$, $f_j(n)/n^2 \to 0$ by a dominated convergence argument, but $f_j(n) \to \infty$ as $n \to \infty$ since the integrand monotone increases to $p^{2(j-1)}|\tilde{g}(p)|^2/(1+m^2)^N$ which gives an infinite integral since $\int [d\rho(m)/(1+m^2)^N] = \infty$. Thus the large n behavior of $n^{2j}f_j(n)$ cannot be compensated by a power behavior bn^{2k}. Therefore for large n, $S(g_n) - S_{m_0}(g_n)$ behaves like

$$\sum_{j=0}^{N}(-1)^j n^{2j} f_j(n) + \sum_{j=0}^{J} c_j n^{2(j+1)} - b\ln n. \tag{9}$$

Using the properties of f_j stated above it follows readily that the quantity (9) cannot remain finite unless all $a_j = 0$ and $N = 0$, i.e., $\int [d\rho/(1+m^2)] < \infty$.

We may now conclude further that $\int_0^\infty d\rho(m) = 1$: If $\int d\rho > 1$ there exist M_1, M_2 such that $0 < M_1 \leq M_2 < \infty$ and

$$1 < \int_{M_1}^{M_2} d\rho(m) < \infty.$$

So

$$S(g_n) - S_{m_0}(g_n)$$
$$\geq \int_{M_1}^{M_2} d\rho(m)\int \frac{d^2p}{(2\Pi)^2}\left(\frac{1}{p^2+m^2} - \frac{1}{p^2+m_0^2}\right)\left|\tilde{g}\left(\frac{p}{n}\right)\right|^2$$
$$+ \left(\int_{M_1}^{M_2} d\rho - 1\right) S_{m_0}(g_n) + c.$$

The first term remains bounded as $n \to \infty$ and the second term goes to $+\infty$ since $S_{m_0}(g_n) \to +\infty$. This contradicts Lemma 3.

Thus $\int d\rho(m) \leq 1$. If $\int d\rho < 1$,

$$S(g_n) - S_{m_0}(g_n) = \int d\rho(m)\int \frac{d^2p}{(2\Pi)^2}\left(\frac{1}{p^2+m^2} - \frac{1}{p^2+m_0^2}\right)\left|\tilde{g}\left(\frac{p}{n}\right)\right|^2$$
$$+ \left(\int d\rho - 1\right) S_{m_0}(g_n) + c.$$

The first term is negative for $m \geq m_0$ and for $m \leq m_0$ the integral is finite since $\int_0^{m_0} d\rho(m)\ln(m_0/m) < \infty$. The second term $\to -\infty$ as $n \to \infty$. Thus again we have a contradiction with Lemma 3, so $\int d\rho(m) = 1$.

We may finally conclude that $S(g_n) - S_{m_0}(g_n)$ *converges* as $n \to \infty$, since

$$S(g_n) - S_{m_0}(g_n)$$
$$= \int_{m_0}^\infty d\rho(m)\int \frac{d^2p}{(2\Pi)^2}\left(\frac{1}{p^2+m^2} - \frac{1}{p^2+m_0^2}\right)\left|\tilde{g}\left(\frac{p}{n}\right)\right|^2$$
$$+ \int_0^{m_0} d\rho(m)\int \frac{d^2p}{(2\Pi)^2}\left(\frac{1}{p^2+m^2} - \frac{1}{p^2+m_0^2}\right)\left|\tilde{g}\left(\frac{p}{n}\right)\right|^2.$$

The second term converges to $(2\pi)^{-1}\int_0^{m_0} d\rho(m)\ln(m_0/m)$. The first term is negative. If $\int_{m_0}^\infty d\rho(m)\ln(m/m_0) < \infty$ the first term converges by a dominated convergence argument. If $\int_{m_0}^\infty d\rho(m)\ln(m/m_0) = \infty$ the first term diverges to $-\infty$ since there is an $a > 0$ such that $|\tilde{g}(p)| > \frac{1}{2}$ for $|p| \leq a$ which implies that the first term is

$$\leq \frac{1}{4}\int_{m_0}^\infty d\rho(m)\int_{|p|\leq na} \frac{d^2p}{(2\Pi)^2}\left(\frac{1}{p^2+m^2} - \frac{1}{p^2+m_0^2}\right)$$

which diverges to $-\infty$ as $n \to \infty$ by a monotone convergence argument. Thus Lemma 3 implies that $\lim_{n\to\infty} S(g_n) - S_{m_0}(g_n)$ exists and is equal to $(2\Pi)^{-1}\int d\rho(m) \times \ln(m_0/m) + c$. ∎

7. THE LIMIT $<:\phi^2:(f)>_\Lambda$, $\Lambda \nearrow R^2$

For completeness we give the argument showing that $\lim_{\Lambda\nearrow R^2}\langle:\phi^2:(f)\rangle_\Lambda$ exists and obtain the necessary bounds. For a more detailed discussion which proves more see Ref. 9. In this section we will denote by $\langle:\phi^2:(f)\rangle_\Lambda^D$, $\langle:\phi^2:(f)\rangle_\Lambda^{HD}$, $\langle:\phi^2:(f)\rangle_\Lambda^F$ the full-Dirichlet, half-Dirichlet, and free boundary condition expectation, and in all cases $:\phi^2:(f)$ denotes Wick ordering with respect to the free covariance, i.e., the covariance of a free field of mass m_0.

Let $0 \le f \in \mathcal{D}$, $\operatorname{supp} f \subset \Lambda_0 \subset \Lambda$. The Griffiths' inequalities[2] imply

$$\langle:\phi^2:(f)\rangle_\Lambda^D \le \langle:\phi^2:(f)\rangle_\Lambda^{HD} \le \langle:\phi^2:(f)\rangle_\Lambda^F. \tag{10}$$

The monotone increase

$$\begin{aligned}&\langle:\phi^2:(f)\rangle_\Lambda^D \le \langle:\phi^2:(f)\rangle_{\Lambda'}^D, \quad \Lambda\subset\Lambda',\\ &\langle:\phi^2:(f)\rangle_\Lambda^{HD} \le \langle:\phi^2:(f)\rangle_{\Lambda'}^{HD}, \quad \Lambda\subset\Lambda',\end{aligned} \tag{11}$$

also follows from the Griffiths' inequalities. Convergence follows from uniform bounds. These can be obtained as follows: Let $0 \le f \in \mathcal{S}$ and denote by f^Λ the restriction of f to Λ, $f_t^\Lambda(x) = f^\Lambda(t,x)$. Finally consider a partition of the real line into unit intervals I_i and for any function $h(x)$ denote by h_i the restriction of h to I_i. Then

$$\langle:\phi^2:(f)\rangle_\Lambda^D = \langle:\phi^2:(f^\Lambda)\rangle_\Lambda^D \le \langle:\phi^2:(f^\Lambda)\rangle_\Lambda^{HD} = \langle:\phi^2:(f)\rangle_\Lambda^{HD}$$

$$= \sum_i \int dt \langle:\phi^2:(f_{ti}^\Lambda)\rangle_\Lambda^{HD} \le \sum_i \int dt \lim_{T\to\infty}\langle:\phi^2:(f_{ti}^\Lambda)\rangle_{L\times T}^F$$

if $\Lambda \subset L\times T$ by (10) and (11).

Now

$$\lim_{T\to\infty}\langle:\phi^2:(f_{ti}^\Lambda)\rangle_{L\times T}^F = (\Omega_L, :\phi^2:(f_{ti}^\Lambda)\Omega_L)$$

where Ω_L is the ground state of $H_L = H_0 + \int_{-L/2}^{L/2}:P(\phi(x,0)):dx$ and

$$|(\Omega_L, :\phi^2:(f_{ti}^\Lambda)\Omega_L)|$$

$$\le c\|(H_L+c)^{-1/2}:\phi^2:(f_{ti}^\Lambda)(H_L+c)^{-1/2}\| \le c\|f_{ti}^\Lambda\|_2$$

by the generalized ϕ-bounds for H_L (e.g., Ref. 20). Thus $\langle:\phi^2:(f)\rangle_\Lambda^{D,HD} \le c\sum_i \int dt \|f_{ti}\|_2$. Since a lower bound can easily be obtained from (11) it follows that $\langle:\phi^2:(f)\rangle_\Lambda^{D,HD}$ converges for any $f\in\mathcal{S}$ and the integration by parts formula can be handled as in Sec. 5 of Ref. 9.

ACKNOWLEDGMENTS

We thank Ira Herbst, Barry Simon, and Arthur Wightman for stimulating discussions. L. J. Landau thanks Jean-Pierre Antoine and David Speiser for the hospitality of the Institut de Physique Theorique.

Note added to proof: J. Glimm, A Jaffe, and T. Spencer have proved existence of phase transitions for ϕ_2^4 (Proc. Int. Colloq. Mathematical Methods of Quantum Field Theory, Marseille, June, 1975).

[1](a) J. Glimm, A. Jaffe, and T. Spencer, "The Particle Structure of the Weakly Coupled $P(\phi)_2$ Model and Other Applications of High Temperature Expansions, Part I" in *Constructive Quantum Field Theory*, edited by G. Velo and A. Wightmann (Springer-Verlag, Berlin, pp. 1973), pp. 132—198; (b) Part II, pp. 199—242.
[2]F. Guerra, L. Rosen, and B. Simon, Ann. Math. **101**, 111 (1975).
[3]R. Baumel, Princeton University Thesis (1973).
[4]J. Glimm and A. Jaffe, "On the Approach to the Critical Point," Ann. Inst. H. Poincaré **22** (1975).
[5]R. Dobrushin and R. Minlos, "Construction of a One-Dimensional Quantum Field Via a Continuous Markov Field," Funct. Anal. Appl. 7, 324 (1973).
[6]J. Glimm and A. Jaffe, Phys. Rev. D 10, 536 (1974).
[7]Note that their interaction is parametrized by $\lambda\phi^4 + \frac{1}{2}\sigma\phi^2$.
[8]Proved by Klein and Landau[9] for half- and full-Dirichlet boundary conditions and by Schrader[10] for small coupling. We remark that in Ref. 9 all Schwinger functions involving $:\phi^2:$ were shown to converge. Since we need here only the properties of $\langle:\phi(x)^2:\rangle$, the discussion can be somewhat simplified and a sketch is given in Sec. 7.
[9]A. Klein and L. J. Landau, "The $:\phi^2:$ Field in the $P(\phi)_2$ Model," Commun. Math. Phys. (to be published).
[10]R. Schrader, "Local Operator Products and Field Equations in $P(\phi)_2$ Theories," Berlin, preprint.
[11]We will not appeal to the canonical commutation relations and will actually prove $\int d\rho = 1$ directly from the inequalities in Sec. 2 (see Sec. 6).
[12]B. Simon, Ann. Math. **101**, 260 (1975).
[13]B. Simon, *The $P(\phi)_2$ Euclidean (Quantum) Field Theory* (Princeton U. P., Princeton, 1974).
[14]We use the notations $\phi(f) = \int d^2x \phi(x) f(x)$, $:\phi^2:(f) = \int d^2x :\phi(x)^2: f(x)$.
[15]Here and in Lemmas 2 and 3 we give the proof for full-Dirichlet boundary conditions. In the half-Dirichlet case there is a slight modification due to the fact that the Wick ordering is done with respect to the free measure (see Ref. 9, proof of Theorem 6).
[16]J. Fröhlich, "Verification of Axims for Euclidean and Relativistic Fields and Haag's Theorem in a class of $P(\phi)_2$ Models," Ann. de l'Inst. Henri Poincaré **21**, 271 (1974).
[17]R. Streater and A. Wightman, *P.C.T., Spin and Statistics and All That* (Benjamin, New York, 1964).
[18]M. Reed and B. Simon, *Methods of Modern Mathematical Physics, Fourier Analysis, Self-Adjointness* (Academic, New York, 1975), Vol. II.
[19]Note that $\int d\rho/(1+m^2)^{N+1} < \infty$ implies that $S^{(N)}$ is a tempered distribution, the additional $1/(1+m^2)$ fall-off coming from $1/(p^2+m^2)$.
[20]B. Simon, "The Glimm—Jaffe ϕ-Bound: A Markov Proof" in *Constructive Quantum Field Theory*, edited by G. Velo and A. Wightman (Springer-Verlag, Berlin, 1973), pp. 125—131.

Spectrum generating group of the symmetric top molecule*

Arno Bohm and R. B. Teese

Center for Particle Theory, Physics Department, University of Texas at Austin, Austin, Texas 78712
(Received 20 May 1975)

The spectrum generating group for the symmetric top is found and applied to derive spectra, energy values and selection rules. All the well-known results are reproduced and some generalizations are obtained as a consequence of the group property, except for the Coriolis splitting for which an additional dependence upon the factor k/j is derived.

I. INTRODUCTION

According to the basic assumptions of quantum mechanics, a quantum mechanical physical system is described by an algebra of operators in a linear space. For many simple nonrelativistic systems such an algebra of observables is given by the enveloping algebra of a group, which is nowadays called a spectrum generating group.[1]

A spectrum generating group is a group which possesses an irreducible representation space that can serve as the space of physical states of the quantum mechanical system. Thus the spectrum generating group has to contain the invariance group of the physical system, and some of its irreducible representations (irreps) must reduce into a direct sum of those irreps of the invariance group which make up the spectrum of the physical system. This requirement does not completely fix the spectrum generating group, as there are in general several groups which have this property. In particular, a simple group and the contraction of this simple group with respect to the invariance group describe the same spectrum, as we shall discuss later for the particular case considered here.

For physical systems with an unlimited number of possible physical states the irrep space must be infinite dimensional and, therefore, the spectrum generating group must be noncompact. The well-known examples of spectrum generating groups are SO(4,1) or SO(4,2) for the hydrogen atom[2] SU(n,1) for the n-dimensional harmonic oscillator[3] and SO(3,1) for the rotator.[4] Another solvable model which is of great importance in molecular and nuclear physics, and whose relativistic generalization may well serve as one of the clues to particle physics, is the symmetric top. Therefore, and in view of the large literature on spectrum generating groups it is surprising that the group theory of the symmetric top has not been uncovered before.

In the succeeding sections the description of the symmetric top by a spectrum generating group will be given. The treatment will be in complete analogy to (and is a generalization of) the description of the rotator[4] that is in fact the model on which the notion of spectrum generating groups was first conceived. As the nonrelativistic symmetric top model is best realized in molecular physics, we will use the symmetric top molecule as the physical system and refer to the book of Herzberg[5] for the numerous concrete examples.

II. REPRESENTATIONS OF THE PARITY-EXTENDED SO (3,2) AND THEIR CONTRACTIONS

In this section the properties of a class of representations of SO(3,2) (covering group of the 3-2-de Sitter group) are described, and their extension by parity are discussed. This class is a subclass of representations of SO(3,2) that have a discrete reduction with respect to SO(3,1), and their properties have been derived in Ref. 6. It is also contained in the set of singleton representations listed in Ref. 7. We will state here the results only.

A. A representation of SO (3,2)

The Lie algebra of SO(3,2) is given by the following commutation relations:

$$[J_{\alpha\beta}, J_{\gamma\delta}] = i[g_{\alpha\delta}J_{\beta\gamma} + g_{\beta\gamma}J_{\alpha\delta} - g_{\alpha\gamma}J_{\beta\delta} - g_{\beta\delta}J_{\alpha\gamma}] \tag{II.1}$$

where the metric $g_{\alpha\beta}$ is given by

$$\begin{aligned} g_{\alpha\beta} &= 1 \quad \text{for } \alpha = \beta = 0 \text{ or } 4, \\ &= -1 \quad \text{for } \alpha = \beta = 1, 2, 3, \\ &= 0 \quad \text{otherwise,} \end{aligned}$$

and $J_{\alpha\beta} = -J_{\beta\alpha}[\alpha, \beta = 0, 1, 2, 3, 4]$ are the generators.

If we make the identification

$$\Gamma_\mu = J_{\mu 4}, \quad \mu = 0, 1, 2, 3, \tag{II.2}$$

we obtain

$$[J_{\mu\nu}, J_{\rho\sigma}] = i[g_{\mu\sigma}J_{\nu\rho} + g_{\nu\rho}J_{\mu\sigma} - g_{\mu\rho}J_{\nu\sigma} - g_{\nu\sigma}J_{\mu\rho}], \tag{II.3a}$$

$$[\Gamma_\mu, J_{\rho\sigma}] = i[g_{\mu\rho}\Gamma_\sigma - g_{\mu\sigma}\Gamma_\rho], \tag{II.3b}$$

$$[\Gamma_\mu, \Gamma_\nu] = -iJ_{\mu\nu}, \tag{II.3c}$$

where the range of the indices μ, ν, ρ, σ is 0, 1, 2, 3.

$J_{\mu\nu}$ are the generators of the proper Lorentz group SO(3,1) and Γ_μ is a Lorentz vector, its vector properties being defined by (II.3b).

The Casimir operators of SO(3,2) are

$$\begin{aligned} \tilde{D}_1 &= -1/2 J_{\alpha\beta}J^{\alpha\beta}, \\ \tilde{D}_2 &= +W_\alpha W^\alpha, \end{aligned} \tag{II.4}$$

where $W_\alpha = \frac{1}{8}\epsilon_{\alpha\beta\gamma\delta\epsilon}J^{\beta\gamma}J^{\delta\epsilon}$. The range of the indices is 0, 1, 2, 3, 4 and $\epsilon_{\alpha\beta\gamma\delta\epsilon}$ is the antisymmetric tensor with $\epsilon_{01234} = +1$. Hence we obtain

$$\tilde{D}_1 = -\{\Gamma_\mu \Gamma^\mu - \tilde{C}_1\}, \tag{II.5a}$$

$$\tilde{D}_2 = W_\mu W^\mu + \tilde{C}_2^{\,2}, \tag{II.5b}$$

where

$$\tilde{C}_1 = -\tfrac{1}{2} J_{\mu\nu} J^{\mu\nu} = K_i K_i - J_i J_i, \tag{II.6a}$$

$$\tilde{C}_2 = \tfrac{1}{8} \epsilon_{\mu\nu\rho\sigma} J^{\mu\nu} J^{\rho\sigma} = K_i J_i, \tag{II.6b}$$

$$W_\mu = \tfrac{1}{2} \epsilon_{\mu\nu\rho\sigma} J^{\nu\rho} \Gamma^\sigma, \tag{II.6c}$$

and $\epsilon_{\mu\nu\rho\sigma}$ is the Levi–Cevita tensor with $\epsilon_{0123} = +1$.

$\tilde{C}_1$ and $\tilde{C}_2$ are the Casimir operators of SO(3,1) and an irreducible representation[8] of SO(3,1) is characterized by (k,c) where

$$C_1 = -k^2 - c^2 + 1, \quad C_2 = ikc \tag{II.7}$$

are the eigenvalues of $\tilde{C}_1$ and $\tilde{C}_2$, respectively. The possible values of (k,c)—for the unitary principle series of representations of $SO(3,1)^g_{J_{\mu\nu}}$—are c is pure imaginary, $ic \geq 0$ and k is any integral or half-integral number.

The reduction of the representation space of (k,c) with respect to the rotation group $SO(3)_{J_{ij}}$, $i,j = 1,2,3$, is given by

$$H(kc) \underset{SO(3)}{\Longrightarrow} \sum_{j=|k|,|k|+1,\ldots}^{\infty} \oplus R^j \tag{II.8}$$

where R^j are the irreducible representation spaces of SO(3).

The basis of eigenvectors of the system of commuting operators

$$\tilde{C}_1, \tilde{C}_2, \quad \mathbf{J}^2 = \tfrac{1}{2} \sum J_{ij} J_{ij}, \quad J_3 = J_{12} \tag{II.9}$$

is denoted $|kc\ jm\rangle$ and the matrix elements of the generators of SO(3,1)

$$\begin{aligned} J(0) &= J_{12}, \\ J(\pm 1) &= \mp \frac{1}{\sqrt{2}} (J_{23} \pm i J_{31}) \\ &= \mp \frac{1}{\sqrt{2}} (J_1 \pm i J_2), \\ K(0) &= J_{03} = K_3, \\ K(\pm 1) &= \mp \frac{1}{\sqrt{2}} (J_{01} \pm i J_{02}) \\ &= \mp \frac{1}{\sqrt{2}} (K_1 \pm i K_2) \end{aligned} \tag{II.10}$$

in this basis are given by[8]

$$J(q)|kcjm\rangle = \sum_{m'} |kcjm'\rangle \langle jm'|1q;jm\rangle J_j, \tag{II.11}$$

$$K(q)|kcjm\rangle = \sum_{j'm'} |kcj'm'\rangle \langle j'm'|1q;jm\rangle [2j+1]^{1/2} K_{j',j}(k,c),$$

where $\langle j'm'|1q;jm\rangle$ are the SO(3) Clebsch–Gordon coefficients [9]

$$J_j = -\sqrt{j(j+1)} \tag{II.12}$$

and

$$K_{j-1,j}(k,c) = -i \frac{[(j^2-k^2)(j^2-c^2)]^{1/2}}{[j(2j-1)(2j+1)]^{1/2}} \quad \text{for } j > |k|$$

$$= 0 \quad \text{for } j = |k|, \tag{II.13a}$$

$$K_{j,j}(k,c) = i \frac{kc}{[(2j+1)j(j+1)]^{1/2}} \quad \text{for } j > 0, \tag{II.13b}$$

$$K_{j+1,j}(k,c) = K_{j,j+1}(k,c). \tag{II.13c}$$

The class of representations of SO(3,2) that we are interested in (subclass I and IV of Ref. 6) are characterized by a fixed value of c. The eigenvalues of the Casimir operators are

$$D_1 = 2(1-c^2), \quad D_2 = c^2(1-c^2) \tag{II.14}$$

and the reduction of the irreducible representation space $H^{(c)}$ with respect to SO(3,1) is given by

$$H^{(c)} \underset{SO(3,1)}{\Longrightarrow} \sum_{k=0,\pm1,\pm2,\ldots}^{\pm\infty} \oplus H(k,c) \quad \text{for } c \neq 0, \tag{II.15}$$

$$H^{(c)} \underset{SO(3,1)}{\Longrightarrow} \sum_{k=\pm1/2,\pm3/2,\ldots}^{\pm\infty} \oplus H(k,c) \quad \text{for } c \neq 0 \tag{II.16}$$

and

$$H^{(0)} \underset{SO(3,1)}{\Longrightarrow} \sum_{k=0,1,2\ldots}^{\infty} \oplus H(k,c) \quad \text{for } c = 0. \tag{II.17}$$

The action of the SO(3,1)-vector operator Γ_μ, i.e.,

$$\Gamma_0, \quad \Gamma(0) = \Gamma_3, \quad \Gamma(\pm 1) = \mp (\Gamma_1 \pm i\Gamma_2), \tag{II.18}$$

is given by

$$\Gamma_0 |ckjm\rangle = \sum_{k'} |ck'jm\rangle \Gamma_j^{k'k}, \tag{II.19}$$

$$\Gamma(q)|ckjm\rangle = \sum_{k'j'm'} |ck'j'm'\rangle \langle j'm'|1q;jm\rangle [2j+1]^{1/2} \Gamma_{j'j}^{k'k} \tag{II.20}$$

where

$$\Gamma_j^{k\pm1,k} = \tfrac{1}{2} \sqrt{(j \mp k)(j \pm k + 1)} \quad \text{for } c \neq 0$$

and for $c = 0$, $k > 1$, (II.21)

$$\Gamma_j^{1,0} = \Gamma_j^{0,1} = \frac{1}{\sqrt{2}} \sqrt{j(j+1)} \tag{II.22}$$

and

$$\Gamma_{j'j}^{k'k} = i(K_{j'j}(k'c)\Gamma_j^{k'k} - \Gamma_{j'}^{k'k} K_{j'j}(k,c)) \tag{II.23}$$

B. Extension of the SO (3,2) representation by parity

The representation space H^c is not only a representation space of SO(3,2) but also a representation space of an extension of SO(3,2) by the parity P [and also a representation space of an extension of SO(3,2) by P and time inversion T]. However, there are other extensions of SO(3,2) by P (and P, T), which require doubling of the representation, and it appears that just these are the representations that are realized in physics. We shall therefore discuss the extension in more detail.

The relations between the operator U_P representing parity P and the operators representing the generators are

$$U_P J_{ij} U_P = J_{ij}, \quad U_P \Gamma_0 U_P = \Gamma_0,$$
$$U_P K_i U_P = -K_i, \quad U_P \Gamma_i U_P = -\Gamma_i \tag{II.24}$$

(U_P is chosen such that $U_P^2 = 1$).[10] The irrep spaces $\mathcal{H}(k,c)$ of SO(3,1) are not representation spaces of extensions of SO(3,1) by P, but one has[8]

$$U_P |ckjm\rangle = (-1)^j \eta_P |c - kjm\rangle \tag{II.25a}$$

or

$$U_P |ckjm\rangle = (-1)^j \eta_P |-ckjm\rangle \tag{II.25b}$$

where

$$\eta_P = +1 \quad \text{or} \quad \eta_P = -1.$$

[Note that $\mathcal{H}(-kc) = \mathcal{H}(k-c)$.[8]] (II.25) can easily be seen from the fact that as a consequence of (II.24)

$$U_P K_i J_i U_P = -K_i J_i, \tag{II.26}$$

so that according to (II.4) U_P has either to change the sign of K or the sign of c.[11] Therefore, an irrep space of the extended SO(3,1) is given by

$$\mathcal{H}(|k|c) = \mathcal{H}(k,c) \oplus \mathcal{H}(-k,c) \tag{II.27}$$

or by

$$\mathcal{H}(k|c|) = \mathcal{H}(k,c) \oplus \mathcal{H}(k-c) \tag{II.28}$$

unless k or c is equal to zero, in which case

$$\mathcal{H}(0c) \text{ and } \mathcal{H}(k0) \tag{II.29}$$

are already irreps of the extended group.

We shall first turn to the extension of SO(3,2) that does not require representation doubling. The $\mathcal{H}^{(c)}$ given by (II.15) is the representation space of the extension so that (II.25b) is impossible. Consequently, U_P is given by (II.25a). The parity eigenstates in the representation space $\mathcal{H}(c \neq 0)$ are then defined by

$$|c, |k|, jm\pm\rangle = \frac{1}{\sqrt{2}}(|c,k,j,m\rangle \pm |c,-k,j,m\rangle), \quad k > 0,$$
$$|c, |k| = 0 jm+\rangle = |k=0,j,m\rangle, \tag{II.30}$$

which have the property

$$P\,||k|jm\pm\rangle = \eta_P(-1)^j(\pm 1)\,||k|jm\pm\rangle,$$
$$P|0,jm\rangle = \eta_P(-1)^j|0,jm\rangle. \tag{II.31}$$

The action of the generators $\mathbf{K}$ and $\mathbf{\Gamma}$ on these parity eigenstates follows from (II.30), (II.11), and (II.20) and is given by

$$\begin{aligned} K(q)|c,|k|,jm\pm\rangle &= |c,|k|j-1m'\pm\rangle\langle j-1,m'|1qjm\rangle(2j+1)^{1/2}K_{j-1,j}(|k|) \\ &+ |c,|k|,j,m'\mp\rangle\langle j,m'|1qjm\rangle(2j+1)^{1/2}K_{jj}(|k|) \\ &+ |c,|k|,j+1,m'\pm\rangle\langle j+1,m'|1qjm\rangle(2j+1)^{1/2} \\ &\times K_{j+1,j}(|k|) \end{aligned} \tag{II.32}$$

where $K_{j'j}(|k|)$ are given by (II.13) [note that $K_{j\pm1,k}(|k|) = K_{j\pm1,k}(k)$]:

$$\Gamma_0|c,|k|,jm\pm\rangle = |c,|k-1|,jm\pm\rangle\Gamma_j^{|k-1|,|k|}$$
$$+ |c,|k+1|,jm\pm\rangle\Gamma_j^{|k+1|,|k|} \tag{II.33}$$

where

$$\Gamma^{|k\pm1||k|} = \Gamma^{k\pm1,k} = \Gamma^{-(k\pm1),-k} \tag{II.34}$$

and

$$\begin{aligned} \Gamma(q)|c,|k|jm\pm\rangle &= (|c,|k+1|,j+1,m'\pm\rangle\Gamma_{j+1j}^{|k+1||k|} + |c,|k-1|,j+1,m'\pm\rangle \\ &\times \Gamma_{j-1j}^{|k-1||k|})\langle j+1,m'|1qjm\rangle(2j+1)^{1/2} \\ &+ (|c,|k+1|,j,m'\mp\rangle\Gamma_{jj}^{|k+1||k|} \\ &+ |c,|k-1|,j,m'\mp\rangle\Gamma_{jj}^{|k-1||k|})\langle jm'|1qjm\rangle(2j+1)^{1/2} \\ &+ (|c|k+1|j-1,m'\pm\rangle\Gamma_{j-1,j}^{|k+1||k|} \\ &+ |c|k-1|,j-1,m'\pm\rangle\Gamma_{j-1j}^{|k-1||k|}) \\ &\times\langle j-1m'|1qjm\rangle(2j+1)^{1/2} \end{aligned} \tag{II.35}$$

where

$$\Gamma_{j'j}^{|k'||k|} = i[k_{j'j}(|k'|)\Gamma_j^{|k'||k|} - \Gamma_j^{|k'||k|}K_{j'j}(|k|)]. \tag{II.36}$$

We note from (II.32), (II.33) and (II.34) that $\mathbf{K}$ and $\mathbf{\Gamma}$ change the parity of the state, as is required of SO(3) vector operators.

$\mathbf{K}$ fulfills the selection rules

$$\Delta j = \pm 1, 0, \quad \Delta|k| = 0 \text{ for } k \neq 0 \tag{II.37a}$$

and

$$\Delta j = \pm 1, \quad \Delta|k| = 0 \quad \text{for } k = 0. \tag{II.37b}$$

(II.37b) can be seen from (II.32) since by (II.13b) $K_{jj}(|k| = 0) = 0$. $\mathbf{\Gamma}$ fulfills the selection rule

$$\Delta|k| = \pm 1, \quad \Delta j = 0, \pm 1. \tag{II.38}$$

Γ_0 fulfills the selection rules

$$\Delta k = \pm 1, \quad \Delta j = 0. \tag{II.39}$$

C. Extension with representation doubling

We now turn to the extension with representation doubling. Then the irrep space of the extension of SO(3,2) by P is

$$\mathcal{H}^D = \mathcal{H}^{(c)} \oplus \mathcal{H}^{(-c)}. \tag{II.40}$$

Without restricting generality we choose $ic \geq 0$. P is represented by

$$U_P|ckjm\rangle = (-1)^j\eta_P|-ckjm\rangle. \tag{II.25b}$$

The parity eigenstates are then given by

$$|Dkjm\pm\rangle = \frac{1}{\sqrt{2}}(|ckjm\rangle \pm |-ckjm\rangle) \tag{II.41}$$

which have the property

$$U_P|Dkjm\pm\rangle = \pm(-1)^j\eta_P|Dkjm\pm\rangle. \tag{II.42}$$

The matrix elements of $K(q)$, Γ_0, and $\Gamma(q)$ in these states are easily obtained from (II.11), (II.19), (II.20):

$$K(q)\,|Dkjm\pm\rangle = \sum_{j'm'} |D,k,j,m',(\pm)(-1)^{1-j+j'}\rangle\langle j'm'|1qjm\rangle \times(2j+1)^{1/2}K_{jj}(k,c), \tag{II.43}$$

$$\Gamma_0\,|Dkjm\pm\rangle = |D,k-1,j,m\pm\rangle\Gamma_j^{k-1k} + |D,k+1,j,m\pm\rangle\Gamma_j^{k+1k}, \tag{II.44}$$

$$\Gamma(q)\,|Dkjm\pm\rangle = \sum_{k'j'm'} |D,k',j',m',(\pm)(-1)^{1-j+j'}\rangle \times\langle j'm'|1qjm\rangle(2j+1)^{1/2}\Gamma_{j'j}^{k'k}(c) \tag{II.45}$$

where $K_{j'j}(k,c)$ is given by (II.13), $\Gamma_j^{k'k}$ is given by (II.21), (II.22), and

$$\Gamma_{j'j}^{k'k}(c) = i(K_{j'j}(k',c)\Gamma_j^{k'k} - \Gamma_{j'}^{k'k}K_{j'j}(k,c). \tag{II.46}$$

(II.43), (II.44), and (II.45) are completely analogous to (II.32), (II.33), (II.35); the distinction between the representation of the extended SO(3,2) given in $\mathcal{H}^{(c)}$ and $\mathcal{H}^D$ of (II.40) is that in $\mathcal{H}^D$ the subspace with a definite value of k is two-dimensional, i.e., the subspace with a definite value of k^2 or $|k|$ is four-dimensional, whereas in $\mathcal{H}^{(c)}$ the subspace with a definite value of k is one-dimensional, i.e., the subspace with a definite value of k^2 or $|k|$ is two-dimensional.

The selection rules of **K**, **Γ**, and Γ_0 in $\mathcal{H}^D$ are obtained from (II.43), (II.44), and (II.45) and are very similar to the selection rules in $\mathcal{H}^{(c)}$. **K** and also **A** fulfill the selection rules

$$\Delta j = 0,\pm1,\quad \Delta k = 0 \text{ for } k\neq 0, \qquad \Delta j = \pm1,\quad \Delta k = 0 \text{ for } k = 0. \tag{II.47}$$

Γ and also **B** fulfill the selection rules

$$\Delta j = 0,\pm1,\quad \Delta k = \pm1, \qquad \Delta j = \pm1,\quad \Delta k = \pm1 \text{ for } k = 0,\ j = 0. \tag{II.48}$$

Γ_0 fulfills the selection rules

$$\Delta j = 0,\quad \Delta k = \pm1. \tag{II.49}$$

Summarizing, we have described above the following irreps of the extended SO(3,2):

$\mathcal{H}^c$ which reduces with respect to irreps of the parity extended SO(3,1) in the following way:

$$\mathcal{H}^c = \sum_{|k|=0,1,2\cdots} \oplus\mathcal{H}(|k|,c)$$

$$\left(\text{or } \mathcal{H}^c = \sum_{|k|=0,1/2,3/2\cdots} \oplus\mathcal{H}(|k|,c)\right) \tag{II.50}$$

and $\mathcal{H}^D$ which reduces with respect to the irreps of the parity-extended SO(3,1) in the way

$$\mathcal{H}^D = \sum_{k=0,\pm1,\pm2,\pm3,\cdots} \oplus(\mathcal{H}(k,c)\oplus\mathcal{H}(k,-c)) \qquad \begin{pmatrix}\text{or the corresponding}\\ \text{for half integer } k\end{pmatrix}. \tag{II.51}$$

Each $\mathcal{H}(|k|,c)$ in (II.50) contains an infinite tower of parity and angular momentum multiplets j^π.

$$\mathcal{H}(|k|,c) = \sum_{j^\pi}\oplus\mathcal{R}(j^\pi)$$

where

$$j^\pi = |k|^\pm, (|k|+1)^\pm \text{ for } |k|\neq 0 \qquad \text{and } j^\pi = 0^+, 1^-, 2^+\cdots \text{ for } |k| = 0$$

and the arbitrary phase η_P (intrinsic parity) is chosen $+1$. Thus schematically $\mathcal{H}^c$ describes the j^π spectrum shown in Fig. 1.

Each $\mathcal{H}(kc)\oplus\mathcal{H}(k,-c)$ in (II.51) contains the following tower of parity and angular momentum multiplets:

$$\mathcal{H}(k,c)\oplus\mathcal{H}(k,-c) = \sum_{\substack{j=|k|,|k|+1\cdots\\ \pi=+,-}} \oplus\mathcal{R}(j^\pi). \tag{II.52}$$

Thus for every value of $k = 0, \pm1, \pm2\cdots$ there are two angular momentum multiplets with value $j = |k|, |k|$

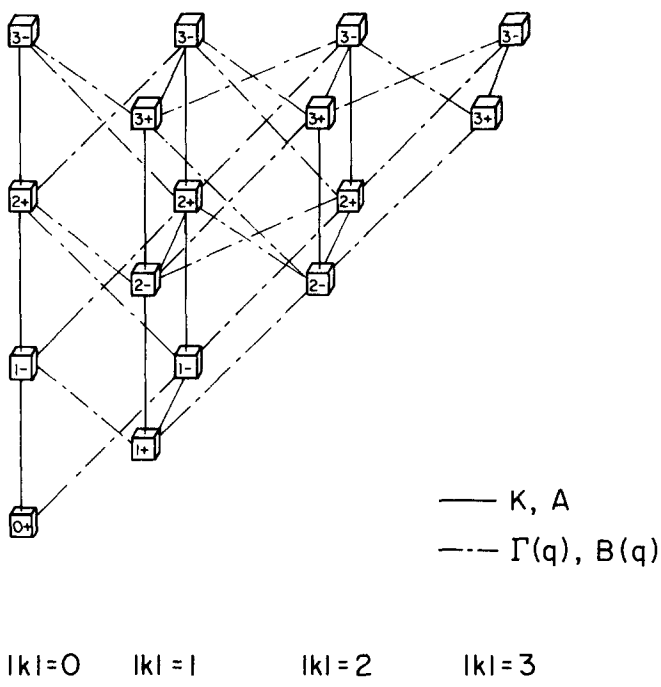


FIG. 1. The spectrum of $\mathcal{H}^c$, shown schematically.

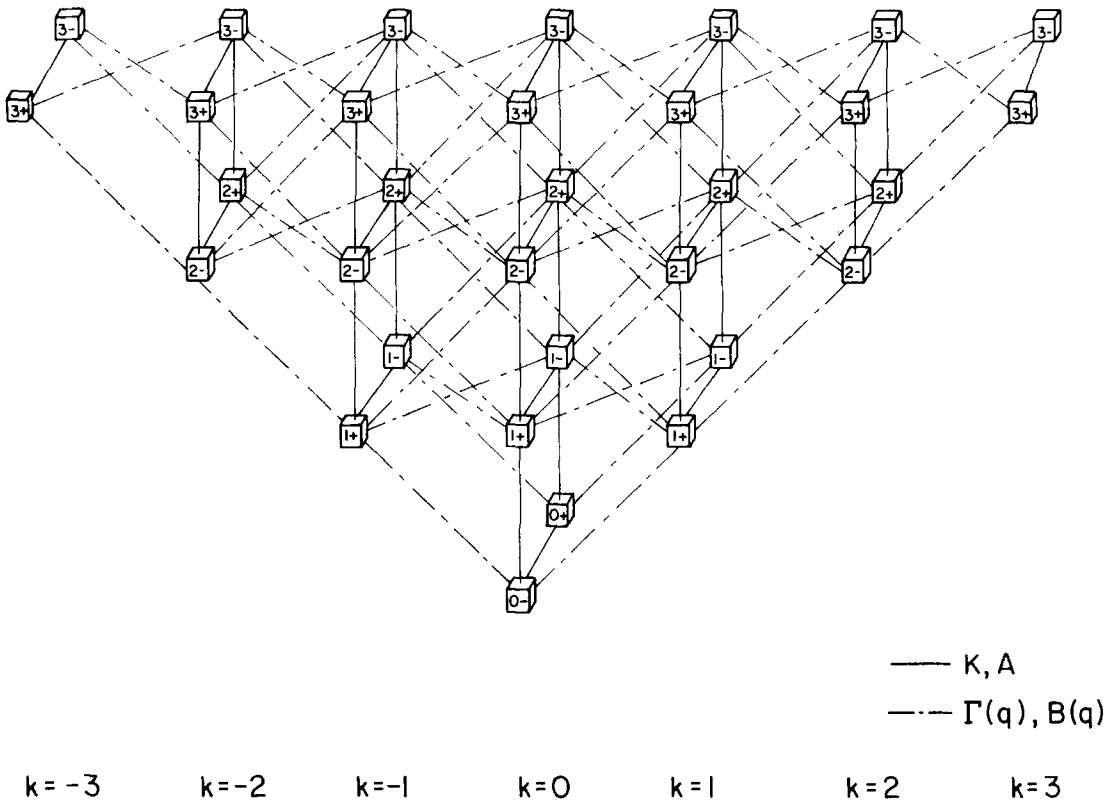


FIG. 2. The spectrum of H^D, shown schematically.

$+1\cdots$ and opposite parity. Or, for every value of $|k|$ there are two parity and angular momentum multiplets with the same value of j^π. Thus schematically H^D describes the j^π spectrum shown in Fig. 2.

The connecting lines between the multiplets j^π in Fig. 1 and Fig. 2 are an expression of the selection rules (II.37), (II.38), (II.39) and (II.47), (II.48), (II.49), respectively.

If one compares the diagram of Fig. 1 and Fig. 2 with the rotational energy levels of a symmetric top molecule in Fig. 3[13], one notices that H^c and H^D are prime candidates for the description of the quantum mechanical symmetric top. We will give a detailed comparison between the spectrum given by H^c, H^D and the rotational energy spectrum of the symmetrical top molecule in the next section.

The objection one may raise to the description of the symmetric top spectrum by the representation of the extended SO(3,2) is that the two SO(3)-vector operators **K** and **Γ** do not commute, whereas the observables that perform the transitions between the energy levels are believed to be represented by SO(3)-vector operators that do commute among themselves, like, e.g., the operator of the permanent dipole moment. We will, therefore, now obtain representations of a different group from the representations (c) and D of SO(3,2), which have the same j^π spectrum but for which one has instead of the two noncommuting vector operators **K** and **Γ** two vector operators whose components commute. These representations are obtained by group contraction.

D. Contraction of SO (3,2) representations with respect to SO (3)

To perform this contraction we introduce

$$A_i^{(\lambda)} = \lambda K_i \quad B_i^{(\lambda)} = \lambda \Gamma_i, \quad S^{(\lambda)} = \lambda \Gamma_0 \tag{II.53}$$

where λ is a positive real parameter, the contraction

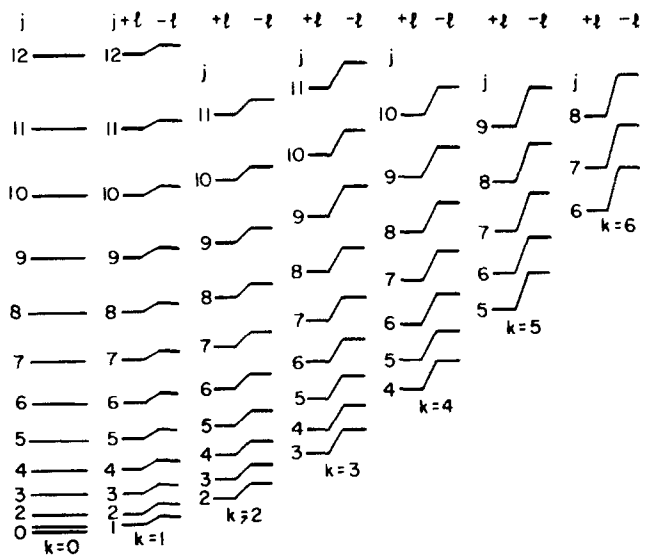


FIG. 3. Rotational energy levels of a symmetric top molecule.[13]

parameter. The contraction in the representation is then performed by the limiting process

$$\lambda \to 0, \quad |ic| \to \infty \text{ such that } ic\lambda \to \epsilon. \tag{II.54}$$

From the commutation relation (II. 1) it follows that

$$\begin{aligned}
&[J_i, J_j] = i\epsilon_{ijk}J_k,\\
&[A_i^{(\lambda)}, J_j] = i\epsilon_{ijk}A_k^{(\lambda)},\\
&[B_i^{(\lambda)}, J_j] = i\epsilon_{ijk}B_k^{(\lambda)},\\
&[S^{(\lambda)}, J_j] = 0,
\end{aligned} \tag{II.55a}$$

$$\begin{aligned}
&[A_i^{(\lambda)}, A_j^{(\lambda)}] = -i\epsilon_{ijk}\lambda^2 J_k,\\
&[B_i^{(\lambda)}, B_j^{(\lambda)}] = -i\epsilon_{ijk}\lambda^2 J_k,\\
&[A_i^{(\lambda)}, B_j^{(\lambda)}] = -ig_{ij}\lambda S^{(\lambda)},\\
&[S^{(\lambda)}, A_i^{(\lambda)}] = i\lambda B_i^{(\lambda)},\\
&[S^{(\lambda)}, B_i^{(\lambda)}] = i\lambda A_i^{(\lambda)}.
\end{aligned} \tag{II.55b}$$

The operators

$$A_i = \lim A_i^{(\lambda)}, \quad B_i = \lim B_i^{(\lambda)}, \quad S = \lim S^{(\lambda)}, \tag{II.56}$$

or in spherical components

$$A(q) = \lim \lambda K(q), \quad B(q) = \lim \lambda \Gamma(q), \tag{II.56'}$$

where the limit is understood in the sense of (II. 54), are for $\mathcal{H}^{(c)}$ obtained from (II. 32), (II. 33), (II. 35):

$$\begin{aligned}
&A(q)|\epsilon, |k|jm\pm\rangle\\
&= |\epsilon, |k|, j-1, m'\pm\rangle\langle j-1, m'|1qjm\rangle(2j+1)^{1/2}a_{j-1j}\\
&\quad \times (|k|, \epsilon)\\
&\quad + |\epsilon|k|j, m'\mp\rangle\langle jm'|1qjm\rangle(2j+1)^{1/2}a_{jj}(|k|\epsilon)\\
&\quad + |\epsilon|k|j+1, m'\pm\rangle\langle j+1, m'|1qjm\rangle(2j+1)^{1/2}\\
&\quad \times a_{j+1j}(|k|, \epsilon)
\end{aligned} \tag{II.57}$$

where

$$a_{j-1,j}(|k|\epsilon) = \lim_{\substack{\lambda|c\to\epsilon\\ \lambda\to 0\\ ic\to\infty}} \lambda K_{j-1j}(|k|, c) = -i\frac{[(j^2-k^2)\epsilon^2]^{1/2}}{[j(2j-1)(2j+1)]^{1/2}}, \tag{II.58}$$

$$a_{j+1,j}(|k|, \epsilon) = a_{j,j+1}(|k|, \epsilon),$$

$$a_{jj}(|k|, \epsilon) = \frac{|k|\epsilon}{[(2j+1)j(j+1)]^{1/2}},$$

$$S|\epsilon|k|jm\pm\rangle = 0, \tag{II.59}$$

$$\begin{aligned}
&B(q)|\epsilon|k|jm\pm\rangle\\
&= \sum_{k'} |\epsilon|k'|j+1, m'\pm\rangle b_{j+1j}^{|k'||k|}\langle j+1m'|1qjm\rangle(2j+1)^{1/2}\\
&\quad + \sum_{k'} |\epsilon|k'|j, m'\mp\rangle b_{jj}^{|k'||k|}\langle jm'|1qjm\rangle(2j+1)^{1/2}\\
&\quad + \sum_{k'} |\epsilon|k'|j-1m'\pm\rangle b_{j-1j}^{|k'||k|}\langle j-1, m'|1qjm\rangle(2j+1)^{1/2}
\end{aligned} \tag{II.60}$$

where

$$b_{j'j}^{|k'||k|}(\epsilon) = i(a_{j'j}(|k'|, \epsilon)\Gamma_j^{|k'||k|} - \Gamma_{j'}^{|k'||k|}a_{j'j}(|k|, \epsilon)). \tag{II.61}$$

Analogous expressions are obtained for the operators in $\mathcal{H}^D$ from (II. 43), (II. 44), (II. 45) [the vectors in (II. 57), (II. 59), and (II. 60) have to be replaced by $|D, k, j, m, \pm\rangle$, $a_{j'j}(|k|\epsilon)$ by $a_{j'j}(k, |\epsilon|)$, and $b_{j'j}^{|k'||k|}(\epsilon)$ by $b_{j'j}^{k'k}(|\epsilon|) = i(a_{j'j}(k', |\epsilon|)\Gamma_j^{j'k} - \Gamma_{j'}^{j'k}K_{j'j}(k, |\epsilon|)).$]

In this limit the commutation relations (II. 55) go over into

$$\begin{aligned}
&[A_i, J_j] = i\epsilon_{ijk}A_k, \quad [B_i, J_j] = i\epsilon_{ijk}B_k\\
&[J_i, J_j] = i\epsilon_{ijk}J_k,
\end{aligned} \tag{II.62a}$$

$$[A_i, A_j] = 0, \qquad [B_i, B_j] = 0, \tag{II.62b}$$

$$[A_i, B_j] = 0.$$

The operator S is according to (II. 59) represented by the zero operator. Thus in the limit (II. 54) the representation (c) of SO(3, 2) has gone into a representation (ϵ) of the semidirect product of SO(3) with the six-dimensional Abelian group which we will call $E^2(3)$ whose generators are defined by (II. 57), (II. 58), (II. 60), and (II. 61). The same is true for the representation of the extended SO(3, 2) in $\mathcal{H}^{(c)}$. The representation of the extended SO(3, 2) in $\mathcal{H}^D$ goes into the representation of the extended $E^2(3)$ in the space $\mathcal{H}^{D(|\epsilon|)} = \mathcal{H}^{(\epsilon)} \oplus \mathcal{H}^{(-\epsilon)}$. $SO(3,1)_{J_iK_i}$ is contracted into the representation (k, ϵ) of the three-dimensional Euclidean group E(3) generated by J_i and A_i. The eigenvalues of the Casimir operators

$$\begin{aligned}
&A_iA_i = -\lim \lambda^2 \tilde{C}_1 = \lim(A_iA_i - \lambda^2 J_iJ_i),\\
&A_iJ_i = \lim \lambda \tilde{C}_2 = \lim \lambda K_iJ_i
\end{aligned} \tag{II.63}$$

of $E(3)_{A_iJ_i}$ obtained from (II. 5) are found to be

$$\begin{aligned}
&\text{eigenvalue } A_iA_i = \epsilon^2,\\
&\text{eigenvalue } A_iJ_i = k\epsilon.
\end{aligned} \tag{II.64}$$

The reduction of the representation space $\mathcal{H}^{(\epsilon)}$ with respect to E(3) is obtained from (II. 15) to be

$$\mathcal{H}^{(a)} \underset{E(3)}{\Rightarrow} \sum_{k=0\pm1\pm2}^{\infty} \oplus \mathcal{H}(k, \epsilon). \tag{II.65}$$

The special case $\epsilon = 0$ is the case in which the A_i and B_i are represented by the zero operator.

The j^P spectrum of $\mathcal{H}^{(\epsilon)}$ of $E^2(3)$ is again given by Fig. 1 and the j^P spectrum of $\mathcal{H}^{D(|\epsilon|)}$ of $E^2(3)$ by Fig. 2. The essential difference between SO(3, 2) and $E^2(3)$ is that the matrix elements of the vector operators A_i and B_i and S are simpler than the matrix elements of K_i, Γ_i, and Γ_0 and in fact the matrix element of S is identically zero. The connecting lines between the multiplets j^P do not exist for $E^2(3)$.

The selection rules for **A** and **B** are the same as the selection rules for **K** and **Γ**, respectively, i.e., in $\mathcal{H}^{(\epsilon)}$ the selection rules for **A** are given by (II. 37), the selection rules for **B** are given by (II. 38), (II. 39), and in $\mathcal{H}^{D(|\epsilon|)}$ the selection rules for **A** are given by (II. 47) and for **B** by (II. 48).

III. SPECTRUM OF THE SYMMETRIC TOP

A. Quantum mechanical dumbbell

It has been known for some time[4] that the representation spaces

$\mathcal{H}(|k|, c)$ of SO(3, 1)

or

$$H(|k|\epsilon)=H(k,\epsilon)\oplus H(-k,\epsilon) \text{ of } E(3) \tag{III.1}$$

can serve as the image of the space of physical states of the rotator. The rotator is a quantum mechanical dumbbell that is realized in nature by diatomic molecules when their external conditions are such that they remain in one and the same vibrational and electronic state. The space

$$H(k=0,c \text{ or } \epsilon)=\sum_{j=0,1\cdots}\oplus R^j \tag{III.2}$$

describes the diatomic molecule whose moment of inertia about the line joining the nuclei is zero. The physical states

$$|k=0jm\rangle$$

are parity eigenstates

$$P|k=0jm\rangle=\eta_P(-1)^j|k=0jm\rangle \tag{III.3}$$

and the rotational levels have positive or negative parity if j is even or odd respectively (for $\eta_P=+1$) or vice versa (for $\eta_P=-1$). Each rotational level has no degeneracy [except for the $(2j+1)$-fold degeneracy in m] and the selection rules for the operator **A** and **K** are according to (II.37b)

$$\Delta j=\pm 1. \tag{III.4}$$

According to (II.57) and (II.7)

$$\mathbf{A}\circ\mathbf{J}=0 \text{ and } \mathbf{K}\cdot\mathbf{J}=0, \tag{III.5}$$

i.e., **J** is perpendicular to **A** or **K**.

There are numerous realizations of this model by diatomic molecules known.[14] The permanent dipole moment **d**, which in the classical picture is a vector that points into the direction of the axis of the dumbbell is in the quantum case assumed to be proportional to the vector operator **A**. With this assumption the algebra of observables is the algebra of $E(3)_{J_i A_i}$ and the space of physical states is the representation space $H(0,\epsilon)$. This view is supported by the idea that the dipole moment is proportional to the position operator Q_i whose components commute. However, one does not seem to arrive at any contradiction to experimental facts if one assumes that **d** is proportional to **K** and the space of physical states is the representation space $H(0c)$ of $SO(3,1)_{J_i K_i}$. The space

$$H(|k|,c \text{ or } \epsilon)=\sum_{j=|k|,|k|+1,\cdots}\oplus R^j(\pi=(-1)^j)\oplus R^j(\pi=(-1)^{j+1}) \tag{III.6}$$

is the space of quantum mechanical states of the physical system whose classical analog is the dumbbell carrying a flywheel on its axis. In this case **J** is not perpendicular to the axis. The component of angular momentum in the direction of the figure axis is given by (II.24) or (II.5):

$$\mathbf{A}\circ\mathbf{J}=k\epsilon\neq 0 \text{ or } \mathbf{K}\circ\mathbf{J}=ick\neq 0. \tag{III.7}$$

The physical states

$$||k|(c \text{ or } \epsilon)jm\pm\rangle$$

$$=\frac{1}{\sqrt{2}}(|k(c \text{ or } \epsilon)jm\rangle\pm|-k(c \text{ or } \epsilon)jm\rangle) \tag{III.8}$$

are parity eigenstates with parity, according to (II.31),

$$\pi=(\pm 1)(-1)^j\eta_P \tag{III.9}$$

and for each value of j there is a positive and negative rotational level [of almost equal energy (Λ-type doubling[14])]. The selection rules for the operators **A** and **K** are according to (II.37)

$$\Delta j=\pm 1,0 \text{ and } \Delta\pi=-1. \tag{III.10}$$

There are numerous realizations of this model, in particular by diatomic molecules in higher electronic states.[14] As the component of the angular momentum in the direction of the figure axis is due to the revolution of the electrons, this must be a constant in a particular electronic state. Therefore, k (for a fixed value of ϵ or ic) is constant and in a particular electronic state the diatomic molecule is a rotor described by $H(|k|)$ of (III.6) or (III.2). However, one and the same molecule can be and is in fact a rotor described by one $H(|k_1|)$ in one electronic state and another $H(|k_2|)$ in another electronic state.[14]

B. Quantum mechanical symmetric top

The dumbbell carrying a flywheel is already an example of a symmetric top; however, this is a symmetric top for which the component of angular momentum along the symmetry axis is a constant (in a particular electronic state). In general, the component of angular momentum along the symmetry axis of a symmetric top can have different values. Consequently, in the space of physical states of the quantum mechanical symmetric top $\mathbf{A}\circ\mathbf{J}$ (or $\mathbf{K}\cdot\mathbf{J}$) should not be an invariant but should have a nontrivial spectrum, as is the case for the representation space $H^{D(c)}$ of SO(3,2) and for the representation space of its contraction $H^{D(\epsilon)}$. To show that H^D plays the same role for the symmetric top as $H(|k|)$ plays for the rotator, we will now make the detailed comparison between the symmetrical top spectrum and the spectrum given by H^D.

A classical nonrelativistic symmetric top[15] is a physical system which has constant moments of inertia I_A, I_B, I_B about its principle axes. It has angular velocity ω, angular momentum **J**, and the axis A of symmetry pointing in the direction of the vector **a**. Its energy is given by

$$E=\tfrac{1}{2}\mathbf{j}\cdot\omega.$$

ω may be expressed by **j** and **a** using[15]

$$\mathbf{j}=I_B\omega+\tilde{\eta}(\mathbf{a}\circ\mathbf{j})\mathbf{a} \tag{III.11}$$

where

$$\tilde{\eta}=\frac{I_A-I_B}{I_A}\frac{1}{|\mathbf{a}|^2}=\eta\frac{1}{|\mathbf{a}|^2},\quad \eta=\frac{I_A-I_B}{I_A}, \tag{III.12}$$

so that the energy is given in terms of **j** and **a** by

$$E=\frac{1}{2I_B}(\mathbf{j}^2-\tilde{\eta}(\mathbf{a}\cdot\mathbf{j})^2) \tag{III.13}$$

According to the basic assumptions of quantum mechanics, a quantum mechanical physical system is de-

scribed by an algebra of linear operators which is generated by some basic physical quantities whose multiplication is defined by some algebraic relations. For a quantum mechanical system with a classical counterpart, the algebraic relations between the observables are conjectured from the relation between the corresponding classical quantities. Consequently, for the quantum mechanical symmetric top the energy operator is given by (III. 13) where the classical observables j_i and a_i are replaced by the linear operators J_i, A_i. I_A, I_B are constant parameters for this physical system; however, for an extended physical system for which the moments of inertia cannot be considered constant (e.g. for a vibrating rotator) I_A and I_B may be operators with a nontrivial spectrum, too. The commutation relations between the J_i are easily conjectured to be

$$[J_i, J_j] = \hbar i \epsilon_{ijk} J_k, \tag{III. 14}$$

whereas the commutation relations of the A_i may require some contemplation. Since in the classical case a_i is a vector, one would assume that for the quantum mechanical top A_i is a vector operator, i.e.,

$$[A_i, J_j] = \hbar i \epsilon_{ijk} A_k. \tag{III. 15}$$

Even if one assumes simplicity and requires that the commutator of A_i and A_j does not lead out of the Lie algebra generated by J_i and A_i, one is left with the three possibilities

$$[A_i, A_j] = \kappa \hbar i \epsilon_{ijk} J_k \tag{III. 16}$$

with $\kappa = \pm 1, 0$. The only "physical" argument one can give against case $\kappa = 1$ [i.e., the case in which the algebra of J_i and A_i is the enveloping algebra of SO(4)] is that then there will be only a finite number of angular momentum values j.[16] One is therefore left with the two possible commutation relations between the A_i:

$$[A_i, A_j] = 0, \tag{III. 17}$$

$$[A_i, A_j] = -\hbar i \epsilon_{ijk} J_k. \tag{III. 18}$$

In case (III. 17) the A_i and J_i generate the enveloping algebra of the group $E(3)$, i.e., $\mathcal{E}(\mathrm{E}(3))$. In case (III. 18) they generate the algebra $\mathcal{E}(\mathrm{SO}(3,1))$.

We can now proceed to make the correspondence between the group representation spaces of Sec. II and the space of physical states of the symmetric top. In case (III. 17) the candidates for the symmetric top spaces are $\mathcal{H}^{(\epsilon)}$ and $\mathcal{H}^{D(|\epsilon|)}$, in case (III. 18) the candidates are $\mathcal{H}^{(c)}$ and $\mathcal{H}^{D}$. Choosing the units $\hbar = 1$ we identify the observable J_i with the generator J_i of Sec. II and, in case (III. 17), the observable A_i with the generator A_i. In case (III. 18) we identify the observable A_i with the generator K_i of Sec. II.

The energy operator H of the quantum mechanical symmetric top is obtained from the classical expression (III. 13) by correspondence. In the case (III. 17) this correspondence is straightforward, because $\mathbf{A}^2$ is an invariant with eigenvalue ϵ^2 [cf. (II. 64)]. Therefore, $\tilde{\eta}$ is an invariant,

$$\tilde{\eta} = \frac{I_A - I_B}{I_A} \frac{1}{\epsilon^2},$$

and the energy operator is

$$H = \frac{1}{2I_B}(\mathbf{J}^2 - \tilde{\eta}(\mathbf{A}\cdot\mathbf{J})^2). \tag{III. 19}$$

Let us first consider the space $\mathcal{H}^{(\epsilon)}$ for which the j^P spectrum is given in Fig. 1. Taking the eigenvalue of the operator H in the states (II. 30) one obtains for the energy spectrum

$$E(|k|, j) = \frac{1}{2I_B}(j+1) - \eta k^2), \tag{III. 20}$$

$$j = |k|, |k|+1, |k|+2 \cdots, \quad |k| = 0, 1, 2 \cdots \quad \text{or} \quad |k| = \tfrac{1}{2}, \tfrac{3}{2} \cdots .$$

As for each $|k|$ except $|k| = 0$ there are two angular momentum multiplets with opposite parity, there is a $2(2j+1)$ degeneracy of every energy value $E(|k|, j)$, whereas for $E(|k| = 0, j)$ there is only the usual $(2j+1)$ degeneracy. Thus taking $\mathrm{E}^2(3)$ as the spectrum generating group one predicts in the representation space $\mathcal{H}^{(\epsilon)}$ the spectra shown in Fig. 4.

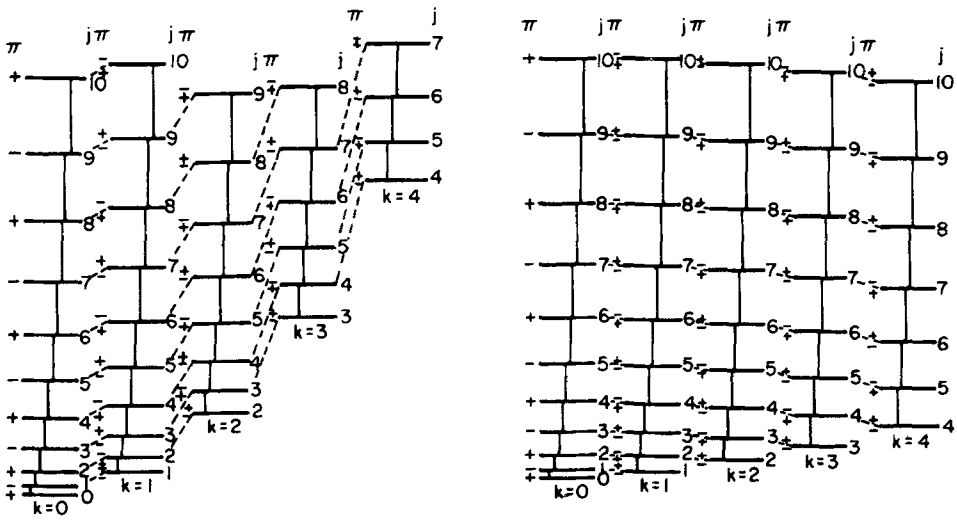


FIG. 4. Energy spectrum given by Eq. (III. 20): (a), $\eta < 0$; (b), $\eta > 0$.

Figure 4(a) shows the spectrum (III.20) for $\eta < 0$, i.e., $I_A < I_B$ (prolate symmetric top) and Fig. 4(b) shows the spectrum (III.20) for $\eta > 0$, i.e., $I_A > I_B$ (oblate symmetric top).

In the space $\mathcal{H}^{D(\epsilon)}$, for which the j^P spectrum is given in Fig. 2, we obtain by applying (III.19) to the states (II.41)

$$E(k,j) = \frac{1}{2I_B}(j(j+1) - \eta k^2). \tag{III.21}$$

These are the same energy values as in (III.20). However, now each energy value $E(k,j)$ with $k \neq 0$ has the degeneracy $4(2j+1)$ and $E(0,j)$ has the degeneracy $2(2j+1)$, or there is a $2(2j+1)$ degeneracy for every value of k. Thus the energy spectrum predicted by the spectrum generating group $E^2(3)$ in $\mathcal{H}^{D(\epsilon)}$ is that shown in Fig. 5 (for the case of a prolate symmetric top).

The spectra in Fig. 4 and Fig. 5 are indeed the energy spectra of the (accidentally) symmetric top molecule. Both spaces $\mathcal{H}^{\epsilon}$ and $\mathcal{H}^{D(|\epsilon|)}$ describe the inversion doubling due to the two possible eigenvalues π of the parity operator (except for $k=0$). The space $\mathcal{H}^{D(|\epsilon|)}$ also describes the l-type doubling,[5] due to the two possible values of sign (k). As the connection between the inversion doubling and the parity operator is clear, we will discuss only the l-type doubling in Sec. IV.

In all cases observed so far, only $\Delta k = 0$ transitions have been observed in the pure rotation spectra (these are indicated by the vertical lines in Fig. 4 and Fig. 5) indicating according to (II.47) that the permanent dipole moment is parallel to the symmetry axis, i.e.,

$\mathbf{M} \sim \mathbf{A}$.

In general, for accidentally symmetric tops the permanent dipole moment may also have a component perpendicular to the symmetry axis:

$$\mathbf{M} = \alpha\mathbf{A} + \beta\mathbf{B}, \tag{III.22}$$

which according to (II.48) can give rise to $\Delta k = \pm 1$ transitions.

So far $\Delta k = \pm 1$ transitions have only been observed in the infrared vibration–rotation spectra, where one has parallel-type bands fulfilling the selection rules

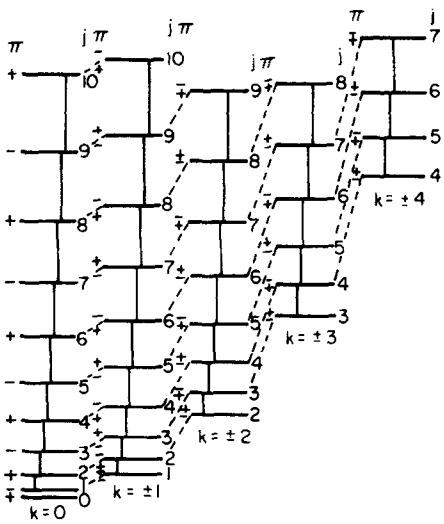


FIG. 5. Energy spectrum given by Eq. (III.21).

(II.47) and perpendicular-type bands fulfilling the selection rules (II.48). We will discuss the combination of rotational and vibrational degrees of freedom only very briefly in Sec. IV when we describe the Coriolis interaction since the primary subjects of this paper are the rotational properties of the symmetric top.

C. A generalization of the quantum mechanical symmetric top

Though the algebra of $E^2(3)$ appears the natural choice for the symmetrical top, as the commuting vector operators **A** and **B** are the quantities which one would associate with observables representing directions in a physical object, we will now establish the consequences of the assumption that the direction of the top axis is represented by the noncommuting operator **K** and the algebra of observables is given by the enveloping algebra of SO(3,2).

In this case, (III.18), the correspondence between the classical expression (III.13) for the energy and the energy operator is not as straightforward as in case (III.17). Substituting in (III.13) for the classical observables j_i, a_i the operators J_i, F_i, respectively, but assuming that $\tilde{\eta}$ is a parameter, will lead to the same spectrum (III.20) with (III.21) and (III.22) as in the case (III.17). Therefore a real generalization is obtained, if one also assumes that $\tilde{\eta}$ is an operator which is obtained from (III.12) by replacing the number $|\mathbf{a}|^2$ by the operator $\mathbf{K}^2$. One obtains for the energy operator

$$E = \frac{1}{2I_B}(\mathbf{J}^2 - \eta(\mathbf{K}^2)^{-1}(\mathbf{K}\circ\mathbf{J})^2). \tag{III.23}$$

The energy eigenvalues for the states (II.30) in the space $\mathcal{H}^{(c)}$ as well as for the states (II.41) in the space $\mathcal{H}^{(D)}$ are obtained from (II.7) and (II.6a):

$$E(k,j) = \frac{1}{2I_B}\left(j(j+1) - \eta\frac{(-c^2)k^2}{(j(j+1) - k^2 - c^2 + 1)}\right) \tag{III.24}$$

where $-c^2$ is a positive integer that characterizes the representation, and therewith the particular physical system. (III.24) may be written

$$E(k,j) = \frac{1}{2I_B}(j(j+1) - \eta k^2) + (\eta/2I_B)k^2\frac{j(j+1)+1-k^2}{j(j+1)+1-k^2-c^2} \tag{III.25}$$

i.e., the energy spectrum differs from the usual symmetric top spectrum by a correction factor which becomes smaller as $(-c^2)$ becomes larger.

The degeneracy in the spaces $\mathcal{H}^{(c)}$ and $\mathcal{H}^{(D)}$ is the same as the degeneracy in $\mathcal{H}^{(\epsilon)}$ and $\mathcal{H}^{D(|\epsilon|)}$, respectively. The selection rules for dipole transitions are also not changed.

IV. *l*-TYPE DOUBLING, CORIOLIS FORCES

Only under very limited circumstances can the symmetrical top molecule, or any other physical top, be considered a perfectly rigid body. In general, when higher energies are involved, one has to consider also the fact that the physical top consists of parts (vibrating nuclei in the molecule, orbiting electrons in the molecule, extra-core particles in the nucleus) which may

move relative to a body-fixed reference frame. We choose as an example for these internal degrees of freedom the vibrations. The symmetrical top molecule is then a combination of the two elementary physical systems oscillator and rotator and, according to the fundamental assumptions of quantum mechanics, its space of physical states is described by the direct product space

$$\mathcal{H}^{\rm osc}\times\mathcal{H}^{\rm rot} \tag{IV.1}$$

where $\mathcal{H}^{\rm osc}$ is the space of states of the oscillator and $\mathcal{H}^{\rm rot}$ is the space of states of the rotator which is given by $\mathcal{H}^{\epsilon}$ or $\mathcal{H}^{D(|\epsilon|)}$. As is well known, the space $\mathcal{H}^{\rm osc}$ is also a representation space of a spectrum generating group, e.g., SU(3,1). However, in this paper we are interested in the rotations, and we do not want to discuss the oscillator spectrum here; we will make use of $\mathcal{H}^{\rm osc}$ only as much as is necessary to discuss the interactions between the oscillations and rotations.

The energy operator of the vibrating and rotating symmetric top molecule is

$$H_{VR}=H^{\rm osc}\times 1+1\times H^{\rm rot}+H'.$$

H' is the rotation—oscillation interaction term caused by the Coriolis force, which gives rise to the l-type doubling and which we want to discuss now. We ignore all other corrections, like centrifugal forces.

We will obtain the quantum mechanical expression for the Coriolis term H' by correspondence from classical mechanics, so we will have to use a few facts from classical mechanics[17] which we shall briefly review now.

Let us assume that the parts of the symmetric top (vibrating nuclei) move with a velocity $\mathbf{v}$ relative to a coordinate system K that is fixed in the top, i.e., that rotates with constant angular velocity ω with respect to the rest system K_0, with which it has the origin in common. The angular momentum of these moving objects $\boldsymbol{l}$ has the same value with respect to both systems K and K_0. The intrinsic energy of the top $E^{(i)}$ in the system K is related to its value $E_0^{(i)}$ in the system K_0 by

$$E_0^{(i)}=E^{(i)}+\boldsymbol{l}\cdot\boldsymbol{\omega}. \tag{IV.2}$$

The total energy is the sum of the rotational energy (caused by the total angular momentum) plus the internal energy measured in the rest system K_0, i.e.,

$$E=E^{(i)}+E^{\rm rot}+\boldsymbol{l}\cdot\boldsymbol{\omega}. \tag{IV.3}$$

ω is related to $\mathbf{j}$ and $\mathbf{a}$ by (III.11) and $\boldsymbol{l}\circ\boldsymbol{\omega}$ is the Coriolis interaction term between the rotation and intrinsic motion (vibration). In general this is rather complicated. We will restrict ourselves to the case that the internal angular momentum $\boldsymbol{l}$ is parallel to the symmetry axis. (This condition is, e.g., always fulfilled in a symmetric top with internal angular momentum originating from the twofold degenerate oscillations. These oscillations must, because of symmetry, lie in the plane perpendicular to $\mathbf{a}$.)

This condition is formulated as the requirement that

$$\mathbf{a}\cdot\boldsymbol{l}=\text{extremum (minimum or maximum)} \tag{IV.4}$$

for a given value of $\boldsymbol{l}$ and $\mathbf{a}$. Then

$$\boldsymbol{l}\cdot\boldsymbol{\omega}=\boldsymbol{l}\cdot\boldsymbol{\omega}_A$$

where ω_A is the component of angular velocity along $\mathbf{a}$ which is easily seen to be given by

$$\boldsymbol{\omega}_A=(\boldsymbol{\omega}\cdot\mathbf{a}^0)\mathbf{a}^0=\frac{1}{I_A}(\mathbf{a}^0\cdot\mathbf{j})\mathbf{a}^0=\frac{1}{I_A}\frac{1}{\mathbf{a}^2}(\mathbf{a}\cdot\mathbf{j})\mathbf{a} \tag{IV.5}$$

where $\mathbf{a}^0$ is the unit vector in the direction $\mathbf{a}$.

Inserting this into (IV.3) we obtain

$$E=E^{(i)}+E^{(\rm rot)}+\frac{1}{I_A}\frac{1}{\mathbf{a}^2}(\mathbf{a}\cdot\mathbf{j})(\mathbf{a}\cdot\boldsymbol{l}). \tag{IV.6}$$

The energy operator of the corresponding quantum mechanical system (i.e., the quantum mechanical vibrating and rotating symmetric top in which all corrections are ignored except the Coriolis force) is obtained from (IV.6) by replacing the classical observables by operators.

Thus

$$H_{VR}=H^{(\rm osc)}+H^{(\rm rot)}+\frac{1}{I_A}\frac{1}{\mathbf{A}^2}(\mathbf{A}\circ\mathbf{J})(\mathbf{A}\circ\mathbf{L}) \tag{IV.7}$$

where L_i is the operator of "vibrational angular momentum" and J_i is the operator of "rotational angular momentum." L_i and J_i are physical quantities only if one of them is zero. E.g., if L_i is zero, as was assumed in the preceding sections, then J_i is the angular momentum of the physical system. The physical quantity is the total angular momentum of the physical system:

$$I_i=L_i\times 1+1\times J_i. \tag{IV.8}$$

Whereas $H^{(\rm osc)}=H^{(\rm osc)}\times 1$ acts nontrivially only in $\mathcal{H}^{\rm osc}$ and $H^{(\rm rot)}=1\times H^{\rm rot}$ acts nontrivially only in $\mathcal{H}^{\rm rot}$, the interaction term H' has nontrivial factors in both spaces $\mathcal{H}^{\rm osc}$ and $\mathcal{H}^{\rm rot}$. With the rotational energy given by (III.19), (IV.7) is written as

$$H_{VR}=H^{(\rm osc)}+\frac{1}{2I_B}\left(\mathbf{J}^2-\eta\frac{1}{\mathbf{A}^2}(\mathbf{A}\cdot\mathbf{J})^2\right)+\frac{1}{I_A}\frac{1}{\mathbf{A}^2}(\mathbf{A}\circ\mathbf{J})(\mathbf{A}\circ\mathbf{L}). \tag{IV.9}$$

The requirement (IV.4) that the internal angular momentum be parallel to the symmetry axis is formulated as

$$\langle|\mathbf{A}\circ\mathbf{L}|\rangle=\text{extremum (maximum or minimum)}, \tag{IV.10}$$

i.e., as a consequence of the fact that (for the molecules under consideration) the oscillations lie in a plane perpendicular to the symmetry axis, only those states $|\,\rangle$ exist for which the subsidiary condition (IV.9) is fulfilled. (We are very grateful to Herbert M. Pickett for advising us of the particular role that the vibrational angular momentum has in the symmetric top molecules).

We will ignore as much as possible the details of the vibrations but we need to discuss briefly the $\mathcal{H}^{\rm osc}$. We assume that we are in an energy subspace of the oscillator space with degenerate energy value $E^{\rm osc}$. We will later use (IV.9) to obtain the doubly degenerate vibrational level, but will presently discuss the more general case of any degree of degeneracy for which the angular momentum l, l_3 of the nonrotating vibrator suffices to label the different states with the same value $E^{\rm osc}$. In

this subspace $\mathcal{H}(E^{\rm osc})$ it is natural to introduce the angular momentum basis[18]

$$|m_{13}ll_3\rangle \tag{IV.11a}$$

where

$$E^{\rm osc}=(m_{13}+3/2)\omega$$

and the spectrum of l for a given value of m_{13} is

$$l=m_{13},\ m_{13}-2,\ m_{13}-4\cdots 0 \quad \text{for } m_{13} \text{ even}, \tag{IV.11b}$$
$$l=m_{13},m_{13}-2\cdots 1 \quad \text{for } m_{13} \text{ odd}.$$

m_{13} and l_3 are connected with the occupation number labels for the spherical oscillations n_0, n_+, n_- by

$$m_{13}=n_0+n_++n_-, \tag{IV.11c}$$
$$l_3=n_--n_+.$$

For the rotations we make use of our results in Secs. II and III.

We first choose for $\mathcal{H}^{\rm rot}=\mathcal{H}^{D(|\epsilon|)}$ and later discuss the choice $\mathcal{H}^{(\epsilon)}$. The basis in $\mathcal{H}^{D(|\epsilon|)}$ is [cf. (II.41)]

$$|Dkjm\pm\rangle=\frac{1}{\sqrt{2}}\left(|\epsilon kjm\rangle\pm|-\epsilon kjm\rangle\right), \tag{IV.12}$$

which according to section II has the property

$$J_3|Dkjm\pm\rangle=m|Dkjm\pm\rangle, \tag{IV.13}$$
$$\mathbf{J}^2|Dkjm\pm\rangle=j(j+1)|Dkjm\pm\rangle,$$
$$\mathbf{A}\cdot\mathbf{J}|Dkjm\pm\rangle=k\epsilon|Dkjm\mp\rangle, \tag{IV.14}$$
$$P|Dkjm\pm\rangle=(-1)^j(\pm 1)\eta_P|Dkjm\pm\rangle, \tag{IV.15}$$
$$\begin{aligned}A(q)|Dkjm\pm\rangle=&|Dkj-1m'\pm\rangle\langle j-1,m'|1qjm\rangle(2j+1)^{1/2}\\&\times a_{j-1j}(k,\epsilon)\\&+|Dkjm'\mp\rangle\langle jm'|1qjm\rangle(2j+1)^{1/2}a_{jj}(k,\epsilon)\\&+|Dkj+1m'\pm\rangle\langle j+1m'|1qjm\rangle(2j+1)^{1/2}\\&\times a_{j+1j}(k,\epsilon)\end{aligned} \tag{IV.16}$$

where

$$a_{jj}(k,\epsilon)=\frac{k\epsilon}{\sqrt{(2j+1)j(j+1)}},\quad a_{j-1j}=\frac{-i\sqrt{(j^2-k^2)\epsilon^2}}{\sqrt{j(2j-1)(2j+1)}}. \tag{IV.17}$$

A basis in the direct product space

$$\mathcal{H}^{\rm osc}\times\mathcal{H}^{\rm rot}$$

is given by the direct product of the basis (IV.11) and (IV.12):

$$|E_{\rm osc}\cdots l_3\rangle\otimes|Dkjm\pm\rangle. \tag{IV.18}$$

This is not a basis of eigenvectors of the total energy operator H_{VR}; furthermore, it is not even a basis of operators that commute with H_{VR}, as the "rotational angular momentum" $J_3,\mathbf{J}^2$ does not commute with H'. The basis vectors (IV.18) are the direct product of "internal angular momentum" and "rotational angular momentum" vectors and are therefore not even approximately physical states. The physical quantity is the total angular momentum of the physical system, and physical states must be total angular momentum states, i.e., eigenvectors of $\mathbf{I}^2$. The total angular momentum states are obtained from (IV.18) by[20]

$$\begin{aligned}&|E_{\rm osc}\cdots D,k,l,j,I,M,\pm\rangle\\&=\sum_{l_3m}|E_{\rm osc}\cdots ll_3\rangle\otimes|D,k,jm\pm\rangle\langle ll_3jm|IM\rangle.\end{aligned} \tag{IV.19}$$

The expectation value of the energy operator (IV.9) in the basis (IV.19) is then

$$\begin{aligned}E=&E^{\rm osc}+\frac{1}{2I_B}(j(j+1)-\eta k^2)\\&+\frac{1}{I_A}\frac{1}{\epsilon^2}k\epsilon\langle\pm MIkjl|L_iA_i|ljkIM\mp\rangle.\end{aligned} \tag{IV.20}$$

L_i is a vector operator with respect to the angular momentum l. A_i is a vector operator with respect to the angular momentum j. Therefore the above matrix element is given by[21]

$$\begin{aligned}&\langle\pm MIkjl|L_iA_i|ljkIM\mp\rangle\\&=(-1)^{l+j+I}\begin{Bmatrix}l&j&I\\j&l&1\end{Bmatrix}\langle l\|L\|l\rangle\langle j\|A\|j\rangle\end{aligned} \tag{IV.21}$$

where $\begin{Bmatrix}ijI\\jil\end{Bmatrix}$ are the 6-j symbols and

$$\langle j\|A\|j\rangle \text{ and } \langle l\|L\|l\rangle$$

are the reduced matrix elements of $\mathbf{A}$ and $\mathbf{L}$, respectively, and

$$\langle l\|L\|l\rangle=\sqrt{(2l+1)l(l+1)} \tag{IV.22}$$

because L_i is also the generator.

Inserting (IV.21), (IV.22) into (IV.20), one obtains

$$\begin{aligned}E=&E^{\rm osc}+\frac{1}{2I_B}(j(j+1)-\eta k^2)\\&+\frac{1}{I_A}\frac{\langle j\|A\|j\rangle}{\epsilon}k(-1)^{l+j+I}\begin{Bmatrix}l&j&I\\j&l&1\end{Bmatrix}\sqrt{(2l+1)l(l+1)}.\end{aligned} \tag{IV.23}$$

For given values of l and j, I can take the values $I=j+l,\ j+l-1,\cdots,j-l$. Thus the energy value

$$E(k,j)=E^{\rm osc}+\frac{1}{2I_B}(j(j+1)-\eta k^2) \tag{IV.24}$$

should split as a consequence of the Coriolis force into $2l+1$ sublevels. However, as a consequence of the subsidary condition (IV.10) not all states (IV.19) occur, but only those with the highest positive and negative value of the matrix element in (IV.21). These are the states with $I=j+l$ and $I=j-l$, as we shall see below. Thus $E(k,j)$ will as a consequence of the Coriolis interaction split into two sublevels (for molecules that have only oscillations in the plane perpendicular to the axis). Inserting the values for the 6-j symbols

$$\begin{Bmatrix}l&j&I\\j&l&1\end{Bmatrix}=(-1)^{I+j+l+1}\frac{[l(l+1)+j(j+1)-I(I+1)]}{2[l(l+1)(2l+1)j(2j+1)(j+1)]^{1/2}} \tag{IV.25}$$

into (IV.23), we obtain

$$\begin{aligned}E=&E(k,j)\\&-k\frac{1}{I_A}\frac{1}{\epsilon}\frac{\langle j\|A\|j\rangle}{\sqrt{j(j+1)(2j+1)}}\tfrac{1}{2}[l(l+1)+j(j+1)-I(I+1)].\end{aligned} \tag{IV.26}$$

The last term takes its extrema for $I=j+l$ ($j \uparrow\uparrow l$) and $I=j-l$ ($j \uparrow\downarrow l$) [or $I=J+l_{max}$ and $I=j-l_{max}$ for a degenerate vibrational state with more than one value of l; cf. (IV.11b)].

Thus for the two possible states,

$$E=E(k,j)+k\frac{1}{I_A}\frac{1}{\epsilon}\frac{\langle j\|A\|j\rangle}{\sqrt{j(j+1)(2j+1)}}\times\begin{cases} lj & \text{for } j\uparrow\uparrow l \\ -l(j+1) & \text{for } j\uparrow\downarrow l. \end{cases} \tag{IV.27}$$

This expression looks very similar to the well-known formula for the energy values with Coriolis splitting (e.g., Eq. IV.42 of Vol. II, Ref. 5).

However, as a consequence of our assumed group property of the vector operator **A**, we can calculate the reduced matrix element $\langle j\|A\|j\rangle$. From (IV.16) and (IV.17) it follows that

$$\begin{aligned}\langle j\|A\|j\rangle &= (-1)(2j+1)a_{jj} \\ &= (-1)k\epsilon\sqrt{(2j+1)/j(j+1)}.\end{aligned} \tag{IV.28}$$

Inserting this into (IV.27), we obtain

$$E=E(k,j)-\frac{1}{I_A}\frac{k^2}{j(j+1)}\times\begin{cases} lj & \text{for } j\uparrow\uparrow l, \\ -l(j+1) & \text{for } j\uparrow\downarrow l. \end{cases} \tag{IV.29}$$

(IV.29) is a surprising result. It agrees with the well-known formula for the Coriolis correction term (including the sign)

$$\mp 2A_{[\text{vibl}]}\zeta_i|k| \quad \text{for} \quad \begin{matrix} j\uparrow\uparrow l \\ j\uparrow\downarrow l \end{matrix}$$

where $A_{[\text{vibl}]}$ and ζ_i may depend upon the vibration, except for the dependence upon the additional factor of approximately k/j, i.e., only for the lowest value of j for a given k does (IV.29) agree with the semiclassical results of Teller and Tissa[19] and Johnston and Dennison.[19]

The origin of this additional factor is clear; it comes from our additional assumption that A_i be a generator of the spectrum generating group. The more general assumption that A_i be a vector operator would only result in an undetermined reduced matrix element in place of the $a_{jj}(k,\epsilon)$ given by (IV.17) and therewith in an undetermined factor on the rhs of (IV.29) instead of our factor proportional to k.

As we have ignored all the details of the oscillator properties, the coefficients that multiply the rotator quantum numbers in (IV.29) may change when all these details are taken into account; in particular, I_A will become dependent upon the vibrational quantum numbers. However, the dependence upon the rotational quantum numbers, in particular the k^2 dependence of $\langle|H'|\rangle$, should not be affected by these oscillator properties.

To illustrate our considerations above we consider the particular case of the first excited vibrational state $m_{13}=1$. Then according to (IV.11b), $l=1$ and

$$E=E(kj)-\frac{1}{I_A}k\begin{cases}\dfrac{k}{j+1} & \text{for } j\uparrow\uparrow l \\ \dfrac{k}{j} & \text{for } j\uparrow\downarrow l\end{cases}. \tag{IV.30}$$

The energy spectrum (IV.29) is schematically depicted in Fig. 3 (the distances between the levels are not drawn to scale).

We will now use for the space of states of the rotator the space $\mathcal{H}^{(\epsilon)}$. Then the basis in $\mathcal{H}^{\text{rot}}=\mathcal{H}^{(\epsilon)}$ is

$$|\epsilon,|k|jm\pm\rangle=\frac{1}{\sqrt{2}}[|\epsilon,k,j,m\rangle\pm|\epsilon,-k,j,m\rangle] \tag{IV.12'}$$

of (II.57). Instead of (IV.17) one has (II.58) and instead of (IV.14) one has

$$\mathbf{A}\circ\mathbf{J}|\epsilon,|k|,j,m,\pm\rangle=(|k|\epsilon)|\epsilon,|k|,j,m,\mp\rangle. \tag{IV.14'}$$

The expectation value of the energy is then again given by (IV.29). Thus both extensions by parity, $\mathcal{H}^{D}$ and $\mathcal{H}^{(\epsilon)}$, give the same energy values; only their degeneracy is different.

The calculations in this section had been undertaken in order to investigate whether the Coriolis force determines which of the spaces $\mathcal{H}^{D(|\epsilon|)}$ and $\mathcal{H}^{(\epsilon)}$ of the spectrum generating group is to be chosen for the space of physical states $\mathcal{H}^{\text{rot}}$. As also the correction term gives the same energy spectrum in $\mathcal{H}^{D(|\epsilon|)}$ and $\mathcal{H}^{(\epsilon)}$, and also the Coriolis splitting can be described by the breaking of the degeneracy in the oscillator part of the physical state vectors, the $\pm k$ degeneracy in $\mathcal{H}^{D(|\epsilon|)}$ is not necessary to describe the splitting, and the space $\mathcal{H}^{(\epsilon)}$ is sufficient to serve as the space of physical states for the rotating quantum mechanical symmetric top.

APPENDIX A

In this appendix we establish the connection between the group theoretical description of this paper and the wavefunction for the symmetric top.

The position of a rigid body is completely described by the three Euler angles (ϕ,θ,ψ) (in addition to the three coordinates of the center of mass, for which we take the origin of the coordinate system and which we can ignore here). The Euler angles are defined such that two of the angles (θ,ϕ) are the polar angles of an axis, for which we take the axis of symmetry in a symmetric top, i.e., the direction of **a**; the third angle ψ describes the rotation around **a**, i.e., it may be the angle between a reference point in the body and the axis that is obtained from the x-axis by a rotation described by the two Euler angles (θ,ϕ). The rotation ρ^{-1} that transforms the direction described by the (unit) vector **a** into the 3-axis is obviously given (see Fig. 6) by

$$\rho^{-1}=R_3^{-1}(-\phi)R_2(-\theta)R_3(-\phi) \tag{A1}$$

i.e.,

$$\rho=R_3(\phi)R_2(\theta)R_3^{-1}(\phi)] \tag{A1'}$$

is the rotation that transforms a unit vector in the direction of the 3-axis into **a**

$$\rho\begin{bmatrix}0\\0\\\epsilon\end{bmatrix}=\begin{bmatrix}a_1\\a_2\\a_3\end{bmatrix}=\epsilon\begin{bmatrix}\sin\theta\cos\phi\\\sin\theta\sin\phi\\\cos\theta\end{bmatrix}. \tag{A2}$$

ρ depends upon **a**, i.e., (ϵ,θ,ϕ). The representation of the rotation R in the space of physical states of the

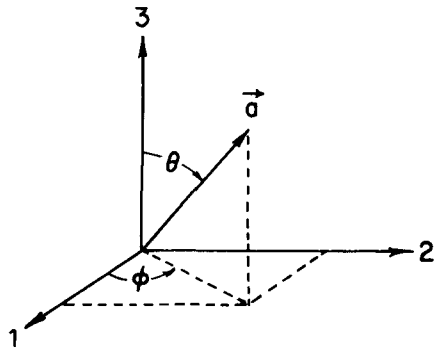


FIG. 6. The Euler angles (θ, ϕ).

quantum mechanical system we call T_R. In the space of quantum mechanical states of the quantum mechanical symmetric top $\mathcal{H}^{D(|\epsilon|)}$ (or $\mathcal{H}^{(\epsilon)}$) we have so far considered the basis system

$$|\epsilon kjm\rangle \text{ (or linear combinations of them)} \tag{A3}$$

of eigenvectors of

$$\mathbf{A}^2, \mathbf{J}\cdot\mathbf{A}, \mathbf{J}^2, J_3 \tag{A4}$$

with eigenvalues

$$\epsilon^2(=1) \quad \epsilon k \quad j(j+1) \quad m, \tag{A5}$$

respectively. (They describe states of definite angular momentum component m, angular momentum j, angular momentum component along the symmetry axis k.) We will now consider generalized eigenvectors

$$|\mathbf{a}\zeta\rangle = |\epsilon, \theta, \phi, \zeta\rangle \tag{A6}$$

of the system of commuting operators

$$A_i \quad \mathbf{J}\cdot\mathbf{A}. \tag{A7}$$

[Note that the K_i, $\mathbf{J}\cdot\mathbf{K}$ do not form a system of commuting operators so that for SO(3,1), SO(3,2) the discussions of this section do not apply.]

$$A_i|\mathbf{a}, \zeta\rangle = a_i|\mathbf{a}, \zeta\rangle, \tag{A8}$$

$$\mathbf{J}\cdot\mathbf{A}|\mathbf{a}, \zeta\rangle = \zeta|\mathbf{a}, \zeta\rangle \tag{A8'}$$

where $\zeta = k\epsilon$ is the eigenvalue of the same operator $\mathbf{A}\cdot\mathbf{J}$. These would describe states of the quantum mechanical top in which the axis A has the sharply defined direction $\mathbf{a}$ and the angular momentum component along this direction has the value ζ. However, these states do not exist physically. One can also consider a generalized basis system

$$|\mathbf{a}, \psi\rangle = |\epsilon, \phi, \theta, \psi\rangle \tag{A9}$$

with sharply defined orientation (position) of the body (which also would describe states which do not exist physically). The transition matrix element between the basis (A3) and the generalized basis (A9) gives the wavefunction; the basis (A6) is usually not considered.

For the normalization of the generalized eigenvectors (A6) and (A9) we choose a suitable δ-function normalization specified below.

We will now study the relations between the basis systems (A3), (A6), and (A9).

We can expand each vector $|\epsilon kjm\rangle$ with respect to the generalized basis system (A6):

$$\begin{aligned}|\epsilon kjm\rangle &= \sum_\zeta \int d^3\mathbf{a}\,|a\zeta\rangle\langle\zeta a|\epsilon kjm\rangle \\ &= \sum_\zeta \int d\epsilon' \epsilon'^2 d\cos\theta d\phi\,|\epsilon', \theta, \phi, \zeta\rangle\langle\zeta, \phi, \theta, \epsilon'|\epsilon kjm\rangle \\ &= \int \epsilon^2 d\cos\theta d\phi\,|\epsilon\theta\phi\,\zeta = \epsilon\circ k\rangle\langle\epsilon k, \phi, k, \epsilon|\epsilon kjm\rangle.\end{aligned} \tag{A10}$$

The volume element here has been chosen in accordance with the "normalization" of the generalized eigenvectors given by

$$\begin{aligned}\langle\mathbf{a}'\zeta'|\zeta\mathbf{a}\rangle &= \delta_{\zeta',\zeta}\delta^3(\mathbf{a}'-\mathbf{a}) \\ &= \delta_{\zeta',\zeta}\delta(\epsilon'-\epsilon)\frac{1}{\epsilon^2}\delta(\cos\theta'-\cos\theta)\delta(\phi'-\phi).\end{aligned} \tag{A11}$$

We want to determine the transition coefficients $\langle\epsilon k, \phi, \theta, \epsilon|\epsilon kjm\rangle$. In order to do this we define the vector

$$\psi_{jmk\epsilon} = \int d\mu(\hat\rho)\bar D^j_{mk}(\hat\rho)T_{\hat\rho}|\epsilon, 0, 0, \zeta = k\circ\epsilon\rangle \tag{A12}$$

where

$$|\epsilon, 0, 0, \zeta\rangle = T_\rho^{-1}|\epsilon, \theta, \phi, \zeta\rangle, \tag{A13}$$

$$\hat\rho = \rho R_3(\gamma) = R_3(\phi)R_2(\theta)R_3(\gamma-\phi) \tag{A14}$$

is a rotation, $D^j_{mk}(\hat\rho) = D^j_{mk}(\phi, \theta, \gamma-\phi) = \langle mj|T_{\hat\rho}|jk\rangle$ is the rotation matrix,[21] $\bar D$ its complex conjugate, $d\mu(\hat\rho) = (1/8\pi^2)\sin\theta d\theta d\phi d(\gamma-\phi)$ is the invariant measure of the rotation group and the integration in (A12) is taken over the whole group volume.

It is easy to see that

$$\mathbf{A}\cdot\mathbf{J}\psi_{jmk\epsilon} = k\epsilon\psi_{jmk\epsilon}, \tag{A15}$$

$$\mathbf{A}^2\psi_{jmk\epsilon} = \epsilon^2\psi_{jmk\epsilon}, \tag{A16}$$

and (using the invariance of the measure)[22]

$$T_R\psi_{jmk\epsilon} = \sum_{m'}\psi_{jm'k\epsilon}D^j_{m'm}(R). \tag{A17}$$

From (A15), (A16), (A17) it follows that $\psi_{jmk\epsilon}$ is an eigenvector of the complete commuting system (A4) as is the vector $|\epsilon kjm\rangle$. Consequently,

$$|\epsilon kjm\rangle = F\psi_{jmk\epsilon} \tag{A18}$$

where F is a normalization factor, which is to be chosen such that

$$\langle m'j'k'\epsilon|\epsilon kjm\rangle = \delta_{kk'}\delta_{jj'}\delta_{mm'}.$$

Thus by (A18) and (A12) with (A14), (A1') we have

$$\begin{aligned}&|\epsilon kjm\rangle \\ &= \frac{F}{8\pi^2}\int d\gamma d\phi\, d\cos\theta\bar D^j_{mk}(\rho R_3(\gamma))T_\rho T_{R_3}(\gamma)|\epsilon, 0, 0, \zeta\rangle \\ &= \frac{F}{8\pi^2}\sum_{m'}\int d\gamma\, d\cos\theta\, d\phi\bar D^j_{mm'}(\rho)D^j_{km'}(R_3(-\gamma))T_\rho \\ &\quad\times\exp(-i\gamma J_3)|\epsilon, 0, 0, \zeta\rangle.\end{aligned} \tag{A19}$$

Here we have used $T_{R_3(\gamma)} = \exp(-i\gamma J_3)$ and

$$\bar D^j_{m',k}(R_3(\gamma)) = D^j_{km'}(R_3(\gamma))^\dagger = D^j_{km'}(R_3(-\gamma)) = \delta_{km'}\exp(i\gamma m). \tag{A20}$$

Because of (A8)

$$\mathbf{J}\cdot\mathbf{A}\,|\epsilon,0,0,\zeta\rangle = J_3\circ\epsilon\,|\epsilon,0,0,\zeta\rangle = \zeta\,|\epsilon,0,0,\zeta\rangle,$$

we have with (A8′)

$$J_3\,|\epsilon,0,0,\zeta\rangle = k\,|\epsilon,0,0,\zeta\rangle, \tag{A8″}$$

i.e., k is the component of J_3 where the axis $\mathbf{a}$ is in the direction of the 3-axis. With (A8″) and (A20) we obtain for (A19) after summation over m' and integration over γ

$$|\epsilon kjm\rangle = \frac{F}{4\pi}\int d\cos\theta\, d\phi\, \bar{D}^j_{mk}(\rho)T_\rho\,|\epsilon,0,0,\zeta\rangle$$

$$= \frac{F}{4\pi}\int d\cos\theta\, d\phi\, \bar{D}^j_{mk}(\rho)\,|\epsilon,\theta,\phi,\mathcal{J}\rangle. \tag{A21}$$

Comparing this with (A10) we obtain the transition coefficients

$$\langle\zeta,\phi,\theta,\epsilon\,|\epsilon,k,j,m\rangle = \frac{F}{4\pi}\frac{1}{\epsilon^2}\bar{D}^j_{mk}(\phi,\theta,-\phi)\delta_{\zeta,\epsilon k}. \tag{A22}$$

To calculate the normalization factor F we form the norm of (A21) and use (A11):

$$\langle m'j'k\epsilon\,|\epsilon kjm\rangle =$$

$$= \frac{F^2}{(4\pi)^2}\frac{1}{\epsilon^2}\int d\cos\theta\, d\phi\, D^{j'}_{m'k}(\phi,\theta,-\phi)\bar{D}^j_{mk}(\phi,\theta,-\phi). \tag{A23}$$

In order to write this in a form that allows us to use the well-known orthogonality properties of the rotation matrices[23]

$$\int d\mu(\hat{\rho})D^{j'}_{m'k}(\hat{\rho})\bar{D}^j_{mk}(\hat{\rho}) = \frac{1}{2j+1}\,\delta_{jj'}\delta_{m'm}, \tag{A24}$$

we use

$$D^{j'}_{m'k}(\rho R_3(\gamma)) = D^{j'}_{m'k}(\rho)\exp(-i\gamma k),$$
$$\bar{D}^j_{mk}(\rho R_3(\gamma)) = \bar{D}^j_{m'k}(\rho)\exp(i\gamma k).$$

Then (A23) becomes

$$\langle m'j'k\epsilon\,|\epsilon kjm\rangle = \frac{F^2}{(4\pi)^2}\frac{1}{\epsilon^2}\frac{1}{2\pi}\int d\gamma\, d\cos\theta\, d\phi\, D^{j'}_{m'k}(\hat{\rho})\bar{D}^j_{mk}(\hat{\rho})$$

which by (A24) gives

$$= \frac{F^2}{4\pi}\frac{1}{\epsilon^2}\frac{1}{2j+1}\,\delta_{j'j}\delta_{m'm}.$$

As on the other hand, the $|\epsilon kjm\rangle$ are assumed to be normalized to 1, we obtain for F (up to an undeterminable phase factor)

$$F = \sqrt{4\pi}\,\sqrt{2j+1}\,|\epsilon|$$

and we obtain from (A22)

$$\langle\zeta,\phi,\theta,\epsilon\,|\epsilon,k,j,m\rangle = \sqrt{\frac{2j+1}{4\pi}}\,\frac{1}{\epsilon}\,\bar{D}^j_{mk}(\phi,\theta,-\phi). \tag{A25}$$

The transition coefficients between the basis (A9) and (A3) are now easily obtained and have in fact already been obtained in the intermediate steps of the previous calculations.

ψ is the angle that describes the rotation $R(\mathbf{n},\psi)$ of the body around the axis $\mathbf{n}=\mathbf{a}/|\mathbf{a}|$ from a reference position $\psi=\psi_0$:

$$T_{R(\mathbf{a},\psi)}\,|\mathbf{a},\psi_0\rangle = |\mathbf{a},\psi_0+\psi\rangle. \tag{A26}$$

It is simplest to consider the state in which $\mathbf{a}$ is along the 3-axis $|\epsilon 00\psi\rangle$. Then

$$\langle\zeta,00\epsilon\,|T_{R_3(\psi)}\,|\epsilon 00\psi_0\rangle = \langle\zeta,00\epsilon\,|\epsilon 00\psi+\psi_0\rangle$$
$$= \langle\zeta,00\epsilon|\,\exp(-i\psi J_3)\,|\epsilon,00\psi_0\rangle \tag{A27}$$
$$= \exp(i\psi k)\langle\zeta,00\epsilon\,|\epsilon 00\psi_0\rangle.$$

Thus the expansion of the vector $|\epsilon,00\zeta\rangle$ with respect to the vector $|\epsilon,00\psi\rangle$ is given by

$$|\epsilon,00\zeta\rangle = \int d\psi'\,|\epsilon,0,0\psi'\rangle\langle\psi'00\epsilon\,|\epsilon 00\zeta\rangle$$
$$= \langle\zeta,0,0,\epsilon\,|\epsilon,0,0,\psi_0\rangle\int d\psi'\,|\epsilon,0,0,\psi'\rangle\exp(i\psi' k), \tag{A28}$$

$$k = \zeta/\epsilon,$$

where the normalization factor depends upon the normalization of $|\epsilon 00\zeta\rangle$ and $|\epsilon 00\psi\rangle$. As

$$\langle\zeta'00\epsilon\,|\epsilon 00\zeta\rangle = \delta_{\zeta'\zeta}$$

and we choose

$$\langle\psi'00\epsilon\,|\epsilon 00\psi\rangle = \delta(\psi'-\psi), \tag{A29}$$

we obtain for the normalization factor (except for a phase)

$$|\langle\zeta 00\epsilon\,|\epsilon 00\psi_0\rangle| = 1/\sqrt{2\pi}. \tag{A30}$$

We insert (A28) into (A21)

$$|\epsilon kjm\rangle = \frac{\sqrt{2j+1}}{\sqrt{8\pi^2}}\,|\epsilon|\int d\cos\theta\, d\phi\, d\psi\,\bar{D}^j_{mk}(\phi,\theta,-\phi)$$
$$\times\exp(i\psi k)T_\rho\,|\epsilon,00\psi\rangle$$
$$= \frac{\sqrt{2j+1}}{\sqrt{8\pi^2}}\,|\epsilon|\int d\cos\theta\, d\phi\, d\psi\,\bar{D}^j_{mk}(\phi,\theta,\psi-\phi)\,|\epsilon,\phi,\theta,\psi\rangle \tag{A31}$$

as

$$T_\rho\,|\epsilon,0,0,\psi\rangle = |\epsilon,\phi,\theta,\psi\rangle. \tag{A32}$$

Changing the integration variable

$$\psi-\phi\to\psi$$

and choosing the reference position suitably, we obtain

$$|\epsilon kjm\rangle = \sqrt{\frac{2j+1}{8\pi^2}}\,|\epsilon|\int d\cos\theta\, d\phi\, d\psi\,\bar{D}^j_{mk}(\phi,\theta,\psi)\,|\epsilon,\phi,\theta,\psi\rangle. \tag{A33}$$

From this we obtain with the normalization

$$\langle\psi'\theta'\phi'\epsilon\,|\epsilon,\phi,\theta,\psi\rangle$$
$$= \frac{1}{\epsilon^2}\,\delta(\cos\theta-\cos\theta')\delta(\phi'-\phi)\delta(\psi'-\psi), \tag{A29′}$$

$$\langle\psi'\theta'\phi'\epsilon'\,|\epsilon\phi\theta\psi\rangle$$
$$= \frac{1}{\epsilon^2}\,\delta(\epsilon'-\epsilon)\delta(\cos\theta-\cos\theta')\delta(\phi'-\phi)\delta(\psi'-\psi), \tag{A29″}$$

$$\langle\psi,\theta,\phi,\epsilon\,|\epsilon,k,j,m\rangle =$$

$$= \sqrt{\frac{2j+1}{8\pi^2}} \frac{1}{|\epsilon|} \bar{D}^j_{mk}(\phi, \theta, \psi). \tag{A34}$$

APPENDIX B

We define the vector operator

$$C_i = \epsilon_{ijk} A_j B_k$$

where A_j, B_k are the generators in (II. 62). From the commutation relations (II. 62) one can calculate the commutator

$$[A_i J_i, B_j J_j] = -iC_i J_i.$$

The operators **AJ, BJ, CJ** are, except for normalization factors, the components of angular momentum relative to the molecular fixed axes which are the quantities commonly used. Their commutation relations have a negative sign as compared to the usual commutation relations of angular momentum. There is no physical distinction between the operators **B, C** or any other vectors orthogonal to **A** (except in cases of additional symmetry properties).

*Supported by the U. S. Atomic Energy Commission Contract No. AT(40-1)3992.

[1]The name spectrum generating algebra was introduced in Y. Dothan, M. Gell-Mann, and Y. Neeman, Phys. Rev. Lett. **17**, 145 (1965). The same concept under different names (dynamical group and noninvariance group, respectively) was first used by A. O. Barut and A. Bohm, Phys. Rev. **139**, B1107 (1965) and N. Mukunda, L. O'Raifeartaigh, and E. C. G. Sudarshan, Phys. Rev. Lett. **15**, 1041 (1965). For a general discussion, see Y. Dothan, Phys. Rev. D **2**, 2944 (1970).

[2]A. O. Barut, P. Budini, and C. Fronsdal, Proc. Roy. Soc. Lond. A **291**, 106 (1966); A. Bohm, Nuovo Cimento 43, 667 (1966); N. Mukunda, L. O'Raifeartaigh, and E. C. G. Sudarshan, Phys. Rev. Lett. **15**, 1041 (1965).

[3]A. O. Barut, Phys. Rev. **139**, 1433 (1965).

[4]A. O. Barut and A. Bohm, Phys. Rev. **139**, 1107 (1965).

[5]G. Herzberg, *Molecular Spectra and Molecular Structure, Vol. II, Infrared and Raman Structure of Polyatomic Molecules* (Van Nostrand Reinhold, New York, 1950).

[6]L. Jaffe, J. Math. Phys. **12**, 882 (1971).

[7]J. B. Ehrman, Proc. Camb. Phil. Soc. **53**, 290 (1957).

[8]M. A. Naimark, *Linear Representation of the Lorentz Group* (Pergamon, New York, 1964); Gel'fand, Minlos, Shapiro, *Representations of the Rotation and Lorentz Groups* (Pergamon, New York).

[9]With the usual phase convention: A. R. Edmonds, *Angular Momentum in Quantum Mechanics* (Princeton U. P., 1960); M. E. Rose, *Elementary Theory of Angular Momentum* (Wiley, New York, 1957); $\langle j, m \mid j_1, m_1, j_2, m\rangle = C(j_1 j_2 j, m_1 m_2 m)$.

[10]E. P. Wigner, in *Group Theoretical Concepts and Methods in Elementary Particle Physics*, edited by F. Gursey (Gordon and Breach, New York, 1964).

[11]The appearance of the factor $(-1)^j$ is well known and the reality of η_p follows from the unitarity of U_p.

[12]According to E. Inonu and E. P. Wigner, Proc. Natl. Acad. Sci. USA **39** (1953), any Lie group can be contracted with respect to any of its continuous subgroups into a semidirect product of S with an Abelian invariant subgroup. However, not every representation of a Lie group can be contracted into a faithful representation with respect to any subgroup. From the properties of $D(C)$, in particular (III. 21), (III. 23) and (III. 30), (III. 33), it can be seen easily that a contraction with respect to $SO(3,1)_{J_i K_i}$, i. e., a contraction of the Γ_μ leaving the K_i unaffected is not possible. We will, therefore, contract $D(c)$ of SO(3, 2) with respect to SO(3) along the lines described in A. O. Barut and A. Bohm, Phys. Rev. **139**, B1107 (1965) for the contraction of the representations of SO(3, 1). This process is so perspicuous that no knowledge of group contraction is required to follow it.

[13]Fig. 117 of Ref. 5, Vol. II.

[14]Herzberg, Ref. 5, Vol. I.

[15]H. C. Corbin and P. Stehle, *Classical Mechanics* (Wiley, New York, 1950), Chap. I.

[16]I. e., there will be a well-defined highest value of angular momentum ($j_{max} = ic - 1$) for a particular physical system. Although, in physics, there is always an upper limit to the observability of any observable, such a highest value is usually not well defined and one prefers to explain this limitation by the breakdown of the model, i. e., one says that a particular physical object is a rotator only for sufficiently low values of angular momentum.

[17]See, e. g., L. D. Landau and E. M. Lifshitz, *Mechanics* (Addison-Wesley, Reading, Mass., 1960), Vol. I, Sec. 39.

[18]See, e. g., J. D. Louck, J. Math. Phys. **6**, 1786 (1965); V. Bargmann and M. Moshinsky, Nucl. Phys. **23**, 127 (1961).

[19]E. Teller and L. Tisza, Z. Phys. **73**, 791 (1932); M. Johnston and D. M. Dennison, Phys. Rev. **48**, 868 (1935).

[20]There are two ways of obtaining total angular momentum states, by coupling (jk) (ll_3) or by coupling (jm) (ll_3). The first corresponds to the rotating frame K in which $\vec{AJ}$, $\vec{BJ}$, $\vec{CJ}$ are angular momentum components (c. f., Appendix B), the second to the rest frame K_0 which we have used as reference frame.

[21]E. g., K. Gottfried, *Quantum Mechanics* (Benjamin, New York, 1966), Sec. 36. 4. We use here the same normalization according to which the reduced matrix element is defined by

$$\langle \cdots jm' \mid A(q) \mid jm \cdots \rangle = \frac{\langle j \| A \| j \rangle}{\sqrt{2j+1}} \langle j, m, 1, q \mid jm' \rangle.$$

[22]Let R be an arbitrary rotation, then

$$T_R \psi_{jj_3 k_0 \epsilon} = \int d\mu(\hat{p}) \bar{D}^j_{j_3 k_0}(\hat{p}) T_{R\hat{p}} \mid 0, 0, \epsilon, k_0 \rangle.$$

Calling $R_{\hat{p}} = \sigma$ and using the invariance of the measure $d\mu(\sigma) = d\mu(\hat{p})$, we calculate

$$T_R \psi_{jj_3 k_0 \epsilon} = \sum_m \int d\mu(\sigma) \bar{D}^j_{j_3 m}(R^{-1}) \bar{D}^j_{mk_0}(\sigma) T_\sigma \mid 0, 0, \epsilon, k_0 \rangle$$
$$= \sum_m D^j_{mj_3}(R) \int d\mu(\sigma) \bar{D}^j_{mk_0}(\sigma) T_\sigma \mid 0, 0, \epsilon, k_0 \rangle.$$

Here we have used the representation property of D:

$$\bar{D}^j_{j_3 m}(R^{-1}) = (D^j_{mj_3}(R^{-1}))^* = ((D^j_{mj_3}(R))^{-1})^* = D^j_{mj_3}(R).$$

(D^* denotes the adjoint matrix.) Inserting (A12) into this we obtain (A17).

[23]E. g., M. E. Rose, *Elementary Theory of Angular Momentum* (Wiley, New York, 1957).

Goldstone's theorem for quantum spin systems of finite range

Walter F. Wreszinski

Institute of Physics, University of São Paulo, São Paulo, Brazil
(Received 11 August 1975)

Goldstone's theorem is proved for quantum spin systems of finite range with rotationally invariant Hamiltonian under the assumptions that the ground state of the infinite system is unique, is invariant under a subset of the translation group (which depends on the system), and has long range order.

In this paper we prove that the spectrum of the physical Hamiltonian of a quantum spin system of finite range with rotationally invariant (finite-region) Hamiltonian has no gap, under three assumptions on the ground state of the infinite system. This result, which is a form of Goldstone's theorem (whose analog in quantum field theory is Ref. 1, to which we also refer for additional and original references), was proved by Streater in Ref. 2 for ferromagnetic Heisenberg systems of finite range. As Streater, we use the results of Ref. 1, but have to generalize his Theorem 7, which relies upon structure specific to the ferromagnet, in particular upon the fact that the physical Hilbert space is the direct sum of dynamically independent "n-magnon sectors," and the restriction of the physical Hamiltonian to each of them is a bounded operator [see especially Ref. 2, Eq. (23)].

In a recent paper, Reeh[3] proved by explicit construction the existence of zero energy ("Goldstone") states in the spectrum of any rotationally invariant antiferromagnetic Hamiltonian, under a number of assumptions on the ground state. Although he did not prove Goldstone's theorem, we shall see that just three of his four assumptions (unicity, translation invariance, and long range order) suffice to provide a proof: His assumption of invariance of the ground state under rotations around the z-axis will not be needed. Furthermore, his proof as given, seems to be restricted to one-dimension while ours holds for any number of dimensions. (Reeh used the fact that the number of terms in the commutator $[H_{\Lambda_1}, \mathbf{S}_\Lambda]$, which is considered in part 1, tends to a finite constant as $\Lambda \to \infty$ (see Part 1 for this notation), which is true only in one-dimension.)

For clarity, we divide the forthcoming proofs into two parts: The results of Part 1 are independent of the above mentioned assumptions on the ground state (which will be more precisely formulated in Part 2), while Part 2 contains the results which explicity depend on them.

1. RESULTS INDEPENDENT OF ASSUMPTIONS ON THE GROUND STATE

We consider a quantum spin system in $\mathbf{Z}^\nu$ in the sense of Refs. 2 and 4. Specifically, to each finite region $\Lambda \subset \mathbf{Z}^\nu$ we associate a Hilbert space $\mathcal{H}_\Lambda = \otimes_{x\in\Lambda}\mathcal{H}_x$, where $\mathcal{H}_x$ are Hilbert spaces of dimensions $(2S_x+1)$, S_x being the magnitude of the total spin associated to the site x, and a matrix algebra $\mathfrak{A}_\Lambda = B(\mathcal{H}_\Lambda)$ generated by the matrices $\{S_x^{(1)}, S_x^{(2)}, S_x^{(3)}\}_{x\in\Lambda}$ satisfying $[S_x^{(1)}, S_{x'}^{(2)}] = 2iS_x^{(3)}\delta_{x,x'}$ plus cyclic permutations. For further use we let $\mathbf{S}_\Lambda = \sum_{x\in\Lambda}\mathbf{S}_x$, and $\mathbf{S}_x \equiv (S_x^{(1)}, S_x^{(2)}, S_x^{(3)})$. For A, B, C finite sets, let

$$S_{ABC} = \prod_{x\in A} S_x^{(1)} \prod_{x\in B} S_x^{(2)} \prod_{x\in C} S_x^{(3)},$$

if none of A, B or C is empty, otherwise $S_{ABC}=1$. To each $\Lambda \subset \mathbf{Z}^\nu$ we associate a rotationally invariant Hamiltonian

$$H_\Lambda = \sum_{A\cup B\cup C\subset\Lambda} c(A,B,C)S_{ABC}, \tag{1}$$

where $c(A,B,C)$ are real functionals of A, B, C such that $c(A,B,C)$ is zero unless A, B and C are mutually disjoint and $c(A,B,C) = c(A',B',C')$ wherever there exists a translation in the lattice simultaneously carrying A into A', B into B' and C into C' (translation invariance). We also assume c has finite range Δ, i.e., the set

$$\Delta = -\Delta = \{x \subset \mathbf{Z}^\nu \text{ such that } X \in \{0\} \text{ and } X = A\cup B\cup C, \text{ with } c(A,B,C) \neq 0\} \tag{2}$$

is finite. We also assume that the functionals $c(A,B,C)$ occurring in H_Λ have a bound independent of A, B and C, i.e.,

$$|c(A,B,C)| \leq \text{const (independent of } A,B,C). \tag{3}$$

Rotational invariance of the Hamiltonian means that for all $\Lambda \subset \mathbf{Z}^\nu$, $[H_\Lambda, \mathbf{S}_\Lambda] = 0$. An example of (1) is the rotationally invariant antiferromagnetic Hamiltonian ("ferrimagnetic" if there are some unequal spins, ferromagnetic if one of A or B is empty):

$$H_\Lambda = -\sum_{x,y\in A\subset\Lambda} [J_1(x-y)\mathbf{S}_x \circ \mathbf{S}_y] - \sum_{x,y\in B\subset\Lambda} [J_2(x-y) \times \mathbf{S}_x \circ \mathbf{S}_y] + \sum_{x\in A\subset\Lambda,\ y\in B\subset\Lambda} [J_3(x-y)\mathbf{S}_x \circ \mathbf{S}_y], \tag{4}$$

where $J_i(x) \geq 0$ and $J_i(x) = 0$ if $|x| > \text{diam}\ (\Delta)$ (diameter of the range Δ of the potential), $i = 1, 2, 3$, and A and B are "sublattices" of Λ, with $A \cup B = \Lambda$ (not necessarily uniquely defined, see Ref. 5. Let $\mathfrak{A}_\Gamma = \cup_{\Lambda\subset\mathbf{Z}^\nu}\mathfrak{A}_\Lambda$ be the local algebra, and $\mathfrak{A}$ be the quasi-local algebra (norm closure of $\mathfrak{A}_\Gamma$). It is shown in Ref. 4 that

$$\exists\ n\text{-}\lim_{\Lambda\to\infty} \exp(-itH_\Lambda)A\exp(itH_\Lambda) \equiv \tau^t(A),\ \forall\ A \in \mathfrak{A}_\Gamma, \tag{5}$$

where the $\lim_{\Lambda\to\infty}$ is in the sense of Van Hove (e.g., see Ref. 6), and that τ^t may be extended to an automorphism of $\mathfrak{A}$. Let Ω_Λ be a ground state of H_Λ in $\mathcal{H}_\Lambda$. We construct a (ground) state φ on $\mathfrak{A}$ by taking the limit along

some subsequence of $(\Omega_\Lambda, .\ \Omega_\Lambda)$. It defines by the G.N.S. construction a representation Π of $\mathfrak{A}$ on a Hilbert space $\mathcal{H}$ with cyclic vector Ω (the physical ground state), such that $\varphi(A)=(\Omega, \Pi(A)\Omega)$, $\forall\ A\in\mathfrak{A}$. From the time translation automorphism, $\mathcal{T}$ and the space translation automorphism (e.g., see Ref. 2) one constructs the unitary groups on $\mathcal{H}$, $t\in\mathbb{R}\to U_t$, $x\in\mathbb{Z}^\nu\to W(x)$, such that $W_x S_0^{(1,2,3)} W(x)^{-1}=S_x^{(1,2,3)}$ and the generator of U is the physical Hamiltonian, a densely defined self-adjoint operator on $\mathcal{H}$ satisfying $H\Omega=0$, $H\geq 0$ (see Ref. 7, which is also the standard reference for ground state representations).

The main result of this part is the following theorem, which is the precise analog of Theorem 7 of Ref. 2 for the ferromagnet, and may be considered as a generalization of that result. For conciseness, we always omit explicit reference to the representation Π, which is, however, implicit whenever operators on $\mathcal{H}$ are considered.

Theorem 1: Let $f\in\mathcal{S}(\mathbb{R})$, $A\in\mathfrak{A}(\Lambda_0)$, and $\mathcal{T}^t$ be the automorphism corresponding to a rotationally invariant Hamiltonian of finite range of form (1). Further, let

$$A(f)\equiv\int dt\, f(t)\ \mathcal{T}^t(A). \tag{6}$$

Then,

$$\lim_{\Lambda\to\infty}(\Omega,[\mathbf{S}_\Lambda, A(f)]\Omega)=(\Omega,[\mathbf{S}_{\Lambda_0},A]\Omega)\int_{-\infty}^{\infty}dt\ f(t). \tag{7}$$

Remark: Integral (6) is a norm integral on $\mathfrak{A}$, which exists since the sutomorphism $\mathcal{T}$ is strongly continuous.[4] In the proof of the above theorem, a crucial role will be played by the following proposition, which is a trivial consequence of the results of Lieb and Robinson.[8]

Proposition: For all $A\in\mathfrak{A}(\Lambda_0)$ there exist constants a and b (depending on A and Λ_0), such that for all $i=1,2,3$:

$$\|[\mathcal{T}^t(S_x^{(i)}),A]\|\leq b\ \exp[-\tfrac{1}{2}\ \mathrm{dist}(x,\Lambda_0)],$$

$$\forall\, t\in\mathbb{R},\quad \text{such that}\quad |t|\leq a|x|. \tag{8}$$

Proof of Theorem 1: since $H\Omega=0$, we have

$$(\Omega,[\mathbf{S}_\Lambda,\ \mathcal{T}^t(A)]\Omega)=(\Omega,[\mathcal{T}^{-t}(\mathbf{S}_\Lambda),A]\Omega). \tag{9}$$

Now, we have[4]

$$\mathcal{T}^t(\mathbf{S}_\Lambda)=\mathbf{S}_\Lambda+i\int_0^t ds\ \ \mathcal{T}^s([H_{\Lambda_1},\mathbf{S}_\Lambda]),$$
$$\forall\,\Lambda_1\ \text{s.t.}\ \Lambda_1\supset(\Lambda+\Delta), \tag{10}$$

as an equation on $\mathfrak{A}$, the above integral, as well as the following ones being norm integrals on $\mathfrak{A}$ (they are applications of $\mathfrak{A}_\Gamma$ into $\mathfrak{A}$). Now,

$$[\mathcal{T}^t(\mathbf{S}_\Lambda),\ A]=[\mathbf{S}_\Lambda,A]+i\int_0^t ds\ [\mathcal{T}^s([H_{\Lambda_1},\mathbf{S}_\Lambda]),\ A]$$
$$=[\mathbf{S}_{\Lambda_0},A]+i\int_0^t ds\,[\mathcal{T}^s([H_{\Lambda_1},\mathbf{S}_\Lambda]),A],$$
$$\forall\,\Lambda_1\ \text{s.t.}\ \ \Lambda_1\supset(\Lambda+\Delta). \tag{11}$$

We now prove that, $\forall\ f\in\mathcal{S}(\mathbb{R})$, $\forall\ A\in\mathfrak{A}(\Lambda_0)$, and H_Λ rotationally invariant

$$\exists\lim_{\Lambda\to\infty}\|\int_{-\infty}^{\infty}dt\ f(t)\int_0^t ds[\mathcal{T}^s([H_{\Lambda_1},\mathbf{S}_\Lambda]),\ A]\|=0. \tag{12}$$

Equations (11) and (12) yield

$$\lim_{\Lambda\to\infty}\|[\mathcal{T}^{-t}(\mathbf{S}_\Lambda),\ A]-[\mathbf{S}_{\Lambda_0},A]\|=0, \tag{13}$$

whence, in particular, (7) results.

To prove (12), we use the proposition. By rotational invariance of H_Λ, $[H_{\Lambda'},\mathbf{S}_\Lambda]=0$ if $\Lambda'\subseteq\Lambda$. Hence, $[H_{\Lambda_1},\mathbf{S}_\Lambda]$ for $\Lambda_1\supset(\Lambda+\Delta)$, is a sum of a number, $N(\Lambda)$ say, of operator monomials of the form $d(A,B,C)S_{ABC}$ concentrated in a region $\partial(\Lambda,\Delta)$ within Δ of the boundary of Λ (i.e., $A\cup B\cup C\subset\partial(\Lambda,\Delta)$ in the above monomials). The coefficients $d(A,B,C)$ depend on the $c(A,B,C)$ occurring in the Hamiltonian (1) and are, by (3), uniformly bounded by a constant independent of A,B,C. Let $r(\Lambda)$ be the diameter of Λ. It easily follows that there exists a constant e independent of Λ such that

$$N(\Lambda)\leq e[r(\Lambda)]^{\nu-1}. \tag{14}$$

Let

$$I_\Lambda^t\equiv\int_0^t ds[\mathcal{T}^s([H_{\Lambda_1},\mathbf{S}_\Lambda]),\ A],$$

and

$$p(\Lambda^0)=\mathrm{dist}(\partial(\Lambda,\Delta),\ \Lambda_0).$$

Clearly, one may find a constant $g>0$ independent of Λ such that

$$p(\Lambda)\geq g\,r(\Lambda). \tag{15}$$

We now write

$$\int_{-\infty}^{\infty}dt\,f(t)\,I_\Lambda^t=I_\Lambda^{(1)}+I_\Lambda^{(2)}, \tag{16}$$

where

$$I_\Lambda^{(1)}\equiv\int_{|t|\leq a\,p(\Lambda)}dt\,f(t)\,I_\Lambda^t, \tag{16a}$$

and

$$I_\Lambda^{(2)}\equiv\int_{|t|\geq a\,p(\Lambda)}dt\,f(t)\,I_\Lambda^t, \tag{16b}$$

and where a is the constant (depending on A and Λ_0) occurring in the proposition. Using now the proposition and Eqs. (14), and (15), we find that there exists a constant h independent of Λ such that

$$\|I_\Lambda^{(1)}\|\leq h[r(\Lambda)]^{\nu-1}\ \exp[-\tfrac{1}{2}gr(\Lambda)]\int_{-\infty}^{\infty}dt\ |t|\ |f(t)| \tag{17}$$

Further, by (14) and (15) there exists a constant s and a positive constant m independent of Λ such that

$$\|I_\Lambda^{(2)}\|\leq s[r(\Lambda)]^{\nu-1}\int_{|t|\geq m\,r(\Lambda)}dt\ |t|\ |f(t)|. \tag{18}$$

By (17), $\lim_{\Lambda\to\infty}I_\Lambda^{(1)}=0$, and by (18) and the property that $f\in\mathcal{S}(\mathbb{R})$, $\lim_{\Lambda\to\infty}I_\Lambda^{(2)}=0$. Those two facts and (16) imply (12). QED

2. RESULTS DEPENDING ON ASSUMPTIONS ON GROUND STATE

In this section, we state more precisely the assumptions on φ (or Ω) mentioned in the Introduction. To do that we restrict ourselves to the model described by the Hamiltonian[8] which includes a large number of cases of physical interest. The same proof is applicable to several Hamiltonians with more complicated interactions and similar structure. We assume that:

(i) φ is pure, i.e., Ω is the unique state in $\mathcal{H}$ such that $H\Omega = 0$;

(ii) Ω is invariant under translations within A or B, namely: $W(x-y)\Omega = \Omega$ for all $x, y \in A$, *or* $x, y \in B$;

(iii) Ω has long range order, i.e.,

$$\lim_{|x-y|\to\infty;\ x,y\in A \text{ or } B} (\Omega,\ S_x^{(3)} S_y^{(3)}\Omega) = \gamma > 0 \tag{19}$$

From these assumptions and Theorem 1 we have as a corollary, "Goldstone's theorem":

Theorem 2: For model (4) and under assumptions (1)–(3), there is no gap in the spectrum of the physical Hamiltonian H.

Proof: Using assumption (1), we may (as in Ref. 2, p. 245), follow the proof of Ref. 1 to prove Ref. 2, Lemma 3 holds as long as the spectrum of H has a gap. This lemma, in conjunction with Theorem 1, implies that, as long as the spectrum of H has a gap, for all $A \in \mathfrak{A}(\Lambda_0)$, that

$$\lim_{\Lambda\to\infty} (\Omega, [\mathbf{S}_\Lambda, A]\Omega) = 0. \tag{20}$$

Specializing in (20) A to $S_0^{(2)}$, and taking the one-component of $\mathbf{S}_\Lambda$, we have

$$0 = \lim_{\Lambda\to\infty} (\Omega, [S_\Lambda^{(1)}, S_0^{(2)}]\Omega) = 2i(\Omega, S_0^{(3)}\Omega). \tag{21}$$

On the other hand, using assumptions (2) and (4) it is a corollary from asymptotic abelianness[3] that

$$(\Omega, S_0^{(3)}\Omega) = \pm\sqrt{\gamma} \neq 0. \tag{22}$$

Equation (22) contradicts (21) and hence the fact that the spectrum of H has a gap. QED

Remarks: (a) Assumption (1) was not known for the ferromagnet at the time of writing of Ref. 2 and was only proved later.[9] Hence it was replaced by ((2), Lemma 1), which was also sufficient. (b) Unicity and long range order are open problems even for the isotropic one-dimensional antiferromagnetic Heisenberg chain, where much is known about the ground state for finite volume.[10] If they were proved, the conjectured[11] absence of an energy gap would follow from Theorem 2. Finally, we mention a related work of Swieca[12] which proves Goldstone's theorem for many-body systems, under an assumption of falloff of the commutators similar but more general than the one in the proposition, which so far, however, has not been rigorously proved.

ACKNOWLEDGMENTS

We should like to thank Professor M. Romerio and Professor P. Huguénin for their hospitality at Neuchâtel, where part of this paper was written.

[1] D. Kastler, D.W. Robinson and J.A. Swieca, Comm. Math. Phys. **2**, 108 (1966).
[2] R.F. Streater, Comm. Math. Phys. **6**, 233 (1967).
[3] H. Reeh, Nuovo Cimento B **17**, 7 (1973).
[4] D.W. Robinson, Comm. Math. Phys. **7**, 337 (1968).
[5] E.H. Lieb and D.C. Mattis, J. Math. Phys. **3**, 749 (1962).
[6] D. Ruelle, *Statistical Mechanics* (Benjamin, New York, 1969).
[7] D. Ruelle, Comm. Math. Phys. **11**, 339 (1969).
[8] E.H. Lieb and D.W. Robinson, Comm. Math. Phys. **28**, 251 (1972).
[9] K. Osterwalder and R. Schrader (unpublished).
[10] C.N. Yang and C.P. Yang, Phys. Rev. **150**, 321 (1966).
[11] E.H. Lieb and D.C. Mattis, Ann. Phys. **16**, 401 (1961).
[12] J.A. Swieca, Comm. Math. Phys. **4**, 1 (1967).

An application of the method of geometrical optics to the scattering of plane electromagnetic waves off cylindrically confined cold plasmas

Gregory A. Kriegsmann

Courant Institute of Mathematical Sciences, New York University, New York, New York 10012
(Received 27 May 1975)

In this paper we apply the method of geometrical optics to study the scattering of plane electromagnetic waves off a cylindrically confined cold plasma. For simplicity, we assume two types of incident polarization. In both cases scalar second order elliptic partial differential equations describe the fields. These problems are studied in the asymptotic limit $a\omega/c \to \infty$ (where a is the radius of the cylinder, ω is the frequency of the incident plane wave, and c is the velocity of light in free space). We furthermore assume a quadratic plasma density. This allows us to calculate explicitly the rays, amplitudes, caustics, and other features which arise in the geometrical optics approach. The assumed density also gives rise to the interesting cutoff and resonance phenomenon. Thus the amplification of the electromagnetic fields is observed. This phenomenon may have applications in laser fusion.

1. INTRODUCTION

The possibility of the amplification of electromagnetic fields within a plasma has recently attracted much attention.[1] For example, the heating of a gaseous plasma by a laser may have application for controlled fusion. In a recent review article, Chen and White[2] discuss certain aspects of this problem in a "cold overdense" plasma slab. Other geometrical configurations are as important in physical application's as the planar situation. The purpose of this paper is to treat, in some detail, the case of cylindrical geometry.

The scattering of plane electromagnetic waves off a cylindrically confined medium has been studied analytically by Alexopoulos.[3,4] In these papers the author performs a modal analysis using a specific index of refraction $(r/a)^m$ (where r is the radial variable and a is the radius of the cylinder). He computes the asymptotic approximation to the backscattered field in the limit $a\omega/c \to \infty$ (here ω is the frequency of the incident plane wave and c is the speed of light in free space). His analysis is made amenable by his particular choice of refractive index which, for $m > 0$, is a reasonable model for a plasma medium.

In this paper we shall study a situation related to that discussed by Chen and White.[2] We assume that the diagonal elements of the "dielectric" tensor vanish on a cylinder $(r = r_0 > 0)$ within the plasma. We further suppose these elements are functions of r alone. This gives rise to interesting cutoff and resonant phenomenon in the media.

In this work we shall suppose ω to be fixed and seek asymptotic approximations to the fields in the limit as $a\omega/c \to \infty$. To simplify our analysis, we shall consider only two types of incident polarization. The first choice is to orient the incident electric field parallel to the axis of the cylinder (the E problem). The second choice is to orient the incident magnetic field parallel to the axis of the cylinder (the B problem). The mathematical formulation of these problems give rise to scalar second order elliptic partial differential equations which describe the appropriate fields.

In this work we apply the method geometrical optics to find asymptotic approximations to these fields. This procedure has been developed to a large degree by J. B. Keller and his colleagues.[5-8] In particular we shall make use of the ideas and techniques described in Keller and Seckler[5] and in Keller and Buchal,[7] where problems involving inhomogeneous media and "caustic boundary layers" are discussed respectively. It is of some interest that, for the models which we study, it is possible to calculate explicitly the rays, amplitudes, caustics, and other features which arise in the geometrical optics approach.

Part I of this paper is primarily concerned with the E problem. Sections 2—9 are devoted to the computation of the electric field, both interior and exterior to the plasma. We find that several separate asymptotic expansions are necessary to describe the fields adequately. In Sec. 10 we examine the scattering of a scalar plane wave off a spherical object whose index of refraction is similar to that of the plasma media. This problem does not have its origins in plasma physics but is recorded here for its mathematical novelty.

Part II of this paper is entirely devoted to the B problem. We find that the results, when interpreted properly, are very similar to those obtained in part I. There is one exceptional region which we discuss in Sec. 12 and 13. In Sec. 14 we compute the electric field at the cutoff cylinder $(\omega = \omega_p)$ and find it unbounded in the absence of collisional damping. We then briefly discuss the addition of damping and find that our results agree with Ref. 2 for the case of planar geometry.

The case, where an externally applied uniform magnetic field B_0 is placed along the axis of the cylinder, is not treated here. Certain aspects of this problem may be found in this author's thesis.[9] Also, a more detailed presentation of Parts I and II of this paper can be found there.

PART I. THE E PROBLEM—AN ILLUSTRATIVE EXAMPLE

2. FORMULATION

A high frequency plane electromagnetic wave impinges

upon a cylindrically confined "cold" plasma of infinite extent and scatters from it. This cylinder is circular in cross section with radius a. We further assume that the incident radiation is polarized with the electric field parallel to the generators of the cylinder. It follows from this assumption and the equations governing the plasma[10] that the time harmonic electric field $\phi \exp(-iwt)$ remains in this direction and satisfies the Helmholtz equation

$$\nabla^2\phi + k^2\mu_0^2\phi = 0. \tag{2.1}$$

In this equation $k = k'a$ and the index of refraction μ_0^2 is given by[10]

$$\mu_0^2 = 1 - 4\pi e^2 N(ar)/m\omega^2, \tag{2.2}$$

where ω is the frequency of the incident plane wave, e is the charge of an electron with mass m, r is the radial variable, and $N(ar)$ is the charge density. Implicit in (2.2) are the assumptions that ω is large enough to neglect ionic motion and that the density is dependent only upon r.

For simplicity we shall now choose the particular density

$$N(ar) = N_0(1-r^2), \quad r<1, \quad N(ar)=0, \quad r>1. \tag{2.3}$$

This implies that

$$\mu_0^2 = (r^2 - r_0^2)/(1-r_0^2), \ r<1, \ \mu_0^2 = 1, \ r>1. \tag{2.4}$$

In (2.4), $r_0^2 = 1 - m\omega^2/4\pi e^2 N_0$ and we restrict the values of N_0 and ω to insure that $0 < r_0 < 1$. This restriction on r_0 is physically realized when an infrared plane wave impinges on an overdense plasma.[2] Although this particular choice of $N(ar)$ cannot be expected to represent an actual plasma concentration, it does possess the correct qualitative behavior. Quadratic densities have been used by other authors in studying related problems.[11]

To complete the mathematical statement of this problem, we must impose further conditions. First we demand that ϕ and $\nabla\phi$ are continuous everywhere. Secondly, the scattered field must satisfy the radiation condition. Finally we choose the x axis to be parallel to the incident wave vector $\mathbf{k}'$ and the z axis to be that of the cylinder.

We shall now suppose that $k \gg 1$ which corresponds to the physical situation mentioned above. Thus we seek an asymptotic approximation of ϕ as $k \to \infty$. This is of course a natural setting for the method of geometrical optics.[5] According to this procedure, one assumes that the function ϕ is of the form

$$\phi(x,y,k) = e^{ik\psi(x,y)}[A(x,y) + O(1/k)] \tag{2.5}$$

as $k \to \infty$. This is suppose to be valid except in certain regions where other expansions are appropriate. Thus our first task is to compute A and ψ and to determine where (2.5) breaks down.

Substituting (2.5) into (2.1) and equating coefficients of like powers of k, we obtain

$$\nabla\psi \cdot \nabla\psi = \mu_0^2 \quad \text{(eikonal equation)}, \tag{2.6}$$

$$2\nabla A \cdot \nabla\psi + A\nabla^2\psi = 0 \quad \text{(transport equation)}. \tag{2.7}$$

3. THE RAYS

The eikonal equation is a nonlinear first order partial differential equation. We will solve it by introducing a one parameter family of curves called "rays" which are everywhere orthogonal to the curves $\psi = \text{const.}$[8] Since the incident field is a plane wave, its rays are horizontal lines. Consider the incident ray which strikes the cylinder at $(\cos\alpha, \sin\alpha)$. It is convenient to use α as the parameter which labels a ray (i.e., α is a constant along a ray). If we introduce the parameteric representation $\mathbf{x}(t;\alpha)$ of a ray, the orthogonality condition can be expressed in the form

$$\frac{d\mathbf{x}}{dt} = 2\nabla\psi. \tag{3.1}$$

It then follows from the eikonal equation and (3.1) that

$$\frac{d^2\mathbf{x}}{dt^2} = \nabla\mu_0^2 = w^2\mathbf{x}, \tag{3.2}$$

where $w^2 = 4/(1 - r_0^2)$.

Let $t = 0$ correspond to the point at which the incident ray enters the plasma. The initial data needed to solve (3.2) is deduced from the definition of α, (3.1), the form of the incident wave $[\exp(ikx)]$, and the continuity of ϕ. We find that

$$\mathbf{x}(0,\alpha) = (\cos\alpha, \sin\alpha), \quad \pi/2 \le \alpha \le 3\pi/2,$$
$$\mathbf{x}_t(0,\alpha) = (2,0), \quad \psi(0,\alpha) = \cos\alpha. \tag{3.3}$$

The solution of (3.2) satisfying (3.3) is easily found to be

$$\mathbf{x} = (x,y) = \cosh wt(\cos\alpha, \sin\alpha) + (2/w)\sinh wt(1,0). \tag{3.4}$$

This vector equation is valid for $0 \le t \le t_e$. The parameter t_e corresponds to the point at which the ray leaves the plasma. We find by setting $x^2(t,\alpha) + y^2(t,\alpha) = 1$ that

$$\tanh wt_e = -4\cos\alpha/(w + 4/w). \tag{3.5}$$

After the ray leaves the plasma, $\mu_0^2 = 1$ and it becomes a straight line. These rays will be determined later.

We shall now show that the rays within the plasma possess several interesting geometrical properties. First by combining (3.4) and (3.5) we have $x(t_e(\alpha), \alpha) \ge 0$ and $y(t_e(\alpha), \alpha) \le 0$. That is a ray which enters the plasma in the second quadrant leaves in that quadrant. Secondly, if we eliminate t from the ray equation we obtain

$$x^2 - 2xy\cot\alpha + y^2(-1 + r_0^2\csc^2\alpha) = r_0^2 - 1. \tag{3.6}$$

Thus the rays are hyperbolas whose axes depend upon α.

Finally, the rays envelope a simple curve. To show this we let $J = x_t y_\alpha - x_\alpha y_t$ be the Jacobian of the transformation $(t,\alpha) \to (x,y)$ given by the rays. By direct computation we find

$$J = w\cosh^2 wt[\tanh wt + (2/w)\cos\alpha]. \tag{3.7}$$

Now $J = 0$ is a necessary condition for the rays to have an envelope. From (3.7) we find that J vanishes only if

$$\tanh wt = -(2/w)\cos\alpha. \tag{3.8}$$

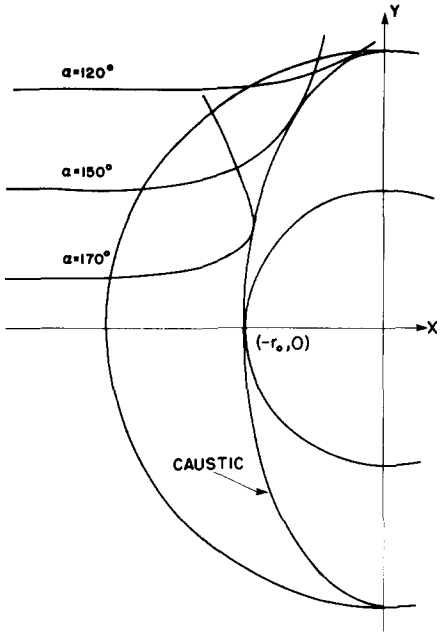


FIG. 1. Typical rays in the irradiated portion of the plasma.

It is easy to show that this equation has a unique real solution $t_c(\alpha)$. When we insert $t_c(\alpha)$ into the ray equations, we obtain the "caustic" curve $(x(t_c(\alpha),\alpha), y(t_c(\alpha),\alpha))$. Upon eliminating α, we find that the caustic is the ellipse $x^2/r_0^2+y^2=1$. Several rays, for various values of α, are shown in Fig. 1.

It is apparent from Fig. 1 that through each point in the "irradiated" portion of the plasma, bounded by the caustic and the unit circle, pass two rays. We call the ray which has (not) reached the caustic the outgoing (incoming) ray. Thus we expect there to be two pairs, (α_i, t_i) and (α_0, t_0), which are mapped by the ray equation into the same point (x,y) in the irradiated portion of the medium.

The veracity of this observation can be shown as follows. Set $l=\sinh wt$ and eliminate α from the ray equation to obtain

$$(x-2l/w)^2+y^2=1+l^2. \tag{3.9}$$

This quadratic equation yields the solutions

$$wt=\operatorname{arcsinh}\{2[-x\pm r_0 w(x^2/r_0^2+y^2-1)^{1/2}/2]/wr_0^2\}. \tag{3.10}$$

Thus for a fixed value of (x,y) there are two values of t. The incoming value of t, t_i, is obtained by using the minus sign while the plus sign yields the outgoing value t_0.

To find the corresponding values of α, we solve the ray equation for $\tan\alpha$ and obtain $\tan\alpha=y/(x-2l/w)$. This coupled with (3.10) yields

$$\alpha=\arctan\{y/[x/r_0^2\mp 2(x^2/r_0^2+y-1)^{1/2}/r_0 w]\}. \tag{3.11}$$

In this formula the negative sign yields α_i while the positive sign gives α_0.

We shall now conclude this section by computing the phase function ψ along each ray. From the eikonal equation and (3.1) it follows that $\psi_t=2\mu_0^2$ along a ray. Combining (2.4) and (3.4) and using (3.3), we obtain

$$\psi=\cos\alpha\cosh^2 wt+(1-w^2/4)t+(1+w^2/4)\sinh 2wt/2w. \tag{3.12}$$

4. THE GEOMETRICAL OPTICS FIELD

To complete the computation of the geometrical optics approximation, one must solve the transport equation. Along the rays this equation reduces to[8]

$$A_t+J_tA/2J=0, \tag{4.1}$$

where J is the Jacobian given by (3.7). From (3.7) at $t=0$ and the assumption that the amplitude of the incident plane wave is one, we find the solution of (4.1) to be

$$A=|2\cos\alpha/J|^{1/2}. \tag{4.2}$$

Now let P be a point in the irradiated portion of the plasma. In the previous section we proved that one incoming and one outgoing ray pass through P. The total electric field ϕ at P will be composed of an incoming and outgoing wave. Thus the field at P is given by (2.5) as

$$\phi\sim|2\cos(\alpha_i)/J_i|^{1/2}\exp(ik\psi_i)+|2\cos(\alpha_i)/J_o|^{1/2}\times\exp(ik\psi_0-i\pi/2). \tag{4.3}$$

Here the subscripts i and o again refer to the incoming and outgoing rays respectively. The phase shift in the outgoing field is due to the well-known fact that, as a ray passes a caustic, the phase of the field is retarded by $\pi/2$ radians.[5] An obvious restriction on (4.3) is that it becomes invalid in a narrow strip about the caustic. This fact is a well-known deficiency of the geometrical optics approximation.

5. THE CAUSTIC BOUNDARY LAYER

To remove this apparent singularity, we assume in a neighborhood of the caustic that $\phi=A(x,y,k)e^{ik\psi(x,y)}$. When this "ansatz" is inserted into (2.1), we find that ψ must satisfy the eikonal equation and that

$$ik(2\nabla A\cdot\Delta\psi+A\nabla^2\psi)+\nabla^2 A=0. \tag{5.1}$$

The local analysis of (5.1) near the caustic is facilitated by using the (α,t) coordinate system. In terms of these coordinates (5.1) becomes

$$ik(JA_t+\tfrac{1}{2}J_tA)+\frac{\partial}{\partial t}\left(\frac{g_{22}A_t}{J}-\frac{g_{12}A_\alpha}{J}\right)+\frac{\partial}{\partial\alpha}\left(-\frac{g_{12}A_t}{J}+\frac{g_{11}A_\alpha}{J}\right)=0, \tag{5.2}$$

where $g_{11}=|\mathbf{x}_t|^2$, $g_{12}=\mathbf{x}_t\cdot\mathbf{x}_\alpha$, and $g_{22}=|\mathbf{x}_\alpha|^2$. These quantities can easily be computed from (3.4).

In a neighborhood of the caustic $|t-t_c|\ll 1$, where t_c is given implicitly by (3.8), we introduce the boundary layer variable

$$\tau=k^{1/3}(t-t_c). \tag{5.3}$$

Then (5.2) reduces to

$$A_{\tau\tau}-\frac{1}{\tau}A_\tau+\frac{i\tau}{g}\left(\tau A_\tau+\tfrac{1}{2}A\right)$$

$$=k^{-1/3}\left[-\frac{4\sin\alpha\,(w^2-4)}{\gamma}A_{\tau\alpha}+\frac{2\sin\alpha\,(w^2-4)}{\gamma\tau}A_\alpha\right.$$

$$\left.+\frac{8\sin\alpha\cos^2\alpha(w^2-4-4\sin^2\alpha)}{\gamma}A_\tau\right] \qquad (5.4)$$

$$+k^{-2/3}(-4\sin^2\alpha)A_{\alpha\alpha}+O(k^{-1}),$$

where $\gamma=w^2-4\cos^2\alpha$, $g=\rho_c^2/w^4$, and ρ_c is the caustic's radius of curvature. If we assume that the α derivatives are $O(1)$ as $k\to\infty$, we may neglect the right-hand side of (5.4) and obtain

$$A_{\tau\tau}-(1/\tau)A_\tau+i\tau(\tau A_\tau+A/2)/g=0. \qquad (5.5)$$

The solution of (5.5) is

$$A=\exp(-\tau^3/6g)[e_1\mathrm{Ai}(-\eta e^{2\pi i/3})+e_2\mathrm{Ai}(-\eta e^{-2\pi i/3})], \qquad (5.6)$$

where $\eta=\tau^2/(4g)^{2/3}$, the e_i are constants, and $Ai(\xi)$ is the Airy function.[12] These results are essentially those given in ref. 7. when $\mu_0^2=1$.

Let (x,y) be an arbitrary point in the caustic boundary layer. Since there are two rays passing through this point, it follows that

$$\phi^{BL}\sim\exp(ik\psi_i-i\tau_i^3/6g_i)[e_1^i\mathrm{Ai}(-\eta_i e^{2\pi i/3})$$

$$+e_2^i\mathrm{Ai}(-\eta_i e^{-2\pi i/3})]+\exp(ik\psi_o-i\tau_o^3/6g_o) \qquad (5.7)$$

$$\times[e_1^o\mathrm{Ai}(-\eta_o e^{2\pi i/3})+e_2^o\mathrm{Ai}(\eta_o e^{-2\pi i/3})].$$

Here again the subscripts i and o refer to the incoming and outgoing rays respectively. Formally, this result is valid for $t-t_c\sim O(k^{-1/3})$. To find the unknown constants e_i^j, we "match" the results (4.3) and (5.7). This matching procedure is based on the assumption that the regions where (4.3) and (5.7) are valid overlap. In this region one assumes that $\eta\gg1$ while $|t-t_c|\ll1$. Matching is performed by comparing the asymptotic expansion of (5.7) for $\eta\to\infty$ with the expansion of (4.3) as $t\to t_c$.[12] We apply this procedure and find that

$$\phi^{BL}\sim\frac{\sqrt{\pi}\,2^{7/6}k^{1/6}}{w^{1/3}}\left[\frac{|\cos\alpha_i|^{1/2}}{\rho_c^{1/3}(\alpha_i)}\exp\left(ik\psi_i+\frac{i\pi}{12}-\frac{i\tau_i^3}{6g_i}\right)\right.$$

$$\times\mathrm{Ai}(-\eta_i e^{-2\pi i/3})+\frac{|\cos\alpha_o|^{1/2}}{\rho_c^{1/3}(\alpha_o)}\exp\left(ik\psi_o-\frac{7i\pi}{12}-\frac{i\tau_o^3}{6g_i}\right)$$

$$\left.\times\mathrm{Ai}(-\eta_o e^{2\pi i/3})\right]. \qquad (5.8)$$

Thus the solution ϕ of (2.1) becomes quite large $O(k^{1/6})$ near the caustic. This is, of course, due to the pinching of neighboring rays as they become tangent to the caustic.

We shall now verify the phase shift stated in (4.3). This will be accomplished by applying a technique developed in Ref. 7. The technique is as follows. First the boundary layer expansion (5.8) is rewritten in terms of n, the normal distance from a point P to the caustic. Thus, negative values of n correspond to points inside the caustic. Next, $k^{2/3}n$ is allowed to become large and negative. There are terms in (5.8) which grow exponentially with k. Since the plasma is being excited by an external plane wave, we suppress these exponentially growing solutions. Finally, the factor $e^{-i\pi/2}$ appearing in (4.3) guarantees the suppression of these growing terms.

In order to carry out this procedure, we let n denote the normal distance from P to the caustic and $\bar\alpha$ correspond to the ray which is tangent to the caustic and orthogonal to the normal (Fig. 2). We deduce from a power series expansion of (3.4) about t_c that

$$n\simeq 2\mu_0^2(P)\,[K_R(t_c(\alpha_i))+K_c(Q)\,]k^{-2/3}\tau_i^2, \qquad (5.9)$$

where K_R and K_c are the curvatures of the ray and caustic respectively at point Q. A similar result holds for τ_0.

We state this result because there is a slight subtlety. We find that from (5.4) that

$$K_R(t_c(\alpha_i))=\gamma_i^{3/2}/4w\sin^2\alpha_i, \qquad (5.10)$$

which becomes singular as $\alpha_i\to\pi$. This apparent singularity is removed from (5.9) by expressing μ_0^2 in terms of τ and α. From the definition of g_{11} and (3.4) we find that

$$\mu_0^2(P)\simeq w^2\sin^2\alpha_i/\gamma_i. \qquad (5.11)$$

The curvature of the caustic is found to be $K_c(Q)=\gamma_i^{3/2}/w(w^2-4)$. When this expression, (5.11), and (5.10) are inserted into (5.9), we obtain

$$n\approx(w^2/2)K_c(Q)k^{-2/3}\tau_i^2. \qquad (5.12)$$

Since the caustic is a smooth function, we may replace the Q in (5.12) by S without introducing any significant error.

With the aid of (5.12) and the identity $\mathrm{Ai}(*\exp(\mp2\pi i/3))=0.5\exp(\mp i\pi/3)\cdot[\mathrm{Ai}(*)\pm i\mathrm{Bi}(*)]$ (Bi is the Airy function of the second kind) we may rewrite (5.8) in terms of n and $\bar\alpha$. Thus we obtain

$$\phi^{BL}\sim\frac{(2k)^{1/6}|\pi\cos\bar\alpha|^{1/2}}{(w\rho_c(\bar\alpha))^{1/3}}\exp(ik\bar\psi-i\pi/4)\mathrm{Ai}(-\chi\bar n), \qquad (5.13)$$

where $\chi=[w^2/2\rho_c(\bar\alpha)]^{1/3}$. From the expansion of Ai for $-n\gg1$ we find the field decays exponentially behind the caustic.

The result (5.13) is formally valid in a small

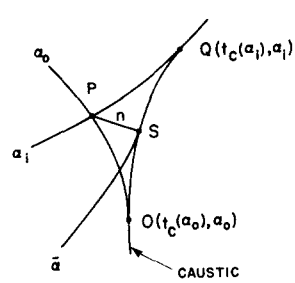


FIG. 2. Local structure of the rays near the caustic.

neighborhood about the caustic. It was obtained with the proviso that all α derivatives were $O(1)$ as $k \to \infty$. Since the approximation (5.13) has the factor $|\cos\alpha|^{1/2}$, this is not the case as $\alpha \to \pi/2$ or $3\pi/2$. Thus we expect neither (4.3) nor (5.13) to be valid in a small neighborhood about the poles.

6. THE POLAR BOUNDARY LAYER

In this section we shall remove the apparent singularity in the α derivatives at $\alpha = \pi/2$. Analogous results hold at $\alpha = 3\pi/2$. To this end we introduce the stretched variable

$$\bar{s} = k^{1/3}(\alpha - \pi/2) \tag{6.1}$$

into (5.4) and obtain

$$A_{\tau\tau} + \frac{4}{r_0^2} A_{\tau\bar{s}} + \frac{4}{r_0^4} A_{\bar{s}\bar{s}} - \frac{1}{\tau}\left(A_\tau + \frac{2A_{\bar{s}}}{r_0^2}\right) + \frac{iw^4\tau}{r_0^4}(\tau A + A/2) \approx 0. \tag{6.2}$$

We find by setting $\bar{\xi} = \gamma\,(\tau - r_0^2\bar{s}/2)$, $\bar{\rho} = \gamma^2\tau^2$, $A = \exp(-\tau^3/6g) f(\bar{\rho}, \bar{\xi})$, and $\gamma^3 = w^4/4r_0^2$ that f must satisfy

$$f_{\bar{\rho}\bar{\rho}} + \bar{\rho} f = -i f_{\bar{\xi}}. \tag{6.3}$$

This is just the time dependent Schrödinger equation with a linear potential. Equations similar to this have been derived and discussed by Zauderer[13] in his work on boundary layers in diffraction problems.

By separation of variables the solution of (6.3) is of the form

$$f = \int_0^\infty H(E) e^{-iE\bar{\xi}} \mathrm{Ai}(-(E+\bar{\rho}) e^{\pm 2\pi i/3})\, dE, \tag{6.4}$$

where $H(E)$ must be determined. The total field is given by

$$\phi^c \sim \exp[ik\psi_i + (2i/3)\bar{\rho}_i^{3/2}] f_i + \exp[ik\psi_o - (2i/3)\bar{\rho}_o^{3/2}] f_o, \tag{6.5}$$

where f_i and f_o satisfy (6.4).

We now consider the incoming amplitude A^i defined by

$$A^i = e^{-i\eta_i^3/6g} \int_0^\infty H(E) e^{-iE\bar{\xi}} \mathrm{Ai}(-(E+\bar{\rho}_i) e^{-2\pi i/3})\, dE. \tag{6.6}$$

When $t = 0$ we find from (5.3) that $\tau_i = -k^{1/3} l_c(\alpha_i) \sim 2\bar{s}_i/w^2$. We further deduce from the definitions of $\bar{\xi}$ and $\bar{\rho}$ that $\bar{\xi}_i = -\gamma \bar{s}_i/2$ and $\bar{\rho}_i = 4\gamma^2 \bar{s}_i^2/w^2$. In order to match the incoming wave we take $A^i = 1$ when $t = 0$. Inserting this data into (6.5), we obtain for $H(E)$ the integral equation

$$\exp(-16i\chi^3 \bar{s}^3/3w^6) = \int_0^\infty H(E) e^{i\chi \bar{s} E/2} A_i(-e^{-2\pi i/3} B)\, dE, \tag{6.7}$$

where $B = E + 4\chi^2 \bar{s}^2/w^4$. We have not solved this equation. However, if we suppose that

$$H(E) \sim C_2 \sqrt{E}\, \exp(iE^2/4\chi) \quad \text{as } E \to \infty, \tag{6.8}$$

where $C_2 = \exp(i\pi/12)(4\pi/r_0^2 w^2)^{1/2}$, we can show that A^i given by (6.6) has the correct asymptotic behavior to match the geometrical optics solution. The details of this analysis are given in the Appendix.

It suffices to note here that in a neighborhood of the polar cap the incoming geometrical optics field reduces to

$$\phi_i \sim |2\bar{s}/-w^2\tau_i|^{1/2} \exp(ik\psi_i). \tag{6.9}$$

A similar assumption on the asymptotic form of $H(E)$ for the outgoing amplitude can be made and the previous analysis can be performed in the same fashion. So, although we cannot compute the solution exactly, we do know its asymptotic behavior.

7. THE SOLUTION WITHIN THE CAUSTIC

Let $\mathcal{E}$ denote the region within the ellipse $x^2/r_0^2 + y^2 = 1$ with $y \geq 0$. This section is concerned with the determination of an approximate value to the field in $\mathcal{E}$. results for $y \leq 0$ follow from the geometrical symmetry of the problem.

Now from (3.10) and (3.11) it follows that a point in $\mathcal{E}$ yields complex values of wt and α. It these values are inserted into (3.12) and (4.2), the phase and amplitude become complex valued functions. Thus by combining this information with (2.5), we may formally continue the geometrical optics solution behind the caustic. Although this procedure is formal, it has yielded correct asymptotic results for several classes of problems whose exact solutions are known.[5]

There is some apparent ambiguity in this procedure which we shall now dispel. First, there are two choices of sign in (3.10). As in Ref. 6 we choose the minus sign which corresponds to continuing the incoming field in $\mathcal{E}$. This will force the imaginary part of ψ to be positive and insure that the field is exponentially small within $\mathcal{E}$ [cf. (5.13)].

To exhibit the other ambiguity, we eliminate $\cos\alpha$ from (3.12) and obtain

$$\psi = (1+l^2)^{1/2}(x + wr_o^2 l/4) - (wr_o^2/4)\ \sinh^{-1}(l), \tag{7.1}$$

where $l = \sinh(wt) = a + ib$. The function $l(x, y)$ is given by (3.10) with the choice of sign noted above. From (7.1) we deduce that ψ is a multivalued function of l and hence of x and y.

We choose the correct branch of ψ by introducing a cut in the region $\mathcal{E}$ from the point $(0, 1)$ to $(0, (1-r_o^2)^{1/2})$ along the y axis. This induces a cut in the complex l plane from $+i$ to $-i$ along the imaginary axis. With this cut l maps $\mathcal{E}$ onto the region

$$l(\mathcal{E}) = \{(w^2 r_o^2/4) a^2 + r_o^2 b^2 \leq 1;\ b \leq 0;\ a = \mathrm{Re}(l);\ b = \mathrm{Im}(l)\}$$

in a one-to-one manner. This region is then taken by the map $\sinh^{-1}(l)$ onto the rectangle $R = \{-\pi \leq \mathrm{Im}(wl) \leq 0,\ 0 \leq \mathrm{Re}(wl) \leq \sinh(2/wr_o)\}$ in similar fashion. The image of the previous cut lies on the imaginary axis from $-\pi/2$ to $-\pi$.

It is now an easy matter to write ψ in terms of the real and imaginary parts of wl and to interpret this result physically. We find the along the lines Re (wl) = const the imaginary part of ψ increases monotonically from 0 to $wr_o^2\pi/4$. These curves correspond to confocal hyperbolas in $\mathcal{E}$ with foci at $(0, \pm(1-r_o^2)^{1/2})$. Hence as one progresses through $\mathcal{E}$ the solution (2.5) decays

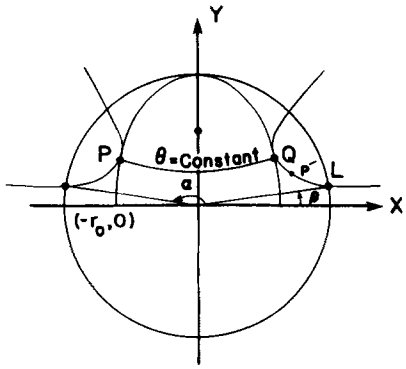


FIG. 3. Symmetry of the two ray systems. On the hyperbola connecting P and Q, $\mathrm{Re}(\psi) = \theta = \text{const}$.

exponentially. It is interesting to note that the real part of ψ remains constant along these hyperbolas.

Taking these results into account, we find that the field is given by

$$\phi \sim e^{-i\pi/4}(2/\rho_c w^2 n)^{1/4}\, e^{-kr_o^2 w\pi/4}\, |\cos\bar{\alpha}|^{1/2} \times e^{ik\bar{\psi}} \cdot \exp[2wn^{3/2}/3(2\rho_c)^{1/2}] \tag{7.2}$$

near the backside of the caustic (i.e., $x^2/r_0^2 + y^2 = 1$, $x \geq 0$). Here n denotes the normal distance from a point in $\mathcal{C}$ to this curve and $n \ll 1$. Now (7.2) becomes singular as $n \to 0$, and, to rectify this condition, we must perform a boundary layer analysis. This analysis would be analogous to that done on the irradiated side of the caustic. In the next section we shall continue ϕ into the region $T = \{(x,y) \mid x^2/r_0^2 + r^2 \geq 1; x^2 + y^2 \leq 1; x \geq 0\}$ with the aid of such a boundary layer.

8. THE TRANSMITTED REGION

We assume in this region (T) that the field ϕ is of the form (2.5). Thus, our task is to compute ψ and A and to determine where (2.5) breaks down. We shall accomplish this by the following indirect method.

Suppose we solve the eikonal equation subject to the initial data

$$(x_o, y_o, p_o, q_o, \psi_o) = (\cos\beta,\ \sin\beta,\ -1, 0,\ -\cos\beta), \tag{8.1}$$

where $|\beta| \leq \pi/2$. The ray equation has the solution

$$x = \cosh w\xi\,(\cos\beta,\ \sin\beta) + (2/w)\sinh w\xi(-1, 0). \tag{8.2}$$

Here β is the parameter which is constant along a ray. This vector equation is valid for $0 \leq \xi \leq \xi_e$, where ξ_e is given implicitly by $\tanh w\xi_e = 4\cos\beta/(w + 4/w)$. If we set $\alpha = \pi - \beta$, then the ray given by (8.2) is the reflection of the α ray given by (3.4) about the y axis (Fig. 3).

Due to this symmetry, the β rays have the curve $x^2/r_0^2 + y^2 = 1$ ($x \geq 0$) as its caustic and envelope. The phase ψ again satisfies along each ray, $d\psi/d\xi = 2\mu_o^2$. The solution is obtained by setting $\alpha = \pi - \beta$ in (3.12).

The computation of A follows from (4.1) with α replaced by $\pi - \beta$ and t replaced by ξ. Now the total field ϕ at a point P' in T differs from (4.3) in the following way. If we allowed an incoming wave at P', then this field could be continued back to L (Fig. 3) and whence into free space ($r > 1$). Clearly this field would not satisfy the radiation condition at infinity. Thus, we allow only an outgoing wave in T which is given by

$$\phi \sim G(\beta_o) e^{ik\psi_o}/\cosh w\,\xi_o(\tanh w\xi_o - 2\cos\beta_o/w)^{1/2}. \tag{8.3}$$

This result is valid for $\xi_c \leq \xi_o \leq \xi_e$, where ξ_c is given implicitly by $\tanh w\xi_c = (2/w)\cos\beta$. This solution becomes invalid as $\xi_o \to \xi_c$ and we are forced to perform a boundary layer analysis in a small region about the caustic. The analysis proceeds in precisely the same manner as on the irradiated side. We find

$$\phi \sim \exp(ik\psi_o)\exp(-i\tau_o^3/6g)\,[e_1 \mathrm{Ai}(-\eta_o e^{2\pi i/3}) + e_2 \mathrm{Ai}(-e^{-2\pi i/3}\eta_o)], \tag{8.4}$$

$$\eta_o = (\xi_o - \xi_c)^2 k^{2/3}/(4g)^{2/3} = \tau_o^2/(4g)^{2/3},$$

where $g = \rho_c^2/w^4$. Here ρ_c is the radius of curvature of the caustic at point Q which by symmetry is the same as at point P (see Fig. 3).

Letting $\tau_o \to \infty$, we match (8.4) into (8.3) as $\xi_o \to \xi_c$. We find that $e_2 = 0$ and

$$e_1 = k^{1/6} G(\beta_o) e^{-i\pi/12} 2^{7/6}\sqrt{\pi}/(w\rho_c)^{1/3}. \tag{8.5}$$

Hence, within the boundary layer we find

$$\phi \sim \frac{e^{ik\psi_o} k^{1/6} G(\beta_o)}{(w\rho_c)^{1/3}}\, e^{-i\pi/12} 2^{7/6} e^{-\tau_o^3/6g} A_i(-\eta_o e^{2\pi i/3}). \tag{8.6}$$

Now to find $G(\beta_o)$ we must continue (8.6) behind the caustic and match it into (7.2) as $n \to 0$. Doing so we obtain

$$G(\bar{\beta}) = |\cos\bar{\beta}|^{1/2} e^{-kr_o^2 w\pi/4}, \tag{8.7}$$

where $\bar{\beta} = \pi - \alpha$. With this choice of G, (8.6) smoothly connects (8.3) with (7.2). The artifice of introducing the initial data (8.1) for the eikonal equation led us conveniently to these results.

We now know the solution to leading order within the plasma column except in small regions about the polar caps and branch cut. It is amusing to note the analogy between this problem and a double turning point problem in ODE's. The irradiated portion has a WKB type of solution. The region $\mathcal{C}$ gives rise to exponentially decaying solutions while the back side of the caustic acts as a second turning point. The region T gives rise to exponentially small transmitted wave. Results analogous to ours for plane geometry are discussed in Ref. 5.

9. THE SCATTERED FIELD

We take up now the description of the field ϕ in the region exterior to the plasma. This is done by continuing the outgoing geometrical optics wave into free space ($r > 1$).

Consider an incident ray corresponding to the angle α. Let L denote the point with coordinates $(\cos\theta_e, \sin\theta_e)$, where the ray again intersects the circumference (see Fig. 4). It follows from (3.11) that α and θ_e are related by

$$\tan\theta_e = (2 - r_o^2)\tan\alpha/r_o^2. \tag{9.1}$$

We next let ψ_e denote the value of the outgoing phase at L. From (3.5), (3.12), and (9.1) we find

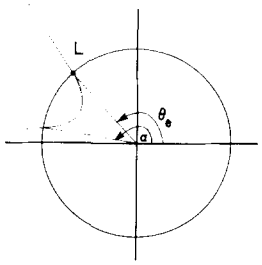


FIG. 4. The refraction angle θ_e.

$$\psi(l_e(\alpha), \alpha) = \psi_e = -wr_o^2 l_e/4. \tag{9.2}$$

To determine the outgoing amplitude A_e at this point, we combine (4.2) and (9.1). We thus obtain

$$A_e = [r_o e^{-i\pi/2}/(2-r_o^2)]\,[1 - 4\sin^2\theta_e(1-r_o^2)^2/(2-r_o^2)^2]. \tag{9.3}$$

The outgoing field at L is then $A_e \exp(ik\psi_e)$.

Now for $r > 1$ we have that $\mu_o^2 = 1$. The parametric solutions of the ray equation (3.2) are thus

$$\left.\begin{aligned} x &= 2p_e T + \cos\theta_e \\ y &= 2q_e T + \sin\theta_e \end{aligned}\right\}, \quad T \geq 0, \tag{9.4}$$

where p_e and q_e are the values of ψ_x and ψ_y respectively at L. We find by combining (3.1), (3.4), and (3.5) that

$$\begin{aligned} p_e &= \frac{2(2-r_o^2)\cos\theta_e}{r_o^2[r_o^4\tan^2\theta_e + (2-r_o^2)^2]^{1/2}} + \left(1 + \frac{16\cos^2\theta_e}{w^2 r_o^4}\right)^{1/2}, \\ q_e &= \frac{2\sin\theta_e}{[r_o^4\tan^2\theta_e + (2-r_o^2)^2]^{1/2}}. \end{aligned} \tag{9.5}$$

Equations (9.4) may be combined into a more aesthetic form. By eliminating T the ray equation becomes

$$y = \sin 2\theta_e(x - \cos\theta_e)/[(1-r_0^2) + \cos 2\theta_e] + \sin\theta_e \tag{9.6}$$

From this result and (9.1) one readily deduces that these rays cover the exterior of the plasma except for the shadow region $S = \{r > 1; x \geq 0; |y| < 1\}$. The rays are, of course, straight lines.

We shall now compute the phase function ψ along each ray. From the eikonal equation and (3.1) it follows that $\partial\psi/\partial T = 2\mu_0^2 = 2$. By combining this fact with (9.2) we obtain

$$\psi = T - wr_o^2 l_e/4. \tag{9.7}$$

Now, to complete the description of the field, we must determine the amplitude. This function must satisfy (4.1) with t replaced by T. We find that

$$A = A_e[J(0, \theta_e)/J(T, \theta_e)]^{1/2}, \tag{9.8}$$

where $J(T, \theta_e)$ is the Jacobian of (9.4). From (9.4) we obtain

$$J(T, \theta_e) = 2(p_e\cos\theta_e + q_e\sin\theta_e) + 4p_e^2 T(q_e/p_e)', \tag{9.9}$$

where the prime denotes differentiation with respect to θ_e.

By combining this result with (9.7) we find that

$$\phi \sim A_e\,|J(0, \theta_e)/J(T, \theta_e)|^{1/2} \exp[ik(8T - wr_o^2 l_e)/4]. \tag{9.10}$$

Now as $r \to \infty$ the scattered field (9.10) takes a simpler form. In this limit (9.4) reduces to $r^2 = x^2 + y^2 \approx 4T^2$. Thus (9.10) simplifies to

$$\phi \sim A_e\,|J(0, \theta_e)/2p_e^2(q_e/p_e)'|^{1/2}\, e^{ikr}/\sqrt{r}. \tag{9.11}$$

Thus ϕ satisfies the radiation condition.

Now these results are not valid in S. Here the theory of geometrical optics predicts that $\phi = 0$. However, we have shown in sec. 8 that an exponentially small field is present in T. Therefore, we may continue this solution into S by extending the outgoing rays into free space. Due to the symmetry of the model, these rays can be found by reflecting (9.6) about the y axis. The scattered field in S is then (9.10) with an additional factor of $\exp(-\pi r_o^2 w/4)$. In this formula α is replaced by $\pi - \beta$ and θ_e with $\pi - \theta_e$.

Our scattered fields are invalid in a small region about the boundaries of S. We shall not discuss this problem here.

10. A SPHERICAL MODEL

In this section we shall briefly discuss the scattering of a scalar plane wave off a spherical object whose index of refraction is given by (2.4). In this expression r^2 is now equal to $x^2+y^2+z^2$. Thus we study (2.1) with the same boundary conditions and initial data as stated in Sec. 2.

We again use the geometrical optics ansatz (2.5) and arrive again at eikonal and transport equations. We shall show that the solutions of these equations in the present case are almost identical with those obtained for the cylinder problem. Thus we obtain with little effort the scattering of a plane wave by a sphere.

To exhibit this connection we first defined the plane P_δ by

$$z\sin\delta - y\cos\delta = 0, \quad 0 \leq \delta \leq \pi. \tag{10.1}$$

The intersection of this plane with the sphere yields the circle C_δ. Let $(\cos\alpha, \cos\beta, \cos\gamma)$ be an arbitrary point on the irradiated portion of this circle's circumference. We note that $\cos^2\alpha + \cos^2\beta + \cos^2\gamma = 1$ and $\cos\delta\cos\beta \equiv \sin\delta\cos\gamma$ at this point.

Now an arbitrary ray impinging on C_δ has an initial direction $(1, 0, 0)$. Using this information and the relationships between α, β, γ, and δ, we find that (3.2) yields

$$\mathbf{x} = [\cos\alpha\cosh wt + (2/w)\sinh wt]\,\hat{\mathbf{x}} + (\sin\alpha\cosh wt)\,\hat{v}, \tag{10.2}$$

where $\hat{x} = (1, 0, 0)$ and $\hat{v} = (0, \sin\delta, \cos\delta)$. Comparing this result with (3.4), we deduce that the rays on C_δ (for any fixed δ) are identical to those described in Sec. 3. Since for each δ the rays envelope an ellipse, we find that (10.2) envelopes the caustic surface $x^2/r_o^2 + y^2 + z^2 = 1$.

Now the Jacobian J_s of (10.2) is readily deduced from its definition $J_s = \mathbf{x}_\delta \cdot (\mathbf{x}_t \times \mathbf{x}_\alpha)$. We find that $J_s = \sin\alpha\cosh wt \cdot J$, where J is given by (3.7). Thus we deduce that the geometrical optics amplitude differs from (4.2) by an additional factor of $(\mathrm{sech}\, wt)^{1/2}$. Furthermore, since $\mu_o^2 = (\mathbf{x}\cdot\mathbf{x} - r_o^2)$ has not changed from

the previous problem and $\psi(t=0)=\cos\alpha$, the phase is still given by (3.12).

Taking into account the additional factor in the amplitude, we see that the field on C_δ is essentially the same as in the cylinder problem. Moreover, we note that these results are independent of δ. This follows from the symmetry of the problem.

The field exterior to the sphere is obtained in a similar manner, and only slight modification are introduced. For brevity we state here the result for $r \gg 1$. We find that

$$\phi \sim \tfrac{1}{2} r_o A_e \left| J(0,\theta_e)/\cos\theta_e p_e^2 (q_e/p_e)' \right|^{1/2} e^{ikr}/r, \qquad (10.3)$$

where A_e, q_e, p_e, θ_e, and $J(0,\theta_e)$ are the same quantities defined in Sec. 9. This result is valid in the far field except in the shadow region.

PART II. THE B PROBLEM—AN ILLUSTRATIVE EXAMPLE

11. FORMULATION AND MAIN RESULTS

In this section we shall examine the scattering problem for a cylindrically confined cold plasma with a different incident polarization. We consider the case where the incident magnetic field is parallel to the generators of the cylinder. From the cold plasma equations[10] we find that the time harmonic magnetic field $\phi \exp(-iwt)$ remains in this direction and satisfies

$$\nabla^2\phi - (\nabla\phi\cdot\nabla\mu_o^2)/\mu_o^2 + k^2\mu_o^2\phi = 0, \qquad (11.1)$$

where μ_o^2 is given by (2.4). The coordinate axes, boundary condition, and restrictions are the same in the E problem.

We assume that ϕ is given by (2.5) and insert this ansatz into (2.1). We find that ψ must satisfy the *same* eikonal equation as studied in the E problem. The transport equation is now modified by the addition of the new term $(-\nabla\psi\cdot\nabla\mu_o^2/\mu_o^2)A$. Along a ray we find that A must satisfy

$$\mathcal{L}(A) = A_t + \frac{1}{2}\frac{J_t}{J}A - \frac{1}{2}\frac{\mu_{o,t}^2}{\mu_o^2}A = 0, \qquad (11.2)$$

where J is given by (3.7). Taking into account the unit value of the incident radiation and $J(t=0)$, we find that now

$$A = \left|\mu_o^2(\cos\alpha)/J\right|^{1/2}. \qquad (11.3)$$

From (11.3) we see that the amplitude becomes singular at the caustic. Thus we are forced to perform a boundary layer analysis. We find now that A must satisfy

$$ik\mathcal{L}(A) + \nabla^2 A - \nabla\mu_o^2\cdot\nabla A/\mu_o^2 = 0, \qquad (11.4)$$

where $\mathcal{L}(A)$ is defined in (11.2). Upon introducing the stretched variable τ as before, we find that the term $\nabla\mu_o^2\cdot\nabla A/\mu_o^2$ is unimportant except as $\alpha\to\pi$. In this limit the rays become arbitrarily close to the circle $r=r_o$ where μ_o^2 vanishes (see Fig. 1). Thus, if we a avoid this region, the boundary layer results are essentially the same as in the E problem.

The results of this problem are almost identical to those obtained in Part I. We need only insert the factor $|\mu_o|$ into the field to obtain the proper result. The exception occurs in an asymptotically small neighborhood of the negative x axis ($\alpha\to\pi$) within the plasma. This is the topic of the next section.

12. AN ANOMALY

We now conjecture that the outgoing geometrical optics wave undergoes a phase shift of π radius in this region. To exhibit this feature, we introduce the stretched variables $\tau = k^{1/3}(t-t_c)$ and $\xi = k^{1/3}(\pi-\alpha)$ into (11.4) and obtain

$$A_{\tau\tau} - (1/\tau)A_\tau - \gamma\tau A_\tau/2H + (i\tau/g)(\tau A_\tau + \tfrac{1}{2}A - \gamma\tau^2 A/4H) = 0, \qquad (12.1)$$

where H is given by $\gamma(\tau^2 + 4\xi^2/\gamma)/4$. The parameters γ and g are the same as in the E problem. If we let

$$A = \sqrt{H}\exp(-i\tau^3/6g) f(\eta,\xi), \qquad (12.2)$$

we find that f must satisfy

$$f_{\eta\eta} + [\eta - \tfrac{3}{4}(\eta+\tilde{\xi})^{-2}]f = 0, \qquad (12.3)$$

where $\eta = \tau^2/(4g)^{2/3}$ and $\tilde{\xi} = 4\xi^2/\gamma^2(4g)^{2/3}$.

To the best of our knowledge (12.3) cannot be solved in terms of tabulated functions when $\tilde{\xi}\neq 0$. However, we can show that (12.3) has the proper structure for matching purposes. This can be seen by letting either $\eta\to\infty$ or $\tilde{\xi}\to\infty$ and thus reducing (12.3) to Airy's equation. The problem lies in the fact that we cannot deduce from this asymptotic behavior the phase of the outgoing geometrical optics wave. This information is obtained by introducing the normal distance n into the explicit solutions of (12.3). We are unable to do so when $\tilde{\xi}\neq 0$.

If we now set $\tilde{\xi}=0$, we find that (12.3) has the solutions $\sqrt{\eta}\,J_{\pm 2/3}(2/3\,\eta^{3/2})$. We are now able to execute the matching scheme discussed in Sec. 5. We find that the phase shift is $+\pi/2$ radians as compared to $-\pi/2$ radians outside of this region. This fact has led us to our conjecture.

This conjecture naturally suggests that we must incorporate into (4.1) certain α derivatives. By introducing the new sketched variable $\xi^* = (\pi-\alpha)\sqrt{k}$ into (4.1) we find that A must satisfy

$$i\cosh^2 wt\,\mathcal{L}(A) = -A_{t^*t^*}. \qquad (12.4)$$

This equation is formally valid in strip of $O(1/\sqrt{k})$ about the negative x axis ($-1 \le x < r_o$). The solution of (12.4) is given formally by

$$A = \left|\mu_o^2/J\right|^{1/2}\int_{-\infty}^{\infty} h(s)e^{is\xi^* - is^2\tanh wt}\,ds. \qquad (12.5)$$

If this A were to represent the incoming amplitude, then $A(t=0)$ being equal to 1 would imply that $h(s)$ is the delta function. Thus the incoming amplitude would be unchanged. Unfortunately, the asymptotic behavior of the outgoing amplitude as $\xi^*\to\infty$ is not enough to determine its corresponding $h(s)$. This information must come from matching (12.5) with the boundary layer solution. We are unable to do so at the present time.

13. THE SCATTERED FIELD

We shall now investigate a consequence of this conjecture. As mentioned earlier, the results of this problem differ from those of the E problem by factor of $|\mu_o|$ "almost" everywhere. This factor is 1 in the region exterior to the plasma. Thus the scattered fields are identical except in a certain region. Due to the apparent phase shift in the outgoing geometrical optics wave, we expect the scattered field to undergo a similar transition in the wedgelike region $W=\{|\theta-\pi|<k^{-1/2},\ \theta=\tan^{-1}(y/x),\ x^2+y^2\geq 1\}$. An analgous statement can be made for the shadow region.

14. THE ELECTRIC FIELD

In closing this section, we shall compute the electric field E within the plasma. From Maxwell's equation and the assumed incident polarization we find that[10]

$$\mathbf{E}=(E_x,E_y,0)\sim(A/\mu_o^2)\,e^{ik\psi}(-q,p,0), \tag{14.1}$$

where A is given by (11.3) and $(p,q)=\nabla\psi$.

Now in the region $\mathcal{E}$ the phase function becomes complex and hence the electric field has an exponentially small factor. However, as r approaches r_o, the field becomes unbounded. On the circle $r=r_o$ we have from (2.2) that $\omega=\omega_p$, and the electrons resonate at the "plasma" frequency. This singularity can be removed by introducing a dissipative mechanism into the model.[2]

When this is done, we find[2] that $\mu_o^2(r_o)=ie\nu/\omega$, where ν is the collision frequency. Although ν/ω is extremely small in many applications, the field (14.1) is still $o(1)$ as $k\to\infty$. This is due to the fact that $\psi=R+iI$ within the region $\mathcal{E}$. thus the field is algebraically singular in ν but exponentially decaying in k. This result agrees with Ref. 2. in the case of planar geometry.

ACKNOWLEDGMENTS

The author would like to thank Professor J. B. Keller for his many helpful suggestions in the preparation and condensation of this work. This research was partially supported by grant from the Alfred P. Sloan Foundation.

APPENDIX

We shall now that the incoming polar cap solution matches into the incoming geometrical optics wave in the proper limit. To this end we let $u=-\tau_i=k\xi$, $\bar{s}=ku^\beta$, and $E=u^\beta I$ where $\frac{1}{2}<\beta<\frac{5}{8}$, k is a fixed constant, ξ is a parameter, and I is a new variable of integration for (6.6). Our entire analysis is asymptotic in the limit as $k\to\infty$. Thus we have $u\to\infty$ and we may replace the Airy function in (6.6) by its asymptotic formula[12]. We find that (6.6) reduces to

$$A^i\sim(u^{3\beta-1/2}/|\chi r_o^2 w^2|^{1/2})\int_0^\infty \sqrt{t}\ \exp(iu^{2\beta}\phi\,dt) \tag{A1}$$

where

$$\Phi=t^2/4\chi+\chi r_o^2\chi\, kl/1+o(1)\ \text{ as } k\to\infty.$$

In obtaining (A1) we have used (6.8).

We now apply the method of stationary phase[12] to the integral in (A1) and find that

$$A^i\sim|2\bar{s}/-w^2\tau_i|^{1/2}\,e^{iu^{2\beta}\Phi^*}. \tag{A2}$$

In this expression $\Phi^*=r_o^4\chi^3k^2/4$. We thus find that the incoming polar cap field reduces to

$$\phi_i^c\sim(2\bar{s}/-w^2\tau_i)^{1/2}\ \exp(ik\psi_i(1+\tau_i^{2\beta}\Phi^*/k\psi_i)) \tag{A3}$$

If we show that $\tau_i^{2\beta}\Phi^*/k\psi_i=O(1)$ as $k\to\infty$, then (A3) matches with (6.9). Now from (3.12) we deduce that $\psi_i\sim c\,\tau_i/k^{1/3}$ as $c\to\pi/2$ with some constant c.

Thus we have

$$\tau_i^{2\beta}\Phi^*/k\psi_i\sim k^{2\beta-5/3}(\Phi^*\xi/c)=o(1)\ \text{ as } k\to\infty, \tag{A4}$$

and our assertion is true.

[1] F.F. Chen, "A New Role for Infrared Lasers," UCLA Plasma Physics Group Report 118, 1972.
[2] F.F. Chen and R.B. White, "Amplification of Electromagnetic Waves in Overdense Plasma," UCLA Plasma Physics Group Report 126, 1972.
[3] N.G. Alexopoulos, J. Opt. Soc. Am. **9**, 1088 (1972).
[4] N.G. Alexopoulos, IEEE Trans. Antennas Propagation AP-**22**, 226 (1974).
[5] J.B. Keller and B.D. Seckler, J. Acoust. Soc. Am. **31**, 192 (1958).
[6] J.B. Keller, Proc. Symp. Appl. Math. **8**, 27 (1958).
[7] J.B. Keller and R.N. Buchal, Commun. Pure Appl. Math. **13**, 85 (1960).
[8] R.M. Lewis, N. Bleistein, and D. Ludwig, Commun. Pure Appl. Math. **20**, 295 (1967).
[9] G.A. Kriegsmann, "An Application of the Method of Geometrical Optics to the Scattering of Plane Electromagnetic Waves off Cylindrically Confined Cold Plasmas," Disseration, UCLA, Sept. 1974.
[10] J. Brandstatter, *An Introduction to Waves, Rays, and Radiation in Plasma Media* (McGraw-Hill, New York, 1963).
[11] A. Baños and D. Kelley, "Electromagnetic Wave Functions for Parabolic Plasma Density Profiles," Plasma Physics Group Report 125, 1972.
[12] E.T. Copson, *Asymptotic Expansions* (Cambridge U.P., Cambridge, 1967).
[13] E. Zauderer, SIAM J. Appl. Math. **19**, 575 (1970).

On a convergent nonlinear perturbation theory without small denominators or secular terms

Charles R. Eminhizer, Robert H. G. Helleman, and Elliott W. Montroll

Institute for Fundamental Studies, Department of Physics and Astronomy, University of Rochester, Rochester, New York 14627
(Received 13 June 1975)

We demonstrate a method for solving the dynamics of systems of extremely nonlinear coupled anharmonic oscillators with polynomial force laws. Several convergent perturbative and iterative techniques are described and applied, yielding the Fourier frequencies and coefficients of the time dependent solution. The usual stumbling block in a series analysis of coupled nonlinear *systems* is the appearance of terms with arbitrarily small denominators resulting in a rapidly divergent series. We avoid this classic "*small denominators problem*" by concentrating on the periodic solutions of a system. We employ the "Poincaré recurrence time" T_r of a system as the period of the solution, i.e., each Fourier frequency can be expressed as $n(2\pi/T_r)$, with some integer n. These periodic solutions are dense among all possible solutions and suffice for our practical calculations (just as the rational numbers suffice for most practical computations). Hence we construct the solutions with rationally *dependent* Fourier frequencies whereas the Kolmogorov–Arnold–Moser theorems demonstrate the existence of certain solutions whose frequencies are rationally *independent*. As an example we treat a two oscillator system equivalent to the "*Duffing equation*," a *driven* anharmonic oscillator appearing in models of the laser, the driven pendulum, etc. In our analysis no small denominators arise. We extend the region of bounded solutions beyond the (numerical) estimates in the literature and add to the list of peculiar phenomena exhibited by the Duffing equation.

1. INTRODUCTION

In several fields of science the basic laws or appealing models lead to *nonlinear* differential and integral equations. Yet, most successful theories and solvable models in physics and applied mathematics employ *linear* differential equations, or linear approximations, where *complete* solutions can be obtained as linear combinations of a set of basis functions. All other equations, to which this principle of "superposition" does not apply, are by definition *non*linear, and often yield new and unexpected phenomena, wherever we can solve them.

The classical style of dealing with systems of nonlinear differential equations has been through perturbation theory. Perturbation theory is encountered in mathematics, engineering, and many branches of physics, from celestial mechanics to quantum mechanics, from linear response theory to field theory. It saw a relative climax in Delaunay's seventh order calculation of the orbit of the moon, with one equation extending over 170 pages, in print. The subject has acquired a well-deserved bad reputation for yielding unbounded (divergent) expressions for some variables, otherwise known to be bounded according to some basic principles. Even in the perturbation solution of a *single* anharmonic oscillator terms appear proportional to powers of the time t, t^2, etc., the so-called "secular" terms. In practice one often truncates such series at a given order with the unfortunate consequence that the highest power of t which remains, automatically leads to an unbounded result as $t \to \infty$. During the 19th century, Stokes, Lindstedt, and Poincaré found a way to remove such terms and derived expansions uniformly valid for all t.[1–3] By systematizing their ideas we have developed, in Ref. 4, a series solution, inherently without secular terms, for an anharmonic oscillator and proved convergence for all t and a range of initial conditions. The frequency of oscillation, ν, was obtained from a perturbation expansion about the frequency ω_0 of the harmonic (linear) approximation to the equation of motion.

In the perturbation theory of *systems* of coupled nonlinear differential equations there exists a grave classical difficulty, *the problem of the "small denominators" or "small divisors."*[5–7] The denominators of terms in the series solution become arbitrarily small, due to nonlinear resonance effects, and the series diverges. It presents a major stumbling block for all theories on the dynamics of N bodies ($N \geq 3$) in classical, celestial, and quantum mechanics. It was not until 1957–1963 that *existence* proofs for some series solutions of certain classical N body problems were obtained, by Kolmogorov, Arnold, and Moser.[5–7] If we could make a Fourier analysis of their ("quasiperiodic") solutions, we would obtain N rationally *independent* Fourier frequencies of oscillation, i.e., $\nu_1, \ldots, \nu_N$ with $m_1\nu_1 + \ldots + m_N\nu_N \neq 0$, for all $m_1, \ldots, m_N$ (except all $m_k = 0$).

In this article we outline a method of *constructing* rapidly convergent series solutions for coupled anharmonic oscillators with polynomial force laws and N rationally *dependent* frequencies. In series analyses this case is usually ruled out explicitly since it yields the worst "resonances", making some of the above-mentioned denominators exactly zero.

Thus we are able to define the "Poincaré recurrence time" of a system and employ it as the period of our solutions with *rationally dependent frequencies*. Hence the period is not simply related to any physical time constant of the problem and is, in general, larger by many orders of magnitude. While we offer no *explicit* solution for the case where the Fourier frequencies are rationally *in*dependent, we can, in our present examples, approximate them arbitrarily well by means of our (periodic) solutions with dependent frequencies, in the same sense as one can approximate any irrational num-

ber arbitrarily well by means of rational numbers. The periodic solutions are in fact dense among all possible solutions of our problems[6,5] and suffice for our purposes here. We will discuss this further in Sec. 4.

These new solutions for systems of coupled anharmonic oscillators form the core of what we would like to communicate in this paper. Because of the need to first develop some technical devices, in Secs. 2 and 3, these systems are not introduced until Sec. 4. There we discuss some historical developments as well as the method used to remove the "small denominators" and the solutions with dependent frequencies. Section 4 can be read without a knowledge of Secs. 2 and 3.

In Sec. 5, we apply our method to a simple system of a coupled harmonic and anharmonic oscillator known to have serious "small denominator" problems, in the usual series analysis. This system is equivalent to the well known "*Duffing equation*," a driven anharmonic oscillator appearing, for example, in certain laser models,[8] in the theory of a pendulum driven by a periodic external force,[8–14] etc. The frequency of the driving force and the main response frequency of the anharmonic oscillator constitute the ν_2 and ν_1 for this problem. In our solutions none of the combination tones or harmonics give rise to a "small denominator" and the series converges. One of the main problems, besides constructing any solutions at all, is to determine the region of values for the amplitude B of the driving force for which the solution is bounded and to find this region as a function of the driving frequency ν_2 and other parameters (the potential energy of the anharmonic oscillator itself has the shape of a "valley on top of a mountain, and for certain B and ν_2 values the physical solution is unbounded). The region of bounded solutions we obtain, in Sec. 5, is considerably larger than the ones in the literature,[9] which were based on numerical analysis. The series solution also indicates why numerical analysis is likely to be unstable in the additional regions we find. In addition to the many peculiar and interesting effects known to exist in the Duffing equation[8–14] we obtain some *bounded* solutions which leave the potential "valley" and go down part of the slope of the "mountain" before returning to the potential valley. In the anharmonic oscillator without a driving force these would be unbounded.

In Secs. 2 and 3 we systematize two more ideas of Poincaré concerning the *existence* of some periodic solutions of a nonlinear differential equation. One embeds the equation of motion in a class of equations characterized by one parameter ϵ, such that for $\epsilon = 1$ one has the given equation of motion.[12] We choose the class of equations such that $\epsilon = 0$ yields a harmonic oscillator with *the same*, as yet unknown, *frequency* of oscillation ν as the solution of the *anharmonic* oscillator. Then we *construct* the solution for $\epsilon = 1$ using perturbation theory and the (consistency) requirement that ν remains unchanged. This device yields excellent "zeroth order" results leading to rapid convergence, even at arbitrarily high energy and nonlinearity. Moreover, it provides us with a quick qualitative insight into the behavior of extremely nonlinear systems of anharmonic oscillators. Typically it yields a "zeroth order" result within 5% of the value of the exact result. There is no trouble going as high as 14th order or higher on a digital computer. Typically this takes 10 seconds of computer time (360-65 IBM) and yields 10 significant digits.

In Sec. 2 we specify the initial conditions and calculate the "renormalized" frequency ν as part of the solution. We expand about a frequency ν_0, *different from the harmonic* frequency ω_0. This expansion has a finite region of convergence which can be located anywhere and thus include any possible initial condition (all our Fourier expansions converge for $0 \leq t < \infty$). In Sec. 3 we change our strategy and *specify* ν, the frequency of oscillation, while *calculating* the initial displacement and velocity via a perturbation series. One can think of this as specifying nonlinear action-angle variables and deriving a (local) transformation back to ordinary variables. We find that *all* bounded solutions are included in one and the same region of convergence of our iteration procedure [another region of convergence yields all (physically) unbounded solutions]. It is this idea that enables us, in Sec. 4, to *specify* rationally dependent frequencies $\nu_1, \ldots, \nu_N$, unaltered by higher order perturbation terms, and to obtain solutions for *systems* of anharmonic oscillators. The germ of this idea is present in the works of Poincaré[2] and of Kolmogorov, Arnold, and Moser,[5,6] which, in hindsight, should not have come as a surprise to us.

Larger systems of coupled anharmonic oscillators will be discussed elsewhere.[15]

2. FREQUENCY RENORMALIZATION OR A LOCALLY CONVERGENT PERTURBATION SCHEME

Consider the cubic anharmonic oscillator

$$\ddot{\phi} = -\omega_0^2\phi + R\phi^3, \quad \text{with } \phi(0) \neq 0,\ \dot{\phi}(0) = 0, \text{ and} \tag{2.1}$$

$$2H \equiv \dot{\phi}^2 + \omega_0^2\phi^2 - \tfrac{1}{2}R\phi^4 = \omega_0^2\phi^2(0) - \tfrac{1}{2}R\phi^4(0), \tag{2.2}$$

the conserved Hamiltonian. Notice that for $R > 0$ the potential has the shape of a "valley on top of a mountain" and that the solutions will be bounded only if $R\phi^2(0)/\omega_0^2 < 1$, cf. Fig. 1(a). For $R \leq 0$ the potential is convex, cf. Fig. 1(b), everywhere, and all solutions will be bounded. For $R > 0$ the restoring force is weaker than the harmonic force. Hence we expect the frequency of oscillation, ν, to be smaller than ω_0 (for $R < 0$, we expect $\nu > \omega_0$). After we treat this case with a cubic nonlinearity, the extension to *polynomial forces containing*

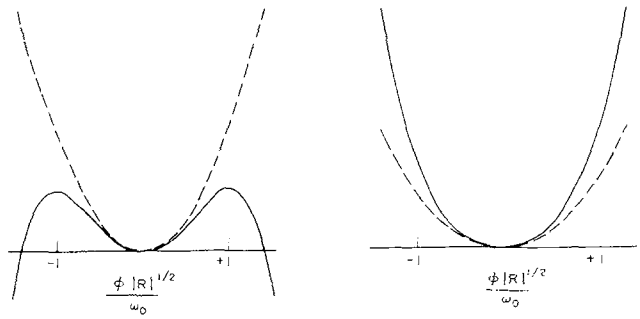


FIG. 1. (a) The shape of the potential energy for $R > 0$. The harmonic part of the potential is indicated by a dashed line. (b) The same for $R < 0$.

odd powers of ϕ will be clear.[16] If the force contains at least one even power of ϕ, the potential is no longer symmetric about $\phi=0$ and the convergence properties of our current schemes will depend more critically on "where we start" our perturbation series. This inconvenience is overcome automatically in another scheme,[16] and we will discuss the "even powered" systems elsewhere. Symmetric potentials are used here to introduce and compare our schemes.

We scale to unit initial displacement and unit harmonic frequency

$$\chi_{\tau\tau}=-\chi+\theta\chi^3, \quad \text{with } \chi(0)=1,\ \chi_\tau\equiv d\chi/d\tau,\ \chi_\tau(0)=0, \text{ and}$$
$$\chi\equiv\phi/\phi(0), \quad \tau\equiv\omega_0 t, \text{ and } \theta\equiv R\phi^2(0)/\omega_0^2; \tag{2.3}$$
$$\text{hence } |\chi(\tau)|\le 1. \tag{2.4}$$

All nonlinear properties and parameter choices can now be characterized by the single parameter θ. It is easy to see from the potential energy and from

$$\chi_{\tau\tau}\equiv-\chi(1-\theta\chi^2) \tag{2.5}$$

that the solutions will be bounded and oscillatory, for all τ, if

$$-\infty<\theta<+1 \tag{2.6}$$

and become unbounded, as $\tau\to\infty$, if $\theta>+1$. In Ref. 4, Secs. 3, 4, and 7, we have given a perturbation theory which converges and provides the exact solution of (2.1) and (2.3) for $|\theta|<+1$. In this section we present a perturbation scheme which converges in some neighborhood of any given θ, with $-\infty<\theta<+1$. (In the next section we obtain a scheme which includes all these θ values within *one and the same* region of convergence).

The theory of Ref. 4 is limited to $|\theta|<1$ as can be seen from the following argument: At $\theta=+1$ the physical solutions become unbounded and change from "periodic" to "real exponential." Therefore, the point $\theta=+1$ must be a singular point in any *periodic* Fourier series representation. Our previous perturbation theory produced Taylor series in θ about $\theta_0=0$ (Ref. 4, Sec. 7), whence the region of convergence $|\theta|<1$. In order to also include values of θ with $\theta<-1$, we must expand about a point $\theta=\theta_0$, where θ_0 is a negative number.

For each value of $\theta<+1$, the solution is periodic with some, as yet unknown, frequency ν. Let us call ν_0 the unknown frequency belonging to the above θ_0 about which we are going to expand. The most convenient way of doing this expansion is to embed the equation of motion (2.1) in a class of equations parametrized by ϵ, and with ν_0 as its main frequency of oscillation,

$$\ddot\phi+\nu_0^2\phi=\epsilon[(\nu_0^2-\omega_0^2)\phi+R\phi^3]. \tag{2.7}$$

We hope that in this way the "perturbation" on the rhs will be "small." Note that for $\epsilon=1$ (2.7) is identical to our Eq. (2.1). We have in effect "relocated" a ν_0^2 part of the linear term to the lhs of the equation of motion and we will call (2.7) the *relocated* version of (2.1). This is the first of three essential devices we introduce in this paper.

We now attempt to solve (2.7) with a θ near θ_0, and hopefully obtain a frequency ν near ν_0. Let us assume a solution of the form

$$\phi(t)=\sum_{j=0}^{\infty}\epsilon^j\phi_j(t)=\sum_{j=0}^{\infty}\epsilon^j\sum_n A_{j,n}\eta^n \quad \text{with } \eta\equiv\exp(i\nu t), \tag{2.8}$$

$$\nu=\nu_0+\epsilon\nu_1+\epsilon^2\nu_2+\cdots, \quad \text{whence } \nu^2=Q_0+\epsilon Q_1+\epsilon^2Q_2+\cdots, \tag{2.9}$$

with

$$Q_j=\sum_{p=0}^{j}\nu_p\nu_{j-p}, \quad j=0,1,\cdots, \qquad \epsilon=1, \tag{2.10}$$

$$A_{j,-n}=A_{j,n}^* \quad \text{and} \quad A_{j,n}\equiv 0, \quad \text{if } |n|>2j+1 \ \ (\text{or } j<0). \tag{2.11}$$

This particular form of the solution is easily obtained by rewriting (2.7) as an integral equation (cf. Sec. 3.4) and iterating the integral equation starting with $\phi_0(t)=\phi(0)\cos(\nu t)$. It is discussed at length in Secs. 2–4 of Ref. 4. A nonperturbative version, i.e., with ϕ a pure Fourier series in η, will be presented in Sec. 5. The initial conditions are met as follows

$$\phi_0(0)=\phi(0) \quad \text{and} \quad \dot\phi_0(0)=\dot\phi(0), \tag{2.12a}$$

while

$$\phi_j(0)=0 \quad \text{and} \quad \dot\phi_j(0)=0 \quad \text{for all } j\ge 1. \tag{2.12b}$$

For our particular initial conditions in (2.1), Eqs. (2.12a) and (2.11) yield

$$A_{0,1}+A_{0,-1}=\phi(0) \quad \text{and} \quad A_{0,1}-A_{0,-1}=0,$$
$$\text{whence } a\equiv A_{0,\pm1}=\tfrac{1}{2}\phi(0). \tag{2.13}$$

Let us now substitute our assumed solution (2.8)—(2.10) into the equation of motion (2.7) and equate like powers of ϵ and η. The latter is required if we wish to satisfy the equation for all t and a range of ϵ values [strictly speaking we only need the solution of (2.7) for $\epsilon=1$]. Equating the coefficients of $\epsilon^j\eta^n$ in (2.7), we obtain our *basic recursion formula*

$$(1-n^2)\nu_0^2A_{j,n}-\delta_{n,\pm1}aQ_j$$
$$=(1-\delta_{j,1})n^2\sum_{k=1}^{j-1}Q_{j-k}A_{k,n}+\Big[(\nu_0^2-\omega_0^2)A_{j-1,n}$$
$$+R\sum_{\mathcal{S}_3(j-1;n)}\cdots\sum A_{j_1,n_1}A_{j_2,n_2}A_{j_3,n_3}\Big], \tag{2.14}$$

where we have moved the lower order terms of ν^2 to the rhs, cf. (2.9), and where the summation proceeds over a set of six integers, i.e., all integers n_k and $j_k\ge 0$ such that[4]

$$\mathcal{S}_3(j;n)\equiv\{j_1,j_2,j_3;n_1,n_2,n_3\}:$$
$$j_1+j_2+j_3=j, \quad n_1+n_2+n_3=n, \quad \text{while } |n_k|\le 2j_k+1. \tag{2.15}$$

Therefore, all terms on the rhs of (2.14) are of order $j-1$ or less. All terms on the lhs of (2.14) are of order j, and (2.14) thus constitutes a recursion formula. For $j=0$ we find that only $A_{0,\pm1}$ can be different from zero. As an example we compute $A_{1,3}$ and Q_1. Consider $j=1$ and $n=3$ in (2.14).

$$(-8)\nu_0^2A_{1,3}=0+(0+RA_{0,1}^3),$$

so that

$$A_{1,3}=-Ra^3/8\nu_0^2. \tag{2.16}$$

Now let $j=1$ and $n=1$ in (2.14),

$$0 - aQ_1 = 0 + [(\nu_0^2 - \omega_0^2)A_{0,1} + R3A_{0,1}^2 A_{0,-1}],$$

so that

$$Q_1 = (\omega_0^2 - \nu_0^2) - 3Ra^2. \tag{2.17}$$

The factor 3 is the number of different permutations of $(+1, +1, -1)$ due to (2.15). So far we have not encountered the θ_0, about which we are expanding, which is mentioned in the argument leading to (2.7). The first order result for Q_1 (2.17) will provide a relation between this θ_0 and the parameter ν_0 used in (2.7). According to the argument leading to (2.7), we expect the best rate of convergence if $\nu^2 \approx \nu_0^2$, i.e., if $Q_1 \approx 0$ to the present (first) order. Since the value of ν_0 is still arbitrary and since ν_0 appears in the expression for Q_1 (2.17), we will *choose* ν_0 *such that* Q_1 *becomes identically zero.* Hence

$$\nu^2 \approx \nu_0^2 = \omega_0^2 - 3Ra^2 = \omega_0^2(1 - \tfrac{3}{4}\theta). \tag{2.18}$$

We defined ν as the frequency corresponding to an "initial condition" θ [θ contains the initial displacement, cf. (2.3)] and ν_0 as the frequency corresponding to a θ_0. In Eq. (2.18) we *set* $\nu_0 \approx \nu$ (to first order). Therefore (to first order) we also have $\theta_0 \approx \theta$. Thus (2.18) provides us with a convenient estimate of how to choose the parameter ν_0 so as to have a good rate of convergence, i.e., so as to have an expansion about a θ_0 *close* to the given value of θ. In the following we will only use the ν_0 and not make explicit use of the corresponding θ_0.

At the same time (2.18) yields an estimate of $\nu = \nu(\theta)$, which is independent of ν_0 and therefore a good approximation over a wide range of θ values, cf. (3.13).

A. Initial conditions

We see that for $n=1$ the forefactor in (2.14) becomes zero and that the $\{A_{j,\pm 1}\}$ are *not* determined by (2.14). The $\{A_{j,\pm 1}\}$ are determined by the initial conditions (2.12b). The condition $\phi_j(0)=0$ means $\sum_n A_{j,n}=0$, or

$$A_{j,1} + A_{j,-1} = -\sum_{\substack{n=-2j-1 \\ n\neq\pm 1}}^{2j+1} A_{j,n} \quad (j \geq 1). \tag{2.19}$$

The condition $\dot{\phi}_j(0)=0$ means $\sum_n \sum_{k=0}^{j} n\nu_{j-k}A_{k,n}=0$, or

$$A_{j,1} - A_{j,-1} = -\left(\sum_{\substack{n=-2j-1 \\ (k;n)\neq(j;\pm 1)}}^{+2j+1} \sum_{k=0}^{j} n\nu_{j-k}A_{k,n}\right) \Big/ \nu_0 \quad (j \geq 1). \tag{2.20}$$

Having obtained all Q's, whence all ν's, cf. (2.9), and all A's up to and including order j (except $A_{j,\pm 1}$), we can solve for the $A_{j,\pm 1}$ from (2.19)–(2.20). The case we consider here, i.e., $\dot{\phi}(0)=0$, is particularly simple: Since all $A_{j,n}$ will be real, cf. (2.13), (2.14), and (2.19), the $A_{j,\pm 1}$ will be identical and can be determined from (2.19). If we let $j=1$, we find

$$2A_{1,1} = -2A_{1,3}$$

so that

$$A_{1,1} = Ra^3/8\nu_0^2 = (a\theta/32)(\omega_0/\nu_0)^2. \tag{2.21}$$

Coefficients with even n cannot result from (2.14) due to the odd power of the nonlinear term.

B. The rate of convergence

We get a first indication of the rate of convergence by comparing the first and zeroth order terms

$$\frac{|A_{0,1}|}{|A_{1,1}| + |A_{1,3}|} = \frac{16}{|\theta|}\left(\frac{\nu_0}{\omega_0}\right)^2 = 16\frac{(1-3\theta/4)}{|\theta|}, \tag{2.22}$$

where (2.18) was used to obtain the last equality. This "rate of convergence" is 4 at $\theta=+1$ and increases to ∞ at $\theta=0$. From infinity (2.22) decreases to 12 as $\theta \to -\infty$. Remember that we obtain these good rates when we choose a *new* value of ν_0 for each new value of θ (using (2.18)). We will see that we are then near the center of the region of convergence of the series with that value of ν_0. However, if ν_0 is fixed and θ is varied we find a "rate" which, in general, is worse than (2.22). Sometimes it is convenient to keep ν_0 fixed since the *coefficients* of our series in θ do not have to be recalculated for each new θ value in that case.

C. Quantitative results

Equations (2.14) and (2.19)–(2.20) constitute the algorithm for obtaining the solution (2.8)–(2.11) of the equation of motion (2.7) or (2.1). With the aid of an electronic computer we evaluate the algorithm to 14th order in ϵ.

For each choice of ν_0 we then find a θ interval on which the series converges numerically. In Fig. 2 we extend each interval to a circle in the *complex* θ plane, with the interval as its diameter. The circle drawn with a thick line indicates the region of convergence for $\nu_0/\omega_0 = 2.0$. The data for the points A through E, evaluated with $\nu_0=2.0$, is displayed in Table I. For points A and B the calculated frequency ν (at 14th order) is approximately equal to ν_0, and we find a rapid rate of convergence [the θ for point A is chosen so as to satisfy our estimate (2.18) of $\nu=\nu(\theta)$]. We indicate a measure of the rate of convergence for the (largest) Fourier coefficient $A_{j,1}$, at $j=4,9$, and 14 and observe that $A_{j,1}/A_{0,1}$ has decreased to $\approx 10^{-11}$ at $j=14$ in both cases. This is entirely consistent with our expectations near (2.7) and (2.18). As we move to points further from the center, such as C, this rate of convergence slows until near the edges, points D and E, the ratio is nearly constant.

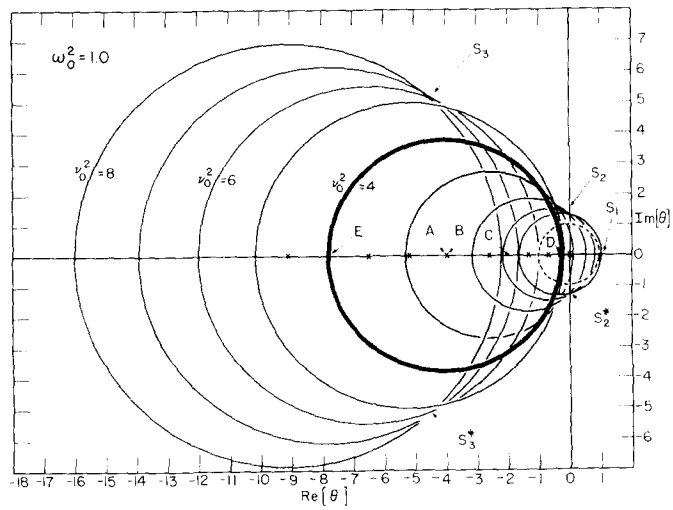


FIG. 2. Regions of convergence in the complex theta plane, of a forward perturbation scheme.

TABLE I. A comparison between the results for one region of convergence (cf. Fig. 2).

Data points cf Fig. 2	Chosen parameters[a,b] ($\omega_0^2=1.0$)	Order of trunc.[c]	ν^2 ($\omega_0^2=1.0$)[d] cf. (2.9)	Max[$\Delta H(t)/H(t)$] over one period	Max[relative error]over one period[e]	$A_{j,1}/A_{0,1}$[f]	Computing time[g]
A	$\theta=-4.0$	$j=4$	3.9047	2×10^{-4}	1×10^{-3}	-3.6×10^{-5}	0.15 s
ν_0^2 and θ	$\nu_0^2=4.0$	l.c.	$\approx-1.6\times10^{-3}$				
as in (2.18)	$A_{0,1}=1.0$	$j=9$	$3.904\,6408$	7×10^{-8}	7×10^{-7}	-4.0×10^{-8}	2.0 s
		l.c.	≈0.0				
	$E_{>2}/E=0.667$	$j=14$	$3.904\,640\,671\,089$	4×10^{-12}	3×10^{-10}	-1.1×10^{-11}	11.0 s
		l.c.	$\approx-2.2\times10^{-10}$				
B	$\theta=-3.96$	$j=4$	3.8759	2×10^{-4}	1×10^{-3}	-4.8×10^{-5}	[all
Near A	$\nu_0^2=4.0$	l.c.	$\approx-1.5\times10^{-3}$				$j=14$
	$A_{0,1}=0.995$	$j=9$	$3.875\,840\,43$	7×10^{-8}	-3×10^{-7}	-4.2×10^{-8}	have
		l.c.	$\approx1.0\times10^{-7}$				the
	$E_{>2}/E=0.664$	$j=14$	$3.875\,840\,33\,104$	4×10^{-11}	3×10^{-10}	-1.9×10^{-11}	same
		l.c.	$\approx-2.1\times10^{-10}$				comp.
C	$\theta=-2.0$	$j=4$	2.4644	6×10^{-3}	6×10^{-3}	-1.0×10^{-3}	time,
Midradius	$\nu_0^2=4.0$	l.c.	$\approx-3.4\times10^{-3}$				etc.
	$A_{0,1}=0.707$	$j=9$	$2.462\,125$	9×10^{-5}	-1×10^{-4}	-1.6×10^{-5}	for
		l.c.	$\approx4.1\times10^{-5}$				$j=4$
	$E_{>2}/E=0.50$	$j=14$	$2.462\,094$	2×10^{-6}	-3×10^{-6}	3.2×10^{-7}	and
		l.c.	$\approx-7.1\times10^{-7}$				$j=9$]
D	$\theta=-0.2$	$j=4$	1.1494	1×10^{-2}	1×10^{-2}	-5.7×10^{-4}	
Lower edge	$\nu_0^2=4.0$	l.c.	$\approx-1.2\times10^{-4}$				
	$A_{0,1}=0.224$	$j=9$	1.1492	2×10^{-3}	2×10^{-3}	-1.1×10^{-4}	
		l.c.	$\approx2.2\times10^{-5}$				
	$E_{>2}/E=0.091$	$j=14$	$1.149\,194$	4×10^{-4}	3×10^{-4}	-2.0×10^{-5}	
		l.c.	$\approx-4.0\times10^{-6}$				
E	$\theta=-7.8$	$j=4$	6.54	2×10^{-1}	-2×10^{-1}	2.5×10^{-2}	
Upper edge	$\nu_0^2=4.0$	l.c.	$\approx-2.0\times10^{-1}$				
	$A_{0,1}=1.40$	$j=9$	6.72	6×10^{-1}	-3	-2.1×10^{-2}	
		l.c.	$\approx-1.5\times10^{-1}$				
	$E_{>2}/E=0.80$	$j=14$	6.53	3×10^{-1}	3	2.9×10^{-2}	
		l.c.	$\approx-2.1\times10^{-1}$				

[a] $A_{0,1}=\phi(0)/2$.

[b] At $t=0$, the kinetic energy is zero; $E_{>2}/E$ is the ratio of the nonquadratic potential energy and the total potential energy at $t=0$.

[c] l.c.: last correction $\sum_n A_{j,n}$.

[d] Underlined numbers mean questionable precision in view of higher-order corrections or machine "round off." The machine allows 14—16 significant digits.

[e] We substitute the solution up to order j into (2.1) and we take all terms to the rhs and define: relative error(t) $\equiv$ rhs (t)/$[\nu_0^2-\omega_0^2]\phi(t)$ when $\phi(t)\neq0$. We scale in $[\nu_0^2-\omega_0^2]\phi(t)$ because it is characteristic of the magnitude of the terms in the equation, cf. (2.7).

[f] This is a measure of the rate of convergence, since $A_{j,1}$ is the largest coefficient at jth order.

[g] Computation time on IBM 360-65, in seconds.

(Notice that although D and E are symmetric about $\theta=-4.0$, there is a difference in the rate of convergence.)

The dotted circle has radius 1, and is centered at $\theta=0$. It is the region of convergence of the perturbation theory for $\nu_0\equiv\omega_0$, given in Ref. 4. We see from Fig. 2, that several circles intersect in approximately the same points S_1, S_2, and S_3. These are in fact the complex θ values for which the solution will be singular. For the subject cubic force, their values could be derived from the solution in closed form, i.e., the Jacobi elliptic cosine, cf. Ref. 4, Appendix B. While the center of the region of convergence is roughly determined by our choice of ν_0, we find that *the radius* of the region is determined also by the location of the nearest singularity of ϕ, which is not known *a priori*, for our series solutions.

The singularities of the solutions could however be found from the singularities of the equations, cf. the arguments leading to Eq. (7.10) of Ref. 4.[12] Figures like Fig. 2 might actually be used as a quick graphical method to locate some singularities of the solution in the complex θ plane, especially for those force laws where the solution is not easily obtainable in closed form.

When the parameter ν_0 is chosen according to our estimate in (2.18) we find, in Table II, that the rate of convergence remains good for all θ values we checked, i.e., $-10^{+5}\leq\theta\lesssim+0.8$. The first order estimate of the rate of convergence, given in (2.22), is consistent with the values of $A_{j,1}/A_{0,1}$ given in Table II. We see in Table II that *the calculated value of* ν^2 ($j=14$) *is a nearly linear function of* θ and observe that the $\nu^2(\theta)$ is virtually identical to the linear result we derived at *first* order, i.e., (2.18). We have this linear dependence from $\theta\approx+0.8$ to as far as we checked, i.e., $\theta=-10^5$.

We see, in Fig. 3, that a deviation sets in at $\theta\approx+0.8$ and that ν^2 bends over to 0 as $\theta\to+1$. The data points denoted by solid dots in Fig. 3 were obtained using the

TABLE II. Results over a range of energies and nonlinearities.

Chosen parameters ($\omega_0^2=1.0$)[a,b]	Order of trunc.[c]	ν^2[d] cf. (2.9)	max[$\Delta H(t)/H(t)$] over one period	max[relative error] over one period[e]	$A_{j,1}/A_{0,1}$[f]	Computing time[g]
$\theta=0.8$	$j=4$	.36001	5×10^{-3}	2×10^{-2}	-5.8×10^{-4}	0.15 s
$\nu_0^2=0.40$	l.c.	$\approx-2.5\times10^{-3}$				
	$j=9$	.359 632	8×10^{-5}	-5×10^{-4}	-2.1×10^{-5}	2.0 s
$A_{0,1}=0.447$	l.c.	≈0.0				
	$j=14$	.359 620 577	1×10^{-8}	9×10^{-6}	-1.8×10^{-7}	11.0 s
$E_{>2}/E=-0.67$	l.c.	$\approx-3.6\times10^{-7}$				
$\theta=-0.5$	$j=4$	1.370 729	-2×10^{-4}	2×10^{-5}	-1.3×10^{-7}	[all $j=14$ have the same comp. time, etc. for $j=4$ and $j=9$]
$\nu_0^2=1.375$	l.c.	$\approx-9.3\times10^{-6}$				
	$j=9$	1.370 729 241	6×10^{-6}	-9×10^{-11}	-4.5×10^{-12}	
$A_{0,1}=0.354$	l.c.	≈0.0				
	$j=14$	1.370 729 241 035 541	$<10^{-15}$	$<10^{-15}$	$<10^{-15}$	
$E_{>2}/E=0.20$	l.c.	$\approx-5.4\times10^{-17}$				
$\theta=-10$	$j=4$	8.2179	5×10^{-4}	6×10^{-3}	-6.9×10^{-5}	
$\nu_0^2=8.5$	l.c.	$\approx-6.3\times10^{-3}$				
	$j=9$	8.217 626	4×10^{-7}	-4×10^{-6}	-1.7×10^{-7}	
$A_{0,1}=1.581$	l.c.	≈0.0				
	$j=14$	8.217 625 6733	1×10^{-11}	5×10^{-9}	-1.1×10^{-10}	
$E_{>2}/E=0.83$	l.c.	$\approx-4.6\times10^{-9}$				
$\theta=-10^3$	$j=4$	7.189×10^2	1×10^{-3}	4×10^{-3}	-1.1×10^{-4}	
$\nu_0^2=751.0$	l.c.	$\approx-9.2\times10^{-1}$				
	$j=9$	$7.188\,16\times10^2$	2×10^{-6}	-9×10^{-6}	-5.3×10^{-7}	
$A_{0,1}=15.81$	l.c.	≈0.0				
	$j=14$	$7.188\,1569\times10^2$	8×10^{-10}	3×10^{-8}	-6.1×10^{-10}	
$E_{>2}/E=0.998$	l.c.	$\approx-2.3\times10^{-6}$				
$\theta=-10^5$	$j=4$	7.1784×10^4	1×10^{-3}	4×10^{-3}	-1.1×10^{-4}	
$\nu_0^2=75001.0$	l.c.	$\approx-9.2\times10^1$				
	$j=9$	$7.177\,808\times10^4$	2×10^{-6}	-9×10^{-6}	-5.4×10^{-7}	
$A_{0,1}=1.58\times10^2$	l.c.	≈0.0				
	$j=14$	$7.177\,8046\times10^4$	5×10^{-9}	3×10^{-8}	-6.2×10^{-10}	
$E_{>2}/E=1.0-2\times10^{-5}$	l.c.	$\approx-2.3\times10^{-4}$				

[a] $A_{0,1}=\phi(0)/2$.

[b] At $t=0$, the kinetic energy is zero, $E_{>2}/E$ is the ratio of the nonquadratic potential energy and the total potential energy at $t=0$.

[c] l.c.: last correction $\sum_n A_{j,n}$.

[d] Underlined numbers mean questionable precision in view of higher-order corrections or machine "round off." The machine allows 14–16 significant digits.

[e] We substitute the solution up to order j into (2.1) and we take all terms to the rhs and define the relative error(t) $\equiv$ rhs $(t)/[\nu_0^2-\omega_0^2]\phi(t)$ when $\phi(t)\neq0$. We scale in $(\nu_0^2-\omega_0^2)\phi(t)$ because it is characteristic of the magnitude of the terms of the equation.

[f] This is a measure of the rate of convergence, since $A_{j,1}$ is the largest coefficient at jth order.

[g] Computation time on IBM 360-65, in seconds.

present scheme to 14th order. Additional points, denoted by ×, were obtained using the methods of Secs. 3 and 5, and of Appendix A of Ref. 16. (Evaluation with the present scheme and Appendix A of Ref. 16 is possible but requires more computer time.)

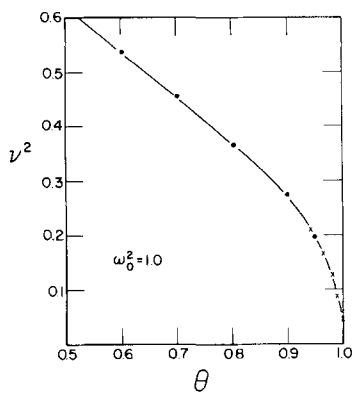


FIG. 3. ν^2 versus θ near $\theta=1$.

Our particular anharmonic oscillator example (2.1) has a solution in closed form, namely the Jacobi elliptic cosine function.[4] This in turn requires a numerical algorithm for its evaluation because of the elliptic integrals involved. The Fourier coefficients are more easily obtained using our procedure since a Fourier analysis of the elliptic functions is a cumbersome procedure.[9]

However, our procedure also applies when no solutions in closed form are available.[16]

3. AMPLITUDE RENORMALIZATION OR A PERTURBATION SCHEME WITH A LARGE REGION OF CONVERGENCE

Here we introduce the *"backward scheme,"* in which we *specify* the frequency of oscillation, ν, (and phase angle δ) and *derive* the initial displacement $\phi(0)$ [and velocity $\dot{\phi}(0)$]. The region of convergence includes all ν values with $0<\nu^2<\infty$, or in terms of the θ of (2.6),

all θ values with $+1>\theta>-\infty$.[17] While this is a considerable improvement over the previous "forward" schemes, the "backward" scheme is far more important to us for another reason. It enables us to obtain *convergent solutions for systems of coupled anharmonic oscillators*, where previous schemes yield asymptotic and strongly divergent, solutions.[4-6] The generalization to systems is made in Secs. 4 and 5 and subsequent papers. Here we introduce a perturbative version of the scheme on our usual single anharmonic oscillator

$$\ddot{\phi}=-\omega_0^2\phi+R\phi^3, \quad \text{with} \quad \phi(0)\neq 0 \text{ and } \dot{\phi}(0)=0, \tag{3.1}$$

Again we rewrite this as

$$\ddot{\phi}+\nu^2\phi=\epsilon[(\nu^2-\omega_0^2)\phi+R\phi^3], \quad (\epsilon=1) \tag{3.2}$$

the "relocated" version of (3.1), cf. (2.7). However, we now keep the ν constant and "renormalize" $\phi(0)$. This amplitude renormalization in a backward scheme is the second of three essential devices introduced in this paper. In other words, we specify nonlinear action-angle variables δ, νt and derive a transformation back to $\mathring{\phi},\phi$ variables. It is our contention that, for *systems* of nonlinear oscillators with rationally dependent frequencies ν_k, cf. Secs. 4 and 5, this *back* transformation has a large region of convergence while the corresponding forward transformations, from $\mathring{\phi}$, ϕ to δ, νt variables, are asymptotic in most cases. The latter part is, of course, well known.[6,5,4]

We attempt a solution of the form

$$\phi(t)=\sum_{j=0}^{\infty}\epsilon^j\sum_n A_j(n)\eta^n \quad \text{with } \eta\equiv\exp[i(\nu t+\delta)], \tag{3.3}$$

$$A_j(-n)\equiv A_j(+n), \ \textit{real},$$

$$A_j(n)\equiv 0 \quad \text{if } |n|>2j+1 \quad (\text{or } j<0). \tag{3.4}$$

In Sec. 5 we shall describe a nonperturbative version in which we use a pure Fourier series without ϵ^j powers. As before, we substitute the series (3.3) in Eq. (3.2), equate the coefficients of $\epsilon^j\eta^n$, and find, for $j\geq 0$, our *basic recursion formula*

$$(1-n^2)\nu^2A_j(n)=(\nu^2-\omega_0^2)A_{j-1}(n)+R\sum\cdots\sum_{\mathcal{S}_3(j-1;n)}A_{j_1}(n_1)A_{j_2}(n_2)A_{j_3}(n_3). \tag{3.5}$$

The set of integers $\mathcal{S}_3$ is defined in (2.15). This recursion formula is simpler than the previous one, (2.14). Also note that the phase angle δ does not enter into these recursion relations. For $j=0$ we see that only the $A_0(\pm 1)\equiv a$ are allowed to be different from zero. For ($j=1$ and) $n=1$ we find the "*resonance condition*"

$$0=(\nu^2-\omega_0^2)A_0(1)+R[3A_0^2(1)A_0(-1)]; \tag{3.6}$$

hence

$$\nu^2-\omega_0^2=-3Ra^2 \quad \text{or} \quad a=[(\omega_0^2-\nu^2)/3R]^{1/2}, \tag{3.7}$$

using (3.4). Hence the value of $A_0(1)$ is now *derived*, at $j=1$. Notice that there are *no* solutions (periodic with frequency ν) for $\nu>\omega_0$ if $R>0$ or for $\nu<\omega_0$ if $R<0$, physically an obvious result, cf. Fig. 1, which in previous schemes required a greater effort, cf. Ref. 4, Eq. (7.18).

The above procedure for $n=1$ is an essential one; otherwise the rhs of (3.2) would have a Fourier component of frequency ν, the same as the harmonic oscillator on the lhs. This would lead to resonance and secular terms, i.e., divergencies.[4] Therefore, we refer to (3.5) with $n=1$ as the "resonance condition."

A. Rate of convergence

As in Sec. 2B we compare magnitudes of the first and zeroth order terms. For $j=1$ and $n=3$ in (3.5) we find

$$A_1(3)=-Ra^3/8\nu^2=a(\nu^2-\omega_0^2)/24\nu^2. \tag{3.8}$$

For $j=2$ and $n=1$ we have the resonance condition

$$0=(\nu^2-\omega_0^2)A_1(1)+R[6A_1(1)A_0(1)A_0(-1)+3A_1(-1)A_0^2(1)+3A_1(3)A_0^2(-1)].$$

From this we *derive*, at $j=2$, the value of $A_1(1)$,

$$A_1(1)=Ra^3/16\nu^2=-a(\nu^2-\omega_0^2)/48\nu^2, \tag{3.9}$$

where we used (3.4)—(3.8). A measure of the "rate of convergence" can now be obtained from

$$|A_0(1)|/\{|A_1(1)|+|A_1(3)|\}=16|\nu^2|/|\nu^2-\omega_0^2|. \tag{3.10}$$

This rate of convergence is infinite for $\nu=\omega_0$. While it decreases as $\nu\to\infty$, it remains larger than 16. It also decreases as $\nu\to 0$ but stays larger than 16 for $\nu^2/\omega_0^2>\frac{1}{2}$. This may serve as a first indication of the high "rate of convergence," which we find in the backward scheme over most of the region of convergence, i.e., $0<\nu^2<\infty$ here. Also it shows why we prefer to use additional "convergence—acceleration" techniques in the region $\nu^2\ll\omega_0^2$.[16]

B. Initial conditions

Here we calculate the initial conditions $\phi(0)$ and $\dot{\phi}(0)$, the end products of a backward scheme, from the given values of ν and δ. Our series solution (3.3) yields

$$\phi(0)=\sum_{j=0}^{\infty}\sum_n A_j(n)\exp(in\delta)=\sum_{j=0}^{\infty}\sum_{n=-2j-1}^{2j+1}A_j(n)\cos(n\delta), \tag{3.11}$$

$$\dot{\phi}(0)=\sum_{j=0}^{\infty}\sum_n in\nu A_j(n)\exp(in\delta)=-2\sum_{j=0}^{\infty}\sum_{n=1}^{2j+1}n\nu A_j(n)\sin(n\delta), \tag{3.12}$$

where we used $\epsilon=1$. In the previous forward scheme, only the $j=0$ terms, with ν_0, were present, cf. (2.12)—(2.13). Let us first consider the case $\delta=0$, when (3.11), (3.12) yield $\dot{\phi}(0)=0$ and $\phi(0)=\sum_j\sum_n A_j(n)$. Knowing that the combined first order terms are at least 16 times smaller than the zeroth order ones in the region $\frac{1}{2}\lesssim\nu^2/\omega_0^2<+\infty$, we have to have a good approximation $\phi(0)\approx 2A_0(1)\equiv 2a$. In the same approximation we calculate the θ of the previous section: $\theta\equiv R[\phi(0)/\omega_0]^2\approx 4R[a/\omega_0]^2$. Hence, using (3.7), we find

$$\nu^2\approx\omega_0^2(1-\tfrac{3}{4}\theta) \tag{3.13}$$

to be a good approximation over a wide range of ν values, i.e., $\frac{1}{2}\lesssim\nu^2/\omega_0^2<+\infty$, cf. (2.18). Table II and Fig. 3 show this range, after 14 iterations, to be about

TABLE III. Results of the "backward scheme" over a wide range of energies and nonlinearities.[a]

Chosen parameters ($\omega_0 = 1.0$)[b]	Order of trunc.[c]	θ cf. (3.12)—(3.13) and (2.4)[d]	max[$\Delta H(t)/H(t)$] over one period	max[relative error] over one period[e]	$\Delta A_j(1)/A(1)$ [f]	Computing time[g]
$\nu^2 = 0.359\,620\,577$	$j=0$	$8.\underline{55}\times10^{-1}$, cf. (3.7)				0.0 s
	$j=4$	$8.\underline{0014}\times10^{-1}$	9×10^{-4}	2×10^{-3}	-6×10^{-4}	0.2 s
$E_{>2}/E = -0.67$	l. c.	$\approx 8.\overline{4}\times10^{-4}$				
	$j=9$	$7.999\,99\underline{8}\times10^{-1}$	3×10^{-7}	-8×10^{-7}	2×10^{-7}	0.9 s
Comparable but	l. c.	$\approx -3.\overline{2}\times10^{-7}$				
not identical to first	$j=14$	$7.999\,998\,9907\times10^{-1}$	1×10^{-10}	3×10^{-10}	-1×10^{-10}	2 s
entry in Table II	l. c.	$\approx -6.\overline{2}\times10^{-10}$				
$\nu^2 = 7.177\,8046\times10^4$	$j=0$	$-9.\underline{57}\times10^4$, cf. (3.7)				[all
	$j=4$	$-9.999\,9\underline{49}\times10^4$	7×10^{-5}	8×10^{-5}	-3×10^{-5}	$j=14$
$E_{>2}/E = 1.0-2\times10^{-5}$	l. c.	$\approx -6.1\times\overline{10}^{-2}$				have
	$j=9$	$-9.999\,999\,\underline{49}\times10^4$	4×10^{-10}	5×10^{-10}	-2×10^{-10}	the
Comparable but not	l. c.	$\approx -1.\overline{3}\times10^{-4}$				same
identical to last	$j=14$	$-9.999\,999\,880\,671\,\underline{20}\times10^4$	4×10^{-15}	3×10^{-15}	-1×10^{-15}	comp.
entry in Table II	l. c.	$\approx -1.7\times10^{-9}$				time,
						etc.
$\nu^2 = 10^8$	$j=0$	$-1.\underline{333}\times10^8$, cf. (3.7)				for $j=4$, 0
	$j=4$	$-1.\underline{393}\,197\times10^8$	7×10^{-5}	8×10^{-5}	-3×10^{-5}	and $j=9$]
$E_{>2}/E = 1.0-1\times10^{-8}$	l. c.	$\approx 7.1\times\overline{10}^3$				
	$j=9$	$-1.393\,203\,915\,\underline{07}\times10^8$	4×10^{-10}	5×10^{-10}	-2×10^{-10}	
	l. c.	$\approx -2.9\times10^{-2}$				
	$j=14$	$-1.393\,203\,915\,116\underline{2}\times10^8$	4×10^{-15}	3×10^{-15}	-1×10^{-15}	
	l. c.	$\approx -2.5\times10^{-6}$				
$\nu^2 = 10^{14}$	$j=0$	$-1.\underline{333}\times10^{14}$, cf. (3.7)				
	$j=4$	$-1.393\,\underline{196}\times10^{14}$	7×10^{-5}	8×10^{-5}	-3×10^{-5}	
	l. c.	$\approx -7.\overline{1}\times10^9$				
$E_{>2}/E = 1.0-1\times10^{-14}$	$j=9$	$-1.393\,203\,929\,\underline{64}\times10^{14}$	4×10^{-10}	5×10^{-10}	-2×10^{-10}	
	l. c.	$\approx -3.7\times10^4$				
	$j=14$	$-1.393\,203\,929\,685\underline{7}\times10^{14}$	4×10^{-15}	3×10^{-15}	-1×10^{-15}	
	l. c.	≈ -2.5				
$\nu^2 = 10^{20}$	$j=0$	$-1.\underline{333}\times10^{20}$, cf. (3.7)				
	$j=4$	$-1.393\,\underline{196}\times10^{20}$	7×10^{-5}	8×10^{-5}	-3×10^{-5}	
	l. c.	$\approx -7.\overline{1}\times10^{15}$				
$E_{>2}/E = 1.0-\underline{1}\times\underline{10}^{-12}$	$j=9$	$-1.393\,203\,929\,\underline{64}\times10^{20}$	4×10^{-10}	5×10^{-10}	-2×10^{-10}	
	l. c.	$\approx -3.7\times10^{10}$				
	$j=14$	$-1.393\,203\,929\,685\,\underline{67}\times10^{20}$	4×10^{-15}	3×10^{-15}	-1×10^{-15}	
	l. c.	$\approx 2.4\times10^6$				

[a] This table was evaluated using the method of Sec. 5 rather than Sec. 3 for reasons of convenience and computing time. The results are virtually identical; however, error and/or time estimates in the last four columns would be slightly worse.

[b] At $t=0$, the kinetic energy is zero, $E_{>2}/E$ is the ratio of the nonquadratic potential energy and the total potential energy at $t=0$.

[c] l. c.: last correction: $\sum_n \Delta A_j(n)$, cf. caption f.

[d] Underlined numbers mean questionable precision in view of higher-order corrections or machine "round off." The machine allows 14—16 significant digits.

[e] We substitute the solution up to order j into (3.2) and we take all terms to the rhs and define the relative error(t) $\equiv$ rhs $(3.2)(t)/[(\nu^2-\omega_0^2)\phi(t)]$ when $\phi(t)\neq 0$. We scale in $(\nu^2-\omega_0^2)\phi(t)$ because it is characteristic of the magnitude of the terms in the equation.

[f] This is a measure of the rate of convergence where $\Delta A_j(n)$ is the difference in $A(n)$ between the j and $j-1$ iteration.

[g] Computation time on IBM 360-65, in seconds.

the same as the above, first order, estimate. Therefore, we now invariably derive our "intuition" about the qualitative behavior of such nonlinear systems from just a single iteration of the backward scheme. The estimate in (3.13) is the first example of such an approximation, not easily obtained through any other form of "intuition", cf. (5.15)—(5.16) and Ref. 16, Eq. (2.18).

The extension to $\delta\neq 0$ and $\dot\phi(0)\neq 0$ is extremely easy in this scheme. The δ does not enter into the recursion relation *or* $A_0(1)$, cf. (3.5)—(3.7). Therefore, while one has to prescribe *both* ν and δ in order to derive $\phi(0)$ and $\dot\phi(0)$, one can obtain all coefficients $A_j(n)$ from ν alone. Once the $\{A_j(n)\}$ are available, we may choose any δ, with $0\leq\delta<2\pi$, in (3.11)—(3.12) and immediately

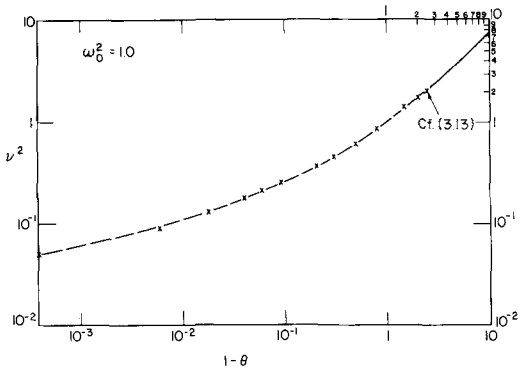


FIG. 4. ν^2 versus $1-\theta$.

obtain all $\phi(0)$, $\dot{\phi}(0)$ values which can possibly give rise to an oscillation with frequency ν.

In the theory of action-angle variables the δ is an integration constant, not affecting the functional form of the other functions of the time and the action. Hence the possibility of choosing the δ at the last moment. We will discuss the connection with "action-angle" variables elsewhere[17] and merely note that $u \equiv \nu t + \delta$, cf. (3.3) and (3.16), is the "angle" variable while $\mathcal{J} \equiv \sum_n n^2 \nu |A_n|^2$ is the "action" variable $[A_n \equiv \sum_j A_j(n)]$.

C. Quantitative results

In Table III, we display some numerical results for a backward scheme. We have used a modification of the backward scheme, introduced in Sec. 5 and Ref. 16 in order to reduce the computation time and, for the region $\nu^2/\omega_0^2 < 0.25$, increase the rate of convergence. The results are not affected by this.

The first two entries of Table III should be compared with the first and last entries in Table II. For these two cases, we specify ν^2 equal to the ν^2 calculated in Table II, using a forward scheme. The θ values calculated in Table III are indeed equal to the ones specified in Table II, within the precision of the ν^2 in Table II.

Notice the excellent rate of convergence, in Table III, even for ν^2/ω_0^2 values up to 10^{20}, where the ratio of the nonquadratic to quadratic potential energy is approximately 10^{20}. In the region $\nu^2/\omega_0^2 = 0$ to 0.25 the rate of convergence is so slow that we prefer the modifications of the scheme, introduced in Ref. 16.

A "log–log" plot of $\nu^2(\theta)$ is shown in Fig. 4. The solid line on the right indicates the linear relation between ν^2 and θ, cf. (3.13), which is still a good approximation in that region. For smaller $1-\theta$ values, corresponding to the "knee" in Fig. 3 the curve is not linear. Graphs of $\chi(u)$ for various values of ν^2 are displayed in Fig. 5. To interpret these graphs, we consider the two extreme cases, $\nu^2 \to 0$ and $\nu^2 \to \infty$. For $R > 0$ the potential has the shape of a "valley on top of a mountain." As $\phi(0)$ approaches the edge of this valley, the force becomes zero, the period goes to ∞, and $\nu^2 \to 0$. For $\nu^2 \approx 0$ the particle spends a large fraction of the period near the edge, and the graph of $\chi(u)$ looks like a square wave, cf. the graphs for $\nu^2 = 0.01$ and 0.05 in Fig. 4. When $\nu^2 \to \omega_0^2$, the graphs change to the familiar cosine waves, cf. $\nu^2/\omega_0^2 = 0.13$ in Fig. 5. As we take $\nu^2 \to \infty$, we find $|\phi(0)| \to \infty$. The χ^4 term in the scaled potential can be written as $\chi^4 \equiv [\phi/\phi(0)]^4$ and becomes more and more peaked at $\chi = \pm 1$, as $|\phi(0)| \to \infty$. In that case the acceleration is sharply peaked near $\chi = \pm 1$, the velocity will be nearly constant for all intermediate χ values, and the graph of $\chi(u)$ will look like a triangular wave, cf. $\nu^2/\omega_0^2 = 10^{20}$ in Fig. 5.

D. The expansion parameters and a convergence proof

In this section we show that the backward series is in fact an expansion in powers of $(1-\mu_0^2)$, where

$$(1-\mu_0^2) = (\nu^2 - \omega_0^2)/\nu^2, \quad \text{with } \mu_0 \equiv \omega_0/\nu. \tag{3.14}$$

We shall also indicate a convergence proof for the region $\mu_0 < 1$, i.e., $\omega_0 < \nu < \infty$. A separate paper will be devoted to the convergence proofs for the full region of this scheme, i.e., $0 < \nu < \infty$, and the other schemes of the present paper.[17]

First we scale the "relocated" equation of motion (3.2) to unit initial displacement and unit harmonic frequency

$$\chi_{uu} + \chi = \epsilon[(1-\mu_0^2)\chi + \mu_0^2 \theta \chi^3],$$

with (3.15)

$$\chi(0) = 1, \quad \chi_u(0) = 0,$$

and $\chi \equiv \phi/\phi(0)$, $u \equiv \nu t + \delta$, and $\theta \equiv R\phi^2(0)/\omega_0^2$, (3.16)

with $\chi_u \equiv d\chi(u)/du$ and $\epsilon = 1$. We use the same symbol χ as before, cf. (2.4), but here for a different function of the time u, a convenient but confusing practice. The $\delta = 0$ in order to obtain $\chi_u(0) = 0$. The lhs of (3.15) is a harmonic oscillator of unit frequency, being driven by the terms on the rhs. In our solution we use Fourier exponentials of the form $\exp(inu)$, cf. (3.3), and the rhs of (3.15) will contain terms with $n = \pm 1, \pm 3, \pm 5$, etc. It is clear that a term with $n = \pm 1$ may not be present

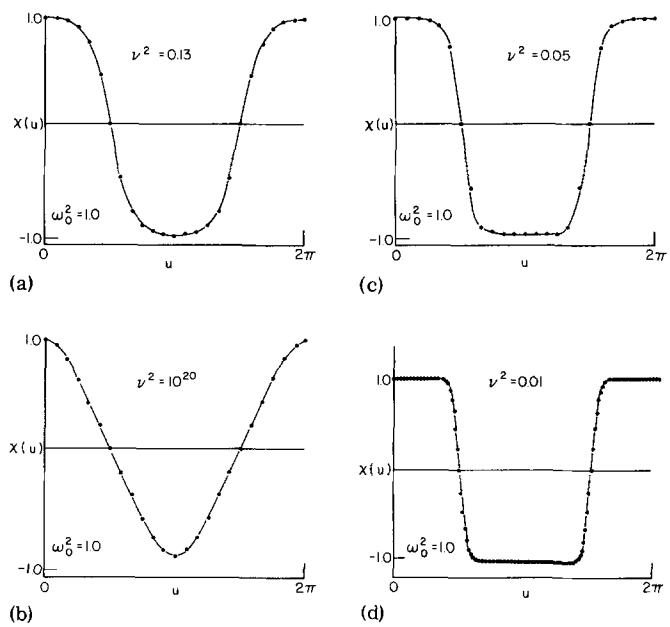


FIG. 5. Waveforms for the cubic anharmonic oscillator, obtained with the backward perturbation scheme.

on the rhs if we want to avoid resonance and unbounded, secular, terms.[4] We therfore equate this Fourier component of the rhs to zero and find the resonance condition

$$\int_0^{2\pi} \cos(u)[(1-\mu_0^2)\chi + \mu_0^2\theta\chi^3]\,du = 0, \tag{3.17}$$

or

$$\theta = (1 - 1/\mu_0^2)\oint\chi / \oint\chi^3, \tag{3.18}$$

where we denote the cosine integral over one period by $\oint$. In the forward schemes of Sec. 2 we, in effect, wrote μ_0 explicitly and obtained $\nu = \nu(\theta)$ from the resonance condition (3.17), i.e., "frequency renormalization." Here, in a backward scheme, we extract $\theta = \theta(\nu)$ in (3.18), i.e., "amplitude renormalization." Upon substitution of this θ in the equation of motion, we finally obtain

$$\chi_{uu} + \chi = \epsilon(1-\mu_0^2)[\chi - \chi^3\oint\chi/\oint\chi^3], \quad \text{with } \epsilon = 1, \tag{3.19}$$

an integro-differential equation. It is easy to verify by direct differentiation that (3.19) is equivalent to

$$\chi(u) = \cos(u) + \epsilon(1-\mu_0^2)\int_0^u dv\,\sin(u-v)[\chi(v) - \chi^3(v)\oint\chi/\oint\chi^3]. \tag{3.20}$$

This integral equation is solved by iteration, i.e., we start with $\chi_0 \equiv \cos(u)$ and evaluate the integrals using this χ_0. On the lhs we then obtain a new approximation $\chi_1 = \chi_0 + \epsilon(1-\mu_0^2)I_0$, where $I_0(u)$ is the evaluated integral. We can repeat the procedure with this new approximation instead of $\cos(u)$ and obtain a (Neumann) series solution of (3.19)—(3.20). The procedure is explained in detail in Secs. 2 and 7 of Ref. 4. It is obvious that each iteration brings in higher powers of $\epsilon(1-\mu_0^2)$ and that $(1-\mu_0^2)$ is in fact the expansion parameter of our solution. We could have seen this from the first order results (3.8)—(3.9) and the recursion relation (3.5). At first glance we thus expect poor convergence when $\mu_0 \to \infty$, i.e., $\nu \to 0$, and reasonable convergence even when $\mu_0 \to 0$, i.e., $\nu \to \infty$. This is precisely the behavior found in the previous section on quantitative results. For the region $0 < (1-\mu_0^2) < 1$, i.e., $\omega_0 < \nu < \infty$, one can easily give a convergence proof which is completely analogous to the one in Sec. 7 and Appendix A of Ref. 4. The essential step is again to show that all iterates χ_j of (3.20) stay bounded, i.e., $|\chi_j(u)| \le 1$, cf. (A.23)—(A.24) of Ref. 4. Translated into θ variables this region corresponds to $0 > \theta > -\infty$, substantially larger than the region of convergence of any forward scheme, cf. Sec. 2C. Convergence proofs for *all* ν, including $0 < \nu < \omega_0$, will be given elsewhere, using an entirely different technique.[17]

If $\nu = \nu(\theta)$, or $\theta = \theta(\nu)$, does not satisfy the resonance condition (3.17)—(3.18), we unavoidably obtain Fourier coefficients which are a function of the time! In the present case they would contain simple powers of t or u (secular terms).[4] In practice one terminates the series at a finite order and the highest power of t that remains automatically leads to an unbounded result as $t \to \infty$. Less obvious but more serious is the fact that one cannot improve this situation by including higher order terms with higher powers of t [as one can, for example, in a Taylor expansion of $\sin(t)$]. In other words, even if one sums out to infinite order the result will, in general, *not* converge for all t, with $0 \le t < \infty$. The transcendental functions in t which solve *nonlinear* differential equations, generally, have singularities in the complex t plane. Hence their power series expansions in t cannot converge for all t with $0 \le t < \infty$. These singularities can be found from the singularities of the differential *equations*, whenever the solution itself is not available, using an argument similar to the one leading to Eq. (7.10) of Ref. 4.[12] Some transcendental functions which solve *linear* differential equations and a few simple nonlinear ones, e.g., $\exp(t), \sin(t)$, etc., do have power series expansions in t which converge for all t with $0 \le t < \infty$. *For most nonlinear systems an infinite series with secular terms will, however,* not *result in a solution which converges for all* t *with* $0 \le t < \infty$ *and a "renormalization"* $\nu = \nu(\theta)$, or $\theta = \theta(\nu)$, *as in* (3.17)—(3.18) *is needed in theory as well as in practice.*

For a doubly periodic solution, with two complex singularities per period, a power series in t and subsequent analytic continuation might suffice if one accepts a poor rate of convergence at some t values. However, these are precisely the Jacobi elliptic functions which can be evaluated more easily using other techniques. Our equation (2.1) is solved by the Jacobi elliptic cosine.[4]

E. Physically unbounded solutions ($\theta > +1$)

When $\theta > +1$, and $\dot\phi(0) = 0$, the "particle" is on the "down slope" of the potential, cf. (2.2)—(2.5) or Fig. 1(a), and escapes to $+\infty$ or $-\infty$. The "backward" scheme converges again if we choose $\nu^2 < 0$, i.e., if we expand the solution in powers of $\exp(pt+\delta)$, with p and δ *real valued*, i.e.,

$$\phi(t) = \sum_j \epsilon^j \sum_n A_j(n)\exp[n(pt+\delta)].$$

A. Bountis, employing a "forward" version, found that the expansion converges over a *finite* time interval $0 \le t < T_{esc}$, with $T_{esc} (< \infty)$ a function of $\phi(0)$ and $\dot\phi(0)$. Clearly the expansion cannot converge for *all* t with $0 \le t < \infty$. This is not an artefact of the method but a true indication that the particle "reaches" infinity in a *finite* time T_{esc}, as we will demonstrate below.

The regions of convergence in the complex θ plane were found to be exactly as in Fig. 2 but mirrored across the line $\mathrm{Re}[\theta] = +1$. This can be derived directly by transforming to $\psi \equiv 1/\phi$. Equation (2.1) becomes

$$\ddot\psi = -\omega_0^2\psi + \{4H\}\psi^3 = -\omega_0^2\psi + \{\omega_0^2\phi^2(0)[2-\theta]\}\psi^3, \tag{3.21}$$

with

$$\psi \equiv \phi^{-1} \quad \text{and} \quad \theta_\psi \equiv \{\omega_0^2\phi^2(0)[2-\theta]\}\psi^2(0)/\omega_0^2, \tag{3.22}$$

the latter analogous to the definition of θ in (2.4). H is the Hamiltonian (2.2).

Equation (3.21) can now be recast as

$$\psi_{\tau\tau} = -\psi + \theta_\psi\psi^3, \quad \text{with } \theta_\psi = 2-\theta \text{ and } \tau \equiv \omega_0 t, \tag{3.23}$$

$$\text{hence} \quad -\infty < \theta_\psi < +1 \quad \text{when } \theta > +1. \tag{3.24}$$

This equation for ψ is exactly the same as the original equation for ϕ, (2.3), if we replace χ with ψ and θ with θ_ψ. Hence, for $\theta > +1$, we can obtain periodic $\psi(\tau)$ solu-

tions exactly as in Secs. 2 and 3 but using θ_ψ instead of θ. Thus the regions of convergence in the complex θ plane are the same as before but mirrored across $\mathrm{Re}[\theta]=1$, due to $\theta_\psi = 2-\theta$. The solution $\psi(t)$ has a period $2\pi/\nu(\theta_\psi)$. From this we can estimate the finite escape time T_{esc}, to $\phi = \pm\infty$, for $\phi(t)$

$$T_{\mathrm{esc}} = \tfrac{1}{4}2\pi/\nu(\theta_\psi) \approx \pi/\omega_0[3\theta - 2]^{1/2}, \tag{3.25}$$

when $\theta > +1$ and where we have used the approximation (3.13) for $\nu(\theta_\psi)$. So (3.25) is certainly a good approximation in the region $1.2 \lesssim \theta < +\infty$.

4. THE PROBLEM OF THE SMALL DENOMINATORS

In the previous sections, and Ref. 4, we were forced to "renormalize" the frequency of oscillation as a function of the initial displacement, i.e., $\nu = \nu(\theta)$, or vice versa $\theta = \theta(\nu)$, in order to obtain a bounded Fourier series representation of the solution, *convergent for all* t with $0 \leq t < \infty$. Any other choice of ν results in Fourier coefficients which are functions of t and a series which does not converge for all t, cf. Sec. 3D.

However in *systems* of coupled nonlinear differential equations there exists yet another classical problem resulting in unbounded representations of variables which one knows to be bounded, the problem of the "small denominators" or "small divisors." While there is some relation to the problem mentioned above, the small denominators problem persists even if we renormalize all frequencies correctly. It presents a major stumbling block for all theories on the dynamics of N bodies ($N \geq 3$) in classical, celestial, quantum, and statistical mechanics.[5-7] Before we indicate how this problem can be avoided, for a large class of solutions, let us see how these small denominators arise. As an example, consider the following simple system of coupled harmonic and anharmonic oscillators:

$$\ddot{\phi} = -\omega_0^2\phi + R\phi^3 + \psi, \quad \text{with} \quad \phi(0)=0, \quad \dot{\phi}(0)=0, \tag{4.1}$$

$$\ddot{\psi} = -\gamma^2\psi, \qquad \text{with} \quad \psi(0)=B, \quad \dot{\psi}(0)=0. \tag{4.2}$$

Since we can solve (4.2) exactly the first equation becomes

$$\ddot{\phi} = -\omega_0^2\phi + R\phi^3 + B\cos(\gamma t), \tag{4.3}$$

with $\phi(0)=\dot{\phi}(0)=0$. This is a form of the "Duffing equation,"[10-14] which appears, for example, in the equation of motion of a pendulum driven by a periodic force[9] and certain laser models.[8] The frequency γ never changes since the $\cos(\gamma t)$ is the solution of a *linear* differential equation. Only the ω_0 will be renormalized to a new frequency ν.

In the literature the small denominators problem is most often discussed in terms of ω_0 and γ rather than ν and γ. However, only for very small values of the driving amplitude B will the main frequency of oscillation, ν, approach ω_0. The small denominators problem is present in both cases; so we use the *renormalized* frequency ν and relocate (4.3) as usual

$$\ddot{\phi} + \nu^2\phi = \epsilon[(\nu^2-\omega_0^2)\phi + R\phi^3] + B\cos\gamma t \qquad \text{with } \epsilon = 1. \tag{4.4}$$

It is obvious that, upon iteration, we obtain terms of the form $\exp[it(n\nu + p\gamma)]$ and that any combination of n and p values will appear if we iterate to high enough order. Nevertheless, the problem originates precisely in assuming a *two*-frequency expansion

$$\phi = \sum_{j=0}^{\infty} \epsilon^j \sum^{2j+1}\cdots\sum_{n,p=-2j-1} A_j(n,p)\exp[it(n\nu+p\gamma)]. \tag{4.5}$$

In this example, we chose $\phi(0)=\dot{\phi}(0)=0$. So, in the current "backward" scheme the quantity to be derived is the driving term

$$B = B_0 + \epsilon B_1 + \epsilon^2 B_2 + \cdots. \tag{4.6}$$

Substitution of our series (4.5) in the equation of motion (4.4) yields a recursion relation

$$[\nu^2 - (n\nu+p\gamma)^2]A_j(n,p) - \tfrac{1}{2}\delta_{n,0}\delta_{p,\pm1}B_j = (\nu^2-\omega_0^2)A_{j-1}(n,p) + R\left(\sum_{\Gamma_3(j-1;n,p)}\cdots\sum A_{j_1}(n_1,p_1)\cdots A_{j_3}(n_3,p_3)\right) \tag{4.7}$$

or

$$A_j(n,p) = [\text{rhs } (4.7)]/[(1-n)\nu - p\gamma][(1+n)\nu + p\gamma], \tag{4.8}$$

when $(n,p) \neq (0,\pm1)$. The denominator, in (4.8), can become arbitrarily small, for large enough (n,p) values, and preclude convergence. Hence the name "small denominators" problem.

Since the above series cannot possibly converge for *all* (ν,γ) choices one might justifiably ask whether there is *any* choice of (ν,γ) for which it will. The mathematics responds "in kind" to our posing a desperate question within a frequently inapplicable formalism. It raises obstacles which depend on whether γ/ν is a *rational* or an *irrational* number! A fascinating insight into, and historic account of, the problems involved can be found in Ref. 5 (e.g., Chap. 1) by Moser. His work and that of Kolmogorov and Arnold demonstrates, for the first time, that such series can converge for certain (ν,γ) choices with γ/ν *irrational*.[5-7] In addition to the problem of actually *constructing* these series there is, in a *"forward"* scheme, the practical problem of deciding whether γ/ν *is* irrational when each higher order term adds further corrections to the value of ν, cf. (2.9). The conditions under which the *K*olmogorov–*A*rnold–*M*oser theorems apply are known only in terms of the values of ν and γ (to be discussed at the end of Sec. 4B). A first advantage of the backward scheme is that ν and γ are not derived but *chosen*, i.e., the angle variables $(\nu t, \gamma t)$ are independent variables while the $\dot{\phi}$, ϕ are dependent ones. The main problem now is to find a transformation from the angle variables $(\nu t, \gamma t)$ back to $\dot{\phi}, \phi$ variables. Below, we present such a transformation for all frequencies ν and γ, with γ/ν *rational*. Remember that the KAM theorems apply to certain ν and γ with γ/ν *irrational*.

A. Removal of the "small denominators" for two rationally dependent frequencies

Consider the case where γ/ν is a *rational* number, different from 1, i.e., $\gamma/\nu \equiv m_2/m_1$ with $m_2 \neq m_1$ and m_1, m_2 positive integers which are relative primes. Then

$$m_2\nu - m_1\gamma = 0 \quad (m_2 \neq m_1). \tag{4.9}$$

At a sufficiently high order, the coefficients of ν and γ, in the denominator of (4.8), will assume the values $\pm m_2$ and $\pm m_1$ for the first time, and the denominator becomes zero exactly. So one was forced to rule out, explicitly, the case with rationally dependent frequencies. [Gustavson, and Harris, transform the Hamiltonian to a "normal" form,[18] i.e., a nonlinear analogue of the "normal modes" transformation in linear systems. Their series are not necessarily convergent,[6] but they do obtain accurate results in some cases.[18,6]]

All this is a manifestation of the fact that the derivation of *the recursion relation* (4.7) *is not valid in this case*. We derive such recursion relations by arguing that we want to satisfy the differential equation for all values of t and equate like powers of $\eta \equiv \exp[it(n\nu + p\gamma)]$, e.g., (2.14). But in the present case we have $\exp[it(m_2\nu)] = \exp[it(m_1\gamma)]$, cf. (4.9), and we should not have distinguished between $A_j(m_2, 0)$ and $A_j(0, m_1)$. The same holds for all $A_j(n + km_2, p)$ and $A_j(n, p + km_1)$. The "rationale" for allowing all these values of (n,p) in the $\{A_j(n,p)\}$ was that all combinations $n\nu + p\gamma$ do indeed appear in an exponent in (4.5). However, the $\{n\nu + p\gamma\}$ are *not independent* in the present case and are in fact all multiples of a "greatest common frequency" ν_r, with

$$\nu_r \equiv \gamma/m_2 = \nu/m_1 \quad (m_1 \neq m_2), \tag{4.10}$$

which we call *the recurrence frequency*. Note that, after m_1 periods of frequency ν, we have also completed m_2 periods of frequency γ and the solution returns exactly as it was. Thus the (Poincaré) recurrence time of the solution is $T_r = 2\pi/\nu_r$.

In order to avoid the above "degeneracies," we have to assume a *single frequency* expansion

$$\phi = \sum_{j=0}^{\infty} \epsilon^j \sum_n A_j(n) \exp[it(n\nu_r)] \quad (\nu = m_1\nu_r, \gamma = m_2\nu_r), \tag{4.11}$$

with

$$A_j(-n) \equiv A_j(+n), \quad \text{real},$$

and

$$A_j(n) \equiv 0 \quad \text{if} \quad |n| > (2j+1)\max(m_1, m_2) \quad \text{or} \quad j < 0. \tag{4.12}$$

Upon substitution of *this* series in the equation of motion (4.4), we obtain the recursion relation

$$(m_1^2 - n^2)\nu_r^2 A_j(n) - \tfrac{1}{2}\delta_{n,\pm m_2} B_j = (\nu^2 - \omega_0^2)A_{j-1}(n) + R\left(\sum_{S_3(j-1;n)} \cdots \sum A_{j_1}(n_1) \cdots A_{j_3}(n_3)\right). \tag{4.13}$$

The relevant denominator now is $|(m_1^2 - n^2)\nu_r^2|$. For $n^2 = m_1^2$ it becomes zero and the $A_j(\pm m_1)$ is arbitrary, at order j. *In all other cases the denominator has a constant lower bound* $\nu_r^2 > 0$, and we expect convergence for a range of ν_r values, exactly as in Sec. 3. [The $A_j(m_1)$ takes over the role of the previous $A_j(1)$. Its value is derived from the resonance condition $n = m_1$ at $j+1$, cf. (3.9).] This single frequency expansion, in terms of a recurrence frequency ν_r, is the third essential "device" we introduce in this paper.

This approach yields the "*periodic*" solutions of the equations. Note that the recurrence "period" $T_r = 2\pi/\nu_r$ is the result of the mathematical construction (4.10) and does not represent any obvious physical time constant of the problem. The T_r can differ from the "apparent" period of oscillation of $\phi(t)$ by any arbitrary amount, e.g., $T_r = 2\pi$ when $(\nu, \gamma) = (1, 3)$ while $T_r = 2\pi 10^{+20}$ when $(\nu, \gamma) = (1 + 10^{-20}, 3)$. Apparently T_r and ν_r $(= 2\pi/T_r)$ are extremely "erratic" functions of (ν, γ) and of the initial "displacements" $\dot{\phi}(0)$, $\phi(0)$, $\dot{\psi}(0)$, and $\psi(0)$, even when ν and γ are "smooth" functions. While ν_r and ν are both constants of the motion of the subsequent time evolution $\phi(t)$, it is the ν_r which enters into *every* term of the series, cf. (4.13). Poincaré and Bruns showed that in (celestial) mechanics most of the constants of the motion *cannot* be analytic or differentiable functions of the $\dot{\phi}, \phi$,[5-7] cf. Ref. 2, Chap. 5 [with the exception of the Hamiltonian and of a few simple or pathological cases]. The implications of this well-known theorem are generally less well known. It provides an explanation of the failure of the "canonical transformation and perturbation theory" of classical mechanics to yield a *convergent* series solution in most cases, (as proven by a theorem of C.L. Siegel[7,6,5]). Such power series in $\dot{\phi}, \phi$ would also express the constants of the motion as power series in $\dot{\phi}, \phi$. If they converged, the constants of the motion would be differentiable functions of $\dot{\phi}, \phi$ and Poincaré's theorem would be violated! In our current example the Fourier series (4.11)–(4.13) for $(\nu, \gamma) = (1, 3)$ differs greatly from the one for $(\nu, \gamma) = (1 + 10^{-20}, 3)$, if only in the number of components per unit frequency. Yet the difference between the *initial* displacements $B = \psi(0)$ for both cases is $\approx 10^{-20}$ only. From the above example we see that such a "discontinuous" change in the *structure* of our series will always be present no matter how small the difference between the ν's (and B's) and that any given series will only converge for one unique value of ν (and B) constant γ). The previous arguments do not apply, of course, in the case of a *single* anharmonic oscillator where there is only one frequency (i.e., $\nu_r = \nu$), since it is an analytic function of $\phi(0)$ for a vast range of $\phi(0)$ values. By using quadrature the latter can be obtained directly from the Hamiltonian being an analytic function of $\dot{\phi}, \phi$, cf. Sec. 7 of Ref. 4. It also shows why a forward series transformation, i.e., $\{\dot{\phi}_k, \phi_k\} \to \{\nu_k, \delta_k\}$, does converge in the case of a *single* oscillator, cf. Sec. 2, while diverging for most *systems*.[6,5,4] Our backward transformation, i.e., $\{\nu_k, \delta_k\} \to \{\dot{\phi}_k, \phi_k\}$, converges in both cases, when all $\{\nu_k\}$ are rationally *dependent*.[17] Systems with N variables will be mentioned in Sec. 4C. We continue the actual solution of the Duffing equation in Sec. 5.

B. Rationally independent frequencies

Consider the case where γ/ν is an *irrational* number. While there is no longer any integer combination $(m_2\nu - m_1\gamma)$ yielding zero exactly, as in (4.9), such combinations come arbitrarily close to zero and the denominator of (4.8) will be "nearly" zero infinitely many times during the iterations, i.e., as $j \to \infty$. Not only is the two-frequency series (4.5) divergent in general,[6] but it is also useless as an (asymptotic) approximation due to the rapid increase in the values of the coefficients, in most nonlinear cases we checked.

While we offer no *explicit* scheme for the removal of the "small denominators" when γ/ν is irrational there now appears a way of sidestepping the problem in that case. Any irrational number $\sigma \equiv \gamma/\nu$ can be "approximated" (i.e., defined) by two infinite sequences of rational numbers and one might well ask whether the solution $\phi(t)$ for an irrational σ, can be similarly "approximated" (, defined, constructed) by two infinite sequences of solutions $\phi(t)$ with rational σ. For our current Duffing problem (4.1)—(4.3) this does appear to be the case. In general we do not know for certain that the "error" in $\phi(t)$, at any large t, will disappear if we let the "error" in σ go to zero. It is certainly not true near $\nu^2 = 0$, since $\nu^2 < 0$ yields unbounded solutions while $\nu^2 > 0$ yields bounded ones. However, $\nu^2 = 0$ is an exceptional point in many other ways already (and $\sigma = \pm\infty$).

In practice one would now "approximate" the irrational σ by a rational number which differs from σ by an amount less than the "round-off" error of any computer employed or less than the error involved in observations on the physical problem at hand over the time interval of the observations. At the moment this is a rather glib argument, in view of the difficulty of estimating the error in $\phi(t)$ for a given "error" in σ, but it does point out that "rational-σ approximations" might even become good enough in practice, similar to the situation in number theory. In general we expect, of course, that the "error" in σ must be chosen smaller and smaller if we require the same "error" in $\phi(t)$ at larger and larger t. [The practical construction of the series (4.11) might seem more complicated and time consuming for the large m_1, m_2 values, needed to obtain a small "error" in σ, than for small m_1 and m_2. A way to avoid this is discussed in an appendix of Ref. 15.]

We can illustrate the previous discussion for the case of the Duffing equation (4.1)—(4.3) as follows.

Let us plot curves of constant $\sigma(\equiv \gamma/\nu = m_2/m_1)$ in a *two-dimensional* graph of the driving amplitude squared, i.e., RB^2, cf. (4.2)—(4.6), versus the driving frequency γ, cf. Figs. 6—9 in Sec. 5. The locus of points with the same rational σ is indeed one continuous curve since *the structure (form) of our series* (4.11) *cannot change if* σ *stays constant.* [The integers m_2 and m_1 are the same for all points on one curve, cf. (4.9)—(4.10).] Only the values of the Fourier coefficients change, and we obtain a B which, at constant σ, is a continuous function of γ, inside the region of convergence. The points on each curve are parametrized continuously by the second frequency, ν, (or by ν_r) and for $\nu = \omega_0$, i.e., the harmonic case, all curves cross the horizontal axis, $B = 0$, at $\gamma = \omega_0\sigma$. It is essential to notice that these *curves of constant* σ *cannot intersect, and must all be "parallel,"* [the solution of the differential equations (4.1)—(4.2) is unique], when $\gamma \neq \nu$. But we also find, below, that B is a continuous function of σ, at constant ν. Hence all points with one and the same *irrational* σ are indeed "locked in" between the same curves of rational σ, making "rational-σ" approximations feasible in the case of the Duffing equation. The "error" in $\phi(t)$ can be estimated via the following continuity argument. Let us transform the Duffing equation (4.3) into

$$\phi_{uu} = [-\omega_0^2\phi + R\phi^3 + B\cos(\sigma u)]/\nu^2, \qquad (4.14)$$

$$\text{with } u \equiv \nu t \text{ and } \sigma \equiv \gamma/\nu. \qquad (4.15)$$

In the theory of differential equations,[12] one shows that bounded solutions $\phi(u)$ depend continuously [analytically] on any parameter σ, if the rhs of (4.14) is a (uniformly) continuous [analytic] function of σ (and ϕ and u). For a bounded solution the latter is certainly true on any finite time interval, and we find that $\phi(u)$ and B [$= \phi_{uu}(0)\nu^2$, cf. (4.1)—(4.3)] are continuous functions of σ. From the proof of the above, in Ref. 12, one can then obtain an estimate of the "error" in $\phi(t)$, given some "error" in σ.

We do not rule out the possibility that for some systems, perhaps less simple than the Duffing equation, the "*independent* frequencies" solutions may be totally different from the (nearest) "dependent frequencies" solutions. We would like to point out that, if so, ordinary numerical experiments, using stepwise integration, lose their relevance in such cases [the converse does not hold]. A numerical integration in "action-angle" variables with dependent frequencies would miss all solutions with independent frequencies and thus not be representative of the latter (under the hypothesis). A numerical integration in ordinary ϕ, $\dot{\phi}$ variables would, within one "round-off" error of the machine, have two classes of solutions which do not resemble each other (under the hypothesis). With each subsequent time step and "round-off," one could move from one class to the other and the final result would be representative of neither. We have applied our method to larger systems of anharmonic oscillators[15] but have not yet investigated the above questions for those systems.

The KAM theorems do apply here, i.e., solely to the case where γ/ν is *irrational*. They state[5—7] that if the frequencies γ and ν have the property that

$$|n\nu - p\gamma| \geq c(|n| + |p|)^{-d}, \qquad (4.16)$$

for all integers $n, p (\neq 0, 0)$, with some positive c and d, then, under certain restrictions on the system there *exists*, "near" a "*given*" solution with those frequencies γ, ν and $R = R_1$ in the equation of motion (5.1), another solution with the same frequencies γ, ν but $R = R_2 \neq R_1$, for sufficiently small $|R_2 - R_1|$. All solutions thus established have *frequencies* γ and ν which, apparently, are *invariant under small perturbations in* R. Normally we would like to take, $R_1 = 0$, then the equation of motion would be linear (hence $\nu = \omega_0$), the "given" solution would be easily obtained, and the KAM theorems would demonstrate the existence of a nonlinear solution ($R_2 \neq 0$) with the same frequencies γ and ω_0 [if γ and ω_0 satisfy (4.16)]. However, our current equation and initial conditions (4.1)—(4.3) do not satisfy one of the additional restrictions we alluded to. Finally we would like to emphasize that (4.16) represents an *infinite* number of inequalities, each of which must be satisfied. Relation (4.16) ensures, among other things, that the denominators in (4.8) will approach zero at a slower rate than the numerators. Hence there is no "small-denominator" problem either if (4.16) is satisfied. A precise discussion of the theorems can be found in Chap. II, Sec. 3, of Ref. 5.[5—7]

C. Systems of *N* variables

The periodic solutions for systems of $N(> 2)$ coupled

polynomial anharmonic oscillators will be discussed at length elsewhere.[15] Here we merely outline the simple modifications, of the method described in Sec. 4A (and 5). The equation of motion for the kth oscillator is transformed to

$$\ddot{\phi}(k) = -\omega_0^2(k) + P_k^L(\phi(1), \ldots, \phi(N)) \quad (k = 1, \ldots N), \tag{4.17}$$

where P_k^L is an Lth order polynomial in $\phi(1), \ldots, \phi(N)$, without constants or linear terms. We specify a set of frequencies $\nu_1, \ldots, \nu_N$ such that these frequencies, *as well as the frequencies in each subset of it*, are rationally dependent. Thus there is again a "greatest common frequency" ν_r and each frequency ν_k can be written as

$$\nu_k = m_k \nu_r \quad (k = 1, \ldots, N), \tag{4.18}$$

where m_k is some positive integer and the m_k's are relative prime (in practice one specifies, of course, the ν_r and m_k's rather than the ν_k's). After adding a relocation term $\nu_k^2 \phi(k)$ to both sides of (4.17), we attempt a solution of the form

$$\phi(k,t) = \sum_j \epsilon^j \sum_n A_j^{(k)}(n) \exp[it(n\nu_r + \delta^{(k)})]. \tag{4.19}$$

Substitution of this series in the relocated version of (4.17) yields recursion relations

$$\begin{aligned}(m_k^2 - n^2)\nu_r^2 A_j^{(k)}(n) &= [\nu_k^2 - \omega_0^2(k)] A_{j-1}^{(k)}(n) \\ &+ \sum_{S_L(j-1;n)} \cdots \sum P_k^L(A_{j_{k_1}}^{(1)}(n_{k_1}), \ldots, A_{j_{k_N}}^{(N)}(n_{k_N})),\end{aligned} \tag{4.20}$$

for $k = 1, \ldots, N$. Having obtained the $A^{(1)}, \ldots, A^{(N)}$ up to order $j-1$, Eq. (4.20) enables one to obtain all jth order terms.[15] The evaluation of the sums over the set $S_L(j-1;n)$, as well as the iteration of (4.20), proceeds analogously to the method described before (or to the simpler method of the next section).

5. AN ITERATION SCHEME WITHOUT SMALL DENOMINATORS FOR A DRIVEN ANHARMONIC OSCILLATOR (THE DUFFING EQUATION)

In the previous section we studied the "small denominators" problem as it appears in the "Duffing equation"

$$\ddot{\phi} + \nu^2 \phi = \epsilon[(\nu^2 - \omega_0^2)\phi + R\phi^3] + B\cos(\gamma t),$$
$$\text{with} \quad \phi(0) = \dot{\phi}(0) = 0, \quad \epsilon = 1, \tag{5.1}$$

here written in "relocated" form, cf. (4.1) and (4.4). In this section we will study its solution. Notice that if $R > 0$, the anharmonic-oscillator part of (5.1) has a potential energy in the shape of a "valley on top of a mountain," cf. Fig. 1(a). Hence, for a large enough driving term B, we expect that the solution will "leave the valley and go down the slope of the mountain," i.e., the solution becomes unbounded (we do obtain, in Sec. 5D, some *bounded* solutions which go *slightly* outside the valley before being driven back). One of the main problems, besides constructing any solutions at all, is to determine the region of B values for which the solution stays bounded and *to find this region as a function of* γ and the other parameters.[8–14] Our results are graphically displayed in Figs. 6–9 and considerably extend estimates of this region in the literature.[9] In general those estimates, based on numerical analysis, are unlikely to provide the *full* "region of bounded solutions," in the case of the Duffing equation, due to the peculiar properties of the analytic solutions near the edge of this region. This will be discussed in Sec. 5D.

We might now proceed using recursion relation (4.13) with (4.11)–(4.12), which we derived before, and we indeed found the above region in this way. However, we take the opportunity to introduce a slightly different version of our scheme which is simpler to handle, close to the integral equations of Sec. 3D, closer to the techniques used to prove convergence and in general facilitates a recognition of the underlying physics. In addition any numerical iteration of *this* scheme uses a great deal less computer memory, time, and money.

A. A Fourier component iteration scheme

While Eq. (5.1) in ϕ variables is most convenient for the study of the region of convergence,[17] the *solution* of (5.1) is more easily obtained if we first scale (5.1) in a rather peculiar way:

$$\chi_{rr} + m_1^2 \chi = \epsilon[(\nu^2 - \omega_0^2)\chi + \omega_0^2 \theta \chi^3]/\nu_r^2 + (m_1^2 - m_2^2)\cos(m_2 r), \tag{5.2}$$

with

$$\nu \equiv m_1 \nu_r, \quad \gamma \equiv m_2 \nu_r, \quad \text{cf. (4.11)–(4.12)}, \quad r \equiv \nu_r t, \tag{5.3}$$

$$K \equiv B/\nu_r^2(m_1^2 - m_2^2), \quad \chi \equiv \phi/K, \tag{5.4}$$

$$\chi_{rr} \equiv d^2\chi(r)/dr^2 \quad \text{and} \quad \theta \equiv RK^2/\omega_0^2, \tag{5.5}$$

i.e., we changed to *a new time scale* r, incorporating the recurrence frequency ν_r, and divided Eq. (5.1) by a factor $\nu_r^2 K$. We will now attempt a solution of the form

$$\chi(r) = \sum_{n=-\infty}^{+\infty} A_n \eta^n, \tag{5.6}$$

with

$$\eta \equiv \exp(i[r + \delta_r]) \quad \text{and} \quad A_{-n} = A_n^*, \tag{5.7}$$

i.e., in this scheme we abandon the power series in ϵ and use a pure Fourier series only. Let us substitute (5.6)–(5.7) in the equation of motion (5.2). Equating the coefficients of η^n, we obtain our *basic recursion formula*

$$(m_1^2 - n^2)A_n' = \epsilon[(\nu^2 - \omega_0^2)A_n + \omega_0^2\theta \sum_{S_3(n)} A_{n_1}A_{n_2}A_{n_3}]/\nu_r^2 + \delta_{n,\pm m_2}\exp(-in\delta_r)(m_1^2 - m_2^2)/2, \tag{5.8}$$

for $m_1 \neq m_2$ and $\epsilon = 1$, which is simpler than (4.15). We have attached a prime to the A_n on the lhs to indicate our iteration procedure: We evaluate the rhs, using some given vector **A**, with only a finite number of elements A_n different from zero, and thus derive a *new* vector **A′** *on the lhs*, containing a larger number of non zero elements, etc. The summation in (5.8) proceeds over a set of three integers

$$S_3(n) \equiv \{n_1, n_2, n_3 \mid n_1 + n_2 + n_3 = n\} \tag{5.9}$$

While this set contains infinitely many (n_1, n_2, n_3) combinations, the sum over $S_3(n)$ will, after a finite number

of iterations, only contain a finite number of terms, since only a finite number of the A_n will then be different from zero.

For $n=m_1$ the lhs of (5.8) becomes zero and we obtain the usual "resonance condition" from which, this time, we solve for θ

$$\theta=-\left(\frac{\nu^2-\omega_0^2}{\omega_0^2}\right)\left(\sum_{S_3(m_1)}A_{n_1}A_{n_2}A_{n_3}\right)^{-1}\times A_{m_1}, \qquad (5.10)$$

cf. (3.18). Under our iteration procedure we first calcu late the above θ, given a "finite" vector **A**, and then *employ this value of* θ when we evaluate (5.8) for all remaining n values. The definition of θ, cf. (5.4)—(5.5), includes the scaling factor K, i.e., the amplitude B of the driving term. Hence the above θ is the principal quantity to be derived, once we specify ν and some initial conditions.

B. Initial conditions

Since the χ of (5.6) is no longer a power series in ϵ, we are also no longer forced to start our iterations with the solutions of the *homogeneous* part of the equation of motion (5.2) as in previous sections. This freedom will be used in some cases to extend the region of convergence and/or improve the "rate of convergence".[16] This freedom can also be introduced in previous schemes if we choose the homogeneous part of the equation in a different way. [Cf. (7.19) of Ref. 4. The "relocation" in (2.7) forms an example.] The current scheme, however, allows us to start the iterations with any vector **A** without altering a single equation.[16] Here we consider the same case as in Sec. 4A and start once again with the solution of the homogeneous equations, i.e., (5.2) and (5.8) for $\epsilon=0$. The recursion formula (5.8) then yields

$$A_{\pm m_1}\neq 0 \quad \text{and} \quad A_{\pm m_2}=\tfrac{1}{2}\exp(\mp i m_2\delta_r) \quad (0\text{ it.}), \qquad (5.11)$$

as the only coefficients which may be different from zero after the "zeroth" iteration, (0 it.). The particular initial conditions of (5.1), in "scaled" variables, are

$$\chi(0)=0, \quad \chi_\tau(0)=0, \quad \text{whence } \chi_{\tau\tau}(0)=m_1^2-m_2^2. \qquad (5.12)$$

These three relations allow us to solve for the three unknown quantities in (5.11), and we obtain

$$\delta_r=0 \quad \text{whence } A_{\pm m_2}=\tfrac{1}{2} \text{ and } A_{\pm m_1}=-\tfrac{1}{2} \quad (0\text{ it.}). \qquad (5.13)$$

The simplicity of these coefficients is the result of the peculiar scaling in (5.2)—(5.5). If we choose $\chi(0)$, $\chi_\tau(0)$ and the driving term differently, we can again use the same procedure. However, one should keep in mind that the $\phi(0)$ *and* $\dot\phi(0)$ *are to be derived* then, since the scaling factor K is a derived quantity!

Our particular initial coefficients result in all A_n being real, cf. (5.8). After each iteration we satisfy the initial condition (5.12) again by choosing

$$A'_{\pm m_1}=0-\tfrac{1}{2}\sum_{n\neq\pm m_1}A'_n\,. \qquad (5.14)$$

The $A'_{\pm m_1}$ can, as usual, not be derived from the recursion formula (5.8) but are found from (5.14) instead. The extension of this condition to the case $\chi(0)\neq 0$ and/or $\chi_\tau(0)\neq 0$ is easily found from Secs. 2A and 3B. We will again derive our "intuition" from the zeroth order results.

C. Zeroth order results

Let us use the coefficients (5.13) and evaluate θ [(5.10)]:

$$\theta=-[(\nu^2-\omega_0^2)/\omega_0^2](3A_{m_1}^2A_{-m_1}+6A_{m_1}A_{m_2}A_{-m_2})^{-1}\times A_{m_1}$$
$$=4(\omega_0^2-\nu^2)/9\omega_0^2 \quad (0\text{ it.}). \qquad (5.15)$$

From the definition of θ, (5.5), and (5.4), we obtain the amplitude of the driving term in

$$RB^2=\omega_0^2\theta(\gamma^2-\nu^2)^2=4(\gamma^2-\nu^2)^2(\omega_0^2-\nu^2)/9$$
$$\equiv 4(\sigma^2-1)\nu^4(\omega_0^2-\nu^2)/9 \quad (0\text{ it.}), \qquad (5.16)$$

where

$$\sigma\equiv\gamma/\nu=m_2/m_1. \qquad (5.17)$$

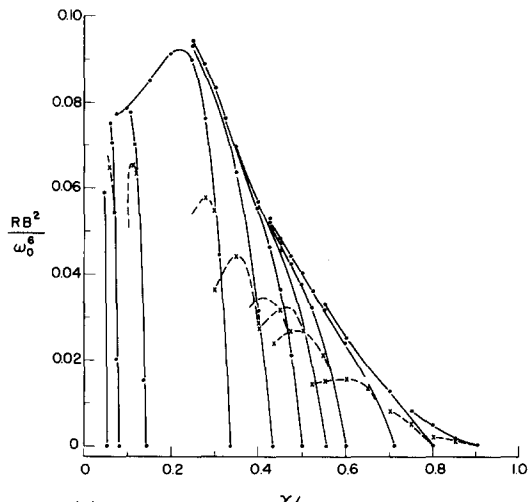


Fig. 6(a)

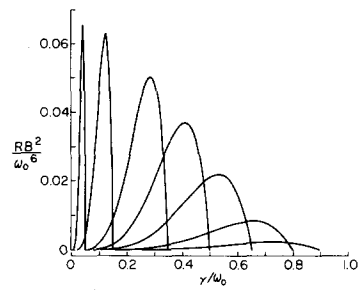


Fig. 6(b)

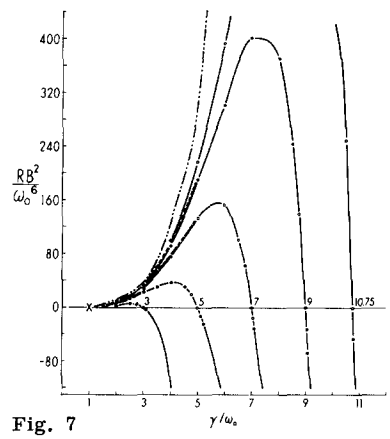


Fig. 7

FIGS. 6 and 7. Plots of RB^2 versus γ at constant $\sigma\equiv\gamma/\nu\ (=m_2/m_1)$. The σ value can be read off from the intersection with the $B=0$ axis, since $\gamma/\omega_0=\sigma$ there. The dotted curves are the zeroth order result (5.16). The dot-dot-dash line is our estimate of the boundary of the bounded and unbounded solutions (i.e., the confluence of all our large-σ lines). To within the precision of Fig. 7, it coincides with the graph of the zeroth order result (5.16) at $\nu=0$, i.e., $RB^2=4\gamma^4\omega_0^2/9$. Fig. 6(b): Same as Fig. 6(a), but only the zeroth order results are plotted.

These "zeroth" order results contain already a great deal of information. First we observe that, as in (3.7), we must have $\nu < \omega_0$ if $R > 0$ and $\nu > \omega_0$ if $R < 0$. Second, Eq. (5.16) provides a good estimate of the relation between the driving amplitude B and the two frequencies γ and ν. It is apparent from (5.16) that $\nu^2 = \nu^2(RB^2)$ *is a two-valued function of* RB^2, when $R > 0$, while RB^2 and θ are single valued functions of ν. Our methods are in fact series expansions and yield one unique result for each choice of parameters inside a region of convergence. Therefore, we expect the "backward" scheme, in which we specify ν rather than B, to have a much larger region of convergence than any "forward" scheme. Using a "forward" scheme, we have only been able to recover one "branch" of the two valued functions mentioned above, cf. Figs. 6—9.

D. Quantitative results

Our results, after 14 or more iterations, are displayed in Figs. 6—9. We plot curves of RB^2 versus γ, *at constant* $\sigma \equiv \gamma/\nu\ (= m_2/m_1)$ [cf. (5.16)–(5.17) and the discussion in Sec. 4B]. The points on each curve are parametrized continuously by ν (or ν_r) and for $\nu = \omega_0$, i.e., the harmonic case, all curves cross the horizontal axis, $B = 0$, at $\gamma = \omega_0\sigma$. These curves of constant rational σ cannot intersect, by definition, and must all be "parallel." For $\nu > \omega_0$ the curves drop below the horizontal axis.

We first present a *heuristic* argument explaining the general features of Figs. 6—9. Notice that there is no bounded solution at $\gamma = \omega_0$, whence $B = 0$ there. In the region near $\gamma = \omega_0$, and for small B values, we expect small displacements, i.e., we expect the solution to approach the solution of the *harmonic* approximation

$$\ddot{\phi} \approx -\omega_0^2\phi + B\cos\gamma t, \quad \text{hence } [(\phi(0) = \dot{\phi}(0) = 0], \tag{5.18}$$

$$\phi(t) \approx [B/(\gamma^2 - \omega_0^2)](\cos\omega_0 t - \cos\gamma t). \tag{5.19}$$

This solution becomes unbounded if we let $\gamma \to \omega_0$. However, in order for the solution of the *an*harmonic oscillator (5.1) to stay inside or near the potential valley [cf. $R > 0$ in Fig. 1(a)], we must have $\theta \equiv R\phi^2(0)/\omega_0^2 \lesssim 1$ or

$$(4R/\omega_0^2)[B/(\gamma^2 - \omega_0^2)]^2 \lesssim 1, \tag{5.20}$$

cf. (2.6). So if we let $\gamma \to \omega_0$ we expect that $B \to 0$ *in order for the solution to stay bounded.* The argument becomes stronger the smaller the expression in (5.20). We call this case, with $\gamma \approx \omega_0$, "ω_0-*resonance.*"

Another notable feature is the large difference in RB^2 values between $\gamma > \omega_0$ and $\gamma < \omega_0$, cf. Figs. 7 and 6 (or 9 and 8). In order to heuristically explain this, we observe that in the actual *an*harmonic oscillator many Fourier frequencies $n\nu_r$ are generated, cf. (5.6). Each of these can exhibit an "ω_0-resonance" similar to the one above (now between $n\nu_r$ and ω_0). Analogous to (5.20) we expect the solution to be unbounded unless

$$\left(\frac{4R}{\omega_0^2}\right)A_n^2 \approx \left(\frac{4R}{\omega_0^2}\right)\left[\frac{B}{(n\nu_r)^2 - \omega_0^2}\right]^2 \lesssim 1. \tag{5.21}$$

The main response frequency $\nu \equiv m_1\nu_r$ can be identical to ω_0 only if $B = 0$, according to (5.21). We do indeed find that all σ curves cross the $B = 0$ axis when $\nu = \omega_0$. Let us follow any given σ curve in Fig. 7. Its points were parametrized by the values of ν_r and, for $R > 0$ as in Fig. 7, we expect (cf. Secs. 2 and 3) and obtain $\nu \lesssim \omega_0$. Going up along a σ curve, the ν and ν_r decrease as larger and larger B values are permitted, according to (5.21). However, when $n\nu_r \approx \omega_0$ for *another* integer n, different from m_1, a *reduction* in B^2 is required to satisfy (5.21). We do indeed see the B^2 increase, along a σ curve, reach a maximum and then fall again as $\nu_r \to 0$. Thus the ν, ν_r, and γ are double valued functions of RB^2 along each σ curve in Fig. 7.

In the region $\gamma < \omega_0$ (Fig. 6) multiples of γ, as well

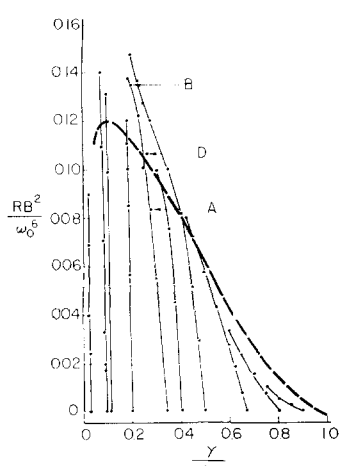


Fig. 8

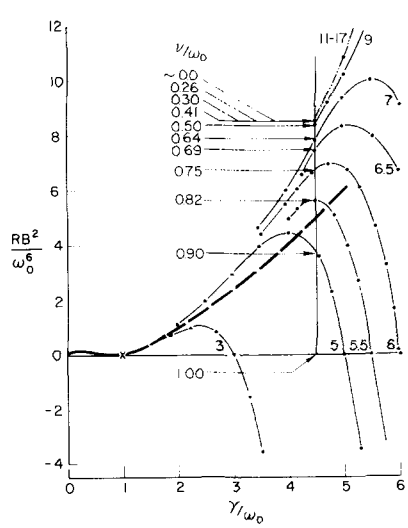


Fig. 9(a)

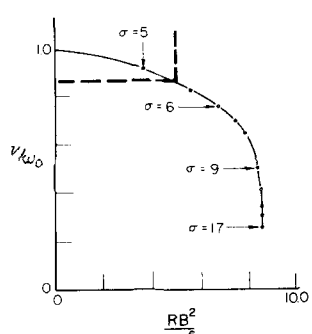


Fig. 9(b)

FIGS. 8 and 9(a). Plots of RB^2 versus γ at constant $\sigma \equiv \gamma/\nu(m_2/m_1)$, with a *sine drive*, i.e., $\sin(\gamma t)$ replacing $\cos(\gamma t)$ in (5.1). The σ values can be read off from the intersection with the $B = 0$ axis. The heavy dashed lines indicate Davis' numerical estimate of the boundary between bounded and unbounded solutions.[9] The dot-dot-dash line indicates our estimate of the boundary between bounded and unbounded solutions. The analog of (5.16), i.e., the zeroth order approximation, is less useful in the case of the sine drive. FIG. 9(b): Plot of ν/ω_0 versus RB^2/ω_0^6 for $\gamma/\omega_0 = 4.5\omega_0$, indicated in Fig. 9(a) by a vertical solid line. Note the precipitous drop with increasing σ values. The region to the left and above the heavy dashed lines indicates Davis' region of bounded solutions. Compare Fig. 9(b) with Fig. 3.

as ν, can come close to ω_0 and give rise to ω_0-resonance. Since $\gamma = m_2\nu_r$, the Fourier coefficient A_{m_2} must be smaller, according to (5.21), than in the case $\gamma > \omega_0$. Thus the main coefficients A_{m_1} and A_{m_2} are *both* severely restricted in the region $\gamma < \omega_0$, whence the difference in RB^2 magnitude between Figs. 6(a) and 7. If we follow a σ-curve in Fig. 6(a), the B^2 can only increase for γ values further away from ω_0, according to (5.20). Hence the σ-curves in Fig. 6(a) bend to the left.

In the region $\gamma > \omega_0$ the zeroth order values of RB^2 [(5.16)], indicated by dashed lines in Fig. 7, virtually coincide with the calculated values over most of the σ curve. The boundary between the region of bounded and unbounded solutions therefore nearly coincides with the graph of Eq. (5.16) at $\nu = 0$, i.e., $RB^2 \approx 4\gamma^4\omega_0^2/9$. Only when γ approaches ω_0 does the zeroth order become discernable from the calculated values. This excellent agreement of calculated and zeroth order values is due to the rapid rate of convergence in the region $\gamma > \omega_0$. While the zeroth order curves do display the general features for $\gamma > \omega_0$, this is *not* the case for $\gamma < \omega_0$. Since the σ curves cannot intersect, they cannot follow very far the zeroth order curves of RB^2 when $\gamma < \omega_0$, which do intersect, cf. Fig. 6(b) and (5.16). In Fig. 6(a) we see the calculated values deviate noticeably from the zeroth order, dashed lines, long before those reach their maximum and extend well above that value. Clearly a better choice of zeroth order terms would be advantageous,[16] when $\gamma < \omega_0$.

Davis, in Ref. 9, used various numerical methods to integrate the equation of motion and determine an estimate of the boundary between the bounded and unbounded solutions for the Duffing equation with a *sine* drive. His estimate is indicated by a heavy dashed line in Figs. 8- and 9(a). Especially for $\gamma > \omega_0$, we find a sizeable additional region of bounded solutions, cf. Fig. 9(a). We suggest the following heuristic explanation of why ordinary numerical integration of the equation of motion might be unstable in the additional region we find. Let us follow the behavior of the solution along a vertical line, e.g., at fixed $\gamma = 4.5\omega_0$, in Fig. 9(a). As RB^2 is increased, the ν decreases, cf. Figs. 9(a) and 9(b). Note that ν decreases little in the region also found by numerical integration while dropping precipitously in the additional region we find. In this additional region the ratio of the two time scales of the problem, i.e., $2\pi/\nu$ and $2\pi/\gamma$, increases faster and faster as we increase RB^2 (along $\gamma = 4.5\omega_0$). A very small error in B will, in this region, result in a large relative error in ν, cf. Fig. 9(b), and, in a numerical integration, the displacement after the next time step will also have a large relative error. (Ordinary numerical integration should proceed in time steps, small compared to the time scales of the problem.) Thus one expects that the more the two internal time scales of our problem diverge, the more erratic the results of such an integration become. In our Fig. 9 this would correspond to an increasingly erratic behavior for larger and larger RB^2 values and a "safe" *numerical* estimate would have to be considerably lower than the actual boundary.

In addition, Davis found a "jump-like" solution, i.e.,

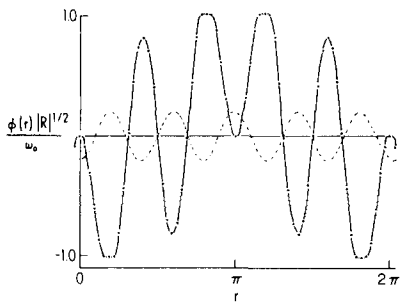


FIG. 10(a). $\gamma/\omega_0 = 1.5$, $\nu/\omega_0 = 0.30$. Hence $m_1 = 1$, $m_2 = 5$, $\sigma = 5$, $\nu_r/\omega_0 = 0.30$ and from our scheme $K = 0.758$, $\theta = 0.574$ and $B = -1.637$.

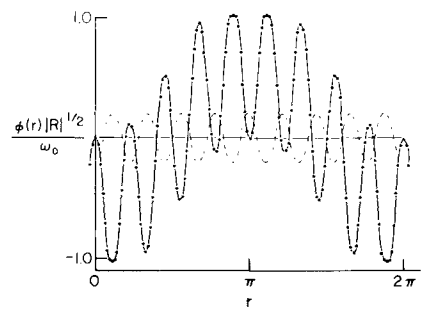


FIG. 10(b). $\gamma/\omega_0 = 4.0$, $\nu/\omega_0 = 4/9$. Hence $m_1 = 1$, $m_2 = 9$, $\sigma = 9$, $\nu_r/\omega_0 = 0.44$, and from our scheme $K = 6.067$, $\theta = 0.368$ and $B = -95.87$.

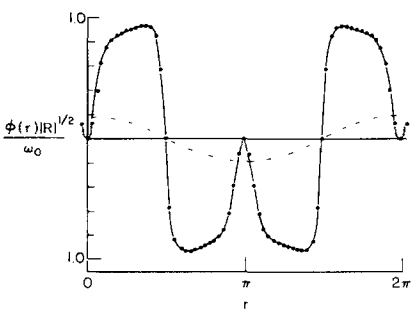


FIG. 10(c). $\gamma/\omega_0 = 0.10$, $\nu/\omega_0 = 0.3$. Hence $m_1 = 3$, $m_2 = 1$, $\sigma = 1/3$, $\nu_r/\omega_0 = 0.10$ and from our scheme $K = 3.516$, $\theta = 12.36$ and $B = 0.281$.

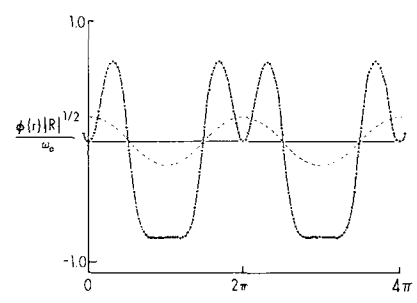


FIG. 10(d). $\gamma/\omega_0 = 0.275$, $\nu/\omega_0 = 0.550$. Hence $m_1 = 2$, $m_2 = 1$, $\sigma = 1/2$, $\nu_r/\omega_0 = 0.275$ and from our scheme $K = 1.310$, $\theta = 1.72$ and $B = 0.298$.

FIG. 10. The waveform $\phi(r)$. The time r is scaled in such a way that the recurrence time is 2π. The dotted curves represent the $B\cos(m_2 r)$ driving term, cf. (5.2). Note that in Figs. 10(a) and 10(b) $|\phi||R|^{1/2}/\omega_0 > 1$ at some times, i.e., the particle is outside the potential "valley" and on the "down slope". All K, θ, and B values are calculated with $\omega_0 = 1$ and $R = 1$.

oscillations which seem to "recur" several times and then, suddenly, become unbounded. Near the data point indicated by him ($RB^2/\omega_0^2 = 0.64/6$ and $\gamma/\omega_0 = 0.27$; point D in Fig. 8) we obtain nothing but bounded solutions, using both "forward" and "backward" schemes. While point D lies inside our additional region of bounded solutions (and just outside his), it is indeed close to the boundary we find, and we suggest the same explanation as above.

Yet we do expect a seemingly similar behavior just inside the region of *un*bounded solutions, i.e., there might be many oscillations, but no exact recurrences, before the "jump." We suggest points "near" those graphed in Figs. 10(a) and (b), which are "bounded." Notice that the displacement here is occasionally larger than 1.0, which places the particle *outside the potential valley on the down-hill slope*. Evidently a particle can periodically leave the valley and be forced back into it by the driving force. Without the driving force the displacement would become unbounded.

Some other representative waveforms are displayed in Figs. 10—11. The driving term, either $\cos(m_2 r)$ or $\sin(m_2 r)$, is depicted by a dotted line (amplitude is not to scale). For a cosine drive with $\gamma > \omega_0$ the apparent "phase" of displacement and driving term are roughly opposite over most of the recurrence period, cf. Figs. 10(a) and 10(b). It is here that we find the bounded solutions leaving the potential valley. For a cosine drive with $\gamma < \omega_0$ we find that the "phases" coincide most of the the time. Near $r = k\pi$ (k integer) a sudden "bounce" occurs, cf. Figs. 10(a—d). The corresponding waveforms for the sine drive seem less interesting in that the apparent "phases" virtually coincide, cf. Figs. 11(a—d).

Whenever the number of significant Fourier coefficients becomes large, the computation becomes longer and more costly. We never evaluated any data point requiring more than a few minutes of computing time. For most cases the behavior of the σ curves is fairly predictable and therefore not profitable to pursue. In the region $\gamma \approx 0$, in Fig. 8, some new "ad hoc" techniques would be required to continue the σ curves, within reasonable computing times. One fairly general technique will be discussed in the next section.

E. Computational techniques

As we proceed to calculate the values of RB^2 along a σ curve starting at $\nu = \omega_0$, the efficiency of the iteration procedure decreases, due to the increase in the number of "significant" Fourier coefficients. In some of those cases we found it advantageous to iterate the recursion relations (5.8) several times without including any new coefficients generated by (5.8). This procedure improves the estimate of the coefficients so far obtained before introducing new ones thereby improving the efficiency of the calculation. We refer to this procedure as "recycling."

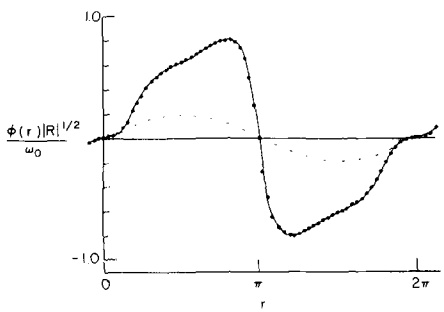


FIG. 11(a) $\gamma/\omega_0 = 0.20$, $\nu/\omega_0 = 0.40$. Hence $m_1 = 2$, $m_2 = 1$, $\sigma = 1/2$, $\nu_r/\omega_0 = 0.20$ and from our scheme $K = 7.861$, $\theta = 10.3$ and $B = 0.942$.

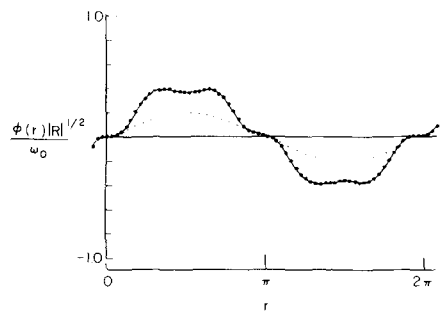


FIG. 11(b). $\gamma/\omega_0 = 0.181$, $\nu/\omega_0 = 0.905$. Hence $m_1 = 5$, $m_2 = 1$, $\sigma = 1/5$, $\nu_r/\omega_0 = 0.181$ and from our scheme $K = 1.042$, $\theta = 0.181$ and $B = 0.820$.

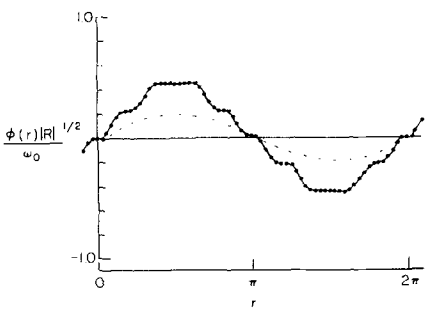


FIG. 11(c). $\gamma/\omega_0 = 0.095$, $\nu/\omega_0 = 0.855$. Hence $m_1 = 9$, $m_2 = 1$, $\sigma = 1/9$, $\nu_r/\omega_0 = 0.095$ and from our scheme $K = 1.227$, $\theta = 0.251$ and $B = 0.887$.

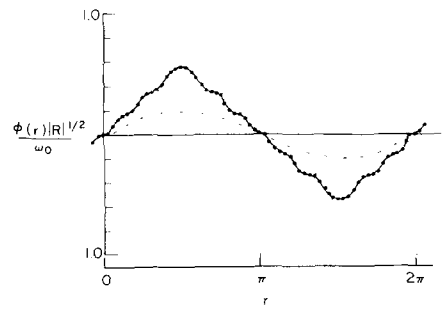


FIG. 11(d). $\gamma/\omega_0 = 0.070$, $\nu/\omega_0 = 0.770$. Hence $m_1 = 11$, $m_2 = 1$, $\sigma = 1/11$, $\nu_r/\omega_0 = 0.070$ and from our scheme $K = 2.341$, $\theta = 0.913$ and $B = 1.377$.

FIG. 11. The waveform $\phi(r)$ for the case of a sine drive. The time r is scaled in such a way that the recurrence time is 2π. The dotted curves represent the $B\sin(m_2 r)$ driving term, cf. (5.2) with cosine replaced by sine. All K, θ, and B values are calculated with $\omega_0 = 1$ and $R = 1/6$.

TABLE IV. Two representative cases for the Duffing equation (cf. Fig. 8).

Chosen parameters[a]	Order/recycles (total iterations)[b]	RB^2 [c]	max{relative error} over one recurrence period[d]	$\Delta A_{m_1}/A_{m_1}$ [e]	$\Delta A_{m_2}/A_{m_2}$ [e]	Computing time IBM 360/65
$\nu_r = 0.275$						
point A in Fig. 8	0	$5.\underline{01}\times10^{-2}$				0.0 s
	4/0	$8.\underline{343}\times10^{-2}$	1.6×10^{-2}	5.2×10^{-3}	1.9×10^{-3}	0.5 s
	l. c.	$\approx -3.\underline{8}\times10^{-4}$				
$m_1 = 3$						
$m_2 = 1$	9/0	$8.348\,9837\times10^{-2}$	-2.9×10^{-6}	5.2×10^{-6}	-3.4×10^{-7}	3 s
$w = 1.0$	l. c.	$\approx 6.\underline{9}\times10^{-8}$				
$\omega_0 = 1.0$						
$\gamma = 0.275$	14/0	$8.348\,982\,\underline{71}\times10^{-2}$	5.2×10^{-10}	1.1×10^{-7}	6.4×10^{-9}	6 s
$\nu = 0.825$	l. c.	$\approx -7.\underline{1}\times10^{-9}$				
$\nu_r = 0.200$	0	$\underline{2.8}\times10^{-2}$				0.0 s
point B in Fig. 8						
	4/0	$1.\underline{351}\times10^{-1}$	2.1×10^{0}	-1.9×10^{-1}	-8.7×10^{-1}	0.5 s
	(4) l. c.	$\approx 5.\underline{6}\times10^{-3}$				
$m_1 = 3$						
$m_2 = 1$	4/4	$1.\underline{360}\times10^{-1}$	-2.6×10^{-1}	2.2×10^{-2}	2.7×10^{-2}	2 s
$w = 0.4$	(8) l. c.	$\approx -3.\underline{3}\times10^{-4}$				
$\omega_0 = 1.0$						
$\gamma = 0.200$	4/9	$1.355\,\underline{17}\times10^{-1}$	7.6×10^{-3}	1.0×10^{-3}	1.8×10^{-3}	4 s
$\nu = 0.600$	(13) l. c.	$\approx -2.\underline{1}\times10^{-6}$				
	4/14	$1.354\,980\underline{1}\times10^{-1}$	-1.5×10^{-3}	-2.8×10^{-6}	5.1×10^{-6}	6 s
	(18) l. c.	$\approx -3.\underline{7}\times10^{-7}$				
	4/19	$1.354\,980\,\underline{40}\times10^{-1}$	-3.6×10^{-5}	-3.1×10^{-6}	-1.4×10^{-7}	8 s
	(23) l. c.	$\approx 3.\underline{7}\times10^{-8}$				
	5/4	$1.354\,981\,\underline{15}\times10^{-1}$	8.2×10^{-6}	-3.1×10^{-7}	-3.9×10^{-8}	10.5 s
	(28) l. c.	$\approx 5.\underline{5}\times10^{-9}$				
	5/9	$1.354\,981\,\underline{236}\times10^{-1}$	-5.9×10^{-7}	-2.2×10^{-8}	-4.6×10^{-9}	12.5 s
	(33) l. c.	$\approx -2.\underline{3}\times10^{-10}$				
	5/14	$1.354\,981\,242\,\underline{25}\times10^{-1}$	-3.8×10^{-8}	-1.4×10^{-9}	-3.1×10^{-10}	14.5 s
	(38) l. c.	$\approx 2.\underline{2}\times10^{-11}$				
	5/19	$1.354\,981\,242\,\underline{633}\times10^{-1}$	4.9×10^{-8}	5.6×10^{-11}	2.4×10^{-11}	16.5 s
	(43) l. c.	$\approx 3.\underline{8}\times10^{-13}$				
	9/0	$1.354\,981\,242\,652\,\underline{45}\times10^{-1}$	9.7×10^{-9}	2.6×10^{-12}	-1.7×10^{-12}	19 s
	(47) l. c.	$\approx -5.\underline{0}\times10^{-14}$				
	14/0	$1.354\,981\,242\,6\underline{525}\times10^{-1}$	1.1×10^{-9}	4.3×10^{-13}	-2.6×10^{-13}	22 s
	(52) l. c.	$\approx 1.\underline{1}\times10^{-14}$				
Final Fourier coefficients after 52 iterations (order of magnitude only)		$A_1 \approx 2.\underline{2}\times10^{-1}$	$A_3 \approx -3.\underline{2}\times10^{-2}$	$A_5 \approx -2.\underline{0}\times10^{-2}$	$A_7 \approx -2.\underline{9}\times10^{-3}$	
		$A_9 \approx 2.\underline{6}\times10^{-5}$	$A_{11} \approx 7.\underline{7}\times10^{-5}$	$A_{13} \approx 7.\underline{2}\times10^{-6}$	$A_{15} \approx -2.\underline{3}\times10^{-6}$	(the others
		$A_{17} \approx -7.\underline{5}\times10^{-7}$	$A_{19} \approx -6.\underline{3}\times10^{-8}$	$A_{21} \approx 1.\underline{3}\times10^{-8}$	$A_{23} \approx 3.\underline{8}\times10^{-9}$	are smaller)

[a] w is the value of the "smoothing" parameter in the text.

[b] At zeroth order we have only two Fourier components A_{m_1}, A_{m_2}. At each successive order additional nonzero coefficients will be generated. "Recycles" refers to the number of iterations performed at a fixed order without using additional coefficients (see text). The number in parentheses is the total number of iterations performed up to this point, e.g., in going to fifth order we generate additional Fourier coefficients and do four recycles (total iterations 28). Then we do five more recycles (total iterations 33). l. c.: last correction ΔRB^2 = last value of RB^2 minus the previous iteration value.

[c] Underlined numbers mean questionable precision in view of higher-order (iteration) corrections or machine "round off."

[d] We substitute the solution, at this point, in the equation of motion (5.1) and we take all terms to the rhs and define the relative error$(t) \equiv$ rhs $(5.1)(t)/[(\nu^2 - \omega_0^2)\phi(t)]$ when $\phi(t) \neq 0$. We scale in $(\nu^2 - \omega_0^2)\phi(t)$ because it is characteristic of the magnitude of the terms in the equation.

[e] These are a measure of the rate of convergence where ΔA_i is the difference in A_i between successive iterations.

In Table IV two representative cases are displayed. For the first case, the rate of convergence is quite good, most of the Fourier coefficients are small and no recycling was used. In the second, more coefficients are significant, cf. Table IV, and nineteen recycles were used at order 4 and 5. Note the improvement in the "relative error," etc., with recycling.

In many cases an additional "smoothing" procedure was used to improve the rate of convergence, which

permitted the averaging of the present and the preceding iteration. The procedure is discussed in Appendix A of Ref. 16 and is, in the theory of iterations of *linear* systems, known as a "relaxation method." The smoothing is characterized by the w in Table IV, with $w=1$ corresponding to no smoothing and $w=\frac{1}{2}$ to equal weighting of the two iterations.

For $\gamma < \omega_0$ the rate of convergence might be improved by proving a better zeroth order estimate, cf. Figs. 6(a) and 6(b), referred to in Ref. 16 as a different "starting" vector $\mathbf{A}$ [with $(\mathbf{A})_n \equiv A_n$]. Here we have not pursued this approach.

6. DISCUSSION

We have demonstrated a method for solving the dynamics of systems of coupled anharmonic oscillators, with polynomial force laws, employing discrete Fourier series. Several convergent[17] perturbative and iterative techniques are described, and applied, yielding the Fourier frequencies and coefficients of those series. Even at arbitrarily large energy and nonlinearity we are able to do the perturbation expansion about a function "close" to the exact solution. Thus we obtain a rapid rate of convergence and, equally important, simple analytic formulas which are quick and accurate approximations to the exact solutions even in extremely nonlinear systems.

We avoid the divergencies connected with the classic "small denominators" problem in the series analysis of coupled systems by concentrating on all periodic solutions, i.e., solutions with rationally dependent Fourier frequencies, cf. Secs. 1 and 4. These periodic solutions are dense among all possible solutions and suffice for our practical calculations here (just as the rational numbers suffice for most practical computations). An application to the Duffing equation, a driven anharmonic oscillator appearing in certain laser models[8] or the driven pendulum,[9] is given in Sec. 5. We will discuss some transport properties of larger systems of anharmonic oscillators elsewhere.[15] Currently we are investigating the corresponding quantum-theoretical perturbation theory.

ACKNOWLEDGMENTS

One of us (R.H.) would like to thank Dr. Jürgen Moser for some illuminating conversations and references on nonlinear differential equations. We thank Dr. Lakshmanan for kindly informing us of his solution in closed form of an anharmonic oscillator with positive ϕ^4 and ϕ^6 terms in the Hamiltonian,[19] cf. Ref. 16. This study was partially supported by the N.S.F. under Grant No. DID71-04010 A02. *Note added in proof*: Dr. Moser informed us that our techniques might also be useful, *in principle*, for the case of the *independent* frequencies, for which he has suggested a similar approach.[20]

[1]G. Stokes, "On the Theory of Oscillatory Waves" (1847), in *Mathematics and Physics Papers* (Cambridge U. P., Cambridge, 1905), Vol. I, p. 197 [originally: Trans. Camb. Phil. Soc. 8, 441].
[2]H. Poincaré, *Les méthodes nouvelles de la mécanique céleste* (1892) (Dover, New York, 1957).
[3]M. Lindstedt, Mém. Acad. Roy. Sci. St. Petersburg, Ser. 7, 31, No. 4, 438 (1883).
[4]R. Helleman and E. Montroll, Physica 74, 22 (1974); Errata:
p. 66 (3rd line from bottom), "integral" should be "integrand."
p. 67, "(A.24a) = (A.24b)" should be "≥."
p. 72, "$b_n = \sum_{k=n-1}$" should be "$\sum_{k=0}$."
[5]J. Moser, *Stable and Random Motions in Dynamical Systems* (Princeton U.P., Princeton, N.J., 1973).
[6]J. Moser, Mem. Am. Math. Soc. 81, 1 (1968).
[7]C. Siegel and J. Moser, *Lectures on Celestial Mechanics* (Springer, New York, 1971).
[8]W. Lamb, "Quantum Mechanical Amplifiers" (especially lecture 3), in *Lectures in Theoretical Physics, Vol. II* (Boulder, Colorado, 1959), edited by Britten and Downs (Interscience, New York, 1960).
[9]H. Davis, *Introduction to Nonlinear Differential and Integral Equations* (Dover, New York, 1962), especially Sec. 12.9.
[10]J. J. Stoker, *Nonlinear Vibrations* (Interscience, New York, 1950); J. Moser, Comm. Pure Appl. Math. 17, 167 (1965).
[11]M. Ablowitz, B. Funk, and A. Newell, Studies Appl. Math. 52, 51 (1973); S. Aks, B. Varga, and J. Sienicki, J. Math. Phys. 16, 1928 (1975); S. Aks, Fortschr. Phys. 15, 661 (1967).
[12]K.O. Friedrichs, *Advanced Ordinary Differential Equations* (Gordon and Breach, New York, 1965).
[13]R. McGehee and K. Meyer, Am. J. Math. 96, 409 (1974).
[14]D. Hsieh, J. Math. Phys. 16, 275 (1975).
[15]C. Eminhizer and R. Helleman, "Fermi—Pasta—Ulam Chains," in preparation.
[16]C. Ginsburg, C. Eminhizer, and R. Helleman, "On a Polynomial Anharmonic Oscillator," preprint (1975); E. Montroll and R. Helleman, in press, to be published in AIP Conference Proceedings, Memorial Volume to Julius Jackson.
[17]R. Helleman (in preparation).
[18]F. Gustavson, Astron. J. 71, 670 (1966); T. Harris, Int. J. Non. Lin. Mech. 5, 491 (1970); J. Diff. Eq. 4, 131 (1968).
[19]M. Lakshmanan, Lett. Nuovo Cimento 8, 743 (1973).
[20]J. Moser, in *Stability of the Solar System*, edited by Kozai (I.A.U., 1974), pp. 1—9.

On the positive energy conjecture*

Pong Soo Jang†

Laboratory for Astrophysics and Space Research and Department of Physics, University of Chicago, Chicago, Illinois 60637
(Received 21 May 1975)

A theorem, to the effect that the positive energy conjecture is true for any space–time possessing a flat spatial hypersurface, is obtained.

I. INTRODUCTION

With any isolated gravitating system in general relativity there is associated a certain number E, whose physical interpretation is "the total energy of the system, including contributions from both the matter and the gravitational field." For some ten years there has stood a conjecture to the effect that, if the local energy density of matter is nonnegative everywhere then this total energy E is also nonnegative.

This conjecture is of interest for several reasons. First, the question of positivity of gravitational energy is an important one of principle: perhaps the first step in understanding the role of energy in general relativity would be to know its sign. Second, the issue might have an impact on the problem of quantization of the gravitational field. The existence of any negative energy state for a gravitating system implies nonexistence of a lower bound for the energy, and this could well cause a runaway problem with any quantization procedure. Third, the issue might even have an application for astrophysical problems: Should the conjecture turn out to be false, so a certain configuration of the gravitational field possesses negative energy, such a configuration might then act as a source which satisfies the enormous energy requirements of certain astrophysical systems.

There are, in fact, two distinctive regimes in which the total energy E can be defined. The first is at spatial infinity, i.e., in the limit as one approaches the asymptotic region "along spacelike directions." In more detail, the Arnowitt–Deser–Misner energy[1] is defined in terms of the behavior of initial data on a spacelike three-dimensional surface which, asymptotically, behaves as a $t=$ constant plane in Minkowski space. The second regime is at null infinity, i.e., in the limit as one approaches the asymptotic region "along null directions." In more detail, the Bondi energy[2] is defined in terms of the rate of approach of the metric to flatness on a null surface which, asymptotically, behaves as a null cone in Minkowski space. Hence, there are in fact, two energy conjectures depending on whether "total energy" refers to the Arnowitt–Deser–Misner (ADM) energy or the Bondi energy.

The conjectures appear to be quite difficult. In the Bondi case, very little is known. In the ADM case, on the other hand, there are several special cases available, in certain of which the conjecture has been answered in the affirmative. For example, Brill[3] showed that the conjecture is true (i.e., the ADM energy is indeed nonnegative) if the system is axially symmetric and, in addition, possesses a moment of time-symmetry.

We shall confine ourselves here to the ADM energy conjecture, i.e., that at spatial infinity. In Sec. II, we formulate the conjecture and discuss special cases. In Sec. III, we obtain a new theorem which asserts that the conjecture is true for any space-time possessing a flat spatial hypersurface. In Sec. IV, we point out the similarity in the structure of the proof of this theorem and Geroch's argument for another special case (Ref. 4), and discuss the possibility of combining these arguments, i.e., the possibility of obtaining a proof for the general case. Finally, in the Appendix, we obtain a theorem which asserts that the conjecture is true in the presence of spherical symmetry.

II. POSITIVE ENERGY CONJECTURE AT SPATIAL INFINITY

It is known[1] that associated with any asymptotically flat initial data for Einstein's equation there is a certain number E, which depends explicitly only on the induced metric of the initial hypersurface. This number E has the physical interpretation of the total energy of the system, including contributions from both the matter and the gravitational field. The positive energy conjecture states that if the matter satisfies a certain local energy condition then this energy E is nonnegative and, furthermore, is zero when and only when the initial data are data for Minkowski space.

Unfortunately, some technical complications in the definitions of asymptotic flatness and of the energy tend to obscure the essential features of the conjecture. However, for a certain class of initial data, the issue of the conjecture can be isolated from these complications. In particular, for initial data whose Riemann tensor and extrinsic curvature have compact support, the conjecture can be formulated without using the detailed definitions of asymptotic flatness and of energy: This data is automatically asymptotically flat and automatically has zero energy E. Hence, the conclusion of the conjecture in this case, is that the initial data is data for Minkowski space. A precise statement of the conjecture for this class of initial data is as follows:

Conjecture 1: Let S be the manifold R^3 and let C be a compact subset of S. Let q_{ab} be a positive-definite metric on S such that $(S-C)$ with metric q_{ab} is isometric to the complement of some compact subset of three-dimensional Euclidean space. Let p^{ab} be a smooth symmetric tensor field on S, whose support is in C. Let $\mu \geq (J^m J_m)^{1/2}$, where μ and J^a are defined by

$$2\mu = R - p^{ab}p_{ab} + \tfrac{1}{2}p^2, \tag{1}$$

$$J^a = D_b p^{ab}, \tag{2}$$

where D_a is the derivative operator associated with q_{ab}, R is the scalar curvature of q_{ab}, and $P = P^{ab}q_{ab}$. Then μ and J^a vanish and

$$R^{ab} = p^{am}p^{b}{}_{m} - \tfrac{1}{2}pp^{ab}, \tag{3}$$

$$D^{[a}(p^{b]c} - \tfrac{1}{2}pq^{b]c}) = 0, \tag{4}$$

[i.e., $(S, q_{ab}, p^{ab}, \mu, J^a)$ are data for Minkowski space].

The above conjecture is purely mathematical, it is stated without reference to space-time. Its connection with space—time, however, is this. The S above represents a three-dimensional hypersurface in the space-time manifold, and q_{ab} represents the metric of S induced from the metric of space—time. The p^{ab} is related to the extrinsic curvature $\Pi^{ab}{}_{,}$ of S by $p^{ab} = \Pi^{ab} - (\Pi^{mn}q_{mn})q^{ab}$. The μ and J^a, obtained by projecting the stress-energy tensor into S, have the physical interpretation of the "local energy density" and the "momentum density of matter," respectively. Einstein's equation on space-time imposes the constraint Eqs. (1) and (2) on the induced quantities on S, i.e., on the q_{ab}, p^{ab}, μ, and J^a. The inequality, $\mu \geq (J^aJ_a)^{1/2}$, is the "local energy condition." Finally, (3) and (4) are the sufficient and necesary condition for initial data to be data for Minkowski space.

Why did we choose $\mu \geq (J^aJ_a)^{1/2}$ as the local energy condition? The truth of our conjecture depends crucially on what condition we impose on the stress—energy tensor. If, for instance, there were permitted matter with negative local energy density μ, then the conjecture so modified would certainly be false. However, since no matter with negative μ has ever been observed, it would not be unreasonable to impose the condition, $\mu \geq 0$. But here we impose an even stronger condition,[5] $\mu \geq (J^aJ_a)^{1/2}$. This choice seems to be an appropriate one for the energy conjecture for the following reasons. First, the nonnegativity of μ is not sufficient to exclude initial data with negative total energy.[6] Second, this condition seems to be the only alternative: There is no other condition on μ and J^a which is stronger than $\mu \geq 0$ and which is still physically reasonable.

What is the point of the assumption that the Riemann tensor and p^{ab} have compact support on S? This assumption simplifies the conjecture to formulate it without explicitly using the definitions of asymptotic flatness and of energy. Even though we restricted the class of initial data by this assumption, the above conjecture nonetheless, should retain the essential features of the full conjecture. To decide whether this is the case, we are led to consider the following two questions: Is a counter example to Conjecture 1 a counter example to the full conjecture?; will a proof of Conjecture 1 yield a proof of the full conjecture? To the first question, the answer is, of course, yes. Furthermore, if there is a counter example to Conjecture 1, then there is even an asymptotically flat system with negative total energy: Brill and Deser have shown that any asymptotically flat initial data which is not a data for Minkowski space has a variation which decreases the total energy.[7] To the second question, the answere is not easy. However, every known proof of a special case of Conjecture 1 has suggested an immediate generalization of a proof of the corresponding special case of the full conjecture. That is to say, the structural features which make Conjectur 1 true are exactly the structural features which make th full conjecture true in those special cases. We suspect that the situation will persist even into the general case. Apparently, therefore, our Conjecture 1, although it is logically a special case of the full conjecture, nonetheless retains perhaps the essential features of the full conjecture.

TABLE I. Special cases of the positive energy conjecture.[a]

	q_{ab}		
P_{ab}	Flat	Conformally flat	General
$p^{ab} = 0$	1 √	4 √	7 ?*
$p^{ab} = (p/3)\, q^{ab}$	2 √	5?	8 ?
General	3 √	6?	9 ?

[a] A check indicates that the conjecture has been proved for that special case, while a question mark indicates that that case is open. A star indicates that the conjecture has been proved for axially sysmmetric q_{ab} and p^{ab} in that special case Finally, for spherically symmetric q_{ab} and p^{ab} the conjecture has been proved for every case of the table.[8]

Conjecture 1 appears to be quite difficult. One can, however, invent a number of special cases; one can impose various additional conditions on q_{ab} and/or p^{ab}. It is hoped that perhaps the proofs of a sufficient number of these special cases could be combined and generalized to yield a proof of Conjecture 1. Some of the special cases are summarized in Table I.

The status of each special case is indicated in the table. As one moves downwards or rightwards in the table, one moves to increasing generality. Case 3 of the table is called the "pure kinetic case." In the next section we shall give a proof for this case. Proofs for cases 1 and 2 are immediate, while case 3 has been treated by Brill and Deser[7] but with the additional assumption that $J^a = 0$.

Case 7 of the table is called the "pure potential case." There are two special cases of 7 in which the conjecture has been proved: That in which the metric is conformally flat[9] (i.e., 4 of the table) and that in which the metric is axially symmetric.[3] Finally, the general spherically symmetric case has been treated by both Misner and Israel[10]. Both of these arguments involve some special choice of three-surface in the space-time, and hence strictly speaking, these arguments are not proofs of the spherically symmetric case of our Conjecture 1. In the Appendix, we shall give an argument in which the theorem and proof are formulated entirely in terms of initial data.

III. THE PURE KINETIC CASE

We now prove the positive energy conjecture in the pure kinetic case, i.e., when the metric q_{ab} is flat and p^{ab} is general.

Theorem: Let p^{ab} be a smooth symmetric tensor field with compact support on R^3 (with its positive-definite flat metric q_{ab}). Let $\mu \geq (J^aJ_a)^{1/2}$ where

$$2\mu = -p^{ab}p^{ab} + \tfrac{1}{2}p^2, \tag{5}$$

$$J^a = D_b p^{ab}. \tag{6}$$

Then $p^{ab} = 0$ everywhere.

Proof: Introduce a family of concentric 2-spheres in R^3 and denote their radii by r. Denote by r^a the gradient of r, a unit radial vector field. Decompose $p^{ab}r_b$ relative to this r^a as follows:

$$p^{ab}r_b = vr^a + S^a, \tag{7}$$

where S^a is orthogonal to r^a. Using this expression, Eq. (5) for μ becomes

$$2\mu = -\tfrac{1}{2}v^2 + v(q^{ab} - r^a r^b)p_{ab} - 2S^a S_a - [(q^{ab} - r^a r^b)(q^{mn} - r^m r^n)p_{am}p_{bn} - \tfrac{1}{2}\{(q^{ab} - r^a r^b)p_{ab}\}^2]. \tag{8}$$

Since the last two terms on the right side of the above equation are non-positive and since μ is nonnegative, S^a must vanish whenever v vanishes. Next, contract each side of (6) with r_a and use (7) to obtain

$$r_a J^a = r^a D_a v + 2r^{-1}v - r^{-1}(q^{ab} - r^a r^b)p_{ab} + \tilde{D}_a S^a, \tag{9}$$

where $\tilde{D}_a$ denotes the intrinsic derivative within a two-sphere.

Now let $\underline{r}$ be the smallest radius outside of which $|rv| \leq 2$. Such an $\underline{r}$ always exists because v is zero outside the compact support. Denote by B, the set of points of R^3 at which $r \geq \underline{r}$ and $v \geq 0$. Combining (8), (9) and $\mu \geq (J^a J_a)^{1/2}$, we have

$$r_a D^a v + \tfrac{3}{2}r^{-1}v \geq -\tilde{D}_a S^a, \tag{10}$$

where we have used $0 \leq rv \leq 2$ in B and $(J_a J^a)^{1/2} \geq -r^a J_a$. For each value of r, denote by $B(r)$ the intersection of B and the sphere of radius r, and denote by $f(r)$ value of the integral of v over $B(r)$, so $f(r) \geq 0$.

Next, take the rate of change of $f(r)$ with respect to r to obtain

$$\frac{df(r)}{dr} = \int_{B(r)} [r^a D_a v + 2r^{-1}v]\,dA \geq \tfrac{1}{2}r^{-1}f(r) - \int_{B(r)} \tilde{D}_a S^a\,dA = \tfrac{1}{2}r^{-1}f(r), \tag{11}$$

where the first step results from using the usual expression for the rate of change of an integral and the fact that v vanishes on the boundary of $B(r)$. The second step results from substituting (10), and the third step results from: (i) using Gauss' law, (ii) noting that v vanishes on the boundary of $B(r)$ and, (iii) recalling that S^a vanishes whenever v vanishes.

We conclude that $df(r)/dr$ is greater than or equal to the non-negative quantity $\frac{1}{2}r^{-1}f(r)$, at least for r greater than $\underline{r}$. But $f(r)$ is non-negative everywhere and vanishes for large r; hence $f(r)$ must vanish for every greater than $\underline{r}$. That is to say, v cannot assume positive values for r greater than $\underline{r}$. Repeating the above argument, but using now the region in which v is negative, we conclude that v cannot assume negative values for r greater than $\underline{r}$. Hence $v = 0$ for r greater than $\underline{r}$. This contradicts the definition of $\underline{r}$ unless $\underline{r} = 0$ and $v = 0$ everywhere. It is now immediate [e.g. from Eq. (8)] that $p^{ab} = 0$ everywhere.

Only minor modifications are required to obtain the same conclusion for initial data for which the metric q_{ab} is still required to be flat, but now p^{ab} is only required to be asymptotically flat.[11] In this case, define $\underline{r}$ and f in the same way as in the above proof: Then (11) still holds for r greater than $\underline{r}$. Now, however, $f(r)$ need not necessarily be zero for large r; it need only be bounded. Nonetheless, we still obtain a contradiction with (11) unless $f(r)$ vanishes for r greater than $\underline{r}$. From this, just as above, we conclude that $p^{ab} = 0$ everywhere.

IV. DISCUSSION

The argument of the energy conjecture in three available cases—the proofs for the pure kinetic and spherically symmetric cases (see the Appendix) and an argument for the pure potential case by Geroch[4]—are all very similar to each other. In each, one introduces a certain family of two-surfaces, topologically two-spheres, and examines the variations, from one surface to the next, of the integral of some quantity over each surface. The only difference is in the choice of the two-surfaces, and of the quantity to integrate. In the pure kinetic case, the surfaces are concentric two-spheres in Euclidean three-space; and the integrand is, essentially, $|v|$ (where $v = p^{ab}r_a r_b$). In the spherically symmetric case, the surfaces are those of the spherical symmetry; and the integrand is $(2\tilde{R} - \tilde{p}^2 + v^2)$, where $\tilde{R}$ and $\tilde{p}$ are the scalar curvature and trace of the extrinsic curvature of a two-surface, respectively. Finally, in the pure potential case, the surfaces are that family which admits parametrization by that scalar field t satisfying $(D^m t D_m t)^{-1/2}\tilde{p} = 1$; and the integrand is $(2\tilde{R} - \tilde{p}^2)$.

The similarity discussed in the above paragraph suggests that the argument for the completely general case might have similar structure: One introduces a certain family of two-surfaces and examines the variation, from one surface to the next, of the integral over each surface of some appropriate quantity. We now discuss the problem of choosing the appropriate surfaces, and the quantity to integrate over each surface. For the first problem, we observe that the surfaces in each case above are determined only by the metric. Furthermore, in each case the surfaces admit a parametrization by the scalar field t satisfying $(D^m t D_m t)^{-1/2}\tilde{p} = 1$. Hence this choice of surface, i.e., that which was used in the pure potential case, seems to be a very promising candidate. However, for the second problem the situation is apparently less clear. The integrand $(2\tilde{R} - \tilde{p}^2 + v^2)$, of the spherically symmetric case seems to be compatible with the integrand $(2\tilde{R} - \tilde{p}^2)$, of the pure potential case because the v^2 term in the former automatically vanishes in the latter (i.e., when $p^{ab} = 0$). But the integrand in the pure kinetic case is $|v|$ rather than the v^2 which would have been suggested by those of the other two case cases (since $2\tilde{R} = \tilde{p}^2$ for the choice of surfaces in the pure kinetic case). Hence, it is difficult to guess the appropriate integrand for the general conjecture from these three cases. This difficulty might be the indication that, in order to make the integrands more compatible with each other, we must choose surfaces which are determined not only by the metric, but also by p^{ab}.

If we could find a way to reconcile the pure kinetic argument with the pure potential one, it would perhaps lead to a general proof. The spherically symmetric case can now be viewed as a means of studying the mechanism of this reconciliation. The cases in which the metric is conformally flat (i.e. those in the second column of the table) can be viewed in the same way. Unfortunately, little is known in these cases. Even for case 5 of the table (i.e., that in which $p^{ab}=(p/3)q^{ab}$ and $q_{ab}=\psi^4 f_{ab}$ where f_{ab} is a flat metric) the conjecture has not yet been resolved. This is perhaps the simplest open special case of the energy conjecture. The constraint equations and the local energy condition can be written entirely in terms of only two functions on Euclidean three-space: the conformal factor ψ and the trace p of p^{ab}. The energy conjecture then takes the following form:

Conjecture 2: Let S be Euclidean three-space with the positive-definite metric f_{ab}. Let C be a compact subset of S. Let p and ψ be smooth scalar fields on S which assume the constant values 0 and 1, respectively, outside C. Finally, let p and ψ satisfy the differential inequality

$$-12\nabla^2\psi + p^2\psi^5 \geq 2\psi^3(\nabla_a p \nabla_b p f^{ab})^{1/2}. \tag{12}$$

Then $p=0$ and $\psi=1$ everywhere.

Here, (12) is the local energy condition which results from eliminating μ and J^n, in favor of p and ψ, by using the constraint Eqs. (1) and (2).]

We observed that the proofs for the pure kinetic case and the spherically symmetric one have similar structure but that they are not compatible as regards to the choice of the integrand. What one would like to know is whether there is a way to rewrite the pure kinetic proof and/or the spherical one so that they look more comparable. One way to approach this question would be to study the confluence of these two cases, (i.e., that in which metric q_{ab} is spherically symmetric while p^{ab} is general). A proof for this case, i.e., a reconciliation of the pure kinetic and the spherically symmetric cases, might give some insight into the structure of a possible proof for the general case.

It might be of interest to study other special cases.

ACKNOWLEDGMENTS

I am deeply grateful to R. Geroch, my advisor, for his guidance, warm encouragement, and for numerous suggestions. I would also like to thank the members of the relativity group of the University of Chicago for stimulating discussions on many topics in general relativity. Finally, I would like to acknowledge valuable discussions with S. Deser and A. Ashtekar.

APPENDIX

Theorem: Conjecture 1 is true with the additional condition that metric q_{ab} and p^{ab} are spherically symmetric.[12]

Proof: Choose a spherically symmetric function on S, i.e., choose t such that the two-dimensional surfaces $t=$constant in S are two-spheres of spherical symmetry with the inner most surface, $t=0$, reducing to the center of the spherical symmetry. Further, choose t, as one may, to satisfy the inequality $r^a D_a t > 0$ everywhere, where r^a denotes the unit normal vector field to these two-surfaces. Denote by $\tilde{q}_{ab}$ and $\tilde{p}^{ab}$ the induced metric and the extrinsic curvature, respectively, of these surfaces. We then have the Gauss–Codazzi equation,

$$\tilde{R} = \tilde{p}^2 - \tilde{p}^{ab}\tilde{p}_{ab} + R - 2R_{ab}r^a r^b, \tag{A1}$$

where $\tilde{R}$ denotes the scalar curvature of the two-spheres and $\tilde{p}$ denotes $\tilde{p}^a_a$.

Next, contract each side of the second constraint Eq. (2) with r_a to obtain

$$r_a J^a = D_b(r_a p^{ab}) - (\tilde{p}/2)(\tilde{q}_{ab}p^{ab}), \tag{A2}$$

where we have used $D_a r_b = (\tilde{p}/2q)\tilde{q}_{ab}$. Since p^{ab} is spherically symmetric, we have $r_a p^{ab} = v r^b$ with some scalar field v. Hence, we may write (A2) as follows:

$$r_a J^a = D_b(v r^b) - (\tilde{p}/2)w, \tag{A3}$$

where we have set $\tilde{q}_{ab}p^{ab} = w$. Since p^{ab} is spherically symmetric, it is uniquely determined by v and w (which are also spherically symmetric). Hence, by the first constraint equation(1), μ can be expressed in terms of these quantities as follows:

$$2\mu = R - \tfrac{1}{2}v^2 + vw. \tag{A4}$$

Now define a positive scalar field ϕ on S by $\phi r^a D_a t = 1$. Let a dot, affixed to a quantity, denote its rate of change with respect to t (i.e., its Lie derivative by ϕr^a). Then we have

$$\dot{\tilde{q}}_{ab} = 2\phi\tilde{p}_{ab}, \tag{A5}$$

$$\dot{\tilde{p}} = -\tfrac{1}{2}\phi\tilde{p}^2 - \tfrac{1}{2}\phi\tilde{p}^{ab}\tilde{p}_{ab} + \tfrac{1}{2}\phi\tilde{R} - \tfrac{1}{2}\phi R. \tag{A6}$$

For each value of t, set

$$f(t) = \int [2\tilde{R} - \tilde{p}^2 + v^2]\, dA \tag{A7}$$

where the integral extends over the surface $t=$const. Using Eqs. (A3), (A4), (A5), (A6) and the Gauss–Bonnet equation, the rate of change of $f(t)$ with respect to t is

$$\frac{df(t)}{dt} = -\frac{1}{2}\phi\tilde{p}f + 2\phi\tilde{p}\int(\mu + \tilde{p}^{-1}vr^a J_a)\, dA. \tag{A8}$$

Next, denote by $\underline{t}$ the smallest t outside of which $|v| \leq \tilde{p}$. (Such $\underline{t}$ always exists since v vanishes and $\tilde{p}$ is positive-definite outside C.) Then the integral on the right of (A8) is non-negative for t greater than $\underline{t}$ because of the local energy condition. Therefore, (A8) reduces, for t greater than $\underline{t}$, to the following inequality:

$$\frac{df(t)}{dt} \geq -\frac{1}{2}\phi\tilde{p}f(t). \tag{A9}$$

It follows from equation (A1) and the fact that $\tilde{p}^{ab} = (\tilde{p}/2)\tilde{q}^{ab}$ on each $=$const surface, that $f(t)$ vanishes both for a small sphere about the center and for a large sphere outside C.

Now, suppose $\underline{t}$ were non-zero. This supposition implies that $\tilde{p} = |v|$ at $t=\underline{t}$ whence $f(t)$ is positive by (A7) and the Gauss–Bonnet equation. The positivity of $f(t)$ at $\underline{t}$ and the above inequality, (A9), for $t \geq \underline{t}$ implies that $f(t)$ remains positive for $t \geq \underline{t}$. But this contradicts the fact that $f(t)$ vanishes for spheres outside C. We conclude that $\underline{t}=0$. Now recall that $f(t)$ vanishes at $t=0$.

This fact and the inequality (A9) is consistent with the fact that $f(t)$ vanishes for spheres outside C, only when $f=0$ everywhere. The μ and J^a must also vanish from equation (A8). From these and the spherical symmetry of q_{ab} and p^{ab}, the conclusion of the theorem follows.[13]

Only minor modifications are required to generalize the theorem to an asymptotically flat initial data set. The argument above, up to (A9), holds for this case too. Now the $f(t)$ need not vanish for large spheres. However notice that $f(t)$ is asymptotically equal to E/r for large r, where E is the ADM energy and r is any typical radial distance. Finally, since $f(t)$ is nonnegative at $t=\underline{t}$ (whether or not $\underline{t}$ represents the center), inequality (A9) implies that f remains non-negative for $t \geq \underline{t}$. Hence E is nonnegative.

*This research has been presented as a thesis to the Department of Physics, University of Chicago, in partial fulfillment of the requirement for the Ph.D. degree.

†Supported in part by the National Science Foundation, Contract No. GP-34721X1.

‡Present address: Syracuse University, Department of Physics, 201 Physics Bldg., Syracuse, New York 13210.

[1]R. Arnowitt, S. Deser and C.W. Misner, Phys. Rev. **118**, 1100 (1960); **121**, 1556 (1961); **122**, 997 (1961); R. Geroch, J. Math. Phys. **13**, 956 (1972).

[2]H. Bondi, M.G.J. Van Der Berg and A.W.K. Metzner, Proc. R. Soc. A **269**, 21 (1962); R. Sachs, Proc. R. Soc. A **270**, 193 (1962).

[3]D. Brill, Ann. Phys. **7**, 466 (1959).

[4]R. Geroch, Ann. N.Y. Acad. Sci. **224**, 108 (1973).

[5]R. Geroch (to appear).

[6]Let the metric q_{ab} be flat and $p^{ab}=(p/3)q^{ab}$. Let p be any nonzero scalar field which has a compact support on S. Then this data automatically satisfies all the conditions of Conjecture 1 but with the local energy condition $\mu \geq (J^a J_a)^{1/2}$ replaced by $\mu \geqslant 0$. However, this data does not satisfy the conclusion of the conjecture: It is not true that μ vanishes everywhere.

[7]D.R. Brill, and S. Deser, Ann. Phys. **50**, 548 (1968); D.R. Brill, T.S. Deser and L. Faddev, Phys. Lett. A **26**, 538 (1968).

[8]It is not necessary to include a row in the table in which $p=0$, a specialization of p^{ab} which is often taken in another context. This row would essentially be the same as the first row; for $p=0$, constraint Eq. (1) reduces to $2\mu+p^{ab}p_{ab}=R \geq 0$, i.e., the case reduces to the $p^{ab}=0$ case.

[9]R. Arnowitt, S. Deser and C.W. Misner, Ann. Phys. (N.Y.) **11**, 116 (1960).

[10]C. Leibovitz, W. Israel, Phys. Rev. D **1**, 3226 (1970); C.W. Misner, *Astrophysics and General Relativity*, edited by M. Chretien, S. Deser and J. Goldstein (Gordon and Breach, New York, 1968).

[11]A p^{ab} on R^3 (with flat metric q_{ab}) is said to be asymptotically flat if $r^4 p^{ab}p_{ab}$ is bounded where r is the radial distance from some origin. For a general definition of asymptotically flat initial data, see Ref. 1.

[12]The metric q_{ab} on the manifold S, is said to be spherically symmetric if there exist three nonzero Killing vectors, l_1^q, l_2^q and l_3^q (i.e., $L_{l_i} q_{ab}=0$ for $i=1$, 2 and 3), which are linearly dependent at each point and which have the following commutation relations:

$$[l_1 l_2]^a = l_3^q, \quad [l_2 l_3]^a = l_1^a \quad [l_3 l_1]^a = l_2^q.$$

The p^{ab} is said to be spherically symmetric if $L_{l_i} p^{ab}=0$ for $i=1$, 2 and 3.

[13]From the Gauss–Codazzi Eq. (A1), the constraint Eq. (1), and $\tilde{p}_{ab}=(\tilde{p}/2)\tilde{q}^{ab}$ we have

$$2\mu - (\tilde{R} - \tfrac{1}{2}\tilde{p}^2 + \tfrac{1}{2}v^2) = 2(R^{ab} - p^{am}p_m^b + (p/2)p^{ab})r_a r_b.$$

Since the left side of the above equation vanishes, the right side also vanishes. Now the spherically symmetric field $(R^{ab} - p^{am}p_m^b + (p/2)p^{ab})$ must vanish because it has vanishing trace (since $\mu=0$) and also has vanishing $r_n r_h$-component, whence we get Eq. (3) of Conjecture 1. For Eq. (4), set $F^{abc} = D^{[a}(p^{b]c} - \frac{1}{2}pq^{b]c})$. Then we have $F^{abc}=F^{[ab]c}$, $F_a^{ba}=J^b=0$, and $F^{[abc]}=0$. The only spherically symmetric F^{abc} having these algebraic properties is the zero one.

Solution of initial-value problems for some infinite chains of harmonic oscillators

W. G. Christian

Morehouse College, Atlanta, Georgia

A. G. Law

University of Regina, Regina, Saskatchewan, Canada

W. F. Martens

University of Alabama, Birmingham, Alabama

A. L. Mullikin and M. B. Sledd

Georgia Institute of Technology, Atlanta, Georgia
(Received 4 March 1975)

Procedures are given for solving the equations of motion of infinite and half-infinite chains of linear spring-mass oscillators with nearest-neighbor coupling. Arbitrary initial displacements and initial velocities may be prescribed for any finite number of the masses in the chains, and external forces may be applied to any finite number of them. The solutions appear as integrals with integrands involving orthogonal polynomials generated by three-term recurrence relations whose coefficients are determined from the equations of motion. Tables of the polynomials needed for solving the equations of motion of a number of special chains are included.

1. INTRODUCTION

In the paragraphs to follow, chains of harmonic oscillators are considered in which each mass is coupled to its nearest neighbors by massless linear springs. The chains are called half-infinite if they extend indefinitely in one direction only (Fig. 1) and fully infinite if they extend indefinitely in both directions (Fig. 2). The springs and masses in a particular chain are not necessarily similar.

The following matters are discussed.

(a) A procedure is described (Sec. 2) for solving the equations of motion of the half-infinite, frictionless chain when arbitrary initial velocities and displacements are prescribed for any finite number of the masses, the remaining masses being initially stationary in their equilibrium positions. The possibility that external forces act on any finite number of the masses is included. In the procedure outlined, a sequence of polynomials is used which is generated by a three-term recurrence relation with coefficients determined by the physical parameters of the chain. It is necessary to determine a distribution (defined later) with respect to which these polynomials are orthogonal.

(b) A method is described (Sec. 3) for solving the equations of motion of the fully infinite, frictionless chain when arbitrary initial velocities and displacements are prescribed for any finite number of the masses and when finitely many masses are subjected to external forces. Here sequences of recursively generated polynomials with matrix coefficients[1] are used, and again it is necessary to produce the appropriate distribution.

(c) The procedures described in Secs. 2 and 3 are modified (Sec. 4) to include the effect of viscous damping when the damping coefficient associated with each inertial element is proportional to the mass of that element.

(d) Some remarks are made (Sec. 5) on the problem of finding a distribution for a sequence of orthogonal polynomials whose recurrence relation is known.

(e) A simplification of the solution for the fully infinite chain is given (Sec. 6) which applies when the chain is physically symmetric about the mass m_0 (Fig. 2); that is, when $k_j = k_{-j+1}$ and $m_j = m_{-j}$ $(j \geq 1)$. As an example, the fully infinite uniform chain (in which all the masses are alike and all the springs are alike) is used; the results deduced can be shown to agree with those obtained by Schrödinger.[2]

(f) A short catalog is given (Sec. 7) of the polynomials and distributions appropriate to chains consisting of various combinations of springs and masses. The polynomials associated with half-infinite and fully infinite uniform chains are included.

Though the ideas discussed are closely related to the theory of linear operators, no explicit use of the relation is made. Questions concerning the existence of solutions are not discussed, because the procedures described lead to the actual display of the solutions whose existence is claimed. Uniqueness can be proved if each component of the solution vector $\mathbf{x}(t)$ of the associated homogeneous system is required to be an analytic function of t (that is, expansible in a Maclaurin's series).

2. THE HALF-INFINITE FRICTIONLESS CHAIN

In the absence of externally applied forces, the equations of motion of the half-infinite frictionless chain (Fig. 1) are

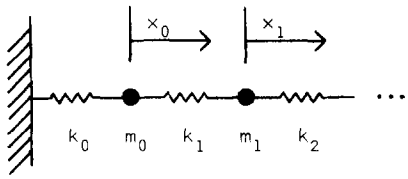


FIG. 1. A half-infinite linear chain of springs and masses.

$$m_0\ddot{x}_0 = -(k_0+k_1)x_0 + k_1x_1,$$
$$m_n\ddot{x}_n = k_nx_{n-1} - (k_n+k_{n+1})x_n + k_{n+1}x_{n+1}, \quad n=1,2,\cdots. \tag{1}$$

Let $\{P_n\}_0^\infty$ be the sequence of polynomials generated by the recurrence relation

$$P_0(x) = 1,$$
$$k_1P_1(x) = -m_0x + k_0 + k_1, \tag{2}$$
$$k_{n+1}P_{n+1}(x) = (-m_nx + k_n + k_{n+1})P_n(x) - k_nP_{n-1}(x), \quad n=1,2,\cdots.$$

Since k_0 is nonnegative and since m_n $(n\geq 0)$ and k_n $(n\geq 1)$ are positive, there exists (see Sec. 5) a bounded, nondecreasing function $\alpha(x)$ which has an infinite number of points of increase and for which

$$\int_{-\infty}^{\infty} P_n(x)P_k(x)\,d\alpha(x) = (m_0/m_n)\delta_{n,k}, \quad n,k=0,1,2,\cdots, \tag{3}$$

in the Lebesgue–Stieltjes sense. A function α satisfying these conditions is called a distribution function for $\{P_n\}_0^\infty$, and $d\alpha$ is called a distribution for $\{P_n\}_0^\infty$. Discussion of some facts concerning the existence, uniqueness, and construction of distribution functions for sequences of polynomials generated by three-term recurrence relations is postponed to Sec. 5.

In terms of the polynomials $P_n(x)$ and the associated distribution $d\alpha$, a solution of Eqs. (1) may be displayed.

Theorem 1: For fixed x and fixed nonnegative integral k, let $u(x,t)$ be the solution of

$$u_{tt}(x,t) + xu(x,t) = 0, \tag{4a}$$
$$u(x,0) = \alpha_k, \quad u_t(x,0) = \beta_k, \tag{4b}$$

and let $\{P_n\}_0^\infty$ be defined by (2). Then

$$x_n(t) = (m_k/m_0)\int_{-\infty}^{\infty} P_n(x)P_k(x)u(x,t)\,d\alpha(x), \quad n=0,1,2,\cdots, \tag{5}$$

is a solution of (1) satisfying

$$x_n(0) = \delta_{n,k}\alpha_k, \quad \dot{x}_n(0) = \delta_{n,k}\beta_k, \quad n=0,1,2,\cdots, \tag{6}$$

where $d\alpha$ is a distribution for $\{P_n\}_0^\infty$.

To verify Theorem 1, notice first that if $u(x,t)$ is a solution of (4a) and if $P_n(x)$ is a solution of (2), then $u(x,t)P_n(x)$ is a solution of (1) in which x appears as a parameter. In order to produce a solution of (1) satisfying the initial conditions (6), solutions $u(x,t)P_n(x)$, $t\geq 0$, are superposed for a continuum of values of x to obtain (5). That $\{x_n(t)\}_0^\infty$ is a solution of (1) now follows from the facts that for each x, $u(x,t)P_n(x)$ is a solution of (1) and that differentiation under the integral sign with respect to t is permissible.[3] The initial conditions (6) are satisfied as a consequence of (4b) and the orthogonality relation (3).

A solution of (1) with nonzero initial conditions imposed on any finite number of masses can be constructed by superposition.

Now suppose that one of the masses (say m_j) is subjected to an external force $f_j(t)$. Then the equations of motion are

$$m_n\ddot{x}_n = k_nx_{n-1} - (k_n+k_{n+1})x_n + k_{n+1}x_{n+1} + \delta_{n,j}f_j,$$
$$n=0,1,2,\cdots, \quad x_{-1}=0, \tag{7}$$

and a solution is given by the following theorem.

Theorem 2: For fixed x and fixed nonnegative integral j, let $w(x,t;j)$ be the solution of

$$w_{tt}(x,t;j) + xw(x,t;j) = f_j(t),$$
$$w(x,0;j) = w_t(x,0;j) = 0,$$

where f_j is piecewise continuous on $0\leq t\leq T<\infty$. Let $\{P_n\}_0^\infty$ be defined by (2). Then

$$x_n(t) = (1/m_0)\int_{-\infty}^{\infty} P_n(x)P_j(x)w(x,t;j)\,d\alpha(x), \quad n=0,1,2,\cdots, \tag{8}$$

is a solution of (7) on $[0,T]$ satisfying

$$x_n(0) = \dot{x}_n(0) = 0, \quad n=0,1,2,\cdots, \tag{9}$$

where $d\alpha$ is a distribution for $\{P_n\}_0^\infty$.

Theorem 2 can be verified by substituting $\{x_n\}_0^\infty$ from (8) into (7).

If external forces are applied to any finite number of masses, a solution of the nonhomogeneous differential equations satisfying homogeneous initial conditions can be constructed by superposition; and by superposing additional solutions (5) of the homogeneous differential equations with nonhomogeneous initial conditions (6), it is possible to solve the initial-value problem with nonzero initial conditions on any finite number of masses and external forces applied to any (possibly different) finite number of masses.

3. THE FULLY INFINITE FRICTIONLESS CHAIN

When no external forces are applied, the equations of motion of the fully infinite frictionless chain (Fig. 2) are

$$m_n\ddot{x}_n = k_nx_{n-1} - (k_n+k_{n+1})x_n + k_{n+1}x_{n+1}, \quad n=\cdots,-1,0,1,\cdots. \tag{10}$$

In this section it is first shown (Lemma 1) that these equations are equivalent to a system of matrix equations. The system of matrix equations is solved (Theorem 3) by a procedure analogous to that used in solving (1), and then a solution of (10) satisfying prescribed initial conditions is derived (Theorem 4) from the solution of the matrix equations.

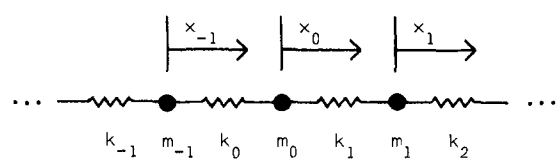


FIG. 2. A fully infinite linear chain of springs and masses.

Let

$$\mathbf{M}_n = \begin{bmatrix} m_{-(n+1)} & 0 \\ 0 & m_n \end{bmatrix}, \quad n \geq 0, \quad \mathbf{K}_n = \begin{bmatrix} k_{-(n+1)} & 0 \\ 0 & k_{n+1} \end{bmatrix}, \quad n \geq 0,$$

$$\mathbf{L}_0 = \begin{bmatrix} k_{-1}+k_0 & -k_0 \\ -k_0 & k_0+k_1 \end{bmatrix}, \tag{11}$$

$$\mathbf{L}_n = \begin{bmatrix} k_{-(n+1)}+k_{-n} & 0 \\ 0 & k_n+k_{n+1} \end{bmatrix}, \quad n \geq 1.$$

Lemma 1: Let k be a fixed nonnegative integer. If

$$\mathbf{X}_n = \begin{bmatrix} x_{-(n+1)} \\ x_n \end{bmatrix}, \quad n \geq 0,$$

is a solution of

$$\mathbf{M}_0\ddot{\mathbf{X}}_0 = \mathbf{K}_0\mathbf{X}_1 - \mathbf{L}_0\mathbf{X}_0,$$
$$\mathbf{M}_n\ddot{\mathbf{X}}_n = \mathbf{K}_n\mathbf{X}_{n+1} - \mathbf{L}_n\mathbf{X}_n + \mathbf{K}_{n-1}\mathbf{X}_{n-1}, \quad n \geq 1, \tag{12a}$$

$$\mathbf{X}_n(0) = \delta_{n,k}\begin{bmatrix} \alpha_{-(k+1)} \\ \alpha_k \end{bmatrix}, \quad \dot{\mathbf{X}}_n(0) = \delta_{n,k}\begin{bmatrix} \beta_{-(k+1)} \\ \beta_k \end{bmatrix}, \quad n = 0, 1, 2, \cdots, \tag{12b}$$

then x_p $(p = \cdots, -1, 0, 1, \cdots)$ is a solution of (10) with

$$x_p(0) = \delta_{p,k}\alpha_k + \delta_{p,-(k+1)}\alpha_{-(k+1)}, \quad p = \cdots, -1, 0, 1, \cdots,$$
$$\dot{x}_p(0) = \delta_{p,k}\beta_k + \delta_{p,-(k+1)}\beta_{-(k+1)}, \quad p = \cdots, -1, 0, 1, \cdots, \tag{13}$$

and conversely.

The lemma is a consequence of the definitions (11) of $\mathbf{K}_n$, $\mathbf{L}_n$, and $\mathbf{M}_n$.

The discussion which now follows, involves the use of matrix polynomials.[4,5] The reader is reminded that the integral of a matrix $\mathbf{M}$ is defined to be the matrix whose elements are the integrals of the elements of $\mathbf{M}$.

Define a sequence $\{\mathbf{P}_n\}_0^\infty$ of matrix polynomials by means of the recurrence relation,

$$\mathbf{P}_0(x) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix},$$
$$\mathbf{K}_0\mathbf{P}_1(x) = -x\mathbf{M}_0 + \mathbf{L}_0, \tag{14}$$
$$\mathbf{K}_n\mathbf{P}_{n+1}(x) = (-x\mathbf{M}_n + \mathbf{L}_n)\mathbf{P}_n(x) - \mathbf{K}_{n-1}\mathbf{P}_{n-1}(x), \quad n \geq 1,$$

and let $\mathbf{P}_n^*(x)$ denote the conjugate transpose of $\mathbf{P}_n(x)$.

Let $\boldsymbol{\alpha}$ be a matrix-valued function

$$\boldsymbol{\alpha}(x) = \begin{bmatrix} \alpha_4(x) & \alpha_3(x) \\ \alpha_2(x) & \alpha_1(x) \end{bmatrix}$$

with the property that

$$\int_{-\infty}^{\infty} \mathbf{P}_n(x)\, d\boldsymbol{\alpha}(x)\mathbf{P}_q^*(x) = \delta_{n,q}\mathbf{M}_q^{-1}\mathbf{M}_0, \quad n, q = 0, 1, 2, \cdots, \tag{15}$$

where

$$d\boldsymbol{\alpha}(x) = \begin{bmatrix} d\alpha_4(x) & d\alpha_3(x) \\ d\alpha_2(x) & d\alpha_1(x) \end{bmatrix}$$

and $d\alpha_2 = d\alpha_3$. The problem of constructing a "distribution matrix" $d\boldsymbol{\alpha}$ is discussed in Sec. 5. In the remainder of this section, it is assumed that $d\boldsymbol{\alpha}$ is known.

Theorem 3: Let $\{\mathbf{P}_n(x)\}_0^\infty$ be defined by (14). Let k be a fixed nonnegative integer, and for fixed x, let

$$\mathbf{U}(x,t) = \begin{bmatrix} u_1(x,t) \\ u_2(x,t) \end{bmatrix}$$

be the solution of

$$\mathbf{U}_{tt}(x,t) = -x\mathbf{U}(x,t),$$
$$\mathbf{U}(x,0) = \begin{bmatrix} \alpha_{-(k+1)} \\ \alpha_k \end{bmatrix}, \quad \mathbf{U}_t(x,0) = \begin{bmatrix} \beta_{-(k+1)} \\ \beta_k \end{bmatrix}.$$

Then

$$\mathbf{X}_n(t) = \int_{-\infty}^{\infty} \mathbf{P}_n(x)\, d\boldsymbol{\alpha}(x)\mathbf{P}_k^*(x)\mathbf{M}_0^{-1}\mathbf{M}_k\mathbf{U}(x,t), \quad n = 0, 1, 2, \cdots,$$

is a solution of (12), where $d\boldsymbol{\alpha}$ is a distribution matrix for $\{\mathbf{P}_n\}_0^\infty$.

The proof parallels that of Theorem 1. Equation (15) is the required orthogonality condition.

Now to construct a solution of (10) satisfying prescribed initial conditions, let $\{P_p\}_{-\infty}^\infty$ and $\{Q_p\}_{-\infty}^\infty$ be solutions of the recurrence

$$k_{p+1}T_{p+1}(x) = (-m_p x + k_p + k_{p+1})T_p(x) - k_p T_{p-1}(x), \quad p = \cdots, -1, 0, 1, \cdots, \tag{16a}$$

corresponding respectively to the initial conditions

$$P_{-1}(x) = 0, \quad P_0(x) = 1 \tag{16b}$$

and

$$Q_{-1}(x) = 1, \quad Q_0(x) = 0. \tag{16c}$$

Then it is easily verified by direct substitution that the solution $\mathbf{P}_n(x)$, $n \geq 0$, of (14) is

$$\mathbf{P}_n(x) = \begin{bmatrix} Q_{-(n+1)}(x) & P_{-(n+1)}(x) \\ Q_n(x) & P_n(x) \end{bmatrix}, \quad n = 0, 1, 2, \cdots. \tag{17}$$

Theorem 4: Let $\{P_p\}_{-\infty}^\infty$ and $\{Q_p\}_{-\infty}^\infty$ be solutions of (16a) satisfying the initial conditions (16b) and (16c) respectively. For fixed x, let $u(x,t)$ be the solution of

$$u_{tt}(x,t) = -xu(x,t),$$
$$u(x,0) = \alpha_j, \quad u_t(x,0) = \beta_j,$$

where j is a fixed integer $(-\infty < j < \infty)$. Then

$$\begin{aligned} x_p(t) = (m_j/\gamma)[&\int_{-\infty}^{\infty} P_p(x)P_j(x)u(x,t)\, d\alpha_1(x) \\ &+ \int_{-\infty}^{\infty} P_p(x)Q_j(x)u(x,t)\, d\alpha_2(x) \\ &+ \int_{-\infty}^{\infty} Q_p(x)P_j(x)u(x,t)\, d\alpha_3(x) \\ &+ \int_{-\infty}^{\infty} Q_p(x)Q_j(x)u(x,t)\, d\alpha_4(x)], \\ & p = \cdots, -1, 0, 1, \cdots, \end{aligned} \tag{18}$$

is a solution of (10) satisfying

$$x_p(0) = \delta_{p,j}\alpha_j, \quad \dot{x}_p(0) = \delta_{p,j}\beta_j, \quad p = \cdots, -1, 0, 1, \cdots,$$

where

$$\gamma = \begin{cases} m_0 & \text{if } j \geq 0, \\ m_{-1} & \text{if } j < 0. \end{cases} \tag{19}$$

Suppose first that $j \geq 0$. Set $j = k$, and let $\alpha_{-(k+1)} = \beta_{-(k+1)} = 0$. Then

$$\mathbf{U}(x,t) = \begin{bmatrix} u_1(x,t) \\ u_2(x,t) \end{bmatrix} = \begin{bmatrix} 0 \\ u(x,t) \end{bmatrix},$$

and the expansion of $\mathbf{X}_n(t)$ in Theorem 3 yields (18), where $p = n$ if $p \geq 0$ and $p = -(n+1)$ if $p < 0$. But by Theorem 3, $\mathbf{X}_n(t)$ is a solution of (12), and by Lemma 1, a solution of (12) yields a solution of (10) satisfying (13). Hence

$$x_p(t), \quad p = \cdots, -1, 0, 1, \cdots,$$

is a solution of (10) satisfying (13), where (13) now takes the form

$$x_p(0) = \delta_{p,j}\alpha_j, \quad \dot{x}_p(0) = \delta_{p,j}\beta_j, \quad p = \cdots, -1, 0, 1, \cdots,$$

since $\alpha_{-(k+1)} = \beta_{-(k+1)} = 0$. Similarly, if $j < 0$, set $j = -(k+1)$, and let $\alpha_k = \beta_k = 0$. Then

$$\mathbf{U}(x, t) = \begin{bmatrix} u(x,t) \\ 0 \end{bmatrix},$$

and the argument proceeds as before.

A solution can be constructed by superposition for a system in which a finite number of masses have nonzero initial conditions.

If one of the masses (say m_r) is subjected to an external force $f_r(t)$, then in analogy to the result stated in Theorem 2 a solution of the system

$$m_n\ddot{x}_n = k_n x_{n-1} - (k_n + k_{n+1})x_n + k_{n+1}x_{n+1} + \delta_{n,r}f_r, \quad n = \cdots, -1, 0, 1, \cdots, \tag{20}$$

is given by the following theorem.

Theorem 5: For fixed x, let $w(x, t; r)$ be the solution of

$$w_{tt}(x, t; r) + xw(x, t; r) = f_r(t),$$
$$w(x, 0; r) = w_t(x, 0; r) = 0,$$

where r is a fixed integer $(-\infty < r < \infty)$ and $f_r(t)$ is piecewise continuous on $0 \le t \le T < \infty$. Let $\{P_n\}_{-\infty}^{\infty}$ and $\{Q_n\}_{-\infty}^{\infty}$ be solutions of (16a) satisfying the conditions (16b) and (16c) respectively. Then

$$\begin{aligned} x_n(t) = (1/\gamma)[& \int_{-\infty}^{\infty} P_n(x)P_r(x)w(x, t; r)\, d\alpha_1(x) \\ & + \int_{-\infty}^{\infty} P_n(x)Q_r(x)w(x, t; r)\, d\alpha_2(x) \\ & + \int_{-\infty}^{\infty} Q_n(x)P_r(x)w(x, t; r)\, d\alpha_3(x) \\ & + \int_{-\infty}^{\infty} Q_n(x)Q_r(x)w(x, t; r)\, d\alpha_4(x)], \end{aligned}$$
$$n = \cdots, -1, 0, 1, \cdots, \tag{21}$$

is a solution of (20) on $[0, T]$ satisfying

$$x_n(0) = \dot{x}_n(0) = 0, \quad n = \cdots, -1, 0, 1, \cdots.$$

With external forces applied to any finite number of masses and with nonzero initial conditions on any finite number of them, a solution may again be constructed by superposition.

4. THE HALF-INFINITE AND FULLY INFINITE CHAINS WITH VISCOUS DAMPING

Suppose that the motion of each mass m_j in Fig. 1 or Fig. 2 is retarded by a viscous frictional force of magnitude $\rho_j|\dot{x}_j|$, where $\rho_j = \beta m_j$, β being a real constant (normally positive). Solutions of the equations of motion of the damped systems can be generated by a modification of the results of Secs. 2 and 3.

Theorem 6: Let β be a real constant. For fixed x and fixed nonnegative integral k, let $u(x, t)$ be the solution of

$$u_{tt}(x, t) + \beta u_t(x, t) + xu(x, t) = 0,$$
$$u(x, 0) = \alpha_k, \quad u_t(x, 0) = \beta_k,$$

and let $\{P_n\}_0^{\infty}$ be defined by (2). Then

$$x_n(t) = (m_k/m_0)\int_{-\infty}^{\infty} P_n(x)P_k(x)u(x, t)\, d\alpha(x), \quad n = 0, 1, 2, \cdots, \tag{22}$$

is a solution of

$$m_0(\ddot{x}_0 + \beta\dot{x}_0) = -(k_0 + k_1)x_0 + k_1x_1,$$
$$m_n(\ddot{x}_n + \beta\dot{x}_n) = k_nx_{n-1} - (k_n + k_{n+1})x_n + k_{n+1}x_{n+1}, \quad n \ge 1,$$
$$x_n(0) = \delta_{n,k}\alpha_k, \quad \dot{x}_n(0) = \delta_{n,k}\beta_k, \quad n \ge 0.$$

The proof is the same as that of Theorem 1.

Theorem 7: Let β be a real constant. Let $\{P_n\}_{-\infty}^{\infty}$ and $\{Q_n\}_{-\infty}^{\infty}$ be solutions of (16a) satisfying (16b) and (16c) respectively. For fixed x, let $u(x, t)$ be the solution of

$$u_{tt}(x, t) + \beta u_t(x, t) + xu(x, t) = 0,$$
$$u(x, 0) = \alpha_k, \quad u_t(x, 0) = \beta_k,$$

where k is a fixed integer $(-\infty < k < \infty)$. Then

$$\begin{aligned} x_n(t) = (m_k/\gamma)[& \int_{-\infty}^{\infty} P_n(x)P_k(x)u(x, t)\, d\alpha_1(x) \\ & + \int_{-\infty}^{\infty} P_n(x)Q_k(x)u(x, t)\, d\alpha_2(x) \\ & + \int_{-\infty}^{\infty} Q_n(x)P_k(x)u(x, t)\, d\alpha_3(x) \\ & + \int_{-\infty}^{\infty} Q_n(x)Q_k(x)u(x, t)\, d\alpha_4(x)], \end{aligned}$$
$$n = \cdots, -1, 0, 1, \cdots,$$

is a solution of

$$m_n(\ddot{x}_n + \beta\dot{x}_n) = k_nx_{n-1} - (k_n + k_{n+1})x_n + k_{n+1}x_{n+1}, \quad n = \cdots, -1, 0, 1, \cdots,$$
$$x_n(0) = \delta_{n,k}\alpha_k, \quad \dot{x}_n(0) = \delta_{n,k}\beta_k, \quad n = \cdots, -1, 0, 1, \cdots.$$

The proof is like that of Theorem 4.

Theorems 2 and 5 have similar analogs which take into account an external force applied to one of the masses.

For damped systems in which a finite number of masses have nonzero initial conditions and a finite number of masses are subjected to externally applied forces, solutions may be constructed by superposition.

5. SOME COMMENTS ON THE CONSTRUCTION OF A DISTRIBUTION FOR A SEQUENCE OF ORTHOGONAL POLYNOMIALS WHOSE RECURRENCE RELATION IS KNOWN

If $\{P_n\}_0^{\infty}$ is the sequence of polynomials generated by the recurrence relation

$$\begin{aligned} & P_{-1}(x) = 0, \\ & P_0(x) = 1, \\ & \kappa_{n+1}P_{n+1}(x) = (-\mu_n x + \lambda_n)P_n(x) - \kappa_n P_{n-1}(x), \quad n \ge 0, \end{aligned} \tag{23}$$

a necessary and sufficient condition[6,7] for the P_n to be orthogonal with respect to some distribution function α is that $\kappa_n^2/\mu_n\mu_{n-1} > 0$ $(n \ge 1)$. If α is normalized so that $\int_{-\infty}^{\infty} d\alpha = 1$, then[8]

$$\int_{-\infty}^{\infty} P_n^2(x)\, d\alpha(x) = \mu_0/\mu_n, \quad n \ge 0. \tag{24}$$

For special cases of (23), practical procedures for constructing distribution functions are known. For

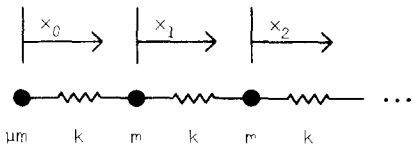


FIG. 3. A half-infinite linear chain with first mass different from the others.

example, conditions on the coefficients κ_n ($n \geq 1$), λ_n ($n \geq 0$), and μ_n ($n \geq 0$) may be given[9] which are necessary and sufficient for the polynomials P_n to constitute a Sturm—Liouville sequence associated with a second-order linear differential equation—that is, for each P_n, $n \geq 0$, to be a solution of a differential equation of the form

$$p(x)y'' + q(x)y' + \nu_n r(x)y = 0, \tag{25}$$

where ν_n is a parameter dependent on n, but $p(x)$, $q(x)$, and $r(x)$ are independent of n. When these conditions are satisfied, the distribution function can be determined from the coefficients in (25). Otherwise the determination of the distribution function may be difficult.

The discussion in the remainder of this section consists of two parts. The first part is an example (from which routine but tedious calculations are omitted) illustrating how the distributions shown in the tables at the end of the paper were constructed when the polynomials P_n did not constitute a Sturm—Liouville sequence. The second part describes a method of finding distributions for the polynomials associated with a fully infinite linear chain when distributions for the polynomials associated with related half-infinite linear chains are known.

First, the example. Consider the linear spring-mass chain shown in Fig. 3, where all the springs are alike and all the masses are alike except perhaps the first, whose mass is μm ($\mu > 0$). Corresponding to Eqs. (1) and (2) of Sec. 2, the equations of motion are

$$\begin{aligned} \mu m\ddot{x}_0 &= -kx_0 + kx_1, \\ m\ddot{x}_n &= kx_{n-1} - 2kx_n + kx_{n+1}, \quad n \geq 1, \end{aligned} \tag{26}$$

and the associated polynomials are generated by the recurrence relation

$$\begin{aligned} P_0(x) &= 1, \\ P_1(x) &= -(\mu m/k)x + 1, \\ P_{n+1}(x) &= [-(m/k)x + 2]P_n(x) - P_{n-1}(x), \quad n \geq 1. \end{aligned} \tag{27}$$

In (27), set $x = (2k/m)(1 - \cos\theta)$, $0 < \theta < \pi$. Then

$$\begin{aligned} P_0(x(\theta)) &= 1, \\ P_1(x(\theta)) &= 1 - 2\mu + 2\mu\cos\theta, \\ P_{n+1}(x(\theta)) &= 2(\cos\theta)P_n(x(\theta)) - P_{n-1}(x(\theta)), \quad n \geq 1. \end{aligned} \tag{28}$$

The solution of the system of difference equations (28) is

$$\begin{aligned} P_n(x(\theta)) &= [\cos(n\theta)\cos(\theta/2) \\ &\quad + (1 - 2\mu)\sin(n\theta)\sin(\theta/2)]/\cos(\theta/2), \quad n \geq 0. \end{aligned} \tag{29}$$

If $\mu = \frac{1}{2}$, then $P_n(x(\theta)) = \cos(n\theta)$, $n \geq 0$; and, except for a multiplicative factor depending on n, $P_n(x)$ is the Tchebichef polynomial of the first kind with argument $1 - (m/2k)x$, for which the distribution $d\alpha(x)$ has the form $d\alpha(x) = \omega(x)\,dx$,[10] where

$$\omega(x) = \begin{cases} (m/\pi)[mx(4k - mx)]^{-1/2}, & 0 < x < 4k/m, \\ 0 & \text{elsewhere.} \end{cases} \tag{30}$$

If $\mu \neq \frac{1}{2}$ nor 1, the polynomials $P_n(x(\theta))$ are not Sturm—Liouville polynomials.[11] Notice, however, that

$$\begin{aligned} P_n(x(\theta)) &= \frac{\mu\cos(n + \frac{1}{2})\theta}{\cos(\theta/2)} + (1 - \mu)\frac{\cos(n - \frac{1}{2})\theta}{\cos(\theta/2)} \\ &= \begin{cases} 1, & n = 0, \\ \mu\beta_n P_n^{(-1/2,\,1/2)}(\cos\theta) \\ \quad + (1 - \mu)\beta_{n-1}P_{n-1}^{(-1/2,\,1/2)}(\cos\theta), & n \geq 1, \end{cases} \end{aligned} \tag{31}$$

where the $P_n^{(-1/2,\,1/2)}(\cos\theta)$ are Jacobi polynomials,[12] β_j ($j \geq 1$) is a factor depending on j but not on θ, and $\beta_0 = 1$.

If $\mu > \frac{1}{2}$, it follows from (29) that $P_n(x(\theta))$ has n zeros in the interval $(0, \pi)$ and hence that for each $n \geq 0$ all the zeros of $P_n(x)$ are in the interval $(0, 4k/m)$. Since the support of a distribution $d\alpha(x)$ for a sequence of orthogonal polynomials is contained in any finite closed interval which contains all the zeros of all the polynomials of the sequence,[13] and since the $P_n(x(\theta))$ are linear combinations of the Jacobi polynomials $P_m^{(-1/2,\,1/2)}(\cos\theta)$, one might conjecture that $\{P_n(x)\}$ has a distribution of the form $d\alpha(x) = \omega(x)\,dx$ with

$$\omega(x) = \rho(x)\sum_{j=0}^{\infty} a_j P_j(x), \tag{32}$$

where

$$\rho(x) = \begin{cases} \dfrac{m}{2k\pi}\left(\dfrac{4k - mx}{mx}\right)^{1/2}, & 0 < x < 4k/m, \\ 0 & \text{elsewhere,} \end{cases}$$

is a weight function for $\{P_n^{(-1/2,\,1/2)}(1 - (m/2k)x)\}$. If the conjecture is correct, then by (31) and (32)

$$\begin{aligned} \omega(x) = \rho(x)\Big(a_0 + \sum_{j=1}^{\infty} a_j \{&\mu\beta_j P_j^{(-1/2,\,1/2)}(1 - (m/2k)x) \\ &+ (1 - \mu)\beta_{j-1}P_{j-1}^{(-1/2,\,1/2)}(1 - (m/2k)x)\}\Big); \end{aligned}$$

and formally the requirement that

$$\int_{-\infty}^{\infty} P_n(x)P_0(x)\omega(x)\,dx = \delta_{0,n} \quad (n = 0, 1, 2, \cdots)$$

yields

$$\begin{aligned} &a_0 + (1 - \mu)a_1 = 1, \\ &(1 - \mu)a_0 + [\mu^2 + (1 - \mu)^2]a_1 + \mu(1 - \mu)a_2 = 0, \\ &\mu(1 - \mu)a_j + [\mu^2 + (1 - \mu)^2]a_{j+1} + \mu(1 - \mu)a_{j+2} = 0, \quad j \geq 1. \end{aligned} \tag{33}$$

The system (33) has a solution

$$\begin{aligned} a_0 &= \frac{\mu^2}{2\mu - 1} \\ a_j &= \frac{\mu}{2\mu - 1}\left(\frac{\mu - 1}{\mu}\right)^j, \quad j \geq 1. \end{aligned}$$

If these values are introduced in (32), and if the trigonometric form (31) of $P_j(x(\theta))$ is used, the resulting series may be summed to yield

$$\omega(x)=\begin{cases}\dfrac{m}{2\pi}\left(\dfrac{4k-mx}{mx}\right)^{1/2}\dfrac{\mu}{k+\mu(\mu-1)mx}, & 0<x<4k/m,\\ 0 & \text{elsewhere.}\end{cases}\tag{34}$$

Since the calculations leading to (34) are based on a conjecture regarding the form (32) of $\omega(x)$, it is necessary to show that (34) is in fact the desired weight function by verifying directly that

$$\int_{-\infty}^{\infty}P_m(x)P_n(x)\omega(x)\,dx=\begin{cases}1, & m=n=0,\\ \mu, & m=n\neq 0,\\ 0, & m\neq n.\end{cases}$$

This step, which involves some tedious integrations in the complex plane, is omitted.

If $0<\mu<\frac{1}{2}$, all except at most one of the zeros of each P_n lie in the interval $(0, 4k/m)$. But for sufficiently large n, each P_n has its largest zero x_n (and no other) in the interval $(4k/m, k/[m\mu(1-\mu)])$. Furthermore,

$$\lim_{n\to\infty}x_n=k/[m\mu(1-\mu)]\triangleq\bar{x}.$$

Now in any finite open interval in which there are only finitely many zeros of the polynomials of a sequence of orthogonal polynomials, a distribution function for the sequence is constant[14]; but at a limit point of zeros of the polynomials of the sequence, the distribution function may have a stepwise discontinuity. So conjecture that a distribution for $\{P_n(x)\}$ has the form $d\alpha(x)=[\omega(x)+c\delta(x-\bar{x})]\,dx$, where $\omega(x)$ is given by (34), c is a constant, and δ is the Dirac delta function. The constant c is determined by the requirement that

$$\int_{-\infty}^{\infty}P_0^2(x)\,d\alpha(x)=1,\quad\text{or}\quad\int_0^{4k/m}\omega(x)\,dx+c=1,$$

from which it follows by use of (34) that

$$c=(1-2\mu)/(1-\mu).$$

Verification that $d\alpha(x)=\{\omega(x)+[(1-2\mu)/(1-\mu)]\delta(x-\bar{x})\}\,dx$ is a distribution for $\{P_n(x)\}$ when $0<\mu<\frac{1}{2}$ is omitted.

Remark: The identity (31) greatly simplified the calculations in this example. Without such an identity, the assumption (for $\mu>\frac{1}{2}$) that $\omega(x)=f(x)\sum_{j=0}^{\infty}a_jP_j(x)$, where

$$f(x)=\begin{cases}1, & 0<x<4k/m,\\ 0 & \text{elsewhere,}\end{cases}$$

would generate an infinite linear system of algebraic equations for the a_j as before; but each equation in this system, with perhaps some exceptions, would have an infinite number of nonzero terms.

Now consider the problem of constructing distributions for the polynomials associated with fully infinite linear chains. Let $\{P_n\}$ and $\{Q_n\}$ be defined by (16). Define

$$a_n=-k_{n+1}/(m_nm_{n+1})^{1/2}$$

and

$$b_n=(k_n+k_{n+1})/m_n,\quad n=\cdots,-1,0,1,\cdots,$$

and let

$$R_n^+(x)=(m_n/m_0)^{1/2}P_n(x),\quad n=-1,0,1,\cdots,$$

$$R_n^-(x)=(m_{-(n+1)}/m_{-1})^{1/2}Q_{-(n+1)}(x),\quad n=-1,0,1,\cdots,$$

$$S_{-1}^+(x)=0,$$

$$S_n^+(x)=-(1/a_{-1})(m_n/m_{-1})^{1/2}Q_n(x),\quad n=0,1,2,\cdots,$$

$$S_{-1}^-(x)=0,$$

$$S_n^-(x)=-(1/a_{-1})(m_{-(n+1)}/m_0)^{1/2}P_{-(n+1)}(x),\quad n=0,1,2,\cdots,$$

$$\mathbf{R}_{-1}(x)=\begin{bmatrix}0&0\\0&0\end{bmatrix},\quad \mathbf{R}_n(x)=\begin{bmatrix}R_n^-(x) & -a_{-1}S_n^-(x)\\ -a_{-1}S_n^+(x) & R_n^+(x)\end{bmatrix},\quad n=0,1,2,\cdots,$$

$$\mathbf{S}_n(x)=\begin{bmatrix}S_n^-(x)&0\\0&S_n^+(x)\end{bmatrix},\quad n=0,1,2,\cdots.$$

Then $\{R_n^+\}$ and $\{S_n^+\}$ are polynomials of the first and second kinds[15] for the recurrence relation

$$a_nT_{n+1}(x)=(x-b_n)T_n(x)-a_{n-1}T_{n-1}(x),\quad n=0,1,2,\cdots;$$

$\{R_n^-\}$ and $\{S_n^-\}$ are polynomials of the first and second kinds for the recurrence relation

$$a_{-(n+2)}T_{n+1}(x)=(x-b_{-(n+1)})T_n(x)-a_{-(n+1)}T_{n-1}(x),\quad n=0,1,2,\cdots;$$

and $\{\mathbf{R}_n\}$ and $\{\mathbf{S}_n\}$ are matrix polynomials of the first and second kinds for the recurrence relation

$$\mathbf{A}_n\mathbf{T}_{n+1}(x)=(x\,\mathbf{I}-\mathbf{B}_n)\mathbf{T}_n(x)-\mathbf{A}_{n-1}\mathbf{T}_{n-1}(x),\quad n=0,1,2,\cdots,$$

where

$$\mathbf{A}_n=\begin{bmatrix}a_{-(n+2)}&0\\0&a_n\end{bmatrix},\quad \mathbf{B}_n=\begin{bmatrix}b_{-(n+1)}&\delta_{0,n}a_{-1}\\ \delta_{0,n}a_{-1}&b_n\end{bmatrix},\quad n=0,1,2,\cdots,$$

and $\mathbf{I}$ is the two-by-two identity matrix.

In the remainder of this section, it is assumed that unique distributions $d\alpha^+(x)$ and $d\alpha^-(x)$ such that

$$\int_{-\infty}^{\infty}R_m^{\pm}(x)R_n^{\pm}(x)\,d\alpha^{\pm}(x)=\delta_{m,n},\quad m,n=-1,0,1,\cdots,$$

are *known* (note that $d\alpha^+$ and $d\alpha^-$ are distributions for associated *half-infinite* chains). The *existence* of a unique matrix function $\alpha(x)$ such that

$$\int_{-\infty}^{\infty}[\mathbf{R}_m(x)\,d\alpha(x)\mathbf{R}_n^*(x)]=\delta_{m,n}\begin{bmatrix}1&0\\0&1\end{bmatrix},\quad m,n\geq 0,$$

is assumed, and how to construct $\boldsymbol{\alpha}(x)$ from $d\alpha^+(x)$ and $d\alpha^-(x)$ is explained.

Conditions sufficient to guarantee the existence of unique distributions are known,[16,17] but discussion of them is omitted. Two lemmas which are needed are now stated; proofs are omitted but are available.[18]

Lemma 2: Suppose that there exists a unique distribution $d\alpha^+(x)$ such that

$$\int_{-\infty}^{\infty}R_m^+(x)R_n^+(x)\,d\alpha^+(x)=\delta_{m,n},\quad m,n=-1,0,1,\cdots,$$

and suppose that $m^+(z)=\int_{-\infty}^{\infty}[d\alpha^+(x)/x-z]$, $\mathrm{Im}(z)\neq 0$, is the Stieltjes transform of $d\alpha^+$. Then $\sum_{n=0}^{\infty}|S_n^+(z)+f(z)R_n^+(z)|^2<\infty$ for each z with $\mathrm{Im}(z)\neq 0$ if and only if $f(z)=m^+(z)$.

If $\{R_n^+\}$, $\{S_n^+\}$, α^+, and m^+ are replaced by $\{R_n^-\}$, $\{S_n^-\}$, α^-, and m^- in Lemma 2, the resulting statement is also true.

Lemma 3: Suppose that there exists a unique matrix function $\boldsymbol{\alpha}(x)$ such that

$$\int_{-\infty}^{\infty}[\mathbf{R}_m(x)\,d\boldsymbol{\alpha}(x)\mathbf{R}_n^*(x)] = \delta_{m,n}\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}, \quad m,n = 0,1,2,\cdots,$$

and suppose that $\mathbf{m}(z) = \int_{-\infty}^{\infty}[d\boldsymbol{\alpha}(x)/x - z]$, $\mathrm{Im}(z) \neq 0$, is the Stieltjes transform of $d\boldsymbol{\alpha}$. Then $\sum_{n=0}^{\infty}[\mathbf{S}_n(z) + \mathbf{R}_n(z)\mathbf{f}(z)]^*[\mathbf{S}_n(z) + \mathbf{R}_n(z)\mathbf{f}(z)]$ converges for each z with $\mathrm{Im}(z) \neq 0$ if and only if $\mathbf{f}(z) = \mathbf{m}(z)$.

By the assumption of the existence and uniqueness of $\boldsymbol{\alpha}(x)$ and by Lemma 3, there exists a unique matrix $\mathbf{m}(z)$ such that

$$\sum_{n=0}^{\infty}[\mathbf{S}_n(z) + \mathbf{R}_n(z)\mathbf{m}(z)]^*[\mathbf{S}_n(z) + \mathbf{R}_n(z)\mathbf{m}(z)] < \infty \tag{35}$$

for each z such that $\mathrm{Im}(z) \neq 0$. Let

$$\mathbf{m}(z) = \begin{bmatrix}M_4(z) & M_2(z)\\ M_2(z) & M_1(z)\end{bmatrix}.$$

It then follows from (35) and the definitions of $\mathbf{R}_n$ and $\mathbf{S}_n$ that $\{M_1(z)R_n^+(z) + [1 - a_{-1}M_2(z)]S_n^+(z)\}_0^{\infty}$ $\{M_2(z)R_n^+(z) - a_{-1}M_4(z)S_n^+(z)\}_0^{\infty}$, $\{M_2(z)R_n^-(z) - a_{-1}M_1(z)S_n^-(z)\}_0^{\infty}$, and $\{M_4(z)R_n^-(z) + [1 - a_{-1}M_2(z)]S_n^-(z)\}_0^{\infty}$ are in $l_2(0,\infty)$ for each z with $\mathrm{Im}(z) \neq 0$ (a sequence $\{c_n\}_0^{\infty}$ is in $l_2(0,\infty)$ if $\sum_{n=0}^{\infty} c_n\bar{c}_n < \infty$). Hence, from the uniqueness of $m^+(z)$ and $m^-(z)$ and by Lemma 2,

$$m^+(z) = M_1(z)/[1 - a_{-1}M_2(z)] = M_2(z)/[-a_{-1}M_4(z)]$$

and

$$m^-(z) = M_2(z)/[-a_{-1}M_1(z)] = M_4(z)/[1 - a_{-1}M_2(z)].$$

Solving these equations yields

$$M_1(z) = m^+(z)/[1 - a_{-1}^2 m^+(z)m^-(z)], \tag{36a}$$

$$M_2(z) = -a_{-1}m^+(z)m^-(z)/[1 - a_{-1}^2 m^+(z)m^-(z)], \tag{36b}$$

$$M_4(z) = m^-(z)/[1 - a_{-1}^2 m^+(z)m^-(z)]. \tag{36c}$$

Hence, if $\alpha^+(x)$ and $\alpha^-(x)$ are known, M_1, M_2, and M_4 may be computed from m^+ and m^-, and then α_1, α_2, and α_4 may be recovered from their Stieltjes transforms by the Stieltjes inversion formula.[19]

To illustrate the procedure just described, suppose that

$$m_n = \begin{cases} m, & n \leq 0,\\ (n+1)^2 m, & n \geq 1,\end{cases} \quad \text{and} \quad k_n = \begin{cases} k, & n \leq 0,\\ n(n+1)k, & n \geq 1.\end{cases}$$

Then (16a) can be written in the form

$$(n+2)T_{n+1}(x) = [-(m/k)x + 2](n+1)T_n(x) - nT_{n-1}(x), \quad n \geq 1, \tag{37a}$$

$$2T_1(x) = [-(m/k)x + 3]T_0(x) - T_{-1}(x), \quad n = 0, \tag{37b}$$

$$T_{n+1}(x) = [-(m/k)x + 2]T_n(x) - T_{n-1}(x), \quad n \leq -1. \tag{37c}$$

By making the substitution $-(m/k)x + 2 = -2\cos\theta$ and by the usual techniques of difference equations, it may be seen that

$$P_n(x(\theta)) = \begin{cases} \dfrac{(-1)^n \cos(n+\frac{1}{2})\theta}{(n+1)\cos(\theta/2)}, & n \geq 0,\\ \dfrac{(-1)^n \sin(n+1)\theta}{\sin\theta}, & n < 0,\end{cases}$$

and

$$Q_n(x(\theta)) = \begin{cases} \dfrac{(-1)^n \sin(n\theta)}{(n+1)\sin\theta}, & n \geq 0,\\ \dfrac{(-1)^n \sin(n\theta)}{\sin\theta}, & n < 0,\end{cases}$$

are solutions of (37a)–(37c) satisfying the conditions

$$P_{-1}(x) = 0, \quad P_0(x) = 1,$$

$$Q_{-1}(x) = 1, \quad Q_0(x) = 0.$$

Furthermore, the unique distributions referred to in Lemma 2 and the sentence following it are

$$d\alpha^+(x) = \begin{cases} \dfrac{m}{2k\pi}\left(\dfrac{mx}{4k - mx}\right)^{1/2} dx, & 0 \leq x \leq 4k/m,\\ 0, & \text{otherwise,}\end{cases}$$

and

$$d\alpha^-(x) = \begin{cases} m(2k^2\pi)^{-1}[mx(4k - mx)]^{1/2}\,dx, & 0 \leq x \leq 4k/m,\\ 0, & \text{otherwise.}\end{cases}$$

The corresponding Stieltjes transforms are

$$m^+(z) = \frac{m}{2k\pi}\int_0^{4k/m} \frac{1}{x - z}\left(\frac{mx}{4k - mx}\right)^{1/2} dx, \quad \mathrm{Im}(z) \neq 0,$$

and

$$m^-(z) = \frac{m}{2k^2\pi}\int_0^{4k/m} \frac{1}{x - z}[mx(4k - mx)]^{1/2}\,dx, \quad \mathrm{Im}(z) \neq 0.$$

The substitution $u = [mx/(4k - mx)]^{1/2}$ produces integrals which can be evaluated by residue theory to yield

$$m^+(z) = \frac{m}{2k}\left[1 + i\left(\frac{mz}{4k - mz}\right)^{1/2}\right]$$

and

$$m^-(z) = \frac{m}{2k^2}\left[(2k - mz) + i(4k - mz)\left(\frac{mz}{4k - mz}\right)^{1/2}\right],$$

where the branch cut is on $[0, 4k/m]$. For these and subsequent calculations, the square root takes its values in the upper half-plane if z is not in $[0, 4k/m]$ and on the nonnegative real half-axis if z is in $[0, 4k/m]$. It then follows from (36a)–(36c) that

$$M_1(z) = \frac{m\{1 + i[mz/(4k - mz)]^{1/2}\}}{(mz + k) + i(mz - 3k)[mz/(4k - mz)]^{1/2}},$$

$$M_2(z) = -\frac{m}{k}\cdot\frac{(mz - k) + i(mz - 3k)[mz/(4k - mz)]^{1/2}}{(mz + k) + i(mz - 3k)[mz/(4k - mz)]^{1/2}},$$

and

$$M_4(z) = -\frac{m}{k}\cdot\frac{(mz - 2k) + i(mz - 4k)[mz/(4k - mz)]^{1/2}}{(mz + k) + i(mz - 3k)[mz/(4k - mz)]^{1/2}}.$$

The calculations that determine $d\alpha_1$ are similar to those that determine $d\alpha_2$ and $d\alpha_4$; so only those for $d\alpha_1$ are included.

The Stieltjes inversion formula states[20] that

$$\alpha_1(x) - \alpha_1(x_0) = \lim_{v \to 0^+}(1/\pi)\int_{x_0}^{x} \mathrm{Im}[M_1(u + iv)]\,du,$$

where it is assumed that the value of α_1 at each of its at most countably many jumps is defined (or redefined) as

$$\alpha_1(x) = [\alpha_1(x+0) + \alpha_1(x-0)]/2.$$

Note now that $M_1(z)$ is analytic except on $[0, 4k/m]$ and at the pole $z_0 = (2+\sqrt{5})(k/m)$, that for any real u in the domain of analyticity $\lim_{v\to 0^+} M_1(u+iv)$ is real, and that because of the way in which the square root was defined, $\lim_{v\to 0^+} M_1(u+iv) = M_1(u)$, $0 \le u \le 4k/m$. Hence, $\alpha_1(x)$ is constant on the intervals $(-\infty, 0)$, $(4k/m, z_0)$, and (z_0, ∞); $\alpha_1(x)$ is continuous at 0 and $4k/m$; and

$$d\alpha_1(x) = -\frac{m}{\pi}\left(\frac{[mx(4k-mx)]^{1/2}}{m^2x^2 - 4kmx - k^2}\right)dx, \quad 0 \le x \le 4k/m.$$

To evaluate the jump in α_1 at z_0, a technique due to Chihara[21] can be used. Let x_0 and x satisfy the inequality $4k/m < x_0 < z_0 < x$, and consider the rectangular contour with vertices $x+iv$, x_0+iv, x_0-iv, and $x-iv$, where v is positive. By the residue theorem

$$\int_x^{x_0} M_1(u+iv)\,du + i\int_v^{-v} M_1(x_0+i\omega)\,d\omega + \int_{x_0}^{x} M_1(u-iv)\,du + i\int_{-v}^{v} M_1(x+i\omega)\,d\omega = 2\pi i \operatorname{Res}[M_1(z)]_{z=z_0}.$$

Since $M_1(\bar{z}) = \overline{M_1(z)}$,

$$\begin{aligned}2\pi i \operatorname{Res}[M_1(z)]_{z=z_0} &= \int_{x_0}^{x}[-M_1(u+iv) + \bar{M}_1(u+iv)]\,du \\ &+ i\int_0^v [-\bar{M}_1(x_0+i\omega) - M_1(x_0+i\omega) + \bar{M}_1(x+i\omega) \\ &+ M_1(x+i\omega)]\,d\omega \\ &= -2i\int_{x_0}^{x} \operatorname{Im}[M_1(u+iv)]\,du + 2i\int_0^v \operatorname{Re}[M_1(x+i\omega) \\ &- M_1(x_0+i\omega)]\,d\omega.\end{aligned}$$

Thus

$$(1/\pi)\lim_{v\to 0^+}\int_{x_0}^{x} \operatorname{Im}[M_1(u+iv)]\,du = -\operatorname{Res}[M_1(z)]_{z=z_0} = 1/\sqrt{5}$$

and

$$d\alpha_1(x) = \left[-\frac{m}{\pi}H(x)H(4k-mx)\left(\frac{[mx(4k-mx)]^{1/2}}{m^2x^2 - 4kmx - k^2}\right) + \frac{1}{\sqrt{5}}\,\delta\left(x - (2+\sqrt{5})\frac{k}{m}\right)\right]dx,$$

where $H(x)$ is the Heaviside unit step function and δ is the Dirac delta function.

Similar calculations yield

$$d\alpha_2(x) = d\alpha_3(x) = \left[\frac{m}{2k\pi}H(x)H(4k-mx)\left(\frac{(mx-3k)[mx(4k-mx)]^{1/2}}{m^2x^2 - 4kmx - k^2}\right) + \frac{1-\sqrt{5}}{2\sqrt{5}}\,\delta\left(x - (2+\sqrt{5})\frac{k}{m}\right)\right]dx,$$

$$d\alpha_4(x) = \left[\frac{m}{2k\pi}H(x)H(4k-mx) \times \left(\frac{(mx-5k)[mx(4k-mx)]^{1/2}}{m^2x^2 - 4kmx - k^2}\right) + \frac{(1-\sqrt{5})^2}{4\sqrt{5}}\,\delta\left(x - (2+\sqrt{5})\frac{k}{m}\right)\right]dx.$$

6. THE PHYSICALLY SYMMETRIC FULLY INFINITE FRICTIONLESS CHAIN

If a fully infinite chain is physically symmetric about one of the masses (which can be called m_0) or about one of the springs (which can be called k_0), the method of solution outlined in Sec. 3 can be simplified. The present section describes the simplified procedure which applies when the chain is symmetric about the mass m_0 (Fig. 2)—that is, when $k_j = k_{-j+1}$ and $m_j = m_{-j}$ $(j \ge 1)$—and when no external forces are applied. For symmetry about the spring k_0, only a slight modification is required, and for either type of symmetry the effect of finitely many applied forces can be included in a way not greatly different from that used in Sec. 3.

Suppose then that the chain in Fig. 2 is symmetric about m_0 and that no external forces are applied. One can easily imagine two modes of oscillation. In the first, which will be called the symmetric mode, the displacements of masses m_j and m_{-j} have the same amplitude and sign, and springs k_0 and k_1 exert numerically equal and similarly directed forces on the mass m_0. The forces and displacements in the chain can then be visualized by imagining that the chain has been folded about the mass m_0 to yield the configuration shown in Fig. 4.

In the second or antisymmetric mode, the displacements of masses m_j and m_{-j} have the same amplitude but opposite signs, and springs k_0 and k_1 exert numerically equal but oppositely directed forces on the mass m_0, which remains stationary. The forces and displacements can then be visualized by considering only the right half of the chain, which is effectively insulated from the left half by the stationary mass m_0 (see Fig. 5). One might conjecture that the displacements of the masses in the actual chain (Fig. 2) can be represented as linear combinations of the displacements occurring in the symmetric and antisymmetric modes. It is now shown that this conjecture is correct.

Under the symmetry assumptions ($k_j = k_{-j+1}$ and $m_j = m_{-j}$), the equations of motion (10) and the initial conditions $x_n(0) = \delta_{n,j}\alpha_j$, $\dot{x}_n(0) = \delta_{n,j}\beta_j$ may be written

$$\begin{aligned}&k_n x_{n-1} - m_n\ddot{x}_n - (k_n + k_{n+1})x_n + k_{n+1}x_{n+1} = 0, \quad n \ge 1,\\ &k_1 x_{-1} - m_0\ddot{x}_0 - 2k_1x_0 + k_1x_1 = 0,\\ &k_{n+1}x_{-(n+1)} - m_n\ddot{x}_{-n} - (k_n + k_{n+1})x_{-n} + k_n x_{-(n-1)} = 0, \quad n \ge 1,\\ &x_p(0) = \delta_{p,j}\alpha_j, \quad p = \cdots, -1, 0, 1, \cdots,\\ &\dot{x}_p(0) = \delta_{p,j}\beta_j, \quad p = \cdots, -1, 0, 1, \cdots,\end{aligned} \tag{38}$$

where j is a fixed integer $(-\infty < j < \infty)$. Set

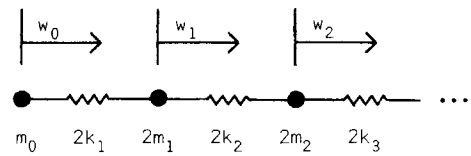


FIG. 4. Configuration for analysis of the symmetric mode of oscillation of a fully infinite chain symmetric about m_0.

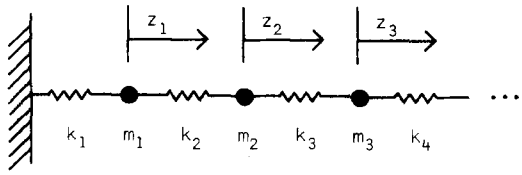


FIG. 5. Configuration for analysis of the antisymmetric mode of oscillation of a fully infinite chain symmetric about m_0.

$$w_n = (x_n + x_{-n})/2, \quad n \geq 0,$$
$$z_n = (x_n - x_{-n})/2, \quad n \geq 0.$$

Then system (38) may be replaced by the equivalent pair of systems

$$\begin{aligned} &-m_0\ddot{w}_0 - 2k_1w_0 + 2k_1w_1 = 0,\\ &k_nw_{n-1} - m_n\ddot{w}_n - (k_n + k_{n+1})w_n + k_{n+1}w_{n+1} = 0, \quad n \geq 1,\\ &w_n(0) = \delta_{n,|j|}(1 + \delta_{0,j})\alpha_j/2, \quad n \geq 0,\\ &\dot{w}_n(0) = \delta_{n,|j|}(1 + \delta_{0,j})\beta_j/2, \quad n \geq 0, \end{aligned} \tag{39}$$

and

$$\begin{aligned} &z_0 \equiv 0,\\ &k_nz_{n-1} - m_n\ddot{z}_n - (k_n + k_{n+1})z_n + k_{n+1}z_{n+1} = 0, \quad n \geq 1,\\ &z_n(0) = (\operatorname{sgn} j)\delta_{n,|j|}(1 - \delta_{0,j})\alpha_j/2, \quad n \geq 0,\\ &\dot{z}_n(0) = (\operatorname{sgn} j)\delta_{n,|j|}(1 - \delta_{0,j})\beta_j/2, \quad n \geq 0. \end{aligned} \tag{40}$$

Attention is invited to the fact that the differential equations in systems (39) and (40) are the equations of motion for the symmetric and antisymmetric modes (see Figs. 4 and 5).

To construct solutions of systems (39) and (40), let $\{R_n(x)\}_0^\infty$ and $\{S_n(x)\}_0^\infty$ be solutions of the recurrence relation

$$k_{n+1}T_{n+1}(x) = (-m_nx + k_n + k_{n+1})T_n(x) - k_nT_{n-1}(x), \quad n \geq 1,$$

satisfying the initial conditions

$$R_0(x) = 1, \quad k_1R_1(x) = -(m_0/2)x + k_1,$$
$$S_0(x) = 0, \quad S_1(x) = 1.$$

Let $\alpha(x)$ and $\beta(x)$ be distribution functions for the sequences $\{R_n(x)\}_0^\infty$ and $\{S_n(x)\}_1^\infty$ so that by (24)

$$\int_{-\infty}^{\infty} R_n(x)R_k(x)\,d\alpha(x) = \begin{cases} \delta_{0,k}, & n = 0, \quad k \geq 0,\\ \dfrac{m_0}{2m_n}\delta_{n,k}, & n > 0, \quad k \geq 0, \end{cases}$$

and

$$\int_{-\infty}^{\infty} S_n(x)S_k(x)\,d\beta(x) = \frac{m_1}{m_n}\delta_{n,k}, \quad n \geq 1, \quad k \geq 1.$$

Then by an argument similar to that used in proving Theorem 1, the solutions of (39) and (40) are

$$w_n(t) = \begin{cases} \int_{-\infty}^{\infty} R_n(x)R_0(x)v(x,t)\,d\alpha(x), & j = 0, \quad n \geq 0,\\ (2m_j/m_0)\int_{-\infty}^{\infty} R_n(x)R_{|j|}(x)v(x,t)\,d\alpha(x), & j \neq 0, \quad n \geq 0, \end{cases}$$

and

$$z_n(t) = (m_j/m_1)\int_{-\infty}^{\infty} S_n(x)S_{|j|}(x)\tilde{v}(x,t)\,d\beta(x), \quad n \geq 0,$$

where, for fixed x, $v(x,t)$ and $\tilde{v}(x,t)$ are solutions of the differential equation

$$y_{tt}(x,t) + xy(x,t) = 0$$

satisfying the initial conditions

$$v(x,0) = w_{|j|}(0), \quad v_t(x,0) = \dot{w}_{|j|}(0),$$

and

$$\tilde{v}(x,0) = z_{|j|}(0), \quad \tilde{v}_t(x,0) = \dot{z}_{|j|}(0),$$

respectively.

Hence a solution of (38) is

$$\begin{aligned} &x_n(t) = w_n(t) + z_n(t), \quad n \geq 0,\\ &x_{-n}(t) = w_n(t) - z_n(t), \quad n \geq 0. \end{aligned} \tag{41}$$

The problem of showing that expressions (18) and (41) are equivalent when the chain is symmetric about the mass m_0 is not considered here. For the special case of a uniform chain, however, calculations based either on (18) or on (41) are simple; they yield the same result, and they are included now as an example.

Suppose then that $k_n = k$ and $m_n = m$, $-\infty < n < \infty$, in the chain shown in Fig. 2. The general method of calculation outlined in Secs. 3 and 5 proceeds as follows. The equations of motion (10) are

$$m\ddot{x}_n = kx_{n-1} - 2kx_n + kx_{n+1}, \quad n = \cdots, -1, 0, 1, \cdots,$$

and the corresponding polynomials $\{P_n\}_{-\infty}^{\infty}$ and $\{Q_n\}_{-\infty}^{\infty}$ [see 16(a)–16(c)] are solutions of the recurrence

$$T_{n+1}(x) = [-(m/k)x + 2]T_n(x) - T_{n-1}(x), \quad n = \cdots, -1, 0, 1, \cdots, \tag{42}$$

with

$$P_{-1}(x) = 0, \quad P_0(x) = 1$$

and

$$Q_{-1}(x) = 1, \quad Q_0(x) = 0.$$

Upon making the substitution $-(m/k)x + 2 = -2\cos\theta$, it follows from (42) by standard techniques of difference equations that

$$\begin{aligned} P_n(x(\theta)) &= \frac{(-1)^n \sin(n+1)\theta}{\sin\theta}, \quad n = \cdots, -1, 0, 1, \cdots,\\ Q_n(x(\theta)) &= \frac{(-1)^n \sin n\theta}{\sin\theta}, \quad n = \cdots, -1, 0, 1, \cdots. \end{aligned} \tag{43}$$

Hence (see Sec. 5),

$$\begin{aligned} R_n^+(x(\theta)) &= R_n^-(x(\theta))\\ &= \frac{(-1)^n \sin(n+1)\theta}{\sin\theta}, \quad n = -1, 0, 1, \cdots,\\ &= \frac{(-1)^n}{2}\left(\frac{2\cdot 4\cdot \cdots \cdot (2n+2)}{1\cdot 3\cdot \cdots \cdot (2n+1)}\right)P_n^{(1/2,1/2)}(\cos\theta)\\ &\qquad n = 0, 1, \cdots, \end{aligned}$$

where $\{P_n^{(1/2,1/2)}(y)\}_0^\infty$ is the sequence of Jacobi polynomials orthogonal with respect to the weight function

$$\rho(y) = \begin{cases} (1-y)^{1/2}(1+y)^{1/2}, & |y| < 1,\\ 0, & |y| \geq 1. \end{cases}$$

Since $\cos\theta = (m/2k)x - 1$, $\{R_n^+(x)\}_{-1}^{\infty}$ and $\{R_n^-(x)\}_{-1}^{\infty}$ are orthogonal with respect to the distribution

$$d\alpha^+(x) = d\alpha^-(x) = \begin{cases} \frac{m}{2k^2\pi}[mx(4k-mx)]^{1/2}\,dx, & 0 \le x \le 4k/m, \\ 0, & \text{otherwise.} \end{cases}$$

By Lemma 2,

$$m^+(z) = m^-(z) = \frac{m}{2k^2\pi}\int_0^{4k/m} \frac{1}{x-z}[mx(4k-mx)]^{1/2}\,dx$$

$$= \frac{m}{2k^2}\left[-(mz-2k) + i(4k-mz)\left(\frac{mz}{4k-mz}\right)^{1/2}\right],$$

where the square root with nonnegative imaginary part is to be used.

Hence, from (36a)–36(c),

$$M_1(z) = M_4(z) = \frac{m\{-(mz-2k) + i(4k-mz)[mz/(4k-mz)]^{1/2}\}}{(4k-mz)\{mz + i(mz-2k)[mz/(4k-mz)]^{1/2}\}}$$

and

$$M_2(z) = \frac{m}{k} \cdot \frac{(mz-2k)^2 - 2k^2 - i(mz-2k)(4k-mz)[mz/(4k-mz)]^{1/2}}{(4k-mz)\{mz + i(mz-2k)[mz/(4k-mz)]^{1/2}\}}.$$

By the Stieltjes inversion formula (Sec. 5)

$$d\alpha_1(x) = d\alpha_4(x) = \begin{cases} (m/\pi)[mx(4k-mx)]^{-1/2}\,dx, & 0 < x < 4k/m, \\ 0, & \text{otherwise,} \end{cases} \tag{44}$$

$$d\alpha_2(x) = \begin{cases} -(m/2k\pi)(mx-2k)[mx(4k-mx)]^{-1/2}\,dx, & \\ & 0 < x < 4k/m, \\ 0, \quad \text{otherwise.} & \end{cases} \tag{45}$$

With $\{P_n\}_{-\infty}^{\infty}$ and $\{Q_n\}_{-\infty}^{\infty}$ as in (43) and with $d\alpha_1$, $d\alpha_2$, $d\alpha_4$ as in (44) and (45), $x_n(t)$, $n = \cdots, -1, 0, 1, \cdots$, as given by (18) is the solution of the equations of motion of the uniform chain with initial conditions $x_n(0) = \delta_{n,j}\alpha_j$, $\dot{x}_n(0) = \delta_{n,j}\beta_j$. If in the expression for $x_n(t)$, one sets $x = (2k/m)(1+\cos\theta)$ and $u((2k/m)(1+\cos\theta), t) = F(\theta, t)$, one finds after some simplification that

$$x_n(t) = [(-1)^{n+j}/\pi]\int_0^{\pi}[\cos(n-j)\theta]F(\theta,t)\,d\theta, \quad n = \cdots, -1, 0, 1, \cdots. \tag{46}$$

The method of solution discussed at the beginning of the present section, which takes advantage of the symmetry of the chain, is considerably shorter. The sequences $\{R_n(x)\}_0^{\infty}$ and $\{S_n(x)\}_0^{\infty}$ are solutions of the recurrence

$$T_{n+1}(x) = [-(m/k)x + 2]T_n(x) - T_{n-1}(x), \quad n \ge 1,$$

with

$$R_0(x) = 1, \quad R_1(x) = -(m/2k)x + 1,$$

$$S_0(x) = 0, \quad S_1(x) = 1.$$

It can be shown that

$$R_n(x) = \frac{(-1)^n 4^n (n!)^2}{(2n)!} P_n^{(-1/2,\,-1/2)}((m/2k)x - 1), \quad n = 0, 1, \cdots,$$

which is a sequence of Jacobi polynomials orthogonal with respect to the weight function

$$\omega(x) = \begin{cases} (m/\pi)[mx(4k-mx)]^{-1/2}, & 0 < x < 4k/m, \\ 0, & \text{otherwise,} \end{cases}$$

and that

$$S_{n+1}(x) = \frac{(-1)^n 4^n n!\,(n+1)!}{(2n+1)!} P_n^{(1/2,\,1/2)}((m/2k)x - 1), \quad n = 0, 1, \cdots,$$

which is a sequence of Jacobi polynomials orthonormal with respect to the weight function

$$\rho(x) = \begin{cases} \frac{m[mx(4k-mx)]^{1/2}}{2\pi k^2}, & 0 < x < 4k/m, \\ 0, & \text{otherwise.} \end{cases}$$

The change of variable $mx = 2k(1+\cos\theta)$ yields

$$R_n(x(\theta)) = (-1)^n \cos n\theta, \quad n = 0, 1, \cdots,$$

with weight function

$$\tilde{\omega}(\theta) = \begin{cases} 1/\pi, & 0 < \theta < \pi, \\ 0, & \text{otherwise,} \end{cases}$$

and

$$S_{n+1}(x(\theta)) = \frac{(-1)^n \sin(n+1)\theta}{\sin\theta}, \quad n = -1, 0, 1, \cdots,$$

with weight function

$$\tilde{\rho}(\theta) = \begin{cases} (2/\pi)\sin^2\theta, & 0 < \theta < \pi, \\ 0, & \text{otherwise.} \end{cases}$$

Then from (41),

$$x_n(t) = \begin{cases} (1/\pi)(-1)^n \int_0^{\pi} \cos(n\theta) v(x(\theta), t)\,d\theta, & n \ge 0, \quad j = 0, \\ (2/\pi)(-1)^{n+j}\int_0^{\pi}\cos(n\theta)\cos(j\theta)v(x(\theta),t)\,d\theta & \\ + (2/\pi)(-1)^{n+j}\int_0^{\pi}\sin(n\theta)\sin|j|\,\theta\tilde{v}(x(\theta),t)\,d\theta, & \\ & n \ge 0, \quad j \ne 0; \end{cases} \tag{47a}$$

$$x_{-n}(t) = \begin{cases} (1/\pi)(-1)^n \int_0^{\pi} \cos(n\theta) v(x(\theta), t)\,d\theta, & n \ge 0, \quad j = 0, \\ (2/\pi)(-1)^{n+j}\int_0^{\pi}\cos(n\theta)\cos(j\theta)v(x(\theta),t)\,d\theta & \\ - (2/\pi)(-1)^{n+j}\int_0^{\pi}\sin(n\theta)\sin|j|\,\theta\tilde{v}(x(\theta),t)\,d\theta, & \\ & n \ge 0, \quad j \ne 0. \end{cases} \tag{47b}$$

Now $v(x(\theta), t) = \tilde{v}(x(\theta), t) = \frac{1}{2}F(\theta, t)$, $j > 0$; $v(x(\theta), t) = F(\theta, t)$, $\tilde{v}(x(\theta), t) = 0$, $j = 0$; and $v(x(\theta), t) = -\tilde{v}(x(\theta), t) = \frac{1}{2}F(\theta, t)$, $j < 0$. Hence the results (47a) and (47b) agree with (46), which completes the example showing that (18) and (41) yield the same result for the uniform chain.

7. POLYNOMIALS AND DISTRIBUTIONS ASSOCIATED WITH VARIOUS LINEAR CHAINS

The polynomials needed in the methods described in Secs. 2–6 are solutions $\{P_n(x)\}_0^{\infty}$ of systems of the form

$$P_0(x) = 1,$$

$$\kappa_1 P_1(x) = (-\mu_0 x + \kappa_0 + \kappa_1)P_0(x), \tag{48}$$

$$\kappa_{n+1}P_{n+1}(x) = (-\mu_n x + \kappa_n + \kappa_{n+1})P_n(x) - \kappa_n P_{n-1}(x), \quad n \ge 1,$$

TABLE I. Orthogonal polynomials $P_n(x)$ generated by recurrence relation (48) when $\kappa_0 = 0$. $d\alpha(x)$ is the associated distribution. $\zeta_n = \int_a^b P_n^2(x)\, d\alpha(x) = \mu_0/\mu_n$, $n \geq 0$. k, m are positive constants.

	μ_n	κ_n	ζ_n	$P_n(x)$	$[a,b]$	$d\alpha(x)$
1.	$m, n \geq 0$	$\kappa_0 = 0$ $\kappa_n = k,\ n \geq 1$	1	$q_n P_n^{(1/2,-1/2)}\left(\frac{mx}{2k} - 1\right)$ $q_n = \frac{(-1)^n 4^n (n!)^2}{(2n)!}$	$[0, 4k/m]$	$\frac{m}{2k\pi}\left(\frac{4k - mx}{mx}\right)^{1/2} dx$
2.	$\mu_0 = m/2$ $\mu_n = m, n \geq 1$	$\kappa_0 = 0$ $\kappa_n = k,\ n \geq 1$	$\zeta_0 = 1$ $\zeta_n = \frac{1}{2},\ n \geq 1$	$q_n P_n^{(-1/2,-1/2)}\left(\frac{mx}{2k} - 1\right)$ $q_n = \frac{(-1)^n 4^n (n!)^2}{(2n)!}$	$[0, 4k/m]$	$\frac{m}{\pi[mx(4k - mx)]^{1/2}} dx$
3.	$\mu_0 = \alpha m,\ \alpha \geq \frac{1}{2}$ $\mu_n = m, n \geq 1$	$\kappa_0 = 0$ $\kappa_n = k,\ n \geq 1$	$\zeta_0 = 1$ $\zeta_n = \alpha, n \geq 1$	$M_n^{(\alpha)}\left(\frac{m}{k}x\right)$	$[0, 4k/m]$	$\frac{\alpha m}{2\pi} \frac{1}{k - \alpha(1-\alpha)mx}\left(\frac{4k - mx}{mx}\right)^{1/2} dx$
4.	$\mu_0 = \alpha m, \alpha < \frac{1}{2}$ $\mu_n = m, n \geq 1$	$\kappa_0 = 0$ $\kappa_n = k,\ n \geq 1$	$\zeta_0 = 1$ $\zeta_n = \alpha, n \geq 1$	$M_n^{(\alpha)}\left(\frac{mx}{k}\right)$	$\left[0, \frac{k}{m\alpha(1-\alpha)}\right]$	$\left[\frac{\alpha m}{2\pi} \frac{1}{k - \alpha(1-\alpha)mx}\left(\frac{4k - mx}{mx}\right)^{1/2} \cdot H(x)H(4k - mx) + \frac{1 - 2\alpha}{1 - \alpha}\delta\left(x - \frac{k}{m\alpha(1-\alpha)}\right)\right] dx$
5.	$(2n+1)m,$ $n \geq 0$	$\kappa_0 = 0$ $\kappa_n = nk,\ n \geq 1$	$\frac{1}{2n+1}$	$(-1)^n P_n^{(0,0)}\left(\frac{mx}{k} - 1\right)$	$[0, 2k/m]$	$\frac{m}{2k} dx$
6.	$(2n+1+\alpha)m,$ $\alpha > -1, n \geq 0$	$\kappa_0 = 0$ $\kappa_n = (n + \alpha\sin^2 n\pi/2)k,$ $\alpha > -1,\ n \geq 1$	$\frac{1+\alpha}{2n+1+\alpha}$	$q_n S_n^{(\alpha)}\left(\frac{mx}{k} - 1\right)$; $q_0 = 1$; for $n \geq 1$, $q_n = \frac{(-1)^n(\alpha+1)(\alpha+3)\cdots(\alpha+2n-1)}{(1+\alpha)(2)(3+\alpha)(4)\cdots(n+\alpha\sin^2 n\pi/2)}$	$[0, 2k/m]$	$\frac{m(1+\alpha)}{2k}\left\lvert\frac{mx}{k} - 1\right\rvert^{\alpha} dx$
7.	$m, n \geq 0$	$k_0 = 0$ $\kappa_n = nk,\ n \geq 1$	1	$L_n^{(0)}\left(\frac{mx}{k}\right)$	$[0, \infty)$	$\frac{m}{k}\exp\left(-\frac{mx}{k}\right) dx$
8.	$(2n+1)^2 m,$ $n \geq 0$	$\kappa_0 = 0$ $\kappa_n = (4n^2 - 1)k,\ n \geq 1$	$\frac{1}{(2n+1)^2}$	$q_n P_n^{(-1/2,1/2)}\left(\frac{mx}{2k} - 1\right)$ $q_n = \frac{(-1)^n 4^n (n!)^2}{(2n+1)!}$	$[0, 4k/m]$	$\frac{m}{2\pi k}\left(\frac{mx}{4k - mx}\right)^{1/2} dx$

for various choices of μ_n and κ_n, $n \geq 0$. The system (48) is constructed in each instance by examination of the particular system of differential equations to be solved. The reader's attention is invited to the use of the coefficients μ_n and κ_n, which are introduced to distinguish the coefficients in the recurrence relation from the physical parameters m_n and k_n, because μ_n is not always equal to m_n (see, for example, the equation $k_1 R_1(x) = -\frac{1}{2}m_0 x + k_1$ in Sec. 6, where $\mu_0 = m_0/2$).

Tables I and II record the solutions $\{P_n\}_0^\infty$ of (48) for the indicated choices of μ_n and κ_n. In each case, the support of the distribution $d\alpha$ is contained in the interval $[a, b]$, and $d\alpha$ is normalized so that

$$\zeta_n = \int_a^b P_n^2(x)\, d\alpha(x) = \mu_0/\mu_n, \quad n \geq 0.$$

In most cases, standard notation is used. $P_n^{(\alpha,\beta)}$ and $L_n^{(\alpha)}$ are respectively the Jacobi and Laguerre polynomials as described by Szegö.[22] $\{S_n^{(\alpha)}\}_0^\infty$ and $\{M_n^{(\alpha)}\}_0^\infty$ are sequences of polynomials that arose in studies made by two of the authors[23,24]; they are not Sturm–Liouville sequences. $S_n^{(\alpha)}(x)$ and $M_n^{(\alpha)}(x)$ have the following forms:

$$S_n^{(\alpha)}(x) = x^n + \sum_{k=1}^{[n/2]}\left[\left(-\frac{1}{2}\right)^k \binom{[n/2]}{k} \cdot \frac{[2n+2\alpha+(-1)^{n+1}-1][2n+2\alpha+(-1)^{n+1}-5]\cdot\cdots\cdot[2n+2\alpha+(-1)^{n+1}-4k+3]}{(2n+\alpha-1)(2n+\alpha-3)\cdot\cdots\cdot(2n+\alpha-2k+1)} x^{n-2k}\right], \quad n \geq 0,$$

where $[r]$ denotes the greatest integer less than or equal to r; and

$$M_0^{(\alpha)}(x) = 1, \qquad M_n^{(\alpha)}(x) = \frac{(-1)^n 4^n [n!]^2}{(2n)!}\left[\alpha P_n^{(1/2,-1/2)}\left(\frac{x}{2} - 1\right) - (1-\alpha)\left(\frac{2n-1}{2n}\right) P_{n-1}^{(1/2,-1/2)}\left(\frac{x}{2} - 1\right)\right], \quad n \geq 1.$$

$H(x)$ is the Heaviside unit step function, and $\delta(x)$ is the Dirac delta (generalized) function.

TABLE II. Orthogonal polynomials $P_n(x)$ generated by recurrence relation (48) when $\kappa_0 \neq 0$. $d\alpha(x)$ is the associated distribution. $\zeta_n = \int_a^b P_n^2(x)\,d\alpha(x) = \mu_0/\mu_n$, $n \geq 0$. k, m are positive constants.

	μ_n	κ_n	ζ_n	$P_n(x)$	$[a,b]$	$d\alpha(x)$
1.	$mr^n, n \geq 0; r > 0$, r constant	$kr^n, n \geq 0; r > 0$, r constant	r^{-n}	$q_n P_n^{(1/2,1/2)}\left(\frac{mx}{2k\sqrt{r}} - \frac{1+r}{2\sqrt{r}}\right)$ $q_n = \frac{(-1)^n 4^n (n!)(n+1)!}{(\sqrt{r})^n (2n+1)!}$	$\left[\frac{k}{m}(1-\sqrt{r})^2, \frac{k}{m}(1+\sqrt{r})^2\right]$	$\frac{m\{[k(1+\sqrt{r})^2 - mx][mx - k(1-\sqrt{r})^2]\}^{1/2} dx}{2\pi k^2 r}$
2.	$\frac{m}{n+1}$, $n \geq 0$	$k, n \geq 0$	$n+1$	$L_n^{(1)}\left(\frac{mx}{k}\right)$	$[0, \infty]$	$\frac{m^2 x}{k^2}\exp\left(-\frac{mx}{k}\right)dx$
3.	$(n+2)^2 m$, $n \geq 0$	$(n+1)(n+2)k$, $n \geq 0$	$\frac{4}{(n+2)^2}$	$q_n P_n^{(1/2,1/2)}\left(\frac{mx}{2k} - 1\right)$ $q_n = \frac{(-1)^n 4^{n+1}[(n+1)!]^2}{(n+2)[(2n+2)!]}$	$[0, 4k/m]$	$\frac{m[mx(4k-mx)]^{1/2}}{2\pi k^2}dx$
4.	$(2n+3)^2 m$, $n \geq 0$	$(2n+1)(2n+3)k$, $n \geq 0$	$\frac{9}{(2n+3)^2}$	$q_n P_n^{(1/2,1/2)}\left(\frac{mx}{2k} - 1\right)$ $q_n = \frac{(-1)^n (3)(4^{n+1})[(n+1)!]^2}{2[(2n+3)!]}$	$[0, 4k/m]$	$\frac{m[mx(4k-mx)]^{1/2}}{2\pi k^2}dx$

In Table I, lines 1 and 2 are special cases of line 3, and line 5 is a special case of line 6. In Table II, lines 3 and 4 are also related—in a way explained in the next paragraph.

Suppose that the sequence of polynomials $\{P_n(x)\}_0^\infty$ satisfies the recurrence relation (48). Let $f_0 = 1$; let f_1 be a constant such that the inequality $0 < f_1 < S_n/(S_n - 1)$ holds for each $n \geq 1$, where

$$S_n = \kappa_1 \sum_{j=0}^{n} \frac{1}{\kappa_{j+1}};$$

let

$$f_{n+1} = \frac{f_1}{f_1(1 - S_n) + S_n}, \quad n \geq 0.$$

Note that the conditions stated assure that $f_n > 0$ $(n \geq 0)$. Let α_0 and α_1 be positive constants such that $\alpha_0\kappa_0 + \alpha_1\kappa_1 = \alpha_1 f_1(\kappa_0 + \kappa_1)$; and let

$$\alpha_{n+1} = f_1\alpha_1/(f_n f_{n+1}), \quad n \geq 0,$$
$$\beta_n = f_1\alpha_1/f_n^2, \quad n \geq 0.$$

Then it is easily verified that the sequence of polynomials $\{f_n P_n(x)\}_0^\infty$ satisfies the recurrence relation

$$f_0 P_0(x) = 1,$$
$$\alpha_1\kappa_1 f_1 P_1(x) = (-\beta_0\mu_0 x + \alpha_0\kappa_0 + \alpha_1\kappa_1) f_0 P_0(x),$$
$$\alpha_{n+1}\kappa_{n+1} f_{n+1} P_{n+1}(x) = (-\beta_n\mu_n x + \alpha_n\kappa_n + \alpha_{n+1}\kappa_{n+1}) f_n P_n(x) - \alpha_n\kappa_n f_{n-1} P_{n-1}(x), \quad n \geq 1.$$

So, just as the polynomials $P_n(x)$ are associated with a chain characterized by recurrence coefficients μ_n and κ_n, the polynomials $f_n P_n(x)$ are associated with a chain characterized by recurrence coefficients $\beta_n\mu_n$ and $\alpha_n\kappa_n$. For example, consider lines 3 and 4 of Table II. Let

$$\alpha_n = \frac{(2n+1)(2n+3)}{(n+1)(n+2)}, \quad n \geq 0,$$

$$\beta_n = \left(\frac{2n+3}{n+2}\right)^2, \quad n \geq 0,$$

$$f_n = \frac{3(n+2)}{2(2n+3)}, \quad n \geq 0.$$

Multiplying the entries μ_n, κ_n, P_n in line 3 by β_n, α_n, f_n, respectively, yields the entries μ_n, κ_n, P_n in line 4. Note that the interval $[a,b]$ and the distribution $d\alpha$ are not changed, because the factors f_n by which the polynomials P_n in line 3 are multiplied do not depend on x. The entries describing other physically different chains could be constructed similarly.

[1] F. V. Atkinson, *Discrete and Continuous Boundary Problems* (Academic, New York, 1964), pp. 150ff.
[2] E. Schrödinger, Ann. Phys. (Leipz.) **44**, 916 (1914).
[3] The permissibility of interchanging differentiation and integration follows from Theorem 14—24, p. 443, in T. M. Apostol, *Mathematical Analysis* (Addison-Wesley, Reading, Massachusetts, 1957). For details see Ref. 5, Appendix B.
[4] Ref. 1, pp. 150ff.
[5] W. G. Christian, *Linear Operators and the Equations of Motion of Infinite Linear Chains* (doctoral dissertation, Georgia Institute of Technology, 1972), pp. 23ff.
[6] J. Favard, C. R. Acad. Sci. (Paris) **200**, 2052 (1935).
[7] A. G. Law, *Solutions of Some Countable Systems of Ordinary Differential Equations* (doctoral dissertation, Georgia Institute of Technology, 1968), p. 18.
[8] Ref. 7, p. 14.
[9] D. V. Ho, J. W. Jayne, and M. B. Sledd, Duke Math. J. **33**, 131 (1966).
[10] G. Szegö, *Orthogonal Polynomials*, AMS Colloquium Publications (Am. Math. Soc., Providence, 1939), Vol. XXIII, p. 60.

[11]Ref. 9, pp. 134-35.
[12]Ref. 10, p. 60.
[13]Ref. 10, p. 111.
[14]Ref. 10, p. 111.
[15]Ju.M. Berezanskii, *Expansions in Eigenfunctions of Self-adjoint Operators*, Translations of Mathematical Monographs (Am. Math. Soc., Providence, 1968), Vol. 17, pp. 503, 520, 560, 577.
[16]Ref. 15, pp. 504, 584.
[17]Ref. 5, pp. 11, 17.
[18]Ref. 15, pp. 521-23, 578, 579.
[19]N.I. Akhiezer, *The Classical Moment Problem*, University Mathematical Monographs (Hafner, New York, 1965), p. 124.
[20]Ref. 19, p. 125.
[21]T.S. Chihara, Proc. Am. Math. Soc. 8, 899 (1957).
[22]Ref. 10, pp. 58, 100.
[23]Ref. 7, p. 32.
[24]W.F. Martens, *Solutions of the Differential Equations of Some Infinite Linear Chains and Two-Dimensional Arrays* (doctoral dissertation, Georgia Institute of Technology, 1971), p. 92. Martens' symbol for $M_n^{(\alpha)}$ is $M_n^{(1,2\alpha)}$. In the expression given in Sec. 7 (above) for $M_n^{(\alpha)}$, a typographical error in Martens' representation has been corrected.

Erratum: A fluid sphere in general relativity [J. Math. Phys. 15, 727 (1974)]

Ronald J. Adler

Departamento de Física, Universidade Federal de Pernambuco, Cidade Universitária, 50.000-Recife-Pe, Brazil
(Received 5 September 1975)

The following corrections should be made. In Eq. (2.3b) the sign of $1/r^2$ on the right side is positive. In Eq. (2.5) the first symbol is τ' not r'. The symbol preceding Eq. (2.5) is τ, not r.

CUMULATIVE AUTHOR INDEX

All authors published so far in the current volume are listed alphabetically with the issue and page numbers following the dash. A cumulative author and subject index covering issues 1 through 12 is published yearly. An (E) *after the page number indicates an Erratum.*

[11]Ref. 9, pp. 134-35.
[12]Ref. 10, p. 60.
[13]Ref. 10, p. 111.
[14]Ref. 10, p. 111.
[15]Ju.M. Berezanskii, *Expansions in Eigenfunctions of Self-adjoint Operators*, Translations of Mathematical Monographs (Am. Math. Soc., Providence, 1968), Vol. 17, pp. 503, 520, 560, 577.
[16]Ref. 15, pp. 504, 584.
[17]Ref. 5, pp. 11, 17.
[18]Ref. 15, pp. 521-23, 578, 579.
[19]N.I. Akhiezer, *The Classical Moment Problem*, University Mathematical Monographs (Hafner, New York, 1965), p. 124.
[20]Ref. 19, p. 125.
[21]T.S. Chihara, Proc. Am. Math. Soc. 8, 899 (1957).
[22]Ref. 10, pp. 58, 100.
[23]Ref. 7, p. 32.
[24]W.F. Martens, *Solutions of the Differential Equations of Some Infinite Linear Chains and Two-Dimensional Arrays* (doctoral dissertation, Georgia Institute of Technology, 1971), p. 92. Martens' symbol for $M_n^{(\alpha)}$ is $M_n^{(1,2\alpha)}$. In the expression given in Sec. 7 (above) for $M_n^{(\alpha)}$, a typographical error in Martens' representation has been corrected.

Erratum: A fluid sphere in general relativity [J. Math. Phys. 15, 727 (1974)]

Ronald J. Adler

Departamento de Física, Universidade Federal de Pernambuco, Cidade Universitária, 50.000-Recife-Pe, Brazil
(Received 5 September 1975)

The following corrections should be made. In Eq. (2.3b) the sign of $1/r^2$ on the right side is positive. In Eq. (2.5) the first symbol is τ' not r'. The symbol preceding Eq. (2.5) is τ, not r.

CUMULATIVE AUTHOR INDEX

All authors published so far in the current volume are listed alphabetically with the issue and page numbers following the dash. A cumulative author and subject index covering issues 1 through 12 is published yearly. An (E) *after the page number indicates an Erratum.*