On subalgebras of the Lie algebra of the extended Poincaré group $\widetilde{\mathrm{P}}(1,n)$

L. F. Barannik and W. I. Fushchich

Institute of Mathematics, Academy of Sciences of Ukrainian SSR, Repin Street 3, Kiev, Union of Soviet Socialist Republics

(Received 7 August 1986; accepted for publication 18 February 1987)

Some general results on the subalgebras of the Lie algebra $\mathrm{A}\widetilde{\mathrm{P}}(1,n)$ of the extended Poincaré group $\widetilde{\mathrm{P}}(1,n)$ $(n\geqslant 2)$ with respect to $\widetilde{\mathrm{P}}(1,n)$ conjugation have been obtained. All subalgebras of $\mathrm{A}\widetilde{\mathrm{P}}(1,4)$ that are nonconjugate to the subalgebras of $\mathrm{AP}(1,4)$ are classified with respect to $\widetilde{\mathrm{P}}(1,4)$ conjugation. The list of representatives of each conjugacy class is presented.

I. INTRODUCTION

The systematic study of subalgebras of quantum mechanics transformation algebras was begun in the fundamental paper by Patera, Winternitz, and Zassenhaus (PWZ) (Ref. 1) in which the general method for classifying the subalgebras of a finite-dimensional Lie algebra with a nontrivial solvable ideal with respect to some group of automorphisms was suggested. This method is applied to classify all subalgebras of Lie algebras of the following groups: the Poincaré group $\mathrm{P}(1,3)$,[1] the extended Poincaré groups $\widetilde{\mathrm{P}}(1,2)$,[2] $\widetilde{\mathrm{P}}(1,3)$,[3] the de Sitter groups $\mathrm{O}(1,4)$,[4] $\mathrm{O}(2,3)$,[5] the optical groups $\mathrm{Opt}(1,2)$,[5] $\mathrm{Opt}(1,3)$,[6] the Euclidean group $\mathrm{E}(3)$,[7] the Schrödinger group $\widetilde{\mathrm{Sch}}(2)$,[8] and the extended Schrödinger group $\mathrm{Sch}(2)$,[8] the Poincaré group $\mathrm{P}(1,4)$,[9–11] the Euclidean group $\mathrm{E}(5)$,[12,13] the Galilei group $\mathrm{G}(3)$,[12] and the extended Galilei group $\widetilde{\mathrm{G}}(3)$.[12] The application of the general method had allowed us to study the subalgebras structure of the Lie algebra of the generalized Euclidean group $\mathrm{E}(n)$ $(n\geqslant 2)$.[13] The subalgebras of the algebras $\mathrm{AP}(1,3)$, $\mathrm{AG}(3)$, and $\mathrm{A}\widetilde{\mathrm{G}}(3)$ were described by another method.[14–17]

The PWZ method needs the development for particular classes of algebras of its generality. In the present paper we give the further development of the PWZ method for extended Poincaré algebras $\mathrm{A}\widetilde{\mathrm{P}}(1,n)$ $(n\geqslant 2)$, denoted also by $\mathrm{ASim}(1,n)$. The necessity in the description of subalgebras of $\mathrm{A}\widetilde{\mathrm{P}}(1,n)$ follows from certain problems of theoretical and mathematical physics.[1] In particular, knowledge of the algebra $\mathrm{A}\widetilde{\mathrm{P}}(1,n)$ subalgebras gives us the possibility to study the symmetry reduction for the relativistically invariant scalar differential equation

$$\Phi(\Box u,(\nabla u)^2,u)=0\,,$$

where

$$\Box u = u_{x_0x_0} - u_{x_1x_1} - \cdots - u_{x_nx_n}\,,$$

$$(\nabla u)^2 = (u_{x_0})^2 - (u_{x_1})^2 - \cdots - (u_{x_n})^2\,,$$

and Φ is a sufficiently smooth function.[18–20] The description of the algebra $\mathrm{A}\widetilde{\mathrm{P}}(1,n)$ subalgebras allows us to solve the problem of the reduction of $\mathrm{A}\widetilde{\mathrm{P}}(1,n)$ algebra representations on its subalgebras.[21,22]

In Sec. II we describe the maximal reducible subalgebras of the algebra $\mathrm{A}\widetilde{\mathrm{O}}(1,n)$, and in Sec. III we describe the completely reducible subalgebras of the algebra $\mathrm{A}\widetilde{\mathrm{O}}(1,n) = \mathrm{AO}(1,n)\oplus\langle\mathbb{D}\rangle$, where $\mathbb{D}$ is the dilatation generator. Section IV is devoted to study of the subalgebras of the extended Galilei algebra $\mathrm{A}\widetilde{\mathrm{G}}(n-1)$, which is one of the important subalgebras of the $\mathrm{AP}(1,n)$ algebra. In Sec. V which is the logical sequel to Sec. IV, a number of assertions on subalgebras of the normalizer of isotropic subspace of the Minkowski space $\mathrm{M}(1,n)$ in algebra $\mathrm{A}\widetilde{\mathrm{P}}(1,n)$ are conceived. Classification of the $\mathrm{A}\widetilde{\mathrm{P}}(1,4)$ algebra subalgebras with respect to the $\widetilde{\mathrm{P}}(1,4)$ conjugation is carried out in Sec. VI. The conclusions are summarized in Sec. VII.

II. MAXIMAL REDUCIBLE SUBALGEBRAS OF THE ALGEBRA $\mathrm{A}\widetilde{\mathrm{O}}(1,n)$

In this section we describe the maximal reducible subalgebras and the maximal Abelian subalgebras of the algebra $\mathrm{A}\widetilde{\mathrm{O}}(1,n)$.

Let R be the real number field; $\langle Y_1,\ldots,Y_s\rangle$ is a vector space or Lie algebra over R with the generators $Y_1,\ldots,Y_s$; R^m is the m-dimensional arithmetical vector space over R; $U=\mathrm{M}(1,n)$ is $(1+n)$-dimensional pseudo-Euclidean space with the scalar product

$$(X,Y) = x_0y_0 - x_1y_1 - \cdots - x_ny_n\,; \tag{2.1}$$

$\mathrm{O}(1,n)$ is the group of the linear transformations of $\mathrm{M}(1,n)$ which conserve (X,X) for every $X\in\mathrm{M}(1,n)$; E_q is the unit matrix of degree q. We suppose that $\mathrm{O}(1,n)$ is realized as the group of the real matrices of degree $n+1$.

We call the extended Poincaré group $\widetilde{\mathrm{P}}(1,n)$ the multiplicative group of the matrices

$$\begin{pmatrix}\lambda\Delta & Y\\ 0 & 1\end{pmatrix},$$

where $\Delta\in\mathrm{O}(1,n)$, $\lambda\in R$, $\lambda>0$, $Y\in R^{n+1}$.

We denote by AG the Lie algebra of the Lie group G. Using the definition of Lie algebra, we find that $\mathrm{AO}(1,n)$ consists of matrices

$$X=\begin{pmatrix}0 & \alpha_{01} & \alpha_{02} & \cdots & \alpha_{0,n-1} & \alpha_{0n}\\ \alpha_{01} & 0 & \alpha_{12} & \cdots & \alpha_{1,n-1} & \alpha_{1n}\\ \alpha_{02} & -\alpha_{12} & 0 & \cdots & \alpha_{2,n-1} & \alpha_{2n}\\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots\\ \alpha_{0,n-1} & -\alpha_{1,n-1} & -\alpha_{2,n-1} & \cdots & 0 & \alpha_{n-1,n}\\ \alpha_{0n} & -\alpha_{1n} & -\alpha_{2n} & \cdots & -\alpha_{n-1,n} & 0\end{pmatrix}. \tag{2.2}$$

 0022-2488/87/071445-14$02.50

Let E_{ik} be the matrix of degree $n+2$ which has the unity on the cross of ith line and k th column and zeros on the other places $(i,k=0,1,...,n+1)$. It is easy to get that the basis of the algebra $A\widetilde{P}(1,n)$ is formed by the matrices

$$\mathbb{D}=E_{00}+E_{11}+\cdots+E_{nn};\quad J_{0a}=-E_{0a}-E_{a0},$$
$$J_{ab}=-E_{ab}+E_{ba};\quad P_0=E_{0,n+1},\quad P_a=E_{a,n+1}$$
$$(a<b,\ a,b=1,...,n).$$

The basis elements satisfy the following commutation relations:

$$\begin{aligned}&[J_{\alpha\beta},J_{\gamma\delta}]=g_{\alpha\delta}J_{\beta\gamma}+g_{\beta\gamma}J_{\alpha\delta}-g_{\alpha\gamma}J_{\beta\delta}-g_{\beta\delta}J_{\alpha\gamma},\\&[P_\alpha,J_{\beta\gamma}]=g_{\alpha\beta}P_\gamma-g_{\alpha\gamma}P_\beta,\quad J_{\beta\alpha}=-J_{\alpha\beta},\\&[P_\alpha,P_\beta]=0,\quad[\mathbb{D},J_{\alpha\beta}]=0,\quad[\mathbb{D},P_\alpha]=P_\alpha,\end{aligned}\tag{2.3}$$

where $g_{00}=-g_{11}=\cdots=-g_{nn}=1$, $g_{\alpha\beta}=0$, when $\alpha\neq\beta$ $(\alpha,\beta=0,1,...,n)$.

The generators of turning $J_{\alpha\beta}$ generate the algebra $AO(1,n)$ and the translation P_α the commutative ideal N, and moreover $A\widetilde{P}(1,n)=N\Subset(AO(1,n)\oplus\langle\mathbb{D}\rangle)$. Let $\widetilde{O}(1,n)=\{\lambda E_{n+1}|\lambda\in R,\lambda>0\}\times O(1,n)$. Evidently, $A\widetilde{O}(1,n)=AO(1,n)\oplus\langle\mathbb{D}\rangle$. It is easy to see that $[X,Y]=X\cdot Y$ for all $X\in A\widetilde{O}(1,n)$, $Y\in N$. Let us identify N and $M(1,n)$ establishing correspondence between P_i and the $(n+1)$-dimensional column with unity on the ith place and zeros on the others $(i=0,1,...,n)$.

Let C be such matrix of degree $n+2$ over R that mapping φ_C: $X\to CXC^{-1}$ is an automorphism of the algebra $A\widetilde{P}(1,n)$. If $C\in G$, where G is a subgroup of $\widetilde{P}(1,n)$, then φ_C is called G automorphism. The subalgebras L and L' of algebra $A\widetilde{P}(1,n)$ are called $\widetilde{P}(1,n)$ conjugated if $\varphi_C(L)=L'$ for some $\widetilde{P}(1,n)$ automorphism φ_C of algebra $A\widetilde{P}(1,n)$. Let us identify φ_C and C.

Let W be a nondegenerate subspace of the space U. This subspace we also consider to be pseudo-Euclidean relative to scalar product defined in U. Let $O(W)$ be the group of isometries of the space W, $\widetilde{O}(W)=O(W)\times\{\lambda E_{n+1}|\lambda\in R,\lambda>0\}$. A subalgebra $F\subset A\widetilde{O}(W)$ is called irreducible if in W there does not exist any F-invariant subspace different from O and W. Otherwise F is called reducible. If for every F-invariant subspace W' in W there exists an F-invariant subspace W'' in W such that $W=W'\oplus W''$ then it is called completely reducible.

Theorem 2.1: The maximal reducible subalgebras of algebra $A\widetilde{O}(1,n)$ are exhausted with respect to $\widetilde{O}(1,n)$ conjugation by the following algebras: (1) $AO(1,n-1)\oplus\langle\mathbb{D}\rangle$; (2) $AO(n)\oplus\langle\mathbb{D}\rangle$; (3) $AO(1,k)\oplus AO'(n-k)\oplus\langle\mathbb{D}\rangle$, where $AO'(n-k)=\langle J_{ab}|a,b=k+1,...,n\rangle$ $(k=2,...,n-2)$; (4) $\langle G_1,...,G_{n-1}\rangle\Subset(AO(n-1)\oplus\langle J_{0n},\mathbb{D}\rangle)$, where $G_a=J_{0a}-J_{an}$ $(a=1,...,n-1)$.

Proof: If L is a maximal subalgebra of the algebra $A\widetilde{O}(1,n)$ then $L=AO(1,n)$ or $L=L_1\oplus\langle\mathbb{D}\rangle$, where L_1 is a maximal subalgebra of the algebra $AO(1,n)$. Let F be a maximal reducible subalgebra of the algebra $AO(1,n)$, U' a subspace of the space U invariant under F. If U' is a degenerate space then it contains one-dimensional F-invariant isotropic space W conjugated under $O(1,n)$ to the space $\langle P_0+P_n\rangle$. In this case

$$F=\{X\in AO(1,n)|X(P_0+P_n)\in\langle P_0+P_n\rangle\}.$$

It is not difficult to show that

$$F=\langle G_1,...,G_{n-1}\rangle\Subset(AO(n-1)\oplus\langle J_{0n}\rangle).$$

If U' is a nondegenerate space of dimension r then it possesses an orthogonal basis consisting of r vectors with nonzero length. Let r_+,r_- be numbers of positive and negative length vectors, in the given basis of the space U', respectively. These numbers are independent of the choice of basis. In accordance with Witt's mapping theorem any two spaces U' and U'_1, for which $r_+=r^1_+$, $r_-=r^1_-$, are mutually conjugate under the group $O(1,n)$. Obviously, $r_+\in\{0,1\}$. Since $U=U'\oplus U'^{\perp}$ and $U'^{\perp}$ is invariant under F therefore F is $O(1,n)$ conjugated to one of the algebras,

$$AO(1,n-1),\ AO(n),\ AO(1,k)\oplus AO'(n-k).$$

The theorem is proved. Let

$$A\widetilde{E}(n)=\langle P_1,...,P_n\rangle\Subset(AO(n)\oplus\langle\mathbb{D}\rangle),$$
$$AE'(n-k)=\langle P_{k+1},...,P_n\rangle\Subset AO'(n-k),$$

and $A\widetilde{G}(n-1)$ is the extended Galilei algebra with the basis

$$M=P_0+P_n,P_0,P_1,...,P_{n-1},G_1,...,G_{n-1},J_{ab}$$
$$(a,b=1,...,n-1).$$

According to Theorem 2.1, the description of subalgebras of the algebra $A\widetilde{P}(1,n)$ is reduced to the description with respect to the $\widetilde{P}(1,n)$ conjugation of irreducible subalgebras of the algebra $AO(1,n)$ and subalgebras of the following algebras:

$$\langle P_0\rangle\Subset A\widetilde{E}(n),\quad(AP(1,k)\oplus AE'(n-k))\Subset\langle\mathbb{D}\rangle,$$
$$A\widetilde{G}(n-1)\Subset\langle J_{0n},\mathbb{D}\rangle\quad(k=2,...,n-1).$$

Let π be the projection of the algebra $A\widetilde{P}(1,n)$ onto $A\widetilde{O}(1,n)$, F a nonzero subalgebra of $A\widetilde{O}(1,n)$, and $\hat{F}$ such subalgebra of $A\widetilde{P}(1,n)$ that $\pi(\hat{F})=F$. If the algebra $\hat{F}$ is $\widetilde{P}(1,n)$ conjugated to the algebra $W\Subset F$, where W is an F-invariant subspace of the space U, then we shall assume $\hat{F}$ to be splitting. If every subalgebra $\hat{F}\subset A\widetilde{P}(1,n)$ satisfying $\pi(\hat{F})=F$ is splitting, we shall say that subalgebra F possesses only splitting extensions in the algebra $A\widetilde{P}(1,n)$. The splittability of subalgebras for other algebras of inhomogeneous transformations is defined by analogy. If nothing is reserved, then the investigation of subalgebras of given algebra for conjugation is carried out with respect to the group of inner automorphisms.

The affine group $IGL(n,R)$ is defined as a group of matrices

$$\begin{pmatrix}B&Y\\0&1\end{pmatrix},\tag{2.4}$$

where $B\in GL(n,R)$, $Y\in R^n$. The Lie algebra $AIGL(n,R)$ of this group consists of matrices

$$\begin{pmatrix}X&Y\\0&0\end{pmatrix},$$

where X is a square matrix of degree n over R. Let 0_a be the zero matrix of degree a, $P_a=E_{a,n+1}$. Let us identify X and $\mathrm{diag}[X,0_1]$, then $AIGL(n,R)=\langle P_1,...,P_n\rangle\Subset AGL(n,R)$. If $m<n$, then we shall assume that $AGL(m,R)$ consists of the matrices $\mathrm{diag}[\overline{X},0_{n+1-m}]$, where $\deg\overline{X}=m$.

Lemma 2.1: Let F be a completely reducible subalgebra of the Lie algebra $\mathrm{AGL}(m,R)$ $(m<n)$, which is not semisimple. If Z is a nonzero central element of the algebra F and $\hat{F}$ is the Lie algebra, which is obtained from F by replacing Z by $Z+P_{m+1}$, then the algebra $\hat{F}$ is nonsplitting in $\mathrm{AIGL}(n,R)$ with respect to $\mathrm{IGL}(n,R)$ conjugation.

Proof: Let X_0 be a square matrix of the degree m, $T=\mathrm{diag}[X_0,0_{n-m}]$, $Z=\mathrm{diag}[T,0_1]$,

$$P_{m+1}=\begin{pmatrix}0_n & Y_{m+1}\\ 0 & 0_1\end{pmatrix}.$$

If $\hat{F}$ is a splitting algebra, then there exists the matrix C of the form (2.4) such that $C(Z+P_{m+1})C^{-1}=\mathrm{diag}[T',0_1]$. It follows that $-BTB^{-1}Y+BY_{m+1}=0$, which implies that $Y_{m+1}=(TB^{-1})Y$. However,

$$TB^{-1}=\begin{pmatrix}X_0 & 0\\ 0 & 0_{n-m}\end{pmatrix}\cdot\begin{pmatrix}B_1 & B_2\\ B_3 & B_4\end{pmatrix}=\begin{pmatrix}X_0B_1 & X_0B_2\\ 0 & 0_{n-m}\end{pmatrix},$$

and therefore

$$(TB^{-1})\cdot Y=\begin{pmatrix}\alpha_1\\ \vdots\\ \alpha_m\\ 0\\ \vdots\\ 0\end{pmatrix}.$$

This contradiction proves the lemma.

Proposition 2.1: Let F be a completely reducible Lie algebra of linear transformations of vector space V over the field R, W is an irreducible F submodule of module V. If $FW\neq 0$, then algebra F possesses only splitting extensions in algebra $W\dotplus F$.

Proof: Since F is a completely reducible subalgebra of the algebra $\mathfrak{gl}(V)$, then $F=Q\oplus Z(F)$, where Q is Levy's factor and $Z(F)$ is the center of F.[23] Using Jacobi identity it is not difficult to conceive that $F=F_1\oplus F_2$, where $F_1W=0$ and every direct summand of algebra F_2 annuls in W only zero subspace. Further we may restrict ourselves only with the case when $F=F_2$.

Let $Q\neq 0$; $\hat{F}$ be such a subalgebra of the algebra $W\dotplus F$ that its projection onto F coincides with F. According to Whitehead's theorem[23] $H^1(Q,W)=0$. From this it follows that the algebra $\hat{F}$ contains Q. Let $J\in Z(F)$, $Y\in W$, $Y\neq 0$, and $J+Y\in\hat{F}$. Since $[Q,Y]\neq 0$, then there exists such an element $X\in Q$ that $[X,Y]\neq 0$. Let $Y_1=[X,Y]$, W_1 be the F submodule of module W, generated by Y_1. Because of the fact that $W_1\neq 0$ and W is the irreducible F module we have $W_1=W$. Hence $J\in\hat{F}$. Therefore, if $Q\neq 0$ then $F\subset\hat{F}$, i.e., $\hat{F}$ is a splitting algebra.

Let $Q=0$, $J\in Z(F)$. Since J annuls in W the only zero subspace is then $[J,W]=W$. Whence for every $Y\in W$ there exists such element $Y'\in W$ that $[J,Y']=Y$. Consequently we may suppose that $J\in\hat{F}$. If $\hat{F}$ contains J_1+Y_1, where $Y_1\in W$ and $Y_1\neq 0$, then $[J,Y_1]\in\hat{F}$ and $[J,Y_1]\neq 0$. Arguing as in the case $Q\neq 0$, we get that $J_1\in\hat{F}$, i.e., $\hat{F}$ is a splitting algebra. The proposition is proved.

Proposition 2.2: Let

$$\mathrm{A}\tilde{\mathrm{E}}(n-1)=\langle G_1,\dots,G_{n-1}\rangle\dotplus(\mathrm{AO}(n-1)\oplus\langle J_{0n}\rangle),$$

where $G_a=J_{0a}-J_{an}$ $(a=1,\dots,n-1)$. The subalgebra $F\subset\mathrm{AO}(n-1)\oplus\langle J_{0n}\rangle$ possesses only splittable extensions in $\mathrm{A}\tilde{\mathrm{E}}(n-1)$ if and only if F is a semisimple algebra or F is not conjugated to a subalgebra of the algebra $\mathrm{AO}(n-2)$.

Proof: Let $W=\langle G_1,\dots,G_{n-1}\rangle$. Since every subalgebra of the algebra $\mathrm{AO}(n-1)$ is completely reducible and $[J_{0n},G_a]=-G_a$, then every subalgebra F of the $\mathrm{AO}(n-1)\oplus\langle J_{0n}\rangle$ algebra is also a completely reducible algebra of linear transformations of space W.

Let $W=W_1\oplus\cdots\oplus W_s$ be the decomposition of W into the direct sum of irreducible F modules. If projection F onto $\langle J_{0n}\rangle$ is nonzero, then $[F,W_i]=W_i$ for every $i=1,\dots,s$. Whence according to Proposition 2.1 F has only splittable extensions in $\mathrm{A}\tilde{\mathrm{E}}(n-1)$. Let us assume that projection of F onto $\langle J_{0n}\rangle$ is equal to 0. If F is a semisimple algebra then by Whitehead's theorem every extension of F in $\mathrm{A}\tilde{\mathrm{E}}(n-1)$ is splitting. Let F not be a semisimple algebra. When $\dim W_i\geqslant 2$ for every $i=1,\dots,s$ we have $[F,W_i]\neq 0$ and in view of Proposition 2.1 F possesses only splitting extensions in $\mathrm{A}\tilde{\mathrm{E}}(n-1)$. When $\dim W_i=1$ $(1\leqslant i\leqslant s)$, the module W_i is annuled by the algebra F and the algebra F is conjugated to a subalgebra of the algebra $\mathrm{AO}(n-2)$. If $Z(F)$ is the center of F and X is a nonzero element of $Z(F)$ then for every nonzero $Y\in W_i$ there exists a subalgebra $\hat{F}$ of the algebra $\mathrm{A}\tilde{\mathrm{E}}(n-1)$, which is obtained from F by replacing X by $X+Y$. By Lemma 2.1 $\hat{F}$ is not splitting. The proposition is proved.

From Theorem 2.1 and properties of solvable subalgebras of algebra $\mathrm{AO}(n)$ it follows that if n is odd then $\mathrm{AO}(1,n)$ possesses with respect to $\mathrm{O}(1,n)$ conjugation only one maximal solvable subalgebra

$$\langle G_1,\dots,G_{n-1},J_{12},J_{34},\dots,J_{n-2,n-1},J_{0n}\rangle.$$

If n is even then $\mathrm{AO}(1,n)$ possesses two maximal solvable subalgebras

$$\langle J_{12},J_{34},\dots,J_{n-1,n}\rangle;$$
$$\langle G_1,\dots,G_{n-1},J_{12},J_{34},\dots,J_{n-3,n-2},J_{0n}\rangle.$$

Since an extension of an Abelian algebra with the help of a solvable algebra is a solvable algebra itself then maximal solvable subalgebras of the algebra $\mathrm{AP}(1,n)$ are of the form $U\dotplus F$, where F is the maximal solvable subalgebra of the algebra $\mathrm{AO}(1,n)$. Maximal solvable subalgebras of the $\mathrm{A}\tilde{\mathrm{P}}(1,n)$ are exhausted by algebras $U\dotplus(F\oplus\langle\mathbb{D}\rangle)$.

Proposition 2.3: Let $\mathrm{AH}(t)$ be the Cartan subalgebra of the algebra $\mathrm{AO}(t)$. The maximal Abelian subalgebras of the algebra $\mathrm{A}\tilde{\mathrm{O}}(1,n)$ are exhausted with respect to $\tilde{\mathrm{O}}(1,n)$ conjugation by the following algebras: $\mathrm{AH}(n-1)\oplus\langle J_{0n},\mathbb{D}\rangle$; $\mathrm{AH}(n)\oplus\langle\mathbb{D}\rangle$ $[n\equiv 0(\mathrm{mod}\,2)]$; $\langle G_1,\dots,G_{n-1},\mathbb{D}\rangle$; $\mathrm{AH}(2a)\oplus\langle G_{2a+1},\dots,G_{n-1},\mathbb{D}\rangle$ $(a=1,\dots,[n-2/2])$. The written algebras are pairwise nonconjugated.

Proof: If F is a maximal Abelian subalgebra of the algebra $\mathrm{A}\tilde{\mathrm{O}}(1,n)$ then from Proposition 2.2 $F=\Omega\oplus L\oplus\langle\mathbb{D}\rangle$, where L is a subalgebra of the algebra $\mathrm{AO}(l)\oplus\langle J_{0n}\rangle$ or the algebra $\mathrm{AO}(n)$ and Ω is a subalgebra of the algebra $\langle G_1,\dots,G_{n-1}\rangle$. If projection L onto $\langle J_{0n}\rangle$ is different from 0 then $\Omega=0$. Let projection L onto $\langle J_{0n}\rangle$ be equal to 0. If $L=\mathrm{AH}(n)$, then $\Omega=0$. If $L=\mathrm{AH}(2a)$, $1\leqslant a\leqslant[n-2/2]$, then $\Omega=\langle G_{2a+1},\dots,G_{n-1}\rangle$. The proposition is proved.

III. COMPLETELY REDUCIBLE SUBALGEBRAS OF THE ALGEBRA $\widetilde{AO}(1,n)$

In this section we shall prove a number of general results on completely reducible subalgebras of the algebra $\widetilde{\mathrm{AO}}(1,n)$ and shall indicate how to search invariant subspaces of space U for these algebras. The main results of this section are Proposition 3.3 and Theorem 3.1.

Proposition 3.1: If $n \geqslant 2$ then any irreducible subalgebra of the algebra $\mathrm{AO}(1,n)$ is semisimple and noncompact.

Proof: Let F be an irreducible subalgebra of the algebra $\mathrm{AO}(1,n)$, $Z(F)$ the center of F. If $Z(F) \neq 0$ then $Z(F) = \langle J \rangle$, where $J^2 = -E_{n+1}$. Let X be an arbitrary element of the form (2.2) of the algebra $\mathrm{AO}(1,n)$. If $X^2 = -E_{n+1}$ then $\alpha_{01}^2 + \alpha_{02}^2 + \cdots + \alpha_{0n}^2 = -1$. This contradiction proves that $Z(F) = 0$.

If F is a compact algebra then there exists such symmetric matrix C that $C^{-1}FC \subset \mathrm{AO}(n+1)$.[24] Since

$$\exp(C^{-1}FC) = C^{-1} \cdot \exp F \cdot C$$

then in $\mathrm{O}(n+1)$ there exists an irreducible subgroup conserving simultaneously

$$x_0^2 + x_1^2 + \cdots + x_n^2 \quad \text{and} \quad \lambda_0^2 x_0^2 - \lambda_1^2 x_1^2 - \cdots - \lambda_n^2 x_n^2$$

($\lambda_0, \lambda_1, \ldots, \lambda_n$ are nonzero real numbers). This contradiction proves the second part of the proposition.

Proposition 3.2: A reducible subalgebra of the algebra $\widetilde{\mathrm{AO}}(1,n)$ is completely reducible if and only if it is conjugated to $L_1 \oplus L_2$ or a subalgebra of algebra $L \oplus \langle \mathbb{D} \rangle$, where L_1 is an irreducible subalgebra of the algebra $\mathrm{AO}(1,k)$ $(k \geqslant 2)$, L_2 is a subalgebra of $\mathrm{AO}'(n-k) \oplus \langle \mathbb{D} \rangle$ and L is one of the algebras, $\mathrm{AO}(n)$, $\mathrm{AO}(n-1) \oplus \langle J_{0n} \rangle$.

Proposition 3.2 follows from Theorem 2.1, Propositions 2.2 and 3.1, and the fact that G_a acts noncompletely reducible onto the space $\langle P_0 + P_n, P_a \rangle$.

Let L be a direct sum of the Lie algebras $L_1, \ldots, L_s$, B a Lie subalgebra of L, and π_i the projection L onto L_i. If $\pi_i(B) = L_i$ for $i = 1, \ldots, s$, then B is called a subdirect sum of $L_1, \ldots, L_s$.

Proposition 3.3: A completely reducible subalgebra $F \subset \widetilde{\mathrm{AO}}(1,n)$ has only splitting extensions in $\widetilde{\mathrm{AP}}(1,n)$ if and only if F is semisimple or F is nonconjugate to subalgebra of one of the algebras, $\mathrm{AO}(n)$ or $\mathrm{AO}(1,n-1)$.

The proof of Proposition 3.3 is analogous to that of Proposition 2.2.

Let A_i be a Lie algebra over R $(i = 1,2)$, $f: A_1 \to A_2$ is an isomorphism, $B = \{(X, f(X)) | X \in A_1\}$. Here B is the Lie algebra over R with "componentwise" operational rules,

$$[(X,f(X)),(X',f(X'))] = ([X,X'], f([X,X'])),$$

$$(X,f(X)) + (X',f(X')) = (X+X', f(X+X')),$$

$$\lambda(X,f(X)) = (\lambda X, f(\lambda X)),$$

where $X, X' \in A_1$, $\lambda \in \mathrm{R}$. Let us denote it as (A_1, A_2, φ). Evidently (A_1, A_2, φ) is the subdirect sum of the algebras A_1 and A_2.

Let W_i be a left A_i module $(i = 1,2)$. It is easy to see that W_i is the B module if we put

$$(X,f(X)) \cdot Y_1 = X \cdot Y_1, \quad (X,f(X)) \cdot Y_2 = f(X) \cdot Y_2,$$

for every $X \in A_1$, $Y_i \in W_i$ $(i = 1,2)$. Let W be a B submodule of the module $W_1 \oplus W_2$. If $W = W_1' \oplus W_2'$, where $W_i' \subset W_i$ $(i = 1,2)$ then W is called a splitting B module. Otherwise the module W is called nonsplitting B module.

Lemma 3.1: Let $B = (A_1, A_2, \varphi)$ and V_i be a left A_i module $(i = 1,2)$. In the B module $V_1 \oplus V_2$ exists a nonsplitting B submodule if and only if the B modules V_1 and V_2 have isomorphic composition factors.

Proof: Let W be a nonsplitting B submodule of the module $V_1 \oplus V_2$. Then W is the subdirect sum of the modules W_1 and W_2, where $W_i \subset V_i$ $(i = 1,2)$. Let $S_i = W \cap V_i$ $(i = 1,2)$. Evidently, S_i is the B submodule of the module W. The module $W/(S_1 \oplus S_2)$ is nonsplitting B submodule of the module $V_1/S_1 \oplus V_2/S_2$. Whence we shall assume that $W \cap V_i = 0$ $(i = 1,2)$.

For every element $Y_1 \in W_1$ there exists only one such element $Y_2 \subset W_2$ such that $(Y_1, Y_2) \in W$. We put $\varphi(Y_1) = Y_2$. The mapping φ is the isomorphism of B modules W_1 and W_2. In this case modules W_1 and W_2 have isomorphic composition factors. The necessity is proved.

Let W_i be a left B submodule of the module V_i $(i = 1,2)$ and let the composition factor W_1/N_1 of the module W_1 be isomorphic to the composition factor W_2/N_2 of the module W_2. We denote as W the vector space over the field R generated by the pairs $(Z_1, 0)$, $(0, Z_2)$, (Y_1, Y_2), where $Z_i \in N_i$, $Y_i \in W_i$ $(i = 1,2)$ and $\varphi(Y_1 + N_1) = Y_2 + N_2$ for the isomorphism φ: $W_1/N_1 \to W_2/N_2$. It is easy to see that W is a nonsplitting B module. The sufficiency of the lemma is proved.

Let Γ: $X \to X$ be the trivial representation of the completely reducible algebra $F \subset \widetilde{\mathrm{AO}}(1,n)$, the projection of which onto $\mathrm{AO}(1,n)$ has not any invariant isotropic subspaces in the space U or annuls the isotropic subspaces. Then Γ is $\mathrm{O}(1,n)$ equivalent to $\mathrm{diag}[\Gamma_1, \ldots, \Gamma_m]$, where Γ_i is an irreducible subrepresentation $(i = 1, \ldots, m)$. One may suppose that algebra $F_i = \{\mathrm{diag}[0, \ldots, \Gamma_i(X), \ldots, 0] | X \in F\}$ is an irreducible subalgebra $\widetilde{\mathrm{AO}}(W_i)$, where

$$W_i = \langle P_{k_{i-1}+1}, P_{k_{i-1}+2}, \ldots, P_{k_i} \rangle$$

$$(k_0 = -1, \quad k_m = n, \quad i = 1, \ldots, m).$$

If $F_i \neq 0$ then we shall call algebra F_i an irreducible part of the algebra F. It is well known that if representations Δ and Δ' of the Lie algebra L by skew-symmetric matrices are equivalent over R, then $C \cdot \Delta(X) \cdot C^{-1} = \Delta'(X)$ for some orthogonal matrix C $(X \in L)$. Whence and from Proposition 3.1 we conclude that if Γ_i and Γ_j are equivalent representations then we can assume that for every $X \in F$ the equality $\Gamma_i(X) = \Gamma_j(X)$ takes place. Having united equivalent nonzero irreducible subrepresentations we shall get a nonzero disjunctive primary subrepresentation of the representation Γ. Corresponding to those subalgebras of the algebra $\widetilde{\mathrm{AO}}(1,n)$ built by the same rule as the irreducible parts of F_i, we shall call them primary parts of the algebra F. If F coincides with its primary part then F is called a primary algebra.

Theorem 3.1: Let $K_1, K_2, \ldots, K_q$ be primary parts of a subalgebra F of the algebra $\widetilde{\mathrm{AO}}(1,n)$, and V a subspace of the space U invariant under F. Then $V = V_1 \oplus \cdots \oplus V_q \oplus \widetilde{V}$, where $V_i = [K_i, V] = [K_i, V_i]$, $[K_j, V_i] = 0$ when $j \neq i$ $(i,j = 1, \ldots, q)$, $\widetilde{V} = \{X \in V | [F,X] = 0\}$. If the primary algebra K is the subdirect sum of the irreducible subalgebras of the algebras $\widetilde{\mathrm{AO}}(W_1)$, $\widetilde{\mathrm{AO}}(W_2), \ldots, \widetilde{\mathrm{AO}}(W_1)$, respectively, then nonzero subspaces W of the space U with the condition

$[K,W]=W$ are exhausted with respect to $\mathrm{O}(1,n)$ conjugation by the spaces W_1, $W_1\oplus W_2,\dots,W_1\oplus W_2\oplus\cdots\oplus W_r$.

Proof: From the complete reducibility of algebra F it follows that $V=V'\oplus V''$, where V'' is the maximal subspace of the space V, annulled by F. Further we shall suppose that $V=V'$. From Proposition 3.1 one can suppose that $F\subset \mathrm{A}\widetilde{\mathrm{O}}(m)$, $m\leqslant n$. Let K_i be a subdirect sum of irreducible parts $K_{i1},\dots,K_{is_i}$, $V_{ij}=[K_{ij},V]$, π_a be a projection of V onto

$$\sum_{j=1}^{s_a}\oplus V_{aj}\,.$$

In view of Lemma 3.1 $\pi_a(V)\subset V$ and that is why

$$V=\sum_{a=1}^{q}\oplus\pi_a(V)\,.$$

Since K_a annuls in $\pi_a(V)$ only the zero subspace, then $[K_a,V]=[K_a,\pi_a(V)]=\pi_a(V)$.

Let primary algebra K be a subdirect sum of irreducible subalgebras of algebras $\mathrm{A}\widetilde{\mathrm{O}}(W_1)$, $\mathrm{A}\widetilde{\mathrm{O}}(W_2),\dots,\mathrm{A}\widetilde{\mathrm{O}}(W_r)$, respectively. If W is a nonzero subspace of the space

$$\Omega=\sum_{j=1}^{r}\oplus W_j$$

and $[K,W]=W$ then in view of Witt's mapping theorem there exists such isometry $B\in\mathrm{O}(\Omega)$ that $B(W)=W_1\oplus\cdots\oplus W_s$ $(1\leqslant s\leqslant r)$ and the space W_i is invariant under BKB^{-1} $(i=1,\dots,s)$. Whence BKB^{-1} is a subdirect sum of irreducible subalgebras of algebras $\mathrm{A}\widetilde{\mathrm{O}}(W_1)$, $\mathrm{A}\widetilde{\mathrm{O}}(W_2),\dots,\mathrm{A}\widetilde{\mathrm{O}}(W_r)$, respectively. Since irreducible parts of the algebra $L\subset\mathrm{A}\widetilde{\mathrm{O}}(n)$ are defined uniquely up to conjugation then one may consider that $BKB^{-1}=K$. The theorem is proved.

On the basis of Theorem 3.1 the description of splitting subalgebras $\widehat{F}\subset\mathrm{A}\widetilde{\mathrm{P}}(1,n)$, for which $\pi(\widehat{F})$ is a completely reducible algebra and has no isotropic invariant subspaces in the space U, reduces to the description of irreducible subalgebras of the algebras $\mathrm{AO}(1,k)$ and $\mathrm{AO}(k)$ $(k=2,3,\dots,n)$. The rest of the cases can be reduced to the case of the algebra $\mathrm{A}\widetilde{\mathrm{G}}(n-1)\,{\subset\!\!\!\!+}\,\langle J_{0n},\mathbb{D}\rangle$.

IV. ON THE SUBALGEBRAS OF THE EXTENDED GALILEI ALGEBRA

The aim of this section is to study subalgebras of the algebra $\mathrm{A}\widetilde{\mathrm{G}}(n-1)$ with respect to $\widetilde{\mathrm{P}}(1,n)$ conjugation. The main result concerning this problem is contained in Theorem 4.1. Theorem 4.2 gives a description of all Abelian subalgebras of the algebra $\mathrm{A}\widetilde{\mathrm{G}}(n-1)$. As a corollary, we obtain the list of maximal Abelian subalgebras and one-dimensional subalgebras of the algebra $\mathrm{A}\widetilde{\mathrm{G}}(n-1)$.

The basis elements of the extended Galilei algebra $\mathrm{A}\widetilde{\mathrm{G}}(n-1)$ satisfy the following commutation relations:

$$[J_{ab},J_{cd}]=g_{ad}J_{bc}+g_{bc}J_{ad}-g_{ac}J_{bd}-g_{bd}J_{ac}\,;$$
$$[P_a,J_{bc}]=g_{ab}P_c-g_{ac}P_b;\quad [P_a,P_b]=0\,;$$
$$[G_a,J_{bc}]=g_{ab}G_c-g_{ac}G_b;\quad [G_a,G_b]=0\,;$$
$$[P_a,G_b]=\delta_{ab}M;\quad [P_a,M]=[G_a,M]=[J_{ab},M]=0\,;$$
$$[P_0,J_{ab}]=[P_0,M]=[P_0,P_a]=0,\quad [P_0,G_a]=P_a$$
$$(a,b,c,d=1,\dots,n-1)\,.$$

Let $V_1=\langle G_1,\dots,G_{n-1}\rangle$ be a Euclidean space with orthonormal basis $G_1,\dots,G_{n-1}$, $V_2=[P_0,V_1]$ $(n\geqslant 3)$, $\mathfrak{M}=V_1+V_2+\langle P_0,M\rangle$. We settle on identifying the group $\mathrm{O}(n-1)$ with the isometry group $\mathrm{O}(V_1)$, $\mathrm{O}(V_2)$. If W is a subspace of V_1 and $\dim W=k$ then according to Witt's theorem for every a, $0\leqslant a\leqslant n-k-1$, there exists an isometry $B_a\in\mathrm{O}(V_1)$ such that

$$B_a(W)=V_1(a+1,a+k)=\langle G_{a+1},G_{a+2},\dots,G_{a+k}\rangle\,.$$

Further, in spaces V_1, V_2 we shall consider only subspaces $V_1(a,b)$, $V_2(a,b)=[P_0,V_1(a,b)]$. We call them elementary spaces. The basis G_a, $G_{a+1},\dots,G_b$ of the space $V_1(a,b)$ and the basis P_a, $P_{a+1},\dots,P_b$ of the space $V_2(a,b)$ we shall call canonical.

Let W_1, W_2 be subspaces of some vector space W over the field R and $W_1\cap W_2=0$. If φ: $W_1\to W_2$ is an isomorphism then we denote as (W_1,W_2,φ) the space $\{Y+\varphi(Y)\,|\,Y\in W_1\}$. As $I(W_1,W_2)$ we denote the isomorphism of elementary spaces W_1 and W_2, by which the canonical basis of W_1 is mapped to the canonical basis of W_2 with numeration of the basis of elements maintained.

Let $\mathrm{AG}(n-1)=\mathrm{A}\widetilde{\mathrm{G}}(n-1)/\langle M\rangle$. For the generators of the $\mathrm{AG}(n-1)$ we preserve the notation of the generators of the algebra $\mathrm{A}\widetilde{\mathrm{G}}(n-1)$. By τ, τ_0, τ_1, and τ_2 we denote the projection of $\mathrm{A}\widetilde{\mathrm{G}}(n-1)$ and $\mathrm{AG}(n-1)$ onto $\mathrm{AO}(n-1)\oplus\langle P_0\rangle$, P_0, V_1, and V_2, respectively.

Let F be a subalgebra of the $\mathrm{AO}(n-1)\oplus\langle P_0\rangle$, $\widehat{F}$ an subalgebra of the $\mathrm{AG}(n-1)$ such that $\tau(\widehat{F})=F$. If algebra $\widehat{F}$ is conjugated to the algebra $W\,{\subset\!\!\!\!+}\,F$, where W is the F-invariant subspace of space V_1+V_2, then $\widehat{F}$ is called splitting in the algebra $\mathrm{AG}(n-1)$. The notion of a splitting subalgebra of the algebra $\mathrm{A}\widetilde{\mathrm{G}}(n-1)$ is defined analogously.

Proposition 4.1: Let L_1 be a subalgebra of the $\mathrm{AO}(n-1)$, L_2 be a subalgebra of the $\langle P_0\rangle$, and F be the subdirect sum of L_1 and L_2. If $P_0\notin F$ then the algebra F only has splitting extensions in the algebra $\mathrm{AG}(n-1)$ if and only if L_1 is a semisimple algebra or L_1 is not conjugated to any subalgebra of the algebra $\mathrm{AO}(n-2)$. When $P_0\in F$, the algebra F only has splitting extensions in the $\mathrm{AG}(n-1)$ if and only if L_1 is not conjugated to any subalgebra of the algebra $\mathrm{AO}(n-2)$.

Proof: If L_1 is a semisimple algebra and $L_2=\langle P_0\rangle$ then by Whitehead's theorem[23] $P_0\in\widehat{F}$. Let us assume that $L_2=\langle P_0\rangle$ and $P_0\notin\widehat{F}$. Let $\widehat{F}$ be an subalgebra of the $\mathrm{AG}(n-1)$ such that $\tau(\widehat{F})=F$. If L_1 is not conjugated to any subalgebra of the $\mathrm{AO}(n-2)$ then by Proposition 2.2 the algebra $\widehat{F}$ is splitting. If L_1 is conjugated to some subalgebra of $\mathrm{AO}(n-2)$ then $F=\langle X\rangle\oplus F_1$ where $X\neq 0$, $\langle X\rangle$, and F_1 are subalgebras of the algebra $\mathrm{AO}(n-2)\oplus\langle P_0\rangle$. The algebra

$$\widehat{F}=\langle P_1,\dots,P_{n-2},P_{n-1},G_1,\dots,G_{n-2},X+G_{n-1}\rangle\,{\subset\!\!\!\!+}\,F_1$$

is not splitting by Lemma 2.1. The case $L_2=0$ can be treated similarly.

Let $P_0\in F$. If $L_1\subset\mathrm{AO}(n-2)$ then algebra $\langle P_0+G_{n-1}\rangle\,{\subset\!\!\!\!+}\,L_1$ is nonsplitting. If L_1 is not conjugated to any subalgebra of the algebra $\mathrm{AO}(n-2)$ then by way of complete reducibility of the algebra L_1 we get that $P_0\in\widehat{F}$ and whence algebra $\widehat{F}$ is splitting. The proposition is proved.

Proposition 4.2: The subalgebra F of the algebra $\mathrm{AO}(n-1)\oplus\langle P_0\rangle$ has only splitting extensions in the

$A\tilde{G}(n-1)$ if and only if F is a semisimple algebra.

Lemma 4.1: Let $W_1 = \langle Y_1,...,Y_m \rangle$, $W_2 = \langle Z_1,...,Z_m \rangle$ be Euclidean spaces over the field R, $O(W_i)$ the isometry group of W_i $(i=1,2)$, $0<\alpha_1 \leqslant \alpha_2 \leqslant \cdots \leqslant \alpha_t$, $S_0^t = 0$, $S_j^t = \langle Z_{t+1},...,Z_{t+j} \rangle$ $(j=1,...,m-t)$. The subspaces of the space $W_1 \oplus W_2$ are exhausted with respect to $O(W_1) \times O(W_2)$ conjugation by the following spaces:

$$O, \langle Y_1,...,Y_r \rangle, \quad \langle Z_1,...,Z_s \rangle,$$
$$\langle Y_1,...,Y_r,Z_1,...,Z_s \rangle \quad (r,s=1,...,m);$$
$$\langle Y_1,...,Y_k,Y_{k+1}+\alpha_1 Z_1,...,Y_{k+t}+\alpha_t Z_t \rangle \oplus S_j^t$$
$$(k=1,...,m-1, \quad t=1,...,m-k, \quad j=0,1,...,m-t);$$
$$\langle Y_1+\alpha_1 Z_1,...,Y_t+\alpha_t Z_t \rangle \oplus S_j^t$$
$$(t=1,...,m, \quad j=0,1,...,m-t).$$

Proof: Let N be a subspace of $W_1 \oplus W_2$ and $N \neq W_1' \oplus W_2'$, where W_i' is a subspace of W_i $(i=1,2)$. If $B_i = N \cap W_i$, N_i is a projection of N onto W_i $(i=1,2)$ and then $N_1/B_1 \cong N_2/B_2$. Let $\dim B_1 = k$. By Witt's theorem the space B_1 is conjugated to the space $\langle Y_1,...,Y_k \rangle$. If $\dim(N_1|B_1) = t$ then N contains elements $Y_{k+j} + \alpha_{1j} Z_1 + \cdots + \alpha_{tj} Z_t$ $(j=1,...,t)$, and moreover the matrix $A = (\alpha_{ij})$ is nonsingular. The matrix A can be represented uniquely in the form CT, where C is an orthogonal matrix and T is a positively definite symmetric matrix.

The isometry $\mathrm{diag}[E_m, C^{-1}, E_{m-t}]$ maps N onto the space to which the matrix $C^{-1}(CT) = T$ corresponds. There exists such orthogonal matrix C_1 that $C_1 T C_1^{-1} = \mathrm{diag}[\lambda_1,...,\lambda_t]$. The isometry $\mathrm{diag}[E_k, C_1, E_{m-k-t}, C_1, E_{m-t}]$ maps N onto the space to which the matrix $C_1 T C_1^{-1}$ corresponds. Therefore N is conjugated to the space

$$B_1 \oplus \langle Y_{k+1} + \alpha_1 Z_1,...,Y_{k+t} + \alpha_t Z_t \rangle \oplus B_2,$$

where $0<\alpha_1 \leqslant \alpha_2 \leqslant \cdots \leqslant \alpha_t$. The lemma is proved.

Let K be the primary subalgebra of the algebra $AO(n-1)$ which is a subdirect sum of irreducible subalgebras of the algebras $AO(V_1(1,q))$, $AO(V_1(q+1,2q))$,..., $AO(V_1((r-1)q+1,rq))$, respectively, and W a nonzero subspace of the space $\mathfrak{M}$ with the property $[K,W] = W$. If $\tau_1(W) = 0$ then by way of Theorem 3.1 W is conjugated to the space $V_2(1,iq)$ $(1 \leqslant i \leqslant r)$. If $\tau_2(W) = 0$ then W is conjugated to $V_1(1,iq)$ $(1 \leqslant i \leqslant r)$. Let us suppose that $\tau_1(W) \neq 0$, $\tau_2(W) \neq 0$. Then W is a subdirect sum of $\tau_1(W)$, $\tau_2(W)$, where $\tau_1(W) = V_1(1,m)$ and $\tau_2(W)$ coincides with $V_2(1,k)$ or $V_2(m+1,m+l)$ or a subdirect sum of $V_2(1,k)$ and $V_2(m+1,m+l)$ $(k \leqslant m)$. Every number of k, m, and l is divisible by q. Let us consider the case when $\tau_2(W)$ is a subdirect sum of $V_2(1,k)$ and $V_2(m+1,m+l)$. In the space W we choose the basis in the following form:

$$G_a + \alpha_a^i P_i, \quad \beta_c^i P_i \tag{4.1}$$
$$(a=1,...,m, \; c=m+1,...,m+t,$$
$$i=1,...,k,m+1,...,m+l).$$

The coefficients of the decomposition we write down as the corresponding columns of the matrix

$$\Gamma = \begin{pmatrix} A_1 & B_1 \\ A_2 & B_2 \end{pmatrix},$$

having $m+t$ columns and $k+l$ lines. We call the matrix Γ a coupling matrix of elementary spaces in the space W. With the coupling matrix we shall carry out the transformations corresponding to definite $O(n-1)$ automorphisms and transformations to new bases of the form (4.1). Let $C_1 \in O(k)$, $C_2 \in O(m-k)$, $C_3 \in O(l)$, $S = \mathrm{diag}[C_1, C_2]$, T be a $t \times m$ matrix, and T_2 a nonsingular matrix of degree t. The most general admissible transformations of the coupling matrix have the form

$$\begin{pmatrix} A_1 & B_1 \\ A_2 & B_2 \end{pmatrix} \to \begin{pmatrix} C_1 A_1 S^{-1} + C_1 B_1 T_1 & C_1 B_1 T_2 \\ C_3 A_2 S^{-1} + C_3 B_2 T_1 & C_3 B_2 T_2 \end{pmatrix}.$$

If $B_2 \neq 0$ then according to Theorem 3.1 for some matrices C_3, T_2, the following equality is correct:

$$C_3 B_2 T_2 = \begin{pmatrix} 0 & 0 \\ 0 & \Delta_1 \end{pmatrix},$$

where $\Delta_1 = \mathrm{diag}[\mu_1 E_q,...,\mu_a E_q]$, $\mu_1 = \cdots = \mu_a = 1$. By this transformation algebra K is left invariant. Applying Theorem 3.1 again we get that with $k=m$ the matrix A_2 can be transformed into matrix

$$\begin{pmatrix} \Delta_2 & 0 \\ 0 & 0 \end{pmatrix},$$

where Δ_2 is a square matrix of degree bq. For simplicity we shall assume that Δ_2 is a coupling matrix of elementary spaces in the subdirect sum of the spaces $V_1(1,bq)$ and $V_2(bq+1,2bq)$. One can admit that

$$K = \mathrm{diag}[A^b, A^b] = \{\mathrm{diag}[\underbrace{X,...,X}_{2b}] \,|\, X \in A\},$$

where A is an irreducible subalgebra of the algebra $AO(q)$. Since for every matrix $Y \in A^b$ the equality $\Delta_2 Y = Y \Delta_2$ takes place then $\Delta_2 = QS$, where S is a symmetric matrix, Q is an orthogonal matrix, and $Y \cdot Q = Q \cdot Y$. Applying the automorphism $\mathrm{diag}[E, Q^{-1}]$ we transform the coupling matrix Δ_2 into S. There exists such matrix $C \in O(bq)$ that

$$CSC^{-1} = \mathrm{diag}[\lambda_1 E_{(1)}, \lambda_2 E_{(2)},...,\lambda_t E_{(t)}],$$

where $\lambda_i \neq \lambda_j$ when $i \neq j$, and $E_{(i)}$ is the unit matrix $(i,j=1,...,t)$. The automorphism $\mathrm{diag}[C,C]$ transforms K into $\mathrm{diag}[CA^bC^{-1}, CA^bC^{-1}]$ and the coupling matrix S into CSC^{-1}. If $Y \in CA^bC^{-1}$ then $Y(CSC^{-1}) = (CSC^{-1})Y$. Whence $Y = \mathrm{diag}[Y_1, Y_2,...,Y_t]$ where $\deg Y_i = \deg E_{(i)}$. The further decomposition of the blocks Y_i by $O(2bq)$ automorphisms $\mathrm{diag}[\overline{C}, \overline{C}]$, where $\overline{C} = \mathrm{diag}[C_1,...,C_t]$, $\deg C_i = \deg E_{(i)}$ does not change the coupling matrix. Since irreducible parts of an algebra are defined uniquely then by the considered transformations of the coupling matrix the algebra K is left invariant. That is why one can suppose that with $k=m$

$$C_3 A_2 S^{-1} + C_3 B_2 T_1 = \begin{pmatrix} \Delta_2 & 0 \\ 0 & 0 \end{pmatrix},$$

where $\Delta_2 = \mathrm{diag}[\lambda_1 E_q,...,\lambda_b E_q]$, $0<\lambda_1 \leqslant \cdots \leqslant \lambda_b$, and $(a+b)q = l$ or $\lambda_1 = \cdots = \lambda_b = 0$ and $aq = l$. If $B_1 \neq 0$ then for some C_1, T_2 we have

$$C_1 B_1 T_2 = \begin{pmatrix} 0 & 0 \\ 0 & \Delta_3 \end{pmatrix},$$

where $\Delta_3 = \mathrm{diag}[E_q,...,E_q]$.

The complete classification of coupling matrices one can get for large n.

Further we shall use the following notation:

$\mathfrak{M} = \langle P_0, M, P_1, \ldots, P_{n-1}, G_1, \ldots, G_{n-1} \rangle; \quad m = [(n-1)/2]$;

$$\Gamma(n-1) = \left\{ \sum_{i=1}^{m} \gamma_i J_{2i-1,2i} | \gamma_i = 0,1 \right\};$$

$X_a \cap X_b = 0$ if $X_a, X_b \in \Gamma(n-1)$ and have no common summand.

Lemma 4.2: Let $T = \alpha_1 X_1 + \cdots + \alpha_k X_k + Z$, $Z = \beta J_{0n} + \gamma \mathbb{D} + \delta P_0$, where $X_i \in \Gamma(n-1)$, $\alpha_i \neq 0$, $\alpha_i^2 \neq \alpha_j^2$, $X_i \cap X_j = 0$ when $i \neq j$ $(i,j = 1,\ldots,k)$. If W is a subspace of the space $\mathfrak{M}$ and $[T,W] \subset W$ then $W = W_1 \oplus \ldots \oplus W_k \oplus \widetilde{W}$, where $W_i = [X_i, W] = [X_i, W_i]$, $[Z, W_i] \subset W_i$, $[X_j, W_i] = 0$ when $j \neq i$, $[X_i, \widetilde{W}] = 0$, $[Z, \widetilde{W}] \subset \widetilde{W}$.

Proof: Let $X = T - Z$, $\mathfrak{M}' = [X, \mathfrak{M}]$, $\widetilde{\mathfrak{M}} = \{Y \in \mathfrak{M} | [X, Y] = 0\}$, W' be a projection of W onto $\mathfrak{M}'$, and $\widetilde{W}$ be a projection of W onto $\widetilde{M}$. Evidently, $\mathfrak{M} = \mathfrak{M}' \oplus \widetilde{\mathfrak{M}}$ (as spaces). Since composition factors of the $\langle Z \rangle$ module $\mathfrak{M}$ are one dimensional, then the composition factors of the $\langle Z \rangle$ module $\widetilde{W}$ are one dimensional, too. Let $\mathfrak{M}(P) = \{P_a \in \mathfrak{M} | [X, P_a] \neq 0\}$. It is easy to see that $\langle \mathfrak{M}(P) \rangle$ and $\mathfrak{M}'/\langle \mathfrak{M}(P) \rangle$ can be represented as direct sums of two-dimensional irreducible $\langle T \rangle$ submodules. Whence the dimensions of composition factors of the $\langle T \rangle$ module W' are equal to 2, too. When we now apply Lemma 3.1 we conclude that $W = W' \oplus \widetilde{W}$.

Let $\mathfrak{M}_i = [X_i, \mathfrak{M}]$ and W_i be a projection of W' onto $\mathfrak{M}_i$. Clearly $\mathfrak{M}' = \mathfrak{M}_1 \oplus \cdots \oplus \mathfrak{M}_k$. At first let us establish that $[Z, W_i] \subset W_i$. Since for any $Y_i \in W_i$ we have $[J_{0n} - \mathbb{D}, Y_i] = -Y_i$, then we may assume that $\beta = 0$. Obviously

$$[T,[T,Y_i]] = -\alpha_i^2 Y_i + 2\alpha_i [X_i,[Z,Y_i]] + \gamma [Z,Y_i].$$

Let

$$Y_i' = 2\alpha_i [X_i,[Z,Y_i]] + \gamma [Z,Y_i],$$

$$Y_i'' = 2\alpha_i [X_i,[Z,Y_i']] + \gamma [Z,Y_i'].$$

The space W_i contains Y_i', Y_i''. It is easy to check that

$$Y_i'' = 4\alpha_i \gamma^2 [X_i,[Z,Y_i]] + \gamma(\gamma^2 - 4\alpha_i^2)[Z,Y_i].$$

The determinant constructed by the coefficients of $[X_i,[Z,Y_i]]$, $[Z,Y_i]$ in Y_i', Y_i'' is equal to $-2\alpha_i \gamma (\gamma^2 + 4\alpha_i^2)$. If $\gamma \neq 0$ then $[Z,Y_i] \in W_i$. If $\gamma = 0$ then W_i contains $Y_i' = [X_i,[\delta P_0, Y_i]]$ and $Y_i'' = [T, Y_i'] = -\alpha_i [\delta P_0, Y_i]$.

In the composition factors of the $\langle T \rangle$ module $\mathfrak{M}_i$ one can choose the basis so that the matrix of the operator T is one of the matrices

$$\begin{pmatrix} \gamma & -\alpha_i \\ \alpha_i & \gamma \end{pmatrix}, \quad \begin{pmatrix} -\beta & -\alpha_i \\ \alpha_i & -\beta \end{pmatrix}.$$

If for $i \neq j$ the modules $\mathfrak{M}_i$ and $\mathfrak{M}_j$ are possessed by isomorphic composition factors then one of the following conditions is satisfied: $\alpha_i^2 = \alpha_j^2$; $2\gamma = -2\beta$, $\gamma^2 + \alpha_i^2 = \beta^2 + \alpha_j^2$. Since it is impossible then on the basis of Lemma 3.1 we conclude that $W' = W_1 \oplus \cdots \oplus W_k$. The lemma is proved.

Proposition 4.3: Let L_1 be a subalgebra of the $AO(n-1)$, $L_2 = \langle \beta J_{0n} + \gamma \mathbb{D} + \delta P_0 \rangle$, and F a subdirect sum of L_1 and L_2. If W is a subspace of $\mathfrak{M}$ and $[F,W] \subset W$ then $[L_j, W] \subset W$ $(j = 1,2)$.

This is proved by virtue of Lemma 4.2.

Theorem 4.1: Let $V_1 = \langle G_1, \ldots, G_{n-1} \rangle$, $V_2 = [P_0, V_1]$, $V_{1,a}$ be a subspace of V_1, $V_{2,a} = [P_0, V_{1,a}]$; $K_1, K_2, \ldots, K_q$ be primary parts of nonzero subalgebra L_1 of the algebra $AO(n-1)$; $\mathfrak{N}$ be the maximal subalgebra of algebra $\mathfrak{M}$, annulled by L_1; and L_2 be a subalgebra of the algebra $\mathfrak{N} \oplus \langle J_{0n}, \mathbb{D} \rangle$. If F is the subdirect sum of L_1 and L_2, and W is a subspace of $\mathfrak{M}$ invariant under F, then $W = W_1 \oplus \cdots \oplus W_q \oplus \widetilde{W}$, where $W_i = [K_i, W] = [K_i, W_i]$, $[L_2, W_i] \subset W_i$, $[K_j, W_i] = 0$ when $j \neq i$, $[K_i, \widetilde{W}] = 0$, $[L_2, \widetilde{W}] \subset \widetilde{W}$ $(i,j = 1,\ldots,q)$.

If a primary algebra K is a subdirect sum of irreducible subalgebras of the algebras $AO(V_{1,1}), \ldots, AO(V_{1,r})$, respectively, then nonzero subspaces W of the space $\mathfrak{M}$ with the property $[K,W] = W$ are conjugated to

$$\sum_{i=1}^{a} V_{1,i}, \quad \sum_{i=1}^{a} V_{2,i} \quad (a = 1,\ldots,r)$$

or to subdirect sums of such spaces

$$\sum_{i=1}^{\tilde{a}} V_{1,i} \text{ and } \sum_{i=1}^{\tilde{b}} V_{2,i}; \quad \sum_{i=1}^{a} V_{1,i} \text{ and } \sum_{i=a+1}^{c} V_{2,i};$$

$$\sum_{i=1}^{a} V_{1,i}, \quad \sum_{i=1}^{b} V_{2,i}, \text{ and } \sum_{i=a+1}^{c} V_{2,i}$$

$$(\tilde{a} = 1,\ldots,r, \quad \tilde{b} = 1,\ldots,\tilde{a}, \quad a = 1,\ldots,r-1,$$
$$b = 1,\ldots,a, \quad c = a+1,\ldots,r).$$

The subdirect sums of the spaces

$$\sum_{i=1}^{a} V_{1,i}, \quad \sum_{i=a+1}^{c} V_{2,i}$$

are exhausted with respect to $O(n-1)$ conjugation by the following spaces:

$$\sum_{i=1}^{a} V_{1,i} \oplus \sum_{j=a+1}^{c} V_{2,j};$$

$$\sum_{i=1}^{b} (V_{1,i}, V_{2,a+1}, \lambda_i I(V_{1,i}, V_{2,a+i}))$$

$$\oplus \sum_{j=b+1}^{a} V_{1,j} \oplus \sum_{k=a+b+1}^{c} V_{2,k}$$

$$(0 < \lambda_1 \leqslant \cdots \leqslant \lambda_b, \ b = 1,\ldots,\min\{a, c-a\}).$$

The written spaces are mutually nonconjugated.

Proof: Let $Q = [L_1, W]$, S be a projection of W onto $\mathfrak{N}$. It is easy to see that W is the subdirect sum of Q and S. Since the composition factors of the L_2 module $\mathfrak{N}$ are one dimensional and the composition factors of the L_1 module $[L_1, \mathfrak{M}]$ have dimension not less than 2 then in view of Lemma 3.1 $W = Q + S$. In virtue of Proposition 4.3 $[L_2, Q] \subset Q$. We can show, as in Theorem 3.1, that $Q = W_1 \oplus \cdots \oplus W_q$, where $W_i = [K_i, Q]$, $W_i = [K_i, W_i]$ $(i = 1,\ldots,q)$. The truthfulness of the further statements is established earlier when considering the transformations of the coupling matrix of elementary spaces in the space W. The theorem is proved.

Theorem 4.2: Let $\alpha_1 \leqslant \alpha_2 \leqslant \cdots \leqslant \alpha_s$, $\alpha_1 = 0$, and $\alpha_s \in \{0,1\}$, $AH(0) = 0$, $AH(2d) = \langle J_{12}, J_{34}, \ldots, J_{2d-1,2d} \rangle$, and L be a nonzero Abelian subalgebra of the algebra $A\widetilde{G}(n-1)$. If the projection $\tau_0(L)$ of the algebra L onto $\langle P_0 \rangle$ is equal to

0 then L is conjugated to the subdirect sum of the algebras L_1, L_2, L_3, and L_4, where $L_1 \subset \mathrm{AH}(2d)$ $(0 \leqslant d \leqslant m)$, $L_2 = 0$ or $L_2 = \langle G_{2d+1} + \alpha_1 P_{2d+1}, G_{2d+2} + \alpha_2 P_{2d+2}, \ldots, G_{2d+s}, + \alpha_s P_{2d+s} \rangle$, $L_3 = 0$ or $L_3 = \langle P_{2d+s+1}, \ldots, P_l \rangle$, $L_4 = 0$ or $L_4 = \langle M \rangle$. If $\tau_0(L) \neq 0$ then L is conjugated to the subdirect sum of the algebras L_1, L_2, L_3, and L_4 where $L_1 \subset \mathrm{AH}(2d)$, $L_2 = \langle P_0 + \alpha G_{2d+1} \rangle$ $(\alpha \in \{0,1\})$, $L_3 = 0$ or $L_3 = \langle P_r, \ldots, P_t \rangle$, $L_4 = 0$ or $L_4 = \langle M \rangle$ $(0 \leqslant d \leqslant m;\ r = 2d+1$ when $\alpha = 0$; $r = 2d+2$ when $\alpha = 1)$.

Proof: Let

$$X_i = G_i + \sum_{j=2d+1}^{2d+s} \beta_{ji} P_j, \quad L = \langle X_{2d+1}, \ldots, X_{2d+s} \rangle .$$

Obviously, $[X_i, X_k] = (\beta_{ki} - \beta_{ik})M$. Since L is an Abelian algebra then $\beta_{ik} = \beta_{ki}$ and therefore $B = (\beta_{ik})$ $(i,k = 2d+1, \ldots, 2d+s)$ is a symmetric matrix. Hence, there exists a matrix $C \in \mathrm{O}(s)$ such that $CBC^{-1} = \mathrm{diag}[\lambda_1, \ldots, \lambda_s]$. Whence we can assume up to conjugacy under $\mathrm{O}(n-1)$ that $X_{2d+j} = G_{2d+j} + \lambda_j P_{2d+j}$ $(j = 1, \ldots, s)$. $\mathrm{O}(n-1)$ automorphisms permit us to change the numeration of generators $G_{2d+1}, \ldots, G_{2d+s}$. That is why we can suppose that $\lambda_1 \leqslant \cdots \leqslant \lambda_s$. Applying the automorphism $\exp(-\lambda_1 P_0)$ we get generators $G_{2d+j} + \mu_j P_{2d+j}$ $(j = 1, \ldots, s)$, where $\mu_1 = 0$, $0 \leqslant \mu_2 \leqslant \cdots \leqslant \mu_s$. If $\mu_s > 0$ then $\mu_s = \exp\theta$ $(\theta \in R)$. Evidently,

$$\exp(-\theta J_{0n})(G_{2d+j} + \mu_j P_{2d+j}) \exp(\theta J_{0n}) = \exp\theta \cdot (G_{2d+j} + \mu_j \exp(-\theta) P_{2d+j}) .$$

Therefore when $\mu_s > 0$ we can assume that $\mu_s = 1$.

The rest of the assertion of the theorem follows from Proposition 4.1. The theorem is proved.

Corollary 1: Let

$$A(r,t) = \langle G_r + \alpha_r P_r, G_{r+1} + \alpha_{r+1} P_{r+1}, \ldots, G_t + \alpha_t P_t, M \rangle ,$$

where $\alpha_r \leqslant \alpha_{r+1} \leqslant \cdots \leqslant \alpha_t$, $\alpha_r = 0$, *and* $\alpha_t = 1$ when $\alpha_t \neq 0$. The maximal Abelian subalgebras of the algebra $\mathrm{A}\widetilde{\mathrm{G}}(n-1)$ are exhausted up to conjugacy under $\widetilde{\mathrm{P}}(1,n)$ by the following algebras:

U; $\mathrm{A}(1,n-1)$; $\mathrm{A}(1,s) \oplus V_2(s+1,n-1)$ $(s = 1, \ldots, n-2)$;

$\langle G_1 + P_0, M \rangle \oplus V_2(2,n-1)$;

$\mathrm{AH}(n-2) \oplus \langle G_{n-1} + P_0, M \rangle$ $[n \equiv 0 (\mathrm{mod}\ 2)]$;

$\mathrm{AH}(2d) \oplus \langle P_0 \rangle \oplus V_2(2d+1,n)$ $(d = 1, \ldots, [(n-1)/2])$;

$\mathrm{AH}(2d) \oplus \mathrm{A}(2d+1,n-1)$ $(d = 1, \ldots, [(n-2)/2])$;

$\mathrm{AH}(2d) \oplus \mathrm{A}(2d+1,s) \oplus V_2(s+1,n-1)$ $(d = 1, \ldots, [(n-3)/2])$;

$\mathrm{AH}(2d) \oplus \langle G_{2d+1} + P_0, M \rangle \oplus V_2(2d+2,n-1)$ $(d = 1, \ldots, [(n-3)/2])$.

The written algebras are not mutually conjugated.

Corollary 2: Let $n \geqslant 3$, $X_t = \alpha_1 J_{12} + \alpha_2 J_{34} + \cdots + \alpha_t J_{2t-1,2t}$;

$\alpha_1 = 1$, $0 < \alpha_2 \leqslant \cdots \leqslant \alpha_t \leqslant 1$; $t = 1, \ldots, [(n-1)/2]$; $s = 1, \ldots, [(n-2)/2]$.

The one-dimensional subalgebras of the algebra $\mathrm{A}\widetilde{\mathrm{G}}(n-1)$ are exhausted with respect to $\widetilde{\mathrm{P}}(1,n)$ conjugation by the following algebras: $\langle P_0 \rangle$; $\langle M \rangle$; $\langle P_1 \rangle$; $\langle G_1 \rangle$; $\langle G_1 + P_2 \rangle$; $\langle G_1 + P_0 \rangle$; $\langle X_t \rangle$; $\langle X_t + P_0 \rangle$; $\langle X_t + M \rangle$; $\langle X_t + P_{2t+1} \rangle$; $\langle X_s + G_{2s+1} \rangle$; $\langle X_s + G_{2s+1} + P_0 \rangle$; $\langle X_r + G_{2r+1} + P_{2r+2} \rangle$ $(r = 1, \ldots, [(n-3)/2])$.

The written algebras are not mutually conjugated. Let

$$\Phi(0) = \langle M \rangle, \quad \Phi(i) = \langle M, P_1, \ldots, P_i \rangle, \quad \Omega(0) = \langle M, P_0 \rangle ,$$
$$\Omega(i) = \langle M, P_0, P_1, \ldots, P_i \rangle, \quad V_2(s,t) = \langle P_s, \ldots, P_t \rangle \quad (s \leqslant t) , \tag{4.2}$$
$$\Lambda_{r+1,k+1}(j) = \langle P_{r+d} + \lambda_d P_{k+d} | d = 1,2, \ldots, j \rangle ,$$

where

$$0 < \lambda_1 \leqslant \lambda_2 \leqslant \cdots \leqslant \lambda_j \quad (1 \leqslant j \leqslant k-r) .$$

Proposition 4.4: Let $L = \langle G_1, \ldots, G_k \rangle$. The subspaces of the space $U = \langle P_0, P_1, \ldots, P_n \rangle$, which are invariant under L, are exhausted with respect to $\widetilde{\mathrm{O}}(1,n)$ conjugation by the following spaces: 0, $\Phi(i)$, $\Omega(k)$, $V_2(k+1,t)$,

$$\Phi(i) \oplus V_2(k+1,t) , \quad \Omega(k) \oplus V_2(k+1,t) ,$$
$$\Phi(r) \oplus \Lambda_{r+1,k+1}(j) ,$$
$$\Phi(r) \oplus \Lambda_{r+1,k+1}(j) \oplus V_2(k+j+1,s) ,$$

where $i = 0,1, \ldots, k$, $t = k+1, \ldots, n-1$, $r = 0,1, \ldots, k-1$, $j = 1, \ldots, k-r$, $s = k+j+1, \ldots, n-1$.

Proof: Let W be a subspace of the space $\Omega(k)$ invariant under L. Since $[P_a, G_a] = M$ then with $W \neq 0$ we have $M \in W$. The normalizer of the algebra L in $\mathrm{O}(n-1)$ contains $\mathrm{O}(k)$. It follows from this and Witt's theorem that if $W \neq \langle M \rangle$ and $P_0 \notin W$ then $W = \Phi(i)$ $(1 \leqslant i \leqslant k)$. If $P_0 \in W$ then $W = \Omega(k)$.

For a description of all subspaces of the space U which are invariant under L we shall use the Goursat twist method.[25] Since by Witt's theorem the nonzero subspaces of the space $V_2(k+1,n-1)$ are exhausted with respect to $\mathrm{O}(n-1)$ conjugation by the spaces $V_2(k+1,t)$ $(t = k+1, \ldots, n-1)$ we need to classify the subdirect sums of the following pairs of spaces: $\Omega(k)$, $V_2(k+1,t)$; $\Phi(i)$, $V_2(k+1,t)$ $(i = 0,1, \ldots, k$, $t = k+1, \ldots, n-1)$.

Let N be the subdirect sum of $\Omega(k)$ and $V_2(k+1,t)$. If $P_0 + \lambda P_{k+1} \in N$ $(\lambda \neq 0)$ then N contains P_1, $P_1 = -[G_1, P_0 + \lambda P_{k+1}]$, and whence it contains M, too. Let

$$N' = \exp(\theta G_{k+1}) \cdot N \cdot \exp(-\theta G_{k+1}) .$$

The space N' contains $P_0 + (\lambda - \theta) P_{k+1} + (\theta^2/2 - \lambda\theta) M$. Since $M \in N'$ then $P_0 + (\lambda - \theta) P_{k+1} \in N'$. Putting $\theta = \lambda$ we get that $P_0 \in N'$ and whence $\Omega(k) \subset N'$. Therefore $N' = \Omega(k) \oplus V_2(k+1,t')$.

Let N be the subdirect sum of $\Phi(i)$ and $V_2(k+1,t)$. If $i = 0$, $M + \lambda P_{k+1} \in N$ $(\lambda \neq 0)$ then N' contains $(1 - \theta\lambda) \times M + \lambda P_{k+1}$. Putting $1 - \theta\lambda = 0$ we get that $N' = V_2(k+1,t)$. If $i \neq 0$ then $M \in N$. Let us assume that $N \neq \Phi(i) \oplus V_2(k+1,t)$. Then $\Phi(i)/S_1 \cong V_2(k+1,t)/S_2$, where $S_1 = N \cap \Phi(i)$, $S_2 = N \cap V_2(k+1,t)$. Let $\dim(\Phi(i)/S_1) = i - r = j$. Within the conjugation we can assume that $S_1 = \Phi(r)$ and $S_2 = 0$ or $S_2 = V_2(k+j+1,s)$ and that is why by means of Lemma 4.1 N is conjugated to one of the spaces,

$\Phi(r) \oplus \Lambda_{r+1,k+1}(j)$;

$\Phi(r) \oplus \Lambda_{r+1,k+1}(j) \oplus V_2(k+j+1,s)$.

The proposition is proved.

V. ON SUBALGEBRAS OF THE NORMALIZER OF ISOTROPIC SPACE

In virtue of Theorem 2.1 the normalizer of the isotropic space $\langle P_0 + P_n \rangle$ in $A\widetilde{P}(1,n)$ coincides with the algebra $K = A\widetilde{G}(n-1) \supset\!\!\!\!+ \rangle J_{0n}, \mathbb{D}\rangle$. In this section we shall establish a number of assertions on subalgebras of the algebra K possessing nonzero projection onto $\langle J_{0n}, \mathbb{D}\rangle$. On the grounds of these results in Theorem 5.1 we describe all Abelian subalgebras of the algebra K that are nonconjugate to the subalgebras of $A\widetilde{G}(n-1)$. As a corollary, we obtain the list of maximal Abelian subalgebras and one-dimensional subalgebras of the algebra K as well as one-dimensional subalgebras of the algebra $A\widetilde{P}(1,n)$.

Further ϵ denotes the projection of K onto $\langle J_{0n}, \mathbb{D}\rangle$ and ξ denotes the projection of K onto $AO(n-1) \oplus \langle J_{0n}, \mathbb{D}\rangle$.

Proposition 5.1: Let $L = \langle G_1, ..., G_k \rangle$ $(1 \leqslant k \leqslant n-1)$, and F be a subdirect sum of L and $\langle \mathbb{D} \rangle$. The algebra F has only splitting extensions in $A\widetilde{P}(1,n)$.

Proof: Let $\hat{F}$ be a subalgebra of $A\widetilde{P}(1,n)$ such that $\pi(\hat{F}) = F$. Up to an $O(n-1)$ automorphism one can assume that $\hat{F}$ contains the generator

$$X_1 = G_1 + \sum_{\nu=0}^{n} \alpha_\nu P_\nu + \gamma \mathbb{D} \quad (\gamma \neq 0).$$

Clearly,

$$\begin{aligned}\exp\left(\sum_{\mu=0}^{n} b_\mu P_\mu\right)\cdot X_1 \cdot \exp\left(-\sum_{\mu=0}^{n} b_\mu P_\mu\right) &= G_1 + \gamma\mathbb{D} + (\alpha_0 - \gamma b_0 + b_1)P_0 \\ &+ (\alpha_1 + b_0 - b_n - \gamma b_1)P_1 + (\alpha_n + b_1 - \gamma b_n)P_n \\ &+ \sum_{i=2}^{n-1} (\alpha_i - \gamma b_i)P_i .\end{aligned}$$

We put

$$\begin{aligned}&\alpha_0 - \gamma b_0 + b_1 = 0, \quad \alpha_1 + b_0 - b_n - \gamma b_1 = 0, \\ &\alpha_n + b_1 - \gamma b_n = 0, \\ &\alpha_i - \gamma b_i = 0 \quad (i = 2,...,n-1).\end{aligned} \tag{5.1}$$

The determinant of coefficients by b_0, b_1, and b_n is equal to $-\gamma^3$. Since $\gamma \neq 0$ then the systems (5.1) has a solution. Therefore one can assume that $X_1 = G_1 + \gamma\mathbb{D}$. Let $a \neq 1$,

$$X_a = G_a + \sum_{\mu=0}^{n} \alpha_\mu P_\mu + \delta \mathbb{D}.$$

Since

$$[X_1, X_a] = -(\alpha_0 - \alpha_n)P_1 - \alpha_1 M + \gamma \sum \alpha_\mu P_\mu,$$

$$[X_1, X_a] - \gamma X_a = -\gamma G_a - \gamma\delta\mathbb{D} - (\alpha_0 - \alpha_n)P_1 - \alpha_1 M,$$

we shall assume that

$$X_a = G_a + \alpha M + \beta P_1 + \delta \mathbb{D}.$$

Then

$$[X_1, X_a] = (\gamma\alpha - \beta)M + \gamma\beta P_1 \quad (2 \leqslant a \leqslant k).$$

If $\gamma\alpha - \beta \neq 0$ then we shall consider that $\alpha = 0, \beta \neq 0$. Since

$$[X_1, [X_1, X_a]] = -2\gamma\beta M + \gamma^2 \beta P_1,$$

then $\hat{F}$ contains $M - \gamma P_1$, $-2M + \gamma P_1$ and whence M, $P_1 \in \hat{F}$. That is why $G_a + \delta\mathbb{D} \in \hat{F}$.

Let $\gamma\alpha - \beta = 0$. If $\beta \neq 0$ then $P_1 \in \hat{F}$. Since $[X_1, P_1] = [G_1 + \gamma\mathbb{D}, P_1] = -M + \gamma P_1$ then $M \in \hat{F}$ and therefore $G_a + \delta\mathbb{D} \in \hat{F}$. If $\beta = 0$ then $\alpha = 0$. It proves that $\hat{F}$ is a splitting algebra. The proposition is proved.

The record F: $W_1, ..., W_s$ means that we deal with the subalgebras $W_1 \supset\!\!\!\!+ F, ..., W_s \supset\!\!\!\!+ F$.

In virtue of Propositions 4.4 and 5.1 we conclude that the subalgebras of the algebra $\mathfrak{M} \supset\!\!\!\!+ \langle \mathbb{D} \rangle$ possessing a nonzero projection onto $\langle \mathbb{D} \rangle$ are exhausted with respect to $\widetilde{P}(1,n)$ conjugation by the following algebras [see notations (4.2)]:

$\langle \mathbb{D} \rangle$: 0, $\Phi(i)$, $V_2(s,t)$

$(i = 0, 1, ..., n-1, \quad s = 0,1, \quad t = s, s+1, ..., n)$;

$\langle G_1 + \alpha_1 \mathbb{D}, ..., G_k + \alpha_k \mathbb{D}, \beta\mathbb{D} \rangle$: 0, $\Phi(i)$, $\Omega(k)$,

$V_2(k+1,t)$, $\Phi(i) \oplus V_2(k+1,t)$, $\Omega(k) \oplus V_2(k+1,t)$,

$\Phi(r) \oplus \Lambda_{r+1,k+1}(j)$,

$\Phi(r) \oplus \Lambda_{r+1,k+1}(j) \oplus V_2(k+j+1,s)$

$(k = 1, ..., n-1, \quad i = 0,1,...,k, \quad t = k+1, ..., n-1,$

$r = 0,1,...,k-1, \quad j = 1,...,k-r,$

$s = k+j+1, ..., n-1)$.

These algebras must then be simplified using transformations contained in the normalizer of each algebra in the group of $O(1,n)$ automorphisms. If, for example, the normalizer contains $\exp(\theta J_{12})$ then instead of $\langle G_1 + \alpha_1 \mathbb{D}, G_2 + \alpha_2 \mathbb{D} \rangle$ we can take $\langle G_1 + \alpha_1 \mathbb{D}, G_2 \rangle$.

Proposition 5.2: Let L be a subalgebra of $AO(n)$, and F be the subdirect sum of L and $\langle \mathbb{D} \rangle$. The algebra F possesses only the splitting extensions in $A\widetilde{P}(1,n)$.

Proposition 5.2 is proved by virtue of Propositions 2.1 and 3.2.

Proposition 5.3: Let L_1 be a subalgebra of $AO(n-1)$, $L_2 = \langle \mathbb{D}, J_{0n} \rangle$ or $L_2 = \langle \mathbb{D} + \gamma J_{0n} \rangle$, where $\gamma = 0$, $\gamma^2 \neq 1$, $2\gamma + 1 \neq 0$. If F is a subdirect sum of the algebras L_1 and L_2 then every subalgebra $\hat{F}$ of the algebra K with the property $\xi(\hat{F}) = F$ is conjugated to the algebra $(W_1 + W_2) \supset\!\!\!\!+ F$, where $W_1 \subset U$, $W_2 \subset V_1 = \langle G_1, ..., G_{n-1} \rangle$.

Proof: Let $L_2 = \langle \mathbb{D}, J_{0n} \rangle$. On the basis of Propositions 2.2 and 5.1 algebra $\hat{F}$ contains the elements

$$X_1 = J_{0n} + \sum_{i=0}^{n} \alpha_i P_i, \quad X_2 = \mathbb{D} + \sum_{j=1}^{n-1} \beta_j G_j .$$

Since $[X_1, X_2] = \Sigma \gamma_i P_i - \Sigma \beta_j G_j$ then $\mathbb{D} + \Sigma \gamma_i P_i \in \hat{F}$. Therefore one can suppose that $\mathbb{D} \in \hat{F}$. Whence $J_{0n} \in \hat{F}$ and $F \subset \hat{F}$.

Let $L_2 = \langle \mathbb{D} + \gamma J_{0n} \rangle$. Since $[\mathbb{D} + \gamma J_{0n}, P_a] = P_a$, $[\mathbb{D} + \gamma J_{0n}, G_a] = -\gamma G_a$ $(a = 1, ..., n-1)$, then by virtue of Proposition 5.2 one can admit that $\hat{F}$ contains the subdirect sum of F and subalgebra of the algebra $\langle P_0, P_n \rangle$. Evidently

$$\exp(\theta_0 P_0 + \theta_n P_n)\cdot(\mathbb{D} + \gamma J_{0n} + \alpha_0 P_0 + \alpha_n P_n)$$
$$\cdot\exp(-\theta_0 P_0 - \theta_n P_n)$$
$$= \mathbb{D} + \gamma J_{0n} + (\alpha_0 - \theta_0 + \gamma\theta_n)P_0$$
$$+ (\alpha_n + \gamma\theta_0 - \theta_n)P_n .$$

Since $\gamma^2 \neq 1$, then coefficients by P_0, P_n can be transformed into zero. On the basis of the conditions $\gamma^2 \neq 1$, $[\mathbb{D} + \gamma J_{0n}, \widehat{F}\cap\mathfrak{M}] \subset \widehat{F}\cap\mathfrak{M}$ it is not difficult to get that $F\subset\widehat{F}$.

Let $W = \widehat{F}\cap\mathfrak{M}$, $Y = \Sigma\,\delta_a G_a + \Sigma\,\rho_i\,P_i \in W$. Since

$$[\mathbb{D} + \gamma J_{0n}, Y] = -\gamma\sum\delta_a G_a$$
$$-\gamma(\rho_0 P_n + \rho_n P_0) + \sum\rho_i\,P_i$$

and $\gamma^2 \neq 1$ then one can assume that $Y = \Sigma\,\delta_a G_a + \rho_0 P_0 + \rho_n P_n$. By the direct calculations we find that

$$[\mathbb{D} + \gamma J_{0n}, Y] = -\gamma\sum\delta_a G_a$$
$$+ (\rho_0 - \gamma\rho_n)P_0 + (\rho_n - \gamma\rho_0)P_n ,$$
$$[\mathbb{D} + \gamma J_{0n}, [\mathbb{D} + \gamma J_{0n}, Y]]$$
$$= \gamma^2\sum\delta_a G_a + (\gamma^2\rho_0 - 2\gamma\rho_n + \rho_0)P_0$$
$$+ (\gamma^2\rho_n - 2\gamma\rho_0 + \rho_n)P_n .$$

The determinant Δ constructed by the coefficients of $\Sigma\,\delta_a G_a$, P_0, P_n in Y and the vectors received is equal to $\gamma(2\gamma + 1)(\rho_n^2 - \rho_0^2)$. If $\Delta \neq 0$ then $\Sigma\,\delta_a G_a$, P_0, $P_n \in W$. If $\Delta = 0$ then $\rho_n = \pm\rho_0$. When $\rho_n = \rho_0$ we get that

$$[\mathbb{D} + \gamma J_{0n}, Y] - (1-\gamma)Y = -\sum\delta_a G_a .$$

If $\rho_n = -\rho_0$ then

$$[\mathbb{D} + \gamma J_{0n}, Y] - (1+\gamma)Y = (-2\gamma - 1)\sum\delta_a G_a .$$

The proposition is proved.

Proposition 5.4: The subalgebras of the algebra $\mathfrak{M}\oplus\langle J_{0n}, \mathbb{D}\rangle$ containing J_{0n} or having the property that their projection F onto $\langle J_{0n}, \mathbb{D}\rangle$ coincides with $\langle\mathbb{D} + \gamma J_{0n}\rangle$, where $\gamma \neq 0$, $\gamma^2 \neq 1$, $2\gamma + 1 \neq 0$, are exhausted with respect to $\widetilde{\mathrm{P}}(1,n)$ conjugation by the following algebras [see notation (4.2)]:

F: 0, $\Phi(a)$, $\Omega(a)$, $V_2(1,d)$
$(a = 0,1,\ldots,n-1,\ d = 1,\ldots,n-1)$;
$(G_1,\ldots,G_k)\oplus F$: 0, $\Phi(i)$, $\Omega(k)$,
$V_2(k+1,t)$, $\Phi(i)\oplus V_2(k+1,t)$,
$\Omega(k)\oplus V_2(k+1,t)$, $\Phi(r)\oplus\Lambda_{r+1,k+1}(j)$,
$\Phi(r)\oplus\Lambda_{r+1,k+1}(j)\oplus V_2(k+j+1,s)$
$(i = 0,1,\ldots,k,\quad t = k+1,\ldots,n-1,$
$r = 0,1,\ldots,k-1,\quad j = 1,\ldots,k-r,$
$s = k+j+1,\ldots,n-1,\ k = 1,\ldots,n-1)$.

The proof of Proposition 5.4 is based on Proposition 5.3.

Proposition 5.5: Let L_1 be a subalgebra of $\mathrm{AO}(n-1)$, $L_2 = \langle 2\mathbb{D} - J_{0n}\rangle$, F a subdirect sum of L_1 and L_2, and $\widehat{F}$ such subalgebra of K that $\xi(\widehat{F}) = F$. The algebra $\widehat{F}$ is conjugated to the algebra $W\oplus F$, where $W\subset\mathfrak{M}$ and satisfies the following condition: if $Y\in W$ and projection of Y onto $V_1 = \langle G_1,\ldots,G_{n-1}\rangle$ is equal to $\Sigma\,\delta_a G_a$ then W contains $\Sigma\,\delta_a G_a + \rho P_0$ and ρM or $\Sigma\,\delta_a G_a + \rho(P_0 - P_n)$.

Proposition 5.6: Let L_1 be a subalgebra of $\mathrm{AO}(n-1)$, $L_2 = \langle\mathbb{D} + J_{0n} + \gamma M\rangle$ $(\gamma\in\{0,1\})$, and F the subdirect sum of L_1 and L_2. If a subspace W of the space $\mathfrak{M}$ is invariant under F then $W = W_1 + W_2$, where $W_1\subset U$, $W_2\subset V_1$.

The proof of Propositions 5.5 and 5.6 is similar to that of Proposition 5.3.

Let $\theta = (\gamma_0 - \gamma_n)/2$. Since

$$\exp(\theta P_0)\cdot(\mathbb{D} + J_{0n} + \gamma_0 P_0 + \gamma_n P_n)\cdot\exp(-\theta P_0)$$
$$= \mathbb{D} + J_{0n} + \tfrac{1}{2}(\gamma_0 + \gamma_n)M ,$$

then further we shall suppose that the projection of the algebra $\widehat{F}\subset\mathrm{A}\widetilde{\mathrm{P}}(1,n)$ onto $\langle\mathbb{D} + J_{0n}, P_0, P_n\rangle$ contains $\mathbb{D} + J_{0n} + \alpha M$ where $\alpha\in\{0,1\}$. Proposition 5.6 gives the considerable information on the structure of such algebras.

Proposition 5.7: Let L_1 be a subalgebra of $\mathrm{AO}(n-1)$, $L_2 = \langle\mathbb{D} - J_{0n} + \gamma P_0\rangle$ $(\gamma\in\{0,1\})$, and F the subdirect sum of the algebras L_1 and L_2. If a subspace W of the space $\mathfrak{M}$ is invariant under F, then W contains its own projection onto $\langle P_0, P_n\rangle$ and $[L_1, W]\subset W$, $[\gamma P_0, W]\subset W$.

Proof: On the basis of Proposition 4.3 $[L_i, W]\subset W$ $(i = 1,2)$. Let $\widetilde{\mathfrak{M}} = \{Y\in\mathfrak{M}\,|\,[L_1, Y] = 0\}$, and $\widetilde{W}$ be a projection of W onto $\widetilde{\mathfrak{M}}$. It is easy to see that the matrix diag$[2,0]$ is the matrix of the operator $\mathbb{D} - J_{0n}$ in the basis $P_0 + P_n$, $P_0 - P_n$ of the space $\langle P_0, P_n\rangle$ and in the basis of the space $\mathfrak{M}\,|\,\langle P_0, P_n\rangle$ the matrix of the same operator is the unit one. Whence by Lemma 3.1 we conclude that $\widetilde{W}$ contains its own projection onto $\langle P_0, P_n\rangle$. It remains for us to note that for arbitrary

$$Y = \sum_{j=1}^{n-1}(\alpha_j P_j + \beta_j G_j)$$

we have $[\mathbb{D} - J_{0n} + \gamma P_0, Y] = Y + [\gamma P_0, Y]$. The proposition is proved.

Proposition 5.8: Let F be a subalgebra of the algebra $\mathrm{AO}(1,n)$ generated by J_{0n} and G_a, where a runs through some subset I of the set $\{1,2,\ldots,n-1\}$. If $\widehat{F}$ is a subalgebra of $\mathrm{AP}(1,n)$ with $\pi(\widehat{F}) = F$, then within the conjugation with respect to the group of translations the algebra $\widehat{F}$ contains elements G_a $(a\in I)$ and $J_{0n} + \Sigma\,\delta_i P_i$ $(i = 1,\ldots,n-1)$.

Proposition 5.9: Let L be a subalgebra of the algebra $\mathrm{AP}(1,n)$, $X = J_{ab} + \delta J_{0n} + \beta P_c$, $Y = G_c + \Sigma\,\gamma_i P_i$ $(i = 1,\ldots,n)$, where $\beta\neq 0$, $\delta\neq 0$, and a, b, and c are different numbers of $\{1,2,\ldots,n-1\}$. If X, $Y\in L$ then L contains G_c.

Theorem 5.1: Let L be an Abelian subalgebra of the algebra K and $\epsilon(L)\neq 0$. If $\epsilon(L) = \langle J_{0n}\rangle$ then L is $\widetilde{\mathrm{P}}(1,n)$ conjugated to the subdirect sum of algebras L_1, L_2, $\langle J_{0n}\rangle$, where $L_1\subset\mathrm{AH}(2d)$, $L_2 = 0$, or $L_2 = \langle P_{2d+1},\ldots,P_{2d+s}\rangle$. If $\epsilon(L) = \langle\mathbb{D}\rangle$ then L is $\widetilde{\mathrm{P}}(1,n)$ conjugated to the subdirect sum of L_1, L_2, $\langle\mathbb{D}\rangle$, where $L_1\subset\mathrm{AH}(2d)$, $L_2 = 0$ or $L_2 = \langle G_{2d+1},\ldots,G_{2d+s}\rangle$. If $\epsilon(L) = \langle\mathbb{D}, J_{0n}\rangle$ or $\epsilon(L) = \langle\mathbb{D} + \gamma J_{0n}\rangle$, where $\gamma\neq 0$, $\gamma^2\neq 1$ then L is $\widetilde{\mathrm{P}}(1,n)$ conjugated to the subdirect sum of algebras $\epsilon(L)$ and $L_1\subset\mathrm{AH}(2d)$. If $\epsilon(L) = \langle\mathbb{D} + J_{0n}\rangle$, then L is conjugated to the subdirect sum of the algebras L_1, L_2, L_3, where $L_1 \subset \mathrm{AH}(2d)$, $L_2 \subset \langle M \rangle$, $L_3 = \langle J_{0n} + \mathbb{D} \rangle$.

Proof: If $\epsilon(L) = \langle J_{0n} \rangle$ then in view of Propositions 2.2 and 4.3 the algebra L contains its own projection onto $\langle M, P_0 - P_n, G_1, \ldots, G_{n-1} \rangle$. Since $[J_{0n}, G_a] = -G_a$, $[J_{0n}, M] = -M$, $[J_{0n}, P_0 - P_n] = P_0 - P_n$ then this projection is equal to zero. Therefore L is the subdirect sum of $L_1 \subset \mathrm{AH}(2d)$ and $L_2 \subset \langle P_{2d+1}, \ldots, P_{n-1} \rangle$. If $L_2 \neq 0$ then by Witt's theorem L_2 is conjugated to $\langle P_{2d+1}, \ldots, P_{2d+s} \rangle$.

If $\varepsilon(L) = \langle \mathbb{D} \rangle$ then in virtue of Propositions 4.3 and 5.2 the projection of L onto U is equal to 0.

If $\epsilon(L) = \langle \mathbb{D}, J_{0n} \rangle$ or $\epsilon(L) = \langle \mathbb{D} + \gamma J_{0n} \rangle$, where $\gamma \neq 0$, $\gamma^2 \neq 1$, $2\gamma + 1 \neq 0$, then by Proposition 5.3 the algebra L is conjugated to the subdirect sum of the algebras $\epsilon(L)$ and $L_1 \subset \mathrm{AH}(2d)$. With $\epsilon(L) = \langle 2\mathbb{D} - J_{0n} \rangle$ Proposition 5.5 is applicable.

Let $\epsilon(L) = \langle \mathbb{D} - J_{0n} \rangle$. On the basis of Propositions 2.2 and 4.3 the projection of L onto $\langle G_1, \ldots, G_{n-1} \rangle$ is equal to 0. Applying the $\mathrm{O}(1,n)$ automorphism corresponding to the matrix $\mathrm{diag}[1, \ldots, 1, -1]$ we get that $\epsilon(L) = \langle \mathbb{D} + J_{0n} \rangle$. According to Proposition 5.2 the projection of L onto $\langle P_1, \ldots, P_{n-1} \rangle$ is equal to 0. Since $[J_{0n} + \mathbb{D}, P_0 + P_n] = 0$, $[J_{0n} + \mathbb{D}, P_0 - P_n] = 2(P_0 - P_n)$ then by Propositions 2.1 and 4.3 the projection of L onto $\langle P_0, P_n \rangle$ belongs to $\langle P_0 + P_n \rangle$. The theorem is proved.

Corollary 1: The maximal Abelian subalgebras of the algebra K with the condition $\epsilon(K) \neq 0$ are exhausted with respect to $\widetilde{\mathrm{P}}(1,n)$ conjugation by the following algebras:

$$\mathrm{AH}(n-1) \oplus \langle J_{0n}, \mathbb{D} \rangle, \quad \mathrm{AH}(n-1) \oplus \langle M, J_{0n}, \mathbb{D} \rangle,$$
$$\mathrm{AH}(2d) \oplus \langle P_{2d+1}, \ldots, P_{n-1}, J_{0n} \rangle,$$
$$\mathrm{AH}(2d) \oplus \langle G_{2d+1}, \ldots, G_{n-1}, \mathbb{D} \rangle \quad (d = 0, 1, \ldots, [(n-2)/2]).$$

The written algebras are not conjugated mutually.

Corollary 2: Let $n \geqslant 3$, $X_t = \alpha_1 J_{12} + \alpha_2 J_{34} + \cdots + \alpha_t J_{2t-1,2t}$; $\alpha_1 = 1$, $0 < \alpha_2 \leqslant \cdots \leqslant \alpha_t \leqslant 1$; $t = 1, \ldots, [(n-1)/2]$; $s = 1, \ldots, [(n-2)/2]$; $\alpha > 0$. The one-dimensional subalgebras of the algebra K with the condition $\epsilon(K) \neq 0$ are exhausted with respect to $\widetilde{\mathrm{P}}(1,n)$ conjugation by the following algebras: $\langle J_{0n} \rangle$; $\langle \mathbb{D} \rangle$; $\langle \mathbb{D} + \alpha J_{0n} \rangle$; $\langle J_{0n} + P_1 \rangle$; $\langle \mathbb{D} + G_1 \rangle$;

$$\langle \mathbb{D} + J_{0n} + M \rangle; \ \langle X_t + \alpha \mathbb{D} + \beta J_{0n} \rangle \ (\beta \geqslant 0);$$
$$\langle X_t + \alpha J_{0n} \rangle; \ \langle X_t + \alpha(\mathbb{D} + J_{0n} + M) \rangle;$$
$$\langle X_s + G_{2s+1} + \alpha \mathbb{D} \rangle; \ \langle X_s + P_{2s+1} + \alpha J_{0n} \rangle.$$

The written algebras are not conjugated mutually.

Proposition 5.10: The one-dimensional subalgebras of the algebra $\mathrm{A}\widetilde{\mathrm{P}}(1,n)$ are exhausted wtih respect to the $\widetilde{\mathrm{P}}(1,n)$ conjugation by the one-dimensional subalgebras of the algebra K and the following algebras:

$$\langle J_{12} + \beta_1 J_{34} + \cdots + \beta_{n/2-1} J_{n-1,n} + \gamma \mathbb{D} \rangle,$$
$$\langle J_{12} + \beta_1 J_{34} + \cdots + \beta_{n/2-1} J_{n-1,n} + P_0 \rangle,$$

where $n \equiv 0 (\mathrm{mod}\ 2)$, $\gamma \geqslant 0$, $0 < \beta_1 \leqslant \cdots \leqslant \beta_{n/2-1} \leqslant 1$.

VI. SUBALGEBRAS OF THE ALGEBRA $\mathrm{A}\widetilde{\mathrm{P}}(1,4)$

In this section we make use of the previous results to provide a classification of all subalgebras of $\mathrm{A}\widetilde{\mathrm{P}}(1,4)$ with respect to $\widetilde{\mathrm{P}}(1,4)$ conjugation.

Let $\hat{F}$ be an subalgebra of $\mathrm{A}\widetilde{\mathrm{P}}(1,4)$ such that $\pi(\hat{F}) = F$. An expression $\hat{F} + W$ means that W is a subspace of U, $[F, W] \subset W$, and $\hat{F} \cap U \subset W$. As concerns the algebras $\hat{F} + W_1, \ldots, \hat{F} + W_s$ we will use the notation $\hat{F}$: $W_1, \ldots, W_s$.

Lemma 6.1: Let $\alpha, \beta, \gamma \in R$, $\alpha > 0$, $\beta \geqslant 0$, $\gamma \neq 0$, and F run through the full system of representatives of the classes of $\mathrm{O}(1,4)$-conjugated subalgebras of the algebra $\mathrm{AO}(1,4)$.[4] The subalgebras of the algebra $\mathrm{AO}(1,4) \oplus \langle \mathbb{D} \rangle$ are exhausted with respect to $\widetilde{\mathrm{O}}(1,4)$ conjugation by the algebras F, $F \oplus \langle \mathbb{D} \rangle$ and the following algebras:

$\langle J_{12} + \alpha \mathbb{D} \rangle$; $\langle J_{12} + c J_{34} + \alpha \mathbb{D} \rangle$ $(0 < c \leqslant 1)$; $\langle J_{04} + \alpha \mathbb{D} \rangle$; $\langle J_{12} + c J_{04} + \alpha \mathbb{D} \rangle$ $(c > 0)$; $\langle G_3 + \mathbb{D} \rangle$; $\langle G_3 - J_{12} + \alpha \mathbb{D} \rangle$; $\langle J_{12} + \alpha \mathbb{D}, J_{34} + \beta \mathbb{D} \rangle$; $\langle J_{04} + \alpha \mathbb{D}, J_{12} + \beta \mathbb{D} \rangle$; $\langle J_{04}, J_{12} + \alpha \mathbb{D} \rangle$; $\langle G_3 + \mathbb{D}, J_{12} + \beta \mathbb{D} \rangle$; $\langle G_3, J_{12} + \alpha \mathbb{D} \rangle$; $\langle G_1 + \mathbb{D}, G_2 \rangle$; $\langle G_3, J_{04} + \gamma \mathbb{D} \rangle$; $\langle G_3, J_{12} + c J_{04} + \gamma \mathbb{D} \rangle$ $(c > 0)$; $\langle G_3, J_{04} + \gamma \mathbb{D}, J_{12} + \beta \mathbb{D} \rangle$; $\langle G_3, J_{04}, J_{12} + \alpha \mathbb{D} \rangle$; $\langle G_1, G_2, J_{12} + \alpha \mathbb{D} \rangle$; $\langle G_1, G_2, J_{04} + \gamma \mathbb{D} \rangle$; $\langle G_1, G_2, J_{12} + c J_{04} + \gamma \mathbb{D} \rangle$ $(c > 0)$; $\langle G_1 + \mathbb{D}, G_2, G_3 \rangle$; $\langle G_1, G_2, G_3 - J_{12} + \alpha \mathbb{D} \rangle$; $\langle J_{03}, J_{04}, J_{34}, J_{12} + \alpha \mathbb{D} \rangle$; $\langle J_{12} + J_{34}, J_{13} - J_{24}, J_{23} + J_{14}, J_{34} + \gamma \mathbb{D} \rangle$; $\langle G_1, G_2, J_{12} + \alpha \mathbb{D}, J_{04} + \delta \mathbb{D} \rangle$; $\langle G_1, G_2, J_{12}, J_{04} + \gamma \mathbb{D} \rangle$; $\langle G_1, G_2, G_3 + \mathbb{D}, J_{12} + \beta \mathbb{D} \rangle$; $\langle G_1, G_2, G_3, J_{12} + \alpha \mathbb{D} \rangle$; $\langle G_1, G_2, G_3, J_{04} + \gamma \mathbb{D} \rangle$; $\langle G_1, G_2, G_3, J_{12} + c J_{04} + \gamma \mathbb{D} \rangle$ $(c > 0)$; $\langle J_{12}, J_{13}, J_{23}, J_{04} + \alpha \mathbb{D} \rangle$; $\langle G_1, G_2, G_3, J_{12} + \alpha \mathbb{D}, J_{04} + \delta \mathbb{D} \rangle$; $\langle G_1, G_2, G_3, J_{12}, J_{13}, J_{23}, J_{04} + \gamma \mathbb{D} \rangle$.

Lemma 6.1 is proved with the Goursat method[25] and the result on the classification of subalgebras of the algebra $\mathrm{AO}(1,4)$.[4]

Theorem 6.1: Let $\Delta(\Gamma)$ be the system of representatives of the classes of conjugated subalgebras of the algebra $\mathrm{A}\widetilde{\mathrm{O}}(1,4)$ [respectively, $\mathrm{AO}(1,4)$] found in Lemma 6.1. The splitting subalgebras of the algebra $\mathrm{A}\widetilde{\mathrm{P}}(1,4)$ are exhausted with respect to $\widetilde{\mathrm{P}}(1,4)$ conjugation by the following algebras:

(1) $W \oplus F$, where $F \in \Gamma$, $W \subset U$, and $[F, W] \subset W$;
(2) $W \oplus \hat{F}$, where $\hat{F} \in \Delta$ and the projection of $\hat{F}$ onto $\mathrm{AO}(1,4)$ coincides with F, $F \in \Gamma$;
(3) $\langle J_{12}, J_{34} + \alpha \mathbb{D} \rangle$: $\langle P_1, P_2 \rangle$, $\langle P_0, P_1, P_2 \rangle$ $(\alpha > 0)$;
(4) $\langle G_1 + \alpha \mathbb{D}, G_2 + \beta \mathbb{D} \rangle$: $\langle M, P_1 \rangle$, $\langle M, P_1 + \omega P_3 \rangle$, $\langle M, P_1, P_3 \rangle$, $\langle M, P_1 + \omega P_3, P_2 \rangle$ $(\omega > 0, \ \alpha \geqslant 0, \ \beta \geqslant 0, \ \alpha^2 + \beta^2 \neq 0)$;
(5) $\langle G_1 + \alpha \mathbb{D}, G_2 + \beta \mathbb{D}, G_3, M, P_1 \rangle$ $(\alpha \geqslant 0, \ \beta \geqslant 0, \ \alpha^2 + \beta^2 \neq 0)$;
(6) $\langle G_1 + \alpha \mathbb{D}, G_2, G_3 + \beta \mathbb{D}, M, P_1, P_2 \rangle$ $(\alpha \geqslant 0, \ \beta \geqslant 0, \ \alpha^2 + \beta^2 \neq 0)$.

Proof: Let $\hat{F}$ be the subdirect sum of $F \in \Gamma$ and $\langle \mathbb{D} \rangle$, and W a subspace of U invariant under $\hat{F}$. Then $[F, W] \subset W$ and on the contrary, if $[F, W] \subset W$ then $[\hat{F}, W] \subset W$. Therefore we can use the results on the classification of the splitting subalgebras of AP(1,4).[9] Only the cases of the algebras $\hat{F} \in \Delta$ simplified by O(1,4) automorphisms demand an additional consideration. Such algebras correspond to the algebra F coinciding with $\langle J_{12}, J_{34} \rangle$, $\langle G_1, G_2 \rangle$, or $\langle G_1, G_2, G_3 \rangle$. If, for example,

$$\hat{F} = \langle G_1 + \alpha_1 \mathbb{D}, G_2 + \alpha_2 \mathbb{D}, G_3 + \alpha_3 \mathbb{D} \rangle$$

then this algebra must be simplified using transformations contained in the normalizer of $\langle M, P_1 \rangle$, $\langle M, P_1, P_2 \rangle$, respectively, in the group of O(1,4) automorphisms. The theorem is proved.

We conceive the classification of nonsplitting subalgebras of AP(1,4) with respect to $\tilde{\mathrm{P}}(1,4)$ conjugation by virtue of the known classification of the nonsplitting subalgebras of AP(1,4) with respect to P(1,4) conjugation.[11] The application of the automorphism $\exp(\theta \mathbb{D})$ allows us to substitute one of the continuous parameters by the translation generators onto 1.

Let $(i_1, ..., i_q) = \langle P_{i_1}, ..., P_{i_q} \rangle$; $(awb) = \langle P_a + wP_b \rangle$ $(\omega > 0)$; $(04) = \langle M \rangle$.

Theorem 6.2: The nonsplitting subalgebras of the algebra $\mathrm{A}\tilde{\mathrm{P}}(1,4)$ are exhausted with respect to $\tilde{\mathrm{P}}(1,4)$ conjugation by the nonsplitting subalgebras of the algebra AP(1,4) and the following algebras:

$\langle J_{04} - \mathbb{D} + P_0 \rangle$: 0, (1), (04), (1,2), (04,1), (1,2,3), (04,1,2), (04,1,2,3);

$\langle J_{12} + c(J_{04} - \mathbb{D} + P_0) \rangle$: 0, (04), (3), (04,3), (1,2), (1,2,3), (04,1,2), (04,1,2,3) $(c > 0)$;

$\langle J_{04} + \mathbb{D} + M, J_{12} + \alpha M \rangle$: 0, (3), (1,2), (1,2,3) $(\alpha > 0)$; $\langle J_{04} + \mathbb{D}, J_{12} + M \rangle$: 0, (3), (1,2), (1,2,3);

$\langle J_{04} + \mathbb{D} + M, J_{12} \rangle$: 0, (3), (1,2), (1,2,3); $\langle J_{04} - \mathbb{D} + P_0, J_{12} + \alpha P_0 \rangle$: (04), (04,3), (04,1,2), (04,1,2,3) $(\alpha \geqslant 0)$;

$\langle J_{04} - \mathbb{D}, J_{12} + P_0 \rangle$: (04), (04,3), (04,1,2), (04,1,2,3);

$\langle J_{04} - 2\mathbb{D}, G_3 + P_0 \rangle$: (04), (04,1), (04,1$\omega$3), (04,3), (04,1$\omega$3,2), (04,1,2), (04,1,3), (04,1,2,3);

$\langle J_{04} - 2\mathbb{D}, G_3 + P_0 - P_4 \rangle$: 0, (1), (1,2); $\langle J_{04} - \mathbb{D}, G_3 + P_1 \rangle$: 0, (04), (04,3), (0,3,4);

$\langle J_{04} - \mathbb{D}, G_3 + P_2 \rangle$: (1), (04,1), (04,1$\omega$3), (04,1,3), (0,1,3,4); $\langle G_3 + \alpha P_1, J_{04} - \mathbb{D} + P_0, M, P_3 \rangle$ $(\alpha > 0)$;

$\langle J_{04} - \mathbb{D} + P_0, G_3 + \alpha P_2, M, P_1, P_3 \rangle$ $(\alpha > 0)$; $\langle G_3, J_{04} - \mathbb{D} + P_0 \rangle$: (04,3), (04,1,3), (04,1,2,3);

$\langle G_3, J_{04} + \mathbb{D} + M \rangle$: 0, (1), (1,2); $\langle G_3 + P_0, J_{12} + c(J_{04} - 2\mathbb{D}) \rangle$: (04), (04,3), (04,1,2), (04,1,2,3) $(c > 0)$;

$\langle G_3 + P_0 - P_4, J_{12} + c(J_{04} - 2\mathbb{D}) \rangle$: 0, (1,2) $(c > 0)$; $\langle G_3, J_{12} + c(J_{04} - \mathbb{D} + P_0) \rangle$: (04,3), (04,1,2,3);

$\langle G_3, J_{12} + c(J_{04} + \mathbb{D} + M) \rangle$: 0, (1,2); $\langle G_3 + P_0, J_{12}, J_{04} - 2\mathbb{D} \rangle$: (04), (04,3), (04,1,2), (04,1,2,3);

$\langle G_3 + P_0 - P_4, J_{12}, J_{04} - 2\mathbb{D} \rangle$: 0, (1,2); $\langle G_3, J_{12} + \alpha P_0, J_{04} - \mathbb{D} + P_0 \rangle$: (04,3), (04,1,2,3) $(\alpha \geqslant 0)$;

$\langle G_3, J_{12} + P_0, J_{04} - \mathbb{D} \rangle$: (04,3), (04,1,2,3); $\langle G_3, J_{12} + \alpha M, J_{04} + \mathbb{D} + M \rangle$: 0,(1,2) $(\alpha \geqslant 0)$;

$\langle G_3, J_{12} + M, J_{04} + \mathbb{D} \rangle$: 0,(1,2); $\langle G_1, G_2 + P_0, J_{04} - 2\mathbb{D} \rangle$: (04,1), (04,1,2), (04,1,2ω3), (04,1,3), (04,1,2,3);

$\langle G_1 + P_3, G_2 + \mu P_2 + \delta P_3, J_{04} - \mathbb{D} \rangle$ $(\mu > 0, \ \delta \geqslant 0)$; $\langle G_1 + P_3, G_2, J_{04} - \mathbb{D} \rangle$;

$\langle G_1, G_2 + P_2 + \delta P_3, J_{04} - \mathbb{D} \rangle$ $(\delta \geqslant 0)$; $\langle G_1, G_2 + P_2, J_{04} - \mathbb{D}, P_3 \rangle$;

$\langle G_1 + P_2 + \lambda P_3, G_2 - P_1 + \mu P_2 + \delta P_3, J_{04} - \mathbb{D}, M \rangle$ $(\mu > 0, \ \lambda > 0 \vee \lambda = 0, \ \delta \geqslant 0)$;

$\langle G_1 + P_2 + \lambda P_3, G_2 - P_1, J_{04} - \mathbb{D}, M \rangle$ $(\lambda \geqslant 0)$; $\langle G_1 + P_3, G_2, J_{04} - \mathbb{D}, M \rangle$;

$\langle G_1 + \lambda P_3, G_2 + P_2 + \delta P_3, J_{04} - \mathbb{D}, M \rangle$ $(\lambda > 0 \vee \lambda = 0; \ \delta \geqslant 0)$;

$\langle G_1 + P_2, G_2 - P_1 + \mu P_2, J_{04} - \mathbb{D}, M, P_3 \rangle$ $(\mu \geqslant 0)$; $\langle G_1, G_2 + P_2, J_{04} - D, M, P_3 \rangle$;

$\langle G_1 + \alpha P_2 + \beta P_3, G_2 + P_3, J_{04} - \mathbb{D}, M, P_1 \rangle$ $(\alpha > 0 \vee \alpha = 0, \ \beta \geqslant 0)$; $\langle G_1 + P_2 + \beta P_3, G_2, J_{04} - \mathbb{D}, M, P_1 \rangle$ $(\beta \geqslant 0)$;

$\langle G_1 + P_3, G_2, J_{04} - \mathbb{D}, M, P_1 \rangle$; $\langle G_1 + \alpha P_2 + \beta P_3, G_2 + P_3, J_{04} - \mathbb{D}, M, P_1 + \omega P_3 \rangle$ $(\omega > 0)$;

$\langle G_1 + P_2 + \beta P_3, G_2, J_{04} - \mathbb{D}, M, P_1 + \omega P_3 \rangle$ $(w > 0)$; $\langle G_1 + P_3, G_2, J_{04} - \mathbb{D}, M, P_1 + \omega P_3 \rangle$ $(\omega > 0)$;

$\langle G_1 + P_3, G_2, J_{04} - \mathbb{D}, M, P_1, P_2 \rangle$; $\langle G_1 + P_2, G_2, J_{04} - \mathbb{D}, M, P_1, P_3 \rangle$; $\langle G_1, G_2 + P_3, J_{04} - \mathbb{D}, M, P_1 + \omega P_3, P_2 \rangle$ $(\omega > 0)$;

$\langle G_1 + P_3, G_2, J_{04} - \mathbb{D}, P_0, P_1, P_2, P_4 \rangle$; $\langle G_1 + \beta P_3, G_2, J_{04} - \mathbb{D} + P_0, M, P_1, P_2 \rangle$ $(\beta \geqslant 0)$;

$\langle G_1, G_2, J_{04} - \mathbb{D} + P_0, M, P_1, P_2, P_3 \rangle$; $\langle G_1, G_2, J_{04} + \mathbb{D} + M \rangle$; $\langle G_1, G_2, J_{04} + \mathbb{D} + M, P_3 \rangle$;

$\langle G_1 + P_2, G_2 - P_1, J_{12} + c(J_{04} - \mathbb{D}) \rangle$: (04), (04,3) $(c > 0)$;

$\langle G_1, G_2, J_{12} + c(J_{04} - \mathbb{D} + P_0), M, P_1, P_2, sP_3 \rangle$ $(c > 0, \ s = 0,1)$;

$\langle G_1, G_2, J_{12} + c(J_{04} + \mathbb{D} + M) \rangle$: 0, (3) $(c > 0)$; $\langle G_1, G_2, J_{12} + P_0, J_{04} - \mathbb{D}, M, P_1, P_2, sP_3 \rangle$ $(s = 0,1)$;

$\langle G_1, G_2, J_{12} + M, J_{04} + \mathbb{D} \rangle$: 0, (3); $\langle G_1, G_2, J_{12} + \delta P_0, J_{04} - \mathbb{D} + P_0, M, P_1, P_2, sP_3 \rangle$ $(\delta \geqslant 0, \ s = 0,1)$;

$\langle G_1 + P_2, G_2 - P_1, J_1, J_{04} - \mathbb{D}, M, sP_3 \rangle$ $(s = 0,1)$; $\langle G_1, G_2, J_{12} + \alpha M, J_{04} + \mathbb{D} + M \rangle$: 0, (3) $(\alpha \geqslant 0)$;

$\langle G_1, G_2, G_3 + P_0, J_{04} - 2\mathbb{D}, M, P_1, P_2, sP_3 \rangle$ $(s = 0,1)$;

$\langle G_1, G_2 + P_2, G_3 + \alpha P_3, J_{04} - \mathbb{D} \rangle$; $\langle G_1, G_2 + P_2, G_3 + \alpha P_3, J_{04} - \mathbb{D}, M \rangle$;

$$\langle G_1+P_2+\beta P_3, G_2-P_1+\mu P_2+\gamma P_3, G_3+\beta P_1+\gamma P_2+\delta P_3, J_{04}-\mathbb{D}, M\rangle \quad (\mu>0,\quad \beta>0\vee\beta=0,\quad \gamma\geqslant 0);$$
$$\langle G_1+P_2+\beta P_3, G_2-P_1, G_3+\beta P_1+\delta P_3, J_{04}-\mathbb{D}, M\rangle \quad (\beta\geqslant 0);$$
$$\langle G_1+\beta P_2, G_2+P_3, G_3-P_2, J_{04}-\mathbb{D}, M, P_1\rangle \quad (\beta\geqslant 0);$$
$$\langle G_1+\beta P_2+\gamma P_3, G_2+P_3, G_3-P_2+\mu P_3, J_{04}-\mathbb{D}, M, P_1\rangle \quad (\mu>0,\quad \beta>0\vee\beta=0,\quad \gamma\geqslant 0);$$
$$\langle G_1+\beta P_2+\gamma P_3, G_2, G_3+P_3, J_{04}-\mathbb{D}, M, P_1\rangle \quad (\beta>0\vee\beta=0,\quad \gamma\geqslant 0);$$
$$\langle G_1+P_2, G_2, G_3, J_{04}-\mathbb{D}, M, P_1\rangle;\quad \langle G_1+P_3, G_2, G_3, J_{04}-\mathbb{D}, M, P_1, P_2\rangle;$$
$$\langle G_1, G_2, G_3, J_{04}-\mathbb{D}+P_0, M, P_1, P_2, P_3\rangle;\quad \langle G_1, G_2, G_3, J_{04}+\mathbb{D}+M\rangle;$$
$$\langle G_1, G_2, G_3+P_0, J_{12}+c(J_{04}-2\mathbb{D}), M, P_1, P_2, sP_3\rangle \quad (c>0,\quad s=0,1);$$
$$\langle G_1+P_2, G_2-P_1, G_3+\beta P_3, J_{12}+c(J_{04}-\mathbb{D}), M\rangle \quad (c>0);$$
$$\langle G_1+P_2, G_2-P_1, G_3, J_{12}+c(J_{04}-\mathbb{D}), M, P_3\rangle \quad (c>0);$$
$$\langle G_1, G_2, G_3+P_3, J_{12}+c(J_{04}-\mathbb{D})\rangle\text{: } 0,\ (04);\quad \langle G_1, G_2, G_3, J_{12}+c(J_{04}-\mathbb{D}+P_0), M, P_1, P_2, P_3\rangle \quad (c>0);$$
$$\langle G_1, G_2, G_3, J_{12}+c(J_{04}+\mathbb{D}+M)\rangle \quad (c>0);\quad \langle J_{12}, J_{13}, J_{23}, J_{04}-\mathbb{D}+P_0\rangle\text{: } 0,\ (04),\ (1,2,3),\ (04,1,2,3);$$
$$\langle G_1, G_2, G_3+P_0, J_{12}, J_{04}-2\mathbb{D}, M, P_1, P_2, sP_3\rangle \quad (s=0,1);\quad \langle G_1, G_2, G_3, J_{12}+P_0, J_{04}-\mathbb{D}, M, P_1, P_2, P_3\rangle;$$
$$\langle G_1, G_2, G_3, J_{12}+\delta P_0, J_{04}-\mathbb{D}+P_0, M, P_1, P_2, P_3\rangle \quad (\delta\geqslant 0);\quad \langle G_1+P_2, G_2-P_1, G_3+\beta P_3, J_{12}, J_{04}-\mathbb{D}, M\rangle;$$
$$\langle G_1+P_2, G_2-P_1, G_3, J_{12}, J_{04}-\mathbb{D}, M, P_3\rangle;\quad \langle G_1, G_2, G_3+P_3, J_{12}, J_{04}-\mathbb{D}\rangle\text{: } 0,\ (04);$$
$$\langle G_1, G_2, G_3, J_{12}+M, J_{04}+\mathbb{D}\rangle;\quad \langle G_1, G_2, G_3, J_{12}+\delta M, J_{04}+\delta+M\rangle \quad (\delta\geqslant 0);$$
$$\langle G_1, G_2, G_3, J_{12}, J_{13}, J_{23}, J_{04}-\mathbb{D}+P_0, M, P_1, P_2, P_3\rangle;\quad \langle G_1, G_2, G_3, J_{12}, J_{13}, J_{23}, J_{04}+\mathbb{D}+M\rangle.$$

VII. CONCLUSIONS

The results of the present paper may be summarized in the following way.

(1) The maximal Abelian subalgebras of the algebra $A\widetilde{P}(1,n)$ have been explicitly found in Corollary 1 to Theorem 4.2 and Corollary 1 to Theorem 5.1.

(2) The full classification of one-dimensional subalgebras of algebra $A\widetilde{P}(1,n)$ is contained in Corollary 2 to Theorem 4.2, Corollary 2 to Theorem 5.1 and Proposition 5.10.

(3) The completely reducible subalgebras of $A\widetilde{O}(1,n)$ which possess only splitting extensions in the algebra $A\widetilde{P}(1,n)$ have been picked out. We have established in Theorem 3.1 that the description of the splitting subalgebras $\widehat{F}$ of $A\widetilde{P}(1,n)$, whose projection F onto $A\widetilde{O}(1,n)$ does not have any invariant isotropic subspaces in the space of translations or annul such subspaces, could be reduced to the description of the irreducible parts of the algebra F.

(4) A number of assertions on the subalgebras of the algebra $U \dotplus K'$ has been proved where K' is the normalizer of $\langle P_0+P_n\rangle$ in $A\widetilde{O}(1,n)$. These assertions concern the following matters: The splittability of all extensions of the subalgebra $L\subset K'$ in $A\widetilde{P}(1,n)$ or in some other algebras (Propositions 4.1, 4.2, 5.1, and 5.2); the decomposition of invariant subspaces into a direct sum of its projections onto certain subspaces (Propositions 5.3, 5.5, 5.6, 5.7, and 5.8); the explicit description of some classes of the conjugated subalgebras of the algebra $A\widetilde{P}(1,n)$ (Theorem 4.1, Propositions 4.4 and 5.4).

(5) The full classification with respect to $\widetilde{P}(1,4)$ conjugation of the nonsplitting subalgebras of $A\widetilde{P}(1,4)$ which are nonconjugate to the subalgebras of AP(1,4) has been carried out.

Note added in proof: In Refs. 26–28 the subalgebras of the algebra AP(1,n) were used to construct the exact solutions of many-dimensional nonlinear d'Alembert and Dirac equations. The invariants of subgroups of the generalized Poincaré group P(1,n) were constructed in Ref. 29. A number of general results on continuous subgroups of pseudo-orthogonal pseudounitary groups had been obtained.[30]

ACKNOWLEDGMENT

We are grateful to the referee for his valuable remarks.

[1] J. Patera, P. Winternitz, and H. Zassenhaus, J. Math. Phys. **16**, 1597 (1975).
[2] J. Patera, P. Winternitz, R. T. Sharp, and H. Zassenhaus, Can. J. Phys. **54**, 950 (1976).
[3] J. Patera, P. Winternitz, and H. Zassenhaus, J. Math. Phys. **16**, 1615 (1975).
[4] J. Patera, P. Winternitz, and H. Zassenhaus, J. Math. Phys. **17**, 717 (1976).
[5] J. Patera, R. T. Sharp, P. Winternitz, and H. Zassenhaus, J. Math. Phys. **18**, 2259 (1977).
[6] G. Burdet, P. Patera, M. Perrin, and P. Winternitz, J. Math. Phys. **19**, 1758 (1978).
[7] J. Beckers, J. Patera, M. Perroud, and P. Winternitz, J. Math. Phys. **18**, 72 (1977).
[8] G. Burdet, J. Patera, M. Perrin, and P. Winternitz, Ann. Sci. Math. Québec **2**, 81 (1978).
[9] V. M. Fedorchuk, Ukrainian Math. J. **31**, 717 (1979).
[10] V. M. Fedorchuk, Ukrainian Math. J. **33**, 696 (1981).
[11] W. I. Fushchich, A. F. Barannik, L. F. Barannik, and V. M. Fedorchuk, J. Phys. A: Math. Gen. **18**, 2893 (1985).
[12] W. I. Fushchich, A. F. Barannik, and L. F. Barannik, "The continuous subgroups of the generalized Galilei group. I," Preprint, Mathematical Institute, Kiev 85.19, 1985.
[13] W. I. Fushchich, A. F. Barannik, and L. F. Barannik, Ukrainian Math. J. **38**, 67 (1986).
[14] W. Lassner, Acta Phys. Slovaca **23**, 193 (1973).
[15] H. Bacry, Ph. Combe, and P. Sorba, Rep. Math. Phys. **5**, 145 (1974).
[16] H. Bacry, Ph. Combe, and P. Sorba, Rep. Math. Phys. **5**, 361 (1974).
[17] P. Sorba, J. Math. Phys. **17**, 941 (1976).

[18]W. I. Fushchich, "The symmetry of mathematical physics problems," in *Algebraic–Theoretical Studies in Mathematical Physics* (Mathematical Institute, Kiev, 1981) (in Russian).
[19]W. I. Fushchich, "On symmetry and particular solutions of some multidimensional equations of mathematical physics," in *Algebraic-Theoretical Methods in Mathematical Physics Problems* (Mathematical Institute, Kiev, 1983) (in Russian).
[20]A. M. Grundland, J. Harnad, and P. Winternitz, J. Math. Phys. **25**, 791 (1984).
[21]A. G. Nikitin, W. I. Fushchich, and I. I. Jurik, Theor. Math. Phys. **26**, 206 (1976).
[22]W. I. Fushchich and A. G. Nikitin, J. Phys. A: Math. Gen. **13**, 2319 (1980).
[23]N. Jacobson, *Lie Algebras* (Dover, New York, 1962).
[24]M. Goto and F. D. Grosshans, *Semisimple Lie Algebras* (Dekker, New York, 1978).
[25]E. Goursat, Ann. Sci. Ecole Norm. Sup. (3) **6**, 9 (1889).
[26]W. I. Fushchich and A. G. Nikitin, *Symmetries of Maxwell's Equations* (Reidel, Dordrecht, 1987).
[27]W. I. Fushchich and N. I. Serov, J. Phys. A: Math. Gen. **16**, 3645 (1983).
[28]W. I. Fushchich and W. M. Shtelen, J. Phys. A: Math. Gen. **18**, 271 (1985).
[29]L. F. Barannik and W. I. Fushchich, "The invariants of subgroups of the generalized Poincaré group $P(1,n)$," preprint, Mathematical Institute, Kiev 86.86, 1986.
[30]A. F. Barannik and W. I. Fushchich, "On continuous subgroups of pseudoorthogonal and pseudounitary groups," preprint, Mathematical Institute, Kiev 86.87, 1986.

Time-ordered operators and Feynman–Dyson algebras

Tepper L. Gill[a), b)]
Department of Physics, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061
W. W. Zachary
Naval Research Laboratory, Washington, D.C. 20375

(Received 28 February 1986; accepted for publication 18 March 1987)

An approach to time-ordered operators based upon von Neumann's infinite tensor product Hilbert spaces is used to define Feynman–Dyson algebras. This theory is used to show that a one-to-one correspondence exists between path integrals and semigroups, which are integral operators defined by a kernel, the reproducing property of the kernel being a consequence of the semigroup property. For path integrals constructed from two semigroups, the results are more general than those obtained by the use of the Trotter–Kato formula. Perturbation series for the Feynman–Dyson operator calculus for time evolution and scattering operators are discussed, and it is pointed out that they are "asymptotic in the sense of Poincaré" as defined in the theory of semigroups, thereby giving a precise formulation to a well-known conjecture of Dyson stated many years ago in the context of quantum electrodynamics. Moreover, the series converge when these operators possess suitable holomorphy properties.

I. INTRODUCTION

It has long been an open question as to what mathematical meaning can be given to the Feynman–Dyson time-ordered operator calculus, which was developed in the 1950's for the study of quantum electrodynamics. In this paper we define Feynman–Dyson algebras and show that they give a natural algebraic framework which allows for the replacement of the *noncommutative structure* of quantum theory with a uniquely defined commutative structure in the time-ordered sense. This approach is analogous to the well-known method in the study of Lie algebras wherein the use of the universal enveloping algebra allows the replacement of a nonassociative structure with a uniquely defined *associative structure* for the development of a coherent representation theory.[1]

The use of this tensor algebra framework allows us to improve upon the customary formal approach to time-ordered operators based upon product integration.

In Sec. II we discuss infinite tensor product Hilbert spaces V and V_ϕ modeled on an arbitrary separable Hilbert space $\mathscr{H}$ and discuss the relationship between algebras of bounded linear operators on these two types of spaces. *It is* shown that V_ϕ *may be assumed separable with no loss in generality* (see also Sec. IV).

In Sec. III we apply these considerations to the discussion of time-ordered integral operators and discuss how this approach leads to unique solutions to the Cauchy problem for the Schrödinger equation with time-dependent Hamiltonians. The use of infinite tensor product Hilbert spaces requires the introduction of a new topology, and so we discuss how uniqueness in the Cauchy problem is to be understood in this framework.

In Sec. IV we discuss the relationship between various algebras of bounded linear operators on infinite tensor product Hilbert spaces and give a mathematically rigorous treatment of algebras of time-ordered operators on these spaces. The latter algebras, called Feynman–Dyson algebras, provide a mathematical treatment of Feynman's operator calculus.[2] Our use of infinite tensor product Hilbert spaces in this connection can be seen to be the mathematical embodiment of the method of Fujiwara[3] in the implementation of Feynman's approach. The definition of these so-called "expansional" operators has been discussed in a Banach algebraic framework different from that of the present paper by Miranker and Weiss[4] and Araki.[5] Related discussions of time-ordered operators have been given by Nelson[6] and Maslov.[7]

In Sec. V we apply our theory of time-ordered operators to the discussion of path integrals of the type first envisioned by Feynman.[8] We show that there exists a one-to-one correspondence between path integrals and semigroups which are integral operators defined by a kernel. In this situation, the reproducing property of the kernel follows from the semigroup property. In this section, path integrals are written for more general Hamiltonians than perturbations of Laplacians by making use of some results of Maslov and Shishmarev[9,10] on hypoelliptic pseudodifferential operators. In those cases in which one is dealing with two semigroups, it is not necessary to assume that the sum of the generators is a generator of a third semigroup. In particular, it is not necessary to assume that one of the two generators is small in some sense relative to the other.

In Sec. VI we discuss perturbation expansions for time-evolution operators. It is shown that these expansions generally do not converge, but are "asymptotic in the sense of Poincaré" as this term is used in the theory of semigroups.[11] This nonconvergence of the perturbation expansions was conjectured in the special case of the renormalized perturbation expansions of quantum electrodynamics in a well-known paper by Dyson.[12] We also prove that these series converge when the semigroups possess suitable holomorphy properties.

Section VII consists of some concluding remarks.

[a)] On leave from the Department of Mathematics, Howard University, Washington, D.C. 20059.

[b)] Present address: Department of Electrical Engineering, Howard University, Washington, D.C. 20059.

 0022-2488/87/071459-12$02.50

II. PRELIMINARIES

Let $J = [-T,T]$, $T > 0$, denote a compact subinterval of the real line and $V = \otimes_{s \in J} \mathscr{H}(s)$ the infinite tensor product Hilbert space, where $\mathscr{H}(s) = \mathscr{H}$ for each $s \in J$ and $\mathscr{H}$ denotes a fixed abstract separable Hilbert space. Here $L[\mathscr{H}]$ and $L[V]$ denote the bounded linear operators on the respective spaces. Here $L[\mathscr{H}(s)]$ is defined by

$$L[\mathscr{H}(s)] = \left\{ B(s) = \overline{\underset{T > r > s}{\otimes} I_r \otimes \widetilde{B} \otimes \left(\underset{s > r > -T}{\otimes} I_r \right)} \middle| \widetilde{B} \in L[\mathscr{H}] \right\} \tag{2.1}$$

where I_r is the identity operator, and $L^{\#}[V]$ is the uniform closure of the algebra generated by the family: $\{L[\mathscr{H}(s)] | s \in J\}$.

Definition 2.1: We say that $\phi = \otimes_s \phi_s$ is *equivalent* to $\psi = \otimes_S \psi_s$ and write $\phi \simeq \psi$ if and only if

$$\sum_s |\langle \phi_s, \psi_s \rangle_s - 1| < \infty , \tag{2.2}$$

where $\langle \ , \ \rangle_s$ denotes the inner product on $\mathscr{H}(s)$. It is to be understood that the sum is meaningful only if at most a countable number of terms are different from zero. The following result is due to von Neumann,[13] but see Guichardet[14] for a simplified proof.

Theorem 2.1: The above relation is an equivalence relation V. If we let V_ϕ denote the closure of the linear spin of all $\psi \simeq \phi$, then (1) ψ not equivalent to ϕ implies $V_\psi \cap V_\phi = \{0\}$; and (2) if we replace J by $\widehat{J} \subset J$, where $\widehat{J}$ is a countable dense subset, in our definition of V [i.e., $V = \widehat{\otimes}_{s \in \widehat{J}}(s)$], then V is a separable Hilbert space.

Let $\mathbb{P}_\phi$ be the projection from V onto V_ϕ.

Theorem 2.2[13]**:** For all $T \in L^{\#}[V]$, the restriction of T to V_ϕ is a bounded linear operator, and

$$\mathbb{P}_\phi T = T \mathbb{P}_\phi . \tag{2.3}$$

Let $\mathbb{C}[V]$ denote the set of closable linear operators on V.

Definition 2.2: An exchange operator $E[t,t']$ is a linear operator defined on $\mathbb{C}[V]$ for pairs $t,t' \in J$ such that

(1) $E[t,t']$ maps $\mathbb{C}[\mathscr{H}(t')]$ onto $\mathbb{C}[\mathscr{H}(t)]$,
(2) $E[t,s]\, E[s,t'] = E[t,t']$,
(3) $E[t,t']\, E[t',t] = I$,
(4) if $s \neq t,t'$, then $E[t,t']A(s) = A(s)$, for all $A(s) \in \mathbb{C}[\mathscr{H}(s)]$.

It should be noted that $E[t,t']$ is linear in the sense that whenever the sum of two closable operators is defined and closable, then $E[t,t']$ maps in the appropriate manner (see Gill[15]). In particular, $E[t,t']$ restricted to $L^{\#}[V]$ is a Banach algebra isomorphism and $E[t,t']\, E[s,s'] = E[s,s']\, E[t,t']$ for distinct pairs (t,t') and (s,s') in J.

Theorem 2.3: If $F = \Pi_{n=1}^{\infty} E[\tau_n, s_n]$, $\{(\tau_n, s_n) \in J \times J | n \in \mathbb{N}\}$ then F is a Banach algebra isomorphism on $L^{\#}[V]$ and

(1) $\|F\|_{\#} = 1$,
(2) $F^{-1} = F$.

Proof: As $\|E[\tau,s]\|_{\#} = 1$, F is a convergent product of algebra isomorphisms and $\|F\|_{\#} \leqslant \|E[\tau_n,s_n]\|_{\#} = 1$. On the other hand, $1 = \|I\|_{\#} = \|F(I)\|_{\#} \leqslant \|F\|_{\#} \|I\|_{\#}$, so that $\|F\|_{\#} = 1$. Since $E[\tau_n,s_n] E[s_n,\tau_n] = I$ and exchange operators for distinct pairs commute, we see that $F^2 = I \Rightarrow F^{-1} = F$.

Definition 2.3: A chronological morphism (or c-morphism) on $L^{\#}[V]$ is any (Banach) algebra isomorphism F on $L^{\#}[V]$ composed of products of exchange operators such that

(1) $\|F\|_{\#} = 1$,
(2) $F^{-1} = F$.

Definition 2.4: Let $\{\widetilde{H}(t) | t \in J\} \subset \mathbb{C}[\mathscr{H}]$ denote a family of densely defined closed self-adjoint operators on $\mathscr{H}$, then the corresponding time-ordered version in $\mathbb{C}[V]$ is defined by

$$H(t) = \overline{\underset{T > s > t}{\otimes} I_s \otimes \widetilde{H}(t) \otimes \left(\underset{t > s > -T}{\otimes} I_s \right)} . \tag{2.4}$$

Definition 2.5: A family $\{H(t) | t \in J\} \subset \mathbb{C}[V]$ is said to be *chronologically continuous (or c-continuous) in the strong sense at* t_0 if there exists an exchange operator $E[t_0,t]$ such that

$$\lim_{t \to t_0} \|E[t_0,t] H(t)\phi - H(t_0)\phi\| = 0 , \tag{2.5}$$

where $\phi \in \otimes_{s \in J} \mathscr{D}[\mathscr{H}(s)]$.

Definition 2.6: The family $\{H(t) | t \in J\}$ is said to be chronologically differentiable (or c-differentiable) in the strong sense at t_0 if there exists an operator $DH(t_0)$ and an exchange operator $E(t_0,t)$ such that

$$\lim_{t \to t_0} \left\| \frac{E(t_0,t)H(t)\phi - H(t_0)\phi}{t - t_0} - DH(t_0)\phi \right\| = 0 ,$$

for all $\phi \in \otimes_{s \in J} \mathscr{D}(H(s))$.

Theorem 2.4: Suppose the family of operators $\{\widetilde{H}(t) | t \in J\}$ have a common domain. Then the corresponding family $\{H(t) | t \in J\}$ is strongly c-continuous iff $\{\widetilde{H}(t) | t \in J\}$ is strongly continuous.

Proof: See Gill.[15]

III. INTEGRALS AND EVOLUTIONS

In the following discussion, all operators of the form $\{\widetilde{A}(t) | t \in J\}$ are closed infinitesimal generators of contraction semigroups, while $\{\widetilde{H}(t) | t \in J\}$ are strongly continuous densely defined linear operators with a common domain, and generate unitary groups. The corresponding operators of the form $\{A(t) | t \in J\}$ [resp. $\{H(t) | t \in J\}$] are the time-ordered versions. Define $A^z(t)$ by

$$A^z(t) = \exp\{zA(t)\} - I/z \tag{3.1}$$

and recall that $\exp\{A^z(t)\}$ is a linear contraction and s-$\lim_{z \downarrow 0} A^z(t) = A(t)$ (strong limit). Similar results hold for $H^z(t)$, with z replaced by iz in (3.1).

Definition 3.1: An integral approximate on $L^{\#}[V]$ is a family of operators of the form $\{Q^z_\lambda[t, -T]$ with $-T \leqslant t \leqslant T, \lambda > 0\}$, where

$$Q^z_\lambda[t, -T] = e^{-2\lambda T} \sum_{n=0}^{\infty} \frac{(2\lambda T)^n}{n!} \sum_{j=1}^{k(n)} \Delta t_j A^z(\tau_j) .$$

For each $n,k = k(n) \geqslant n$ and $\{\mathbb{P}_k = \{-T = t_1 < t_2 \cdots < t_k = t\}, n,k \in \mathbb{N}\}$ is a family of partitions of $[-T,t]$ such that $\lim_{n \to \infty} |\mathbb{P}_k| = 0$ and we take $\tau_j \in [t_{j-1}, t_j)$.

Definition 3.2: Let $\{Q^z_\lambda[t, -T]\}$ and $\{\bar{Q}^z_\lambda[t, -T]\}$ be any two families of integral approximates. We say Q^z_λ is c-

equivalent to $\overline{Q}_\lambda^z$ and write $Q_\lambda^z \overset{c}{\simeq} \overline{Q}_\lambda^z$ (in the uniform sense) if and only if there exists a c-morphism $F=F[Q_\lambda^z,\overline{Q}_\lambda^z]$ such that

$$\lim_{\lambda\to\infty} \|Q_\lambda^z[t,-T] - F\overline{Q}_\lambda^z[t,-T]\| = 0 . \tag{3.2}$$

Theorem 3.1: The relation $\overset{c}{\simeq}$ is an equivalence relation on the set of all integral approximates on $L^{\#}[V]$.

Proof: Reflexivity is obvious. To prove symmetry, we note that

$$\|Q_\lambda^z - F\overline{Q}_\lambda^z\| = \|F^{-1}Q_\lambda^z - \overline{Q}_\lambda^z\|$$

since $\|F\|=1$, and $F=F^{-1}$. Hence $Q_\lambda^z \overset{c}{\simeq} \overline{Q}_\lambda^z$ implies $\overline{Q}_\lambda^z \overset{c}{\simeq} Q_\lambda^z$. To prove transitivity, suppose F_1 and F_2 exist such that

$$\lim_{\lambda\to\infty} \|Q_\lambda^z - F_1\overline{Q}_\lambda^z\| = 0, \quad \lim_{\lambda\to\infty} \|\overline{Q}_\lambda^z - F_2\overline{\overline{Q}}_\lambda^z\| = 0 .$$

Setting $F=F_1F_2$ we have

$$\begin{aligned}\|Q_\lambda^z - F\overline{\overline{Q}}_\lambda^z\| &= \|Q_\lambda^z - F_1\overline{Q}_\lambda^z + F_1\overline{Q}_\lambda^z - F_1F_2\overline{\overline{Q}}_\lambda^z\| \\ &\leqslant \|Q_\lambda^z - F_1\overline{Q}_\lambda^z\| + \|\overline{Q}_\lambda^z + F_2\overline{\overline{Q}}_\lambda^z\| ,\end{aligned}$$

hence $\lim_{\lambda\to\infty} \|Q_\lambda^z - F\overline{\overline{Q}}_\lambda^z\| = 0$, so that $Q_\lambda^z \overset{c}{\simeq} \overline{\overline{Q}}_\lambda^z$.

Here $Q^z[t,-T] = \text{s-lim}_{\lambda\to\infty} Q_\lambda^z[t,-T]$ is called the *time-ordered integral operator* associated with the family $\{A^z(t)\,|\,t\in J\} \subset L^{\#}[V]$ if the above limit exists.

Theorem 3.2 (existence): For the family $\{H^z(t)\,|\,t\in J\}$ we have (1) $\text{s-lim}_{\lambda\to\infty} Q_\lambda^z[t,-T] = Q^z[t,-T]$ exists and

$$Q^z[t,-T] = Q^z[t,s] + Q^z[s,-T] , \quad -T\leqslant s<t ,$$

(2) $\text{s-lim}_{z\downarrow 0} Q^z[t,-T] = Q[t,-T]$ exists, is a densely defined generator of a unitary group on V, and

$$Q[t,-T] = Q[t,s] + Q[s,-T] ,$$

and

(3) $\text{s-lim}_{\lambda\to\infty}[\text{s-lim}_{z\downarrow 0} Q_\lambda^z[t,-T]] = \text{s-lim}_{z\downarrow 0}[\text{s-lim}_{\lambda\to\infty} Q_\lambda^z[t,-T]]$.

Proof: See Gill,[16] Theorems (1.1) and (1.2).

From now on, our results assume that we are working with the family $\{H^z(t)\,|\,t\in J\}$.

Theorem 3.3: Let $Q_\lambda^z[t,-T]$ and $\overline{Q}_\lambda^z[t,-T]$ be two integral approximates with the same family of partitions but different points $\tau_j,s_j\in[t_{j-1},t_j)$ ("place values"). Then $Q_\lambda^z \overset{c}{\simeq} \overline{Q}_\lambda^z$ (in the strong sense).

Proof: Define

$$F = \prod_{n=1}^{\infty}\left(\prod_{j=1}^{n} E[\tau_j,s_j]\right)$$

so that

$$F\overline{Q}_\lambda^z = e^{-2\lambda T}\sum_{n=0}^{\infty}\frac{(2\lambda T)^n}{n!}\sum_{j=1}^{k}\Delta t_j E[\tau_j,s_j]H^z(s_j) . \tag{3.3}$$

By Theorem 2.3, we see that F is a c-morphism and

$$\begin{aligned}&\|Q_\lambda^z\phi - F\overline{Q}_\lambda^z\phi\| \\ &\quad\leqslant e^{-2\lambda T}\sum_{n=0}^{\infty}\frac{(2\lambda T)^n}{n!} \\ &\qquad\times\sum_{j=1}^{k}\Delta t_j\|H^z(\tau_j)\phi - E[\tau_j,s_j]H^z(s_j)\phi\| .\end{aligned}$$

We now note that strong c-continuity of $H(t)$ (cf. definition 2.5) implies strong c-continuity of $H^z(t)$ so, given $\epsilon>0$, there exists $\delta>0$ such that $|\tau-s|<\delta$ implies for $\phi\in V$, $\|H^z(\tau)\phi - E[\tau,s]H^z(s)\phi\| < \epsilon/(t+T)$. Now, choose N so large that $n\geqslant N$ implies $|\mathbb{P}_k|<\delta$, then

$$\begin{aligned}\|Q_\lambda^z\phi - F\overline{Q}_\lambda^z\phi\| &\leqslant e^{-2\lambda T}\sum_{n=0}^{N-1}\frac{(2\lambda T)^n}{n!}\sum_{j=1}^{k}\Delta t_j\|H^z(\tau_j)\phi - E[\tau_j,s_j]H^z(s_j)\phi\| \\ &\quad + e^{-2\lambda T}\sum_{n=N}^{\infty}\frac{(2\lambda T)^n}{n!}\sum_{j=1}^{k}\Delta t_j\|H^z(\tau_j)\phi - E[\tau_j,s_j]H^z(s_j)\phi\| \\ &\leqslant e^{-2\lambda T}\sum_{n=0}^{N-1}\frac{(2\lambda T)^n}{n!}\sum_{j=1}^{k}\Delta t_j\|H^z(\tau_j)\phi - E[\tau_j,s_j]H^z(s_j)\phi\| + \left(e^{-2\lambda T}\sum_{n=N}^{\infty}\frac{(2\lambda T)^n}{n!}\right)\epsilon \\ &< e^{-2\lambda T}\sum_{n=0}^{N-1}\frac{(2\lambda T)^n}{n!}\sum_{j=1}^{k}\Delta t_j\|H^z(\tau_j)\phi - E[\tau_j,s_j]H^z(s_j)\phi\| + \epsilon .\end{aligned}$$

If we now let $\lambda\to\infty$, we obtain $\lim_{\lambda\to\infty}\|Q_\lambda^z\phi - F\overline{Q}_\lambda^z\phi\| < \epsilon$. Since ϵ was arbitrary we are done. ||

Let us note that in Theorem 3.3 it is not necessary to require that $s_j,\tau_j\in[t_{j-1},t_j)$. It suffices to assume that for n sufficiently large, $|s_j-\tau_j|<\delta$, $1\leqslant j\leqslant k(n)$ [i.e., $\lim_{n\to\infty}|s_j-\tau_j|=0$ $\forall_j$, $1\leqslant j\leqslant k(n)$].

Let $\overline{Q}_\lambda^z$ and $\overline{\overline{Q}}_\lambda^z$ be two integral approximates generated from arbitrary families of partitions $\{\overline{\mathbb{P}}_{l_1}\}$, $\{\overline{\overline{\mathbb{P}}}_{l_2}\}$ with respective place values $\overline{\tau}_l\in(\overline{t}_{l-1},\overline{t}_l)$, $1\leqslant l\leqslant l_1(n)$, and $\overline{\overline{\tau}}_l\in[\overline{\overline{t}}_{l-1},\overline{\overline{t}}_l)$, $1\leqslant l\leqslant l_z(n)$. Define a new family of partitions $\mathbb{P}_k = \overline{\mathbb{P}}_{l_1}\cup\overline{\mathbb{P}}_{l_2}$ and integral approximate Q_λ^z with $\tau_j\in[t_{j-1},t_j)$.

Since $\overline{\mathbb{P}}_{l_1}\subset\mathbb{P}_k$,

$$\overline{Q}_\lambda^z = e^{-2\lambda T}\sum_{n=0}^{\infty}\frac{(2\lambda T)^n}{n!}\sum_{l=1}^{l_1}\overline{\Delta t_l}\,H^z(\overline{\tau}_l)$$

may be reindexed to give

$$\overline{Q}_\lambda^z = e^{-2\lambda T}\sum_{n=0}^{\infty}\frac{(2\lambda T)^n}{n!}\sum_{j=1}^{k}\Delta t_j H^z(s_j) , \tag{3.4}$$

where $s_j = \bar{\tau}_l$ for $\bar{t}_{l-1} \leqslant t_{j-1} < t_j \leqslant \bar{t}_l$. Thus $\bar{Q}^z_\lambda$ and Q^z_λ have the same family of partitions, but different place values.

Theorem 3.4: $\bar{Q}^z_\lambda \overset{c}{\simeq} \bar{\bar{Q}}^z_\lambda$.

Proof: We first show that $\bar{Q}^z_\lambda \overset{c}{\simeq} Q^z_\lambda$. From the above remarks, it suffices to show that $|\tau_j - s_j| \to 0$, $n \to \infty$. To see this, recall that $\tau_j \in [t_{j-1}, t_j)$ and $s_j = \bar{\tau}_l$ for $\bar{t}_{l-1} \leqslant t_{j-1} < t_j \leqslant \bar{t}$, hence $|\tau_j - s_j| < \overline{\Delta t_l} \to 0$ as $n \to \infty$. Therefore $\bar{Q}^z_\lambda \overset{c}{\simeq} Q^z_\lambda$ by Theorem 3.3. The same argument with $\bar{Q}^z_\lambda$ replaced by $\bar{\bar{Q}}^z_\lambda$ shows that $\bar{\bar{Q}}^z_\lambda \overset{c}{\simeq} Q^z_\lambda$. We now use the transitivity of $\overset{c}{\simeq}$ to conclude that $\bar{Q}^z_\lambda \overset{c}{\simeq} \bar{\bar{Q}}^z_\lambda$. ‖

Definition 3.3: A time-ordered integral operator is said to be *chronologically unique* (or *c-unique*) if every integral approximate is c-equivalent.

Let $Q[t, -T] = \text{s-lim}_{z \downarrow 0} Q^z[t, -T]$.

Theorem 3.5: (1) $Q^z[t, -T]$ is c-unique.

(2) $Q[t, -T]$ is a generator of a unitary group (densely defined and closed).

Proof: (1) is clear; (2) is in Gill.[16]

The uniqueness property in part (1) of this theorem is an important feature of our theory. There are path integrals which depend upon the choice of partition. See Ref. 17 for a discussion.

Theorem 3.6: $U^z[t, -T] = \exp\{-iQ^z[t, -T]\}$ satisfies

$$(1)\ U^z[t, -T] = U^z[t,s]U^z[s, -T], \quad -T \leqslant s \leqslant t,$$

$$(2)\ i\frac{\partial U^z[t, -T]}{\partial t} = H^z(t)U^z[t, -T],$$

$$(3)\ U[t, -T] = \underset{z \downarrow 0}{\text{s-lim}}\, U^z[t, -T] = \exp\{-iQ[t, -T]\}$$

satisfies

$$U[t,s]U[s, -T] = U[t, -T], \quad -T \leqslant s \leqslant t,$$

$$(4)\ i\frac{\partial U[t, -T]}{\partial t} = H(t)U[t, -T].$$

Proof: See Gill.[16] The derivatives are in the strong chronological sense. This theorem allows us to give a complete solution to the Cauchy problem. Recall that if $\phi_0 \in D(\tilde{H}(t)) \subset \mathscr{H}$ for $t \in J$, then the initial value problem

$$i\frac{\partial f(t)}{\partial t} = \tilde{H}(t)f(t), \quad f(-T) = \phi_0,$$

has a unique solution $f(t)$ provided a few additional assumptions are made. For a direct proof with explicit statements of the required additional assumptions, see Tanabe.[18] We prove a similar result in the Hilbert space V with no additional assumptions.

Theorem 3.7: Let $\phi_s = \phi_0$, $\|\phi_0\| = 1$, $s \in J$, and set $\phi = \otimes_s \phi_s$. Then $\phi(t) = U(t, -T)\phi$ is the c-unique solution to

$$i\frac{\partial \phi(t)}{\partial t} = H(t)\phi(t), \quad \phi(-T) = \phi,$$

where the derivatives are interpreted in the strong chronological sense.

Proof: Follows from Theorems 3.5 and 3.6.

IV. OPERATOR ALGEBRAS

Let us recall from Theorem 2.1 that if we replace J by $\bar{J} \subset J$, where $\bar{J}$ is a dense subset and construct $\bar{V} = \hat{\otimes}_{s \in \bar{J}} \mathscr{H}(s)$ then $\bar{V}_\phi$ (the closure of the linear span of all $\psi \simeq \phi$) is a separable (Hilbert) subspace. The next theorem is quite interesting in view of the fact that $\bar{V}$ and V are not related as spaces.

Theorem 4.1: $L^{\#}[\bar{V}] \subset L^{\#}[V]$ (i.e., is an injection into).

Proof: From (2.1), it is easy to see that $L[\mathscr{H}(s)]$ is a closed subalgebra of $L^{\#}[V]$ for each $s \in J$ (a detailed proof is in von Neumann[13]). This is also true for each $s \in \bar{J}$, so the result follows trivially, since $L^{\#}[\bar{V}]$ is generated by $\{L[\mathscr{H}(s)] | s \in \bar{J}\}$, and $L[\mathscr{H}(s)] \subset L^{\#}[V]$, $s \in \bar{J}$.

Let us note that the existence and uniqueness of $Q^z[t, -T]$ and $U[t, -T]$ do not change if we restrict $\{\tau_j | 1 \leqslant j \leqslant k(n), n \in \mathbb{N}\}$, to lie in $\bar{J}$ in defining Q^z_λ and U_λ. This means that the following holds.

Theorem 4.2:

(1) $Q^z[t, -T]$ and $U[t, -T]$ belong to $L^{\#}[\bar{V}]$,

$$(2)\ Q^z[t, -T]|_{\bar{V}_\phi} \in L[\bar{V}_\phi], \tag{4.1}$$

$$(3)\ U[t, -T]|_{\bar{V}_\phi} \in L[\bar{V}_\phi]. \tag{4.2}$$

Proof: (1) is obvious while (2) and (3) follows from Theorem 2.3.

The above result shows that both $U[t, -T]$ and $Q[t, -T]$ are well defined (and the same operators as in V_ϕ) when restricted to $\bar{V}_\phi$, which is a separable Hilbert space. This means that all of standard quantum theory can be formulated in our setting.

We now turn to some other important properties of $L^{\#}[V]$. First, let us establish some notation. If $\{\tilde{B}(t), t \in J\}$ denotes an arbitrary family of opertors in $L[\mathscr{H}]$, the operator $\Pi_{t \in J}\tilde{B}(t)$ (when defined) is understood in its natural order:

$$\prod_{T \geqslant t \geqslant -T} \tilde{B}(t). \tag{4.3}$$

It is easy to see that every operator A in $L^{\#}[V]$ that depends on a countable number of elements in J may be written as

$$A = \sum a_i \prod_{k=1}^{n_i} A_i(t_k), \tag{4.4}$$

where $A_i(t_k) \in L[\mathscr{H}(t_k)]$, $t_1, t_2, \ldots, t_{n_i}$ for all i. Define dT: $L^{\#}[V] \to L[\mathscr{H}]$ by

$$dT[A] = \sum_{i=1}^{\infty} a_i \prod_{n_i \geqslant k \geqslant 1} \tilde{A}_i(t_k). \tag{4.5}$$

Lemma 4.2: The map dT is a bounded linear map which is surjective but not injective.

Proof: The proof is trivial. To see that dT is not injective, note that (for example) $dT[E[t,s]A(s)] = dT[A(s)]$ yet $A(s) \in L[\mathscr{H}(s)]$ while $E[t,s]A(s) \in L[\mathscr{H}(t)]$ so that these operators are not equal when $t \neq s$.

From Theorem 2.2, we know that the algebras $L[\mathscr{H}(t)]$ and $L[\mathscr{H}]$ are isomorphic as Banach algebras so that for each $t\in J$, there exists an isomorphism $t\theta$: $L[\mathscr{H}]\to L[\mathscr{H}(t)]$. Now $t\theta^{-1}$: $L[\mathscr{H}(t)]\to L[\mathscr{H}]$; and since $L[\mathscr{H}(t)]$ is a closed subalgebra of $L^{\#}[V]$, we know that dT restricted to $L[\mathscr{H}(t)]$ is an algebra homomorphism.

Theorem 4.3: $dT_{L[\mathscr{H}(t)]} = t\theta^{-1}$.

Proof: It is clear that $t\theta^{-1}[A(t)] = \widetilde{A}(t)$ and $dT[A(t)] = \widetilde{A}(t)$, $A(t)\in L[\mathscr{H}(t)]$, so we need only show that dT is injective when restricted to $L[\mathscr{H}(t)]$. If $A(t)$ and $B(t)$ belong to $L[\mathscr{H}(t)]$ and $dT[A(t)] = dT[B(t)]$, then $\widetilde{A}(t) = \widetilde{B}(t)$ (by definition of dT) so that $A(t) = B(t)$ by definition of $L[\mathscr{H}(t)]$.

Definition 4.1: The map dT is called the *disentanglement morphism.*

Definition 4.2: The quadruple $(\{t\theta\,|\,t\in J\}$, $L[\mathscr{H}]$, $dT, L^{\#}[V])$, is called a *Feynman–Dyson algebra* (FD *algebra*) over $\mathscr{H}$ for the parameter set J.

We now show that the FD algebra is universal for time ordering in the following sense.

Theorem 4.4: Given any family $\{\widetilde{B}(t)\,|\,t\in J\}\in(L[\mathscr{H}])^J$ there is a unique family $\{B(t)\,|\,t\in J\}\subset L^{\#}[V]$ such that the following conditions hold.

(1) $B(t)\in L[\mathscr{H}(t)]$, $t\in J$.

(2) $dT[B(t)] = \widetilde{B}(t)$, $t\in J$.

(3) For an arbitrary family $\{\{\tau_j\,|\,1\leqslant j\leqslant n\}\,|\,n\in N\}$, $\tau_j\in J$ (distinct) the map

$$\mathop{X}_{n=1}^{\infty}(\widetilde{B}(\tau_n),\ldots,\widetilde{B}(\tau_1))\to\sum_{n=1}^{\infty}a_n\prod_{n\geqslant j\geqslant 1}\widetilde{B}(\tau_j)$$

from

$$\mathop{X}_{n=1}^{\infty}\left\{\mathop{X}_{j=1}^{n}L[\mathscr{H}]\right\}\text{ to }L[\mathscr{H}],$$

has a unique factorization through $L^{\#}[V]$ so that

$$\sum_{n=0}^{\infty}a_n\prod_{n\geqslant j\geqslant 1}\widetilde{B}(\tau_j)$$

corresponds to

$$\sum_{n=0}^{\infty}a_n\prod_{j=1}^{n}B(\tau_j)\,.$$

Here we naturally assume that $\{a_n\}$ is such that

$$\sum_{n=0}^{\infty}a_n\prod_{n\geqslant j\geqslant 1}\widetilde{B}(\tau_j)\in L[\mathscr{H}]\,.$$

Proof: $B(t) = t\theta[\widetilde{B}(t)]$, $\forall t\in J$, gives (1). By Theorem 4.3 we have $dT[B(t)] = t\theta^{-1}[B(t)] = \widetilde{B}(t)$ which gives (2). To prove (3), note that

$$\Theta\colon\mathop{X}_{n=1}^{\infty}\left\{\mathop{X}_{j=1}^{n}L[\mathscr{H}]\right\}\to\mathop{X}_{n=1}^{\infty}\mathop{X}_{j=1}^{n}L[\mathscr{H}(\tau_j)]$$

defined by

$$\Theta[X_{n=1}^{\infty}(\widetilde{A}_n,\widetilde{A}_{n-1},\ldots,\widetilde{A}_1)] = X_{n=1}^{\infty}(\tau_n\theta[\widetilde{A}_n],\ \tau_{n-1}\theta[\widetilde{A}_{n-1}],\ldots,\tau_1\theta[\widetilde{A}_1])$$

is one-to-one and onto $(\tau_j\theta[\widetilde{A}_j] = A(\tau_j)\in L[\mathscr{H}(\tau_j)])$. The map

$$\mathop{X}_{n=1}^{\infty}(B(\tau_n),\ldots,B(\tau_1))\to\sum_{n=1}^{\infty}a_n\prod_{j=1}^{\infty}B(\tau_j)\in L^{\#}[V]$$

factors through the tensor algebra $\oplus_{n=1}^{\infty}\{\otimes_{j=1}^{n}L[\mathscr{H}(\tau_j)]\}$ via the universal property of that object (Hu,[19] p. 179). We now note that $\oplus_{n=1}^{\infty}\{\otimes_{j=1}^{n}L[\mathscr{H}(\tau_j)]\}\subset L^{\#}[V]$. In diagram form we have

$$\begin{array}{ccc}
\mathop{X}\limits_{n=1}^{\infty}(\widetilde{B}(\tau_n),\ldots,\widetilde{B}(\tau_1))\in\mathop{X}\limits_{n=1}^{\infty}\left\{\mathop{X}\limits_{j=1}^{n}L[\mathscr{H}]\right\} & \xrightarrow{f} & \sum\limits_{n=1}^{\infty}a_n\prod\limits_{n\geqslant j\geqslant 1}\widetilde{B}(\tau_j)\in L[\mathscr{H}]\\
\Big\downarrow\theta & & \Big\uparrow dT\\
\mathop{X}\limits_{n=1}^{\infty}(B(\tau_n),\ldots,B(\tau_1))\in\mathop{X}\limits_{n=1}^{\infty}\left\{\mathop{X}\limits_{j=1}^{n}L[\mathscr{H}(\tau_j)]\right\} & \xrightarrow{f_\otimes} & \sum\limits_{n=1}^{\infty}a_n\prod\limits_{j=1}^{n}B(\tau_j)\in L^{\#}[V]
\end{array}$$

so that $dT\circ f_\otimes\circ\theta = f$.

Example 1: Let

$$A(t) = \overline{\mathop{\otimes}_{T\geqslant\tau>t}I_\tau\otimes\widetilde{A}\otimes\left(\mathop{\otimes}_{t>\tau\geqslant -T}I_\tau\right)}\,,$$

$$B(s) = \overline{\mathop{\otimes}_{T\geqslant\tau>s}I_\tau\otimes\widetilde{B}\otimes\left(\mathop{\otimes}_{s>\tau\geqslant -T}I_\tau\right)}\,,$$

where $\widetilde{A}$ and $\widetilde{B}$ are bounded on $\mathscr{H}$. If $s<t$, then by Lemma 2.3 in Ref. 15, we have

$$A(t)B(s) = B(s)A(t) = \overline{\mathop{\otimes}_{T\geqslant\tau>t}I_\tau\otimes\widetilde{A}\otimes\left(\mathop{\otimes}_{t>\tau>s}I_\tau\right)\otimes\widetilde{B}\otimes\left(\mathop{\otimes}_{s>\tau\geqslant -T}I_\tau\right)}\,,$$

so that $dT[A(t)B(s)] = dT[B(s)A(t)] = \widetilde{A}\widetilde{B}$, while

$$dT[A(t)B(s) - B(t)A(s)] = dT[A(t)B(s) - A(s)B(t)] = \widetilde{A}\widetilde{B} - \widetilde{B}\widetilde{A}\,.$$

Example 2: Let $\{\widetilde{H}(t)\,|\,t\in J\}$ be strongly continuous (with common dense domain), and suppose this family generates a product integral (Dollard and Friedman[20]). Choose any family $\{\mathbb{P}_n\,|\,n\in\mathbb{N}\}$ of partitions such that

$$\lim_{n\to\infty}\prod_{j=1}^{n}\exp\{-i\Delta t_j\widetilde{H}(\tau_j)\} = \widetilde{U}[t,-T]\,,$$

then $\lim_{\lambda\to\infty}\widetilde{U}_\lambda[t,-T] = \widetilde{U}[t,-T]$, where

$$\widetilde{U}_\lambda[t,-T] = e^{-2\lambda T}\sum_{n=0}^{\infty}\frac{(2\lambda T)^n}{n!}\prod_{j=1}^{n}\exp\{-i\Delta t_j\widetilde{H}(\tau_j)\}\,.$$

This follows from the fact that Borel summability is regular. For the same family $\{\mathbb{P}_n\,|\,n\in\mathbb{N}\}$, construct

$$U_\lambda[t,-T] = e^{-2\lambda T}\sum_{n=0}^{\infty}\frac{(2\lambda T)^n}{n!}\exp\left\{-i\sum_{j=1}^{n}\Delta t_j H(\tau_j)\right\}.$$

As in Ref. 16, we see that $U[t,-T]=\lim_{\lambda\to\infty}U_\lambda[t,-T]$ exists in $L^{\#}[V]$. Furthermore, $dT\{U[t,-T]\} = dT\lim_{\lambda\to\infty}U_\lambda[t,-T] = \lim_{\lambda\to\infty}dT\{U_\lambda[t,-T]\} = \widetilde{U}[t,-T]$. We can interchange limits since dT is a closed linear operator on $L^{\#}[V]$. It should be noted that the above limit can exist even if the standard product integral does not. This result will be discussed in a subsequent paper (see Gill and Zachary[21]).

V. APPLICATIONS TO THE CONSTRUCTION OF PATH INTEGRALS

In the present section we consider time-ordered operators in more detail, and discuss the proposition that there exists a one-to-one correspondence between path integrals and semigroups which are integral operators defined by a kernel. We apply our formulation of time-ordered operators to the discussion of path integrals of the type first considered by Feynman.[8] There have been many approaches to the mathematical construction of time-ordered operators and path integrals in recent years. We will not be using any of these approaches, so we content ourselves with offering the following admittedly incomplete list of references[7,9,10,22–25] from which the reader can trace these developments.

Let us consider the time-independent self-adjoint generator $\widetilde{H}_0$ of a unitary group defined on $\mathscr{H}$ in terms of a transition kernel $\widetilde{K}$ which satisfies the Chapman–Kolmogorov equation.

If we replace the operator $\widetilde{H}_0$ by its time-ordered version $\{H_0(t)\colon t\in J\}$, we induce a natural family of kernels $K(x(t),t;y(s),s)$ via Theorem 3.2. To see this, note that

$$\begin{aligned}
\overline{U}_n(t,-T)\phi_0 &= \exp\left[-i\sum_{j=1}^{n}\int_{t_{j-1}}^{t_j}E(\tau_j,\tau)H_0(\tau)d\tau\right]\phi_0\\
&= \prod_{j=1}^{n}\overline{\left[\left(\bigotimes_{t\geqslant s>\tau_j}I_s\right)\otimes\exp[-i(t_j-t_{j-1})\widetilde{H}_0]\otimes\left(\bigotimes_{\tau_j>s\geqslant -T}I_s\right)\right]}\phi_0\\
&= \prod_{j=1}^{n}\overline{\left[\left(\bigotimes_{t\geqslant s>\tau_j}I_s\right)\otimes\int_{\mathbb{R}^k}\widetilde{K}(x_j,t_j;x_{j-1},t_{j-1})dx_{j-1}\otimes\left(\bigotimes_{\tau_j>s\geqslant -T}I_s\right)\right]}\phi_0\\
&= \prod_{j=1}^{n}\int_{\mathbb{R}^k}\mathbb{K}_{\tau_j}(x_j,t_j;x_{j-1},t_{j-1})\,dx_{j-1}\,\phi_0,
\end{aligned}\tag{5.1}$$

where $\phi_0=\otimes_{s\in J}\phi(s)$, $J=[-T,T]$. In (5.1), $x_j=x(t_j)$ and the index τ_j on $\mathbb{K}$ is used to indicate the time at which $\mathbb{K}$ acts. Combination of (5.1) with Theorem 3.2 shows that $\overline{U}_\lambda(t,-T)$ may be represented in the form

$$\overline{U}_\lambda(t,-T)\phi_0 = e^{-2\lambda T}\sum_{n=0}^{\infty}\frac{(2\lambda T)^n}{n!}\times\prod_{j=1}^{n}\int_{\mathbb{R}^k}\mathbb{K}_{\tau_j}(x_j,t_j;x_{j-1},t_{j-1})\,dx_{j-1}\,\phi_0.\tag{5.2}$$

Since $\overline{U}_\lambda(t,-T)$ exists as a well-defined bounded operator, and

$$\lim_{\lambda\to\infty}\overline{U}_\lambda(t,-T)=U_0(t,-T)$$

exists in the uniform operator topology, $U_0(t,-T)$ has a natural representation as an operator-valued path integral:

$$U_0(t,-T)=\int_{\mathscr{X}(t,-T)}\mathbb{K}(x(t),t;x(s),s)\,\mathscr{D}[x(s)],\tag{5.3}$$

where $\mathscr{X}(t,-T)=\mathbb{R}^{k(t,-T)}$ denotes the set of *all* functions from $[t,-T]$ to $\mathbb{R}^k$. In (5.3) we have used a formal "functional measure" notation, although a measure generally does not exist, as we discuss in more detail below.

In recent years many authors have attempted to bypass the difficulty that Feynman-type path integrals cannot generally be written in terms of countably additive measures,[25] as is the case for its closest relative, the Wiener integral. In the present paper we take the point of view that integration theory, as contrasted with measure theory, is the appropriate vehicle to be considered for a theory of path integration. An essential ingredient in our approach is the idea that it is possible to define path integrals by giving up the requirement of the existence of a countably additive measure. This idea has a precursor in the theory of integration in Euclidean spaces. *That is, it is possible to define a consistent theory of integration, which generalizes Lebesgue integration, in which the integrals are finitely additive, but are generally not countably additive.*[26] Indeed, Henstock[27] has already discussed the Feynman integral from this point of view.

Returning now to our discussion of (5.3), we note that many authors have sought to restrict consideration to continuous functions in the definition of path integrals. The best known example is undoubtedly the Wiener integral.[28] However, the fact that we must see $\mathscr{X}(t,-T)$ follows naturally from the time-ordered operator calculus, and such a restriction is probably neither possible nor desirable in our theory. This means that our approach does not encourage attempts at the standard measure theoretic formulations with countably additive measures. In previous work by one of us,[15] the Riemann-complete (generalized Riemann) integral of Henstock and Kurzweil[26] was employed, because the time-ordered integrals need not be absolutely integrable, even in the bounded operator case. These issues will be studied in greater depth at another time. We note in passing that this failure of absolute integrability also plays an important role in the path integral theory of Albeverio and Høegh-Krohn,[23] and also in more recent developments (see, e.g., Ref. 24). Our theory, to be discussed in the remainder of the present section, allows for more general Hamiltonians.

Before proceeding to a discussion of these results, we pause to discuss some examples. The first one is well known—the familiar Laplacian operator. Our purpose in discussing it here is to show how our theory works in a familiar case.

Let $\widetilde{H}_0 = -\Delta/2$ so that $H_0(t) = -\Delta_t/2$, where the subscript t indicates the time slot at which this operator is to be evaluated. We have

$$\widetilde{K}(x,t;y,s) = (2\pi i(t-s))^{-k/2}\exp[i|x-y|^2/2(t-s)]\,.$$

In this case it is easy to see that

$$\begin{aligned}\overline{U}_n(t,-T)\phi_0 &= \prod_{j=1}^{n}\int_{\mathbf{R}^k}\exp\left[\frac{i(t_j-t_{j-1})}{2}\left|\frac{x_j-x_{j-1}}{t_j-t_{j-1}}\right|^2(\tau_j)\right]\\ &\quad\times\frac{dx_{j-1}\,\phi_0}{[2\pi i(t_j-t_{j-1})]^{k/2}} \qquad (5.4)\\ &=\int_{\mathbf{R}^{kn}}\exp\left[i\sum_{j=1}^{n}\frac{1}{2}(t_j-t_{j-1})\left|\frac{x_j-x_{j-1}}{t_j-t_{j-1}}\right|^2(\tau_j)\right]\\ &\quad\times\prod_{j=1}^{n}\frac{dx_{j-1}}{[2\pi i(t_j-t_{j-1})]^{k/2}}\phi_0\,. \qquad (5.5)\end{aligned}$$

By analogy with the definition of $H_0(t)$ given above, the (τ_j) are used to remind us that the corresponding functions in (5.4) and (5.5) are not ordinary exponentials because they have a specific time slot at which they are evaluated. This is our version of the occurrence of expansionals in the usual approach.[3,5] Using (5.5) with (5.2), we have

$$\begin{aligned}\overline{U}_\lambda^0(t,-T)\phi_0 &= e^{-2\lambda T}\sum_{n=0}^{\infty}\frac{(2\lambda T)^n}{n!}\\ &\quad\times\int_{\mathbf{R}^{kn}}\exp\left[i\sum_{j=1}^{n}\frac{1}{2}(t_j-t_{j-1})\left|\frac{x_j-x_{j-1}}{t_j-t_{j-1}}\right|^2(\tau_j)\right]\\ &\quad\times\prod_{j=1}^{n}D(x_{j-1})\phi_0\,,\end{aligned}$$

where $D(x_{j-1}) = (2\pi i(t_j-t_{j-1}))^{-k/2}\,dx_{j-1}$. This means that $U^0(t,-T)$ may be represented by

$$U^0(t,-T)\phi = \int_{\mathscr{P}(t,-T)}\exp\left[\frac{1}{2}i\int_{-T}^{t}\left|\frac{dx}{ds}\right|^2 ds\right]\times\prod_{t>s>-T}D(x(s))\phi_0\,.$$

As our second example, we consider the operator $\widetilde{H} = \sqrt{-\Delta+\omega^2}$. It was shown by Pursey[29] that the Bargmann–Wigner equation for a relativistic particle of any physically allowed spin value $s = 0,\frac{1}{2},1,\frac{3}{2},\ldots$ is unitarily equivalent to the equation defining the Cauchy problem for this square root operator. Foldy and Wouthuysen[30] showed that this operator is nonlocal with effective spatial extension equal to a Compton wavelength. Our interest here is to show that it is an integral operator defined by a kernel $\widetilde{K}$.

The method of pseudodifferential operators can be used to show that a kernel exists and, under reasonable conditions, can provide a phase space representation as we discuss in detail more general operators later in this section. However, if we desire a direct representation, then other methods are required. In our case, we have found that the method of fractional powers of operator semigroups allows us to solve the problem in a simple manner. By using results on pp. 281 and 302 of Ref. 31, p. 260 of Ref. 32, and p. 498 of Ref. 11, it can be shown that the semigroup generated by the closure of $\sqrt{-\Delta+\omega^2}$, $T(\tau)$, can be written in the form

$$T(\tau)\phi(x) = \frac{i\omega^2}{2\pi^2\tau}\int_{\mathbf{R}^3}\frac{K_2[\omega\tau\sqrt{|x-y|^2/\tau^2-1}]}{|x-y|^2/\tau^2-1}\phi(y)dy\,, \qquad (5.6)$$

where $K_2(\cdot)$ denotes the modified Bessel function of the third kind of order 2. It is clear that $T(\tau)$ is holomorphic. From (5.6) we see that we have an example of a semigroup with a kernel that is not of the form

$$\exp\left[i\frac{m}{2}\left|\frac{x_j-x_{j-1}}{t_j-t_{j-1}}\right|^2(t_j-t_{j-1})\right]. \qquad (5.7)$$

Since (5.7) is appropriate for the nonrelativistic regime, we cannot expect it to have general validity. However, if the argument of the Bessel function is large, we should expect the kernel in (5.6) to approximate (5.7) when $|(x_j-x_{j-1})/(t_j-t_{j-1})|$ is small compared to unity (= speed of light). Since $K_2(z)\sim\sqrt{\pi/2}\,e^{-z}/z$ for large argument, we see that we may approximate the kernel in (5.6) by (using $\sqrt{v^2-1}\to i\sqrt{1-v^2}$)

$$\begin{aligned}\widetilde{K}(x_j,t_j;x_{j-1},t_{j-1}) &\cong \frac{i\omega^2}{2\pi^2(t_j-t_{j-1})}\\ &\quad\times\sqrt{\frac{\pi}{2}}\frac{\exp[-i\omega(t_j-t_{j-1})\sqrt{1-v^2}]}{\sqrt{i\omega(t_j-t_{j-1})\sqrt{1-v^2}(1-v^2)}}\,,\end{aligned}$$

where $v = |(x_j-x_{j-1})/(t_j-t_{j-1})|$. Now, letting $v\to 0$ in the denominator and approximating the square root in the numerator, we obtain

$$\begin{aligned}\widetilde{K}(x_j,t_j;x_{j-1},t_{j-1}) &\cong +i\left(\frac{\omega}{2\pi i(t_j-t_{j-1})}\right)^{3/2}\exp[-i\omega(t_j-t_{j-1})]\\ &\quad\times\exp\left[i\frac{\omega}{2}\left|\frac{x_j-x_{j-1}}{t_j-t_{j-1}}\right|^2(t_j-t_{j-1})\right]. \qquad (5.8)\end{aligned}$$

Thus we see that the kernel in (5.6) reduces to the nonrelativistic limit except for the extra phase factor which corresponds to a rest mass term in the standard approaches. It is important to realize, however, that two distinct assumptions are required to obtain (5.8). The first corresponds to observations far removed from the particle, while the second involves the nonrelativistic approximation. In order to see the effect of the first assumption, we need only note that for small z,

$$K_2(z)\sim 2z^{-2}\,. \qquad (5.9)$$

It is also of interest to investigate the limit $\omega \to 0$ corresponding to a massless particle. In this case we replace $K_2(z)$ by (5.9) to obtain

$$\widetilde{K}(x_j,t_j;x_{j-1},t_{j-1}) \cong \frac{+i}{\pi^2(t_j-t_{j-1})^3}\left[1-\left|\frac{x_j-x_{j-1}}{t_j-t_{j-1}}\right|^2\right]^{-2}. \quad (5.10)$$

It is very interesting to note that both (5.8) and (5.10) are propagators for unitary groups.

In order to describe path integrals for more general situations than covered thus far in the present section, we consider the case of two families of self-adjoint time-ordered operators $\{H_0(t): t \in J\}$ and $\{H_1(t): t \in J\}$ with respective domains D_0 and D_1 which are dense in V_ϕ. It is assumed that both families are strongly c-continuous generators of unitary groups. Consider a partition P_n of $[-T,t]$ as in Definition 3.1 and let $\tau_j,s_j \in [t_{j-1},t_j)$. We then define

$$U_n(t,-T) = \exp\left[\sum_{j=1}^n (t_j-t_{j-1})\{H_0(\tau_j)+H_1(s_j)\}\right],$$

$$U_n^0(t,-T) = \exp\left[\sum_{j=1}^n (t_j-t_{j-1})H_0(\tau_j)\right],$$

$$U_n^1(t,-T) = \exp\left[\sum_{j=1}^n (t_j-t_{j-1})H_1(s_j)\right].$$

Since we do not assume any relationship between D_0 and D_1, $U_n(t,-T)$ is well defined except when $\tau_j = s_j$ for some j. In the contrary case we have

$$U_n(t,-T) = U_n^0(t,-T)U_n^1(t,-T) = U_n^1(t,-T)U_n^0(t,-T).$$

Now, defining $U_\lambda(t,-T)$, $U_\lambda^0(t,-T)$, and $U_\lambda^1(t,-T)$ by combining the notations of Theorems 3.2 and 3.6, we have the following theorem.

Theorem 5.4[16]**:**

(1) $\lim_{\lambda\to\infty} U_\lambda(t,-T) = U(t,-T)$ exists a.s.,

(2) $U(t,-T) = U^1(t,-T)U^0(t,-T) = U^0(t,-T)U^1(t,-T)$ a.s.

By specializing the partition P_n by choosing $t_j - t_{j-1} = 1/n$, $1 \leqslant j \leqslant n$, we have

$$U_\lambda(t,-T) = e^{-2\lambda T}\sum_{n=0}^\infty \frac{(2\lambda T)^n}{n!}\left[\prod_{j=1}^n \exp\left\{\frac{1}{n}H_0(\tau_j)\right\}\right] \times\left[\prod_{j=1}^n \exp\left\{\frac{1}{n}H_1(s_j)\right\}\right].$$

This is reminiscent of the Trotter–Kato product formula,[31,33] but is more general due to our weak restrictions on the two self-adjoint operator families and our use of the Borel summability procedure. For example, it is not necessary to assume that H_0+H_1 is self-adjoint as in Ref. 33. This means that, in particular, it is not necessary to assume that one of the operators, $\widetilde{H}_1$ say, is small in some sense relative to the other, $\widetilde{H}_0$. The fact that Theorem 5.4 does not depend on the domains is anticipated by the work of Chernoff[34] on the "generalized additivity" of generators of semigroups arising from Trotter–Kato-type product formulas. This author has given an extensive discussion of these formulas for quite arbitrary domains. See also Kato[35] for a discussion of the case of two positive self-adjoint operators on a Hilbert space when the intersection of their domains may be arbitrary.

We remind the reader at this point that the Trotter–Kato formula is one of the standard methods for formal derivations of Feynman's formula for the nonrelativistic time-evolution operator.[23,36] Similarly, Theorem 5.4 is the basis for our treatment of the Feynman integral which, however, is completely rigorous.

We now discuss the results in Theorem 5.4 from a slightly different point of view. We see from this theorem that

$$U(t,-T) = \exp\left[-i\int_{-T}^t \{H_0(\tau)+H_1(\tau)\}\,d\tau\right]$$

exists a.e. and

$$U(t,-T) = \lim_{\lambda\to\infty} \overline{U}_\lambda(t,-T),$$

where

$$\overline{U}_\lambda(t,-T) = e^{-2\lambda T}\sum_{n=0}^\infty \frac{(2\lambda T)^n}{n!} \times\exp\left[-i\sum_{j=1}^n \int_{t_{j-1}}^{t_j} \{E(\tau_j,\tau)H_0(\tau) + E(s_j,\tau)H_1(\tau)\}\,d\tau\right] \quad (5.11)$$

with $\tau_j,s_j \in [t_{j-1},t_j)$. If we use (5.5), the exponent in (5.11) can be replaced by

$$i\sum_{j=1}^n \left\{(t_j-t_{j-1})\left|\frac{x_j-x_{j-1}}{t_j-t_{j-1}}\right|^2(\tau_j) - \int_{t_{j-1}}^{t_j} E(s_j,\tau)H_1(x(\tau),\tau)d\tau\right\}.$$

Taking limits, we have

$$U(t,-T) = \iint_{\mathscr{P}(t,-T)} \exp\left[i\int_{-T}^t \left\{\frac{1}{2}\left|\frac{dx}{ds}\right|^2 - H_1(x(s),s)\right\}ds\right]\prod_{t>s>-T} D(x(s)). \quad (5.12)$$

It is clear that our conditions on the family $\widetilde{H}_1(x,s)$ are sufficiently general to cover most cases of interest in nonrelativistic quantum theory. We can now write (5.12) in the form originally envisioned by Feynman, namely,

$$U(t,-T) = \int_{\mathscr{P}(t,-T)} \exp\left[i\int_{-T}^t L(\dot{x}(s),x(s),s)ds\right] \times\prod_s D(x(s)),$$

where $x(s) = dx/ds$ and

$$L(\dot{x}(s),x(s),s) = \frac{1}{2}\left|\frac{dx}{ds}\right|^2 - H_1(x(s),s)$$

denotes the Lagrangian.

We now generalize the representation (5.12) by considering more general choices for the operator $\widetilde{H}_0(\tau)$. For these operators we choose the class of hypoelliptic pseudodifferential operators studied by Shishmarev.[10] In this way, we are able to derive a representation for $U(t, -T)$ analogous to (5.12) which will include cases useful for studies in relativistic quantum mechanics such as, e.g., perturbations of the square root operator studied earlier.

Let $\widetilde{H}(x,p)$ denote a $k \times k$ matrix operator $[\widetilde{H}_{ij}(x,p)]$, $i,j = 1,2,...,k$, whose components are pseudodifferential operators with symbols $h_{ij}(x,\eta)\in C^{\infty}(\mathbb{R}^N \times \mathbb{R}^N)$ and we have, for any multi-indices α and β,

$$|h_{ij(\beta)}^{(\alpha)}(x,\eta)| \leqslant C_{\alpha\beta}(1 + |\eta|)^{m - \xi|\alpha| + \delta|\beta|}, \qquad (5.13)$$

where

$$h_{ij(\beta)}^{(\alpha)}(x,\eta) = \partial^{\alpha} p^{\beta} h_{ij}(x,\eta)$$

with $\partial_l = \partial/\partial\eta_l$, and $p_l = (1/i)(\partial/\partial x_l)$. The multi-indices are defined in the usual manner by $\alpha = (\alpha_1,...,\alpha_N)$ for integers $\alpha_j \geqslant 0$, and $|\alpha| = \Sigma_{j=1}^N \alpha_j$, with similar definitions for β. The notation for derivatives is $\partial^{\alpha} = \partial_1^{\alpha_1} \cdots \partial_N^{\alpha_N}$ and $p^{\beta} = p_1^{\beta_1} \cdots p_N^{\beta_N}$. Here, m, β, and δ are real numbers satisfying $0 \leqslant \delta < \xi$. Equation (5.13) states that each $h_{ij}(x,\eta)$ belongs to the symbol class[37] $S_{\xi,\delta}^m$.

Let $h(x,\eta) = [h_{ij}(x,\eta)]$ be the matrix-valued symbol for $\widetilde{H}(x,p)$, and let $\lambda_1(x,\eta),...,\lambda_k(x,\eta)$ denote its eigenvalues. If $|\cdot|$ denotes a norm in the space of $k \times k$ matrices, we suppose that the following conditions are satisfied by $h(x,\eta)$: For $|\eta| > c_0 > 0$ and $x\in\mathbb{R}^N$ we have

(1) $|h_{(\beta)}^{(\alpha)}(x,\eta)| \leqslant C_{\alpha\beta}|h(x,\eta)|(1 + |\eta|)^{-\xi|\alpha| + \delta|\beta|}$

(hypoellipticity),

(2) $\lambda_0(x,\eta) = \max_{1\leqslant j \leqslant k} \operatorname{Re}\lambda_j(x,\eta) < 0$,

(3) $\dfrac{|h(x,\eta)|}{|\lambda_0(x,\eta)|} = O((1 + |\eta|)^{(\xi - \delta)/(2k - \epsilon)}), \quad \epsilon > 0.$

We assume that $\widetilde{H}(x,p)$ is a self-adjoint generator of a unitary group, so that

$$U(t,0)\psi_0(x) = \exp[-it\widetilde{H}(x,p)]\psi_0(x) = \psi(x,t)$$

solves the Cauchy problem

$$i\frac{\partial\psi}{\partial t} = H(x,p)\psi, \psi(x,0) = \psi_0(x). \qquad (5.14)$$

Definition 5.1: We say that $Q(x,t,\eta,0)$ is a *symbol for the Cauchy* problem (5.14) if $\psi(x,t)$ may be represented as

$$\psi(x,t) = (2\pi)^{-N/2}\int_{\mathbb{R}^N} e^{i(x,\eta)}Q(x,t,\eta,0)\hat{\psi}_0(\eta)d\eta. \quad (5.15)$$

It suffices to assume that ψ_0 belongs to the Schwartz space $\mathscr{S}(\mathbb{R}^N)$, which is contained in the domain of $\widetilde{H}(x,p)$, in order that (5.15) makes sense.

Following Shishmarev,[10] and using the theory of Fourier integral operators, we define an operator-valued kernel for $U(t,0)$ by

$$\widetilde{K}(x,t;y,0) = (2\pi)^{-N/2}\int_{\mathbb{R}^N} e^{i(x-y,\eta)}Q(x,t,\eta,0)d\eta,$$

so that

$$U(t,0)\psi_0(x) = \psi(x,t) = \int_{\mathbb{R}^N}\widetilde{K}(x,t;y,0)\psi_0(y)\,dy. \quad (5.16)$$

The following results are due to Shishmarev.[10]

Theorem 5.5: Suppose $\widetilde{H}(x,p)$ is a self-adjoint generator of a strongly continuous unitary group with a domain which is dense in $L^2(\mathbb{R}^N)$ and contains $\mathscr{S}(\mathbb{R}^N)$, such that conditions (1)–(3) are satisfied. Then there exists precisely one symbol $Q(x,t,\eta,0)$ for the Cauchy problem (5.14).

Theorem 5.6: Suppose one replaces condition (3) in Theorem 5.5 by the condition

(3′) $\dfrac{|h(x,\eta)|}{|\lambda_0(x,\eta)|} = O((1 + |\eta|)^{(\xi - \delta)/(3k - 1 - \epsilon)}),$

$\epsilon > 0, \quad |\eta| > c_0.$

Then the symbol $Q(x,t,\eta,0)$ of the Cauchy problem (5.14) has the following asymptotic behavior as $t \to 0$:

$$Q(x,t,\eta,0) = \exp[-ith(x,\eta)] + o(1),$$

uniformly for $x,\eta\in\mathbb{R}^N$.

Now, using Theorem 5.6 we see that under the strengthened condition (3′) the kernel $\widetilde{K}(x,t;y,0)$ satisfies

$$\widetilde{K}(x,t;y,0) = \int_{\mathbb{R}^N}\exp[i\{(x-y,\eta) - th(x,\eta)\}]\frac{d\eta}{(2\pi)^N} + \int_{\mathbb{R}^N}\exp[i(x-y,\eta)]\frac{d\eta}{(2\pi)^N}\,o(1).$$

We now apply the results discussed earlier in this section to construct the path integral associated with $\widetilde{H}(x,p)$. The group property of $U(t,0)$ insures that $\widetilde{K}$ has the reproducing property expressed by the Chapman–Kolmogorov equation. In our time-ordered version, we obtain

$$K_\tau(x,t;y,0) = \int_{\mathbb{R}^{N}(\tau)}\exp[i\{(x-y,\eta) - th_\tau(x,\eta)\}] \times\frac{d\eta}{(2\pi)^N} + o(1).$$

This representation leads to the Feynman phase space version of the path integral.

We can now obtain more general path integrals than (5.12) by replacing (5.5) by (5.16). It follows from Theorems 5.4–5.6 that path integrals exist which are generalizations of (5.12). These new path integrals correspond, of course, to Hamiltonian operators which are perturbations of the operators described in Theorems 5.5 and 5.6, rather than to Hamiltonians which are perturbations of Laplacians. These path integrals constitute a very large class which contain most integrals of interest in mathematical physics.

VI. PERTURBATION EXPANSIONS

In this section we discuss the Feynman–Dyson operator calculus for $U(t, -T)$. It is shown that the corresponding perturbation expansions do not converge in general, but are "asymptotic in the sense of Poincaré" in the sense used in the theory of semigroups.[11] On the other hand, if we assume that the semigroups possess certain holomorphy properties, then the perturbation series converge. Previous investigations of these perturbation expansions have been confined to the interaction representation in the framework of nonrelativistic scattering by time-dependent potentials[38] and external field problems in quantum field theory.[39]

Our results of this section pertaining to the asymptotic nature of these perturbation expansions affirms a well-known conjecture of Dyson[12] made in the context of the special case of the renormalized perturbation expansions in quantum electrodynamics on the basis of a simple physical argument. Although presently, many people *believe* quantum electrodynamics should be formulated in a Hilbert space with an indefinite metric (see, e.g., Ref. 40 and the works cited therein), Dyson made no such assumptions. In our concluding remarks to this section, we make explicit our basic assumptions and argue that they certainly cover conditions that physicists believe QED should satisfy.

Consider the infinite tensor product Hilbert space $V = \hat{\otimes}_{s\in J}\mathscr{H}(s)$ of Sec. II, where $J = [-T,T]$, $\mathscr{H}(s) = \mathscr{H}$ for each $s\in J$, and $\mathscr{H}$ denotes a fixed abstract separable Hilbert space. For a family $\{\hat{H}(t): t\in J\}$ of densely defined strongly continuous self-adjoint operators on $\mathscr{H}$, the corresponding time-ordered family $\{H(t): t\in J\}$ is defined on V by (2.4). Let $U(t,-T)$ denote the corresponding time-evolution operator whose existence is guaranteed by Theorem 3.2.

Let

$$Q(t,-T) = -i\int_{-T}^{t} H(s)ds$$

denote the time-ordered integral of the family $\{-iH(t): t\in J\}$. Then the closure of $Q(t,-T)$, which we will also denote by $Q(t,-T)$, generates the strongly c-continuous unitary group $U(t,-T) = \exp[Q(t,-T)]$ on V. We also have the following.

Theorem 6.1: Suppose $\phi\in D(H^N(s))$ for $-T\leqslant s\leqslant t$. Then $U(t,-T)\phi$ can be written in the form

$$U(t,-T)\phi = \sum_{k=0}^{N-1}\frac{1}{k!}(Q(t,-T))^k\phi + R_N(t,-T)\phi, \tag{6.1}$$

with the following representations for the remainder term:

$$R_N(t,-T)\phi = \int_0^1 dv(1-v)^{N-1}\exp[vQ(t,-T)] \times\frac{(Q(t,-T))^N}{(N-1)!}\phi, \tag{6.2}$$

and

$$R_N(t,-T)\phi = (-i)^N\int_{-T}^{t}d\tau_N\cdots\int_{-T}^{\tau_2}d\tau_1 \times H(\tau_N)\cdots H(\tau_1)U(\tau_1,-T)\phi. \tag{6.3}$$

Proof: It follows from a result of Hille and Phillips (Ref. 11, p. 354) that (6.1) holds with the remainder term given by (6.2). The equality of the latter with (6.3) is a consequence of the following result, which establishes a Fubini-type theorem for the Feynman–Dyson operator calculus.

Lemma 6.1: For any $N = 1,2,\ldots$, we have

$$\frac{1}{N!}\left[\int_{-T}^{t}H(\tau)d\tau\right]^N = \int_{-T}^{t}d\tau_N\int_{-T}^{\tau_N}d\tau_{N-1}\cdots\int_{-T}^{\tau_2}d\tau_1\, H(\tau_N)\cdots H(\tau_1).$$

Proof: Recall that the bounded operators

$$H_z(\tau) = [\exp(zH(\tau)) - I]/z, \quad z>0,$$

converge as $z\downarrow 0$ to $H(\tau)$ on $D(H(\tau))$ uniformly in τ on compact sets. We can therefore, without loss in generality, assume that $H(\tau)$ is bounded for each τ. The proof can then be completed by a bounded operator version of the usual integration by parts procedure for functions.

In the remainder of this section we discuss the problem of approximating the various terms in the expansion (6.1). For this purpose we use the form (6.2) for the remainder term.

Using the fact that $Q(t,-T)$ generates the strongly c-continuous unitary group $U(t,-T)$, we find from the theory of semigroups[11,41] that

$$P_z(t,-T) = (\exp[zQ(t,-T)] - I)/z, \quad z>0,$$

converges to $Q(t,-T)$ on $D(Q(t,-T))$ as $z\downarrow 0$. More generally, we have the following.

Lemma 6.2: Fix some $r\in\{1,2,\ldots\}$ and take $f\in D(\{Q(t,-T)\}^r)$. Then

$$\operatorname*{s-lim}_{z\downarrow 0}\{P_z(t,-T)\}^r f = \{Q(t,-T)\}^r f.$$

Proof: From p. 99 of Ref. 41 we have

$$(P_z^r - Q^r)\phi = \frac{1}{r!}\sum_{j=1}^{r}(-1)^{r-j}j^r\binom{r}{j} \times\left\{\frac{r!}{(jz)^r}\left[e^{jzQ}\phi - \sum_{k=0}^{r-1}\frac{(jz)^k}{k!}Q^k\phi\right] - Q^r\phi\right\},$$

so that

$$\|(P_z^r - Q^r)\phi\| \leqslant \sup_{0\leqslant u\leqslant rz}\|(e^{uQ} - I)Q^r\phi\|,$$

from which the proof readily follows.

Let us now define the bounded operators

$$U_z(t,-T) = \exp[P_z(t,-T)] = \sum_{k=0}^{N-1}\frac{[P_z(t,-T)]^k}{k!} + R_N^z,$$

where

$$R_N^z(t,-T) = \int_0^1 dv(1-v)^{N-1}\exp[vP_z(t,-T)] \times\frac{[P_z(t,-T)]^N}{(N-1)!}. \tag{6.4}$$

The boundedness of these operators follows from the estimates,

$$\|\{P_z(t,-T)\}^r\| \leqslant (2/z)^r, \quad r = 1,2,\ldots, \tag{6.5}$$

which are, in turn, consequences of the fact that $Q(t,-T)$ generates a contractive semigroup.

Now we have the following Theorem.

Theorem 6.2:

(a) $\operatorname*{s-lim}_{z\downarrow 0} U_z(t,-T) = U(t,-T)$,

(b) $\operatorname*{s-lim}_{z\downarrow 0} R_N^z(t,-T)\phi = R_N(t,-T)\phi, \quad \phi\in D(\{Q(t,-T)\}^N)$.

Proof: (a) follows from the fact that $U(t, -T)$ is a strongly c-continuous unitary group on V and Hille's first exponential formula (see, e.g., Ref. 41, Theorem 1.2.2).

To prove (b) we write, using (6.2) and (6.4),

$$(R_N - R_N^z)\phi = \int_0^1 dv \frac{(1-v)^{N-1}}{(N-1)!} [e^{vQ}Q^N - e^{vP_z}P_z^N]\phi$$

so that

$$\|(R_N - R_N^z)\phi\| \leqslant \frac{1}{N!} \sup_{v\in[0,1]} \|(e^{vQ}Q^N - e^{vP_z}P_z^N)\phi\|$$

$$\leqslant \frac{1}{N!} \sup_{v\in[0,1]} [\|(e^{vQ} - e^{vP_z})Q^N\phi\| + \|e^{vP_z}(Q^N - P_z^N)\phi\|] . \tag{6.6}$$

For the first term on the right-hand side of (6.6) we use the fact that, by Theorem 1.2.2 of Ref. 41,

$$\|(e^{vQ} - e^{vP_z})Q^N\phi\| \to 0 \quad \text{as } z\downarrow 0 ,$$

for $\phi\in D(Q^N)$ uniformly with respect to $v\in[0,1]$. The vanishing of the remaining term in (6.6) as $z\downarrow 0$ follows from Lemma 6.2 coupled with the estimate

$$\|\exp[vP_z(t, -T)]\| \leqslant 1 , \tag{6.7}$$

which in turn follows from Hille's first exponential formula and the fact that $U(t, -T)$ is unitary. ||

We see from (6.4) that R_N^z is a bounded operator, and we find with the help of (6.5) and (6.7),

$$\|R_N^z\| \leqslant (1/N!)(2/z)^N .$$

Now, using this estimate and Theorem 6.2, we obtain an estimate for the remainder term of the perturbation series:

$$\|R_N\phi\| \leqslant \|R_N^z\phi\| + \|(R_N - R_N^z)\phi\|$$

$$\leqslant (1/N!)(2/z)^N\|\phi\| + \epsilon , \tag{6.8}$$

where, for N fixed and given $\epsilon > 0$, we choose $z_0 > 0$ sufficiently small that

$$\|(R_N - R_N^z)\phi\| < \epsilon , \quad \phi\in D(Q^N) ,$$

for $z < z_0$. However, it does not follow from the estimate (6.8) that $R_N\phi \to 0$ as $N\to\infty$ because z_0 cannot be chosen independently of N. Thus the perturbation series does not converge.

It does follow from the above results, however, that the perturbation expansion is "asymptotic in the sense of Poincaré." Compare the definition of this concept on p. 487 of Ref. 11 with Theorem 2.2.13 of Ref. 41.

We can use techniques similar to those discussed in the present section to obtain results for the perturbation series for the scattering operator, since $\lim_{T\to\infty} U_\lambda[T,t] = U[\infty,t]$ and $\lim_{T\to\infty} U_\lambda[t, -T] = U[t, -\infty]$; $S[\infty, -\infty] = U[\infty,t]U[t, -\infty]$.

We now make a few remarks concerning the convergence of the perturbation expansions when the corresponding semigroup is holomorphic. The semigroup that we have been considering is $U(t, -T) = \exp\{Q(t, -T)\}$, which we now rewrite in the form

$$U(t, -T) = \exp[\tau\{Q(t, -T)/\tau\}]$$

in terms of a parameter τ. We say that $U(t, -T)$ is *holomorphic* if, as a function of τ, it can be continued into a neighborhood of unity in the complex τ-plane (compare with Ref. 32, p. 254). It then follows from the general theory of semigroups that the perturbation series (6.1) converges. The proof is similar to that of Theorem 1.1.11 in Ref. 41.

In conclusion, it is important to note that our only assumptions are (1) $\widetilde{H}(t) = \int_{\mathbb{R}} \widetilde{H}(t,\mathbf{x})d\mathbf{x}$ is the generator of a unitary group on $\mathscr{H}$ for each t [where $\widetilde{H}(t,\mathbf{x})$ is the field energy density on $\mathbb{R}^n$]; (2) the set of operators $\{\widetilde{H}(t)\,|\,t\in J\}$ is strongly continuous with common dense domain; and (3) $\mathscr{H}$ is a separable Hilbert space. It could be argued that the assumption of a common dense domain for the Hamiltonians is too strong for any formulation of QED; however, this assumption is not necessary for our theory to apply. This will be taken up at a later time when we consider applications to nonlinear formulations.

VII. CONCLUDING REMARKS

In this paper we have used an algebraic approach to time-ordered operators based upon von Neumann's infinite tensor product Hilbert spaces to define path integrals which appear to include most cases of interest in mathematical physics. We have proved that there exists a one-to-one correspondence between path integrals and semigroups which are integral operators defined by a kernel. The reproducing property of the kernel is a consequence of the semigroup property.

The generality of our construction is intimately connected with the fact that our tensor product Hilbert spaces are constructed using an *abstract* separable Hilbert space as a base. This allows application to many different physical problems according to different choices of this base Hilbert space. We will consider some of these applications in future work.

We have shown that our treatment is a generalization of the customary approach to time-ordered operators and path integration by means of product integrals. Moreover, when Hamiltonians which are sums of two parts (in a certain well-defined sense) are considered, our results do not depend upon the domains of the latter operators.

We have also shown that our approach leads to unique solutions to the Cauchy problem for Schrödinger equations with time-dependent Hamiltonians. This is clearly of interest for mathematics as well as physics, since one is concerned here with linear time-evolution equations.

We have advanced the point of view that it is unnatural to try to force path integrals into a description by means of countably additive measures. The viewpoint has been expressed that the theory of integration, rather than measure theory, is the appropriate vehicle for a *general* formulation of path integration. Thus, although path integrals can be written in terms of countably additive measures in certain special cases, this is not the situation in general.

We have also discussed perturbation expansions for time-evolution operators. It has been shown that these expansions generally do not converge, but are asymptotic in a certain well-defined sense. On the other hand, these series converge when the semigroups possess suitable holomorphy properties. It should also be noted that our approach shows that the general belief expressed in Ref. 39, to the effect that the Dyson expansion can only hold with $H(t)$ bounded, is not quite correct (see p. 283 of that reference).

[1]J. E. Humphreys, *Introduction to Lie Algebras and Representation Theory* (Springer, Heidelberg, 1972); J. Diximier, *Algèbres Enveloppantes* (Gauthier–Villars, Paris, 1974).

[2]R. P. Feynman, Phys. Rev. **84**, 108 (1951).

[3]I. Fujiwara, Prog. Theor. Phys. **7**, 433 (1952).

[4]W. L. Miranker and B. Weiss, SIAM Rev. **8**, 224 (1966).

[5]H. Araki, Ann. Sci. École Norm. Sup. **6**, 67 (1973).

[6]E. Nelson, in *Functional Analysis and Related Fields*, edited by F. E. Browder (Springer, Heidelberg, 1970), pp. 172–187.

[7]V. P. Maslov, *Operational Methods* (Nauka, Moscow, 1973) [English translation (Mir, Moscow, 1976)].

[8]R. P. Feynman, Rev. Mod. Phys. **20**, 367 (1948).

[9]V. P. Maslov and I. A. Shishmarev, J. Soviet Math. **13**, 81 (1980).

[10]I. A. Shishmarev, Sov. Math. Izv. **20**, 577 (1983).

[11]E. Hille and R. S. Phillips, *Functional Analysis and Semigroups* (Am. Math. Soc., Providence, RI, 1957).

[12]F. J. Dyson, Phys. Rev. **85**, 631 (1952).

[13]J. von Neumann, Comp. Math. **6**, 1 (1938).

[14]A. Guichardet, Ann. Sci. École Norm. Sup. **83**, 1 (1966).

[15]T. L. Gill, Trans. Am. Math. Soc. **266**, 161 (1981). On p. 163, the last line of definition 1.6 should read "...where $\{J_t\}$ is a fixed family of duality maps for $\{\phi_t\}$."

[16]T. L. Gill, Trans. Am. Math. Soc. **279**, 617 (1983). On p. 617, paragraph 3 should read: "...We assume that unless otherwise stated, all operators of the form $\widetilde{A}(s)$ are strongly space continuous operators in the sense of definition 2.6 in I, while... ." On p. 618, Theorem 1.1 should read (1) $\text{s-lim}_{\lambda\to\infty} Q^z_\lambda[t,0] = \text{s-lim}_{\lambda\to\infty} \overline{Q}^z_\lambda[t,0]$ exists and $Q^z[t,0] = Q^z[t,s] + Q^z[s,0]$, $0\leqslant s<t$. On p. 619, the eighth line from the bottom of the page should read $\lambda\geqslant\lambda_2$. The series in the seventh line from the bottom should be truncated for sufficiently large N so that z_0 in the sixth line from the bottom may be chosen independent of n and τ_j. On p. 630, at the end of the second paragraph the last line should read: "In our approach we assume no special domain relationship between domains other than that required for strong space continuity."

[17]F. Langouche, D. Roekaerts, and E. Tirapegui, *Functional Integration and Semiclassical Expansions* (Reidel, Dordrecht, 1982).

[18]H. Tanabe, *Equations of Evolution* (Pitman, London, 1979).

[19]S. T. Hu, *Elements of Modern Algebra* (Holden–Day, San Francisco, 1965).

[20]J. D. Dollard and C. N. Friedman, *Product Integration with Applications to Differential Equations, Encyclopedia of Mathematics and Its Applications*, Vol. 10 (Addison–Wesley, Reading, MA, 1979).

[21]T. L. Gill and W. W. Zachary, "General theory of linear evolutions," in preparation.

[22]A. Truman, J. Math. Phys. **19**, 1742 (1978); C. DeWitt-Morette, A. Maheshwari, and B. Nelson, Phys. Rep. **50**, 255 (1979); D. Fujiwara, Duke Math. J. **47**, 559 (1980); *Feynman Path Integrals, Lecture Notes in Physics*, Vol. 106, edited by S. Albeverio, P. Combe, R. Høegh-Krohn, G. Rideau, M. Sirugue-Collin, M. Sirugue, and R. Stora (Springer, Heidelberg, 1979); *Functional Integration, Theory and Applications*, edited by J. P. Antoine and E. Tirapegui (Plenum, New York, 1980).

[23]S. A. Albeverio and R. J. Høegh-Krohn, *Mathematical Theory of Feynman Path Integrals, Lecture Notes in Mathematics*, Vol. 523 (Springer, Heidelberg, 1976).

[24]R. H. Cameron and D. A. Storvick, Mem. Am. Math. Soc. **288** (1983).

[25]R. H. Cameron, J. Math. Phys. (Cambridge, Mass.) **39**, 126 (1960).

[26]J. Kurzweil, Czech. Math. J. **7**, 418 (1957); R. Henstock, *Theory of Integration* (Butterworths, London, 1983); *Linear Analysis* (Butterworths, London, 1967); R. M. McLeod, *The Generalized Riemann Integral* (Mathematical Association of America, Washington, D.C., 1980); V. G. Chelidze and A. G. Dzhvarshishvili, *Theory of the Denjoy Integral and Some of its Applications* (Izdat. Tbilisi U., Tbilisi, 1978) (in Russian).

[27]R. Henstock, Proc. London Math. Soc. **27**, 317 (1973).

[28]N. Wiener, A. Siegel, W. Rankin, and W. T. Martin, *Differential Space, Quantum Systems, and Prediction* (MIT, Cambridge, MA, 1966).

[29]D. L. Pursey, Nucl. Phys. **53**, 174 (1964).

[30]L. L. Foldy and S. A. Wouthuysen, Phys. Rev. **78**, 29 (1950).

[31]T. Kato, *Perturbation Theory for Linear Operators* (Springer, Heidelberg, 1966).

[32]K. Yosida, *Functional Analysis* (Springer, Heidelberg, 1980), 6th ed.

[33]H. F. Trotter, Proc. Am. Math. Soc. **10**, 545 (1959).

[34]P. R. Chernoff, Mem. Am. Math. Soc. No. 140 (1974).

[35]T. Kato, Proc. Jpn. Acad. **50**, 694 (1974); T. Kato and K. Masuda, J. Math. Soc. Jpn. **30**, 169 (1978).

[36]L. S. Schulman, *Techniques and Applications of Path Integration* (Wiley, New York, 1981).

[37]For a discussion of these symbol classes, see, e.g., F. Treves, *Introduction to Pseudodifferential and Fourier Integral Operators: Pseudodifferential Operators* (Plenum, New York, 1980), Vol. 1.

[38]M. Reed and B. Simon, *Methods of Modern Mathematical Physics II: Fourier Analysis, Self-Adjointness* (Academic, New York, 1975), Sec. X.12.

[39]M. Reed and B. Simon, *Methods of Modern Mathematical Physics III: Scattering Theory* (Academic, New York, 1979), Sec. XI.15.

[40]F. Strocchi and A. S. Wightman, J. Math. Phys. **15**, 2198 (1974).

[41]P. L. Butzer and H. Berens, *Semigroups of Operators and Approximation* (Springer, New York, 1967).

A uniqueness theorem for an inverse Sturm–Liouville problem

Joyce R. McLaughlin
Mathematical Sciences Department, Rensselaer Polytechnic Institute, Troy, New York 12181
William Rundell
Department of Mathematics, Texas A&M University, College Station, Texas 77843

(Received 8 December 1986; accepted for publication 11 March 1987)

A new uniqueness theorem is established for the inverse Sturm–Liouville problem. It is shown that the measurement of a particular eigenvalue for an infinite set of different boundary conditions is sufficient to determine the unknown potential.

I. INTRODUCTION

In this paper a new uniqueness theorem will be established for the inverse Sturm–Liouville problem. We will show that the measurement of a particular set of eigenvalues is sufficient to determine the unknown potential.

Specifically, let us consider the eigenvalue problem

$$\begin{aligned} &y''(x) + (\lambda - q(x))y = 0, \quad 0<x<1, \\ &y(0) = 0, \quad y'(1) + \beta y(1) = 0, \end{aligned} \tag{1.1}$$

where $q(x) \in L^2(0,1)$ has to be determined in addition to $y(x)$.

In the typical formulation of the inverse Sturm–Liouville problem one seeks to recover both $q(x)$ and the constant β, by giving the eigenvalues $\lambda_j(q,\beta)$, for $j = 0,1,2,\ldots$ for (1.1), together with another piece of spectral data. These data can take several forms, leading to many versions of the problem, some of which can be shown to be equivalent.

In the Gelfand–Levitan formulation,[1] if $y_j(x,q,\lambda_j)$ denotes the jth eigenfunction of (1.1), then one gives, in addition to the eigenvalues λ_j, the values of the *norming constants* ρ_j,

$$\rho_j = \frac{\|y_j(\cdot,q,\lambda_j)\|_2}{|y_j'(0,q,\lambda_j)|}, \quad j = 0,1,2,\ldots. \tag{1.2}$$

Here $\|\cdots\|_2$ denotes the $L^2(0,1)$ norm, and one usually gives the normalization $y_j'(0,q,\lambda_j) = 1$.

Another possible set of norming constants that leads to a unique determination of both $q(x)$ and β is

$$\rho_j = \log\left\{(-1)^j \frac{y_j'(1,q,\lambda_j)}{y_j'(0,q,\lambda_j)}\right\}, \quad j = 0,1,2,\ldots, \tag{1.3}$$

as given by Dahlberg and Trubowitz.[2] (See also Levinson[3] and Isaacson–Trubowitz.[4])

If the boundary condition at $x = 1$ is changed, say to the Dirichlet condition $y(1) = 0$, and the corresponding set of eigenvalues $\tilde{\lambda}_k(q)$, $k = +,1,2,\ldots$, is also given, then at most one pair (q,β) can satisfy this data. This is the classical two spectrum version of the inverse Sturm–Liouville problem studied by Borg.[5] See also Refs. 6 and 7. We will use this last result in the proof of our theorem.

Finally, if it is known *a priori* that $q(x)$ is symmetric about the midpoint of the interval, that is, $q(x) = q(1-x)$, then this information together with a knowledge of the eigenvalues $\lambda_j(q,\beta)$ is sufficient to determine $q(x)$ and β.[6,8] Further information on these results may be obtained from the survey article, Ref. 9.

Recently, other kinds of spectral data have been considered and uniqueness results have been established. In particular, it has been shown by McLaughlin,[10] that the potential $q(x) \in L^2(0,1)$ in the Sturm–Liouville problem

$$y'' + (\lambda - q)y = 0, \quad y(0) = y(1) = 0 \tag{1.4}$$

is uniquely determined by knowledge of the position of one node (or zero) of each eigenfunction.

In the result to be given here we suppose that we can measure a single eigenvalue, for example the second eigenvalue, for a fixed unknown potential $q(x)$ but be able to do so for a countable number of different boundary conditions. That is, in the notation above we suppose that for a fixed j we can measure $\lambda_j(q,\beta_k)$ for distinct β_k, $-\infty < \beta_k < \infty$, $k = 1,2,\ldots$. We establish that at most one potential $q(x)$ is determined by these measurements.

II. THE UNIQUENESS THEOREM

In this section we will establish the result that at most one potential $q(x)$ can be determined from $\lambda_j(q,\beta_k)$, where j is fixed and β_k, $k = 1,2,\ldots$, are all distinct.

We require two preliminary lemmas, which we shall state without proof. This first is a well-known oscillation theorem for the eigenvalues of a Sturm–Liouville problem; the second is a precise statement of the two spectrum inverse Sturm–Liouville problem mentioned in the Introduction.

Lemma 1: Let $q(x) \in L^2(0,1)$. Then

$$\tilde{\lambda}_j(q) < \lambda_j(q,\beta) < \tilde{\lambda}_{j+1}(q) \tag{2.1}$$

for all β, $-\infty < \beta < \infty$.

Lemma 2: If for two values of β in (1.1), say β_1 and β_2, and for q_1 and $q_2 \in L^2(0,1)$, the eigenvalues of problem (1.1) satisfy

$$\begin{aligned} &\lambda_j(q_1,\beta_1) = \lambda_j(q_2,\beta_1), \quad j = 0,1,2,\ldots, \\ &\lambda_k(q_1,\beta_2) = \lambda_k(q_2,\beta_2), \quad k = 0,1,2,\ldots, \end{aligned}$$

then $q_1 = q_2$ a.e.

The uniqueness theorem is as follows.

Theorem: Let $q_1(x)$ and $q_2(x) \in L^2(0,1)$. Fix j, a positive integer. Suppose that β_k for $k = 1,2,\ldots$ are distinct real numbers and

$$\lambda_j(q_1,\beta_k) = \lambda_j(q_1,\beta_k), \quad k = 1,2,\ldots, \tag{2.2}$$

then $q_1(x) = q_2(x)$ a.e.

Proof: For each λ we let $y_2(x,q_i,\lambda)$ be the solution of the initial value problem

$$y'' + (\lambda - q_i)y = 0, \quad y(0) = 0, \quad y'(0) = 1. \tag{2.3}$$

 0022-2488/87/071471-02$02.50

Then for each λ we have the Sturm identity

$$
\begin{aligned}
0 &= y_2(x,q_1,\lambda)\{y_2''(x,q_2,\lambda) + (\lambda - q_2)y_2(x,q_2,\lambda)\} \\
&\quad - y_2(x,q_2,\lambda)\{y_2''(x,q_1,\lambda) + (\lambda - q_1)y_2(x,q_1,\lambda)\} \\
&= (q_1 - q_2)y_2(x,q_1,\lambda)y_2(x,q_2,\lambda) \\
&\quad + \{y_2(x,q_1,\lambda)y_2'(x,q_2,\lambda) - y_2(x,q_2,\lambda)y_2'(x,q_1,\lambda)\}'.
\end{aligned} \tag{2.4}
$$

We shall denote by $\tilde{\lambda}_j(q_i)$ the eigenvalues of the problem (1.1) when homogeneous Dirichlet boundary conditions are imposed at $x = 1$, that is, $y_2(1,q_i,\tilde{\lambda}_j) = 0$.

We now use the simplified notation

$$\mu_k = \lambda_j(q_1,\beta_k) = \lambda_j(q_2,\beta_k), \quad k = 1,2,\ldots .$$

Setting $\lambda = \mu_k$ in (2.3) and integrating from 0 to 1, we see that the term in brackets on the final line has value zero at $x = 0$ and $x = 1$ for each $k = 1,2,\ldots$, and we are left with

$$\int_0^1 (q_1 - q_2)y_2(x,q_1,\mu_k)y_2(x,q_2,\mu_k)dx = 0, \quad k = 1,2,\ldots . \tag{2.5}$$

It is now observed from Lemma 1 that the sequence $\{\mu_k\}_{k=1}^{\infty}$ forms a bounded set on the real line and consequently has at least one finite accumulation point. Further, since for fixed x, $y_2(x,q_i,\lambda)$ is an analytic function of λ, we can show that

$$F(\lambda) = \int_0^1 (q_1 - q_2)y_2(x,q_1,\lambda)y_2(x,q_2,\lambda)dx \tag{2.6}$$

is also an analytic function of λ. However, since $F(\lambda) = 0$ at an infinite set of values of λ with a finite accumulation point, then

$$F(\lambda) \equiv 0 \tag{2.7}$$

for all complex λ.

We now seek to show that all of the eigenvalues of (1.1) with $\beta = 0$ and all of the eigenvalues of (1.4) are the same for the function $q(x)$ set equal to $q_1(x)$ or $q_2(x)$, that is, we will show that

$$\tilde{\lambda}_n(q_1) = \tilde{\lambda}_n(q_2), \quad n = 1,2,\ldots, \tag{2.8}$$

$$\lambda_m(q_1,0) = \lambda_m(q_2,0), \quad m = 0,1,2,\ldots . \tag{2.9}$$

From Lemma 2 we would then be able to conclude that $q_1 = q_2$ a.e.

In order to prove (2.8) and (2.9), we return to the identity (2.4) and recall that when $\lambda = \lambda_m(q_1,0)$, then $y_2(1,q_1,\lambda_m(q_1,0)) \neq 0$, while $y_2'(1,q_1,\lambda_m(q_1,0)) = 0$. Integrating (2.4) from 0 to 1 when $\lambda = \lambda_m(q_1,0)$ and using (2.7), we must have

$$y_2'(1,q_2,\lambda_m(q_1,0)) = 0, \quad m = 1,2,\ldots .$$

This implies that each $\lambda_m(q_1,0)$ is an eigenvalue for (1.1) and (1.2) when (q,β) is chosen to be $(q_2,0)$. From the asymptotic forms that $\lambda_m(q_2,0)$ must satisfy, it follows that (2.8) holds.

Similarly set $\lambda = \tilde{\lambda}_n(q_1)$ in the identity (2.4) and note that $y_2(1,q_1,\tilde{\lambda}_n(q_1)) = 0$ while $y_2'(1,q_1,\tilde{\lambda}_n(q_1)) \neq 0$. Then, again using (2.4) and integrating from 0 to 1, we must have

$$y_2'(1,q_2,\tilde{\lambda}_n(q_1)) = 0, \quad n = 1,2,\ldots .$$

The proof is now complete.

Remark: The uniqueness theorem also holds when $j = 0$, provided that we have the existence of an M, $M > -\infty$ such that $M < \lambda_0(q_i,\beta_k)$, for all $k = 1,2,\ldots$.

ACKNOWLEDGMENTS

The work of J. McLaughlin was supported in part by the Office of Naval Research under Grant No. ONR N00014-84-5-0516 and in part by the Air Force Office of Scientific Research under Grant No. AFOSR 86-0180.

[1] I. M. Gelfand and B. M. Levitan, "On the determination of a differential equation from its spectrum," Am. Math. Soc. Trans. **1**, 253 (1955).

[2] B. E. J. Dahlberg and E. Trubowitz, "The inverse Sturm–Liouville problem III," Commun. Pure Appl. Math. **37**, 255 (1984).

[3] N. Levinson, "The inverse Sturm–Liouville problem," Mat. Tidsskr. **3**, 25 (1949).

[4] E. L. Isaacson and E. Trubowitz, "The inverse Sturm–Liouville problem I," Commun. Pure Appl. Math. **36**, 767 (1983).

[5] G. Borg, "Eine umkehrung der Sturm–Liouvilleschen eigenwertaufgabe," Acta Math. **75**, 1 (1946).

[6] H. Hochstadt, "The inverse Sturm–Liouville problem," Commun. Pure Appl. Math. **26**, 715 (1973).

[7] B. M. Levitan, "On the determination of an inverse Sturm–Liouville equation by two spectra," Am. Math. Soc. Trans. Ser. 2 **68**, 1 (1968).

[8] O. H. Hald, "The inverse Sturm–Liouville problem with symmetric potentials," Acta Math. **141**, 263 (1978).

[9] J. R. McLaughlin, "Analytical methods for recovering coefficients in differential equations from spectral data," SIAM Rev. **28**, 53 (1986).

[10] J. R. McLaughlin, "The inverse Sturm–Liouville problem using nodal positions as data" (to be published).

Ovsiannikov's method and the construction of partially invariant solutions

C. C. A. Sastri
Mathematical Sciences Institute, Cornell University, Ithaca, New York 14853 and Department of Mathematics, Statistics, and Computing Science, Dalhousie University, Halifax, Nova Scotia B3H 3J5, Canada

K. A. Dunn
Department of Mathematics, Statistics, and Computing Science, Dalhousie University, Halifax, Nova Scotia B3H 3J5, Canada

D. R. K. S. Rao
Department of Applied Mathematics, Andhra University, Waltair 530003, A.P., India

(Received 19 November 1985; accepted for publication 4 March 1987)

A plausibility argument presented by the first two authors in an earlier paper [J. Math. Phys. **26**, 3042 (1985)] concerning the existence of partially invariant solutions for some equations of the Fokker–Planck type is made precise by the explicit construction of one such solution. In the process a substantial simplification of Ovsiannikov's method for finding partially invariant solutions is achieved. In addition, the class of partially invariant solutions obtained by Ovsiannikov for the equations of transonic flow of a gas is enlarged.

I. INTRODUCTION

In a recent paper[1] a plausibility argument for the existence of a certain class of solutions for some equations of the Fokker–Planck type was presented. That argument is made precise in the present paper by the construction of one such solution, thereby showing that the class referred to is not empty. At the same time, the method of Ovsiannikov, used in that construction, is substantially simplified. Finally, the class of solutions of the same type as above, obtained by Ovsiannikov for the nonlinear equations of transonic flow of a gas, is enlarged.

The notation and definitions used here are the same as those of Refs. 1 and 2, which in turn are the same as those of Ovsiannikov.[3,4] The reader is referred to the work of Ovsiannikov[3,4] for a full discussion of the ideas involved.

The method used here is that of Ovsiannikov, which differs from the standard method of finding similarity solutions in that it regards a partial differential equation as a system of first-order equations rather than a single higher-order equation. To be specific, let us consider one of the equations discussed in some detail in Ref. 1, namely, the one-dimensional heat equation

$$\frac{\partial u}{\partial t}=\frac{\partial^2 u}{\partial x^2} \quad (-\infty<x<\infty, \quad t>0). \tag{1}$$

In Ovsiannikov's method one considers, instead of Eq. (1), the equivalent system

$$u_x=v, \quad u_t=v_x, \tag{2}$$

and proceeds to construct its group of Lie symmetries. It can be shown[3,4] that Eq. (1) and the system (2) have the same group of Lie symmetries and that it is infinite dimensional. The infinite-dimensional component arises from the fact that the heat equation, like any other linear homogeneous equation, is invariant under translations in u by solutions of the equation. If one takes the quotient of the full group by this infinite-dimensional component due to translation, which forms a normal subgroup, one obtains a six-dimensional group G whose infinitesimal generators are

$$\begin{aligned}
X_1 &= \frac{\partial}{\partial t},\\
X_2 &= t\frac{\partial}{\partial t}+\frac{x}{2}\frac{\partial}{\partial x}-\frac{v}{2}\frac{\partial}{\partial v},\\
X_3 &= t^2\frac{\partial}{\partial t}+tx\frac{\partial}{\partial x}-\left(\frac{t}{2}+\frac{x^2}{4}\right)u\frac{\partial}{\partial u}\\
&\quad -\left(\frac{ux}{2}+\frac{3tv}{2}+\frac{x^2v}{4}\right)\frac{\partial}{\partial v},\\
X_4 &= \frac{\partial}{\partial x},\\
X_5 &= t\frac{\partial}{\partial x}-\frac{xu}{2}\frac{\partial}{\partial u}-\left(\frac{u}{2}+\frac{xv}{2}\right)\frac{\partial}{\partial v},
\end{aligned}$$

and

$$X_6=u\frac{\partial}{\partial u}+v\frac{\partial}{\partial v}.$$

Let H be a subgroup of G. A solution $f(t,x)$ of Eq. (1) is said to be invariant under H if the surface $u=f(t,x)$ is invariant under H. Solutions that are invariant with respect to (w.r.t.) some subgroup of G are simply called invariant or similarity solutions. A solution $f(t,x)$ is said to be partially invariant[3,4] w.r.t. H if the surface $u=f(t,x)$ is contained in some surface invariant under H.

The invariant solutions of Eq. (1) can be found, and have been found,[5] by the standard method. The question is, does Eq. (1) have partially invariant solutions w.r.t. some subgroup of G which are not invariant w.r.t. *any* subgroup of G, and if so, how does one construct them? It has been shown[1] that the standard method of finding similarity solutions cannot be used to construct them even if they exist; the presence of the extra variable v in the system (2), the "superfluous variable" in Ovsiannikov's terminology, turns out to

 0022-2488/87/071473-04$02.50

be essential. The question raised is answered in the affirmative in the present paper.

At this stage, a criterion for checking whether a given solution $f(t,x)$ is invariant or not w.r.t. a subgroup H would be useful. This is provided by the theorem on p. 31 of Ref. 3, which says that if X denotes the infinitesimal generator of H and if $F(t,x,u) \equiv u - f(t,x)$, then $f(t,x)$ is an invariant solution w.r.t. H if and only if $XF(t,x,u) = 0$ whenever $F(t,x,u) = 0$.

II. EXAMPLES

Let H denote the one-dimensional subgroup of G whose infinitesimal operator is $\alpha X_1 + \beta X_6$, $\alpha,\beta \neq 0$. Then it has been shown[1] that solutions partially invariant w.r.t. H are of the form

$$u = e^{\beta t/\alpha} f(\lambda,\mu), \tag{3}$$

where $\lambda = x$ and $\mu = ve^{-\beta t/\alpha}$ and f obeys the partial differential equation (PDE)

$$\begin{aligned} &\mu^2 f_{\mu\mu} + 2\mu f_\mu f_{\lambda\mu} - 2\mu f_\lambda f_{\mu\mu} \\ &\quad - 2 f_\lambda f_\mu f_{\lambda\mu} + f_\lambda^2 f_{\mu\mu} + f_\mu^2 f_{\lambda\lambda} \\ &\quad + (\beta/\alpha) u f_\mu^3 - (\beta/\alpha) f f_\mu^2 = 0. \end{aligned} \tag{4}$$

It is in the derivation of this PDE that Ovsiannikov's algorithm (as described in the example on p. 286 of Ref. 3) can be substantially simplified. Following that algorithm, Eq. (4) was derived in Ref. 1 by the imposition of the compatibility criteria $v_{tx} = v_{xt}$ and $f_{\mu\lambda} = f_{\lambda\mu}$ and of the condition that u,v must satisfy the system (2). The calculations, though straightforward, are tedious and take about six pages when written in longhand. The easier (and quicker) method still involves assuming that u,v satisfy the system (2) and that $f_{\lambda\mu} = f_{\mu\lambda}$ but dispenses with the assumption that the superfluous variable v satisfies the compatibility condition $v_{xt} = v_{tx}$; instead, it is assumed that $v_{xx} = v_t$. What this amounts to is a weakening of the assumption that u is thrice continuously differentiable, which implies that $v_{xt} = v_{tx}$, to the assumption that u is twice continuously differentiable, which implies that $v_{xx} = v_t$. Interestingly, this weakening of the assumption reduces the work involved in deriving the PDE for f by a factor of at least 6. This is true for any subgroup of G, but since Eq. (4) was derived earlier[1] using Ovsiannikov's algorithm, let us derive it using the easier method.

We have $u = e^{\beta t/\alpha} f(\lambda,\mu)$, where $\lambda = x$, $\mu = ve^{-\beta t/\alpha}$, so that $\lambda_x = 1$, $\lambda_t = 0$, $\mu_x = v_x e^{-\beta t/\alpha}$, $\mu_t = (v_t - v\beta/\alpha)e^{-\beta t/\alpha}$. So

$$\begin{aligned} u_x &= f_\mu v_x + f_\lambda e^{\beta t/\alpha}, \\ u_t &= f_\mu v_t - f_\mu v(\beta/\alpha) + (\beta/\alpha) f e^{\beta t/\alpha}, \\ u_{xx} &= (f_{\mu\lambda} + f_{\mu\mu} v_x e^{-\beta t/\alpha}) v_x + f_\mu v_{xx} \\ &\quad + (f_{\lambda\lambda} + f_{\lambda\mu} v_x e^{-\beta t/\alpha}) e^{\beta t/\alpha}. \end{aligned}$$

Use now of the conditions $u_t = u_{xx}$ and $v_t = v_{xx}$ as well as the relation $v_x = (1/f_\mu)(v - f_\lambda(v/\mu))$ gives

$$\begin{aligned} &f_{\mu\mu}(\mu - f_\lambda)^2 + 2 f_\mu f_{\lambda\mu}(\mu - f_\lambda) + f_\mu^2 f_{\lambda\lambda} \\ &\quad + (\beta/\alpha) f_\mu^2 (\mu f_\mu - f) = 0, \end{aligned} \tag{5}$$

provided $f_\mu \neq 0$. Equation (5) is just another form of Eq. (4). Although the above simplification is specific to the heat equation, for which the assumption $v_t = v_{xx}$ is meaningful, it seems reasonable to believe that similar assumptions, appropriate to the equation under consideration, would lead to a simplification of the Ovsiannikov method. Of course one could use computer software such as MAPLE and MACSYMA to carry out these calculations, but even when such packages are used, shortcuts would be helpful in reducing the amount of computing time used.

It is interesting to note a simple pattern that emerges in the case of the heat equation. For the subgroup with the generator $\alpha X_1 + \beta X_6$, we have

$$v = A v_x + B, \quad v_x = A v_t + C, \tag{6}$$

where $A = f_\mu$, $B = f_\lambda e^{\beta t/\alpha}$, $C = (\beta/\alpha) f e^{\beta t/\alpha} - f_\mu v(\beta/\alpha)$. Equations of the same form hold for different subgroups of G, with different A, B, and C. Use of the heat equation and the condition $v_{xx} = v_t$ leads to the equation

$$A(C - B_x) = A_x(v - B).$$

This equation contains Eq. (4).

Going back to Eq. (3), one way to find u is to solve Eq. (4) for f and use that solution in

$$v_x = (e^{\beta t/\alpha}/f_\mu)(ve^{-\beta t/\alpha} - f_\lambda) \tag{7}$$

and

$$v_t = v\left(\frac{1}{f_\mu^2} + \frac{\beta}{\alpha}\right) - \frac{e^{\beta t/\alpha}}{f^2}\left(f_\lambda + \frac{\beta}{\alpha} f f_\mu\right) \tag{8}$$

and substitute the resulting quantities in Eq. (3).

Two remarks concerning Eq. (4) are in order. First of all, since it is derived by a process of differentiation, Eq. (4) is necessary but not sufficient for u given by Eq. (3) to satisfy the system (2); it is conceivable–in fact it is easy to show–that solutions of Eq. (4), which contain arbitrary functions, will not in general be such that (s.t.) u, given by Eq. (3), satisfies the system (2) unless the arbitrary functions are chosen appropriately. Thus one may regard Eqs. (3) and (4) as a source for finding functions u, *some* of which turn out to be solutions of Eq. (2) as well. The second remark is that although Eq. (4) is substantially harder to solve than Eq. (2), it is not the general solution of Eq. (4) we are interested in but particular ones which lead to solutions of the specific form (3) of Eq. (2). In other words, the method of partially invariant solutions leads to the construction of *trial* solutions of Eq. (2).

It is easy to check that $f(\lambda,\mu) = e^{a\lambda} + c\mu$, where $a^2 = \beta/\alpha$ and c is an arbitrary constant, is a solution of Eq. (5). Choosing c so that $c \neq 0$ and $ac \neq 1$, one can construct the following solution of Eq. (1):

$$u(t,x) = g(t,x) + h(t,x) + k(t,x), \tag{9}$$

where

$$g(t,x) = (q/a) e^{\beta t/\alpha + ax}, \tag{10}$$

$$h(t,x) = Ae^{t/c^2} e^{x/c}, \tag{11}$$

and

$$k(t,x) = Be^{t/c^2} e^{-x/c}, \tag{12}$$

where A, B are arbitrary constants, and $q = p/(1/c^2 - \beta/\alpha)$, with $p = a/c^2 + \beta/\alpha c$.

Observe that each of g, h, k is a solution of Eq. (1); therefore so is $g + h + k$. Observe also that each of g, h, k is an invariant solution w.r.t. a subgroup of G consisting of simultaneous translations in t and x and stretching in the dependent variable. Nevertheless, we claim that $g + h + k$ is not an invariant solution w.r.t. any subgroup of G other than the identity group. Assuming that this claim has been established, one may ask, considering the simple nature of $g + h + k$, whether one cannot construct such noninvariant solutions by some *ad hoc* means using the standard approach. The answer to that is twofold. Our aim has been to show that the Ovsiannikov method yields, in a natural way, partially invariant solutions which are not invariant; the example being presented here shows that. Second, using the simplified Ovsiannikov method, the authors have derived for several subgroups of G the PDE's that lead to partially invariant solutions; these will be listed later.

It is sufficient to consider the action of G on the (t,x,u) space in order to check whether a given solution is invariant or not. Now any subgroup of G is generated by some linear combination of $\{X_i\}$, $1 \leqslant i \leqslant 6$. So, in view of the theorem mentioned earlier, it suffices to show that if

$$F(t,x,u) = u - g(t,x) - h(t,x) - k(t,x),$$

then for any constants $c_1, c_2, \ldots, c_6$, the statement

$$\left(\sum_{i=1}^{6} c_i X_i\right) F(t,x,u) = 0 \quad \text{whenever } F(t,x,u) = 0$$

implies that $c_i = 0$ for all i s.t. $1 \leqslant i \leqslant 6$. Operating $\Sigma_{i=1}^{6} c_i X_i$ on $F(t,x,u)$ and setting $F(t,x,u) = 0$ *after* differentiation, one immediately finds that $c_2 = c_3 = c_5 = 0$ because of the linear independence of g, h, and k. One then obtains

$$\begin{pmatrix} -\beta/\alpha & -a & 1 \\ -1/c^2 & -1/c & 1 \\ -1/c^2 & 1/c & 1 \end{pmatrix} \begin{pmatrix} c_1 \\ c_4 \\ c_6 \end{pmatrix} = 0.$$

The determinant of the coefficient matrix is $(2/c)(\beta/\alpha - 1/c^2)$, which is different from zero if $c^2 \neq \alpha/\beta$. Hence, for the choice of c, $c_1 = c_4 = c_6 = 0$. This proves that the solution given by Eq. (9) is noninvariant.

We shall now list some examples which can be worked out in a similar manner.

(1) Generator: $X_1 + \alpha X_4 + \beta X_6$;

$$\text{Invariants: } I^1 = x - \alpha t, \quad I^2 = ue^{-\beta t},$$

$$I^3 = ve^{-\beta t}, \quad u = e^{\beta t} f(\lambda, \mu);$$

where $\lambda = I^1$, $\mu = I^3$, and f satisfies the PDE

$$f_\mu^2 f_{\lambda\lambda} + \mu^2 R^2 f_{\mu\mu} + 2\mu R f_\mu f_{\lambda\mu} - f_\mu^2 (\beta f - \alpha f_\lambda - \beta\mu f_\mu) = 0, \tag{13}$$

with $R = 1 - f_\lambda/\mu$.

(2) Generator: $X_2 + \alpha X_1 + \beta X_6$;

$$\text{Invariants: } I^1 = ux^{-2\beta}, \quad I^2 = vx^{1-2\beta},$$

$$I^3 = x^2/(\alpha + t), \quad u = x^{2\beta} f(\lambda, \mu);$$

where $\lambda = I^3$, $\mu = I^2$, and

$$2\mu f_\mu^2 + 2\beta(2\beta - 3) f f_\mu^2 + \lambda f_\lambda f_\mu^2 (\lambda - 2 + 8\beta) + 4\lambda^2 f_{\lambda\lambda} f_\mu^2 + R^2 f_{\mu\mu} + 4\lambda R f_\mu f_{\lambda\mu} - 2(1-\beta)(1-2\beta)\mu f_\mu^3 = 0, \tag{14}$$

with $R = \mu - 2\beta f - 2\lambda f_\lambda$.

(3) Generator: $\alpha X_1 + \beta X_2 + \gamma X_4$;

$$\text{Invariants: } I^1 = u, \qquad I^2 = \frac{\alpha + \beta t}{(\beta x/2 + \gamma)^2},$$

$$I^3 = v^2(\alpha + \beta t), \quad u = f(\lambda, \mu);$$

where $\lambda = I^2$, $\mu = I^3$, and

$$2\beta^2 \lambda^3 f_\mu^2 f_{\lambda\lambda} - 4\beta\sqrt{\mu}\lambda^{3/2} R f_\mu f_{\lambda\mu} + (2\mu f_{\mu\mu} + f_\mu) R^2 + \beta\lambda(3\beta\lambda - 2) f_\lambda f_\mu^2 - 2\beta\mu f_\mu^3 = 0, \tag{15}$$

with $R = 1 + (\beta\lambda^{3/2}/\sqrt{\mu}) f_\lambda$.

(4) Generator: $\alpha X_4 + \beta X_5 + \gamma X_6$;

$$\text{Invariants: } I^1 = t, \quad I^2 = v/u + \beta x/2(\alpha + \beta t),$$

$$I^3 = ue^{g(x,t)},$$

where $g(x,t) = (x/(\alpha + \beta t))(\beta x/4 - \gamma)$,

$$u = e^{-g(x,t)} f(\lambda, \mu);$$

where $\lambda = I^1$, $\mu = I^2$, and

$$f[4\mu^2(\alpha + \beta\lambda)^2 + (\alpha + \beta\lambda)(\beta - 8\gamma\mu) + 2\gamma^2] - 2(\alpha + \beta\lambda)^2 f_{\mu\mu} L^2 - 2\beta\mu(\alpha + \beta\lambda) f_\mu + 2(\alpha + \beta\lambda)^2 f_\lambda = 0, \tag{16}$$

with $L = (f/f_\mu)(\mu - \gamma/(\alpha + \beta\lambda))$.

In all the above examples, it is assumed that $f_\mu \neq 0$. If $f_\mu = 0$, the solutions reduce to invariant ones. Observe that example 1 is a more general case of the example discussed in detail earlier. Although Eqs. (13)–(16) are very difficult to solve, special cases of them may be amenable to solution. For example, if f depends only on μ, Eq. (13) becomes

$$\mu^2 f''(\mu) - \beta f'^2(\mu)(f - \mu f'(\mu)) = 0, \tag{17}$$

while Eq. (14) becomes

$$2\mu f'^2(\mu) + 2\beta(2\beta - 3) f f'^2(\mu) + (\mu - 2\beta f)^2 f''(\mu) - 2(1-\beta)(1-2\beta)\mu f'^3(\mu) = 0. \tag{18}$$

If β is chosen to be $\frac{1}{2}$ and if it is assumed that $\mu - f \neq 0$ (the case $\mu = f$ is not interesting), Eq. (18) becomes

$$2f'^2(\mu) + (\mu - f) f''(\mu) = 0. \tag{19}$$

Equations (17) and (19) are typical of the second-order, nonlinear ordinary differential equations (ODE's) one obtains in attempting to construct partially invariant solutions. Rather than solve those equations, we shall sketch here the solution of a similar second-order ODE derived earlier[2] by one of the authors in trying to enlarge the class of partially invariant solutions obtained by Ovsiannikov[3,4] for the equations of the transonic flow of a gas.

The equations are

$$\left.\begin{aligned} uu_x &= -v_y \\ u_y &= v_x \end{aligned}\right\}, \quad x,y \in \mathbb{R}. \tag{20}$$

It has been shown[2] that solutions of Eq. (20) partially invariant w.r.t. a certain one-dimensional subgroup (parametrized by a constant c) of the group of Lie symmetries of Eq. (20) are given by

$$v = k(\lambda,\mu) - c \ln y, \tag{21}$$

$$u_x = \frac{(\lambda/y)k_\lambda + (c/y) - (\lambda/x)k_\lambda k_\mu}{\mu + k_\mu^2}, \tag{22}$$

$$u_y = \frac{(\lambda\mu/x)k_\lambda + (\lambda/y)k_\lambda k_\mu + (ck_\mu/y)}{\mu + k_\mu^2}, \tag{23}$$

where $\lambda = x/y$, $\mu = u$, and $k(\lambda,\mu)$ satisfies a complicated, second-order nonlinear PDE. The PDE for k as well as Eqs. (22) and (23) are derived[2] under the assumption that $\mu + k_\mu^2 \neq 0$.

Now the assumptions that $c = 0$ and that k depends only on μ lead[2] to the partially invariant solutions given by

$$\mu + k'^2(\mu) = 0, \tag{24}$$

which gives $k(\mu) = \pm\frac{2}{3}(-\mu)^{3/2} + a_1$ with a_1 arbitrary, and

$$v = k(\mu), \tag{25}$$

$$u_y = \mp(-u)^{1/2} u_x. \tag{26}$$

The solutions given by Eqs. (24)–(26) are the ones found on p. 293 of Ref. 3.

Suppose now that k still depends only on μ but that $c \neq 0$. Then one obtains[2] additional partially invariant solutions given by

$$\mu + k'^2(\mu) + ck''(\mu) = 0, \tag{27}$$

$$v = k(\mu) - c \ln y, \tag{28}$$

$$u_y = u_x k'(\mu). \tag{29}$$

Equations (27) and (29) are consequences of the PDE for k, derived under the assumption that $\mu + k_\mu^2 \neq 0$. Equation (27), which is similar to Eqs. (17) and (19), was derived in Ref. 2 but not solved there. We shall sketch its solution now. The transformation[6] $z = \exp((1/c) \int k' \, d\mu)$ changes Eq. (27) into

$$z'' + (\mu/c^2)z = 0, \tag{30}$$

whose solution can be expressed[7] in terms of Bessel functions as follows:

$$z(\mu) = \sqrt{\mu}\,[AJ_{1/3}((2/3c)\mu^{3/2}) + BJ_{-1/3}((2/3c)\mu^{3/2})], \tag{31}$$

where A, B are arbitrary constants. Use of this solution and of the relation $k(\mu) = c \int (z'/z) d\mu$ gives the desired result.

III. CONCLUDING REMARKS

Suppose that H is a subgroup of G. If the infinitesimal operator X of H does not contain $\partial/\partial u$, then whenever $f(t,x)$ is an invariant solution w.r.t. H, so is $f(t,x) + c$, where c is an arbitrary constant. If, however, the operator X does contain $\partial/\partial u$, the claim is false in the sense that if $f(t,x)$ is an H-invariant solution, then $f(t,x) + c$ will be an invariant solution not w.r.t. H but w.r.t. some subgroup of the full group which includes translations in u as well. For example,[1] let H be the subgroup of G whose infinitesimal operator is $X = \alpha(\partial/\partial t) + \beta(\partial/\partial x) + \gamma u(\partial/\partial u)$. It is easy to check that $f(t,x) = ace^{(ax+t)/a^2}$, where a and c are nonzero arbitrary constants, is an H-invariant solution of Eq. (1) provided that $\gamma = \alpha/a^2 + \beta/a$. Now let b be an arbitrary constant. Then $f(t,x) = ace^{(ax+t)a^2} + b$ is an invariant solution w.r.t. the group H' whose infinitesimal generator is $X = \alpha(\partial/\partial t) + \beta(\partial/\partial x) + (\gamma u + \delta)(\partial/\partial u)$, provided that $\gamma = \alpha/a^2 + \beta/a$ as before and $\delta = -b\gamma$. Observe that H' is not a subgroup of G, but of the full group, since it includes translations in u also.

We remark that since the method of partially invariant solutions essentially provides trial solutions, one can sometimes bypass the PDE for f, which is a second-order equation, and actually work with the PDE for v such as Eq. (6), which is a first-order equation, by judiciously guessing the form of f. This is because the PDE for f is obtained by differentiating Eq. (6).

Observe that the Ovsiannikov method is quite general and applies to any PDE. Its simplification, as indicated earlier, depends on the equation under consideration. Finally, as far as the authors are aware, there are no physically significant examples in which partially invariant solutions, as opposed to invariant solutions, play an essential role. True, Ovsiannikov considers[3,4] the transonic flow equations, but no specific problem with initial/boundary conditions is considered. It is our next goal to find such an example.

ACKNOWLEDGMENTS

The authors wish to thank the referee of an earlier version of the paper for pointing out that a criterion they had derived for checking the invariance of a solution, thinking it might be new, was in fact well known.

The work was supported in part by Grants No. A 4825 and No. A 8732 from the Natural Sciences and Engineering Research Council of Canada as well as by the U. S. Army Research Office through the Mathematical Sciences Institute of Cornell University. The work of the third author was carried out while he visited Dalhousie University. He wishes to take this opportunity to thank the university for its hospitality.

[1]C. C. A. Sastri and K. A. Dunn, J. Math. Phys. **26**, 3042 (1985).
[2]C. C. A. Sastri, Contemp. Math. **54**, 35 (1986).
[3]L. V. Ovsiannikov, *Group Analysis of Differential Equations*, English translation edited by W. F. Ames (Academic, New York, 1982).
[4]L. V. Ovsiannikov, *Group Properties of Differential Equations*, English translation by G. W. Bluman (1967); perhaps available from the Dept. of Applied Math. at Caltech; original Russian edition published by the Siberian Section of the Academy of Sciences, U.S.S.R., Novosibirsk (1962).
[5]G. W. Bluman and J. D. Cole, J. Math. Mech. (now Indiana Univ. Math. J.) **18**, 1025 (1969).
[6]W. F. Ames, *Nonlinear Ordinary Differential Equations in Transport Processes* (Academic, New York, 1968).
[7]E. Kamke, *Differentialgleichungen (Losungsmethoden und Losungen)* (Akad. Verlagsges., Leipzig, 1956), Vol. I.

On the stability of the telegraph equation

J. H. Swart
Department of Mathematics, Western Michigan University, Kalamazoo, Michigan 49008

(Received 30 June 1986; accepted for publication 11 March 1987)

The linear first-order boundary conditions that will lead to a stable (well-posed) problem for the telegraph equation in quarter space are established.

I. INTRODUCTION

We consider the telegraph equation with initial square integrable Cauchy data prescribed on $t = 0$, and then make use of a technique introduced by Hersh[1] to establish suitable linear first-order boundary data on $x = 0$ which will ensure a stable solution in $x,t > 0$. It is found that instability may occur when $GL = RC$ (together with some auxiliary conditions) which is of some interest since under this condition minimum attenuation and distortionless transmission will take place.

In Sec. II we give a brief summary of the procedure introduced by Hersh while in Sec. III we apply this theory to the problem under consideration, namely,

$$C\frac{\partial V}{\partial t} + \frac{\partial I}{\partial x} + GV = 0,$$
$$L\frac{\partial I}{\partial t} + \frac{\partial V}{\partial x} + RI = 0, \tag{1}$$

with suitable Cauchy data on $t = 0$ and to-be-determined boundary conditions on $x = 0$. The constants C, G, L, and R are all positive and have their usual physical meaning.

II. BOUNDARY CONDITIONS AND STABLE SOLUTIONS

Following Hersh, we consider an arbitrary system of n linear differential equations, of any order, in $m + 2$ independent variables, namely,

$$P(D_t,D_x,D_{y_j})U = 0,$$

where $l \leqslant j \leqslant m$, U is an n-vector, and P an $n\times n$ matrix of differential polynomials with constant coefficients. We suppose that P is correct in the sense of Petrovsky, i.e., for all real ξ and η_j there exists a single fixed constant M_0 such that all the roots τ of det $P(\tau,i\xi,i\eta_j) = 0$ satisfy Re $\tau < M_0$. It is well known that the associated Cauchy problem has a unique square integrable solution for all square integrable initial data if and only if P is Petrovsky correct.

We consider the following mixed initial-boundary-value problem: U is to satisfy $PU = 0$ on $t > 0$, $x > 0$, y_j unbounded. On $t = 0$, $x > 0$, square integrable Cauchy data are prescribed, while on $x = 0$, $t > 0$, are given a k-vector $F(t,y_j)$ and a $k \times n$ matrix $B(D_t,D_x,D_{y_j})$ of differential polynomials with constant coefficients, such that $BU|_{x=0} = F$.

Our aim is to find all B for which the problem is correctly set in L_2, determination of k being part of our task. The analysis is simplified by two assumptions: that the Cauchy data are all identically zero, and that the boundary data are all delta functions, which entails no loss of generality, as arbitrary data can be treated in the usual "Green's matrix" fashion.

A solution of $PU = 0$ is called an exponential solution if it is of the form

$$\exp\left[\tau t + i\sum_i^m n_j y_j\right] \sum C_{r,k} x^r \exp[\xi_k x],$$

where the sum is over a number of terms $\leqslant d$, the number of roots ξ_k of det $P(\tau,\xi,i\eta) = 0$, i.e., the degree of ξ in det P; $C_{r,k}$ is a set of constants and if ξ_k has multiplicity m_k, r runs from 0 to $m_k - 1$. For each τ,η the set of exponential solutions forms a vector space of dimension d, while those solutions corresponding to ξ with real part negative form a subspace E^-, with dimension d^-.

The set of n vectors U which satisfy both $PU = 0$ in $t,x > 0$ *and* the homogeneous boundary conditions $BU = 0$ on $x = 0$, $t > 0$ forms a vector space N. The pair B,N will be called *unstable* if N contains exponential solutions in E^- for η real and Re τ arbitrarily large, since in this case a sequence of solutions can be chosen for which the values at any point where $t > 0$ grow arbitrarily large even when the initial values are uniformly bounded. The original boundary value problem is then *not* well posed.

If B,N is not unstable, it is called *stable* (a definition validated by the existence theorem soon to be stated). *Stability means that for some number M_1 it is true that $\{N\cap E^-\}\equiv\phi$ for all real η and* Re $\tau > M_1$. Let $W(\tau,i\eta)$ be a matrix whose columns form a basis for E^-, i.e., W is an $n\times d^-$ matrix. We denote the $k\times d^-$ matrix formed by $BW|_{x=0}$ as $e^{\tau t + i\eta y}\widetilde{B}$. It is not difficult to show that if one of the columns of W lies in N, then the function space spanned by the columns of $\widetilde{B}$ would have dimension less than d^-. Thus stability amounts to possession of B of rank d^- for all real η and Re $\tau > \{M_1,M_0\}$. In particular, it is clear that this is possible only if $\widetilde{B}$, and therefore B, has at least d^- rows, so that k must be $\geqslant d^-$.

We may now state the fundamental theorem established by Hersh: If B has d^- rows and is stable, then there exists exactly one U such that $PU = 0$ and $e^{-Mt}U$ is a tempered distribution in $t,x > 0$; $BU = \delta I$ on $x = 0$, $t > 0$, and having zero Cauchy data on $t = 0$, $x > 0$. Here U is a distribution of finite order, and has the representation

$$U = -i(2\pi)^{-m-1}\int_{M-i\infty}^{M+i\infty} d\tau \int_{-\infty}^{\infty} W\widetilde{B}^{-1}\, d\eta_1\cdots d\eta_m,$$

where W is a column basis for E^-, and $M > \max(M_0,M_1)$. To establish stability for any problem we therefore need only establish whether B is stable in the sense described above.

 0022-2488/87/071477-04$02.50

III. STABILITY OF THE TELEGRAPH EQUATION

It is well known that the system (1) is Petrovsky correct, so that we may apply the procedure outlined in Sec. II. We rewrite (1) as

$$P(D_t,D_x)U=0,$$

where

$$P(D_t,D_x)=\begin{pmatrix} CD_t+G & D_x \\ D_x & LD_t+R \end{pmatrix}$$

and U^T is the vector (V,I). It follows easily that $\det P(\tau,\xi)=0$ leads to

$$CL\tau^2+(CR+GL)\tau+GR-\xi^2=0.$$

This equation obviously has only one root ξ_1 with negative real part, i.e.,

$$\xi_1=-\{CL\tau^2+(CR+GL)\tau+GR\}^{1/2}, \qquad (2)$$

so that the solution space E^- is one dimensional, and we are led to the study of boundary conditions of the form

$$aV+bI+e\frac{\partial V}{\partial x}+d\frac{\partial I}{\partial x}=f(t) \qquad (3)$$

on $x=0$. We shall establish the following theorem.

Theorem: The initial-boundary-value problem associated with (1), with suitable Cauchy data on $t=0$ and (3) prescribed on $x=0$, will have unique, stable solutions in $x,t>0$ except in the cases where

$$d\sqrt{C}+e\sqrt{L}=0$$

and either

$$GL-RC=0, \quad b\sqrt{C}+a\sqrt{L}=0,$$

or

$$d^2(GL-RC)-2L(be-ad)=0.$$

Proof: We rewrite (3) as

$$BU=f,$$

where $B=(a+eD_x,b+dD_x)$. A basis of E^- is easily obtained, viz.,

$$W=\begin{pmatrix}-\xi_1\\ C\tau+G\end{pmatrix}e^{\tau t+\xi_1 x},$$

from which follows

$$BW|_{x=0}=\widetilde{B}e^{\tau t},$$

where

$$\widetilde{B}=-a\xi_1-\xi_1^2e+bC\tau+bG+d(C\tau+G)\xi_1.$$

On setting $\xi_1=p+iq$, $\tau=w+iv$, it follows immediately from $\det\widetilde{B}=0$ that

$$-ap-ep^2+eq^2+bCw+bG+dCwp$$
$$-dCvq+dGp=0, \qquad (4)$$

$$-aq-2pqe+bCv+dCwq+dCvp+dGq=0. \qquad (5)$$

These two equations must be taken in conjunction with (2), which can be rewritten as

$$p^2-q^2-CL(w^2-v^2)-(CR+GL)w-GR=0, \qquad (6)$$

$$2pq-2CLwv-(CR+GL)v=0. \qquad (7)$$

Instability will occur iff we can find "suitable" solutions to the system (4)–(7), viz., v, w, p, and q, which are such that $p<0$ and w can be made arbitrarily large positive.

From (5)

$$q=(bCv+dCpv)[a+2ep-dCw-dG]^{-1},$$

and substitution into (7) leads to

$$2p(bC+dCp)v[a+2ep-dCw-dG]^{-1}-2CLwv$$
$$-(CR+GL)v=0. \qquad (8)$$

Obviously (8) can be satisfied only by $v=0$ or by

$$2pC(b+dp)-(a+2ep-dCw-dG)$$
$$\times(2CLw+CR+GL)=0. \qquad (9)$$

We first consider the simple case of $v=0$: The system collapses, since from (7) it follows that $q=0$ (since we desire $p<0$). We need only consider (4) and (6):

$$-ap-ep^2+bCw+bG+dCwp+dGp=0, \qquad (4')$$

$$p^2-CLw^2-(CR+GL)w-GR=0. \qquad (6')$$

By eliminating w we obtain

$$p^4C(d^2C-e^2L)+p^3[2C(bdC-aeL)$$
$$+(LG-CR)ed]+p^2[C(b^2C-a^2L)$$
$$+C(eb+ad)(GL-CR)]+Cab(GL-CR)p=0. \qquad (10)$$

From (6') it is obvious that w can be made arbitrarily large only if p is, so that the restriction (10), which would ensure that p is finite, must fall away, i.e., all coefficients must be identically zero. This occurs iff

$$d^2C-e^2L=0, \quad GL-CR=0,$$
$$b^2C-a^2L=0, \quad bdC-aeL=0. \qquad (11)$$

If these conditions are met, it follows without difficulty that the sought-after solution of the system is given by $v=q=0$, p arbitrarily large negative and

$$W=-p/\sqrt{CL}-G/C,$$

where the conditions (11) reduce to

$$d\sqrt{C}+e\sqrt{L}=0, \quad b\sqrt{C}+a\sqrt{L}=0,$$
$$GL-CR=0. \qquad (12)$$

We now turn to the other alternative in (8), viz., that $v\neq 0$ and (9) holds. We used (5) and (7) to obtain (8), so let us now consider (4) and (6). On substituting q from (5) into (4), we obtain

$$[a+2ep-dCw-dG]^2[-ap-ep^2+bCw+bG+dCwp+dGp]$$
$$+eC^2v^2(b+dp)^2-dC^2v^2(b+dp)(a+2ep-dCw-dG)=0, \qquad (4'')$$

while similarly (6) becomes

$$[p^2-CL(w^2-v^2)-(CR+GL)w-GR][a+2ep-dCw-dG]^2-(b+dp)^2C^2v^2=0. \qquad (6'')$$

On eliminating v^2 from these two equations we obtain

$$[-eC^2(b+dp)^2+dC^2(b+dp)(a+2ep-dCw-dG)][p^2-CLw^2-(CR+GL)w-GR]$$
$$-[-ap-ep^2+bCw+bG+dCwp+dGp][(b+dp)^2C^2-CL(a+2ep-dCw-dG)^2]=0. \tag{13}$$

Obviously (13) and (9) have to be compatible for solutions to exist. For convenience we set

$$W=2w+(RC+GL)C^{-1}L^{-1}, \quad \beta=CR-GL, \tag{14}$$

so that (9) becomes

$$dCLW^2-(d\beta+2aL+4epL)W+4p(b+dp)=0, \tag{9'}$$

while (13) becomes

$$\tfrac{1}{4}LC^3d^2(b+dp)W^3+\{\tfrac{1}{4}CL(b+db)(-3adC+Cbe-5edCp-2d^2CL^{-1}\beta)-\tfrac{1}{4}C^2LD^2(a+ep)p\}W^2$$
$$+\{(b+dp)[(2e^2LC-d^2C^2)p^2+p(2LaCe-C^2bd+2edC\beta)+\tfrac{1}{4}d^2C\beta^2L^{-1}+adC\beta+\tfrac{1}{2}LCa^2-\tfrac{1}{2}C^2b^2]$$
$$+LdCp(a+ep)(a+2ep+\tfrac{1}{2}d\beta L^{-1})\}W+\{(b+dp)[2edCp^3+p^2(2ad+Cd^2\beta pL^{-1}-2e^2\beta)$$
$$+p(abC-2ae\beta+bdC\beta L^{-1}-\tfrac{3}{4}ed\beta^2L^{-1})-\tfrac{1}{4}(ad+be)\beta^2L^{-1}+\tfrac{1}{2}(Cb^2-a^2L)\beta L^{-1}]$$
$$-p(a+ep)(a+2ep+\tfrac{1}{2}d\beta L^{-1})^2\}=0. \tag{13'}$$

We now subtract $\frac{1}{4}C^2d(b+dp)W$ times (9′) from (13′), arriving at

$$-\tfrac{1}{4}LC^2[2ed^2p^2+p(2ad^2+d^3\beta L^{-1})+b(ad-be+d^2\beta L^{-1})]W^2$$
$$+[(4LCe^2d-2d^3C^2)p^3+p^2(\tfrac{5}{2}d^2Ce\beta+5LaCde+2LCe^2b-4d^2C^2b)$$
$$+p(2bCde\beta+2abCeL-\tfrac{5}{2}dC^2b^2+\tfrac{3}{2}LCa^2d+\tfrac{3}{2}ad^2C\beta+\tfrac{1}{4}d^3C\beta^2L^{-1})$$
$$+(\tfrac{1}{4}bd^2C\beta^2L^{-1}+adCb\beta+\tfrac{1}{2}LCa^2b-\tfrac{1}{2}C^2b^3)]W$$
$$+[(2ed^2C-4e^3L)p^4+(2edCb+2aCd^2-8Le^2a-4e^2d\beta+Cd^3\beta L^{-1})p^3$$
$$+(-ed^2\beta^2L^{-1}+2bCd^2\beta L^{-1}-5aed\beta-2e^2b\beta+3abCd-5a^2eL)p^2$$
$$+(abC-La^3+\tfrac{3}{2}b^2dC\beta L^{-1}-2aeb\beta-\tfrac{3}{2}a^2d\beta-edb\beta^2L^{-1}-\tfrac{1}{2}ad^2\beta^2L^{-1})p$$
$$-\tfrac{1}{4}(adb+eb^2)\beta^2+\tfrac{1}{2}(b^3C-a^2Lb)\beta L^{-1}]=0. \tag{15}$$

The two equations (9′) and (15) may be regarded as quadratic equations in W, which we can rewrite, respectively, as

$$\alpha W^2+\sigma W+\gamma=0, \tag{9''}$$
$$AW^2+BW+C=0, \tag{15'}$$

where the symbols have the obvious meaning. It is well known that two such quadratic equations will have a common root iff

$$(\alpha C-\gamma A)^2=(\gamma B-\sigma C)(\sigma A-\alpha B), \tag{16}$$

which, in this case, becomes an eighth degree polynomial in p. If it has roots they will be finite, and the corresponding values of W will be finite, contrary to our desire. Hence the only possibility that could lead to arbitrarily large W is that the condition (16) vanishes identically, i.e., every coefficient be zero.

The construction of (16) is routine, but extremely laborious. To reduce the calculations, let us first calculate only the coefficients of p^8 and p^0, leading to

$$16c^2d^2L(d^2C-e^2L)^3p^8+\cdots+\tfrac{1}{4}ab^3C^2L\beta$$
$$\times[2b^2C-2a^2L-(ad+be)\beta]$$
$$\times[aeL-Cbd+ed\beta]=0. \tag{17}$$

It is therefore necessary that

$$d^2C-e^2L=0 \tag{18}$$

and either

$$\beta=0 \tag{19}$$

or

$$2b^2C-2a^2L-(ad+be)\beta=0 \tag{20}$$

or

$$aeL-Cbd+ed\beta=0. \tag{21}$$

We shall assume (18), from which it follows that (20) can be rewritten as

$$(ad+be)d^{-2}[-d^2\beta+2L(be-ad)]=0. \tag{22}$$

For convenience we shall introduce the notation

$$T\equiv-d^2\beta+2L(eb-ad).$$

On making use of (18) we can now write the terms of (16) as

$$\sigma A-\alpha B=LC[deCTp^2+\{-\tfrac{1}{2}Cd^2\beta(ad+2be)-\tfrac{1}{2}a^2d^2LC-abCdeL+3b^2e^2CL\}p$$
$$+\tfrac{1}{4}Cb\{-\beta d(ad+be)+2b(-aeL+Cbd)\}],$$
$$\gamma B-\sigma C=4edCTp^4+[-bd^2e^2\beta^2+(20d^2bCe-2bad^2C)\beta+44abCdeL-18LCe^2b^2-2bLCa^2d^2]p^3$$
$$+[-ed^3L^{-1}\beta^3-(2bde^2+qad^2e)\beta^2+(14b^2Cde+5abe^2L-21^2aedL)\beta+12ab^2CeL$$

$$+ 14a^2bCdL - 14a^3eL^2 - 10db^3C^2 - 2e^2b^3d]p^2 + [-(ed^2b - \tfrac{1}{2}ad^3)L^{-1}\beta^3$$
$$+ (\tfrac{3}{2}b^2e^2 - 5abde - \tfrac{5}{2}a^2d^2)\beta^2 + (8ab^2Cd - 4a^3dL + 2b^2Ce - 6a^2beL)\beta$$
$$+ 4a^2b^2CL - 2C^2b^4 - 2a^4L^2]p + \tfrac{1}{4}(a\beta + 2aL)\beta bL^{-1}[-(ad+be)\beta + 2(b^2C - a^2L)],$$
$$\alpha C - \gamma A = CL[2e^2Tp^3 + \{-d^3e\beta^2L^{-1} - 5aed^2\beta + 2bde^2\beta + 6abCd^2 - 5a^2deL - Cb^2de\}p^2$$
$$+ \{2ab^2Cd - La^3d - Cb^3e + \tfrac{5}{2}b^2e^2\beta - 2aebd\beta - \tfrac{3}{2}a^2d^2\beta - \tfrac{1}{2}ad^3L^{-1}\beta^2$$
$$- e^2dbL^{-1}\beta^2\}p - \tfrac{1}{4}\beta bL^{-1}[-(ad+be)\beta + 2(b^2C - a^2L)]]. \tag{23}$$

Making use of (23) we may now determine (16), which can be shown to be

$$4C^2Le^2(Le^2 - d^2C)T^2p^6 + 2C^3edT^3p^5 + T[\cdots]p^4 + T[\cdots]p^3 + T[\cdots]p^2$$
$$+ T[\cdots]p + \tfrac{1}{4}ab^3C^2L\beta[aeL - Cbd + ed\beta](ae+bd)d^{-2}T = 0. \tag{24}$$

Obviously a necessary and sufficient condition for (24), and hence (16), to vanish identically is (18) and

$$T = 0. \tag{25}$$

From (9″) and (15′) it then follows that

$$W = (\alpha C - \gamma A)(\tau A - \alpha B)^{-1},$$

and if $T = 0$ in (23) this expression reduces to

$$W = 2ped^{-1}C^{-1},$$

so that from (14)

$$W = pC^{-1}ed^{-1} - (RC + GL)(2CL)^{-1}. \tag{26}$$

Hence we have the final condition, viz., $ed^{-1} < 0$, so that for arbitrary large negative P we can make W arbitrarily large positive, and instability will occur. This means that (18) should be written as

$$d\sqrt{C} + e\sqrt{L} = 0, \tag{18′}$$

and the theorem is proved.

[We remark that the condition (25) could be more elegantly written as

$$dG\sqrt{2} + \mathrm{Re}\sqrt{C} + 2b\sqrt{C} + 2a\sqrt{L} = 0.]$$

[1]R. Hersh, "Boundary conditions for the equations of evolution," Arch. Rat. Mech. Anal. 16, 243 (1964).

Multidimensional inverse scattering: An orthogonalization formulation

Andrew E. Yagle
Communications and Signal Processing Laboratory, Department of Electrical Engineering and Computer Science, The University of Michigan, Ann Arbor, Michigan 48109

(Received 12 June 1986; accepted for publication 4 March 1987)

The three-dimensional Schrödinger equation inverse scattering problem is solved using an orthogonalization approach. The plane waves propagating in free space are orthogonalized with respect to an inner product defined in terms of a Jost operator. The resulting integral equation is identical to the generalized Gel'fand–Levitan equation of Newton, although the present derivation is simpler and more physical than that of Newton. Newton's generalized Marchenko equation is derived from the defining integral equation for the Jost operator. These integral equations are shown to be solved by fast algorithms derived directly from the properties of their solutions. This paper thus presents a simple interpretation of Newton's two integral equations, two fast algorithms for solving these integral equations, and relations between the various approaches. This is a generalization of previously obtained results, which are also reviewed here, for the one-dimensional inverse scattering problem.

I. INTRODUCTION

The inverse scattering problem for the Schrödinger equation in three dimensions with a time-independent, local, non-spherically-symmetric potential has a wide variety of applications. For example, the inverse seismic problem of reconstructing the density and wave speed of an inhomogeneous isotropic acoustic medium from surface measurements of the medium response to a harmonic excitation can be formulated as a Schrödinger equation inverse scattering problem, as was done by Coen *et al.*[1]

A major breakthrough in obtaining an *exact* solution to the three-dimensional Schrödinger equation inverse scattering problem was made by Newton.[2] In Ref. 2 Newton presented generalized versions of two integral equations obtained for the *one-dimensional* inverse problem by Marchenko[3] and Gel'fand and Levitan.[4] These *generalized* Marchenko and Gel'fand–Levitan integral equations reconstruct the scattered field in the vicinity of the scattering potential from far-field data, just as their one-dimensional namesakes do (for details of the one-dimensional problem integral equations, see Refs. 5 and 6). The scattering potential is then recovered from the scattered field using an equation Newton calls the "miracle" equation. This completes the solution of the inverse scattering problem. In Ref. 1 this procedure was applied to the inverse seismic problem noted above.

Recently it has been noted that the derivation of the generalized Gel'fand–Levitan integral equation in Ref. 2 relies implicitly on the existence of a so-called "regular" solution. It was not firmly established in Ref. 2 that this regular solution is always well defined. However, this does not invalidate the results of Ref. 2; it merely limits their applicability to situations for which the regular solution *does* exist. In this paper the inverse scattering problem is restricted to situations in which the regular solution exists and is well defined; this is expected to cover most physical inverse scattering problems. Since a major goal of this paper is to underscore ways in which one-dimensional results generalize to three dimensions, this is an acceptable limitation.

Although Ref. 2 is a highly significant contribution to inverse scattering theory, the derivations contained therein shed little insight into the actual mechanism of the inversion process. Several recent papers have presented much simpler derivations of Newton's Marchenko integral equation. In Ref. 7 the frequency-domain Schrödinger equation was transformed into a time-domain plasma wave equation, and the interpretation of various frequency-domain properties (e.g., analyticity in the upper half-plane) as time-domain properties (e.g., causality) lended some physical insight into the inversion process. Newton's Marchenko integral equation was derived in Ref. 8 using a representation theorem, and was derived in Ref. 9 using a generalized Radon transform; both of these derivations are much simpler than Newton's derivation. However, there are no such simpler derivations as yet for Newton's generalized Gel'fand–Levitan integral equation.

For the one-dimensional inverse problem the integral equation procedures of Refs. 3–6 are known to have *differential* counterparts, which are called *layer stripping* algorithms (in the seismic literature they are known as "downward continuation" algorithms). These algorithms may be derived by exploiting the Toeplitz or Hankel structure of the kernel of integral equations[10]; however, derivations that are more physical and insightful result if basic physical principles such as causality are exploited.[11] Since they exploit the inherent structure of the inverse scattering problem, which manifests itself in the structure of the kernel of the integral equation, these algorithms require significantly fewer computations than would solving the integral equations; hence they are referred to as "fast" algorithms. An important point is that these differential, layer stripping algorithms are intimately related to the integral equation procedures; these relations are discussed in Ref. 11.

Layer stripping algorithms for the three-dimensional Schrödinger equation inverse scattering problem have been proposed in Refs. 9 and 12. Although the numerical performance of these algorithms is unknown at present, their computational complexity is significantly less than that of the

 0022-2488/87/071481-11$02.50

integral equation procedures of Newton. A relation between the algorithm of Ref. 9 and Newton's Marchenko integral equation procedure was presented in Ref. 9; this relation involved a generalized Radon transform. However, this relation did not extend to Newton's generalized Gel'fand–Levitan integral equation, and a differential fast algorithm for this integral equation has not been obtained previously.

In this paper Newton's generalized Gel'fand–Levitan integral equation is rederived by treating the inverse scattering problem as an orthogonalization problem. A Gram–Schmidt orthonormalization is performed on the free-space form of the wave field, which is a probing plane wave in a given direction of incidence. The orthogonalization is performed with respect to an inner product defined in terms of a multidimensional Jost operator, and the associated orthogonality principle results in Newton's generalized Gel'fand–Levitan integral equation. This is the first derivation of this equation other than that of Ref. 2. Newton's generalized Marchenko integral equation is also derived from the integral equation defining the Jost operator.

Two differential fast algorithms that also solve these integral equations are given. One of these algorithms is the algorithm of Ref. 12; the other is a generalized Levinson-like algorithm that is new, although it bears some resemblance to a fast algorithm derived in Ref. 13 for the problem of computing the filter for the linear, least-squares estimate of a homogeneous, anisotropic random field.

This paper thus provides a unified derivation of two multidimensional integral equations and two multidimensional fast algorithms, all of which solve the inverse scattering problem for the three-dimensional Schrödinger equation. It is thus a generalization of results for the one-dimensional inverse problem presented in Refs. 11 and 14, and illustrates how all of these procedures are connected.

The paper is organized as follows. Results for one dimension are quickly summarized in Sec. II, which contains some results from Refs. 11 and 14. The new results for three dimensions are contained in Sec. III, and the ways in which the one-dimensional results generalize to three dimensions are emphasized. The main results of Sec. III are Newton's generalized Gel'fand–Levitan and Marchenko integral equations. In Sec. IV the differential, layer-stripping algorithms are presented and related to the integral equations of Sec. III. Some connections between multidimensional inverse scattering and linear, least-squares estimation of homogeneous, anisotropic random fields are also noted. Finally, Sec. V concludes by summarizing the results of the paper and noting directions in which further research is needed.

II. THE ONE-DIMENSIONAL PROBLEM

This section derives the Gel'fand–Levitan and Marchenko integral equations for the one-dimensional inverse scattering problem using an orthogonalization procedure, following Ref. 14. It also derives differential fast algorithms that solve the inverse scattering problem and require fewer computations than would solving the integral equations. The purpose of this section is to review these concepts in a simple setting before proceeding to the more complex three-dimensional inverse problem, and to demonstrate how the concepts for the one-dimensional case generalize to the three-dimensional case.

A. The fundamental solutions

The one-dimensional inverse scattering problem considered in this section is as follows. The wave field $u(x,k)$ satisfies the Schrödinger equation

$$\left(\frac{d^2}{dx^2} + k^2 - V(x)\right)u(x,k) = 0\,, \tag{2.1}$$

where the scattering potential $V(x)$ is real valued, smooth, and has compact support. Two different initial conditions for this differential equation will be considered, resulting in two different solutions. These correspond to two different inverse scattering problems: the reflection problem and the regular problem. The names of these problems come from the names of their solutions, as will be explained shortly.

The time-domain version of the Schrödinger equation (2.1) is the *plasma wave equation*

$$\left(\frac{\partial^2}{\partial x^2} - \frac{\partial^2}{\partial t^2} - V(x)\right)\check{u}(x,t) = 0\,. \tag{2.2}$$

Solutions of (2.2) are related to solutions of (2.1) by a Fourier transform. In the sequel we will switch freely from the time domain to the frequency domain and back again.

First, some solutions to the *reflection* problem are defined. The wave field $u(x,k)$ is split into two waves traveling in the $+x$ and $-x$ directions, and two different reflection problems (probing from $-\infty$ and $+\infty$) are considered. This results in four solutions, which are then arranged in a 2×2 matrix $\Psi(x,k)$ and termed the *Jost* solution. The components of the Jost solution $\Psi(x,k) = [\psi(x,k,+),\ \psi(x,k,-)]^T$ to (2.1) are defined by their behavior at $\pm\infty$. Specifically,

$$\psi(x,k+) = [e^{-ikx},\ R_L(k)e^{ikx}]^T \quad \text{as } x\to-\infty\,, \tag{2.3a}$$

$$\psi(x,k,+) = [T(k)e^{-ikx},\ 0]^T \quad \text{as } x\to\infty\,, \tag{2.3b}$$

$$\psi(x,k,-) = [0,\ T(k)e^{ikx}]^T \quad \text{as } x\to-\infty\,, \tag{2.3c}$$

$$\psi(x,k,-) = [R_R(k)e^{-ikx},\ e^{ikx}]^T \quad \text{as } x\to\infty\,. \tag{2.3d}$$

Physically, the solution $\psi(x,k,+)$ results from a problem in which the scattering potential is probed from the left, in the $+x$ direction, resulting in a transmitted wave $T(k)e^{-ikx}$ and a reflected wave $R_L(k)e^{ikx}$. The solution $\psi(x,k,-)$ results from a problem in which probing takes place from the right, in the $-x$ direction. Here $R_L(k)$ and $R_R(k)$ are the *reflection coefficients* for the two problems, and $T(k)$ is the *transmission coefficient*, which by reciprocity is the same for both problems. The first component of each solution is the rightward traveling wave, and the second component is the leftward traveling wave. The situation is illustrated in Fig. 1. Note that the complete Jost solution $\Psi(x,k)$ is thus a 2×2 matrix. Since the data for these problems consists of the reflection coefficient $R_L(k)$ or $R_R(k)$, the inverse scattering problem that results in the Jost solution $\Psi(x,k)$ is termed the *reflection problem*. Note that given either $R_L(k)$ or $R_R(k)$ it is possible to reconstruct the other reflection coefficient and $T(k)$; see Ref. 6.

Next, some solutions to the *regular* problem are defined. The wave field $u(x,k)$ is again split into two waves traveling

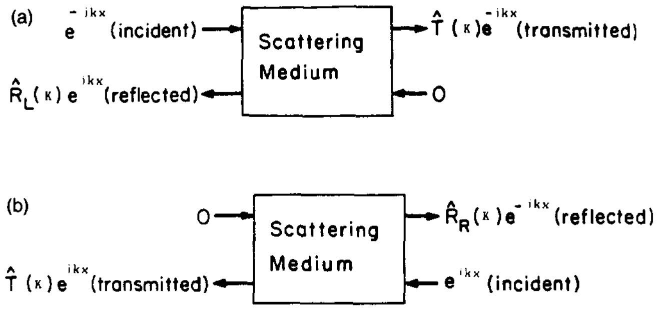

FIG. 1. (a) The reflection problem for an impulsive plane wave incident from the left. (b) The reflection problem for an impulsive plane wave incident from the right.

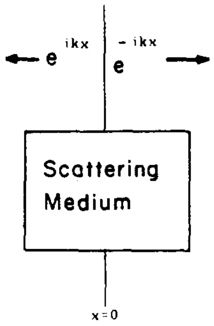

FIG. 2. The regular problem. For the 1-D problem: an impulsive boundary condition at $x=0$. For the 3-D problem: an impulsive boundary condition on the plane $e_i \cdot x = 0$.

in the $+x$ and $-x$ directions; however, the boundary conditions are changed. Instead of specifying the behavior of the wave field at $\pm\infty$, the behavior is specified at the origin $x=0$. Since each wave must be initialized, this again results in a 2×2 matrix. The *regular* solution $\Phi(x,k) = [\phi(x,k,+),\phi(x,k,-)]^T$ to (2.1) is defined by the initial conditions

$$\Phi(0,k) = I_2, \quad \frac{d}{dx}\Phi(0,k) = \mathrm{diag}[ik, -ik]. \tag{2.4}$$

In the time domain, this corresponds to introducing an impulse at the origin $x=0$. Thus in the time domain the regular solution is actually a noncausal impulse response relating the field at the origin to the field at x. This is discussed in more detail in Ref. 15. The term "regular solution" was introduced by Newton in Ref. 2, and has become standard; hence we use it here. The inverse scattering problem resulting in the regular solution $\Phi(x,k)$ is termed the *regular problem*, and it is illustrated in Fig. 2.

Since the reflection and regular solutions are linearly independent, they are related by a *Jost function* $J(k)$, which is also a 2×2 matrix. We have

$$\Phi(x,k) = \Psi(x,k)J(k) \tag{2.5}$$

and at $x=0$ we also have

$$\Psi(0,k) = \Phi(0,k)J^{-1}(k) = J^{-1}(k). \tag{2.6}$$

Since the total field $u(x,k)$ is the sum of the leftgoing and rightgoing waves at x, we have

$$u(0,k) = [1,1]J^{-1}(k). \tag{2.7}$$

All of these equations generalize directly to the three-dimensional case, as we shall see in Sec. III.

Since the one-dimensional problem is defined on the entire real line, and the potential $V(x)$ has compact support, we may without loss of generality restrict its support to the half-line $x \geqslant 0$. Then the Jost solution condition at $-\infty$ may be replaced by a similar condition at $x=0$. Equations (2.3) and (2.6) then yield

$$J^{-1}(k) = \begin{bmatrix} 1 & 0 \\ R_L(k) & T(k) \end{bmatrix}. \tag{2.8}$$

This explicit representation of the Jost function will not be available in the three-dimensional case, since that problem is radial, i.e., defined on $|\mathbf{x}| \geqslant 0$.

B. Orthogonalization

It is well known that the Jost solutions $\Psi(x,k)$ are orthonormal on the real line with respect to the usual L^2 matrix inner product, i.e., that

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} \Psi(x,k)\Psi(y,k)^{\mathrm{H}}\,dk = \delta(x-y), \tag{2.9}$$

where the superscript H denotes Hermitian transpose. This naturally suggests that the reconstruction of the field resulting from a scattering problem might be regarded as an orthogonalization procedure. However, such a procedure would clearly have to start from a point and proceed outward, and for the Jost solutions there is no clear place to start. The regular solutions $\Phi(x,k)$ would be an ideal candidate for such a procedure, since they are formed starting at $x=0$ and propagate outward in the $\pm x$ directions, but they are not orthonormal. But the regular solutions *are* orthonormal with respect to the inner product with weighting matrix $(J^{\mathrm{H}}J)^{-1}(k)$, since

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} \Phi(x,k)(J^{\mathrm{H}}J)^{-1}(k)\Phi(y,k)^{\mathrm{H}}\,dk = \frac{1}{2\pi}\int_{-\infty}^{\infty} \Psi(x,k)\Psi(y,k)^{\mathrm{H}}\,dk = \delta(x-y). \tag{2.10}$$

Note that

$$(J^{\mathrm{H}}J)^{-1}(k) = \begin{bmatrix} 1 & R_L^*(k) \\ R_L(k) & 1 \end{bmatrix}, \tag{2.11}$$

which follows from (2.8) and the conservation of energy relation

$$|R_L(k)|^2 + |T(k)|^2 = 1. \tag{2.12}$$

This suggests that the solutions $\Phi(x,k)$ may be constructed from the scattering data from the left, $R_L(k)$, as follows.

The quantities to be orthogonalized are, in the time domain, the free-space leftgoing and rightgoing impulsive plane waves resulting from the impulse introduced at the origin. In the frequency domain, these waves have the form $e^{\pm ikx}$, and arranging them into a 2×2 matrix as was done with the reflection and regular solutions results in the *free-space* solutions

$$E(x,k) = \begin{bmatrix} e^{-ikx} & 0 \\ 0 & e^{ikx} \end{bmatrix}. \tag{2.13}$$

In the absence of a scattering potential these would constitute the regular solution to the Schrödinger equation (2.1), so that we *would* have $\Phi(x,k) = E(x,k)$.

Since there is a scattering potential, the solution $\Phi(x,k)$ is formed by *orthogonalizing* $E(x,k)$ in increasing $|x|$. This is done by projecting $E(x,k)$ onto the subspace of already-orthogonalized $\Phi(x,k)$, which is $\text{span}\{\Phi(y,k), |y| < |x|\} = \text{span}\{E(y,k), |y| < |x|\}$. The projection onto a subspace is a linear combination of the elements of the subspace; here it takes the form

$$\mathscr{P}\{E(x,k)\} = -\int_{-x}^{x} M(x,y)E(y,k)dy\,, \tag{2.14}$$

where $M(x,y)$ is a matrix kernel to be specified momentarily. Note that the linear combination has been taken over the elements of $\text{span}\{E(y,k), |y| < |x|\}$, rather than $\text{span}\{\Phi(y,k), |y| < |x|\}$; since the orthogonalization of a subspace does not change its span these two subspaces are equal, and the projection can be taken to be a linear combination of the elements of either subspace.

The *error* $E(x,k) - \mathscr{P}\{E(x,k)\}$ is then orthogonal to the above subspace, and we take the *error* to be $\Phi(x,k)$. We now recognize $M(x,t)$ to be the *smooth* part of the inverse Fourier transform of $\Phi(x,k)$,

$$\Phi(x,k) = \text{diag}[e^{-ikx},e^{ikx}] + \int_{-x}^{x} M(x,t)e^{-ikt}\,dt\,, \tag{2.15}$$

so that $M(x,t)$ is the *scattered* part of the regular solution to the plasma wave equation (2.2), which is the Schrödinger equation in the time domain.

Writing out the condition that the error $\Phi(x,k)$ be orthogonal to $E(y,k)$ with respect to the inner product defined in (2.9) for $|y| < |x|$ results in the following integral equation for the scattered field $M(x,t)$:

$$\begin{bmatrix} 0 & R(x+t) \\ R(x+t) & 0 \end{bmatrix} + M(x,t) + \int_{-t}^{x} M(x,y)\begin{bmatrix} 0 & R(y+t) \\ R(y+t) & 0 \end{bmatrix} dy = 0\,, \tag{2.16}$$

where

$$R(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} R_L(k)e^{ikt}\,dk \tag{2.17}$$

is the inverse Fourier transform of $R_L(k)$. Note that $R(t)$ is a *causal* function, which accounts for the lower limit of the integral in (2.16). The centrosymmetry of (2.16) implies that $M(x,t)$ is a centrosymmetric matrix, i.e., that

$$M_{11}(x,t) = M_{22}(x,t)\,, \quad M_{12}(x,t) = M_{21}(x,t) \tag{2.18}$$

[note that this also follows on purely physical grounds from the definition of $\Phi(x,k)$]. This implies that the *scattered field* $\check{u}_s(x,t)$, which is the sum of the waves traveling in the $\pm x$ directions, i.e.,

$$\check{u}_s(x,t) = M_{11}(x,t) + M_{21}(x,t)\,, \tag{2.19}$$

satisfies the *Gel'fand–Levitan integral equation*

$$R(x+t) + \check{u}_s(x,t) + \int_{-t}^{x} \check{u}_s(x,y)R(y+t)dy\,, \quad -x \leqslant t \leqslant x\,. \tag{2.20}$$

Equation (2.20) is a Gel'fand–Levitan equation since the unknown scattered field $\check{u}_s(x,t)$ arising from a *regular* problem has finite support $-x \leqslant t \leqslant x$, resulting in a finite interval of integration.

The *Marchenko* equation for this problem has the same form, except that the range of validity is changed to $t \geqslant x$. This follows since it has been assumed that the potential $V(x)$ has support on the half-line $x \geqslant 0$. Thus there is no difference between the regular and reflection *problems*, so that the only difference between the regular and reflection (Jost) *solutions* is their supports, which are complementary.

The half-line assumption was necessary in order to obtain an explicit representation of the inverse Jost function $J^{-1}(k)$. In the three-dimensional case an explicit representation of $J^{-1}(k)$ will not be available, and the distinction between the two integral equations will become important. This distinction is also important in the one-dimensional inverse problem on the *full* (*real*) *line*.

For both the regular and Jost solutions, the potential $V(x)$ may be obtained from the jump in the scattered field at the wave front, as follows. The solution to the plasma wave equation (2.2) can be written as

$$\check{u}(x,t) = \delta(t-x) + \check{u}_s(x,t)1(t-x) \tag{2.21a}$$

for the reflection problem, and

$$\check{u}(x,t) = \delta(t-x) + \check{u}_s(x,t)(1(t+x) - 1(t-x)) \tag{2.21b}$$

for the regular problem, where $\check{u}_s(x,t)$ is the smooth part of the scattered field and $1(\cdot)$ is the unit step or Heaviside funciton. Inserting (2.21) in (2.2) and equating orders of singularities yields[16]

$$V(x) = \pm 2\frac{d}{dx}\check{u}_s(x,x)\,, \tag{2.22}$$

where the $+$ applies for the regular problem and the $-$ for the reflection problem. Equation (2.22) in conjunction with the integral equation (2.20) completes the solution of the inverse scattering problem.

C. Fast algorithms

An alternative to solving the integral equation (2.20) is to propagate the scattered field $\check{u}_s(x,t)$ for all t recursively in x, obtaining $V(x)$ from (2.22) as we go. This is the essence of a *layer stripping* algorithm, which recursively reconstructs the scattered field and potential and strips away their effects. However, the layer stripping algorithms for the regular and reflection problems, although superficially similar in appearance, are actually quite different. The difference is due to the complementary nature of the support of the scattered fields for the two problems, as illustrated in Figs. 3 and 4. The regular solution in the time domain, which is $\text{diag}[\delta(t-x),\delta(t+x)] + M(x,t)$, has support in t in the interval $[-x,x]$. The reflection solution in the time domain has support in t in the interval $[x,\infty]$ for the problem in which probing takes place from the left, and has support in t in the interval $[-\infty,-x]$ for probing from the right. This produces a major difference in the manner in which (2.22) is implemented in the algorithms.

A fast algorithm that recursively reconstructs the potential and scattered field for the *reflection* problem is as follows.[11]

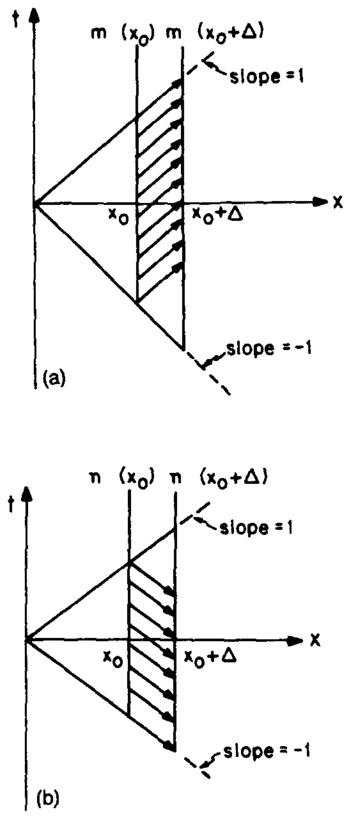

FIG. 3. (a) Recursion pattern for updating $m(x,t)$ in the fast algorithm for the regular problem. (b) Recursion pattern for updating $n(x,t)$ in the fast algorithm for the regular problem.

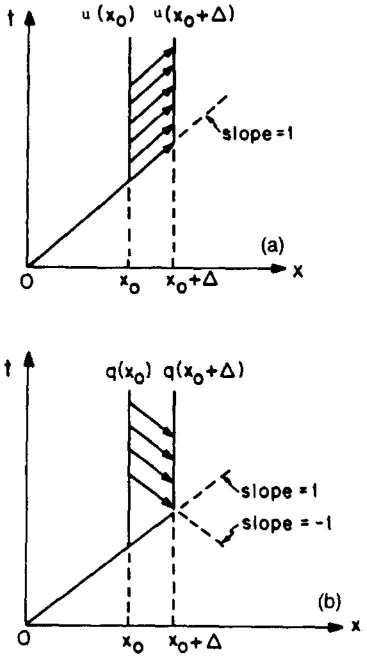

FIG. 4. (a) Recursion pattern for updating $u(x,t)$ in the fast algorithm for the reflection problem. (b) Recursion pattern for updating $q(x,t)$ in the fast algorithm for the reflection problem.

(1) Initialize the algorithm with

$$\check{u}(0,t) = R(t)\ , \quad \check{q}(0,t) = 2\frac{d}{dt}R(t)\ . \tag{2.23}$$

(2) Propagate the following equations recursively in x and t, for $t > x$:

$$\left(\frac{\partial}{\partial x} + \frac{\partial}{\partial t}\right)\check{u}(x,t) = \check{q}(x,t)\ , \tag{2.24a}$$

$$\left(\frac{\partial}{\partial x} - \frac{\partial}{\partial t}\right)\check{q}(x,t) = V(x)\check{u}(x,t)\ , \tag{2.24b}$$

$$V(x) = -2\check{q}(x,x)\ . \tag{2.24c}$$

The recursion pattern for this algorithm is illustrated in Fig. 4. Note that this amounts to successively truncating the potential—at each recursion, the region to the left of x has been replaced by free space $[V(y) = 0$ for $y < x]$. Thus the algorithm is successively reconstructing the potential and then stripping away its effects; hence the name "layer stripping" algorithm.

A fast algorithm that recursively reconstructs the potential and scattered field for the *regular* problem is as follows. For convenience let the *scalars* $m(x,t)$ and $n(x,t)$ constitute the first column of the *matrix* $M(x,t)$ of (2.14), i.e., $m(x,t) = M_{11}(x,t)$ and $n(x,t) = M_{21}(x,t)$. Then proceed as follows.

(1) Initialize the algorithm with

$$m(0,t) = n(0,t) = 0\ . \tag{2.25}$$

(2) Propagate the following equations recursively in x and t, for $-x < t < x$:

$$\left(\frac{\partial}{\partial x} + \frac{\partial}{\partial t}\right)m(x,t) = n(x,t)\ , \tag{2.26a}$$

$$\left(\frac{\partial}{\partial x} - \frac{\partial}{\partial t}\right)n(x,t) = V(x)m(x,t)\ , \tag{2.26b}$$

$$m(x,t = -x) = 0\ , \tag{2.26c}$$

$$\begin{aligned} V(x) &= 2n(x,x) \\ &= -4\frac{d}{dt}R(2x) - 2R(2x)m(x,x) \\ &\quad + \int_{-x}^{x} m(x,y)\,\frac{d}{dt}R(x+y)dy \\ &\quad + \int_{-x}^{x} R(x+y)\left\{n(x,y) - \frac{d}{dt}R(x+y)\right. \\ &\quad \left. - \int_{-y}^{x} m(x,z)\,\frac{d}{dy}R(z+y)dz\right\}dy\ , \end{aligned} \tag{2.26d}$$

where (2.26d) follows from applying (2.26a) to the integral equation (2.20).

The recursion pattern for this algorithm is illustrated in Fig. 3. Note that for the regular problem the support in t of $m(x,t)$ and $n(x,t)$ is the interval $[-x,x]$, so that the data $R(t)$ enters into the algorithm not in the initialization, but in the computation of $V(x)$ at each recursion. Thus this algorithm solves a *boundary value* problem, while the reflection problem algorithm solves an *initial value* problem. This is why the additional computation of (2.26d) is necessary for the regular problem algorithm, but not for the reflection problem algorithm.

Let the region where $V(x)$ has support be discretized into N subintervals. Then each of these algorithms requires $O(N)$ multiplication-and-add operations at each recursion, for a total of $O(N^2)$ operations to reconstruct $V(x)$. Solution of the integral equations by Gaussian elimination requires $O(N^3)$ operations to reconstruct $V(x)$. The fast algorithms require fewer computations because they exploit the causal structure of the inverse scattering problem. Both of these algorithms have their three-dimensional problem counterparts, which are given in Sec. IV.

It should be noted that other differential fast algorithms exist; see Ref. 11. In particular, the more familar continuous-parameter fast Cholesky and Krein–Levinson algorithms can be derived by reformulating the Schrödinger equation as a two-component wave system parametrized by a reflectivity function. Details are given in Ref. 11.

In this section the Gel'fand–Levitan integral equation has been derived by considering the inverse scattering problem as an orthogonalization problem with respect to the inner product defined in (2.10). This result has appeared previously in Refs. 14 and 17; it has been reviewed here in order to make apparent the ways in which this approach generalizes to the three-dimensional problem. In the next section the three-dimensional problem is treated using a similar approach, and generalized Gel'fand–Levitan and Marchenko integral equations identical to those of Ref. 2 are obtained.

III. THE THREE-DIMENSIONAL PROBLEM

In this section the main results of this paper are presented. The generalized Gel'fand–Levitan and Marchenko integral equations derived in Ref. 2 are here derived using an orthogonalization procedure similar to that used above for the one-dimensional problem. This is a much simpler derivation than the one used in Ref. 2, and it clarifies the difference between the solutions of the two integral equations. It also illustrates how the one-dimensional results presented above generalize to three dimensions.

A. The fundamental solutions

The inverse scattering problem considered in this section is as follows. The wave field $u(\mathbf{x},k)$ satisfies the Schrödinger equation

$$(\Delta + k^2 - V(\mathbf{x}))u(\mathbf{x},k) = 0\,, \tag{3.1}$$

where $\mathbf{x}\in\mathbb{R}^3$ and the potential $V(\mathbf{x})$ is real valued, smooth, and has compact support. It is also assumed that $V(\mathbf{x})$ does not induce bound states; a sufficient condition for this is for $V(\mathbf{x})$ to be non-negative. It should be noted that bound states are treated in Ref. 2; we omit them in the present derivation for simplicity and to emphasize the parallels with the one-dimensional problem. The time-domain version of (3.1) is again the plasma wave equation[7,8]

$$\left(\Delta - \frac{\partial^2}{\partial t^2} - V(\mathbf{x})\right)\check{u}(\mathbf{x},t) = 0\,, \tag{3.2}$$

where solutions to (3.1) and (3.2) are related by a Fourier transform. As before, we will switch freely from the time domain to the frequency domain, and back again.

As in the one-dimensional problem, two different sets of boundary conditions are specified, resulting in two different solutions. To emphasize the parallels with the one-dimensional problem, we use the same notation as in Sec. II.

Let $\psi(\mathbf{x},k,\mathbf{e}_i)$ be the solution to (3.1) with boundary condition

$$\psi(\mathbf{x},k,\mathbf{e}_i) = e^{-ik\mathbf{e}_i\cdot\mathbf{x}} + (e^{-ik|\mathbf{x}|}/4\pi|\mathbf{x}|)A(k,\mathbf{e}_s,\mathbf{e}_i) + O(|\mathbf{x}|^{-2})\,, \tag{3.3}$$

where the scattering amplitude is defined by

$$A(k,\mathbf{e}_s,\mathbf{e}_i) = -\int e^{-ik\mathbf{e}_s\cdot\mathbf{y}}V(\mathbf{y})\psi(\mathbf{y},k,\mathbf{e}_i)d\mathbf{y} \tag{3.4}$$

and $\mathbf{e}_i$ and $\mathbf{e}_s$ are unit vectors. The solutions $\psi(\mathbf{x},k,\mathbf{e}_i)$ can be considered as a generalizaton of the one-dimensional Jost solutions (as in Ref. 2), with the ensemble of directions $\{\mathbf{e}_i\}$ replacing the directions $\pm x$. These solutions also result from a reflection problem in which an incident impulsive plane wave in the direction $\mathbf{e}_i$ is used to probe the scattering potential, and the data consists of the far-field reflection response in the form of the scattering amplitude. Equations (3.3) and (3.4) have their time-domain counterparts that specify solutions to the plasma wave equation (3.2); see Refs. 7–9. Note that in the present formulation the factor of 4π is incorporated in (3.3) instead of (3.4), as in Refs. 7–9.

Let $\phi(\mathbf{x},k,\mathbf{e}_i)$ be the solution to (3.2) that is an entire analytic function of k, is of exponential order $|\mathbf{e}_i\cdot\mathbf{x}|$, and has a value of 1 along the plane $\mathbf{e}_i\cdot\mathbf{x} = 0$. More specifically, $\phi(\mathbf{x},k,\mathbf{e}_i)$ is specified by the boundary conditions

$$\phi(\mathbf{x},k,\mathbf{e}_i) = 1\,; \quad \nabla\phi(\mathbf{x},k,\mathbf{e}_i) = ik\mathbf{e}_i \quad \text{for } \mathbf{e}_i\cdot\mathbf{x} = 0\,; \tag{3.5}$$

$\phi(\mathbf{x},k,\mathbf{e}_i)$ is thus a generalization of the regular solution (2.4) to three dimensions. It is also the regular solution referred to in Ref. 2.

In Ref. 17 it was pointed out that the regular solution defined in Ref. 2 cannot be guaranteed to exist. This is because the regular solution in Ref. 2 was *defined* by a Jost operator [Eq. (3.6) below; compare to (2.5) for the one-dimensional problem], and thus it cannot be guaranteed to be of exponential order $|\mathbf{e}_i\cdot\mathbf{x}|$. This implies that the Povsner–Levitan relation (7.3) used in Ref. 2 may be incorrect. Here, however, we *assume* that this regular solution exists.

It should be noted that the existence of the regular solution in general is still an unsolved problem. However, the corrections made to the results of Ref. 2 in Refs. 18 and 19 obfuscate an already complicated inverse scattering procedure still further, and as noted in Ref. 18, the results of Ref. 2 are "probably correct" in any case. In the sequel we simply restrict our attention to situations in which it does exist.

We further assume that $\phi(x,k,e_i) - e^{-ik\mathbf{e}_i\cdot\mathbf{x}}$ is square integrable in k. Then, using the Paley–Wiener theorem, as in Ref. 2, it follows that $\check{\phi}(\mathbf{x},t,\mathbf{e}_i) = \mathscr{F}^{-1}\{\phi(\mathbf{x},k,\mathbf{e}_i)\}$ has support in t in the interval $[-\mathbf{e}_i\cdot\mathbf{x},\mathbf{e}_i\cdot\mathbf{x}]$ (compare this to the one-dimensional support interval $[-x,x]$). Thus $\phi(\mathbf{x},k,\mathbf{e}_i)$ has the Povsner–Levitan representation [compare with (7.3) in Ref. 2 and (2.14) above]

$$\phi(\mathbf{x},k,\mathbf{e}_i) = e^{-ik\mathbf{e}_i\cdot\mathbf{x}} - \int_{-\mathbf{e}_i\cdot\mathbf{x}}^{\mathbf{e}_i\cdot\mathbf{x}} m(\mathbf{x},t,\mathbf{e}_i)e^{-ikt}\,dt \tag{3.6}$$

[the impulse in the $-\mathbf{e}_i\cdot\mathbf{x}$ direction is included in $\phi(\mathbf{x},k,-\mathbf{e}_i)$] so that $m(\mathbf{x},t,\mathbf{e}_i)$ is the nonimpulsive part of the regular solution $\check{\phi}(\mathbf{x},t,\mathbf{e}_i)$. Note that in the time domain, the solutions $\check{\phi}(\mathbf{x},t,\mathbf{e}_i)$ and $\check{\psi}(\mathbf{x},t,\mathbf{e}_i)$ have complementary support in that the former has support in t on the interval $[-\mathbf{e}_i\cdot\mathbf{x},\mathbf{e}_i\cdot\mathbf{x}]$, while the latter has support in t on the interval $[\mathbf{e}_i\cdot\mathbf{x},\infty]$.

The solutions $\psi(\mathbf{x},k,\mathbf{e}_i)$ and $\phi(\mathbf{x},k,\mathbf{e}_i)$ are related by a *Jost operator* $J(k)$. This is an operator on the space $L^2(S^2)$ (S^2 is the unit sphere) with kernel $J(k,\mathbf{e}_1,\mathbf{e}_2)$. The 2×2 matrix multiplication (2.5) becomes

$$\phi(\mathbf{x},k,\mathbf{e}_i) = \int_{S^2} \psi(\mathbf{x},k,\mathbf{e}_s) J(k,\mathbf{e}_s,\mathbf{e}_i)\, d\mathbf{e}_s\,. \tag{3.7}$$

The Jost operator has inverse[2] $J^{-1}(k)$ with a kernel defined as above. Setting $\mathbf{x}=\mathbf{0}$ results in

$$\psi(\mathbf{0},k,\mathbf{e}_i) = \int_{S^2} \phi(\mathbf{0},k,\mathbf{e}_s) J^{-1}(k,\mathbf{e}_s,\mathbf{e}_i)\, d\mathbf{e}_s = \int_{S^2} J^{-1}(k,\mathbf{e}_s,\mathbf{e}_i)\, d\mathbf{e}_s = \mathbf{1}J^{-1}(k)\,, \tag{3.8}$$

where the effect of the operator $\mathbf{1}$ is a generalization of premultiplication by the vector $[1,1]$ in (2.7). This confirms that the Jost operator defined here matches the one defined in Ref. 2.

In Sec. II the potential was required to have support in the half-space $x>0$, allowing an explicit representation (2.8) of the Jost function to be determined. Unfortunately, this will not work for the three-dimensional problem, since the present problem is defined over all of $\mathbb{R}^3$. It is noted in Ref. 2 that the Jost operator satisfies

$$J(-k) = QS(k)J(k)Q\,, \tag{3.9}$$

where $S(k)$ is the scattering operator with kernel

$$S(k,\mathbf{e}_s,\mathbf{e}_i) - I = -(k/2\pi i)A(k,\mathbf{e}_s,\mathbf{e}_i) \tag{3.10}$$

and Q is the operator such that $QA(k,\mathbf{e}_s,\mathbf{e}_i) = A(k,-\mathbf{e}_s,\mathbf{e}_i)$. In Ref. 2 the relation (3.9) leads to a Marchenko integral equation for the kernel $J(k,\mathbf{e}_s,\mathbf{e}_i)$. We now derive a similar equation for the kernel $J^{-1}(k,\mathbf{e}_s,\mathbf{e}_i)$.

From (3.9) we have that

$$J^{-1}(-k) = QJ^{-1}(k)S^H(k)Q\,, \tag{3.11}$$

where the well-known unitarity of the scattering operator $S(k)$ has been used. Repeating the derivation of Ref. 2 (p. 1707) for (3.11) instead of (3.9) leads to a Marchenko integral equation for the kernel $J^{-1}(k,\mathbf{e}_s,\mathbf{e}_i)$, as follows. Since both ψ and ϕ contain impulses in the time domain, J^{-1} does also, and $J^{-1}(k)-1$ is square integrable (see Ref. 2). Therefore we may write

$$J^{-1}(k,\mathbf{e}_s,\mathbf{e}_i) = 1 + \int_0^\infty L(t,\mathbf{e}_s,\mathbf{e}_i)e^{-ikt}\,dt \tag{3.12}$$

and, following Ref. 2, this leads to the following Marchenko integral equation for $L(t,\mathbf{e}_s,\mathbf{e}_i)$:

$$L(t,\mathbf{e}_s,\mathbf{e}_i) = G(t,-\mathbf{e}_s,\mathbf{e}_i) + \int_0^\infty \int_{S^2} L(\tau,-\mathbf{e}_s,\mathbf{e}')G(t+\tau,\mathbf{e}',\mathbf{e}_i)\,d\mathbf{e}'\,d\tau\,, \tag{3.13}$$

where $G(t,\mathbf{e}_i,\mathbf{e}_s)$ is defined by

$$G(t,\mathbf{e}_i\mathbf{e}_s) = \frac{1}{2\pi}\int_{-\infty}^{\infty} [S(k,\mathbf{e}_s,-\mathbf{e}_i)-1]e^{ikt}\,dk = \frac{1}{4\pi^2}\int_{-\infty}^{\infty} ikA(k,\mathbf{e}_s,-\mathbf{e}_i)e^{ikt}\,dk \tag{3.14}$$

(note the transposition of $\mathbf{e}_i$ and $\mathbf{e}_s$ caused by the Hermitian operator).

It is indeed unfortunate that the solution of the generalized Gel'fand–Levitan equation requires the prior solution of this Marchenko equation in order to obtain the inverse Jost operator kernel $J^{-1}(k,\mathbf{e}_s,\mathbf{e}_i)$, but there is no other known way to obtain this kernel. However, in Sec. III C below it will be shown that the generalized Marchenko equation for the scattered field resulting from a reflection problem can be derived from (3.13) and (3.14).

B. Orthonormalization

It is well known that in the absence of bound states the solutions $\psi(\mathbf{x},k,\mathbf{e}_i)$ are orthonormal, in that

$$\frac{1}{(2\pi)^3}\int_0^\infty \int_{S^2} \psi(\mathbf{x},k,\mathbf{e})\psi^*(\mathbf{y},k,\mathbf{e})k^2\,d\mathbf{e}\,dk = \delta(\mathbf{x}-\mathbf{y})\,. \tag{3.15}$$

As in the one-dimensional case, the solutions $\{\psi(\mathbf{x},k,\mathbf{e}_i)\}$ are inappropriate candidates for the result of an orthogonalization procedure, since they are initiated in the far field. The solutions $\{\phi(\mathbf{x},k,\mathbf{e}_i)\}$ are ideal candidates for such a procedure, since they are generated in increasing $|\mathbf{e}_i\cdot\mathbf{x}|$ in the time domain, and from (3.6) and (3.15) they are orthonormal with respect to the inner product

$$\langle u_1(\mathbf{x},k,\mathbf{e}),u_2(\mathbf{y},k,\mathbf{e})\rangle = \frac{1}{(2\pi)^3}\int_0^\infty \int_{S^2}\int_{S^2} u_1(\mathbf{x},k,\mathbf{e}_1)(J^{\mathrm{H}}J)^{-1}(k,\mathbf{e}_1,\mathbf{e}_2) \times u_2^*(\mathbf{y},k,\mathbf{e}_2)k^2\,d\mathbf{e}_1\,d\mathbf{e}_2\,dk\,. \tag{3.16}$$

However, the region $\{\mathbf{y}\in\mathbb{R}^3:\ -\mathbf{e}_i\cdot\mathbf{x}\leqslant\mathbf{e}_i\cdot\mathbf{y}\leqslant\mathbf{e}_i\cdot\mathbf{x}\}$ in which the orthogonalization takes place is still not compact, so a further transformation is necessary. Since the time-domain solution $\check{\phi}(\mathbf{x},t,\mathbf{e}_i)$ is only defined for $t>0$, we may regard its smooth part $m(\mathbf{x},t,\mathbf{e}_i)$ as the *Radon transform* of a function $h(\mathbf{x},\mathbf{y})$ (Ref. 2):

$$\mathscr{R}\{h(\mathbf{x},\mathbf{y})\} = \int h(\mathbf{x},\mathbf{y})\delta(t-\mathbf{e}_i\cdot\mathbf{y})\,d\mathbf{y} = m(\mathbf{x},t,\mathbf{e}_i)\operatorname{sgn}[\mathbf{e}_i\cdot\mathbf{x}]\,. \tag{3.17}$$

Note that the support of $h(\mathbf{x},\mathbf{y})$ in $\mathbf{y}$ is the interior of the sphere of radius $|\mathbf{x}|$: $\{|\mathbf{y}|<|\mathbf{x}|\}$. This is the triangularity property that makes an integral equation procedure possible; we see here that this property follows from time causality. Using the projection-slice property of the Radon transform, the Fourier transform relation (3.6) becomes

$$\phi(\mathbf{x},k,\mathbf{e}_i) = e^{-ik\mathbf{e}_i\cdot\mathbf{x}} - \int h(\mathbf{x},\mathbf{y})e^{-ik\mathbf{e}_i\cdot\mathbf{y}}\,d\mathbf{y} = \mathscr{F}\{\delta(\mathbf{x}-\mathbf{y})-h(\mathbf{x},\mathbf{y})\}\,. \tag{3.18}$$

From this point on the argument matches that given in Sec. II for the one-dimensional problem. The free-space solutions $\{e^{-ik\mathbf{e}_i\cdot\mathbf{x}}\}$ are orthogonalized in increasing $|\mathbf{x}|$. The projection of $e^{-ik\mathbf{e}_i\cdot\mathbf{x}}$ on $\text{span}\{\phi(\mathbf{y},k,\mathbf{e}_i),\ |\mathbf{y}|<|\mathbf{x}|\} = \text{span}\{e^{-ik\mathbf{e}_i\cdot\mathbf{y}},\ |\mathbf{y}|<|\mathbf{x}|\}$ takes the form [compare to (2.14)]

$$\mathscr{P} = \int_{|\mathbf{y}|<|\mathbf{x}|} h(\mathbf{x},\mathbf{y})e^{-ik\mathbf{e}_i\cdot\mathbf{y}}\,d\mathbf{y}\,. \tag{3.19}$$

The reason that the kernel of the projection (3.19) is $h(\mathbf{x},\mathbf{y})$ is as follows. As in the one-dimensional case, we take the *error*

$$e^{-ik\mathbf{e}_i\cdot\mathbf{x}} - \mathscr{P} = e^{-ik\mathbf{e}_i\cdot\mathbf{x}} - \int_{|\mathbf{y}|<|\mathbf{x}|} h(\mathbf{x},\mathbf{y})e^{-ik\mathbf{e}_i\cdot\mathbf{y}}\,d\mathbf{y} = \phi(\mathbf{x},k,\mathbf{e}_i) \tag{3.20}$$

to be the regular solution at $\mathbf{x}$, since by the orthogonality principle the error is orthogonal to this subspace, and thus may be used to expand it. Comparing (3.18) and (3.20) proves that the kernel of the projection (3.19) is precisely $h(\mathbf{x},\mathbf{y})$. The kernel $h(\mathbf{x},\mathbf{y})$ should be compared with the matrix kernel $M(x,y)$ in the projection (2.14). The difference is that $h(\mathbf{x},\mathbf{y})$ is the *inverse Radon transform* of the smooth part of the regular solution in the time domain, while $M(x,y)$ is simply the smooth part of the regular solution in the time domain.

We now derive a generalized Gel'fand–Levitan integral equation identical to that of Ref. 2. For convenience the notation of Ref. 2 is adopted. Writing out the condition that the error $\phi(\mathbf{x},k,\mathbf{e}_i)$ be orthogonal to the subspace element $e^{-ik\mathbf{e}_i\cdot\mathbf{y}}$ for $|\mathbf{y}|\leqslant|\mathbf{x}|$, with respect to the inner product defined by (3.16), results in

$$h_0(\mathbf{x},\mathbf{y}) = h(\mathbf{x},\mathbf{y}) + \int_{|\mathbf{z}|<|\mathbf{x}|} h(\mathbf{x},\mathbf{z})h_0(\mathbf{z},\mathbf{y})d\mathbf{z}\,, \tag{3.21}$$

which is Eq. (8.4) in Ref. 2. Here

$$\begin{aligned} h_0(\mathbf{x},\mathbf{y}) &= \frac{1}{(2\pi)^3}\int_0^\infty \int_{S^2}\int_{S^2} M(k,\mathbf{e}_1,\mathbf{e}_2) \\ &\quad\times e^{-ik(\mathbf{e}_1\cdot\mathbf{x}-\mathbf{e}_2\cdot\mathbf{y})}k^2\,d\mathbf{e}_1\,d\mathbf{e}_2\,dk \\ &= \mathscr{F}^{-1}_{k,\mathbf{e}_2}\int_{S^2} M(k,\mathbf{e}_1,\mathbf{e}_2)e^{-ik\mathbf{e}_1\cdot\mathbf{x}}\,d\mathbf{e}_1\,, \end{aligned} \tag{3.22}$$

where $M(k,\mathbf{e}_1,\mathbf{e}_2) = ((J^{\mathrm{H}}J)^{-1}-I)(k,\mathbf{e}_1,\mathbf{e}_2)$ is the perturbation of the spectral function $(J^{\mathrm{H}}J)^{-1}$ away from its free-space representation. Equations (3.21) and (3.22) should be compared to the one-dimensional problem Eqs. (2.16) and (2.17).

The key fact here is the *triangularity* of $h(\mathbf{x},\mathbf{y})$ in (3.21). This follows from the support of the regular solution, although it has also been established rigorously.[19] Taking the *partial* inverse Radon transform[19] of (3.21), and using (3.17) and the projection-slice theorem results in the *generalized Gel'fand–Levitan integral equation*[2]

$$\begin{aligned} &\operatorname{sgn}[\mathbf{e}_i\cdot\mathbf{x}]\,m(\mathbf{x},t,\mathbf{e}_i) \\ &= \int_{S^2} M(t+\mathbf{e}_i\cdot\mathbf{x},\mathbf{e}_s,\mathbf{e}_i)d\mathbf{e}_s - \int_{S^2}\int_{-|\mathbf{e}_s\cdot\mathbf{x}|}^{|\mathbf{e}_s\cdot\mathbf{x}|} \operatorname{sgn}[\mathbf{e}_s\cdot\mathbf{x}] \\ &\quad\times m(\mathbf{x},\tau,\mathbf{e}_s)M(t+\tau,\mathbf{e}_s,\mathbf{e}_i)d\tau\,d\mathbf{e}_s\,, \end{aligned} \tag{3.23}$$

where $M(t,-\mathbf{e}_s,\mathbf{e}_i) = \mathscr{F}^{-1}\{M(k,\mathbf{e}_s,\mathbf{e}_i)\}$.

Once the integral equation (3.23) has been solved, the potential $V(x)$ is then recovered from $m(\mathbf{x},t,\mathbf{e}_i)$ using the *miracle*[2] or *fundamental identity*[7]

$$V(x) = 2\mathbf{e}_i\cdot\nabla m(\mathbf{x},t=\mathbf{e}_i\cdot\mathbf{x},\mathbf{e}_i)\,, \tag{3.24}$$

which is the three-dimensional analog of (2.22) and is derived in the same way. Note the sign change in (3.24) as compared to the equation in Refs. 2 and 7; this is due to the use of the regular solution instead of $\psi(\mathbf{x},k,\mathbf{e}_i)$. In Refs. 18 and 19 the gradient of the *jump* in the scattered field must be used in (3.24), since the regular solution as defined in those papers is not known to satisfy $m(\mathbf{x},t,\mathbf{e}_i)=0$ for $t>\mathbf{e}_i\cdot\mathbf{x}$. However, in the present case this anticausality follows from the support of the regular solution.

As in the one-dimensional case, the generalized Gel'fand–Levitan equation has a finite interval of integration, which is an advantage over the generalized Marchenko integral equation to be derived next. However, it is necessary to solve the Marchenko equation (3.13) for the generalized Jost function $J^{-1}(k,\mathbf{e}_1,\mathbf{e}_2)$ first, which is most inconvenient.

C. Generalized Marchenko equation

In the one-dimensional case the inverse Jost function was related to the reflection problem scattered field at the origin by (2.7). Since the scattered field was known at the origin, an explicit representation of J^{-1} could be found. For the three-dimensional case, the reflection problem scattered field is not known at the origin, and J^{-1} must be found from the integral equation (3.13). However, the integral equation (3.13) can be transformed into an integral equation for the scattered field at the origin, and then into an integral equation for the reflection problem scattered field *anywhere*, using an observation made in Ref. 2. This integral equation is identical to the generalized Marchenko equation of Ref. 2.

Integrating (3.13) with respect to $\mathbf{e}_s$ over the unit sphere S^2 and using (3.8) and (3.12) results in

$$\begin{aligned} \check{u}_s(\mathbf{0},t,\mathbf{e}_i) &= \int_{S^2} G(t,\mathbf{e}_s,\mathbf{e}_i)d\mathbf{e}_s \\ &\quad+ \int_0^\infty\int_{S^2} G(t+\tau,\mathbf{e}',\mathbf{e}_i)\check{u}_s(\mathbf{0},t,-\mathbf{e}')d\mathbf{e}'\,d\tau\,, \end{aligned} \tag{3.25}$$

where $\check{u}_s(\mathbf{0},t,\mathbf{e}_i)$ is the scattered field at the origin for the reflection problem with probing impulsive plane wave in the direction $\mathbf{e}_i$. This integral equation is equivalent to the generalized Marchenko equation of Ref. 2 with $\mathbf{x}=\mathbf{0}$, since it is identical to (4.14) of Ref. 7 with $\mathbf{x}=\mathbf{0}$. Here $G(t,\mathbf{e}_i,\mathbf{e}_s)$ is the time derivative of the inverse Fourier transform of the scattering amplitude $A(k,\mathbf{e}_s,\mathbf{e}_i)$ [note the transposition of $\mathbf{e}_i$ and $\mathbf{e}_s$, and compare with (4.11) of Ref. 7].

We now make use of an observation made in Ref. 2. If the potential $V(\mathbf{x})$ is *shifted* by a translation $\mathbf{x}'$, becoming $V(\mathbf{x}-\mathbf{x}')$, then the solution $\psi(\mathbf{x},k,\mathbf{e}_i)$ becomes $\psi(\mathbf{x}-\mathbf{x}',k,\mathbf{e}_i)e^{-ik\mathbf{e}_i\cdot\mathbf{x}'}$ and thus the scattering amplitude $A(k,\mathbf{e}_i,\mathbf{e}_s)$ becomes $A(k,\mathbf{e}_i,\mathbf{e}_s)e^{-ik(\mathbf{e}_s-\mathbf{e}_i)\cdot\mathbf{x}'}$. Therefore to compute the scattered field $\check{u}_s(\mathbf{x}',t,\mathbf{e}_i)$ at $\mathbf{x}'$ resulting from a potential $V(\mathbf{x})$, we compute the field at the origin $\mathbf{x}=\mathbf{0}$ [using (3.25)] resulting from a shifted potential $V(\mathbf{x}-\mathbf{x}')$.

This merely requires that we replace the scattering amplitude, and hence $G(t,\mathbf{e}_i,\mathbf{e}_s)$, with its shifted version. This yields

$$G(t,\mathbf{e}_i,\mathbf{e}_s,\mathbf{x}) = -(2\pi)^{-2}\int_{-\infty}^{\infty} e^{ik(t-(\mathbf{e}_s-\mathbf{e}_i)\cdot\mathbf{x})} ikA(k,\mathbf{e}_s,\mathbf{e}_i)\,dk \tag{3.26}$$

and the integral equation (3.25) is modified to

$$\check{v}_s(\mathbf{x},t,\mathbf{e}_i) = \int_{S^2} G(t,\mathbf{e}_s,\mathbf{e}_i,\mathbf{x})\,d\mathbf{e}_s + \int_0^\infty \int_{S^2} G(t+\tau,\mathbf{e}',\mathbf{e}_i,\mathbf{x})\check{v}_s(\mathbf{x},t,-\mathbf{e}')\,d\mathbf{e}'\,d\tau, \tag{3.27}$$

where

$$\check{v}_s(\mathbf{x},t,\mathbf{e}_i) = \check{u}_s(\mathbf{x},t-\mathbf{e}_i\cdot\mathbf{x},\mathbf{e}_i) \tag{3.28}$$

is simply the delayed scattered field. Equations (3.26) and (3.27) are identical to (4.11) and (4.14) of Ref. 7, which in turn are equivalent to the generalized Marchenko integral equation of Ref. 2.

It has been shown that the generalized Gel'fand–Levitan integral equation of Ref. 2 can be interpreted as an orthogonality condition for the construction of the solutions $\phi(\mathbf{x},k,\mathbf{e}_i)$ with respect to the inner product (3.16). The construction of the inverse Jost operator requires the solution of a Marchenko equation, and this equation can be extended to the generalized Marchenko integral equation of Ref. 2. This shows the relation between the two integral equations, and how this relation is a generalization of the relation that exists between them in one dimension.

IV. FAST ALGORITHMS FOR THE THREE-DIMENSIONAL INVERSE SCATTERING PROBLEM

In this section differential, layer stripping fast algorithms for solving the three-dimensional inverse scattering problem are presented. These algorithms require fewer computations than solving the integral equations presented above, but they reconstruct $V(\mathbf{x})$, $\psi(\mathbf{x},t,\mathbf{e}_i)$, and $\phi(\mathbf{x},t,\mathbf{e}_i)$ just as the integral equations do. They are also generalizations of the algorithms presented in Sec. II.

A. The reflection problem

A major distinction between the one-dimensional and three-dimensional reflection problems is that for the one-dimensional problem near-field and far-field data are identical (save for a time shift), while for the three-dimensional problem the extrapolation of the near-field scattered field from the far-field scattering amplitude is a nontrivial problem. For the reflection problem differential algorithms it is assumed that the scattered field is observed in the *near field*. Since in many inverse scattering problems (e.g., inverse seismic problems) data are actually taken in the near field, this assumption is not only tenable, but realistic.

A differential algorithm for solving the reflection problem is as follows.[12] For convenience let $z=\mathbf{e}_i\cdot\mathbf{x}$ be the axis normal to the incident impulsive plane wave, and let $\mathbf{y}$ be the two directions perpendicular to z, so that any function $f(\mathbf{x})$ of $\mathbf{x}$ can be written as a function $f(z,\mathbf{y})$ of z and $\mathbf{y}$.

(1) Initialize the algorithm on the plane $z=\mathbf{e}_i\cdot\mathbf{x}=0$ with observations of the scattered field and its derivative on this plane.

(2) Propagate the following equations recursively in $z=\mathbf{e}_i\cdot\mathbf{x}$ and t, for $t\geqslant z$ and for all $\mathbf{y}$:

$$\left(\frac{\partial}{\partial z}+\frac{\partial}{\partial t}\right)u(z,\mathbf{y},t) = q(z,\mathbf{y},t), \tag{4.1a}$$

$$\left(\frac{\partial}{\partial z}-\frac{\partial}{\partial t}\right)q(z,\mathbf{y},t) = (V(\mathbf{x})-\Delta_\mathbf{y})u(z,\mathbf{y},t), \tag{4.1b}$$

$$V(\mathbf{x}) = -2q(z,\mathbf{y},t=z), \tag{4.1c}$$

where $\Delta_\mathbf{y}$ is the Laplacian operator with respect to $\mathbf{y}$, which is also the *transverse* Laplacian operator with respect to $\mathbf{x}$. The recursion patterns for this algorithm are the same as for its one-dimensional counterpart, and are illustrated in Fig. 4.

Note that (4.1c) follows using the same argument used to derive (3.24) (see Ref. 12) and is comparable to (2.22). Also note that this algorithm requires $O(N^6)$ operations to reconstruct $V(\mathbf{x})$, while the solution of the generalized Marchenko integral equation requires $O(N^{12})$ operations. Some details on ways to implement this algorithm numerically are given in Ref. 12.

The computational simplicity of this algorithm as compared to the solution of the generalized Marchenko integral equation (and the algorithm for the regular problem given below) results from the inherent causal structure of the reflection problem, which is fully exploited by this algorithm. Instead of attempting to reconstruct the scattered field all at once in one huge operation, the algorithm recursively reconstructs both the scattered field and the potential as the wave front penetrates the region where $V(\mathbf{x})$ has support. It then strips away the effects of the reconstructed region, reducing the size of the problem and obviating the need to store information about the reconstructed region to process the data associated with the unknown region. Another important feature is the use of near-field data, which avoids the coupling between the scattered fields associated with different $\mathbf{e}_i$ that makes the generalized Marchenko equation so computationally intensive to solve.

B. The regular problem

The regular problem lacks the causal structure of the reflection problem, which is why it is harder to solve using either the generalized Gel'fand–Levitan equation or a differential algorithm. Two different differential algorithms for the regular problem are presented. The second algorithm is similar to an algorithm proposed for estimation of random fields in Ref. 13, illustrating some connections between inverse scattering in three dimensions and estimation of random fields. This generalizes the connections between these two topics that exists in one dimension (e.g., Ref. 20).

A new differential algorithm for solving the regular problem is as follows.

(1) Initialize the algorithm on the plane $z=\mathbf{e}_i\cdot\mathbf{x}=0$ using

$$m(z=0,\mathbf{y},t=0) = n(z=0,\mathbf{y},t=0) = 0. \tag{4.2}$$

(2) Propagate the following equations recursively in z and t, for $-z \leqslant t \leqslant z$ and for all $\mathbf{y}$:

$$\left(\frac{\partial}{\partial z}+\frac{\partial}{\partial t}\right)m(z,\mathbf{y},t)=n(z,\mathbf{y},t)\,, \tag{4.3a}$$

$$\left(\frac{\partial}{\partial z}-\frac{\partial}{\partial t}\right)n(z,\mathbf{y},t)=(V(\mathbf{x})-\Delta_{\mathbf{y}})m(z,\mathbf{y},t)\,, \tag{4.3b}$$

$$m(z,\mathbf{y},t=-z)=0\,, \tag{4.3c}$$

$$V(\mathbf{x})=2n(z,\mathbf{y},t=z) \tag{4.3d}$$

obtained from (3.23). The recursion patterns for this algorithm are the same as for its one-dimensional counterpart, and are illustrated in Fig. 3. Note that $n(z,\mathbf{y},t=z)$ for the *regular* problem must be obtained from the values of $n(z,\mathbf{y},t\neq z)$ using the integral equation (3.23). This is analogous to (2.26d) for the one-dimensional problem, for which $n(x,t=x)$ is obtained from the integral equation (2.20).

Aside from the computation of (4.3d), a major problem with this algorithm is that the region in which the computations are to be carried out has infinite extent in $\mathbf{y}$. This can be avoided by using the inverse Radon transform, as in (3.15), which maps the region in which computations are performed into the interior of a sphere. Taking the inverse Radon transform of the Schrödinger equation (3.1) in the time domain and using (3.18) results in

$$(\Delta_{\mathbf{x}}-\Delta_{\mathbf{y}})h(\mathbf{x},\mathbf{y})=V(\mathbf{x})h(\mathbf{x},\mathbf{y})\,, \tag{4.4}$$

where $\Delta_{\mathbf{x}}$ is again the Laplacian operator with respect to $\mathbf{x}$. An equation similar to (4.4) was encountered in the problem of deriving a fast algorithm for the linear least-squares estimation of a homogeneous random field,[13] and a variation of the algorithm presented in Ref. 13 is useful here.

Another differential algorithm for solving the regular problem is as follows.

(1) Initialize the algorithm at the origin using

$$h(0,0)=g(0,0)=0\,. \tag{4.5}$$

(2) Propagate the following equations recursively in $r=|\mathbf{x}|$ and $s=|\mathbf{y}|$, for $0\leqslant s\leqslant r$:

$$\left(\frac{\partial}{\partial r}+\frac{\partial}{\partial s}\right)h(\mathbf{x},\mathbf{y})=g(\mathbf{x},\mathbf{y})\,, \tag{4.6a}$$

$$\left(\frac{\partial}{\partial r}-\frac{\partial}{\partial s}\right)g(\mathbf{x},\mathbf{y})=H(\mathbf{x},\mathbf{y})\,, \tag{4.6b}$$

$$H(\mathbf{x},\mathbf{y})=V(\mathbf{x})h(\mathbf{x},\mathbf{y})+(\Delta^0_{\mathbf{y}}-\Delta^0_{\mathbf{x}})h(\mathbf{x},\mathbf{y})\,, \tag{4.6c}$$

$$h(\mathbf{x},0)\ \text{obtained from}\ \frac{\partial}{\partial s}h(\mathbf{x},\mathbf{y}=0)=0\,, \tag{4.6d}$$

$$V(\mathbf{x})=-2g(\mathbf{x},|\mathbf{y}|=|\mathbf{x}|)/r^2 \tag{4.6e}$$

is obtained from (3.21).

Here $\Delta^0_{\mathbf{x}}$ is the *transverse* radial Laplacian operator in spherical coordinates, which is

$$\Delta^0_{\mathbf{x}}=\frac{1}{r^2\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial}{\partial\theta}\right)+\frac{1}{r^2\sin^2\phi}\frac{\partial^2}{\partial\phi^2}. \tag{4.7}$$

The quantity $h(\mathbf{x},\mathbf{y})$ being computed in this algorithm is actually $rsh(\mathbf{x},\mathbf{y})$, where $h(\mathbf{x},\mathbf{y})$ is defined in (3.20) as the inverse Radon transform of the scattered field $m(\mathbf{x},t,\mathbf{e}_i)$. Multiplication by $rs=|\mathbf{x}|\,|\mathbf{y}|$ is a normalization that results in better numerical behavior near the origin.

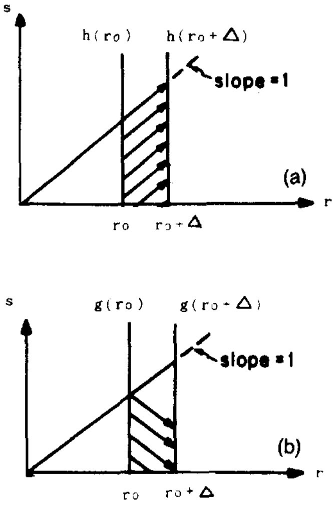

FIG. 5. (a) Recursion pattern for updating $h(r,s)$ in the fast algorithm for the 3-D regular problem. (b) Recursion pattern for updating $g(r,s)$ in the fast algorithm for the 3-D regular problem.

The recursion pattern for this algorithm is illustrated in Fig. 5. Note that since the radii r and s are both non-negative, the recursion pattern differs from the previous algorithm in that s is required to be non-negative. The only other significant difference is that computations need only be performed over the interior of the sphere of radius r, rather than over the infinite slab $-\mathbf{e}_i\cdot\mathbf{x}\leqslant t\leqslant\mathbf{e}_i\cdot\mathbf{x}$. This is a considerable advantage over the two preceding algorithms, both of which require computations over an infinite region in $\mathbf{y}$. However, (4.6e) still requires a considerable amount of computation at each recursion, although now the simpler integral equation (3.21) is used to compute $g(\mathbf{x},|\mathbf{x}|)$ from values of $g(\mathbf{x},\mathbf{y})$. This computation is absent in the reflection problem algorithm, since this problem has a causal structure that is more easily exploited.

The amount of computation required by the above algorithm for the regular problem is $O(N^8)$ operations. This is a significant reduction from the $O(N^{12})$ operations required to solve the generalized Gel'fand–Levitan integral equation. Note that the ratio of the exponents of the orders of computations required for the integral equation procedure to the differential procedure is the same in both one and three dimensions, viz., $\frac{12}{8}=\frac{3}{2}$. Also note that the layer stripping algorithm for the reflection problem requires only $O(N^6)$ computations. This is because the layer stripping reflection problem algorithm is initialized using near-field data, while the regular problem procedures all use far-field data in the form of the scattering amplitude [in order to compute the Jost function $J(k)$].

This algorithm is quite similar to the algorithm given in Ref. 13 for computation of the optimal filter for the linear, least-squares estimation of a homogeneous random field. Since the integral equation (3.21) looks much like a multidimensional Wiener–Hopf equation, this is not surprising. The form of (3.21) suggests that the well-known connection between inverse scattering and linear least-squares estimation that exists in one dimension[20] extends to higher dimensions. Details of this connection are given in Ref. 21 for isotropic random fields and spherically symmetric potentials, and in Ref. 22 for a more general class of random fields and nonspherically symmetric potentials.

V. CONCLUSION

This paper has presented a unified treatment of various differential and integral equation procedures for solving three-dimensional inverse scattering problems. The relation between the generalized Gel'fand–Levitan and Marchenko integral equations of Ref. 2 has been explored by noting that the former can be interpreted as an orthogonality principle with respect to an inner product defined in terms of a weighting function computed using an integral equation equivalent to the latter. The problems solved by the two integral equations, and the resulting scattering solutions, are complementary in their support. This is emphasized by the differential counterparts to the integral equation procedures, which require less computation since they directly exploit the causal structure of the inverse scattering problem.

An important feature of this presentation is the emphasis on how results for the one-dimensional inverse problem generalize to three dimensions. The parallels between Secs. II and III are remarkable, considering the greater complexity of the three-dimensional problem. These strong parallels in the derivations of both the integral equation procedures and their differential, fast algorithm counterparts suggest that the approach taken in this paper may be particularly insightful for further research.

Several topics developed in this paper require further research. The most important one is the connection between multidimensional inverse scattering and linear least-squares estimation of random fields. A useful starting point would be the characterization of the class of covariance functions that can be put in the form of (3.22). Connections between other exact inverse problem procedures and those of Ref. 2 should also be explored, in the spirit of Ref. 9; this could result in further insights and more fast algorithms. Finally, the numerical performances of all of these procedures need to be investigated.

[1]S. Coen, M. Cheney, and A. Weglein, "Velocity and density of a two-dimensional acoustic medium from point source surface data," J. Math. Phys. **25**, 1857 (1984).

[2]R. G. Newton, "Inverse scattering. II. Three dimensions," J. Math. Phys. **21**, 1698 (1980).

[3]Z. S. Agranovich and V. A. Marchenko, *The Inverse Problem of Scattering Theory* (Gordon and Breach, New York, 1963).

[4]I. M. Gel'fand and B. M. Levitan, "On the determination of a differential equation from its spectral function," Am. Math. Soc. Trans. Ser. 2, **1**, 253 (1955).

[5]L. D. Faddeev, "The inverse problem of quantum theory of scattering," J. Math. Phys. **4**, 72 (1963).

[6]K. Chadan and P. C. Sabatier, *Inverse Problems in Quantum Scattering Theory* (Springer, New York, 1977).

[7]J. H. Rose, M. Cheney, and B. DeFacio, "The connection between time and frequency domain three-dimensional inverse scattering methods," J. Math. Phys. **25**, 2995 (1984).

[8]J. H. Rose, M. Cheney, and B. DeFacio, "Three-dimensional inverse scattering: Plasma and variable-velocity wave equations," J. Math. Phys. **26**, 2803 (1985).

[9]A. E. Yagle, "Differential and integral methods for multidimensional inverse scattering problems," J. Math. Phys. **27**, 2584 (1986).

[10]B. D. O. Anderson and T. Kailath, "Fast algorithms for the integral equations of the inverse scattering problem," Int. J. Eqns. Operator Theory **1**, 132 (1978).

[11]A. M. Bruckstein, B. C. Levy, and T. Kailath, "Differential methods in inverse scattering," SIAM J. Appl. Math. **45**, 312 (1985).

[12]A. E. Yagle and B. C. Levy, "Layer stripping solutions of multidimensional inverse scattering problems," J. Math. Phys. **27**, 1701 (1986).

[13]A. E. Yagle, "A fast algorithm for linear estimation of three-dimensional homogeneous anisotropic random fields," in Proceedings of the International Conference on Acoustics, Speech, and Signal Processing, Dallas, Texas, 1987.

[14]B. C. Levy, "An orthonormalization formulation of the inverse scattering problem for two-component wave equations," Technical Report, Laboratory for Information and Decision Systems, MIT, Cambridge, MA, 02139.

[15]R. Burridge, "The Gelfand–Levitan, the Marchenko, and the Gopinath–Sondhi integral equations of inverse scattering theory, regarded in the context of inverse impulse-response problems," Wave Motion **2**, 305 (1980).

[16]G. N. Balanis, "The plasma inverse problem," J. Math. Phys. **13**, 1001 (1972).

[17]R. Carroll, "The integral equations of inverse scattering theory and the underlying transmutations," to appear in Acta Appl. Math.

[18]R. G. Newton, "Inverse scattering. III. Three dimensions, continued," J. Math. Phys. **22**, 2191 (1981).

[19]R. G. Newton, "Inverse scattering. IV. Three dimensions: Generalized Marchenko construction with bound states, and generalized Gel'fand–Levitan equations," J. Math. Phys. **23**, 594 (1982).

[20]P. Dewilde, J. T. Fokkema, and I. Widya, "Inverse scattering and linear prediction, the time continuous case," in *Stochastic Systems: Mathematics of Filtering and Identification Applications* (Reidel, Dordrecht, Holland, 1981).

[21]B. C. Levy and J. Tsitsiklis, "A fast algorithm for linear estimation of two-dimensional isotropic random fields," IEEE Trans. Inf. Theory. **IT-31**, 635 (1985).

[22]A. E. Yagle, "Connections between three-dimensional inverse scattering and the linear least-squares estimation of random fields," submitted to Acta Appl. Math.

Any physical, monopole equation of motion structure uniquely determines a projective inertial structure and an $(n-1)$-force

Robert Alan Coleman
Interdisciplinary Studies in Mathematical Physics and Philosophy of Science, University of Regina, Regina, Saskatchewan, Canada S4S 0A2

Herbert Korte
Interdisciplinary Studies in Philosophy of Science and Mathematical Physics, University of Regina, Regina, Saskatchewan, Canada S4S 0A2

(Received 30 July 1986; accepted for publication 26 November 1986)

It is proved that, in the context of a conformal causal structure, (a) any acceleration field decomposes uniquely into the sum of an affine structure that is compatible with the conformal structure and an n-force, and (b) any directing field, such that the n-force of the corresponding family of acceleration fields is due to tensor fields and is orthogonal to the n-velocity, uniquely decomposes into a projective structure that is compatible with the conformal structure and an $(n-1)$-force. Moreover, if there are no second clock effects and variable rest masses do not exist, there exists a unique pseudo-Riemannian metric on space-time that determines the unique standard of no acceleration for all massive monopoles. It follows from this that a non-null result for the Eötvös experiment entails the existence of a *fifth force* rather than a violation of the universality of free fall.

I. INTRODUCTION

We prove that the conformal causal structure of space-time reduces to a Weyl structure provided that there exists on space-time *either* an acceleration field $A^i_2(x^i, \gamma^i_1)$ *or* a directing field $\Xi^\alpha_2(x^i, \xi^\alpha_1)$ such that the n-force of the corresponding family of acceleration fields is due to tensor fields and is orthogonal to the n-velocity. Moreover, the Weyl structure and hence its affine and projective structures are unique unless there are second clock effects or variable rest masses or both. To put the significance of these results in perspective, we briefly outline some of our previous results.

According to the principle of the universality of free fall (UFF), the motions of all neutral monopole particles are governed by one common path structure. In a previous paper,[1] we formulated this principle as follows.

UFF: The set of all actually existing equation of motion structures for massive monopoles constitutes a one-parameter family of directing fields of the form

$$\Xi^\alpha_2(x^i, \xi^\alpha_1) = W^\alpha(x^i, \xi^\alpha_1) + (Q/m)\mathbb{F}^\alpha(x^i, \xi^\alpha_1), \qquad (1)$$

where W is a specific directing field and Q/m is the electromagnetic charge to mass ratio. The principle UFF does not require that the universal equation of motion structure $W^\alpha(x^i, \xi^\alpha_1)$ be geodesic, that is, cubic[2,3] in the variables ξ^α_1, denoting the three-velocity. However, if the special theory of relativity is valid in every sufficiently small region of space-time, then at every point of space-time the first-order part of the microsymmetry (invariance) group of the field W must contain a subgroup that is isomorphic to the Lorentz group. In our paper,[1] we proved that any second-order equation of motion structure, either an acceleration field $A^i_2(x^i, \gamma^i_1)$ or a directing field $\Xi^\alpha_2(x^i, \xi^\alpha_1)$, that satisfies this microsymmetry condition and is C^1 in its velocity variables (γ^i_1 or ξ^α_1), must be geodesic. Hence the field W that governs the motion of all neutral monopoles must be a projective structure. A theorem proved by Ehlers, Pirani, and Schild (EPS)[4] asserts that any projective structure that is causally compatible with the conformal structure of space-time determines a reduction of the conformal structure to a Weyl structure; that is, the two structures jointly determine a unique symmetric linear connection on space-time. A weaker theorem proved by Weyl[5,6] asserts that the conformal and projective structures determined by a given Weyl structure, in turn, uniquely determine that Weyl structure. See also Ref. 7, Sec. 8.

In Sec. II, we discuss the additional constraints on physical acceleration and directing fields due to the fact that the n-velocity of material particles must be future timelike, and define the concepts of n-force and $(n-1)$-force. The decomposition theorems for acceleration and directing fields are presented in Secs. III and IV, respectively. The significance of our results for the constructive axiomatics of the general theory of relativity (GTR) is noted in Sec. V. Finally, in Sec. VI, we discuss the implications that our results have for the interpretation of the Eötvös experiment and for the existence of a fifth force.

II. PHYSICAL ACCELERATION AND DIRECTING FIELDS

Let M denote the space-time manifold and let $\gamma: \mathbb{R} \to M$ be a curve in M such that $\gamma(0) = p$. Then the k-jet $j^k_0\gamma$ is the kth-order curve element determined by γ at $p \in M$. The set of such elements at $p \in M$ for all curves through p is denoted by $J^k(\mathbb{R}_0, M_p)$. The space of all curve elements forms an associated fiber bundle

$$\mathscr{J}^k(M) = \langle J^k(\mathbb{R}_0, M), \pi_k, M, J^k(\mathbb{R}_0, \mathbb{R}^n_0), G^k_n \rangle, \qquad (2)$$

where $J^k(\mathbb{R}_0, \mathbb{R}^n_0)$ is the typical fiber and G^k_n is the Lie group of k-jets $j^k_0 f$ of diffeomorphisms $f: \mathbb{R}^n \to \mathbb{R}^n$ such that $f(\mathbf{0}) = \mathbf{0}$. The natural projections

 0022-2488/87/071492-07$02.50

$$\pi^k_l: J^l(\mathbb{R}_0,M) \to J^k(\mathbb{R}_0,M), \quad \text{for} \quad k<l, \tag{3}$$

are defined by truncating the l-jets. An acceleration field is a map $A: J^1(\mathbb{R}_0,M) \to J^2(\mathbb{R}_0,M)$ such that

$$\pi^1_2 \circ A = \mathrm{id}_{J^1(\mathbb{R}_0,M)}. \tag{4}$$

Denote by (x^i,γ^i_1) and $(x^i,\gamma^i_1,\gamma^i_2)$ the coordinates of $j^1_p\gamma$ and $j^2_p\gamma$ with respect to some local chart $(U,x)_p$. Then the field A is determined locally by

$$A(x^i,\gamma^i_1) = (A^i_0(x^i,\gamma^i_1),A^i_1(x^i,\gamma^i_1),A^i_2(x^i,\gamma^i_1))$$
$$= (x^i,\gamma^i_1,A^i_2(x^i,\gamma^i_1)). \tag{5}$$

The corresponding second-order equation of motion for the curve $\gamma: \mathbb{R} \to M$ is

$$\ddot{\gamma}(\lambda) = A^i_2(\gamma(\lambda),\dot{\gamma}(\lambda)). \tag{6}$$

The causal structure of space-time is determined by a conformal structure that determines an equivalence class of pseudo-Riemannian metrics up to an arbitrary, possibly nonintegrable choice of gauge

$$\{e^{\alpha(x^i)}g_{ij}(x^i)\}. \tag{7}$$

Curve elements of world lines that describe the motion of physical particles must satisfy the condition that the n-velocity γ^i_1 be timelike and future pointing; that is,

$$g_{ij}\gamma^i_1\gamma^j_1 > 0 \quad \text{and} \quad \gamma^0_1 > 0. \tag{8}$$

The open subbundle of $\mathscr{J}^k(M)$ defined by this condition will be denoted by $\mathscr{H}^k(M)$. A *physical* acceleration field is thus defined by a map $A: H^1(M) \to H^2(M)$ such that (4) is satisfied.

Given two elements $(x^i,\gamma^i_1,\gamma^i_{2A})$ and $(x^i,\gamma^i_1,\gamma^i_{2B})$ of $H^2(M)$, one can form the geometric object $(x^i,\gamma^i_1,\Delta\gamma^i_2)$, where

$$\Delta\gamma^i_2 = \gamma^i_{2B} - \gamma^i_{2A}. \tag{9}$$

Under a change of space-time coordinates γ^i_1 and $\Delta\gamma^i_2$ transform in the same way; however, under a change of parameter, they transform according to

$$\tilde{\gamma}^i_1 = D\mu\gamma^i_1 \tag{10}$$

and

$$\Delta\tilde{\gamma}^i_2 = (D\mu)^2\Delta\gamma^i_2. \tag{11}$$

The bundle of geometric objects $(x^i,\gamma^i_1,\Delta\gamma^i_2)$ will be denoted by $\mathscr{F}(M)$. An n-force is determined by a map F: $H^1(M) \to F(M)$ such that

$$\pi^{H^1}_F \circ F = \mathrm{id}_{H^1}, \tag{12}$$

where

$$\pi^{H^1}_F(x^i,\gamma^i_1,\Delta\gamma^i_2) = (x^i,\gamma^i_1). \tag{13}$$

If $\mu: \mathbb{R} \to \mathbb{R}$ is a diffeomorphism such that $\mu(0) = 0$, then G^k_1 is the Lie group of k-jets $j^k_0\mu$ of such diffeomorphisms. This group acts on the fibers of $H^k(M)$ in a natural way. The equivalence classes determined by this group action are the elements of the bundle $\mathscr{D}^k(M) = \mathscr{H}^k(M)/G^k_1$ of kth-order *path* elements with total space $\mathbb{D}^k(M)$. If the space-time coordinates are $x^i = (t,x^\alpha)$, where t is the timelike coordinate, then a kth-order path element is described by $(t,x^\alpha,\xi^\alpha_1,\ldots,\xi^\alpha_k)$, where the coordinate ξ^α_r corresponds to the rth derivative of x^α with respect to t. The elements of $\mathbb{D}^k(M)$ satisfy the condition

$$g_{00} + 2g_{0\rho}\xi^\rho_1 + g_{\rho\sigma}\xi^\rho_1\xi^\sigma_1 > 0, \tag{14}$$

where $x^0 = t$.

A *physical* directing field is defined by a map Ξ: $\mathbb{D}^1(M) \to \mathbb{D}^2(M)$ such that

$$\pi^{\mathbb{D}^1}_{\mathbb{D}^2} \circ \Xi = \mathrm{id}_{\mathbb{D}^1}. \tag{15}$$

Such a field is locally described by

$$\Xi(x^i,\xi^\alpha_1) = (x^i,\xi^\alpha_1,\Xi^\alpha_2(x^i,\xi^\alpha_1)) \tag{16}$$

and determines the second-order equation of motion

$$\frac{d^2\xi^\alpha}{dt^2}(t) = \Xi^\alpha_2\left(t,\xi^\alpha(t),\frac{d\xi^\alpha}{dt}(t)\right) \tag{17}$$

for the path $t \to (t,\xi^\alpha(t))$.

Given two elements $(x^i,\xi^\alpha_1,\xi^\alpha_{2A})$ and $(x^i,\xi^\alpha_1,\xi^\alpha_{2B})$ of $\mathbb{D}^2(M)$, one can form the geometric object $(x^i,\xi^\alpha_1,\Delta\xi^\alpha_2)$, where

$$\Delta\xi^\alpha_2 = \xi^\alpha_{2B} - \xi^\alpha_{2A}. \tag{18}$$

Under a change of space-time coordinates, ξ^α_1 and $\Delta\xi^\alpha_2$ transform according to

$$\bar{\xi}^\alpha_1 = (\bar{X}^\alpha_0 + \bar{X}^\alpha_\beta\xi^\beta_1)/(\bar{X}^0_0 + \bar{X}^0_\gamma\xi^\gamma_1) \tag{19}$$

and

$$\Delta\bar{\xi}^\alpha_2 = \frac{\bar{X}^\alpha_\beta\Delta\xi^\beta_2(\bar{X}^0_0 + \bar{X}^0_\rho\xi^\rho_1) - \bar{X}^0_\beta\Delta\xi^\beta_2(\bar{X}^\alpha_0 + \bar{X}^\alpha_\rho\xi^\rho_1)}{(\bar{X}^0_0 + \bar{X}^0_\gamma\xi^\gamma_1)^3}. \tag{20}$$

The total space of the bundle of geometric objects $(x^i,\xi^\alpha_1,\Delta\xi^\alpha_2)$ will be denoted by $\mathbb{F}(M)$. An $(n-1)$-force is determined by a map $\mathbb{F}$: $\mathbb{D}^1(M) \to \mathbb{F}(M)$ such that

$$\pi^{\mathbb{D}^1}_{\mathbb{F}} \circ F = \mathrm{id}_{\mathbb{D}^1}. \tag{21}$$

Although a curve structure need not determine a path structure, to every path structure there corresponds a family of curve structures (Ref. 1, Theorem 3.1).

Theorem 1: An acceleration field A determines a directing field Ξ iff A is of the form

$$A^i_2(x^i,\gamma^i_1) = P(x^i,\gamma^i_1)\gamma^i_1 + I^i_2(x^i,\gamma^i_1), \tag{22}$$

where $I^i_2(x^i,\gamma^i_1)$ is orthogonal to γ^i_1 and

$$I^i_2(x^i,\lambda\gamma^i_1) = \lambda^2 I^i_2(x^i,\gamma^i_1). \tag{23}$$

In much of the following analysis, the variables x^i play the role of spectators. Functional dependence on these variables will be suppressed and indicated derivatives are with respect to the variables γ^i_1 unless explicitly stated otherwise.

It is useful to have a concrete example in mind. Consider the acceleration field defined by

$$A^i_2(\gamma^i_1) = (g_{rs}\gamma^r_1\gamma^s_1)^{1/2}T^i_{j_1}\gamma^{j_1}_1 + T^i_{j_1j_2}\gamma^{j_1}_1\gamma^{j_2}_1$$
$$+ (g_{rs}\gamma^r_1\gamma^s_1)^{-1/2}T^i_{j_1j_2j_3}\gamma^{j_1}_1\gamma^{j_2}_1\gamma^{j_3}_1 + \cdots$$
$$+ (g_{rs}\gamma^r_1\gamma^s_1)^{-(k-2)/2}T^i_{j_1\cdots j_k}\gamma^{j_1}_1\cdots\gamma^{j_k}_1. \tag{24}$$

This field is homogeneous of degree 2 in the variables γ^i_1 and therefore determines a directing field. Under a change of coordinates, $-T^{i}_{j_1 j_2}$ transforms like a symmetric linear connection and the other coefficients transform like tensor fields. The usual family of electromagnetic acceleration fields is the special case in which

$$T^{i}_{j_1 j_2} = -\Gamma^{i}_{j_1 j_2}, \tag{25}$$

where $\Gamma^{i}_{j_1 j_2}$ is the symmetric linear connection determined by the space-time metric g_{ij}, and

$$T^{i}_{j_1} = \mu F^{i}_{j_1}, \tag{26}$$

where $F^{i}_{j_1}$ is the electromagnetic field tensor and $\mu = Q/m$ is the specific charge.

For the acceleration field (24) with $T^{i}_{j_1 j_2} = -\Gamma^{i}_{j_1 j_2}$ the n-force is clearly

$$F^{i}(\gamma^{j}_{1}) \equiv A^{i}_{2}(\gamma^{j}_{1}) + \Gamma^{i}_{j_1 j_2}\gamma^{j_1}_{1}\gamma^{j_2}_{1}. \tag{27}$$

In order to rule out variable rest masses, the n-force must be orthogonal to the n-velocity,

$$g_{ab}\gamma^{a}_{1}F^{b}(\gamma^{j}_{1}) = 0. \tag{28}$$

It is easy to satisfy this condition by requiring that the fields $T_{ij_1\cdots j_r}$ (for $r\neq 2$) have the symmetry obtained by antisymmetrizing on the first two indices and then symmetrizing on the last r indices. This symmetry condition also guarantees that

$$F^{a}_{,a}(\gamma^{j}_{1}) = 0, \tag{29}$$

and therefore that

$$A^{a}_{2,a}(\gamma^{j}_{1}) = -2\Gamma_{k}\gamma^{k}_{1}, \tag{30}$$

where $\Gamma_k = \Gamma^{a}_{ak}$.

Even if the n-force has a more general form than that given by (24) and (27), it must still satisfy the orthogonality condition (28) because this condition is *necessary* (but not sufficient) for rest masses to be constant.

III. THE DECOMPOSITION OF PHYSICAL ACCELERATION FIELDS

In this section, it is shown that the conformal structure of space-time and an acceleration field together determine a symmetric linear connection and hence determine the free fall and force components of the acceleration field. The analysis does not apply to every mathematically conceivable acceleration field, but it does apply to a very large class of acceleration fields that includes all acceleration fields with force terms due to tensor fields.

At each space-time point, a conformal structure determines a pseudo-Riemannian metric up to an arbitrary, possibly nonintegrable choice of gauge

$$\{e^{\alpha(x^{i})}g_{ij}(x^{i})\}, \tag{31}$$

and an equivalence class of symmetric linear connections, such that the conformal structure is preserved under the parallel transport of each of these connections, namely,

$$\Gamma^{i}_{jk} = K^{i}_{jk} + (1/n)(\delta^{i}_{j}\Gamma_{k} + \delta^{i}_{k}\Gamma_{j} - g_{jk}g^{ir}\Gamma_{r}), \tag{32}$$

where the trace of the connection Γ_k is arbitrary.

Remark: The geometric object determined by the Γ_k may reasonably be called a "volume connection" since the Γ_k determine a principal connection on the bundle of volume elements. ■

The conformal connection coefficients are given by

$$2K^{i}_{jk} = g^{ir}(g_{rj,k} + g_{rk,j} - g_{jk,r}) - (1/n)(\delta^{i}_{j}\varphi_{k} + \delta^{i}_{k}\varphi_{j} - g_{jk}g^{ir}\varphi_{r}), \tag{33}$$

where

$$\varphi_k = g^{rs}g_{rs,k}, \tag{34}$$

and the indicated derivatives are with respect to the suppressed variables x^{i}. As the standard representative of the equivalence class (31), we choose $\wp_{ij}(x^{i})$, such that $\det(\wp_{ij}(x^{i})) = -1$. Let $C(\gamma^{j}_{1})$ transform as $\Gamma_{k}\gamma^{k}_{1}$ does under a change of local coordinates; that is,

$$\bar{C}(\bar{\gamma}^{j}_{1}) = C(\gamma^{j}_{1}) - \bar{X}_{s}^{-1r}\bar{X}^{s}_{rk}\gamma^{k}_{1}. \tag{35}$$

In view of the structure of (32), it is reasonable to require that the field that governs free motion be of the form

$$B^{i}_{2}(\gamma^{j}_{1}) \equiv -K^{i}_{jk}\gamma^{j}_{1}\gamma^{k}_{1} - (2/n)\gamma^{i}_{1}C_{,k}(\gamma^{j}_{1})\gamma^{k}_{1} + (1/n)(g_{jk}\gamma^{j}_{1}\gamma^{k}_{1})g^{ir}C_{,r}(\gamma^{j}_{1}). \tag{36}$$

Such a field gives the correct equations of motion for light rays and transforms in the same way as an acceleration field $A^{i}_{2}(\gamma^{j}_{1})$ does; namely,

$$\bar{A}^{i}_{2}(\bar{\gamma}^{j}_{1}) = \bar{X}^{i}_{j}A^{j}_{2}(\gamma^{j}_{1}) + \bar{X}^{i}_{jk}\gamma^{j}_{1}\gamma^{k}_{1}. \tag{37}$$

A general acceleration field may then be written as the sum of the field (36) and an n-force,

$$A^{i}_{2}(\gamma^{j}_{1}) \equiv B^{i}_{2}(\gamma^{j}_{1}) + F^{i}(\gamma^{j}_{1}). \tag{38}$$

One obtains from (36), (38), and the orthogonality condition (28) the equation

$$C_{,k}(\gamma^{j}_{1})\gamma^{k}_{1} = -n\left[\frac{g_{ab}\gamma^{a}_{1}K^{b}_{jk}\gamma^{j}_{1}\gamma^{k}_{1}}{g_{pq}\gamma^{p}_{1}\gamma^{q}_{1}} + P(\gamma^{j}_{1})\right], \tag{39}$$

where

$$P(\gamma^{j}_{1}) \equiv \frac{g_{ab}\gamma^{a}_{1}A^{b}_{2}(\gamma^{j}_{1})}{g_{pq}\gamma^{p}_{1}\gamma^{q}_{1}}. \tag{40}$$

Theorem 2: Given an acceleration field $A^{i}_{2}(\gamma^{j}_{1})$, define

$$C(\gamma^{j}_{1}) \equiv -n\left[\frac{g_{ab}\gamma^{a}_{1}K^{b}_{jk}\gamma^{j}_{1}\gamma^{k}_{1}}{g_{pq}\gamma^{p}_{1}\gamma^{q}_{1}} + E(\gamma^{j}_{1})\right], \tag{41}$$

where

$$E(\gamma^{j}_{1}) = S(\gamma^{j}_{1}) + \int_{0}^{1}\frac{P(\lambda\gamma^{j}_{1})}{\lambda}\,d\lambda \tag{42}$$

and $S: H^{1}(M)\to\mathbb{R}$ satisfies

$$S(\lambda\gamma^{j}_{1}) = S(\gamma^{j}_{1}). \tag{43}$$

If the integral in (42) exists, then the acceleration field $B^{i}_{2}(\gamma^{j}_{1})$ determined by (36), (41), and (42) determines an n-force given by

$$F^{i}(\gamma^{j}_{1}) \equiv A^{i}_{2}(\gamma^{j}_{1}) - B^{i}_{2}(\gamma^{j}_{1}), \tag{44}$$

which satisfies the orthogonality condition

$$g_{ab}\gamma^{a}_{1}F^{b}(\gamma^{j}_{1}) = 0. \tag{45}$$

Proof: It must be shown that (41) is the most general solution of (39). Since

$$\frac{d}{d\lambda}C(\lambda\gamma^{j}_{1}) = \gamma^{k}_{1}C_{,k}(\lambda\gamma^{j}_{1}), \tag{46}$$

it follows that

$$\frac{d}{d\lambda}C(\lambda\gamma_1^i) = -n\left[\frac{g_{ab}\gamma_1^a K^b_{jk}\gamma^j_1\gamma_1^k}{g_{pq}\gamma_1^p\gamma_1^q} + \frac{P(\lambda\gamma_1^i)}{\lambda}\right]. \tag{47}$$

Define

$$E(\gamma_1^i) \equiv C(\gamma_1^i) + n\frac{g_{ab}\gamma_1^a K^b_{jk}\gamma^j_1\gamma_1^k}{g_{pq}\gamma_1^p\gamma_1^q}. \tag{48}$$

Then,

$$\frac{d}{d\lambda}E(\lambda\gamma_1^i) = \frac{P(\lambda\gamma_1^i)}{\lambda}, \tag{49}$$

and

$$E(\lambda\gamma_1^i) = \text{const} + \int_0^\lambda \frac{P(\mu\gamma_1^i)}{\mu}\,d\mu, \tag{50}$$

where const may depend on the *direction* from which $\gamma_1^i \to \mathbf{0}$; that is, const may be a function that satisfies (43) and hence satisfies the homogeneous equation

$$\gamma_1^k S_{,k}(\gamma_1^i) = 0. \tag{51}$$

Note that $S(\gamma_1^i)$ contributes to the field $B^i_2(\gamma_1^i)$ only if $S_{,k}(\gamma_1^i) \neq 0$. Hence, the general solution of Eq. (39) is given by (48), where

$$E(\gamma_1^i) = S(\gamma_1^i) + \int_0^1 \frac{P(\lambda\gamma_1^i)}{\lambda}\,d\lambda. \tag{52}$$

Finally, note that the field $P(\gamma_1^i)$ transforms according to

$$\overline{P}(\bar{\gamma}_1^i) = P(\gamma_1^i) + \frac{g_{ab}\gamma_1^a \overline{X}^b_{jk}\gamma^j_1\gamma_1^k}{g_{pq}\gamma_1^p\gamma_1^q}. \tag{53}$$

Since

$$\overline{S}(\bar{\gamma}_1^i) = S(\gamma_1^i) \tag{54}$$

and

$$\int_0^1 \frac{1}{\lambda}\frac{g_{ab}\lambda\gamma_1^a \overline{X}^b_{jk}\lambda\gamma^j_1\lambda\gamma_1^k}{g_{pq}\lambda\gamma_1^p\lambda\gamma_1^q}\,d\lambda = \frac{g_{ab}\gamma_1^a \overline{X}^b_{jk}\gamma^j_1\gamma_1^k}{g_{pq}\gamma_q^p\gamma_1^q}, \tag{55}$$

it follows that $E(\gamma_1^i)$ obeys the same transformation law that $P(\gamma_1^i)$ obeys. ■

Remark: For the class of acceleration fields (24) that satisfies (28),

$$P(\gamma_1^i) = g_{ab}\gamma_1^a T^b_{jk}\gamma^j_1\gamma_1^k / g_{pq}\gamma_1^p\gamma_1^q. \tag{56}$$

Moreover, the field $B^i_2(\gamma_1^i)$ will in this case be an affine structure provided that $S(\gamma_1^i)$ is chosen to be zero in which case $E(\gamma_1^i) = P(\gamma_1^i)$. ■

The acceleration field $B^i_2(\gamma_1^i)$ is not uniquely determined by the condition that the n-force be orthogonal to the n-velocity. However, if this field governs force-free motion and therefore represents the inertial structure of space-time, which is an aspect of the geometry of space-time in the general theory of relativity, then it must satisfy additional constraints. The essential features of the special theory of relativity are incorporated into the general theory of relativity by the requirement that at each point $p\in M$, the microsymmetry group[1,2] of the space-time metric is isomorphic to the Lorentz group, $SO(1,n-1)$. A consequence of this requirement is that derivative geometric structures, such as the conformal, affine, and projective structures, have microsymmetry groups at each $p\in M$, the first-order part of which contains a *subgroup* that is isomorphic to the group $SO(1,n-1)$.

Definition: A space-time geometric structure field is compatible with the special theory of relativity iff its microsymmetry group at each point $p\in M$ has a first-order part that contains a subgroup that is isomorphic to the group $SO(1,n-1)$. ■

In a previous paper,[1] we proved the following result.

Theorem 3: If an acceleration field $B^i_2(\gamma_1^i)$ is compatible with the special theory of relativity, then it is geodesic.

Remark: Our proof requires that $B^i_2(\gamma_1^i)$ be C^1 on its domain of definition (8) and that

$$\lim_{\lambda\to 0} B^i_{2,k}(\lambda\gamma_1^i) = 0. \tag{57}$$

These conditions are satisfied by the directing field (24). ■

If it is required that the field $B^i_2(\gamma_1^i)$ in the decomposition Theorem 2 above represent the geometric inertial field, then it must be compatible with the special theory of relativity; consequently, by Theorem 3, it must be geodesic. It follows that $C_{,k}(\gamma_1^i)$ can not depend on the variables γ_1^i, that $S(\gamma_1^i)$ can not depend on the variables γ_1^i (and therefore may be set to zero without affecting the decomposition), and that $P(\gamma_1^i)$ must be homogeneous of degree 1 in the variables γ_1^i. With this additional restriction on the class of acceleration fields, we have the following result.

Theorem 4: Let $A^i_2(\gamma_1^i)$ be an acceleration field such that the right-hand side of (39) is linear in the variables γ_1^i. Then, this acceleration field together with the conformal causal structure of space-time uniquely determines a symmetric linear connection that is compatible with the conformal structure and an n-force that is orthogonal to the n-velocity γ_1^i.

IV. THE DECOMPOSITION OF PHYSICAL DIRECTING FIELDS

It has been noted in Theorem 1 of Sec. II that a directing field Ξ determines only a family of acceleration fields. Since the term proportional to γ_1^i is arbitrary, the field $P(x^i,\gamma_1^i)$ given by (40) is not known; consequently, the method used in Theorem 2 of Sec. III to define $C(\gamma_1^i)$ is not applicable in the directing field case. However, the analysis can be modified, for an important class of directing fields to be described below, to yield a unique decomposition of a directing field into the sum of a projective structure Π and an $(n-1)$-force $\mathbf{F}$.

According to Theorem 1 of Sec. II, the field $I^i_2(x^i,\gamma_1^i)$, which satisfies

$$g_{ij}\gamma_1^i I^j_2(x^i,\gamma_1^i) = 0, \tag{58}$$

may be used as the standard representative of the equivalence class of acceleration fields determined by the directing field Ξ. However, the transformation law of the field $I^i_2(x,\gamma_1^i)$ is rather complicated and it is preferable to use the field

$$\Psi^i_2(\gamma_1^i) \equiv I^i_2(\gamma_1^i) - [1/(n+1)]I^a_{2,a}(\gamma_1^i)\gamma_1^i, \tag{59}$$

which satisfies

$$\Psi^a_{2,a}(\gamma^i_1) = 0, \tag{60}$$

as the standard representative. Under a change of local coordinates, the field Ψ transforms according to

$$\overline{\Psi}^i_2(\overline{\gamma}^i_1) = \overline{X}^i_t[\Psi^t_2(\gamma^i_1) + \{\overline{X}_s^{-1t}\overline{X}^s_{jk} - [1/(n+1)] \times(\delta^t_j\overline{X}_s^{-1r}\overline{X}^s_{rk} + \delta^t_k\overline{X}_s^{-1r}\overline{X}^s_{rj})\}\gamma^j_1\gamma^k_1]. \tag{61}$$

This transformation law directly corresponds to the well-known transformation law for the coefficients of a projective structure.

A projective structure Π determines an equivalence class of symmetric linear connections that are compatible with the projective structure. These symmetric linear connections are given by

$$\Gamma^i_{jk} = \Pi^i_{jk} + [1/(n+1)](\delta^i_j\Gamma_k + \delta^i_k\Gamma_j), \tag{62}$$

where the trace of the connection Γ_k is arbitrary. From (32) and (62), it follows that a conformal structure determines an equivalence class of projective structures given by

$$\Pi^i_{jk} = K^i_{jk} + [1/n(n+1)](\delta^i_j\Gamma_k + \delta^i_k\Gamma_j) - (1/n)g_{jk}g^{ir}\Gamma_r. \tag{63}$$

It is straightforward to show that

$$C(\gamma^i_1) \equiv \frac{n(n+1)}{n-1}\left[\frac{g_{ab}\gamma^a_1(K^b_{jk}\gamma^j_1\gamma^k_1 + \Psi^b_2(\gamma^i_1))}{g_{pq}\gamma^p_1\gamma^q_1}\right] \tag{64}$$

transforms according to (35). The expression (63) motivates the requirement that the field that governs force-free motion be of the form

$$\chi^i_2(\gamma^i_1) \equiv -K^i_{jk}\gamma^j_1\gamma^k_1 - [2/n(n+1)]\gamma^i_1C_{,k}(\gamma^i_1)\gamma^k_1 + (1/n)(g_{jk}\gamma^j_1\gamma^k_1)g^{ir}C_{,r}(\gamma^i_1). \tag{65}$$

The field $\chi^i_2(\gamma^i_1)$ has the same transformation law (61) as $\Psi^i_2(\gamma^i_1)$ does. One can readily show that the n-force defined by

$$F^i(\gamma^i_1) \equiv \Psi^i_2(\gamma^i_1) - \chi^i_2(\gamma^i_1) \tag{66}$$

satisfies the orthogonality condition

$$g_{ab}\gamma^a_1F^b(\gamma^i_1) = 0. \tag{67}$$

Unfortunately, the decomposition (66) is not unique. The projective transformation

$$\tilde{\Psi}^i_2(\gamma^i_1) = \Psi^i_2(\gamma^i_1) + [1/(n+1)]\gamma^i_1\omega(\gamma^i_1), \tag{68}$$

where $\omega: H^1(M) \to \mathbb{R}$ and

$$\omega(\lambda\gamma^i_1) = \lambda\omega(\gamma^i_1), \tag{69}$$

does not change the directing field Ξ, which is given by

$$\Xi^\alpha_2(\xi^\alpha_1) = [\gamma^0_1\Psi^\alpha_2(\gamma^i_1) - \gamma^\alpha_1\Psi^0_2(\gamma^i_1)]/(\gamma^0_1)^3. \tag{70}$$

One finds, however, that

$$\tilde{C}(\gamma^i_1) = C(\gamma^i_1) + [n/(n+1)]\omega(\gamma^i_1), \tag{71}$$

$$\tilde{\chi}^i_2(\gamma^i_1) = \chi^i_2(\gamma^i_1) - [2/(n-1)(n+1)]\gamma^i_1\omega(\gamma^i_1) + [1/(n-1)](g_{jk}\gamma^j_1\gamma^k_1)g^{ir}\omega_{,r}(\gamma^i_1) \tag{72}$$

and

$$\tilde{F}^i(\gamma^i_1) = F^i(\gamma^i_1) + [1/(n-1)]\gamma^i_1\omega(\gamma^i_1) - [1/(n-1)](g_{jk}\gamma^j_1\gamma^k_1)g^{ir}\omega_{,r}(\gamma^i_1). \tag{73}$$

The terms proportional to γ^i_1 do not affect the decomposition of the directing field Ξ, but the terms proportional to $g^{ir}\omega_{,r}(\gamma^i_1)$ do. Fortunately, this arbitrariness in the decomposition of a directing field can be eliminated for a large and important class of directing fields, namely, the class of directing fields given by

$$\Xi^\alpha_2(\xi^\alpha_1) = [\gamma^0_1A^\alpha_2(\gamma^i_1) - \gamma^\alpha_1A^0_2(\gamma^i_1)]/(\gamma^0_1)^3, \tag{74}$$

where

$$A^i_2(\gamma^i_1) = (\wp_{rs}\gamma^r_1\gamma^s_1)^{1/2}\mathscr{T}^i_{j_1}\gamma^{j_1}_1 + T^i_{j_1j_2}\gamma^{j_1}_1\gamma^{j_2}_1 + (\wp_{rs}\gamma^r_1\gamma^s_1)^{-1/2}\mathscr{T}^i_{j_1j_2j_3}\gamma^{j_1}_1\gamma^{j_2}_1\gamma^{j_3}_1 + \cdots + (\wp_{rs}\gamma^r_1\gamma^s_1)^{-(k-2)/2}\mathscr{T}^i_{j_1\cdots j_k}\gamma^{j_1}_1\cdots\gamma^{j_k}_1, \tag{75}$$

the $\wp_{ij}$ are the conformal coefficients and the $\mathscr{T}_{ij_1j_2\cdots j_r}$ for $r \neq 2$ are tensor densities of appropriate weight that have first been antisymmetrized on the first two indices and then symmetrized on the last r indices. Directing fields that do not belong to this class are matematically conceivable; however, they do not seem plausible from a physical point of view. The following theorem summarizes the above discussion.

Theorem 5: Let Ξ be a directing field given by (74) and (75). This directing field uniquely determines and is uniquely determined by the standard representative $\Psi^i_2(\gamma^i_1)$, which satisfies (60). If $C(\gamma^i_1)$ given by (64) is linear in the variables γ^i_1, then the field $\Psi^i_2(\gamma^i_1)$ has the unique decomposition

$$\Psi^i_2(\gamma^i_1) = \chi^i_2(\gamma^i_1) + F^i(\gamma^i_1), \tag{76}$$

where the projective structure standard representative χ is given by (65) and satisfies

$$\chi^a_{2,a}(\gamma^i_1) = 0, \tag{77}$$

and the n-force $F^i(\gamma^i_1)$ satisfies the conditions (67) and

$$F^a_{,a}(\gamma^i_1) = 0. \tag{78}$$

To the decomposition (76), there corresponds the unique decomposition

$$\Xi^\alpha_2(\xi^\alpha_1) = \Pi^\alpha_2(\xi^\alpha_1) + \mathbb{F}^\alpha(\xi^\alpha_1), \tag{79}$$

where Π is the projective structure corresponding to χ and $\mathbb{F}$ is the $(n-1)$-force corresponding to the n-force F.

Remark: Note that for the class of directing fields considered, the condition (78) that ensures the uniqueness of the decomposition (79) is a consequence of the orthogonality condition (67). Also, the field $C(\gamma^i_1)$ will fail to be linear in the variables γ^i_1 only if the projective part of the directing field is incompatible[4] with the conformal structure. ■

V. THE CONSTRUCTIVE AXIOMATICS OF GTR

Ehlers, Pirani, and Schild have proposed a set of constructive axioms for the general theory of relativity.[4] One of their axioms, the projective axiom, asserts the existence of a path structure $\mathscr{P}_f$, the members of which are the possible world line paths of "freely falling" particles. They were, however, unable to provide an effective and noncircular procedure for measuring the geodesic directing field Π which uniquely determines and is uniquely determined by the path structure $\mathscr{P}_f$. We presented the solution to this difficulty in a terse form in Ref. 2 and in greater detail in Ref. 8. Our solution provides noncircular, empirical procedures for the identification of monopoles, for the separation of monopole particles into distinct classes each of which corresponds to a particular path structure, for the measurement of these path structures and for the testing of a given path structure for geodesicity. Thus our results show that the projective axiom of EPS is directly testable and hence truly constructive provided that there exists a class of *neutral* monopoles governed by the projective structure of space-time. In our more detailed presentation (Ref. 8, p. 171), we show that the projective structure of space-time can be measured by measuring the directing fields of electrically charged monopoles for at least two distinct charge to mass ratios.

Theorem 5 of Sec. IV greatly simplifies the problem of measuring the projective structure of space-time. One need only measure the directing field Ξ corresponding to any *one* kind of monopole particle regardless of the type of charge(s) this kind of particle may have. Thus, rather than having to measure the projective structure as part of a general parametric (charge to mass ratio) analysis of the family of directing fields, one can separately decompose *each* directing field into the sum of the projective structure and an $(n-1)$-force field. The parametric structure of the family of force fields so obtained may then be analyzed as a separate problem.

VI. IMPLICATIONS FOR THE EÖTVÖS EXPERIMENT AND THE FIFTH FORCE

Our results radically modify the interpretation of the Eötvös experiment. The traditional view is that a null result for this experiment establishes the principle of the universality of free fall. But, the analysis of Sec. IV shows that the equation of motion structure of a massive monopole uniquely decomposes into the sum of an $(n-1)$-force and a geodesic directing field or projective structure that is causally compatible with the conformal structure of space-time. Thus one of the two following possibilities holds: (1) the equation of motion structures of monopole particles *all* have the same projective component; or (2) there are at least two distinct projective structures which are the projective components of an equation of motion structure for some monopole particles.

If case (1) holds, there exists a unique projective structure on space-time that is compatible with the conformal structure of space-time and hence there exists a unique Weyl structure on space-time. Moreover, if there is no second clock effect, the Weyl structure reduces to a Riemannian structure.

In connection with the second possibility, note that if Γ is an affine structure that is causally compatible with the conformal structure of space-time, then any other such affine structure is a member of the one-parameter (λ) family of acceleration fields given by

$$A^{i}_{2}(\gamma^{i}_{1}) = -\Gamma^{i}_{jk}\gamma^{j}_{1}\gamma^{k}_{1} + \lambda S^{i}_{jk}\gamma^{j}_{1}\gamma^{k}_{1}, \tag{80}$$

where S^{i}_{jk} is a tensor field of the form

$$S^{i}_{jk} = (1/n)(\delta^{i}_{j}\omega_k + \delta^{i}_{k}\omega_j - g_{jk}g^{ir}\omega_r), \tag{81}$$

where ω_k is a covector field. Every acceleration field in the family given by (80) and (81) determines a distinct Weyl structure. If $S^{i}_{jk} \neq 0$ and does not have the form (81), the equation of motion determined by the field (80) for $\lambda \neq 0$ has solution curves that "break the light barrier."[1,4]

Possibility (2) asserts that there exist at least two members of the family given by (80) and (81). Note also that the family is determined by any two of its members. There are three cases to consider. First, if none of these Weyl structures reduces to a Riemannian structure, then no member of the family (80) is in any way singled out, and hence there is *no* natural zero point for the parameter λ and no reason to decompose the affine structures into a particular affine structure and an n-force. Moreover, in this case there is a *second clock effect* for each λ. Second, if one of the symmetric linear connections satisfies

$$\Gamma_{j,k} - \Gamma_{k,j} = 0, \tag{82}$$

but

$$\omega_{j,k} - \omega_{k,j} \neq 0, \tag{83}$$

then the connection Γ^{i}_{jk} is Riemannian and determines the natural zero point for the parameter λ. It is then natural to decompose the other members of the family into this Riemannian affine structure and an n-force. Such a one-parameter family of acceleration fields cannot be ruled out by simply demanding Lorentz microcovariance any more than the one-parameter family (25) and (26) of electromagnetic acceleration fields can be so ruled out. Unlike the electromagnetic case, however, the rest mass of any of the monopoles for which $\lambda \neq 0$ varies as judged by the standard provided by the Riemannian metric. The rate of change of the rest mass is proportional to

$$g_{ij}\gamma^{i}_{1}\gamma^{j}_{1}\omega_k\gamma^{k}_{1}. \tag{84}$$

Thus the phenomenon of variable rest mass exists unless ω_k vanishes identically in which case the Riemannian metric is unique. Finally, if (82) holds and

$$\omega_{j,k} - \omega_{k,j} = 0 \tag{85}$$

as well, then there is again no natural zero point for the parameter λ. All of the Weyl structures are, in this case, Riemannian structures. The phenomenon of variable rest mass still occurs in a mutual form unless $\omega_k = \varphi_{,k}$ vanishes identically.

Thus if there is no second clock effect and there are no variable rest masses, and if, moreover, the result of the Eötvös experiment is *not* null, then case (1) holds and there necessarily exists a fifth force because there exist at least two nonelectromagnetic equation of motion structures at most one of which can be purely geodesic. Whether or not one would choose to call such a fifth force "gravitational" depends on the details of the model proposed. In any case, the principle UFF, the existence of a universal standard of no acceleration, is not in question.

Remark: Our work is clearly relevant to Shiff's conjecture, which states that the principle UFF entails the Einstein equivalence principle. See Lightman and Lee[9] and Ni.[10] ■

Moreover, the recent reanalysis[11] of the experimental results of Eötvös, Pekar, and Fekete[12] (EPF) indicate that the result of the Eötvös experiment is not null. They suggest that there may be an additional, intermediate range, vector coupling to hypercharge. Such a coupling could also account for some other subtle effects in the $K^0 - \bar{K}^0$ system; however, a coupling to baryon number, which is conserved, would also account for the EPF data. Moffat[13] has recently pointed out that his theory of gravitation can also account for the EPF data. All of these proposals are in agreement with case (1).

ACKNOWLEDGMENT

We wish to express our appreciation to Professor Jürgen Ehlers for his thoughtful criticism of an earlier version of this paper which led us to substantially reformulate and strengthen our arguments.

[1]R. A. Coleman and H. Korte, J. Math. Phys. **25**, 3513 (1984).

[2]R. A. Coleman and H. Korte, J. Math. Phys. **21**, 1340 (1980); **23**, 345 (E) (1982).

[3]J. Ehlers and E. Köhler, J. Math. Phys. **18**, 2014 (1977).

[4]J. Ehlers, F. A. E. Pirani, and A. Schild, "The geometry of free fall and light propagation," in *General Relativity, Papers in Honour of J. L. Synge*, edited by L. O'Raifeartaigh (Oxford U. P., Oxford, 1972).

[5]H. Weyl, Nach. Ges. Wiss. Goettingen, Math.-Phys. Kl. **1921**, 99; *Gesammelte Abhandlungen Band II*, edited by K. Chandrasekharan (Springer, Berlin, 1968), Vol. II, No. 43, p. 195.

[6]H. Weyl, *Mathematische Analyse des Raumproblems* (Springer, Berlin, 1923).

[7]R. A. Coleman and H. Korte, J. Math. Phys. **22**, 2598 (1981).

[8]R. A. Coleman and H. Korte, "A realist field ontology of the causal-inertial structure (the refutation of geometric conventionalism)," University of Regina preprint, 1981, extended version, 1984 (192 pages); final version, entitled "The philosophical and mathematical foundations of space-time theories," *The Synthese Library* (Reidel) Dordrecht, to appear.

[9]A. P. Lightman and D. L. Lee, Phys. Rev. D **8**, 364 (1973).

[10]W.-T. Ni, Phys. Rev. Lett. **38**, 301 (1977).

[11]E. Fishbach, D. Sudarsky, A. Szafer, C. Talmadge, and S. H. Aronson, Phys. Rev. Lett. **56**, 3 (1986).

[12]R. V. Eötvös, D. Pekár, and E. Fekete, Ann. Phys. (Leipzig) **68**, 11 (1922).

[13]J. W. Moffat, "Test particle motion in the nonsymmetric theory," to be published in Phys. Rev.

Canonical structures for dispersive waves in shallow water

Fahrünisa Neyzi and Yavuz Nutku
TÜBİTAK Research Institute for Basic Sciences, Gebze, Kocaeli, Turkey and Bosphorus University, Istanbul, Turkey

(Received 6 March 1985; accepted for publication 14 January 1987)

The canonical Hamiltonian structure of the equations of fluid dynamics obtained in the Boussinesq approximation are considered. New variational formulations of these equations are proposed and it is found that, as in the case of the KdV equation and the equations governing long waves in shallow water, they are degenerate Lagrangian systems. Therefore, in order to cast these equations into canonical form it is again necessary to use Dirac's theory of constraints. It is found that there are primary and secondary constraints which are second class and it is possible to construct the Hamiltonian in terms of canonical variables. Among the examples of Boussinesq equations that are discussed are the equations of Whitham–Broer–Kaup which Kupershmidt has recently expressed in symmetric form and shown to admit tri-Hamiltonian structure.

I. INTRODUCTION

The equations governing the propagation of long waves in shallow water consist of a pair of coupled first-order partial differential equations which can be interpreted as a Hamiltonian system in several different ways. First, Luke's variational principle[1] for these equations was cast into canonical form by Zakharov,[2] Broer,[3] and Miles.[4] But with this approach it was not possible to obtain an explicit expression for the exact Hamiltonian in terms of canonical variables. Recently a new formulation[5] of these equations in terms of potentials led to the construction of the requisite Hamiltonian through the use of Dirac's theory of constraints.[6,7] Dirac's theory plays a crucial role for casting an overwhelming majority of the equations of fluid dynamics into canonical form because they turn out to be degenerate Lagrangian systems. In particular, the Hamiltonian for the KdV equation[8] was obtained by an application of Dirac's theory of constraints and in this paper we shall show that it can be used to cast the Boussinesq equations into canonical form as well. An alternative approach to the Hamiltonian structure of fluid equations is based on the Poisson bracket Gardner[9] has introduced for the KdV equation which, as Macfarlane,[10] Bergvelt and DeKerf,[11] and Lund[12] have shown, is equivalent to the Dirac bracket. This generalized Hamiltonian formalism has led to interesting new developments such as the theory of bi-Hamiltonian structure through the work of Lenard,[13] Olver,[14] Magri,[15] Gel'fand and Dorfman,[16] and Fokas and Fuchssteiner,[17] but it has some strange features from the vantage point of field theory. It does not, for example, employ the full set of canonical variables. The gap between these two approaches is bridged by Dirac's theory of constraints. Most recently Olver[18] has clarified this relationship by giving a proof of Darboux's theorem for first-order Hamiltonian operators.

The prototype of a field theory where all these structures emerge is the theory of long surface waves in shallow water. It will be of interest to find out which properties of this system of equations are stable in the sense that they survive in an appropriately generalized form when the equations are modified to take into account new effects. To this end we shall now consider the theory of dispersive waves in shallow water which are grouped together under the title of "Boussinesq equations." We shall briefly discuss the variety of fluid equations which are obtained in the Boussinesq approximation and choose two sets of equations on which we shall concentrate our attention for the rest of this paper. Among them are the equations of Whitham,[19] Broer,[20] and Kaup[21] (hereafter to be referred to as WBK). Recently Kupershmidt[22] has found a transformation whereby these equations assume a symmetric form. Kupershmidt's equations admit tri-Hamiltonian structure.[22] We shall construct new variational principles for these equations and find that they are degenerate Lagrangian systems. Applying Dirac's theory we obtain primary and secondary constraints all of which are second class and construct Dirac's total Hamiltonian which yields the canonical formulation of these Boussinesq equations. All of the equations of fluid dynamics in the Boussinesq approximation can be cast into canonical form using Dirac's theory of constraints, but Kupershmidt's equations occupy a privileged position among them because only in this case does the rich structure of the shallow water equations survive intact in every respect.

II. BOUSSINESQ EQUATIONS

We refer to Whitham,[23] Bona and Smith,[24] and Olver[25] for complete discussions of the issues in the Boussinesq approximation. The choice of equations of motion for a fluid in this approximation may be summarized in the following two points. The horizontal velocity field u is a function of the depth y as well as being a function of t,x while the surface elevation h depends only on t,x. In place of y we shall use θ which is normalized so that the undisturbed depth is given by $\theta = 1$. The form of the dispersive terms in the equations of motion changes depending on the choice of θ. The second ambiguity stems from the fact that we are making an approximation which is not disturbed by adding to the terms of second order, second derivatives in either t or x of the first-order terms. This gives rise to 12 free parameters after fixing the depth θ. There are physical and mathematical criteria for cutting down these possibilities. First of all we shall require that the equations of motion must be derivable from a variational principle. If we were to allow time derivatives of an order higher than the first which we find in the limit of zero dispersion, then the character of these equations changes

 0022-2488/87/071499-06$02.50

drastically as far as the initial-value problem is concerned. This is a particularly important point as we are primarily interested in the Hamiltonian formulation of these equations. Thus we shall further require that the Cauchy data for these equations should essentially be the same as that of the shallow water theory. This eliminates equations which contain second- or higher-order time derivatives in dispersive terms. Physically the most important criterion for the selection of an equation is its dispersion relation. The requirements above leave room for equations which are physically interesting.

The first set of Boussinesq equations we shall consider corresponds to the choice $\theta = 0$ (Ref. 26)

$$h_t + (hu)_x = 0, \tag{2.1a}$$

$$u_t + uu_x + h_x + \nu h_{3x} = 0, \tag{2.1b}$$

where subscripts will denote partial derivatives and after the second one we shall indicate the number of derivatives by a numerical prefix. The shallow water equations on a flat bottom are obtained when the constant ν is set equal to zero. A variant of Eq. (2.1a) where h_{3x} is replaced by h_{xtt} is discussed in Ref. 27. Next we shall consider the WBK equations which are given by

$$\tilde{u}_t + \tilde{u}\tilde{u}_x + \tilde{h}_x + \tilde{\nu}\tilde{u}_{xx} = 0, \tag{2.2a}$$

$$\tilde{h}_t + \tilde{h}\tilde{u}_x + \tilde{u}\tilde{h}_x + \tilde{\sigma}\tilde{u}_{3x} - \tilde{\nu}\tilde{h}_{xx} = 0, \tag{2.2b}$$

where $\tilde{\nu},\tilde{\sigma}$ are constants. By a change of variables[22] Kupershmidt has written these equations as

$$\tilde{u}_t + \tilde{u}\tilde{u}_x + \tilde{h}_x + \tilde{\nu}\tilde{u}_{xx} = 0, \tag{2.2a}$$

$$\tilde{h}_t + \tilde{h}\tilde{u}_x + \tilde{u}\tilde{h}_x + \tilde{\sigma}\tilde{u}_{3x} - \tilde{\nu}\tilde{h}_{xx} = 0, \tag{2.2b}$$

which is the form we shall use in this paper. Our knowledge of Eqs. (2.3) is much better than that of Eqs. (2.1) so that we shall not need to refer to the limit of zero dispersion and consequently an arbitrary constant coefficient of the dispersive terms has been scaled out.

We shall start with a reformulation of these equations in terms of potentials. For this purpose note that Eqs. (2.1) are the conditions for the one-forms,

$$\alpha = h\,dx - hu\,dt, \tag{2.4a}$$

$$\omega = u\,dx - (\tfrac{1}{2}u^2 + h + \nu h_{xx})dt, \tag{2.4b}$$

to be closed

$$d\alpha = 0, \quad d\omega = 0. \tag{2.5}$$

Therefore, using Poincaré's lemma we have locally

$$\alpha = d\Psi, \quad \omega = d\Phi, \tag{2.6}$$

where Ψ and Φ are scalar potentials. In terms of components, Eqs. (2.5) and (2.6) yield the relations

$$\begin{aligned}\Phi_x &= u, \quad \Phi_t = -(\tfrac{1}{2}u^2 + h + \nu h_{xx}),\\ \Psi_x &= h, \quad \Psi_t = -uh,\end{aligned} \tag{2.7}$$

between the phenomenological fields u,h and the potentials Φ,Ψ. The integrability conditions of Eqs. (2.7) yield the original equations of motion, and their compatibility requires that

$$\Psi_t + \Phi_x\Psi_x = 0, \tag{2.8a}$$

$$\Phi_t + \tfrac{1}{2}\Phi_x^2 + \Psi_x + \nu\Psi_{3x} = 0, \tag{2.8b}$$

which are nonlinear partial differential equations satisfied by the potentials. Equations (2.1) can be derived from an action principle

$$\delta I = 0, \quad I = \int \mathscr{L}\,dx\,dt,$$

where

$$\mathscr{L}_1 = \Phi_t\Psi_x + \Phi_x\Psi_t + \Phi_x^2\Psi_x + \Psi_x^2 - \nu\Psi_{xx}^2 \tag{2.9}$$

is the Lagrangian density.

The introduction of potentials for Kupershmidt's equations follows along similar lines. The closed one-forms are now given by

$$\alpha = u\,dx - (\tfrac{1}{2}u^2 + h - \tfrac{1}{2}u_x)dt, \tag{2.10a}$$

$$\omega = h\,dx - (uh + \tfrac{1}{2}h_x)dt, \tag{2.10b}$$

and the potentials satisfy

$$\Phi_t + \tfrac{1}{2}\Phi_x^2 + \Psi_x - \tfrac{1}{2}\Phi_{xx} = 0, \tag{2.11a}$$

$$\Psi_t + \Phi_x\Psi_x + \tfrac{1}{2}\Psi_{xx} = 0. \tag{2.11b}$$

The elimination of Ψ from these equations results in Kaup's equation[21]

$$\Phi_{tt} + 2\Phi_x\Phi_{xt} + (\Phi_t + \tfrac{3}{2}\Phi_x^2)\Phi_{xx} - \Phi_{4x} = 0 \tag{2.12a}$$

for which there is a companion[28]

$$\begin{aligned}&\Psi_x^2\Psi_{tt} - 2\Psi_x\Psi_t\Psi_{xt} + (\Psi_t^2 - 2\Psi_t\Psi_x + \Psi_x^3)\Psi_{xx}\\ &\quad + 3\Psi_{xx}^2 - 2\Psi_x\Psi_{xx}\Psi_{3x} + \Psi_x^2\Psi_{4x} = 0.\end{aligned} \tag{2.12b}$$

Kaup was led to Eq. (2.12a) by seeking an equation for which he could formulate and solve the inverse scattering problem where the AKNS potentials depend linearly on the eigenvalue. Kaup's results can be used to solve Eq. (2.12b) as well. Finally,

$$\mathscr{L}_2 = \Phi_t\Psi_x + \Psi_t\Phi_x + \Phi_x^2\Psi_x + \Psi_x^2 - \Psi_{xx}^2 \tag{2.13}$$

is the Lagrangian for Kupershmidt's equations.

III. CANONICAL FORMULATION OF BOUSSINESQ EQUATIONS

For passing to a Hamiltonian formulation of Eqs. (2.1) we shall start with an alternative form of the Lagrangian (2.9) which depends at most on the first derivatives of all the fields. This can be accomplished by introducing another potential Υ. We can readily verify that the Euler–Lagrange equations for

$$\begin{aligned}\mathscr{L}_3 = {}&\Phi_t\Psi_x + \Psi_t\Phi_x + \Phi_x^2\Psi_x\\ &- 2\kappa\Upsilon_x\Psi_x - 2\Upsilon\Psi_x + \epsilon\Upsilon^2 + (1+\epsilon)\Psi_x^2,\\ &\epsilon = \mathrm{sgn}(\nu), \quad \kappa = |\nu|^{1/2},\end{aligned} \tag{3.1}$$

yield Eqs. (2.1) together with

$$\Upsilon = \epsilon\Psi_x - \epsilon\kappa\Psi_{xx}, \tag{3.2}$$

which serves as the definition of Υ.

The Lagrangian (3.1) is degenerate. That is, the canonical momenta

$$\Pi_\Phi = \Psi_x, \quad \Pi_\Psi = \Phi_x, \quad \Pi_\Upsilon = 0, \tag{3.3}$$

cannot be inverted for the velocities and we need to use Dirac's theory of constraints in order to cast this system into canonical form. Therefore we introduce

$$C_1 = \Pi_\Phi - \Psi_x, \quad C_2 = \Pi_\Psi - \Phi_x, \quad C_3 = \Pi_\Upsilon, \tag{3.4}$$

as primary constraints. Using the canonical Poisson brackets $\{\,,\,\}$ between the potentials and their conjugate momenta we find that

$$\{C_1(x), C_2(x')\} = -2\delta_x(x-x'), \tag{3.5}$$

is the only nonvanishing one among the Poisson brackets of the constraints. The primary constraints are therefore second class. The total Hamiltonian

$$H = \int \mathscr{H}\, dx, \quad \mathscr{H} = \mathscr{H}_0 + \mathscr{H}', \tag{3.6a}$$

will be given by

$$\mathscr{H}_0 = \Pi_\Phi \Phi_t + \Pi_\Psi \Psi_t + \Pi_\Upsilon \Upsilon_t - \mathscr{L}_3, \tag{3.6b}$$

$$\mathscr{H}' = \lambda C_1 + \sigma C_2 + \beta C_3, \tag{3.6c}$$

where λ, σ, and β are Lagrange multipliers. These multipliers will be determined from the requirement that the Poisson bracket of the Hamiltonian with each one of the constraints should vanish. But we find that

$$\{C_3, H\} = 2(\epsilon\Upsilon - \Psi_x + \kappa\Psi_{xx})$$

cannot be set equal to zero because it is independent of the multipliers. Therefore we introduce a secondary constraint

$$\chi = 2(\epsilon\Upsilon - \Psi_x + \kappa\Psi_{xx}) \tag{3.7}$$

and modify Eq. (3.6c),

$$\mathscr{H}' = \lambda C_1 + \sigma C_2 + \beta C_3 + \mu\chi, \tag{3.6c'}$$

where μ is another multiplier. The Poisson brackets of the secondary constraint with the primary constraints are given by

$$\begin{aligned} &\{\chi, C_1\} = 0, \\ &\{\chi, C_2\} = -2\delta_x(x-x') + 2\kappa\delta_{xx}(x-x'), \\ &\{\chi, C_3\} = 2\epsilon\delta(x-x'), \end{aligned} \tag{3.8}$$

and now we must check to see if there are any tertiary constraints. It turns out that there are no further constraints in this problem because with the choice

$$\begin{aligned} &\sigma = -\Phi_x\Psi_x, \\ &\mu = \Upsilon - \epsilon\Psi_x + \epsilon\kappa\Psi_{xx}, \\ &\beta = -\epsilon(\Phi_x\Psi_x)_x + \epsilon\kappa(\Phi_x\Psi_x)_{xx}, \\ &\lambda = -\tfrac{1}{2}\Phi_x^{\,2} - \Psi_x - \nu\Psi_{3x}, \end{aligned} \tag{3.9}$$

the Poisson brackets of the Hamiltonian with C_1, C_2, C_3, and χ all vanish. From Eqs. (3.4), (3.6a), (3.6b), (3.6c'), (3.7), and (3.9) we find the total Hamiltonian density

$$\begin{aligned} \mathscr{H} = {} & -\Phi_x^{\,2}\Psi_x + 2\kappa\Upsilon_x\Psi_x + 2\Upsilon\Psi_x - \epsilon\Upsilon^2 - (1+\epsilon)\Psi_x^{\,2} \\ & - (\Pi_\Phi - \Psi_x)(\tfrac{1}{2}\Phi_x^{\,2} + \Psi_x + \nu\Psi_{3x}) \\ & - (\Pi_\Psi - \Phi_x)\Phi_x\Psi_x + \Pi_\Upsilon\epsilon[\Phi_x\Psi_x - \kappa(\Phi_x\Psi_x)_x]_x \\ & + 2\epsilon(\Upsilon - \epsilon\Psi_x + \epsilon\kappa\Psi_{xx})^2, \end{aligned} \tag{3.10a}$$

which can be simplified to the form

$$\begin{aligned} \mathscr{H} = {} & \tfrac{1}{2}\Phi_x^{\,2}\Psi_x + \epsilon\Psi_x^{\,2} + \nu\Psi_{xx}^{\,2} + \epsilon\Upsilon^2 - 2\Upsilon\Psi_x \\ & + 2\kappa\Upsilon\Psi_{xx} - \Pi_\Phi(\tfrac{1}{2}\Phi_x^{\,2} + \Psi_x + \nu\Psi_{3x}) - \Phi_x\Psi_x\Pi_\Psi \\ & - \epsilon[\Phi_x\Psi_x - \kappa(\Phi_x\Psi_x)_x]_x\Pi_\Upsilon \end{aligned} \tag{3.10b}$$

by discarding a divergence. The fact that the constraints are second class enables us to use them in order to further simplify the Hamiltonian. First eliminating all reference to Υ from Eqs. (3.2) and (3.3c) we find

$$\mathscr{H} = \tfrac{1}{2}\Phi_x^{\,2}\Psi_x - (\tfrac{1}{2}\Phi_x^{\,2} + \Psi_x + \nu\Psi_{3x})\Pi_\Phi - \Phi_x\Psi_x\Pi_\Psi. \tag{3.10c}$$

Finally using Eqs. (3.3a) and (3.3b) we can express the Hamiltonian in terms of the potentials and from Eqs. (2.7) we obtain

$$\mathscr{H} = \tfrac{1}{2}u^2h + \tfrac{1}{2}h^2 + \nu hh_{xx} + \tfrac{1}{2}\nu h_x^{\,2}, \tag{3.10d}$$

where we have again discarded a divergence. As we shall reconfirm later, this is the energy density for Eqs. (2.1).

We shall now turn to a discussion of the generalized Hamiltonian structure of Eqs. (2.1). The phase space consists of the set (u,h) of infinitely differentiable functions and $E = (E_u, E_h)$ will denote the variational derivative with respect to the variable indicated by its subscript. If A,B are two smooth functions of these variables, the Poisson bracket is defined by

$$[A,B] = \int E(A)JE(B)dx, \tag{3.11a}$$

where

$$J = -\begin{pmatrix} 0 & D \\ D & 0 \end{pmatrix}, \quad D = \frac{\partial}{\partial x}, \tag{3.11b}$$

is the Hamiltonian operator. With this definition of the Poisson bracket, Jacobi's identity is satisfied. The Hamiltonian function is given by an integral over the density (3.10d) so that for z standing for u or h, Hamilton's equations

$$z_t = [z,H] \tag{3.12}$$

reduce to Eqs. (2.1). An integral of motion P will satisfy

$$[P,H] = 0, \tag{3.13a}$$

which reduces to the following condition on P:

$$E_u[E_u(P)] - hE_h[E_h(P)] + \nu E_u\{[E_u(P)]_{xx}\} = 0, \tag{3.13b}$$

generalizing the result of Benney[29] for shallow water waves.

Equations (2.1) are already in the form of a pair of conservation laws but they admit two further ones corresponding to the conservation of momentum and energy

$$\begin{aligned} & (uh)_t + (u^2h + \tfrac{1}{2}h^2 + \nu hh_{xx} - \tfrac{1}{2}\nu h_x^{\,2})_x = 0, \\ & (\tfrac{1}{2}u^2h + \tfrac{1}{2}h^2 + \nu hh_{xx} + \tfrac{1}{2}\nu h_x^{\,2})_t \\ & \qquad + (\tfrac{1}{2}u^3h + uh^2 + \nu hh_{xx} - \nu hh_{xt})_x = 0, \end{aligned} \tag{3.14}$$

respectively. We can read off P from these equations and verify that it satisfies Eq. (3.13b) in each case. In particular, the conserved quantity for the latter of Eqs. (3.14) is the Hamiltonian (3.10d). Furthermore, setting $\nu = 0$ we obtain Benney's conserved quantities. Correspondence with the well-known results in the limit of vanishing dispersion shows that there are no further conservation laws for this system. In order to see this let us consider a possible fifth conservation law. Such a correspondence principle will require that this conservation law must be of the form

$$(\tfrac{1}{3}u^3h + uh^2 + f)_t + (\tfrac{1}{4}u^4h + \tfrac{3}{2}u^2h^2 + \tfrac{1}{3}h^3 + g)_x = 0, \tag{3.15a}$$

where f,g may depend on $u, h, h_x, h_t, h_{xx}, \ldots$ and vanish in the limit $\nu \to 0$. Then we find

$$f_t + g_x = \nu(u^2h + h^2)h_{3x} \tag{3.15b}$$

and using Eqs. (2.1) repeatedly this can be cast into the form

$$\begin{aligned} f_t + g_x &= \nu(h_x h_t)_t + \nu(u^2hh_{xx} + h_x(uh)_t + h^2h_{xx} - \tfrac{1}{2}h_t^2)_x \\ &\quad - \nu hh_x h_{xx} + \nu^2 hh_{xx}h_{3x}; \end{aligned} \tag{3.15c}$$

the first two groups of terms on the right-hand side of Eq. (3.15c) are of the desired form but it is not possible to write either one of the last two terms as a total divergence. It will be sufficient to prove this only for one of them, say hh_xh_{xx}. In order to express this term as a divergence we consider all possible divergences which can result in such an expression. Thus we write

$$hh_xh_{xx} = \alpha(hh_x^2)_x + \beta(h^2h_{xx})_x + \gamma(h^2h_x)_{xx}, \tag{3.16}$$

where α, β, and γ are constants which must be chosen so as to make this an identity. Note that it is unnecessary to include $(h^3)_{3x}$ in Eq. (3.16) since it reduces to the last term above. From the coefficients of all linearly independent functions in Eq. (3.16) we obtain a system of linear equations for α, β, and γ. This system of equations has no solution, which makes it impossible to express hh_xh_{xx} as a total divergence. Therefore there is no fifth conservation law for Eqs. (2.1) which reduces to Benney's result in the limit of zero dispersion.

IV. THE CANONICAL FORMULATION OF KUPERSHMIDT'S EQUATIONS

Kupershmidt's equations, given by Eq. (2.3), may be derived from a variational principle with the Lagrangian density,

$$\begin{aligned} \mathscr{L}_4 &= \Phi_t\Psi_x + \Psi_t\Phi_x + \Phi_x{}^2\Psi_x + \Psi_x{}^2 + \tfrac{1}{2}\Upsilon\Xi_x \\ &\quad - \tfrac{1}{2}\Xi\Upsilon_x + \Upsilon_x\Psi_x + \Xi_x\Phi_x. \end{aligned} \tag{4.1}$$

Here Υ and Ξ have been introduced to avoid terms with higher-order derivatives. Using the following definitions for Υ and Ξ which are obtained as the Euler–Lagrange equations corresponding to Eq. (4.1),

$$\Upsilon = -\Phi_x, \quad \Xi = \Psi_x, \tag{4.2}$$

the other two equations of motion reduce to Eq. (2.1). Since we cannot eliminate velocities in terms of momenta, the constraint equations are

$$\begin{aligned} C_1 &= \Pi_\Phi - \Psi_x, \quad C_3 = \Pi_\Upsilon, \\ C_2 &= \Pi_\Psi - \Phi_x, \quad C_4 = \Pi_\Xi. \end{aligned} \tag{4.3}$$

The only nonzero Poisson bracket between the constraints turns out to be

$$\{C_1(x), C_2(x')\} = -2\delta_x(x - x'). \tag{4.4}$$

The total Hamiltonian density consists of

$$\mathscr{H}_0 = \Pi_\Phi\Phi_t + \Pi_\Psi\Psi_t + \Pi_\Upsilon\Upsilon_t + \Pi_\Xi\Xi_t - \mathscr{L}_4, \tag{4.5a}$$

$$\mathscr{H}' = \alpha C_1 + \beta C_2 + \sigma C_3 + \mu C_4, \tag{4.5b}$$

where α, β, σ, and μ are Lagrange multipliers. When we calculate the Poisson bracket of each one of the constraints with the total Hamiltonian, we find that

$$\begin{aligned} \{C_3, H\} &= \Xi_x - \Psi_{xx} \equiv \chi_a, \\ \{C_4, H\} &= -\Upsilon_x - \Phi_{xx} \equiv \chi_b, \end{aligned} \tag{4.6}$$

cannot be set equal to zero by any choice of multipliers. Therefore, χ_a and χ_b are secondary constraints, whose nonvanishing Poisson brackets with the primary constraints are

$$\begin{aligned} \{\chi_a(x), C_2(x')\} &= -\delta_{xx}(x - x'), \\ \{\chi_a(x), C_4(x')\} &= \delta_x(x - x'), \\ \{\chi_b(x), C_1(x')\} &= -\delta_{xx}(x - x'), \\ \{\chi_b(x), C_3(x')\} &= -\delta_x(x - x'). \end{aligned} \tag{4.7}$$

We incorporate these secondary constraints into the Hamiltonian through

$$\mathscr{H}' = \alpha C_1 + \beta C_2 + \sigma C_3 + \mu C_4 + \nu\chi_a + \tau\chi_b, \tag{4.5b$'$}$$

where ν and τ are again multipliers. Finally we see that with the following choice for the six multipliers:

$$\begin{aligned} \alpha &= -\tfrac{1}{2}(\Phi_x{}^2 + 2\Psi_x + \Upsilon_x), \\ \beta &= -\tfrac{1}{2}(\Xi_x + 2\Psi_x\Phi_x), \\ \sigma &= -\alpha_x, \\ \mu &= \beta_x, \\ \nu &= \Phi_x + \Upsilon, \\ \tau &= \Xi - \Psi_x, \end{aligned} \tag{4.8}$$

the Poisson bracket of H_T with each one of the constraints vanishes. Using Eqs. (4.3), (4.5a), (4.5b′), (4.6), and (4.8) we find the total Hamiltonian density for Kupershmidt's equation is

$$\begin{aligned} H_T &= \tfrac{1}{2}\Phi_x{}^2\Psi_x - \Pi_\Phi(\tfrac{1}{2}\Phi_x{}^2 + \Psi_x - \tfrac{1}{2}\Phi_{xx}) \\ &\quad - \Pi_\Psi(\Phi_x\Psi_x + \tfrac{1}{2}\Psi_{xx}), \end{aligned} \tag{4.9a}$$

where we have eliminated the auxiliary potentials Ξ and Υ. Once again, since the constraints are second class, this expression can be reduced to one involving only the potentials which through Eqs. (4.3) becomes

$$\mathscr{H}_T = \tfrac{1}{2}(u^2h + h^2 + uh_x). \tag{4.9b}$$

In the framework of the generalized Hamiltonian formalism where the Poisson bracket is defined according to Eqs. (3.11) we find that the Hamiltonian is given by Eqs. (4.9b). We have also the conserved momentum density

$$\mathscr{P}_0 = uh, \tag{4.10a}$$

and Eqs. (2.3) admit an infinite family of conserved quantities $P_k = \int \mathscr{P}_k \, dx$ with

$$\mathscr{P}_1 = \mathscr{H}_T \tag{4.10b}$$

satisfying

$$E_u[E_u(P)] - hE_h[E_h(P)] - \tfrac{1}{2}[E_h\{[E_u(P)]_x\} + E_u\{[E_h(P)]_x\}] = 0, \quad (4.11)$$

which reduce to Benney's results in the limit of zero dispersion. The existence of these conserved quantities is best understood in terms of the tri-Hamiltonian structure of Eqs. (2.3). Kupershmidt[22] found that in addition to the Hamiltonian operator (3.11b) (which will henceforth be denoted J_1) there exist two further operators such that

$$J_l E(P_k) = J_1 E(P_{k+l-1}), \quad l = 2,3; \; k = 0,1,2,..., \quad (4.12)$$

and repeated applications of the recursion operator D_{12} where

$$D_{ij} = J_j J_i^{-1} \quad (4.13)$$

and[22]

$$J_2 = \begin{pmatrix} 2D & Du - D^2 \\ uD + D^2 & hD + Dh \end{pmatrix} \quad (4.14)$$

yields infinitely many conserved quantities starting with P_0. Here D_{13} generates the same conserved quantities in steps of two. The equations of gas dynamics in 1 + 1 dimensions exhibit a similar symplectic structure.[31]

Kupershmidt's equation is of Painlevé type II. This can be established from an analysis of the invariance properties of Eqs. (2.3) which enables us to reduce them to ordinary differential equations. We shall use the formalism of Lakshmanan and Kaliappan[32] and start by writing Eqs. (2.3) in the form

$$\begin{aligned} H_1(u,h,h_x,...) &\equiv u_t + uu_x + h_x - \tfrac{1}{2}u_{xx} = 0, \\ H_2(u,h,h_x,...) &\equiv h_t + uh_x + hu_x + \tfrac{1}{2}h_{xx} = 0, \end{aligned} \quad (4.15)$$

and furthermore let

$$u = \theta(x,t), \quad h = \lambda(x,t) \quad (4.16)$$

be a solution of this system. If these equations are invariant under the infinitesimal transformations

$$\begin{aligned} x &= x + \epsilon\zeta(x,t,u,h), \\ t &= t + \epsilon\tau(x,t,u,h), \\ u &= u + \epsilon\eta(x,t,u,h), \\ h &= h + \epsilon\sigma(x,t,u,h), \end{aligned} \quad (4.17)$$

where ϵ is an infinitesimal parameter, then

$$XH_1\Big|_{\substack{u=\theta(x,t)\\ h=\lambda(x,t)}} = 0, \quad XH_2\Big|_{\substack{u=\theta(x,t)\\ h=\lambda(x,t)}} = 0,$$

where

$$X = \zeta\frac{\partial}{\partial x} + \tau\frac{\partial}{\partial t} + \eta\frac{\partial}{\partial u} + \sigma\frac{\partial}{\partial h} + \{\eta_t\}\frac{\partial}{\partial u_t} + \{\sigma_t\}\frac{\partial}{\partial h_t}\cdots \quad (4.18)$$

with $\{\eta_t\},\{\sigma_t\},...$ denoting the first-order changes in the derivatives of $u_t,h_t,...$. In terms of θ,λ,η,τ, the explicit expressions for the higher extensions $\{\eta_t\},\{\sigma_t\}$ can be obtained by using Eqs. (4.17) and a typical extension is

$$\{\sigma_t\} = \sigma_t + \lambda_t(\sigma_h - \tau_t) + \theta_t(\sigma_u - \tau_u\lambda_t) - \zeta_t\lambda_x - \zeta_h\lambda_t\lambda_x - \zeta_u\theta_t\lambda_x - \tau_h\lambda_t^2.$$

From Eqs. (4.18) we get

$$\begin{aligned} &\{\eta_t\} + \eta\theta_x + \theta\{\eta_x\} + \{\sigma_x\} - \tfrac{1}{2}\{\eta_{xx}\} = 0, \\ &\{\sigma_t\} + \theta\{\sigma_x\} + \sigma\theta_x + \lambda\{\eta_x\} + \eta\lambda_x + \tfrac{1}{2}\{\sigma_{xx}\} = 0, \end{aligned} \quad (4.19)$$

into which we must substitute extensions. Equations (4.19) are algebraic equations for the variables θ,λ and their derivatives. Since these variables are linearly independent we require that their coefficients which depend on η,ζ,τ,σ must vanish separately. This leads to

$$\zeta = at + b, \quad \eta = a, \quad \tau = c, \quad \sigma = 0, \quad (4.20)$$

where a, b, and c are arbitrary constants. Then from

$$\frac{dx}{\zeta} = \frac{dt}{\tau} = \frac{du}{\eta} = \frac{dh}{\sigma} \quad (4.21)$$

we find

$$c\,dx = (at + b)dt, \quad (4.22a)$$

$$a\,dt = c\,du, \quad (4.22b)$$

$$a\,dx = (at + b)du, \quad (4.22c)$$

$$\sigma = \text{const}, \quad (4.22d)$$

which can be integrated readily.

Integrating Eq. (4.22a) we obtain the invariant variable ξ,

$$((at + b)/c)^2 - 2ax/c = \text{const} \equiv \xi, \quad (4.23)$$

and from Eqs. (4.22b) and (4.22c) we find

$$\xi - \text{const} = (u - (at + b)/c)^2,$$

which suggests that we define

$$g(\xi) \equiv u - (at + b)/c \quad (4.24)$$

as an invariant function. Finally we shall take the right-hand side of Eq. (4.21d) as $f(\xi)$ and change from the variables (u,h,x,t) to $(f(\xi), g(\xi), \xi)$. In this way we can write Eqs. (4.15) as a pair of coupled ordinary differential equations for f and g,

$$\begin{aligned} gg' + f' + (a/c)g'' &= \tfrac{1}{2}, \\ gf' + fg' - (a/c)f'' &= 0, \end{aligned} \quad (4.25)$$

where the prime denotes differentiation with respect to ξ. These equations can be decoupled to yield a second-order equation and letting

$$\begin{aligned} g &= (a^2/2c^2)^{1/6}w, \\ z &= -(2c^2/a^2)^{1/3}(\tfrac{1}{2}\xi + C_2), \end{aligned} \quad (4.26)$$

we find that it is Painlevé II[33]

$$w'' = w^3 + 2zw + \mu, \quad (4.27)$$

where

$$\mu = (2c/a)^{1/6}.$$

ACKNOWLEDGMENTS

We thank M. Gürses and P. J. Olver for interesting conversations.

This work was in part supported by the Turkish Scientific Research Council TÜBITAK and NATO Research Grant RG 86/0055.

APPENDIX: KUPERSHMIDT'S TRANSFORMATION

The transformation Kupershmidt has found to pass from Eqs. (2.2) to (2.3) is best understood by an examination of the Hamiltonian structure of these equations. Equations (2.2) are Hamilton's equations (3.12) for the Hamiltonian function

$$\widetilde{H} = \tfrac{1}{2}\tilde{u}^2\tilde{h} + \tfrac{1}{2}\tilde{h}^2 - \tfrac{1}{2}\tilde{\sigma}\tilde{u}_x^{\,2} + \tilde{\nu}\tilde{u}_x\tilde{h} \tag{A1}$$

with the Poisson bracket defined according to Eqs. (3.11). Kupershmidt's transformation

$$\tilde{u} = u, \quad \tilde{h} = h + \gamma u_x \tag{A2}$$

applied to Eq. (A1) yields

$$H = \tfrac{1}{2}u^2 h + \tfrac{1}{2}h^2 + \tfrac{1}{2}(\gamma^2 - \tilde{\sigma})u_x^{\,2} + (\tilde{\nu} + \gamma)u_x h \tag{A3}$$

up to a divergence. The choice

$$\gamma^2 = \tilde{\sigma}, \quad \mu = \tilde{\nu} + \gamma = \tilde{\nu} \pm \sqrt{\tilde{\sigma}} = \tfrac{1}{2} \tag{A4}$$

gives the Hamiltonian function of Kupershmidt's equations. We note that there is a restriction on $\tilde{\sigma}$, namely $\tilde{\sigma}$ must be positive for the Hamiltonian (A3) to be real. This corresponds to a negative definite contribution in the original Hamiltonian (A1) but the total Hamiltonian is positive.

[1] J. C. Luke, J. Fluid Mech. **27**, 395 (1967).
[2] V. E. Zakharov, J. Appl. Mech. Tech. Phys. (USSR) **9**, 190 (1968).
[3] L. J. F. Broer, Physica (Utrecht) **76**, 364 (1974).
[4] J. W. Miles, J. Fluid Mech. **83**, 153 (1977).
[5] Y. Nutku, J. Phys. A **16**, 4195 (1983).
[6] P. A. M. Dirac, *Lectures on Quantum Mechanics*, Belfer Graduate School Monograph Series No. 3 (Yeshiva U.P., New York, 1964).
[7] A. Hanson, T. Regge, and C. Teitelboim, Accad. Naz. dei Lincei, Roma, 1976.
[8] Y. Nutku, J. Math. Phys. **25**, 2007 (1984).
[9] C. S. Gardner, J. Math. Phys. **12**, 1548 (1971).
[10] A. J. Macfarlane, CERN preprint TH 3289, 1982.
[11] M. J. Bergvelt and B. A. DeKerf, Lett. Math. Phys. **10**, 18 (1985).
[12] F. Lund, Physica (Utrecht) D **18**, 420 (1986).
[13] A. Lenard, reported in P. D. Lax, SIAM Rev. **18**, 351 (1976).
[14] P. J. Olver, J. Math. Phys. **18**, 1212 (1977).
[15] F. Magri, J. Math. Phys. **19**, 1156 (1978).
[16] I. M. Gel'fand and I. Ya. Dorfman, Funk. Anal. **13**, 13 (1979).
[17] A. S. Fokas and B. Fuchssteiner, Lett. Nuovo Cimento **28**, 299 (1980); Physica (Utrecht) D **4**, 47 (1981).
[18] P. J. Olver, J. Math. Phys. **27**, 2495 (1986).
[19] G. B. Whitham, Proc. R. Soc. London Ser. A **299**, 6 (1967).
[20] L. J. F. Broer, Appl. Sci. Res. **31**, 377 (1975).
[21] D. J. Kaup, Prog. Theor. Phys. **54**, 396 (1975).
[22] B. A. Kupershmidt, Commun. Math. Phys. **99**, 51 (1985).
[23] G. B. Whitham, *Linear and Non-linear Waves* (Wiley, New York, 1974).
[24] J. L. Bona and R. Smith, Math. Proc. Cambridge Philos. Soc. **79**, 167 (1976).
[25] P. J. Olver, Contemp. Math. **28**, 231 (1984).
[26] G. B. Whitham, Proc. R. Soc. London Ser. A **283**, 238 (1965).
[27] F. Neyzi, thesis, Bosphorus University, 1984.
[28] Y. Nutku, J. Math. Phys. **26**, 1237 (1985).
[29] D. J. Benney, Stud. Appl. Math. **52**, 45 (1973).
[30] P. J. Olver, *Applications of Lie Groups to Differential Equations* (Springer, Berlin, 1986).
[31] Y. Nutku, to be published in J. Math. Phys.
[32] M. Lakshmanan and P. Kaliappan, J. Math. Phys. **24**, 795 (1983).
[33] M. J. Ablowitz and H. Segur, "Solitons and the inverse scattering transform" in *SIAM Studies in Applied Mathematics* (SIAM, Philadelphia, 1981).

The Dirac wave equation in the presence of an external field

Jinane Abounadi and Teymour Darkhosh
Department of Physics, Bryn Mawr College, Bryn Mawr, Pennsylvania 19010

(Received 31 October 1986; accepted for publication 4 March 1987)

The method of characteristics is applied to the Dirac wave equation in the presence of an external field. The retarded Green's function for the minimal coupling to an external electromagnetic field is calculated explicitly, and a general coupling is discussed.

I. INTRODUCTION

In studying the wave properties of the Dirac equation, it is customary to convert it to the Klein–Gordon type of equation by squaring it.[1] However, we feel that this method does not reveal the hidden structure of the Dirac equation. Therefore, in this paper we apply the method of characteristics directly to the Dirac equation and calculate the retarded Green's function. In so doing, we establish an existence theorem since the necessary and sufficient condition for the existence of the solution is the existence of the fundamental solution or the Green's function.[2]

We begin by considering the minimal coupling to an electromagnetic field where we expand the solution in the neighborhood of the characteristic surface and show that the terms can be calculated to any order. Finally, we apply the same method to analyze a general interaction. Our notation is that of Bjorken and Drell.[3]

II. THE MINIMAL COUPLING

Consider the Dirac particle minimally coupled to an external electromagnetic field,

$$(i\not{D} - m)\psi = 0\,, \tag{2.1}$$

where D^μ is defined as $\partial^\mu - ieA^\mu$, A^μ is the electromagnetic potential, ψ is a Dirac spinor, and $\not{D} \equiv \gamma^\mu D_\mu$. The spinor indices are suppressed.

As we mentioned in the Introduction, to study Eq. (2.1) it is sufficient to analyze the solution of

$$(i\not{D} - m)S(x,y) = \delta^4(x-y)\,, \tag{2.2}$$

which is the equation for the Green's function. In particular, we are interested in the retarded solution, i.e.,

$$S(x-y) = 0 \quad \text{for} \quad (x^0 - y^0) < 0\,. \tag{2.3}$$

For convenience, we choose our origin at y.

If Eq. (2.2) is a hyperbolic partial differential equation, it has to accept a solution in the following form[4,5]:

$$S(x) = \sum_{n=N}^{0} \sum_{j=1}^{m} \delta^n(u_j) E^n(x) + \sum_{j=1}^{m} \theta(u_j) \sum_{n=0}^{\infty} \frac{u^n}{n!} G^n(x)\,, \tag{2.4}$$

$u_j = 0$ are the characteristic surfaces, $\delta^n(u_j)$ is the nth derivative of the Dirac delta function, and $\theta(u_j)$ is the step function.

Upon substitution of (2.4) into (2.2) and separation of coefficients of different singular terms, we find

$$\begin{aligned}
&in\cdot\gamma E^N = 0\,,\\
&in\cdot\gamma E^{N-1} + (i\not{D} - m)E^N = 0\,,\\
&\vdots \qquad\qquad \vdots\\
&in\cdot\gamma G^0 + (i\not{D} - m)E^0 = \delta^3(r)\,,\\
&in\cdot\gamma G^1 + (i\not{D} - m)G^0 = 0\,,\\
&\vdots \qquad\qquad \vdots
\end{aligned} \tag{2.5}$$

Here $\partial^\mu u = n^\mu$ is defined as the normal vector to the characteristic surface, $u = 0$.

Consider the first equation in (2.5), i.e.,

$$in\cdot\gamma E^N = 0\,, \tag{2.6}$$

where E^N is a 4×4 matrix and, for it to exist, $n\cdot\gamma$ has to be a singular matrix for which the necessary condition is

$$\det|in\cdot\gamma| = -(n^2)^2 = 0\,. \tag{2.7}$$

This implies that there is a twofold degenerate characteristic surface which is the light cone. Unlike higher spin ($s \geqslant 1$) equations, there is no ill effect. The reason, of course, is that there are no constraints, and the characteristic matrix does not depend on the external field.[6,7]

The solution to Eq. (2.6) has the following form:

$$E^N = \sum_{s=1}^{2} r^s \sigma^{s,N}\,, \tag{2.8}$$

where the r^s are two linearly independent right-hand solutions of the characteristic matrix, $n\cdot\gamma$, and are given by

$$r^1 = n\cdot\gamma\hat{e}^1 \quad \text{and} \quad r^2 = n\cdot\gamma\hat{e}^2\,. \tag{2.9}$$

Here $\hat{e}^1$ and $\hat{e}^2$ are defined as

$$\hat{e}^1 = \begin{pmatrix}1\\0\\0\\0\end{pmatrix} \quad \text{and} \quad \hat{e}^2 = \begin{pmatrix}0\\1\\0\\0\end{pmatrix}, \tag{2.10}$$

and the $\sigma^{s,N}$ are unknown spinors to be determined later. There are two linearly independent left-hand solution spinors given by

$$l^1 = \hat{e}^{1T} n\cdot\gamma \quad \text{and} \quad l^2 = \hat{e}^{2T} n\cdot\gamma\,. \tag{2.11}$$

Here T indicates the transpose.

If we multiply the second equation in (2.5) by a left vector, $l^{s'}$, and use the anticommutation relation

$$\{\not{D}, n\cdot\gamma\} = 2n\cdot\partial + (\partial\cdot n) - 2ieA\cdot n\,, \tag{2.12}$$

we find

$$n_0(2n\cdot\partial + \partial\cdot n - 2ieA\cdot n)\sigma^{s',N} = 0\,, \tag{2.13}$$

 0022-2488/87/071505-03$02.50

which within a constant, determines $\sigma^{s',N}$. Since we are interested in the future cone, $u = t - r$, Eq. (2.13) becomes an ordinary differential equation along the bicharacteristics. Therefore, $\sigma^{s',N}$ has the following form:

$$\sigma^{s,N} = (1/r)e^{ie\phi}C^{s,N}, \tag{2.14}$$

where r is the radius vector measured from the origin, ϕ is given by

$$\phi = \int_0^r A\cdot n\, dr', \tag{2.15}$$

and the constant $C^{s,N}$ is determined from the initial condition.

Now, substituting Eq. (2.14) into the second equation in (2.5), and following the same procedure, we find

$$E^{N-1} = \sum_{s=1}^{2} r^s\sigma^{s,N-1} - i(i\not{D} + m)\sum_{s=1}^{2} \hat{e}^s\sigma^{s,N}. \tag{2.16}$$

Since Eqs. (2.5) are singular, it is not always possible to determine E^{N-1} as is the case with the Rarita–Schwinger equation.[8]

The equation determining E^{N-1} is given by

$$\begin{aligned} &in\cdot\gamma E^{N-2} - n\cdot\gamma(i\not{D} + m)\sum_{s=1}^{2} \hat{e}^s\sigma^{s,N-1} \\ &\quad - i(i\not{D} - m)(i\not{D} + m)\sum_{s=1}^{2} \hat{e}^s\sigma^{s,N} \\ &\quad + i[2n\cdot\partial + (\partial\cdot n) - 2ieA\cdot n]\sum \hat{e}^s\sigma^{s,N-1} = 0 \text{ or } \delta^3(r)\,. \end{aligned} \tag{2.17}$$

Contracting Eq. (2.17) with $l^{s'}$ yields

$$\begin{aligned} &in_0(2n\cdot\partial + \partial\cdot n - 2ieA\cdot n)\sigma^{s',N-1} \\ &\quad + il^{s'}(D^2 + m^2 - (e/2)\sigma\cdot F)\sum \hat{e}^s\sigma^{s,N} \\ &\quad = 0 \quad \text{or} \quad l^{s'}\delta^3(r)\,. \end{aligned} \tag{2.18}$$

By $\sigma\cdot F$ we mean $\sigma^{\mu\nu}F_{\mu\nu}$, $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$, and $\sigma^{\mu\nu} = (i/2)[\gamma^\mu,\gamma^\nu]$. To remove the singularity from both sides of Eq. (2.18), we have to choose $C^{s,N}$ to be equal to $-(i/4\pi n_0)l^s$. This completely determines E^N, i.e.,

$$E^N = -\frac{i}{4\pi r}\sum_{s=1}^{2} r^s l^s e^{ie\phi} = -\frac{i}{4\pi r}n\cdot\gamma e^{ie\phi}. \tag{2.19}$$

It also determines N, which is equal to 1, i.e., the expansion begins with the first derivative of the delta function.

Having obtained E', we go back to the equation determining E^0, namely,

$$in\cdot\gamma E^0 + (i\not{D} - m)(e^{ie\phi}/4\pi r)(-in\cdot\gamma) = 0\,. \tag{2.20}$$

Solving for E^0, we get

$$E^0 = \sum_{s=1}^{2} r^s\sigma^{s,0} - (i\not{D} + m)\frac{e^{ie\phi}}{4\pi r}. \tag{2.21}$$

To determine $\sigma^{s,0}$, we consider the next equation in (2.5), i.e.,

$$\begin{aligned} &in\cdot\gamma G^0 - n\cdot\gamma(i\not{D} + m)\sum \hat{e}^s\sigma^{s,0} - (i\not{D} - m)(i\not{D} + m)\frac{e^{ie\phi}}{4\pi r} \\ &\quad + i[2n\cdot\partial + (\partial\cdot n) - 2ieA\cdot n]\sum \hat{e}^s\sigma^{s,0} = 0\,. \end{aligned} \tag{2.22}$$

Multiplying it by a left vector $l^{s'}$, we obtain

$$\begin{aligned} \sigma^{s,0} = &\frac{i}{8\pi n_0} l^s e^{ie\phi}\frac{1}{r} \\ &\times\int_0^r r' e^{-ie\phi}\left(D^2 + m^2 - \frac{e}{2}\sigma\cdot F\right)\frac{e^{ie\phi}}{r'}dr'\,. \end{aligned} \tag{2.23}$$

Therefore, E^0 is given by

$$E^0 = (e^{ie\phi}/r)\, n\cdot\gamma f^0(r) - (i\not{D} + m)(e^{ie\phi}/4\pi r)\,, \tag{2.24}$$

where

$$f^0(r) = \frac{i}{8\pi}\int_0^r r' e^{-ie\phi}\left(D^2 + m^2 - \frac{e}{2}\sigma\cdot F\right)\frac{e^{ie\phi}}{r'}dr'\,.$$

Now it is clear that all the terms in expansion (2.4) can be determined. For example, we can write the nth term as follows:

$$G^n = (e^{ie\phi}/r)n\cdot\gamma g^n(r) - i(i\not{D} + m)(e^{ie\phi}/r)g^{n-1}(r) \tag{2.25}$$

and

$$\begin{aligned} g^n(r) = &-\frac{1}{2}\int_0^r r' e^{-ie\phi}\left(D^2 + m^2 - \frac{e}{2}\sigma\cdot F\right) \\ &\times\frac{e^{ie\phi}}{r'}g^{n-1}(r')dr'\,, \end{aligned} \tag{2.26}$$

with $g^n(0) = 0$.

Thus we have established the existence of the solution explicitly and shown that the Dirac equation in the presence of an external electromagnetic field is hyperbolic without relying on the wave properties of the Klein–Gordon equation.

III. A GENERAL INTERACTION

In this section, we shall use the same method to discuss a general interaction. The equation we are concerned with is

$$(i\not{\partial} + B(x))S(x) = \delta^4(x)\,, \tag{3.1}$$

where $B(x)$ is a general 4×4 matrix that can be written as follows:

$$\begin{aligned} B(x) = &a(x) + \gamma^5 b(x) + \gamma^\mu c_\mu(x) \\ &+ \gamma^5\gamma^\mu d_\mu(x) + \tfrac{1}{4}\sigma^{\mu\nu}f_{\mu\nu}(x)\,. \end{aligned}$$

We assume that the coefficients are smooth functions of x, and that $f^{\mu\nu}$ is an antisymmetric tensor.

For Eq. (3.1) to be hyperbolic, it has to admit a solution in the form of Eq. (2.4). The leading singular term, as in the previous case, is the coefficient of $\delta'(u)$, and we write it as

$$E^1 = \frac{e^{i\phi}}{r}\sum_{s=1}^{2} r^s\sigma^{s,1}\,, \tag{3.2}$$

where ϕ is $\int_0^r c\cdot n\, dr'$, the r^s are right-hand solutions of the characteristic matrix, $n\cdot\gamma$, and r is the position vector.

The next term, E^0, is determined from

$$in\cdot\gamma E^0 + (i\not{\partial} + B)\frac{e^{i\phi}}{r}\sum_{s=1}^{2} r^s\sigma^{s,1} = 0\,, \tag{3.3}$$

which can be written as

$$in\cdot\gamma E^0 - n\cdot\gamma(i\partial\!\!\!/ - \widetilde{B})\frac{e^{i\phi}}{r}\sum_{s=1}^{2}\hat{e}^s\sigma^{s,1}$$

$$+ i[2n\cdot\partial + (\partial\cdot n) - 2in\cdot c]\frac{e^{i\phi}}{r}\sum_{s=1}^{2}\hat{e}^s\sigma^{s,1}$$

$$+ 2(\gamma^5 n\cdot d + \gamma\cdot f\cdot n)\frac{e^{i\phi}}{r}\sum_{s=1}^{2}\hat{e}^s\sigma^{s,1} = 0\,. \tag{3.4}$$

Here $\widetilde{B}(x)$ is defined as $a(x) + \gamma^5 b(x) - \gamma^\mu c_\mu(x) - \gamma^5\gamma^\mu d_\mu(x) + \frac{1}{4}\sigma^{\mu\nu}f_{\mu\nu}$.

Equations that determine $\sigma^{s,1}$ are ordinary differential equations along bicharacteristic lines, and are given by

$$l^{s'}L\sum_{s=1}^{2}\hat{e}^s\sigma^{s,1} = 0\,, \tag{3.5}$$

where L is defined as

$$2i\frac{d}{dr} + 2(\gamma^5 n\cdot d + \gamma\cdot f\cdot n)\,.$$

Using Eq. (3.5), we can write (3.4) as

$$in\cdot\gamma E^0 - n\cdot\gamma\left[(i\partial\!\!\!/ + M)\sum_{s=1}^{2}\hat{e}^s\sigma^{s,1} + \sum_{s=1}^{2}\alpha^s\sigma^{s,1}\right] = 0\,, \tag{3.6}$$

with

$$M = -\widetilde{B}(x) + (2/n_0)\,[n\cdot d - (f\cdot n)_3\gamma^5]\,,$$

$$\alpha^1 = -(2i/n_0)(f\cdot n)_+\gamma^5\hat{e}^2\,,$$

and

$$\alpha^2 = -(2i/n_0)(f\cdot n)_-\gamma^5\hat{e}^1\,,$$

where the $(f\cdot n)_\pm$ are defined as $(f\cdot n)_1 \pm i(f\cdot n)_2$.

Factoring the singular matrix, $n\cdot\gamma$, we can solve for E^0,

$$E^0 = \frac{e^{i\phi}}{r}\sum r^s\sigma^{s,0} - i(i\partial\!\!\!/ + M)\sum_{s=1}^{2}\hat{e}^s\sigma^{s,1} + \sum_{s=1}^{2}\alpha^s\sigma^{s,1}\,. \tag{3.7}$$

Continuing to the next term, we find that G^0 satisfies

$$in\cdot\gamma G^0 + (i\partial\!\!\!/ + B)\,(e^{i\phi}/r)\sum r^s\sigma^{s,0}$$

$$- i(i\partial\!\!\!/ + B)(i\partial\!\!\!/ + M)\sum\hat{e}^s\sigma^{s,1}$$

$$+ (i\partial\!\!\!/ + B)\sum\alpha^s\sigma^{s,1} = \delta^3(r)\,. \tag{3.8}$$

This equation forces us to assign the following value to $\sigma^{s,1}$ at $r=0$:

$$\sigma^{s,1} = -(i/4\pi n_0)l^s\,.$$

With the above initial condition and the differential equation, (3.5), the $\sigma^{s,1}$ are uniquely determined.

Multiplying Eq. (3.8) by $l^{s'}$ yields

$$l^{s'}L\sum_{s=1}^{2}\hat{e}^s\sigma^{s,0} = -l^{s'}(i\partial\!\!\!/ + B)(i\partial\!\!\!/ + M)\sum\hat{e}^s\sigma^{s,1}$$

$$+ (i\partial\!\!\!/ + B)\sum\alpha^s\sigma^{s,1}\,, \tag{3.9}$$

which determines $\sigma^{s,0}$ uniquely when the condition, $\sigma^{s,0}(0) = 0$ is imposed. We choose this initial condition to avoid any extra delta singularities.

Therefore, in general, we can write

$$G^n = \frac{e^{i\phi}}{r}\sum_{s=1}^{2}(r^s\sigma^{s,n} + O^{n,s}\sigma^{s,n-1}) \tag{3.10}$$

and relate $\sigma^{s,n}$ to $\sigma^{s,n-1}$ by

$$l^{s'}L\left[\sum_{s=1}^{2}(\hat{e}^s\sigma^{s,n}) + \sum_{s=1}^{2}(\Lambda^{s's}\sigma^{s,n-1})\right] = 0\,, \tag{3.11}$$

where $O^{n,s}$ is a known operator and $\Lambda^{s's}$ is a known matrix. The above analysis shows that even in the case of a general interaction, the Dirac equation preserves its hyperbolicity, and that the Green's function can be calculated to any order. The analysis strongly supports the existence theorem for the Dirac equation in the presence of an external field.

[1] A. S. Wightman, "Relativistic wave equations as singular hyperbolic systems," in *Proceedings of the Symposium on Pure Mathematics* (Am. Math. Soc., Providence, RI, 1973), Vol. XXIII.

[2] A. S. Wightman, in *Aspects of Quantum Theory*, edited by A. Salam and E. P. Wigner (Cambridge U. P., Cambridge, 1972), p. 95.

[3] J. Bjorken and S. Drell, *Relativistic Quantum Fields* (McGraw-Hill, New York, 1965).

[4] T. Darkhosh, J. Math. Phys. **20**, 1466 (1979).

[5] D. Zwanziger, in *Invariant Wave Equations*, proceedings of the "Ettore Majorana" International School of Mathematical Phyiscs, Erice, Italy, 1977, *Lecture Notes in Physics*, Vol. 73, edited by G. Velo and A. S. Wightman (Springer, New York, 1978), p. 160.

[6] G. Velo and D. Zwanziger, Phys. Rev. **186**, 1337 (1969).

[7] G. Velo and D. Zwanziger, Phys. Rev. **188**, 2218 (1969).

[8] T. Darkhosh, Phys. Rev. D **32**, 3251 (1985).

Extensions of Wigner's distribution to particles with spin $\frac{1}{2}$

Basil V. Landau
Department of Mathematics, The Papua New Guinea University of Technology, Lae, Papua New Guinea

(Received 16 September 1986; accepted for publication 4 March 1987)

For particles of spin $\frac{1}{2}$ a class of distributions closely related to Wigner's is introduced on dynamical grounds. It is found that they may be refined to give correct expectation values of higher powers of spin components, but depart somewhat from a criterion that has been used in characterizing the Wigner distribution. For a special choice amongst this class, a more subtle refinement is possible satisfying this criterion exactly. This requires, however, a dubious distinction at every point between positive spin about a direction and negative spin about the reverse direction.

I. INTRODUCTION: RELATION OF THIS PAPER TO PREVIOUS EXTENSIONS

The Wigner[1] distribution, expressible for a single particle as

$$w(\mathbf{x},\mathbf{p}) = w[\psi(\mathbf{x})] = (2\pi)^{-3}\int \psi^*\Big(\mathbf{x} - \frac{1}{2}\hbar\mathbf{y}\Big)e^{-i\mathbf{p}\cdot\mathbf{y}}\psi\Big(\mathbf{x} + \frac{1}{2}\hbar\mathbf{y}\Big)d\mathbf{y}, \tag{1}$$

is a density that, like a probability density function, integrates to 1 over the whole space on which it is defined, provided the wave function ψ is normalized. Through it several aspects of quantum mechanics—expectations of an important class of functions of $\mathbf{x}$ and $\mathbf{p}$, also certain dynamical features[1,2]—can be expressed in an essentially classical manner, and without recourse to the orthodox operator formalism.

Stratonovich[3] introduced a distribution over spin direction that succeeds in treating in a comparable way expectations of spin components. However, when higher powers are treated by his formalism essentially nonclassical features appear which amount to a partial retention of the operator formalism, discussed in Sec. V. Some more recent formulations[4,5] incorporating spin $\frac{1}{2}$ into the Wigner distribution ignore his distribution over spin direction but take up his observation that spin and space aspects can be treated more or less independently—in his language the kernel of the combined representations is the product of the separate kernels. These formulations can consequently be described as direct products of a Wigner treatment of spatial features with an orthodox operator treatment of spin. While this is useful for some purposes the present paper is concerned rather with replacing as completely as possible the operator formalism by a density one. Of recent work that of Scully[6] perhaps comes closest to this aim, but has some unsatisfactory features. An amended form of his approach, given in the Appendix, turns out to be in agreement with the body of this paper.

This paper also differs from the four others referred to in the previous paragraph by first considering dynamical aspects of proposed distributions.

II. A PROPOSED EXTENSION, AND DYNAMICAL ASPECTS THEREOF

Given a spinor wave function with components ψ^+, ψ^- with respect to $0xyz$, and an orthonormal triad with third member $\mathbf{e}$, denote its spin-up component with respect to the triad by $\psi_{\mathbf{e}}$. The dependence on the other members of the triad is omitted; it affects $\psi_{\mathbf{e}}$ only by a multiplying constant of modulus unity, which is unimportant here. At a point where ψ^+ and ψ^- are not both zero, the direction $\mathbf{e}$ such that $\psi_{-\mathbf{e}}$ vanishes will be called the principal spin direction at the point.

The extension discussed first can be written as

$$f(\mathbf{x},\mathbf{p},\mathbf{e}) = (2\pi)^{-1}w[\psi_{\mathbf{e}}], \tag{2}$$

where the constant has been chosen to make the integral of f over all $\mathbf{x}$, $\mathbf{p}$, $\mathbf{e}$, equal to 1.

Dynamical aspects of the density functions w and f can be conveniently discussed using language appropriate to a material substratum underlying the particle. What may be called Wigner's dynamical principle[1] says that in the evolution of the distribution for a free particle satisfying Schrödinger's equation every portion of the substratum moves classically, without interaction with any other portion. Knowing that Wigner's principle holds for w, one easily sees that it holds for f also; for the Schrödinger equation satisfied by the spinor wave function of a free particle of spin $\frac{1}{2}$ implies that, for each $\mathbf{e}$, $\psi_{\mathbf{e}}$ satisfies that for a spinless particle of the same mass. That f satisfies this dynamical principle, has an isotropic definition, and seems to be the simplest distribution over $\mathbf{x}$, $\mathbf{p}$, $\mathbf{e}$ with these properties, is what has led the author to study this distribution further.

A second dynamical principle,[2] weaker but more widely applicable, governs the evolution of w: Weaker, because it does not forbid interaction between portions of the substratum in different regions of phase space provided they are in the same region of configuration space and conserve material and momentum; more widely applicable, because it is not limited to the free particle. For a particle of mass m subject to a force field $\mathbf{F}$ the portion of substratum occupying $d\mathbf{x}\,d\mathbf{p}$ is considered to have mass $mw\,d\mathbf{x}\,d\mathbf{p}$, momentum $\mathbf{p}w\,d\mathbf{x}\,d\mathbf{p}$, etc., and to be subject to an external force $\mathbf{F}w\,d\mathbf{x}\,d\mathbf{p}$, as well as possibly to actions from portions with different $\mathbf{p}$ but in

 0022-2488/87/071508-04$02.50

the same interval $d\mathbf{x}$ of configuration space. In a field derivable from a scalar potential, classical evolution of w is now to be understood as involving such local interaction that (1) continues to hold with suitably varying ψ; the second dynamical principle asserts this evolution to be in accordance with Schrödinger's equation.

For forces independent of spin the extension to f of this second principle is straightforward, so attention focuses on motion in a magnetic field $\mathbf{H}$ of an electrically neutral particle of spin $\frac{1}{2}$ and magnetic moment $\gamma\hbar/2$. [Neutrality is assumed for convenience here because Lorentz force would otherwise demand a modification[2] to formula (1) for w.] We suppose the spin angular momentum and magnetic moment of the portion of substratum occupying $d\mathbf{x}\, d\mathbf{p}\, d\mathbf{e} = d\Omega$ to be $S\hbar f\, \mathbf{e}\, d\Omega$ and $\gamma S\hbar\, \mathbf{e}\, d\Omega$, respectively, where S is a number discussed below. The force and moment acting on this portion are $\gamma S\hbar f \mathbf{e}\cdot\nabla\mathbf{H}\, d\Omega$ and $\gamma S\hbar f\mathbf{e}\wedge\mathbf{H}\, d\Omega$. The author found that with curl $\mathbf{H} = 0$ the classical interactive evolution of f (now also requiring conservation of spin angular momentum in local interaction), such that (2) continues to hold, is indeed in agreement with the spinor Schrödinger equation in a magnetic field. In the special case of a uniform magnetic field the motion turns out to be classical also in the original Wignerian sense, i.e., without interaction between portions of the substratum.

III. A REFINEMENT OF f TO GIVE CORRECT EXPECTATIONS

Equating any component of the total spin $\int S\mathbf{e}f\, d\Omega$ to its expected value as ordinarily calculated from the spinor wave function requires $S = \frac{3}{2}$, a surprise perhaps, but scarcely a difficulty. More serious is the failure of corresponding equalities for higher powers of spin components. Consider, for example, a case where $\psi^+ = \psi(\mathbf{x})$, and $\psi^- = 0$ everywhere, i.e., one of positive, and definite, spin about the z direction.

Then $\psi_{\mathbf{e}} = \psi\cos(\theta/2)$, whence, by (1) and (2), f is $(2\pi)^{-1}\cos^2(\theta/2)\, w[\psi]$, leading to

$$\int f\, d\mathbf{x}\, d\mathbf{p} = \frac{1+\cos\theta}{4\pi}. \tag{3}$$

Then the natural way to obtain $\langle s_z{}^n\rangle$ from f is to evaluate

$$\int (S\cos\theta)^n f\, d\Omega = \int \frac{S^n}{4\pi}\cos^n\theta(1+\cos\theta)\sin\theta\, d\theta\, d\phi$$
$$= \begin{cases} S^n/(n+1), & n \text{ even}, \\ S^n/(n+2), & n \text{ odd}. \end{cases}$$

This agrees with the correct result $(\frac{1}{2})^n$ only when $n = 0$ or 1. A clue to removing this discrepancy is afforded by the more general result for $\langle s_\alpha{}^n\rangle$, where s_α is the spin component in a direction making angle α with the z axis. One finds that for every α the discrepancy is by just the same factors, viz., $3^n/(n+1)$ for n even, $3^n/(n+2)$ for n odd, as in the $\alpha = 0$ case above. This suggests that we might replace f by a distribution F over $\mathbf{x}$, $\mathbf{p}$, $\mathbf{e}$, and spin intensity s, of the specially simple form $f(\mathbf{x}, \mathbf{p}, \mathbf{e})g(s)$. In that case $\int(s\cos\theta)^n \times F\, d\mathbf{x}\, d\mathbf{p}\, d\mathbf{s}$, with $d\mathbf{s} = s^2\sin\theta\, ds\, d\theta\, d\phi$, factorizes as $\int\cos^n\theta f\, d\Omega \int s^{n+2}g(s)ds$. Now equating this product to $\langle s_z{}^n\rangle$ in the above $\psi^- = 0$ case gives

$$2^n\int s^{n+2}g(s)ds = \begin{cases} n+1, & n \text{ even}, \\ n+2, & n \text{ odd}. \end{cases} \tag{4}$$

The intended range for s here was $[0,\infty)$—but then (4) has no solution. Over $(-\infty,\infty)$, however, one can apply the two-sided Laplace transform to find g, thus arriving, provisionally, at

$$F(\mathbf{x},\mathbf{p},\mathbf{e},s) = -\pi^{-1}w[\psi_{\mathbf{e}}]$$
$$\times\{\delta'(s-\tfrac{1}{2}) + \delta(s-\tfrac{1}{2}) + \delta(s+\tfrac{1}{2})\}.$$

To limit s to $(0,\infty)$ we can treat spin $-\frac{1}{2}$ about $\mathbf{e}$ as spin $\frac{1}{2}$ about $-\mathbf{e}$ and so replace the product $w[\psi_{\mathbf{e}}]\delta(s+\frac{1}{2})$ occurring above by $w[\psi_{-\mathbf{e}}]\delta(s-\frac{1}{2})$. This gives

$$-\pi F = w[\psi_{\mathbf{e}}]\delta'(s-\tfrac{1}{2})$$
$$+ \{w[\psi_{\mathbf{e}}] + w[\psi_{-\mathbf{e}}]\}\delta(s-\tfrac{1}{2}). \tag{5}$$

This refinement of f to F to include spin intensity as well as spin direction does not affect essentially the arguments that the two dynamical principles are satisfied, though the generalized functions make precise formulation of these principles more difficult.

IV. THE WIGNER–O' CONNELL PROBLEM FOR f and F, AND THE STRATONOVICH DISTRIBUTION

Though satisfactory dynamically, and in the case of F probabilistically also, f and F may be criticized in relation to what may be called, on account of its importance in their joint paper,[7] the Wigner–O'Connell property. This property concerns the integral, over the distribution space, of the product of distributions corresponding to two states; it asserts that this integral is proportional to the transition probability between the states. But for f and F the integral is in fact proportional to $A + B$, where, using (u,v) for $\int u^*(\mathbf{x})v(\mathbf{x})d\mathbf{x}$ and $|u,v|^2$ for $|(u,v)|^2$,

$$A = |(\psi_1^+,\psi_2^+) + (\psi_1^-,\psi_2^-)|^2$$

and

$$B = |\psi_1^+,\psi_2^+|^2 + |\psi_1^-,\psi_2^-|^2 + |\psi_1^+,\psi_2^-|^2 + |\psi_1^-,\psi_2^+|^2.$$

For the property to apply unmodified, B would have to be absent. Features of $A + B$ are that it vanishes if and only if the spinor wave functions are "strongly orthogonal," i.e., each component of the first is spatially orthogonal to each component of the second; that it attains its maximum 2 if and only if the states are not only identical but also have definite spin; and that it has value 1 for the orthogonal (but not strongly orthogonal) case when the states have identical space factors and definite but opposite spins (e.g., $\psi_1^+ = \psi$, $\psi_1^- = 0$, $\psi_2^+ = 0$, $\psi_2^- = \psi$). In the case of F, of course, the product to be integrated does not strictly exist; however, using for the δ function any of the usual smooth approximations, and for δ' its derivative, makes the terms involving $\int\{\delta'(s-\frac{1}{2})\}^2 ds$ swamp the others, leading in the limit to proportionality (by an "infinite constant") to $A + B$.

Many properties of $f(\mathbf{x}, \mathbf{p}, \mathbf{e})$ apply just as well, or with slight changes, to $f^\lambda = \lambda f + (1-\lambda)\bar{f}$, where $\bar{f}$ is obtained from f at any $\mathbf{x}$, $\mathbf{p}$ by averaging over $\mathbf{e}$. Some examples follow.

(i) The two dynamical principles are satisfied, with $S = 3/(2\lambda)$ in the second.

(ii) If we integrate f^λ over $\mathbf{p}$ at a given $\mathbf{x}$, we get a simple distribution over $\mathbf{e}$ which is symmetrical about the principal direction of spin at $\mathbf{x}$ and which has the same shape for every $\mathbf{x}$. In particular, where the principal axis is in the $\pm z$ direction, the density of this distribution is proportional to $1 \pm \lambda \cos\theta$.

(iii) The Wigner–O'Connell integral is proportional to a linear combination of A and B.

(iv) The right expectations for powers of spin components can be obtained by refining f^λ to F^λ via $f^\lambda g^\lambda(s)$—with the difference, important in the discussion below, that in the general case, as opposed to the $\lambda = 1$ case above, $\delta'(s+\frac{1}{2})$ as well as $\delta'(s-\frac{1}{2})$ appears in $g^\lambda(s)$.

The case $\lambda = \sqrt{3}$ is if special interest. Then f^λ is the expression, though not in the formalism Stratonovich[3] used, for his distribution. And, as we shall see, f^λ, though not F^λ, then satisfies the Wigner–O'Connell property. It should be said that in Stratonovich's approach $S\ (=\sqrt{3}/2)$ was in effect chosen before λ and in no sense represents a kind of spin intensity differing from $\frac{1}{2}$ as might be said of the value $S = \frac{3}{2}$ associated earlier with the $\lambda = 1$ case. Here $\sqrt{3}$ appears in the coefficient of $\cos\theta$ in the function representing $\hat{s}_z$ because of certain orthonormality conditions which make use of the equality of the integrals of $(\sqrt{3}\cos\theta)^2$ and of 1 over the sphere. From (ii) above one can see that in the case of spatially similar states of definite but opposite spin referred to earlier, and which make $A = 0$ and $B = 1$, the Wigner–O'Connell integral for f has as a factor the integral of $(1+\sqrt{3}\cos\theta)\ (1-\sqrt{3}\cos\theta)$ over the sphere, and hence vanishes. This implies that in the linear combination referred to in (iii) above, B must be absent, so Stratonovich's distribution satisfies the Wigner–O'Connell property.

On the other hand, in F^λ, obtained from $f^\lambda g^\lambda(s)$ by recasting negative values of s as positive ones associated with the opposite direction, $\delta'(s-\frac{1}{2})$ multiplies a linear combination of $f^\lambda(\mathbf{x},\mathbf{p},\mathbf{e})$ and $f^\lambda(\mathbf{x},\mathbf{p},-\mathbf{e})$, compared with $f^\lambda(\mathbf{x},\mathbf{p},\mathbf{e})$ only in the earlier $\lambda = 1$ case. This complication becomes important when two such distributions are multiplied, and prevents the simple Wigner–O'Connell property for $f^{\sqrt{3}}$ being carried over to $F^{\sqrt{3}}$. Indeed, if one accepts the grounds for adopting F^λ rather than f^λ, the Wigner–O'Connell approach favors F rather than $F^{\sqrt{3}}$, for, though neither satisfies the desired property, F leads to a simpler modification of it.

V. INTUITIVE AND OTHER ASPECTS IN STRATONOVICH'S FORMALISM AND THE PRESENT TREATMENT

The problem solved earlier through replacing f by F does not arise in Stratonovich's formalism which does not require $\cos^n\theta$ to appear in the function associated with $s_z{}^n$. Though the association of $\frac{1}{2}\sin\theta\cos\phi$, $\frac{1}{2}\sin\theta\sin\phi$, $\frac{1}{2}\cos\theta$ with s_x, s_y, s_z was suggested by "the direct physical meaning of the concept 'spin,' " he remains free within his formalism to obtain correct values for expectations by associating with $s_z{}^n$ the expressions $(\frac{1}{2})^n\cos\theta$ or $(\frac{1}{2})^n$ when n is odd or even, which comes close to direct use of the operator equality $\hat{s}_z{}^2 = \frac{1}{4}$. There is precedent of a sort for this in the Wigner distribution already in the one-dimensional case; when considering, say, $\langle (p^2+x^2)^n \rangle$, only for $n = 0, 1$ should $(p^2+x^2)^n$ be used in conjunction with the distribution. On the other hand the distribution was constructed so that for the basic variables x and p (and indeed for linear combinations of them) a naive connection of this sort held for all positive integer n. The fact that $\hat{s}_z{}^n$ is equal to a scalar multiple of $\hat{s}_z$ or of the identity does not mean that such a naive program is unsuitable when we include spin—it just means that the distributions will be such that multiplying them by $(s\cos\theta)^n$ and integrating produces the same results as multiplying by $(\frac{1}{2})^n$ or by $(\frac{1}{2})^{n-1}s\cos\theta$ and integrating.

While Stratonovich's formalism is not naive in the above sense, some of his statements are compatible with a naive view. Thus, "the 'representation distributions' of course do not give an entirely classical interpretation of quantum theory, but they provide a basis for that interpretation of quantum theory which has maximum closeness to classical ideas and thus has the greatest physical-intuitive meaning." And, characterizing the limitations of Moyal's notions,[8] insofar as Moyal had hoped they would apply to variables with no classical analog, Stratonovich says: "...with a discrete set of characteristic values of the 'basis' operators one must not restrict the distribution to the discrete values only, but must include the continuous spectrum in the treatment."

Using a three-dimensional instead of a two-dimensional manifold for the spin vector takes these ideas a stage further than Stratonovich himself did. In the actual form of F, or of $F^{\sqrt{3}}$, his restriction of s to $\frac{1}{2}$ is in a sense vindicated; but the presence of the δ' function makes a crucial difference from the sense in which he accepted this restriction.

VI. POSTSCRIPT: THE DOUBLY SPINNING ELECTRON

In Sec. III negative values of s were temporarily allowed as a device to obtain a solution for $g(s)$. To obtain F, negative spin about $\mathbf{e}$ was understood there as merely another way of referring to positive spin about $-\mathbf{e}$. Whether this equivalence amounted to a complete identity was of no consequence in the dynamical and probabilistic features then under discussion. But for the Wigner–O'Connell property the integrand is bilinear in the distribution and it does matter whether (i) s is limited, as in Secs. III and IV, to $[0,\infty)$ or (ii) s ranges, as in this section, over $(-\infty,\infty)$ while F^λ is taken as a product $f^\lambda(\mathbf{x},\mathbf{p},\mathbf{e})g^\lambda(s)$.

Alternative (ii), which might be called the doubly spinning electron, is capable, in contrast to (i), of satisfying the Wigner–O'Connell property; this is achieved by taking the Stratonovich value $\lambda = \sqrt{3}$. There then appears to be a good case for adopting this as the definitive extension of the Wigner distribution to a particle of spin $\frac{1}{2}$.

However, once we grant a mysterious distinction in spin space between $(\mathbf{e}, s)$ and $(-\mathbf{e}, -s)$, both of which correspond to the same point of (s_x, s_y, s_z) space, why should the distinction be allowed only *within* the distribution? For example, why should $f^\lambda(\mathbf{x},\mathbf{p},-\mathbf{e})g^\lambda(-s)$ not represent a different state from $f^\lambda(\mathbf{x},\mathbf{p},\mathbf{e})g^\lambda(s)$ even though they correspond to the same spinor wave function? In that case we would have two distinct states where we now admit one, and for an n electron system 2^n states where we now have one. We should then need a stronger exclusion mechanism than Fermi–Dirac statistics to generate Pauli's principle. The real question may be, is the doubly spinning electron merely a formal devices to save the Wigner–O'Connell property, or can it offer new insights?

ACKNOWLEDGMENT

The author wishes to thank a referee for acquainting him with the paper of Scully discussed below.

APPENDIX: THE SCULLY AND SCULLY–MOYAL KERNELS

Scully's paper[6] contains a recipe for a distribution over s_x, s_y, and s_z. For ease of comparison with a formulation containing explicitly as few hidden variables as his purpose will allow, he concentrates on functions of s_x and s_z. With this restriction his recipe leads in the spin-up case to a probability distribution assigning equal probability to values $(\pm\frac{1}{2},\frac{1}{2})$ for (s_x, s_z), and zero probability to other (s_x, s_z). As this gives correct expectations for $s_x{}^n$ and $s_z{}^n$, and for powers $n = 0,1,2$ of spin components about other directions in the xz plane, it may serve a worthwhile purpose in the context in which Scully introduces it. However, it fails to give correct expectations for higher powers of spin components about oblique directions.

When s_x, s_y, and s_z are all included, his recipe yields, in the spin-up case, equal probabilities for $(\pm\frac{1}{2}, \pm\frac{1}{2}, \frac{1}{2})$ where the two $\pm$ signs are independent, and zero probability for other (s_x, s_y, s_z). This is more obviously unacceptable, as it treats unequally different directions making the same angle with the z axis. Furthermore, when taken in conjunction with spatial variation so as to deal with states with different principal spin directions at different points, it does not satisfy the first dynamical principle (Wigner's) of Sec. II, though it does satisfy the second.

Scully's procedure involves at a certain stage the kernel

$$e^{i\xi\sigma_x}e^{i\eta\sigma_y}e^{i\zeta\sigma_z}$$

or better, such a product averaged over the different orders of the factors. Here σ_x, σ_y, σ_z are Pauli matrices. In considering one-dimensional motion a somewhat similar averaged product is

$$\tfrac{1}{2}(e^{i\alpha\hat{x}}e^{i\beta\hat{p}} + e^{i\beta\hat{p}}e^{i\alpha\hat{x}}).$$

Taking the trace of its product with the density matrix and applying a suitable Fourier transformation from α, β to x, p yields the one-dimensional Wigner distribution.

In this case the averaged product is identical with

$$e^{i(\alpha\hat{x} + \beta\hat{p})}.$$

But, owing to the commutation relations between pairs of σ_x, σ_y, σ_z being of a different form from that connecting $\hat{x}$ and $\hat{p}$, the Scully kernel above is not the same as

$$e^{i(\xi\sigma_x + \eta\sigma_y + \zeta\sigma_z)}.$$

This last expression may be called the Scully–Moyal kernel, being similar in some ways to expressions used in conjunction with Fourier transformations by Moyal[8] to yield Wigner-type distribution functions. We can apply appropriate Fourier transformation, and selection of the 1,1 element (i.e., taking the trace with density matrix $\left[\begin{smallmatrix}1 & 0\\ 0 & 0\end{smallmatrix}\right]$, for the spin-up case) to the Scully–Moyal kernel, as Scully did to his kernel. When this is done one finds, instead of the unsatisfactory distribution mentioned above, a distribution over spin magnitude s and spin direction $\mathbf{e}$, proportional to

$$(1 + \cos\theta)\delta'(s - \tfrac{1}{2}) + 2\delta(s - \tfrac{1}{2}),$$

where θ is the angle between $\mathbf{e}$ and the z direction. This is identical with that associated with the distribution denoted by F in the body of the paper [see Sec. III, Eq. (5)]. Thus, indirectly, Scully's work can be said to lend support to the kind of distribution considered in this paper; and in particular to the choice of $\lambda = 1$ rather than $\lambda = \sqrt{3}$ in F^λ, suggested at the end of Sec. IV.

[1]E. Wigner, Phys. Rev. **40**, 749 (1932); and in "Quantum mechanical distributions revisited," in *Perspectives in Quantum Theory*, edited by W. Yourgrau and A. van der Merwe (MIT, Cambridge, MA, 1971).

[2]B. V Landau, Found. Phys. **14**, 507 (1984).

[3]R. L. Stratonovich, Sov. Phys. JETP **4**, 891 (1957).

[4]P. Carruthers and F. Zachariasen, Rev. Mod. Phys. **55**, 245 (1983).

[5]R. F. O'Connell and E. P. Wigner, Phys. Rev. A **30**, 2613 (1984).

[6]M. O. Scully, Phys. Rev. D **28**, 2477 (1983).

[7]R. F. O'Connell and E. P. Wigner, Phys. Lett. A **83**, 145 (1981).

[8]J. E. Moyal, Proc. Cambridge Philos. Soc. **45**, 99 (1949).

Witten index, axial anomaly, and Krein's spectral shift function in supersymmetric quantum mechanics

D. Bollé
Instituut voor Theoretische Fysica, Universiteit Leuven, B-3030 Leuven, Belgium

F. Gesztesy[a]
LPTHE–Orsay,[b] Université de Paris-Sud, F-91405 Orsay, France

H. Grosse
Institut für Theoretische Physik, Universität Wien, A-1090 Wien, Austria

W. Schweiger
Institut für Theoretische Physik, Universität Graz, A-8010 Graz, Austria

B. Simon
Division of Physics, Mathematics and Astronomy, Caltech, Pasadena, California 91125

(Received 13 November 1986; accepted for publication 11 February 1987)

A new method is presented to study supersymmetric quantum mechanics. Using relative scattering techniques, basic relations are derived between Krein's spectral shift function, the Witten index, and the anomaly. The topological invariance of the spectral shift function is discussed. The power of this method is illustrated by treating various models and calculating explicitly the spectral shift function, the Witten index, and the anomaly. In particular, a complete treatment of the two-dimensional magnetic field problem is given, without assuming that the magnetic flux is quantized.

I. INTRODUCTION

Since the first observation of fractionally charged states in certain field theoretic soliton models,[1] various techniques to obtain a more detailed understanding of that phenomenon have been developed.[2] Furthermore, the possible phenomenological realization of these states in one-dimensional polymers such as polyacetylene strongly stimulated this development.[3–6]

Among the different existing approaches[2] the treatment of external field problems offers the simplest possibility to study fractional charge quantum numbers. In this context, one starts from a Dirac operator with some external potential with nontrivial asymptotics. For example, in one dimension this can be realized in the easiest way by considering the following operator, acting on two-component wave functions:

$$Q_m = \begin{pmatrix} m & A^* \\ A & -m \end{pmatrix}, \quad A = \frac{d}{dx} + \phi, \tag{1.1}$$

where $\phi(x)$ and $m(x)$ are space-dependent "mass" terms. Nontrivial (solitonlike) asymptotics is then expressed by $\lim_{x\to\pm\infty} \phi(x) = \phi_\pm$, in comparison with the trivial case $\lim_{x\to\pm\infty} \phi(x) = \phi_0$. Since, in a field theoretic context, the transition from one case to the other corresponds to the passage from one representation of the canonical anticommutation relations to an inequivalent one, the relative charge is usually defined through a regularization procedure. It turns out that under suitable conditions on the Dirac Hamiltonian, the charge is given by half of the associated η_m invariant.[2,7–12]

The method described above (for $m = 0$) is closely connected with supersymmetry, a subject of current interest in different fields of physics.[13,14] Indeed, the Hamiltonian defined as

$$H = Q^2 = \begin{pmatrix} A^*A & 0 \\ 0 & AA^* \end{pmatrix} \tag{1.2}$$

represents two Schrödinger operators, A^*A and AA^*, which are non-negative and which have the same spectrum, except perhaps for zero modes. The investigation of such supersymmetric quantum mechanical models is important. They serve as a laboratory to test and to understand supersymmetry breakdown in realistic field theories.[2,14–16] Furthermore, they provide a simple recipe for generating partner potentials, which can be used successfully in many physical problems. See Ref. 13 and references therein.

To study supersymmetric systems, Witten[16] introduced a quantity Δ, counting the difference in the number of bosonic and fermionic zero-energy modes of the Hamiltonian. This quantity, called the Witten index, has to be regularized if the threshold of the continuous spectrum of A^*A (AA^*) extends down to zero (see, e.g., Refs. 2 and 16–19). Here we will use the resolvent regularization, viz., Ref. 17,

$$\begin{aligned} \Delta &= \lim_{z\to 0} \Delta(z)\,, \\ \Delta(z) &= -z\,\mathrm{Tr}[(A^*A - z)^{-1} - (AA^* - z)^{-1}]\,. \end{aligned} \tag{1.3}$$

When A is Fredholm (i.e., if and only if the infimum of the essential spectrum of A^*A is strictly positive), this index Δ equals the Fredholm index $i(A) \equiv [\dim \mathrm{Ker}(A) - \dim \mathrm{Ker}(A^*)]$. When A is not Fredholm, this equality is, in general, destroyed and Δ can become noninteger; in fact, it can be any arbitrary real number,[20] due to threshold effects.

[a] On leave of absence from Institut für Theoretische Physik, Universität Graz, A-8010 Graz, Austria.
[b] Laboratoire associé au CNRS.

 0022-2488/87/071512-14$02.50

Fractionization of Δ has been seen explicitly in a number of examples.[2,8,20–25]

In this paper, we develop a new method to study supersymmetric quantum mechanics without assuming the Fredholm property for the operator A. This method, based on relative scattering techniques (Levinson theorem-type arguments, etc.), has the advantage of being simple and mathematically rigorous at the same time. In particular, we derive a relationship between Krein's spectral shift function[26–28] and the Witten index Δ. Furthermore, we show how the topological invariance of the (resolvent) regularized Witten index leads to the corresponding invariance of the spectral shift function itself. These new results offer a useful tool for explicit model calculations. To illustrate this, we discuss several examples in detail. A short account of this work has appeared in Ref. 20.

The rest of this paper is organized as follows: In Sec. II, we recall the basic properties of Krein's spectral shift function, $\xi(\lambda)$, λ the energy, and its connection with (modified) Fredholm determinants.[29–31] In Sec. III, we consider supersymmetric quantum mechanical systems. We prove that under certain conditions on the Hamiltonian, the Witten index Δ is given as (minus) the jump of the spectral shift function $\xi(\lambda)$ at $\lambda = 0$ and that the axial anomaly $\mathscr{A}$ (Refs. 17 and 32) is equal to the limit of $\xi(\lambda)$ as $\lambda \to \infty$. Furthermore, we use the topological invariance of the resolvent regularized Witten index under "sufficiently small" perturbations to derive the corresponding invariance of Krein's spectral shift function itself. Finally, we discuss the spectral asymmetry η_m associated with Q_m in terms of $\xi(\lambda)$. Section IV illustrates the power of our method in explicit calculations by treating a number of models. Using the connection between Fredholm determinants and Wronskians[33] or exploiting the topological invariance discussed in Sec. III, we calculate in a straightforward way Krein's spectral shift function, the Witten index, and the anomaly for various examples on the line and on the half-line. Furthermore, we analyze the supersymmetric system describing a particle in a two-dimensional magnetic field without assuming the magnetic flux to be quantized. In this case, our method is the first rigorous and nonperturbative one that shows that the spectral shift function is piecewise constant, and thus that both the anomaly $\mathscr{A}$ and minus the index Δ are equal to the flux. Also, the spectral asymmetry for the corresponding two-dimensional Q_m model is calculated.

We end this introduction with the remark that Secs. III and IV are completely self-contained, so that they may be read independently of Sec. II, which offers a full account of the more technical results needed in the paper.

II. FREDHOLM DETERMINANTS AND KREIN'S SPECTRAL SHIFT FUNCTION

In this section, we present a full account of those basic, more technical results on Krein's spectral shift function and its connection with Fredholm determinants that we need in the rest of the paper. We start by introducing the following hypotheses. For any result, only some of the hypotheses will be assumed.

Hypothesis (i): Let $\mathscr{H}$ be some (complex, separable) Hilbert space, let H_j, $j = 1,2$, be two self-adjoint operators in $\mathscr{H}$ such that $(H_1 - z_0)^{-1} - (H_2 - z_0)^{-1} \in \mathscr{B}_1(\mathscr{H})$ for some $z_0 \in \rho(H_1) \cap \rho(H_2)$.

[Here $\mathscr{B}_p(\mathscr{H})$, $p \in [1,\infty)$ denote the usual trace ideals[31] and $\rho(\cdot)$ denotes the resolvent set.]

Hypothesis (ii): In addition to Hypothesis (i), assume that H_j, $j = 1,2$, are bounded from below. Suppose that $H_1 = H_2 \dotplus V_{12}$ (here $\dotplus$ denotes the form sum), where V_{12} can be split into two parts, $V_{12} = v_{12} u_{12}$ such that $u_{12}(H_2 - z)^{-1} v_{12}$ is analytic with respect to $z \in \rho(H_2)$ in the $\mathscr{B}_1(\mathscr{H})$ topology and such that $u_{12}(H_2 - z_0)^{-1}$, $(H_2 - z_0)^{-1} v_{12} \in \mathscr{B}_2(\mathscr{H})$ for some $z_0 \in \rho(H_2)$.

Clearly, Hypothesis (ii) resembles the Rollnik trick of splitting a self-adjoint multiplication operator $V(x)$ into $V(x) = |V(x)|^{1/2} |V(x)|^{1/2} \operatorname{sgn} V(x)$.[34] Next, we introduce a "high-energy" assumption of the following type.

Hypothesis (iii): Assume Hypothesis (ii) and

$$\lim_{\substack{|z| \to \infty \\ \operatorname{Im} z \neq 0}} \det[1 + u_{12}(H_2 - z)^{-1} v_{12}] = 1 .$$

Finally, we introduce two assumptions which will allow generalizations in the sense that the Fredholm determinant used later on can be replaced by a modified one. This generalization is critical in higher-dimensional systems where Hypothesis (iii) is known to fail (cf., e.g., Refs. 35 and 36).

Hypothesis (iv): Suppose Hypothesis (ii) is satisfied except that $u_{12}(H_2 - z)^{-1} v_{12}$ is now assumed to be analytic with respect to $z \in \rho(H_2)$ in the $\mathscr{B}_2(\mathscr{H})$ norm.

Hypothesis (v): Assume Hypothesis (iv) and

$$\lim_{\substack{z \to \infty \\ \operatorname{Im} z \neq 0}} \det_2[1 + u_{12}(H_2 - z)^{-1} v_{12}] = 1 .$$

We first recall the following.

Lemma 2.1: Assume Hypothesis (i). Then there exists a real-valued measurable function ξ_{12} on $\mathbb{R}$ (Krein's spectral shift function[26–28]) unique a.e. up to a constant with

(a) $(1 + |\cdot|^2)^{-1} \xi_{12} \in L^1(\mathbb{R})$; (2.1)

(b)
$$\operatorname{Tr}[(H_1 - z)^{-1} - (H_2 - z)^{-1}] = -\int_{\mathbb{R}} d\lambda\, \xi_{12}(\lambda)(\lambda - z)^{-2}, \quad z \in \rho(H_1) \cap \rho(H_2); \tag{2.2}$$

(c) if $S_{12}(\lambda)$ denotes the on-shell scattering operator for the pair (H_1, H_2), then

$$\det S_{12}(\lambda) = e^{-2\pi i \xi_{12}(\lambda)} \quad \text{for a.e. } \lambda \in \sigma_{\text{ac}}(H_j) \tag{2.3}$$

[$\sigma_{\text{ac}}(\cdot)$ denotes the absolutely continuous spectrum].

For a proof, see, e.g., Refs. 37 and 38. For an appropriate class of $C^1(\mathbb{R})$ functions Φ with $\Phi(H_1) - \Phi(H_2) \in \mathscr{B}_1(\mathscr{H})$, one gets similarly

$$\operatorname{Tr}[\Phi(H_1) - \Phi(H_2)] = \int_{\mathbb{R}} d\lambda\, \xi_{12}(\lambda) \Phi'(\lambda) \tag{2.4}$$

(cf. Refs. 37–39). Finally, the invariance principle for wave operators can be used to relate ξ_{12} associated with (H_1, H_2) and ξ_{12}^{Φ} corresponding to $(\Phi(H_1), \Phi(H_2))$ by[37]

$$\xi_{12}(\lambda) = \xi_{12}^{\Phi}(\Phi(\lambda)) \operatorname{sgn}(\Phi'(\lambda)) . \tag{2.5}$$

If $H_j, j=1,2$, are bounded from below, we define $\xi_{12}(\lambda)=0$ to the left of the spectra of H_1 and H_2 in order to guarantee uniqueness for ξ_{12}. For connections between Levinson's theorem and ξ_{12}, see, e.g., Refs. 28, 40, and 41.

Example 2.2: Let H_1 denote the Friedrichs extension of $(-d^2/dx^2+\alpha/x^2)|_{C_0^\infty(\mathbb{R}\setminus\{0\})}$ in $L^2(\mathbb{R})$, $\alpha\geqslant -\frac{1}{4}$ and

$$H_2=-\frac{d^2}{dx^2}\Big|_{H^{2,2}(\mathbb{R})}. \tag{2.6}$$

Then the on-shell scattering operator $S_{12}(\lambda)$ in $\mathbb{C}^2$ reads[42]

$$S_{12}(\lambda)=\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix}e^{-i\pi[(\alpha+1/4)^{1/2}+1/2]}, \quad \lambda>0. \tag{2.7}$$

Thus

$$\xi_{12}(\lambda)=\begin{cases}0, & \lambda<0,\\ (\alpha+\frac{1}{4})^{1/2}, & \lambda>0,\end{cases} \tag{2.8}$$

and, e.g.,

$$\mathrm{Tr}[(H_1-z)^{-1}-(H_2-z)^{-1}] = (\alpha+\tfrac{1}{4})^{1/2}z^{-1}, \quad z\in\mathbb{C}\setminus[0,\infty). \tag{2.9}$$

By a Laplace transform, Eq. (2.9) is equivalent to a result of Ref. 43. If H_1 equals the Neumann instead of Friedrich's extension of $(-d^2/dx^2+\alpha/x^2)|_{C_0^\infty(\mathbb{R}\setminus\{0\})}$, $\alpha>-\frac{1}{4}$, one obtains[42]

$$\xi_{12}(\lambda)=\begin{cases}0, & \lambda<0,\\ -(\alpha+\frac{1}{4})^{1/2}, & \lambda>0.\end{cases}$$

Next, we recall[28,29] the following.

Lemma 2.3: (a) Let $U,G\subset\mathbb{C}$ be open, $A\in\mathscr{B}_p(\mathscr{H})$ for some $p\in[1,\infty)$, and $\sigma(A)\subset U\subset G$, where ∂U is compact and consists of a finite number of closed rectifiable Jordan curves (cf., e.g., Ref. 44) oriented in the positive sense. [Here $\sigma(\cdot)$ denotes the spectrum and ∂U denotes the boundary of the set U.] Let f: $G\to\mathbb{C}$ be analytic with $f(0)=0$. Then $f(A)\in\mathscr{B}_p(\mathscr{H})$.

(b) Let A: $[a,b]\to\mathscr{B}_1(\mathscr{H})$ be continuously differentiable in the $\mathscr{B}_1(\mathscr{H})$ norm. Let $\cup_{t\in[a,b]}\sigma(A(t))\subset G$, where $G\subset\mathbb{C}$ is open. Let f: $G\to\mathbb{C}$ be analytic with $f(0)=0$. Then

$$\frac{d}{dt}\mathrm{Tr}[f(A(t))]=\mathrm{Tr}\left[f'(A(t))\frac{dA(t)}{dt}\right], \quad t\in(a,b). \tag{2.10}$$

(c) Let $G\subset\mathbb{C}$ be open, and A: $G\to\mathscr{B}_1(\mathscr{H})$ be analytic in the $\mathscr{B}_1(\mathscr{H})$ norm. Then $\det[1+A(z)]$ is analytic with respect to $z\in G$ and

$$\frac{d}{dz}\ln\det[1+A(z)] = \mathrm{Tr}\left\{[1+A(z)]^{-1}\frac{dA(z)}{dz}\right\}, \quad -1\notin\sigma(A(z)), \quad z\in G. \tag{2.11}$$

Lemma 2.3 immediately implies the following.

Lemma 2.4: Assume Hypothesis (ii). Then

$$\mathrm{Tr}[(H_1-z)^{-1}-(H_2-z)^{-1}] = -\frac{d}{dz}\ln\det[1+u_{12}(H_2-z)^{-1}v_{12}], \quad z\in\rho(H_1)\cap\rho(H_2). \tag{2.12}$$

Proof: By Lemma 2.3, cyclicity of the trace, and the resolvent equation one gets

$$\begin{aligned}\frac{d}{dz}\ln\det[1+u_{12}(H_2-z)^{-1}v_{12}] &= \mathrm{Tr}\{[1+u_{12}(H_2-z)^{-1}v_{12}]^{-1}u_{12}(H_2-z)^{-2}v_{12}\}\\ &= \mathrm{Tr}\{(H_2-z)^{-1}v_{12}[1+u_{12}(H_2-z)^{-1}v_{12}]^{-1}\\ &\quad\times u_{12}(H_2-z)^{-1}\}\\ &= -\mathrm{Tr}[(H_1-z)^{-1}-(H_2-z)^{-1}],\\ &\quad z\in\rho(H_1)\cap\rho(H_2).\end{aligned}$$ □

In order to connect Krein's spectral shift function with Fredholm determinants, we formulate the following.

Lemma 2.5: Assume Hypothesis (iii) and assume that $(1+|\cdot|)^{-1}\xi_{12}\in L^1(\mathbb{R})$. Then

$$\int_{\mathbb{R}}d\lambda\,\xi_{12}(\lambda)(\lambda-z)^{-1} = \ln\det[1+u_{12}(H_2-z)^{-1}v_{12}], \quad z\in\rho(H_1)\cap\rho(H_2). \tag{2.13}$$

If, in addition, ξ_{12} is bounded and piecewise continuous on $\mathbb{R}$, then

$$[\xi_{12}(\lambda_+)+\xi_{12}(\lambda_-)]/2 = \frac{1}{2\pi i}\lim_{\epsilon\to 0_+}\ln\frac{\det[1+u_{12}(H_2-\lambda-i\epsilon)^{-1}v_{12}]}{\det[1+u_{12}(H_2-\lambda+i\epsilon)^{-1}v_{12}]}, \quad \lambda\in\mathbb{R}. \tag{2.14}$$

Proof: By Lemma 2.4, we have

$$\begin{aligned}-\frac{d}{dz}\ln\det[1+u_{12}(H_2-z)^{-1}v_{12}] &= \mathrm{Tr}[(H_1-z)^{-1}-(H_2-z)^{-1}]\\ &= -\frac{d}{dz}\int_{\mathbb{R}}d\lambda\,\xi_{12}(\lambda)(\lambda-z)^{-1}, \quad z\in\rho(H_1)\cap\rho(H_2).\end{aligned}$$

Thus Eq. (2.13) holds up to a constant. By Hypothesis (iii), this constant equals zero. Equation (2.14) results from standard properties of the Poisson kernel (cf., e.g., Ref. 45). □

Without the piecewise continuity of ξ_{12}, Eq. (2.14) holds a.e. in $\lambda\in\mathbb{R}$. Hypothesis (iii) is, in general, valid for one-dimensional systems (cf. Sec. IV) but breaks down in higher dimensions. Thus we formulate the following.

Lemma 2.6: Let $G\subset\mathbb{C}$ be open, and A: $G\to\mathscr{B}_2(\mathscr{H})$ be analytic in $\mathscr{B}_2(\mathscr{H})$ topology. Then the modified Fredholm determinant $\det_2[1+A(z)]$ is analytic with respect to $z\in G$ and

$$\begin{aligned}\frac{d}{dz}\ln\det{}_2[1+A(z)] &= \mathrm{Tr}\left\{([1+A(z)]^{-1}-1)\frac{dA(z)}{dz}\right\}\\ &= -\mathrm{Tr}\left\{[1+A(z)]^{-1}A(z)\frac{dA(z)}{dz}\right\},\\ &\quad -1\notin\sigma(A(z)), \quad z\in G.\end{aligned} \tag{2.15}$$

Proof: Obviously Eq. (2.15) holds for $A(z)\in\mathscr{B}_1(\mathscr{H})$, $z\in G$. The general case follows by a limiting argument. □

Lemma 2.7: Assume Hypothesis (iv). Then

$$\begin{aligned}\mathrm{Tr}[&(H_1-z)^{-1}-(H_2-z)^{-1}\\&+(H_2-z)^{-1}V_{12}(H_2-z)^{-1}]\\&=-\frac{d}{dz}\ln \det{}_2[1+u_{12}(H_2-z)^{-1}v_{12}]\,,\\&\quad z\in\rho(H_1)\cap\rho(H_2)\,.\end{aligned}\tag{2.16}$$

Proof: By Lemma 2.6 one gets

$$\begin{aligned}\frac{d}{dz}&\ln \det{}_2[1+u_{12}(H_2-z)^{-1}v_{12}]\\&=\mathrm{Tr}\{([1+u_{12}(H_2-z)^{-1}v_{12}]^{-1}-1)\\&\quad\times u_{12}(H_2-z)^{-2}v_{12}\}\\&=-\mathrm{Tr}\{(H_1-z)^{-1}-(H_2-z)^{-1}\\&\quad+(H_2-z)^{-1}V_{12}(H_2-z)^{-1}\}\,,\\&\quad z\in\rho(H_1)\cap\rho(H_2)\,.\end{aligned}$$ □

For related work, see also Ref. 46.

Next, we assume the existence of some η_{12}: $[\lambda_0,\infty)\to\mathbb{R}$ such that

$$\begin{aligned}\mathrm{Tr}[&(H_2-z)^{-1}V_{12}(H_2-z)^{-1}]\\&=\int_{\lambda_0}^{\infty}d\lambda\,\eta_{12}(\lambda)(\lambda-z)^{-2},\quad z\in\rho(H_2)\,,\end{aligned}\tag{2.17}$$

and we define

$$\tilde{\xi}_{12}(\lambda)=\begin{cases}\xi_{12}(\lambda)-\eta_{12}(\lambda)\,, & \lambda>\lambda_0\,,\\ \xi_{12}(\lambda)\,, & \lambda<\lambda_0\,.\end{cases}\tag{2.18}$$

Lemma 2.8: Assume Hypothesis (v) and assume that $(1+|\cdot|)^{-1}\tilde{\xi}_{12}\in L^1(\lambda_0,\infty)$. Then

$$\begin{aligned}\int_{\mathbb{R}}&d\lambda\,\tilde{\xi}_{12}(\lambda)(\lambda-z)^{-1}\\&=\ln\det{}_2[1+u_{12}(H_2-z)^{-1}v_{12}]\,,\\&\quad z\in\rho(H_1)\cap\rho(H_2)\,.\end{aligned}\tag{2.19}$$

If, in adddition, $\tilde{\xi}_{12}$ is piecewise continuous and bounded on $\mathbb{R}$, then

$$\begin{aligned}[\tilde{\xi}_{12}&(\lambda_+)+\tilde{\xi}_{12}(\lambda_-)]/2\\&=\frac{1}{2\pi i}\lim_{\epsilon\to 0_+}\ln\frac{\det_2[1+u_{12}(H_2-\lambda-i\epsilon)^{-1}v_{12}]}{\det_2[1+u_{12}(H_2-\lambda+i\epsilon)^{-1}v_{12}]}\,.\end{aligned}\tag{2.20}$$

Proof: Similar to that of Lemma 2.5. □

Example 2.9: Let $|V_{12}|^{1+s}\in L^1(\mathbb{R}^2)$, $(1+|\cdot|^s)V_{12}\in L^1(\mathbb{R}^2)$ for some $s>0$, respectively, $V_{12}\in L^1(\mathbb{R}^3)\cap R$ (R the Rollnik class,[34] i.e.,

$$\int_{\mathbb{R}^6}d^3x\,d^3y|V(x)||V(y)||x-y|^{-2}<\infty)$$

and define in $L^2(\mathbb{R}^n)$: $H_1=-\Delta\dotplus V_{12}$ and $H_2=-\Delta|_{H^{2,2}(\mathbb{R}^n)}$, $n=2,3$. Then

$$\begin{aligned}\mathrm{Tr}[&(H_2-z)^{-1}V_{12}(H_2-z)^{-1}]\\&=-\frac{1}{4\pi}\int_{\mathbb{R}^n}d^nx\,V_{12}(x)\\&\quad\times\begin{cases}z^{-1}\,, & n=2\,,\\(2\sqrt{-z})^{-1}\,, & n=3\,,\end{cases}\quad z\in\mathbb{C}\setminus[0,\infty)\end{aligned}\tag{2.21}$$

and hence $\lambda_0=0$ and (cf., e.g., Refs. 35 and 36)

$$\eta_{12}(\lambda)=\begin{cases}0\,, & \lambda<0\,,\\ -\dfrac{1}{4\pi}\displaystyle\int_{\mathbb{R}^n}d^nx\,V_{12}(x)\begin{cases}1\,, & n=2,\\ \sqrt{\lambda}/\pi\,, & n=3\,,\end{cases} & \lambda>0\,.\end{cases}\tag{2.22}$$

Finally, assume Hypothesis (i) and define, for some $M\in\mathbb{R}$,

$$\begin{aligned}\Delta_M(z)&=-(z-M)\,\mathrm{Tr}[(H_1-z)^{-1}-(H_2-z)^{-1}]\,,\\&\quad z\in\rho(H_1)\cap\rho(H_2)\,.\end{aligned}\tag{2.23}$$

Furthermore, define

$$\Delta_M=\lim_{\substack{z\to M\\|\mathrm{Re}\,z-M|\leqslant C_0|\mathrm{Im}\,z|}}\Delta_M(z)\tag{2.24}$$

and, if in addition H_j, $j=1,2$, are bounded from below,

$$\mathscr{A}=-\lim_{\substack{z\to\infty\\|\mathrm{Re}\,z|\leqslant C_1|\mathrm{Im}\,z|}}\Delta_M(z)\tag{2.25}$$

(C_0,C_1 positive constants). Then one has the following.

Lemma 2.10: Assume Hypothesis (i).

(a) Let $M\in\mathbb{R}$ and suppose that ξ_{12} is bounded on $\mathbb{R}$ and piecewise continuous in $(M-2\delta,M+2\delta)$ for some $\delta>0$. Then

$$\Delta_M=\xi_{12}(M_-)-\xi_{12}(M_+)\,.\tag{2.26}$$

(b) If H_j, $j=1,2$, are bounded from below and if ξ_{12} is bounded and $\lim_{\lambda\to\infty}\xi_{12}(\lambda)=\xi_{12}(\infty)$ exists, then

$$\mathscr{A}=\xi_{12}(\infty)\,.\tag{2.27}$$

Proof: Choose $\epsilon>0$ sufficiently small,

$$\begin{aligned}\Delta_M(z)&=(z-M)\int_{M-\epsilon}^{M+\epsilon}d\lambda\,\xi_{12}(\lambda)(\lambda-z)^{-2}+O(z-M)\\&=\xi_{12}(M_-)-\xi_{12}(M_+)+(z-M)\int_{M}^{M+\epsilon}d\lambda\,[\xi_{12}(\lambda)-\xi_{12}(M_+)](\lambda-z)^{-2}\\&\quad+(z-M)\int_{M-\epsilon}^{M}d\lambda\,[\xi_{12}(\lambda)-\xi_{12}(M_-)](\lambda-z)^{-2}+O(z-M)\,.\end{aligned}$$

Now

$$\int_M^{M+\epsilon} d\lambda\, [\xi_{12}(\lambda) - \xi_{12}(M_+)](z-M)(\lambda - z)^{-2}$$

$$= \int_M^{M+\epsilon} d\lambda\, [\xi_{12}(\lambda) - \xi_{12}(M_+)]\{(\mathrm{Re}\, z - M)[(\lambda - \mathrm{Re}\, z)^2 - (\mathrm{Im}\, z)^2] - 2(\lambda - \mathrm{Re}\, z)(\mathrm{Im}\, z)^2\}$$

$$\times [(\lambda - \mathrm{Re}\, z)^2 + (\mathrm{Im}\, z)^2]^{-2} + i\int_M^{M+\epsilon} d\lambda\, [\xi_{12}(\lambda) - \xi_{12}(M_+)]\{(\mathrm{Im}\, z)[(\lambda - \mathrm{Re}\, z)^2 - (\mathrm{Im}\, z)^2]$$

$$+ 2(\lambda - \mathrm{Re}\, z)(\mathrm{Re}\, z - M)\mathrm{Im}\, z\}[(\lambda - \mathrm{Re}\, z)^2 + (\mathrm{Im}\, z)^2]^{-2}. \tag{2.28}$$

For example, the real part in Eq. (2.28) yields

$$\int_{-\infty}^{\infty} d\mu \left\{\left[\mu^2 \frac{|\mathrm{Re}\, z - M|}{|\mathrm{Im}\, z|}\right] - \left[\frac{|\mathrm{Re}\, z - M|}{|\mathrm{Im}\, z|}\right] - 2\mu\right\}(\mu^2 + 1)^{-2}$$

$$\times [\xi_{12}((\mu|\mathrm{Im}\, z| + |\mathrm{Re}\, z - M|)\mathrm{sgn}(\mathrm{Re}\, z - M) + M) - \xi_{12}(M_+)]\chi_{I(M,z)}(\mu) \to 0$$

$$\text{as } z \to M \text{ and } |\mathrm{Re}\, z - M| \leqslant C_0 |\mathrm{Im}\, z|,$$

$$I(M,z) = [-|\mathrm{Re}\, z - M|/|\mathrm{Im}\, z|,\quad [\epsilon\, \mathrm{sgn}(\mathrm{Re}\, z - M) - |\mathrm{Re}\, z - M|]/|\mathrm{Im}\, z|]$$

by dominated convergence. (Here χ_I denotes the characteristic function of the interval $I \subset \mathbb{R}$.) The same analysis applies for the imaginary part in Eq. (2.28), proving Eq. (2.26). Similarly one proves Eq. (2.27). □

III. SUPERSYMMETRY AND KREIN'S SPECTRAL SHIFT FUNCTION

In this section we consider general supersymmetric quantum mechanical systems and we establish a basic relationship betwen Krein's spectral shift function $\xi_{12}(\lambda)$ and the Witten index, and between $\xi_{12}(\lambda)$ and the axial anomaly. Furthermore, we discuss the topological invariance of the (regularized) Witten index and the spectral shift function. Finally, the spectral asymmetry for Q_m-type models [cf. Eq. (1.1)] is related to $\xi_{12}(\lambda)$.

Let A be a closed, densely defined operator in $\mathscr{H}$ and define the "bosonic," respectively, "fermionic" Hamiltonian H_1 and H_2, by

$$H_1 = A^*A, \quad H_2 = AA^*. \tag{3.1}$$

The corresponding supercharge Q and the supersymmetric Hamiltonian H in $\mathscr{H} \oplus \mathscr{H}$ are, respectively,

$$Q = \begin{pmatrix} 0 & A^* \\ A & 0 \end{pmatrix}, \quad H = Q^2 = \begin{pmatrix} H_1 & 0 \\ 0 & H_2 \end{pmatrix}. \tag{3.2}$$

Assuming Hypothesis (i) throughout this section, Witten's (resolvent) regularized index $\Delta(z)$ is defined by[17]

$$\Delta(z) = -z\, \mathrm{Tr}\, [(H_1 - z)^{-1} - (H_2 - z)^{-1}],$$

$$z \in \mathbb{C} \setminus [0, \infty), \tag{3.3}$$

and Witten's index Δ (Ref. 16) is given by (cf. Sec. II)

$$\Delta = \lim_{\substack{z \to 0 \\ |\mathrm{Re}\, z| \leqslant C_0 |\mathrm{Im}\, z|}} \Delta(z) \tag{3.4}$$

(for some $C_0 > 0$) whenever the limit exists. Instead of the regularization (3.3), one could as well consider a (heat kernel) regularization $\widetilde{\Delta}(s)$ of the type

$$\widetilde{\Delta}(s) = \mathrm{Tr}[e^{-sH_1} - e^{-sH_2}], \quad s \geqslant 0, \tag{3.5}$$

and define Witten's index by

$$\Delta = \lim_{s \to \infty} \widetilde{\Delta}(s). \tag{3.6}$$

In order to avoid technicalities, we restrict ourselves to Callias's regularization (3.3).

As a first result, we try to relate Δ and the Fredholm index $i(A)$ of A: We recall an operator is Fredholm[47] iff A is a closed operator with a closed range such that dim Ker(A) and dim Ker(A^*) are finite. The Fredholm index $i(A)$ is then given by

$$i(A) = \dim \mathrm{Ker}(A) - \dim \mathrm{Ker}(A^*). \tag{3.7}$$

We remark that A is Fredholm iff A^* (or A^*A) is.[47] In addition

$$\dim \mathrm{Ker}(A) = \dim \mathrm{Ker}(A^*A) \tag{3.8}$$

implying that

$$i(A) = \dim \mathrm{Ker}(H_1) - \dim \mathrm{Ker}(H_2). \tag{3.9}$$

Thus $i(A)$ describes precisely the difference of bosonic and fermionic zero-energy states (counting multiplicities).

We emphasize that we shall also use definition (3.7) for $i(A)$ in case A is not Fredholm. Of course, in this case $i(A)$ might lose some of the typical properties of an index.

We state the following.

Theorem 3.1: Assume Hypothesis (i) and suppose A is Fredholm. Then

$$\Delta = i(A). \tag{3.10}$$

Proof: We only sketch the major step. The fact that H_j, $j = 1,2$, are Fredholm guarantees an expansion of the type

$$-z[(H_1 - z)^{-1} - (H_2 - z)^{-1}]$$

$$= P_1 - P_2 - z\sum_{n=0}^{\infty} z^n [T_1^{n+1} - T_2^{n+1}] \tag{3.11}$$

valid in the $\mathscr{B}_1(\mathscr{H})$ norm. Here P_j denotes the projection onto the eigenvalue zero of H_j, $j = 1,2$, and T_j is the reduced resolvent, viz., Ref. 47,

$$T_j = n - \lim_{z \to 0} (H_j - z)^{-1}[1 - P_j], \quad j = 1,2. \tag{3.12}$$

Taking the trace in Eq. (3.11) and observing that

$$\mathrm{Tr}[P_1 - P_2] = i(A) \tag{3.13}$$

completes the proof. □

What happens if A is not a Fredholm operator? Before trying to answer this question, let us consider an equivalent definition of the Fredholm property of A. Since $A^*A \geqslant 0$ and A is Fredholm iff A^*A is, we get the criterion that A is Fredholm iff $\inf \sigma_{\mathrm{ess}}(A^*A) > 0$ [$\sigma_{\mathrm{ess}}(\cdot)$ denotes the essential spectrum]. The examples of the next section show that, in general, equality (3.10) is violated if A is not Fredholm. In fact, Δ may take on half-integer values in the first four examples of Sec. IV, whereas in the fifth example it can even take on arbitrary real values (see also Ref. 20).

To study also these non-Fredholm cases we now introduce Krein's spectral shift function ξ_{12} associated with (H_1,H_2) as discussed in Sec. II. We always assume Hypothesis (vi). Assume that ξ_{12} (or $\tilde{\xi}_{12}$) is bounded and piecewise continuous on $\mathbb{R}$ and $\xi_{12}(\lambda) = 0$ for $\lambda < 0$.

As can be seen from Lemma 2.5 (Lemma 2.8), this essentially requires continuity of the trace-norm (Hilbert–Schmidt norm) limits $u_{12}(H_2 - \lambda \mp i0)^{-1}v_{12}$ with respect to $\lambda \in \mathbb{R}$. This can be checked explicitly in concrete examples (cf., e.g., Sec. IV).

Let us denote the threshold of H_j by

$$\Sigma = \inf \sigma_{\mathrm{ess}}(H_1) = (\inf \sigma_{\mathrm{ess}}(H_2)). \tag{3.14}$$

We observe that H_1 and H_2 are essentially isospectral[49] (cf. also Ref. 50), i.e.,

$$\sigma(H_1)\setminus\{0\} = \sigma(H_2)\setminus\{0\}$$

and

$$H_1 f = Ef, \quad E \neq 0$$

implies $H_2(Af) = E(Af)$, $f \in \mathscr{D}(H_1)$,

$$H_2 g = E'g, \quad E' \neq 0$$

implies $H_1(A^*g) = E'(A^*g)$, $g \in \mathscr{D}(H_2)$, (3.15)

with multiplicities preserved. Under the additional assumption that

$$\Sigma = \inf \sigma_{ac}(H_1) \quad [= \inf \sigma_{ac}(H_2)] \tag{3.16}$$

and that, e.g., $u_{12}(H_2 - \lambda - i\epsilon)^{-1}v_{12}$, $\lambda \geqslant \Sigma$, has $\mathscr{B}_2(\mathscr{H})$-valued limits as $\epsilon \to 0_+$ and that the exceptional set

$$\delta = \{\lambda \geqslant \Sigma \mid \exists f \in \mathscr{H},\, f \neq 0 \text{ with } u_{12}(H_2 - \lambda - i0)^{-1}v_{12} f = -f\} \tag{3.17}$$

is discrete (cf., e.g., Refs. 31 and 51), we get

$$\xi_{12}(\lambda) = \begin{cases} 0, & \lambda < 0, \\ \xi_{12}(0_+), & 0 < \lambda < \Sigma, \\ -(2\pi i)^{-1} \ln \det S_{12}(\lambda), & \lambda > \Sigma. \end{cases} \tag{3.18}$$

The simple structure in Eq. (3.18) follows from the fact that the effects of all nonzero bound states of H_1 and H_2 cancel since they occur with the same multiplicity in both H_1 and H_2.[49] Under suitable conditions on V_{12},[37,51] the on-shell S matrix $S_{12}(\lambda)$ is continuous in trace norm in $\lambda > \Sigma$ [with $\det S_{12}(\lambda) \neq 0$], implying continuity of ξ_{12} for $\lambda > \Sigma$. [If $\Sigma = 0$, then the second line of the rhs of Eq. (3.18) should be omitted.]

If we define the axial anomaly $\mathscr{A}$ by (cf. Refs. 17 and 32)

$$\mathscr{A} = -\lim_{\substack{z \to \infty \\ |\mathrm{Re}\, z| < C_1 |\mathrm{Im}\, z|}} \Delta(z) \tag{3.19}$$

(for some $C_1 > 0$) we obtain from Lemma 2.10 the following.

Theorem 3.2: Assume Hypotheses (i) and (vi). Then

$$\Delta = -\xi_{12}(0_+). \tag{3.20}$$

If, in addition, $\lim_{\lambda \to \infty} \xi_{12}(\lambda) \equiv \xi_{12}(\infty)$ exists, then

$$\mathscr{A} = \xi_{12}(\infty). \tag{3.21}$$

If $\Sigma > 0$, then $-\xi_{12}(0_+)$ describes precisely the difference of zero-energy bound states of H_1 and H_2 (counting multiplicity) since $\xi_{12}(\lambda) = 0$ for $\lambda < 0$. Thus $-\xi_{12}(0_+) = i(A)$ in agreement with Theorem 3.1. If $\Sigma = 0$, then $\xi_{12}(0_+)$ might be fractional due to threshold resonances or bound states of H_1 or H_2 or due to relative long-range interactions as shown in Sec. IV.

We also recall that by Lemma 2.5, ξ_{12} can be recovered from the Fredholm determinants by

$$[\xi_{12}(\lambda_+) + \xi_{12}(\lambda_-)]/2 = \frac{1}{2\pi i} \lim_{\epsilon \to 0_+} \ln \frac{\det[1 + u_{12}(H_2 - \lambda - i\epsilon)^{-1}v_{12}]}{\det[1 + u_{12}(H_2 - \lambda + i\epsilon)^{-1}v_{12}]} \tag{3.22}$$

assuming Hypotheses (iii) and (vi) and $(1 + |\cdot|)^{-1}\xi_{12} \in L^1(\mathbb{R})$. Under the same assumptions, $\Delta(z)$ is given by [cf. Eq. (2.13)]

$$\begin{aligned}\Delta(z) &= -z\,\mathrm{Tr}[(H_1 - z)^{-1} - (H_2 - z)^{-1}] \\ &= z\frac{d}{dz}\int_{\mathbb{R}} d\lambda\, \xi_{12}(\lambda)(\lambda - z)^{-1} \\ &= z\frac{d}{dz} \ln \det[1 + u_{12}(H_2 - z)^{-1}v_{12}], \quad z \in \mathbb{C}\setminus[0,\infty).\end{aligned} \tag{3.23}$$

We omit the corresponding generalizations based on Hypothesis (v) in terms of modified Fredholm determinants. If an expansion of the type

$$\det[1 + u_{12}(H_2 - z)^{-1}v_{12}] = z^{\alpha}[1 + O(z)] \quad \text{as } z \to 0 \tag{3.24}$$

holds, then obviously

$$\Delta = \alpha. \tag{3.25}$$

In the same way, a high-energy expansion determines the anomaly $\mathscr{A}$.

Next, we turn to an important invariance property of $\Delta(z)$ under sufficiently small perturbations of A. Let B be another closed operator in $\mathscr{H}$ infinitesimally bounded with respect to A, and introduce on $\mathscr{D}(A)$,

$$A_\beta = A + \beta B, \quad \beta \in \mathbb{R}. \tag{3.26}$$

The quantities $H_{1,\beta}$, $H_{2,\beta}$, $u_{12,\beta}$, $v_{12,\beta}$, $\xi_{12,\beta}$, and $\Delta(\beta,z)$ then result after replacing A by A_β. We have[48] the following.

Theorem 3.3: Fix $z_0 \in \mathbb{C}\setminus[0,\infty)$ and assume that

(i) $(H_{1,\beta} - z_0)^{-1} - (H_{2,\beta} - z_0)^{-1} \in \mathscr{B}_1(\mathscr{H})$ for all $\beta \in \mathbb{R}$;

(ii) $B^*B(H_1 - z_0)^{-1}$, $BB^*(H_2 - z_0)^{-1} \in \mathscr{B}_\infty(\mathscr{H})$,

$$[A^*B + B^*A](H_1 - z_0)^{-1},$$
$$[AB^* + BA^*](H_2 - z_0)^{-1} \in \mathscr{B}_\infty(\mathscr{H});$$

(iii) $(H_1 - z_0)^{-1}B^*B(H_1 - z_0)^{-1}$,
$(H_2 - z_0)^{-1}BB^*(H_2 - z_0)^{-1} \in \mathscr{B}_1(\mathscr{H})$,
$(H_1 - z_0)^{-1}[A^*B + B^*A](H_1 - z_0)^{-1}$,
$(H_2 - z_0)^{-1}[AB^* + BA^*](H_2 - z_0)^{-1} \in \mathscr{B}_1(\mathscr{H})$;

(iv) $(H_1 - z_0)^{-M}B^*(H_2 - z_0)^{-M} \in \mathscr{B}_1(\mathscr{H})$ for some $M \in \mathbb{N}$.

[Here $\mathscr{B}_1(\mathscr{H})$ and $\mathscr{B}_\infty(\mathscr{H})$ denote trace class and compact operators in $\mathscr{H}$, respectively.] Then

$$\Delta(\beta,z) = \Delta(z), \quad z \in \mathbb{C}\setminus[0,\infty), \quad \beta \in \mathbb{R}, \tag{3.27}$$

i.e., the regularized Witten index is invariant against small perturbations B of the above type.

Since a more general result (where A acts between different Hilbert spaces $\mathscr{H}$ and $\mathscr{H}'$) has been proven in Ref. 48, we only formally indicate the proof: By conditions (i)–(iii) one proves that the function

$$F(\beta,z) = \mathrm{Tr}[(H_{1,\beta} - z)^{-1} - (H_{2,\beta} - z)^{-1}],$$
$$z \in \mathbb{C}\setminus[0,\infty), \tag{3.28}$$

is differentiable with respect to β with derivative

$$\frac{\partial}{\partial\beta}F(\beta,z) = -\mathrm{Tr}\{(H_{1,\beta} - z)^{-1}[A_\beta^*B + B^*A_\beta](H_{1,\beta} - z)^{-1} - (H_{2,\beta} - z)^{-1}[A_\beta B^* + BA_\beta^*](H_{2,\beta} - z)^{-1}\}. \tag{3.29}$$

Using the commutation formulas[49]

$$(A_\beta^*A_\beta - z)^{-1}A_\beta^* \subseteq A_\beta^*(A_\beta A_\beta^* - z)^{-1},$$
$$(A_\beta A_\beta^* - z)^{-1}A_\beta \subseteq A_\beta(A_\beta^*A_\beta - z)^{-1}, \quad z \in \mathbb{C}\setminus[0,\infty), \tag{3.30}$$

and cyclicity of the trace, the two terms on the rhs of Eq. (3.29) cancel. Thus

$$\frac{\partial}{\partial\beta}F(\beta,z) = 0, \quad \beta \in \mathbb{R}, \quad z \in \mathbb{C}\setminus[0,\infty), \tag{3.31}$$

implying the desired result $F(\beta,z) = F(0,z)$. Conditions (iii) and (iv) enter in a rigorous derivation of Eq. (3.31).[48]

The result (3.27) yields the topological invariance of the regularized index $\Delta(z)$ in the concrete examples of Sec. IV (cf. also Ref. 52). Moreover, it proves the topological invariance of Δ and $\mathscr{A}$ whenever the limits $z \to 0$ and $z \to \infty$ of $\Delta(z)$ exist. In the case where A is Fredholm, the invariance of the Fredholm index $i(A)$ (and thus of Δ by Theorem 3.1), i.e.,

$$i(A + \beta B) = i(A), \quad \beta \in \mathbb{R}, \tag{3.32}$$

under relatively compact perturbations B with respect to A is a standard result.[47] Equation (3.27) works without assuming A to be Fredholm, but needs much stronger assumptions on the "smallness" of B than just relative compactness.

Another application of Eq. (3.27) concerns the invariance of Krein's spectral shift function. In fact, we get the following.

Theorem 3.4: Assume Hypothesis (vi) with A replaced by A_β and $(1 + |\cdot|)^{-1}[\xi_{12,\beta} - \xi_{12}] \in L^1(\mathbb{R})$ for all $\beta \in \mathbb{R}$. If conditions (ii)–(iv) of Theorem 3.3 hold, then

$$[\xi_{12,\beta}(\lambda_+) - \xi_{12}(\lambda_+)] + [\xi_{12,\beta}(\lambda_-) - \xi_{12}(\lambda_-)] = 0, \tag{3.33}$$

for all $\beta, \lambda \in \mathbb{R}$. In particular if $\xi_{12,\beta}$, $\beta \in \mathbb{R}$, and ξ_{12} are continuous at a point $\lambda \in \mathbb{R}$ then

$$\xi_{12,\beta}(\lambda) = \xi_{12}(\lambda), \quad \beta \in \mathbb{R}. \tag{3.34}$$

Proof: Equations (2.2) and (3.27) together with the Lebesgue dominated convergence theorem imply

$$0 = \int_{\mathbb{R}} d\lambda\,[\xi_{12,\beta}(\lambda) - \xi_{12}(\lambda)](\lambda - z)^{-2} = \frac{d}{dz}\int_{\mathbb{R}} d\lambda\,[\xi_{12,\beta}(\lambda) - \xi_{12}(\lambda)](\lambda - z)^{-1} \tag{3.35}$$

and hence

$$\int_{\mathbb{R}} d\lambda\,[\xi_{12,\beta}(\lambda) - \xi_{12}(\lambda)](\lambda - z)^{-1} = 0$$

by taking $|z| \to \infty$, $\mathrm{Im}\, z \neq 0$. Thus Eq. (3.33) results from standard properties of the Poisson kernel (cf., e.g., Ref. 45). □

In the first four examples of the next section, $\xi_{12,\beta}(\lambda)$ coincides with a multiple of the relative phase shift between H_1 and H_2 and the Fredholm determinants in Eq. (3.22) are expressed in terms of Wronski determinants. In these cases the topological invariance property of $\Delta(z)$ and $\xi_{12}(\lambda)$ can be established by simple and explicit calculations.

Finally, we note that the following family of operators in $\mathscr{H} \oplus \mathscr{H}$:

$$Q_m = \begin{pmatrix} m & A^* \\ A & -m \end{pmatrix},$$
$$H_m = Q_m^2 = \begin{pmatrix} H_1 + m^2 & 0 \\ 0 & H_2 + m^2 \end{pmatrix}, \quad m \in \mathbb{R}\setminus\{0\}, \tag{3.36}$$

can be treated analogously. In order to illustrate a simple application of the above results, we briefly discuss the invariance of the spectral asymmetry η_m (Refs. 7 and 9) under "small" perturbations. Under suitable conditions on H_m [cf., e.g., Eq. (3.17)], the (regularized) and spectral asymmetry can be defined by

$$\eta_m = \lim_{t \to 0_+} \eta_m(t), \tag{3.37}$$
$$\eta_m(t) = \mathrm{Tr}[Q_m H_m^{-1/2} e^{-tH_m}], \quad m \in \mathbb{R}\setminus\{0\}. \tag{3.38}$$

(This definition resembles the ones available in the literature, e.g., in Refs. 2, 8, 12, 53, and 54.) Since

$$\mathrm{Tr}[Q_m(H_m + z^2)^{-1}e^{-tH_m}] = m\,\mathrm{Tr}[(H_1 + m^2 + z^2)^{-1}e^{-t(H_1 + m^2)} - (H_2 + m^2 + z^2)^{-1}e^{-t(H_2 + m^2)}], \quad t > 0, \tag{3.39}$$

we can rewrite Eq. (3.38) in the form

$$\eta_m(t) = m\,\mathrm{Tr}[(H_1 + m^2)^{-1/2}e^{-t(H_1 + m^2)} - (H_2 + m^2)^{-1/2}e^{-t(H_2 + m^2)}] \tag{3.40}$$

and, using Eq. (2.4),

$$\eta_m(t) = m\int_0^\infty d\lambda\, \xi_{12}(\lambda)\, \frac{d}{d\lambda}\left[(\lambda+m^2)^{-1/2} e^{-t(\lambda+m^2)}\right]. \tag{3.41}$$

This implies

$$\eta_m = -\frac{m}{2}\int_0^\infty d\lambda\, \xi_{12}(\lambda)(\lambda+m^2)^{-3/2}. \tag{3.42}$$

Obviously, Eqs. (3.41) and (3.42) imply the invariance of η_m with respect to the substitution $A \to A_\beta = A + \beta B$ as a consequence of Theorem 3.4.

IV. SPECIFIC MODELS

We present a series of examples of explicit model calculations which illustrate the practical use of the abstract results of the foregoing section.

Example 4.1: Let $\mathscr{H} = L^2(\mathbb{R})$ and

$$A = \left(\frac{d}{dx} + \phi\right)\Big|_{H^{2,1}(\mathbb{R})}, \tag{4.1}$$

where ϕ fulfills the following requirements:

$$\begin{aligned} &\phi, \phi' \in L^\infty(\mathbb{R}) \text{ are real valued} \\ &\lim_{x\to\pm\infty} \phi(x) = \phi_\pm \in \mathbb{R}, \quad \phi_-^2 \leqslant \phi_+^2, \\ &\int_{\mathbb{R}} dx(1+|x|^2)|\phi'(x)| < \infty, \\ &\pm\int_0^{\pm\infty} dx(1+|x|^2)|\phi(x)-\phi_\pm| < \infty. \end{aligned} \tag{4.2}$$

In this case, H_1 and H_2 explicitly read

$$H_j = \left(-\frac{d^2}{dx^2} + \phi^2 + (-1)^j\phi'\right)\Big|_{H^{2,2}(\mathbb{R})}, \quad j=1,2. \tag{4.3}$$

Then

$$\Delta(z) = [\phi_+(\phi_+^2 - z)^{-1/2} - \phi_-(\phi_-^2 - z)^{-1/2}]/2, \quad z \in \mathbb{C}\setminus[0,\infty), \tag{4.4}$$

and hence

$$\Delta = [\operatorname{sgn}(\phi_+) - \operatorname{sgn}(\phi_-)]/2, \quad \mathscr{A} = 0, \tag{4.5}$$

$$\begin{aligned} \xi_{12}(\lambda) &= \pi^{-1}\{\theta(\lambda-\phi_+^2)\arctan[(\lambda-\phi_+^2)^{1/2}/\phi_+] \\ &\quad - \theta(\lambda-\phi_-^2)\arctan[(\lambda-\phi_-^2)^{1/2}/\phi_-]\} \\ &\quad + \theta(\lambda)[\operatorname{sgn}(\phi_-) - \operatorname{sgn}(\phi_+)]/2, \\ &\quad \phi_- \neq 0, \quad \phi_+ \neq 0, \\ \xi_{12}(\lambda) &= \pi^{-1}\theta(\lambda-\phi_+^2)\arctan[(\lambda-\phi_+^2)^{1/2}/\phi_+] \\ &\quad - \theta(\lambda)[\operatorname{sgn}(\phi_+)]/2, \\ &\quad \phi_- = 0, \quad \phi_+ \neq 0; \quad \lambda \in \mathbb{R}. \end{aligned} \tag{4.6}$$

[Here $\theta(x) = 1$ for $x \geqslant 0$ and $\theta(x) = 0$ for $x < 0$ and $\operatorname{sgn}(x) = \pm 1$ for $x \gtrless 0$ and $\operatorname{sgn}(0) = 0$.] Equations (4.4)–(4.6) clearly demonstrate the topological invariance of these quantities as discussed in Sec. III since they only depend on the asymptotic values $\phi_\pm$ of $\phi(x)$ and not on its local properties. In fact, replace $\phi(x)$ by $\phi(x) + \beta\psi(x)$, $\beta \in \mathbb{R}$, where

$$\begin{aligned} &\psi, \psi' \in L^\infty(\mathbb{R}) \text{ are real valued,} \\ &\psi(x), \psi'(x) = O(|x|^{-3-\epsilon}) \quad \text{for some } \epsilon > 0 \text{ as } |x| \to \infty. \end{aligned} \tag{4.7}$$

Then the perturbation B [cf. Eq. (3.26)] given by multiplication with ψ leaves the regularized index invariant since the hypotheses of Theorem 3.3 are satisfied.

Concerning zero-energy properties of H_j, $j = 1,2$, see Table I.

These zero-energy results easily follow from the fact that the equations

$$Af = 0, \quad A^*g = 0 \tag{4.8}$$

have the solutions

$$\begin{aligned} f(x) &= f(0)\exp\left(-\int_0^x dt\phi(t)\right) \\ &= O(e^{-\phi_\pm x}) \quad \text{as } x \to \pm\infty, \end{aligned} \tag{4.9}$$

$$g(x) = g(0)\exp\left(\int_0^x dt\phi(t)\right) = O(e^{\phi_\pm x}) \quad \text{as } x \to \pm\infty.$$

In order to derive Eq. (4.4), we introduce Jost solutions $f_{j\pm}(z,x)$ associated with H_j, $j = 1,2$,

TABLE I. Zero-energy properties of H_1 and H_2 in example 4.1.

	Zero-energy resonance of H_1	of H_2	Zero-energy bound state $\sigma_p(H_1)\cap\{0\}$	$\sigma_p(H_2)\cap\{0\}$	Δ	$i(A)$
$\phi_- < 0 < \phi_+$	no	no	$\{0\}$	ϕ	1	1
$\phi_+ < 0 < \phi_-$	no	no	ϕ	$\{0\}$	-1	-1
$\phi_+,\phi_- > 0$ or $\phi_+,\phi_- < 0$	no	no	ϕ	ϕ	0	0
$\phi_- = 0, \phi_+ \neq 0$	yes	no	ϕ	ϕ	$\frac{1}{2}\operatorname{sgn}(\phi_+)$	0
$\phi_- = \phi_+ = 0$	yes	yes	ϕ	ϕ	0	0

$$f_{j\pm}(z,x) = e^{\pm ik_\pm x} - \int_x^{\pm\infty} dx'\, k_\pm^{-1} \sin[k_\pm(x-x')] \times [\phi^2(x') - \phi_\pm^2 + (-1)^j \phi'(x')] f_{j\pm}(z,x'), \quad z\in\mathbb{C}, \quad j=1,2, \tag{4.10}$$

where

$$k_\pm(z) = (z-\phi_\pm^2)^{1/2}, \quad \operatorname{Im} k_\pm \geqslant 0. \tag{4.11}$$

The corresponding Fredholm integral equation reads

$$f_{1\pm}(z,x) = [T_{12}(z)]^{-1} f_{2\pm}(z,x) - \int_{\mathbb{R}} dx'\, g_2(z,x,x')[-2\phi'(x')] f_{1\pm}(z,x'), \quad z\in\mathbb{C}\setminus\sigma_p(H_2), \quad z\neq\phi_-^2, \tag{4.12}$$

where

$$g_2(z,x,x') = -[W(f_{2-}(z),f_{2+}(z))]^{-1} \times \begin{cases} f_{2+}(z,x) f_{2-}(z,x'), & x\geqslant x', \\ f_{2-}(z,x) f_{2+}(z,x'), & x\leqslant x', \end{cases} \quad z\in\mathbb{C}\setminus\sigma_p(H_2), \quad z\neq\phi_-^2, \tag{4.13}$$

$$g_2(z) = (H_2 - z)^{-1}, \quad z\in\rho(H_2),$$

and $T_{12}(z)$ denotes

$$T_{12}(z) = W(f_{2-}(z),f_{2+}(z))/W(f_{1-}(z),f_{1+}(z)), \quad z\in\mathbb{C}\setminus\sigma_p(H_2), \quad z\neq\phi_-^2. \tag{4.14}$$

Here

$$W(F,G)_x = F(x)G'(x) - F'(x)G(x) \tag{4.15}$$

denotes the Wronskian of F and G. (For more details on one-dimensional systems with nontrivial spatial asymptotics, cf. Ref. 23.) As can be seen, e.g., from Eq. (4.12), the relative interaction V_{12} reads

$$V_{12}(x) = -2\phi'(x). \tag{4.16}$$

Our first main step to derive Eq. (4.4) now consists of the observation that

$$\frac{W(f_{1-}(z),f_{1+}(z))}{W(f_{2-}(z),f_{2+}(z))} = \det[1 - 2|\phi'|^{1/2} \operatorname{sgn}(\phi') g_2(z) |\phi'|^{1/2}], \quad z\in\rho(H_2), \quad z\neq\phi_-^2, \tag{4.17}$$

such that (cf. Lemma 2.4)

$$\operatorname{Tr}[(H_1-z)^{-1} - (H_2-z)^{-1}] = -\frac{d}{dz} \ln \frac{W(f_{1-}(z),f_{1+}(z))}{W(f_{2-}(z),f_{2+}(z))}, \quad z\in C\setminus[0,\infty). \tag{4.18}$$

Equality (4.17) can be proved along the lines of Ref. 33 using Eqs. (4.10) and (4.12) (cf. Ref. 23).

Next, we note that Eq. (3.15) also holds for distributional-type (e.g., Jost) solutions of H_1 and H_2. In fact, assume that $f_1(z,x)$, $z\neq 0$, is normalized according to Eq. (4.10), i.e.,

$$f_{1\pm}(z,x) = e^{\pm ik_\pm x} + o(1) \quad \text{as } x\to\pm\infty,$$

then $(Af_{1\pm})(z,x)$ asymptotically fulfills

$$(Af_{1\pm})(z,x) = (\pm ik_\pm + \phi_\pm) e^{\pm ik_\pm x} + o(1) \quad \text{as } x\to\pm\infty.$$

Thus

$$\begin{cases} f_{1\pm}(z,x) \\ f_{2\pm}(z,x) = (\pm ik_\pm + \phi_\pm)^{-1} (Af_{1\pm})(z,x), \quad z\neq 0, \end{cases} \tag{4.19}$$

are correctly normalized Jost solutions for H_1 and H_2. Equation (4.17) thus becomes

$$\det[1 - 2|\phi'|^{1/2} \operatorname{sgn}(\phi') g_2(z) |\phi'|^{1/2}] = (-ik_- + \phi_-)(ik_+ + \phi_+) W(f_{1-}(z),f_{1+}(z)) \times [W((Af_{1-})(z),(Af_{1+})(z))]^{-1}, \quad z\in\mathbb{C}\setminus[0,\infty). \tag{4.20}$$

Finally, a straightforward computation yields

$$W((Af)(z),(Ag)(z)) = zW(f(z),g(z)), \quad z\in\mathbb{C}, \tag{4.21}$$

where f,g are distributional solutions of

$$(A^*A\psi(z))(x) = -\psi''(z,x) + [\phi^2(x) - \phi'(x)]\psi(z,x) = z\psi(z,x), \quad z\in\mathbb{C}. \tag{4.22}$$

Consequently, Eq. (4.20) becomes

$$\det[1 - 2|\phi'|^{1/2} \operatorname{sgn}(\phi') g_2(z) |\phi'|^{1/2}] = (-ik_- + \phi_-)(ik_+ + \phi_+)/z, \quad z\in\mathbb{C}\setminus[0,\infty) \tag{4.23}$$

and Eq. (4.4) follows from Eqs. (3.23) and (4.23).

The result (4.4) was first derived by Callias,[17] and since then by numerous authors.[2,10,11,18,21,22,25,55] While our derivation is close to that in Ref. 22, it seems to be the shortest one since the trick based on Eq. (4.21) explicitly exploits supersymmetry and avoids the use of an additional comparison Hamiltonian in the approach of Ref. 22.

Next, we discuss an example on the half-line $(0,\infty)$.

Example 4.2: Let $\mathscr{H} = L^2(0,\infty)$ and

$$A = \left(\frac{d}{dr} + \tilde\phi(r)\right)\bigg|_{H_0^{2,1}(0,\infty)}, \tag{4.24}$$

where $\tilde\phi$ fulfills the following requirements:

$$\begin{gathered} \tilde\phi, \tilde\phi' \in L^\infty(0,\infty) \text{ are real valued}, \\ \lim_{r\to\infty} \tilde\phi(r) = \tilde\phi_+ \in\mathbb{R}, \quad \lim_{r\to 0+} \tilde\phi(r) = \tilde\phi_0\in\mathbb{R}, \\ \int_0^\infty dr\, r(1+r)|\tilde\phi'(r)| < \infty, \\ \int_0^\infty dr\, r(1+r)|\tilde\phi(r) - \tilde\phi_+| < \infty. \end{gathered} \tag{4.25}$$

In this case, H_1 and H_2 read

$$H_1 = \left(-\frac{d^2}{dr^2} + \tilde\phi^2 - \tilde\phi'\right)_F, \tag{4.26}$$

where F denotes the Friedrichs extension of the corresponding operator restricted to $C_0^\infty(0,\infty)$ and

$$H_2 = -\frac{d^2}{dr^2} + \tilde{\phi}^2 + \tilde{\phi}',$$

$$\mathscr{D}(H_2) = \{g \in L^2(0,\infty) | g,g' \in AC_{\text{loc}}(0,\infty); \tag{4.27}$$
$$g'(0_+) - \tilde{\phi}_0 g(0_+) = 0; \quad g'' \in L^2(0,\infty)\},$$

With $AC_{\text{loc}}(a,b)$ the set of locally absolutely continuous functions on (a,b). Then

$$\Delta(z) = (z/2)(\tilde{\phi}_+^2 - z)^{-1/2}[\tilde{\phi}_+ + (\tilde{\phi}_+^2 - z)^{1/2}]^{-1},$$
$$z \in \mathbb{C} \setminus [0,\infty), \tag{4.28}$$

and hence

$$\Delta = \begin{cases} -[1 - \text{sgn}(\tilde{\phi}_+)]/2, & \tilde{\phi}_+ \neq 0, \\ -\frac{1}{2}, & \tilde{\phi}_+ = 0, \end{cases} \quad \mathscr{A} = \tfrac{1}{2}, \tag{4.29}$$

$$\xi_{12}(\lambda) = \begin{cases} \pi^{-1}\theta(\lambda - \tilde{\phi}_+^2) \arctan[(\lambda - \tilde{\phi}_+^2)^{1/2}/\tilde{\phi}_+] + \theta(\lambda)\theta(-\tilde{\phi}_+), & \tilde{\phi}_+ \neq 0, \\ \theta(\lambda)/2, \quad \tilde{\phi}_+ = 0; \quad \lambda \in \mathbb{R}. \end{cases} \tag{4.30}$$

Again, Eqs. (4.28)–(4.30) exhibit the topological invariance of all these quantities since only $\tilde{\phi}_+$ enters. [The arguments in connection with Eq. (4.7) can easily be extended to the present situation.] Concerning zero-energy properties, see Table II.

In order to derive Eq. (4.28), we introduce the Jost solutions

$$f_{j\pm}(z,r) = e^{\pm ik_+ r} - \int_r^\infty dr'\, k_+^{-1} \sin[k_+(r-r')] \times [\tilde{\phi}^2(r') - \tilde{\phi}_+^2 + (-1)^j \tilde{\phi}'(r')] f_{j\pm}(z,r'),$$
$$z \in \mathbb{C}, \quad j = 1,2, \tag{4.31}$$

where

$$k_+(z) = (z - \tilde{\phi}_+^2)^{1/2}, \quad \text{Im}\, k_+ \geqslant 0, \tag{4.32}$$

and the regular solutions

$$\psi_1(z,r) = k_+^{-1} \sin k_+ r + \int_0^r dr'\, k_+^{-1} \sin[k_+(r-r')] \times [\tilde{\phi}^2(r') - \tilde{\phi}_+^2 - \tilde{\phi}'(r')]\, \psi_1(z,r'),$$

$$\psi_2(z,r) = \cos k_+ r + \tilde{\phi}_0 k_+^{-1} \sin k_+ r + \int_0^r dr'\, k_+^{-1} \sin[k_+(r-r')] \times [\tilde{\phi}^2(r') - \tilde{\phi}_+^2 + \tilde{\phi}'(r')]\, \psi_2(z,r'), \quad z \in \mathbb{C}. \tag{4.33}$$

Using again Eq. (3.15), we assume that $f_{1\pm}(z,r)$, $z \neq 0$ is normalized according to Eq. (4.31), i.e.,

$$f_{1\pm}(z,r) = e^{\pm ik_+ r} + o(1) \quad \text{as } r \to \infty.$$

Then $(Af_{1\pm})(z,r)$ fulfills

$$(Af_{1\pm})(z,r) = (\pm ik_+ + \tilde{\phi}_+) e^{\pm ik_+ r} + o(1) \quad \text{as } r \to \infty$$

such that the Jost functions

$$\begin{cases} f_{1\pm}(z,r), \\ f_{2\pm}(z,r) = (\pm ik_+ + \tilde{\phi}_+)^{-1}(Af_{1\pm})(z,r), \quad z \neq 0, \end{cases} \tag{4.34}$$

are correctly normalized. Similarly, we assume that $\psi_1(z,r)$, $z \neq 0$ fulfills

$$\psi_1(z,r) = r + o(r) \quad \text{as } r \to 0_+.$$

Then

$$(A\psi_1)(z,r) = 1 + \tilde{\phi}_0 r + o(r) \quad \text{as } r \to 0_+$$

and thus

$$\begin{cases} \psi_1(z,r), \\ \psi_2(z,r) = (A\psi_1)(z,r), \quad z \neq 0, \end{cases} \tag{4.35}$$

are correctly normalized regular solutions of H_1 and H_2. The rest is now identical to the treatment of example 4.1. First of all, one derives, as in Eq. (4.18) (cf., e.g., Ref. 30)

$$\text{Tr}[(H_1 - z)^{-1} - (H_2 - z)^{-1}] = \frac{d}{dz} \ln \frac{W(\psi_2(z), f_{2+}(z))}{W(\psi_1(z), f_{1+}(z))}, \quad z \in \mathbb{C} \setminus [0,\infty). \tag{4.36}$$

Then one calculates, as in Eq. (4.21), that

$$W((A\psi_1)(z), (Af_{1+})(z)) = zW(\psi_1(z), f_{1+}(z)), \quad z \in \mathbb{C}. \tag{4.37}$$

We now consider a generalization of this example which allows us to discuss n-dimensional spherically symmetric systems (cf., e.g., Refs. 2 and 13).

Example 4.3: Let $\mathscr{H} = L^2(0,\infty)$ and

$$A = \overline{\left(\frac{d}{dr} + \phi\right)\Big|_{C_0^\infty(0,\infty)}}, \tag{4.38}$$

where ϕ fulfills the following requirements:

$$\phi(r) = \phi_0 r^{-1} + \tilde{\phi}(r), \quad \phi_0 \leqslant -\tfrac{1}{2}, \quad r > 0,$$
$$\tilde{\phi}, \tilde{\phi}' \in L^\infty(0,\infty) \text{ are real valued},$$

$$\lim_{r \to \infty} \tilde{\phi}(r) = \tilde{\phi}_+ \in \mathbb{R},$$

$$\int_0^\infty dr\, W_{\phi_0}(r)(|\tilde{\phi}'(r)| + r^{-1}|\tilde{\phi}(r) - \tilde{\phi}_+|) < \infty, \tag{4.39}$$

$$\int_0^\infty dr\, W_{\phi_0}(r)|\tilde{\phi}(r) - \tilde{\phi}_+| < \infty,$$

and the weight function W_{ϕ_0} is defined by

TABLE II. Zero-energy properties of H_1 and H_2 in example 4.2.

	Zero-energy resonance of H_1	Zero-energy resonance of H_2	Zero-energy bound state $\sigma_p(H_1) \cap \{0\}$	Zero-energy bound state $\sigma_p(H_2) \cap \{0\}$	Δ	$i(A)$
$\tilde{\phi}_+ > 0$	no	no	ϕ	ϕ	0	0
$\tilde{\phi}_+ < 0$	no	no	ϕ	$\{0\}$	-1	-1
$\tilde{\phi}_+ = 0$	no	yes	ϕ	ϕ	$-\frac{1}{2}$	0

$$W_{\phi_0}(r)=\begin{cases} r(1+r) & \text{if } \phi_0<-\frac{1}{2}, \\ r(1+|\ln r|^2), \quad 0<r\leqslant\frac{1}{2} & \text{if } \phi_0=-\frac{1}{2}, \\ r(1+r), \quad r\geqslant\frac{1}{2}. \end{cases} \tag{4.40}$$

Now H_1 and H_2 are given by

$$H_j=\Big(-\frac{d^2}{dr^2}+\phi^2+(-1)^j\phi'\Big)_{\rm F}, \quad j=1,2. \tag{4.41}$$

Explicitly, we have

$$\phi^2(r)\mp\phi'(r)=(\phi_0^2\pm\phi_0)r^{-2}+2\phi_0\tilde\phi_+r^{-1}+\tilde\phi_+^2+\tilde\phi^2(r)-\tilde\phi_+^2\mp\tilde\phi'(r)+2\phi_0[\tilde\phi(r)-\tilde\phi_+]r^{-1}, \quad r>0. \tag{4.42}$$

Then

$$\Delta(z)=(z/2)(\tilde\phi_+^2-z)^{-1/2}[\tilde\phi_+-(\tilde\phi_+^2-z)^{1/2}]^{-1}, \quad z\in\mathbb{C}\setminus[0,\infty) \tag{4.43}$$

and hence

$$\Delta=\begin{cases}[1+\operatorname{sgn}(\tilde\phi_+)]/2, & \tilde\phi_+\neq 0, \\ \frac{1}{2}, & \tilde\phi_+=0,\end{cases} \quad \mathscr{A}=-\tfrac{1}{2}, \tag{4.44}$$

$$\xi_{12}(\lambda)=\begin{cases}\pi^{-1}\theta(\lambda-\tilde\phi_+^2)\arctan[(\lambda-\tilde\phi_+^2)^{1/2}/\tilde\phi_+]-\theta(\lambda)\theta(\tilde\phi_+), & \tilde\phi_+\neq 0, \\ -\theta(\lambda)/2, & \tilde\phi_+=0; \quad \lambda\in\mathbb{R}. \end{cases} \tag{4.45}$$

The topological invariance in Eqs. (4.43)–(4.45) is obvious. (See Table III.) If $\tilde\phi_+=0$, the result $\Delta=\frac{1}{2}$ is not due to a zero-energy (threshold) resonance, but due to the long-range nature of the relative interaction $V_{12}(r)=2\phi_2r^{-2}+o(r^{-2})$ as $r\to\infty$. Since Eq. (4.43) is independent of ϕ_0, this result holds in any dimension $\geqslant 2$ and for any value of the angular momentum.

In order to derive Eq. (4.43), one could follow the strategy of example 4.2 step by step since formula (4.36) remains valid in the present case for suitably normalized Jost and regular solutions (although we are dealing with a long-range problem!). To shorten the presentation, we will use instead a different approach based on the topological invariance property of $\Delta(z)$ and $\xi_{12}(\lambda)$ (this approach obviously also works in example 4.2). Indeed, because of Theorem 3.3, it suffices to choose $\tilde\phi(r)=\tilde\phi_+$, $r\geqslant 0$ in example 4.3. Then

$$H_j=\Big(-\frac{d^2}{dr^2}+[\phi_0^2-(-1)^j\phi_0]r^{-2}+2\phi_0\tilde\phi_+r^{-1}+\tilde\phi_+^2\Big)_{\rm F}, \quad j=1,2 \tag{4.46}$$

[cf. Eq. (4.42)] and hence[56]

$$S_j(\lambda)=\frac{\Gamma(2^{-1}+2^{-1}(-1)^j-\phi_0+i(\phi_0\tilde\phi_+/k_+))}{\Gamma(2^{-1}+2^{-1}(-1)^j-\phi_0-i(\phi_0\tilde\phi_+/k_+))}\times e^{i\pi[2^{-1}-(-1)^j2^{-1}+\phi_0]}, \quad \lambda>\tilde\phi_+^2, \quad j=1,2 \tag{4.47}$$

[$k_+(\lambda)$ defined in Eq. (4.32)] implying

$$S_{12}(\lambda)=S_1(\lambda)S_2(\lambda)^{-1}=(\tilde\phi_+-ik_+)/(\tilde\phi_++ik_+), \quad \lambda>\tilde\phi_+^2. \tag{4.48}$$

Equation (4.48) proves Eq. (4.45). Now Eq. (4.43) follows by explicit integration (Ref. 57, p. 556) in Eq. (3.23).

The result (4.43), in the special case $\tilde\phi(r)\equiv 0$, has been discussed in Ref. 21 by different methods.

Next, we briefly discuss nonlocal interactions.

Example 4.4: Let $\mathscr{H}=L^2(0,\infty)$ and

$$A=\frac{d}{dr}\Big|_{H_0^{2,1}(0,\infty)}+B, \tag{4.49}$$

where

$$B, A^*B, AB^*\in\mathscr{B}_1(L^2(0,\infty)). \tag{4.50}$$

In this case the assumptions of Theorem 3.3 are trivially fulfilled, and hence Eqs. (4.28)–(4.30), in the special case $\phi(r)\equiv 0$, hold. In particular

$$\Delta(z)=\Delta=-\tfrac{1}{2}, \quad z\in\mathbb{C}\setminus[0,\infty), \quad \mathscr{A}=\tfrac{1}{2}. \tag{4.51}$$

In order to illustrate the possible complexity of zero-energy properties of H_1 and H_2 in spite of the simplicity of Eq. (4.51), it suffices to treat the following rank 2 example:

$$B=\alpha(f,\cdot)f+\beta(g,\cdot)g, \quad \alpha,\beta\in\mathbb{R},$$
$$f,g\in C_0^1(0,\infty), \quad f\geqslant 0, g\geqslant 0, \quad f\neq g. \tag{4.52}$$

By straightforward calculations, one obtains the information contained in Table IV. Here the following case distinction has been used:

TABLE III. Zero-energy properties of H_1 and H_2 in example 4.3.

	Zero-energy resonance of H_1	Zero-energy resonance of H_2	Zero-energy bound state $\sigma_p(H_1)\cap\{0\}$	Zero-energy bound state $\sigma_p(H_2)\cap\{0\}$	Δ	$i(A)$
$\tilde\phi_+>0$	no	no	$\{0\}$	ϕ	1	1
$\tilde\phi_+<0$	no	no	ϕ	ϕ	0	0
$\tilde\phi_+=0$	no	no	ϕ	ϕ	$\frac{1}{2}$	0

TABLE IV. Zero-energy properties of H_1 and H_2 in example 4.4.

	Zero-energy resonance of H_1	Zero-energy resonance of H_2	Zero-energy bound state $\sigma_p(H_1)\cap\{0\}$	Zero-energy bound state $\sigma_p(H_2)\cap\{0\}$	Δ	$i(A)$
Case I	no	yes	ϕ	ϕ	$-\frac{1}{2}$	0
Case II	yes	no	ϕ	$\{0\}$	$-\frac{1}{2}$	-1
Case III	no	yes	$\{0\}$	$\{0\}$	$-\frac{1}{2}$	0

case I, $\Psi(\alpha,\beta)\neq 0$;
case II, $\Psi(\alpha,\beta)=0$,
$\alpha\neq 2G(\infty)\{F(\infty)[(f,G)-(g,F)]\}^{-1}$;
case III, $\Psi(\alpha,\beta)=0$,
$\alpha=2G(\infty)\{F(\infty)(f,G)-(g,F)]\}^{-1}$;

where

$$F(x)=\int_0^x dx'\,f(x')\,,\quad G(x)=\int_0^x dx'\,g(x')\,,$$
$$\Psi(\alpha,\beta)=[1+\alpha(f,F)][1+\beta(g,G)]-\alpha\beta(f,G)(g,F)\,. \tag{4.53}$$

Finally, we consider in detail the following two-dimensional magnetic field problem.

Example 4.5: Let $\mathscr{H}=L^2(\mathbb{R}^2)$ and

$$A=\overline{[(-i\partial_1-a_1)+i(i\partial_2+a_2)]|_{C_0^\infty(\mathbb{R}^2)}}\,, \tag{4.54}$$

where

$$a=(\partial_2\phi,-\partial_1\phi)\,,\quad \partial_j\equiv\frac{\partial}{\partial x_j}\,,\quad j=1,2\,, \tag{4.55}$$

and ϕ fulfills the following requirements:

$$\begin{aligned}&\phi\in C^2(\mathbb{R}^2)\ \text{is real valued}\,,\\ &\phi(x)=-F\ln|x|+C+O(|x|)^{-\epsilon}\,,\\ &(\nabla\phi)(x)=-F|x|^{-2}x+O(|x|^{-1-\epsilon})\,,\\ &\qquad C,F\in\mathbb{R}\,,\quad \epsilon>0\ \text{as}\ |x|\to\infty\,,\\ &(\Delta\phi)^{1+\delta},\ (1+|\cdot|^{\delta})(\Delta\phi)\in L^1(\mathbb{R}^2)\quad\text{for some }\delta>0\,.\end{aligned} \tag{4.56}$$

Then

$$H_j=[(-i\nabla-a)^2-(-1)^j b\,]|_{H^{2,2}(\mathbb{R}^2)}\,,\quad j=1,2\,, \tag{4.57}$$

where

$$b(x)=(\partial_1a_2-\partial_2a_1)(x)=-(\Delta\phi)(x)\,. \tag{4.58}$$

Introducing the magnetic flux F by

$$F=(2\pi)^{-1}\int_{\mathbb{R}^2}d^2x\,b(x) \tag{4.59}$$

we obtain

$$\Delta(z)=\Delta=-F\,,\quad z\in\mathbb{C}\setminus[0,\infty)\,,\quad \mathscr{A}=F\,, \tag{4.60}$$
$$\xi_{12}(\lambda)=F\theta(\lambda)\,,\quad \lambda\in\mathbb{R}\,. \tag{4.61}$$

Moreover, we have

$$\begin{aligned}&i(A)\operatorname{sgn}(F)\\ &\quad=\theta(-F)\dim\operatorname{Ker}(A)-\theta(F)\dim\operatorname{Ker}(A^*)\\ &\quad=\begin{cases}-N\ \text{if}\ |F|=N+\epsilon, & 0<\epsilon<1\,,\\ -(N-1)\ \text{if}\ |F|=N, & N\in\mathbb{N}\,.\end{cases}\end{aligned} \tag{4.62}$$

Since Eq. (4.62) has been derived in Ref. 58 (cf. also Refs. 8, 24, and 59–62), we concentrate on Eqs. (4.60) and (4.61). For this purpose we first study a special example (treated in Ref. 63). Let

$$\phi(R,r)=\begin{cases}-(Fr^2/2R^2)\,, & r\leqslant R\,,\\ -(F/2)[1+\ln(r^2/R^2)]\,, & r\geqslant R\,,\quad R>0\,,\end{cases} \tag{4.63}$$

and denote the corresponding Hamiltonian in (4.57) by $H_j(R)$, $j=1,2$. Next, define U_ϵ, $\epsilon\geqslant 0$, to be the unitary group of dilations in $L^2(\mathbb{R}^2)$, viz.,

$$(U_\epsilon g)(x)=\epsilon^{-1}g(x/\epsilon)\,,\quad \epsilon>0\,,\quad g\in L^2(\mathbb{R}^2)\,. \tag{4.64}$$

Then a simple calculation yields

$$U_\epsilon H_j(R)U_\epsilon^{-1}=\epsilon^2H_j(\epsilon R)\,,\quad \epsilon,R>0\,,\quad j=1,2\,. \tag{4.65}$$

If we denote by $S_{12}(R)$, the scattering operator in $L^2(\mathbb{R}^2)$ associated with the pair $(H_1(R),H_2(R))$, then $S_{12}(R)$ is decomposable with respect to the spectral representation of $H_2(R)P_{\rm ac}(H_2(R))$ [$P_{\rm ac}(\cdot)$ is the projection onto the absolutely continuous spectral subspace]. Let $S_{12}(\lambda,R)$ in $L^2(S^1)$ denote the fibers of $S_{12}(R)$, then Eq. (4.65) implies

$$\begin{aligned}&S_{12}(\lambda,R)=S_{12}(\epsilon^2\lambda,R/\epsilon)\,,\\ &\xi_{12}(\lambda,R)=\xi_{12}(\epsilon^2\lambda,R/\epsilon)\,,\quad \lambda>0\,.\end{aligned} \tag{4.66}$$

Applying now Theorem 3.4, we infer that $\xi_{12}(\lambda)$ cannot depend on $R>0$ as long as F is kept fixed in Eq. (4.63). Thus Eq. (4.63) implies $\xi_{12}(\lambda)=\xi_{12}(\epsilon^2\lambda)$, $\lambda>0$, which in turn implies that ξ_{12} is energy independent.

We will give two methods of computing this constant value of ξ_{12}, the first using heat kernels, the second, resolvents.

Method 1: By Eq. (2.4)

$$\begin{aligned}\operatorname{Tr}(e^{-tH_1}-e^{-tH_2})&=-t\int_0^\infty e^{-t\lambda}\xi_{12}(\lambda)d\lambda\\ &=-\xi_{12}\,.\end{aligned} \tag{4.67}$$

Let $H_0=-\Delta_{H^{2,2}(\mathbb{R}^2)}$. We will prove that

$$\lim_{t\downarrow 0}[\operatorname{Tr}(e^{-tH_1}-e^{-tH_0})]=-\tfrac{1}{2}F\,. \tag{4.68}$$

This, with the analogous calculation for H_2, yields

$$\xi_{12}=F\,. \tag{4.69}$$

To prove (4.68), we expand e^{-tH_1} perturbatively (Du Hamel expansion) and obtain

$$\operatorname{Tr}(e^{-tH_1}-e^{-tH_0})=\alpha+\beta\,, \tag{4.70}$$

$$\begin{aligned}&\alpha=-t\operatorname{Tr}(e^{-tH_0}b)\,,\\ &\beta=\int_0^t s\operatorname{Tr}(e^{-sH_0}be^{-(t-s)H_1}b)ds\,.\end{aligned} \tag{4.71}$$

Since $(e^{-tH_0})(x,x)=(4\pi t)^{-1}$, we have

$$\alpha=-t(4\pi t)^{-1}\int_{\mathbb{R}^2}b(x)d^2x=-\tfrac{1}{2}F \tag{4.72}$$

so we need only show that

$$\lim_{t\downarrow 0}\beta=0\,. \tag{4.73}$$

By the Schwarz inequality

$$\operatorname{Tr}(e^{-sH_0}be^{-(t-s)H_1}b)\leqslant\gamma^{1/2}\delta^{1/2}\,, \tag{4.74}$$

$$\begin{aligned}&\gamma=\operatorname{Tr}(e^{-2sH_0}b^2)=(8\pi s)^{-1}\int_{\mathbb{R}^2}b^2\,d^2x\,,\\ &\delta=\operatorname{Tr}(e^{-2(t-s)H_1}b^2)\leqslant e^{2(t-s)\|b\|_\infty}\operatorname{Tr}(e^{-2(t-s)H_0}b^2)\\ &\quad=e^{2(t-s)\|b\|_\infty}(8\pi(t-s))^{-1}\int_{\mathbb{R}^2}b^2\,d^2x\,,\end{aligned} \tag{4.75}$$

where we have used the diamagnetic inequalities (see Ref. 59 and references therein). Thus

$$\beta \leqslant \left(\int_{\mathbb{R}^2} b^2 \, d^2x \right) e^{+2t\|b\|_\infty} (8\pi)^{-1} \int_0^t s^{1/2} (t-s)^{-1/2} \, ds \tag{4.76}$$

goes to zero as $t \downarrow 0$.

Method 2: This is essentially the Laplace transform of method 1. Since $\Delta(z) = \Delta$ is independent of z, we can calculate it in the $z \to \infty$ limit. To do this, we infer from the proof of Lemma 2.7 that

$$\begin{aligned} \Delta(z) &= z \operatorname{Tr} [(H_2 - z)^{-1} V_{12} (H_2 - z)^{-1}] \\ &\quad - z \operatorname{Tr}\{[1 + u_{12}(H_2 - z)^{-1} v_{12}]^{-1} u_{12} (H_2 - z)^{-1} \\ &\quad \times v_{12} u_{12} (H_2 - z)^{-2} v_{12}\}, \quad z \in \mathbb{C} \setminus [0, \infty). \end{aligned} \tag{4.77}$$

Next, we employ the resolvent equation giving

$$\begin{aligned} (H_2 - z)^{-1} V_{12} (H_2 - z)^{-1} &= (H_0 - z)^{-1} V_{12} (H_0 - z)^{-1} \\ &\quad - (H_2 - z)^{-1} V_2 (H_0 - z)^{-1} V_{12} (H_0 - z)^{-1} \\ &\quad - (H_0 - z)^{-1} V_{12} (H_0 - z)^{-1} V_2 (H_2 - z)^{-1} \\ &\quad + (H_2 - z)^{-1} V_2 (H_0 - z)^{-1} \\ &\quad \times V_{12} (H_0 - z)^{-1} V_2 (H_2 - z)^{-1}, \quad z \in \mathbb{C} \setminus [0, \infty), \end{aligned} \tag{4.78}$$

where

$$\begin{aligned} &H_0 = -\Delta|_{H^{2,2}(\mathbb{R}^2)}, \quad V_{12}(x) = 2b(x), \\ &V_2 = 2ia\nabla + i(\nabla a) + a^2 - b. \end{aligned} \tag{4.79}$$

Then estimates of the type[35]

$$\begin{aligned} &\|w(H_0 - z)^{-1}\|_2^2 \leqslant C \|w\|_2^2 |z|^{-1}, \\ &\operatorname{Im} z^{1/2} > 0, \quad w \in L^2(\mathbb{R}^2), \end{aligned} \tag{4.80}$$

and, e.g.,

$$\begin{aligned} &\|(H_2 - z)^{-1} V_2 (H_0 - z)^{-1} V_{12} (H_0 - z)^{-1}\|_1 \\ &\quad \leqslant \|(H_2 - z)^{-1/2}\| \, \|(H_2 - z)^{-1/2} V_2\| \\ &\qquad \times \|(H_0 - z)^{-1} u_{12}\|_2 \, \|v_{12} (H_0 - z)^{-1}\|_2 \\ &\quad \leqslant C |z|^{-1} |\operatorname{Im} z|^{-1/2}, \quad |\operatorname{Re} z| \leqslant C_1 |\operatorname{Im} z| \end{aligned} \tag{4.81}$$

imply [cf. Eq. (2.21)] that

$$\begin{aligned} &\lim_{\substack{|z| \to \infty \\ |\operatorname{Re} z| \leqslant C_1 |\operatorname{Im} z|}} z \operatorname{Tr}[(H_2 - z)^{-1} V_{12} (H_2 - z)^{-1}] \\ &\quad = \lim_{\substack{|z| \to \infty \\ |\operatorname{Re} z| \leqslant C_1 |\operatorname{Im} z|}} z \operatorname{Tr}[(H_0 - z)^{-1} V_{12} (H_0 - z)^{-1}] \\ &\quad = -(2\pi)^{-1} \int_{\mathbb{R}^2} d^2x \, b(x) = -F. \end{aligned} \tag{4.82}$$

Similarly, we get

$$\begin{aligned} &\|[1 + u_{12}(H_2 - z)^{-1} v_{12}]^{-1} u_1 (H_2 - z)^{-1} \\ &\quad \times v_{12} u_{12} (H_{12} - z)^{-2} v_{12}\|_1 \\ &\quad \leqslant C \|u_{12}(H_2 - z)^{-1} v_{12}\| \, \|(H_0 - z)(H_2 - z)^{-1}\|^2 \\ &\qquad \times \|u_{12}(H_0 - z)^{-1}\|_2 \, \|(H_0 - z)^{-1} v_{12}\|_2 \\ &\quad \leqslant C' |z|^{-1} \|u_{12}(H_2 - z)^{-1} v_{12}\| = o(|z|^{-1}) \\ &\qquad \text{as } |z| \to \infty, \quad |\operatorname{Re} z| \leqslant C_1 |\operatorname{Im} z|. \end{aligned} \tag{4.83}$$

Inequality (4.83) follows from the fact that

$$\|u_{12}(H_0 - z)^{-1} v_{12}\|_2 \xrightarrow[|\operatorname{Re} z| \leqslant C_1 |\operatorname{Im} z|]{|z| \to \infty} 0, \tag{4.84}$$

which in turn is a consequence of the Hankel function estimate

$$\begin{aligned} &|H_0^{(1)}(\sqrt{z}|x - y|)|^2 \\ &\quad \leqslant d_1 + d_2 (\ln|x - y|)^2, \quad \operatorname{Im} \sqrt{z} \geqslant \mu > 0, \end{aligned} \tag{4.85}$$

and dominated convergence. Relation (4.80) then shows

$$\|u_{12}(H_2 - z)^{-1} v_{12}\|_2 \xrightarrow[|\operatorname{Re} z| \leqslant C_1 |\operatorname{Im} z|]{|z| \to \infty} 0, \tag{4.86}$$

where we have again used the resolvent equation and Eq. (4.84). Thus we have shown that $\Delta(\infty) = -\mathscr{A} = -F$, which completes the derivation of Eq. (4.60).

The result of Aharonov–Casher[58] implies that $\dim \operatorname{Ker}(H_1) - \dim \operatorname{Ker}(H_2)$ differs from Δ by at most 1. It would be nice to know why this is true.

We remark that the result (4.60) has been obtained in Ref. 24 by using certain approximations in a path integral approach. The above treatment seems to be the first rigorous and nonperturbative one.

To complete this discussion, we still mention that the (regularized) spectral asymmetry, $\eta_m(t)$, associated with this magnetic field example (4.5) after replacing H_j by $H_j + m^2$ [Q by Q_m, cf. Eq. (3.36)] can be calculated using the result (4.61) and Eq. (3.41). One easily gets

$$\eta_m(t) = \operatorname{sgn}(m) F e^{-tm^2}, \quad m \in \mathbb{R} \setminus \{0\}, \quad t > 0, \tag{4.87}$$

containing in the limit $t \to 0_+$ the known result for η_m (cf., e.g., Ref. 2).

ACKNOWLEDGMENTS

We gratefully acknowledge stimulating discussions with Professor R. Blankenbecler. F. G. is indebted to Professor K. Chadan for the warm hospitality extended to him at LPTHE–Orsay, and to E. Stone and D. Wales for hospitality at Caltech. D. B. is an Onderzoeksleider of the National Fonds voor Wetenschappelijk Onderzoek, Belgium.

The work of F. G. and H. G. is partially supported by Fonds zur Förderung der Wissenschaftlichen Forschung in Österreich, Project No. P5588. The research of B. S. is partially supported by the USNSF under Grant No. DMS-84-16049.

[1]R. Jackiw and C. Rebbi, Phys. Rev. D **13**, 3398 (1976).

[2]A. J. Niemi and G. W. Semenoff, Phys. Rep. **135**, 100 (1986).

[3]R. Jackiw and J. R. Schrieffer, Nucl. Phys. B **190** [FS3], 253 (1981).

[4]Proceedings of the International Conference on Physics and Chemistry of Conducting Polymers, Les Arcs, France, J. Phys. (Coll.) C **3**, 44 (1983).

[5]W. P. Su, in *Handbook of Conducting Polymers*, edited by T. Skotheim (Dekker, New York, 1984).

[6]W. P. Su, J. R. Schrieffer, and A. J. Heeger, Phys. Rev. Lett. **42**, 1698 (1979); Phys. Rev. B **22**, 2099 (1980).

[7]M. Atiyah, V. Patodi, and I. Singer, Proc. Cambridge Philos. Soc. **77**, 42 (1975); **78**, 405 (1975); **79**, 71 (1976).

[8]R. Blankenbecler and D. Boyanovsky, Phys. Rev. D **31**, 2089 (1985).

[9]T. Eguchi, P. B. Gilkey, and A. J. Hanson, Phys. Rep. **66**, 213 (1980).

[10]M. Hirayama, Phys. Rev. D **33**, 1079 (1986).

[11]J. Lott, Commun. Math. Phys. **93**, 533 (1984).

[12]Z. Q. Ma, H. T. Nieh, and R. K. Su, Phys. Rev. D **32**, 3268 (1985).

[13]T. D. Imbo and U. P. Sukhatme, Phys. Rev. Lett. **54**, 2184 (1985).

[14]Proceedings of the Conference on Supersymmetry in Physics, Los Alamos

edited by V. A. Kostelecky and D. A. Campbell, Physica D **15**, 1 (1985).
[15]B. Sathiapalan, Phys. Rev. Lett. **55**, 669 (1985).
[16]E. Witten, Nucl. Phys. B **202**, 253 (1982).
[17]C. Callias, Commun. Math. Phys. **62**, 213 (1978).
[18]M. Hirayama, Prog. Theor. Phys. **70**, 1444 (1983).
[19]E. J. Weinberg, Phys. Rev. D **20**, 936 (1979).
[20]D. Bollé, F. Gesztesy, H. Grosse, and B. Simon, Lett. Math. Phys. **13**, 127 (1987).
[21]R. Akhoury and A. Comtet, Nucl. Phys. B **246**, 253 (1984).
[22]D. Boyanovsky and R. Blankenbecler, Phys. Rev. D **30**, 1821 (1984).
[23]F. Gesztesy, "Scattering theory for one-dimensional systems with nontrivial spatial asymptotics, in Schrödinger Operators, Aarhus 1985", *Lecture Notes in Mathematics*, Vol. 1218, edited by E. Balslev (Springer, Berlin, 1986), p. 93.
[24]A. Kihlberg, P. Salomonson, and B. S. Skagerstam, Z. Phys. C **28**, 203 (1985).
[25]M. Stone, Ann. Phys. (NY) **155**, 56 (1984).
[26]M. S. Birman and M. G. Krein, Sov. Math. Dokl. **3**, 740 (1962).
[27]M. G. Krein, Sov. Math. Dokl. **3**, 707 (1962).
[28]M. G. Krein, in *Topics in Differential and Integral Equations and Operator Theory*, edited by I. Gohberg (Birkhäuser, Basel, 1983), p. 107.
[29]I. C. Gohberg and M. G. Krein, "Introduction to the theory of linear nonself adjoint operators," *AMS-Translations of Mathematical Monographs*, Vol. 18 (Am. Math. Soc., Providence, RI, 1969).
[30]R. G. Newton, *Scattering Theory of Waves and Particles* (Springer, New York, 1982), 2nd ed.
[31]B. Simon, "Trace ideals and their applications," *London Mathematical Society Lecture Notes Series*, Vol. 35 (Cambridge U. P., Cambridge, 1979).
[32]D. Boyanovsky and R. Blankenbecler, Phys. Rev. D **31**, 3234 (1985).
[33]R. G. Newton, J. Math. Phys. **21**, 493 (1980).
[34]B. Simon, *Quantum Mechanics for Hamiltonians Defined as Quadratic Forms* (Princeton U. P., Princeton, NJ, 1971).
[35]M. Cheney, J. Math. Phys. **25**, 1449 (1984).
[36]R. G. Newton, J. Math. Phys. **18**, 1348 (1977).
[37]H. Baumgärtel and M. Wollenberg, *Mathematical Scattering Theory* (Akademie, Berlin, 1983).
[38]K. B. Sinha, "On the theorem of M. G. Krein," preprint, University of Geneva, 1975.
[39]A. Jensen and T. Kato, Commun. Part. Diff. Eq. **3**, 1165 (1978).
[40]T. Dreyfus, J. Math. Anal. Appl. **64**, 114 (1978).
[41]M. Wollenberg, Math. Nachr. **78**, 369 (1977).
[42]C. Macedo, PhD. thesis, University of Bielefeld, 1985.
[43]C. Callias, Commun. Math. Phys. **88**, 357 (1983).
[44]N. Dunford and J. T. Schwartz, *Linear operators II: Spectral theory* (Interscience, New York, 1963).
[45]H. Bremermann, *Distributions, Complex Variables and Fourier Transforms* (Addison–Wesley, New York, 1965).
[46]L. S. Koplienko, Sib. Math. J. **25**, 735 (1984).
[47]T. Kato, *Perturbation Theory for Linear Operators* (Springer, Berlin, 1980).
[48]F. Gesztesy and B. Simon, "Topological invariance of the Witten index," preprint 1987.
[49]P. A. Deift, Duke Math. J. **45**, 267 (1978).
[50]U. W. Schmincke, Proc. R. Soc. Edinburgh A **80**, 67 (1978).
[51]M. Reed and B. Simon, *Methods of Modern Mathematical Physics III: Scattering Theory* (Academic, New York, 1979).
[52]M. D. Bernstein and L. S. Brown, Phys. Rev. D **31**, 1515 (1985).
[53]A. J. Niemi and G. W. Semenoff, Phys. Rev. D **30**, 809 (1984).
[54]A. J. Niemi and G. W. Semenoff, Nucl. Phys. B **269**, 131 (1986).
[55]C. Imbimbo and S. Mukhi, Nucl. Phys. **242**, 81 (1984).
[56]F. Gesztesy, W. Plessas, and B. Thaller, J. Phys. A **13**, 2659 (1980).
[57]I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series and Products* (Academic, New York, 1965), p. 556.
[58]Y. Aharonov and A. Casher, Phys. Rev. A **19**, 2461 (1979).
[59]H. L. Cycon, R. Froese, W. Kirsch, and B. Simon, "Schrödinger operators with application to quantum mechanics and global geometry," *Texts and Monographs in Physics* (Springer, New York, 1987).
[60]H. J. deVega, Phys. Rev. D **18**, 2932 (1978).
[61]R. Jackiw, Phys. Rev. D **29**, 2375 (1984).
[62]K. C. Miller, PhD. thesis, Princeton University, 1982.
[63]J. Kiskis, Phys. Rev. D **15**, 2329 (1977).

An SU(8) model for the unification of superconductivity, charge, and spin density waves

Allan I. Solomon[a] and Joseph L. Birman[b]
Institute for Scientific Interchange, 10133 Torino, Italy

(Received 26 September 1986; accepted for publication 11 March 1987)

A model Hamiltonian for a many-electron system which unifies superconductivity, charge density waves, and spin density waves is analyzed. It is shown that the spectrum generating algebra for this system is su(8), and all 63 generators of this Lie algebra are identified. The seven symmetry operators that are broken in transition to the condensed state are identified, together with 56 order operators, whose expectations give the order parameters of the various phases present in the model. The discrete symmetry properties of these operators are tabulated. A chain of subalgebras of submodels with corresponding decoupled phases is constructed. Finally, how the finite temperature Green's functions may be obtained and used to solve the problem of self-consistency of the order parameters in the model is indicated.

I. INTRODUCTION

The pioneering experiments of Sooryakumar and Klein[1] on the coexistence of superconductivity and charge density wave phases, and many subsequent investigations, both theoretical and experimental,[2] have sparked interest in those systems for which the coexistence of these and other phases, such as ferro- and antiferromagnetic, are possible. In this paper we give a purely theoretical description, based on the approach of Lie algebras, to a system capable of embracing the phenomena of superconductivity and density waves. The model we analyze incorporates conventional homogeneous singlet superconductivity—and, perforce as a consequence of algebraic consistency, homogeneous triplet superconductivity. The density wave phenomena are those of charge density waves and spin density waves (antiferromagnetism)—within the same algebraic framework it is also possible to include ferromagnetic effects.

The approach we adopt is that of the spectrum-generating Lie algebra (SGA). Our model will be described by a Hamiltonian H given in terms of fermion creation and annihilation operators $a^\dagger_{k\sigma}$, $a_{k'\sigma'}$ for electrons constituting the electron gas in the system. Under suitable approximations, which we detail, H becomes a sum of bilinears in these operators; and so the terms of H generate a compact Lie algebra, the SGA of the model. For a model sufficiently general to include the physical phenomena noted above, the algebra is su(8).

The advantages of this algebraic approach are manifold. First, the various phenomena are synthesized into a single structure in which their relationships are transparent. The most striking example of this is the relationship between the existence of singlet superconductivity and density waves on the one hand, and triplet superconductivity on the other.[3] Another example is the description of the large number of "order operators"—these are operators whose expectations give the order parameters—which it would otherwise be difficult to classify. Second, although such a complex system does not lend itself easily to explicit calculation, the existence of low-dimensional faithful representations (8×8 matrices in the case of the full system, smaller matrices in the case of subsystems) simplifies explicit calculation of such physical quantities as spectra and phase coexistence boundaries, as we have previously illustrated in the simpler superconductivity-charge density wave su(4) case,[4] as well as selection rules[5] for various transition processes.

Third, this model may be regarded as unifying a variety of submodels, obtainable as subalgebras of su(8), which describe interesting physical systems of one or more phases, many of which have been previously treated separately in the literature.[6] Finally, within the context of mean field theory, where our model is firmly situated, finite temperature effects may be treated using the thermal Green's function method, and problems of self-consistency may also be tackled in this manner. We touch upon these questions in the final section of this paper.

II. MODEL HAMILTONIAN

Our starting Hamiltonian is a conventional sum of contributions from kinetic energy, superconducting, and density wave terms, thus

$$H = H_{\rm KE} + H_{\rm SC} + H_{\rm DW}, \tag{2.1}$$

where

$$H_{\rm KE} = \sum \epsilon(k) a^\dagger_{k\sigma} a_{k\sigma}, \tag{2.2}$$

$$H^{\rm O}_{\rm SC} = \sum \Delta^*_0(k) a_{k\uparrow} a_{-k\downarrow} + \text{H.c.}, \tag{2.3}$$

$$H_{\rm DW} = \sum \gamma_\mu(k) a^\dagger_{k+Q} \sigma_\mu a_k + \text{H.c.} \tag{2.4}$$

In the above, $a^\dagger_{k\sigma}$ is the fermion creation operator for an electron in the Bloch state labeled by wave vector k with spin σ and energy $\epsilon(k)$. We have the anticommutation rule

$$\{a_{k\sigma}, a^\dagger_{k'\sigma'}\} = \delta_{kk'}\delta_{\sigma\sigma'} \tag{2.5}$$

[a] Permanent address: Faculty of Mathematics, Open University, Milton Keynes, MK7 6AA, United Kingdom.
[b] Permanent address: Department of Physics, City College of the City University of New York, New York, New York 10031.

 0022-2488/87/071526-09$02.50

with other anticommutators zero. The BCS parameter $\Delta_0(k)$ may be taken complex, as may the density wave coupling constants $\gamma_\mu(k)$. Here $Q\equiv 2k_F$ is the characteristic wave vector of antiferromagnetic order, where k_F is the Fermi level. We have implicitly summed over the spin indices (understood) in H_{DW}, and over the index $\mu = 0,1,2,3$; we include $\mu = 0$ corresponding to a γ_0 charge-density wave coupling, while γ_i $(i = 1,2,3)$ is the spin-density wave term.

In principle, the summations in the above terms are over all k values. However, we now effect a considerable simplification, which leads to a decoupling and eventual algebraic solvability, by assuming that our model is quasi-one-dimensional, with no contributions from terms for which $|k| > Q$. The first two terms (2.2) and (2.3) may then be rearranged by use of the identity

$$\sum_{-Q}^{Q} f(k) = \sum_{0}^{k_F} \{f(k) + f(-k) + f(\bar{k}) + f(-\bar{k})\},$$

where $\bar{k}\equiv k - Q$; and a similar reduction of (2.4) leads to the model Hamiltonian $H = \Sigma_{k=0}^{k_F} H(k)$, where

$$\begin{aligned} H(k) &= \epsilon(k)(a^\dagger_{k\sigma}a_{k\sigma} + a^\dagger_{-k\sigma}a_{-k\sigma}) \\ &\quad + \epsilon(\bar{k})(a^\dagger_{\bar{k}\sigma}a_{\bar{k}\sigma} + a^\dagger_{-\bar{k}\sigma}a_{-\bar{k}\sigma}) \\ &\quad + \Delta_0^* a_{k\uparrow}a_{-k\downarrow} + \Delta_0^* a_{-k\uparrow}a_{k\downarrow} \\ &\quad + \Delta_0'^* a_{\bar{k}\uparrow}a_{-\bar{k}\downarrow} + \Delta_0'^* a_{-\bar{k}\uparrow}a_{\bar{k}\downarrow} + \text{H.c.} \\ &\quad + \gamma_\mu a^\dagger_{k\alpha}\sigma_\mu^{\alpha\beta}a_{\bar{k}\beta} + \gamma_\mu a^\dagger_{-\bar{k}\alpha}\sigma_\mu^{\alpha\beta}a_{-k\beta} + \text{H.c.} \end{aligned} \tag{2.6}$$

Here, as throughout the paper, we sum over repeated indices. We allow a k dependence of the BCS singlet gap parameter Δ_0, and so write Δ_0 for $\Delta_0(k)$, and Δ_0' for $\Delta_0(\bar{k})$.

We note that $[H(k),H(k')] = 0$ for $k,k'\in[0,k_F]$ so we have decoupled the Hamiltonian into a direct sum. As in Ref. 7, where we treated H_{DW} in more detail, we now define the set $\{B_i(k)\}$ $(i = 1,2,\ldots,8)$ by

$$\{B_i(k)\} = \{a_{k\uparrow}, a^\dagger_{-k\downarrow}, a_{\bar{k}\uparrow}, a^\dagger_{-\bar{k}\downarrow}; a_{k\downarrow}, a^\dagger_{-k\uparrow}, a_{\bar{k}\downarrow}, a^\dagger_{-\bar{k}\uparrow}\}. \tag{2.7}$$

From (2.5) we have $\{B_i,B_j^\dagger\} = \delta_{ij}$ whence the operators $X_{ij} \equiv B_i^\dagger B_j$ generate the Lie algebra gl(8) with commutation relations $[X_{ij},X_{kl}] = \delta_{jk}X_{il} - \delta_{il}X_{kj}$. The Hamiltonian $H(k)$ in (2.6) is a linear sum of Hermitian combinations and has trace zero since $\epsilon(k) = \epsilon(-k)$; therefore $H(k)$ may be considered as an element of su(8). The spectrum-generating algebra (SGA) of the model Hamiltonian H is thus a subalgebra

$$\oplus_k g_{(k)} \subset \oplus_k \mathrm{su}(8)_{(k)},$$

with each $g_{(k)}$ isomorphic to a fixed Lie algebra g (which we shall call the SGA of our model). We shall determine g later; we show that the presence of singlet superconductivity and spin density waves is sufficient to generate the whole su(8) algebra. This very rich rank-7 algebra possesses, in a Cartan basis, seven mutually commuting operators, which we interpret as conserved quantitites (above the transition temperatures) that are no longer conserved in the various phases present in the model below the appropriate transition temperatures; and 56 other basis elements which are putative order operators, whose expectations are order parameters for the corresponding phases.[8]

The bulk of this paper will be devoted to exploiting the algebraic consequence of this system of operators. We commence by introducing some notation. Define the Pauli matrices

$$\tau_0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \tau_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$

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

and the 4×4 matrices

$$\begin{aligned} &S_i = \tfrac{1}{2}\tau_0\times\tau_i, \quad T_i = \tfrac{1}{2}\tau_1\times\tau_i, \\ &U_i = \tfrac{1}{2}\tau_2\times\tau_i, \quad W_i = \tfrac{1}{2}\tau_3\times\tau_i, \\ &E_i = \tfrac{1}{2}\tau_i\times\tau_0 \quad (i = 1,2,3). \end{aligned} \tag{2.8}$$

The set (2.8) provides a "Nambu basis" for su(4) (Ref. 5). The basis for su(8) that we shall use is given by

$$\{\mathbf{S}\times\tau_\mu, \mathbf{T}\times\tau_\mu, \mathbf{U}\times\tau_\mu, \mathbf{W}\times\tau_\mu, \mathbf{E}\times\tau_\mu, I\times\boldsymbol{\tau}\} \quad (\mu = 0,1,2,3). \tag{2.9}$$

Here I is the 4×4 identity matrix. This is effectively a triple Nambu representation. The algebra su$(8)_{(k)}$ is generated by $\{\Sigma_{ij} B_i^\dagger(k)M^r_{ij}B_j(k)\}$ $(r = 1,\ldots,63)$, where M^r_{ij} is one of the 63 Hermitian matrices defined in (2.9).

If we take the standard representation of the gl(8) algebra generated by $X_{ij}(k) \equiv B_i^\dagger(k)B_j(k)$,

$$\hat{X}_{ij}(k) = e_{ij},$$

where

$$(e_{ij})_{lm} = \delta_{il}\delta_{jm} \quad (i,j,l,m = 1,2,\ldots,8),$$

then (2.9) is a basis for a representation of su$(8)_{(k)}$; we shall consistently denote this representation by a circumflex $\hat{}$. In this representation the number operator $N = \Sigma N(k)$, where

$$\begin{aligned} N(k) = \sum_{\alpha=\uparrow,\downarrow} &(a^\dagger_{k\alpha}a_{k\alpha} + a^\dagger_{-k\alpha}a_{-k\alpha} \\ &+ a^\dagger_{\bar{k}\alpha}a_{\bar{k}\alpha} + a^\dagger_{-\bar{k}\alpha}a_{-\bar{k}\alpha}) \end{aligned} \tag{2.10}$$

is given by

$$\hat{N}(k) = I\times\tau_3, \quad \text{where } I \text{ is the } 4\times 4 \text{ unit matrix.}$$

The spin operator $\Sigma_k\, \boldsymbol{\sigma}(k)$, where

$$\begin{aligned} \boldsymbol{\sigma}(k) = \sum_{\alpha,\beta} &(a^\dagger_{k\alpha}\boldsymbol{\sigma}^{\alpha\beta}a_{k\beta} + a^\dagger_{-k\alpha}\boldsymbol{\sigma}^{\alpha\beta}\boldsymbol{\sigma}_{-k\beta} \\ &+ a^\dagger_{\bar{k}\alpha}\boldsymbol{\sigma}^{\alpha\beta}a_{\bar{k}\beta} + a^\dagger_{-\bar{k}\alpha}\boldsymbol{\sigma}^{\alpha\beta}a_{-\bar{k}\beta}), \end{aligned}$$

is given by

$$(\hat{\sigma}_1(k),\hat{\sigma}_2(k),\hat{\sigma}_3(k)) = (E_1\times\tau_3,\ E_2\times\tau_3,\ E_3\times\tau_0). \tag{2.11}$$

(The spin matrices σ_μ are defined as usual by $\sigma_\mu \equiv \frac{1}{2}\tau_\mu$.)

Introduce the operator

$$\begin{aligned} S(k) = \frac{1}{2}\sum_{\alpha=\uparrow,\downarrow} &[a^\dagger_{k\alpha}a_{k\alpha} + a^\dagger_{-k\alpha}a_{-k\alpha} \\ &- (a^\dagger_{\bar{k}\alpha}a_{\bar{k}\alpha} + a^\dagger_{-\bar{k}\alpha}a_{-k\alpha})] \end{aligned}$$

represented by

$$\hat{S}(k) = S_3\times\tau_3. \tag{2.12}$$

We may now rewrite the Hamiltonian (2.6) as

$$H(k) = \tfrac{1}{2}(\epsilon + \epsilon')N(k) + (\epsilon - \epsilon')S(k) - \Delta_0 D_0(k) + \Delta_0' D_0'(k) + \text{H.c.} + \gamma_\mu \Gamma_\mu(k) + \text{H.c.} \qquad (2.13)$$

In (2.13) we have introduced a scalar, complex superconducting order operator

$$D_0(k) = a_{k\uparrow}^\dagger a_{-k\downarrow}^\dagger + a_{-k\uparrow}^\dagger a_{k\downarrow}^\dagger, \quad [\boldsymbol{\sigma}(k), D_0(k)] = 0, \qquad (2.14)$$

with a similar expression for $D_0'(k)$ in which k is replaced by $\bar{k}$. We have also introduced a complex charge-spin density wave order operator $\Gamma_\mu(k)$, defined by

$$\Gamma_\mu(k) = a_{k\alpha}^\dagger \sigma_\mu^{\alpha\beta} a_{\bar{k}\beta} + a_{-\bar{k}\alpha}^\dagger \sigma_\mu^{\alpha\beta} a_{-k\beta}. \qquad (2.15)$$

The $\mu = 0$ scalar component is the charge density part, while the $\mu = 1,2,3$ vector components refer to the spin density wave. The real and imaginary parts of $\Gamma_\mu(k)$ are two of a quartet of density wave order operators $\Gamma_\mu^{(\alpha)}$, fully defined in Sec. III, which satisfy

$$[\boldsymbol{\sigma}, \Gamma_0^{(\alpha)}] = 0, \quad [\sigma_l, \Gamma_m^{(\alpha)}] = i e_{lmn} \Gamma_n^{(\alpha)}, \qquad (2.16)$$

where e_{lmn} is the permutation symbol on $l,m,n = 1,2,3$, $\alpha = 1,2,3,4$.

In the representation with basis (2.9) and number and spin operators repesented by (2.10) and (2.11), respectively, these order operators are given by

$$\widehat{D}_0 = (E_3 + W_3)\times\tfrac{1}{2}(\tau_1 + i\tau_2), \quad \widehat{D}_0' = (E_3 - W_3)\times\tfrac{1}{2}(\tau_1 + i\tau_2) \qquad (2.17)$$

and

$$\{\widehat{\Gamma}_0, \widehat{\Gamma}_1, \widehat{\Gamma}_2, \widehat{\Gamma}_3\} = \{\tfrac{1}{2}(S_1 - iS_2)\times\tau_3, \tfrac{1}{2}(T_1 + iT_2)\times\tau_3, \tfrac{1}{2}(U_1 + iU_2)\times\tau_3, \tfrac{1}{2}(W_1 + iW_2)\times\tau_0\}. \qquad (2.18)$$

We may now rewrite our starting Hamiltonian (2.6) in the representation with basis (2.9) as

$$\widehat{H} = \widehat{H}_{\text{KE}} + \widehat{H}_{\text{SC}} + \widehat{H}_{\text{SDW}} + \widehat{H}_{\text{CDW}}, \qquad (2.19)$$

$$\widehat{H}_{\text{KE}} = (\epsilon + \epsilon')(\tfrac{1}{2} I \times \tau_3) + (\epsilon - \epsilon')S_3\times\tau_3, \qquad (2.20)$$

$$\widehat{H}_{\text{SC}} = -(\alpha + \alpha')(E_3\times\tau_1) + (\alpha' - \alpha)(W_3\times\tau_1) + (\beta + \beta')(E_3\times\tau_2) + (\beta - \beta')(W_3\times\tau_2), \qquad (2.21)$$

$$\widehat{H}_{\text{SDW}} = \text{Re}\,\gamma_1(T_1\times\tau_3) + \text{Re}\,\gamma_2(U_1\times\gamma_3) + \text{Re}\,\gamma_3(W_1\times\tau_0) - \text{Im}\,\gamma_1(T_2\times\tau_3) - \text{Im}\,\gamma_2(U_2\times\tau_3) - \text{Im}\,\gamma_3(W_2\times\tau_0), \qquad (2.22)$$

$$\widehat{H}_{\text{CDW}} = \text{Re}\,\gamma_0(S_1\times\tau_3) - \text{Im}\,\gamma_0(S_2\times\tau_3). \qquad (2.23)$$

In (2.21) $\Delta_0 = \alpha + i\beta$, $\Delta_0' = \alpha' + i\beta'$. The expressions (2.22) and (2.23) give the spin density wave and charge density wave terms, respectively. The operators in (2.20)–(2.23) are only part of a full system of order operators for this model. We define and examine the full system of order operators in the next section. In the meanwhile we write down for reference the matrix for the Hamiltonian (2.19) in the basis (2.9):

$$\widehat{H} = \begin{bmatrix} \epsilon & -\Delta_0 & \frac{1}{2}(\gamma_0+\gamma_3) & 0 & 0 & 0 & \frac{1}{2}(\gamma_1 - i\gamma_2) & 0 \\ -\Delta_0^* & -\epsilon & 0 & -\frac{1}{2}(\gamma_0-\gamma_3) & 0 & 0 & 0 & -\frac{1}{2}(\gamma_1 - i\gamma_2) \\ \frac{1}{2}(\gamma_0^*+\gamma_3^*) & 0 & \epsilon' & -\Delta_0' & \frac{1}{2}(\gamma_1^* - i\gamma_2^*) & 0 & 0 & 0 \\ 0 & -\frac{1}{2}(\gamma_0^*-\gamma_3^*) & -\Delta_0'^* & -\epsilon' & 0 & -\frac{1}{2}(\gamma_1^*-i\gamma_2^*) & 0 & 0 \\ 0 & 0 & \frac{1}{2}(\gamma_1+i\gamma_2) & 0 & \epsilon & \Delta_0 & \frac{1}{2}(\gamma_0-\gamma_3) & 0 \\ 0 & 0 & 0 & -\frac{1}{2}(\gamma_1+i\gamma_2) & \Delta_0^* & -\epsilon & 0 & -\frac{1}{2}(\gamma_0+\gamma_3) \\ \frac{1}{2}(\gamma_1^*+i\gamma_2^*) & 0 & 0 & 0 & \frac{1}{2}(\gamma_0^*-\gamma_3^*) & 0 & \epsilon' & \Delta_0' \\ 0 & -\frac{1}{2}(\gamma_1^*+i\gamma_2^*) & 0 & 0 & 0 & -\frac{1}{2}(\gamma_0^*+\gamma_3^*) & \Delta_0'^* & -\epsilon' \end{bmatrix}.$$

III. THE ORDER OPERATORS

We now analyze the Lie algebra su(8) with basis (2.9). This rank-7 algebra has seven Cartan (diagonal) elements and 56 off-diagonal elements. If h is a Cartan, e is a typical nondiagonal element satisfying the canonical rules

$$[h,e] = \lambda e \quad (\lambda \neq 0),$$

we see that in an eigenstate $|\rangle$ of h, $\langle|e|\rangle = 0$. The root vectors e, and linear combinations of such root vectors, are order operators for eigenstates of h. Their expectations are the order parameters which vanish in states for which h is a conserved operator. The eight Cartan elements for the u(8) algebra generated by the $B_i^\dagger B_j$ of (2.7) may simply be written $B_i^\dagger B_i$ $(i = 1,...,8)$; or more physically $n_{K\sigma}$, the number operator for K,σ $(K = \pm k, \pm \bar{k}, \sigma = \uparrow,\downarrow)$. In terms of the basis (2.9), the Cartan elements are

$$\widehat{N} = I\times\tau_3, \; \widehat{P} = 2S_3\times\tau_0, \; \widehat{S} = S_3\times\tau_3, \; \widehat{F} = 2E_3\times\tau_0 \;\; (=2\hat{\sigma}_3), \; E_3\times\tau_3, \;\; W_3\times\tau_0, \;\; W_3\times\tau_3. \qquad (3.1)$$

We have already introduced the number operator N, the difference of $k,\bar{k}$ number $S \equiv \frac{1}{2}(N_k - N_{\bar{k}})$, and the third component of spin σ_3 in Sec. II. (This last plays the role of a ferromagnetic order parameter F.) The matrix $\widehat{P}$ represents the momentum operator. [In the case of u(8) we would have additionally the unit matrix $I \times \tau_0$.]

We now illustrate a useful algebraic method for obtaining the order operators Q_i corresponding to a given quantum observable h. The operator h is assumed to be one of the operators conserved in the lower symmetry phase; we take it to be one of the elements of the Cartan subalgebra, and therefore diagonal in our representation. From the above remarks, the $\hat{Q}_i$ are the elements of the Cartan basis that do not commute with $\hat{h}$. Defining the centralizer of $\hat{h}$ as

$$C_{\mathrm{su}(8)}(\hat{h}) = \{x \in \mathrm{su}(8):\ [x,\hat{h}] = 0\},$$

we see that the set of order operators we seek is precisely the complement in su(8) of this centralizer, $C'_{\mathrm{su}(8)}(\hat{h})$. In addition, one may readily obtain such centralizers by the following method[9]: Let the matrix M in the defining representation of the group $U(n)$ be diagonal, with the eigenvalue multiplicities $m_1, m_2, \ldots, m_s$, where $m_1 + m_2 + \cdots + m_S = n$. The little group of M is $\mathrm{U}(m_1) \otimes \mathrm{U}(m_2) \otimes \cdots \otimes \mathrm{U}(m_S)$. Translating this result to the present Lie algebra context, if the diagonal matrix $\hat{h}$ has eigenvalues with multiplicities $m_1, m_2, \ldots, m_S$, where $m_1 + m_2 + \cdots + m_S = 8$, then

$$C_{\mathrm{u}(8)}(\hat{h}) = \mathrm{u}(m_1) \oplus \mathrm{u}(m_2) \oplus \cdots \otimes \mathrm{u}(m_S).$$

For the case of su(8) the corresponding result is

$$C_{\mathrm{su}(8)}(\hat{h}) = \mathrm{s}(\mathrm{u}(m_1) \oplus \cdots \oplus \mathrm{u}(m_s))$$
$$\sim \mathrm{u}(1) \oplus \mathrm{su}(m_1) \oplus \cdots \oplus \mathrm{su}(m_s).$$

As an example, take for our quantum observable h the number operator N. This is represented by the matrix $\hat{N} = I \times \tau_3$ [Eq. (2.10)]; we have $m_1 = m_2 = 4$, and so

$$C_{\mathrm{su}(8)}(\hat{N}) = \mathrm{u}(1) \oplus \mathrm{su}(4) \oplus \mathrm{su}(4).$$

Taking the complement, we find by this means 32 N-nonconserving operators $C'(\hat{N})$, which split into 16 superconducting D operators $C'(\hat{N}) \cap C(\hat{P})$, and 16 anomalous A operators $C'(\hat{N}) \cap C'(\hat{P})$. There are 16 density-wave Γ operators $C(\hat{N}) \cap C'(\hat{P})$, and finally eight ferromagnetic F operators $C(\hat{N}) \cap C(\hat{P}) \cap C'(\hat{F})$. The first three sets of operators divide naturally into scalar plus vector quartets as follows:

superconducting order operators:

$$\begin{aligned}
\hat{D}_\mu^{(1)} &= (E_3 \times \tau_1,\ -E_2 \times \tau_1,\ E_1 \times \tau_1,\ \tfrac{1}{2} I \times \tau_2),\\
\hat{D}_\mu^{(2)} &= (E_3 \times \tau_2,\ -E_2 \times \tau_2,\ E_1 \times \tau_2,\ \tfrac{1}{2} I \times \tau_1),\\
\hat{D}_\mu^{(3)} &= (-W_3 \times \tau_1,\ U_3 \times \tau_1,\ -T_3 \times \tau_1,\ -S_3 \times \tau_2),\\
\hat{D}_\mu^{(4)} &= (-W_3 \times \tau_2,\ U_3 \times \tau_2,\ -T_3 \times \tau_2,\ S_3 \times \tau_1);
\end{aligned} \tag{3.2}$$

charge-spin density wave operators:

$$\begin{aligned}
\hat{\Gamma}_\mu^{(1)} &= (-S_2 \times \tau_3,\ T_1 \times \tau_3,\ U_1 \times \tau_3,\ W_1 \times \tau_0),\\
\hat{\Gamma}_\mu^{(2)} &= (S_1 \times \tau_3,\ T_2 \times \tau_3,\ U_2 \times \tau_3,\ W_2 \times \tau_0),\\
\hat{\Gamma}_\mu^{(3)} &= (S_1 \times \tau_0,\ T_2 \times \tau_0,\ U_2 \times \tau_0,\ W_2 \times \tau_3),\\
\hat{\Gamma}_\mu^{(4)} &= (-S_2 \times \tau_0,\ T_1 \times \tau_0,\ U_1 \times \tau_0,\ W_1 \times \tau_3);
\end{aligned} \tag{3.3}$$

anomalous-order operators:

$$\begin{aligned}
\hat{A}_\mu^{(1)} &= (W_2 \times \tau_2,\ -U_1 \times \tau_1,\ T_1 \times \tau_1,\ S_1 \times \tau_2),\\
\hat{A}_\mu^{(2)} &= (W_1 \times \tau_1,\ -U_2 \times \tau_2,\ T_2 \times \tau_2,\ -S_2 \times \tau_1),\\
\hat{A}_\mu^{(3)} &= (-W_2 \times \tau_1,\ -U_1 \times \tau_2,\ T_1 \times \tau_2,\ -S_1 \times \tau_1),\\
\hat{A}_\mu^{(4)} &= (-W_1 \times \tau_2,\ -U_2 \times \tau_1,\ T_2 \times \tau_1,\ S_2 \times \tau_2);
\end{aligned} \tag{3.4}$$

ferromagnetic order parameters:

$$\{E_1, E_2, T_3, U_3\} \times \{\tau_0, \tau_3\} \tag{3.5}$$

are simply the off-diagonal elements of the ferromagnetic subalgebra which is their closure, namely

$$\{E_1, E_2, E_3, T_3, U_3, W_3\} \times \{\tau_0, \tau_3\}. \tag{3.6}$$

This algebra is the $\mathrm{su}(2) \oplus \mathrm{su}(2) \oplus \mathrm{su}(2) \oplus \mathrm{su}(2)$ generated by

$$\{a_k^\dagger \boldsymbol{\sigma} a_k,\ a_{-k}^\dagger \boldsymbol{\sigma} a_{-k},\ a_{\bar{k}}^\dagger \boldsymbol{\sigma} a_{\bar{k}},\ a_{-\bar{k}}^\dagger \boldsymbol{\sigma} a_{-\bar{k}}\},$$

four independent spin algebras. Corresponding to four linearly independent combinations of these spins, we may define the operators $\boldsymbol{\sigma}$ from (2.11) and $\boldsymbol{\sigma}^{(1)}$, $\boldsymbol{\sigma}^{(2)}$, $\boldsymbol{\sigma}^{(3)}$ from (5.10) below.

IV. DISCRETE SYMMETRIES

A. Parity inversion π

This is defined by $\pi a_{k\sigma} \pi^\dagger = a_{-k\sigma}$, where π is a unitary, linear operator. Acting on the B basis (2.7) we have

$$\pi(B_1, B_2, B_3, B_4)\pi^\dagger = B_6^\dagger, B_5^\dagger, B_8^\dagger, B_7^\dagger).$$

We may represent this action as an 8×8 matrix

$$\pi B_i \pi^\dagger = \sum_j A_{ij} B_j^\dagger,$$

where

$$A = \tau_1 \times \tau_0 \times \tau_1.$$

The action on a bilinear in B, $\Sigma_{ij}\, m_{ij} B_i^\dagger B_j$ (with tr $m = 0$) is easily calculated to be

$$\pi(B^\dagger m B)\pi^\dagger = -B^\dagger A \tilde{m} A B,$$

where $\tilde{m}$ is the transpose of the 8×8 matrix m. Thus in the 8×8 representation of Sec. II (2·8 *et seq.*) parity inversion corresponds to

$$m \to -A \tilde{m} A.$$

B. Time inversion $\mathscr{T}$

This is defined by $\mathscr{T} a_{k\sigma} \mathscr{T}^\dagger = \Sigma_{\sigma'} (i\tau_2)_{\sigma\sigma'} a_{-k\sigma'}$, where $\mathscr{T}$ is a unitary, antilinear operator. Acting on the B basis of (2.7) we have

$$\mathscr{T}(B_1, B_2, B_3, B_4; B_5, B_6, B_7, B_8)\mathscr{T}^\dagger$$
$$= (B_2^\dagger, -B_1^\dagger, B_4^\dagger, -B_3^\dagger; -B_6^\dagger, B_5^\dagger, -B_8^\dagger, B_7^\dagger).$$

We may represent this action as an 8×8 matrix,

$$\mathscr{T} B_i \mathscr{T}^\dagger = \sum_j T_{ij} B_j^\dagger,$$

where

$$T = i\tau_3 \times \tau_0 \times \tau_2.$$

The action on a bilinear $\Sigma m_{ij} B_i^\dagger B_j$ (with tr $m = 0$) is readily evaluated to give

$$\mathscr{T}(B^\dagger m B)\mathscr{T}^\dagger = B^\dagger T m^\dagger T B.$$

In our 8×8 representation, time reversal corresponds to

$$m \to T m^\dagger T.$$

C. Charge conjugation $\mathscr{C}$

From the action $\psi_\sigma(x) \to \mathscr{C} \psi_\sigma(x) \mathscr{C}^\dagger = \psi_\sigma^\dagger(x)$, we define charge conjugation to act on the electron destruction operator $a_{k\sigma}$ by $\mathscr{C} a_{k\sigma} \mathscr{C}^\dagger = a_{-k\sigma}^\dagger$, where $\mathscr{C}$ is a unitary, linear operator. On the B basis, we have $\mathscr{C}(B_1, B_2, B_3, B_4)\mathscr{C}^\dagger = (B_6, B_5, B_8, B_7)$, whence

TABLE I. Parity, time reversal, and charge conjugation properties.

Scalars

	$D_0^{(1)}$	$D_0^{(2)}$	$D_0^{(3)}$	$D_0^{(4)}$	;	$\Gamma_0^{(1)}$	$\Gamma_0^{(2)}$	$\Gamma_0^{(3)}$	$\Gamma_0^{(4)}$	;	$A_0^{(1)}$	$A_0^{(2)}$	$A_0^{(3)}$	$A_0^{(4)}$	;	N	S	P
π	+	+	+	+	;	−	+	−	+	;	−	+	−	+	;	+	+	−
$\mathscr{T}$	+	−	+	−	;	+	+	−	−	;	−	+	+	−	;	+	+	−
$\mathscr{C}$	−	+	−	+	;	−	−	+	+	;	+	−	−	+	;	−	−	+

Vectors

	$\mathbf{D}^{(1)}$	$\mathbf{D}^{(2)}$	$\mathbf{D}^{(3)}$	$\mathbf{D}^{(4)}$	;	$\mathbf{\Gamma}^{(1)}$	$\mathbf{\Gamma}^{(2)}$	$\mathbf{\Gamma}^{(3)}$	$\mathbf{\Gamma}^{(4)}$	;	$\mathbf{A}^{(1)}$	$\mathbf{A}^{(2)}$	$\mathbf{A}^{(3)}$	$\mathbf{A}^{(4)}$	;	$\boldsymbol{\sigma}$	$\boldsymbol{\sigma}^{(1)}$	$\boldsymbol{\sigma}^{(2)}$	$\boldsymbol{\sigma}^{(3)}$
π	−	−	−	−	;	+	−	+	−	;	−	+	−	+	;	+	−	−	+
$\mathscr{T}$	−	+	−	+	;	−	−	+	+	;	−	+	+	−	;	−	+	+	−

$$\mathscr{C} B_i C^\dagger = \sum A_{ij} B_j,$$

where A is the same matrix as in part (i). The action on a bilinear in B_i is therefore given by

$$\mathscr{C}(B^\dagger m B)\mathscr{C}^\dagger = B^\dagger (AmA) B;$$

in our representation the effective action of charge conjugation corresponds to

$$m \to AmA.$$

We append a table of the discrete transformation properties of the 15 scalars and 48 vectors of this model (see Table I).

V. COMMUTATORS

In (3.2), (3.3), and (3.4), writing $X_\mu = D_\mu^{(\alpha)}, A_\mu^{(\alpha)}$, or $\Gamma_\mu^{(\alpha)}$, the zero-component operators are the scalar quantities satisfying

$$[\boldsymbol{\sigma}, X_0] = 0. \tag{5.1}$$

Thus $D_0^{(\alpha)}$ ($\alpha = 1,2,3,4$) are the ordinary superconducting singlet order operators occurring in the (2.22), while the $\Gamma_0^{(\alpha)}$ are the charge-density order operators, of which the two even-time-reversal scalars appear in (2.4). The triplet operators satisfy

$$[\sigma_i, X_j] = i e_{ijk} X_k, \tag{5.2}$$

$$[X_i, X_j] = i e_{ijk} \sigma_k, \tag{5.3}$$

so that, for example, the $\Gamma_j^{(\alpha)}$ are spin-density order operators, of which the two odd-time-reversal triplets appear in (2.23). The operators in (3.2)–(3.4) satisfy

$$[D_0^{(\alpha)}, \Gamma_0^{(\alpha)}] = i A_0^{(\alpha)} \tag{5.4}$$

and

$$[D_i^{(\alpha)}, \Gamma_j^{(\alpha)}] = i e_{ijk} A_k^{(\alpha)}, \tag{5.5}$$

with two similar sets of commutators obtained by cyclic permutation.

The N,P, and S operators (3.1) move one quartet of order operators to the next, for example,

$$\begin{aligned} &[\tfrac{1}{2} N, D_\mu^{(1)}] = i D_\mu^{(2)}, \quad [\tfrac{1}{2} N, D_\mu^{(3)}] = i D_\mu^{(4)}, \\ &[S, D_\mu^{(2)}] = i D_\mu^{(3)}, \quad [S, D_\mu^{(4)}] = i D_\mu^{(1)}. \end{aligned} \tag{5.6}$$

The analogous commutators for $\Gamma_\mu^{(\alpha)}$ and $A_\mu^{(\alpha)}$ are

$$\begin{aligned} &[\tfrac{1}{2} P, \Gamma_\mu^{(1)}] = i\Gamma_\mu^{(2)}, \quad [\tfrac{1}{2} P, \Gamma_\mu^{(4)}] = i\Gamma_\mu^{(3)}, \\ &[S, \Gamma_\mu^{(1)}] = i\Gamma_\mu^{(3)}, \quad [S, \Gamma_\mu^{(4)}] = i\Gamma_\mu^{(2)}; \end{aligned} \tag{5.7}$$

$$\begin{aligned} &[\tfrac{1}{2} N, A_\mu^{(1)}] = i A_\mu^{(3)}, \quad [\tfrac{1}{2} N, A_\mu^{(4)}] = i A_\mu^{(2)}, \\ &[\tfrac{1}{2} P, A_\mu^{(3)}] = i A_\mu^{(2)}, \quad [\tfrac{1}{2} P, A_\mu^{(1)}] = i A_\mu^{(4)}. \end{aligned} \tag{5.8}$$

The singlet and triplet components of the order operators are related as follows:

$$\begin{aligned} &[D_0^{(\alpha)}, \mathbf{D}^{(\alpha)}] = i\boldsymbol{\sigma}^{(1)}, \\ &[\Gamma_0^{(\alpha)}, \mathbf{\Gamma}^{(\alpha)}] = i\boldsymbol{\sigma}^{(2)}, \\ &[A_0^{(\alpha)}, \mathbf{A}^{(\alpha)}] = i\boldsymbol{\sigma}^{(3)} \quad (\text{no sum over } \alpha). \end{aligned} \tag{5.9}$$

These pseudospin triplets are represented by

$$\begin{aligned} &\hat{\boldsymbol{\sigma}}^{(1)} = (E_1 \times \tau_0,\ E_2 \times \tau_0,\ E_3 \times \tau_3), \\ &\hat{\boldsymbol{\sigma}}^{(2)} = (T_3 \times \tau_0,\ U_3 \times \tau_0,\ W_3 \times \tau_3), \\ &\hat{\boldsymbol{\sigma}}^{(3)} = (-T_3 \times \tau_3,\ -U_3 \times \tau_3,\ -W_3 \times \tau_0). \end{aligned} \tag{5.10}$$

These triplets have the following commutation relations:

$$\begin{aligned} &[\sigma_i^{(\alpha)}, \sigma_j^{(\alpha)}] = i e_{ijk} \sigma_k, \\ &[\sigma_i, \sigma_j^{(\alpha)}] = i e_{ijk} \sigma_k^{(\alpha)}, \\ &[\sigma_i^{(\alpha)}, \sigma_j^{(\beta)}] = -i e_{ijk} e^{\alpha\beta\gamma} \sigma_k^{(\gamma)} \quad (i,j,k;\ \alpha,\beta,\gamma = 1,2,3). \end{aligned} \tag{5.11}$$

The $\sigma_j^{(\alpha)}$ connect triplet components with singlet, for example:

$$[\boldsymbol{\sigma}^{(1)}, D_0^{(\alpha)}] = i\mathbf{D}^{(\alpha)}$$

and

$$[\boldsymbol{\sigma}^{(1)}, \mathbf{D}^{(\alpha)}] = i D_0^{(\alpha)}, \tag{5.12}$$

with similar relations for $\Gamma_\mu^{(\alpha)}$ and $A_\mu^{(\alpha)}$.

VI. THE SPECTRUM GENERATING ALGEBRA

We may write our starting Hamiltonian (2.6) and (2.19)–(2.23) in terms of the order operators (3.2) and (3.3) as

$$\begin{aligned} \hat{H} = \tfrac{1}{2}(\epsilon + \epsilon')\hat{N} + (\epsilon - \epsilon')\hat{S} + \Delta_0^{(\alpha)} \hat{D}_0^{(\alpha)} \\ + \gamma_\mu^{(1)} \hat{\Gamma}_\mu^{(1)} + \gamma_\mu^{(2)} \hat{\Gamma}_\mu^{(2)} \end{aligned} \tag{6.1}$$

(with summation over μ and σ), where

$$\begin{aligned} &\{\Delta_0^{(1)}, \Delta_0^{(2)}, \Delta_0^{(3)}, \Delta_0^{(4)}\} \\ &\quad \equiv \{-\mathrm{Re}(\Delta_0 + \Delta_0'), \mathrm{Im}(\Delta_0, \Delta_0'), \\ &\qquad \mathrm{Re}(\Delta_0 - \Delta_0'), \mathrm{Im}(\Delta_0' - \Delta_0)\} \end{aligned}$$

and

$$\gamma_\mu^{(1)} = \{\mathrm{Im}\,\gamma_0, \mathrm{Re}\,\gamma\}, \quad \gamma_\mu^{(2)} = \{\mathrm{Re}\,\gamma_{0'} - \mathrm{Im}\,\gamma\}. \tag{6.2}$$

From the form of the Hamiltonian given in (6.1), using the commutation relations of the previous section, it is a straightforward matter to determine the spectrum generating algebra g for this system; that is, the algebra generated by the elements of (6.1). Since these are all elements of su(8), the SGA must be a subalgebra of su(8). In fact, we now demonstrate that the algebraic closure g of the operators occurring in (6.1) is all of su(8). This has the consequence that all of the 63 operators of the theory will appear in the time evolution of the order operators already present in the Hamiltonian (6.1). Whether they give rise to physical phases will depend on an evaluation of their expectations in the eigenstates of (6.1) or on a self-consistent analysis.

The generation of all su(8), for example, from (6.1) may be seen in the following stages.

(i) Since $\Gamma_\mu^{(1)},\Gamma_\mu^{(2)}\in H$ and $S\in H$, using (5.7) we have that all $\Gamma_\mu^{(\alpha)}\in g$.

(ii) Also $[\Gamma_i^{(1)},\Gamma_j^{(2)}] = ie_{ijk}\sigma_k$, so $\boldsymbol{\sigma}\in g$.

(iii) Evaluating $[\Gamma_i^{(1)},\Gamma_j^{(4)}] = ie_{ijk}\sigma_k^{(1)}$ gives $\boldsymbol{\sigma}^{(1)}\in g$.

(iv) From $D_0^{(\alpha)},\boldsymbol{\sigma}^{(1)}\in g$, using (5.12) gives all $D_\mu^{(\alpha)}\in g$.

(v) Using (5.4) and (5.5) we see that $D_\mu^{(\alpha)},\Gamma_\mu^{(\alpha)}$ generate $A_\mu^{(\alpha)}$.

(vi) As in (iii) $[D_i^{(1)},D_j^{(3)}] = ie_{ijk}\sigma_k^{(3)}$ and $[A_i^{(1)},A_j^{(2)}] = ie_{ijk}\sigma_k^{(2)}$ imply that all $\boldsymbol{\sigma}^{(\alpha)}\in g$. The 60 operators $D_\mu^{(\alpha)},\Gamma_\mu^{(\alpha)},A_\mu^{(\alpha)},\boldsymbol{\sigma}^{(\alpha)},\boldsymbol{\sigma}$ together with the 3 remaining Cartan operators S, N, and $P = (2/i)[\Gamma_\mu^{(1)},\Gamma_\mu^{(2)}]$ exhaust su(8).

We may note at this point that the commutation relation

$$[D_0^{(3)},\boldsymbol{\Gamma}^{(2)}] = -i\mathbf{A}^{(3)}$$

generates an odd-parity odd-time inversion anomalous triplet term from singlet superconductivity ($T = -1$) and a spin-density term ($T = -1$). The production of such an anomalous term has been previously noted in the literature.[10]

However, even more striking is the generation of conventional ($Q = 0$) triplet superconductivity from the interaction of singlet superconductivity and density waves. A simplified model[3] exhibiting this phenomenon may be obtained from (6.1) by choosing $\Delta_0 = \Delta_0'$ (real) and $\gamma_\mu^{(2)} = 0$ in (6.2). It is also sufficient to choose axes so that only $\gamma^{(2)}\to\gamma_3^{(2)}$. It may be shown that the SGA of this submodel is so(4) $\oplus$ so(4). The even-time-reversal triplet superconductivity order operator $D_3^{(3)}$ is generated as a second-order effect of the interaction between the singlet superconductivity, and the charge and spin density waves; it has nonzero expectation in the ground state of the Hamiltonian and may therefore be considered as an observable phase.[3]

VII. SUBALGEBRAS AND SUBMODELS

It is a fairly straightforward matter to obtain the spectrum generating algebras corresponding to submodels of the Hamiltonian (6.1). These algebras are generated by subsets of the 63 su(8) operators, (3.1)–(3.5). The components of the algebras generated by the order operator terms may most easily be calculated by taking centralizers; to these one must add the other terms of the Hamiltonian (such as kinetic energy N,S). We illustrate this method by obtaining the spectrum generating algebras of some previously noted submodels.

A. Superconducting models

The order operators for superconducting systems are defined as those which conserve momentum, but do not conserve number. As in Sec. III, we obtain the set $\{D_\mu^{(\alpha)}\}$, represented by the matrices (3.2). These may be succinctly written as

$$\{\mathbf{E},T_3,U_3,W_3,S_3,I\}\times\{\tau_1,\tau_2\} \tag{7.1}$$

and in this form we see that they generate the subalgebra

$$\{\mathbf{E},T_3,U_3,W_3\}\times\tau_\mu\cup\{S_3,I\}\times\boldsymbol{\tau}, \tag{7.2}$$

which is isomorphic to su(4) $\oplus$ su(4). This algebra is the semisimple component of the centralizer of momentum P in su(8);

$$C_{\mathrm{su}(8)}(P)\sim \mathrm{u}(1)\oplus\mathrm{su}(4)\oplus\mathrm{su}(4).$$

As this su(4) $\oplus$ su(4) algebra also contains the appropriate kinetic energy terms N and S, this is the spectrum generating algebra corresponding to a two-component $(k,\bar{k})$ superconducting fermion system. Each su(4) corresponds to a mixed triplet–singlet superconductor as previously obtained[11]—one for k and the other for $\bar{k}$. This may be made explicit as follows: define

$$\tau_\uparrow = \tfrac{1}{2}(\tau_0+\tau_3), \quad \tau_\downarrow = \tfrac{1}{2}(\tau_0-\tau_3).$$

Then the two commuting su(4) algebras are

$$\begin{aligned} &k \text{ component: } \mathrm{su}(4)\sim\{\boldsymbol{\tau}\times\tau_\uparrow\times\tau_\mu,\ \tau_0\times\tau_\uparrow\times\boldsymbol{\tau}\},\\ &\bar{k} \text{ component: } \mathrm{su}(4)\sim\{\boldsymbol{\tau}\times\tau_\downarrow\times\tau_\mu,\ \tau_0\times\tau_\downarrow\times\boldsymbol{\tau}\}. \end{aligned} \tag{7.3}$$

Conventional singlet superconductivity may be obtained as the centralizer of the spin operator in either of the above su(4) models, thus

$$\begin{aligned} C_{\mathrm{su}(4)}(\boldsymbol{\sigma}) &= \{\tau_3\times\tau_\uparrow\times\tau_1,\ \tau_3\times\tau_\uparrow\times\tau_2,\ \tau_0\times\tau_\uparrow\times\tau_3\}\\ &\sim\mathrm{so}(3), \end{aligned} \tag{7.4}$$

which is the spectrum-generating algebra of the singlet superconductor. In the notation of the previous section, the $\mathrm{so}(3)_{(k)}\oplus\mathrm{so}(3)_{(\bar{k})}$ singlet subalgebra has basis $\{N,S,D_0^{(\alpha)}\}$. The spin-1, pure triplet case corresponds to the $\mathrm{so}(5)_{(k)}\oplus\mathrm{so}(5)_{(\bar{k})}$ subalgebra with basis $\{N,S,\mathbf{D}^{(\alpha)},\boldsymbol{\sigma},\boldsymbol{\sigma}^{(3)}\}$, in the notation of the previous section. Each so(5) algebra is also the SGA for superfluid He III,[12,13] or a spin-1 superconductor.

B. Density wave models

The order operators for density wave systems are defined as those which conserve number, but do not conserve momentum. As in Sec. III, we obtain the set $\{\Gamma_\mu^{(\alpha)}\}$, represented by the matrices (3.3). We may rewrite this set as

$$\tau_\mu\times\{\tau_1,\tau_2\}\times\{\tau_0,\tau_3\}. \tag{7.5}$$

As in (i) above, under commutation these generate

$$\boldsymbol{\tau}\times\tau_\mu\times\{\tau_0,\tau_3\}\cup\tau_0\times\boldsymbol{\tau}\times\{\tau_0,\tau_3\}, \tag{7.6}$$

which is again isomorphic to an su(4) $\oplus$ su(4) algebra. To obtain the spectrum generating algebra (SGA) of the density wave Hamiltonian containing the $\Gamma_\mu^{(\alpha)}$ order operators we must adjoin the number operator N, which is not present in (7.6). Thus the SGA for a mixed spin and charge density wave model is $u(1)\oplus su(4)\oplus su(4)$, as previously obtained.[7] This algebra may be most simply obtained as the centralizer of the number operator N in su(8),

$$C_{su(8)}(N)\sim u(1)\oplus su(4)\oplus su(4). \tag{7.7}$$

The centralizer of spin $\boldsymbol{\sigma}$ in (7.7) gives the CDW algebra generated by $\{P,S,N,\Gamma_0^{(\alpha)}\}$ which is $u(1)\oplus so(4)$; we have previously obtained this directly from a model charge-density wave Hamiltonian.[14] The spin-1 part of (7.7), the pure spin density wave algebra, is generated by the nonspin conserving elements of (7.5), and has for basis $\{P,S,N,\boldsymbol{\Gamma}^{(\alpha)}, \boldsymbol{\sigma},\boldsymbol{\sigma}^{(1)}\}$. This is the algebra $u(1)\oplus so(5)\oplus SO(5)$, as calculated previously from a specific density wave Hamiltonian.[7]

If we consider only the parity-invariant elements of the above subalgebras, where the parity operator π is as defined in Sec. IV A, then we obtain subalgebras as follows:

$$\begin{aligned}&\text{CDW} \quad \{N,S,\Gamma_0^{(2)},\Gamma_0^{(4)}\}\sim u(2),\\ &\text{SDW} \quad \{N,S,\boldsymbol{\Gamma}^{(1)},\boldsymbol{\Gamma}^{(3)},\boldsymbol{\sigma}\}\sim u(1)\oplus so(5).\end{aligned} \tag{7.8}$$

These are the spectrum generating algebras for the model Hamiltonian (6.1) in the absence of superconductivity taking the coupling constants (6.2) real, and considering pure scalar and pure vector, respectively.

C. Singlet model

We obtain a spin-0 model by taking the centralizer of the spin operator $\boldsymbol{\sigma}$ in our su(8) algebra; thus

$$C_{su(8)}(\boldsymbol{\sigma})\sim su(4)$$

with basis

$$\{I\times\tau_3,\ \mathbf{S}\times\tau_0,\ \mathbf{S}\times\tau_3,\ \mathbf{W}\times\tau_1,\mathbf{W}\times\tau_2,\ E_3\times\tau_1,\ E_3\times\tau_2\}. \tag{7.9}$$

This su(4) is isomorphic to that obtained previously for the spectrum generating algebra of a model Hamiltonian exhibiting the coexistence of superconductivity and charge density waves.[4] To see this isomorphism more readily, write the set (7.9) in the form

$$\tau_0\times\tau_\mu\times\{\tau_0,\tau_3\}\cup\tau_3\times\tau_\mu\times\{\tau_1,\tau_2\} \tag{7.10}$$

[where in (7.10) we have actually considered $C_{u(8)}(\boldsymbol{\sigma})\sim u(4)$ for simplicity; we can always discard the central element $\tau_0\times\tau_0\times\tau_0$ later]. The set (7.10) is clearly isomorphic to

$$\{\tau_0\times\tau_0,\ \tau_0\times\tau_3,\ \tau_3\times\tau_1,\ \tau_3\times\tau_2\}\times\tau_\mu,$$

which in turn is isomorphic to $\tau_\mu\times\tau_\mu$. This is the set of 15 generations $\{\mathbf{E},\mathbf{S},\mathbf{T},\mathbf{U},\mathbf{W}\}$ of su(4) (together with the unit element $\tau_0\times\tau_0$) of Ref. 4.

D. Spin models

The eight spin order operators (3.5) generate the algebra with basis $\{\boldsymbol{\sigma}^{(\alpha)}\}$ ($\alpha=1,2,3,4$, $\boldsymbol{\sigma}^{(4)}\equiv\boldsymbol{\sigma}$); as remarked in Sec. III this is equivalent to four independent su(2) algebras. We may obtain the spectrum generating algebras for spin model Hamiltonians by adjoining the kinetic energy terms N and S. For example, the even parity spins give an algebra $\{N,S,\boldsymbol{\sigma},\boldsymbol{\sigma}^{(3)}\}$. This splits up into $\{\tau_0\times\tau_\uparrow\times\tau_3,\ \tau_1\times\tau_\uparrow\times\tau_3,\ \tau_2\times\tau_\uparrow\times\tau_3,\ \tau_3\times\tau_\uparrow\times\tau_0\}\sim u(2)$, and $\{\tau_0\times\tau_\downarrow\times\tau_3,\ \tau_1\times\tau_\downarrow\times\tau_3,\ \tau_2\times\tau_\downarrow\times\tau_3,\ \tau_3\times\tau_\downarrow\times\tau_0\}\sim u(2)$; two independent u(2) spin models, for the k and $\bar{k}$ systems respectively.

We show the descent from su(8) to the subgroups corresponding to models (A)–(D) in Fig. 1.

VIII. GENERAL HAMILTONIAN AND SELF-CONSISTENCY

We may now write down the most general Hamiltonian for the coexistence of superconductivity (singlet and triplet) and density waves (charge and spin) within the context of our su(8) algebra. This will generate the expression (6.1), thus

$$\begin{aligned}H&=\sum_k H(k),\\ H(k)&=\tfrac{1}{2}(\epsilon+\epsilon')N+(\epsilon-\epsilon')S+pP\\ &\quad+\Delta_\mu^{(\rho)}D_\mu^{(\rho)}\quad\text{(superconducting terms)}\\ &\quad+\gamma_\mu^{(\rho)}\Gamma_\mu^{(\rho)}\quad\text{(density waves)}\\ &\quad+\alpha_\mu^{(\rho)}A_\mu^{(\rho)}\quad\text{(anomalous terms)}\\ &\quad+\mathbf{H}_{\text{ext}}^{(\rho)}\cdot\boldsymbol{\sigma}^{(\rho)}\quad\text{(magnetic field terms)}.\end{aligned} \tag{8.1}$$

We sum over repeated indices in (8.1): $\mu=0,1,2,3$ and $\rho=1,2,3,4$. We have written $\boldsymbol{\sigma}^{(4)}\equiv\boldsymbol{\sigma}$ for conciseness; and have included a momentum term pP, where $\hat{P}=2S_3\times\tau_0$, in order to attain the full complement of 63 operators. The magnetic field terms in (8.1) enable calculations of susceptibilities, as has been carried out for the SDW subalgebra of su(8) (Ref. 7).

The expression (8.1) has the virtue of explicitness; however, a more concise, if less transparent, form of the mean-field Hamiltonian H is given by

$$H=\sum_k m_{ij}(k)X_{ij}(k) \tag{8.2}$$

(summation over repeated indices i,j) in terms of the operators $X_{ij}(k)\equiv B_i^\dagger(k)B_j(k)$ introduced in Sec. II. These satisfy the commutation relations

$$[X_{ij}(k),X_{rs}(k')]=\delta_{kk'}(\delta_{jr}X_{is}(k)-\delta_{is}X_{rj}(k)). \tag{8.3}$$

We may consider the mean-field Hamiltonian (8.2) to have arisen from a pairing Hamiltonian H^{red} in the following way. We require that H^{red} conserve number N, momentum P, etc., in fact, all seven Cartan operators that are broken in the passage to the lower symmetry, mean-field system. These operators have the form $\sum_k\lambda_i(k)X_{ii}(k)$ ($i=1,2,\dots,8$) (adding in the identity) and it is straightforward to verify that the Hamiltonian

$$H^{\text{red}}=\frac{1}{2}\sum_{i,j,k,k'}g_{ij}(k,k')X_{ij}(k)X_{ij}(k')^\dagger \tag{8.4}$$

conserves these quantities. Thus (8.4) is a suitable choice of pairing Hamiltonian. If we choose

$$g_{ij}(k,k)=2\epsilon_i(k)\delta_{ij} \tag{8.5}$$

and note that $[X_{ii}(k)]^2=X_{ii}(k)$, then the kinetic energy terms are also included in (8.4). With this choice the coupling constants satisfy

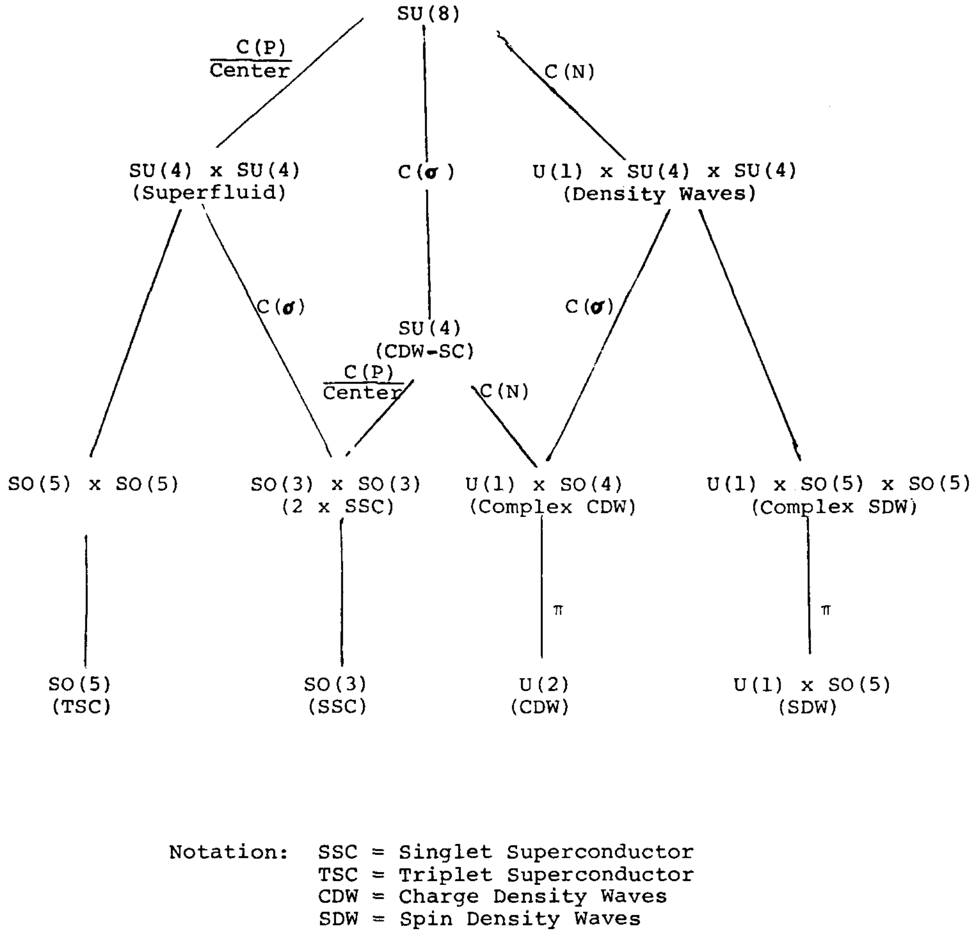

FIG. 1. Subgroup descent from SU(8).

$$g_{ij}(k,k')=g_{ji}(k',k). \tag{8.6}$$

In addition, from the Hermiticity of $H^{\rm red}$ we have

$$g^*_{ij}(k,k')=g_{ij}(k',k). \tag{8.7}$$

Now define

$$m_{ij}(k)=\left\langle\!\!\left\langle\sum_{k'} g_{ij}(k,k')X^\dagger_{ij}(k')\right\rangle\!\!\right\rangle \quad (i\neq j) \tag{8.8}$$

(no summation over ij), where $\langle\langle\ \rangle\rangle$ refers to a thermal average with respect to the pairing Hamiltonian (8.4),

$$\langle\langle Q\rangle\rangle \equiv \mathrm{tr}\{\exp(-\beta H^{\rm red})Q\}/\mathrm{tr}\exp(-\beta H^{\rm red}). \tag{8.9}$$

We now apply a Hartree–Fock linearization to $H^{\rm red}$, and obtain as an approximation the mean-field form (8.2), using relations (8.6) and (8.7). We now introduce the thermal Green's functions,[15,16] and

$$G_{ij}(k,\tau)=-\langle\langle T_\tau\{B_i(k,\tau)B^\dagger_j(k,0)\}\rangle\rangle, \tag{8.10}$$

where, at the level of mean-field theory, the thermal average is with respect to the mean-field Hamiltonian H of (8.2), as is the Heisenberg τ evolution

$$B_i(k,\tau)=\exp(H\tau)B_i(k)\exp(-H\tau).$$

Here T_τ is the τ-ordering operator, so that

$$G_{ij}(k,0^-)=\langle\langle X^\dagger_{ij}(k)\rangle\rangle. \tag{8.11}$$

Writing the conventional ω transform of the Green's function, and replacing $\Sigma_{k'}$ in (8.8) by the integral, we obtain the self-consistent equation

$$m_{ij}(k)=\frac{1}{\beta}\sum_n\int\frac{d^3k'}{(2\pi)^3}g_{ij}(k,k')G_{ij}(k',\omega_n). \tag{8.12}$$

In mean-field approximation the Green's function is explicitly known[16,17]; in matrix form

$$G(k,\omega_n)=(i\omega_n I-\widehat{H}(k))^{-1}, \tag{8.13}$$

where $\widehat{H}(k)$ is the 8×8 representation of (8.2), $\widehat{H}(k)_{ij}\equiv m_{ij}(k)$. Thus Eq. (8.12) becomes

$$m_{ij}(k)=\frac{1}{\beta}\sum_n\int\frac{d^3k'}{(2\pi)^3}g_{ij}(k,k')\mathrm{tr}[e_{ji}G(k,\omega_n)], \tag{8.14}$$

where e_{ij} is the same matrix as was introduced in Sec. II. A slightly more conventional form of (8.14) is obtained by using the Hamiltonian (8.1) in the triple-Nambu representation (2.9), thus—for simplicity—taking $g_{ij}(k,k')=-g$ $(i\neq j)$ independent of k,k'

$$m_{abc}=-\frac{g}{\beta}\sum_n\int\frac{d^3k}{(2\pi)^3}\mathrm{tr}[(\tau_a\times\tau_b\times\tau_c)G(k,\omega_n)], \tag{8.15}$$

where we have written $\widehat{H}=\Sigma m_{abc}\tau_a\times\tau_b\times\tau_c$. Thus, for example, using (2.8) and (3.2) to determine the coefficient

$\Delta_0^{(1)}$ of $D_0^{(1)}$ in (8.1), and taking $\Delta = \Delta_0 = \Delta_0'$ real in (6.2) we have

$$m_{301} = -\Delta = \frac{-g}{\beta} \sum_n \int \frac{d^3k}{(2\pi)^3} \times \mathrm{tr}[(\tau_3 \times \tau_0 \times \tau_1) G(k,\omega_n)],$$

a self-consistent equation for the singlet superconducting gap in this formalism.

IX. CONCLUSIONS

Starting with the simple model Hamiltonian (2.1) of Sec. II, we are led by algebraic closure of the operators therein, to the general Hamiltonian of (8.1). This new system includes many new phenomena not present in the original system, involving as it does 63 parameters against the original 14. Two questions concerning the algebraically generated operators arise naturally: (i) Is it really necessary to include them in the theory?; and (ii) Do they give rise to physically observable phenomena? The answer to (i) is that even if the new operators are not present in the original Hamiltonian, they will be generated by the time evolution of the dynamics acting on the operators already present; and so they must be included for completeness. The physical detection of the corresponding order parameters will depend on their not vanishing in the ground state of the system; this requires diagonalization of the Hamiltonian. This calculation has been carried out for a simplified so(4) ⊗ so(4) version[3] of the complete su(8) model, where it was found that a new operator (triplet $Q = 0$ superconductor) *not* present in the *original* Hamiltonian.[18] These questions may also be examined by conventional self-consistent methods; and we sketched this approach in Sec. VIII.

ACKNOWLEDGMENTS

The authors would like to thank Professor Mario Rasetti and the Institute for Scientific Interchange at Torino for the hospitality extended to them while this work was completed, and Professor M. Berger and Professor L. Michel for the hospitality of the Institut des Hautes Etudes Scientifiques, Bures, where some of the earlier work was done.

Partial support at City College by the PSC-BHE Faculty Research Award Program is acknowledged. We also acknowledge partial support of this work by NATO Grant No. RG.86/0223.

[1] R. Sooryakumar and M. V. Klein, Phys. Rev. Lett. **45**, 660 (1980); Phys. Rev. B **23**, 3213 (1981).

[2] C. Y. Huang, Crit. Rev. Solid State Mat. Sci. **12**, 78 (1984); M. Ishikawa, Contemp. Phys. **23**, 442 (1982).

[3] A. I. Solomon and J. L. Birman, "Mechanism for triplet superconductivity" (to be published).

[4] J. L. Birman and A. I. Solomon, Phys. Rev. Lett. **49**, 230 (1982).

[5] J. L. Birman and A. I. Solomon, "State labels and selection rules in superconductors and charge-density-wave superconductors in the dynamical algebra," in *Proceedings of XIIIth International Coloquium on Group Theoretical Methods in Physics*, edited by W. W. Zachary (World Scientific, Singapore, 1985), p. 445.

[6] For a review of some of these systems, see J. L. Birman and A. I. Solomon, "Spectrum generating algebras in condensed matter Physics," in *Spectrum Generating Algebras in Physics*, edited by A. Bohm and Y. Ne'eman (World Scientific, Singapore, 1986).

[7] A. I. Solomon and J. L. Birman, Phys. Lett. A **104**, 235 (1984).

[8] A. I. Solomon, *Differential Geometric Methods in Mathematical Physics*, edited by S. Sternberg (Reidel, Dordrecht, 1984), p. 279.

[9] L. O'Raifeartaigh, Rep. Prog. Phys. **42**, 159 (1979).

[10] E. W. Fenton and G. C. Psaltakis, Solid Sate Commun. **45**, 5 (1982); see also G. C. Psaltakis and E. W. Fenton, J. Phys. C **16**, 3913 (1983); E. W. Fenton, Prog. Theor. Phys. Suppl. **80**, 94 (1984).

[11] A. I. Solomon, Ann. NY Acad. Sci. **337**, 161 (1980).

[12] A. I. Solomon, J. Phys. A **14**, 2177 (1981).

[13] Y. Hasegawa, T. Usugawa, and F. Iwamoto, Prog. Theor. Phys. **62**, 1458 (1979).

[14] A. I. Solomon and J. L. Birman, Phys. Lett. A **88**, 413 (1982).

[15] T. Matsubara, Prog. Theor. Phys. **14**, 351 (1955); A. L. Fetter and J. D. Walecka, *Quantum Theory of Many Particle Systems* (McGraw-Hill, New York, 1971).

[16] J. L. Birman and A. I. Solomon, Prog. Theor. Phys. (Kyoto), Suppl. **80**, 62 (1984).

[17] A. I. Solomon and J. L. Birman, Phys. Lett. A **111**, 423 (1985).

[18] Another simple model is examined in papers by A. Solomon and J. L. Birman, "Discrete symmetries and selection rules in unified su(8) for superconductivity and density waves," in *Proceedings of the XVth International Colloquim on Group Theory in Physics*, edited by R. Gilmore and D. H. Feng (World Scientific, Singapore, 1987).

Separability of the Killing–Maxwell system underlying the generalized angular momentum constant in the Kerr–Newman black hole metrics

B. Carter
Départment d'Astrophysique Relativiste et Cosmologie, Centre National de la Recherche Scientifique, Observatoire de Paris, 92 Meudon, France

(Received 24 December 1986; accepted for publication 25 February 1987)

The concept of a Killing–Maxwell system may be defined by the relation $\hat{A}_{[\mu;\nu];\rho} = (4\pi/3)\hat{j}_{[\mu}g_{\nu]}$. In such a system the one-form $\hat{A}_\mu$ is interpretable as the four-potential of an electromagnetic field $\hat{F}_{\mu\nu}$, whose source current $\hat{j}^\mu$ is an ordinary Killing vector. Such a system determines a canonically associated duality class of source-free electromagnetic fields, its own dual being a Killing–Yano tensor, such as was found by Penrose [Ann. N.Y. Acad. Sci. **224**, 125 (1973)] (with Floyd) to underlie the generalized angular momentum conservation law in the Kerr black hole metrics, the existence of the Killing–Yano tensor being also a sufficient condition for that of the Killing–Maxwell system. In the Kerr pure vacuum metric and more generally in the Kerr–Newman metrics for which a member of the associated family of source-free fields is coupled in gravitationally, it is shown that the gauge of the Killing–Maxwell one-form may be chosen so that it is expressible (in the standard Boyer–Lindquist coordinates) by $\frac{1}{2}(a^2\cos^2\theta - r^2)dt + \frac{1}{2}a(r^2 - a^2)\sin^2\theta\, d\phi$, the corresponding source current being just $(4\pi/3)(\partial/\partial t)$. It is found that this one-form (like that of the standard four-potential for the associated source-free field) satisfies the special requirement for separability of the corresponding coupled charged (scalar or Dirac spinor) wave equations.

I. INTRODUCTION

Although it is well known that the charged black hole uniqueness and no hair theorems[1–7] allow only two electromagnetic degrees of freedom (or just one if a magnetic monopole moment is deemed to be physically unrealistic) for regular electromagnetic perturbations that are *source-free* and *asymptotically vanishing*, the dropping of these latter restrictions permits one to envisage many other possibilities. Among these, one particular example is specially singled out (if not for any obvious astrophysical relevance, at least for its remarkable mathematical properties), namely what we shall refer to as the *Killing–Maxwell* field. It is demonstrated in this paper that if this field is taken seriously, in the sense of being considered to act in the usual way on charged scalar or spinor fields and discrete classical particles on the black hole background, then the resulting coupled systems have the same kind of very special separability properties as have already been found, respectively,[8–10] when such charged fields and particles are coupled to the familiar *source-free* electromagnetic perturbations allowed by the no hair theorems.

The existence of a Killing–Maxwell system in the sense to be defined below is an equivalent (necessary and sufficient) condition to the existence—in four dimensions—of a *second degree Killing–Yano* tensor,

$$f_{\lambda\mu} = f_{[\lambda\mu]}, \quad f_{\lambda(\mu;\rho)} = 0 \tag{1.1}$$

(using a semicolon for covariant differentiation, with square and round brackets for symmetrization and antisymmetrization of tensor indices). It was the culmination of a systematic attempt (using two-spinor methods) by several coworkers[11–14] to obtain (from the Weyl tensor degeneracy property that was the basis of Kerr's original discovery of his metric[15]) a simple underlying reason for the remarkable integrability properties of so many kinds of systems in the Kerr (and Kerr–Newman[16]) black hole metrics[8,10,17–21] that the existence of such a tensor in these metrics was first brought to light by Penrose[22] (with Floyd). Much further work (including the use of a Debever-type bivector formalism[23] for transcription of earlier two-spinor results into equivalent but more widely readable tensorial form) has explored the general properties of such systems, essentially confirming that the remarkable properties just referred to can indeed be considered as automatic consequences of (1.1). A recent summary and guide to many relevant references, of which only a sample can be mentioned here,[24–30] has been given by Kamran and Marck.[31] This body of work together with earlier results[8] soon made it clear that the existence of a (nonzero) solution of (1.1) is by itself sufficient to characterize the Kerr (or Kerr–Newman) solution uniquely among asymptotically flat pure vacuum Einstein (or source-free Einstein–Maxwell) solutions (and likewise for the author's asymptotically de Sitter black hole solutions [2,8,32]—though it remains a teasing mystery why the solutions of the (global) black hole problem should turn out to belong to this (locally) privileged class.

We start by collecting some essential conclusions that can be drawn directly from (1.1) (without recourse to Einstein or any other equations) by straightforward tensor analysis. Among the most basic of these results is the existence of an ordinary (symmetric) *Killing tensor* (whose presence in the case of the Kerr solutions was directly implied by the original discovery[10] of a quadratic generalized angular momentum constant of the motion)

$$a_{\lambda\mu} = f_\lambda{}^\rho f_{\rho\mu}, \quad a_{(\lambda\mu;\rho)} = 0, \tag{1.2}$$

together with the existence of what we shall refer to as *the primary and the secondary killing vector* (giving rise to linear constants of motion, interpretable as linear combinations of energy and axial angular momentum), the first defined (using the alternating tensor) as the dual of the (necessarily antisymmetric) covariant derivative of the Killing–Yano tensor,

$$k^{\lambda} = (1/3!)\epsilon^{\lambda\mu\rho\sigma} f_{\mu\rho;\sigma}, \quad k_{(\lambda;\mu)} = 0, \tag{1.3}$$

and the second given in terms of the first by

$$h^{\lambda} = a^{\lambda\rho} k_{\rho}, \quad h_{(\lambda;\mu)} = 0. \tag{1.4}$$

Furthermore, as well as having the Killing vector property of generating symmetries of the metric, $g_{\lambda\mu}$, these two vector fields also generate symmetries of the Killing–Yano tensor itself (and hence of the system as a whole, which entails in particular that they must commute) in the sense that the Lie derivatives of the Killing–Yano tensor (and hence of anything constructed directly from it) with respect to each of these vectors must vanish. (The primary Killing vector has the additional special property that the corresponding covariant derivatives along it must also vanish.)

II. THE CONCEPT OF A KILLING–MAXWELL SYSTEM

The basic defining equation of what we refer to henceforth as a *Killing–Maxwell system* may be taken to be

$$\hat{A}_{[\mu;\nu];\rho} = (4\pi/3)\hat{j}_{[\mu} g_{\nu]\rho}, \tag{2.1}$$

where $\hat{A}_{\mu}$ is a four-potential one-form, associated with a four-current vector $\hat{j}^{\mu}$, and $g_{\mu\nu}$ is the metric of the background space-time (and where we have introduced a circumflex to distinguish quantities pertaining to the Killing–Maxwell field from the analogous quantities pertaining to the closely related source-free Maxwell field to be mentioned below). Such a system evidently satisfies the (much less highly restrictive) ordinary Maxwell equations for the corresponding electromagnetic field tensor

$$\hat{F}_{\mu\nu} = 2\hat{A}_{[\nu;\mu]} \tag{2.2}$$

since the contraction of (2.1) leads directly to the source equation

$$\hat{F}^{\rho\mu}{}_{;\rho} = 4\pi\hat{j}^{\mu}. \tag{2.3}$$

By straightforward tensor algebra and the use of the Maxwell–Faraday integrability condition for (2.2),

$$\hat{F}_{[\mu\nu;\rho]} = 0, \tag{2.4}$$

it can easily be checked that the systems (1.1) and (2.1) are equivalent (modulo gauge transformations in the latter) since one can be constructed from the other and vice versa by the simple duality relation

$$f_{\mu\nu} = {}^{*}\hat{F}_{\mu\nu} \equiv \tfrac{1}{2}\epsilon_{\mu\nu}{}^{\rho\sigma}\hat{F}_{\rho\sigma}, \tag{2.5}$$

which evidently entails that the current is to be identified, modulo a rationalization factor, with the primary Killing vector:

$$k^{\mu} = (4\pi/3)\hat{j}^{\mu}. \tag{2.6}$$

For many purposes it is convenient to work with the corresponding complex self-dual Killing–Maxwell Yano tensor

$${}^{+}\hat{F}_{\lambda\mu} = \hat{F}_{\lambda\mu} + i f_{\lambda\mu}. \tag{2.7}$$

Using the fact that by (2.1) its contraction with the primary Killing vector is a pure gradient,

$$k^{\rho}\,{}^{+}\hat{F}_{\rho\mu} = -\tfrac{1}{8}({}^{+}\hat{F}_{\rho\sigma}\,{}^{+}\hat{F}^{\sigma\rho})_{,\mu}, \tag{2.8}$$

it is straightforward to check that one can construct a new (complex) proportionally related self-dual field of the form

$${}^{+}F_{\lambda\mu} = Cs^{-3}\,{}^{+}\hat{F}_{\lambda\mu}, \tag{2.9}$$

which will satisfy the full set of *source-free* Maxwell equations, whose complex form is

$${}^{+}F^{\mu\rho}{}_{;\rho} = 0,$$

for an arbitrary value of the complex (charge) constant C, provided that s is taken to be the scalar field given in terms of the scalar invariants of the Killing–Maxwell field by

$$4s^2 = {}^{+}\hat{F}_{\rho\sigma}\,{}^{+}\hat{F}^{\sigma\rho}. \tag{2.10}$$

The fact, mentioned above, that the Killing–Yano tensor and hence also the Killing–Maxwell system (not to mention the associated source-free field that has just been constructed) will be invariant under the action generated by the primary Killing vector can at this stage be seen directly by combining the gradient property (2.8) with the condition [obtained by contracting the Killing vector with the dual of its defining relation (1.3)], which leads to a pair of equations

$$\hat{F}_{\lambda\mu;\rho}k^{\rho} = 0, \quad 2\hat{F}_{\rho[\lambda}\,k_{\mu]}{}^{;\rho} = 0 \tag{2.11}$$

which add up to the condition that the Lie derivative with respect to k^{μ} vanishes. For the secondary Killing vector h^{μ}, we do not have analogs of the separate equations (2.11) but we can nevertheless obtain the combination expressing the corresponding invariance condition,

$$\hat{F}_{\lambda\mu;\rho}h^{\rho} + 2\hat{F}_{\rho[\lambda}\,h_{\mu]}{}^{;\rho} = 0, \tag{2.12}$$

from (2.4), using the fact that the imaginary (magnetic) part of (2.8) implies a corresponding real (electric but not magnetic) gradient property for the effective electric (but not the magnetic) field as defined with respect to the secondary Killing vector:

$$h^{\rho}\hat{F}_{\rho\mu} = -\tfrac{1}{32}\{(\hat{F}_{\rho\sigma}f^{\sigma\rho})^2\}_{,\mu}. \tag{2.13}$$

We can use (2.12) together with (1.3) to see that the *Killing bivector* $2k^{[\lambda}h^{\mu]}$ has a dual two-form given by

$$\epsilon_{\lambda\mu\rho\sigma}k^{\rho}h^{\sigma} = 2f_{\rho[\lambda}\,h_{\mu]}{}^{;\rho}, \tag{2.14}$$

which enables us to derive the equations

$$h_{[\lambda;\mu}k_{\rho}h_{\sigma]} = 0, \quad k_{[\lambda;\mu}k_{\rho}h_{\sigma]} = 0, \tag{2.15}$$

of which the first is an obvious consequence directly of (2.15), while the second can be obtained from (2.11), which evidently entails a formally identical pair of equations with $f_{\lambda\mu}$ in place of $\hat{F}_{\lambda\mu}$. The same considerations also, respectively, imply

$$k^{\rho}h^{\sigma}f_{\rho\sigma} = 0, \quad k^{\rho}h^{\sigma}\hat{F}f_{\rho\sigma} = 0. \tag{2.16}$$

It can be seen that (2.15) and (2.16) are the same *circularity conditions* as those deduced from the generalized Papapetrou theorem in the black hole problem[2,33] from quite a different starting point (involving Einstein curvature equations and global boundary conditions) instead of the very simple equations (1.1) or equivalently (2.1), which is all that we have assumed here. In particular (2.15) is interpretable as the Frobenius integrability condition for the two-surface elements orthogonal to the Killing bivector to be themselves two-surface forming.

III. THE KILLING–MAXWELL ONE-FORM

We have so far mainly been collecting results that (although rather dispersed about the literature quoted above, and derived by perhaps more devious routes and in more specialized notation than the ordinary tensor calculus used here) are nevertheless for the most part, in principle, "well," albeit not "widely," known by now. However, we shall now concentrate our attention on what is in a sense the most fundamental element of all in the foregoing tree of relationships, which does not yet seem to have had the attention it deserves (or even to have been considered explicitly at all), namely what we have dubbed as the Killing–Maxwell one-form, $\hat{A}_\mu$. Once it has been specified (assuming that the metric tensor is also known) all the other quantities can be constructed by successive differentiations (the Killing–Yano tensor at first order, the ordinary Killing vectors at second order, and so on.) One reason for the neglect of the zero-order element at the base of the tree may be that to make it explicit one must, of course, make some specific choice of the gauge. In practice, however, there is no real ambiguity because there turns out to be a *canonical gauge* that imposes itself naturally (just as I found long ago[8,10] to be the case for what can now be interpreted as the canonically associated source-free fields).

To pin down the gauge we start by requiring that the four-potential one-form $\hat{A}_\mu$ should have the same properties of invariance under the action of the Killing vectors as the field $\hat{F}_{\lambda\mu}$ itself, properties which are simultaneously compatible in consequence of the commutation relation

$$h^\mu{}_\rho k^\rho - k^\mu{}_\rho h^\rho = 0 \tag{3.1}$$

that follows from the fact that the secondary Killing vector is constructed from quantities known [by (2.11)] to be invariant under the action of the primary Killing vector. We can thus obtain

$$\hat{A}_{\mu;\rho} k^\rho + \hat{A}_\rho k^\rho{}_{;\mu} = 0, \tag{3.2}$$

and

$$\hat{A}_{\mu;\rho} h^\rho + \hat{A}_\rho h^\rho{}_{;\mu} = 0. \tag{3.3}$$

Using the real (electric) part of (2.8) we see from (3.2) that it is possible by a further minor adjustment to arrange to have

$$\hat{A}_\rho k^\rho = -\tfrac{1}{4}\hat{F}_{\rho\sigma}\hat{F}^{\sigma\rho}, \tag{3.4}$$

while similarly, by (2.13) [again bearing in mind the compatibility property (3.1)], we see from (3.3) that it is possible also to arrange to have

$$\hat{A}_\rho h^\rho = -\tfrac{1}{32}(\hat{F}_{\rho\sigma} f^{\sigma\rho})^2. \tag{3.5}$$

Finally, leaving aside the possibility of degenerate limit cases in which the primary and secondary Killing vectors might not be independent, it can be seen that (as in the analogous stage in the black hole problem[2]) the orthogonal transitivity and field circularity properties, (2.15) and (2.16), allow us to impose the gauge circularity condition

$$\hat{A}_{[\lambda}\, k_\mu h_{\rho]} = 0, \tag{3.6}$$

which now ties down the gauge completely. Although there is now no longer any freedom to impose further gauge restrictions, it is apparent that (3.2), (3.3), and (3.6) together are sufficient to ensure automatically that the standard Lorentz gauge condition

$$\hat{A}_\rho{}^{;\rho} = 0 \tag{3.7}$$

is also satisfied.

IV. ALGEBRAICALLY PREFERRED COORDINATES AND SEPARABILITY

Up to this stage we have kept to fully covariant terminology, but it is now useful (at the price of leaving aside degenerate limit cases in which the two Maxwellian scalar invariants do not vary independently) to bring in *algebraically preferred* coordinates of the kind introduced by the present author[8] and commonly used in studies of the general cases[26,29,30] (as opposed to the more particular physical black hole case, for which the slightly different geometrically preferred coordinates of the type introduced by Boyer and Lindquist[34] are usually chosen). Within the present approach the algebraically preferred system may be specified to consist of two nonignorable coordinates, r and q say, given in terms of the Killing–Maxwell invariants by

$$r^2 - q^2 = \tfrac{1}{2}F_{\rho\sigma}F^{\rho\sigma}, \quad 2rq = \tfrac{1}{2}\hat{F}_{\rho\sigma}f^{\sigma\rho}, \tag{4.1}$$

together with two ignorable coordinates, $\tilde{t}$ and $\tilde{\phi}$ say, taken to be constant on the orthogonal hypersurfaces whose existence is established by (2.15) and such that the primary and secondary Killing vectors, k^μ and h^μ, can be identified, respectively, with the operators $\partial/\partial\tilde{t}$ and $\partial/\partial\tilde{\phi}$. It can be seen that the specification (4.1) is satisfied simply by taking r and q as the real and imaginary parts of the scalar field defined by (2.9), i.e., we have

$$s = r + iq. \tag{4.2}$$

In this system the gauge conditions imposed at the end of the previous section lead unambiguously to the explicit expression

$$\hat{A}_\rho\ dx^\rho = \tfrac{1}{2}(q^2 - r^2)d\tilde{t} - \tfrac{1}{2}r^2q^2\ d\tilde{\phi}. \tag{4.3}$$

Nothing in the preceding line of reasoning makes it obvious in advance that this field should share the already known property of the associated source-free Maxwell field of satisfying the author's condition[2,8] for separability of the Klein–Gordon wave equation (and hence *a fortiori* the corresponding classical charged orbit equations) for a charged scalar field coupled to an electromagnetic field. In the present terminology this very restrictive condition is expressible as the requirement that the four-potential one-form should have the form

$$\hat{A}_\rho\ dx^\rho = \frac{\hat{X}_+(r)(d\tilde{t} + q^2 d\tilde{\phi}) - \hat{X}_-(q)(d\tilde{t} - r^2\ d\tilde{\phi})}{r^2 + q^2}, \tag{4.4}$$

where $\hat{X}_+(r)$ is a function of r only, and $\hat{X}_-(q)$ is a function of q only. It transpires nevertheless that in the gauge (4.3) the Killing–Maxwell one-form does indeed satisfy this condition, the two single variable functions having the simplest form imaginable on dimensional grounds, namely

$$\hat{X}_+(r) = \tfrac{1}{2}r^4, \quad \hat{X}_-(q) = \tfrac{1}{2}q^4. \tag{4.5}$$

For comparison, it may be recalled that the analogous functions for the family (2.9) of source-free associated fields (including those coupled gravitationally in the Kerr–Newman[16] solutions) are correspondingly expressible[2,8,10] in terms of the real (electric charge) part Q and the imaginary (magnetic monopole) part P of the complex charge parameter C appearing in (2.8) by

$$X_+(r) = Qr, \quad X_-(q) = Pq \quad (P + iQ = C). \tag{4.6}$$

The significance of the property of being expressible in the form (4.4) is strengthened by the recent work of Kamran and McLenaghan,[28] which shows that the condition (4.4) is sufficient to ensure (undecoupled Chandrasekhar-type[19]) separability in the case where the charged scalar is replaced by a charged Dirac spinor. Although such separability properties can be studied more easily in the algebraically preferred coordinates used here, they are, of course, preserved by the transformation to the standard geometrically preferred Boyer–Lindquist[34] coordinates according to the prescription

$$r \to r, \ q \to a \cos\theta, \ \tilde{\phi} \to a^{-1}\phi, \ \tilde{t} \to t - a\phi. \tag{4.7}$$

(I would insist, by the way, that contrary to a widespread myth that has been implicitly perpetuated by a recent major treatise on the subject[35] the transformation to Boyer–Lindquist coordinates does *not* imply any need to transform to a noncanonical—e.g. Kinnersley-type[36] —tetrad in place of the maximally symmetric one.[8,29,37–40]) It is also to be remarked that the separability condition (4.4) is preserved by the trivial gauge changes corresponding to addition of *constant* multiples of $d\tilde{t}$ and $d\tilde{\phi}$. The Boyer–Lindquist form of the Killing–Maxwell potential quoted in the abstract does in fact differ from (4.3) by such a separability-preserving adjustment.

Despite the fact that the corresponding constants of the motion could have been constructed in advance as eigenfunctions of corresponding operators in both the scalar[41] and Dirac spinor[25] cases, the fact that these constants are associated with full separability still seems somewhat miraculous. In the simplest case, that of a classical particle with charge to mass ratio e/m on an orbit whose unit tangent vector u^μ evolves according to

$$u^\mu{}_{;\rho} u^\rho = (e/m) F^\mu{}_\rho u^\rho, \tag{4.8}$$

our original postulate (1.1) implies that the generalized (specific) angular momentum vector and scalar, defined by

$$l^\mu = f^\mu{}_\rho u\rho, \quad l^\rho l_\rho = a_{\rho\sigma} u^\rho u^\sigma, \tag{4.9}$$

will satisfy corresponding precessing translation and conservation laws,

$$l^\mu{}_{;\rho} u^\rho = (e/m) F^\mu{}_\rho l^\rho, \quad (l^\rho l_\rho)_{,\sigma} u^\sigma = 0, \tag{4.10}$$

for *any* field $F_{\lambda\mu}$ given by an expression of the form (2.9) whatever the field s may be. Now although any field satisfying the separability condition (4.4) will have the form (2.9) for some scalar field s, the converse requirement is highly restrictive.[1] It is therefore remarkable that such a requirement (which in this case is manifestly not necessary for the conservation law to apply) should turn out to hold both for the source-free solutions [with s given by (2.10) or (4.3)] and for the Killing–Maxwell field (with s uniform so that $Cs^{-3} = 1$). Indeed even in the Minkowski space limit, for which the Killing–Maxwell field is interpretable as that within a uniform spherical charge distribution, the spherical symmetry of which is *broken* by the superposition of a uniform magnetic field, the (scalar and Dirac) separability that has been revealed was hardly obvious in advance.

ACKNOWLEDGMENTS

The author would like to thank D. Lynden-Bell, J-A. Marck, R. G. McLenaghan, and R. Penrose for stimulating and informative conversations on various recent and past occasions.

[1] W. Israel, Commun. Math. Phys. **8**, 1776 (1968).
[2] B. Carter, in *Black Holes*, Les Houches, 1972, edited by B. V. and C. De Witt (Gordon and Breach, New York, 1973).
[3] D. C. Robinson, Phys. Rev. D **10**, 458 (1974).
[4] B. Carter, in *General Relativity*, edited by S. W. Hawking and W. Israel (Cambridge U.P., Cambridge, 1979).
[5] G. Bunting, Ph.D. thesis, University of New England, Armidale, New South Wales, 1963.
[6] P.O. Mazur, J. Phys. A **15**, 3173 (1962).
[7] B. Carter, Commun. Math. Phys. **99**, 563 (1985).
[8] B. Carter, Commun. Math. Phys. **10**, 280 (1968).
[9] D. N. Page, Phys. Rev. D **14**, 1509 (1976).
[10] B. Carter, Phys. Rev. **174**, 1559 (1968).
[11] M. Walker and R. Penrose, Commun. Math. Phys. **18**, 265 (1970).
[12] L. P. Hughston, R. Penrose, P. Sommers, and M. Walker, Commun. Math. Phys. **27**, 303 (1972).
[13] L. P. Hughston and P. Sommers, Commun. Math. Phys. **32**, 147 (1973).
[14] L. P. Hughston and P. Sommers, Commun. Math. Phys. **33**, 129 (1973).
[15] R. P. Kerr, Phys. Rev. Lett. **11**, 238 (1963).
[16] E. T. Newman, E. Crouch, R. Chinapared, A. Exton, A. Prakash, and R. Torrence, J. Math. Phys. **6**, 918 (1965).
[17] W. Unruh, Phys. Rev. Lett. **31**, 1265 (1973).
[18] S. A. Teukolsky, Astrophys. J. **185**, 283 (1973).
[19] S. Chandrasekhar, Proc. R. Soc. London Ser. A **349**, 571 (1976).
[20] R. Guven, Phys. Rev. D **22**, 2327 (1980).
[21] J-A. Marck, Proc. R. Soc. London Ser. A **385**, 431 (1983).
[22] R. Penrose, Ann. N.Y. Acad. Sci. **224**, 125 (1973).
[23] R. Debever, Bull. Cl. Sci. Acad. Belg. **LXII**, 662 (1976).
[24] C. D. Collinson, Int. J. Theor. Phys. **15**, 311 (1976).
[25] R. Cattenacci and F. Salmestraro, J. Math. Phys. **19**, 2047 (1978).
[26] B. Carter and R. G. McLenaghan, Phys. Rev. D **19**, 1093 (1979).
[27] W. Dietz and R. Rudiger, Proc. R. Soc. London Ser. A **375**, 361 (1981).
[28] N. Kamran and R. G. McLenaghan, J. Math. Phys. **25**, 1019 (1984).
[29] R. Debever, N. Kamran, and R. G. McLenaghan, J. Math. Phys. **25**, 1955 (1984).
[30] N. Kamran and R. G. McLenaghan, Bull. Cl. Sci. Acad. Belg. **LXX**, 596 (1984).
[31] N. Kamran and J-A. Marck, J. Math. Phys. **27**, 1589 (1986).
[32] G. Gibbons and S. W. Hawking, Phys. Rev. D **15**, 2738 (1977).
[33] B. Carter, J. Math. Phys. **10**, 70 (1969).
[34] R. H. Boyer and R. W. Lindquist, J. Math. Phys. **8**, 265 (1966).
[35] S. Chandrasekhar, *The Mathematical Theory of Black Holes* (Clarendon, Oxford, 1983).
[36] W. Kinnersley, J. Math. Phys. **10**, 1195 (1969).
[37] R. Debever, Bull. Soc. Math. Belg. **XXIII**, 360 (1971).
[38] R. L. Znajek, Mon. Not. R. Astron. Soc. **179**, 457 (1977).
[39] B. Carter and R. G. McLenaghan, in *Recent Developments in General Relativity*, edited by R. Ruffini (North-Holland, Amsterdam, 1982).
[40] R. Debever, R. G. McLenaghan, and N. Tarig, Gen. Relativ. Gravit. **10**, 853 (1979).
[41] B. Carter, Phys. Rev. D **16**, 3414 (1977).

A new self-similar space-time

L. K. Chi
Computer Science Department, U. S. Naval Academy, Annapolis, Maryland 21402-5002

(Received 5 March 1986; accepted for publication 25 February 1987)

A new self-similar solution of the Einstein field equations is presented. In the new space-time, the density is zero at time zero and follows an inverse square law for large t. The new solution may have interesting astrophysical applications since it has the same reference lengths as that of the Friedmann universe.

I. INTRODUCTION

Recently, Wesson[1] proposed a dimensional cosmological principle and a dimensional perfect cosmological principle that can lead to both Friedmann and non-Friedmann models. These dimensional principles require that the physical properties of the universe such as density, pressure, and mass appear only in combination with the gravitational constant G, the speed of light c, and the coordinates r and t as dimensionless functions that solely depend on the epoch. These principles were later modified by Chi[2] who showed that the coordinate $r(R,t)$ can be replaced by the comoving coordinate R. Chi also pointed out that the self-similar space-times found by Henriksen and Wesson[3] can be derived from these modified principles by using R/c as the reference time instead of a constant reference time. Self-similarity has wide applications in hydrodynamics.[4] In recent years, self-similarity found many astrophysical applications.[5,6] In this paper, we apply the modified dimensional cosmological principle to Einstein's field equations to find a new self-similar space-time.

II. THEORY

We assume a spherically symmetric metric of the form

$$ds^2 = c^2 e^{\sigma} dt^2 - e^{\omega} dR^2 - r^2(d\theta^2 + \sin^2\theta\, d\phi^2), \quad (1)$$

where R is a comoving radial coordinate, and $r = r(R,t)$ is not comoving. The functions $\sigma(R,t)$ and $\omega(R,t)$ are dimensionless. Let $m(R,t)$ be the total mass inside the radius R, $p(R,t)$ the total pressure, and $\rho(R,t)c^2$ the total energy density. Then the Einstein field equations become[7]

$$2Gm/(c^2 r) = 1 + e^{-\sigma}\dot{r}^2/c^2 - e^{-\omega}r'^2, \quad (2a)$$

$$\dot{m} = 4\pi P r^2 \dot{r}/c^2, \quad (2b)$$

$$m' = +4\pi\rho r^2 r', \quad (2c)$$

$$\sigma' = -2P'/(P + \rho c^2), \quad (2d)$$

$$\dot{\omega} = -2\dot{\rho} c^2/(P + \rho c^2) - 4\dot{r}/r, \quad (2e)$$

where a dot means $\partial/\partial t$ and a prime means $\partial/\partial R$. Following the modified cosmological principle, we require that the properties of the universe be made dimensionless as follows:

$$\begin{aligned} &8\pi G l_\rho^2 \rho/c^2 = N(\xi), \quad 8\pi G l_P^2 P/c^4 = Q(\xi), \\ &2Gm/(c^2 l_m) = M(\xi), \quad r = RS(\xi), \quad \xi = ct/R, \end{aligned} \quad (3)$$

where l_ρ, l_P, and l_m are all functions of R and ct. The reference time in Eq. (3) is chosen to be R/c, since we are looking for self-similar solutions of Eq. (2). The self-similar space-times developed by Henriksen and Wesson were obtained from Eqs. (2) and (3) by choosing $l_\rho = l_P = l_m = R$. Here we choose

$$l_\rho = l_P = ct, \quad l_m = R^3/(c^2t^2). \quad (4)$$

It was shown in Ref. 2 that the choice of reference lengths (4) gives rise to the Friedmann universes. We are therefore interested to know what kind of self-similar space-times can have the same reference lengths (4).

With the choice of Eqs. (3) and (4), Eqs. (2) become

$$M = \xi^2 S[1 + e^{-\sigma}S'^2 - e^{-\omega}(S - \xi S')^2], \quad (5a)$$

$$3M - \xi M' = NS^2(S - \xi S'), \quad (5b)$$

$$M' - 2M/\xi = -QS^2S', \quad (5c)$$

$$\omega' = -2(N' - 2N/\xi)(Q + N) - 4S'/S, \quad (5d)$$

$$\sigma' = 2Q'/(Q + N), \quad (5e)$$

where a prime means $d/d\xi$. For simplicity, we look for the dust solutions of Eqs. (5). Assume $Q = 0$. Then Eq. (5c) gives $M = M_0\xi^2$, where M_0 is a constant. And Eq. (5e) can be integrated to give $e^{-\sigma} = 1$. Equation (5d) implies that $e^{-\omega} = N^2S^4/\xi^4$, provided the integration constant is properly chosen. Substituting $M = M_0\xi^2$ into Eq. (5b), we find that

$$M_0\xi^2 = NS^2(S - \xi S'),$$

and consequently Eqs. (5a) becomes

$$M_0 = S[1 - S'^2 - M_0^2].$$

In short, the dimensionless dust solutions of Eqs. (5) are given by

$$\begin{aligned} &Q = 0, \quad M = M_0\xi^2, \quad e^{-\sigma} = 1, \quad e^{-\omega} = N^2S^4/\xi^4, \\ &SS'^2 + S(1 - M_0^2) - M_0 = 0, \\ &M_0\xi^2 = NS^2(S - \xi S'). \end{aligned} \quad (6)$$

The above solution differs from that found by Henriksen and Wesson in that the dimensionless density $M = M_0\xi^2$ is not constant.

Consider the case $M_0 = 1$. Then we find from Eq. (6) that

$$S = [3(\xi_0 \pm \xi)/2]^{2/3}, \quad (7)$$

and that N satisfies

$$\xi^2 = NS^2(S - \xi S'). \quad (8)$$

Eliminating S in Eqs. (7) and (8), we find that

$$\xi^2 = 9N(\xi_0 \pm \xi)(\xi_0 \pm \xi/3)/4.$$

Consequently,

$$\rho = \frac{c^2}{18\pi G R^2 \xi_0^2} \frac{(ct/R)^2}{[1 \pm (ct/R\xi_0)][1 \pm ct/(3R\xi_0)]}. \quad (9)$$

The solution (6), more specifically the expression (9), has some interesting behavior. For small $|ct/R|$, $\rho \sim \xi^4/t^2$, and ρ equals zero at $t = 0$. For large $|ct/R|$, $\rho \sim (\xi/t)^2$, and ρ has the behavior of an inverse square law. Along the lines $|ct/R| = \text{const}$, the density is inversely proportional to t^2 for large t.

III. CONCLUSION

In conclusion, we demonstrate the usefulness of the new dimensional cosmological principle by using it to find a new self-similar space-time. Since the Friedmann universes are solutions of the Einstein's field equations with the length scales (4), we expect the self-similar solutions having the same length scales to have interesting astrophysical applications.

[1] P. S. Wesson, "A new look at the cosmological principles," Astron. Astrophys. **68**, 131 (1978).

[2] L. K. Chi, "On the dimensional cosmological principles," Astrophys. J. **289**, 443 (1985).

[3] R. N. Henriksen and P. S. Wesson, "Self-similar space-times I: Three solutions," Astrophys. Space Sci. **53**, 429 (1978).

[4] G. I. Barenblatt and Y. B. Zeldovich, "Self-similar solutions as intermediate asymptotics," Annu. Rev. Fluid Mech. **4**, 285 (1972).

[5] R. N. Henriksen and P. S. Wesson, "Self-similar space-times II: Perturbation," Astrophys. Space. Sci. **53**, 445 (1978).

[6] P. S. Wesson, "Improved standard cosmology," Phys. Rev. D **23**, 2137 (1981).

[7] M. A. Podurets, "On one form of Einstein's equations for a spherically symmetrical motion of a continuous medium," Sov. Astron. **8**, 19 (1984).

On the solution of the Tolman–Oppenheimer–Volkov equation with the ultrarelativistic equation of state

Tuomo Toimela
Research Institute for Theoretical Physics, University of Helsinki, Siltavuorenpenger 20 C, 00170 Helsinki, Finland

(Received 31 October 1986; accepted for publication 25 February 1987)

The Tolman–Oppenheimer–Volkov equation is studied in the case of the ultrarelativistic equation of state. The original system of two first-order differential equations is turned into one first-order equation that is independent of the central density, plus an integral. It is shown how the physical solutions are related to the analytically known infinite central density solution. The results are further generalized into the arbitrary γ-law equation of state, $p(\gamma-1)\rho$. Finally, the case of a nonzero bag constant is briefly discussed.

I. INTRODUCTION

The properties of neutron stars have been studied extensively since the pioneering works by Landau[1] and Oppenheimer and Volkov.[2] At the turn of the 1950's the accepted knowledge on the behavior of matter at such high densities was established in the Harrison–Wakano–Wheeler (HWW) equation of state.[3] Since then, the neutron star calculations based on that equation of state have been refined along with the growing understanding of phenomena in particle physics.[4]

In higher densities the γ-law equations of state

$$\rho \sim n^{\gamma}, \quad p=(\gamma-1)\rho \tag{1}$$

(where $1 \leqslant \gamma \leqslant 2$ is the physically meaningful range) have been used as a suitable approximation. If it is inserted in the Tolman–Oppenheimer–Volkov (TOV) equation,[5]

$$p'(r) = -(\rho+p)(pr^3+m/4\pi)/r(r-m/2\pi), \tag{2}$$

where

$$m = 4\pi \int_0^r \rho(r')r'^2\,dr'. \tag{3}$$

It is known in this case that the TOV equation has an analytic solution[6]

$$\begin{aligned}\rho &= 2(\gamma-1)/(\gamma^2+4\gamma-4)r^2\\ &= 3/14r^2, \quad \text{for } \gamma=\tfrac{4}{3}.\end{aligned} \tag{4}$$

Although the matter in any star cannot obey the equations of state of the type given by Eq. (1) without any r dependence in γ, the solution (4) can still be used as a high density limit of the star core. The procedure is then as follows: the equilibrium configuration given by Eq. (4) is smoothly matched with some low energy configuration (for example, the one obtained by the HWW equation of state) at some transition regime. In this way both the γ-law equation of state and the infinite central density solution (4) have their importance. In this paper we shall consider the physical solutions obeying the γ-law equations of state (especially, when $\gamma=\frac{4}{3}$). By "physical" we mean solutions that have finite central density. We shall examine their behavior and show how they are related to the infinite central density solution (4).

II. REFORMULATION OF THE TOV EQUATION

Let us start with the special (but the most interesting) case, $\gamma=\frac{4}{3}$, which corresponds to ideal gas of ultrarelativistic particles. We insert the appropriate equation of state, $p=\rho/3$, to the TOV equation, which then reads

$$\rho'(r) = -4\rho(\rho r^3+3m/4\pi)/3r(r-m/2\pi). \tag{5}$$

If we differentiate Eq. (5) with respect to r and use Eq. (3) to eliminate m from the differentiated equation, we obtain

$$\begin{aligned}&2f(2f+3)r^2f'' - (2f+9)r^2f'^2 - 8f(4f-3)rf'\\ &\quad + 8f^2(14f-3)=0,\end{aligned} \tag{6}$$

where

$$f=\rho(r)r^2. \tag{7}$$

Now we make the substitutions

$$t=\ln r, \quad x(t)=f(r), \tag{8}$$

which transforms Eq. (6) to a form having no explicit r dependence

$$\begin{aligned}&2x(2x+3)x'' - (2x+9)x'^2 - 18x(2x-1)x'\\ &\quad + 8x^2(14x-3)=0.\end{aligned} \tag{9}$$

If we then define

$$x'(t)=y(x), \tag{10}$$

we get an equation for $y(x)$,

$$\begin{aligned}&2x(2x+3)yy' - (2x+9)y^2 - 18x(2x-1)y\\ &\quad + 8x^2(14x-3)=0.\end{aligned} \tag{11}$$

The appropriate initial value for this equation can be obtained by

$$y(x=0) = x'(t)|_{x=0} = rf'(r)|_{r=0} = 0. \tag{12}$$

Suppose the function $y=y(x)$ is solved from Eq. (11). Then the solution of the TOV equation (5) is expressed in the form

$$\ln\frac{r}{\epsilon} = \int_{\rho(0)\epsilon^2}^{\rho(r)r^2} \frac{dx}{y(x)}, \tag{13}$$

where the limit $\epsilon \to 0$ is understood.

By the above manipulation we have turned the original TOV equation, which is essentially a system of two first-order differential equations [Eqs. (5) and (3)], into one

 0022-2488/87/071541-03\$02.50

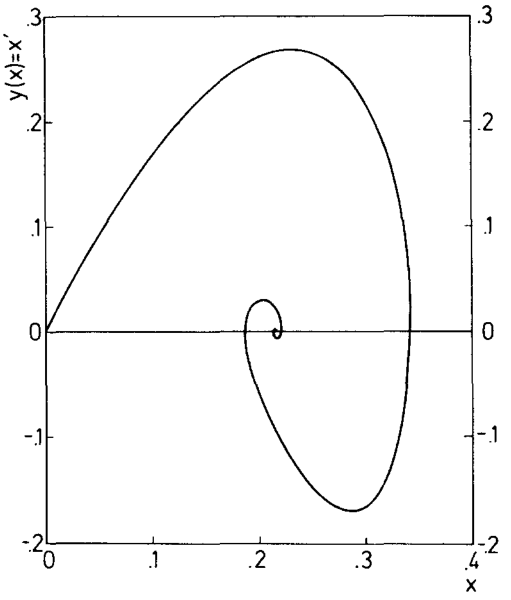

FIG. 1. The numerical solution of the function $y = y(x)$.

first-order equation (11) plus an integral (13). Note that the initial value that fixes the solution of the TOV equation, namely the central density, $\rho(0)$, does not exist in Eq. (11) nor in its boundary value, Eq. (12), but enters only in the integral (13). Thus, the function $y(x)$ is independent of $\rho(0)$ and can be used in Eq. (13), once obtained, for every central density.

It should be noted here that the possibility of expressing in this case the solution of the TOV equation in form (13) is equivalent to the notion that by the substitution (8) the TOV equation can be reduced to a plane autonomous system.[7]

We note from Eq. (11) that it has a trivial solution $x = \frac{3}{14}$, which corresponds to the Misner–Zapolsky (MZ) solution (4) for $\gamma = \frac{4}{3}$. However, if we consider the physical solutions with finite central density, we should use the initial value $y(0) = 0$ obtained by Eq. (12). The numerical solution appropriate for this initial value is shown in Fig. 1. Starting from the origin the curve rapidly spirals to the point $x = \frac{3}{14}$, $y = 0$. For the energy density this means that it approaches the MZ energy density oscillating with decreasing amplitude around it, when $r \to 0$. The behavior of corresponding energy density is shown schematically in Fig. 2, from which we obtain that the energy density configuration continues to infinity. This was expected, since it is compatible with the known fact that gravitation can never produce hydrostatic equilibrium in a *finite* fluid, if the pressure is proportional to the energy density.[8]

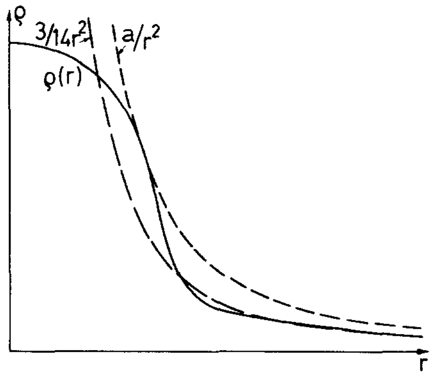

FIG. 2. Schematically drawn behavior of the energy density configuration as a function of the radius. The Misner–Zapolsky solution and the outermost envelope curve are also shown.

From Eq. (11) we find that near the origin $y \sim 2x$. This means that also the right-hand side of Eq. (13) behaves like $\sim -\ln \epsilon$ at small ϵ, which guarantees that the limit $\epsilon \to 0$ is well defined. An interesting fact can be extracted from Fig. 1. The largest value of x is obtained from the point where the curve crosses the x axis for the first time after the origin. The explicit value is $x = a = 0.342$. Because $x \leqslant a$ everywhere, this means $\rho \leqslant a/r^2$ and the mass inside radius r is $m(r) < 4\pi a r$. The curve $\rho = a/r^2$ is drawn also in Fig. 2. The physical solution touches this curve at one point. As a matter of fact, this curve is the envelope of the family of the physical solutions parametrized by the central density. It is worth noting that also all the other points where the curve $y(x)$ intersects the x axis define an envelope. Hence, the family of solutions given by Eq. (13) have an infinite number of envelopes of type $\sim \alpha_n r^{-2}$, the constant of proportionality α_n given by the point where the curve $y(x)$ intersects the x axis for the nth time after the origin ($\alpha_1 = a$ above).

III. GENERALIZATION TO OTHER EQUATIONS OF STATE

We can generalize our procedure into the case, where the equation of state is of the arbitrary γ-law type (1). (This is because our manipulation above is not based on the ultrarelativistic equation of state but on the proportionality of the pressure and energy density.) For an arbitrary γ law, Eq. (6) becomes

$$\begin{aligned}\gamma(\gamma-1)(1+2(\gamma-1)f)fr^2f''&\\ -(\gamma-1)(3\gamma-2+4(\gamma-1)^2f)r^2f'^2&\\ +(5\gamma-4)(\gamma-1)(2+(\gamma-4)f)frf'&\\ -(5\gamma-4)(2(\gamma-1)-(\gamma^2+4\gamma-4)f)f^2=0.&\end{aligned} \tag{14}$$

Moreover, if we use the same substitutions (8) and definition (10) as before, Eq. (11) changes to

$$\begin{aligned}\gamma(\gamma-1)(1+2(\gamma-1)x)xyy'&\\ -(\gamma-1)(3\gamma-2+4(\gamma-1)^2x)y^2&\\ +(\gamma-1)(9\gamma-8+(3\gamma^2-22\gamma+16)x)xy&\\ -(5\gamma-4)(2(\gamma-1)-(\gamma^2+4\gamma-4)x)x^2=0.&\end{aligned} \tag{15}$$

From this equation the general MZ solution, $x = 2(\gamma-1)/(\gamma^2+4\gamma-4)$, can be found immediately. The solutions $y_\gamma(x)$ with the initial value $y_\gamma(0) = 0$, which give the physical solutions of the TOV equation by the integral (13), are qualitatively similar to the $\gamma = \frac{4}{3}$ case, starting from the origin with $y \sim 2x$ and spiraling to the focal point that is given by the MZ equation. Consider the outermost envelope defined by the first intersection of function $y_\gamma(x)$ with the x axis. We find that this point $\alpha_1^{(\gamma)} = a_\gamma$ increases with increasing γ up to $a_2 = 0.352$ (which is only slightly larger than $a_{4/3}$).

As a final note, let us also consider the case of a nonzero bag constant. Unfortunately the MZ solution cannot be generalized to that case. There is, however, one exception. That is, if the Zel'dovich equation of state, $p = \rho$ with a bag constant B, is accepted, we find that the TOV equation has in that case an exact solution

$$\rho = 1/4r^2 + B, \quad p = 1/4r^2 - B. \tag{16}$$

The energy density is infinite in the center of the star and falls to $2B$ at the surface, where $p = 0$. The radius of such a star is

$$R = 1/2\sqrt{B} \tag{17}$$

and the mass

$$m = 2\pi/3\sqrt{B}. \tag{18}$$

If $B^{1/4} = 135$ MeV (in the usual units), the radius and the mass become

$$R \approx 19 \text{ km}, \quad m \approx 4.2 M_\odot. \tag{19}$$

The nontrivial components of the metric of this star are given by

$$g_{rr} = -(1 - m/2\pi r)^{-1} = 2(1 - r^2/3R^2)^{-1}, \tag{20}$$

$$g_{tt} = r^2/3R^2, \tag{21}$$

which match the metric components of the Schwartzschild exterior solution at the surface. The red shift of a photon emitted from the surface of such a star is then

$$z = \sqrt{3} - 1 = 0.732. \tag{22}$$

Equation (16) represents an extreme solution of the TOV equation, as the equation of state of quarks *inside* the bag is the "hardest" possible, $p = \rho$ (corresponding to the sound velocity that equals the velocity of light) and the energy density has an infinite limit at the center.

ACKNOWLEDGMENT

I am indebted to K. Kajantie for discussions.

[1] L. D. Landau, Phys. Z. Sov. **1**, 285 (1932).

[2] J. R. Oppenheimer and G. Volkov, Phys. Rev. **55**, 374 (1939).

[3] B. K. Harrison, K. S. Thorne, M. Wakano, and J. A. Wheeler, *Gravitation Theory and Gravitational Collapse* (Univ. of Chicago P., Chicago, 1965).

[4] See S. L. Shapiro and S. A. Teukolsky, *Black Holes, White Dwarfs and Neutron Stars* (Wiley, New York, 1983), and references quoted therein.

[5] We have used a system of units, where $4\pi G = 1$; i.e., the nonrelativistic gravitation potential of mass M at distance r is $\phi = -M/4\pi r$. Moreover, $c = 1$.

[6] C. W. Misner and H. S. Zapolsky, Phys. Rev. Lett. **12**, 635 (1964).

[7] C. B. Collins, J. Math. Phys. **18**, 1374 (1977); **26**, 2268 (1985); see also B. K. Harrison, Phys. Rev. **137**, B 1644 (1965).

[8] S. Weinberg, *Gravitation and Cosmology* (Wiley, New York, 1972).

Spin fluids in stationary axis-symmetric space-times

J. P. Krisch
Department of Physics, The University of Michigan, Ann Arbor, Michigan 48109

(Received 2 September 1986; accepted for publication 4 February 1987)

The relations establishing the equivalence of an ordinary perfect fluid stress-energy tensor and a spin fluid stress-energy tensor are derived for stationary axis-symmetric space-times in general relativity. Spin fluid sources for the Gödel cosmology and the van Stockum metric are given.

I. INTRODUCTION

The search for new and significant solutions to the field equations has long been an important aspect of general relativity. With the realization that the stress-energy content of a given geometry is not unique, this aspect of general relativity has grown to include the search for new and significant sources to known geometries.

The relation between a viscous-heat conducting fluid and a perfect fluid was derived by King and Ellis[1] in their paper on tilted space-times. The equivalence of electromagnetic fields and some viscous fluids has been discussed by Tupper[2] and Raychaudhuri and Saha.[3] Tupper[4] has also derived the equivalence relations for perfect fluid space-times and space-times with viscous-magnetohydrodynamical matter content. Carot and Ibanez[5] have shown that the interior of a Schwarzschild sphere could contain a viscous heat conducting fluid as well as a simple perfect fluid.

In this paper we extend the possible alternatives to simple perfect fluid sources by considering the stress-energy tensor for a perfect fluid with spin in a stationary axis-symmetric space-time. The metric we treat is

$$ds^2 = -f\,dt^2 - 2k\,d\phi\,dt + l\,d\phi^2 + e^b(dr^2 + dz^2). \tag{1}$$

We do not assume that l is harmonic. This will allow us to discuss the Gödel cosmology. In this space-time a simple perfect fluid has a stress-energy tensor

$$T_{\mu\nu} = \bar{\epsilon} U_\mu U_\nu + \bar{P}(g_{\mu\nu} + U_\mu U_\nu), \tag{2}$$

where $\bar{\epsilon}$ is the energy density, $\bar{P}$ is the pressure, and U_μ is the fluid velocity. We work in the comoving frame, where U_μ is the timelike component of the tetrad $a^\mu_{(i)}$ that diagonalizes the metric:

$$U^\mu = a^\mu_0. \tag{3}$$

The tetrad is

$$\begin{aligned}
&a^\mu_0 = (1/\sqrt{f},0,0,0), \quad a^\mu_0 = (\sqrt{f},0,0,k/\sqrt{f}),\\
&a^\mu_1 = (0,e^{-b/2},0,0), \quad a^1_\mu = (0,e^{b/2},0,0),\\
&a^\mu_2 = (0,0,e^{-b/2},0), \quad a^2_\mu = (0,0,e^{b/2},0),\\
&a^\mu_3 = (-k/D\sqrt{f},0,0,\sqrt{f}/D), \quad a^3_\mu = (0,0,0,D/\sqrt{f}),
\end{aligned} \tag{4}$$

where $D^2 = fl + k^2$. The tetrad indices are in parentheses or are numerical indices. We use coordinate labels, i.e., (t,x,y,z) for the space-time indices. Tetrad indices are raised and lowered by $\eta_{ij} = (-1,+1,+1,+1)$.

The general stress-energy tensor for a spin fluid was given by Ray and Smalley[6,7] and has two parts:

$$T_{\mu\nu} = T_{\mu\nu}(\text{fluid}) + T_{\mu\nu}(\text{spin}). \tag{5}$$

The fluid portion of Eq. (5) is the spin-fluid counterpart of Eq. (1):

$$T_{\mu\nu}(\text{fluid}) = \epsilon U_\mu U_\nu + P(g_{\mu\nu} + U_\mu U_\nu). \tag{6}$$

The spin contribution to the stress-energy tensor is

$$\begin{aligned}
T_{\mu\nu}(\text{spin}) = {} & U_{(\mu}S_{\nu)\alpha}\dot{U}^\alpha + [U_{(\nu}S_{\mu)\alpha}]_{;\alpha}\\
& + \omega_{\mu(\alpha}S_{\nu)\alpha} + U_{(\mu}S_{\nu)\alpha}\omega^{\alpha\beta}U_\beta,
\end{aligned} \tag{7}$$

where $\dot{U}_\mu = U_{\mu;\nu}U^\nu$ is the acceleration of the fluid and $\omega_{\mu\nu}$ is the angular velocity tensor associated with the spin. This angular velocity is defined in terms of the tetrad given in Eq. (4):

$$\omega_{\mu\nu} = \tfrac{1}{2}[\dot{a}^{(\alpha)}_\mu a_{(\alpha)\nu} - \dot{a}^{(\alpha)}_\nu a_{(\alpha)\mu}]. \tag{8}$$

The spin density obeys the Weyssenhoff condition

$$U^\mu S_{\mu\nu} = 0. \tag{9}$$

In Sec. II we derive the equations that establish the equivalence between the sources described by Eqs. (1) and (5). Some metric applications are given in Sec. III.

II. EQUIVALENCE RELATIONS

Equating the stress-energy tensors in Eqs. (1) and (5), we find

$$\begin{aligned}
&\bar{\epsilon} U_\mu U_\nu + \bar{P}(g_{\mu\nu} + U_\mu U_\nu)\\
&\quad = \epsilon U_\mu U_\nu + P(g_{\mu\nu} + U_\mu U_\nu) + (U_\mu S_{\nu l} + U_\nu S_{\mu l})\dot{U}^l\\
&\qquad + \tfrac{1}{2}(U_\mu W_\nu + U_\nu W_\mu) + \tfrac{1}{2}(S_\mu{}^\alpha U_{\nu;\alpha} + S_\nu{}^\alpha U_{\mu;\alpha})\\
&\qquad + \tfrac{1}{2}(S_\mu{}^\alpha \omega_{\nu\alpha} + S_\nu{}^\alpha \omega_{\mu\alpha}),
\end{aligned} \tag{10}$$

where W^μ is the spin divergence,

$$W^\mu = (1/\sqrt{-g})(\sqrt{-g}S^{\mu\nu})_{,\nu}. \tag{11}$$

The equations to be satisfied are generated from (10) by running through the possible index combinations. We will eventually want some of the equivalence relations with tetrad indices, but several useful equations result from considering coordinate indices first.

The equivalence expressed by Eq. (10) assumes the same fluid velocity in the perfect fluid as in the spin fluid. We could have used different velocities as, for example, Tupper[2] did in adding fluid viscosity and shear. This would introduce more parameters into the equivalence description. Since the spin-fluid stress-energy tensor is lengthy, we choose the simplest workable equivalence.

 0022-2488/87/071544-03$02.50

For the space-time described by Eq. (1), we find the kinematic parameters of the spin fluid are

$$\omega_{tr} = -f_r/2\sqrt{f}, \quad \omega_{r\phi} = k_r/2\sqrt{f}, \quad \dot{U}_r = f_r/2f, \qquad (12)$$
$$w_{tz} = -f_z/2\sqrt{f}, \quad \omega_{z\phi} = k_z/2\sqrt{f}, \quad \dot{U}_z = f_z/2f,$$

where $f_r = \partial_r f$, etc. The spin divergences are calculated to be

$$W_t = (\bar{e}^b/D^2)S_{\phi r}(fk_r - kf_r) + (\bar{e}^b/D^2)S_{\phi z}(fk_z - kf_z),$$
$$W_\phi = \bar{e}^b[\partial_r S_{\phi r} + \partial_z S_{\phi z} + (S_{\phi r}/2D^2)(lf_r - fl_r) + (S_{\phi z}/2D^2)(lf_z - fl_z), \qquad (13)$$
$$W_r = (1/D)\partial_z(\bar{e}^b DS_{rz}),$$
$$W_z = (1/D)\partial_r(\bar{e}^b DS_{zr}).$$

The tr, tz, $r\phi$, and $z\phi$ components of the stress-energy tensor are

$$T_{tr} = (-\sqrt{f}/2D)(\bar{e}^b DS_{rz})_{,z} - (3/4\sqrt{f})f_z\bar{e}^b S_{rz},$$
$$T_{tz} = (-\sqrt{f}/2D)(e^b DS_{zr})_{,r} - (3/4\sqrt{f})f_r\bar{e}^b S_{zr},$$
$$T_{r\phi} = -(k/2D\sqrt{f})(\bar{e}^b DS_{rz})_{,z} - (\bar{e}^b S_{rz}/4f^{3/2})(kf_z + 2fk_z), \qquad (14)$$
$$T_{z\phi} = (-k/2D\sqrt{f})(\bar{e}^b DS_{zr})_{,r} - \bar{e}^b(S_{zr}/4f^{3/2})(kf_r + 2fk_r).$$

These stress-energy components are zero and Eqs. (14) determine S_{rz} and set the condition for a consistent solution to exist. We find

$$S_{rz} = e^b A/Df^{3/2}, \qquad (15)$$

with $A \neq 0$ if f is proportional to k and $A = 0$ otherwise. We will find this is a very restrictive condition which eliminated the S_{rz} spin component in all of the examples we found. The rz component of the stress-energy tensor establishes another strong condition on the spins:

$$T_{rz} = (1/2D^2\sqrt{f})[S_{r\phi}(fk_z - kf_z) + S_{z\phi}(fk_r - kf_r)], \qquad (16)$$

which is also zero. Many useful and symmetric solutions depend only on one coordinate. In this case Eq. (16) causes a second spin component to be zero. The stress-energy components that are left are used to determine the remaining spin density and matter content of the spin fluid:

$$T_{t\phi} = \epsilon k - \frac{3}{4}S_{\phi r}\frac{e^{-b}f_r}{\sqrt{f}} - \frac{3}{4}S_{\phi z}\frac{\bar{e}^b f_z}{\sqrt{f}} - \frac{\sqrt{f}W_\phi}{2} - \frac{k}{2\sqrt{f}}W_t,$$
$$T_{tt} = f\epsilon - \sqrt{f}W_t, \qquad (17)$$
$$T_{rr} = P_r e^b + (k_r\sqrt{f}/D^2)S_{r\phi} - (kf_r/D^2\sqrt{f})S_{r\phi},$$
$$T_{zz} = P_z e^b + (k_z\sqrt{f}/D^2)S_{z\phi} - (kf_z/D^2\sqrt{f})S_{z\phi},$$
$$T_{\phi\phi} = \frac{\epsilon k^2}{f} + \frac{P_\phi D^2}{f} - \frac{kW_\phi}{\sqrt{f}} - \frac{S_{\phi r}\bar{e}^b}{2f^{3/2}}(kf_r + 2fk_r), - \frac{S_{\phi z}\bar{e}^b}{2f^{3/2}}(kf_z + 2fk_z).$$

We have allowed for anisotropic pressures in the spin fluid.

Equations (15) and (16) are useful as they stand. The remaining stress-energy components are more convenient to use with tetrad indices. Using Eq. (4) we find the tetrad indexed stress-energy components are

$$T_{00} = \epsilon - W_t/\sqrt{f}, \qquad (18)$$
$$T_{11} = P_r + S_{r\phi}(\bar{e}^b/D^2\sqrt{f})(fk_r - kf_r), \qquad (19)$$
$$T_{22} = P_z + S_{z\phi}(\bar{e}^b/D^2\sqrt{f})(fk_z - kf_z), \qquad (20)$$
$$T_{33} = P_\phi + (S_{\phi r}\bar{e}^b/D^2\sqrt{f})(kf_r - fk_r) + (S_{\phi z}\bar{e}^b/D^2\sqrt{f})(kf_z - fk_z), \qquad (21)$$

$$T_{03} = 0 = \frac{-W_3}{2} - \frac{3}{4}\frac{S_{\phi r}\bar{e}^b f_r}{D\sqrt{f}} - \frac{3}{4}\frac{S_{\phi z}\bar{e}^b f_z}{D\sqrt{f}}, \qquad (22)$$

$$W_3 = \frac{\bar{e}^b}{\sqrt{f}}\left[\left(\frac{S_{\phi r}f}{D}\right)_{,r} + \left(\frac{S_{\phi z}f}{D}\right)_{,z}\right] \qquad (23)$$

The procedure is simply to check Eqs. (15) and (16) for a possible zero and then to use Eqs. (18)–(23) to generate the description of the spin fluid source.

III. APPLICATIONS

A. The Gödel cosmology

We have

$$ds^2 = -(dt + e^{ax}\,dy)^2 + dx^2 + \tfrac{1}{2}e^{2ax}\,dy^2 + dz^2, \qquad (24)$$

with $(t,r,\phi,z) \to (t,x,y,z)$. This space-time has $\bar{\epsilon} = \bar{p} = \frac{1}{2}a^2$. For this space-time we have $f = 1$, $k = e^{ax}$, $l = -\frac{1}{2}e^{2ax}$, $b = 0$, $D^2 = \frac{1}{2}e^{2ax}$. Clearly f is not proportional to k, so

$$S_{xz} = 0. \qquad (25)$$

From Eq. (16),

$$S_{zy} = 0. \qquad (26)$$

The only nonzero spin is S_{xy}, a spin along the z axis of rotation. Equation (22) determines the functional form of the spin as

$$S_{yx} = Ae^{ax}. \qquad (27)$$

Using this spin and Eqs. (18)–(21) we find the energy density and pressure to be

$$\tfrac{1}{2}a^2 = \epsilon - 2aA, \quad \tfrac{1}{2}a^2 = P_r - 2aA,$$
$$\tfrac{1}{2}a^2 = P_z, \quad \tfrac{1}{2}a^2 = P_\phi - 2aA. \qquad (28)$$

There is a rotational correction to the usual isotropic Gödel pressures. There is no pressure change along the rotational axis. This spin fluid has a timelike divergence. The divergence along the spatial tetrad components is zero.

B. The Van Stockum solution[8]

We have

$$ds^2 = -(dt - a\rho^2\,d\phi)^2 + \rho^2\,d\phi^2 + e^{-\alpha^2\rho^2}(d\rho^2 + dz^2),$$
$$f = 1, \quad k = -a\rho^2, \quad l = \rho^2 - a^2\rho^4, \tag{29}$$
$$b = -\alpha^2\rho^2, \quad D^2 = \rho^2.$$

This space-time has a zero pressure and $\bar{\epsilon} = 4a^2 e^{\alpha^2\rho^2}$.[9] We have identified a and α in $\bar{\epsilon}$.

As in the previous example, we find

$$S_{\rho z} = S_{z\phi} = 0. \tag{30}$$

The nonzero spin component is determined by the vanishing of T_{03}:

$$S_{r\phi} = A\rho. \tag{31}$$

The pressures and energy densities are

$$\bar{\epsilon} = \epsilon + 2aAe^{\alpha^2\rho^2}, \quad P_r = 2aAe^{\alpha^2\rho^2},$$
$$P_z = 0, \quad P_\phi = 2aAe^{\alpha^2\rho^2}. \tag{32}$$

The nonzero pressures can again be interpreted as the rotational action of the spin about the axis of rotation. The Van Stockum spin source has only a timelike tetrad divergence, as in the Gödel cosmology.

Both of the examples considered thus far have only a timelike divergence component. The last example, which is a dust metric due to Hoenselaers and Vishveshwara,[10] develops a spatial component to the spin divergence.

C. An example with spatial divergence

The rotating dust solution of Hoenselaers and Vishveshwara[10] has a metric

$$\begin{aligned} ds^2 &= e^{2(cx+d)}(dx^2 + dy^2) + \frac{1}{\Psi}\left(a^2x^2 - \frac{1}{2}\right)dz^2 \\ &\quad + 2\left[\frac{\Omega}{\Psi}\left(\frac{1}{2} - a^2x^2\right) + ax\right]dz\,dt \\ &\quad + \Psi\,dt^2\left(\left(1 - \frac{a\Omega x}{4}\right)^2 - \frac{\Omega^2}{2\Psi}\right), \end{aligned} \tag{33}$$
$$f = \frac{\Omega^2}{2\Psi} - \left[1 - \frac{a\Omega x}{\Psi}\right]^2\Psi, \quad l = \frac{1}{\Psi}\left(a^2x^2 - \frac{1}{2}\right),$$
$$k = \frac{\Omega}{\Psi}\left(a^2x^2 - \frac{1}{2}\right) - ax,$$
$$b = 2(cx + d), \quad (t,r,z,\phi) \to (t,x,y,z),$$

where Ω, Ψ, c, and d are constants determined by boundary matching. This metric, like the previous examples, has only a single nonzero-spin component:

$$S_{xy} = 0, \quad S_{yz} = 0. \tag{34}$$

The nonzero component S_{xz} is functionally determined by the T_{03} component of the stress-energy tensor:

$$T_{03} = -W_3/2 + 3(S_{xz}\bar{e}^b f_x/4\sqrt{f}D) = 0, \tag{35}$$
$$S_{xz} = AD/f^{5/2}. \tag{36}$$

This spin fluid is the first example to have a spatial divergence component. The divergence is

$$W_t = (S_{zx}\bar{e}^b a/D^2\Psi)[\Omega^2/2 + (\Psi - a\Omega x)^2],$$
$$W_3 = -\tfrac{3}{2}(S_{zx}\bar{e}^b f_x/\sqrt{f}D). \tag{37}$$

The energy density is related to the perfect fluid energy density by

$$\epsilon = \bar{\epsilon} + W_t/\sqrt{f}. \tag{38}$$

The pressures are

$$P_x = P_z, \quad P_y = 0,$$
$$P_z = +S_{xz}(\bar{e}^b a/\sqrt{f}\Psi)[\Omega^2 + 2(\Psi - \Omega xa)^2]. \tag{39}$$

In summary, we have given the relations establishing spin fluid sources for axis-symmetric stationary space-times. The space-times used for illustration seem quite different, with some, for example, having fluid accelerations and some not; however, there are similarities between the spin fluid sources. All of the source examples are polarized, have some nonzero component of the spin divergence, and exhibit anisotropic pressure. All of the pressures are, however, symmetric about the axis of rotation. The spin density is in general required to Fermi–Walker transport. For the three space-times considered, this is equivalent to

$$\dot{S}_{\mu\nu} = 0.$$

The spin is constant along the flow lines. These examples provide another alternative to simple perfect fluids or fluids with viscosity and heat conduction.

[1] A. R. King and G. F. R. Ellis, Commun. Math. Phys. **31**, 209 (1973).
[2] B. O. Tupper, J. Math. Phys. **22**, 2666 (1981).
[3] A. K. Raychaudhuri and S. K. Saha, J. Math. Phys. **22**, 2237 (1981); **23**, 2554 (1982).
[4] B. O. J. Tupper, Gen. Relativ. Gravit. **15**, 47, 849 (1983).
[5] J. Carot and J. Ibanez, J. Math. Phys. **26**, 2282 (1985).
[6] J. R. Ray and L. L. Smalley, Phys. Rev. Lett. **49**, 1059 (1982).
[7] J. R. Ray and L. L. Smalley, Phys. Rev. D **27**, 1383 (1983).
[8] W. J. Van Stockum, Proc. R. Soc. Edinburgh **57**, 135 (1937).
[9] A nice discussion of this metric is found in W. B. Bonner, J. Phys. A **10**, 1673 (1977).
[10] C. Hoenselaers and C. V. Vishveshwara, J. Phys. A **12**, 209 (1979).

Homogeneous space-times of Gödel-type in higher-derivative gravity

A. J. Accioly
Instituto de Física, Universidade de São Paulo, Caixa Postal 20516, 01498-São Paulo, Brazil
A. T. Gonçalves
Instituto de Física, Universidade Federal do Rio de Janeiro, Caixa Postal 68528, 21944-Rio de Janeiro, Brazil

(Received 7 November 1986; accepted for publication 11 February 1987)

A general theorem concerning any Gödel-type solution of higher-derivative gravity field equations, which may be produced by any reasonable physical source with a constant energy-momentum tensor, is analyzed. The resulting class of metrics depends on two parameters, one of which is related to the vorticity. A general class of solutions of Gödel-type space-time-homogeneous universes in the context of the higher-derivative theory is exhibited. This is the most general higher-derivative solution of such type of metric and includes all known solutions of Einstein's equations related to these geometries as a special case. A number of completely causal rotating models is also obtained. Some of them present the interesting feature of having no analogs in the framework of general relativity.

I. INTRODUCTION

General relativity with higher-derivative terms has been considered[1–4] as a very attractive candidate for a theory of quantum gravity. The theory is defined by the action

$$I=\int d^4x\sqrt{-g}\left[\frac{R}{2\varkappa}-\frac{\Lambda}{\varkappa}+\alpha R^2+\beta R_{\mu\nu}R^{\mu\nu}+L_m\right], \tag{1.1}$$

where α and β are dimensionless coupling constants (in natural units), $\varkappa$ and Λ are the Einstein and cosmological constants, respectively, and L_m is the matter Lagrangian density. The corresponding field equations are given by

$$H_{\mu\nu} = -T_{\mu\nu}, \tag{1.2}$$

$$\begin{aligned}H_{\mu\nu} &= (1/\varkappa)(R_{\mu\nu}-\tfrac{1}{2}Rg_{\mu\nu})+(\Lambda/\varkappa)g_{\mu\nu}\\ &\quad+\alpha(-R^2g_{\mu\nu}+4RR_{\mu\nu}-4g_{\mu\nu}\Box R+4\nabla_\nu\nabla_\mu R)\\ &\quad+\beta(-2\Box R_{\mu\nu}-R_{\rho\Theta}R^{\rho\Theta}g_{\mu\nu}\\ &\quad+4R_{\mu\rho\Theta\nu}R^{\rho\Theta}-g_{\mu\nu}\Box R+2\nabla_\nu\nabla_\mu R),\end{aligned} \tag{1.3}$$

with trace

$$T=R/\varkappa+4\Lambda/\varkappa+4(3\alpha+\beta)\Box R. \tag{1.4}$$

For the quantum field theorist this higher-derivative theory has the great advantage of being renormalizable by power counting,[1] whereas, as it is well known, classical general relativity is clearly perturbatively nonrenormalizable by power counting in four dimensions.[5,6] In the pure classical framework, the aforementioned theory may be considered as a possible generalization of Einstein's general relativity, in the sense that it respects the geometrical nature of gravity as well as its gauge symmetry (invariance under general coordinate transformations). Recent work has shown[4,7–12] that the presence of a ghost responsible for a pseudononunitarity of the theory, which was considered its Achilles's heel, is no more a vulnerable point of it. The reason is that the ghost is unstable. In spite of the previously mentioned virtues, comparatively little is known about fourth-order gravity theory. Of course a better understanding of its behavior is of vital interest to those working on quantum gravity, and in particular, quantum cosmology. Consequently, the investigation of cosmological models in the framework of higher-derivative gravity is well suited.

Here we wish to focalize the so-called Gödel-type universes.[13] These models are defined by the line element

$$ds^2=[dt^2+H(x)dy]^2-D^2(x)dy^2-dx^2-dz^2, \tag{1.5}$$

and are such that in case

$$H=e^{mx},\quad D=e^{mx}/\sqrt{2}, \tag{1.6}$$

we recover Gödel's universe,[14] which is a solution of Einstein's equations with an energy-momentum tensor given by

$$\begin{aligned}T_{\mu\nu}&=\rho v_\mu v_\nu,\quad v^\alpha=\delta^\alpha{}_0,\\ m^2&=-2\Lambda=\varkappa\rho=2\Omega^2,\end{aligned} \tag{1.7}$$

where ρ is the constant density of matter, v^α is the fluid four-velocity, and Ω is the rate of rigid rotation of matter. Our choice for Gödel-type models is dictated, first of all, by their simplicity, which will allow us to accomplish the formidable task of finding exact solutions of higher-derivative gravity field equations, in the case of models that are homogeneous in space and time (ST homogeneous). And second, because this analysis will give us the opportunity of answering a very interesting question, i.e., what happens to the causal pathologies of these universes when quantum corrections are introduced in the standard general relativity theory?

We organize the paper in the following way. In Sec. II, we present a general theorem concerning any Gödel-type solution of fourth-order gravity field equations with constant energy-momentum tensor. The resulting class of metrics is characterized by two parameters, one of which is related to the rotation of the matter relative to the compass of inertia. Of course, any reasonable physical source will put restrictions on these parameters through the higher-derivative equations. Taking into account the last consideration,

 0022-2488/87/071547-06$02.50

we show in Sec. III that a geometry having as its source a perfect fluid plus a massless scalar field and an electromagnetic field can fit the parameters of the ST homogeneous Gödel-type universes. This is the most general higher-derivative solution concerning this type of metric and includes all known solutions of Einstein's equations related to such geometries as a special case. On the other hand, contrary to what generally happens in Einstein's theory, the restrictions on the parameters of the ST-homogeneous Gödel-type models, imposed by the sources through the higher-derivative equations, will provide us with a number of solutions which contain no closed timelike lines, i.e., that are completely causal. We will look into this subject in a comprehensive way in the last section.

II. A GENERAL THEOREM

In order to facilitate our calculations, we shall use a class of locally stationary observers represented by the vectors $e^{(A)}{}_{\alpha}$ defined by $\Theta^A = e^{(A)}{}_{\alpha}\, dx^{\alpha}$, wherein the one-forms Θ^A are given by

$$\Theta^0 = dt + H(x)dy, \quad \Theta^1 = dx, \qquad \Theta^2 = D(x)dy, \quad \Theta^3 = dz. \tag{2.1}$$

(Capital letters are tetrad indices and vary from 0 to 3 and Greek indices are tensor indices.) As a consequence, the vectors $e^{(A)}{}_{\alpha}$ assume the form

$$e^{(0)}{}_0 = e^{(1)}{}_1 = e^{(3)}{}_3 = 1, \quad e^{(0)}{}_2 = H, \quad e^{(2)}{}_2 = D, \tag{2.2}$$

and the geometry (1.5) may be written as

$$ds^2 = \eta_{AB}\Theta^A\Theta^B, \tag{2.3}$$

where $\eta_{AB} = \mathrm{diag}(+,-,-,-)$.

On the other hand, taking into account that $dx^{\alpha} = e^{\alpha}{}_{(A)}\Theta^A$, we immediately get

$$e^0{}_{(0)} = e^1{}_{(1)} = e^3{}_{(3)} = 1, \qquad e^0{}_{(2)} = -H/D, \quad e^2{}_{(2)} = D^{-1}. \tag{2.4}$$

We shall also need the Ricci coefficients of rotation defined by

$$\gamma^A_{BC} = -e^{(A)}{}_{\alpha;\beta}e^{\alpha}{}_{(B)}e^{\beta}{}_{(C)}. \tag{2.5}$$

[We use the comma for partial derivative, the semicolon for covariant derivative, and the bar for tetrad components of covariant derivatives. For instance, $R_{AB|C} = R_{AB;\alpha}e^{\alpha}{}_{(C)} = R_{AB,\alpha}e^{\alpha}{}_{(C)}$.] From (2.2) and (2.4) together with (2.5) we get the following nonvanishing components concerning these coefficients:

$$\begin{aligned} \gamma^0{}_{12} &= \gamma^1{}_{02} = \gamma^1{}_{20} = -\gamma^2{}_{01} \\ &= -\gamma^2{}_{10} = -\gamma^0{}_{21} = H'/2D, \\ \gamma^2{}_{12} &= -\gamma^1{}_{22} = D'/D, \end{aligned} \tag{2.6}$$

where the prime denotes differentiation with respect to x.

In the local inertial frame defined by $\Theta^A = e^{(A)}{}_{\alpha}\, dx^{\alpha}$ the higher-derivative gravity field equations, Eqs. (1.2) and (1.3), take the form

$$H_{AB} = -T_{AB}, \tag{2.7}$$

$$\begin{aligned} H_{AB} = {} & (1/\varkappa)(R_{AB} - \tfrac{1}{2}R\eta_{AB}) + (\Lambda/\varkappa)\eta_{AB} \\ & + \alpha[-R^2\eta_{AB} + 4RR_{AB} - 4\eta_{AB}\eta^{CD}(R_{|C|D} - \gamma^M{}_{CD}R_{|M}) + 4(R_{|A|B} - \gamma^M{}_{AB}R_{|M})] \\ & + \beta\{-R^{CD}R_{CD}\eta_{AB} + 4R_{ACDB}R^{CD} - \eta_{AB}\eta^{CD}(R_{|C|D} - \gamma^M{}_{CD}R_{|M}) + 2(R_{|A|B} - \gamma^M{}_{AB}R_{|M}) \\ & - 2\eta^{CD}[R_{AB|C|D} - (\gamma^M{}_{AC}R_{MB|D} + \gamma^M{}_{BC}R_{AM|D} \\ & + \gamma^M{}_{AC|D}R_{MB} + \gamma^M{}_{BC|D}R_{AM}) - \gamma^M{}_{AD}(R_{MB|C} - \gamma^N{}_{MC}R_{NB} - \gamma^N{}_{BC}R_{MN}) \\ & - \gamma^M{}_{BD}(R_{AM|C} - \gamma^N{}_{AC}R_{NM} - \gamma^N{}_{MC}R_{AN}) - \gamma^M{}_{CD}(R_{AB|M} - \gamma^N{}_{AM}R_{NB} - \gamma^N{}_{BM}R_{AN})]\}. \end{aligned} \tag{2.8}$$

We are ready now to demonstrate the following general result.

Theorem: Any Gödel-type solution of higher-derivative gravity field equations $H_{AB} = -T_{AB}$, having as the source of the geometry any field with T_{AB} independent of the points of the space-time, is space-time-homogeneous up to a local Lorentz transformation.

Proof: The only surviving components of H_{AB} [Eq. (2.7)] related to Gödel-type metrics [Eq. (1.5)] are

$$\begin{aligned} H_{00} = {} & \frac{1}{\varkappa}\left[-\frac{1}{2}\left(\frac{H'}{D}\right)^2 - \frac{R}{2}\right] + \frac{\Lambda}{\varkappa} + \alpha\left[-R^2 - 2R\left(\frac{H'}{D}\right)^2 + \frac{4D'R'}{D} + 4R''\right] + \beta\left\{-\frac{15}{4}\left(\frac{H'}{D}\right)^4\right. \\ & \left. - 3\frac{H'}{D}\left(\frac{H'}{D}\right)'' - \frac{3}{2}\left[\left(\frac{H'}{D}\right)'\right]^2 - 3\frac{D'}{D}\frac{H'}{D}\left(\frac{H'}{D}\right)' + 6\frac{D''}{D}\left(\frac{H'}{D}\right)^2 - 2\left(\frac{D''}{D}\right)^2 - 2\left(\frac{D''}{D}\right)'' - 2\frac{D'}{D}\left(\frac{D''}{D}\right)'\right\}, \end{aligned} \tag{2.9}$$

$$\begin{aligned} H_{11} = {} & \frac{1}{\varkappa}\left[-\frac{1}{2}\left(\frac{H'}{D}\right)^2 + \frac{D''}{D} + \frac{R}{2}\right] - \frac{\Lambda}{\varkappa} + \alpha\left[-R^2 - R\left(\frac{H'}{D}\right)^2 - \frac{4D'R'}{D}\right] \\ & + \beta\left\{-\frac{9}{4}\left(\frac{H'}{D}\right)^4 - 5\frac{D'}{D}\frac{H'}{D}\left(\frac{H'}{D}\right)' + 4\frac{D'}{D}\left(\frac{D''}{D}\right)' + \frac{1}{2}\left[\left(\frac{H'}{D}\right)'\right]^2 - \frac{H'}{D}\left(\frac{H'}{D}\right)'' + 4\frac{D''}{D}\left(\frac{H'}{D}\right)^2 - 2\left(\frac{D''}{D}\right)^2\right\}, \end{aligned} \tag{2.10}$$

$$H_{22}=\frac{1}{\varkappa}\left[-\frac{1}{2}\left(\frac{H'}{D}\right)^2+\frac{D''}{D}+\frac{R}{2}\right]-\frac{\Lambda}{\varkappa}+\alpha\left[-R^2-4R''-R\left(\frac{H'}{D}\right)^2\right]$$
$$+\beta\left\{-\frac{9}{4}\left(\frac{H'}{D}\right)^4-\frac{9}{2}\left[\left(\frac{H'}{D}\right)'\right]^2+4\left(\frac{D''}{D}\right)''-2\left(\frac{D''}{D}\right)^2-\frac{H'}{D}\frac{D'}{D}\left(\frac{H'}{D}\right)'-5\frac{H'}{D}\left(\frac{H'}{D}\right)''+4\left(\frac{H'}{D}\right)^2\frac{D''}{D}\right\}, \quad (2.11)$$
$$H_{33}=\frac{1}{\varkappa}\frac{R}{2}-\frac{\Lambda}{\varkappa}+\alpha\left[R^2-4R''-4\frac{D'}{D}R'\right]+\beta\left\{\frac{3}{4}\left(\frac{H'}{D}\right)^4+2\left(\frac{D''}{D}\right)^2\right.$$
$$\left.-\frac{3}{2}\left[\left(\frac{H'}{D}\right)'\right]^2-2\frac{D''}{D}\left(\frac{H'}{D}\right)^2-\frac{H'}{D}\left(\frac{H'}{D}\right)''+2\left(\frac{D''}{D}\right)''-\frac{D'}{D}\frac{H'}{D}\left(\frac{H'}{D}\right)'+2\frac{D'}{D}\left(\frac{D''}{D}\right)'\right\}, \quad (2.12)$$
$$H_{02}=-\frac{1}{2\varkappa}\left(\frac{H'}{D}\right)'+\alpha\left[-2\left(\frac{H'}{D}\right)'R-\frac{2H'R'}{D}\right]$$
$$+\beta\left\{-\left(\frac{H'}{D}\right)'''-9\left(\frac{H'}{D}\right)^2\left(\frac{H'}{D}\right)'+3\left(\frac{H'}{D}\right)'\frac{D''}{D}+4\frac{H'}{D}\left(\frac{D''}{D}\right)'+\left(\frac{H'}{D}\right)'\left(\frac{D'}{D}\right)^2-\frac{D'}{D}\left(\frac{H'}{D}\right)''\right\}; \quad (2.13)$$

where

$$R=\frac{1}{2}\left(\frac{H'}{D}\right)^2-2\frac{D''}{D}. \quad (2.14)$$

The assumption of T_{AB} constant, however, implies that H_{AB} is constant, too. On the other hand, it is not difficult to see from the above equations that H_{AB} is constant in case we have

$$H'/D=\text{const}\equiv 2\Omega,\quad D''/D=\text{const}\equiv m^2. \quad (2.15)$$

These are precisely the necessary and sufficient conditions for a Gödel-type metric to be space-time homogeneous.[15,16] Q.E.D.

Thus the whole class of solutions with T_{AB} constant is characterized by the two independent parameters m and Ω. It is not difficult to show that the last parameter is related to the vorticity. In fact, in the local frame considered, the rotation may be written as

$$\omega_{AB}=\tfrac{1}{2}[(\gamma^0{}_{BA}-\gamma^0{}_{AB})+(\gamma^0{}_{AB}\delta^0{}_B-\gamma^0{}_{B0}\delta^0{}_A)], \quad (2.16)$$

for a velocity field given by

$$v^A=e^{(A)}{}_0=\delta^A{}_0. \quad (2.17)$$

As a consequence, the vorticity assumes the form

$$\omega_{12}=-\tfrac{1}{2}(H'/D). \quad (2.18)$$

It follows then that the vorticity vector $\omega^A=\frac{1}{2}\epsilon^{ABCD}\omega_{BC}v_D$ is given by

$$\omega^A=(0,0,0,\Omega),$$

where $2\Omega=H'/D$. (2.19)

III. A CLASS OF HIGHER-DERIVATIVE GÖDEL-TYPE SOLUTIONS

It is reasonable to question, *ab initio*, what material content we may consider as source of our geometry, in order to obtain the most general higher-derivative Gödel-type solution, i.e., a solution that includes all known solutions of Einstein's equations related to such geometries. The answer is straightforward if we appeal to a recent work of Rebouças and Tiomno.[16] There, they exhibit a remarkable class of exact solutions of Einstein–Maxwell-scalar field equation which is the most general solution of a Gödel-type ST-homogeneous metric. So we consider a rotating universe ($\Omega\neq 0$) for which the material content is a perfect fluid of density ρ and pressure p plus a source-free electromagnetic field F_{AB} and a massless scalar field S. Consequently, the energy-momentum tensor in the tetrad frame becomes

$$T_{AB}=\rho v_A v_B-p(\eta_{AB}-v_A v_B)+T^{(S)}_{AB}+T^{(EM)}_{AB}, \quad (3.1)$$

where

$$T^{(EM)}_{AB}=\tfrac{1}{4}F_{CD}F^{CD}\eta_{AB}-F_{AM}F_B{}^M,$$
$$T^{(S)}_{AB}=S_{|A}S_{|B}-\tfrac{1}{2}\eta_{AB}S_{|M}S_{|N}\eta^{MN}, \quad (3.2)$$

and v^A is given by Eq. (2.17).

The Maxwell equations concerning the source-free electromagnetic field are given by

$$F^{AB}{}_{|B}+\gamma^A{}_{MB}F^{MB}+\gamma^B{}_{MB}F^{AM}=0,$$
$$F_{[AB|C]}+2F_{M[C}\gamma^M{}_{AB]}=0, \quad (3.3)$$

whereas the zero-mass scalar field equation is as follows:

$$\eta^{AB}S_{|A|B}-\gamma^M{}_{AB}\eta^{AB}S_{|M}=0. \quad (3.4)$$

The brackets denote total antisymmetrization.

On the other hand, the fact we are requiring space-time homogeneity of the Gödel-type models implies that T_{AB} is constant (cf. the theorem in the last section). We remark also that we have a preferred direction in our universe determined by the rotation. Taking into account the above considerations we can seek solutions of (3.3) and (3.4), respectively, related to our model. Let us first consider the electromagnetic field. Since it is not a pure test field but also acts as source of curvature, it must then be compatible with the space-time symmetries. As a consequence, we are led to take both **E** and **B** along the direction of rotation. Thus the only nonvanishing components of F_{AB} are

$$F_{30}=-F_{03}=E(z),\quad F_{12}=-F_{21}=B(z). \quad (3.5)$$

Using (2.6) and (3.5), Eqs. (3.3) reduce, respectively, to

$$E_{,3}+(H'/D)B=0,\quad B_{,3}-(H'/D)E=0. \quad (3.6)$$

But, since $H'/D=2\Omega$ (ST homogeneity), the general solution of Eqs. (3.6) can be written as

$$E = E_0 \cos[2\Omega(z - z_0)],$$
$$B = E_0 \sin[2\Omega(z - z_0)], \tag{3.7}$$

where E_0 and z_0 are constants. In the case of the massless scalar field it is trivial to show that if we take

$$S = az + b, \tag{3.8}$$

where a and b are constants, we can satisfy Eq. (3.4) as well as the space-time symmetries.

Now, the non-null components of H_{AB} for the ST homogeneous Gödel-type metric are

$$H_{00} = (1/\varkappa)[-3\Omega^2 + m^2] + (\Lambda/\varkappa) + \alpha[-20\Omega^4 - 4m^4 + 24\Omega^2 m^2] + \beta[-60\Omega^4 + 24m^2\Omega^2 - 2m^4], \tag{3.9}$$

$$H_{11} = H_{22} = (1/\varkappa)[-\Omega^2] - (\Lambda/\varkappa) + \alpha[-12\Omega^4 - 4m^4 + 16\Omega^2 m^2] + \beta[-36\Omega^4 + 16m^2\Omega^2 - 2m^4], \tag{3.10}$$

$$H_{33} = (1/\varkappa)[\Omega^2 - m^2] - (\Lambda/\varkappa) + \alpha[4\Omega^4 + 4m^4 - 8\Omega^2 m^2] + \beta[12\Omega^4 + 2m^4 - 8m^2\Omega^2]. \tag{3.11}$$

As a result, the higher-derivative gravity field equations reduce to the following set of three equations:

$$\rho = \frac{E_0^2}{2} - \frac{3}{2}a^2 - \frac{\Lambda}{\varkappa} - 2m^4(2\alpha + \beta) + 4\Omega^4(\alpha + 3\beta) + \Omega^2/\varkappa, \tag{3.12}$$

$$p = -\frac{E_0^2}{2} + \frac{a^2}{2} + \frac{\Lambda}{\varkappa} + 2m^4(2\alpha + \beta) + 12\Omega^4(\alpha + 3\beta) - 16\Omega^2 m^2(\alpha + \beta) + \Omega^2/\varkappa, \tag{3.13}$$

$$m^2/\varkappa = 16\Omega^4(\alpha + 3\beta) + 4m^4(2\alpha + \beta) - 24m^2\Omega^2(\alpha + \beta) + 2(\Omega^2/\varkappa) - E_0^2 + a^2. \tag{3.14}$$

The positivity of energy and pressure is guaranteed if the cosmological constant satisfies the relation

$$-12\Omega^4(\alpha + 3\beta) - 2m^4(2\alpha + \beta) + 16m^2\Omega^2(\alpha + \beta) + \frac{E_0^2}{2} - \frac{\Omega^2}{2} - \frac{a^2}{2} \leqslant \frac{\Lambda}{\varkappa} \leqslant 4\Omega^4(\alpha + 3\beta) - 2m^4(2\alpha + \beta) + \frac{E_0^2}{2} - \frac{3}{2}a^2 + \frac{\Omega^2}{\varkappa}, \tag{3.15}$$

which implies that

$$8\Omega^4(\alpha + 3\beta) + \Omega^2[1/\varkappa - 8m^2(\alpha + \beta)] - a^2/2 \geqslant 0, \tag{3.16}$$

the equality having as its consequence

$$\frac{\Lambda}{\varkappa} = \frac{E_0^2}{2} - \frac{2\Omega^2}{\varkappa} - 20\Omega^4(\alpha + 3\beta) + 24m^2\Omega^2(\alpha + \beta) - 2m^4(2\alpha + \beta). \tag{3.17}$$

Equations (3.12) and (3.13) imply in an equation of state $p = \gamma\rho$ for the cosmic fluid, wherein γ is a constant. The Lichnerowicz condition, $0 \leqslant \gamma \leqslant 1$, will be ensured if

$$\Lambda/\varkappa \leqslant E_0^2/2 - a^2 - 2m^4(2\alpha + \beta) - 4\alpha^4(\alpha + 3\beta) + 8\Omega^2 m^2(\alpha + \beta), \tag{3.18}$$

which is consistent with (3.15) and (3.16).

In the integration of Eqs. (3.12)–(3.14) three cases arise, according as m^2 is $>$, $<$, or $= 0$. In order to make easier the comparison of our results with those of the literature, we express our solutions in cylindrical coordinates. Of course, the Gödel-type metric in cylindrical coordinates, i.e.,

$$ds^2 = [dt + H(r)d\Phi]^2 - D^2(r)d\Phi^2 - dr^2 - dz^2, \tag{3.19}$$

is precisely of the form (1.5). The Gödel universe corresponds to

$$H(r) = (2\sqrt{2}/m)\sinh^2(mr/2),$$
$$D(r) = \sinh(mr)/m, \tag{3.20}$$

where Φ is an angular coordinate. We also call attention to the fact that the theorem of Sec. II is valid *mutatis mutandis*.

In case I

$$[16\Omega^4(\alpha + 3\beta) + 4m^4(2\alpha + \beta) - 24m^2\Omega^2(\alpha + \beta) + 2\Omega^2/\varkappa - E_0^2 + a^2 = m^2 > 0],$$

we obtain

$$ds^2 = \left[dt + \frac{4\Omega}{m^2}\sinh^2\left(\frac{mr}{2}\right)d\Phi\right]^2 - \frac{1}{m^2}\sinh^2(mr)d\Phi^2 - dr^2 - dz^2. \tag{3.21}$$

Here Φ is to be regarded as an angular coordinate. In fact, Eqs. (3.21) satisfy Maitra's conditions for regularity near the origin,[17] i.e.,

$$H = r^2 \times \text{const}, \quad D = r.$$

We also have that the relation

$$\Omega^2/\varkappa > E_0^2/4 - 8\Omega^4(\alpha + 3\beta) + 10\Omega^2 m^2(\alpha + \beta) - m^4(2\alpha + \beta) \tag{3.22}$$

holds.

Case II

$$[16\Omega^4(\alpha + 3\beta) + 4n^4(2\alpha + \beta) + 24n^2\Omega^2(\alpha + \beta) + 2\Omega^2/\varkappa - E_0^2 + a^2 = -n < 0, \quad m^2 \equiv -n^2 < 0]$$

corresponds to the following metric:

$$ds^2 = \left[dt + \frac{4\Omega}{n^2}\sin^2\left(\frac{nr}{2}\right)d\Phi\right]^2 - \frac{\sin^2 nr}{n^2}d\Phi^2 - dr^2 - dz^2. \tag{3.23}$$

The relation

$$E_0^2 > 2a^2 + 8n^2\Omega^2(\alpha + \beta) + 4n^4(2\alpha + \beta) \tag{3.24}$$

holds. Equation (3.23) is an analytical extension of Eq. (3.21) with $m \rightarrow in$. We remark that our coordinates are true cylindrical coordinates, i.e., they satisfy Maitra's conditions.

The remaining case, $m^2 = 0$, may be considered as a limit of the first ($m^2 \to 0$) and the second ($n^2 \to 0$) cases, respectively. The metric is given by

$$ds^2 = [dt + \Omega r^2 \, d\Phi]^2 - r^2 \, d\Phi^2 - dr^2 - dz^2. \quad (3.25)$$

In this case the following relation holds:

$$4\Omega^2/\varkappa + 32\Omega^4(\alpha+\beta) \geqslant E_0^2 = 2\Omega^2/\varkappa + a^2 + 16\Omega^4(\alpha + 3\beta) \geqslant 2a^2. \quad (3.26)$$

We have thus succeeded in deriving the most general higher-derivative solution concerning ST-homogeneous Gödel-type universes. As we have anticipated, our solutions are such that we can recover from them all known solutions of Einstein's equations concerning such geometries. Indeed, as $\alpha,\beta \to 0$, we obtain the Rebouças and Tiomno solution,[16] which includes all known solutions of Einstein's equations related to these geometries. (For instance, when $\alpha,\beta,E_0,a \to 0$, we get the Gödel solution[14] with $m^2 = 2\Omega^2$ [cf. Eq. (3.20)]. If $\alpha,\beta,a,m \to 0$ we recover the Som–Raychaudhuri metric.[18] The Banerjee–Banerji[19] as well as Rebouças[20] solutions are obtained when $\alpha,\beta,a \to 0$, noting that the first one concerns a charged fluid and thus the electromagnetic field is different from that of the second one, but both have the same $T_{AB}^{(EM)}$, and so on.[21–24]) We also remark that Riemannian Gödel-type ST-homogeneous metrics with the same value of m^2 and Ω are isometric.[16]

We have analyzed so far the fourth-order gravity solutions from a classical point of view. In this sense, the parameters α and β are quite arbitrary. However, in the framework of quantum field theory, the situation is rather different. In fact, the higher-derivative theory contains two mass scales,[12–14,25] associated with the spin-0 and spin-2 particles present in the linearized theory. They are given, respectively, by

$$m_0^2 = 1/4\varkappa(3\alpha+\beta), \quad (3.27)$$

and

$$m_2^2 = -1/2\varkappa\beta. \quad (3.28)$$

The spin-0 particle has significance even in the nonlinear sector.[26]

Thus nontachyionic spin-0 and spin-2 particles require $(3\alpha+\beta)$ to be positive and β to be negative, respectively. Consequently, these restrictions on the parameters α and β must be included in our solutions.

IV. ROTATING GÖDEL-TYPE UNIVERSE WITHOUT VIOLATION OF CAUSALITY IN HIGHER-DERIVATIVE GRAVITY

It is interesting to consider the question of closed timelike lines in our solutions. To accomplish this we write Eq. (2.38) in the form

$$ds^2 = dt^2 + 2H \, d\Phi \, dt - L \, d\Phi^2 - dr^2 - dz^2, \quad (4.1)$$

where

$$L(r) = D^2 - H^2. \quad (4.2)$$

Clearly, if $L(r)$ becomes negative at $r_1 < r < r_2$, then the curve defined by $r,t,z = \text{const}$ is a closed timelike trajectory. The existence of such curves poses a difficult problem related to the possibility of violation of the well-established causality principle.

In our case, when $m^2 > 0$, Eq. (3.21) leads to

$$L(r) = \frac{4}{m^2}\sinh^2\left(\frac{mr}{2}\right)\left[1 - \left(\frac{4\Omega^2}{m^2} - 1\right)\sinh^2\left(\frac{mr}{2}\right)\right]. \quad (4.3)$$

Consequently, unless

$$m^2 \geqslant 4\Omega^2, \quad (4.4)$$

$L(r)$ will become negative for

$$\sinh^2\left(\frac{mr_1}{2}\right) > \left(\frac{4\Omega^2}{m^2} - 1\right)^{-1}. \quad (4.5)$$

Thus the limiting case in which the noncausal region will disappear corresponds to $m^2 = 4\Omega^2$. On the other hand, a straightforward calculation gives the following relation in the case of our solutions with $m^2 > 0$:

$$4\Omega^2/\varkappa \geqslant m^2/\varkappa - 32\Omega^4(\alpha + 3\beta) + 40m^2\Omega^2(\alpha+\beta) - 4m^4(2\alpha+\beta). \quad (4.6)$$

Undoubtedly, the solution $m^2 = 4\Omega^2$ is compatible with the preceding inequality. It follows then from (3.14) and (3.16) that

$$2\Omega^2/\varkappa = 16(3\alpha+\beta)\Omega^4 - E_0^2 + a^2, \quad (4.7)$$

$$-16(3\alpha+\beta)\Omega^4 + 2\Omega^2/\varkappa - a^2 \geqslant 0. \quad (4.8)$$

Consequently,

$$E_0^2 = 0, \quad 2\Omega^2/\varkappa = a^2 + 16(3\alpha+\beta)\Omega^4. \quad (4.9)$$

Now, from Eq. (3.17) we get

$$\Lambda/\varkappa = 4(3\alpha+\beta)\Omega^4 - 2\Omega^2/\varkappa, \quad (4.10)$$

and from Eqs. (3.12) and (3.13)

$$\rho = p = 0. \quad (4.11)$$

Admitting that $16\varkappa^2a^2(3\alpha+\beta) < 1$ and taking into account that $(3\alpha+\beta)$ must be positive in order to avoid the tachyonic spin-0 particle, we obtain from Eq. (4.9) the following values concerning Ω^2:

$$\Omega_{(c)}^2 = \left[1 - \sqrt{1 - 16\varkappa^2a^2(3\alpha+\beta)}\,\right]/16(3\alpha+\beta)\varkappa, \quad (4.12)$$

$$\Omega_{(q)}^2 = \left[1 + \sqrt{1 - 16\varkappa^2a^2(3\alpha+\beta)}\,\right]/16(3\alpha+\beta)\varkappa. \quad (4.13)$$

When $(3\alpha+\beta) \to 0$, $\Omega_{(c)}^2 \to a^2\varkappa/2$, and we recover the Rebouças and Tiomno solution,[16] which is the only known exact Gödel-type solution of Einstein's equations describing a completely causal space-time homogeneous rotating universe.

We have thus succeeded in finding two completely causal rotating solutions. We should like to mention that the solution concerning $\Omega_{(q)}^2$ has no classical analogs, and it is, as far as we know, the first known exact solution of higher-derivative gravity field equations with this characteristic. "Classical" here means "from the point of view of general relativity." On the other hand, it is not difficult to show that in case $m \leqslant 0$ we cannot have completely causal solutions.

Last but not least, it is interesting to question if the causal pathologies of these universes can be avoided in the absence of the scalar field. The answer is yes. Indeed, our previous results provide us with the completely causal rotating solution, in case $a^2 = 0$ and $m^2 > 0$:

$$\Omega^2 = \frac{1}{8(3\alpha + \beta)\varkappa} = \frac{m^2}{4}, \quad \Lambda = -\frac{3}{2}\Omega^2, \quad \rho = p = 0. \tag{4.14}$$

We point out that the above solution has no similar one in the framework of general relativity.

ACKNOWLEDGMENT

The authors gratefully acknowledge financial support from the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

[1]K. S. Stelle, Phys. Rev. D **16**, 953 (1977).
[2]N. H. Barth and S. M. Christensen, Phys. Rev. D **28**, 1876 (1983).
[3]S. M. Christensen, in *Quantum Structure of Space and Time*, edited by M. J. Duff and C. J. Isham (Cambridge U. P., Cambridge, 1982), p. 71.
[4]I. Antoniadis and E. T. Tomboulis, Phys. Rev. D **33**, 2756 (1986).
[5]G. 't Hooft and M. Veltman, Ann. Inst. H. Poincaré **20**, 69 (1974).
[6]S. Deser and P. van Nieuwenhuizen, Phys. Rev. D **10**, 401, 411 (1974).
[7]E. T. Tomboulis, Phys. Lett. B **97**, 77 (1980).
[8]E. T. Tomboulis, Phys. Lett. B **70**, 361 (1977).
[9]L. Smolin, Nucl. Phys. B **208**, 439 (1982).
[10]M. Kaku, Phys. Rev. D **27**, 2819 (1983).
[11]E. T. Tomboulis, Phys. Rev. Lett. **52**, 1173 (1984).
[12]D. G. Boulware, G. T. Horowitz, and A. Strominger, Phys. Rev. Lett. **50**, 1726 (1983).
[13]M. J. Rebouças and J. Tiomno, Nuovo Cimento B **90**, 204 (1985).
[14]K. Gödel, Rev. Mod. Phys. **21**, 447 (1949).
[15]A. K. Raychaudhuri and S. N. G. Thakurta, Phys. Rev. D **22**, 802 (1980).
[16]M. J. Rebouças and J. Tiomno, Phys. Rev. D **28**, 1251 (1983).
[17]S. C. Maitra, J. Math. Phys. **7**, 1025 (1966).
[18]M. M. Som and A. K. Raychaudhuri, Proc. R. Soc. London Ser. A **304**, 81 (1968).
[19]A. Banerjee and S. Banerji, J. Phys. A **1**, 1 (1968).
[20]M. J. Rebouças, Phys. Lett. A **70**, 161 (1979).
[21]M. Novello and M. J. Rebouças, Phys. Rev. D **19**, 2850 (1979).
[22]C. Hoenselaers and C. V. Vishveshwara, Gen. Relativ. Gravit. **1**, 43 (1979).
[23]S. K. Chakraborty, Gen. Relativ. Gravit. **12**, 925 (1980).
[24]F. Bampi and C. Zordan, Gen. Relativ. Gravit. **9**, 393 (1978).
[25]B. Whitt, Phys. Rev. D **33**, 379 (1985).
[26]B. Whitt, Phys. Lett. B **145**, 176 (1984).

Symmetric tensor spherical harmonics on the N-sphere and their application to the de Sitter group SO(N,1)

Atsushi Higuchi
Physics Department, Yale University, New Haven, Connecticut 06511 and Department of Physics, University of Wisconsin–Milwaukee, Milwaukee, Wisconsin 53201[a]

(Received 21 November 1986; accepted for publication 18 March 1987)

The symmetric tensor spherical harmonics (STSH's) on the N-sphere (S^N), which are defined as the totally symmetric, traceless, and divergence-free tensor eigenfunctions of the Laplace–Beltrami (LB) operator on S^N, are studied. Specifically, their construction is shown recursively starting from the lower-dimensional ones. The symmetric traceless tensors induced by STSH's are introduced. These play a crucial role in the recursive construction of STSH's. The normalization factors for STSH's are determined by using their transformation properties under SO($N+1$). Then the symmetric, traceless, and divergence-free tensor eigenfunctions of the LB operator in the N-dimensional de Sitter space-time which are obtained by the analytic continuation of the STSH's on S^N are studied. Specifically, the allowed eigenvalues of the LB operator under the restriction of unitarity are determined. Our analysis gives a group-theoretical explanation of the forbidden mass range observed earlier for the spin-2 field theory in de Sitter space-time.

I. INTRODUCTION

A renewed interest in field theories in de Sitter space-time[1] has been aroused among particle physicists and cosmologists since inflationary cosmologies were proposed.[2–4] Since bosons of definite spin are described by totally symmetric, traceless, and divergence-free tensors in Minkowski space-time we are led to study such tensors also in de Sitter space-time. (Note, however, that particles with arbitrary spin can be studied by using Weinberg-type fields.[5–7]) It is particularly interesting to find the conditions for such tensors to form unitary representations of SO(4,1), which is the isometry group of de Sitter space-time.[8] (Recall that unitarity is necessary to avoid negative probabilities.) The author has observed before that certain values of (mass)^2 are forbidden for the spin-2 field theory because of the appearance of negative-norm states.[9] This implies that the representations of SO(4,1) corresponding to those values of (mass)^2 are nonunitary. (The mass gap discussed in Ref. 9 is different in nature from the well-known mass discontinuity in Minkowski space-time,[10,11] which rules out the possibility of the graviton being the massless limit of a massive spin-2 particle.)

The N-dimensional de Sitter space-time is the maximally symmetric solution of the Einstein equation with a positive cosmological constant Λ,

$$R_{\mu\nu} - \tfrac{1}{2} g_{\mu\nu} R + \Lambda g_{\mu\nu} = 0. \tag{1.1}$$

The metric can be written as

$$ds^2 = -dt^2 + \cosh^2 t \, ds^2_{N-1}, \tag{1.2}$$

where ds^2_{N-1} is the line element of S^{N-1}. We have adopted the unit in which

$$2\Lambda/(N-1)(N-2) = 1. \tag{1.3}$$

The metric (1.2) is related to that of S^N,

$$ds^2 = d\chi^2 + \sin^2 \chi \, ds^2_{N-1}, \tag{1.4}$$

by

$$\chi = \pi/2 - it. \tag{1.5}$$

Consequently, the totally symmetric, traceless, and divergence-free tensor eigenfunctions of the Laplace–Beltrami (LB) operator $\nabla_\alpha \nabla^\alpha$ in de Sitter space-time can be obtained by the analytic continuation of the symmetric tensor spherical harmonics (STSH's) on S^N, which are defined here as the symmetric tensor eigenfunctions $h_{\mu\nu\kappa\cdots\lambda}$ of the LB operator on S^N satisfying

$$\nabla^\alpha h_{\alpha\nu\kappa\cdots\lambda} = g^{\alpha\beta} h_{\alpha\beta\kappa\cdots\lambda} = 0, \tag{1.6}$$

where $g^{\mu\nu}$ is the inverse of the metric $g_{\mu\nu}$ of S^N. (It must be kept in mind that the symmetric tensors are not enough to obtain all the possible bosonic representations for $N>4$.)

With this observation in mind we first study the STSH's on S^N. The STSH's of rank $r \leqslant 2$ have been studied by Chodos and Myers[12] and those of arbitrary rank by Rubin and Ordóñez[13] by using polynomials of the Cartesian coordinates in $(N+1)$-dimensional Euclidean space. (We leave N arbitrary because no difficulty arises by doing so.) Here we show how to construct them in terms of associated Legendre functions. This construction is more suitable for obtaining the symmetric, traceless, and divergence-free tensor eigenfunctions of the LB operator in N-dimensional de Sitter space-time ($N \geqslant 3$) by analytic continuation. Now these eigenfunctions form a representation of SO(N,1). But it must be unitary if they are to describe particles in de Sitter space-time. So we introduce an inner product among these eigenfunctions and determine the allowed eigenvalues for the LB operator $\nabla_\alpha \nabla^\alpha$ under the restriction of unitarity.

The rest of the paper is organized as follows. In Sec. II we construct scalar spherical harmonics on S^N. In Sec. III we discuss the STSH's on S^2. In Sec. IV we rewrite the equa-

[a] Present address.

0022-2488/87/071553-14$02.50

tions satisfied by STSH's in such a way that they are adequate for the recursive construction of the STSH's on S^N from the lower-dimensional ones. In Sec. V the symmetric traceless tensors induced by STSH's are defined and some of their properties are derived. These tensors play an important role in the construction of STSH's. Then in Sec. VI the equations derived in Sec. IV are examined and the STSH's on S^N are given in terms of those on S^{N-1} for $N\geqslant 3$. Since the STSH's on S^2 are constructed in Sec. III, this enables one to construct the STSH's on S^N with arbitrary N. In Sec. VII we analyze the transformation properties of a certain class of STSH's under SO$(N+1)$ and use them to determine the normalization factors of STSH's in Sec. VIII. Then we go on from S^N to N-dimensional de Sitter space-time. In Sec. IX we study the symmetric, traceless, and divergence-free tensor eigenfunctions of the LB operator in N-dimensional de Sitter space-time ($N\geqslant 3$), which are obtained by the analytic continuation of the STSH's on S^N. There we define an inner product among those tensor eigenfunctions and find the allowed eigenvalues of the LB operator by requiring unitarity, i.e., the positive-definiteness of the norm. We encounter zero-norm eigenfunctions, which are identified with zero in the unitary representations, for certain values of the LB operator. We show in Sec. X that these eigenfunctions are obtained by the analytic continuation of symmetric traceless tensors induced by STSH's. In Sec. XI we summarize the results obtained in this paper.

II. SCALAR SPHERICAL HARMONICS

Scalar spherical harmonics on S^N have been studied by many authors.[12–14] Here we write them down in terms of associated Legendre functions.

Let us parametrize the line element ds_N^2 of S^N as follows:

$$ds_1^2 = d\theta_1^2, \tag{2.1}$$

$$ds_n^2 = d\theta_n^2 + \sin^2\theta_n\, ds_{n-1}^2 \quad (n=2,\dots,N). \tag{2.2}$$

Let $\Box_n$ be the LB operator on S^n

$$\Box_n = (1/\sqrt{g^{(n)}})\partial_i(\sqrt{g^{(n)}}g^{(n)ij}\partial_j), \tag{2.3}$$

where i and j run from θ_1 to θ_n and $g^{(n)ij}$ is the inverse of the metric tensor $g_{ij}^{(n)}$ on S^n and

$$\sqrt{g^{(n)}} = \sin^{n-1}\theta_n \sin^{n-2}\theta_{n-1}\cdots\sin\theta_2. \tag{2.4}$$

Then the scalar (0th-rank tensor) eigenfunctions of the LB operator $\Box_N$ on S^N, i.e., the scalar spherical harmonics on S^N, are given by

$$Y_{l_N\cdots l_1}(\theta_N,\dots,\theta_1) = \left[\prod_{n=2}^{N} {}_n\overline{P}{}^{\,l_{N-1}}_{l_n}(\theta_n)\right]\frac{1}{\sqrt{2\pi}}e^{il_1\theta_1}, \tag{2.5}$$

where $l_1,l_2,\dots,l_N$ are integers that satisfy

$$l_N\geqslant l_{N-1}\geqslant\cdots\geqslant l_2\geqslant|l_1|. \tag{2.6}$$

Here ${}_n\overline{P}{}^{\,l}_L(\theta)$ is defined by

$${}_n\overline{P}{}^{\,l}_L(\theta) = {}_nc^l_L(\sin\theta)^{-(n-2)/2}P^{-(l+(n-2)/2)}_{L+(n-2)/2}(\cos\theta), \tag{2.7}$$

where $P_\nu^{-\mu}(x)$ is the associated Legendre function of the first kind defined by[15]

$$P_\nu^{-\mu}(x) = \frac{1}{\Gamma(1+\mu)}\left(\frac{1-x}{1+x}\right)^{\mu/2} \times F\left(-\nu,\nu+1;1+\mu;\frac{1-x}{2}\right). \tag{2.8}$$

Here $F(\alpha,\beta;\gamma;z)$ is the hypergeometric function and ${}_nc^l_L$ is a normalization constant determined by requiring

$$\int d\theta_1\cdots d\theta_n\sqrt{g^{(n)}}Y_{l_n\cdots l_1}Y^*_{l'_n\cdots l'_1} = \delta_{l_nl'_n}\cdots\delta_{l_1l'_1}, \tag{2.9}$$

as

$${}_nc^l_L = \left[\frac{2L+n-1}{2}\frac{(L+l+n-2)!}{(L-l)!}\right]^{1/2}. \tag{2.10}$$

The following formula proved in Appendix A has been used to find the above equation:

$$\int_{-1}^{1} P_\nu^{-\mu}(x)P_{\nu'}^{-\mu}(x)dx = \frac{2\sin\pi(\nu'-\nu)}{(\nu'-\nu)(\nu'+\nu+1)\pi}\,\frac{\Gamma(1-\mu+\nu')}{\Gamma(1+\mu+\nu')}. \tag{2.11}$$

One has

$$\left[\frac{\partial^2}{\partial\theta^2} + (N-1)\cot\theta\frac{\partial}{\partial\theta} - \frac{l(l+N-2)}{\sin^2\theta}\right]{}_N\overline{P}{}^{\,l}_L(\theta) = -L(L+N-1)\,{}_N\overline{P}{}^{\,l}_L(\theta). \tag{2.12}$$

This implies

$$\Box_n Y_{l_N\cdots l_1} = -l_n(l_n+n-1)Y_{l_N\cdots l_1}. \tag{2.13}$$

In particular, the eigenvalues of the LB operator on S^N for scalar spherical harmonics are $-L(L+N-1)$ ($L=0,1,\dots$).

III. STSH'S ON S^2

In this section we study the STSH's on S^2. The metric of S^2 is

$$ds^2 = d\theta^2 + \sin^2\theta\, d\phi^2. \tag{3.1}$$

The scalar spherical harmonics are well known and have been given in the previous section. The vector spherical harmonics (the STSH's of rank 1) $Y_\mu^{(lm)}(\theta,\phi)$ are given by[16]

$$Y_\mu^{(lm)}(\theta,\phi) = [1/\sqrt{l(l+1)}]\epsilon_{\mu\nu}\partial^\nu Y_{lm}(\theta,\phi) \quad (l\geqslant 1), \tag{3.2}$$

where the totally antisymmetric tensor $\epsilon_{\mu\nu}$ is defined by

$$\epsilon_{\theta\theta} = \epsilon_{\phi\phi} = 0, \tag{3.3a}$$

$$\epsilon_{\theta\phi} = -\epsilon_{\phi\theta} = \sin\theta. \tag{3.3b}$$

It is covariantly constant, i.e.,

$$\nabla_\mu\epsilon_{\nu\lambda} = 0. \tag{3.4}$$

The vector spherical harmonic $Y_\mu^{(lm)}(\theta,\phi)$ satisfies

$$\nabla^\alpha\nabla_\alpha Y_\mu^{(lm)}(\theta,\phi) = [-l(l+1)+1]Y_\mu^{(lm)}(\theta,\phi) \tag{3.5}$$

and

$$\int d\Omega_2\, g^{\mu\nu}Y_\mu^{(lm)}(\theta,\phi)Y_\nu^{(l'm')}(\theta,\phi) = \delta_{ll'}\delta_{mm'}, \tag{3.6}$$

where $d\Omega_2$ is the volume element of S^2.

It is well known that the STSH's of rank $r\geqslant 2$ do not exist on S^2 (see Refs. 13 and 16) as we will show below. Suppose that $h_{\mu_1\cdots\mu_r}$ is a symmetric, traceless, and divergence-free tensor of rank $r\geqslant 2$. Define

$$\nabla_{[\mu}h_{\nu]\mu_1\cdots\mu_{r-1}} = \epsilon_{\mu\nu} f_{\mu_1\cdots\mu_{r-1}}, \tag{3.7}$$

where $[\cdots]$ indicates antisymmetrization. The trace of the left-hand side (lhs) with respect to ν and μ_1 is zero because $h_{\mu_1\cdots\mu_r}$ is both traceless and divergence-free. Thus we have

$$f_{\mu_1\cdots\mu_{r-1}} = 0. \tag{3.8}$$

Therefore

$$\nabla_{[\mu}h_{\nu]\mu_1\cdots\mu_{r-1}} = 0. \tag{3.9}$$

By multiplying this equation by ∇^μ we find

$$\nabla_\alpha\nabla^\alpha h_{\nu\mu_1\cdots\mu_{r-1}} = rh_{\nu\mu_1\cdots\mu_{r-1}}. \tag{3.10}$$

Since $\nabla_\alpha\nabla^\alpha$ is a negative-definite operator on a compact manifold, this equation implies that $h_{\mu_1\cdots\mu_r}=0$.

IV. EQUATIONS FOR STSH'S

Let $h_{\mu_1\cdots\mu_r}$ be an STSH of rank r on S^N. Then it satisfies by definition

$$g^{\alpha\beta}h_{\alpha\beta\mu_1\cdots\mu_{r-2}} = 0, \tag{4.1}$$

$$\nabla^\alpha h_{\alpha\mu_1\cdots\mu_{r-1}} = 0, \tag{4.2}$$

$$\Box h_{\mu_1\cdots\mu_r} = \nabla^\alpha\nabla_\alpha h_{\mu_1\cdots\mu_r} = [-L(L+N-1)+r]h_{\mu_1\cdots\mu_r}, \tag{4.3}$$

where $g^{\alpha\beta}$ is the inverse of the metric $g_{\alpha\beta}$ of S^N. Here L is known to be an integer larger than or equal to r.[13] We write the metric of S^N here as follows:

$$ds^2 = d\chi^2 + \sin^2\chi\,\eta_{ij}\,d\theta_i\,d\theta_j, \tag{4.4}$$

where η_{ij} is the metric tensor of S^{N-1}. We denote the covariant derivative on S^{N-1} by $\tilde\nabla_k$. Below we rewrite Eqs. (4.1)–(4.3) as tensor equations on S^{N-1} with χ dependence written explicitly. This is the first step toward the construction of the STSH's on S^N from those on S^{N-1}.

It is convenient to rescale $h_{\mu_1\cdots\mu_r}$ as follows:

$$h_{\chi\cdots\chi i_1\cdots i_m} = (\sin\chi)^{2m-r} f_{i_1\cdots i_m}, \tag{4.5}$$

where $i_1,\ldots,i_m$ run from θ_1 to θ_{N-1}. Then Eqs. (4.1)–(4.3) can be written, after a tedious calculation, as follows:

$$\eta^{jk}f_{jki_1\cdots i_m} = -[1/(\sin^2\chi)]f_{i_1\cdots i_m}, \tag{4.6}$$

$$\tilde\nabla^k f_{ki_1\cdots i_m} = -\left[\frac{\partial}{\partial\chi} + (N+m-2)\cot\chi\right] f_{i_1\cdots i_m}, \tag{4.7}$$

$$\begin{aligned}&\left[\frac{\partial^2}{\partial\chi^2} + (N-1)\cot\chi\frac{\partial}{\partial\chi} + \frac{\tilde\Box - m}{\sin^2\chi}\right] f_{i_1\cdots i_m}\\ &\quad + 2m[(\cot\chi/(\sin^2\chi)]\tilde\nabla_{(i_1}f_{i_2\cdots i_m)}\\ &\quad + m(m-1)[(\cot^2\chi)/(\sin^2\chi)]\eta_{(i_1i_2}f_{i_3\cdots i_m)}\\ &\quad = -L(L+N-1)f_{i_1\cdots i_m},\end{aligned} \tag{4.8}$$

where $\tilde\Box = \tilde\nabla^i\tilde\nabla_i$. Notice that these equations do not depend on r.

V. SYMMETRIC TRACELESS TENSORS INDUCED BY STSH'S

Let $\tilde f_{i_1\cdots i_m}$ be an STSH of rank m on S^{N-1} $(N\geqslant 3)$. It is known that[13]

$$\tilde\Box\tilde f_{i_1\cdots i_m} = [-l(l+N-2)+m]\tilde f_{i_1\cdots i_m}, \tag{5.1}$$

where l $(\geqslant m)$ is an integer. We define the nth symmetric traceless tensor $T^{(n)}_{i_1\cdots i_{m+n}}$ induced by $\tilde f_{i_1\cdots i_m}$ as follows: $T^{(n)}_{i_1\cdots i_{m+n}}$ is a linear combination of symmetric tensors of the form

$$\eta_{(i_1i_2}\cdots\eta_{i_{2k-1}i_{2k}}\tilde\nabla_{i_{2k+1}}\cdots\tilde\nabla_{i_n}\tilde f_{i_{n+1}\cdots i_{m+n})},$$

where η_{ij} is the metric tensor of S^{N-1} and $\tilde\nabla_i$ is the covariant derivative there. It is traceless and the coefficient of $\tilde\nabla_{(i_1}\cdots\tilde\nabla_{i_n}\tilde f_{i_{n+1}\cdots i_{m+n})}$ is 1.

For example,

$$T^{(0)}_{i_1\cdots i_m} = \tilde f_{i_1\cdots i_m}, \tag{5.2a}$$

$$T^{(1)}_{i_1\cdots i_{m+1}} = \tilde\nabla_{(i_1}\tilde f_{i_2\cdots i_{m+1})}. \tag{5.2b}$$

$$\begin{aligned}T^{(2)}_{i_1\cdots i_{m+2}} &= \tilde\nabla_{(i_1}\tilde\nabla_{i_2}\tilde f_{i_3\cdots i_{m+2})} + [(l-m)(l+m+N-2)\\ &\quad\times(N+2m-1)^{-1}]\eta_{(i_1i_2}\tilde f_{i_3\cdots i_{m+2})}.\end{aligned} \tag{5.2c}$$

Our analysis here is on S^{N-1} instead of S^N merely for later convenience. We will find in the next section that the STSH's on S^N depend on those on S^{N-1} through tensors $T^{(n)}_{i_1\cdots i_{m+n}}$.

We introduce the following shorthand notation for a totally symmetric tensor $T_{i_1\cdots i_m}$:

$$T_m = T_{i_1\cdots i_m}, \tag{5.3a}$$

$$\eta^k\tilde\nabla^{n-2k}T_m = \eta_{(i_1i_2}\cdots\eta_{i_{2k-1}i_{2k}}\tilde\nabla_{i_{2k+1}}\cdots\tilde\nabla_{i_n}T_{i_{n+1}\cdots i_{m+n})}, \tag{5.3b}$$

$$\tilde\nabla\cdot T_m = \tilde\nabla^k T_{ki_1\cdots i_{m-1}}, \tag{5.3c}$$

$$\mathrm{Tr}\,T_m = \eta^{kl}T_{kli_1\cdots i_{m-2}}. \tag{5.3d}$$

Let us first derive some useful equations using the elementary formula

$$[\tilde\nabla_j,\tilde\nabla_k]A^{i_1\cdots i_n} = \sum_{s=1}^{n} R^{i_s}{}_{ljk}A^{i_1\cdots \overset{l}{\underset{s}{}}\cdots i_n}, \tag{5.4}$$

where $R^i{}_{jkl}$ is the Riemann tensor. The Riemann tensor for a sphere is

$$R^i{}_{jkl} = \delta^i_k\eta_{jl} - \delta^i_l\eta_{jk}. \tag{5.5}$$

Then by calculating the commutator $[\tilde\nabla_{i_1},\tilde\Box]$ we have

$$\begin{aligned}\tilde\Box\tilde\nabla_{i_1}T_{i_2\cdots i_n} &= \tilde\nabla_{i_1}\tilde\Box T_{i_2\cdots i_n} + (N-2)\tilde\nabla_{i_1}T_{i_2\cdots i_n}\\ &\quad + 2\sum_{s=2}^{n}(\tilde\nabla_{i_s}T_{i_2\cdots \underset{s}{i_1}\cdots i_n} - \eta_{i_1i_s}\tilde\nabla^k T_{i_2\cdots \underset{s}{k}\cdots i_n}),\end{aligned} \tag{5.6}$$

where $T_{i_1\cdots i_{n-1}}$ is a totally symmetric tensor. By symmetrizing this equation we obtain

$$\begin{aligned}\tilde\Box\tilde\nabla_{(i_1}T_{i_2\cdots i_n)} &= \tilde\nabla_{(i_1}\tilde\Box T_{i_2\cdots i_n)}\\ &\quad + (2n+N-4)\tilde\nabla_{(i_1}T_{i_2\cdots i_n)}\\ &\quad - 2(n-1)\eta_{(i_1i_2}\tilde\nabla^k T_{i_3\cdots i_n)k},\end{aligned} \tag{5.7}$$

which becomes, in the abbreviated notation,

$$\tilde{\Box}\tilde{\nabla}T_{n-1} = \tilde{\nabla}\tilde{\Box}T_{n-1} + (2n+N-4)\tilde{\nabla}T_{n-1} - 2(n-1)\eta\tilde{\nabla}T\cdot T_{n-1}. \quad (5.8)$$

In a similar manner we have

$$(n+1)\tilde{\nabla}\cdot(\tilde{\nabla}T_n) = [\tilde{\Box} + n(N+n-3)]T_n - n(n-1)\eta \operatorname{Tr} T_n + n\tilde{\nabla}\tilde{\nabla}\cdot T_n. \quad (5.9)$$

Since $\tilde{\Box}T^{(n)}_{m+n}$ and $\tilde{\nabla}\cdot T^{(n)}_{m+n}$ are traceless, they are proportional to $T^{(n)}_{m+n}$ itself and $T^{(n-1)}_{m+n-1}$, respectively. We will prove below that

$$\tilde{\Box}T^{(n)}_{m+n} = a^{(n)}T^{(n)}_{m+n}, \quad (5.10a)$$

$$\tilde{\nabla}\cdot T^{(n)}_{m+n} = c^{(n)}T^{(n-1)}_{m+n-1}, \quad (5.10b)$$

where

$$a^{(n)} = n(2m+n+N-4) - l(l+N-2) + m + n, \quad (5.11a)$$

$$c^{(n)} = \frac{n}{m+n}\,\frac{N+2m+n-4}{N+2(m+n)-5} \times[(m+n-1)(m+n+N-3) - l(l+N-2)]. \quad (5.11b)$$

The proof is by induction. It is obvious for $n=0$. Suppose that these formulas are valid for $n\to 1,\dots,n-1$. We first note that

$$T^{(n)}_{m+n} = \tilde{\nabla}T^{(n-1)}_{m+n-1} - b^{(n)}\eta T^{(n-2)}_{m+n-2}, \quad (5.12)$$

where $b^{(n)}$ is a constant, because $\operatorname{Tr}\operatorname{Tr}\tilde{\nabla}T^{(n-1)}_{m+n-1} = 0$. By taking the trace of this equation we find

$$b^{(n)} = \{(m+n-1)/[N+2(m+n)-5]\}c^{(n-1)}. \quad (5.13)$$

Then $\tilde{\Box}T^{(n)}_{m+n}$ and $\tilde{\nabla}\cdot T^{(n)}_{m+n}$ can be evaluated by using (5.8) and (5.9). We obtain

$$a^{(n)} = a^{(n-1)} + 2(m+n) + N - 4, \quad (5.14)$$

$$c^{(n)} = [1/(m+n)][a^{(n-1)} + (m+n-1) \times(N+m+n-4) + (m+n-1)c^{(n-1)} - 2b^{(n)}]. \quad (5.15)$$

By substituting the explicit expressions of $a^{(n-1)}$, $c^{(n-1)}$, and $b^{(n)}$ we obtain (5.11a) and (5.11b).

As long as $m+n\geqslant l$, $c^{(n)}$ is nonzero. (Notice that $c^{(2)}=0$ for $m=0$ if $N=2$. This is why we excluded $N=2$.) We conclude from this that $T^{(n)}_{m+n}$ is nonzero for $n\leqslant l-m$. If $m+n=l+1$, $T^{(n)}_{m+n}$ is both traceless and divergence-free. This implies that $T^{(n)}_{m+n}=0$, which can be proved as follows. Let us define the inner product of two tensors $A_{i_1\cdots i_m}$ and $B_{i_1\cdots i_m}$ by

$$(A,B) = \int d\Omega_{N-1}\, A^*_{i_1\cdots i_m}B^{i_1\cdots i_m}, \quad (5.16)$$

where $d\Omega_{N-1}$ is the volume element of S^{N-1}. Obviously (A,A) is positive definite and $(A,A)=0$ implies $A_{i_1\cdots i_m}=0$. Since $T^{(n)}_{m+n}$ is traceless,

$$(T^{(n)},T^{(n)}) = \int d\Omega_{N-1}\, T^{(n)*}_{i_1\cdots i_{m+n}}\tilde{\nabla}^{i_1}\cdots\tilde{\nabla}^{i_n}\tilde{f}^{i_{n+1}\cdots i_{n+m}}. \quad (5.17)$$

For $m+n=l+1$ we obtain $(T^{(n)},T^{(n)})=0$ by partial integration because then $T^{(n)}_{m+n}$ is divergence-free as well. Thus $T^{(l-m+1)}_{l+1}=0$.

Now this implies that $\tilde{\nabla}^n\tilde{f}_m$ can be expressed as a linear combination of tensors of the form $\eta^kT^{(n-2k)}_{m+n-2k}$ $(k>0)$ if $m+n=l+1$. Then it is clear that $\tilde{\nabla}^{n+j}\tilde{f}_m$ with any nonnegative integer j can be expressed in a similar manner. Since one cannot make a traceless tensor as a linear combination of the tensors of the form $\eta^kT^{(n)}_{m+n}$ $(k>0,\ m+n\leqslant l)$, $T^{(n)}_{m+n}$ must be zero for $m+n>l$.

Finally, it can readily be shown that two induced symmetric traceless tensors are orthogonal to each other if they are induced by STSH's of different ranks or by those which are of the same rank but orthogonal to each other.

VI. ANALYSIS OF THE EQUATIONS FOR THE STSH'S

In this section we derive the formulas which give the STSH's on S^N in terms of those on S^{N-1} starting from the equations derived in Sec. IV.

Suppose that the first nonzero component of an STSH $h_{\mu_1\cdots\mu_r}$ is $h_{\chi\cdots\chi i_1\cdots i_s}$, that is, $h_{\chi\cdots\chi i_1\cdots i_m}=0$ for $m<s$ and $h_{\chi\cdots\chi i_1\cdots i_s}\neq 0$. Then from (4.8) we obtain

$$\left[\frac{\partial^2}{\partial\chi^2} + (N-1)\cot\chi\frac{\partial}{\partial\chi} + \frac{\tilde{\Box}-s}{\sin^2\chi}\right]f_s = -L(L+N-1)f_s, \quad (6.1)$$

where we have used the abbreviated notation introduced in the previous section. That is,

$$f_s = f_{i_1\cdots i_s}. \quad (6.2)$$

Because of the completeness of the STSH's on S^{N-1} we can assume without loss of generality that

$$f_s \propto \tilde{f}^{l\sigma}_s(\theta_{N-1},\dots,\theta_1), \quad (6.3)$$

where $\tilde{f}^{l\sigma}_s$ is an STSH on S^{N-1} of rank s with

$$\tilde{\Box}\tilde{f}^{l\sigma}_s = [-l(l+N-2)+s]\tilde{f}^{l\sigma}_s, \quad (6.4)$$

where σ represents the labels other than l. The nonsingular solution of Eq. (6.1) is

$$f_s = {}_N\bar{P}^l_L(\chi)\tilde{f}^{l\sigma}_s \quad (L\geqslant l\geqslant s). \quad (6.5)$$

(We do not specify the normalization in this section.)

To find the other components which accompany f_s given above we postulate that

$$f_{s+n} = c^{(n)}_0T^{(n)}_{s+n} - c^{(n)}_1\eta T^{(n-2)}_{s+n-2} + \cdots + (-1)^kc^{(n)}_k\eta^kT^{(n-2k)}_{s+n-2k} + \cdots, \quad (6.6)$$

where the tensors $T^{(n-2k)}_{s+n-2k}$ are the symmetric traceless tensors induced by $\tilde{f}^{l\sigma}_s$. The coefficients $c^{(n)}_k$ are functions of χ only. Since $T^{(n)}_{s+n}$ is zero if $s+n>l$, we assume here that $l\geqslant r$ and will discuss the cases where $l<r$ later in this section.

Now let us derive the equations satisfied by $c^{(n)}_k$. First we examine Eq. (4.6), which can be written as

$$\operatorname{Tr}f_{m+2} = -[1/(\sin^2\chi)]f_m. \quad (6.7)$$

The trace of (6.6) is

$$\mathrm{Tr} f_{s+n} = \sum_{k=1}^{[n/2]} (-1)^k c_k^{(n)} \eta^{k-1} \times \frac{2k[N-1+2(s+n-2k)]}{(s+n-2k)(s+n-2k-1)} T_{s+n-2k}^{(n-2k)}. \tag{6.8}$$

Equation (6.7) is satisfied if

$$c_k^{(n)} = \frac{(s+n)(s+n-1)}{2k[N+2(s+n-k)-3]} \frac{1}{\sin^2\chi} c_{k-1}^{(n-2)}. \tag{6.9}$$

By iterating this equation we find

$$c_k^{(n)} = \frac{1}{2^k k!} \frac{(s+n)!}{(s+n-2k)!} \times \left[\prod_{a=1}^{k} \frac{1}{N+2(s+n-2k+a-1)}\right] \frac{c_0^{(n-2k)}}{\sin^{2k}\chi}. \tag{6.10}$$

Equation (4.7) can be written as

$$\tilde{\nabla}\cdot f_{m+1} = -\left[\frac{\partial}{\partial\chi} + (N+m-2)\cot\chi\right] f_m. \tag{6.11}$$

To calculate the divergence of (6.6) we first note that

$$\tilde{\nabla}\cdot(\eta^k T_{s+n'}^{(n')}) = \frac{2k}{s+n} \eta^{k-1} \tilde{\nabla} T_{s+n'}^{(n')} + \frac{s+n'}{s+n} \eta^k \tilde{\nabla}\cdot T_{s+n'}^{(n')}, \tag{6.12}$$

where $n' = n - 2k$. We find from (5.10b)

$$\tilde{\nabla}\cdot T_{s+n'}^{(n')} = \frac{n}{s+n} \frac{N+2s+n-4}{N+2(s+n)-5} \times [(s+n-1)(s+n+N-3) - l(l+N-2)] T_{s+n'-1}^{(n'-1)}. \tag{6.13}$$

Here $\tilde{\nabla} T_{s+n'}^{(n')}$ can be written by using (5.12) as

$$\tilde{\nabla} T_{s+n'}^{(n')} = T_{s+n'+1}^{(n'+1)} + \alpha\eta T_{s+n'-1}^{(n'-1)}, \tag{6.14}$$

where

$$\alpha = \{n'(N+2s+n'-4)/[(N+2s+2n'-3) \times (N+2s+2n'-5)]\}[(s+n'-1) \times (s+n'+N-3) - l(l+N-2)]. \tag{6.15}$$

Hence we have

$$\tilde{\nabla}\cdot(\eta^k T_{s+n'}^{(n')}) = \frac{2k}{s+n} \eta^{k-1} T_{s+n'+1}^{(n'+1)} + \frac{n'(N+2s+n'-4)}{(N+2s+2n'-3)(N+2s+2n'-5)} \times [(s+n'-1)(s+n'+N-3) - l(l+N-2)] \eta^k T_{s+n'-1}^{(n'-1)}. \tag{6.16}$$

By using this formula Eq. (6.11) can be reduced to

$$\frac{(n'+1)(N+2s+n'-3)[N+2(s+n-k)-1]}{(s+n+1)(N+2s+2n'-3)(N+2s+2n'-1)} \times [(s+n')(s+n'+N-2) - l(l+N-2)] c_k^{(n+1)} - \frac{2(k+1)}{s+n+1} c_{k+1}^{(n+1)} = -\left[\frac{\partial}{\partial\chi} + (N+s+n-2)\cot\chi\right] c_k^{(n)}. \tag{6.17}$$

Then substitution of (6.10) yields

$$\frac{(n'+1)(N+2s+n'-3)}{(s+n'+1)(N+2s+2n'-3)} \times [(s+n')(s+n'+N-2) - l(l+N-2)] c_0^{(n'+1)} - \frac{s+n'}{N+2s+2n'-3} \frac{1}{\sin^2\chi} c_0^{(n'-1)} = -\left[\frac{\partial}{\partial\chi} + (N+s+n'-2)\cot\chi\right] c_0^{(n')}. \tag{6.18}$$

Notice that n and k appear only through $n' = n - 2k$. Otherwise this equation would be inconsistent. By letting

$$c_0^{(0)} = {}_N\bar{P}_L^l(\chi) \tag{6.19}$$

and $c_0^{(-1)} = 0$, $c_k^{(n)}$ for arbitrary n and k can be obtained from (6.10) and (6.18) as long as $s + n \geqslant l$, which is the case for all n here because we have assumed $l \geqslant r$. Then they satisfy Eqs. (6.7) and (6.11).

Now let us show that Eq. (4.8), i.e.,

$$K_m = \left[\frac{\partial^2}{\partial\chi^2} + (N-1)\cot\chi \frac{\partial}{\partial\chi} + \frac{\tilde{\Box} - m}{\sin^2\chi}\right] f_m + 2m \frac{\cot\chi}{\sin^2\chi} \tilde{\nabla} f_{m-1} + m(m-1) \frac{\cot^2\chi}{\sin^2\chi} \eta f_{m-2} + L(L+N-1) f_m = 0, \tag{6.20}$$

is satisfied by f_m obtained by the procedure explained above. The proof is by induction with respect to m. Obviously this equation is satisfied for $m \leqslant s$. Suppose it is satisfied for $m \to m-1$ and $m-2$. Now K_m can be expanded as

$$K_m = \sum_{k=0}^{[(m-s)/2]} g_k(\chi) \eta^k T_{m-2k}^{(m-s-2k)}. \tag{6.21}$$

It is sufficient to show that the coefficients $g_k(\chi)$ are zero. If $\mathrm{Tr}\, K_m = 0$, then $g_k(\chi) = 0$ for $k \neq 0$. If $\tilde{\nabla}\cdot K_m$ is also zero, then $g_0(\chi) = 0$. After a tedious calculation we find

$$\mathrm{Tr}\, K_m = -[1/(\sin^2\chi)] K_{m-2}, \tag{6.22}$$

$$\tilde{\nabla}\cdot K_m = -\left[\frac{\partial}{\partial\chi} + (N+m-3)\cot\chi\right] K_{m-1}, \tag{6.23}$$

which are zero by the assumption of induction.

Next let us examine what happens if we let $l < r$. What we will find is that there are no solutions in this case. Since f_{s+n} with $s + n \leqslant l$ can be constructed without any problem, we examine the component with $s + n = l + 1$. Let us write f_{l+1} as follows:

$$f_{l+1} = f'_{l+1} + f''_{l+1}, \tag{6.24}$$

where

$$f''_{l+1} = -c_1^{(l-s+1)} \eta T_{l-1}^{(l-s-1)} + \cdots + (-1)^k c_k^{(l-s+1)} \eta^k T_{l-2k+1}^{(l-s-2k-1)} + \cdots. \tag{6.25}$$

Here f'_{l+1} would be $c_0^{(l-s+1)} T_{l+1}^{(l-s+1)}$ if $T_{l+1}^{(l-s+1)}$ was nonzero. The trace and the divergence of f'_{l+1} can be found by considering what those of $c_0^{(l-s+1)}T_{l+1}^{(l-s+1)}$ would be. Thus we have

$$\mathrm{Tr} f'_{l+1} = 0, \tag{6.26}$$

$$\tilde{\nabla}\cdot f'_{l+1} = \left\{\frac{1}{N+2l-3}\frac{c_0^{(l-s-1)}}{\sin^2\chi} - \left[\frac{\partial}{\partial\chi} + (N+l-2)\cot\chi\right] c_0^{(l-s)}\right\} T_l^{(l-s)}. \tag{6.27}$$

The coefficient of $T_l^{(l-s)}$ on the rhs is nonzero as is shown in Appendix B.

Now we find

$$\int d\Omega_{N-1}\, T_{i_1\cdots i_l}^{(l-s)*}\tilde{\nabla}_k f'^{ki_1\cdots i_l} = 0 \tag{6.28}$$

by partial integration because

$$\tilde{\nabla} T_l^{(l-s)} \propto \eta T_{l-1}^{(l-s-1)}. \tag{6.29}$$

(Recall that $T_{l+1}^{(l-s+1)} = 0$.) Therefore one cannot find f'_{l+1} satisfying Eq. (6.27). [In general one can add terms proportional to induced symmetric traceless tensors other than $T_l^{(l-s)}$ on the rhs of Eq. (6.27), but the above proof of the absence of the tensor f'_{l+1} satisfying this equation is still valid because of the orthogonality of induced symmetric traceless tensors.] Therefore one cannot construct an STSH starting from $f_s = {}_N\bar{P}_L^l(\chi)\tilde{f}_s^{l\sigma}$ with $l<r$. Hence

$$0 \leqslant s \leqslant r \leqslant l \leqslant L. \tag{6.30}$$

We have given a prescription for constructing an STSH on S^N from one on S^{N-1}. Now it is possible to construct an STSH on S^N starting from one on S^2 which is labeled by $[l_2,r_2;l_1]$, where r_2 is its rank and l_2 is the angular momentum on S^2 and l_1 is the angular momentum on S^1. From this one can construct an STSH on S^3 of rank r_3 with angular momentum l_3. Then one can construct an STSH of rank r_4 with angular momentum l_4 on S^4 from the resulting STSH on S^3, and so forth. An STSH on S^N thus constructed can be labeled by $[l_N,r_N;...;l_2,r_2;l_1]$. It can readily be shown that these STSH's are mutually orthogonal with respect to the inner product defined by (5.16).

The STSH's on S^N form a unitary representation of SO($N+1$). Equation (6.30) constitutes the branching rule under the decomposition SO($N+1$)$\supset$SO(N). From this branching rule one can see that the STSH's on S^N of rank r with $\Box = -L(L+N-1)+r$ form the unitary representation which corresponds to the Young diagram labeled by $[L,r,0,...,0]$ (see Ref. 13).

VII. TRANSFORMATION PROPERTIES OF STSH'S UNDER SO(N+1)

In this section we investigate transformation properties of a certain class of STSH's under SO($N+1$). We will use the result of this section to find the normalization factors for STSH's in the next section. Since the normalization factors for $N=2$ are already calculated in Sec. III, we restrict N to be larger than or equal to 3.

Let $h_{\mu_1\cdots\mu_r}^{(mlq\sigma)}$ be an STSH of rank r with $\Box = -L(L+N-1)+r$ which satisfies the following conditions:

$$\text{(i)}\quad h_{\chi\cdots\chi i_1\cdots i_{m'}}^{(mlq\sigma)} = 0 \quad (m'<m); \tag{7.1a}$$

$$\text{(ii)}\quad h_{\chi\cdots\chi\underbrace{\theta\cdots\theta}_{m}}^{(mlq\sigma)} = (\sin\chi)^{2m-r}(\sin\theta)^{-m}Y_{Llq\sigma}(\chi,\boldsymbol{\theta}), \tag{7.1b}$$

where $Y_{Llq\sigma}(\chi,\boldsymbol{\theta})$ is a scalar spherical harmonic on S^N of which the angular momenta on S^N, S^{N-1}, and S^{N-2} are L, l, and q, respectively, and $\boldsymbol{\theta} = (\theta,\theta_{N-1},...,\theta_1)$; and (iii) $h_{\chi\cdots\chi i_1\cdots i_m}^{(mlq\sigma)}$ is an STSH on S^{N-1} that satisfies

$$\tilde{\Box} h_{\chi\cdots\chi i_1\cdots i_m}^{(mlq\sigma)} = [-l(l+N-2)+m] h_{\chi\cdots\chi i_1\cdots i_m}^{(mlq\sigma)}. \tag{7.1c}$$

Here σ represents the labels other than those explicitly written. Clearly, $h_{\mu_1\cdots\mu_r}^{(mlq\sigma)}$ can be labeled by $[L,r;l,m;q,0;...]$. (See the end of the previous section.)

Next let us parametrize S^N in ($N+1$)-dimensional Euclidean space as

$$\begin{aligned} x_1 &= \sin\chi\sin\theta\sin\theta_{N-2}\cdots\sin\theta_2,\\ x_2 &= \sin\chi\sin\theta\sin\theta_{N-2}\cdots\cos\theta_2,\\ &\vdots\\ x_N &= \sin\chi\cos\theta,\\ x_{N+1} &= \cos\chi. \end{aligned} \tag{7.2}$$

The Killing vectors that form the group SO($N+1$) are given by

$$X_{(ij)} = x_i\frac{\partial}{\partial x_j} - x_j\frac{\partial}{\partial x_i}. \tag{7.3}$$

The SO($N+1$) transformations of STSH's are conveniently described by their Lie derivatives with respect to the Killing vectors. The Lie derivative of a tensor $h_{\mu_1\cdots\mu_r}$ with respect to a vector Y^μ is defined by

$$L_Y = Y^\lambda\partial_\lambda + \partial_{\mu_1}Y^\lambda h_{\lambda\mu_2\cdots\mu_r} + \partial_{\mu_2}Y^\lambda h_{\mu_1\lambda\cdots\mu_r} + \cdots + \partial_{\mu_r}Y^\lambda h_{\mu_1\cdots\mu_{r-1}\lambda}. \tag{7.4}$$

If Y^μ is a Killing vector, tensor operators such as ∇_μ or $g_{\mu\nu}$ commute with L_Y. Therefore, if $h_{\mu_1\cdots\mu_r}$ is an STSH with $\Box = -L(L+N-1)+r$, so is $L_Y h_{\mu_1\cdots\mu_r}$. Below we will study the transformation of STSH's $h_{\mu_1\cdots\mu_r}^{(mlq\sigma)}$ generated by the Killing vector $X = X_{(N+1,N)}$, which can be written as

$$X = \cos\theta\frac{\partial}{\partial\chi} - \cot\chi\sin\theta\frac{\partial}{\partial\theta}. \tag{7.5}$$

The result is Eq. (7.28).

Let us define, for a given tensor $h_{\mu_1\cdots\mu_r}$,

$$h_{(m)} = h_{\chi\cdots\chi\underbrace{\theta\cdots\theta}_{m}}. \tag{7.6}$$

By definition of the Lie derivative (7.4) we find

$$\begin{aligned}\delta h_{(m)} = L_X h_{(m)} &= \left(\cos\theta\frac{\partial}{\partial\chi} - \cot\chi\sin\theta\frac{\partial}{\partial\theta}\right)h_{(m)}\\ &\quad - m(\sin\theta h_{(m-1)} + \cot\chi\cos\theta h_{(m)})\\ &\quad + (r-m)\frac{\sin\theta}{\sin^2\chi}h_{(m+1)}. \end{aligned} \tag{7.7}$$

We define $f_{(m)}$ by

$$h_{(m)} = (\sin\chi)^{2m-r} f_{(m)}. \tag{7.8}$$

Then

$$\delta f_{(m)} = \left(\cos\theta \frac{\partial}{\partial\chi} - \cot\chi \sin\theta \frac{\partial}{\partial\theta}\right) f_{(m)} + (r-m)(\sin\theta f_{(m+1)} - \cot\chi\cos\theta f_{(m)}) + m[(\sin\theta)/(\sin^2\chi)] f_{(m-1)}. \tag{7.9}$$

First by letting $f_{(m)} \to f^{(mlq\sigma)}_{(m-1)}$ on the lhs of this equation we obtain

$$\delta f^{(mlq\sigma)}_{(m-1)} = (r-m+1)\sin\theta f^{(mlq\sigma)}_{(m)}. \tag{7.10}$$

Since

$$f^{(mlq\sigma)}_{(m)} = (\sin\theta)^{-1} f^{(m-1,lq\sigma)}_{(m-1)}, \tag{7.11}$$

one has

$$\delta f^{(mlq\sigma)}_{(m-1)} = (r-m+1) f^{(m-1,lq\sigma)}_{(m-1)}. \tag{7.12}$$

Thus we obtain

$$L_\chi h^{(mlq\sigma)}_{\mu_1\cdots\mu_r} = (r-m+1) h^{(m-1,lq\sigma)}_{\mu_1\cdots\mu_r} + \text{terms of the form } h^{(m',\ldots)}_{\mu_1\cdots\mu_r} \quad (m'\geqslant m). \tag{7.13}$$

STSH's $h^{(m',\ldots)}_{\mu_1\cdots\mu_r}$ with $m'<m-1$ are absent on the rhs because $h^{(mlq\sigma)}_{\chi\cdots\chi i_1\cdots i_{m'}}$ with $m'<m$ are zero and L_χ changes at most one index of the tensor. Since STSH's $h^{(mlq\sigma)}_{\mu_1\cdots\mu_r}$ form an orthogonal basis of a representation of SO$(N+1)$ as was remarked in the previous section, we must have $h^{(m+1,lq\sigma)}_{\mu_1\cdots\mu_r}$ on the rhs because of the unitarity of the representation. There is no other term proportional to $h^{(m'lq\sigma)}_{\mu_1\cdots\mu_r}$ with $m'>m$ for the same reason. Thus we expect

$$L_\chi h^{(mlq\sigma)}_{\mu_1\cdots\mu_r} = (r-m+1) h^{(m-1,lq\sigma)}_{\mu_1\cdots\mu_r} + c_d h^{(m+1,lq\sigma)}_{\mu_1\cdots\mu_r} + \text{terms of the form } h^{(m,\ldots)}_{\mu_1\cdots\mu_r}. \tag{7.14}$$

To determine the terms of the form $h^{(m,\ldots)}_{\mu_1\cdots\mu_r}$ we examine $\delta f^{(mlq\sigma)}_{(m)}$, which is

$$\delta f^{(mlq\sigma)}_{(m)} = \left(\cos\theta\frac{\partial}{\partial\chi} - \cot\chi\sin\theta\frac{\partial}{\partial\theta}\right) f^{(mlq\sigma)}_{(m)} + (r-m)[\sin\theta f^{(mlq\sigma)}_{(m+1)} - \cot\chi\cos\theta f^{(mlq\sigma)}_{(m)}]. \tag{7.15}$$

By substituting $c_0^{(1)}$ obtained from (6.18) with $s=m$ in

$$f^{(mlq\sigma)}_{(m+1)} = c_0^{(1)} \frac{\partial}{\partial\theta} (\sin\theta)^{-m} Y_{lq\sigma}(\boldsymbol{\theta}), \tag{7.16}$$

we find

$$f^{(mlq\sigma)}_{(m+1)} = \frac{m+1}{(l-m)(N+m+l-2)} \times\left[\frac{\partial}{\partial\chi} + (N+m-2)\cot\chi\right]\frac{\partial}{\partial\theta} f^{(mlq\sigma)}_{(m)}. \tag{7.17}$$

The contribution from $(r-m+1)h^{(m-1,lq\sigma)}_{\mu_1\cdots\mu_r}$ in (7.15) is

$$(r-m+1) f^{(m-1,lq\sigma)}_{(m)} = \frac{m(r-m+1)}{(l-m+1)(N+m+l-3)} \times\left[\frac{\partial}{\partial\chi} + (N+m-3)\cot\chi\right] \times\frac{\partial}{\partial\theta}(\sin\theta f^{(mlq\sigma)}_{(m)}). \tag{7.18}$$

Subtraction of this from (7.15) leads to

$$\delta f^{(mlq\sigma)}_{(m)} - (r-m+1) f^{(m-1,lq\sigma)}_{(m)} = \frac{(\sin\theta)^{-m}}{2l+N-2}\left[\frac{(l-r)(l+1)}{(l-m)(l-m+1)} T^{(-)} - \frac{(l+N+r-2)(l+N-3)}{(l+N+m-2)(l+N+m-3)} T^{(+)}\right] Y_{Llq\sigma}, \tag{7.19}$$

where

$$T^{(-)} = -\left[\frac{\partial}{\partial\chi} + (N+l-2)\cot\chi\right] \times\left(\sin\theta\frac{\partial}{\partial\theta} - l\cos\theta\right), \tag{7.20a}$$

$$T^{(+)} = -\left(\frac{\partial}{\partial\chi} - l\cot\chi\right) \times\left[\sin\theta\frac{\partial}{\partial\theta} + (N+l-2)\cos\theta\right]. \tag{7.20b}$$

One can show by using the formulas given in Appendix C that

$$T^{(-)} Y_{Llq\sigma} = k_- Y_{L,l-1,q\sigma}, \tag{7.21a}$$

$$T^{(+)} Y_{Llq\sigma} = k_+ Y_{L,l+1,q\sigma}, \tag{7.21b}$$

where

$$k_- = [((2l+N-2)/(2l+N-4)) \times(l-q)(L+q+N-3) \times(L-l+1)(L+l+N-2)]^{1/2}, \tag{7.22a}$$

$$k_+ = [((2l+N-2)/(2l+N)) \times(l-q+1)(L+q+N-2) \times(L-l)(L+l+N-1)]^{1/2}. \tag{7.22b}$$

Thus we obtain

$$L_\chi h^{(mlq\sigma)}_{(m)} - (r-m+1) h^{(m-1,lq\sigma)}_{(m)} = c_+ h^{(m,l+1,q\sigma)}_{(m)} + c_- h^{(m,l-1,q\sigma)}_{(m)}, \tag{7.23}$$

where

$$c_- = \frac{(l+N+r-2)(l+N-3)}{(l+N+m-2)(l+N+m-3)} \times\left[\frac{(l-q)(l+q+N-3)}{(2l+N-2)(2l+N-4)} \times(L-l+1)(L+l+N-2)\right]^{1/2}, \tag{7.24a}$$

$$c_+ = -\frac{(l-r)(l+1)}{(l-m)(l-m+1)} \times\left[\frac{(l-q+1)(l+q+N-2)}{(2l+N)(2l+N-2)} \times(L-l)(L+l+N-1)\right]^{1/2}. \tag{7.24b}$$

To find the coefficient c_d in (7.14) we use

$$\delta f^{(mlq\sigma)}_{(m+1)} = \left(\cos\theta\frac{\partial}{\partial\chi} - \cot\chi\sin\theta\frac{\partial}{\partial\theta}\right) f^{(mlq\sigma)}_{(m+1)} + (r-m-1)[\sin\theta f^{(mlq\sigma)}_{(m+2)} - \cot\chi\cos\theta f^{(mlq\sigma)}_{(m+1)}] - (m+1)[(\sin\theta)/(\sin^2\chi)] f^{(mlq\sigma)}_{(m)} \tag{7.25}$$

and

$$c_d f^{(m+1,lq\sigma)}_{(m+1)} = \delta f^{(mlq\sigma)}_{(m+1)} - (r-m+1) f^{(m-1,lq\sigma)}_{(m+1)} - c_+ f^{(m,l+1,q\sigma)}_{(m+1)} - c_- f^{(m,l-1,q\sigma)}_{(m+1)}. \tag{7.26}$$

Here $f^{(mlq\sigma)}_{(m+2)}$ and $f^{(m-1,lq\sigma)}_{(m+1)}$ can be obtained by using the formulas in the previous section. Functions $f^{(mlq\sigma)}_{(m+1)}$ and $f^{(m,l\pm 1,q\sigma)}_{(m+1)}$ can be found by Eq. (7.17). After a tedious calculation, of which the details are given in Appendix D, we find

$$c_d = \frac{(m+1)(N+r+m-3)(N+m-4)(q-m)(q+m+N-3)}{(N+2m-2)(N+2m-4)(l-m)(l+m+N-2)(l-m+1)(l+m+N-3)} \times [(m-1)(m+N-2) - L(L+N-1)]. \tag{7.27}$$

Thus

$$L_\chi h^{(mlq\sigma)}_{\mu_1\cdots\mu_r} = (r-m+1) h^{(m-1,lq\sigma)}_{\mu_1\cdots\mu_r} + c_d h^{(m+1,lq\sigma)}_{\mu_1\cdots\mu_r} + c_+ h^{(m,l+1,q\sigma)}_{\mu_1\cdots\mu_r} + c_- h^{(m,l-1,q\sigma)}_{\mu_1\cdots\mu_r}, \tag{7.28}$$

where c_d, c_-, and c_+ are given in (7.27), (7.24a), and (7.24b), respectively.

VIII. NORMALIZATION FACTORS FOR STSH'S

In this section we calculate the normalization factors for STSH's by using Eq. (7.28). Let $\tilde f^{(ml\sigma)}_{i_1\cdots i_m}$ be an STSH on S^{N-1} that satisfies the normalization condition

$$(\tilde f^{(ml\sigma)}, \tilde f^{(m'l'\sigma')}) = \delta_{mm'}\delta_{ll'}\delta_{\sigma\sigma'}. \tag{8.1}$$

As we have seen in Sec. VI, an STSH $h^{(rLml\sigma)}_{\mu_1\cdots\mu_r}$ on S^N can be constructed by letting

$$h^{(rLml\sigma)}_{\chi\cdots\chi i_1\cdots i_m} = {}_N\overline{P}^l_L(\chi) \tilde f^{(ml\sigma)}_{i_1\cdots i_m}, \tag{8.2a}$$

$$h^{(rLml\sigma)}_{\chi\cdots\chi i_1\cdots i_{m'}} = 0 \quad (m' < m), \tag{8.2b}$$

and deriving the other components from constraint equations. We denote the normalized STSH by $\tilde h^{(rLml\sigma)}_{\mu_1\cdots\mu_r}$ and define the normalization factor $c^{(N)}_{rLml}$ by

$$\tilde h^{(rLml\sigma)}_{\mu_1\cdots\mu_r} = c^{(N)}_{rLml} h^{(rLml\sigma)}_{\mu_1\cdots\mu_r}. \tag{8.3}$$

We will determine $c^{(N)}_{rLml}$ below by requiring that

$$(\tilde h^{(rLml\sigma)}, \tilde h^{(rL'm'l'\sigma')}) = \delta_{LL'}\delta_{mm'}\delta_{ll'}\delta_{\sigma\sigma'}. \tag{8.4}$$

Due to the invariance of the inner product under SO$(N+1)$ we have

$$(L_\chi \tilde h^{(rLml\sigma)}, \tilde h^{(rL,m+1,l\sigma)}) + (\tilde h^{(rLml\sigma)}, L_\chi \tilde h^{(rL,m+1,l\sigma)}) = 0. \tag{8.5}$$

By using (7.28) in this equation we find

$$|c^{(N)}_{rLml}|^{-2} |c^{(N-1)}_{ml0q}|^{-2} = -\frac{r-m+1}{c_d(m\to m-1)} |c^{(N)}_{rL,m-1,l}|^{-2} |c^{(N-1)}_{m-1,l0q}|^{-2}, \tag{8.6}$$

where $c_d(m\to m-1)$ is what one obtains by replacing m by $m-1$ in (7.27). The factors $|c^{(N-1)}_{ml0q}|^{-2}$ and $|c^{(N-1)}_{m-1,l0q}|^{-2}$ are present here because STSH's on S^{N-1} were not normalized in Sec. VII. Iteration of this equation yields

$$\begin{aligned}|c^{(N)}_{rLml}|^{-2}|c^{(N-1)}_{ml0q}|^{-2} &= \binom{r}{m} \frac{(N+2m-4)[(N+2m-6)\cdots(N+2)N]^2(N-2)}{(N+r+m-4)\cdots(N+r-2)(N+r-3)} [(N+m-5)\cdots(N-3)(N-4)]^{-1} \\ &\times [(q-m+1)(q+m+N-4)\cdots(q-1)(q+N-2)q(q+N-3)]^{-1}(l-m+1)(l+m+N-3) \\ &\times [(l-m+2)(l+m+N-4)\times\cdots\times(l-1)(l+N-1)l(l+N-2)]^2(l+1)(l+N-3) \\ &\times [(L-m+2)(L+m+N-3)\cdots(L+1)(L+N-2)]^{-1}|c^{(N)}_{rL0l}|^{-2}|c^{(N-1)}_{0l0q}|^{-2}.\end{aligned} \tag{8.7}$$

Here $|c^{(N-1)}_{0l0q}|^{-2}$ is 1 because the scalar spherical harmonics are already normalized. Then $|c^{(N)}_{rL0l}|^{-2}$ and $|c^{(N-1)}_{ml0q}|^{-2}$ can be found as follows. Let us set $r = m$ in (8.6). Since $|c^{(N)}_{rLrl}|^{-2} = 1$, we have

$$|c^{(N-1)}_{ml0q}|^{-2} = -\frac{1}{c_d(m\to m-1)} |c^{(N)}_{mL,m-1,l}|^{-2} |c^{(N-1)}_{m-1,l0q}|^{-2}. \tag{8.8}$$

Here $|c^{(N)}_{mL,m-1,l}|^{-2}$ can be directly calculated by using

$$\tilde h^{(mL,m-1,l\sigma)}_{\chi i_1\cdots i_{m-1}} = \sin^{m-2}\chi\, {}_N\overline{P}^l_L(\chi) \tilde f^{(m-1,l\sigma)}_{i_1\cdots i_{m-1}}, \tag{8.9a}$$

$$\tilde h^{(mL,m-1,l\sigma)}_{i_1\cdots i_m} = -\frac{m\sin^m\chi}{(m-1)(m+N-3) - l(l+N-2)} \left[\frac{\partial}{\partial\chi} + (N+m-3)\cot\chi\right] {}_N\overline{P}^l_L(\chi) \tilde\nabla_{(i_1} \tilde f^{(m-1,l\sigma)}_{i_2\cdots i_m)}. \tag{8.9b}$$

The result is

$$|c_{mL,m-1,l}^{(N)}|^{-2}=\frac{m(L-m+2)(L+N+m-3)}{(l-m+1)(l+N+m-3)}. \tag{8.10}$$

By substituting this in (8.8) we have

$$|c_{ml0q}^{(N-1)}|^{-2}=\frac{(N+2m-6)(l-m+2)(l+m+N-4)}{(N+m-5)(q-m+1)(q+N+m-4)}|c_{m-1,l0q}^{(N-1)}|^{-2}. \tag{8.11}$$

Iteration of this expression yields

$$|c_{ml0q}^{(N-1)}|^{-2}=\frac{(N+2m-6)\cdots(N-2)(N-4)}{(N+m-5)\cdots(N-3)(N-4)} \times\frac{(l-m+2)(l+m+N-4)\cdots l(l+N-2)(l+1)(l+N-3)}{(q-m+1)(q+m+N-4)\cdots(q-1)(q+N-2)q(q+N-3)}. \tag{8.12}$$

We find $|c_{rL0l}^{(N)}|^{-2}$ from this equation by replacing r, L, l, and N by m, l, q, and $N-1$, respectively. By substituting these in (8.7) we obtain

$$|c_{rLml}^{(N)}|^{-2}=\binom{r}{m}\frac{(N+2r-5)\cdots(N+2m-1)(N+2m-3)}{(N+r+m-4)\cdots(N+2m-2)(N+2m-3)} \times\frac{(L-m+1)(L+m+N-2)\cdots(L-r+2)(L+r+N-3)}{(l-m)(l+m+N-2)\cdots(l-r+1)(l+r+N-3)}. \tag{8.13}$$

IX. APPLICATION OF STSH'S TO THE DE SITTER GROUP SO(N,1)

One can obtain the metric of the N-dimensional de Sitter space-time from that of S^N [Eq. (1.3)] by putting $\chi=\pi/2-it$ as we have shown in the Introduction. The corresponding analytic continuation of STSH's on S^N yields symmetric, traceless, and divergence-free tensors $h_{\mu_1\cdots\mu_r}$ in de Sitter space-time that satisfy

$$\Box h_{\mu_1\cdots\mu_r}=[-L(L+N-1)+r]h_{\mu_1\cdots\mu_r}. \tag{9.1}$$

These tensors form a representation of the de Sitter group SO(N,1), or rather the de Sitter algebra. (One can write $\chi=\pi/2+it$ as well but the resulting representations are equivalent to those for $\chi=\pi/2-it$.) There are no restrictions on the value of L. However, it will be restricted once one requires the unitarity of the representation. Unitarity is very important in applying group-theoretical methods to various problems. Especially when one wants to use these tensors to describe particles in de Sitter space-time, they must form a unitary representation. In this section we will determine the values of L for which the representation is unitary. All the unitary representations of the SO(N,1) algebra have been classified by Ottoson.[17] (For $N=3$, 4, and 5, see also Ref. 18.) But it has not been clear so far which representations are realized by the tensor eigenfunctions of the LB operator in de Sitter space-time.

Unitarity is equivalent to the positive-definiteness of the norm. As the inner product that gives rise to a norm, we take

$$\langle h^{(2)},h^{(1)}\rangle=-i\int d\theta\sqrt{-g}\,[h^{(2)}_{\mu_1\cdots\mu_r}{}^{*}\nabla^0 h^{(1)\mu_1\cdots\mu_r} - h^{(1)\mu_1\cdots\mu_r}\nabla^0 h^{(2)}_{\mu_1\cdots\mu_r}{}^{*}], \tag{9.2}$$

where ∇_0 is the covariant derivative with respect to t and $dt\,d\theta\sqrt{-g}$ is the volume element of de Sitter space-time. This inner product is defined only for tensors with the same L.

Let us show that this inner product is independent of t and de Sitter invariant if $L(L+N-1)$ is real. Define

$$V^\alpha=h^{(2)}_{\mu_1\cdots\mu_r}{}^{*}\nabla^\alpha h^{(1)\mu_1\cdots\mu_r}-h^{(1)\mu_1\cdots\mu_r}\nabla^\alpha h^{(2)}_{\mu_1\cdots\mu_r}{}^{*}. \tag{9.3}$$

Then one has

$$\nabla_\alpha V^\alpha=\frac{1}{\sqrt{-g}}\partial_\alpha(\sqrt{-g}V^\alpha)=0. \tag{9.4}$$

Hence

$$\frac{\partial}{\partial t}\int d\theta\sqrt{-g}V^0=-\int d\theta\,\partial_i(\sqrt{-g}V^i)=0. \tag{9.5}$$

The invariance of this inner produce under SO(N,1) can be shown as follows. Since the Lie derivative with respect to a Killing vector Y^μ commutes with covariant derivatives, the change in V^α under the transformation corresponding to Y^μ is

$$\delta V^\alpha=L_Y V^\alpha=Y^\mu\nabla_\mu V^\alpha-(\nabla_\mu Y^\alpha)V^\mu = \nabla_\mu(Y^\mu V^\alpha-Y^\alpha V^\mu). \tag{9.6}$$

This is a conserved current but its corresponding charge vanishes because

$$\sqrt{-g}\delta V^0=\partial_i[\sqrt{-g}(Y^iV^0-Y^0V^i)]. \tag{9.7}$$

Now we start analyzing the tensors obtained by analytic continuation of STSH's. We restrict ourselves to the case where $N\geqslant 3$. We will not discuss the $N=2$ case because it is much different from other cases and a separate analysis would be necessary.

Since the quadratic Casimir invariant $L(L+N-1)+r(r+N-3)$ is real for a unitary representation, L is either a real number or $L=-(N-1)/2+i\tau$ (τ: real). One can assume without loss of generality that $L\geqslant-(N-1)/2$ when it is real and that $L=-(N-1)/2+i\tau$ ($\tau>0$) when it is imaginary since the tensor eigenfunctions do not change by letting $L\to-L-N+1$. This is because[15]

$$P_\nu^{-\mu}(x) = P_{-\nu-1}^{-\mu}(x). \tag{9.8}$$

Let us start with the scalar case. Define

$${}_N\hat{P}^l_L(t) = (\cosh t)^{-(N-2)/2} P^{-(l+(N-2)/2)}_{L+(N-2)/2}(i \sinh t). \tag{9.9}$$

Next define

$$\hat{Y}_{Ll\sigma}(t,\boldsymbol{\theta}) = {}_N\hat{P}^l_L(t) Y_{l\sigma}(\boldsymbol{\theta}). \tag{9.10}$$

The inner products can be calculated by using[15]

$$P_\nu^{-\mu}(0) = \frac{\sqrt{\pi}}{2^\mu \Gamma([(\nu+\mu)/2]+1)\Gamma((-\nu+\mu+1)/2)}, \tag{9.11a}$$

$$\left.\frac{dP_\nu^{-\mu}(x)}{dx}\right|_{x=0} = \frac{\sin[\frac{1}{2}(\nu-\mu)\pi]\Gamma((\nu-\mu)/2+1)}{2^{\mu-1}\sqrt{\pi}\Gamma((\nu+\mu+1)/2)}. \tag{9.11b}$$

The result is

$$\langle \hat{Y}_{Ll\sigma}, \hat{Y}_{Ll'\sigma'} \rangle = \frac{\pi}{\Gamma(l-L)\Gamma(l+L+N-1)} \delta_{ll'}\delta_{\sigma\sigma'}. \tag{9.12}$$

Therefore the representation is unitary if

$$L = -(N-1)/2 + i\tau \quad (\tau > 0) \tag{9.13}$$

or

$$-(N-1)/2 \leqslant L < 0. \tag{9.14}$$

When $L \geqslant 0$, there are negative-norm eigenfunctions unless L is an integer. If

$$L = 0,1,2,..., \tag{9.15}$$

then the norms of the eigenfunctions with $l \leqslant L$ are zero. Zero-norm eigenfunctions do not transform into positive-norm eigenfunctions under SO(N,1) transformations because their norms must remain zero. Therefore one can identify them with zero without causing any inconsistency. Then the positive-norm eigenfunctions form a unitary representation.

Now let us go on to the case where $r > 0$. Let $\hat{h}^{(rLml\sigma)}_{\mu_1\cdots\mu_r}$ be the tensor obtained from the normalized STSH $\tilde{h}^{(rLml\sigma)}_{\mu_1\cdots\mu_r}$ by replacing ${}_N\bar{P}^l_L(\chi)$ by ${}_N\hat{P}^l_L(t)$ and substituting $\chi = \pi/2 - it$. The inner product $\langle \hat{h}^{(rLml\sigma)}, \hat{h}^{(rLm'l'\sigma')}\rangle$ can be determined by using its invariance under the transformation generated by the Killing vector X'^μ defined by

$$X' = \cos\theta \frac{\partial}{\partial t} - \tanh t \sin\theta \frac{\partial}{\partial \theta}, \tag{9.16}$$

as in the case of the STSH's on S^N. The result is

$$\langle \hat{h}^{(rLml\sigma)}, \hat{h}^{(rLm'l'\sigma')}\rangle = \frac{(-1)^{r-m}|c^{(N)}_{rLml}|^{-2}\pi}{\Gamma(l-L)\Gamma(l+L+N-1)} \delta_{mm'}\delta_{ll'}\delta_{\sigma\sigma'}, \tag{9.17}$$

where $|c^{(N)}_{rLml}|^{-2}$ is given by (8.13). There is a factor of $(-1)^{r-m}$ because X'^μ corresponds to $-iX^\mu$ instead of $X^\mu = X^\mu_{(N+1,N)}$ given by (7.5). Let us extract the L-dependent factors as follows:

$$\langle \hat{h}^{(rLml\sigma)}, \hat{h}^{(rLml\sigma)}\rangle = \kappa \frac{[(m-1)(m+N-2) - L(L+N-1)]\cdots[(r-2)(r+N-3) - L(L+N-1)]}{\Gamma(l-L)\Gamma(l+L+N-1)}, \tag{9.18}$$

where κ is a positive constant which depends on N, r, l, and m. Thus if

$$L(L+N-1) < -(N-2), \tag{9.19}$$

then the representation is unitary. This condition can be written as

$$L = -(N-1)/2 + i\tau \quad (\tau > 0), \tag{9.20a}$$

$$-(N-1)/2 \leqslant L < -1. \tag{9.20b}$$

Now, if

$$L = -1,0,...,r-2, \tag{9.21}$$

then the norms of the tensors $\hat{h}^{(rLml\sigma)}_{\mu_1\cdots\mu_r}$ with $m \leqslant L+1$ are zero. Since zero-norm tensors do not transform into positive-norm tensors, they can be identified with zero. Then the positive-norm tensors form a unitary representation if $N \geqslant 4$. For $N=3$, m is either 0 or 1. Thus only $L = -1$ is allowed in (9.21) regardless of the rank of the tensors.

Finally we note that the unitary representations we have obtained here constitute a part of those given by Ottoson,[17] as they should. [The representations realized by scalars with $-(N-1)/2 \leqslant L < 0$ are not listed in the final result of Ref. 15 for odd N. But the condition for unitarity which the author writes down is satisfied by them.]

X. ZERO-NORM TENSORS

We have seen in the previous section that there are zero-norm solutions to Eq. (9.1) if $r=0$ and $L=0,1,...$, or if $r>0$ and $L = -1,0,...,r-2$. In this section we will show that those with $r>0$ are given by analytic continuation of symmetric traceless tensors induced by STSH's studied in Sec. V.

Let us consider the symmetric traceless tensor $T^{(n)}_{s+n}$ on S^N induced by an STSH h_s satisfying

$$\Box h_s = [-L(L+N-1)+s]h_s. \tag{10.1}$$

From Eq. (5.10b) we find

$$\nabla \cdot T^{(n)}_{s+n} = 0 \tag{10.2}$$

if $L = s+n-1$. [Note that one has to replace N by $N+1$ in (5.11b) because the analysis was on S^{N-1} instead of S^N in Sec. V.] This led us to conclude that $T^{(n)}_{s+n} = 0$ for $n > L-s$.

Now we put $\chi = \pi/2 - it$. Then Eqs. (10.1) and (10.2) become equations in N-dimensional de Sitter space-time. (We denote the tensor obtained by the analytic continuation of $T^{(n)}_{s+n}$ also by $T^{(n)}_{s+n}$.) The angular momentum on S^{N-1} can take any integer value larger than s because $T^{(n)}_{s+n}$ is allowed to have a singularity at $\chi = \pi$ [$t = (\pi/2)i$]. If $L-l$ is not a non-negative integer, then $T^{(n)}_{s+n}$ is nonzero even for $s+n>L$. This can be shown by studying the singularity at $\chi = \pi$ as follows. Since the component $h_{\chi\cdots\chi i_1\cdots i_r}$ of h_s with the largest number of χ behaves like $(\pi-\chi)^{2r-s-l}$ near $\chi = \pi$, we have

$$\underbrace{\nabla_\chi \cdots \nabla_\chi}_{n} h_{\chi\cdots\chi i_1\cdots i_{r'}}$$
$$= \left[(\sin\chi)^{r'} \frac{\partial}{\partial\chi} (\sin\chi)^{-r'} \right]^n h_{\chi\cdots\chi i_1\cdots i_{r'}}$$
$$\sim (\pi-\chi)^{2r'-s-l-n}. \tag{10.3}$$

This gives the leading singularity of $T^{(n)}_{\chi\cdots\chi i_1\cdots i_{r'}}$ because other terms have at least one factor of $\sin^2\chi$ which comes from the metric tensor and, therefore, are less singular. Hence $T^{(n)}_{s+n} \neq 0$. Thus $T^{(n)}_{s+n}$ is a nonzero symmetric traceless tensor if $L-l$ is not a non-negative integer. From (5.10a) with $N \to N+1$ we have

$$\Box T^{(n)}_{s+n} = [n(2s+n+N-3) - L(L+N-1)+s+n] T^{(n)}_{s+n}. \tag{10.4}$$

Let $L = s+n-1$ so that $T^{(n)}_{s+n}$ is divergence-free as well. Then

$$\Box T^{(n)}_{s+n} = [-(s-1)(s+N-2)+s+n] T^{(n)}_{s+n}. \tag{10.5}$$

Thus the symmetric traceless tensor $T^{(n)}_{s+n}$ induced by $\hat{h}^{(s,s+n-1,ml\sigma)}_{s}$ satisfies the same equations as the tensor $\hat{h}^{(s+n,s-1,ml\sigma)}_{s}$. Hence the zero-norm tensor $\hat{h}^{(rLml\sigma)}_{r}$ with $-1 \leqslant L \leqslant r-2$ and $0 \leqslant m \leqslant L+1$ is nothing but the $(r-L-1)$th symmetric traceless tensor induced by $\hat{h}^{(L+1,r-1,ml\sigma)}_{L+1}$ $(l \geqslant r)$ up to a numerical factor. Especially, for $L=r-2$ a zero-norm solution of Eq. (9.1) $h_{\mu_1\cdots\mu_r}$ can be written as

$$h_{\mu_1\cdots\mu_r} = \nabla_{(\mu_1}\Lambda_{\mu_2\cdots\mu_r)}, \tag{10.6}$$

where $\Lambda_{\mu_1\cdots\mu_{r-1}}$ is both traceless and divergence-free and satisfies

$$\Box\Lambda_{\mu_1\cdots\mu_{r-1}} = [-(r-1)(N+r-2)+r-1]\Lambda_{\mu_1\cdots\mu_{r-1}}. \tag{10.7}$$

Thus the symmetric, traceless, and divergence-free tensors with $L=r-2$ are analogous to the corresponding tensors with $(\text{mass})^2 = 0$ in flat space-time.

XI. SUMMARY AND DISCUSSIONS

In this paper we showed how to construct the symmetric tensor spherical harmonics (STSH's) on S^N. The symmetric traceless tensors induced by STSH's were defined. It was found that the STSH's on S^N are most simply expressed in terms of symmetric traceless tensors induced by the STSH's on S^{N-1}.

We found the symmetric, traceless, and divergence-free tensor eigenfunctions of the Laplace–Beltrami operator in N-dimensional de Sitter space-time by the analytic continuation of the STSH's on S^N. We determined the allowed eigenvalues of the Laplace–Beltrami operator under the restriction of unitarity. Those values for the STSH's of rank r (for $N \geqslant 4$) are given, by defining $M^2 = -L(L+N-1)+(r-2)(r+N-3)$ so that $M^2=0$ for $L=r-2$, as

$$M^2 > (r-1)(r+N-4) \quad (m=0,1,\ldots,r), \tag{11.1}$$

$$M^2 = (r-n)(r+n+N-5) \quad (1 \leqslant n \leqslant r) \quad (m=n,\ldots,r), \tag{11.2}$$

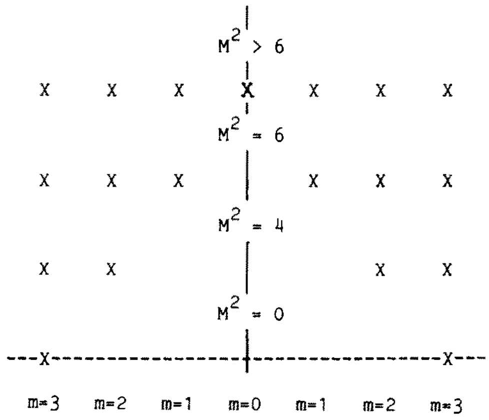

FIG. 1. The states in the unitary representation of the de Sitter group for $N=4$ and $r=3$ with fixed l.

where

$$\{\Box - [M^2-(r-2)(r+N-3)+r]\} h_{\mu_1\cdots\mu_r} = 0. \tag{11.3}$$

Here m is the rank of the SO(N) tensors in the unitary representation. For $N=3$ only $M^2=0$ is allowed in (11.2). The cosmological constant Λ is given by (1.3) in our unit.

As an example, the states with fixed angular momentum on the spatial section S^3 in the unitary representation for $N=4$ and $r=3$ are given in Fig. 1, whereas the corresponding states in flat space-time are given in Fig. 2.

ACKNOWLEDGMENTS

The author is grateful to Feza Gürsey for encouragement and useful discussions. He also thanks Yoram Alhassid, Fabio Dellagiacoma, and Soonkeon Nam for valuable discussions.

This work was supported in part by the Department of Energy under Contract No. DE-AC-02-76ER03075 and by the National Science Foundation under Grant No. PHY-8603173.

APPENDIX A: PROOF OF EQ. (2.11)

Notice first that

$$\left\{ \frac{d}{dx}\left[(1-x^2)\frac{d}{dx}\right] - \frac{\mu^2}{1-x^2} \right\} P_\nu^{-\mu}(x)$$
$$= -\nu(\nu+1) P_\nu^{-\mu}(x), \tag{A1}$$

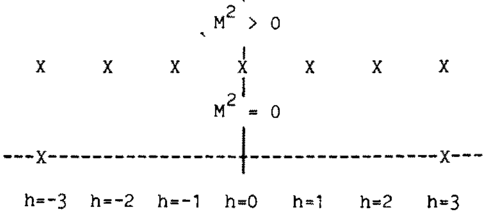

FIG. 2. The states in the spin-3 representation of the Poincaré group with fixed momentum, where h is the helicity.

and that $P_\nu^{-\mu}(x)$ behaves like $(1-x)^{\mu/2}$ for $x\sim 1$. (Here we are assuming that μ is positive.) Then we have

$$[\nu'(\nu'+1)-\nu(\nu+1)]\int_{-1+\epsilon}^{1} P_\nu^{-\mu}(x)P_{\nu'}^{-\mu}(x)dx$$
$$= -\left[(1-x^2)\frac{d}{dx}P_\nu^{-\mu}(x)\cdot P_{\nu'}^{-\mu}(x) - (1-x^2)P_\nu^{-\mu}(x)\frac{d}{dx}P_{\nu'}^{-\mu}(x)\right]_{x=-1+\epsilon}, \quad \text{(A2)}$$

where ϵ is a small positive number. Now let $\nu-\mu$ be a nonnegative integer. Then we find

$$P_\nu^{-\mu}(x)\sim\frac{(-1)^{\nu-\mu}}{\Gamma(1+\mu)}\left(\frac{1+x}{2}\right)^{\mu/2} \quad (x\to -1) \quad \text{(A3)}$$

by using (2.8) and the following formula[15]:

$$F(\alpha,\beta;\gamma;z)=\frac{\Gamma(\gamma)\Gamma(\gamma-\alpha-\beta)}{\Gamma(\gamma-\alpha)\Gamma(\gamma-\beta)}\times F(\alpha,\beta;\alpha+\beta-\gamma+1;1-z)$$
$$+(1-z)^{\gamma-\alpha-\beta}\frac{\Gamma(\gamma)\Gamma(\alpha+\beta-\gamma)}{\Gamma(\alpha)\Gamma(\beta)}\times F(\gamma-\alpha,\gamma-\beta;\gamma-\alpha-\beta+1;1-z). \quad \text{(A4)}$$

By using the same formulas we find

$$P_{\nu'}^{-\mu}(x)\sim\frac{\Gamma(\mu)}{\Gamma(1+\mu+\nu')\Gamma(\mu-\nu')}\left(\frac{1+x}{2}\right)^{-\mu/2}. \quad \text{(A5)}$$

By substituting (A4) and (A5) in (A2) and taking the limit $\epsilon\to 0$ we find Eq. (2.11).

APPENDIX B: PROOF THAT THE rhs OF EQ. (6.27) IS NONZERO

It is sufficient to show that

$$\left[\frac{\partial}{\partial\chi}+(N+s+n-2)\cot\chi\right]c_0^{(n)} - \frac{s+n}{N+2s+2n-3}\frac{1}{\sin^2\chi}c_0^{(n-1)}\neq 0, \quad \text{(B1)}$$

for $n\leqslant l-s$. We will prove this by showing that the leading term of the lhs in the limit $\chi\to 0$ is nonzero. First we find from (6.18),

$$c_0^{(0)}\sim\alpha_0\chi^l \quad (\chi\sim 0), \quad \text{(B2)}$$

where α_0 is a positive number. Let us define α_n by

$$c_0^{(n)}=\alpha_n\chi^{l-n}+O(\chi^{l-n+1}), \quad \text{(B3)}$$

where α_n is a constant.

By substituting this in (6.19) we have

$$\frac{(n+1)(N+2s+n-3)}{(s+n+1)(N+2s+2n-3)}\times[l(l+N-2)-(s+n)(s+n+N-2)]\alpha_{n+1}$$
$$=(N+s+l-2)\alpha_n-\frac{s+n}{N+2s+2n-3}\alpha_{n-1}. \quad \text{(B4)}$$

Assume that

$$(N+s+l-2)\alpha_n-\frac{s+n}{N+2s+2n-3}\alpha_{n-1} > \frac{1}{2}(N+s+l-2)\alpha_n, \quad \text{(B5)}$$

which is obviously the case for $n=0$ because $\alpha_{-1}=0$. Then one can show by using (B4) that, unless $N=4$ and $s=0$, (B5) with $n\to n+1$ is also satisfied as long as $n<l-s$. Hence we conclude by induction that Eq. (B5) holds for $n\leqslant l-s$. Then from (B5) we find $\alpha_n>0$ and

$$(N+s+l-2)\alpha_n-\frac{s+n}{N+2s+2n-3}\alpha_{n-1}>0 \quad \text{(B6)}$$

for $n\leqslant l-s$ unless $N=4$ and $s=0$. But this is nothing but the coefficient of the leading term on the lhs of (B1). For $N=4$ and $s=0$, one can show that

$$(l+2)\alpha_n-\frac{n}{2n+1}\alpha_{n-1}>\frac{n+1}{2n+1}(l+2)\alpha_n \quad \text{(B7)}$$

for $n\leqslant l-s$ in a similar manner. The rest of the proof is exactly the same as the other cases.

APPENDIX C: RAISING AND LOWERING OPERATORS FOR LEGENDRE FUNCTIONS

The function $P_\nu^{-\mu}(x)$ satisfies[15]

$$\left[(1-x^2)\frac{\partial}{\partial x}+\nu x\right]P_\nu^{-\mu}(x)=(\nu-\mu)P_{\nu-1}^{-\mu}(x), \quad \text{(C1)}$$

$$\left[(1-x^2)\frac{\partial}{\partial x}-(\nu+1)x\right]P_\nu^{-\mu}(x) = -(\nu+\mu+1)P_{\nu+1}^{-\mu}(x), \quad \text{(C2)}$$

$$\left[(1-x^2)^{1/2}\frac{\partial}{\partial x}-\frac{\mu x}{(1-x^2)^{1/2}}\right]P_\nu^{-\mu}(x) = -P_\nu^{-\mu+1}(x), \quad \text{(C3)}$$

$$\left[(1-x^2)^{1/2}\frac{\partial}{\partial x}+\frac{\mu x}{(1-x^2)^{1/2}}\right]P_\nu^{-\mu}(x) = (\nu-\mu)(\nu+\mu+1)P_\nu^{-\mu-1}(x). \quad \text{(C4)}$$

APPENDIX D: CALCULATION OF THE CONSTANT c_d

Recall that

$$f_{(m+1)}^{(m+1,lq\sigma)}=(\sin\theta)^{-m-1}Y_{Llq\sigma}(\chi,\boldsymbol{\theta}). \quad \text{(D1)}$$

Thus we have to retain only the terms proportional to $(\sin\theta)^{-m-1}Y_{Llq\sigma}(\chi,\boldsymbol{\theta})$ in calculating the rhs of (7.26). Now each term in (7.26) is of the form (differential operator) $\times Y_{Llq\sigma}(\chi,\boldsymbol{\theta})$. We first eliminate second derivatives with respect to χ or θ by using the differential equations satisfied by $Y_{Llq\sigma}(\chi,\boldsymbol{\theta})$. Then, since the terms which have $\partial/\partial\chi$ or $\partial/\partial\theta$ directly applied to $Y_{Llq\sigma}(\chi,\boldsymbol{\theta})$ must add up to zero, we can make the following substitution:

$$\frac{\partial}{\partial\chi}\to -(N+m-2)\cot\chi, \quad \text{(D2)}$$

$$\frac{\partial}{\partial\theta}\to m\cot\theta. \quad \text{(D3)}$$

We can also let

$$[1/(\sin^2\chi)]Y_{Llq\sigma}(\chi,\boldsymbol{\theta})\rightarrow 0, \tag{D4}$$

$$(\sin\theta)^{-m+1}Y_{Llq\sigma}(\chi,\boldsymbol{\theta})\rightarrow 0. \tag{D5}$$

Then we find that we can let

$$\left(\cos\theta\frac{\partial}{\partial\chi}-\cot\chi\sin\theta\frac{\partial}{\partial\theta}\right)f^{(mlq\sigma)}_{(m+1)}\rightarrow 0, \tag{D6}$$

$$\cot\chi\cos\theta f^{(mlq\sigma)}_{(m+1)}\rightarrow 0, \tag{D7}$$

$$[(\sin\theta)/(\sin^2\chi)]f^{(mlq\sigma)}_{(m)}\rightarrow 0. \tag{D8}$$

Then $f^{(mlq\sigma)}_{(m+2)}$ can be found by using Eq. (6.6) with $s=m$ and $n=2$ and Eqs. (6.18) and (6.10) as follows:

$$f^{(mlq\sigma)}_{(m+2)}=D_1T_{(m+2)}-D_2f^{(mlq\sigma)}_{(m)}, \tag{D9}$$

where

$$T_{(m+2)}=\left[\frac{\partial^2}{\partial\theta^2}-\frac{m(m+N-2)-l(l+N-2)}{N+2m-1}\right]f^{(mlq\sigma)}_{(m)}, \tag{D10}$$

$$D_1=\frac{(m+1)(m+2)}{2}\frac{N+2m-1}{N+2m-2}\frac{1}{(m-l)(m+l+N-2)(m-l+1)(m+l+N-1)}$$
$$\times\left[\frac{\partial}{\partial\chi}+(N+m-1)\cot\chi\right]\left[\frac{\partial}{\partial\chi}+(N+m-2)\cot\chi\right], \tag{D11}$$

and

$$D_2\propto 1/\sin^2\chi. \tag{D12}$$

By following the procedure explained above we obtain

$$\sin\theta f^{(mlq\sigma)}_{(m+2)}\rightarrow\frac{(m+1)(m+2)}{2}\frac{N+2m-1}{N+2m-2}$$
$$\times\frac{C_{NLmq}}{(m-l)(m+l+N-2)(m-l+1)(m+l+N-1)}(\sin\theta)^{-m-1}Y_{Llq\sigma}(\chi,\boldsymbol{\theta}), \tag{D13}$$

where

$$C_{NLmq}=(q-m)(m+q+N-3)[(m-1)(N+m-2)-L(L+N-1)]. \tag{D14}$$

Similarly, we have

$$f^{(m-1,lq\sigma)}_{(m+1)}\rightarrow\frac{m(m+1)}{2}\frac{N+2m-3}{N+2m-4}$$
$$\times\frac{C_{NLmq}}{(m-l-1)(m+l+N-3)(m-l)(m+l+N-2)}(\sin\theta)^{-m-1}Y_{Llq\sigma}(\chi,\boldsymbol{\theta}). \tag{D15}$$

By using (7.17) and (7.21a) we find

$$-c_-f^{(m,l-1,q\sigma)}_{(m+1)}=\frac{(\sin\theta)^{-m}}{2l+N-2}\frac{(m+1)(l-r)(l+1)}{(l-m-1)(N+m+l-3)(l-m)(l-m+1)}$$
$$\times\left[\frac{\partial}{\partial\chi}+(N+m-2)\cot\chi\right]\left[\frac{\partial}{\partial\chi}+(N+l-2)\cot\chi\right]$$
$$\times\left(\frac{\partial}{\partial\theta}-m\cot\theta\right)\left(\sin\theta\frac{\partial}{\partial\theta}-l\cos\theta\right)Y_{Llq\sigma}(\chi,\boldsymbol{\theta}). \tag{D16}$$

Then

$$-c_-f^{(m,l-1,q\sigma)}_{(m+1)}\rightarrow-\frac{(m+1)(l-r)(l+1)}{(2l+N-2)(l-m-1)(N+m+l-3)(l-m)(l-m+1)}$$
$$\times C_{NLmq}(\sin\theta)^{-m-1}Y_{Llq\sigma}(\chi,\boldsymbol{\theta}). \tag{D17}$$

Similarly, we find

$$-c_+f^{(m,l+1,q\sigma)}_{(m+1)}\rightarrow-\frac{(m+1)(l+N+r-2)(l+N-3)}{(2l+N-2)(l-m+1)(N+m+l-1)(N+m+l-2)(N+m+l-3)}$$
$$\times C_{NLmq}(\sin\theta)^{-m-1}Y_{Llq\sigma}(\chi,\boldsymbol{\theta}). \tag{D18}$$

By substituting (D13), (D15), (D17), and (D18) in Eqs. (7.25) and (7.26) we obtain c_d.

[1]See, e.g., S. W. Hawking and G. F. R. Ellis, *The Large Scale Structure of Space-Time* (Cambridge U.P., Cambridge, 1973).
[2]A. H. Guth, Phys. Rev. D **23**, 347 (1981).
[3]A. D. Linde, Phys. Lett. B **108**, 389 (1982).
[4]A. Albrecht and P. J. Steinhardt, Phys. Rev. Lett. **48**, 1220 (1982).
[5]S. Weinberg, Phys. Rev. **133**, B1318 (1964); **134**, B882 (1964).

[6]G. Grensing, J. Phys. A: Gen. Phys. **10**, 1687 (1977).
[7]N. D. Birrell, J. Phys. A: Gen. Phys. **12**, 337 (1979).
[8]See, e.g., F. Gürsey, in *Relativity, Groups and Topology*, edited by B. S. DeWitt and C. DeWitt (Gordon and Breach, New York, 1964).
[9]A. Higuchi, Nucl. Phys. B **282**, 397 (1987).
[10]H. Van Dam and M. Veltman, Nucl. Phys. B **22**, 397 (1970); Gen.Relativ. Gravit. **3**, 215 (1972); Y. Iwasaki, Phys. Rev. D **2**, 2255 (1970); V. I. Zakharov, JETP Lett. **12**, 312 (1970).
[11]M. Veltman, in *Methods in Field Theory, Les Houches 1975*, edited by R. Balian and J. Zinn-Justin (North-Holland, Amsterdam, 1976).
[12]A. Chodos and E. Myers, Ann. Phys. (NY) **156**, 412 (1984).
[13]M. A. Rubin and C. R. Ordóñez, J. Math. Phys. **25**, 2888 (1984).
[14]B. Ratra, Phys. Rev. D **31**, 1931 (1985).
[15]I. S. Gradshteyn and I. M. Ryzhik, *Tables of Integrals, Series, and Products*, corrected and enlarged edition, edited by A. Jeffrey (Academic, New York, 1980).
[16]T. Regge and J. Wheeler, Phys. Rev. **108**, 1063 (1957).
[17]U. Ottoson, Commun. Math. Phys. **8**, 228 (1968).
[18]M. A. Neumark, Am. Math. Soc. Transl. (2) **6**, 337 (1957) ($N=3$); L. H. Thomas, Am. Math. **42**, 113 (1941); J. Dixmier, Bull. Soc. Math. France **89**, 9 (1961) ($N=4$); J. G. Kuriyan, N. Mukunda, and E. C. G. Sudarshan, Commun. Math. Phys. **8**, 204 (1968) ($N=5$).

Exact solutions in 1+1 dimensions of the general two-velocity discrete Illner model

Henri Cornille
Service de Physique Théorique, CEN–Saclay, 91191 Gif-sur-Yvette Cedex, France

(Received 29 August 1986; accepted for publication 4 February 1987)

The Illner model is the most general two-velocity discrete model of a Boltzmann equation in one spatial dimension which satisfies an H-theorem. It includes, as particular cases, both the Carleman and the McKean models. "Solitons" (one-dimensional solutions) and "bisolitons" (two-dimensional, space-plus-time, solutions), which are defined as rational fractions, and solutions with one or two exponential variables are determined. The model is treated as a nonintegrable nonlinear one, and from the solitons the possible class of bisolitons is guessed. Two classes of physically acceptable bisolitons are found. The first class is distributions positive only along one semiaxis and identically zero outside. These are interpreted physically by introducing elastic walls plus source or sink terms which become negligible at infinite time. The second class is periodic solutions which can be seen as damped sound waves. Essentially the same tools are used as in a companion paper for the six-velocity Broadwell model, where the two bisoliton classes mentioned above also exist. This suggests that general methods for obtaining nontrivial exact solutions could exist for the hyperbolic semilinear discrete Boltzmann models.

I. INTRODUCTION

At the present time, many people think that discrete models can provide useful examples for the existing problems in kinetic theory. Due to the numerous difficulties appearing in the study of the continuous Boltzmann equation (BE), they hope that discrete velocity models may shed some light on the solutions of these problems. In this framework, it seems to me a reasonable preliminary condition that these discrete models could be studied with similar methods providing fairly general results.

In a companion paper[1] we determine a class of exact $1+1$-dimensional (space-x-plus-time-t) solutions of the six-velocity discrete Broadwell[2] model with three independent distribution functions. It is the aim of the present paper, for a two-velocity discrete model, to see whether methods and tools similar to those used for the Broadwell model lead to comparable results.

The most popular two-velocity discrete models are the Carleman and McKean[3] models. The Carleman model has been extensively studied: global existence, uniqueness, boundedness, oscillations, and asymptotic behaviors of the solutions.[4] Some of these results have been extended to more complex discrete models, for instance a 14-velocity model.[5] For obvious reasons, the mathematical results were obtained for smooth and relatively small initial data, for instance the distributions are integrable when $|x|\to\infty$.

For two-velocity discrete models let us call f_η, $\eta=\pm$, the distributions of particles moving in opposite directions of the x axis, with velocities ± 1. The most general binary collision term $\mathrm{Col}(f_+,f_-)$ is a quadratic form of the products $f_\eta f_{\eta'}$: $f_\pm^2, f_+ f_-$ with three arbitrary parameters. Illner[6] clarified these mathematical models by supplementing them with the physical restriction that they should be compatible with the H-theorem. He obtained a two-parameter family of models

$$(\partial_t+\eta\,\partial_x)f_\eta=\eta\,\mathrm{Col}(f_+,f_-)=\eta(af_+^2-(a+c)f_+f_-+cf_-^2),\tag{1.1}$$

where the restrictions $a\leqslant 0$, $c\geqslant 0$ ensure that

$$\partial_t\sum f_\eta\log f_\eta+\partial_x(f_+\log f_+-f_-\log f_-)\leqslant 0,$$
$$\forall f_+>0,\quad \forall f_->0$$

(which is a stronger condition than the usual $\forall t\geqslant 0$).

Throughout the paper, we restrict our study to $a\leqslant 0$, $c\geqslant 0$. The values $-a=c=1$ lead to the Carleman model and $a=0$, $c=1$ to the McKean one. Let us introduce $N=f_++f_-$ the density of particles, and $J=f_+-f_-$ the stream velocity (1.1) can be rewritten as

$$\begin{aligned}&\partial_tN+\partial_xJ=0,\\&\partial_xN+\partial_tJ=J\,[(a+c)J+N(a-c)]\,,\end{aligned}\tag{1.2}$$

the first linear equation expressing the mass conservation. We notice that these models fail to conserve momentum.

We first want to obtain algebraic exact solutions N, J of (1.2) and second, building $N\pm J$, *verify the positivity of the distributions f_η*. Note that the conditions[7] $N\geqslant 0$, $N-J\geqslant 0$ are not sufficient to preserve positivity (as we shall see explicitly later). We will not consider the homogeneous formalism because, for this simple model, we think that at least the spatial dependence must be present. The known exact solutions are mainly those of the Carleman model[7–10] if we accept the Platkowski results.[10] All these solutions satisfy the ordinary differential equations (ODE's) which can be transformed into integrable linear differential equations. Self-similar solutions, with only one variable, are of this type, and further they are in one dimension. For the Carleman model a more subtle periodic solution[9] has been obtained by Bobylev, and clarified and extended by Wick. Assuming that N and J

 0022-2488/87/071567-13$02.50

are conjugate harmonic functions, then $f(z=x+iy) = N+iJ$ is an analytic function from which the reduction to integrable differential equations is possible. In this way, a periodic two-dimensional solution was obtained. Summarizing the results obtained so far, the goal has been to extract from the nonintegrable hyperbolic system (1.1) [or (1.2)] particular cases which could be reduced to integrable linear differential equations.

Here we want to face (1.2) as a genuine nonintegrable system and try methods which have been successful for other nonintegrable nonlinear partial differential equations (NLPDE's). The continuous homogeneous BE becomes through the generalized Laplace transform[11] a NLPDE of this type. Exact solutions were obtained, called "solitons" and "bisolitons" because they were defined as rational fractions with one exponential variable $w = d\exp(\gamma x + \rho t)$ or two $w_i = d_i \exp(\gamma_i x + \rho_i t)$. (With this definition we can recover the soliton and bisolitons of the two-dimensional completely integrable NLPDE.) Further it turned out[12] that the homogeneous BE was generic of a whole class (not including the semilinear hyperbolic systems) of nonintegrable NLPDE's sharing common properties both for their linear operators and their bisolitons. Similarly here, for the semilinear hyperbolic systems, by investigating solitons and bisolitons for different discrete models, we hope that common properties will emerge for another class of nonintegrable equations.

For (1.2), the one-dimensional soliton solutions, being self-similar solutions, can be obtained directly from integrable differential equations. However, they are useful because the bisolitons when $d_j \to 0$ must reduce to the solitons $w = w_i$. Consequently they provide some guesses for the search of bisolitons. They are obtained as rational solutions with denominators $\Delta = 1 + w$ (Sec. II A). Then asymptotic $|x| \to \infty$ positivity properties are interesting, although different from the Broadwell ones. There always exists one of the limiting x axis where $\lim f_-$, $\lim f_+$ have the two opposite signs of a and c. Consequently in general ($ac<0$) they cannot be positive solutions on the full x axis. (This holds for the Carleman model $a+c=0$, contrary to the Dukek and Nonnenmacher[7] analysis.) However, the particular $a=0$ or $c=0$ models can avoid this difficulty and sometimes these soliton solutions can be physically interpreted as shock waves.[10] Finally we notice that in all $a \leqslant 0$, $c \geqslant 0$ cases, a lot of positive soliton solutions exist for x on a semiaxis.

The simplest possible ansatz bisolitons [see Eq. (2.5)] have denominators of the type $\Delta = 1 + \Sigma w_i + \mu w_1 w_2$. The constant μ in Δ represents the coupling between the solitons w_1 and w_2. We check the consistency of such solutions with the constraints coming from the linear mass conservation in (1.2) and the most singular term (proportional to Δ^{-2}) of the nonlinear part of (1.2) (see the Appendix). We find that the bisolitons without soliton coupling in Δ or $\mu = 0$ are not possible. This result is not surprising as it is a consequence of the previous nonintegrable NLPDE study[12] of the $\mu = 0$ bisolitons. It was shown that the operator of the linear part must be factorized, a property not existing for the hyperbolic semilinear equations. For the $\mu \neq 0$ case we find that only $\mu = 1$ is allowed which means $\Delta = (1+w_1)(1+w_2)$ or equivalently the bisolitons can be written as a sum of two solitons. Starting with

$$N = n_0 + \sum \frac{n_i}{\Delta_i}, \quad J = j_0 + \sum \frac{j_i}{\Delta_i}, \quad \Delta_i = 1 + w_i, \tag{1.3}$$

all parameters can be algebraically determined. (Notice that in general $N \pm J \to \text{const}$ when $|x| \to \infty$ and so are not integrable.)

The linear mass conservation law in (1.2) can always support a superposition of soliton solutions. On the contrary, the nonlinear part gives the conditions for the coupling of solitons

$$(a+c) + (c-a)(\gamma_i/\rho_i + \gamma_j/\rho_j)/2 = 0, \quad i \neq j, \tag{1.4}$$

and only two different γ_i/ρ_i values are compatible. "Multisolitons" with more than two solitons are not possible; this is the main difference between the present class of nonintegrable equations and completely integrable ones (the same distinction occurs for the Broadwell model). In conclusion the bisoliton parameters satisfy the relations of each soliton component plus the coupling relation (1.4). It appears useful to define two new parameters ν_1 which reduce to γ_i/ρ_i for the Carleman model and such that the coupling relation (1.4) becomes $\nu_1 + \nu_2 = 0$. We arbitrarily choose $\nu = \nu_1$ as the remaining parameter and the algebraic bisolitons become rational fractions of the w_i with two parameters n_0 (or j_0) and ν. In the remaining sections of the paper we study both γ_i, ρ_i real or complex (ν real or purely imaginary) and look at the positivity properties.

In Sec. III we show that for γ_i, ρ_i real no positive bisolitons exist along the full x axis (this is the main difference with the Broadwell model[1]). However, we can find bisolitons compatible with positivity if we restrict to the semi-($x \geqslant 0$)-axis. We establish the class of possible positive solutions following the $a \leqslant 0$, $c \geqslant 0$ values of the parameters of the Illner model. Two classes must be distinguished depending whether in (1.3) $j_0 = 0$ or $j_0 \neq 0$. Invariance properties allow a reduction of the studied intervals for the parameters.

We distinguish between the asymptotic positivity requirements $x \to \infty$, $t \to \infty$ and the $t \geqslant 0$, $x \geqslant 0$ positivity for which we give sufficient conditions on the constants d_i of the $w_i = d_i \exp(\gamma_i x + \rho_i t)$ soliton components. Specular boundary at the $x=0$ condition is possible only for the Carleman $a+c=0$ model.

In Sec. IV the soliton components w_i are complex conjugate $w_2 = \bar{w}_1$ and we define $w = w_1 = d\exp(\gamma x + \rho t)$ with d, γ, ρ complex. The above ν parameters becomes $i\nu_{\mathrm{I}}$ and the discussion occurs with the two parameters n_0 (or j_0) and ν_{I}. Invariance properties still allow a simplification of the parameter study. We must distinguish between the two cases where the spatial part of w has γ complex or purely imaginary.

(i) If γ is complex, then no positive solution exists on the full x axis. However, positive solutions on a semiline, say $x \geqslant 0$ exist.

(ii) If γ is purely imaginary (see Sec. IV A), then the solutions are periodic. For all parameter values a,c ($ac \neq 0$) of the Illner model there exist positive periodic solutions (exact periodic solutions exist also for the Broadwell model). For the positive (i) and (ii) solutions, sufficient conditions on the d parameter of w ensure the positivity for $t \geqslant 0$.

Let us compare the periodic solutions obtained here for the Illner model from the bisoliton method, with the Bobylev–Wick periodic solution of the Carleman model. For the parameter values $a + c \neq 0$ of the Illner model it is shown that the N, J functions associated with the bisolitons are not conjugate harmonic functions. The same result is true in general for the bisolitons of the Carleman model $a + c = 0$. However, there exists a particular restriction on the parameters $|\nu_I| = c$ for which this property for N and J holds and then the solution coincides with the Bobylev–Wick one.

For the physical interpretation of the periodic solutions we distinguish between two cases. For $a + c \neq 0$ they are propagating waves with an absorption factor. They can be seen as damped sound waves. However, the absorption can be so strong that in fact only one or two oscillations in time can be seen. We define as a criterion the ratio of the real to imaginary part of ρ. We show analytically and observe numerically that when the modulus of this ratio decreases, then the number of effective oscillations in time increases. On the contrary for the Carleman $a + c = 0$ model, the waves are nonpropagating with time and the solutions can only describe damped oscillations in the space variable.

In Sec. V we try to give a physical meaning for the solutions positive only along a semiline ($x \geqslant 0$ for instance). Similarly as what was done[13] for the inhomogeneous Kac model, we define new distributions $\tilde{f}_\eta$, identical to f_η for $x \geqslant 0$ and identically zero for $x < 0$. The new distributions are a solution of kinetic equations of the Illner type, with two additional terms at $x = 0$, that we must interpret physically. For the Carleman model, only one supplementary term exists which can be interpreted as an elastic wall at $x = 0$. For other Illner models, $a + c \neq 0$, besides this elastic wall, another term is present which, for the w_i real solutions, can be interpreted either as a source or as a sink, decreasing exponentially in time and becoming negligible compared to the elastic wall at infinite time. For the w_i complex conjugate solutions, the second term, while decreasing exponentially and becoming negligible when $t \to \infty$, changes sign with t. It acts like an oscillating source, a sink term with an interpretation not as simple as above.

In Sec. VI we illustrate the results of the paper with numerical calculations. For some examples, we plot the $N + \eta J = 2 f_\eta$ relaxation curves for the class of solutions positive along $x \geqslant 0$ (Fig. 1) and for periodic solutions (Figs. 2–4) the $N \pm J, N$ relaxation curves and different time oscillations when $\rho_R \rho_I^{-1}$ decreases.

II. RATIONAL SOLUTIONS WITH EXPONENTIAL VARIABLES

We seek solutions of the Illner model ($a \leqslant 0, c \geqslant 0$) which are rational fractions of either one exponential variable $w = d \exp(\gamma x + \rho t)$ (solitons) or two variables $w_i = d_i \exp(\gamma_i x + \rho_i t), i = 1,2$ (bisolitons). The study of the bisolitons will be performed in two successive stages. First, in this section we determine the algebraic structure of these solutions, and second, in the following sections, we look both at the asymptotic positivity constraints $|x| \to \infty, t \to \infty$ and at the positivity $t \geqslant 0$. *A priori* the denominators of the rational fractions solutions of (1.2) are of the type Δ^q, Δ being a polynomial in w (or w_i) and q unknown. However, we remark that the quadratic nonlinearity is associated with linear first-order differential operators. If the most singular part of the solution, which comes from the Δ^{-q} factor, is determined by a balance between linear and nonlinear parts, then necessarily $q = 1$.

A. Solitons

The solitons, self-similar one-dimensional solutions, with the variable $\gamma x + \rho t$, are completely integrable solutions of ODE's. Despite their simplicity, for pedagogical reasons, we briefly present the results (Table I). They will provide some hints in the search of the possible bisolitons. Further the study of their properties, particularly the positivity, will be a guide for the corresponding bisoliton properties. We start with the ansatz

$$N = n_0 + n/\Delta, \quad J = j_0 + j/\Delta, \quad \Delta = 1 + w, \qquad w = d \exp(\gamma x + \rho t), \quad d > 0, \tag{2.1}$$

$n_0, n, j_0, j, \gamma, \rho, d$ being real constants, that we substitute into the Illner system (1.2). We obtain four relations (Table I,

TABLE I. Solitons.

A:	Ansatz $N = n_0 + n/\Delta, \quad J = j_0 + j/\Delta, \quad \Delta = 1 + w, \quad w = d \exp(\gamma x + \rho t), \quad d > 0$
B:	Relations (1) $j\gamma + n\rho = 0$, (2) $j_0((a+c)j_0 + (a-c)n_0) = 0$, (3) and (4) $(n\gamma + j\rho) = -2 j_0 j (a+c) + (a-c)(j_0 n + j n_0) = j[(a+c)j + (a-c)n], \quad a \leqslant 0, \quad c \geqslant 0$
C:	Algebraic solitons definition $\gamma/\rho = (2\nu - a - c)/(c - a)$
C1:	$j_0 = 0$: $N - J = n_0(1 + (-1 + a/\nu)/\Delta)$, $N + J = n_0(1 + (-1 + c/\nu)/\Delta), \quad \rho = n_0 (c-a)^3 / 4(a - \nu)(\nu - c)$
C2:	$j_0 \neq 0$: $N - J = 2 j_0 (a + (\nu - a)/\Delta)/(c - a)$ $N + j = 2 j_0 (c + (\nu - c)/\Delta)/(c - a), \quad \rho = j_0 \nu (c-a)^2 / 2(\nu - a)(\nu - c)$
D:	Physical positive solitons on $x \in [-\infty, +\infty], \quad t \geqslant 0, \quad$ condition $\lvert x \rvert \to \infty$
(i)	$a = 0$: (i.1) $j_0 = 0, n_0 > 0, \nu > 0$; $N \mp J \to n_0$, n_0 and 0, $n_0 c/\nu$ (i.2) $j_0 > 0, \nu > 0$; $(N \mp J)/2 j_0 \to 0, 1$ and $\nu/c, \nu/c$
(ii)	$c = 0$: (i.1) $j_0 = 0, n_0 > 0, \nu < 0$; $N \mp J \to n_0$, n_0 and $n_0 a/\nu, 0$. (i.2) $j_0 < 0, \nu < 0$; $(N \mp J)/2 j_0 \to 1, 0$ and $\nu/a, \nu/a$
E:	Physical positive soliton on $x \geqslant 0, t \geqslant 0$, with $a \neq 0, c \neq 0$:
E1:	$j = 0$: $n_0 > 0 \ (a+c)/2 < \nu < c, d > \sup(\lvert a/\nu \rvert, c/\lvert \nu \rvert), \rho > 0, \gamma > 0, j\nu = n_0 (c-a)/2\Delta > 0$
E2:	$j_0 > 0$: $\sup(0, (a+c)/2) < \nu < c, d > \lvert \nu/a \rvert, \rho < 0, \gamma < 0, J = j_0 w/\Delta > 0$.

part B) easily solved. Two classes of solutions occur depending on whether $j_0 = 0$ or $j_0 \neq 0$. It appears convenient to introduce a parameter ν linked to the ratio γ/ρ and identical to it for the Carleman model $c = -a = 1$,

$$\gamma/\rho = (a + c - 2\nu)/(a - c). \tag{2.2}$$

The solutions, written down in Table I, part C are rewritten here in a form appropriate for the positivity discussion

$$\begin{aligned}(N - J)\Delta/n_0 &= w + a/\nu,\\ (N + J)\Delta/n_0 &= w + c/\nu, \quad j_0 = 0,\end{aligned} \tag{2.3a}$$

$$\begin{aligned}(N - J)(c - a)\Delta/2j_0 &= aw + \nu,\\ (N + J)(c - a)\Delta/2j_0 &= cw + \nu, \quad j_0 \neq 0.\end{aligned} \tag{2.3b}$$

These solutions depend on two arbitrary parameters n_0, ν (or j_0, ν) which are important for the asymptotic positivity and one more integration constant $d > 0$ $(\Delta > 0)$ important for the positivity at $t = 0$. When $|x| \to \infty$ or $t \to \infty$, there exist for w only two possibilities: either $w \to \infty$ or $w \to 0$ which when substituted into (2.3a) and (2.3b) will provide asymptotic constraints.

We discuss possible solutions on the full x axis and begin with the asymptotic $|x| \to \infty$ positivity constraints. For the $j_0 = 0$ solution we note that $\lim(N - J, N + J)$ is either $n_0(1,1)$ on one side or $(n_0/\nu)(a,c)$ on the other side. Due to $ac \leqslant 0$ we see that positivity is violated if $ac \neq 0$. For $j_0 \neq 0$ we find both limits $(2j_0/(a - c))(a,c)$ and $(2j_0\nu/(c - a))(1,1)$ for $N \mp J$, leading to the positivity violation if $ac \neq 0$. In conclusion, positive soliton solutions cannot exist on the full x axis if $ac < 0$. (In particular, this result holds for the Carleman model $a + c = 0$.) There remains the possibility $ac = 0$. We begin with $a = 0$, $c > 0$: from (2.3a) and (2.3b) it follows that positivity ($x \in [-\infty, +\infty]$, $t \geqslant 0$) is satisfied for the $j_0 = 0$ solution with $n_0 > 0$, $\nu > 0$ as well as for the $j_0 > 0$ one with $\nu > 0$. Similarly, for the model $c = 0$, $a < 0$ we find a positive solution when $j_0 = 0$, $n_0 > 0$, $\nu < 0$ as well as for $j_0 < 0$, $\nu < 0$. The soliton solutions with the variable $x + t\rho\gamma^{-1}$ are candidates as shock waves if $|\rho/\gamma| \leqslant 1$ and the total mass N has a jump between the two limits $|x| \to \infty$. These two limits are $n_0(1, (a + c)/\nu)$ for the $j_0 = 0$ solution and $(j_0(c - a)^{-1})(a + c, 2\nu)$ for $j_0 \neq 0$. From (2.2), $|\rho/\gamma| \leqslant 1$ means $\nu/c < 0$ or $\nu/c > 1$ for the $a = 0$ model and $\nu/a < 0$ or $\nu/a > 1$ for the $c = 0$ one. Of course the distributions $N \pm J$ must be positive. It is why our asymptotic $|x| \to \infty$ positivity condition, leading to $ac = 0$, is a first step in the determination of shock-wave solutions. For these models $ac = 0$ and one of $N \pm J$, there exists one of the two limits $|x| \to \infty$ which is zero (see above) representing an infinite-Mach-number shock wave. (See the Platkowski paper[10] for a discussion.)

If we restrict our study to a semiline, for instance $x \geqslant 0$, then $N \pm J > 0$ $(t \geqslant 0)$ exist also for $a \neq 0$, $c \neq 0$. First for $j_0 = 0$ (2.3a), in order to satisfy positivity when x or $t \to \infty$, necessarily ρ and γ must be such that $(N \mp J) \to n_0$, $\rho > 0$, and $\gamma > 0$. Looking at γ in (2.2) and ρ written down in Table I, part C1 we find $(a + c)/2 < \nu < c$. On the other hand, for any $t \geqslant 0$ we notice that (2.3a) leads to the lower bound $(N - J)/n_0 > d - |a/\nu|$, $(N + J)/n_0 > d - c/|\nu|$, so that positivity is satisfied for $|d| > \sup(|a/\nu|, c/|\nu|)$. Second, for the $j_0 \neq 0$, $j_0 > 0$ solution, a similar analysis when $t \to \infty$, $x \to \infty$ leads to $\gamma < 0, \rho < 0$. Then (2.2) and Table I, part C2 give the constraint $c > \nu > \sup(0, (a + c)/2)$ from which we see that $N + J$ is always positive for $t \geqslant 0$. The lower bound $(N - J)(c - a)\Delta/2j_0 > \nu - |a|d$ gives the last constraint $d > |\nu/a|$. For the $a = 0$ or $c = 0$ models exist of course as solutions on the semi-x axis and we have disregarded here these particular cases. In Table I, part D E_1E_2 a summary of the positivity discussion is written down.

B. Possible bisolitons

We introduce the two-exponential variables $w_i = d_i \exp(\gamma_i x + \rho_i t)$ and require

$$\rho_1\gamma_2 - \rho_2\gamma_1 \neq 0 \tag{2.4}$$

for a true two-dimensional solution. We remark that when $d_2 = 0$ (or $d_1 = 0$), the bisolitons must reduce to the solitons studied above, with $w = w_1$ (or w_2). We notice also that if we write the solitons of the above subsection like N/D then N and D are linear in w.

Let us look at the possible denominators Δ of the bisoliton. It must be of the type $1 + \Sigma w_i$ plus terms at least quadratic in w_1, w_2. However, pure power terms w_1^p (or w_2^p), $p > 1$, alone, which do not vanish when $d_2 = 0$ (or $d_1 = 0$) are not present because they do not exist in the soliton case. In other words Δ must reduce to $1 + w_i$ when $d_j = 0$, $i \neq j$. Thus Δ must be of the type $\Delta = 1 + \Sigma w_i + w_1 w_2 P(w_1, w_i)$, P being a polynomial in w_1, w_2. For simplicity we assume the simplest choice $P \equiv \mu$, a constant.

We write down a possible class of ansatz bisolitons,

$$N = n_0 + \frac{n}{\Delta}, \quad J = j_0 + \frac{j}{\Delta}, \quad \Delta = 1 + \sum_i w_i + \mu w_1 w_2, \tag{2.5}$$

with $n = n_{00} + \Sigma n_i w_i$, $j = j_{00} + \Sigma j_i w_i$. We notice that supplementary terms proportional to $w_1 w_2$ in the numerators n and j do not enlarge the class of ansatz because the ratios with Δ still lead to (2.5). The study is done in the Appendix. Two different possibilities occur depending upon whether $\mu \neq 0$ or $\mu = 0$. We tackle the constraints coming from two relations: mass conservation and nonlinear terms proportional to Δ^{-2} which necessarily factorize Δ,

$$\begin{aligned}&N_t + J_x = 0,\\ &n\Delta_x + j\Delta_t - j[(a + c)j + n(a - c)] = 0(\Delta).\end{aligned} \tag{2.6}$$

The calculations are tedious but the results are simple.

(i) The assumption $\mu \neq 0$ leads necessarily to $\mu = 1$ (Sec. 1 in the Appendix) means that Δ is the product $(1 + w_1)(1 + w_2)$. Furthermore the numerators are of the type $n = \Sigma n_i(1 + w_i)$, $j = \Sigma j_i(1 + w_i)$. Consequently, at this stage where only the nonlinear constraint (2.6) is taken into account, the class of possible $\mu \neq 0$ ansatz (2.5) is reduced to

$$N = n_0 + \sum \frac{n_i}{\Delta_i}, \quad J = j_0 + \sum \frac{j_i}{\Delta_i}, \quad \Delta_i = 1 + w_i, \tag{2.7}$$

where the constants $n_0, n_i, j_0, j_i, \gamma_i, \rho_i$ have to be determined from the other constraints of the nonlinear equation in (1.2).

(ii) The $\mu = 0$ case corresponds in Δ to the vanishing of the coupling between the solitons. Such bisolitons appear naturally in the study of the continuous homogeneous BE. The class of NLPDE noncompletely integrable leading to such bisolitons has been investigated.[12] They correspond to factorizations of the operators associated with the linear part of the nonlinear equation. For the inhomogeneous discrete BE such factorization of the linear operator does not occur and we expect that such bisolitons do not exist here. The calculations are done in Sec. 2 of the Appendix. Taking into account the constraints (2.6) and with a lot of cumbersome calculations, we find that bisolitons with $\mu = 0$ do not exist.

In conclusion, among the class (2.5), then (2.7) is the only possible subclass of bisolitons. In the following subsections we shall first substitute the ansatz (2.7) into the nonlinear Illner system (1.2) and determine explicitly the parameters of the bisolitons. Second, with the help of invariance properties we shall show that it is sufficient to consider the parameter values in reduced intervals. Finally in the other sections we shall establish the asymptotic $|x| \to \infty$, $t \to \infty$, and $t \geqslant 0$ positivity constraints.

C. Algebraic forms of the bisolitons

The substitution of the ansatz (2.7) into the Illner system and the vanishing of the coefficients of const Δ_i^{-1}, Δ_i^{-2}, $(\Delta_1\Delta_2)^{-1}$ provide five relations (some of them are double) which are written down in Table II, part B. The four first relations are the soliton ones associated with the soliton parts w_i of the bisolitons. Notice that the bisoliton ansatz (2.7) is formally written as a linear superposition of two solitons w_i. The last relation [see also (1.4)] coming from $(\Delta_1\Delta_2)^{-1}$ represents the coupling between the two solitons such that the bisoliton exists. If, as in the soliton case (2.2), we introduce for each soliton a new parameter ν_i associated with the ratio ρ_i/γ_i (and reducing it for the Carleman model), i.e., $\gamma_i/\rho_i = (a + c - 2\nu_i)/(a - c)$, then the condition for the existence of a double soliton (called bisoliton) is simply $\nu_1 + \nu_2 = 0$. We arbitrarily choose

$$2\nu = 2\nu_1 = a + c - (a - c)\gamma_1/\rho_1 \tag{2.8}$$

as the new parameter and ν_2 becomes $-\nu$. This simple condition for the existence of bisolitons traduces the fact that discrete Boltzmann models represent weakly nonlinear models called semilinear by the mathematicians. It is much more difficult to obtain similar objects in the continuous BE.

Once the coupling between the two solitons has been found then the algebraic determination of the solution in terms of ν, j_0, n_0 is easily done. We still have two classes depending on whether $j_0 = 0$ or $j_0 \neq 0$, written down in Table II, part C1-2, and rewritten here in a slightly different way:

$$(N - J)\Delta_1\Delta_2/n_0 = -1 + w_1w_2 + a(w_2 - w_1)/\nu,$$
$$\rho_1 = n_0(c - a)^3/4(\nu - c)(a - \nu), \tag{2.9a}$$
$$(N + J)\Delta_1\Delta_2/n_0 = -1 + w_1w_2 + c(w_2 - w_1)/\nu, \quad j_0 = 0,$$

$$(N - J)\Delta_1\Delta_2(c - a)/2j_0$$
$$= (-a)(1 - w_1w_2) + \nu(w_2 - w_1),$$
$$\rho_1 = j_0\nu(a - c)^2/2(\nu - a)(\nu - c), \tag{2.9b}$$
$$(N + J)\Delta_1\Delta_2(c - a)/2j_0 = c(w_1w_2 - 1) + \nu(w_2 - w_1),$$
$$j_0 \neq 0,$$

convenient for the positivity study. In both the $j_0 = 0$ and $j_0 \neq 0$ cases we have from (2.8) $\gamma_1 = \rho_1(a + c - 2\nu)/(a - c)$ and for the second component part of the bisoliton the simple relations $\nu \to -\nu$: $\rho_2 = \rho_1(-\nu)$, $\gamma_2 = \gamma_1(-\nu)$. The fact that these solutions (2.9) sustain a two-dimensional space [or (2.4) is satisfied] is verified by the relation $\gamma_1(\nu)/\rho_1(\nu) \neq \gamma_2/\rho_2 = \gamma_1(-\nu)/\gamma_2(-\nu)$. We check now that if one soliton component vanishes then the bisoliton is reduced to the soliton of the other component. Starting from Table II, part C2 for $j_0 \neq 0$, let us define $w = w_1$, $n_0 = 2j_0\nu/(a - c)$, $\rho = \rho_1$ and perform the limit $d_2 \to 0$ or $w_2 \to 0$; then the bisoliton is reduced to the soliton Table I, part C2 for $j = 0$. Conversely, starting from Table II, part C1 for $j_0 = 0$, defining $w = w_1$, $j_0 = (a - c)n_0/2\nu$, and letting $d_2 \to 0$ we find the soliton $j_0 \neq 0$ of Table II, part C2. In conclusion the bisolitons obtained are really the two-dimensional extension of the solitons of Sec. II A.

D. Invariance properties

The bisolitons (2.9a) and (2.9b) depend on the two parameters $n_0(j_0)$, ν which are the coefficients of the w_i's and on the two arbitrary constants d_i contained only in the w_i. In fact, two invariance properties for $N \pm J$ allow a reduction of the $n_0(j_0)$, ν intervals which must be studied.

(i) For the bisolitons $j_0 = 0$ or $j_0 \neq 0$ let us define the

TABLE II. Bisolitons.

A:	Ansatz $N = n_0 + \Sigma n_i/\Delta_i$, $J = j_0 + \Sigma j_i/\Delta_i$, $\Delta_{.i} = 1 + w_i$ $w_i = d_i \exp(\gamma_i x + \rho_i t)$
B:	Relations (1) $j_i\gamma_i + n_i\rho_i = 0$, (2) $j_0[(a + c)j_0 + (a - c)n_0] = 0$, (3) and (4) $(n_i\gamma_i + j_i\rho_i) = -2j_0j_i(a + c) - (a - c)(j_0n_i + j_in_0) = j_i\{(a + c)j_i + (a - c)n_i\}$, (5) $a + c + (a - c)/(\gamma_1/\rho_1 + \gamma_2/\rho_2)/2 = 0$
C:	Algebraic solutions definition: $\gamma_1/\rho_1 = (-2\nu + a + c)/(a - c)$, $\gamma_2/\rho_2 = (2\nu + a + c)/(a - c)$
C1:	$j_0 = 0$, $N - J = n_0\{1 + (-1 + a/\nu)/\Delta_1 - (1 + a/\nu)/\Delta_2\}$, $\rho_1 = n_0(c - a)^3/4(\nu - c)(a - \nu)$ $N + J = n_0\{1 + (-1 + c/\nu)/\Delta_1 - (1 + c/\nu)/\Delta_2\}$, $\rho_2 = \rho_1(-\nu)$
C2:	$j_0 \neq 0$, $N - J = 2j_0[a + (\nu - a)/\Delta_1 - (\nu + a)/\Delta_2]/(c - a)$, $\rho_1 = j_0\nu(a - c)^2/2(\nu - a)(\nu - c)$, $N + J = 2j_0[c + (\nu - c)/\Delta_1 - (\nu + c)/\Delta_2]/(c - a)$, $\rho_2 = \rho_1(-\nu)$
D:	Invariances: $\mathscr{T}_{\rm I}$: $\{n_0 \to -n_0$ (or $j_0 \to -j_0$), $d_i \to d_i^{-1}$, ν fixed$\}$, $\mathscr{T}_{\rm II}$: $\{n_0$ (or j_0) fixed, $d_1 \rightleftarrows d_2$, $\nu \to -\nu\}$
E:	Positive solutions on a semiline $x \geqslant 0$ with real Δ_i
E1:	$j_0 = 0$, $a + c < 0$, $0 < \nu < \inf\{c, -(a + c)/2\}$, $1 < c/\nu < d_2 \leqslant d_1$, $\rho_1 > \rho_2 > 0$, $\gamma_1 > \gamma_2 > 0$, $2J = n_0(c - a)(w_2 - w_1)/\nu\Delta_1\Delta_2 \leqslant 0$ if $d_1 = d_2$, $N \to n_0$
E2:	$j_0 > 0$, $a + c < 0$, $0 < \nu < \inf\{c, -(a + c)/2\}$, $1 < c/\nu < d_2$, $d_1 \leqslant d_2$, $\rho_1 < 0, \rho_2 > 0, \rho_1 + \rho_2 < 0, \gamma_1 < 0, \gamma_2 > 0, \gamma_1 + \gamma_2 < 0$, $J = j_0(w_1w_2 - 1)/\Delta_1\Delta_2 \leqslant 0$, $N \to 2j_0\nu/(a - c)$

transforms $\mathscr{T}_{\rm I}$ with the following changes in the parameters:

$$\mathscr{T}_{\rm I}\{n_0\to -n_0 \;(\text{or}\, j_0\to -j_0),\, d_i\to d_1^{-1},\, \nu \text{ fixed}\}.$$

With the transforms $\mathscr{T}_{\rm I}$ we obtain $\rho_i\to-\rho_i$, $\gamma_i\to-\gamma_i$, $w_i\to w_i^{-1}$, $\Delta_i\to w_i^{-1}\Delta_i$, and finally $\mathscr{T}_{\rm I}(N\pm J)\to N\pm J$.

(ii) Similarly let us define a second class of transforms $\mathscr{T}_{\rm II}$,

$$\mathscr{T}_{\rm II}\{n_0\to n_0,\, d_1\rightleftarrows d_2,\, \nu\to-\nu\}.$$

With $\mathscr{T}_{\rm II}$ we find $\rho_1\rightleftarrows\rho_2$, $\gamma_1\rightleftarrows\gamma_2$, $\Delta_1\rightleftarrows\Delta_2$ and finally $\mathscr{T}_{\rm II}(N\pm J)\to N\pm J$. Without loss of generality we can, later on, restrict our study to $n_0>0$ ($j_0>0$) and $\nu>0$. The properties corresponding to other domains $n_0<0$ ($j_0<0$), $\nu<0$ can be obtained by applying $\mathscr{T}_{\rm I}$ and $\mathscr{T}_{\rm II}$.

III. BISOLITONS WITH REAL EXPONENTIAL VARIABLES

The properties discussed up to now for the bisolitons were algebraic and valid for γ_i,ρ_i real or complex. Here we assume *that γ_i,ρ_i are real $n_0>0$ ($j_0>0$), $\nu>0$, $d_i>0$ (or $\Delta_i\neq0$) and discuss the physically acceptable solutions.*

A. Physical solutions on the full line $x\in[-\infty,+\infty]$

We study the asymptotic positivity $|x|\to\infty$ noticing that if one $\Delta_i\to\infty$ for one limit, necessarily the same $\Delta_i\to1$ for the other. Two possibilities can occur.

(i) For both Δ_i the limits are $+\infty$ and 1. For $j_0=0$ we find $(N\mp J)\to n_0$ for one limit and $-n_0$ for the other, for $j_0\neq0$ we find for $N+J$ the two opposite limiting values $\pm 2j_0c/(c-a)$ and for $N-J$ the two opposite limiting values $\pm 2j_0a/(c-a)$. In all cases the positivity is violated.

(ii) Let us assume that $\Delta_1\to\infty$, $\Delta_2\to1$ for one asymptotic limit and $\Delta_1\to1$, $\Delta_2\to\infty$ for the other. Then for $j_0=0$, $N-J$ (or $N+J$) has two opposite limiting values $\pm n_0a/\nu$ (or $\pm n_0c/\nu$), for $j_0\neq0$, similarly we obtain $N-J$ (or $N+J)\to\pm 2j_0\nu/(c-a)$, the only escape $\nu=0$ leading to $\gamma_i=\rho_i=0$, $w_i=$ const.

In conclusion, the bisolitons with real γ_i,ρ_i cannot be physical solutions on the full line. This means that the escape $ac=0$ for the solitons do not hold for the bisolitons.

B. Physical solutions on a semiline $x\geqslant0$ (we recall that $n_0>0$, $\nu>0$, $d_i>0$)

1. Solutions $j_0=0$

First, looking at the asymptotic positivity for either $t\to\infty$ or $x\to\infty$ we show that necessarily $\gamma_i>0,\rho_i>0$. It is sufficient to notice that when $x\to\infty$, (i) if $\gamma_1>0,\gamma_2>0$, then $N\mp J\to n_0$; (ii) if $\gamma_1>0$, $\gamma_2>0$, $N-J\to -n_0a/\nu$, $N+J\to n_0c/\nu<0$, except for the particular case $c=0$; (iii) if $\gamma_1<0$, $\gamma_2<0$, $N\mp J\to -n_0<0$; (iv) if $\gamma_1<0$, $\gamma_2<0$, $N-J\to n_0a/\nu<0$ except for the particular case $a=0$. The same proof works for ρ_1,ρ_2. Can we have solutions for these particular cases $c=0$ in (ii) and $a=0$ in (iv)? Looking at the explicit expressions (2.9a) of the γ_i,ρ_i as functions of n_0, ν we find that the γ_i,ρ_i do not have the signs prescribed in these particular cases. On the contrary we find ν values for which the ρ_i and the γ_i are positive. We obtain $\rho_1>0$ if $a<\nu<c$, $\rho_2>0$ if $-c<\nu<-a$, $\gamma_1>0$ if $(a+c)/2<\nu<c$, $\gamma_2>0$ if $-c<\nu<-(a+c)/2$. Consequently on the one hand we have the restriction $a+c<0$ on the parameters of the Illner model and on the other hand the condition $0<\nu<\inf(c,-(a+c)/2)$ on the ν parameter.

Second, for the last positivity constraint $N\pm J>0$ for any $t\geqslant0$ values we seek the signs of $\rho_1-\rho_2$, $\gamma_1-\gamma_2$ for n_0, ν, a, c satisfying the previous conditions: from (2.8) and (2.9a) we have

$$\rho_1-\rho_2=\frac{n_0(c-a)^3\nu(a+c)}{2(\nu^2-c^2)(a^2-\nu^2)},\qquad \gamma_1-\gamma_2=\frac{n_0(c-a)^2\nu(\nu^2-(a^2+c^2)/2)}{(\nu^2-c^2)(a^2-\nu^2)},\tag{3.1}$$

and find $\rho_1>\rho_2>0$, $\gamma_1>\gamma_2>0$ in the ν assumed interval. Then from (2.9a) we easily obtain two lower bounds $\forall x\geqslant0$, $\forall t\geqslant0$,

$$\Delta_1\Delta_2(N-J)/n_0>-1+d_1d_2-a(d_1-d_2)/\nu,\qquad \Delta_1\Delta_2(N+J)/n_0>-1+d_2c/\nu-d_1(d_2-c/\nu),\tag{3.2}$$

from which we see that the positivity $\forall x\geqslant0$, $\forall t\geqslant0$ is satisfied if $d_1\geqslant d_2>c/\nu>1$. In conclusion there exists solutions $j_0=0$ positive on the semiline $x\geqslant0$.

2. Solutions $j_0>0$

First we look at the asymptotic positivity when either $t\to\infty$ or $x\to\infty$ and show that necessarily $\gamma_1<0$, $\rho_1<0$, $\gamma_2>0,\rho_2>0$. It is sufficient to notice that when $x\to\infty$: (i) if $\gamma_1>0$, $\gamma_2>0$, $N+J\to 2j_0c/(c-a)>0$, $N-J\to 2j_0a/(c-a)<0$ (except if $a=0$); (ii) if $\gamma_1>0$, $\gamma_2<0$, $N\pm J\to-2j_0\nu/(c-a)<0$; (iii) if $\gamma_1<0$, $\gamma_2<0$, $N-J\to-2j_0a/(c-a)>0$, $N+J\to-2j_0c/(c-a)<0$ (except if $c=0$); (iv) $\gamma_1<0$, $\gamma_2>0$, $N-J$ [or $(N+j)]\to 2j_0\nu/(c-a)>0$. Concerning the particular $a=0$ and $c=0$ cases, from the explicit expression (2.9b) of the γ_i,ρ_i, we do not find ν values such that the ρ_i,γ_i have the appropriate signs. On the contrary for $ac<0$, we find ν such that γ_1,ρ_1 are negative and γ_2,ρ_2 are positive. From (2.9b), for the above ρ_i signs we see that ν must satisfy $0<\nu<\inf(c,-a)$ and for the γ_i signs, $0<\nu<\inf(c,-(a+c)/2)$. Finally we have the same restrictions as for the j_0 solution, i.e., $a+c<0$ and $0<\nu<\inf(c,-(a+c)/2)$.

Second, from (2.9b) we want to deduce lower bounds such that restrictions on the d_i will give sufficient positivity conditions $\forall x\geqslant0$, $\forall t\geqslant0$. Here we must have information on the signs of $\rho_1+\rho_2$, $\gamma_1+\gamma_2$,

$$\rho_1+\rho_2=\frac{j_0\nu^2(a-c)^2(a+c)}{(\nu^2-a^2)(\nu^2-c^2)},\qquad \gamma_1+\gamma_2=\frac{j_0\nu^2(c-a)(2\nu^2-a^2-c^2)}{(\nu^2-a^2)(\nu^2-c^2)},\tag{3.1'}$$

and we find $0<\rho_2<-\rho_1$, $0<\gamma_2<-\gamma_1$. The two lower bounds are easily deduced

$$\begin{aligned}&\Delta_1\Delta_2(N-J)\\&\quad>[(1-d_1d_2)(-a)+\nu(d_2-d_1)]2J_0c/(c-a),\\&\Delta_1\Delta_2(N+J)\\&\quad>[-1+\nu d_2/c+w_1(d_2-\nu/c)]2j_0c/(c-a),\end{aligned}\tag{3.2'}$$

from which we obtain that the positivity $\forall x \geqslant 0$, $\forall t \geqslant 0$ is satisfied if $d_1 \leqslant d_2$, $d_1 d_2 \leqslant 1$, $d_2 > c/v > 1$. A summary of the results of this subsection is quoted in Table II, Part E 1.2. For both $j_0 = 0$ and $j_0 \neq 0$ solutions the sufficient positivity conditions and the restrictions on the parameters belonging to other intervals $n_0 < 0$, $j_0 < 0$, $v < 0$ can be deduced with the help of the $\mathscr{T}_{\mathrm{I}}$ and $\mathscr{T}_{\mathrm{II}}$ transforms.

C. Specular reflection boundary condition at $x = 0$

Let us require $N - J = N + J$ at $x = 0$ or $J = 0$ at $x = 0$. Necessarily we consider the $j_0 = 0$ solution and from (2.9a) obtain $\Delta_1(x = 0,t) = \Delta_2(x = 0,t)$ or $d_1 = d_2$ and $\rho_1 = \rho_2$. From (3.1) we see that for the bisolitons the only possibility is $a + c = 0$ for which $\gamma_1 + \gamma_2 = 0$. This means that for the present class of solutions, the Carleman model is the only possible one with this type of boundary.

D. Multisolitons?

Can we have more than bisolitons or can we find solutions containing N solitons with components $w_i = d_i \exp \rho_i (t + x\gamma_i/\rho_i)$, $N > 2$, γ_i/ρ_i different values? Requiring that when $d_j \to 0$, $j > 2$ the solution reduces to the previous bisoliton, the simplest ansatz is $N = n_0 + \Sigma n_i/\Delta_i$, $J = j_0 + \Sigma j_i/\Delta_i$, $\Delta_i = 1 + w_i$. The linear mass conservation gives $j_i\gamma_i + n_i\rho_i = 0$ and from the coefficient of $(\Delta_i\Delta_k)^{-1}$ in the nonlinear Illner equation we find

$$a + c + \frac{(a-c)}{2}\left(\frac{\rho_i}{\gamma_i} + \frac{\rho_k}{\gamma_k}\right) = 0, \quad \forall i, \forall k, c \neq k.$$

For the bisoliton $N = 2$ there is only one relation which expresses the coupling between the two soliton components. On the contrary for $N > 2$ the set of relations leads $\forall i$ to $\gamma_i/\rho_i =$ const independent of i or to an impossibility.

IV. BISOLITONS WITH COMPLEX CONJUGATE EXPONENTIAL VARIABLES

Let us come back to the algebraic bisolitons written down in (2.3a) and (2.3b) (or Table I, part C1.2), *assume that the v parameter is purely imaginary $v = i v_{\mathrm{I}}$* and recall that $\rho_2 = \rho_1(-v)$, $\gamma_2 = \gamma_1(-v)$. We have $\rho_1 = \rho_1(i v_{\mathrm{I}})$, $\rho_2 = \rho_1(-i v_{\mathrm{I}}) = \rho_1^*$, $\gamma_2 = \gamma_1^*$. Let us define $\rho = \rho_1$, $\gamma = \gamma_1$, $w = w_1$, $d = d_1$, $w = d \exp(\gamma x + \rho t)$, and choose $d_2 = d^*$. It follows that $w_2 = w^*$, $\Delta_2 = \Delta^*$, and $N \pm J$ are real. We notice that these solutions could as well be obtained directly starting with an ansatz $N = n_0 + 2\,\mathrm{Re}(n/\Delta)$, $J = j_0 + 2\,\mathrm{Re}(j/\Delta)$, $\Delta = 1 + w$, n, j, γ, ρ being complex.

The invariance of Sec. II D is still valid and the transforms can be written $\mathscr{T}_{\mathrm{I}}\{n_0 \to -n_0$ (or $j_0 \to -j_0$), $d \to d^{-1}$, $\mathscr{T}_{\mathrm{I}}$ fixed$\}$ and $\mathscr{T}_{\mathrm{II}}$ $\{n_0$ (or j_0) fixed, $d \to d^*$, $v_{\mathrm{I}} \to -v_{\mathrm{I}}\}$. Consequently we still restrict our study to $n_0 > 0$ (or $j_0 > 0$) and $v_{\mathrm{I}} > 0$. We still restrict our study to $n_0 > 0$ (or $j_0 > 0$) and $v_{\mathrm{I}} > 0$. We still have two classes of solutions: $j_0 = 0$, $j_0 \neq 0$, however, the last one violates positivity as we show. The solutions $j_0 \neq 0$ can be written as

$$(N - J)|\Delta|^2 = 2j_0(a(|w|^2 - 1) + 2v_{\mathrm{I}} w_{\mathrm{I}})/(c - a),$$
$$w_{\mathrm{I}} = \mathrm{Im}\, w, \tag{4.1}$$
$$(N + J)|\Delta|^2 = 2j_0(c(|w|^2 - 1) + 2v_{\mathrm{I}} w_{\mathrm{I}})/(c - a),$$
$$j_0 \neq 0.$$

When either t or $|x|$ go to infinity there exist two possible asymptotic behaviors: (i) either $|\omega| \to \infty$, $N - J \to 2j_0 a/(c - a)$ which violates positivity unless $a = 0$, (ii) or $|w| \to 0$, $N + J \to -2j_0 c/(c - a)$ still violating positivity unless $c = 0$. Further if $a = 0$ (or $c = 0$), then $N - J$ (or $N + 1$), proportional to w_{I} changes sign for x and t varying.

Hereafter we always consider the $j_0 = 0$ solution that we write (see also Table III, part B)

$$\begin{aligned}
(N - J)|\Delta|^2 &= n_0(-1 + |w|^2 - 2a w_{\mathrm{I}}/v_{\mathrm{I}}),\\
(N + J)|\Delta|^2 &= n_0(-1 + |w|^2 - 2c w_{\mathrm{I}}/v_{\mathrm{I}}),\\
\rho &= \frac{n_0(c - a)^3(v_{\mathrm{I}}^2 - ac - i v_{\mathrm{I}}(a + c))}{4(c^2 + v_{\mathrm{I}}^2)(a^2 + \nabla_{\mathrm{I}}^2)},\\
\gamma &= \frac{n_0(c - a)^2((a + c)(v_{\mathrm{I}}^2 + ac) + i v_{\mathrm{I}}(2v_{\mathrm{I}}^2 + a^2 + c^2))}{4(c^2 + v_{\mathrm{I}}^2)(a^2 + v_{\mathrm{I}}^2)},
\end{aligned} \tag{4.2}$$

with $\rho = \rho_R + i\rho_{\mathrm{I}}$, $\gamma = \gamma_R + i\gamma_{\mathrm{I}}$. Let us show that if $\gamma_R \neq 0$, physical positive solutions on the full x axis are not possible. When $|x| \to \infty$, on one side $|w| \to \infty$, $N - J \to n_0$ while on the

TABLE III. Bisolitons with complex Δ: $v = i v_{\mathrm{I}}$.

A:	Ansatz $N = n_0 + 2\,\mathrm{Re}\, n/\Delta$, $J = j_0 + 2\,\mathrm{Re}\, j/\Delta$, $\Delta = 1 + d\exp(\gamma x + \rho t)$ n_0, j_0 real, n, j, γ, ρ, d complex
B:	Algebraic solutions $j_0 = 0$ $N - J = n_0[1 - 2\,\mathrm{Re}\,\Delta^{-1} + (2a/v_{\mathrm{I}})\mathrm{Im}\,\Delta^{-1}]$ $N + J = n_0[1 - 2\,\mathrm{Re}\,\Delta^{-1} + (2c/v_{\mathrm{I}})\mathrm{Im}\,\Delta^{-1}]$ $\rho = \dfrac{n_0(c - a)^3[v_{\mathrm{I}}^2 - ac - i v_{\mathrm{I}}(a + c)]}{4(v_{\mathrm{I}}^2 + c^2)(v_{\mathrm{I}}^2 + a_2)}$, $\gamma = \rho(a + c - 2i v_{\mathrm{I}})/(a - c)$
C:	Invariances $\mathscr{T}_{\mathrm{II}}$: $\{n_0 \to -n_0$ (or $j_0 \to -j_0$), $d \to d^{-1}$, v_{I} fixed$\}$, $\mathscr{T}_{\mathrm{II}}$: $\{n_0$ (or j_0) fixed, $d \to d^*$, $v_{\mathrm{I}} \to -v_{\mathrm{I}}\}$.
D:	Periodic solutions $n_0 > 0$, $v_{\mathrm{I}} > 0$
D1:	$v_{\mathrm{I}} = \pm\sqrt{-ac}$, $\rho_R = n_0(c - a)/2 > 0$ for $n_0 > 0$, for $v_{\mathrm{I}} > 0$, $\rho_{\mathrm{I}} = n_0(a^2 - c^2)/4\sqrt{-ac}$, $\gamma_{\mathrm{I}} = n_0(a - c)^2/4\sqrt{-ac}$, $N \mp J \underset{t \to \infty}{\to} n_0 > 0$, $n \pm J > 0$ if $\lvert d\rvert > \sup(\sqrt{\alpha_i} + \sqrt{1 + \alpha_i})$, $\alpha_1 = \lvert c/a\rvert$, $\alpha_2 = \lvert a/c\rvert$; for $n_0 < 0$, $\rho_R < 0$, $N \pm J > 0$ if $\lvert d\rvert < \inf(-\sqrt{\alpha_i} + \sqrt{1 + \alpha_i})$.
D2:	$a + c = 0$, $\gamma_{\mathrm{I}}/v_{\mathrm{I}} = 2n_0 c^2/(c^2 + v_I^2) = \rho_R/c$, $\gamma_R = \rho_{\mathrm{I}} = 0$, $N \pm J \underset{t \to \infty}{\to} n_0 > 0$ $N \pm J > 0$ if $\lvert d\rvert > \lvert c/v_{\mathrm{I}}\rvert + \sqrt{1 + \lvert c/v_{\mathrm{I}}\rvert^2}$; If $v_{\mathrm{I}} = c$, N and J harmonic conjugate
E:	Positive soliton on a semiline $x \geqslant 0$ $n_0 > 0$, $\rho_R > 0$, $\gamma_R > 0$, no restriction on v_{I} if $a = 0$ or $c = 0$, $v_{\mathrm{I}}^2 > -ac$ if $a + c < 0$, $v_{\mathrm{I}}^2 < -ac$ if $a + c < 0$, $\lvert d\rvert > \sup \beta_i + \sqrt{1 + \beta_i^2}$, $\beta_1 = \lvert a/v_{\mathrm{I}}\rvert$, $\beta_2 = \lvert c/v_{\mathrm{I}}\rvert$, $J(x = 0,t) = n_0(c - a)v_{\mathrm{I}}^{-1}\,\mathrm{Im}\,\Delta^{-1}$ changes sign, $N \to n_0$.

other side $|w|\to 0, N-J\to -n_0$. There remains two possibilities: either $\gamma_R=0$, and at fixed t, the solutions are periodic on the full x axis, or $\gamma_R\neq 0$ and we must restrict our study to some semi-x axis (for instance $x\geqslant 0$).

A. Periodic solutions $\gamma_R=0$

From (3.2) we see that $\gamma_R=0$ either if $v_1^2+ac=0$ or if $a+c=0$.

1. $v_1^2=-ac$ $(a+c\neq 0, a\neq 0, c\neq 0)$

We choose $v_1=\sqrt{-ac}>0$, and begin with $n_0>0$ and find

$$\rho_R=n_0(c-a)/2>0,\quad \rho_1=n_0(a^2-c^2)/4\sqrt{-ac},$$
$$\gamma_1=n_0(a-c)^2/4\sqrt{-ac} \tag{4.3}$$

($\rho_R>0$ even if $v_1<0$). When $t\to\infty$, $|w|\to\infty$, $N\mp J\to n_0>0$, let us seek sufficient conditions on d such that the positivity is satisfied for any $t\geqslant 0$. From (4.2) we deduce two lower bounds, $(N\mp J)|\Delta|^2/n_0+1-|w|^2$ greater than either $-2|w|\sqrt{|a/c|}$ or $-2|w|\sqrt{|c/a|}$ from which we obtain

$$N\pm J>0 \text{ if } |d|\exp\rho_R t>\sup_i\,(\sqrt{\alpha_i}+\sqrt{1+\alpha_i}),$$
$$\alpha_1=|a/c|,\quad \alpha_2=|c/a|,\quad n_0>0. \tag{4.4}$$

Recalling $\rho_R>0$, it is sufficient that the inequality (4.4) holds at $t=0$ for $|d|$ alone. For the other sign choice $v_1<0$, with the help of the transform $\mathscr{T}_{\rm II}$, we must replace d by d^* leading to the same result for $|d|$. We notice that in (4.4) the sup is larger than 1 so that $|d|>1$ avoids the possibility $\Delta=0$ for some x, t values.

For $n_0<0$, we want to check the validity of our invariance properties applied to the positivity constraints. First as above, we deduce directly the sufficient positivity condition: $n_0<0$ leads to $\rho_R<0$, $N\pm J\to -n_0$ and we deduce from (4.2) two lower bounds $(N\mp J)/(-n_0)-1+|w|^2$ greater than either $-2|w|\sqrt{|a/c|}$ or $-2|w|\sqrt{|c/a|}$. It follows that

$$N\pm J>0 \text{ if } |d|\exp\rho_R t>\sup_i\,(\sqrt{\alpha_i}+\sqrt{1+\alpha_i}),$$
$$n_0<0,\quad t\geqslant 0,\quad x\geqslant 0, \tag{4.4$'$}$$

with the same α_i as in (4.4). Here also it is sufficient to apply (4.4′) at $t=0$ for $|d|$. Second we apply the transform $\mathscr{T}_{\rm I}$ to our above result obtained for $n_0>0$. In (4.4), for $t=0$ we replace d by d^{-1}. It is trivial to verify that we obtain the same condition (4.4′) for $t=0$. We notice also that in (4.4′) the inf being less than 1 avoids the possibility $\Delta=0$.

2. $a+c=0$, Carleman model

We begin with $n_0>0$ and find

$$\gamma_1 v_1^{-1}=2n_0c^2/(v_1^2+c^2)=\rho_R c^{-1}>0,$$
$$\gamma_R=\rho_1=0,\quad w=d\exp\gamma_1(c\gamma_1^{-1}t+ix) \tag{4.5}$$

and $N\pm J\underset{t\to\infty}{\to} n_0$. Lower bounds, easily deduced from (3.2), lead to sufficient $t\geqslant 0$, $x\geqslant 0$, positivity conditions,

$$(N\pm J)|\Delta|^2n_0^{-1}>-1+|d|^2-2c|d|\,|v_1^{-1}>0$$
$$\text{if } |d|>|c/v_1|+\sqrt{1+|c/v_1|^2}\,. \tag{4.6}$$

For the other choices $v_1<0$, $n_0<0$ we use the transforms $\mathscr{T}_{\rm I}$, $\mathscr{T}_{\rm II}$ and proceed as above. For instance for $n_0<0$ we find $|d|<-|c/v_1|+\sqrt{1+|c/v_1|^2}$.

Let us consider the restricted class of solutions such that $v_1=c$. Then $\gamma_1=\rho_R$ and the condition $\rho^2+\gamma^2=0$ is satisfied. In the case N and J are harmonic conjugate functions. In (4.6) the positivity condition becomes $|d|>1+\sqrt{2}$ for $n_0>0$. If we require further that d is real then the solution becomes identical to the one determined previously by Wick.[9] However, let us recall that the solution exists for any v_1 real values for which N and J are not, in general, conjugate harmonic. Further when v_1 crosses the value c, nothing special happens for the solutions.

3. Physical interpretation of the periodic solutions

We write down the total mass N and the current J and look at their large time behavior

$$N/n_0-1=2\,{\rm Re}(1+i(a+c)/2v_1)\Delta^{-1}$$
$$\simeq e^{-\rho_R t}A_N\cos(\gamma_1 x+\rho_1 t+\phi_N),$$
$$J/n_0=2\,{\rm Re}(a-c)(2v_1\Delta)^{-1}$$
$$\simeq e^{-\rho_R t}A_j\cos(\gamma_1 x+\rho_1 t+\phi_J), \tag{4.7}$$

with A_N, A_J being positive constants and ϕ_N, ϕ_J constant phase factors. They represent damped ($\rho_R>0$) oscillating and propagating (if $\rho_1\neq 0$) waves. Notice that for the current $|J|\to 0$ when $t\to\infty$ and in the mean the flux of particles is equivalent to zero. We discuss the two different cases $a+c\neq 0$ (see Sec. IV A 1) and $a+c=0$ (see Sec. IV A 2).

(i) $a+c\neq 0$. The solutions correspond to damped sound waves. $Nn_0^{-1}-1\simeq(\exp-\rho_R t)\cos(\gamma_1(x-v_0t)+\phi_N)$, where the sound speed $|v_0|=|a+c|/(c-a)$, dependent only on the Illner parameter values a, c, is the same for all solutions [ρ_R, γ_1 are written down in (4.3)]. For a periodic solution, the asymptotic mass value n_0 is fixed and we have only one sound mode. In the dispersion relation γ_1 wave number versus ρ_1 frequency we have only one value. If the ratio of the absorption coefficient ρ_R by the frequency ρ_1 is not small, we have very few effective oscillations when the time is growing. A good criterion for sound waves (not too strongly damped) is

$$|\rho_R/\rho_1|=|2\sqrt{-ac}/(a+c)|\ll 1. \tag{4.8}$$

For c fixed taking Illner models with $-a$ increasing we must observe more and more effective oscillations in time of N for a fixed x value. Choosing $c=1$, and $a=-9$, -225, -625, -2500 with ratios $\rho_R/\rho_1=0.75$, 0.13, 0.08, 0.04 we observe an increasing number 1, 3, 6, 13 of oscillations (see Sec. VI).

(ii) For the Carleman model $\rho_1=0$, the oscillations are not propagating. The periodic solutions cannot correspond to sound waves, they are damped oscillating waves.

B. Physical solutions on a semiline $x\geqslant 0$

We assume $n_0>0$ (for $n_0<0$ we use $\mathscr{T}_{\rm I}$). In order to avoid a positivity violation when x or t goes to infinity [see

(4.2)], we must, in both limits have $|w| \to \infty$ or $\rho_R > 0$, $\gamma_R > 0$ ($N \mp J \to n_0 > 0$). However, ρ_R is always positive while, with γ_R having the sign of $(a+c)(v_I^2 + ac)$, we find $\gamma_R > 0$ either for $a + c > 0$, $v_I^2 > -ac$ or for $a+c<0$, $v_I^2 < -ac$. In the following, we assume that these restrictions are satisfied, still choosing $n_0 > 0$.

In order to find sufficient $t \geqslant 0$, $x \geqslant 0$ positivity conditions we deduce two lower bounds from (4.2): $(N-J)|\Delta|^2 n_0^{-1} + 1 - |w|^2$ greater than either $-2|wa/v_I|$ or $-2|wc/v_I|$, with $|w| = |d|\exp(\gamma_R x + \rho_R t)$. We find $N \pm J > 0$, $\forall t \geqslant 0$, $\forall x \geqslant 0$ if

$$|d| > \sup_i \beta_i + \sqrt{1+\beta_i^2}, \quad \beta_1 = |a/v_I|, \quad \beta_2 = |c/v_I|. \tag{4.9}$$

We emphasize that such solutions exist for all parameters values $a \leqslant 0$, $c \geqslant 0$ of the Illner model. Note that for $a = 0$ or $c = 0$ we have no restrictions on v_I while in the other cases the restrictions written above exist. For $a = 0$ or $a + c > 0$, the sup in (4.9) is given by the second term, while if $c = 0$ or $a + c < 0$, it is the first one. The Carleman model solution (Sec. IV A 2) for which $\rho_R > 0$, $\gamma_R > 0$ is for $x \geqslant 0$ a particular case of the present class of solutions.

Can we have a *specular reflection boundary at* $x = 0$ or $J(x=0,t) = 0$? We must choose the $j_0 = 0$ solution and from (4.2) we find that $\operatorname{Im} w = 0$ or d real $\rho_I = 0$. Once more, the only possible model is the Carleman one $(a+c) = 0$.

C. Conjugate harmonic *N,J* functions

Let us assume that in addition to the mass conservation law $N_t + J_x = 0$, N and J satisfy $N_t - J_x = 0$ or $(\partial_{t^2}^2 + \partial_{x^2}^2)(N,J) = (0,0)$. Then N and J become harmonic conjugate functions. For the ansatz (2.7) this means $\gamma_i^2 + \rho_i^2 = 0$, which is not possible for real γ_i, ρ_i. In the present section, where the γ_i and the ρ_i are complex conjugate, we must have $\rho^2 + \gamma^2 = 0$ or the two conditions $\rho_R \rho_I + \gamma_R \gamma_I = 0$, $\rho_R^2 + \gamma_R^2 = \rho_I^2 + \gamma_I^2$.

Let us check on the explicit ρ, γ written down in (4.2) for these two conditions. From the first condition we find two possibilities: either $a + c = 0$ (Carleman model) or $v_I^4 + 2v_I^2 ac + ac(a^2 + c^2 - ac) = 0$. The second condition for $a + c = 0$ gives $|v_I| = c$ for the Carleman model (see Sec. IV A 1). On the contrary, the second condition, for $a + c \neq 0$ in the first condition, leads after some algebraic calculation to the result $v_I^2 = ac$ which is impossible for the Illner model $ac < 0$. Finally the possibility of constructing explicit N and J harmonic conjugate functions exists only for a particular value of the parameter of the Carleman model. For instance we can directly check, for the periodic solutions $v_I^2 = -ac$ of III A 1, that $\gamma^2 + \rho^2 \neq 0$.

V. PHYSICAL INTERPRETATION OF THE POSITIVE SOLUTIONS ON A SEMILINE, $x \geqslant 0$

In analogy with a previous physical interpretation[13] of the inhomogeneous Kac model solution, which was positive only inside a well-defined interval, we define new distributions $\tilde{f}_\eta = \theta(x) f_\eta$ which are identically zero for $x < 0$ (outside the semiaxis $x \geqslant 0$). Starting with the kinetic equations for f_η: $(\partial_t + \eta\,\partial_x) f_\eta = \eta \operatorname{col}(f_+, f_-)$ we deduce the corresponding kinetic equations for $\tilde{f}_\eta$

$$\begin{aligned} &(\partial_t + \eta\,\partial_x)\tilde{f}_\eta = \eta \operatorname{Col}(\tilde{f}_+, \tilde{f}_-) + \eta S + S_0, \\ &\eta S + S_0 = \eta \delta(x) f_\eta (x = 0,t), \\ &2S = \delta(x) N(x=0,t), \quad 2S_0 = \delta(x) J(x=0,t). \end{aligned} \tag{5.1}$$

We must interpret physically the two supplementary terms S_0 and ηS. For a positive additional term, like a gain term in the collision term, it is interpreted as a source term at $x = 0$, while a negative term, like a loss term, is interpreted as a sink at $x = 0$.

(i) The term $\eta\delta(x)N(x=0,t)/2$, which, due to the positivity of N, has the sign $\eta = \pm 1$, is interpreted as a source for particles of velocity $+1$ and a sink for particles of velocity -1. The amount of incoming and outgoing particles being the same, ηS can be viewed as an elastic wall at $x = 0$. In general $N(x = 0,t) \to \text{const}$ when $t \to \infty$, so that the elastic wall is always present. For a perfect specular reflection boundary condition at $x = 0$, then $J \equiv 0$ or $S_0 \equiv 0$, the elastic wall is the only supplementary term on the rhs of (5.1). (This happens only for the Carleman model choosing $d_1 = d_2$ if Δ is real and d real for Δ complex.)

(ii) The second term $\delta(x)J(x=0,t)/2$, which is the same for both particles of velocities ± 1, has not necessarily a definite sign.

If $J(x=0,t)$ does not change sign, then S_0 can be interpreted as a source ($J > 0$) or as a sink ($J < 0$). If it changes sign for different time values $t_1, t_2, \ldots$, then it acts like a source (or a sink) during the time intervals $[t_i, t_{i+1}]$ in which it is positive (negative). Nevertheless, in general, we find that it decreases exponentially in time, such that for infinite time, it becomes negligible compared with the elastic wall ηS.

We discuss the different possibilities with the restrictions $n_0 > 0$ (or $j_0 > 0$) $v > 0$ (or $v_I > 0$). For the other sign cases we must use the transforms $\mathscr{T}_I$ and $\mathscr{T}_{II}$.

A. Solutions with real γ_i, ρ_i and $j_0 = 0$ (see Sec. III B 1)

From (2.9a) we find for J and N

$$\begin{aligned} 2J &= \frac{n_0(c-a)}{v}\left(\frac{1}{\Delta_1} - \frac{1}{\Delta_2}\right) \\ &= \frac{n_0(c-a)(w_2 - w_1)}{v\Delta_1\Delta_2}, \quad N \to n_0, \quad t \to \infty. \end{aligned} \tag{5.2a}$$

We recall that $a + c < 0$, $0 < v < \inf(c, -(a+c)/2)$, $\rho_1 > \rho_2 > 0$, or $w_2 w_1^{-1} \to 0$ when $t \to \infty$. If we choose $d_1 = d_2$, then $w_2 - w_1 \leqslant 0$, $j \leqslant 0$ while $|J| \simeq 0(\exp - \rho_2 t)$. Thus S_0 looks like a sink which, when the time increases, becomes negligible compared with the elastic wall ηS, $|S_0/\eta S| \simeq 0(\exp - \rho_2 t)$.

B. Solutions with real γ_i, ρ_i and $j_0 > 0$ (see Sec. III B 2)

From (2.9b) we find for J, N

$$\begin{aligned} &J = j_0\left(1 - \frac{1}{\Delta_1} - \frac{1}{\Delta_2}\right) = j_0 \frac{(-1 + w_1 w_2)}{\Delta_1 \Delta_2}, \\ &N \to 2j_0 v/(c-a) t \to \infty. \end{aligned} \tag{5.2b}$$

We recall that for $a + c$, v we have the same restrictions as above. Further here $\rho_1<0$, $\rho_2>0$, $\rho_1+\rho_2<0$, $d_1 d_2 \leqslant 1$, or $J\leqslant 0$. As above S_0 looks like a sink and due to $|J| \simeq 0(\exp-\rho_2 t)$ it becomes, at infinite time, negligible compared with the elastic wall, $|S_0/\eta S| \simeq 0(\exp-\rho_2 t)$.

C. Solutions with complex conjugate γ_i, ρ_i (see Sec. IV B)

From (4.2) for the solutions with $j_0=0$ we have

$$J = n_0(c-a)\nu_{\rm I}^{-1}\,{\rm Im}\,\Delta^{-1} = \frac{-n_0(c-a)}{\nu_{\rm I}|\Delta|^2}\, e^{\rho_R t}\,{\rm Im}(de^{i\rho_{\rm I} t}), \quad N\to n_0. \tag{5.3}$$

We recall that $\rho_R>0$, $\gamma_R>0$, and $|J| = 0(\exp-\rho_R t)$. However, when t increases, J changes sign an increasing number of time. So S_0 acting alternatively like a source or like a sink has dropped oscillations and still becomes negligible compared with the elastic wall: $|S_0/\eta S| \simeq 0(\exp-\rho_R t)$.

D. Solitons on a semiaxis $x\geqslant 0$ for $a\neq 0$ $c\neq 0$ (see Sec. II A)

For the $j_0=0$ solution, we recall that $n_0>0$, $\rho>0$, $(a+c)/2<\nu<c$. $N(x=0,t)\to n_0$ when $t\to\infty$, $J=n_0(c-a)/2\nu\Delta$ having the same sign as ν and $|J|\simeq 0(\exp-\rho t)$. So S_0 can be seen as a source ($\nu>0$) or a sink ($\nu<0$) which becomes negligible $|S_0/\eta S| \simeq 0(\exp-\rho t)$ when $t\to\infty$.

For the $j_0>0$ solution, we recall that $\rho<0$, $\sup(0,(a+c)/2)<\nu<c$, $J=j_0 w/\Delta>0$, $|J|=0(\exp\rho t)$, $N\to 2j_0\nu(c-a)$. In this case S_0 can be viewed as a source which becomes negligible $|S_0/\eta S| \simeq 0(\exp \rho t)$.

Let us define the average flow velocity, $\langle V\rangle = J(x,t)/N(x,t)$; we remark that for the four above cases studied in Sec. V A–V D we have $|\langle V\rangle| \simeq 0(\exp-{\rm const}\, t)\to 0$ when $t\to\infty$.

VI. NUMERICAL CALCULATIONS FOR BISOLITONS

In Fig. 1 we quote the relaxation curves for distributions $N\pm J>0$ on a semiline $x\geqslant 0$. In Fig. 1(a) we present a model $j_0=4$, γ_i and ρ_i reals, for the values $a=-3$, $c=1$ of the Illner model with $\nu=0.5$, $d_1=0.5$, $d_2=2$. In Fig. 1(b) the γ_i, ρ_i are complex conjugate and $j_0=0$. We choose the McKean model $a=0$, $c=1$ with $n_0=1$, $\nu_{\rm I}=0.5$, and $d=4+i1.5$. In Fig. 1(c), the γ_i, ρ_i are real but $j=0$. The Illner parameters are $a=-4$, $c=1$ while $n_0=1$, $\nu=0.5$, $d_1=d_2=1.5$.

In Figs. 2–4 we quote the relaxation curves for periodic solutions (damped sound waves) with the period spatial variable $x'=\gamma_{\rm I} x/2\in[0,1]$, where $c=1$ is fixed and $-a$ increasing, -9, -225, -625. We plot different relaxation curves for $N-J$ in (a), $N+J$ in (b), N in (c). In (d), the spatial variable $x=0$ being fixed we plot different oscillations of N $(x=0,t)$ which are effective. When the ratio $\rho_R/\rho_{\rm I}$ decreases 0.75, 0.13, 0.08, the damping becomes less important, allowing the possibility to observe an increasing number of oscillations.

VII. CONCLUSIONS

Two different one spatial dimensional discrete velocity models, the two-velocity Illner model and the six-velocity

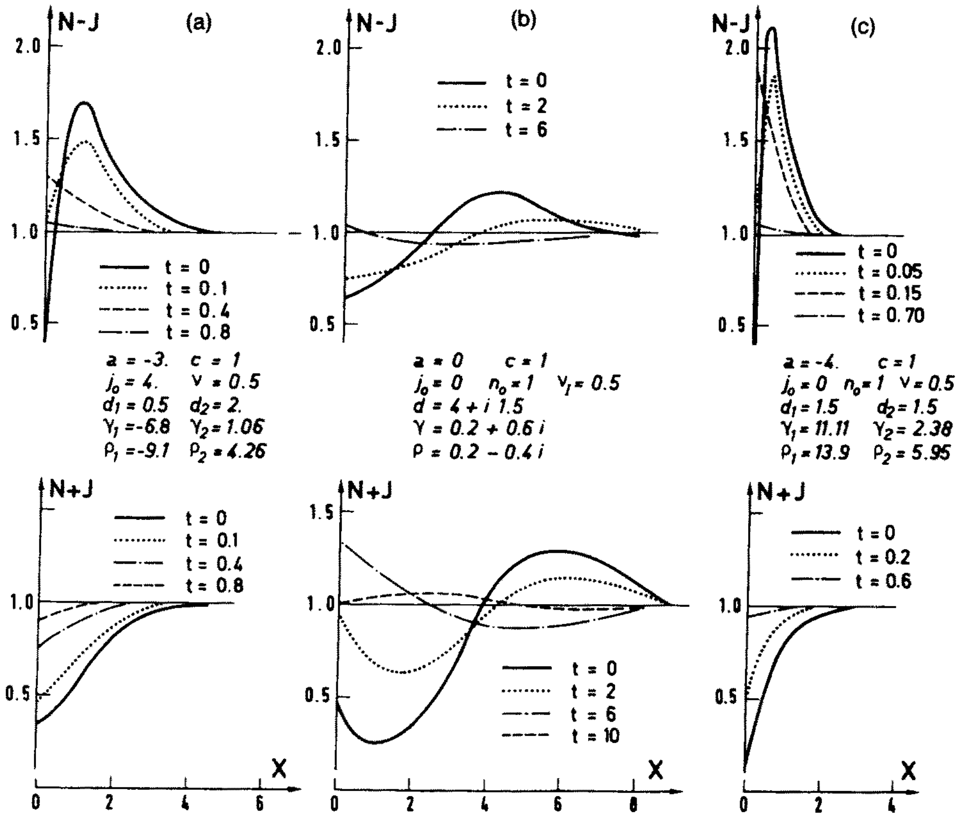

FIG. 1. Plots of $N\pm J$ against $x\in[0,\infty]$ for different t values (a) $a=-3$, $c=1$, $j_0=4$, $\nu=0.5$, $d_1=0.5$, $d_2=2$, $\gamma_1=-6.8$, $\gamma_2=1.06$, $\rho_1=-9.1$, $\rho_2=4.26$; (b) $a=0$, $c=1$, $j_0=0$, $n_0=1$, $\nu_{\rm I}=0.5$, $d=4+i1.5$, $\gamma=0.2+i0.6$, $\rho=0.2-i0.4$; (c) $a=-4$, $c=1$, $j_0=0$, $n_0=1$, $\nu=0.5$, $d_1=d_2=1.5$, $\rho_1=13.9$, $\rho_2=5.25$, $\gamma_1=11.11$, $\gamma_2=2.28$.

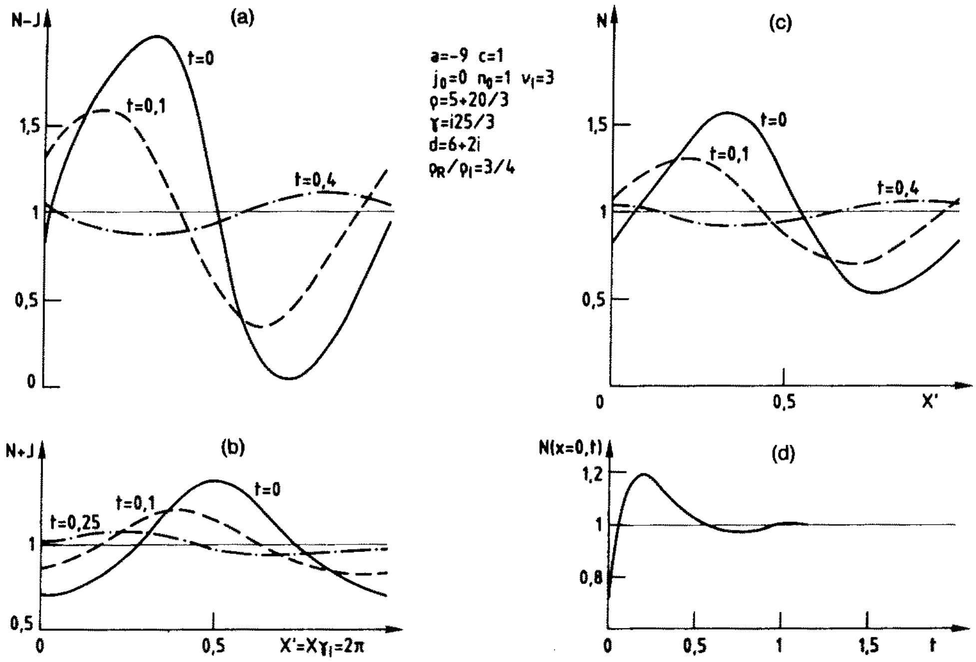

FIG. 2. Plots of the periodic solutions against $x' = \gamma_I x(2\pi)^{-1} \in [0.1]$, for the Illner model $c = 1$, $j_0 = 0$, $n_0 = 1$, $\nu_I = \sqrt{-a}$ and different $-a$ values. (a), (b), (c) $N - J$, $N + J$, N vs x' for different t values, (d) N, $a = -9$, $\rho = 5 + i20/3$, $\gamma = i25/3$, $d = 6 + 2i$, $\rho_R \rho_I^{-1} = 0.75$.

Broadwell model, have been studied with the same method. We have found two types of exact two-dimensional positive solutions: either periodic or half-space solutions. In addition, for the Broadwell model, there also exists positive exact nonperiodic solutions on the full x axis. It could be interesting to investigate more complex discrete models and see whether this method of solitons and bisolitons leads to similar classes of physically acceptable exact solutions.

For the Broadwell model, there also exist models in two and three spatial dimensions. I am investigating the two spatial one and, although the algebraic resolution is very tedious, I hope to find acceptable physical solutions.

Another interesting by-product of the determination of exact solutions is the possibility to check the fluid dynamical limit with the introduction of the mean free path into the collision term.

ACKNOWLEDGMENTS

It is a great pleasure to thank Pr. R. Balian for his great help in the physical interpretation of the solutions.

APPENDIX: POSSIBLE ANSATZ BISOLITONS

We study the class of possible bisolitons (rational functions)

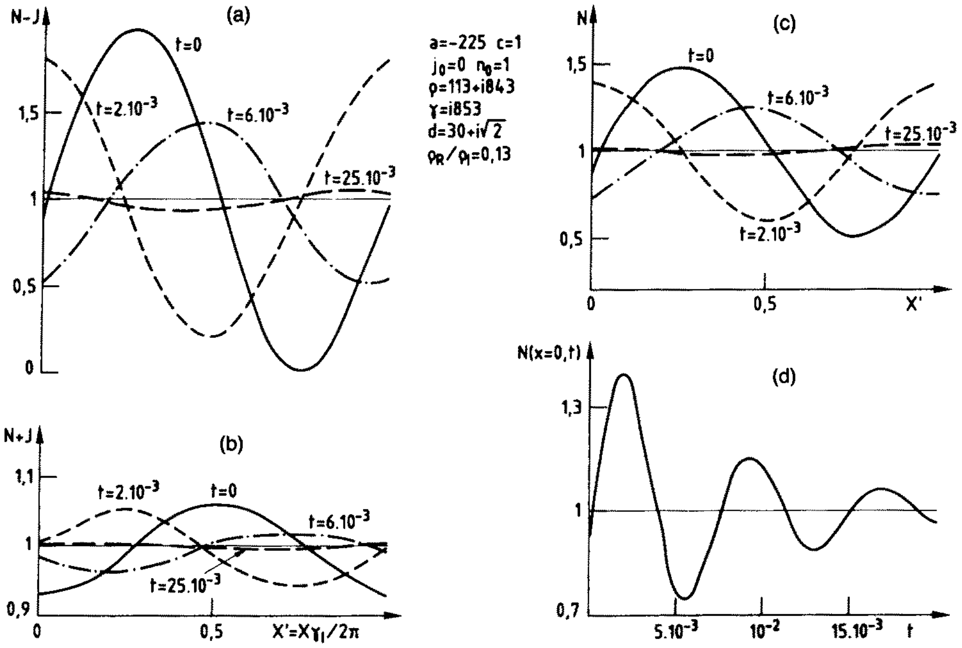

FIG. 3. Plots of the periodic solutions against $x' = \gamma_I x(2\pi)^{-1} \in [0.1]$, for the Illner model $c = 1$, $j_0 = 0$, $n_0 = 1$, $\nu_I = \sqrt{-a}$ and different $-a$ values. (a), (b), (c) $N - J$, $N + J$, N vs x' for different t values, (d) N, for $x = 0$ fixed, against t. $a = -225$, $\rho = 113 + i843$, $\gamma = i853$, $d = 30 + i\sqrt{2}$, $\rho_R \rho_I^{-1} = 0.13$.

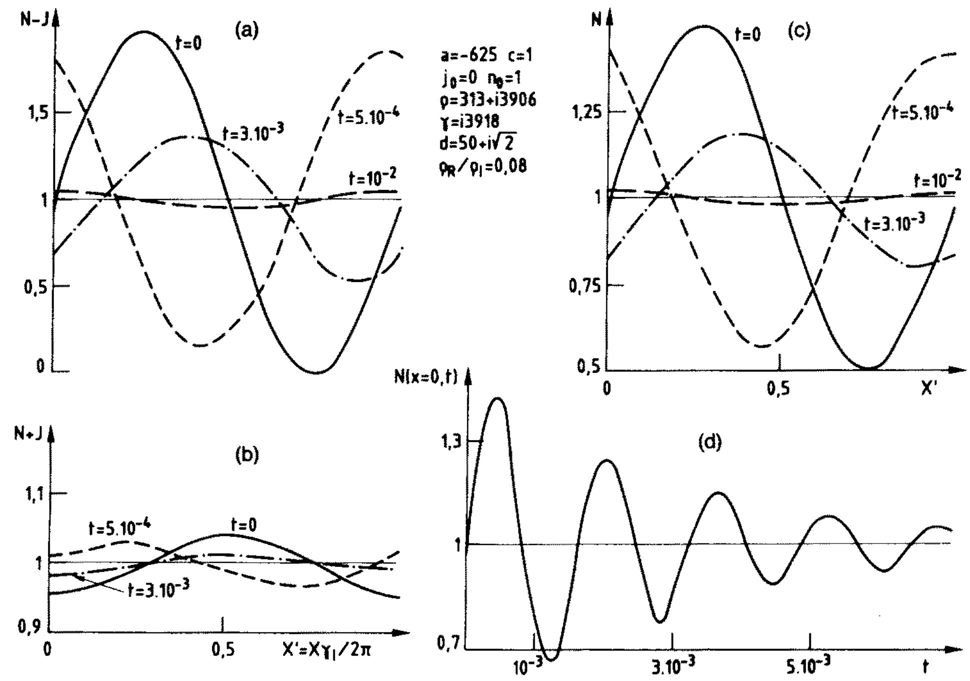

FIG. 4. Plots of the periodic solutions against $x' = \gamma_1 x(2\pi)^{-1} \in [0.1]$, for the Illner model $c = 1$, $j_0 = 0$, $n_0 = 1$, $\nu_1 = \sqrt{-a}$ and different $-a$ values. (a), (b), (c) $N-J$, $N+J$, N vs x' for different t values, (d) N, for $x = 0$ fixed, against t. $a = -625$, $\rho = 313 + i3906$, $\gamma = i3918$, $d = 50 + i\sqrt{2}$, $\rho_R \rho_I^{-1} = 0.08$.

$$N = n_0 + \frac{n}{\Delta}, \quad J = j_0 + \frac{j}{\Delta}, \quad 1 + \sum w_i + \mu w_1 w_2 = \Delta, \tag{A1}$$

$$n = n_{00} + \sum n_i w_i, \quad J = j_{00} + \sum j_i w_i,$$

$$w_i = d_i \exp(\gamma_i x + \rho_i t),$$

solutions of the Illner system

$$N_t + J_x = 0, \quad N_x + J_t = J[(a+c)J + N(a-c)],$$
$$a < 0, \quad c > 0, \tag{A2}$$

$n_0, j_0, n_{00}, j_{00}, n_i, j_i, \gamma_i, \rho_i$ being constants. In a two-dimensional space we must require

$$\gamma_1 \rho_2 - \gamma_2 \rho_1 \neq 0. \tag{A3}$$

We assume $\mu \neq 0$ in Sec. 1, $\mu = 0$ in Sec. 2. We investigate the full constraints coming from the linear differential equation (A2) and for the nonlinear one, we retain only that the term proportional to Δ^{-2} must factorize Δ,

$$j[(a+c)j + (a-c)n] + n\Delta_x + j\Delta_t = 0(\Delta). \tag{A2'}$$

We find that $\mu \neq 0$ requires $\mu = 1$ while $\mu = 0$ is not possible.

1. $\mu \neq 0$

1.1: $N_t + J_x = 0$ *leads to the identity* $(n_t + j_x)\Delta - (n\Delta_t + j\Delta_x) = 0$. We prescribe that the coefficients of w_i, w_i^2, $w_1 w_2$, $w_i^2 w_j$ into this identity are zero. We obtain

$$n = \sum n_i (1 + w_i), \quad j = -\sum \frac{n_i (1 + w_i)\rho_k}{\gamma_k}, \quad k \neq i, \tag{A4}$$

and two possibilities, either $\mu = 1$ or $\mu \neq 1$ and $n_1 \gamma_1 - n_2 \gamma_2 = 0$. We look at the possibility $\mu \neq 1$, so that we can rewrite (A4),

$$n = \bar{n}_1 \Big[\sum (\gamma_i + w_i \gamma_k) \Big], \quad k = -\bar{n}_1 \Big[\sum (\rho_i + w_i \rho_k) \Big],$$

$$\bar{n}_1 = n_1/\gamma_2 \tag{A3'}$$

and substituting into (A2') we want to know whether the assumption $\mu \neq 1$ is still possible.

1.2: Nonlinear constraints. Equation (A12') can be rewritten

$$j((a+c)j + n(a-c)) + n\Delta_x + j\Delta_t \equiv \Big(b_0 + \sum b_i w_i\Big)\Delta \bar{n}_1,$$

$$\bar{n}_1 = n_1/\gamma_2, \tag{A2''}$$

with b_0, b_i unknown to be determined. We prescribe that in (A2'') the coefficients of const, w_i, $w_1 w_2$, w_i^2, $w_i^2 w_j$ are zero. We obtain b_0, b_i: $b_0 = \bar{n}_1 \Sigma \rho_i ((a+c)\Sigma \rho_i - (a-c)\Sigma \gamma_i)$, $b_i = \gamma_j(\gamma_1 + \gamma_2) - \rho_j(\rho_1 + \rho_2)$ and three relations where we define new variables $x_i = \rho_i/\gamma_i$ [(A3) means $x_1 \neq x_2$],

$$\bar{n}_1 = (x_i^2 - 1)/(a + c + x_I(c-a)), \quad i = 1,2,$$

$$(\mu - 1)(2(1 - x_1 x_2) + \bar{n}_1(2(a+c) + (c-a)(x_1 x_2))) = 0. \tag{A5}$$

We discard the $\mu = 1$ case and the coefficient of $(\mu - 1)$ gives another expression for $\bar{n}_1$. Still assuming $x_1 \neq x_2$ we find

$$(a+c)x_i + (c-a)(1 + x_i^2)/2 = 0,$$
$$(a+c)(x_1 + x_2) + (a-c)(1 + x_1 x_2) = 0.$$

If $a + c \neq 0$ the only possibility is $x_1 = x_2 = \pm 1$, while if $a + c = 0$ we have $x_1 = x_2 = \pm i$. Finally, (A3) cannot be satisfied for $\mu \neq 1$.

2. $\mu = 0$

Without loss of generality we redefine the constants in (A1),

$$N = n_0 + \frac{n}{\Delta}, \quad J = j_0 + \frac{j}{\Delta}, \quad \Delta = 1 + \sum w_i,$$
$$n = n_{00} + n_1 w_1, \quad j = j_{00} + j_2 w_2. \tag{A1'}$$

2.1: $N_t + J_x = 0$. From the vanishing of the coefficients of w_i, w_i^2, $w_1 w_2$, $w_i^2 w_j$ we find

$$n=\bar{n}_1(w_1(\gamma_2-\gamma_1)+\gamma_2),\quad j=\bar{n}_1(w_2(\rho_2-\rho_1)-\rho_1),$$
$$\bar{n}_1=n_1/(\gamma_2-\gamma_1). \tag{A6}$$

2.2: Nonlinear constraint. We use (A2″), $\bar{n}_1$ being defined in (A6). The coefficients of const, w_i, w_1w_2, w_i^2 determine b_0, b_i and give three expressions for $\bar{n}_i$,

$$b_0=\bar{n}_1\rho_1((a+c)\rho_1-(a-c)\gamma_2),\quad b_1=\gamma_1(\gamma_2-\gamma_1),$$
$$b_2=\bar{n}_1(a+c)(\rho_2-\rho_1)^2+\rho_2(\rho_2-\rho_1), \tag{A7}$$

$$\bar{n}_1=\frac{(\rho_2-\rho_1)^2-(\gamma_1-\gamma_2)^2}{(\rho_2-\rho_1)((a-c)(\gamma_2-\gamma_1)-(a+c)(\rho_2-\rho_1))}$$
$$=\frac{\rho_i^2-\gamma_i^2}{\rho_i((a-c)\gamma_i-(a+c)\rho_i)},\quad i=1,2. \tag{A8}$$

Introducing $x_i=\rho_i/\gamma_i$ and first assuming $x_i^2\neq 1$, we find either $x_1=x_2$ or

$$\begin{aligned}&(a+c)(x_1+x_2)+(c-a)(1+x_1x_2)=0,\\&2x_i(a+c)+(c-a)(1+x_i^2)=0.\end{aligned} \tag{A9}$$

If $a+c\neq 0$ the only possibility is $x_i^2=1$ while if $a+c=0$ we have $x_1=x_2=\pm i$. Second, if both $x_i^2=1$, the denominators of the last Eq. (A8) vanish leading to $a=c=0$. [If only $x_2^2=1$ the last $i=2$ relation in (A8) gives $x_2=(a-c)/(a+c)$ while the other relations lead to $x_1^2=1$.] Finally, (A3), for the Illner model, cannot be satisfied for $\mu=0$.

2.3: Instead of the Illner model we consider the most general quadratic nonlinearity.[14] This means $\pm(aN_+^2+2bN_+N_-+cN_-^2)$ in (1.1) or $N_t+N_x=AN^2+BJ^2+CNJ$ in (1.2) with $A=b+(a+c)/2$, $B=-b+(a+c)/2$, $C=a-c$. (The Illner model corresponds to $2b+a+c=0$, $a<0$, $c>0$.) We want to show that if we restrict our study, as here, to nonlinear Δ^{-2} terms, the only possible $\mu=0$ solutions require $a=c=0$, $b\neq 0$. For the ansatz (A1′), then (A6) is still valid. Equation (A2″) becomes $j^2B+n^2A+njC+n\Delta_x+j\Delta_t=(b_0+\Sigma b_iw_i)\Delta\bar{n}_1$ and the vanishing of the wA_i powers, as above, give

$$\begin{aligned}&b_0=\bar{n}_1(B\rho_1^2+A\gamma_2^2-\rho_1\gamma_2C),\\&b_1=\gamma_1(\gamma_2-\gamma_1)+\bar{n}_1A(\gamma_1-\gamma_2)^2,\\&b_2=\rho_2(\rho_2-\rho_1)+\bar{n}_1B(\rho_2-\rho_1)^2,\end{aligned} \tag{A7′}$$

$$\bar{n}_1=\frac{-(\rho_2-\rho_1)^2+(\gamma_2-\gamma_1)^2}{(A(\gamma_2-\gamma_1)^2+B(\rho_2-\rho_1)^2-C(\rho_2-\rho_1)(\gamma_2-\gamma_1))}=\frac{-\rho_i^2+\gamma_i^2}{A\gamma_i^2+B\rho_i^2-C\gamma_i\rho_i}. \tag{A8′}$$

As above, introducing $x_i=\rho_i/\gamma_i$ and first assuming $x_i^2\neq 1$, we find either $x_1=x_2$ or

$$\begin{aligned}&(x_j+x_i)(A+B)-C(1+x_ix_j)=0,\\&-2x_i(A+B)+C(x_i^2+1)=0.\end{aligned} \tag{A9′}$$

We can check in the Illner model $A=0$, $B=a+c$, $C=a-c$ that (A7′)–(A9′) reduce to (A7′)–(A9). If the determinant of (A9′) is different than zero then $A+B=C=0$ or $a=c=0$ and the only possibility to satisfy (A3) is to have a term like N_+N_- in the collision term (without N_+^2 terms). The model of Ruijgrok and Wu[15] is of this type and there exist for this model solutions with $\mu=0$. If the determinant of (A9′) is zero, then either $x_i=x_j$ or $x_i^2=1$. Second, if both $x_i^2=1$, the denominators of the last (A8′) relations vanish leading to $A+B=Cx_i$ or $x_1=x_2$. [If only $x_2^2=1$, the last (A8′) relation gives $x_2C=A+B$, and substituting into the other (A8′) relations we find $x_1^2=1$.]

In conclusion only quadratic terms of the type N_+N_- could allow $\mu=0$ bisolitons.

[1] H. Cornille, submitted to J. Phys. A.

[2] J. E. Broadwell, Phys. Fluids 7, 1243 (1964).

[3] H. McKean, Israel J. Math. **21**, 54 (1975); Commun. Pure Appl. Math. **28**, 435 (1975).

[4] I. Kolodner, Ann. Math. Pure Appl. **63**, 11 (1963); R. Teman, Arch. Rational Mech. Anal. **35**, 351 (1969); R. Illner, Math. Methods Appl. Sci. **1**, 187 (1979); M. Shinbrot, "Oberwelfach Conference" (1979), edited by D. Pach and H. Neuzert, p. 11, Kaper, Leaf, and Reich, Math Methods Appl. Sci. **2**, 203 (1980); R. Illner and M. Reed, *ibid.* **3**, 121 (1981); L. C. Tartar, *Lecture Notes in Physics*, Vol. 195 (Springer, Berlin, 1984), p. 384; D. McLaughlin, G. C. Papanicolaou, and L. C. Tartar, *Lecture Notes in Physics*, Vol. 230 (Springer, Berlin, 1985), p. 277; L. C. Tartar, to appear in the *Proceedings of "Dynamics of Infinite Dimensional Systems,"* Lisboa, May 1986, with many references therein.

[5] H. Cabannes, "Oberwelfach Conference," 1979, edited by D. Pack and H. Neunzert, p. 25.

[6] R. Illner, Math Methods Appl. Sci. **1**, 187 (1979); Trans. Theory Stat. Phys. **13**, 431 (1984).

[7] G. Dukek and T. F. Nonnenmacher, Physica A **135**, 167 (1986). We notice that in this paper their solution $\rho=H^{-1}, j$ defined in Eqs. (13) and (14) and satisfying $\rho\geqslant 0$, $\rho-j\geqslant 0$ (solutions called here solitons) cannot stay positive for both limits $x\to\pm\infty$.

[8] I. Kolodner, Ann. Math. Pure Appl. **63**, 11 (1963).

[9] A. V. Bobylev, Math. Congress Warsaw, 1983 (Book Abstract B29); J. Wick, Math. Methods Appl. Sci. **6**, 515 (1984).

[10] T. Platkowski, J. Mec. Theor. Appl. **4**, 555 (1985). In this article, a result due to K. Piechor is quoted.

[11] K. Krook and T. T. Wu, Phys. Rev. Lett. **16**, 1107 (1976); H. Cornille and A. Gervois, J. Stat. Phys. **26**, 181 (1981).

[12] H. Cornille and A. Gervois, Physica D **6**, 29 (1982); H. Cornille, J. Math. Phys. **25**, 1335 (1984), with many references therein.

[13] H. Cornille, J. Phys. A **18**, 1209 (1985).

[14] The results of subsection 2.3 of the Appendix have been obtained in collaboration with T. T. Wu.

[15] Th. W. Ruijgrok and T. T. Wu, Physica A **113**, 401 (1982).

The effect of a one-dimensional potential of finite range on the statistical parameters of an incident ensemble of particles. I. Pure states

J. E. G. Farina
Department of Mathematics, University of Nottingham, Nottingham NG7 2RD, England

(Received 23 August 1986; accepted for publication 25 February 1987)

If an ensemble of particles moving parallel to the x axis in the positive direction impinges on a piecewise continuous potential confined to the interval $[-a,a]$ $(a>0)$ it will divide into a transmitted ensemble and a reflected ensemble. It is shown that the classical results for the means, variances, and covariance of position and velocity of the transmitted and reflected ensembles hold in quantum mechanics if the incident state is assumed to be pure and defined by a C_∞ wave function whose Fourier transform has bounded, positive support, on which the modulus and argument of the transmission and reflection coefficients are C_∞.

I. INTRODUCTION

We shall consider the problem of an ensemble of particles initially moving freely parallel to the x axis in the positive direction, which subsequently impinges on a piecewise continuous potential V which vanishes outside some finite interval $[-a,a]$ $(a>0)$. If the particles obey CM (classical mechanics) the velocity $v_{\rm in}(t)$ and position $x_{\rm in}(t)$ of a particle of the ensemble at time $t\sim-\infty$ are given by

$$v_{\rm in}(t)=v, \tag{1.1a}$$

$$x_{\rm in}(t)=x_{\rm in}+vt, \tag{1.1b}$$

where v, $x_{\rm in}$ are real constants and $v>0$. An observable A is a (possibly time-dependent) function on the two-dimensional phase space Ω, and its expectation value $\langle A\rangle_{\rm in}$ is obtained by averaging over some time-independent probability density $\mu_{\rm in}$ on Ω. The following results for the incident ensemble follow easily from the fact that expectation is a linear functional:

$$\text{mean velocity}=\langle v_{\rm in}(t)\rangle_{\rm in}=\langle v\rangle_{\rm in}, \tag{1.2a}$$

$$\text{mean position}=\langle x_{\rm in}(t)\rangle_{\rm in}=\langle x_{\rm in}\rangle_{\rm in}+\langle v\rangle_{\rm in}t. \tag{1.2b}$$

If we write $\mathrm{Var}_{\rm in}\,A$ for the variance $\langle(A-\langle A\rangle_{\rm in})^2\rangle_{\rm in}$, then

$$\text{variance of velocity}=\mathrm{Var}_{\rm in}\,v_{\rm in}(t)=\mathrm{Var}_{\rm in}\,v. \tag{1.2c}$$

If we write $\mathrm{Cov}_{\rm in}(A,B)$ for the covariance $\langle(A-\langle A\rangle_{\rm in})(B-\langle B\rangle_{\rm in})\rangle_{\rm in}$, then

covariance of position and velocity

$$=\mathrm{Cov}_{\rm in}(x_{\rm in}(t),v_{\rm in}(t))$$

$$=\mathrm{Cov}_{\rm in}(x_{\rm in},v)+t\,\mathrm{Var}_{\rm in}\,v, \tag{1.2d}$$

variance of position of the incident ensemble

$$=\mathrm{Var}_{\rm in}\,x_{\rm in}(t)$$

$$=\mathrm{Var}_{\rm in}\,x_{\rm in}+2t\,\mathrm{Cov}_{\rm in}(x_{\rm in},v)+t^2\,\mathrm{Var}_{\rm in}\,v. \tag{1.2e}$$

Suppose now the particles obey QM (quantum mechanics). An observable A is now a self-adjoint operator in the Hilbert space $\mathscr{H}=L_2(\mathbb{R})$. In particular, if f is a real-valued function on $\mathbb{R}$, multiplication by f is self-adjoint; we shall denote this operator by $\hat f$. So also is $F^*\hat fF$, where F is the Fourier transformation; we shall denote $F^*\hat fF$ by $\tilde f$. In particular, the position operator is $\hat x$. If k is the wave number, m is the mass, $v=\hbar k/m$ is the velocity, and $\tilde v=F^*\hat vF$ is the velocity operator, then k is the momentum in units of $\hbar$.

To obtain the results in QM corresponding to the results (1.1) in CM, we use the Heisenberg picture. The velocity and position observables at time $t\sim-\infty$ are then given by

$$v_{\rm in}(t)=\tilde v, \tag{1.3a}$$

$$x_{\rm in}(t)=x_{\rm in}+\tilde vt, \tag{1.3b}$$

respectively, where $x_{\rm in}=\hat x$. The results (1.2) (with v replaced by $\tilde v$ on the right-hand sides) for the statistical parameters of the incident ensemble can be shown to be valid in QM, provided $\mathrm{Var}_{\rm in}\,\hat x$ and $\mathrm{Var}_{\rm in}\,\tilde v$ exist.[1]

We shall obtain the formulas corresponding to (1.2) when $t\to+\infty$ for the transmitted and reflected particles. First we do this for the CM case, and in subsequent sections for the QM case. For simplicity we shall assume that, in the QM case, the statistical state of the incident ensemble is a pure state described by a C_∞ wave function $\psi_{\rm in}$ whose Fourier transform $F\psi_{\rm in}$ has compact support in the positive interval $\{k\in\mathbb{R}\colon k>0\}$; thus $\langle A\rangle_{\rm in}=\langle\psi_{\rm in}|A|\psi_{\rm in}\rangle$. In a subsequent paper we shall show that this assumption is not so restrictive as it appears.

The establishment of (1.3) and (1.2) is straightforward given the assumption that $\psi_{\rm in}$ is C_∞—in fact, a member of the set $\mathscr{S}$ of Schwarz testing functions. For now by elementary Fourier analysis $\tilde v=-i\hbar D/m$ on $\mathscr{S}$, where D is the differentiation operator. Also the free evolution operator $U_t=F^*\exp(-i\hat\omega t)F$, where $\exp(-i\hat\omega t)$ is multiplication by $\exp(-i\omega t)$ $(\omega=\hbar k^2/2m)$. Thus

$$v(t)=U_t^*\tilde vU_t=\tilde v,$$

which is (1.1a), while

$$x(t)=U_t^*\hat xU_t=F^*\exp(i\hat\omega t)iD\exp(-i\hat\omega t)F$$

$$=F^*iDF+F^*\hat vtF=\hat x+\tilde vt,$$

which is (1.1b). Equations (1.2) now follow from the linearity of $\langle\cdot\rangle_{\rm in}$, if the covariance $\mathrm{Cov}_{\rm in}(A,B)$ of two (possibly time-dependent) observables A,B of the incident ensemble is defined by

$$\langle(A-\langle A\rangle_{\rm in})\circ(B-\langle B\rangle_{\rm in})\rangle_{\rm in},$$

where $C\circ D$ denotes the symmetric product $\frac12(CD+DC)$ of two operators C and D.

 0022-2488/87/071580-03$02.50

II. CLASSICAL RESULTS

We are assuming that the potential is piecewise continuous, hence bounded above. It follows that there is a threshold velocity v_{thresh} such that if the velocity v of a particle in the incident ensemble exceeds v_{thresh} it will be transmitted, while if $v < v_{\text{thresh}}$ it will be reflected. If no particle initially has the exact velocity v_{thresh} the incident ensemble divides into two subensembles, one of which is transmitted, the other reflected.

Consider a particle with $v > v_{\text{thresh}}$. When $t \sim -\infty$ its orbit in Ω is described by (1.1). When $t \sim +\infty$ the particle has been transmitted. Its position $x_{\text{tr}}(t)$ and velocity $v_{\text{tr}}(t)$ at time $t \sim +\infty$ after transmission are now given by

$$v_{\text{tr}}(t) = v, \tag{2.1a}$$

$$x_{\text{tr}}(t) = x_{\text{tr}} + vt, \tag{2.1b}$$

where x_{tr} is a *shifted* time zero position. In fact, if t_{tr} is the time delay due to transmission—that is, the difference between the times spent in the interval $[-a,a]$ with, and without, the potential—then

$$x_{\text{tr}} = x_{\text{in}} - vt_{\text{tr}}, \tag{2.2}$$

where t_{tr} is a function of v.

If A is a classical observable, then its expectation value $\langle A\rangle_{\text{tr}}$ over the transmitted particles is obtained by averaging over μ_{tr}, the probability density on Ω of the transmitted particles. Here μ_{tr} is obtained by multiplying μ_{in} by the probability of transmission, which equals unity if $v > v_{\text{thresh}}$ and zero if $v < v_{\text{thresh}}$, and then normalizing. The results corresponding to (1.2) are easily derived from (2.1) and the linearity of expectation. They may be written down from (1.2) by replacing $v_{\text{in}}(t)$ by $v_{\text{tr}}(t)$, $x_{\text{in}}(t)$ by $x_{\text{tr}}(t)$, x_{in} by x_{tr}, $\langle\cdot\rangle_{\text{in}}$ by $\langle\cdot\rangle_{\text{tr}}$, Var_{in} by Var_{tr}, where $\text{Var}_{\text{tr}}\, A = \langle(A - \langle A\rangle_{\text{tr}})^2\rangle_{\text{tr}}$, and Cov_{in} by Cov_{tr}, where $\text{Cov}_{\text{tr}}(A,B) = \langle(A - \langle A\rangle_{\text{tr}})(B - \langle B\rangle_{\text{tr}})\rangle_{\text{tr}}$.

If $v < v_{\text{thresh}}$ the particle is reflected. For $t \sim +\infty$, when reflection is complete, its velocity and position are, respectively, given by

$$v_{\text{re}}(t) = -v, \tag{2.3a}$$

$$x_{\text{re}}(t) = x_{\text{re}} - vt, \tag{2.3b}$$

where

$$x_{\text{re}} = -x_{\text{in}} + vt_{\text{re}}, \tag{2.4}$$

t_{re} being the reflection delay time—that is, the difference between the time spent by the particle in $[-a,a]$ with, or without, the potential. The results for the reflected particles for the statistical parameters are obtained from (1.2) by replacing $v_{\text{in}}(t)$ by $v_{\text{re}}(t)$, $x_{\text{in}}(t)$ by $x_{\text{re}}(t)$, v by $-v$, $\langle\cdot\rangle_{\text{in}}$ by $\langle\cdot\rangle_{\text{re}}$, x_{in} by x_{re}, Var_{in} by Var_{re}, and Cov_{in} by Cov_{re}, where $\langle A\rangle_{\text{re}}$ is the expectation value of A over the probability density μ_{re} of the reflected particles, etc. Here μ_{re} is obtained from μ_{in} by multiplication by the probability of reflection and normalizing.

III. QUANTUM TRANSMISSION

Let τ be the transmission coefficient, a function of the wave number k. The unnormalized transmitted wave packet is $F^*\hat{\tau}F\psi_{\text{in}} = \tilde{\tau}\psi_{\text{in}}$ (using the notation introduced in Sec. I). The probability of transmission is $w_{\text{tr}} = \|\tilde{\tau}\psi_{\text{in}}\|^2$. A freely evolving observable is represented by the self-adjoint operator $A(t) = U_t^* A U_t$, where U_t is the evolution operator for free motion. The expectation value of $A(t)$ for the transmitted particles (when $t \sim +\infty$) is

$$w_{\text{tr}}^{-1}\langle\tilde{\tau}\psi_{\text{in}}|A(t)|\tilde{\tau}\psi_{\text{in}}\rangle. \tag{3.1}$$

Let $S_{\text{tr}} = F^* \exp(i \arg \hat{\tau})F$ be the unitary operator of multiplication by $\exp(i \arg \tau)$ in the k representation, and $|\tilde{\tau}| = F^*|\hat{\tau}|F$ be the operator of multiplication by $|\tau|$ in the k representation. Then $\tilde{\tau} = S_{\text{tr}}|\tilde{\tau}|$ and so (3.1) can be written

$$\langle A_{\text{tr}}(t)\rangle_{\text{tr}} \equiv \langle\psi_{\text{tr}}|A_{\text{tr}}(t)|\psi_{\text{tr}}\rangle, \tag{3.2}$$

where

$$A_{\text{tr}}(t) = S_{\text{tr}}^* A(t) S_{\text{tr}} \tag{3.3}$$

and ψ_{tr} is the normalized wave function defined by

$$\psi_{\text{tr}} = w_{\text{tr}}^{-1/2}|\tilde{\tau}|\psi_{\text{in}}. \tag{3.4}$$

Here S_{tr} produces a unitary transformation of the space of observables, mapping $A(t)$ into $A_{\text{tr}}(t)$ according to (3.3). For reasons which should become clear in the sequel we shall call it *the transmission shift operator.*

The shifted velocity at time t is $v_{\text{tr}}(t)$, where, by (3.3),

$$v_{\text{tr}}(t) = S_{\text{tr}}^* v_{\text{in}}(t) S_{\text{tr}}.$$

Since $v_{\text{in}}(t) = \tilde{v}$ and S_{tr}, $\tilde{v}$ commute, we obtain

$$v_{\text{tr}}(t) = \tilde{v}. \tag{3.5a}$$

Equation (3.5a) is the QM analog of (2.1a)—the velocity observable is unaffected by transmission.

The shifted position operator at time t is $x_{\text{tr}}(t)$, where, by (3.3),

$$x_{\text{tr}}(t) = S_{\text{tr}}^* x_{\text{in}}(t) S_{\text{tr}}.$$

Using (1.3b) we get

$$x_{\text{tr}}(t) = x_{\text{tr}} + \tilde{v}t, \tag{3.5b}$$

where

$$x_{\text{tr}} = S_{\text{tr}}^* x_{\text{in}} S_{\text{tr}}. \tag{3.6}$$

Equation (3.5b) is formally analogous to the classical result (2.1b). In order to show that it is indeed the QM analog we must show that x_{tr} is given by (2.2) with t_{tr} the transmission delay time. We can do this very easily if we make the following assumption: $|\tau|$ and $\arg \tau$ are C_∞ on the support of $F\psi_{\text{in}}$. For this assumption means that we can express $x_{\text{in}} = \hat{x}$ as F^*iDF, and then by (3.6)

$$x_{\text{tr}} = S_{\text{tr}}^* F^* iDF S_{\text{tr}}.$$

Since $S_{\text{tr}} = F^* \exp(i \arg \hat{\tau}) F$, where $\exp(i \arg \hat{\tau})$ is the operator of multiplication by $\exp(i \arg \tau)$, we easily obtain

$$x_{\text{tr}} = x_{\text{in}} - \tilde{v}\tilde{t}_{\text{tr}}, \tag{3.7}$$

where

$$t_{\text{tr}} = v^{-1} D \arg \tau. \tag{3.8}$$

[By hypothesis the support of $F\psi_{\text{in}}$ consists only of positive numbers, and so (3.8) is well defined.]

The quantity t_{tr} is none other than the well-known Eisenbud–Wigner delay time. In nonrigorous language it is the time delay of the peak of the transmitted wave packet relative to the peak of the incident wave packet.[2] Thus (3.7) is indeed the QM analog of (2.2).

The results (1.2) now follow for the transmitted particles in QM, on using (3.5) and the linearity of $\langle\cdot\rangle_{tr}$; in them $x_{in}(t)$ and $v_{in}(t)$ are replaced by $x_{tr}(t)$ and $v_{tr}(t)$, respectively, x_{in} by x_{tr}, $\langle\cdot\rangle_{in}$ by $\langle\cdot\rangle_{tr}$, Var_{in} by Var_{tr}, Cov_{in} by Cov_{tr}, and v by $\tilde{v}$.

In the CM case discussed in Sec. II the probability density μ_{tr} used in calculating $\langle\cdot\rangle_{tr}$ was obtained from μ_{in} by multiplying by the probability of transmission and normalizing. In the QM case discussed in this section $\langle\cdot\rangle_{tr}$ is calculated from a wave function ψ_{tr} obtained by multiplying the momentum amplitude $F\psi_{in}$ by the square root $|\tau|$ of the probability of transmission and normalizing.

IV. QUANTUM REFLECTION

Let ρ be the reflection coefficient, again a function of k. The unnormalized reflected wave packet is $F^*P\hat{\rho}F\psi_{in}$, where P is the parity operator. Since P commutes with F and F^* this can be written $P\tilde{\rho}\psi_{in}$. The probability of reflection is $w_{re} = \|P\tilde{\rho}\psi_{in}\|^2 = \|\tilde{\rho}\psi_{in}\|^2$. The expectation value for the reflected particles of $A(t)$ is

$$w_{re}^{-1}\langle P\tilde{\rho}\psi_{in}|A(t)|P\tilde{\rho}\psi_{in}\rangle. \tag{4.1}$$

Define the unitary operator S_{re} by

$$S_{re} = F^*\exp(i\arg\hat{\rho})F, \tag{4.2}$$

where $\exp(i\arg\hat{\rho})$ is the operator of multiplication by $\exp(i\arg\rho)$, and let $|\tilde{\rho}| = F^*|\hat{\rho}|F$ in the usual notation. Then $\tilde{\rho} = S_{re}|\tilde{\rho}|$, and so (4.1) can be written

$$\langle A_{re}(t)\rangle_{re} \equiv \langle\psi_{re}|A_{re}(t)|\psi_{re}\rangle, \tag{4.3}$$

where

$$A_{re}(t) = S_{re}^*PA(t)PS_{re} \tag{4.4}$$

and ψ_{re} is the normalized wave function defined by

$$\psi_{re} = w_{re}^{-1/2}|\tilde{\rho}|\psi_{in}. \tag{4.5}$$

Here (4.4) shows that $A_{re}(t)$ is obtained from $A(t)$ by a reflection followed by a unitary transformation. Again we shall show that the physical meaning of this unitary transformation is a shift, and so S_{re} may be called the *reflection shift operator*.

The velocity observable becomes

$$v_{re}(t) = S_{re}^*Pv_{in}(t)PS_{re},$$

which reduces to

$$v_{re}(t) = -\tilde{v}. \tag{4.6a}$$

Equation (4.6a) is the QM analog of (2.3a)—reflection simply reverses the sign of the velocity.

Similarly the position observable becomes

$$x_{re}(t) = S_{re}^*Px_{in}(t)PS_{re}$$

and use of (1.3b) reduces this to

$$x_{re}(t) = x_{re} - \tilde{v}t, \tag{4.6b}$$

where

$$x_{re} = S_{re}^*Px_{in}PS_{re}. \tag{4.7}$$

As with transmission (4.6b) will be the QM analog of the CM result (2.3b) provided the QM analog of (2.4) can be proved. We now assume $|\rho|$ and $\arg\rho$ are C_∞ on the support of $F\psi_{in}$. Then $x_{in} = \hat{x} = F^*iDF$, and since $Px_{in}P = P\hat{x}P = -\hat{x} = -x_{in}$, (4.7) yields

$$x_{re} = -S_{re}^*F^*iDFS_{re}.$$

If we use (4.2) this reduces to

$$x_{re} = -x_{in} + \tilde{v}\tilde{t}_{re}, \tag{4.8}$$

where

$$t_{re} = v^{-1}D\arg\rho. \tag{4.9}$$

(Again our assumption about the support of $F\psi_{in}$ means that t_{re} is well defined.) The quantity t_{re} is the Eisenbud–Wigner delay time for the reflected particle—that is, the time delay of the peak of the reflected wave packet relative to the peak of the incident wave packet. Thus (4.8) is the QM analog of (2.4).

Comparison of (4.6) with (1.1) shows that the results (1.2) are valid for the reflected particles obeying QM if $\langle\cdot\rangle_{in}$ becomes $\langle\cdot\rangle_{re}$, $v_{in}(t)$ and $x_{in}(t)$ become $v_{re}(t)$ and $x_{re}(t)$, respectively, x_{in} becomes x_{re}, v becomes $-\tilde{v}$, Var_{in} becomes Var_{re}, and Cov_{in} becomes Cov_{re}. Since $\mathrm{Var}_{re}(-\tilde{v}) = \mathrm{Var}_{re}\,\tilde{v}$ and Cov_{in} is linear in $\tilde{v}$, we obtain

mean velocity $= -\langle\tilde{v}\rangle_{re}$,

mean position $= \langle x_{re}\rangle_{re} - \langle\tilde{v}\rangle_{re}t$,

variance of velocity $= \mathrm{Var}_{re}\,\tilde{v}$,

covariance of position and velocity

$$= -\mathrm{Cov}_{re}(x_{re},\tilde{v}) + t\,\mathrm{Var}_{re}\,\tilde{v},$$

variance of position

$$= \mathrm{Var}_{re}\,x_{re} - 2t\,\mathrm{Cov}_{re}(x_{re},\tilde{v}) + t^2\,\mathrm{Var}_{re}\,\tilde{v}.$$

[1] J. E. G. Farina, Int. J. Theor. Phys. **21**, 83 (1982).
[2] E. P. Wigner, Phys. Rev. **98**, 145 (1955).

The effect of a one-dimensional potential of finite range on the statistical parameters of an incident ensemble of particles. II. Localized states

J. E. G. Farina
Department of Mathematics, University of Nottingham, Nottingham NG7 2RD, England

(Received 23 September 1986; accepted for publication 25 February 1987)

If an ensemble of particles moving parallel to the x axis in the positive direction impinges on a piecewise continuous potential confined to the interval $[-a,a]$ $(a>0)$ it will divide into a transmitted ensemble and a reflected ensemble. It is shown that the classical results for the means, variances, and covariance of position and velocity of the transmitted and reflected ensembles hold in quantum mechanics if the position and momentum probability densities of the incident ensemble are assumed to be sufficiently localized, both in momentum and position. It is also assumed that the support of the probability density of momentum is bounded and positive, and on it the moduli and arguments of the transmission and reflection coefficients are sufficiently well behaved.

I. INTRODUCTION

In the previous paper[1] (whose notation we shall use here, and to which we shall refer to as 'I') we discussed the effect on the statistical parameters of an ensemble of particles moving parallel to Ox of their passage across a potential V confined to the interval $[-a,a]$. We assumed there that the statistical state of the incident ensemble was pure. Such an assumption, however, cannot be justified for an ensemble occurring in nature. Such an ensemble would have the property that every subensemble would have the same statistics, and it is difficult to believe that is so. More importantly, we only have limited information about such ensembles—for example, the shape of the momentum probability density $p_{\text{mom,in}}$ may be known, or the region of space outside which the position probability density $p_{\text{pos,in}}$ is negligible. Infinitely many states will be consistent with our information, and not all of these will be pure. It therefore becomes important to replace the assumption that the state of the incident ensemble is pure by one which may be reasonably supposed to hold for an actual ensemble. The assumption which we will make is the very reasonable one that $|x|^\nu p_{\text{pos,in}}(x)$ and $|k|^\nu p_{\text{mom,in}}(k)$ ($k = mv/\hbar$ = momentum in units of $\hbar$) are bounded for all $\nu \geqslant 0$. (For example, if $p_{\text{pos,in}}$ and $p_{\text{mom,in}}$ were Gaussians this assumption would be satisfied.)

We also assumed that ψ_{in} was C_∞. We shall show that, although the state may not be described by a wave function (if it is not pure), nevertheless all wave functions occurring in the Gleason expression[2] of the state as a convex linear combination of pure states are C_∞—indeed, as we shall show, they are testing functions in the sense of Schwarz.

In I an assumption was made about the support of $F\psi_{\text{in}}$, viz. that it was bounded and positive. We shall make the same assumption about the support of $p_{\text{mom,in}}$. That is, we shall assume that Supp $p_{\text{mom,in}}$ is contained in some interval $[k_1,k_2]$ $(0<k_1<k_2<\infty)$.

This last assumption is quite natural. For the particles of the incident ensemble are moving to the right, and so their velocities are positive. Classically the lower bound $v_1 = \hbar k_1/m$ on the velocity ensures that the time a particle would spend in $[-a,a]$ in the absence of the potential is bounded. In practice it is consistent with our information to assume that the probability of the velocity of an incident particle being greater than some velocity $v_2 = \hbar k_2/m$ is zero.

We shall also assume that points where $|\rho|$, $\arg\rho$, $|\tau|$, or $\arg\tau$ are singular are not in supp $p_{\text{mom,in}}$. For example, at a resonance when $D\arg\tau$ becomes infinite we might well expect the variance of position of the transmitted ensemble to become infinite. In practice the assumption that $|\rho|$, etc., are C_∞ where they are not singular is satisfied, so our assumptions about them are reasonable.

II. EXPECTATION VALUE IN QUANTUM MECHANICS

A state μ in QM is a probability measure on the lattice $\mathscr{E}(\mathscr{H})$ of projectors on a Hilbert space $\mathscr{H}$. A set of weightings $\{w_j\}_{j=1}^N$ is a countable set of positive numbers that sum to unity; its cardinal number N is either a positive integer, or ∞. Gleason's theorem[2] states that given a state $\mu \exists$ a set of weightings $\{w_j\}_{j=1}^N$ and a corresponding set of unit vectors $\{\psi_j\}_{j=1}^N$ in $\mathscr{H}$ such that, if $\mu(E)$ is the probability assigned by μ to E in $\mathscr{E}(\mathscr{H})$, then

$$\mu(E) = \sum_{j=1}^{N} w_j \|E\psi_j\|^2. \tag{2.1}$$

The sets $\{w_j\}_{j=1}^N$ and $\{\psi_j\}_{j=1}^N$ determine the state μ, but the converse is not necessarily true.

Definition 2.1: We shall call the pair $(\{w_j\}_{j=1}^N, \{\psi_j\}_{j=1}^N)$ *a representation* of μ. □

Definition 2.2: J denotes a positive integer such that $J \leqslant N$. □

Observables in QM, which we shall denote by A, B, C, etc., are self-adjoint operators in $\mathscr{H}$. We shall denote the expectation value of A in the state μ by $\langle A \rangle$, etc.

Proposition 2.1: If $\langle A^2 \rangle$ exists so does $\langle A \rangle$. Further each ψ_j is in the domain $\mathscr{D}(A)$ of A, and

$$\langle A \rangle = \sum_{j=1}^{N} w_j \langle \psi_j | A | \psi_j \rangle, \tag{2.2a}$$

$$\langle A^2 \rangle = \sum_{j=1}^{N} w_j \|A\psi_j\|^2. \tag{2.2b}$$

 0022-2488/87/071583-06$02.50

Proof: See Ref. 3, Chap. 3, or Ref. 4. □

Proposition 2.2: Suppose $\psi_j \in \mathscr{D}(A)$ for each j. Then if

$$\sum_{j=1}^{N} w_j \|A\psi_j\|^2$$

is convergent both $\langle A \rangle$ and $\langle A^2 \rangle$ exist and are given by (2.2).

Proof: Let E_I be the member of the spectral family of A corresponding to the finite interval I of $\mathbb{R}$. Then $E_I A E_I$ is a bounded self-adjoint operator, and

$$\langle (E_I A E_I)^2 \rangle = \sum_{j=1}^{N} w_j \|E_I A E_I \psi_j\|^2 \leqslant \sum_{j=1}^{N} w_j \|A\psi_j\|^2,$$

so when I expands to fill up $\mathbb{R}$ $\langle (E_I A E_I)^2 \rangle$ has a unique limit; that is, $\langle A^2 \rangle$ exists. By Proposition 2.1 so does $\langle A \rangle$, and $\langle A \rangle$ and $\langle A^2 \rangle$ are given by (2.2). □

If $A \circ B = \frac{1}{2}(AB + BA)$ is an observable, $\psi_j \in \mathscr{D}(A \circ B)$ for each j, and $N < \infty$, its expectation value is given by

$$\langle A \circ B \rangle = \sum_{j=1}^{N} w_j \langle \psi_j | A \circ B | \psi_j \rangle,$$

which can be written (since A and B are observables)

$$\langle A \circ B \rangle = \sum_{j=1}^{N} w_j \operatorname{Re}\langle A\psi_j | B\psi_j \rangle. \tag{2.3}$$

These assumptions are, of course, too restrictive so we take (2.3) as a *definition* of $\langle A \circ B \rangle$.

Definition 2.3: We say that $\langle A \circ B \rangle$ exists if the right-hand side of (2.3) is convergent, in which case (2.3) defines the value of $\langle A \circ B \rangle$. □

Proposition 2.3: If $\langle A^2 \rangle$ and $\langle B^2 \rangle$ exist then so does $\langle A \circ B \rangle$.

Proof:

$$\sum_{j=1}^{J} w_j |\operatorname{Re}\langle A\psi_j | B\psi_j \rangle|$$

$$\leqslant \sum_{j=1}^{J} w_j \|A\psi_j\| \, \|B\psi_j\| \quad \text{(Schwarz's inequality)}$$

$$= \sum_{j=1}^{J} (w_j^{1/2} \|A\psi_j\|)(w_j^{1/2} \|B\psi_j\|)$$

$$\leqslant \left(\sum_{j=1}^{J} w_j \|A\psi_j\|^2 \right)^{1/2} \left(\sum_{j=1}^{J} w_j \|B\psi_j\|^2 \right)^{1/2}$$

(Schwarz's inequality)

$$\leqslant \langle A^2 \rangle^{1/2} \langle B^2 \rangle^{1/2} \quad \text{[by (2.2b)]}.$$

Thus the right-hand side of (2.3) is absolutely convergent, so $\langle A \circ B \rangle$ exists by Definition 2.3. We also obtain

$$|\langle A \circ B \rangle| \leqslant \langle A^2 \rangle^{1/2} \langle B^2 \rangle^{1/2}. \tag{2.4}$$

III. COVARIANCE IN QUANTUM MECHANICS

In Ref. 4 certain results were obtained on the hypothesis that $aA + bB$ is an observable for a pair of nonzero real numbers a and b. These results are not of the right form for our purposes here. In this section we shall derive these results in a modified form, ready for subsequent application.

Definition 3.1: We shall say that two observables A and B are *equal on* μ if they satisfy the following condition: if $(\{w_j\}_{j=1}^N, \{\psi_j\}_{j=1}^N)$ is a representation of μ then, for each value of j, $\psi_j \in \mathscr{D}(A) \cap \mathscr{D}(B)$, and $A\psi_j = B\psi_j$. We shall write this equality as $A =_\mu B$. □

Proposition 3.1: If $\langle A^2 \rangle$ and $\langle B^2 \rangle$ exist and $C =_\mu aA + bB$ then $\langle C \rangle$ and $\langle C^2 \rangle$ exist and

$$\langle C \rangle = a\langle A \rangle + b\langle B \rangle, \tag{3.1a}$$

$$\langle C^2 \rangle = a^2 \langle A^2 \rangle + 2ab \langle A \circ B \rangle + b^2 \langle B^2 \rangle. \tag{3.1b}$$

Proof: Recalling Definition 2.2, $C =_\mu aA + bB$ implies that

$$\sum_{j=1}^{J} w_j \|C\psi_j\|^2 = a^2 \sum_{j=1}^{J} w_j \|A\psi_j\|^2 + 2ab \times \sum_{j=1}^{J} w_j \operatorname{Re}\langle A\psi_j | B\psi_j \rangle + b^2 \sum_{j=1}^{J} \|B\psi_j\|^2.$$

If $N < \infty$, set $J = N$, otherwise let $J \to \infty$ and use Propositions 2.1 and 2.3: in either case we obtain

$$\sum_{j=1}^{N} w_j \|C\psi_j\|^2 = a^2 \langle A^2 \rangle + 2ab \langle A \circ B \rangle + b^2 \langle B^2 \rangle. \tag{3.2}$$

The right-hand side of (3.2) is finite, hence so is the left-hand side. Thus $\langle C^2 \rangle$ exists by Proposition 2.2, and (3.1b) is obtained from (3.2) using (2.2b) and (2.3). Now (3.1a) follows by the facts that $C =_\mu aA + bB$ implies that $\langle C \rangle = \langle aA + bB \rangle$, and $\langle \cdot \rangle$ is linear. □

Proposition 3.2: If $\langle A^2 \rangle$ and $\langle B^2 \rangle$ exist and $C =_\mu aA + bB$ then, if $\langle D^2 \rangle$ also exists,

$$\langle C \circ D \rangle = a\langle A \circ D \rangle + b\langle B \circ D \rangle, \tag{3.3a}$$

$$\langle D \circ C \rangle = a\langle D \circ A \rangle + b\langle D \circ B \rangle. \tag{3.3b}$$

Proof: Since $C =_\mu aA + bB$,

$$\sum_{j=1}^{J} w_j \operatorname{Re}\langle C\psi_j | D\psi_j \rangle = a \sum_{j=1}^{J} w_j \operatorname{Re}\langle A\psi_j | D\psi_j \rangle + b \sum_{j=1}^{J} w_j \operatorname{Re}\langle B\psi_j | D\psi_j \rangle.$$

If $N < \infty$ set $J = N$, otherwise let $J \to \infty$. In either case we get, using Proposition 2.3,

$$\sum_{j=1}^{N} w_j \operatorname{Re}\langle C\psi_j | D\psi_j \rangle = a\langle A \circ D \rangle + b\langle B \circ D \rangle. \tag{3.4}$$

The left-hand side of (3.4) therefore exists, hence (3.3a) follows by the definition of $\langle C \circ D \rangle$. Then (3.3b) follows from (3.3a) by the fact that $\langle C \circ D \rangle = \langle D \circ C \rangle$, etc. □

Remark 3.1: Proposition 3.2 establishes the bilinearity of $\langle \cdot \circ \cdot \rangle$. □

Proposition 3.3: If $\operatorname{Var} A$ and $\operatorname{Var} B$ exist, $C =_\mu aA + bB$, and a,b are both nonzero then

$$\langle (A - \langle A \rangle) \circ (B - \langle B \rangle) \rangle$$

is independent of the representation of μ.

Proof: If we subtract (3.1a) from $C =_\mu aA + bB$ we get

$$C - \langle C \rangle =_\mu a(A - \langle A \rangle) + b(B - \langle B \rangle). \tag{3.5}$$

Since $\operatorname{Var} A$ and $\operatorname{Var} B$ exist we can replace A, B, and C in (3.1b) by $A - \langle A \rangle$, $B - \langle B \rangle$, and $C - \langle C \rangle$, respectively, to obtain

$$\operatorname{Var} C = a^2 \operatorname{Var} A + 2ab \langle (A - \langle A \rangle) \circ (B - \langle B \rangle) \rangle + b^2 \operatorname{Var} B. \tag{3.6}$$

Since $ab \neq 0$ we can use (3.6) to express

$\langle (A - \langle A \rangle) \circ (B - \langle B \rangle) \rangle$ in terms of quantities which are independent of the representation of μ. □

Definition 3.2: If $\langle (A - \langle A \rangle) \circ (B - \langle B \rangle) \rangle$ is independent of the representation of μ the covariance of A and B, $\mathrm{Cov}(A,B)$, may be defined by

$$\mathrm{Cov}(A,B) = \langle (A - \langle A \rangle) \circ (B - \langle B \rangle) \rangle. \tag{3.7}$$

Proposition 3.4: If $\mathrm{Var}\,A$ and $\mathrm{Var}\,B$ exist, and $C =_\mu aA + bB$, where a and b are nonzero, then

$$\mathrm{Var}\,C = a^2\,\mathrm{Var}\,A + 2ab\,\mathrm{Cov}(A,B) + b^2\,\mathrm{Var}\,B. \tag{3.8}$$

Proof: Definition 3.2 is applicable, so insert (3.7) into (3.6). □

Proposition 3.5: If $\mathrm{Var}\,A$ and $\mathrm{Var}\,B$ exist, and $C =_\mu aA + bB$, where a and b are nonzero, then

$$\mathrm{Cov}(C,B) = a\,\mathrm{Cov}(A,B) + b\,\mathrm{Var}\,B. \tag{3.9}$$

Proof: Since $\mathrm{Var}\,A$ and $\mathrm{Var}\,B$ exist and $C =_\mu aA + bB$, (3.5) is valid, so again replace A, B, and C in Propositions 3.1 and 3.2 by $A - \langle A \rangle$, $B - \langle B \rangle$, and $C - \langle C \rangle$, respectively. Thus $\mathrm{Var}\,C$ exists and by (3.3a) with $D = B - \langle B \rangle$

$$\begin{aligned}\langle (C - \langle C \rangle) \circ (B - \langle B \rangle) \rangle &= a \langle (A - \langle A \rangle) \circ (B - \langle B \rangle) \rangle \\ &\quad + b \langle (B - \langle B \rangle) \circ (B - \langle B \rangle) \rangle.\end{aligned}$$

Since $C =_\mu aA + bB$, where a and b are nonzero, Proposition 3.3 enables us to use (3.7) for the first term on the right-hand side. Also $\langle (B - \langle B \rangle) \circ (B - \langle B \rangle) \rangle = \langle (B - \langle B \rangle)^2 \rangle = \mathrm{Var}\,B$ so we get

$$\langle (C - \langle C \rangle) \circ (B - \langle B \rangle) \rangle = a\,\mathrm{Cov}(A,B) + b\,\mathrm{Var}\,B. \tag{3.10}$$

The right-hand side of (3.10) is independent of the representation of μ, so the left-hand side equals $\mathrm{Cov}(C,B)$, and (3.9) results. □

IV. LOCALIZED STATES IN ONE DIMENSION

The results derived in Secs. II and III are general; in this section we return to the special case of one degree of freedom.

Definition 4.1: A state μ is *localized* if $|x|^\nu p_{\mathrm{pos}}(x)$ and $|k|^\nu p_{\mathrm{mom}}(k)$ are bounded for all $\nu \geqslant 0$. □

Remark 4.1: As we pointed out in the Introduction it is reasonable to suppose that a state μ, which can be taken to describe an ensemble of particles in nature, is localized. □

Proposition 4.1: Let $\mathscr{S}$ denote the set of Schwarz testing functions $\mathbb{R} \to \mathbb{C}$; then, if μ is localized, for $j = 1,\ldots,N$, ψ_j and $F\psi_j$ belong to $\mathscr{S}$ while

$$\hat{x}\psi_j = F^* iDF\psi_j. \tag{4.1}$$

Proof: First we note that,[5] a.e.,

$$p_{\mathrm{pos}}(x) = \sum_{j=1}^{N} w_j |\psi_j(x)|^2, \tag{4.2a}$$

$$p_{\mathrm{mom}}(k) = \sum_{j=1}^{N} w_j |F\psi_j(k)|^2. \tag{4.2b}$$

Since $w_j > 0$ for each value of j, (4.2) implies that

$$|\psi_j(x)|^2 \leqslant w_j^{-1} p_{\mathrm{pos}}(x) \quad \text{a.e.};$$

$$|F\psi_j(k)|^2 \leqslant w_j^{-1} p_{\mathrm{mom}}(k) \quad \text{a.e.}$$

The hypothesis that μ is localized now implies that, if $\nu \geqslant 0$, $|x|^\nu |\psi_j(x)|^2$ and $|k|^\nu |F\psi_j(k)|^2$ are essentially bounded. It follows (Ref. 6, Chap. III) that ψ_j and $F\psi_j$ are equal a.e. to C_∞ functions; they can therefore be redefined so that they *are* C_∞ functions, and (4.1) is valid pointwise. □

The following result is obvious.

Proposition 4.2: If μ is localized then $\langle \hat{x}^2 \rangle$, $\langle \tilde{v}^2 \rangle$, $\mathrm{Var}\,\hat{x}$, and $\mathrm{Var}\,\tilde{v}$ all exist.

V. FREE MOTION IN ONE DIMENSION

In the Heisenberg picture the state μ is fixed at its time zero value, while an observable represented at time zero by A is represented at time t by $A(t) = U_t^* A U_t$, where U_t is the free evolution operator. Thus

$$U_t = F^* \exp(-i\widehat{\omega t})F, \tag{5.1}$$

where $\exp(-i\widehat{\omega t})$ is the operator of multiplication by $\exp(-i\omega t)$ $(\omega = \hbar k^2/2m = mv^2/2\hbar)$.

Proposition 5.1: If μ is localized the velocity observable $v(t)$ at time t and position observable $x(t)$ at time t satisfy

$$v(t) =_\mu \tilde{v}, \tag{5.2a}$$

$$x(t) =_\mu \hat{x} + \tilde{v}t. \tag{5.2b}$$

Proof: See Sec. I of I. □

Remark 5.1: The result expressed by Proposition 5.1 for the free evolution of velocity and position in the Heisenberg picture is, of course, well known. We have proved it here for localized states. □

Proposition 5.2: If μ is localized then the means, variances, and covariance of position and velocity at time t are given by

$$\langle v(t) \rangle = \langle \tilde{v} \rangle, \tag{5.3a}$$

$$\langle x(t) \rangle = \langle \hat{x} \rangle + \langle \tilde{v} \rangle t, \tag{5.3b}$$

$$\mathrm{Var}\,v(t) = \mathrm{Var}\,\tilde{v}, \tag{5.3c}$$

$$\mathrm{Cov}(x(t),v(t)) = \mathrm{Cov}(\hat{x},\tilde{v}) + t\,\mathrm{Var}\,\tilde{v}, \tag{5.3d}$$

$$\mathrm{Var}\,x(t) = \mathrm{Var}\,\hat{x} + 2t\,\mathrm{Cov}(\hat{x},\tilde{v}) + t^2\,\mathrm{Var}\,\tilde{v}. \tag{5.3e}$$

Proof: Since μ is localized $\langle \tilde{v}^2 \rangle$ and $\langle \hat{x}^2 \rangle$ exist, so we can use the result proved in Ref. 4. □

Remark 5.2: The results (5.3) are valid under more general conditions,[4] but the hypotheses of Proposition 5.2 are adequate for our purposes. □

Remark 5.3: Proposition 5.2 may be proved directly by use of (5.2) combined with (3.1a), (3.8), and (3.9). □

VI. ASSUMPTIONS

We now return to the problem of an ensemble incident on a piecewise continuous potential confined to $[-a,a]$.

Assumption 6.1: The initial state is localized and has representation

$$(\{w_{\mathrm{in},j}\}_{j=1}^N, \{\psi_{\mathrm{in},j}\}_{j=1}^N).$$

Assumption 6.2: The probability density of momentum, $p_{\mathrm{mom,in}}$, of the incident ensemble, has support contained in $[k_1,k_2]$, where $0 < k_1 < k_2 < \infty$.

Assumption 6.3: If

$$w_{\mathrm{tr}} = \sum_{j=1}^{N} w_{\mathrm{in},j} \|\tilde{\tau}\psi_{\mathrm{in},j}\|^2, \tag{6.1a}$$

then w_{tr} is the probability of transmission, and the state of the transmitted ensemble has representation

$$(\{w_{tr}^{-1}w_{in,j}\|\tilde{\tau}\psi_{in,j}\|^2\}_{j=1}^N, \{\|\tilde{\tau}\psi_{in,j}\|^{-1}\tilde{\tau}\psi_{in,j}\}_{j=1}^N).$$

Remark 6.1: This is a natural generalization of Sec. III of I.

Assumption 6.4: If

$$w_{re} = \sum_{j=1}^N w_{in,j}\|\tilde{\rho}\psi_{in,j}\|^2, \tag{6.1b}$$

then w_{re} is the probability of reflection, and the state of the reflected particles has the representation

$$(\{w_{re}^{-1}w_{in,j}\|\tilde{\rho}\psi_{in,j}\|^2\}_{j=1}^N, \{\|\tilde{\rho}\psi_{in,j}\|^{-1}P\tilde{\rho}\psi_{in,j}\}_{j=1}^N).$$

Remark 6.2: This is a natural generalization of Sec. IV of I.

Assumption 6.5: $|\tau|$, $\arg\tau$, $|\rho|$, and $\arg\rho$ are C_∞ on $\operatorname{Supp} p_{mom,in}$, the support of $p_{mom,in}$.

Remark 6.3: It follows from Assumption 6.1 and Proposition 4.1 that $\psi_{in,j}\in\mathscr{S}$ ($j=1,...,N$). Therefore by (4.2b) applied to the initial state $\operatorname{Supp} F\psi_{in,j}\subseteq\operatorname{Supp} p_{mom,in}$ ($j=1,...,N$). Hence by Assumption 6.5 $\tilde{\tau}\psi_{in,j}$ and $\tilde{\rho}\psi_{in,j}$ also belong to $\mathscr{S}$ ($j=1,...,N$).

VII. THE INCIDENT ENSEMBLE

We shall denote by μ_{in} the state of the incident ensemble, and by $=_{in}$ "equality on μ_{in}" (cf. Definition 3.1). We shall also denote by $\langle\cdot\rangle_{in}$, $\operatorname{Var}_{in}\cdot$, and $\operatorname{Cov}_{in}(\cdot,\cdot)$ the expectation value, variance, and covariance over μ_{in}, respectively. The conditions of Proposition 5.1 are satisfied so we get the following proposition.

Proposition 7.1: The velocity observable $v_{in}(t)$ and position observable $x_{in}(t)$ at time t satisfy

$$v_{in}(t) =_{in} \tilde{v}, \tag{7.1a}$$

$$x_{in}(t) =_{in} x_{in} + \tilde{v}t, \tag{7.1b}$$

where $x_{in} = \hat{x}$.

The conditions of Proposition 5.2 are also satisfied so we have the next proposition.

Proposition 7.2: The means, variances, and covariance of position and velocity at time t of the incident ensemble are given by

$$\langle v_{in}(t)\rangle_{in} = \langle\tilde{v}\rangle_{in}, \tag{7.2a}$$

$$\langle x_{in}(t)\rangle_{in} = \langle x_{in}\rangle_{in} + \langle\tilde{v}\rangle_{in}t, \tag{7.2b}$$

$$\operatorname{Var}_{in} v_{in}(t) = \operatorname{Var}_{in}\tilde{v}, \tag{7.2c}$$

$$\operatorname{Cov}_{in}(x_{in}(t),v_{in}(t)) = \operatorname{Cov}_{in}(x_{in},\tilde{v}) + t\operatorname{Var}_{in}\tilde{v}, \tag{7.2d}$$

$$\operatorname{Var}_{in} x_{in}(t) = \operatorname{Var}_{in} x_{in} + 2t\operatorname{Cov}_{in}(x_{in},\tilde{v}) + t^2\operatorname{Var}_{in}\tilde{v}. \tag{7.2e}$$

VIII. EXPECTATION VALUES OVER THE TRANSMITTED ENSEMBLE

Suppose an observable is represented by $A(t) = U_t^*AU_t$ at time t, and the expectation value of $[A(t)]^2$ over the transmitted particles is finite. Then by Proposition 2.1 and Assumption 6.3 the expectation values of $A(t)$ and $[A(t)]^2$ over the transmitted particles are given by the expressions

$$\sum_{j=1}^N w_{tr}^{-1}w_{in,j}\langle\tilde{\tau}\psi_{in,j}|A(t)|\tilde{\tau}\psi_{in,j}\rangle, \tag{8.1a}$$

$$\sum_{j=1}^N w_{tr}^{-1}w_{in,j}\|A(t)\tilde{\tau}\psi_{in,j}\|^2. \tag{8.1b}$$

Using the transmission shift operator S_{tr} defined in I (8.1) can be written

$$\sum_{j=1}^N w_{tr,j}\langle\psi_{tr,j}|A_{tr}(t)|\psi_{tr,j}\rangle, \tag{8.2a}$$

$$\sum_{j=1}^N w_{tr,j}\|A_{tr}(t)\psi_{tr,j}\|^2, \tag{8.2b}$$

where

$$w_{tr,j} = w_{tr}^{-1}w_{in,j}\|\tilde{\tau}\psi_{in,j}\|^2, \tag{8.3}$$

$$\psi_{tr,j} = \|\tilde{\tau}\psi_{in,j}\|^{-1}|\tilde{\tau}|\psi_{in,j}, \tag{8.4}$$

$$A_{tr}(t) = S_{tr}^*A(t)S_{tr}. \tag{8.5}$$

The set $\{w_{tr,j}\}_{j=1}^N$ is a set of weightings, and $\{\psi_{tr,j}\}_{j=1}^N$ is a set of unit vectors in $\mathscr{H}$.

Definition 8.1: μ_{tr} is the state defined by the representation

$$(\{w_{tr,j}\}_{j=1}^N, \{\psi_{tr,j}\}_{j=1}^N);$$

expectation value, variance, and covariance over μ_{tr} are denoted by $\langle\cdot\rangle_{tr}$, $\operatorname{Var}_{tr}\cdot$, and $\operatorname{Cov}_{tr}(\cdot,\cdot)$, respectively, and equality over μ_{tr} by $=_{tr}$. □

Expressions (8.2) and Definition 8.1 lead to the following proposition.

Proposition 8.1: If the expectation value of $[A(t)]^2$ exists over the transmitted particles then the expectation values of $A(t)$ and $[A(t)]^2$ over the transmitted particles are $\langle A_{tr}(t)\rangle_{tr}$ and $\langle[A(t)]^2\rangle_{tr}$, respectively, where $\langle\cdot\rangle_{tr}$ is defined by Definition 8.1 and $A_{tr}(t)$ is the observable defined by (8.5). □

Proposition 8.1 has a converse.

Proposition 8.2: If $\langle[A_{tr}(t)]^2\rangle_{tr}$ exists then $A(t)$ and $[A(t)]^2$ have expectation values over the transmitted particles, and these are given by $\langle A_{tr}(t)\rangle_{tr}$ and $\langle[A_{tr}(t)]^2\rangle_{tr}$, respectively.

Proof: If $\langle[A_{tr}(t)]^2\rangle_{tr}$ exists then by Proposition 2.1 the series (8.2b) converges. It follows that the series (8.1b) converges, and so by Proposition 2.2 $A(t)$ and $[A(t)]^2$ have expectation values over the transmitted particles, given by (8.1), and therefore by (8.2); that is, by $\langle A_{tr}(t)\rangle_{tr}$ and $\langle[A_{tr}(t)]^2\rangle_{tr}$. □

Now S_{tr} and U_t commute, and $A(t) = U_t^*AU_t$, so (8.5) gives

$$A_{tr}(t) = U_t^*A_{tr}U_t, \tag{8.6}$$

where

$$A_{tr} = S_{tr}^*AS_{tr}. \tag{8.7}$$

Proposition 8.3: If $\operatorname{Cov}_{tr}(A_{tr}(t), B_{tr}(t))$ is defined then it is the covariance of $A(t)$ and $B(t)$ over the transmitted particles.

Proof: $\operatorname{Cov}_{tr}(A_{tr}(t), B_{tr}(t))$ is given by

$$\sum_{j=1}^N w_{tr,j}\operatorname{Re}\langle(A_{tr}(t) - \langle A_{tr}(t)\rangle_{tr})\psi_{tr,j}|(B_{tr}(t) - \langle B_{tr}(t)\rangle_{tr})\psi_{tr,j}\rangle.$$

This may be rewritten

$$\sum_{j=1}^{N} w_{\rm tr}^{-1} w_{{\rm in},j} \operatorname{Re}\langle (A(t) - \langle A_{\rm tr}(t)\rangle_{\rm tr})\tilde{\tau}\psi_{{\rm in},j} | (B(t) - \langle B_{\rm tr}(t)\rangle_{\rm tr})\tilde{\tau}\psi_{{\rm in},j}\rangle. \tag{8.8}$$

By Proposition 8.1 $\langle A_{\rm tr}(t)\rangle_{\rm tr}$ and $\langle B_{\rm tr}(t)\rangle_{\rm tr}$ are the expectation values of $A(t)$ and $B(t)$ over the transmitted particles. Since $\operatorname{Cov}_{\rm tr}(A_{\rm tr}(t), B_{\rm tr}(t))$ is defined the expression (8.8) is independent of the representation, and hence by Assumption 6.3 defines the covariance of $A(t)$ and $B(t)$ over the transmitted ensemble. □

IX. STATISTICAL PROPERTIES OF POSITION AND MOMENTUM FOR THE TRANSMITTED PARTICLES

For $j = 1,...,N$, $\psi_{{\rm tr},j} \in \mathscr{S}$, $\operatorname{Supp} F\psi_{{\rm tr},j} \subseteq [k_1,k_2]$, while $|\tau|$ and $\arg\tau$ are C_∞ on $\operatorname{Supp} F\psi_{{\rm tr},j}$. Hence the proofs of Sec. III of I relating to the shifted position and velocity observables go over, and we obtain the following results:

$$v_{\rm tr}(t) =_{\rm tr} \tilde{v}, \tag{9.1a}$$

$$x_{\rm tr}(t) =_{\rm tr} x_{\rm tr} + \tilde{v}t, \tag{9.1b}$$

where

$$x_{\rm tr} = S_{\rm tr}^{*}\hat{x}S_{\rm tr}, \tag{9.1c}$$

$$x_{\rm tr} =_{\rm tr} x_{\rm in} - \tilde{v}\tilde{t}_{\rm tr}, \tag{9.1d}$$

where

$$t_{\rm tr} = v^{-1} D \arg\tau. \tag{9.1e}$$

Proposition 9.1: $\operatorname{Var}_{\rm tr}\tilde{v}$, $\operatorname{Var}_{\rm tr}\tilde{v}\tilde{t}_{\rm tr}$, $\operatorname{Var}_{\rm tr}\hat{x}$, and $\operatorname{Var}_{\rm tr} x_{\rm tr}$ all exist.

Proof: The probability density $p_{\rm mom,tr}$ for momentum for the state $\mu_{\rm tr}$ is given by

$$p_{\rm mom,tr}(k) = \sum_{j=1}^{N} w_{{\rm tr},j} |F\psi_{{\rm tr},j}(k)|^2 \quad \text{(Definition 8.1)}$$
$$= \sum_{j=1}^{N} w_{\rm tr}^{-1} w_{{\rm in},j} |\tau(k)|^2 |F\psi_{{\rm in},j}(k)|^2$$
$$\text{[by (8.3) and (8.4)]}.$$

Now by (4.2b)

$$p_{\rm mom,in}(k) = \sum_{j=1}^{N} w_{{\rm in},j} |F\psi_{{\rm in},j}(k)|^2 \tag{9.2}$$

is the probability density of momentum in the initial state $\mu_{\rm in}$, so

$$p_{\rm mom,tr}(k) = w_{\rm tr}^{-1} |\tau(k)|^2 p_{\rm mom,in}(k). \tag{9.3}$$

Equation (9.3) shows that $\operatorname{Supp} p_{\rm mom,tr} \subseteq \operatorname{Supp} p_{\rm mom,in}$, and so by Assumption 6.2 $\operatorname{Supp} p_{\rm mom,tr}$ is bounded, hence v is bounded on this support; thus $\langle\tilde{v}^2\rangle_{\rm tr}$ exists, and hence so does $\operatorname{Var}_{\rm tr}\tilde{v}$.

By (9.1e) $v\, t_{\rm tr} = D\arg\tau$, and so by Assumption 6.5 is bounded on $\operatorname{Supp} p_{\rm mom,tr}$, hence $\langle(\tilde{v}\tilde{t}_{\rm tr})^2\rangle_{\rm tr}$ exists. Thus $\operatorname{Var}_{\rm tr}\tilde{v}\tilde{t}_{\rm tr}$ exists.

To prove that $\langle\hat{x}^2\rangle_{\rm tr} < \infty$, first note that $\hat{x} =_{\rm tr} F^* iDF$, so by (8.4)

$$\|\hat{x}\psi_{{\rm tr},j}\|^2 = \|\tilde{\tau}\psi_{{\rm in},j}\|^{-2} \|D|\tau|F\psi_{{\rm in},j}\|^2;$$

using (8.3), this gives

$$w_{{\rm tr},j}\|\hat{x}\psi_{{\rm tr},j}\|^2 = w_{\rm tr}^{-1} w_{{\rm in},j} \|(D|\tau|)F\psi_{{\rm in},j} + |\tau|DF\psi_{{\rm in},j}\|^2.$$

Now Assumption 6.5 implies that $|\tau|$ and $D|\tau|$ are bounded on $\operatorname{Supp} p_{\rm mom,in} \supseteq \operatorname{Supp} F\psi_{{\rm in},j}$. They are therefore both less than or equal to some positive number M, which is independent of j. Thus

$$\begin{aligned} w_{{\rm tr},j}\|\hat{x}\psi_{{\rm tr},j}\|^2 &= w_{\rm tr}^{-1} w_{{\rm in},j} [\|(D|\tau|)F\psi_{{\rm in},j}\|^2 \\ &\quad + 2\operatorname{Re}\langle (D|\tau|)F\psi_{{\rm in},j} | |\tau| DF\psi_{{\rm in},j}\rangle + \||\tau|DF\psi_{{\rm in},j}\|^2] \\ &\leqslant w_{\rm tr}^{-1} w_{{\rm in},j} M^2 (\|F\psi_{{\rm in},j}\|^2 + 2\|F\psi_{{\rm in},j}\|\|DF\psi_{{\rm in},j}\| \\ &\quad + \|DF\psi_{{\rm in},j}\|^2) \quad \text{(using Schwarz's inequality)} \\ &= w_{\rm tr}^{-1} w_{{\rm in},j} M^2 (1 + 2\|\hat{x}\psi_{{\rm in},j}\| + \|\hat{x}\psi_{{\rm in},j}\|^2), \end{aligned} \tag{9.4}$$

since $\hat{x} =_{\rm in} F^* iDF$.

Summing the inequality (9.4) for $j = 1,...,N$ using

$$\sum_{j=1}^{N} w_{{\rm in},j} = 1, \tag{9.5}$$

we obtain

$$\sum_{j=1}^{N} w_{{\rm tr},j}\|\hat{x}\psi_{{\rm tr},j}\|^2 \leqslant w_{\rm tr}^{-1} M^2 \Big(1 + 2\sum_{j=1}^{N} w_{{\rm in},j}^{1/2}\cdot w_{{\rm in},j}^{1/2}\|\hat{x}\psi_{{\rm in},j}\| + \sum_{j=1}^{N} w_{{\rm in},j}\|\hat{x}\psi_{{\rm in},j}\|^2\Big)$$

and so by Schwarz's inequality and (9.5)

$$\sum_{j=1}^{N} w_{{\rm tr},j}\|\hat{x}\psi_{{\rm tr},j}\|^2 \leqslant w_{\rm tr}^{-1} M^2 \Big[1 + 2\Big(\sum_{j=1}^{N} w_{{\rm in},j}\|\hat{x}\psi_{{\rm in},j}\|^2\Big)^{1/2} + \sum_{j=1}^{N} w_{{\rm in},j}\|\hat{x}\psi_{{\rm in},j}\|^2\Big]$$
$$= w_{\rm tr}^{-1} M^2 [1 + 2(\langle\hat{x}^2\rangle_{\rm in})^{1/2} + \langle\hat{x}^2\rangle_{\rm in}] < \infty.$$

Therefore $\langle\hat{x}^2\rangle_{\rm tr}$ exists by Proposition 2.2. The existence of $\langle(x_{\rm tr})^2\rangle_{\rm tr}$ now follows by (9.1d) since $x_{\rm in} = \hat{x}$. □

Proposition 9.2: The means, variances, and covariance of position and velocity at time t of the transmitted ensemble all exist and are given by

$$\langle v_{\rm tr}(t)\rangle_{\rm tr} = \langle\tilde{v}\rangle_{\rm tr}, \tag{9.6a}$$

$$\langle x_{\rm tr}(t)\rangle_{\rm tr} = \langle x_{\rm tr}\rangle_{\rm tr} + \langle\tilde{v}\rangle_{\rm tr} t, \tag{9.6b}$$

$$\operatorname{Var}_{\rm tr} v_{\rm tr}(t) = \operatorname{Var}_{\rm tr}\tilde{v}, \tag{9.6c}$$

$$\operatorname{Cov}_{\rm tr}(x_{\rm tr}(t), v_{\rm tr}(t)) = \operatorname{Cov}_{\rm tr}(x_{\rm tr},\tilde{v}) + t\operatorname{Var}_{\rm tr}\tilde{v}, \tag{9.6d}$$

$$\operatorname{Var}_{\rm tr} x_{\rm tr}(t) = \operatorname{Var}_{\rm tr} x_{\rm tr} + 2t\operatorname{Cov}_{\rm tr}(x_{\rm tr},\tilde{v}) + t^2\operatorname{Var}_{\rm tr}\tilde{v}. \tag{9.6e}$$

Proof: By Proposition 9.1 $\langle\tilde{v}^2\rangle_{\rm tr}$ exists, hence the right-hand sides of (9.6a) and (9.6c) exist. Thus by (9.1a) $\langle v_{\rm tr}(t)\rangle_{\rm tr}$ and $\operatorname{Var}_{\rm tr} v_{\rm tr}(t)$ exist for all t, and are given by (9.6a) and (9.6c). Thus $\langle[v_{\rm tr}(t)]^2\rangle_{\rm tr}$ exists so by Proposition 8.2 these yield the mean and variance of velocity of the transmitted ensemble at time t.

Equation (9.6b) follows from (9.1b) by (3.1a), since $\langle\tilde{v}^2\rangle_{\rm tr}$ and $\langle(x_{\rm tr})^2\rangle_{\rm tr}$ exist by Proposition 9.1.

Equation (9.6d) follows from (9.1a) and (9.1b) using Proposition 3.5. That the left-hand side equals the covariance of position and velocity of the transmitted particles follows from Proposition 8.3.

Finally (9.6e) follows from (9.1a) and (9.1b) using Proposition 3.4. □

Proposition 9.3:

$$\mathrm{Cov}_{\mathrm{tr}}(x_{\mathrm{tr}},\tilde{v}) = \mathrm{Cov}_{\mathrm{tr}}(x_{\mathrm{in}},\tilde{v}) - \mathrm{Cov}_{\mathrm{tr}}(\tilde{v}\tilde{\tau}_{\mathrm{tr}},\tilde{v}). \quad (9.7)$$

Proof: It is easy to show from (9.1d) using Proposition 3.2 that

$$\begin{aligned}\langle (x_{\mathrm{tr}} - \langle x_{\mathrm{tr}}\rangle_{\mathrm{tr}})\circ(\tilde{v} - \langle\tilde{v}\rangle_{\mathrm{tr}})\rangle_{\mathrm{tr}} &= \langle (x_{\mathrm{in}} - \langle x_{\mathrm{in}}\rangle_{\mathrm{tr}})\circ(\tilde{v} - \langle\tilde{v}\rangle_{\mathrm{tr}})\rangle_{\mathrm{tr}} \\ &\quad - \langle (\tilde{v}\tilde{t}_{\mathrm{tr}} - \langle \tilde{v}\tilde{t}_{\mathrm{tr}}\rangle_{\mathrm{tr}})\circ(\tilde{v} - \langle\tilde{v}\rangle_{\mathrm{tr}})\rangle_{\mathrm{tr}}. \end{aligned} \quad (9.8)$$

The left-hand side of (9.8) is $\mathrm{Cov}_{\mathrm{tr}}(x_{\mathrm{tr}},\tilde{v})$, and the second term on the right-hand side equals

$$\int (vt_{\mathrm{tr}}(k) - \langle \tilde{v}\tilde{t}_{\mathrm{tr}}\rangle_{\mathrm{tr}})(v - \langle\tilde{v}\rangle_{\mathrm{tr}})p_{\mathrm{mom,tr}}(k)dk,$$

which is clearly independent of the representation of μ_{tr}. It follows that the first term on the right-hand side of (9.8) is independent of the representation of μ_{tr}, and so defines $\mathrm{Cov}_{\mathrm{tr}}(x_{\mathrm{in}},\tilde{v})$ by Definition 3.2. Now (9.7) follows. □

X. THE REFLECTED PARTICLES

The statistical properties of the reflected particles are derived in a similar way, and we state the results without proof (cf. Sec. IV of I).

Definition 10.1: μ_{re} is the state defined by the representation

$$(\{w_{\mathrm{re},j}\}_{j=1}^{N},\{\psi_{\mathrm{re},j}\}_{j=1}^{N}),$$

where

$$w_{\mathrm{re},j} = w_{\mathrm{re}}^{-1}w_{\mathrm{in},j}\|\tilde{\rho}\psi_{\mathrm{re},j}\|^2, \quad (10.1)$$

$$\psi_{\mathrm{re},j} = \|\tilde{\rho}\psi_{\mathrm{in},j}\|^{-1}|\tilde{\rho}|\psi_{\mathrm{in},j}; \quad (10.2)$$

expectation value, variance, and covariance over μ_{re} are denoted by $\langle\cdot\rangle_{\mathrm{re}}$, $\mathrm{Var}_{\mathrm{re}}\ \cdot$, and $\mathrm{Cov}_{\mathrm{re}}(\cdot,\cdot)$, respectively, and equality over μ_{re} by $=_{\mathrm{re}}$. □

Definition 10.2:

$$A_{\mathrm{re}}(t) = S_{\mathrm{re}}^{*}PA(t)PS_{\mathrm{re}}. \quad (10.3)$$

Proposition 10.1: If the expectation value of $[A(t)]^2$ exists over the reflected particles then the expectation value of $A(t)$ and $[A(t)]^2$ over the reflected particles are $\langle A_{\mathrm{re}}(t)\rangle_{\mathrm{re}}$ and $\langle [A_{\mathrm{re}}(t)]^2\rangle_{\mathrm{re}}$, respectively.

Proposition 10.2: If $\langle [A_{\mathrm{re}}(t)]^2\rangle_{\mathrm{re}}$ exists then $A(t)$ and $[A(t)]^2$ have expectation values over the reflected particles, and these are given by $\langle A_{\mathrm{re}}(t)\rangle_{\mathrm{re}}$ and $\langle [A_{\mathrm{re}}(t)]^2\rangle_{\mathrm{re}}$, respectively.

Proposition 10.3:

$$v_{\mathrm{re}}(t) =_{\mathrm{re}} -\tilde{v}, \quad (10.4a)$$

$$x_{\mathrm{re}}(t) =_{\mathrm{re}} x_{\mathrm{re}} - \tilde{v}t, \quad (10.4b)$$

where

$$x_{\mathrm{re}} = S_{\mathrm{re}}^{*}Px_{\mathrm{in}}PS_{\mathrm{re}}, \quad (10.4c)$$

$$x_{\mathrm{re}} =_{\mathrm{re}} -x_{\mathrm{in}} + \tilde{v}\tilde{t}_{\mathrm{re}}, \quad (10.4d)$$

where

$$t_{\mathrm{re}} = v^{-1}D\arg\rho. \quad (10.4e)$$

Proposition 10.4: $\mathrm{Var}_{\mathrm{re}}\ \tilde{v}$, $\mathrm{Var}_{\mathrm{re}}\ \tilde{v}\tilde{t}_{\mathrm{re}}$, $\mathrm{Var}_{\mathrm{re}}\ \hat{x}$, and $\mathrm{Var}_{\mathrm{re}}\ x_{\mathrm{re}}$ all exist.

Proposition 10.5: The means, variances, and covariance of position and velocity at time t of the reflected particles all exist, and are given by

$$\langle v_{\mathrm{re}}(t)\rangle_{\mathrm{re}} = -\langle\tilde{v}\rangle_{\mathrm{re}}, \quad (10.5a)$$

$$\langle x_{\mathrm{re}}(t)\rangle_{\mathrm{re}} = \langle x_{\mathrm{re}}\rangle_{\mathrm{re}} - \langle\tilde{v}\rangle_{\mathrm{re}}t, \quad (10.5b)$$

$$\mathrm{Var}_{\mathrm{re}}\ v_{\mathrm{re}}(t) = \mathrm{Var}_{\mathrm{re}}\ \tilde{v}, \quad (10.5c)$$

$$\mathrm{Cov}_{\mathrm{re}}(x_{\mathrm{re}}(t),v_{\mathrm{re}}(t)) = -\mathrm{Cov}_{\mathrm{re}}(x_{\mathrm{re}},\tilde{v}) + t\,\mathrm{Var}_{\mathrm{re}}\ \tilde{v}, \quad (10.5d)$$

$$\mathrm{Var}_{\mathrm{re}}\ x_{\mathrm{re}}(t) = \mathrm{Var}_{\mathrm{re}}\ x_{\mathrm{re}} - 2t\,\mathrm{Cov}_{\mathrm{re}}(x_{\mathrm{re}},\tilde{v}) + t^2\,\mathrm{Var}_{\mathrm{re}}\ \tilde{v}. \quad (10.5e)$$

Proposition 10.6:

$$\mathrm{Cov}_{\mathrm{re}}(x_{\mathrm{re}},\tilde{v}) = -\mathrm{Cov}_{\mathrm{re}}(x_{\mathrm{in}},\tilde{v}) + \mathrm{Cov}_{\mathrm{re}}(\tilde{v}\tilde{t}_{\mathrm{tr}},\tilde{v}).$$

XI. DISCUSSION

Assumption 6.1 could be relaxed by not requiring $p_{\mathrm{pos,in}}$ and $p_{\mathrm{mom,in}}$ to fall off to zero so fast at infinity; however, this assumption is physically reasonable as it stands.

Assumption 6.2 could also be relaxed, by removing the upper bound on $\mathrm{Supp}\,p_{\mathrm{mom,in}}$ and replacing it by a requirement of sufficiently rapid fall-off when $k\to\infty$. However, as pointed out elsewhere,[5] the initial state is experimentally indistinguishable from one in which $p_{\mathrm{mom,in}}$ satisfies Assumption 6.2, so further relaxation of this assumption is not required by physical considerations.

Rigorous justification of Assumptions 6.3 and 6.4 are outside the scope of the present paper.

Finally, the requirements imposed by Assumption 6.5 can be relaxed. Our arguments remain valid if we only require $|\tau|$, $\arg\tau$, $|\rho|$, $\arg\rho$, and their first derivatives, to be defined and bounded on $\mathrm{Supp}\,p_{\mathrm{mom,in}}$. In this case $\tilde{\tau}\psi_{\mathrm{in},j}$ and $\tilde{\rho}\psi_{\mathrm{in},j}$ may no longer be C_∞, but they are still differentiable and L_2.

[1] J. E. G. Farina, J. Math. Phys. **28**, 1580 (1987).
[2] A. M. Gleason, J. Math. Mech. **6**, 885 (1957).
[3] E. G. Beltrametti and G. Cassinelli, *The Logic of Quantum Mechanics* (Addison–Wesley, Reading, MA, 1981).
[4] J. E. G. Farina, Int. J. Theor. Phys. **21**, 83 (1982).
[5] J. E. G. Farina, Int. J. Theor. Phys. **23**, 1141 (1984).
[6] E. C. Titchmarsh, *Introduction to the Theory of Fourier Integrals* (Clarendon, Oxford, 1937).

On partial differential equations related to Lorenz system

D. Y. Hsieh
Division of Applied Mathematics, Brown University, Providence, Rhode Island 02912

(Received 6 October 1986; accepted for publication 11 March 1987)

Partial differential equations are constructed such that a truncation scheme as adopted by Lorenz [J. Atmos. Sci. **20**, 130 (1963)] will lead to Lorenz equations. The partial differential equations are much simpler than those of the Rayleigh–Bernard problem and are essentially of the mixed type. Various aspects of the partial differential equations are explored. It is suggested that the switching back and forth between the elliptic and hyperbolic regimes represents the chaotic behavior of the system in the context of partial differential equations.

I. INTRODUCTION

When Lorenz[1] first proposed the set of equations that would later bear his name, he obtained those equations by a drastic truncation to three spatially periodic modes from the complicated system of nonlinear partial differential equations of the Rayleigh–Bernard problem. Without going into the details of the development, it is sufficient for our purposes to state that for the Rayleigh–Bernard problem, we have two coupled high order and highly nonlinear partial differential equations for the streamfunction $\psi(x,z,t)$ and temperature $\theta(x,z,t)$. Lorenz proposed, in essence, to let

$$\psi = X(t)\sin ax \sin z\,, \tag{1}$$

$$\theta = Y(t)\cos ax \sin z - [Z(t)/\sqrt{2}]\sin 2z\,, \tag{2}$$

and then substitute into the partial differential equations. By retaining only those spatial modes as represented in Eqs. (1) and (2), he obtained the following Lorenz equations:

$$\begin{aligned}\frac{dX}{dt} &= -\sigma X + \sigma Y\,,\\ \frac{dY}{dt} &= -Y + rX + XZ\,, \qquad\qquad (\mathrm{L}),(\mathrm{L}_\beta)\\ \frac{dZ}{dt} &= -(1+3\beta)Z + XY\,,\end{aligned}$$

where σ, r, and $(1+3\beta)$ are positive parameters. We shall refer to this system by (L) or (L_β) when β is assigned some particular value.

The rich contents of the Lorenz equations have been explored extensively.[2] The bifurcations of the solution, the appearance of strange attractors, the chaotic behavior of the solutions, and the phenomena of period doubling and intermittency have all been associated with the study of Lorenz equations. Thus the Lorenz equations can stand alone without their historical association with the Rayleigh–Bernard problem. However, it is still legitimate to ask the question in what sense and to what degree the Lorenz equations represent a valid description of the original Rayleigh–Bernard problem. In a larger context, since most physical problems are formulated in terms of partial differential equations, we may also ask in what sense the approximation with a finite number of modes by truncation or other means can reveal the behavior of the original problem. Numerical solutions of the Rayleigh–Bernard problem[3] have shown that some of the salient features of the Lorenz system disappear when more modes are retained than the number Lorenz retained. It seems also that the occurrence of certain phenomena depends on the number and which modes are retained.

Now the Rayleigh–Bernard problem Lorenz studied originally is a system of two highly nonlinear partial differential equations of three independent variables (x,z,t). One equation is first order in t and fourth order in (x,z), while the other is first order in t and second order in (x,z). Except for the linearlized problem, it is difficult to analyze those equations in any way besides the numerical computation. Therefore, in order to clarify the larger question raised above, it would be useful if we could have a manageable partial differential equation with an interesting reduced system of ordinary differential equations. Then we shall be able to make a comparative study of the original partial differential equation and its reduced systems.

With the celebrated Lorenz equation in mind, we are thus led to construct such a system of partial differential equations. We shall try to construct a system that is as simple as can be found. Most importantly, if we use a truncation scheme similar to (1) and (2), as envisaged by Lorenz, we should again obtain the same Lorenz equations. Then the rich contents of the Lorenz system are presumably also contained in the solution of these partial differential equations.

In the following, we shall first construct such a system of partial differential equations, then various properties of the partial differential equations will be studied.

II. THE PARTIAL DIFFERENTIAL EQUATIONS

Consider the following system of partial differential equations for $\psi(x,t)$ and $\theta(x,t)$:

$$\begin{aligned}\frac{\partial\psi}{\partial t} &= -\sigma\psi - \sigma\frac{\partial\theta}{\partial x}\,,\\ \frac{\partial\theta}{\partial t} &= -(1-\beta)\theta + \beta\frac{\partial^2\theta}{\partial x^2} + r\frac{\partial\psi}{\partial x} + 2\psi\frac{\partial\theta}{\partial x}\,,\end{aligned} \tag{G_β}$$

where σ and r are positive parameters, and $0\leqslant\beta\leqslant 1$.

Let us now take

$$\psi = (1/\sqrt{2})X(t)\sin x\,, \tag{3}$$

$$\theta = (1/\sqrt{2})\,Y(t)\cos x + \tfrac{1}{2}\,Z(t)\cos 2x\,, \tag{4}$$

and substitute (3) and (4) into the system (G_β). If we retain only the coefficients of $\sin x$, $\cos x$, and $\cos 2x$, it is

 0022-2488/87/071589-09$02.50

straightforward to verify that we obtain the Lorenz system (L_β).

The variable θ can be eliminated from the system (G_β) by differentiating the second equation with respect to x, and we obtain

$$\frac{\partial^2\psi}{\partial t^2} + \sigma(r-\beta)\frac{\partial^2\psi}{\partial x^2} - 2\psi\frac{\partial^2\psi}{\partial x\,\partial t} - \beta\frac{\partial^3\psi}{\partial x^2\,\partial t} + (\sigma+1-\beta)\frac{\partial\psi}{\partial t} - 2\frac{\partial\psi}{\partial x}\frac{\partial\psi}{\partial t} - 4\sigma\psi\frac{\partial\psi}{\partial x} + \sigma(1-\beta)\psi = 0. \qquad (H_\beta)$$

The partial differential equation (H_β) is a third-order equation. However, for the case $\beta = 0$, Eq. (H_0) is a second-order mixed-type equation. The third-order term $-\beta(\partial^3\psi/\partial x^2\,\partial t)$ has the regularization effect of smoothing things out, while the underlying feature may be largely represented through the mixed-type behavior. For the case of $\beta = 1$, Eq. (H_1) can also be simplified in some cases.

It is possible to detect some traces of the Rayleigh–Bernard system in (G_β). But (G_β) is a system of partial differential equations of only two independent variables (x,t) and the order of the equations are also much lower than the Rayleigh–Bernard system. It is a much simpler system, yet is still a rich system, at least as rich as the Lorenz system.

III. THE LINEARIZED PROBLEM

It may be recalled that the Lorenz system (L) has the following equilibrium points:

$$(0)\colon X = Y = Z = 0.$$

$$(C_1,C_2)\colon X = Y = \pm[(1+3\beta)(r-1)]^{1/2}, \quad Z = r-1.$$

The equilibrium point (0) is stable for $r<1$, and becomes unstable for $r>1$. The equilibrium points (C_1,C_2) emerge only for $r>1$. The linearized problem of the system (G_β) or Eq. (H_β) would thus correspond to the behavior of the solution in the neighborhood of (0) for the Lorenz system (L).

Take Eq. (H_β). The linearized equation is

$$\frac{\partial^2\psi}{\partial t^2} + \sigma(r-\beta)\frac{\partial^2\psi}{\partial x^2} - \beta\frac{\partial^3\psi}{\partial x^2\,\partial t} + (\sigma+1-\beta)\frac{\partial\psi}{\partial t} + \sigma(1-\beta)\psi = 0. \qquad (5)$$

Let us use the method of normal modes and take $\psi = w(t)e^{ikx}$, thus (5) becomes

$$\frac{d^2w}{dt^2} + B\frac{dw}{dt} + Cw = 0, \qquad (6)$$

where

$$B = \sigma + (1-\beta) + \beta k^2 \qquad (7)$$

and

$$C = \sigma[(1-\beta) - (r-\beta)k^2]. \qquad (8)$$

Thus

$$w = a_1e^{\nu_1 t} + a_2e^{\nu_2 t}, \qquad (9)$$

and

$$\nu_{1,2} = \tfrac{1}{2}[-B \pm (B^2-4C)^{1/2}]. \qquad (10)$$

Now since σ and r are positive and $0\leqslant\beta\leqslant 1$, it is clear that $B>0$. Thus ν_1 and ν_2 both have negative real parts if and only if $C\geqslant 0$. In other words, the problem is linearly stable if and only if

$$r \leqslant \beta + (1-\beta)/k^2. \qquad (11)$$

Therefore the system is linearly stable if $r\leqslant\beta$, whatever the value of k is. The larger the value of k, the more unstable is the system. When $k=1$, which corresponds to the Lorenz problem, the criterion (11) becomes

$$r \leqslant 1. \qquad (12)$$

The criterion (12) agrees with the stability criterion for the equilibrium point (0) of the Lorenz system.

IV. THE STEADY STATES

The equilibrium points of Lorenz system (L) would correspond to the steady states for Eqs. (H_β) or (G_β). Let us consider the time-independent solutions of (H_β). Take $\psi = \psi(x)$, then (H_β) becomes

$$\sigma(r-\beta)\frac{d^2\psi}{dx^2} - 4\sigma\psi\frac{d\psi}{dx} + \sigma(1-\beta)\psi = 0. \qquad (13)$$

Equation (13) can be rewritten as two coupled first-order equations. Introduce $\phi(x)$ by

$$\frac{d\psi}{dx} = \phi. \qquad (14)$$

Thus (13) becomes

$$\frac{d\phi}{dx} = \frac{1}{(r-\beta)}[4\varphi - (1-\beta)]\psi. \qquad (15)$$

The system (14) and (15) can be readily integrated, since

$$\frac{d\psi}{d\phi} = \frac{(r-\beta)\varphi}{[4\varphi-(1-\beta)]\psi}, \qquad (16)$$

which is separable, and we obtain

$$\left(\frac{2}{r-\beta}\right)\psi^2 - \varphi - \frac{(1-\beta)}{4}\ln\left|1-\frac{4\phi}{1-\beta}\right| = A. \qquad (17)$$

It may be seen from (17) or the system (14) and (15) that a particular solution is

$$\varphi = (1-\beta)/4. \qquad (18)$$

For large $|\varphi|$, the orbit as given by (17) is a parabola given approximately by

$$[2/(r-\beta)]\psi^2 - \varphi = \text{const}. \qquad (19)$$

For $r<\beta$, there is no closed orbit, and the orbits are schematically shown in Fig. 1.

For $r>\beta$, the orbits for $\varphi<(1-\beta)/4$ are closed, while those for $\varphi>(1-\beta)/4$ are not. They are schematically shown in Fig. 2.

The qualitative features can also be seen by taking a closer look at the equilibrium point $\psi=\varphi=0$. For small $|\varphi|$, if we expand the log function in (17) in terms of power series in φ, Eq. (17) is given approximately by

$$[2/(r-\beta)]\psi^2 + [2/(1-\beta)]\varphi^2 = A. \qquad (20)$$

Thus the equilibrium point $\varphi=\psi=0$ is a center for $r>\beta$ and a saddle point for $r<\beta$.

The closed orbits represent the periodic solutions for

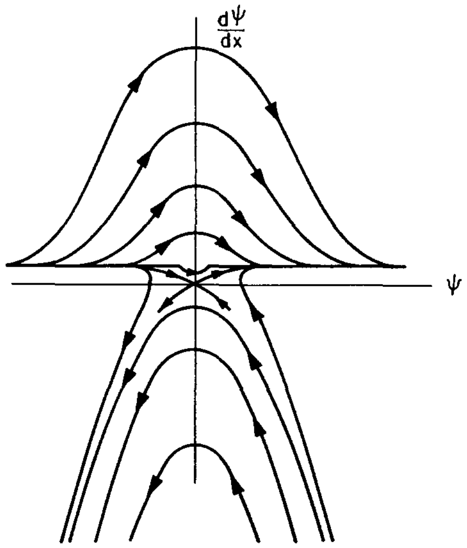

FIG. 1. Schematic orbits of the steady states for $r<\beta$.

$\psi(x)$, which in turn correspond to equilibrium points of the Lorenz system. Combined with the results from Sec. III, we see that the system (H_β) is linearly stable for $r<\beta$, and linearly unstable for $r>\beta$. However, for $r<\beta$, there is no bounded nontrivial steady state solution. Bounded spatially periodic steady state solutions emerge only when r exceeds β. So there is a correspondence between these qualitative features in system (L) and system (H_β).

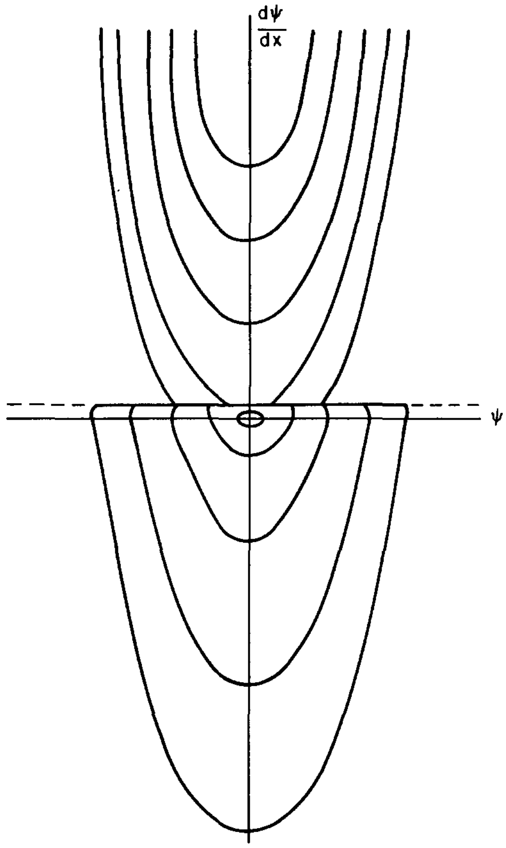

FIG. 2. Schematic orbits of the steady states for $r>\beta$.

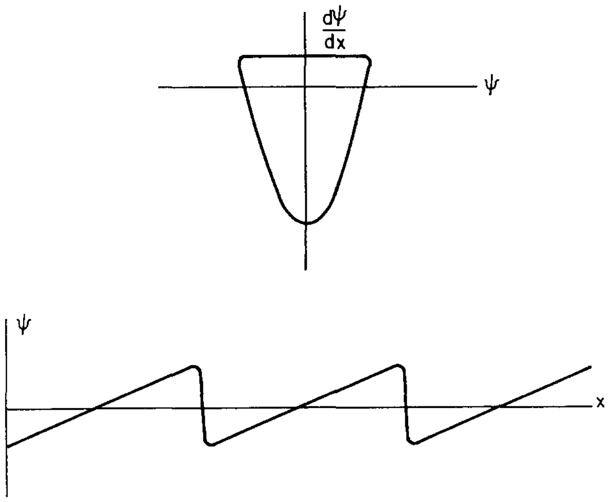

FIG. 3. Schematic solution of a steady state for $r>\beta$.

Making use of (17), we can obtain $\varphi(x)$ and $\psi(x)$ by direct integration. Denote

$$\Phi = 4\varphi/(1-\beta)\,,$$

then Eq. (17) can be rewritten as

$$\psi^2 = [(1-\beta)/8(r-\beta)] \times [4A/(1-\beta) + \Phi + \ln|1-\Phi|]\,. \tag{21}$$

Thus we obtain from (15) and (21) that

$$\frac{d\Phi}{dx} = -\left[\frac{2(1-\beta)}{(r-\beta)}\right]^{1/2}(1-\Phi) \times\left(\frac{4A}{1-\beta} + \Phi + \ln|1-\Phi|\right)^{1/2}. \tag{22}$$

Equation (22) can be directly integrated. Let us concentrate our attention on periodic solutions, i.e., $r>\beta$ and $\Phi<1$. A typical solution is represented in Fig. 3. Let ψ_M and φ_M be the maximum value of $|\psi|$ and $|\varphi|$, respectively. Then we see from (17) and (20) that

$$\psi_M \sim [\tfrac{1}{2}(r-\beta)A]^{1/2},$$

$$\phi_M \sim \begin{cases} [\tfrac{1}{2}(1-\beta)A]^{1/2}, & \text{for } A \text{ small}, \\ A, & \text{for } A \text{ large}. \end{cases}$$

For $A \ll 1$, we see from Eq. (20) that the (φ,ψ) orbit is an ellipse and the solutions for ψ and φ are

$$\psi = [\tfrac{1}{2}(r-\beta)A]^{1/2}\sin([(1-\beta)/(r-\beta)]^{1/2}x)\,, \tag{23}$$

$$\varphi = [\tfrac{1}{2}(1-\beta)A]^{1/2}\cos([(1-\beta)/(r-\beta)]^{1/2}x)\,. \tag{24}$$

The wavelength of these periodic solutions λ is given by

$$\lambda = 2\pi[(r-\beta)/(1-\beta)]^{1/2}\,. \tag{25}$$

For the Lorenz system, we have $\lambda = 2\pi$. Thus only for $r \approx 1$ can a close correspondence between the steady state solutions of the system (H_β) and the equilibrium points (C_1,C_2) of the system (L) be established.

As A increases, the wavelength λ increases also. For $A \gg 1$, it may be established that

$$\lambda = O(\psi_M) = O([(r-\beta)A]^{1/2})\,. \tag{26}$$

Thus for large r, the wavelength of the periodic solutions will be proportional to $r^{1/2}$, and cannot be maintained around the value of 2π, which is implied in the Lorenz system.

V. STABILITY OF THE STEADY STATES

Let us now investigate whether the steady states, in particular the spatially periodic steady state, are stable. Take $\psi_0(x)$ to be the steady state solution, we may discuss the stability problem from two approaches.

(i) The regular perturbation approach: Write

$$\psi(x,t) = \psi_0(x) + \psi_1(x,t) . \tag{27}$$

Substitute (27) in (H_β). Since $\psi_0(x)$ satisfies Eq. (13), we obtain, after retaining only terms linear in ψ_1, the following equation:

$$\frac{\partial^2\psi_1}{\partial t^2} + \sigma(r-\beta)\frac{\partial^2\psi_1}{\partial x^2} - 2\psi_0\frac{\partial^2\psi_1}{\partial x\,\partial t} - \beta\frac{\partial^3\psi_1}{\partial x^2\,\partial t} + (\sigma+1-\beta)\frac{\partial\psi_1}{\partial t} - 2\frac{\partial\psi_0}{\partial x}\frac{\partial\psi_1}{\partial t} - 4\sigma\psi_0\frac{\partial\psi_1}{\partial x} - 4\sigma\frac{\partial\psi_0}{\partial x}\psi_1 + \sigma(1-\beta)\psi_1 = 0 . \tag{28}$$

Although Eq. (28) is a linear equation, it is not easy to solve it analytically. Since $\psi_0(x)$ is a periodic function with period λ, we may also look for solutions that are spatially periodic with period λ. The term $\beta(\partial^3\psi_1/\partial x^2\,\partial t)$ has a tendency to smooth out the solution. If we neglect this term, Eq. (28) is either elliptic or hyperbolic depending on the relative magnitude of $\sigma(r-\beta)$ and ψ_0^2, because the characteristics of the equation are given by

$$\frac{dx}{dt} = \psi_0 \pm [\psi_0^2 - \sigma(r-\beta)] . \tag{29}$$

Thus the equation is elliptic where $\psi_0^2 < \sigma(r-\beta)$, and hyperbolic where $\psi_0^2 > \sigma(r-\beta)$. Ellipticity is usually associated with instability. Since the maximum of $|\psi_0|$ is roughly $[\frac{1}{2}(r-\beta)A]^{1/2}$ as indicated by (17), thus the equation is elliptic when A is small. On the other hand, lower-order terms have a damping effect. Hence the ellipticity needs to overcome the damping mechanisms to cause instability. But it is perhaps reasonable to state that for large $\sigma(r-\beta)$ and small A, the system is unstable. The term $\beta(\partial^3\psi_1/\partial x^2\,\partial t)$ may smooth things out eventually. But perhaps it will not alter the general qualitative behavior in the short run.

For large A, the maximum of ψ_0^2 will be larger than $\sigma(r-\beta)$. Thus in much of the region the equation is hyperbolic, which will imply stability. But there are still regions of ellipticity in the neighborhood of $\psi_0(x)=0$. Numerical studies are needed to study the behavior of the solution of this mixed type equation.

(ii) The multiple scale expansion approach: The spatially periodic steady state solution ψ_0 can be written as $\psi_0(x;A)$, where A is a measure of amplitude. We can look for a solution of the type $\psi_0\{x/\lambda(t);A(t)\}$, where $\lambda(t)$ and $A(t)$ are slowly varying function of t. Whether $\lambda(t)$ and $A(t)$ remain in the neighborhood of constant values of λ and A as t increases determines the stability of the spatially periodic steady state.

The analysis can be carried out by a multiple scale expansion. The details are presented in the Appendix. The stability criterion can be expressed in the form of Eq. (A14),

$$\frac{d^2A}{dt^2} + G(A;r,\beta,\sigma)\frac{dA}{dt} = 0 ,$$

where G can be computed in terms of $\psi_0(x;A)$.

For finite A, since $\psi_0(x;A)$ is not a simple function, numerical computation is needed to calculate G. However, for small A, since ψ_0 is given by (23), it may be shown that the spatially periodic steady state is stable.

VI. THE SPATIALLY PERIODIC SOLUTIONS

It is reasonable to expect that the connection between the partial differential equations (G_β) or (H_β) and the Lorenz system is to be found in spatially periodic solutions. Let the period be λ. Denote

$$\langle\psi\rangle = \int_\lambda \psi(x,t)dx . \tag{30}$$

Other averages over the period λ, $\langle\theta\rangle$, $\langle\psi^2\rangle$, and so on, can be defined likewise. Let us now consider the case that both ψ and θ and their derivatives are all periodic in x with period λ. Let us integrate the first equation of (G_β) with respect to x over the period λ. We then obtain

$$\frac{d}{dt}\langle\psi\rangle = -\sigma\langle\psi\rangle . \tag{31}$$

If we multiply the first equation of (G_β) by ($2\psi/\sigma$), add to the second equation of (G_β), and then integrate with respect to x over the period λ, we obtain

$$\frac{d}{dt}[\langle\psi^2\rangle + \sigma\langle\theta\rangle] = -2\sigma\Big[\langle\psi^2\rangle + \frac{(1-\beta)}{2}\langle\theta\rangle\Big] . \tag{32}$$

From (31), we see that $\langle\psi\rangle \to 0$ as $t\to\infty$. Thus asymptotically not only ψ is periodic in x, but $\int_x \psi\,dx$ is also periodic in x. However, it does not imply that $|\psi|\to 0$ as $t\to\infty$.

To discuss the implication of (32), let us denote

$$p = \langle\psi^2\rangle + \sigma\langle\theta\rangle, \quad q = \langle\psi\rangle + [(1-\beta)/2]\,\langle\theta\rangle .$$

Thus (32) can be rewritten as

$$\frac{dp}{dt} = -2\sigma q .$$

In Figs. 4 and 5, we have plotted lines of constant p and q in the $\langle\psi^2\rangle - \langle\theta\rangle$ plane for the cases $\sigma > (1-\beta)/2$ and $\sigma < (1-\beta)/2$, respectively. At any point where $q>0$, p has a tendency to decrease; while at any point where $q<0$, p has a tendency to increase. It is not clear how p decreases or increases. However, when the state reaches a point on $q=0$, it tends to stay there. Take the case of $\sigma > (1-\beta)/2$ (Fig. 4). Consider any point a for which $q<0$. The motion tends to remain in the region D, which is bounded by the lines $p=p(a)$, $\langle\psi^2\rangle = 0$, and $q=0$. Thus for subsequent motions we have

$$|\langle\theta\rangle| \leqslant (1/\sigma)|p(a)| , \tag{33}$$

$$\langle\psi^2\rangle \leqslant |[(1-\beta)/\{2\sigma-(1-\beta)\}]p(a)| . \tag{34}$$

Thus for $\sigma > (1-\beta)/2$ any motion initiated with $q<0$ tends to be bounded all the time with the bound given by (34).

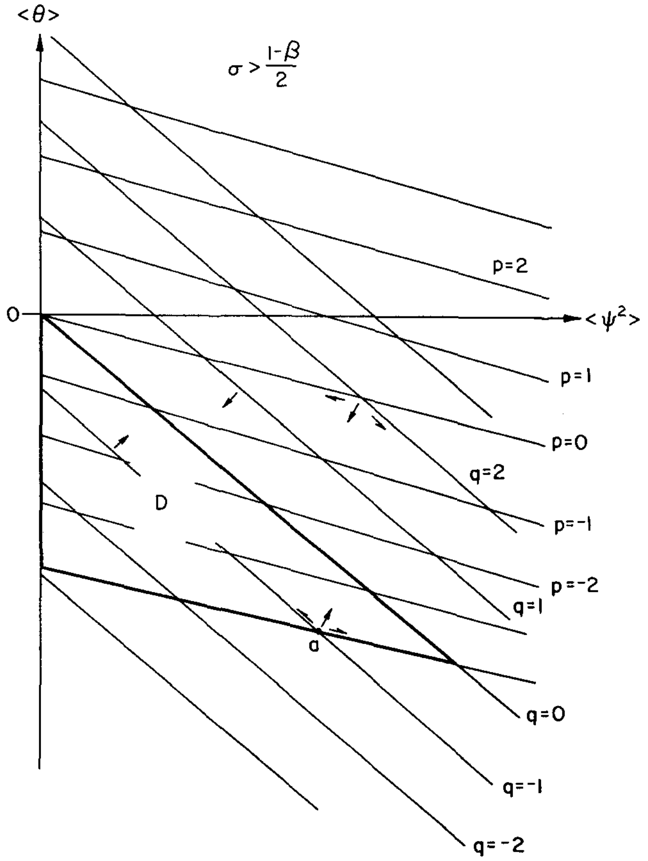

FIG. 4. Tendency of motion of the system in the $(\langle\theta\rangle, \langle\psi^2\rangle)$ plane for $\sigma > (1-\beta)/2$.

The same result is valid for any motion initiated with $q > 0$ when $\sigma < (1-\beta)/2$, as represented in Fig. 5.

The state $q = 0$, i.e., $\langle\psi^2\rangle + [(1-\beta)/2]\langle\theta\rangle = 0$, if ever reached, will be the asymptotic state for large t. It may be mentioned that for steady states

$$\varphi = [1/(r-\beta)]\,[2\psi^2 + (1-\beta)\theta]\,. \tag{35}$$

Thus $q = 0$ is equivalent to $\langle\varphi\rangle = 0$, which is indeed the case for periodic steady states.

VII. ELLIPTICITY AND HYPERBOLICITY

Following the original study of Lorenz, most numerical studies of the Lorenz equations (L) set the parameters $\sigma = 10$ and $\beta = \frac{8}{3}$, while varying the parameter r. Except for the requirement that $\sigma > 2 + 3\beta$, most of the qualitative behaviors of the Lorenz system seem to be insensitive to the variation of β or σ. However, for Eq. (H_β), it is clear that the problem is somewhat qualitatively different when $\beta = 0$ or $\beta = 1$. When $\beta = 1$, from Eqs. (13)–(17), it may be seen that no bounded periodic steady state solution exists. On the other hand, when $\beta = 0$, the order of Eq. (H_β) is reduced by 1. The term $\partial^3\psi/2x^2\,\partial t$ in (H_β) is a "good" term, whose presence can smooth out irregularities that may arise otherwise. Still it may be illuminating to consider the case that $\beta \ll 1$ or even when $\beta = 0$, since the underlying irregularities, if they exist, may reveal some intrinsic properties of the system.

If we set $\beta = 0$, then Eq. (H_β) is of the mixed type. The characteristics of Eq. (H_0) are given by

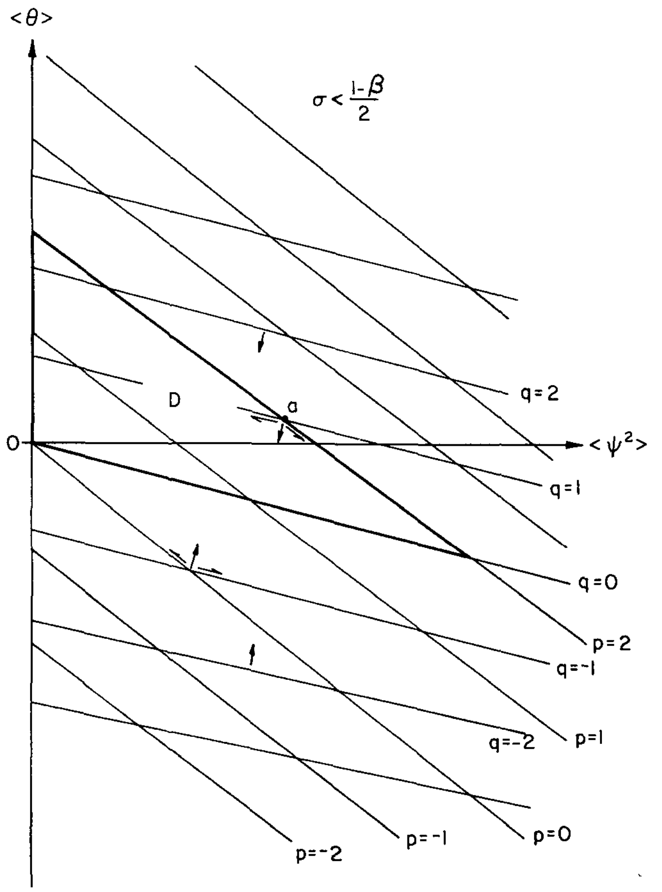

FIG. 5. Tendency of motion of the system in the $(\langle\theta\rangle, \langle\psi^2\rangle)$ plane for $\sigma < (1-\beta)/2$.

$$\frac{dx}{dt} = -\psi \pm [\psi^2 - \sigma r]^{1/2}\,. \tag{36}$$

Thus the equation is elliptic when $\psi^2 < \sigma r$ and hyperbolic when $\psi^2 > \sigma r$. When r is large or when ψ is small, Eq. (H_0) is elliptic. Ellipticity is associated with instability of the system. More precisely, for initial value problems, solutions of the elliptic equations tend to depend sensitively on initial data. The instability due to ellipticity will make ψ^2 grow with t. But when ψ^2 exceeds the value of σr, Eq. (H_0) becomes hyperbolic. Hyperbolicity implies stability of the system. Therefore the initial unstable growth will be arrested. The solution will start to exhibit some wavelike features. However, the system (H_0) is dissipative. The dissipative mechanisms in the lower-order terms in Eq. (H_0) will tend to diminish the magnitude of ψ^2, and drive the system back to the elliptic regime. When the equation is elliptic, the instability mechanism will operate again and raise the magnitude of ψ^2 over the value of σr, thus push the system to the hyperbolic regime. This type of switching back and forth from ellipticity to hyperbolicity may be what corresponds to the chaotic behavior of the Lorenz system. The sensitive dependence on initial data inherent in the elliptic equations is certainly consistent with the trademark of the chaos. The boundedness of the solution as established in Sec. IV may indicate the existence of some strange attractors. What the meanings of these concepts are in the context of partial differential equations is still not clear.

Numerical studies of Eq. (H_β) should shed light on these intriguing questions. To carry out numerical studies, it may be noted that the boundary line between ellipticity and hyperbolicity for Eq. (H_0) is given by $\psi = (\sigma r)^{1/2}$. It can also be shown that on the hyperbolic side the characteristics form cusps at the boundary line. In the hyperbolic regime, shocks may appear. Thus from a numerical point of view, it is better to work with Eq. (H_β) with $\beta > 0$. Whether finite values of β would add too much diffusion is also a point worth noting.

VIII. A REDUCED SYSTEM AND HODOGRAPH TRANSFORMATION

The method of hodograph transformation is sometimes useful in dealing with mixed type partial differential equations. Let us consider a reduced system by dropping the two damping terms in (G_0):

$$\frac{\partial\psi}{\partial t} = -\sigma\frac{\partial\theta}{\partial x}, \quad \frac{\partial\theta}{\partial t} = r\frac{\partial\psi}{\partial x} + 2\psi\frac{\partial\theta}{\partial x}. \qquad (G_R)$$

The corresponding single partial differential equation is

$$\frac{\partial^2\psi}{\partial t^2} + \sigma r\frac{\partial^2\psi}{\partial x^2} - 2\psi\frac{\partial^2\psi}{\partial x\,\partial t} - 2\frac{\partial\psi}{\partial x}\frac{\partial\psi}{\partial t} = 0, \qquad (H_R)$$

and the corresponding reduced Lorenz system is

$$\frac{dX}{dt} = \sigma Y, \quad \frac{dY}{dt} = rX - XZ, \quad \frac{dZ}{dt} = XY. \qquad (L_R)$$

Equation (H_R) is also of the mixed type. But the lower-order damping terms are not present. System (L_R) has the following first integrals:

$$2\sigma Z - X^2 = D \qquad (37)$$

and

$$(Z-r)^2 + Y^2 = B^2. \qquad (38)$$

These integrals define a two-parameter family of periodic solutions in t.

To solve the system (G_R), let us introduce the hodograph transformation by interchanging the roles of (x,t) and (ψ,θ) as independent and dependent variables. Denote the Jacobian of the transformation:

$$J = \frac{\partial(x,t)}{\partial(\psi,\theta)},$$

we obtain

$$\frac{\partial\psi}{\partial x} = \frac{1}{J}\frac{\partial t}{\partial\theta}, \quad \frac{\partial\theta}{\partial x} = -\frac{1}{J}\frac{\partial t}{\partial\psi},$$

$$\frac{\partial\psi}{\partial t} = -\frac{1}{J}\frac{\partial x}{\partial\theta}, \quad \frac{\partial\theta}{\partial t} = \frac{1}{J}\frac{\partial x}{\partial\psi}.$$

Thus system (G_R) becomes

$$\frac{\partial x}{\partial\theta} = -\sigma\frac{\partial t}{\partial\psi}, \quad \frac{\partial x}{\partial\psi} = r\frac{\partial t}{\partial\theta} - 2\psi\frac{\partial t}{\partial\psi}. \qquad (G_h)$$

If we eliminate t, then we obtain

$$\frac{\partial^2 x}{\partial\psi^2} + \frac{r}{\sigma}\frac{\partial^2 x}{\partial\theta^2} - \frac{2}{\sigma}\psi\frac{\partial^2 x}{\partial\theta\,\partial\psi} - \frac{2}{\sigma}\frac{\partial x}{\partial\theta} = 0. \qquad (H_h)$$

Equations (G_h) and (H_h) are both linear, but they are still of the mixed type. The boundary between the region of ellipticity and hyperbolicity is again $\psi^2 = \sigma r$.

There are particular solutions of the form

$$x(\psi,\theta) = f(\psi)e^{a\theta}. \qquad (39)$$

If we substitute (39) in (H_h), we obtain

$$\frac{d^2f}{d\psi^2} - \frac{2a}{\sigma}\psi\frac{df}{d\psi} + \left(\frac{ra^2}{\sigma} - \frac{2a}{\sigma}\right)f = 0. \qquad (40)$$

Denote

$$\omega = (a/\sigma)\psi \qquad (41)$$

and

$$2n = \sigma r - 2\sigma/a,$$

or

$$a = 2\sigma/(\sigma r - 2n). \qquad (42)$$

Then Eq. (40) becomes

$$\frac{d^2f}{d\omega^2} - 2\omega\frac{df}{d\omega} + 2nf = 0, \qquad (43)$$

which is the Hermite equation. When n is an integer, a solution is the Hermite polynomial of order nH_n. Therefore a particular solution of (H_h) is

$$x_n = H_n(2\psi/(\sigma r - 2n))e^{2\sigma\theta/(\sigma r - 2n)}. \qquad (44)$$

Any linear combination of x_n's is also a solution. Using (G_h), we may obtain particular solutions for t:

$$t_n = -\frac{2}{\sigma r - 2n}e^{2\sigma\theta/(\sigma r - 2n)}\int^\psi H_n\left(\frac{2\psi}{\sigma r - 2n}\right)d\psi. \qquad (45)$$

These particular solutions should be helpful to guide the numerical studies of the problem especially in the neighborhood of the transition between elliptic and hyperbolic regimes.

IX. ASYMPTOTIC STATES FOR LARGE r

For the Lorenz system (L), Robbins[4] has found a stable periodic solution as the parameter r becomes very large. Let us introduce a small parameter ϵ and set

$$t = \epsilon\tau, \quad r = 1/\epsilon^2. \qquad (46)$$

Now let

$$X(t,\epsilon) = (1/\epsilon)\,[X_0(\tau) + \epsilon X_1(\tau) + \cdots],$$

$$Y(t,\epsilon) = (1/\epsilon^2)\,[Y_0(\tau) + \epsilon Y_1(\tau) + \cdots],$$

$$Z(t,\epsilon) = (1/\epsilon^2)\,[Z_0(\tau) + 1 + \epsilon Z_1(\tau) + \cdots],$$

then the zeroth order of the Lorenz system becomes

$$\frac{dX_0}{d\tau} = \sigma Y_0, \quad \frac{dY_0}{dv} = -X_0Z_0, \quad \frac{dZ_0}{d\tau} = X_0Y_0. \qquad (L_A)$$

We may note that (L_A) is really the same as (L_R) in the last section, if we let $Z - r = Z_0$. By carrying out the next-order expansions, Robbins has shown that a stable periodic solution exists for finite r, and as r decreases, period-doubling bifurcations appear.

Let us take the same scaling for Eq. (H_β) and let

$$\psi(x,t,\epsilon) = (1/\epsilon)[\psi_0(x,\tau) + \epsilon\psi_1(x,\tau) + \cdots]. \qquad (47)$$

Then the zeroth- and first-order equations of (H_β) are

$$\frac{\partial^2\psi_0}{\partial\tau^2} + \sigma\frac{\partial^2\psi_0}{\partial x^2} - 2\psi_0\frac{\partial^2\psi_0}{\partial x\,\partial\tau} - 2\frac{\partial\psi_0}{\partial x}\frac{\partial\psi_0}{\partial\tau} = 0, \qquad (48)$$

$$\frac{\partial^2\psi_1}{\partial\tau^2}+\sigma\frac{\partial^2\psi_1}{\partial x^2}-2\psi_0\frac{\partial^2\psi_1}{\partial x\,\partial\tau}-2\frac{\partial^2\psi_0}{\partial x\,\partial\tau}\psi_1$$
$$-2\frac{\partial\psi_0}{\partial x}\frac{\partial\psi_1}{\partial\tau}-2\frac{\partial\psi_0}{\partial\tau}\frac{\partial\psi_1}{\partial x}$$
$$=\beta\frac{\partial^3\psi_0}{\partial x^2\,\partial\tau}-(\sigma+1-\beta)\frac{\partial\psi_0}{\partial\tau}+4\sigma\psi_0\frac{\partial\psi_0}{\partial x},\dots. \quad (49)$$

Equation (48) is the same as Eq. (H_R) of the last section when $r=1$. In the last section, we have explored some aspects of this equation. Now to make connection with the periodic solutions of (L_A), we should look for solutions periodic both in x and t for Eq. (48). It is not clear whether such solutions exist. Even if there are such doubly periodic solutions, what would be the corresponding period-doubling bifurcation in the context of partial differential equations is also not clear. Much work is still needed to clarify these issues.

Now if, instead of (46), we introduce a different scaling and let

$$t=\epsilon\tau, \quad r=1/\epsilon, \quad (50)$$

and further, let

$$X(t,\epsilon)=(1/\epsilon)\,[X_0(\tau)+\epsilon X_1(\tau)+\cdots],$$
$$Y(t,\epsilon)=1/\epsilon^2\,[Y_0(\tau)+\epsilon Y_1(\tau)+\cdots],$$
$$Z(t,\epsilon)=1/\epsilon^2\,[Z_0(\tau)+\epsilon Z_1(\tau)+\cdots],$$

then the zeroth order or the Lorenz system again is (L_A). However, if we now apply the scaling (50) and the expansion (47) to Eq. (H_β), the zeroth-order equation becomes

$$\frac{\partial^2\psi_0}{\partial\tau^2}-2\psi_0\frac{\partial^2\psi_0}{\partial x\,\partial t}-2\frac{\partial\psi_0}{2x}\frac{\partial\psi_0}{\partial\tau}=0. \quad (51)$$

Equation (51) is different from Eq. (48). While (48) is a mixed-type equation, (50) is a hyperbolic equation. In fact, Eq. (48) can be integrated once to give

$$\frac{\partial\psi_0}{\partial\tau}-2\psi_0\frac{\partial\psi_0}{\partial x}=C(x). \quad (52)$$

To make connection with the Lorenz system, the function $C(x)$ in (52) should be a periodic function. Equation (52) has been investigated recently by Salas *et al.*[5] They studied also specifically the case that $C(x)=\sin 2x$, and found multiple steady states for characteristic initial value problems. Whether there are solutions periodic both in x and t is again not clear.

If we use another different set of scaling and let

$$t=\epsilon^2 t', \quad x=\epsilon x', \quad r=1/\epsilon, \quad \psi=\psi'/\epsilon+\cdots,$$

then the leading-order equation of (H_β) becomes

$$\frac{\partial^2\psi'}{\partial t'^2}-2\psi'\frac{\partial^2\psi'}{\partial x'\,\partial t'}-\beta\frac{\partial^3\psi'}{\partial x'^2\partial t'}-2\frac{\partial\psi'}{\partial x'}\frac{\partial\psi'}{\partial t'}=0. \quad (53)$$

Equation (53) can be integrated once to yield

$$\frac{\partial\psi'}{\partial t'}-2\psi'\frac{\partial\psi'}{\partial x'}-\beta\frac{\partial^2\psi'}{\partial x'^2}=C(x). \quad (54)$$

The homogeneous equation of (54) is the well-known Burgers equation. This equation has recently been investigated by Kreiss and Kreiss.[6]

X. THE TRAVELING WAVE SOLUTIONS

Traveling wave solutions are solutions such that

$$\psi(x,t)=\psi(\eta), \quad \theta(x,t)=\theta(\eta), \quad (55)$$

and

$$\eta=x-ct, \quad (56)$$

where c is a real constant. If there exist periodic traveling wave solution, then the solutions will be periodic in both x and t.

When we substitute (55) in (G_β), we obtain

$$c\frac{d\psi}{d\eta}=\sigma\psi+\sigma\frac{d\theta}{d\eta}, \quad (57)$$

$$-c\frac{d\theta}{d\eta}=-1(-\beta)\theta+\beta\frac{d^2\theta}{d\eta^2}+r\frac{d\psi}{d\eta}+2\psi\frac{d\theta}{d\eta}. \quad (58)$$

We may recast the system (57) and (58) into the following first-order system:

$$\frac{d\psi}{d\eta}=\frac{\sigma}{c}(\psi+\chi), \quad \frac{d\theta}{d\eta}=\chi,$$
$$\beta\frac{d\chi}{d\eta}=-\frac{\sigma r}{c}\psi+(1-\beta)\theta-\left(\frac{\theta r}{c}+c\right)\chi-2\psi\chi. \quad (G_T)$$

The system (G_T) has only one equilibrium point at $\psi=\theta=\chi=0$. To investigate the stability about this equilibrium point, consider the linearized equation, and take solutions proprotional to $e^{\nu\eta}$. Then we found that

$$F(\nu)\equiv\beta\nu^3+[(\sigma/c)(r-\beta)+c]\nu^2$$
$$-[(1-\beta)+\sigma]\nu+(\sigma/c)(1-\beta)=0. \quad (59)$$

Again, we are only interested in cases where σ and r are positive, $0\leqslant\beta\leqslant 1$, and $r>\beta$. Take $c>0$, then after investigating the behavior of $F(\nu)$, $F'(\nu)$, and $F''(\nu)$, it is readily concluded that for the three characteristic roots of (59), say ν_1,ν_2,ν_3, we have

$$\nu_1<0, \quad Rl\nu_2>0, \quad Rl\nu_3>0. \quad (60)$$

Therefore, except for exceptional cases, the equilibrium point is unstable. When $\beta=0$ or $\beta=1$, the third-order system is reduced to a second-order system. For $\beta=0$, the characteristic condition (60) becomes

$$\nu_1<0, \quad \nu_2>0. \quad (61)$$

For the case $\beta=1$, then (60) becomes

$$\nu_1<0, \quad \nu_2=0, \quad \nu_3>0. \quad (62)$$

For $\beta=1$, the system (G_T) reduces to a second-order system for ψ and χ. Although there is another equilibrium point at $\psi=-\chi=-c/2$, which is stable, yet since $d\theta/d\eta=\chi$, it is not an equilibrium point for (ψ,θ,χ).

Some preliminary numerical computation failed to find any periodic solution for the system (G_T).

XI. DISCUSSIONS

We have constructed a system of partial differential equations (G_β) or the partial differential equation (H_β) which are among the simplest that contain Lorenz system (L) in some approximation. We have analyzed and discussed various aspects of these partial differential equations. Certain qualitative features of these partial differential equations correspond with those of the Lorenz system. Others are difficult to say without entensive numerical studies. A notable feature of these partial differential equations is that in the limit of $\beta \to 0$, the equations are of the mixed type. It is suggested that the switching back and forth from the ellipticity and hyperbolicity may correspond to the chaotic behavior of the reduced system of Lorenz. Again, detailed numerical studies are need to confirm this suggestion, and work is in progress towards this goal.

Qualitatively for $r > \beta$, Eq. (H_β) has those properties: When the amplitude is small, the system is not stable due to the ellipticity. For large amplitude, it becomes hyperbolic. The system has lower-order damping terms as t progresses. It also has spatially diffusive mechanism. The ellipticity tends to make the system unstable and the hyperbolicity to cause wavy behavior, while the damping and diffusion would smooth out the irregularities. These are indeed also qualitative features contained in the Lorenz system.

Given the Lorenz system, there is of course, no unique "simplest" parent partial differential equation. In fact, the system

$$\begin{aligned} \frac{\partial \psi}{\partial t} &= -\sigma\psi - \frac{\partial \theta}{\partial x}, \\ \frac{\partial \theta}{\partial t} &= -(1-\beta)\theta + \beta \frac{\partial^2 \theta}{\partial x^2} + r\frac{\partial \psi}{\partial x} + \frac{\partial}{\partial x}(\theta\psi), \end{aligned} \tag{K_β}$$

if we take

$$\psi = \sqrt{2}\, X(t) \sin x, \tag{63}$$

$$\theta = \sqrt{2}\, Y(t) \cos X + 2Z \cos 2x, \tag{64}$$

and use the same truncation scheme, will again lead to the Lorenz system (L). The system (K_β) may be simpler than (G_β) in some aspects, but may be less simple in other aspects. For one thing, it is difficult to obtain a single partial differential equation as simple as (H_β).

Finally, it may be pointed out that one can detect the qualitative similarities between the system (G_β) and the partial differential equations of the Rayleigh–Bernard problem. Physical problems can also be found that are described approximately by the system (G_β).

APPENDIX: MULTIPLE SCALE ANALYSIS OF STABILITY OF STEADY STATES

Let us write (H_β) as follows:

$$\begin{aligned} \sigma(r-\beta)\frac{\partial^2\psi}{\partial x^2} - 4\sigma\psi\frac{\partial\psi}{\partial x} + \sigma(1-\beta)\psi &= -\frac{1}{\epsilon}\frac{\partial^2\psi}{\partial t^2} + 2\psi\frac{\partial^2\psi}{\partial x\,\partial t} + \beta\frac{\partial^3\psi}{\partial x^2\,\partial t} \\ &\quad - (\sigma+1-\beta)\frac{\partial\psi}{\partial t} + 2\frac{\partial\psi}{\partial x}\frac{\partial\psi}{\partial t}, \end{aligned} \tag{A1}$$

where the order parameter ϵ is introduced artifically in the first term on the right-hand side. Let us assume the following expansion of ψ:

$$\psi = \psi_0(y,\tau) + \epsilon\psi_1(y,\tau) + \cdots, \tag{A2}$$

where

$$\tau = \epsilon t, \tag{A3}$$

and

$$y = k(\tau)x. \tag{A4}$$

Substituting (A2) into (A1), we obtain, to the successive orders of ϵ, that

$$O(1):\ \sigma(r-\beta)k^2\frac{\partial^2\psi_0}{\partial y^2} - 4\sigma k\,\psi_0\frac{\partial\psi_0}{\partial y} + \sigma(1-\beta)\psi_0 = 0, \tag{A5}$$

$$O(\epsilon):\ L[\psi_1] = F_1, \tag{A6}$$

$$\vdots,$$

where

$$L[\psi_1] \equiv k^2\frac{\partial^2\psi_1}{\partial y^2} - \frac{4k}{(r-\beta)}\psi_0\frac{\partial\psi_1}{\partial y} + \frac{[(1-\beta) - 4k(\partial\psi_0/\partial y)]}{(r-\beta)}\psi_1, \tag{A7}$$

and

$$F_1 \equiv \frac{1}{\sigma(r-\beta)}\left[-\frac{\partial^2\psi_0}{\partial\tau^2} + 2k\psi_0\frac{\partial^2\psi_0}{\partial y\,\partial\tau} + k^2\frac{\partial^3\psi_0}{\partial y^2\,\partial\tau} - (\sigma+1-\beta)\frac{\partial\psi_0}{\partial\tau} + 2k\frac{\partial\psi_0}{\partial y}\frac{\partial\psi_0}{\partial\tau}\right]. \tag{A8}$$

The solution of (A5), as given by (22), can be represented as

$$\psi_0 = \psi_0(y/k(\tau) - x_0(\tau), A(\tau)), \tag{A9}$$

and for the case of our interest with $r > \beta$, ψ_0 is periodic in y. Moreover, the function $\int^y \psi_0\,dy$ is also periodic in y. Here $k(\tau)$ is chosen so that the period in y is independent of τ. Thus $k = k(A)$ and the period can generally be set to be 2π.

It may be verified that $\partial\psi_0/\partial y$ and $\partial\psi_0/\partial A$ are two linearly independent solutions of the homogeneous equation $L[\psi_1] = 0$.

Let us introduce

$$\hat\psi_1 = \psi_1 \exp\left(-\frac{2}{(r-\beta)k}\int^y \psi_0\,dy\right), \tag{A10}$$

then (A6) can be written as

$$\hat L[\hat\psi_1] = F_1 \exp\left(-\frac{2}{(r-\beta)k}\int^y \psi_0\,dy\right), \tag{A11}$$

where

$$\hat L[\hat\psi_1] = \frac{\partial^2\hat\psi_1}{\partial y^2} + \frac{1}{(r-\beta)k^2}\left[(1-\beta) - 2k\frac{\partial\psi_0}{\partial y} - \frac{4}{(r-\beta)}\psi_0^2\right]\hat\psi_1. \tag{A12}$$

Following the method of Kuzmak–Luke,[7] the condition that $\hat\psi_1$ or ψ_1 will again be a periodic function of y with period 2π is

$$\int_0^{2\pi}\frac{\partial\psi_0}{\partial y}F_1\exp\left(-\frac{1}{(r-\beta)k}\int^y\psi_0\,dy\right)dy = 0. \tag{A13}$$

Let us take x_0 to be constant in (A9), then, since

$$\frac{\partial \psi_0}{\partial \tau} = \frac{\partial \psi_0}{\partial A}\frac{dA}{\partial \tau} + \frac{\partial \psi_0}{\partial k}\frac{dk}{d\tau},$$

etc. and $k = k(A)$, (A13) will be a differential equation of the form

$$\frac{d^2 A}{d\tau^2} + G(A;r,\beta,\sigma)\frac{dA}{d\tau} = 0. \tag{A14}$$

The coefficient G will determine whether A, a measure of the amplitude of ψ_0, will stay in the neighborhood of a constant value.

[1] E. N. Lorenz, "Deterministic non-periodic flows," J. Atmos. Sci. **20**, 130 (1963).
[2] C. Sparrow, *The Lorenz Equations: Bifurcations, Chaos, and Strange Attractors* (Springer, New York, 1982).
[3] J. H. Curry, J. R. Herring, J. Loncaric, and S. A. Orszag, "Order and Disorder in two- and three-dimensional Bérnard convection," J. Fluid Mech. **147**, 1 (1984).
[4] K. Robbins, "Periodic solutions and bifurcation structure at high r in the Lorenz system," SIAM J. Appl. Math. **36**, 457, (1979).
[5] M. D. Salas, S. Abarbanel, and D. Gottlieb, "Multiple steady states for characteristic initial value problems," Appl. Num. Math. **2**, 193 (1986).
[6] G. Kreiss and H. O. Kreiss, "Convergence to steady state of solutions of Burgers' equations," Appl. Num. Math. **2**, 161 (1986).
[7] J. Kevorkian and J. D. Cole, *Perturbation Methods in Applied Mathematics* (Springer, New York, 1981).

Quasiprimary composite fields and null vectors in critical Ising-type models

P. Furlan
Dipartimento di Fisica Teorica dell' Università di Trieste, International School for Advanced Studies (ISAS), Trieste, Italy and Istituto Nazionale di Fisica Nucleare (INFN), Sezione di Trieste, Italy

G. M. Sotkov and I. T. Todorov
Institute for Nuclear Research and Nuclear Energy, Sofia, Bulgaria[a] *and International School for Advanced Studies (ISAS), Trieste, Italy*

(Received 2 December 1986; accepted for publication 4 February 1987)

Operator product expansions (OPE's) are studied in unitary "minimal conformal models" on the circle. The presence of null vectors at a certain level (signaled by the the vanishing of Kac's determinant) leads to linear relations among SU (1,1)-covariant ("quasiprimary") fields at that level. In the Ising model this is shown to imply the proportionality of two composite tensor currents of dimension 6 and the vanishing of 4 among the first five composite fields expected to appear in the OPE of the stress energy tensor and the canonical Fermi field (of weight $\frac{1}{2}$). For the supersymmetric tricritical model it only implies the vanishing of the quasiprimary Fermi current of dimension $\frac{9}{2}$.

I. INTRODUCTION

In a paper that opened a new avenue in the study of two-dimensional critical models, Belavin, Polyakov, and Zamolodchikov[1] found an infinite series of models for which the conformally invariant Green's functions satisfy linear partial differential equations and can, in principle, be evaluated.[1,2] Such "minimal theories" (in the terminology of Ref. 1) correspond to Virasoro central charge,

$$c = c_m = 1 - 6/(m+2)(m+3), \tag{1.1}$$

with rational m, and an associated finite set of minimal weights derived from Kac's determinant formula.[3] It was conjectured[4] and later proved[5,6] that the subset of values (1.1) for positive integer m's and lowest weights (LW's),

$$\Delta_{r,t}(m) = \{[r(m+3) - t(m+2)]^2 - 1\} \times [4(m+2)(m+3)]^{-1}, \quad 1 \leqslant r \leqslant m+1, \quad 1 \leqslant t \leqslant m+2 \quad (m = 1,2,...), \tag{1.2}$$

satisfying the symmetry condition

$$\Delta_{r,t}(m) = \Delta_{m+2-r,m+3-t} \tag{1.3}$$

(already singled out in Ref. 1), correspond to unitary LW representations of the *Virasoro algebra* Vir.

Recently, modular invariant partition functions were constructed for these models[7] in terms of Feigin and Fuchs character formulas.

In the framework of a field theoretic approach[8,9] (in which the problem has been reduced to studying conformal quantum fields on the circle) composite quasiprimary fields have been constructed and conformal OPE's have been written down for a wide range of models.[10] (We recall that a field $\phi(z)$ of dimension Δ is called *quasiprimary*, if it transforms homogeneously under the projective conformal group SU (1,1); in an infinitesimal form its transformation law reads

$$[L_n, \phi(z)] = z^n\{z\phi'(z) + (n+1)\Delta\phi(z)\}, \tag{1.4}$$

for $n = 0, \pm 1$. The field ϕ is called *primary* if it satisfies (1.4) for all integer n's. We say that the quasiprimary fields $\phi_{\Delta+n}$ of dimension $\Delta + n$, $n = 0,1,2,...$, belong to the *conformal family* of the primary field ϕ_Δ, if they appear in OPE's of expressions of the type $T(z_1)\cdots T(z_k)\phi_\Delta(z)$. Here $T(z)$ is the stress-energy tensor

$$T(z) = \sum_{n\in\mathbb{Z}} L_n z^{-n-2} \quad (L_n^* = L_{-n}), \tag{1.5}$$

where L_n satisfy the commutation relations of Vir,

$$[L_n, L_k] = (n-k)L_{n+k} + (c/12)n(n^2-1)\delta_{n,-k}. \tag{1.6}$$

One advantage of using quasiprimary fields stems from the fact that they provide an orthogonal basis for OPE's: if ϕ_1 and ϕ_2 are quasiprimary fields of different dimensions, then $\langle\phi_1(z_1)\phi_2(z_2)\rangle = 0$.)

It was pointed out in Ref. 11 that the minimal unitary models [corresponding to (1.1) and (1.2) for $m = 1,2,...$]—the c_m models, for short—provide, for each m, representations of an associative algebra $\mathfrak{A}_m$, which includes the enveloping of the Virasoro algebra. Here $\mathfrak{A}_m$ can be defined as the OPE algebra generated by the stress energy tensor $T(z)$ and by the primary field $F(z)$ [$= F(z,m)$] of (half-) integer dimension

$$\Delta_{m+1,1}(m) = \Delta_{1,m+2}(m) = \tfrac{1}{4}m(m+1) \equiv s_m \tag{1.7}$$

(which can also be viewed as "spin"). The fusion rules of Ref. 1 guarantee that the conformal family of $F(z)$ is only coupled to the family of the unit operator. Since $T(z_1)F(z_2)$ belongs by definition to the family of F, whatever the primary field F, the preceding statement just says that $F(z+\epsilon/2)\,F(z-\epsilon/2)$ is expanded in quasiprimary fields $T_{2n}^F(z)$, $n = 0,1,2,...$, of the family of the unit operator, so that $T_0 = 1$, $T_2 = T$ (we shall display in Sec. II the proof of this result for the simplest special case of the Ising model, $m = 1$).

The main purpose of this paper is to develop a technique

[a] Permanent address.

 0022-2488/87/071598-10$02.50

for making efficient use of (global) OPE's in the c_m models. Computing normalization factors from known four-point functions, we are able to compare in particular the quasiprimary fields T^F_{2n}, which appear in the OPE of two F fields, with the composite fields T^T_{2n} made in a similar fashion out of two T's. They have identical conformal properties; however, in general, they do not coincide, since the quasiprimary states have a finite multiplicity (Sec. II). Only levels 2 and 4 are multiplicity-free for an arbitrary central change c. (In fact, $T^T_2 = T^F_2 = T$, while T^T_4 is a multiple of T^F_4.) The presence of a null vector at level $(m+1)(m+2)$ reduces the number of independent quasiprimary fields in the c_m model at this level. In the case of the (c_1) Ising model we deduce (in Sec. II) that level 6 is also multiplicity-free. An explicit computation, involving the canonical Majorana–Weyl field $F(z,1) \equiv \psi(z)$, shows that, in fact, $T^T_6 = 8T^\psi_6$ (Sec. IV). The presence of null vectors in the conformal family of ψ (discussed in Sec. II) leads to the vanishing of the composite quasiprimary fields of dimensions $\frac{5}{2}$, $\frac{7}{2}$, and $\frac{11}{2}$ appearing in the OPE of $T(z+\epsilon)\ \psi(z)$ (Sec. IV). Similar results are obtained for the composite fields $G_{n+3/2}(z)$ in the c_2 model. The computations of Sec. IV are preceded by a general discussion of the $\mathfrak{A}_m$ algebra in Sec. III where the normalization of T^T_{2n} is derived from the four-point function of T for any value of c, while the normalizations of T^F_{2n} and of F_{s+n} (a composite of F and T) are related to the four-point function $\langle TTFF\rangle$ for the c_m models. The new point in the set of propositions of Sec. III is precisely these numerical coefficients. Their knowledge is essential for exhibiting the relations between composite quasiprimary fields at levels involving a null vector and for computing higher correlation functions. The rather lengthy computations of these constant factors are summarized in the Appendix. (Only the first such factor, the one that multiplies the normalized primary field in each conformal family appearing in an OPE, has been evaluated previously, see Ref. 2.)

II. NULL VECTORS, FUSION RULES, AND DEGENERACIES: THE (TRI) CRITICAL ISING MODEL

A. Null vectors and quasiprimary states

For each LW vector $|\Delta_{r,t}\rangle = |c_m, \Delta_{r,t}(m)\rangle$ (1.2) such that

$$L_n|\Delta_{r,t}\rangle = 0, \quad n = 1,2,..., \quad (L_0 - \Delta_{r,t})|\Delta_{r,t}\rangle = 0, \tag{2.1}$$

there exist homogeneous "polynomials" of the type $P_N = a_0 L_{-N} + a_1 L_{1-N} L_{-1} + \cdots$ of "degree" $N = r\cdot t$ and $N = (m+2-r)(m+3-t)$ such that the vectors $P_N|\Delta_{r,t}\rangle$ are *null vectors*. They are characterized by the property $L_n P_N|\Delta_{r,t}\rangle = 0$ $[(L_0 - \Delta_{r,t} - N)|\Delta_{r,t}\rangle = 0]$ for $n \geqslant 1$, which implies that they are orthogonal to all vectors of the Verma module $\mathscr{V}_{c_m,\Delta_{r,t}}$ and can, therefore, be set consistently equal to zero.

In the family $\{1\}$ of the unit operator the null vectors appear at level 1—as $L_{-1}|0\rangle = 0$— and $(m+1)(m+2)$. One can derive from here the absence of quasiprimary fields at odd levels (for $n \leqslant 7$) and a linear dependence among such fields at level $(m+1)(m+2)$.

In order to clarify the last statement we shall estimate the number of quasiprimary fields $T_{2n}(z) \in \{1\}$ for small n's. To each $T_{2n}(z)$ corresponds a *quasiprimary state* (shorthand for a lowest weight quasiprimary state) $T_{2n}(0)|0\rangle$ satisfying

$$(L_0 - 2n)T_{2n}(0)|0\rangle = 0 = L_1 T_{2n}(0)|0\rangle. \tag{2.2}$$

There are unique quasiprimary states at levels 2 and 4 (proportional to $L_{-2}|0\rangle$ and to $(L^2_{-2} - \frac{3}{5}L_{-4})|0\rangle$). For higher n's we expect that the number of independent T_{2n}'s does not exceed[12] $p(n) - 2p(n-1) + p(n-2)$, where $p(n)$ is the "partition function"—i.e., the number of ways in which the positive integer n can be presented as a sum of positive integers (see Table I). Indeed, if we only have a null vector $(L_{-1}|0\rangle)$ at level 1, then the number of independent vectors at level n will be $M^{(n)} = p(n) - p(n-1)$. If, on the other hand, $M^{(n)}$ is the multiplicity of all states at level n, then the multiplicity of quasiprimary states at level n will be $M^{(n)} - M^{(n-1)}$.

We see, in particular, that the first quasiprimary state of odd dimension in the family of $\{1\}$ may appear at level 9. [Quasiprimary composite fields of the type T_{2n+1}, contributing to the OPE of a primary field $\phi(z)$ with itself, would not appear in the *symmetric* expansion $\phi(z+\epsilon/2)\phi(z-\epsilon/2)$.] (Indeed, the vector

$$\{5L^3_{-3} - 8L_{-9} + 6L_{-7}L_{-2} - 12L_{-6}L_{-3} + 8L_{-5}L^2_{-2} - 12L_{-4}L_{-3}L_{-2}\}|0\rangle$$

is annihilated by L_1.) We also find exactly two independent quasiprimary vectors at level 6 [say, $(L^2_{-3} - \frac{8}{5}L_{-4}L_{-2} - \frac{4}{5}L_{-6})|0\rangle$ and $(L^3_{-2} - \frac{9}{5}L_{-4}L_{-2} - \frac{6}{7}L_{-6})|0\rangle$]. For the c_1-Ising model a linear combination of these two vectors should be a null vector. Thus, for $c = \frac{1}{2}$ any two quasiprimary fields (of the family $\{1\}$) of dimension 6 should be proportional to each other. We shall display an implication of this statement—concerning composite quasiprimary fields—in Sec. IV.

B. Null vectors and fusion rules in the c_1 model

An important application of the presence of null vectors in the c_m models is the derivation of the fusion rules[1] for OPE's involving primary fields. We shall reproduce this derivation for the simple case of the product of the canonical Fermi field $\psi(z)$ of weight $\Delta_{2,1} = \Delta_{1,3} = \frac{1}{2}$ in the c_1 model, since some of the intermediate formulas will be also useful for the interpretation of our results in Sec. IV (concerning the vanishing of certain quasiprimary composite fields).

The general L_1-invariant vector at level 2 in the Verma module $\mathscr{V}_{c,\Delta}$ is proportional to

$$|c,\Delta+2\rangle = ([(2\Delta+1)/3]L_{-2} - \tfrac{1}{2}L^2_{-1})|c,\Delta\rangle. \tag{2.3}$$

Indeed, we have (for any Δ and c) $L_1|c,\Delta+2\rangle = \{(2\Delta+1)L_{-1} - L_0L_{-1} - L_{-1}L_0\}|c,\Delta\rangle = 0$. In order to secure also $L_2|c,\Delta+2\rangle = 0$ (which is equivalent to demanding that $|c,\Delta+2\rangle$ is a null vector—given that $L_1|c,\Delta+2\rangle = 0$) we must set

$$c = 18\Delta(1+2\Delta)^{-1} - 8\Delta$$

or

TABLE I. Values of the partition function $p(n)$ and its first two finite differences for $0 \leqslant n \leqslant 15$.

n	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
$p(n)$	1	1	2	3	5	7	11	15	22	30	42	56	77	101	135	176
$p(n) - p(n-1)$		0	1	1	2	2	4	4	7	8	12	14	21	24	34	41
$p(n) - 2p(n-1) + p(n-2)$			1	0	1	0	2	0	3	1	4	2	7	3	10	7

$$16\Delta = 5 - c \pm \sqrt{(25-c)(1-c)}. \tag{2.4}$$

For the c_m series (1.1) Eq. (2.4) gives

$$4\Delta_{2,1}(m) = 1 + 3/(m+2), \quad 4\Delta_{1,2}(m) = 1 - 3/(m+3) \tag{2.5}$$

in accord with (1.2). For $m = 1$ we find $\Delta_{2,1} = \frac{1}{2}$. Similarly, the general L_1-invariant vector at level 3 is a multiple of

$$|c, \Delta+3\rangle = \{\tfrac{1}{2}\Delta(\Delta+1)L_{-3} - (\Delta+1)L_{-2}L_{-1} + \tfrac{1}{2}L^{3}_{-1}\}|c,\Delta\rangle. \tag{2.6}$$

The requirement $L_2|c, \Delta+3\rangle = 0$ is only satisfied for $c = 12\Delta(\Delta+1)^{-1} - 3\Delta - 2$, which, for $\Delta = \frac{1}{2}$, again gives $c = \frac{1}{2}$. We shall now demonstrate that the null vector conditions

$$(\tfrac{2}{3}L_{-2} - \tfrac{1}{2}L^{2}_{-1})|\tfrac{1}{2},\tfrac{1}{2}\rangle = 0 = (\tfrac{1}{4}L_{-3} - L_{-2}L_{-1} + \tfrac{1}{3}L^{3}_{-1})|\tfrac{1}{2},\tfrac{1}{2}\rangle \tag{2.7}$$

imply the fusion rule for two ψ fields.

The conformal family of the primary field $\phi(z)$ (of dimension Δ) appears in the OPE of $\psi(z+\epsilon/2)\psi(z-\epsilon/2)$ iff ϕ has a nonvanishing three point function with a pair of ψ's, consistent with the null vector conditions. For such a ϕ we have $\langle\frac{1}{2},\frac{1}{2}|\phi(z)|\frac{1}{2},\frac{1}{2}\rangle = Az^{-\Delta}$ $(A \neq 0)$. The first equation (2.7) then implies

$$\begin{aligned} A^{-1}&\left\langle \frac{1}{2},\frac{1}{2}\right| \left(\frac{2}{3}L_2 - \frac{1}{2}L_1^2\right)\phi(z)\left|\frac{1}{2},\frac{1}{2}\right\rangle \\ &= z^2\left\{\frac{2}{3}z\frac{d}{dz} + 2\Delta - \frac{1}{2}\left(z\frac{d}{dz} + 2\Delta + 1\right)\right. \\ &\quad \left.\times\left(z\frac{d}{dz} + 2\Delta\right)\right\}z^{-\Delta} = 0, \\ &\Delta(\tfrac{4}{3} - (\Delta+1)/2) = 0. \end{aligned} \tag{2.8}$$

Both roots of (2.8) can be presented in the form (1.2): $\Delta_{1,1} = 0$, $\Delta_{3,1} = \frac{5}{3}$, but for the second one r ($= 3$) lies beyond the allowed range. The same function should, however, also satisfy the second equation (2.7),

$$\langle\tfrac{1}{2},\tfrac{1}{2}|(\tfrac{1}{4}L_3 - L_1L_2 + \tfrac{1}{3}L_1^3)\phi(z)|\tfrac{1}{2},\tfrac{1}{2}\rangle = 0$$

or

$$\begin{aligned} z^3&\left\{\frac{1}{4}z\frac{d}{dz} + \Delta - \left(z\frac{d}{dz} + 3\Delta + 1\right)\left(z\frac{d}{dz} + 2\Delta\right)\right. \\ &+ \frac{1}{3}\left(z\frac{d}{dz} + 2\Delta + 2\right) \\ &\left.\times\left(z\frac{d}{dz} + 2\Delta + 1\right)\left(z\frac{d}{dz} + 2\Delta\right)\right\}z^{-\Delta} = 0, \end{aligned} \tag{2.9a}$$

i.e.,

$$\Delta(\tfrac{1}{3}(\Delta+1)(\Delta+2) - 2\Delta - \tfrac{1}{4}) = 0$$

or

$$\Delta_{1,1} = 0, \quad \Delta_{1,3} = \tfrac{1}{2}, \quad \Delta_{1,5} = \tfrac{5}{2}. \tag{2.9b}$$

The only common root of (2.8) and (2.9) is $\Delta_{1,1} = 0$. Thus, only the family of $\phi(z) = 1$ should appear in the OPE of two ψ's.

We now proceed to spell out the implications of this analysis for the degeneracy of composite quasiprimary fields in the OPE of $T(z_1)\psi(z_2)$ (see Sec. IV).

First of all, we note that if $|\Delta\rangle$ is a LW vector then, for positive Δ, there is no quasiprimary state $|\Delta+1\rangle$ in the family of $|\Delta\rangle$. Indeed, any vector of dimension $\Delta+1$ in the Verma module $\mathscr{V}_{c,\Delta}$ should be proportional to $L_{-1}|\Delta\rangle$. But $L_{-1}|\Delta\rangle$ could only be quasiprimary for $\Delta = 0$, since $L_1L_{-1}|\Delta\rangle = 2\Delta|\Delta\rangle$.

There could be no more than one quasiprimary state at level 2 and 3, proportional to (2,3) and to (2.6). If these are null vectors—as in the case of the Ising model for $\Delta = \frac{1}{2}$—we expect to find no quasiprimary fields of dimension $\Delta + 2$ and $\Delta + 3$ (in the family of $|\Delta\rangle$) Moreover, the existence of null vectors at levels 2 and 3 also implies the presence of such a vector at level 5,

$$\begin{aligned} L_{-5}|c,\Delta\rangle &\equiv [L_{-2}, L_{-3}]\,|c,\Delta\rangle \\ &= \left\{L_{-2}\left(\frac{2}{\Delta}L_{-2}L_{-1} - \frac{1}{\Delta(\Delta+1)}L^{3}_{-1}\right)\right. \\ &\quad \left. - \frac{3}{2(2\Delta+1)}L_{-3}L^{2}_{-1}\right\}|c,\Delta\rangle. \end{aligned} \tag{2.10}$$

No independent null vectors can be obtained at higher levels in this fashion. (We owe the last remark—explaining the absence of a quasiprimary field of dimension $\frac{11}{2} = \frac{1}{2} + 5$ in the Ising model—to Trifonov.)

C. Superquasiprimary states in the c_2 model

For $m = 2$ the fields $T(z)$ and $G(z) = F(z,2)$ acting on the vacuum generate the Neveu–Schwarz (NS) superalgebra.[13] If we set (for the NS sector)

$$G(z) = \sum_{n\in\mathbf{Z}} G_{n+1/2}z^{-n-2}, \tag{2.11}$$

then the super-Virasoro commutation relations are

$$[G_\rho, G_\sigma]_+ = 2L_{\rho+\sigma} + (c/3)(\rho^2 - \tfrac{1}{4})\delta_{\rho,-\sigma}, \quad (2.12a)$$

$$[G_\rho, L_n] = (\rho - n/2)G_{\rho+n} \quad (\rho\in\mathbb{Z}+\tfrac{1}{2}). \quad (2.12b)$$

[The Virasoro commutation relations (1.6) follow from (2.12) and from the super Jacobi identiy.]

A NS superfield $\phi_\Delta(z,\theta) = \varphi_\Delta(z) + \theta\varphi_{\Delta+1/2}(z)$, where θ is a Grassmann variable ($\theta^2 = 0$), is said to be *superquasiprimary* if for any choice of the anticommuting parameter ϵ,

$$[\epsilon G_{k+1/2}, \phi_\Delta(z,\theta)] = \epsilon z^k\left\{z\left(\frac{\partial}{\partial\theta} - \theta\frac{\partial}{\partial z}\right) - 2(k+1)\Delta\theta\right\}\phi_\Delta(z,\theta), \quad (2.13a)$$

$$[L_n, \phi_\Delta(z,\theta)] = z^n\left\{z\frac{\partial}{\partial z} + (n+1)\left(\Delta + \frac{1}{2}\theta\frac{\partial}{\partial\theta}\right)\right\}\phi_\Delta(z,\theta), \quad (2.13b)$$

for $k = 0, -1$ and $n = 0, \pm 1$ [which span the osp(2/1) Lie superalgebra]; ϕ_Δ is called *superprimary* if (2.13) is valid for all $k,n\in\mathbb{Z}$. If $\phi_\Delta(z,\theta)$ is a superquasiprimary field, then $|\Delta\rangle = \varphi_\Delta(0)|0\rangle$ is a *superquasiprimary state* in the sense that

$$G_{1/2}|\Delta\rangle = 0 \quad [= L_1|\Delta\rangle = (L_0 - \Delta)|\Delta\rangle]. \quad (2.14)$$

An example of a superquasiprimary field is given by the odd supercurrent (with $\Delta = \frac{3}{2}$)

$$W(z,\theta) = \tfrac{1}{2}G(z) + \theta T(z). \quad (2.15)$$

We are interested in the question of whether supersymmetry gives additional relations between quasiprimary fields and states. The answer is no. What it does say on the level of quadratic functions of T and G, of dimension $2n$, is to organize the quasiprimary fields into parts of two superfields; one—odd—of dimension $2n$-$\frac{1}{2}$ and another—even—of dimension $2n$. They provide two orthogonal linear combinations of T^G_{2n} and T^T_{2n}. We shall illustrate here the situation in terms of quasiprimary states at the lowest nontrivial level, $2n = 6$. (For $2n = 2$ and 4 the quasiprimary states are multiplicity-free.)

The two quasiprimary states, proportional to $T^T_6(0)|0\rangle$ and $T^G_6(0)|0\rangle$, are

$$|6\rangle_T = (20L_{-6} + 56L_{-4}L_{-2} - 35L^2_{-3})|0\rangle, \quad (2.16a)$$

$$|6\rangle_G = (20L_{-6} + 21G_{-9/2}G_{-3/2} - 35G_{-7/2}G_{5/2})|0\rangle. \quad (2.16b)$$

(The term involving L_{-6} in $|6\rangle_G$ can be though as coming from the anticommutators $[G_{-7/2}, G_{-5/2}]_+ = 2L_{-6} = [G_{-3/2}, G_{-9/2}]_+$.) One readily verifies that $|6\rangle_T$ and $|6\rangle_G$ are indeed lowest weight quasiprimary states,

$$L_1|6\rangle_T = 0 = L_1|6\rangle_G. \quad (2.17)$$

On the other hand,

$$G_{1/2}|6\rangle_T = 28(5G_{-11/2} + 3L_{-4}G_{-3/2} - 5L_{-3}G_{-5/2} + 5G_{-7/2}L_{-2})|0\rangle = 2G_{1/2}|6\rangle_G \equiv 28|\tfrac{11}{2}\rangle_W. \quad (2.18)$$

The common state $|\frac{11}{2}\rangle_W$ obtained in this manner is a superquasiprimary state, i.e., it satisfies

$$G_{1/2}|\tfrac{11}{2}\rangle_W = 0 \quad (2.19)$$

and its superpartner

$$G_{-1/2}|\tfrac{11}{2}\rangle_W = \{2(5L_{-6} + 8L_{-4}L_{-2} - 5L^2_{-3}) + \tfrac{3}{2}(3G_{-9/2}G_{-3/2} - 5G_{-7/2}G_{-5/2})\}|0\rangle \equiv |6\rangle_W \quad (2.20)$$

is, by construction, orthogonal to the superquasiprimary state

$$|6\rangle_S = 2|6\rangle_G - |6\rangle_T = (20L_{-6} - 56L_{-4}L_{-2} + 35L^2_{-3} + 42G_{-9/2}G_{-3/2} - 70G_{-7/2}G_{-5/2})|0\rangle \quad (\neq 0), \quad (2.21)$$

which satisfies, in view of (2.18),

$$G_{1/2}|6\rangle_S = 0. \quad (2.22)$$

(Note that we have been using non-normalized quasiprimary states throughout.)

As a corollary of the absence of quasiprimary states at level 5 and 7—see Table I—it follows that

$$|\tfrac{9}{2}\rangle_W = 0 = |\tfrac{13}{2}\rangle_W \quad (2.23)$$

(otherwise we would have had, e.g., $G_{-1/2}|\frac{13}{2}\rangle_W = |7\rangle_W \neq 0$). On the other hand, $|\frac{7}{2}\rangle_W = (3G_{-7/2} - 4L_{-2}G_{-3/2})|0\rangle \neq 0$, since there is a nonvanishing $|4\rangle_W = G_{-1/2}|\frac{7}{2}\rangle_W$.

III. THE OPE ALGEBRA $\mathfrak{A}_m$

A. OPE for the stress-energy tensor. Normalization of T^T_{2n}

As demonstrated in Refs. 9 and 10 the bilocal vector-valued function $T_T(z+\epsilon/2, z-\epsilon/2)|\Omega\rangle$, where $|\Omega\rangle$ is a finite energy state and

$$T_T(z_1,z_2) = \tfrac{1}{2}z^2_{12}\{T(z_1)T(z_2) - (c/2)z_{12}^{-4}\} \quad (3.1)$$

is analytic in ϵ ($= z_{12}$) and gives rise to the OPE

$$T_T\left(z+\frac{\epsilon}{2}, z-\frac{\epsilon}{2}\right) = \frac{3}{4}\int_{-1}^{1} d\lambda(1-\lambda^2)T\left(z+\lambda\frac{\epsilon}{2}\right) + \sum_{n=1}^{\infty}\epsilon^{2n}\int_{-1}^{1} d\lambda\, p_{2n+2}(\lambda)T^T_{2n+2}\left(z+\lambda\frac{\epsilon}{2}\right) \quad (3.2a)$$

$$= T(z) + \sum_{k=1}^{\infty}\frac{3}{2k+3}\frac{T^{(2k)}_{(z)}}{(2k+1)!}\left(\frac{\epsilon}{2}\right)^{2k} + \sum_{n=1}^{\infty}\epsilon^{2n}\left\{T^T_{2n+2}(z) + \sum_{k=1}^{\infty}\frac{(2k+1)!!(4n-1)!!}{(4n+2k-1)!!}\frac{T^{T(2k)}_{2n+2}(z)}{(2k)!}\left(\frac{\epsilon}{2}\right)^{2k}\right\}, \quad (3.2b)$$

where $p_k(\lambda)$ are the normalized weights

$$p_k(\lambda) = \frac{(2k-1)!!}{2^k(k-1)!}(1-\lambda^2)^{k-1}, \quad \int_{-1}^{1} d\lambda\, p_k(\lambda) = 1. \quad (3.3)$$

Conversely, the quasiprimary fields T^T_{2n} are expressed in terms of the bilocal field (3.1) as

$$(2n-2)!T^T_{2n}(z) = \lim_{z_1,z_2\to z} D^{(2,2)}_{2n-2}(\partial_1,\partial_2)T_T(z_1,z_2), \tag{3.4}$$

where $D^{(\delta_1,\delta_2)}_n(\alpha,\beta)$ is a homogeneous polynomial in (α,β) (related to a Jacobi polynomial),

$$\binom{2n+\delta_1+\delta_2-2}{n} D^{(\delta_1,\delta_2)}_n(\alpha,\beta) = \sum_{k=0}^{n}\binom{n+\delta_1-1}{n-k}\binom{n+\delta_2-1}{k}\alpha^{n-k}(-\beta)^k \tag{3.5a}$$

$$= (\alpha+\beta)^n P^{(\delta_1-1,\delta_2-1)}_n((\alpha-\beta)/(\alpha+\beta)). \tag{3.5b}$$

We have, in particular,

$$T^T_2(z) = \lim_{z_1,z_2\to z} T_T(z_1,z_2) = T(z), \tag{3.6a}$$

$$\Lambda(z) \equiv 2T^T_4(z) = \lim_{z_1,z_2\to z}\left\{\frac{\partial_1^2+\partial_2^2-3\partial_1\partial_2}{10} \times\left(z_{12}^2 T(z_1)T(z_2) - \frac{c}{2z_{12}^2}\right)\right\} \equiv N(T^2(z)) \tag{3.6b}$$

(the last expression defining the renormalized normal product of two T's).

We now proceed to determine the normalization of the two-point function of T^T_{2n}. The starting point will be the four-point function of T,

$$\langle T(z_1)T(z_2)T(z_3)T(z_4)\rangle = (c/2)^2((z_{12}^4z_{34}^4)^{-1} + (z_{13}^4z_{24}^4)^{-1} + (z_{14}^4z_{23}^4)^{-1}) + \langle T(z_1)T(z_2)T(z_3)T(z_4)\rangle^{\rm tr} \tag{3.7}$$

(where the superscript tr stands for "truncated part"); it is determined by conformal invariance, analyticity, and the *Ward–Takahashi identity* (WTI)

$$[T^{(-)}_{(z_1)}, T_{(z_2)}] = c/2z_{12}^4 + 2T(z_2)/z_{12}^2 + T'(z_2)/z_{12}, \tag{3.8}$$

where $T^{(-)}$ is the negative frequency part[8] of T,

$$T^{(-)}_{(z)} = \sum_{n\geqslant -1} L_n z^{-n-2}, \quad T^{(-)}(z)|0\rangle = 0. \tag{3.9}$$

We shall give here, for the sake of completeness, this simple derivation. (A similar argument was used in a slightly more complicated situation in the proof of Proposition 2.1 of Ref. 11.)

Lemma 3.1: The truncated four-point function in (3.7) has the form of three one-loop graphs,

$$\langle T(z_1)T(z_2)T(z_3)T(z_4)\rangle^{\rm tr} = c/z_{12}^2z_{23}^2z_{34}^2z_{14}^2 + c/z_{13}^2z_{34}^2z_{24}^2z_{12}^2 + c/z_{13}^2z_{23}^2z_{24}^2z_{14}^2. \tag{3.10}$$

Proof: Consider the auxiliary function

$$\Phi(z_1) = \Phi(z_1;z_2,z_3,z_4) = z_{12}z_{13}z_{14}\langle T(z_1)T(z_2)T(z_3)T(z_4)\rangle^{\rm tr}. \tag{3.11}$$

Conformal invariance fixes the large z behavior of an n point function of $T(z)$ at z^{-4} (since T is a quasiprimary field of dimension 2). The WTI (3.8) combined with a general analyticity argument [taking into account the fact that we have subtracted the disconnected part of the four-point function (3.7)] tells us that $\Phi(z_1)$ can only have simple poles at the points z_2, z_3, and z_4. Thus, using known analyticity and asymptotic behavior we can write

$$\Phi(z_1) = A/z_{12} + B/z_{13} + C/z_{14}, \tag{3.12}$$

where the numerators may depend on z_2, z_3, and z_4. Multiplying by z_{12} and applying (3.8) we find

$$A = \lim_{z_1\to z_2} z_{12}\Phi(z_1) = 2z_{23}z_{24}\langle T(z_2)T(z_3)T(z_4)\rangle = 2c/z_{23}z_{24}z_{34}^2 \tag{3.13a}$$

(where we have used the expression for the three-point function of T obtained from conformal invariance and the WTI); similarly

$$B = \lim_{z_1\to z_3} z_{13}\Phi(z_1) = 2c/z_{23}z_{24}^2z_{34}, \qquad C = \lim_{z_1\to z_4} z_{14}\Phi(z_1) = 2c/z_{23}^2z_{24}^2z_{34}. \tag{3.13b}$$

Inserting (3.13) into (3.12) and using (3.11) as well as the identity

$$(z_{12}^2z_{34}^2z_{13}z_{14}z_{23}z_{24})^{-1} + (z_{14}^2z_{23}^2z_{12}z_{13}z_{24}z_{34})^{-1} = (z_{12}^2z_{23}^2z_{34}^2z_{14}^2)^{-1} \tag{3.14}$$

we arrive at (3.10).

Proposition 3.2: The three-point function of T^T_{2n} with two T's is given by

$$\langle T(z_1)T(z_2)T^T_{2n}(z_3)\rangle = K^T_{2n}[cz_{12}^{2n-4}/(z_{13}z_{23})^{2n}], \tag{3.15}$$

where

$$K^T_{2n} = \frac{((2n-1)!)^2}{(4n-2)!} \times\left\{\frac{c}{144}\frac{(2n+2)!}{(2n-4)!} + 2(4n^2-2n-1)\right\}. \tag{3.16}$$

The z dependence in (3.15) is a consequence of conformal invariance. The coefficient K^T_{2n} is obtained by a computation summarized in the Appendix.

Corollary 3.3: The two-point function of T^T_{2n} is given by

$$\langle T^T_{2n}(z_1)T^T_{2n}(z_2)\rangle = K^T_{2n}(c/2z_{12}^{4n}), \tag{3.17}$$

with K^T_{2n} again given by (3.16).

Proof of the Corollary: It follows from (3.5 b) that

$$\left(\frac{\alpha-\beta}{2}\right)^n = D^{(\delta_1,\delta_2)}_n(\alpha,\beta) + \sum_{k=0}^{n-1} b_{nk}(\alpha+\beta)^{n-k}D^{(\delta_1,\delta_2)}_k(\alpha,\beta), \tag{3.18}$$

with some real coefficients b_{nk}. Because of the orthogonality property

$$(n-k)\langle T_{2k}(z_1)T_{2n}(z_2)\rangle = 0 \tag{3.19}$$

(valid for any pair of quasiprimary fields—as noted in the Introduction) we deduce from (3.4) and (3.18) that

$$(2n-2)!\langle T^T_{2n}(z_2)T^T_{2n}(z_3)\rangle = \lim_{z_1\to z_2}\left(\frac{\partial_1-\partial_2}{2}\right)^{2n-2}\frac{z_{12}^2}{2}\langle T(z_1)T(z_2)T^T_{2n}(z_3)\rangle, \tag{3.20}$$

which, together with (3.15), implies (3.17).

B. OPE's involving $F(z)$ in a c_m model

We normalize the two-point function of F by

$$\langle F(z_1)F(z_2)\rangle = c/sz_{12}^{2s} \quad [s = s_m = m(m+1)/4]. \tag{3.21}$$

Then the counterpart of the bilocal operator (3.1) is the composite field

$$T_F(z_1,z_2) = (1/2z_{12}^2)\{z_{12}^{2s}F(z_1)F(z_2) - c/s\}. \tag{3.22}$$

Its OPE is again given by (3.2) and (3.3) with T_{2n}^T substituted by T_{2n}^F. In particular, the analog of (3.6a) is still true,

$$T_2^F(z) = \lim_{z_1,z_2\to z} T_F(z_1,z_2) = T(z). \tag{3.23}$$

Indeed, using the WTI we find the conformal three-point function

$$\langle F(z_1)F(z_2)T(z_3)\rangle = (sz_{12}^2/z_{13}^2z_{23}^2)\langle F(z_1)F(z_2)\rangle = cz_{12}^{2-2s}/z_{13}^2z_{23}^2. \tag{3.24}$$

From (3.22)–(3.24) we reproduce the correct two-point function of $T(z)$. To prove (3.23) it remains to use the uniqueness (up to normalization) of the field of dimension 2 in the family of the unit operator (see the last line of Table I).

The properties of T_{2n}^F can be read from the four-point functions of F. One of them is easy to find for arbitrary m.

Lemma 3.4: The four-point function of two F and two T is

$$\langle T(z_1)T(z_2)F(z_3)F(z_4)\rangle = c^2/2sz_{12}^4z_{34}^{2s} + \langle T(z_1)T(z_2)F(z_3)F(z_4)\rangle^{\text{tr}}, \tag{3.25}$$

where

$$\langle T(z_1)T(z_2)F(z_3)F(z_4)\rangle^{\text{tr}} = \frac{cz_{34}^{2-2s}}{z_{12}^2z_{13}z_{24}z_{14}z_{23}}\left(\frac{sz_{12}^2z_{34}^2}{z_{13}z_{24}z_{23}z_{14}} + 2\right). \tag{3.26}$$

Proof: The argument proving Lemma 3.1 applies if we supplement the WTI (3.8) by

$$[T^{(-)}(z_1),F(z_2)] = sF(z_2)/z_{12}^2 + F'(z_2)/z_{12}. \tag{3.27}$$

Proposition 3.5: The three-point function of T_{2n}^F with two T's is

$$\langle T(z_1)T(z_2)T_{2n}^F(z_3)\rangle = K_{2n}^{FT}[cz_{12}^{2n-4}/(z_{13}z_{23})^{2n}] \tag{3.28a}$$

with

$$K_{2n}^{FT} = 2[((2n-1!)^2/(4n-2)!][s(2n+1)(n-1)+1]. \tag{3.28b}$$

The necessary steps in the calculation of K_{2n}^{FT} are again presented in the Appendix.

Corollary 3.6: The mixed two-point function of T_{2n}^T and T_{2n}^F is given by

$$\langle T_{2n}^T(z_1)T_{2n}^F(z_2)\rangle = K_{2n}^{FT}(c/2z_{12}^{4n}). \tag{3.29}$$

The *proof* is the same as for Corollary 3.3.

Corollary 3.7: The field $T_{2n}^\perp(z)$ defined by

$$[((2n-1)!)^2/(4n-2)!]T_{2n}^\perp(z) = K_{2n}^{FT}T_{2n}^T(z) - K_{2n}^T T_{2n}^F(z) \tag{3.30}$$

is orthogonal to T_{2n}^T. The uniqueness of T_4 (see Table I) implies that

$$T_4^\perp = 0 \quad [= \langle T_4^\perp(z_1)T_4^\perp(z_2)\rangle]. \tag{3.31}$$

This will be verified in Sec. IV for the cases in which the four-point function of four F's is also computed.

Lemma 3.4 also allows to evaluate the two- and three-point functions of the quasiprimary composite fields

$$F_{n+s}(z) = \lim_{z_1,z_2\to z}(1/5\cdot n!)D_n^{(2s-2,2)}(\partial_1,\partial_2)\{z_{12}^2T(z_1)F(z_2)\}, \tag{3.32}$$

which appear in the operator product expansion[10]

$$\epsilon^2T(z+\epsilon)F(z) = s\sum_{n=0}^\infty \epsilon^n\int_0^1 F_{n+s}(z+u\epsilon)p_n^{(2s-2,2)}(u)du, \tag{3.33}$$

where

$$p_n^{(\delta_1,\delta_2)}(u) = [\Gamma(2n+\delta_1+\delta_2)/\Gamma(n+\delta_1)\Gamma(n+\delta_2)] \times u^{\delta_2+n-1}(1-u)^{\delta_1+n-1}, \tag{3.34}$$

$$\int_0^1 p_n^{(\delta_1,\delta_2)}(u)du = 1;$$

in particular,

$$sF_s(z) \equiv \lim_{\epsilon\to0}\{\epsilon^2T(z+\epsilon)F(z)\} = sF(z). \tag{3.35}$$

Proposition 3.8: The three-point function $\langle TF_{n+s}F\rangle$ obtained from (3.25) and (3.26) is

$$\langle T(z_1)F_{n+s}(z_2)F(z_3)\rangle = B_{ns}(cz_{13}^{n-2}/z_{12}^{n+2}z_{23}^{n+2s-2}), \tag{3.36a}$$

where

$$B_{ns} = \frac{(-1)^{2s}}{s}\binom{2s+2n-2}{n+1}^{-1} \times\left\{(-1)^n[s(n+1)(2s+n-2)-2] + \frac{c}{6}n(n-1)\binom{2s+n}{n} + 2\binom{2s+n-2}{n+1} \times\left[n + \frac{n+1}{2s+n-2}\right]\right\}. \tag{3.36b}$$

The proof uses once again formulas collected in the Appendix as well as the following identity among hypergeometric functions:

$$(c-a-1)\,{}_2F_1(a,b;c;z) + a\,{}_2F_1(a+1,b;c;z) - (c-1)\,{}_2F_1(a,b;c-1;z) = 0$$

and the well-known integral representation valid for $\operatorname{Re} c > \operatorname{Re} b > 0$,

$${}_2F_1(a,b;c;z) = \Gamma(c)[\Gamma(b)\Gamma(c-b)]^{-1} \times\int_0^1 dt\, t^{b-1}(1-t)^{c-b-1}(1-tz)^{-a}.$$

Corollary 3.9: Noting that $B_{os} = 1$ we verify that

$$\langle T(z_1)F_s(z_2)F(z_3)\rangle = \langle T(z_1)F(z_2)F(z_3)\rangle.$$

Since, on the other hand, every quasiprimary field of (minimal) weight s in the conformal family of the primary field F (of the same weight s) is proportional to F, it follows that

$F_s = F$, thus proving (3.35). Since $B_{1s} = 0$, we obtain a direct proof of the statement that there is no quasiprimary field of dimension $s+1$ in the family of F (cf. Sec. II).

IV. THE c_1 AND c_2 ISING MODELS

A. Degeneracies for $m=1$

The knowledge of the (free) four-point function of the Majorana–Weyl field $\psi(z) = F(z,1)$ (of dimension $\frac{1}{2}$) in the c_1 model,

$$\langle \psi(z_1)\psi(z_2)\psi(z_3)\psi(z_4)\rangle = (z_{12}z_{34})^{-1} - (z_{13}z_{24})^{-1} + (z_{14}z_{23})^{-1} \tag{4.1}$$

allows us to compute the normalization of T^{ψ}_{2n}.

Proposition 4.1: The three-point function of T^{ψ}_{2n} and a pair of ψ's is

$$\langle \psi(z_1)\psi(z_2)T^{\psi}_{2n}(z_3)\rangle = ([(2n-1)!]^2/(4n-2)!)[z_{12}^{2n-1}/(z_{13}z_{23})^{2n}]. \tag{4.2}$$

The two-point function of T^{ψ}_{2n} is then obtained as a corollary,

$$\langle T^{\psi}_{2n}(z_1)T^{\psi}_{2n}(z_2)\rangle = \tfrac{1}{2}([(2n-1)!]^2/(4n-2)!)(1/z_{12}^{4n}). \tag{4.3}$$

The reader will reconstruct the proof using formulas of the Appendix.

Corollary 4.2: Comparing (3.29) (for $s=c=\frac{1}{2}$) with (4.3) we find that

$$T^{\psi\perp}_{2n} = T^{\psi}_{2n} - [2/(2n^2-n+1)]T^{T}_{2n} \tag{4.4}$$

is orthogonal to T^{ψ}_{2n},

$$\langle T^{\psi\perp}_{2n}(z_1)T^{\psi}_{2n}(z_2)\rangle = 0, \tag{4.5}$$

while comparison with (3.16) and (3.17) gives

$$\langle T^{\psi\perp}_{2n}(z_1)T^{T}_{2n}(z_2)\rangle = \frac{[(2n-1)!]^2}{2\cdot(4n-2)!}\left\{n^2 - \frac{n-1}{2} - \frac{1}{2n^2-n+1} \times\left[\frac{1}{288}\frac{(2n+2)!}{(2n-4)!} + 2(4n^2-2n-1)\right]\right\}z_{12}^{-4n}. \tag{4.6}$$

We see that the quantity in the braces vanishes for $n=1,2,3$. For $n=1$ this just says that $T^{\psi}_{2} = T^{T}_{2} = T$, which is true for any (quasi)primary field ψ [cf. (3.6) and (3.23)]. For $n=2$ we find

$$T^{T}_{4} = \tfrac{7}{2}T^{\psi}_{4}. \tag{4.7}$$

This confirms the general statement (made in Sec. II) that there is just one quasiprimary field of dimension 4 in the family of the unit operator [and is a special case of (3.31)]. For $n=3$, however, the result

$$T^{T}_{6} = 8T^{\psi}_{6} \quad \text{for } c=\tfrac{1}{2}\ (=s_1) \tag{4.8}$$

is characteristic for the Ising model. It verifies an implication of the degeneracy at level 6 for $m=1$ noted in Sec. VI. Equation (4.6) also shows that, for $2n \geqslant 8$, $T^{\psi\perp}_{2n} \neq 0$. Thus T^{ψ}_{2n} (for $n \geqslant 1$) and $T^{\psi\perp}_{2n}$ (for $n \geqslant 4$) form an orthogonal basis for OPE's of both $\psi\times\psi$ and $T\times T$. It is interesting to find out whether one can also expand the product $\varphi * \varphi$ in this basis, for the magnetization field φ of dimension $\frac{1}{16}$ in the Ising model (we are using complex S^1 fields for noninteger 2Δ, where Δ is the field's dimension—see Ref. 11).

Going to the composite fields of the ψ family we first derive, as a simple corollary of (3.36), the relation

$$\langle \psi^{(z_1)}_{n+1/2}\psi^{(z_2)}_{n+1/2}\rangle = \frac{1}{4}\binom{2n-1}{n+1}^{-1}\left[(-1)^n\frac{n(n^2-1)}{4!} + \frac{n^2-1}{4} - 1\right] \times z_{12}^{-2n-1}, \quad \text{for } n=2,3,\dots . \tag{4.9}$$

We see that the two-point function—and hence, also the field $\psi_{n+1/2}$—vanishes for $n=2,3,5$ since

$$(-1)^n[n(n^2-1)/4!] + (n^2-1)/4 - 1 = (1/4!)[(-1)^n n - 2]\,[(-1)^n n + 3] \times[(-1)^n n + 5],$$

thus

$$(\psi_{3/2}) = \psi_{5/2} = \psi_{7/2} = \psi_{11/2} = 0 \quad (\text{for } c=\tfrac{1}{2}), \tag{4.10}$$

as anticipated in Sec. II.

B. Summing an OPE

The set of global OPE's for the (quasi)primary fields completely characterizes a theory. To display the efficiency of such a characterization we shall reconstruct the (free) four-point function (4.1) [from the OPE (Ref. 10)]

$$\frac{1}{2\epsilon^2}\left[\epsilon\psi\left(z+\frac{\epsilon}{2}\right)\psi\left(z-\frac{\epsilon}{2}\right) - 1\right] = \frac{3}{4}\int_{-1}^{1} d\lambda(1-\lambda^2)T\left(z+\lambda\frac{\epsilon}{2}\right) + \sum_{n=1}^{\infty}\epsilon^{2n}\int_{-1}^{1} d\lambda\, p_{2n+2}(\lambda)T^{\psi}_{2n+2}\left(z+\lambda\frac{\epsilon}{2}\right), \tag{4.11}$$

where $p_k(\lambda)$ is given by (3.3) and from the knowledge of the three-point function (4.2). Indeed, setting $z_{3,4} = \pm\epsilon/2$, we find

$$\left\langle\psi(z_1)\psi(z_2)\psi\left(\frac{\epsilon}{2}\right)\psi\left(-\frac{\epsilon}{2}\right)\right\rangle - \frac{1}{z_{12}\epsilon} = 2\sum_{n=1}^{\infty}\epsilon^{2n-1}\int_{-1}^{1} d\lambda\, p_{2n}(\lambda)\left\langle\psi(z_1)\psi(z_2)T^{\psi}_{2n}\left(\lambda\frac{\epsilon}{2}\right)\right\rangle = 2\{I(\epsilon;z_1,z_2) - I(-\epsilon;z_1,z_2)\}, \tag{4.12}$$

where I is obtained by summing up the power series in the integrand

$$I(\epsilon;z_1,z_2)=\int_{-1}^{1}d\lambda\left\{\frac{2\epsilon z_{12}(1-\lambda^2)}{[\epsilon z_{12}-4z_1z_2+2(z_1+z_2)\epsilon\lambda-\epsilon(z_{12}+\epsilon)\lambda^2]^2}-\frac{1}{\epsilon z_{12}-4z_1z_2+2(z_1+z_2)\epsilon\lambda-\epsilon(z_{12}+\epsilon)\lambda^2}\right\}=\frac{2}{(2z_1+\epsilon)(2z_2+\epsilon)}; \tag{4.13}$$

inserting (4.13) into (4.12) we recover (4.1).

C. Linear independence of T_{2n}^G and T_{2n}^T for $n\geqslant 3$ in the c_2 model. Vanishing of $G_{9/2}(z)$

The supersymmetry property (2.13a) applied to the supercurrent (2.15) gives, in particular,

$$[G^{(-)}(z_1),G(z_2)]_+=(2/z_{12})T(z_2)+2c/3z_{12}^3. \tag{4.14}$$

Since the three-point function $\langle TGG\rangle$ is known [it is given by (3.24) for $s=\frac{3}{2}$] we can compute from (4.14) the four-point function of G with the result

$$\langle G(z_1)G(z_2)G(z_3)G(z_4)\rangle^{\text{tr}}=2c/z_{12}z_{13}z_{14}z_{23}z_{24}z_{34}. \tag{4.15}$$

(The most economic derivation of this result that we are aware of follows the lines of the proof of Lemma 3.1.) We then have the following.

Proposition 4.3: The three-point function of T_{2n}^G with a pair of G's is

$$\langle G(z_1)G(z_2)T_{2n}^G(z_3)\rangle=K_{2n}^G\,[cz_{12}^{2n-3}/(z_{13}z_{23})^{2n}], \tag{4.16a}$$

where

$$K_{2n}^G=\frac{[(2n-1)!]^2}{(4n-2)!}\left\{2+\frac{c}{9}\frac{(2n+1)!}{(2n-3)!}\right\}\quad\left(c=\frac{7}{10}\right). \tag{4.16b}$$

As a consequence,

$$\langle T_{2n}^G(z_1)T_{2n}^G(z_2)\rangle=K_{2n}^G(c/2z_{12}^{4n}). \tag{4.17}$$

Corollary 4.4: It follows from (3.28)–(3.30) and (4.17) that

$$\langle T_{2n}^{\perp}(z)T_{2n}^G(0)\rangle$$
$$=c\frac{[(2n-1)!]^2}{(4n-2)!}\left\{(6n^2-3n-1)^2-\left[\frac{c}{144}\frac{(2n+2)!}{(2n-4)!}+2(4n^2-2n-1)\right]\times\left[2+\frac{c}{9}\frac{(2n+1)!}{(2n-3)!}\right]\right\}. \tag{4.18}$$

For $n=2$ the quantity in the braces,

$$17^2-2(5c+22)(1+\tfrac{20}{3}c)=5(4c+21)(7-10c),$$

vanishes for $c=c_2=\frac{7}{10}$ (and for the nonunitary point $c=-\frac{21}{4}$).

It follows from (3.36b) that the normalization factor $B_{n\,3/2}$ of the three-point function $\langle T(z_1)G_{n+3/2}(z_2)G(z_3)\rangle$,

$$B_{n\,3/2}=3\frac{n!}{2n+3}\frac{(n+1)!}{(2n+1)!}\times\left\{\frac{c}{36}(-1)^n(n+3)(n+2)(n^2-1)n+\frac{3}{2}(n+1)^2-2+2(-1)^n(n+1)\right\},$$
$$\left(c=\frac{7}{10}\right), \tag{4.19}$$

only vanishes for $n=3$ (in which case the expression in the braces reduces to $14-20c$). This result is again in line with the general discussion of null vectors in Sec II, since the field G is represented by the pair [3,1] ($\sim$[1,4]). The degeneracy at level 4, however, is not felt by (4.19) since, in fact, not all quasiprimary vectors of dimension $\frac{11}{2}$ ($=4+\frac{3}{2}$) have zero norm.

It would be also interesting to classify the composite quasiprimary fields according to their supersymmetry properties. The analysis of Sec. II C indicates that for the class of the unit operator certain linear combinations of T_{2n}^T and T_{2n}^G will have simple commutation relations with the supercharge $G_{-1/2}$.

ACKNOWLEDGMENTS

Two of us (G. S. and I. T.) would like to thank Professor P. Budinich for his hospitality at the International School for Advanced Studies in Trieste where most of this work was done.

G. S. also acknowledges financial support by I. N. F. N.

APPENDIX: THREE-POINT FUNCTIONS IN TERMS OF LIMITS OF FOUR-POINT FUNCTIONS

In this Appendix we provide some auxiliary formulas which we found useful in computing the normalization constants K_{2n}^T, K_{2n}^F, and B_{ns} which summarize the content of Propositions 3.2, 3.5, and 3.8.

The main tool used in the text is the Leibnitz rule for evaluating higher derivatives of the product of two functions,

$$\frac{d^n}{dz^n}[f(z)g(z)]=\sum_{k=0}^{n}\binom{n}{k}\left[\frac{d^kf(z)}{dz^k}\right]\frac{d^{n-k}g(z)}{dz^{n-k}}, \tag{A1}$$

together with the elementary relations

$$\frac{\partial^k}{\partial z_1^k}z_{12}^{\alpha}=\begin{cases}[\Gamma(\alpha+1)/\Gamma(\alpha-k+1)]z_{12}^{\alpha-k}, & \alpha>0, \quad \text{(A2a)}\\ (-1)^k[\Gamma(k-\alpha)/\Gamma(-\alpha)]z_{12}^{\alpha-k}, & \alpha<0. \quad \text{(A2b)}\end{cases}$$

The following formula is also used to get some of the results of the main text,

$$
\begin{aligned}
&\partial_4^k (z_{14}^{-2} z_{24}^{-2} z_{34}^r) \\
&\quad = \Gamma(k+1)\Gamma(r+1) z_{12}^{-2} \\
&\qquad \times \sum_{s=0}^{k} \frac{(-1)^s}{\Gamma(r+1-s)\Gamma(s+1)} z_{34}^{r-s} \\
&\qquad \times [(k+1-s) z_{14}^{-2-k+s} \\
&\qquad + 2 z_{12}^{-1} z_{14}^{-1-k+s} + (1 \leftrightarrow 2)]; \qquad \text{(A3a)}
\end{aligned}
$$

for $k \geqslant r$ it reduces to

$$
\begin{aligned}
&\partial_4^k (z_{14}^{-2} z_{24}^{-2} z_{34}^r) \\
&\quad = (-1)^r \Gamma(k+1) z_{12}^{-3} z_{14}^{-2-k} z_{13}^{r-1} \{ [(k+1) z_{12} \\
&\qquad + 2 z_{14}] z_{13} - r z_{12} z_{14} \} + (1 \leftrightarrow 2). \qquad \text{(A3b)}
\end{aligned}
$$

In order to show the kind of computations involved in getting the constants K_{2n}^{T}, K_{2n}^{FT}, and B_{ns}, we explicitly derive in the following the constant F_{2n}^{FT} appearing in Eq. (3.28a).

The quasiprimary field T_{2n}^{F} is given by a relation analogous to Eq. (3.4) for T_{2n}^{T}, i.e.,

$$
(2n-2)! T_{2n}^{F}(z) = \lim_{z_1, z_2 \to z} D_{2n-2}^{(2,2)}(\partial_1, \partial_2) T_F(z_1, z_2), \qquad \text{(A4)}
$$

with $D_{2n-2}^{(2,2)}(\alpha,\beta)$ *and* $T_F(z_1,z_2)$ given by Eqs. (3.5a) and (3.22), respectively. Then, taking into account Eqs. (3.25) and (3.26), we have

$$
\begin{aligned}
\langle T(z_1) T(z_2) T_{2n}^{F}(z_3) \rangle = {}& \frac{c}{2} \frac{\Gamma(2n+1)[\Gamma(2n)]^2}{\Gamma(4n-1)} \lim_{z_4 \to z_3} \sum_{k=0}^{2n-2} \frac{(-1)^k}{\Gamma(k+2)\Gamma(2n-1-k)\Gamma(2n-k)\Gamma(k+1)} \\
& \times \partial_3^{2n-2-k} \partial_4^k \{ s z_{13}^{-2} z_{14}^{-2} z_{23}^{-2} z_{24}^{-2} z_{34}^{2} + 2 z_{12}^{-2} z_{13}^{-1} z_{14}^{-1} z_{23}^{-1} z_{24}^{-1} \}. \qquad \text{(A5)}
\end{aligned}
$$

Let us define

$$
(\mathrm{I}) := \lim_{z_4 \to z_3} \sum_{k=0}^{2n-2} \frac{(-1)^k}{\Gamma(k+2)\Gamma(2n-1-k)\Gamma(2n-k)\Gamma(k+1)} \cdot \partial_3^{2n-2-k} \partial_4^k [z_{13}^{-2} z_{23}^{-2} z_{14}^{-2} z_{24}^{-2} z_{34}^{2}] \qquad \text{(A6a)}
$$

and

$$
(\mathrm{II}) := \lim_{z_4 \to z_3} \sum_{k=0}^{2n-2} \frac{(-1)^k}{\Gamma(k+2)\Gamma(2n-1-k)\Gamma(2n-k)\Gamma(k+1)} \cdot \partial_3^{2n-2-k} \partial_4^k [z_{12}^{-2} z_{13}^{-1} z_{23}^{-1} z_{14}^{-1} z_{24}^{-1}]. \qquad \text{(A6b)}
$$

Then, applying (A3) for $r=2$

$$
\begin{aligned}
\partial_4^k [z_{14}^{-2} z_{24}^{-2} z_{34}^{2}] = {}& 2 z_{12}^{-3} (\delta_{ko} z_{14} - \delta_{k1}) + \delta_{ko} z_{12}^{-3} (z_{12} - 4 z_{13}) \\
& + 2\Gamma(k+1) z_{12}^{-3} z_{13} z_{23} z_{14}^{-1-k} + \Gamma(k+2) z_{12}^{-2} z_{13}^{2} z_{14}^{-2-k} + (1 \leftrightarrow 2), \qquad \text{(A7)}
\end{aligned}
$$

we find

$$
\begin{aligned}
(\mathrm{I}) = {}& \lim_{z_4 \to z_3} \sum_{k=0}^{2n-2} \frac{(-1)^k}{\Gamma(k+2)\Gamma(2n-1-k)\Gamma(2n-k)} \partial_3^{2n-2-k} [2 z_{12}^{-3} z_{14}^{-1-k} z_{13}^{-1} z_{23}^{-1} \\
& + (k+1) z_{12}^{-2} z_{14}^{-2-k} z_{23}^{-2}] + (1 \leftrightarrow 2) \\
= {}& -z_{12}^{-4} \sum_{t=1}^{2n-1} \frac{(-1)^t}{\Gamma(2n+1-t) t!} \left\{ 2 \left[z_{23}^{-2n} \left(\frac{z_{23}}{z_{13}} \right)^t - z_{13}^{-2n} \right] + z_{12}^{2} z_{13}^{-1} z_{23}^{-2n-1} t(2n-t) \left(\frac{z_{23}}{z_{13}} \right)^t \right\} + (1 \leftrightarrow 2) \\
= {}& \frac{4(2n+1)(n-1)}{\Gamma(2n+1)} z_{12}^{-4} \left(\frac{z_{12}}{z_{13} z_{23}} \right)^{2n}. \qquad \text{(A8)}
\end{aligned}
$$

Similarly we have

$$
\begin{aligned}
(\mathrm{II}) = {}& z_{12}^{-4} \sum_{k=0}^{2n-2} \frac{(-1)^k}{\Gamma(k+2)\Gamma(2n-k)} (z_{23}^{-2n+1+k} - z_{13}^{-2n+1+k})(z_{23}^{-k-1} - z_{13}^{-k-1}) \\
= {}& \frac{2}{\Gamma(2n+1)} z_{12}^{-4} \left(\frac{z_{12}}{z_{13} z_{23}} \right)^{2n}. \qquad \text{(A9)}
\end{aligned}
$$

Inserting Eqs. (A8) and (A9) into (A5), we obtain the result (3.28).

[1] A. A. Belavin, A. M. Polyakov, and A. B. Zamolodchikov, "Infinite conformal symmetry in two dimensional quantum field theory," Nucl. Phys. **B 241**, 333 (1984).

[2] Vl. D. Dotsenko and V. A. Fateev, "Conformal algebra and multi-point correlation functions in 2D statistical models," Nucl. Phys. B **240**, 312 (1984); "Four-point correlation functions and the operator algebra in 2D conformal invariant theories with central charge $c \leqslant 1$," Nucl. Phys. B **251**, 691 (1985); Phys. Lett. B **154**, 291 (1985).

[3] V. G. Kac, "Contravariant form for infinite dimensional Lie algebras and superalgebras," *Lecture Notes in Physics*, Vol. 94 (Springer, Berlin, 1979), pp. 441–445.

[4] D. Friedan, Z. Qiu, and S. Shenker, "Conformal invariance, unitarity, and

critical exponents in two dimensions," Phys. Rev. Lett. **52**, 1575 (1984).
[5]P. Goddard, A. Kent, and D. Olive, "Unitary representations of the Virasoro and super-Virasoro algebras," Commun. Math. Phys. **103**, 105 (1986).
[6]V. G. Kac and M. Wakimoto, "Unitarizable highest weight representations of the Virasoro, Neveu–Schwarz, and Ramond algebras," IHES preprint, Bures-sur-Yvette, November 1985 (to be published in Lect. Notes Phys.).
[7]C. Itzykson and J. B. Zuber, "Two-dimensional conformal invariant theories on a torus," Nucl. Phys. B **275**, 561 (1986).
[8]I. T. Todorov, "Current algebra approach to conformal invariant two-dimensional models," Phys. Lett. B **153**, 77 (1985); "Algebraic approach to conformal invariant 2-dimensional models," Bulg. J. Phys. **12**, 3 (1985); "Infinite Lie algebras in 2-dimensional conformal field theory," in *Differential Geometric Methods in Theoretical Physics*, edited by H. D. Doebner and T. D. Palev (World Scientific, Singapore, 1986), pp. 297–347.
[9]I. T. Todorov, "Infinite dimensional Lie algebras in conformal QFT models," ISAS preprint 35/86/EP Trieste (to be published in Lect. Notes Phys.).
[10]G. M. Sotkov, I. T. Todorov, and V. Yu. Trifonov, "Quasiprimary composite fields and operator product expansions in 2-dimensional conformal models," Lett. Math. Phys. **12**, 127 (1986).
[11]G. M. Sotkov, I. T. Todorov, M. S. Stanishkov, and V. Yu. Trifonov, "Higher symmetries in conformal QFT models," Lecture at the Symposium on Topological and Geometrical Methods in Field Theory, Espoo, Finland, June 1986 (World Scientific, Singapore, to be published).
[12]V. Trifonov, private communication and paper in preparation.
[13]A. B. Zamolodchikov (unpublished); D. Friedan, Z. Qiu, and S. Shenker, "Superconformal invariance in two dimensions and the tricritical Ising model," Phys. Lett. B **151**, 37 (1985).

Superinvariant chiral and vector fields

Haralambos Panagopoulos
Istituto di Fisica, Università di Pisa, Piazza Torricelli 2, Pisa 56100, Italy
Luc Vinet
Laboratoire de Physique Nucléaire, Université de Montréal, Montréal, Québec H3C 3J7, Canada

(Received 21 October 1986; accepted for publication 4 February 1987)

This work is concerned with the characterization of field supermultiplets on four-dimensional Minkowski space that are stable under the action of subgroups of the superconformal group SU(2,2/1). The most general scalar and vector superfields whose Lie derivative with respect to a fermionic tangent vector vanishes are determined. Invariance under subgroups of SU(2,2/1) with more than one odd generator is also discussed.

I. INTRODUCTION

The methods for characterizing tensor fields of various types[1] and connection one-forms[2] that are stable under ordinary space-time transformations are by now well understood. The same situation does not prevail, however, when supergroups are considered. In this paper, we initiate a study of superfields invariant under supersymmetry transformations. For definiteness, we shall consider field supermultiplets on four-dimensional Minkowski space and as a first step, we shall determine the most general chiral and vector superfields that are invariant under the transformations generated by the odd elements of the superconformal algebra SU(2,2/1).

The usefulness of ordinary invariant fields has long been recognized. They have been used in particular to obtain solutions to nonlinear field equations like the Einstein or Yang–Mills equations.[3] They are also instrumental in performing the dimensional reduction of theories formulated in higher dimensions.[4] Clearly, superinvariant fields should have similar applications.

Global, as well as infinitesimal, techniques for obtaining scalar densities, vector fields, and metrics that are invariant under (bosonic) subgroups of the conformal group are discussed in detail in Ref. 1. Spinor fields are studied in Ref. 5. Here, we therefore need to concentrate only on the transformations generated by the fermionic charges. As a rule, in characterizing invariant fields, global methods tend to be simpler when the invariance group is large, while infinitesimal techniques are to be preferred in the opposite case. We have found that supersymmetry imposes severe constraints and that nontrivial fields can rarely admit invariance subgroups with many odd generators; because of that (and for simplicity), we elected to use the infinitesimal approach. The most general fields whose Lie derivative with respect to a single ferionic generator vanishes will be made explicit first and the solution of the invariance conditions that result if two or more such derivatives are set to zero will be examined subsequently.

This paper is organized as follows. In Sec. II, we present some introductory material, establish the notation, and review the field representations of the superconformal algebra SU(2,2/1). Superinvariant chiral and vector fields are presented in Secs. III and IV, respectively. Concluding remarks will be found in Sec. V.

II. THE SUPERCONFORMAL ALGEBRA AND ITS FIELD REPRESENTATIONS

Supersymmetric massless field theories in Minkowski space are usually invariant under the superconformal group. Hence we have the motivation for looking for fields invariant under subgroups of this supergroup of space-time transformations. We shall give in this section a short but self-contained review of the $N=1$ superconformal algebra and of its action on superfields. Unless stated otherwise, we shall stick with the conventions of Wess and Bagger.[6] We shall therefore use the metric $g_{\mu\nu}=\mathrm{diag}(-1,1,1,1)$ and work with Weyl spinors in the Van der Waerden notation.

We shall, respectively, denote by $M_{\mu\nu}$, P_μ, D, and K_μ ($\mu=0,1,2,3$) the generators of infinitesimal homogeneous Lorentz transformations, translations, dilatations, and special conformal transformations. Together they close under the ordinary Lie bracket to form an algebra isomorphic to O(4,2), or equivalently SU(2,2).[1,7] As shown by Haag *et al.*,[8] this algebra can be extended to a superalgebra by adding two spinor charges Q_α and S_α ($\alpha=1,2$) and their Hermitian conjugates $\bar{Q}_{\dot\alpha}$ and $\bar{S}_{\dot\alpha}$. We shall refer to Q_α as the supertranslations generator and to S_α as the superconformal generator. To achieve closure under the graded Lie product, one further needs to introduce an additional bosonic generator Π which we shall call the chiral charge. The structure relations of the $N=1$ conformal superalgebra read as follows:

$$\begin{aligned}&\{Q_\alpha,\bar{Q}_{\dot\alpha}\}=2\sigma^\mu_{\alpha\dot\alpha}P_\mu,\quad \{S_\alpha,\bar{S}_{\dot\alpha}\}=-2\sigma^\mu_{\alpha\dot\alpha}K_\mu,\\ &\{Q_\alpha,\bar{S}_{\dot\alpha}\}=0,\quad \{Q_{\dot\alpha},S_\alpha\}=0,\\ &\{Q_\alpha,S^\beta\}=-2[\sigma^{\mu\nu}M_{\mu\nu}+D-2i\Pi]_\alpha{}^\beta,\\ &\{\bar{Q}^{\dot\alpha},\bar{S}_{\dot\beta}\}=2[\bar\sigma^{\mu\nu}M_{\mu\nu}+D+2i\Pi]^{\dot\alpha}{}_{\dot\beta};\end{aligned}\tag{2.1a}$$

$$\begin{aligned}&[Q_\alpha,D]=(i/2)Q_\alpha,\quad [\bar{Q}_{\dot\alpha},D]=(i/2)\bar{Q}_{\dot\alpha},\\ &[S_\alpha,D]=(-i/2)S_\alpha,\quad [\bar{S}_{\dot\alpha},D]=(-i/2)\bar{S}_{\dot\alpha},\\ &[Q_\alpha,P_\mu]=[\bar{Q}_{\dot\alpha},P_\mu]=0,\quad [S_\alpha,K_\mu]=[\bar{S}_{\dot\alpha},K_\mu]=0,\\ &[Q_\alpha,K^\mu]=-i\sigma^\mu_{\alpha\dot\alpha}\bar{S}^{\dot\alpha},\quad [\bar{Q}^{\dot\alpha},K_\mu]=i\bar\sigma^{\dot\alpha\alpha}_\mu S_\alpha,\\ &[S_\alpha,P^\mu]=i\sigma^\mu_{\alpha\dot\alpha}\bar{Q}^{\dot\alpha},\quad [\bar{S}^{\dot\alpha},P_\mu]=-i\bar\sigma^{\dot\alpha\alpha}_\mu Q_\alpha,\end{aligned}\tag{2.1b}$$

 0022-2488/87/071608-11$02.50

$$[Q_\alpha,\Pi]=\tfrac{3}{4}Q_\alpha,\quad [\bar Q_{\dot\alpha},\Pi]=-\tfrac{3}{4}\bar Q_{\dot\alpha},$$
$$[S_\alpha,\Pi]=-\tfrac{3}{4}S_\alpha,\quad [\bar S_{\dot\alpha},\Pi]=\tfrac{3}{4}\bar S_{\dot\alpha},$$
$$[M^{\mu\nu},Q_\alpha]=i(\sigma^{\mu\nu}Q)_\alpha,\quad [M^{\mu\nu},\bar Q^{\dot\alpha}]=i(\bar\sigma^{\mu\nu}\bar Q)^{\dot\alpha},$$
$$[M^{\mu\nu},S_\alpha]=i(\sigma^{\mu\nu}S)_\alpha,\quad [M^{\mu\nu},\bar S^{\dot\alpha}]=i(\bar\sigma^{\mu\nu}\bar S)^{\dot\alpha};$$
$$[P_\mu,P_\nu]=0,\quad [K_\mu,K_\nu]=0,$$
$$[P_\mu,K_\nu]=-2iM_{\mu\nu}+2ig_{\mu\nu}D,$$
$$[P_\mu,D]=iP_\mu,\quad [K_\mu,D]=-iK_\mu,\quad [M_{\mu\nu},D]=0,$$
$$[M_{\mu\nu},P_\rho]=-ig_{\mu\rho}P_\nu+ig_{\nu\rho}P_\mu,\tag{2.1c}$$
$$[M_{\mu\nu},K_\rho]=-ig_{\mu\rho}K_\nu+ig_{\nu\rho}K_\mu,$$
$$[M_{\mu\nu},M_{\rho\sigma}]=i(M_{\nu\rho}g_{\mu\sigma}-M_{\mu\rho}g_{\nu\sigma}+M_{\sigma\nu}g_{\mu\rho}-M_{\sigma\mu}g_{\nu\rho}),$$
$$[\Pi,P_\mu]=[\Pi,K_\mu]=[\Pi,M_{\mu\nu}]=[\Pi,D]=0.$$

The matrices $\sigma^\mu_{\alpha\dot\alpha}$ and $\bar\sigma^{\mu\dot\alpha\alpha}$ are defined by

$$\sigma^0=\begin{pmatrix}-1&0\\0&-1\end{pmatrix},\quad \sigma^1=\begin{pmatrix}0&1\\1&0\end{pmatrix},$$
$$\sigma^2=\begin{pmatrix}0&-i\\i&0\end{pmatrix},\quad \sigma^3=\begin{pmatrix}1&0\\0&-1\end{pmatrix},\tag{2.2a}$$
$$\bar\sigma^0=\sigma^0,\quad \bar\sigma^{1,2,3}=-\sigma^{1,2,3}.$$

We also used, for the generators of the Lorentz group in the spinor representation,

$$(\sigma^{\mu\nu})_\alpha{}^\beta=\tfrac{1}{4}(\sigma^\mu_{\alpha\dot\alpha}\bar\sigma^{\mu\dot\alpha\beta}-\sigma^\nu_{\alpha\dot\alpha}\bar\sigma^{\mu\dot\alpha\beta}),$$
$$(\bar\sigma^{\mu\nu})^{\dot\alpha}{}_{\dot\beta}=\tfrac{1}{4}(\bar\sigma^{\mu\dot\alpha\alpha}\sigma^\nu_{\alpha\dot\beta}-\bar\sigma^{\nu\dot\alpha\alpha}\sigma^\mu_{\alpha\dot\beta}).\tag{2.2b}$$

Useful identities involving these matrices can be found in Appendix A of Ref. 6. The above superalgebra is identified as SU(2,2/1), with SU(2,2) ⊕ U(1) as its bosonic subalgebra [see Eqs. (2.1c)]. It possesses an important subalgebra, namely, the 14-dimensional Poincaré superalgebra generated by P_μ, $M_{\mu\nu}$, Q_α, and $\bar Q_{\dot\alpha}$.

Minkowski superspace M can be defined as the coset super-Poincaré group/Lorentz group. The elements of this coset can be parametrized as

$$G(x,\theta,\bar\theta)=\exp[i(-x^\mu P_\mu+\theta^\alpha Q_\alpha-\bar\theta^{\dot\alpha}\bar Q_{\dot\alpha})].\tag{2.3}$$

Points of M are therefore labeled by the four space-time coordinates x^μ and four anticommuting spinor coordinates θ_α and $\bar\theta_{\dot\alpha}$. (Implicit in the definition of these coordinates is the existence of an underlying Grassmann algebra $\mathscr{I}$. The coordinates x take their values in the even part of $\mathscr{I}$ and θ and $\bar\theta$ belong to the odd part of $\mathscr{I}$.[9]) The action of the translations and supertranslations on superspace is defined by left multiplication. One finds that P_μ, Q_α, and $\bar Q_{\dot\alpha}$ are represented by the following vector fields:

$$Q_\alpha=\frac{\partial}{\partial\theta^\alpha}-i\sigma^\mu_{\alpha\dot\alpha}\bar\theta^{\dot\alpha}\partial_\mu,$$
$$\bar Q_{\dot\alpha}=-\frac{\partial}{\partial\bar\theta^{\dot\alpha}}+i\theta^\alpha\sigma^\mu_{\alpha\dot\alpha}\partial_\mu;\tag{2.4a}$$
$$P_\mu=i\,\partial_\mu.\tag{2.4b}$$

We shall also need subsequently the right translation vector fields; they are given by

$$D_\alpha=\frac{\partial}{\partial\theta^\alpha}+i\sigma^\mu_{\alpha\dot\alpha}\bar\theta^{\dot\alpha}\partial_\mu,$$
$$\bar D_{\dot\alpha}=-\frac{\partial}{\partial\bar\theta^{\dot\alpha}}-i\theta^\alpha\sigma^\mu_{\alpha\dot\alpha}\partial_\mu,\tag{2.5}$$

and satisfy the following anticommutation relations:

$$\{D_\alpha,\bar D_{\dot\alpha}\}=-2i\sigma^\mu_{\alpha\dot\alpha}\partial_\mu,$$
$$\{D_\alpha,D_\beta\}=\{\bar D_{\dot\alpha},\bar D_{\dot\beta}\}=0,\tag{2.6a}$$
$$\{D_\alpha,Q_\beta\}=\{D_\alpha,\bar Q_{\dot\beta}\}=\{\bar D_{\dot\alpha},Q_\alpha\}=\{\bar D_{\dot\alpha},\bar Q_{\dot\beta}\}=0.\tag{2.6b}$$

Superfields are multispinor functions $\phi_{\alpha\beta\cdots\dot\alpha\cdots}(x,\theta,\bar\theta)$ on superspace which transform under supersymmetry as coordinate scalars and Lorentz multispinors. They should be understood in terms of their power series expansions in θ and $\bar\theta$. In order to obtain their transformation properties under the action of the superconformal generators, we have used the method of induced representations. Here, the isotropy or little algebra has for its basis $M_{\mu\nu}$, D, Π, K_μ, S_α, and $\bar S_{\dot\alpha}$. Of these elements, only the first three can be nontrivially represented at the origin. Given the expressions (2.4) for P_μ, Q_α, and $\bar Q_{\dot\alpha}$, we arrived at the following realization of SU(2,2/1) in terms of differential operators acting on superfields:

$$S_\alpha=-x_\mu\sigma^\mu_{\alpha\dot\alpha}\bar Q^{\dot\alpha}-2i\theta\theta\frac{\partial}{\partial\theta^\alpha}-2i\theta_\alpha\bar\theta^{\dot\beta}\frac{\partial}{\partial\bar\theta^{\dot\beta}}+\sigma^\mu_{\alpha\dot\beta}\bar\theta^{\dot\beta}\theta\theta\,\partial_\mu+2\theta^\beta(\sigma^{\nu\mu}\Sigma_{\mu\nu}-d+2i\pi)_{\beta\alpha},$$
$$\bar S_{\dot\alpha}=x_\mu\epsilon_{\dot\alpha\dot\beta}\bar\sigma^{\mu\dot\beta\alpha}Q_\alpha-2i\bar\theta\bar\theta\frac{\partial}{\partial\bar\theta^{\dot\alpha}}+2i\bar\theta_{\dot\alpha}\theta^\alpha\frac{\partial}{\partial\theta^\alpha}-\epsilon_{\dot\alpha\dot\beta}\bar\sigma^{\mu\dot\beta\alpha}\theta_\alpha\bar\theta\bar\theta\,\partial_\mu-2\bar\theta_{\dot\beta}(\bar\sigma^{\nu\mu}\Sigma_{\mu\nu}-d-2i\pi)^{\dot\beta}{}_{\dot\alpha};\tag{2.7}$$

$$M_{\mu\nu}=i(x_\mu\,\partial_\nu-x_\nu\,\partial_\mu)+i\theta^\alpha(\sigma_{\nu\mu})_\alpha{}^\beta\frac{\partial}{\partial\theta^\beta}+i\bar\theta_{\dot\alpha}(\bar\sigma_{\nu\mu})^{\dot\alpha}{}_{\dot\beta}\frac{\partial}{\partial\bar\theta_{\dot\beta}}+\Sigma_{\mu\nu},$$
$$D=ix^\mu\,\partial_\mu+\frac{i}{2}\theta^\alpha\frac{\partial}{\partial\theta^\alpha}+\frac{i}{2}\bar\theta^{\dot\alpha}\frac{\partial}{\partial\bar\theta^{\dot\alpha}}+d,$$
$$K_\mu=-ix^2\,\partial_\mu+2ix_\mu x^\nu\,\partial_\nu+2x_\mu d+2x^\nu\Sigma_{\mu\nu}+ix_\mu\theta^\alpha\frac{\partial}{\partial\theta^\alpha}+ix_\mu\bar\theta^{\dot\alpha}\frac{\partial}{\partial\bar\theta^{\dot\alpha}}-2ix^\nu\Big(\theta\sigma_{\mu\nu}\frac{\partial}{\partial\theta}\Big)-2ix^\nu\Big(\bar\theta\bar\sigma_{\mu\nu}\frac{\partial}{\partial\bar\theta}\Big)\tag{2.8}$$
$$+\bar\theta\bar\theta\Big(\theta\sigma_\mu\frac{\partial}{\partial\bar\theta}\Big)+\theta\theta\Big(\bar\theta\bar\sigma_\mu\frac{\partial}{\partial\theta}\Big)-i\theta\theta\bar\theta\bar\theta\,\partial_\mu-4\pi(\bar\theta\bar\sigma_\mu\theta)+i\bar\theta(\bar\sigma^{\lambda\rho}\Sigma_{\lambda\rho})\bar\sigma_\mu\theta+i\bar\theta\bar\sigma_\mu(\sigma^{\lambda\rho}\Sigma_{\lambda\rho})\theta,$$
$$\Pi=\frac{3}{4}\theta^\alpha\frac{\partial}{\partial\theta^\alpha}-\frac{3}{4}\bar\theta^{\dot\alpha}\frac{\partial}{\partial\bar\theta^{\dot\alpha}}+\pi.$$

We have introduced above the SL(2,C) invariant antisymmetric symbols

$$\epsilon^{\alpha\beta} = \epsilon_{\beta\alpha} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \quad \epsilon^{\alpha\beta} = \epsilon^{\dot\alpha\dot\beta},$$

which define a metric on the space of spinors and are used to raise and lower indices. The following summation convention is also understood:

$$\psi\xi = \epsilon_{\alpha\beta}\psi^\alpha\xi^\beta = \psi^\alpha\xi_\alpha, \quad \bar\psi\bar\xi = \epsilon^{\dot\alpha\dot\beta}\bar\psi_{\dot\alpha}\bar\xi_{\dot\beta} = \bar\psi_{\dot\alpha}\bar\xi^{\dot\alpha}.$$

In Eqs. (2.7) and (2.8), $\Sigma_{\mu\nu}$, d, and π are, respectively, the matrix pieces of the generators $M_{\mu\nu}$, D, and Π. The eigenvalue d is the canonical dimension of the superfield. It is straightforward to check that the vector fields (2.4), (2.7), and (2.8) indeed satisfy the structure relations (2.2).

III. INVARIANT CHIRAL SUPERFIELDS

We shall consider in this section the invariance of chiral superfields under symmetry transformations. A superfield $\phi_{\beta\cdots\dot\beta\cdots}$ is said to be chiral if it satisfies the constraint

$$\bar D_{\dot\alpha}\phi_{\beta\cdots\dot\beta\cdots} = 0, \tag{3.1}$$

which amounts to the requirement that ϕ only depend on the variables y^μ and θ^α with

$$y^\mu = x^\mu + i\theta\sigma^\mu\bar\theta. \tag{3.2}$$

The most general (conformal) supertransformation can be specified by two anticommuting Weyl spinors ξ and ζ which represent a total of eight odd parameters. Let us denote by

$$\delta_{\xi,\zeta}\phi = (\xi Q + \bar\xi\bar Q + \zeta S + \bar\zeta\bar S)\phi \tag{3.3}$$

the corresponding transformation of a generic superfield. The chirality condition provides field multiplets that are irreducible[10] under (3.3) up to possible separation of real and imaginary parts (see Sec. IV). Since Q_α and $\bar Q_{\dot\alpha}$ anticommute with $\bar D_{\dot\alpha}$, supertranslations manifestly transform chiral fields into chiral fields. Now, we also have

$$\{S_\alpha,\bar D_{\dot\beta}\} = 4i\theta_\alpha\bar D_{\dot\beta}, \tag{3.4a}$$

$$\{\bar S_{\dot\alpha},\bar D_{\dot\beta}\} = 4i\bar\theta_{\dot\beta}\bar D_{\dot\alpha} + 2\epsilon_{\dot\beta\dot\gamma}(\bar\sigma^{\mu\nu}\Sigma_{\mu\nu} + d + 2i\pi)^{\dot\gamma}{}_{\dot\alpha}. \tag{3.4b}$$

Clearly, the second term on the rhs of (3.4b) must vanish when acting on ϕ for superconformal transformations to preserve condition (3.1). This proves to be the case for all known theories.

The representations (2.4a) and (2.7) of the supersymmetry generators take a simpler form when they act on chiral superfields. Indeed, in terms of the variables y and θ, one finds

$$Q_\alpha = \frac{\partial}{\partial\theta^\alpha}, \tag{3.5a}$$

$$\bar Q_{\dot\alpha} = 2i\theta^\alpha\sigma^\mu_{\alpha\dot\alpha}\frac{\partial}{\partial y^\mu}, \tag{3.5b}$$

$$S_\alpha = 2iy_\mu\sigma^\mu_{\alpha\dot\alpha}\bar\sigma^{\nu\dot\alpha\beta}\theta_\beta\frac{\partial}{\partial y^\nu} - 2i\theta\theta\frac{\partial}{\partial\theta^\alpha} + 2\theta^\beta(\sigma^{\nu\mu}\Sigma_{\mu\nu} - d + 2i\pi)_{\beta\alpha}, \tag{3.5c}$$

$$\bar S_{\dot\alpha} = y_\mu\epsilon_{\dot\alpha\dot\beta}\sigma^{\mu\dot\beta\alpha}\frac{\partial}{\partial\theta^\alpha} - 2\bar\theta_{\dot\beta}(\bar\sigma^{\nu\mu}\Sigma_{\mu\nu} - d - 2i\pi)^{\dot\beta}{}_{\dot\alpha}. \tag{3.5d}$$

We shall now concentrate more specifically on scalar fields. In this case the chirality condition entails the following decomposition for ϕ:

$$\begin{aligned}\phi(y,\theta) &= A(y) + \sqrt{2}\theta\psi(y) + \theta\theta F(y)\\ &= A(x) + i\theta\sigma^\mu\bar\theta\,\partial_\mu A(x) + \tfrac14\theta\theta\bar\theta\bar\theta\,\Box A(x)\\ &\quad + \sqrt{2}\theta\psi(x) - (i/\sqrt{2})\theta\theta\,\partial_\mu\psi(x)\sigma^\mu\bar\theta + \theta\theta F(x).\end{aligned} \tag{3.6}$$

Scalar fields have canonical dimension -1. In our notation it means that $d = i$. According to the previous discussion, this in turn imposes that $\pi = -\frac12$. The condition $\delta_{\xi,\zeta}\phi = 0$ that ϕ must satisfy in order to be invariant under the supertransformation parametrized by ξ and ζ can now easily be made explicit. Using Eqs. (3.5), one finds

$$\delta_{\xi,\zeta}\phi = \Big((\xi + \bar\zeta\bar\sigma^\mu y_\mu - 2i\theta\theta\zeta)\frac{\partial}{\partial\theta} - 2i(\bar\xi - \zeta\sigma^\nu y_\nu)\bar\sigma^\mu\theta\frac{\partial}{\partial y^\mu} - 2i\zeta\theta\Big)\phi = 0, \tag{3.7}$$

which can be resolved into components. One finds, in agreement with Wess and Zumino,[11] the following invariance conditions:

$$\delta_{\xi,\zeta}A = \sqrt{2}(\xi^\alpha + \bar\zeta_{\dot\alpha}\bar\sigma^{\mu\dot\alpha\alpha}y_\mu)\psi_\alpha = 0, \tag{3.8a}$$

$$\delta_{\xi,\zeta}\psi_\alpha = \sqrt{2}(\xi_\alpha - y_\mu\sigma^\mu_{\alpha\dot\alpha}\bar\zeta^{\dot\alpha})F + i\sqrt{2}\sigma^\mu_{\alpha\dot\alpha}(\bar\xi^{\dot\alpha} + y_\nu\bar\sigma^{\nu\dot\alpha\beta}\zeta_\beta)\frac{\partial}{\partial y^\mu}A - 2i\sqrt{2}\zeta_\alpha A = 0, \tag{3.8b}$$

$$\delta_{\xi,\zeta}F = -2i\sqrt{2}(\bar\xi_{\dot\alpha} - \zeta^\alpha\sigma^\nu_{\alpha\dot\alpha}y_\nu)\bar\sigma^{\mu\dot\alpha\beta}\frac{\partial}{\partial y^\mu}\psi_\beta = 0. \tag{3.8c}$$

We shall now proceed to solve Eqs. (3.8). To this end, we need to specify further the nature of the parameters ξ and ζ. We shall take them both proportional to a single real Grassmann parameter α; as will be discussed in Sec. V, superfields invariant under transformations associated to such parameters are the building blocks for the most general superinvariant fields. Factoring this generator α out of Eqs. (3.8) we can then consider parameters ξ and ζ as commuting Weyl spinors. In particular, we have $\xi^\alpha\xi_\alpha = 0$. Condition (3.8a), being purely algebraic, is immediately solved. One finds

$$\psi_\alpha = \epsilon_{\alpha\beta}(\xi^\beta + \bar\zeta_{\dot\beta}\bar\sigma^{\mu\dot\beta\beta}y_\mu)q = (\xi_\alpha - y_\mu\sigma^\mu_{\alpha\dot\alpha}\bar\zeta^{\dot\alpha})q, \tag{3.9}$$

with q a scalar function. Upon substituting (3.9) into (3.8c) one obtains

$$(\bar\xi - \zeta\sigma^\nu y_\nu)\bar\sigma^\mu(\xi - y_\rho\sigma^\rho\bar\zeta)\frac{\partial}{\partial y^\mu}q + 4(\bar\xi - \zeta\sigma^\nu y_\nu)\bar\zeta q = 0 \tag{3.10}$$

using $\sigma^\mu_{\alpha\dot\alpha}\bar\sigma^{\dot\beta\beta}_\mu = -2\delta_\alpha{}^\beta\delta^{\dot\beta}{}_{\dot\alpha}$. The solution to Eq. (3.10) can be obtained by the method of the characteristics.

Define the variables

$$r^\mu = (\bar\xi - \zeta\sigma^\nu y_\nu)\bar\sigma^\mu(\xi - y_\rho\sigma^\rho\bar\zeta). \tag{3.11}$$

The first step is to determine the integral curves $y^\mu(s)$ of the vector field $r^\mu\,\partial/\partial y^\mu$. This is done by solving

$$\frac{dy^\mu}{ds} = r^\mu. \tag{3.12}$$

Equations (3.11) and (3.12) are decoupled by going to the variables

$$\bar\zeta\bar\sigma^\mu\xi y_\mu,\quad \zeta\sigma^\mu\bar\xi y_\mu,\quad \xi\sigma^\mu\bar\xi y_\mu,$$

and (3.13)

$$\sigma=\bar\xi\bar\zeta+\zeta\xi-2\zeta\sigma^\mu\bar\zeta y_\mu,$$

which satisfy the following differential equations:

$$\frac{d\sigma}{ds}=\sigma^2-(\bar\xi\bar\zeta-\zeta\xi)^2, \tag{3.14a}$$

$$\frac{d}{ds}(\bar\zeta\bar\sigma^\mu\xi y_\mu)=(\bar\zeta\bar\sigma^\mu\xi y_\mu)(\sigma+\bar\xi\bar\zeta-\zeta\xi), \tag{3.14b}$$

$$\frac{d}{ds}(\zeta\sigma^\mu\bar\xi y_\mu)=(\zeta\sigma^\mu\bar\xi y_\mu)(\sigma+\zeta\xi-\bar\xi\bar\zeta), \tag{3.14c}$$

$$\frac{d}{ds}(\xi\sigma^\mu\bar\xi y_\mu)=-2(\xi\sigma^\nu\bar\zeta y_\nu)(\zeta\sigma^\rho\bar\xi y_\rho). \tag{3.14d}$$

The integration of Eqs. (3.14) yields

$$\sigma=-(\bar\xi\bar\zeta-\zeta\xi)\coth(\bar\xi\bar\zeta-\zeta\xi)s, \tag{3.15a}$$

$$\bar\zeta\bar\sigma^\mu\xi y_\mu=-2ic_1/\{1-\exp[-2(\bar\xi\bar\zeta-\zeta\xi)s]\}, \tag{3.15b}$$

$$\zeta\sigma^\mu\bar\xi y_\mu=-2ic_1^*/\{1-\exp[-2(\bar\xi\bar\zeta-\zeta\xi)s]\}, \tag{3.15c}$$

$$\xi\sigma^\mu\bar\xi y_\mu=c_2-2c_1c_1^*\sigma/(\bar\xi\bar\zeta-\zeta\xi)^2, \tag{3.15d}$$

with c_1, c_1^*, and c_2 integration constants. (The special case $\zeta\xi=0$ will be treated separately.) Eliminating the parameter s, we thus find that the following functions are constant along the curves generated by $r^\mu(\partial/\partial y^\mu)$:

$$c_1=[i\bar\zeta\bar\sigma^\mu\xi y_\mu/(\bar\xi\bar\zeta-\zeta\xi-\sigma)](\bar\xi\bar\zeta-\zeta\xi), \tag{3.16a}$$

$$c_1^*=[-i\zeta\sigma^\mu\bar\xi y_\mu/(\bar\xi\bar\zeta-\zeta\xi+\sigma)](\bar\xi\bar\zeta-\zeta\xi), \tag{3.16b}$$

$$c_2=\xi\sigma^\mu\bar\xi y_\mu-\frac{2\sigma(\bar\zeta\bar\sigma^\mu\xi)y_\mu(\zeta\sigma^\nu\bar\xi)y_\nu}{(\bar\xi\bar\zeta-\zeta\xi)^2-\sigma^2}. \tag{3.16c}$$

Now we see that Eq. (3.10) can be written as

$$\frac{dq}{ds}+2(\bar\xi\bar\zeta-\zeta\xi+\sigma)q=0, \tag{3.17}$$

which is easily integrated. After elimination of s, one finds

$$q=[1/(\bar\xi\bar\zeta-\zeta\xi-\sigma)^2]Q(c_1,c_1^*,c_2), \tag{3.18}$$

where Q is an arbitrary complex function of c_1, c_1^*, and c_2. From (3.9), we see that this completely determines ψ, which is thus found to be

$$\psi=[(\xi-y_\mu\sigma^\mu\bar\zeta)/(\bar\xi\bar\zeta-\zeta\xi-\sigma)^2]Q(c_1,c_1^*,c_2). \tag{3.19}$$

Now only (3.8b) remains to be solved. By contracting with $(\xi^\alpha+\bar\zeta_{\dot\alpha}\bar\sigma^{\nu\dot\alpha\alpha}y_\nu)$ we obtain the following for A:

$$(\xi+\bar\zeta\bar\sigma^\nu y_\nu)\sigma^\mu(\bar\xi+y_\rho\bar\sigma^\rho\zeta)\frac{\partial A}{\partial y^\mu}-2(\xi\zeta+\bar\zeta\bar\sigma^\mu\zeta y_\mu)A=0. \tag{3.20}$$

It is not difficult to see that (3.20) can be rewritten in the form

$$\frac{dA}{ds}=(-\sigma+\bar\xi\bar\zeta-\zeta\xi)A \tag{3.21}$$

using the same variables as before.

The solution is simply

$$A=[1/(\sigma+\bar\xi\bar\zeta-\zeta\xi)]\mathscr{A}(c_1,c_1^*,c_2), \tag{3.22}$$

with $\mathscr{A}$ an arbitrary function of its arguments. In order to obtain F, one contracts Eq. (3.8b) with ζ^α to find that

$$F=\frac{-i}{\zeta(\xi-y_\mu\sigma^\mu\bar\zeta)}\left(\zeta\sigma^\mu(\bar\xi+y_\nu\bar\sigma^\nu\zeta)\frac{\partial}{\partial y^\mu}A\right). \tag{3.23}$$

Substituting the expression (3.21) for A, one arrives at the following formula for F:

$$F=\frac{-2\zeta\xi}{(\bar\xi\bar\zeta-\zeta\xi-\sigma)^2}\left((\bar\xi\bar\zeta-\zeta\xi)\frac{\partial}{\partial c_1}\mathscr{A}-2c_1^*\frac{\partial}{\partial c_2}\mathscr{A}\right). \tag{3.24}$$

The superinvariant chiral multiplet (A,ψ,F) thus obtained can be written in a compact way by returning to the superfield formalism. Observe that $A(y)+\theta\theta F(y)$ can be expressed as $A(z)$ with

$$z^\mu=y^\mu-\theta\theta i\zeta\sigma^\mu(\bar\xi+y_\rho\bar\sigma^\rho\zeta)/\zeta(\xi-y_\nu\sigma^\nu\bar\zeta). \tag{3.25}$$

Then,

$$\begin{aligned}\phi(y,\theta)&=A(y)+\sqrt{2}\theta\psi(y)+\theta\theta F(y)\\&=A(z)+\sqrt{2}\theta\psi(z)\\&=\frac{1}{\sigma(z)+\bar\xi\bar\zeta-\zeta\xi}\mathscr{A}[c_1(z),c_1^*(z),c_2(z)]\\&\quad+\sqrt{2}\theta\frac{(\xi-z_\mu\sigma^\mu\bar\zeta)}{(\sigma(z)-\bar\xi\bar\zeta+\zeta\xi)^2}\\&\quad\times Q[c_1(z),c_1^*(z),c_2(z)]\\&=\frac{1}{\sigma(z)+\bar\xi\bar\zeta-\zeta\xi}\mathscr{A}[c_1(z),c_1^*(z),c_2(z),\gamma(z)],\end{aligned} \tag{3.26}$$

where

$$\begin{aligned}\gamma(z)&=\sqrt{2}\theta(\xi-z_\mu\sigma^\mu\bar\zeta)(\sigma(z)+\bar\xi\bar\zeta-\zeta\xi)\\&\quad\times(\sigma(z)-\bar\xi\bar\zeta+\zeta\xi)^{-2}.\end{aligned} \tag{3.27}$$

Thus we see that ϕ initially defined on chiral superspace with four bosonic and two fermionic coordinates has been constrained by the invariance condition (3.7) to a superspace with three bosonic and one fermionic coordinates. Note that these three bosonic variables have a nontrivial θ dependence in contradistinction with what occurred in standard dimensional reductions of supersymmetric theories.

We now consider the case where $\zeta\xi=0$. This situation occurs when we consider invariance under a pure supertranslation ($\zeta=0$), a pure superconformal transformation ($\xi=0$), or a transformation with $\zeta=m\xi$, $m\in\mathbb{C}\neq0$. In these instances, it is obvious that the set of variables (3.13) is no longer adequate to decouple the characteristic equations since it then consists of a single variable. A special treatment is thus necessary. We shall first take $\zeta=m\xi$, $m\neq0$. In this case, the characteristic equations $dy^\mu/ds=r^\mu$ can be decoupled by adjoining to the variable

$$\sigma=-2|m|^2\xi\sigma^\mu\bar\xi y_\mu \tag{3.28}$$

the following spinorial variable:

$$\bar\chi=\bar\xi-m\xi\sigma^\mu y_\mu. \tag{3.29}$$

We already have

$$\frac{d\sigma}{ds}=\sigma^2 \tag{3.30a}$$

and it is not difficult to check that $\bar{\chi}$ satisfies

$$\frac{d\bar{\chi}}{ds} = \sigma\bar{\chi}. \tag{3.30b}$$

Integration of Eqs. (3.30) is immediate and yields

$$\sigma = -1/s, \quad \bar{\chi} = (1/s)\bar{\chi}_0, \tag{3.31}$$

with $\bar{\chi}_0$ a constant Weyl spinor. To eliminate the parameter s we make use of the fact that when $\zeta = m\xi$,

$$\frac{d}{ds} y^2 = 2y^\mu r_\mu = -\frac{1}{|m|^2}(1 - |m|^2 y^2)\sigma. \tag{3.32}$$

[In deriving (3.32) we have used the identities

$$(\sigma^\mu\bar{\sigma}^\nu + \sigma^\nu\bar{\sigma}^\mu)_\alpha{}^\beta = -2g^{\mu\nu}\delta_\alpha{}^\beta,$$

$$(\bar{\sigma}^\mu\sigma^\nu + \bar{\sigma}^\nu\sigma^\mu)^{\dot\alpha}{}_{\dot\beta} = -2g^{\mu\nu}\delta^{\dot\alpha}{}_{\dot\beta},$$

and

$$\sigma^\nu\bar{\sigma}^\mu\sigma^\rho + \sigma^\rho\bar{\sigma}^\mu\sigma^\nu = 2(g^{\nu\rho}\sigma^\mu - g^{\mu\rho}\sigma^\nu - g^{\mu\nu}\sigma^\rho).]$$

Equation (3.32) can be integrated using (3.31); we thus find that

$$c = [s(1 - |m|^2y^2)]^{-1} \tag{3.33}$$

is invariant along an integral curve of our generator. This allows us to take as characteristic variables the functions

$$c = -\frac{\sigma}{1 - |m|^2y^2} = \frac{2|m|^2\xi\sigma^\mu\bar{\xi}y_\mu}{1 - |m|^2y^2}, \tag{3.34a}$$

$$c\bar{\chi}_0 = \frac{\bar{\chi}}{1 - |m|^2y^2} = \frac{\bar{\xi} - m\xi\sigma^\mu y_\mu}{1 - |m|^2y^2}. \tag{3.34b}$$

While this apparently gives five real variables, it should observe that only three are independent (as it should be) in view of the linear relation

$$c\bar{\chi}_0\bar{\xi} = -m\frac{\xi\sigma^\mu\bar{\xi}y_\mu}{1 - |m|^2y^2} = -\frac{1}{2m^*}c. \tag{3.35}$$

From here the solution of the invariance equation (3.8), specialized to the case $\zeta = m\xi$, proceeds exactly as in the generic situation. We thus obtain

$$\psi = [(\xi - m^*\sigma^\mu\bar{\xi}y_\mu)/(1 - |m|^2y^2)^2]Q(c,c\bar{\chi}_0), \tag{3.36a}$$

$$A = [(1 - |m|^2y^2)]^{-1}\mathscr{A}(c,c\bar{\chi}_0), \tag{3.36b}$$

$$F = -\frac{2m}{\sigma} i[\xi\sigma^\mu\bar{\xi} + \xi\sigma^\mu\bar{\sigma}^\nu\xi y_\nu]\frac{\partial}{\partial y^\mu}A, \tag{3.36c}$$

where Q and $\mathscr{A}$ are arbitrary functions of their arguments. The cases $\xi = 0$ and $\zeta = 0$ are treated analogously. When $\zeta = 0$, i.e., when we look for invariance under a supertranslation, the following functions may serve as characteristic variables:

$$\bar{\rho}_{\dot\alpha} = \xi^\alpha\sigma^\mu_{\alpha\dot\alpha}y_\mu. \tag{3.37}$$

Note again that together these functions represent a set of three linearly independent real variables. The invariant chiral multiplet is given by

$$\psi = \xi Q(\bar{\rho}_{\dot\alpha}), \tag{3.38a}$$

$$A = \mathscr{A}(\bar{\rho}_{\dot\alpha}), \tag{3.38b}$$

$$F = -i\frac{\varphi\sigma^\mu\bar{\xi}}{\varphi\xi}\frac{\partial}{\partial y^\mu}A, \tag{3.38c}$$

where Q and $\mathscr{A}$ are arbitrary functions and φ is an arbitrary constant spinor such that $\psi\xi \neq 0$.

When invariance under a pure superconformal transformation is considered, i.e., when $\xi = 0$, we can take as characteristic variables

$$\bar{\rho}_{\dot\alpha} = (\zeta\sigma^\mu)_{\dot\alpha}y_\mu/y^2 \tag{3.39}$$

and the invariant fields take the form, in analogy with (3.38),

$$\psi = -[\sigma^\mu\bar{\zeta}y_\mu/(y^2)^2]Q(\bar{\rho}_{\dot\alpha}), \tag{3.40a}$$

$$A = (1/y^2)\mathscr{A}(\bar{\rho}_{\dot\alpha}), \tag{3.40b}$$

$$F = -i\frac{\bar{\varphi}\bar{\sigma}^\mu\zeta}{\bar{\varphi}\bar{\zeta}}\frac{\partial}{\partial y^\mu}A. \tag{3.40c}$$

This concludes the characterization of scalar superfields whose Lie derivative with respect to a single fermionic tangent vector vanishes. We shall now examine the constraints that result when we require these fields to possess further supersymmetries. Let the Weyl spinors ξ' and ζ' parametrize another superconformal transformation in addition to the one considered until now. Remember that the invariance conditions for the fermionic field ψ were of the form

$$\bar{\varphi}_{\dot\alpha}\bar{\sigma}^{\mu\dot\alpha\beta}\frac{\partial}{\partial y^\mu}\psi_\beta = 0, \tag{3.41}$$

with $\varphi^\alpha = \xi^\alpha + \bar{\zeta}_{\dot\alpha}\bar{\sigma}^{\mu\dot\alpha\alpha}y_\mu$. From the first of these equations we had ψ_α proportional to φ_α. Requiring ψ_α to satisfy similar equations, with ξ and ζ replaced by ξ' and ζ', forces ψ to vanish, unless $\varphi'\varphi = 0$, that is, unless $\varphi' = \alpha\varphi$, with α a complex function. We see indeed that if ψ satisfies the above equations for some given ξ and ζ, it also automatically verifies the invariance condition associated to the supertransformation with parameters

$$\xi' = i\xi, \quad \zeta' = -i\zeta. \tag{3.42}$$

Any other supersymmetry, however, requires ψ to be trivial. The origin of the fact that $\delta_{i\xi,-i\zeta}A = \delta_{i\xi,-i\zeta}F = 0$ if $\delta_{\xi,\zeta}A = \delta_{\xi,\zeta}F = 0$ is easily found. Note that with $\delta_\pi\phi = \Pi\phi$,

$$\begin{aligned}[\delta_{\xi,\zeta},\delta_\pi] &= [\xi Q + \bar{\xi}\bar{Q} + \zeta S + \bar{\zeta}\bar{S},\Pi] \\ &= \tfrac{3}{4}(\xi Q - \bar{\xi}\bar{Q} - \zeta S - \bar{\zeta}\bar{S}) = -\tfrac{3}{4}i\delta_{i\xi,-i\zeta}.\end{aligned} \tag{3.43}$$

Now when acting on chiral superfields Π takes the form

$$\Pi = \frac{3}{4}\theta^\alpha\frac{\partial}{\partial\theta^\alpha} - \frac{1}{2}. \tag{3.44}$$

From $\delta_\pi\phi = \Pi\phi$ we then get the following chirality transformation for the component fields:

$$\delta_\pi A = -\tfrac{1}{2}A, \quad \delta_\pi\psi = \tfrac{1}{4}\psi, \quad \delta_\pi F = F. \tag{3.45}$$

Since $\delta_{\xi,\zeta}A$ and $\delta_{\xi,\zeta}F$ are linear expressions in ψ only, it is then obvious that the invariance conditions $[\delta_{\xi,\zeta},\delta_\pi](A$ or $F) = 0$ are identical to (3.8a) and (3.8b). Requiring that $\delta_{i\xi,-i\zeta}\psi = 0$ in addition to $\delta_{\xi,\zeta}\psi = 0$ imposes constraints on A and F which are easily solved. For instance, in the generic case $\xi\zeta \neq 0$, it implies that $F = 0$, and from (3.24) we then see that the function $\mathscr{A}$ of (3.22) must now only depend on c_1^* and $c_2 + 2c_1c_1^*/(\bar{\xi}\bar{\zeta} - \zeta\xi)$.

Consider now an additional supersymmetry other than $[\delta_{\xi,\zeta},\delta_\pi]$. Let us then examine for simplicity invariance under two supertranslations with parameters ξ and ξ' such that $\xi'\xi \neq 0$. We know that $\psi = 0$. The only remaining invariance equations are

$$\xi F + i\sigma^\mu \bar{\xi} \frac{\partial}{\partial y^\mu} A = 0, \quad \xi' F + i\sigma^\mu \bar{\xi}' \frac{\partial}{\partial y^\mu} A = 0, \qquad (3.46)$$

which imply that A has vanishing Lie derivative with respect to the following three vector fields:

$$\xi\sigma^\mu\bar{\xi}\frac{\partial}{\partial y^\mu}, \quad \xi'\sigma^\mu\bar{\xi}'\frac{\partial}{\partial y^\mu}, \quad (\xi'\sigma^\mu\bar{\xi} + \xi\sigma^\mu\bar{\xi}')\frac{\partial}{\partial y^\mu}. \qquad (3.47)$$

When $\xi'\xi \neq 0$, it is not difficult to check that these are linearly independent. The invariant A will thus be an arbitrary function of a single variable, namely,

$$A = \mathscr{A}((\xi'\sigma^\mu\bar{\xi} - \xi\sigma^\mu\bar{\xi}')y_\mu) \qquad (3.48)$$

and the corresponding F will be

$$F = -2i\bar{\xi}\bar{\xi}'\mathscr{A}', \qquad (3.49)$$

where $\mathscr{A}'$ denotes the derivative of $\mathscr{A}$. We may add that invariance under three supertranslations can only be achieved for a trivial constant field A, as the reader will easily convince himself.

As a final example we shall determine the most general scalar supermultiplet which is invariant under the full de Sitter supergroup.[12] This subgroup of SU(2,2/1) has for fermionic generators the following four elements: $Q_\alpha + mS_\alpha$, $\bar{Q}_{\dot\alpha} + m^*\bar{S}_{\dot\alpha}$. We want to require the invariance under the transformations generated by these operators. This amounts to setting $\zeta = m\xi$ in (3.8) and considering simultaneously the invariance conditions associated to the parameters ξ, $i\xi$, ξ', and $i\xi'$, with $\xi'\xi \neq 0$. Once again $\psi = 0$. From the invariance condition (3.8b) associated to the parameters ξ and $i\xi$ we find the following equations:

$$\xi_\beta\left(F\delta_\alpha{}^\beta + im\sigma^\mu_{\alpha\dot\alpha}y_\nu\bar{\sigma}^{\nu\dot\alpha\beta}\frac{\partial}{\partial y^\mu}A - 2im\delta_\alpha{}^\beta A\right) = 0, \qquad (3.50a)$$

$$\bar{\xi}^{\dot\alpha}\left(-m^*y_\mu\sigma^\mu_{\alpha\dot\alpha}F + i\sigma^\mu_{\alpha\dot\alpha}\frac{\partial}{\partial y^\mu}A\right) = 0. \qquad (3.50b)$$

Separating out the independent parts we obtain, for A and F, the equations

$$(F - 2imA) - imy^\mu\frac{\partial}{\partial y^\mu}A = 0, \qquad (3.51a)$$

$$y_\nu\sigma^{\mu\nu}\frac{\partial}{\partial y^\mu}A = 0, \qquad (3.51b)$$

$$i\frac{\partial}{\partial y^\mu}A - m^*y_\mu F = 0, \qquad (3.51c)$$

which are independent of ξ. Solution of (3.51) will therefore give the super de Sitter invariant multiplet. In fact it is found to be

$$A = c/(1 - |m|^2y^2), \quad F = 2im\,c/(1 - |m|^2y^2)^2, \qquad (3.52)$$

with c an arbitrary complex constant. In superfield notation we may also write this chiral multiplet in the form

$$\phi(y,\theta) = c/(1 - |m|^2y^2 - 2im\theta^2). \qquad (3.53)$$

This concludes our discussion of superinvariant scalar superfields. In Sec. IV, we shall study the case of vector superfields.

IV. INVARIANT VECTOR SUPERFIELDS

In the context of supersymmetric theories, gauge fields can be described by Hermitian superfields $V(x^\mu,\theta,\bar\theta)$, $V^\dagger = V$, taking values in the algebra of the gauge group. In the present investigation, we will restrict our attention to the Abelian case.

In terms of its components, one writes $V(x^\mu,\theta,\bar\theta)$ as

$$V(x^\mu,\theta,\bar\theta) = C(x) + i\theta\chi(x) - i\bar\theta\bar\chi(x) + (i/2)\theta\theta(M(x) + iN(x)) - (i/2)\bar\theta\bar\theta(M(x) - iN(x)) - \theta\sigma^\mu\bar\theta v_\mu(x)$$
$$+ i\theta\theta\bar\theta(\bar\lambda(x) + (i/2)\bar\sigma^\mu\partial_\mu\chi(x)) + i\bar\theta\bar\theta\theta(\lambda(x) + (i/2)\sigma^\mu\partial_\mu\bar\chi(x)) + \tfrac{1}{2}\theta\theta\bar\theta\bar\theta(D(x) + \tfrac{1}{2}\Box C(x)), \qquad (4.1)$$

where, for Hermiticity, the functions C, D, M, N, and v_μ are real. Acting on $V(x^\mu,\theta,\bar\theta)$, the superconformal generators again take the form (2.4), (2.7), and (2.8), with $\Sigma_{\mu\nu} = 0$. In addition, to preserve Hermiticity, we require that

$$[(\zeta S + \bar\zeta\bar S)V]^\dagger = (\zeta S + \bar\zeta\bar S)V, \qquad (4.2)$$

which sets $\mathrm{Re}(d) = \mathrm{Re}(\pi) = 0$. Another consideration which further restricts the values of canonical dimension and chirality to $d = \pi = 0$ is that of gauge invariance.

A (super) gauge transformation effected by

$$V \to V + \phi + \phi^\dagger, \quad \bar{D}\phi = 0, \qquad (4.3)$$

with ϕ and arbitrary scalar superfield as in Eq. (3.6), leaves invariant the component D and the photino field λ, while it changes the photon field v_μ by an ordinary gauge transformation

$$v_\mu \to v_\mu - i\,\partial_\mu(A - A^*), \qquad (4.4)$$

so that the field strength $v_{\mu\nu} = \partial_\mu v_\nu - \partial_\nu v_\mu$ is gauge invariant. Now, for gauge invariance to be compatible with superconformal symmetry, it must be that a superconformal transformation gives no gauge-dependent contributions to the gauge invariant fields. To illustrate this, consider the superconformal variation of the photon field:

$$\delta_{\xi,\zeta}(-\theta\sigma^\mu\bar\theta v_\mu)$$
$$\equiv(\xi Q - \bar\xi\bar Q + \zeta S + \bar\zeta\bar S)V|_{\theta\bar\theta}$$
$$= (-\theta\sigma^\mu\bar\theta)[i(\bar\xi - \zeta\sigma^\nu x_\nu)\bar\sigma_\mu\lambda + i(\xi - \bar\zeta\bar\sigma^\nu x_\nu)\sigma_\mu\bar\lambda$$
$$+ \partial_\mu((\xi - \zeta\sigma^\nu x_\nu)\bar\chi + (\xi + \bar\zeta\bar\sigma^\nu x_\nu)\chi)$$
$$- i(-d + 2i\pi)\zeta\sigma_\mu\bar\chi - i(d + 2i\pi)\bar\zeta\bar\sigma_\mu\chi],$$

implying

$$\delta_{\xi,\zeta}v_{\mu\nu} = \partial_\mu[i(\bar\xi - \zeta\sigma^\rho x_\rho)\bar\sigma_\nu\lambda + i(\xi + \bar\zeta\bar\sigma^\rho x_\rho)\sigma_\nu\bar\lambda$$
$$- i(-d + 2i\pi)\zeta\sigma_\mu\bar\chi - i(d + 2i\pi)\bar\zeta\bar\sigma_\mu\chi]$$
$$- (\mu\leftrightarrow\nu). \qquad (4.5)$$

The part of $\delta v_{\mu\nu}$ coming from χ, being gauge dependent and at the same time not removable by a gauge transformation on $v_{\mu\nu}$, cannot be tolerated; we therefore must require that $d = \pi = 0$. As for the chiral scalar field, we thus recover the correct canonical dimension and chirality for the photon field.

Having fixed the action of S and $\bar{S}$ on V, we now present it in component notation:

$$\begin{aligned}
\delta_{\xi,\zeta} C &= i(\xi + \bar{\zeta}\bar{\sigma}^\mu x_\mu)\chi - i(\bar{\xi} - \zeta\sigma^\mu x_\mu)\bar{\chi},\\
\delta_{\xi,\zeta}\chi &= (\xi - x_\mu\sigma^\mu\bar{\zeta})(M + iN) + i\sigma^\nu(\bar{\xi} + x_\mu\bar{\sigma}^\mu\zeta)v_\nu\\
&\quad + \sigma^\nu(\bar{\xi} + x_\mu\bar{\sigma}^\mu\zeta)\partial_\nu C,\\
\delta_{\xi,\zeta}(M + iN) &= 2(\bar{\xi} - \zeta\sigma^\mu x_\mu)\bar{\lambda}\\
&\quad + 2i(\bar{\xi} - \zeta\sigma^\mu x_\mu)\bar{\sigma}^\nu\,\partial_\nu\chi - 4i\zeta\chi,\\
\delta_{\xi,\zeta} v_\mu &= i(\bar{\xi} - \zeta\sigma^\nu x_\nu)\bar{\sigma}_\mu\lambda + i(\xi + \bar{\zeta}\bar{\sigma}^\nu x_\nu)\sigma_\mu\bar{\lambda}\\
&\quad + \partial_\mu[(\bar{\xi} - \zeta\sigma^\nu x_\nu)\bar{\chi} + (\xi + \bar{\zeta}\bar{\sigma}^\nu x_\nu)\chi],\\
\delta_{\xi,\zeta}\lambda &= i(\xi - x_\mu\sigma^\mu\bar{\zeta})D\\
&\quad - \tfrac{1}{2}\sigma^\rho\bar{\sigma}^\nu(\xi - x_\mu\sigma^\mu\bar{\zeta})(\partial_\nu v_\rho - \partial_\rho v_\nu),\\
\delta_{\xi,\zeta} D &= (\bar{\xi} - \zeta\sigma^\mu x_\mu)\bar{\sigma}^\nu\,\partial_\nu\lambda - (\xi + \bar{\zeta}\bar{\sigma}^\mu x_\mu)\sigma^\nu\,\partial_\nu\bar{\lambda}.
\end{aligned} \tag{4.6}$$

We observe that indeed all the gauge invariant components $v_{\mu\nu}$, λ, and D transform among themselves only. To study these components one introduces a superfield W_α given by

$$W_\alpha = -\tfrac{1}{4}\bar{D}\bar{D}D_\alpha V. \tag{4.7}$$

By construction this field obeys

$$\bar{D}_\beta W_{\dot\alpha} = 0, \quad \bar{D}_{\dot\alpha}\bar{W}^{\dot\alpha} = D^\alpha W_\alpha. \tag{4.8}$$

Being chiral, W_α can be expressed as a function of the chiral superspace coordinates (y^μ, θ_α), yielding the following component expansion:

$$\begin{aligned}
W_\alpha &= -i\lambda_\alpha(y) + D(y)\theta_\alpha\\
&\quad - (i/2)(\partial_\mu v_\nu(y) - \partial_\nu v_\mu(y))(\sigma^\mu\bar{\sigma}^\nu\theta)_\alpha\\
&\quad + \theta\theta(\sigma^\mu\,\partial_\mu\bar{\lambda}(y))_\alpha.
\end{aligned} \tag{4.9}$$

We see that W_α precisely contains all the gauge invariant components of V; it becomes particularly useful in its generalization for non-Abelian gauge groups. In our case, working with fields defined over chiral superspace is a further advantage, since the problem of finding superconformally invariant configurations entails computing characteristic curves inside a more economical superspace, as in Sec. III. Again one must find the constants $\Sigma_{\mu\nu}$, π, d and $\bar{\Sigma}_{\mu\nu}$, $\bar{\pi}$, $\bar{d}$ appropriate to W_α and $\bar{W}^{\dot\alpha}$. Requiring compatibility with constraints (4.8),

$$\begin{aligned}
&\bar{D}_\beta[(\zeta S + \bar{\zeta}\bar{S})_\alpha{}^\beta W_\beta]) = 0,\\
&D_\beta[(\zeta S + \bar{\zeta}\bar{S})^{\dot\alpha}{}_\beta \bar{W}^\beta] = 0,\\
&D^\alpha[(\zeta S + \bar{\zeta}\bar{S})_\alpha{}^\beta W_\beta] = \bar{D}_{\dot\alpha}[(\zeta S + \bar{\zeta}\bar{S})^{\dot\alpha}{}_\beta \bar{W}^\beta]
\end{aligned} \tag{4.10}$$

[the additional indices on $\zeta S + \bar{\zeta}\bar{S}$ come from nonzero values of $(\Sigma_{\mu\nu})_\alpha{}^\beta$, $(\bar{\Sigma}_{\mu\nu})^{\dot\alpha}{}_\beta$ when acting on W_α, $\bar{W}^{\dot\alpha}$, respectively] completely fixes these constants to

$$\begin{aligned}
&\Sigma_{\mu\nu} = -i\sigma_{\mu\nu}, \quad \bar{\Sigma}_{\mu\nu} = -i\bar{\sigma}_{\mu\nu},\\
&d = \bar{d} = 3i/2, \quad \pi = -\bar{\pi} = -\tfrac{3}{4}.
\end{aligned} \tag{4.11}$$

Thus the action of S and $\bar{S}$ on W_α is given by

$$\begin{aligned}
(S_\alpha)_\beta{}^\gamma W_\gamma &= \Big(2i(\sigma^\mu\bar{\sigma}^\nu\theta)_\alpha y_\mu\frac{\partial}{\partial y^\nu} - 2i\theta\theta\frac{\partial}{\partial\theta^\alpha} - 6i\theta_\alpha\Big)W_\beta\\
&\quad + 2i\epsilon_{\beta\alpha}\theta^\gamma W_\gamma + 2i\theta_\beta W_\alpha,
\end{aligned} \tag{4.12}$$

$$(\bar{S}^{\dot\alpha})_\beta{}^\gamma W_\gamma = \bar{\sigma}^{\dot\alpha\alpha}_\mu y^\mu\frac{\partial}{\partial\theta^\alpha}W_\beta.$$

We want to study the invariance equation,

$$(\xi Q + \bar{\xi}\bar{Q} + \zeta S + \bar{\zeta}\bar{S})_\alpha{}^\beta W_\beta = 0,$$

i.e.,

$$\begin{aligned}
&\Big[\Big((\xi + \bar{\zeta}\bar{\sigma}^\mu y_\mu)\frac{\partial}{\partial\theta} - 2i(\bar{\xi} - \zeta\sigma^\mu y_\mu)\bar{\sigma}^\nu\theta\frac{\partial}{\partial y_\nu}\\
&\quad - 2i\theta\theta\zeta\frac{\partial}{\partial\theta} - 6i\zeta\theta\Big)\delta_\alpha{}^\beta + 2i\zeta_\alpha\theta^\beta + 2i\zeta^\beta\theta_\alpha\Big]W_\beta = 0.
\end{aligned} \tag{4.13}$$

In terms of the components (4.9), Eq. (4.13) becomes

$$\begin{aligned}
\delta_{\xi,\zeta}W_\alpha|_0 &= (\xi - y_\rho\sigma^\rho\bar{\zeta})_\alpha D\\
&\quad - (i/2)[\sigma^\mu\bar{\sigma}^\nu(\xi - y_\rho\sigma^\rho\bar{\zeta})_\alpha]\\
&\quad \times(\partial_\mu v_\nu - \partial_\nu v_\mu) = 0,
\end{aligned} \tag{4.14a}$$

$$\begin{aligned}
\delta_{\xi,\zeta}W_\alpha|_{\theta_\beta} &= (\xi - y_\rho\sigma^\rho\bar{\zeta})_\beta(\sigma^\mu\,\partial_\mu\bar{\lambda})_\alpha\\
&\quad + [\sigma^\mu(\bar{\xi} + y_\rho\bar{\sigma}^\rho\zeta)]_\beta\,\partial_\mu\lambda_\alpha\\
&\quad - 2\zeta_\beta\lambda_\alpha - 2\zeta_\alpha\lambda_\beta = 0,
\end{aligned} \tag{4.14b}$$

$$\begin{aligned}
\delta_{\xi,\zeta}W_\alpha|_{\theta\theta} &= [\sigma^\nu(\bar{\xi} + y_\rho\bar{\sigma}^\rho\zeta)]_\alpha\,\partial_\nu D\\
&\quad + i[\sigma^\mu(\bar{\xi} + y_\rho\bar{\sigma}^\rho\zeta)]_\alpha\\
&\quad \times\partial^\nu(\partial_\mu v_\nu - \partial_\nu v_\mu) - 4\zeta_\alpha D = 0,
\end{aligned} \tag{4.14c}$$

together with the requirement that D and v_μ be real. With this requirement, (4.14c) follows automatically from (4.14a).

We now proceed to solve Eqs. (4.14a) and (4.14b). As in Sec. III, invariance under just one generator means setting ξ and ζ proportional to one real Grassman variable $\xi = \alpha\xi'$, $\zeta = \alpha\zeta'$, with ξ' and ζ' commuting spinors. Integrating out the α variable leaves Eqs. (4.14) unchanged. We thus work with the spinors ξ' and ζ' for the rest of this section, omitting the primes to avoid cumbersome notation.

We solve (4.14a) first for the fields D and v_μ; this equation is equivalent to the following set of equations:

$$(\bar{\xi} - \zeta\sigma^\rho y_\rho)\bar{\sigma}^\mu(\xi - y_\rho\sigma^\rho\bar{\zeta})(\partial_\mu v_\nu - \partial_\nu v_\mu) = 0, \tag{4.15a}$$

$$\begin{aligned}
&D\zeta(\xi - y_\mu\sigma^\mu\bar{\zeta}) - (i/2)\zeta\sigma^\rho\bar{\sigma}^\nu\\
&\quad \times(\xi - y_\mu\sigma^\mu\bar{\zeta})(\partial_\rho v_\nu - \partial_\nu v_\rho) = 0.
\end{aligned} \tag{4.15b}$$

Equation (4.15a) represents a set of real-valued differential equations for v_μ; given a solution to this set, (4.15b) serves to define D in terms of v_μ. One can also check that D thus defined will be real valued by virtue of (4.15a). We find it convenient to choose for v_μ the following gauge:

$$\bar{\zeta}\bar{\sigma}^\mu\zeta v_\mu = 0, \tag{4.16}$$

which is an axial gauge (along a null axis) that permits a residual gauge fixing which in fact will allow us to set a second component of v_μ to zero. To see this, consider (4.15a); defining

$$r^\mu \equiv (\bar\xi - \zeta\sigma^\rho y_\rho)\bar\sigma^\mu(\xi - y_\alpha\sigma^\alpha\bar\zeta),$$

we have

$$(\bar\zeta\bar\sigma^\nu\zeta)r^\mu(\partial_\mu v_\nu - \partial_\nu v_\mu) = -(\bar\zeta\bar\sigma^\nu\zeta)r^\mu\partial_\nu v_\mu = 0, \tag{4.17a}$$

from where we find

$$\begin{aligned}(\bar\zeta\bar\sigma^\nu\zeta)\partial_\nu(r^\mu v_\mu) &= (\bar\zeta\bar\sigma^\nu\zeta)v_\mu\,\partial_\nu r^\mu \\ &= v_\mu(\bar\zeta\bar\sigma^\nu\zeta)[-\zeta\sigma_\nu\bar\sigma^\mu(\xi - y_\rho\sigma^\rho\bar\zeta) \\ &\quad - (\bar\xi - \zeta\sigma^\rho y_\rho)\bar\sigma^\mu\sigma_\nu\bar\zeta] = 0,\end{aligned}$$

i.e.,

$$r^\mu v_\mu = \varphi(\bar\zeta\bar\sigma^\nu\zeta y_\nu,\ \bar\zeta\bar\sigma^\nu\xi y_\nu,\ \bar\xi\bar\sigma^\nu\zeta y_\nu).$$

Thus in this gauge $r^\mu v_\mu$ depends only on the variables transverse to the chosen axis; we now show that $r^\mu v_\mu$ can be set to zero using the residual gauge freedom:

$$v_\mu \to v_\mu + \partial_\mu\Omega, \quad \Omega = \Omega(\bar\zeta\bar\sigma^\nu\zeta y_\nu, \bar\zeta\bar\sigma^\nu\xi y_\nu, \bar\xi\bar\sigma^\nu\zeta y_\nu). \tag{4.18}$$

Under this gauge transformation, we have

$$\begin{aligned}\varphi &= r^\mu v_\mu \to r^\mu\,\partial_\mu\Omega + \varphi \\ &= -2(\bar\xi - \zeta\sigma^\rho y_\rho)\bar\zeta\cdot\zeta(\xi - y_\rho\sigma^\rho\bar\zeta)\Omega^{(1)} \\ &\quad + 2(\bar\xi - \zeta\sigma^\rho y_\rho)\bar\zeta\cdot\xi\sigma^\kappa\bar\zeta y_\kappa\Omega^{(2)} \\ &\quad + 2\zeta\sigma^\rho\bar\xi y_\rho\cdot\zeta(\xi - y_\kappa\sigma^\kappa\bar\zeta)\Omega^{(3)} + \varphi,\end{aligned} \tag{4.19}$$

where $\Omega^{(i)}$ denotes derivative of Ω with respect to its ith argument. We see that expression (4.19) depends only on the transverse variables; setting it to zero gives an equation for Ω which can be solved for an arbitrary function φ of these variables. In this way, we have the simultaneous gauge conditions

$$\bar\zeta\bar\sigma^\mu\zeta v_\mu = 0, \quad r^\mu v_\mu = 0. \tag{4.20}$$

As the two remaining components of v_μ we can choose $v \equiv \bar\zeta\bar\sigma^\mu\xi v_\mu$ and v^*. From (4.15) and (4.20) they will obey

$$\begin{aligned}r^\nu\partial_\nu v &= \bar\zeta\bar\sigma^\mu\xi r^\nu\,\partial_\nu v_\mu = \bar\zeta\bar\sigma^\mu\xi r^\nu\,\partial_\mu v_\nu = -\bar\zeta\bar\sigma^\mu\xi(\partial_\mu r^\nu)v_\nu \\ &= -\bar\zeta\bar\sigma^\mu\xi[-\zeta\sigma_\mu\bar\sigma^\nu(\xi - y_\kappa\sigma^\kappa\bar\zeta) \\ &\quad - (\bar\xi - \zeta\sigma^\rho y_\rho)\bar\sigma^\nu\sigma_\mu\bar\zeta]v_\nu \\ &= [-2\bar\zeta\bar\sigma^\nu\xi\cdot\zeta(\xi - y_\kappa\sigma^\kappa\bar\zeta) - 2\bar\zeta\bar\sigma^\nu\zeta\cdot\bar\zeta\bar\sigma^\kappa\xi y_\kappa]v_\nu \\ &= -2\zeta(\xi - y_\kappa\sigma^\kappa\bar\zeta)v.\end{aligned} \tag{4.21}$$

Equation (4.21) is of the same type as that encountered in Sec. III; its solution is given by

$$v = [(\bar\xi - \zeta\sigma^\mu y_\mu)\bar\zeta]^{-1}\mathscr{V}(c_1, c_1^*, c_2) \tag{4.22}$$

in terms of an arbitrary function $\mathscr{V}$ of the characteristic variables c_1, c_1^*, and c_2 defined in (3.16). From (4.20) and (4.22) the four components of v_μ can be reconstructed and we find after some linear algebra,

$$v^\mu = \frac{(\bar\xi - \zeta\sigma^\rho y_\rho)\bar\sigma^\mu\zeta}{2(\zeta\xi)[(\bar\xi - \zeta\sigma^\kappa y_\kappa)\bar\zeta]^2}\mathscr{V} + \frac{\bar\zeta\bar\sigma^\mu(\xi - y_\rho\sigma^\rho\bar\zeta)}{2(\bar\xi\bar\zeta)[\zeta(\xi - y_\kappa\sigma^\kappa\bar\zeta)]^2}\mathscr{V}^*. \tag{4.23a}$$

For the field D, we obtain from (4.15b),

$$D = \frac{i}{2}\left[\frac{-\zeta\sigma^\rho(\bar\xi + y_\rho\sigma^\rho\zeta)\partial_\rho}{(\zeta\xi)[(\bar\xi - \zeta\sigma^\kappa y_\kappa)\bar\zeta]^2}\mathscr{V} + \frac{\bar\zeta\bar\sigma^\rho(\xi - y_\mu\sigma^\mu\bar\zeta)\partial_\rho}{(\bar\xi\bar\zeta)[\zeta(\xi - y_\kappa\sigma^\kappa\bar\zeta)]^2}\mathscr{V}^*\right]. \tag{4.23b}$$

Equation (4.23a) is the most general expression for a photon field invariant under the action of an odd superconformal generator. Before looking at examples and limiting cases let us first also solve (4.14b) for invariant photino fields.

The main difficulty encountered in (4.14b) is that it involves both the photino λ_α and its complex conjugate $\bar\lambda_{\dot\alpha}$, while at the same time separating this field into its real and imaginary parts would not be a Lorentz invariant operation. To proceed then, we project λ_α along two independent directions:

$$\begin{aligned}\lambda_\alpha = {}&[\zeta(\xi - y_\mu\sigma^\mu\bar\zeta)]^{-1}[-\zeta_\alpha(\xi + \bar\zeta\bar\sigma^\mu y_\mu)\lambda \\ &+ (\xi - y_\mu\sigma^\mu\bar\zeta)_\alpha(\zeta\lambda)].\end{aligned} \tag{4.24}$$

We will solve (4.14b) in terms of the two projections $(\xi + \bar\zeta\bar\sigma^\mu y_\mu)\lambda$ and $\zeta\lambda$. These, being Lorentz scalars, can be separated in real and imaginary parts when necessary.

To begin, we project (4.14b) onto $(\xi + \bar\zeta\bar\sigma^\rho y_\rho)^\beta$, giving

$$r^\mu\,\partial_\mu\lambda_\alpha - 2(\xi + \bar\zeta\bar\sigma^\rho y_\rho)\zeta\lambda_\alpha - 2(\xi + \bar\zeta\bar\sigma^\rho y_\rho)\lambda\zeta_\alpha = 0, \tag{4.25}$$

with r^μ defined as before. In this projected equation $\bar\lambda_{\dot\alpha}$ has dropped out. To decouple the two components of λ_α we project (4.25) onto ζ^α and $(\xi + \bar\zeta\bar\sigma^\mu y_\mu)^\alpha$:

$$\begin{aligned}&r^\mu\,\partial_\mu[(\xi + \bar\zeta\bar\sigma^\mu y_\mu)\lambda] \\ &\quad + (2\bar\zeta\bar\xi - 4\xi\zeta - 2\bar\zeta\bar\sigma^\rho\zeta y_\rho)(\xi + \bar\zeta\bar\sigma^\mu y_\mu)\lambda = 0,\end{aligned} \tag{4.26a}$$

$$r^\mu\,\partial_\mu(\zeta\lambda) - 2(\xi + \bar\zeta\bar\sigma^\rho y_\rho)\zeta(\zeta\lambda) = 0. \tag{4.26b}$$

The solutions to Eqs. (4.26) are

$$(\xi + \bar\zeta\bar\sigma^\mu y_\mu)\lambda = \{\zeta(\xi - y_\mu\sigma^\mu\bar\zeta)/[(\bar\xi - \zeta\sigma^\mu y_\mu)\bar\zeta]^2\}\Omega, \tag{4.27a}$$

$$\zeta\lambda = [(\bar\xi - \zeta\sigma^\mu y_\mu)\bar\zeta]^{-1}\Omega', \tag{4.27b}$$

where Ω and Ω' are arbitrary complex functions of the characteristic variables c_1, c_1^*, and c_2 defined previously. From (4.24) and (4.27) we have for λ,

$$\lambda_\alpha = -\frac{\zeta_\alpha\Omega}{[(\bar\xi - \zeta\sigma^\mu y_\mu)\bar\zeta]^2} + \frac{(\xi - y_\mu\sigma^\mu\bar\zeta)_\alpha\Omega'}{(\bar\xi - \zeta\sigma^\nu y_\nu)\bar\zeta\cdot\zeta(\xi - y_\rho\sigma^\rho\bar\zeta)}. \tag{4.28}$$

We must still inject this expression into the unprojected equation (4.14b). The resulting equations for Ω and Ω' are

$$\begin{aligned}&2(\Omega' + \Omega'^*)(\zeta\xi - \bar\xi\bar\zeta) + \{[\zeta(\xi - y_\mu\sigma^\mu\bar\zeta)/(\bar\xi - \zeta\sigma^\rho y_\rho)\bar\zeta] \\ &\quad\times\zeta\sigma^\nu(\bar\xi + y_\kappa\bar\sigma^\kappa\zeta)\partial_\nu\Omega \\ &\quad - [(\bar\xi - \zeta\sigma^\mu y_\mu)\bar\zeta/\zeta(\xi - y_\rho\sigma^\rho\bar\zeta)] \\ &\quad\times\bar\zeta\bar\sigma^\nu(\xi - y_\kappa\sigma^\kappa\bar\zeta)\partial_\nu\Omega^*\} = 0, \\ &\zeta\sigma^\nu\bar\zeta\,\partial_\nu\Omega^* - [\zeta(\xi - y_\mu\sigma^\mu\bar\zeta)/(\bar\xi - \zeta\sigma^\rho y_\rho)\bar\zeta] \\ &\quad\times\zeta\sigma^\nu(\bar\xi + y_\kappa\bar\sigma^\kappa\zeta)\partial_\nu(\Omega' - \Omega'^*) = 0.\end{aligned} \tag{4.29}$$

The derivation of Eqs. (4.29) and their solution are rather longwinded and we summarize them in the Appendix. Here we present the most general solution, given in terms of two real arbitrary functions χ and χ' of the characteristic variables:

$$\Omega = [(\bar{\xi} - \zeta\sigma^\mu y_\mu)\bar{\zeta}/\zeta(\xi - y_\rho\sigma^\rho\bar{\zeta})] \times \bar{\zeta}\bar{\sigma}^\nu(\xi - y_\kappa\sigma^\kappa\bar{\zeta})\partial_\nu\chi, \tag{4.30a}$$

$$\Omega' = \tfrac{1}{2}\zeta\sigma^\nu\bar{\zeta}\,\partial_\nu\chi + i\chi'.$$

Thus, finally, from (4.28) we write for the photino,

$$\lambda_\alpha = [(\bar{\xi} - \zeta\sigma^\mu y_\mu)\bar{\zeta}\cdot\zeta(\xi - y_\nu\sigma^\nu\bar{\zeta})]^{-1} \times[-\zeta_\alpha\bar{\zeta}\bar{\sigma}^\rho(\xi - y_\mu\sigma^\mu\bar{\zeta})\partial_\rho\chi + (\xi - y_\mu\sigma^\mu\bar{\zeta})_\alpha \times(\tfrac{1}{2}\zeta\sigma^\nu\bar{\zeta}\,\partial_\nu\chi + i\chi')]. \tag{4.30b}$$

Let us now check for simultaneous invariance under more than one odd generator. We first observe that Eqs. (4.14a) and (4.14c) remain unchanged under $\xi \to i\xi$, $\zeta \to -i\zeta$ so that a set of fields $v_{\mu\nu}$, D invariant under a generator G will also be automatically invariant under $[G,\Pi]$. This is not true for the photino λ_α, as is evident from (4.14b). At the same time, invariance of (4.14a) under any additional generator (ξ',ζ') is impossible:

$$(\xi' + \bar{\zeta}'\bar{\sigma}^\rho y_\rho)^\alpha\cdot\delta_{\xi,\zeta}W_\alpha|_0 - (\xi + \bar{\zeta}\bar{\sigma}^\rho y_\rho)^\alpha\cdot\delta_{\xi',\zeta'}W_\alpha|_0 = 2(\xi' + \bar{\zeta}'\bar{\sigma}^\rho y_\rho)(\xi - y_\kappa\sigma^\kappa\bar{\zeta})D = 0$$

implies that

$$D = 0$$

and

$$\sigma^\mu\bar{\sigma}^\nu(\xi - y^\rho\sigma_\rho\bar{\zeta})(\partial_\mu v_\nu - \partial_\nu v_\mu) = 0,$$
$$\sigma^\mu\bar{\sigma}^\nu(\xi' - y_\rho\sigma^\rho\bar{\zeta}')(\partial_\mu v_\nu - \partial_\nu v_\mu) = 0$$

imply that

$$\partial_\mu v_\nu - \partial_\nu v_\mu = 0. \tag{4.31}$$

For the photino field, simultaneous invariance under more than one generator is still possible, but very difficult to compute in the general case. We note in passing that many of the properties of the photino are the same as those of the bosonic fields in the scalar multiplet discussed in Sec. III. This can be traced back to the fact that λ and A represent the lowest component in the superfield W_α and ϕ. Similarly, the invariance properties of the photon resemble those of the fermion in ϕ. A dissimilarity between the two superfields, however, is that λ_α in fact cannot be made super-de Sitter invariant, unlike (A,F).

To conclude, we present the limiting case of a super-Poincaré generator $\zeta = 0$. We also take, for definiteness (and without loss of generality),

$$\xi_\alpha = \begin{pmatrix}1\\0\end{pmatrix}, \quad r^\mu \equiv \bar{\xi}\bar{\sigma}^\mu\xi = (-1,0,0,-1). \tag{4.32a}$$

Equation (4.15a) becomes

$$(\partial_0 v_\nu - \partial_\nu v_0) + (\partial_3 v_\nu - \partial_\nu v_3) = 0. \tag{4.32b}$$

Choosing the gauge $v_0 = 0$, we find from (4.32b) that

$$\partial_0 v_3 = 0, \tag{4.32c}$$

which allows us to use the residual gauge freedom to set also $v_3 = 0$. For the remaining two components of v, (4.31a) gives

$$(\partial_0 + \partial_3)v_1 = (\partial_0 + \partial_3)v_2 = 0, \tag{4.32d}$$

yielding

$$v_1 = v_1(x^0 - x^3, x^1, x^2),\quad v_2 = v_2(x^0 - x^3, x^1, x^2). \tag{4.32e}$$

As expected from the general case, this solution is expressed in terms of two arbitrary real functions of the three characteristic variables $x^0 - x^3$, x^1, and x^2.

For the field D, we get

$$D = \partial_1 v_2 - \partial_2 v_1, \tag{4.32f}$$

while for the photino, solving (4.14b) we find

$$\lambda_\alpha = \begin{pmatrix}\partial_3\Omega + i\Omega'\\(\partial_1 + i\,\partial_2)\Omega\end{pmatrix}, \tag{4.32g}$$

with Ω and Ω' real arbitrary functions of $(x^0 - x^3, x^1, x^2)$.

V. CONCLUDING REMARKS

Let us summarize our findings. We have obtained the most general scalar superfields $\phi(y,\theta)$ and vector superfields $W_\alpha(y,\theta)$ that are invariant under one-parameter fermionic subgroups of the superconformal group. We observed that the invariance requirement restricts the arbitrary functions involved in the definition of these fields to depend on three bosonic and one fermionic variables. We then examined the constraints resulting from additional supersymmetries. The result of this analysis was that generically, stability under two (or more) supersymmetries forces $\phi|_\theta = 0 = W_\alpha|_\theta$. The only case where higher supersymmetry does not require these components to be trivial occurs for simultaneous invariance under the transformations generated by the following pair of vector fields: $G = \delta_{\xi,\zeta}$ and $[\pi,G]$. We also noted the existence of a nontrivial scalar multiplet which is invariant under the full super-de Sitter group and the absence of a vector superfield with analogous invariance. Finally, we specialized our results to the super-Poincaré group.

We would now like to expand on the following important point concerning the Grassmann structure of the transformation parameters. In presenting our solutions we have taken the odd parameters ξ and ζ both proportional to a single real Grassmann generator α. In other words, we considered spinors ξ and ζ of the form $\xi = \alpha\xi'$, $\zeta = \alpha\zeta'$, with ξ' and ζ' elements of the even part of the underlying Grassmann algebra with nonzero bodies.[9] With X a generic superfield, the invariance condition $\delta_{\xi,\zeta}X = 0$ became $\alpha\delta_{\xi',\zeta'}X = 0$ under this assumption. It allowed us to factor out α and to consider the transformation parameters as commuting spinors. We should point out here that given a solution $X^{(0)}$ to $\delta_{\xi',\zeta'}X = 0$, the superfield $X = X^{(0)} + \alpha X^{(1)}$ (with $X^{(1)}$ an arbitrary superfield of grading opposite to that of X) satisfies the invariance condition $\delta_{\xi,\zeta}X = 0$. Now factoring a single Grassman generator out of the transformation parameters is certainly not the only possibility; rather, it presents a prototype for the solution of more complicated invariance conditions, to which we come next.

To illustrate what these more complicated invariance conditions might be, let us suppose that the parameters ξ and

ζ can be decomposed in terms of two independent real Grassmann generators α_1 and α_2; i.e., $\xi = \alpha_1\xi_1 + \alpha_2\xi_2$, $\zeta = \alpha_1\zeta_1 + \alpha_2\zeta_2$, with ξ_i, ζ_i $(i = 1,2)$ four commuting Weyl spinors. The invariance condition $\delta_{\xi,\zeta} X = 0$ then becomes

$$\alpha_1 \delta_{\xi_1,\zeta_1} X + \alpha_2 \delta_{\xi_2,\zeta_2} X = 0. \tag{5.1}$$

Now expanding X in α_i,

$$X = X^{(0)} + \alpha_1 X_1^{(1)} + \alpha_2 X_2^{(1)} + \alpha_1\alpha_2 X^{(2)} \tag{5.2}$$

and resolving (5.1) into independent components, one finds the following invariance conditions for the superfields $X_m^{(n)}$ (chosen with appropriate gradings):

$$\begin{aligned}
&\delta_{\xi_1,\zeta_1} X^{(0)} = \delta_{\xi_2,\zeta_2} X^{(0)} = 0,\\
&\delta_{\xi_1,\zeta_1} X_2^{(1)} - \delta_{\xi_2,\zeta_2} X_1^{(1)} = 0,\\
&X^{(2)} \text{ arbitrary.}
\end{aligned} \tag{5.3}$$

Since ξ_i, ζ_i are commuting spinors the above variations can be treated exactly as in Secs. III and IV. On the one hand, we note that $X^{(0)}$ should be simultaneously invariant under the supertransformations associated to the parameters (ξ_1,ζ_1) and (ξ_2,ζ_2). We therefore see that the solutions of the invariance conditions stemming from the factorization of a single Grassmann generator out of the transformation parameters are prerequisite to solving cases where the parameters involve more than one Grassman generator. On the other hand, we observe that the invariance conditions for $X_i^{(1)}$ $(i = 1,2)$ are milder than the conditions that we have solved in Secs. III and IV. Clearly, however, using the solutions that we have obtained in Secs. III and IV, that is, setting $\delta_{\xi_1,\zeta_1} X_2^{(1)} = \delta_{\xi_2,\zeta_2} X_1^{(1)} = 0$, will give a particular solution to (5.1). All these considerations obviously generalize to situations where the transformation parameters are expressible in terms of an arbitrary number of independent Grassmann generators.

To conclude, let us point out some directions for future research. We have considered superinvariant scalar and U(1)-vector superfields. It would evidently be desirable to examine other sets of fields like gravity supermultiplets or non-Abelian vector superfields. In the latter case the existence of gauge transformations which can accompany the supersymmetry transformations[2] might open interesting new possibilities. Another natural extension of the present work would be to consider the invariance of superfields defined over extended superspace, the relevant transformation supergroup then being SU$(2,2/N)$, $N > 1$. With the enlargement of the underlying supermanifold, it is clear that invariance under subgroups with fairly large numbers of fermionic generators should be possible. Inasmuch as applications are concerned, we envisage to use these superinvariant fields as *Ansätze* to obtain solutions to the equations of motion of supersymmetric field theories. We would also like to investigate what types of dimensional reductions can be achieved using supersymmetries. We hope to report on some of these questions in forthcoming publications.

ACKNOWLEDGMENTS

HP would like to thank the Université de Montréal and LV the Università di Pisa for their hospitality while this work was underway. We are grateful to J-M. Lina for his assistance with the preparation of the manuscript.

One of us (HP) acknowledges a fellowship from the Istituto Nazionale di Fisica Nucleare (INFN). This work is supported in part through funds provided by the Natural Science and Engineering Research Council (NSERC) of Canada and the Fonds FCAR of the Québec Ministry of Education.

APPENDIX

We are after the most general solution to

$$\begin{aligned}
&(\xi - y_\rho \sigma^\rho \bar\zeta)_\beta (\sigma^\mu \partial_\mu \bar\lambda)_\alpha + [\sigma^\mu(\bar\xi + y_\rho \bar\sigma^\rho \zeta)]_\beta \partial_\mu \lambda_\alpha\\
&\qquad - 2\zeta_\beta \lambda_\alpha - 2\zeta_\alpha \lambda_\beta = 0.
\end{aligned} \tag{A1}$$

As seen in Sec. IV, by considering contractions of Eq. (A1), with $(\xi + \bar\zeta\bar\sigma^\rho y_\rho)^\beta \zeta^\alpha$ and $(\xi + \bar\zeta\bar\sigma^\rho y_\rho)^\beta (\xi + \bar\zeta\bar\sigma^\rho y_\rho)^\alpha$, we find that λ_α must be in the form,

$$\begin{aligned}
\lambda_\alpha = {}& -\zeta_\alpha \Omega / [(\bar\xi - \zeta\sigma^\mu y_\mu)\bar\zeta]^2\\
&+ (\xi - y_\mu \sigma^\mu \bar\zeta)_\alpha \Omega' / (\bar\xi - \zeta\sigma^\mu y_\mu)\bar\zeta\zeta(\xi - y_\mu\sigma^\mu\bar\zeta),
\end{aligned} \tag{A2}$$

with Ω and Ω' arbitrary complex functions of the characteristic variables. To see what further restrictions Ω and Ω' must obey, we inject (A2) back to (A1). Now, the full equation (A1) is equivalent to the set of its contraction with $\epsilon^{\beta\alpha}$ and with $\epsilon^{\alpha\gamma}(\sigma^{\mu\nu})_\gamma{}^\beta$, $\forall \mu,\nu$. Contracting with $\epsilon^{\beta\alpha}$ gives

$$\begin{aligned}
&2(\Omega' + \Omega'^*)(\zeta\xi - \bar\xi\bar\zeta)\\
&\quad + \{[\zeta(\xi - y_\mu\sigma^\mu\bar\zeta)/(\bar\xi - \zeta\sigma^\mu y_\mu)\bar\zeta]\\
&\quad \times \zeta\sigma^\nu(\bar\xi + y_\mu\bar\sigma^\mu\zeta)\partial_\nu\Omega\\
&\quad - [(\bar\xi - \zeta\sigma^\mu y_\mu)\bar\zeta / \zeta(\xi - y_\mu\sigma^\mu\bar\zeta)]\\
&\quad \times \bar\zeta\sigma^\nu(\xi - y_\mu\sigma^\mu\bar\zeta)\partial_\nu\Omega^*\} = 0,
\end{aligned} \tag{A3a}$$

while the contractions with $\epsilon^{\alpha\gamma}(\sigma^{\mu\nu})_\gamma{}^\beta$ all lead to the following equation, irrespective of the values of μ [using also (A3a) to simplify]:

$$\begin{aligned}
&\zeta\sigma^\nu\bar\zeta\,\partial_\nu\Omega^* - [\zeta(\xi - y_\mu\sigma^\mu\bar\zeta)/(\bar\xi - \zeta\sigma^\mu y_\mu)\bar\zeta]\\
&\quad \times \zeta\sigma^\nu(\bar\xi + y_\mu\bar\sigma^\mu\zeta)\partial_\nu(\Omega' + \Omega'^*) = 0.
\end{aligned} \tag{A3b}$$

At this stage, it is very convenient to rewrite Eqs. (A3a) and (A3b) in terms of the characteristic variables and of derivatives with respect to them. We work with the real characteristic variables

$$\begin{aligned}
z_1 &= \frac{\bar\zeta\bar\sigma^\mu\xi y_\mu}{\zeta(\xi - y_\mu\sigma^\mu\bar\zeta)} + \frac{\bar\xi\bar\sigma^\mu\zeta y_\mu}{(\bar\xi - \zeta\sigma^\mu y_\mu)\bar\zeta},\\
z_2 &= i\,\frac{\bar\zeta\bar\sigma^\mu\xi y_\mu}{\zeta(\xi - y_\mu\sigma^\mu\bar\zeta)} - i\,\frac{\bar\xi\bar\sigma^\mu\zeta y_\mu}{(\bar\xi - \zeta\sigma^\mu y_\mu)\bar\zeta},\\
z_3 &= \xi\sigma^\mu\bar\xi y_\mu + \tfrac{1}{2}(\bar\xi\bar\zeta + \zeta\xi - 2\zeta\sigma^\mu\bar\zeta y_\mu)\\
&\qquad \times \bar\zeta\bar\sigma^\rho\xi y_\rho \zeta\sigma^\nu\bar\xi y_\nu / (\bar\xi - \zeta\sigma^\mu y_\mu)\bar\zeta\zeta(\xi - y_\mu\sigma^\mu\bar\zeta).
\end{aligned}$$

Thus $\Omega = \Omega(z_1,z_2,z_3)$, $\Omega' = \Omega'(z_1,z_2,z_3)$, and the above equations become

$$\begin{aligned}
&2(\zeta\xi - \bar\xi\bar\zeta)(\Omega' + \Omega'^*) + 2\zeta\xi(D_1 + iD_2)\Omega\\
&\qquad - 2\bar\xi\bar\zeta(D_1 - iD_2)\Omega^* = 0,
\end{aligned} \tag{A4a}$$

$$\bar\xi\bar\zeta D_3\Omega^* + (i/2)(\zeta\xi - \bar\xi\bar\zeta)(D_1 + iD_2)(\Omega' + \Omega'^*) = 0, \tag{A4b}$$

with

$$D_1 = \partial^{(1)} - (iz_2/4)(\zeta\xi - \bar{\xi}\bar{\zeta})\partial^{(3)},$$
$$D_2 = \partial^{(2)} + (iz_1/4)(\zeta\xi - \bar{\xi}\bar{\zeta})\partial^{(3)},$$
$$D_3 = (i/2)(\zeta\xi - \bar{\xi}\bar{\zeta})\partial^{(3)} = [D_1, D_2],$$

and

$$\partial^{(i)} = \frac{\partial}{\partial z_i}.$$

Now given an arbitrary complex function Ω, one can always find another complex function X such that

$$\Omega = 2\bar{\xi}\bar{\zeta}(D_1 - iD_2)X. \tag{A5a}$$

With this, (A4a) becomes

$$\begin{aligned} 2(\zeta\xi - \bar{\xi}\bar{\zeta})(\Omega' + \Omega'^*) &= -4\zeta\xi\bar{\xi}\bar{\zeta}[(D_1 + iD_2)(D_1 - iD_2)X \\ &\quad - (D_1 - iD_2)(D_1 + iD_2)X^*] \\ &= 8i\zeta\xi\bar{\xi}\bar{\zeta}[D_3(\operatorname{Re} X) - (D_1^2 + D_2^2)(\operatorname{Im} X)]. \end{aligned} \tag{A5b}$$

Substituting for $\Omega' + \Omega^*$ in (A4b), we find that $(\operatorname{Im} X)$ must satisfy

$$(D_1 + iD_2)(D_1 + iD_2)(D_1 - iD_2)(\operatorname{Im} X) = 0, \tag{A5c}$$

while there are no constraints on $\operatorname{Re} X$. We will now show that the unique solution to this equation is $\operatorname{Im} X = 0$ up to functions whose contributions to the values of Ω and Ω' can be reproduced by an opportune redefinition of $\operatorname{Re} X$.

Now, Since $(D_1 + iD_2)(D_1 - iD_2)\operatorname{Im} X$ is annihilated by $D_1 + iD_2$, we can again apply the method of characteristic curves to find that

$$\begin{aligned} &(D_1 + iD_2)(D_1 - iD_2)\operatorname{Im} X \\ &\quad = \varphi\left(z_1 + iz_2, z_3 + \tfrac{1}{8}(\zeta\xi - \bar{\xi}\bar{\zeta})(z_1^2 + z_2^2)\right), \end{aligned} \tag{A6a}$$

with φ an arbitrary complex function of its two variables. It follows that

$$\begin{aligned} D_3(\operatorname{Im} X) &= i(D_1 + iD_2)(D_1 - iD_2)\operatorname{Im} X \\ &\quad - i(D_1 - iD_2)(D_1 + iD_2)\operatorname{Im} X \\ &= i\varphi\left(z_1 + iz_2, z_3 + \tfrac{1}{8}(\zeta\xi - \bar{\xi}\bar{\zeta})(z_1^2 + z_2^2)\right) \\ &\quad - i\varphi^*\left(z_1 - iz_2, z_3 - \tfrac{1}{8}(\zeta\xi - \bar{\xi}\bar{\zeta})(z_1^2 + z_2^2)\right). \end{aligned} \tag{A6b}$$

Integrating, we find

$$\begin{aligned} \operatorname{Im} X = i[&\tilde{\varphi}\left(z_1 + iz_2, z_3 + \tfrac{1}{8}(\zeta\xi - \bar{\xi}\bar{\zeta})(z_1^2 + z_2^2)\right) \\ &- \tilde{\varphi}\left(z_1 - iz_2, z_3 - \tfrac{1}{8}(\zeta\xi - \bar{\xi}\bar{\zeta})(z_1^2 + z_2^2)\right)] \\ &+ \bar{\varphi}(z_1, z_2) \end{aligned}$$

in terms of an arbitrary function $\tilde{\varphi}$ and a real function $\bar{\varphi}$. To see what further conditions, if any, $\tilde{\varphi}$ and $\bar{\varphi}$ must satisfy, we reinject (A6b) into (A5c):

$$\partial_1(\partial_1^2 + \partial_2^2)\bar{\varphi} = \partial_2(\partial_1^2 + \partial_2^2)\bar{\varphi} = 0, \tag{A7a}$$

with the solution

$$\bar{\varphi} = \varphi_1(z_1 + iz_2) + \varphi_1^*(z_1 - iz_2) + (z_1^2 + z_2^2)c \tag{A7b}$$

(c is a real constant and φ_1 is an arbitrary function of $z_1 + iz_2$). We see that φ_1 can be absorbed inside $\tilde{\varphi}$. This solves completely the equations (A5b), so that we may write

$$\begin{aligned} \Omega &= 2\bar{\xi}\bar{\zeta}(D_1 - iD_2)[\operatorname{Re} X + i(i(\tilde{\varphi} - \tilde{\varphi}^*) + c(z_1^2 + z_2^2))] \\ &= -2\bar{\xi}\bar{\zeta}(D_1 - iD_2)(\operatorname{Re} X - (\tilde{\varphi} + \tilde{\varphi}^*) \\ &\quad + 8icz_3/(\zeta\xi - \bar{\xi}\bar{\zeta})) + i(i(\tilde{\varphi} - \tilde{\varphi}^*) + c(z_1^2 + z_2^2))] \\ \Omega^* &= 2\zeta\xi(D_1 + iD_2)[\operatorname{Re} X - i(i(\tilde{\varphi} - \tilde{\varphi}^*) + c(z_1^2 + z_2^2))] \\ &= 2\zeta\xi(D_1 + iD_2)[\operatorname{Re} X - (\tilde{\varphi} + \tilde{\varphi}^*) \\ &\quad + 8icz_3/(\zeta\xi - \bar{\xi}\bar{\zeta})], \end{aligned} \tag{A8}$$

and we find that nonzero solutions of (A5c) for $\operatorname{Im} X$ can be absorbed in $\operatorname{Re} X$ by shifting (A5c) by the real quantity $-(\tilde{\varphi} + \tilde{\varphi}^*) + [8ic/(\zeta\xi - \bar{\xi}\bar{\zeta})]z_3$.

To recapitulate, we have found that the most general solution to the invariance equation for the photino (A1) is given in terms of two real arbitrary functions of the characteristic variables $\operatorname{Im}\Omega'$ [which was not constrained by (A1)] and $\operatorname{Re} X$. In terms of these functions, we have

$$\begin{aligned} \Omega &= 2\zeta\xi(D_1 - iD_2)(\operatorname{Re} X), \\ \Omega^* &= 2\bar{\xi}\bar{\zeta}(D_1 + iD_2)(\operatorname{Re} X), \\ \Omega' + \Omega^* &= [4i\zeta\xi\bar{\xi}\bar{\zeta}/(\zeta\xi - \bar{\xi}\bar{\zeta})]D_3(\operatorname{Re} X), \end{aligned} \tag{A9}$$

and, going back to Minkowski variables, we write for the photino,

$$\begin{aligned} \lambda_\alpha = {} & [(\bar{\xi} - \zeta\sigma^\mu y_\mu)\bar{\zeta}\cdot\zeta(\xi - y_\mu\sigma^\mu\bar{\zeta})]^{-1} \\ & \times[-\zeta_\alpha\bar{\zeta}\bar{\sigma}^\rho(\xi - y_\mu\sigma^\mu\bar{\zeta})\partial_\rho(\operatorname{Re} X) \\ & + (\xi - y_\mu\sigma^\mu\bar{\zeta})_\alpha(\tfrac{1}{2}\zeta\sigma^\nu\bar{\zeta}\,\partial_\nu(\operatorname{Re} X) + i\operatorname{Im}\Omega')]. \end{aligned} \tag{A10}$$

[1] J. Beckers, J. Harnad, M. Perroud, and P. Winternitz, J. Math. Phys. **19**, 2126 (1978).
[2] J. Harnad, S. Shnider, and L. Vinet, J. Math. Phys. **21**, 2719 (1980); P. Forgács and N. Manton, Commun. Math. Phys. **72**, 15 (1980).
[3] See, for instance, J. Harnad, S. Shnider, and L. Vinet, "Solutions to Yang–Mills equations on $\bar{M}^4$ invariant under subgroups of $O(4,2)$," in *Complex Manifold Techniques in Theoretical Physics*, edited by D. E. Lerner and P. D. Sommers (Pitman, New York, 1979).
[4] N. Manton, Nucl. Phys. B **158**, 141 (1979); J. Harnad, S. Shnider, and J. Tafel, Lett. Math. Phys. **4**, 107 (1980); N. Manton, Ann. Phys. (NY) **167**, 328 (1986).
[5] J. Doneux, Y. Saint-Aubin, and L. Vinet, Phys. Rev. D **25**, 484 (1982); M. Légaré, J. Math. Phys. **24**, 1219 (1983).
[6] J. Wess and J. Bagger, *Supersymmetry and Supergravity* (Princeton U. P., Princeton, NJ, 1983).
[7] G. Mack and A. Salam, Ann. Phys. (NY) **53**, 174 (1969).
[8] R. Haag, J. Łopuszanski, and M. Sohnius, Nucl. Phys. B **88**, 257 (1975).
[9] B. de Witt, *Supermanifolds* (Cambridge U. P., Cambridge, 1984).
[10] S. J. Gates, Jr., M. T. Grisaru, M. Roček, and W. Siegel, *Superspace* (Benjamin, Reading, MA, 1983).
[11] J. Wess and B. Zumino, Nucl. Phys. B **70**, 39 (1974).
[12] N. S. Baaklini, Nucl. Phys. B **129**, 354 (1977); E. A. Ivanov and A. S. Sorin, Theor. Math. Phys. **39**, 394 (1979).

Green's functions for nonlinear Klein–Gordon kink perturbation theory

R. J. Flesch[a)] and S. E. Trullinger
Department of Physics, University of Southern California, Los Angeles, California 90089-0484

(Received 21 October 1986; accepted for publication 18 February 1987)

A Green's function is defined for nonlinear Klein–Gordon theories in terms of the solutions to the eigenvalue equation obtained by linearizing the nonlinear wave equation about a static kink waveform. Analytic forms in terms of "modified" Lommel functions of two variables are derived for the sine–Gordon, phi-4, and double quadratic potentials. Asymptotic forms for the Green's functions are obtained by investigating the asymptotic behavior of the modified Lommel functions. Methods for calculating the Lommel functions are also outlined.

I. INTRODUCTION

The influence of perturbations on the dynamics of kink solutions of nonlinear Klein–Gordon (NKG) models in $1+1$ dimensions is a subject of particular importance in condensed matter contexts,[1] where many different types of internal and/or external agents are responsible for spoiling the otherwise perfect (and boring!) propagation of kinks through the system of interest. Examples include impurities or other physical imperfections in the system,[2] dissipative forces[2,3] and coupling to other degrees of freedom,[3] external driving forces such as electric[4] or magnetic fields, stress fields,[5] etc. In most cases of interest, the kinks involved carry some physically significant signature such as electric charge or spin, and hence can carry currents of various kinds which are important for the behavior of the system as a whole (e.g., conductivity[1,4]). Many of the physical systems of interest are modeled (sometimes justifiably) by nonlinear Klein–Gordon (NKG) Lagrangians such as the sine–Gordon (SG), ϕ^4, or double-quadratic (DQ) cases,[1] among many others.

As a consequence of the importance of being able to determine the motion of kinks under perturbing influences such as those above, there have been several investigations over the last few years of either a general nature or having limited application to rather specific perturbations. One of the more useful approaches[3] has been to regard the kink of interest as an extended "particle" which obeys Newtonian dynamics at the classical level. Although there has been some controversy[6–9] regarding whether in fact the kink behaves as a Newtonian particle, this question has largely been resolved and one can adopt this Newtonian picture if care is taken to properly treat the behavior of the regions of the system far from the position of the kink.

The modern approach[10–12] based on these ideas is to regard the kink position as a collective coordinate and to perform a canonical transformation[12,13] to new coordinates, one of which is the kink "center-of-mass" position. The deviation of the full field from the pure kink profile is regarded as small (if the perturbing influence is small) and a systematic perturbation theory is employed in which successively higher powers of this deviation are included. The actual deviation of the kink position from its unperturbed value is not required to be small in this collective coordinate method, thus removing some secularities which occur in early versions[3] of this particlelike approach.

In the perturbation expansion method, it is convenient to employ a Green's function technique[11,12] based on knowledge of the exact solutions for the small oscillations about the kink in an unperturbed system.[1,3] Until now this approach has been hindered by the lack of an analytic form for such Green's functions since they involve integrals not found in the tables. In this paper we remedy this situation by reporting our closed-form evaluation of the Green's functions for the three example systems mentioned above, namely SG, ϕ^4, and DQ. These forms involve modified Lommel functions of two variables[14] and since many of their properties have not to our knowledge been discussed in the literature, we examine some of the more useful of these, such as asymptotic expansions, in the present paper.

The remainder of the paper is organized as follows. Section II contains an introduction to the NKG models of interest, their kink solutions, and the nature of small oscillations about the kinks in the pure system. The small oscillation solutions are then used in Sec. III to construct explicit, closed-form expressions for the Green's functions of the three example systems in turn. In Sec. IV we discuss the asymptotic behavior of the Green's functions by first investigating the asymptotic properties of the Lommel functions of two variables. Some of these results are new and are presented for the first time, to our knowledge, in this paper. In Sec. V we display and discuss some representative plots of the SG Green's function as an example. Appendix A contains our evaluation of a generalized form of Hardy's integral for Lommel functions. Appendix B collects some of the properties of the modified Lommel functions of two variables while Appendix C describes some aspects of the numerical evaluation of modified Lommel functions and their asymptotic forms.

II. NONLINEAR KLEIN–GORDON KINKS AND THEIR SMALL OSCILLATIONS

In this section we briefly review the main features of solutions to the nonlinear Klein–Gordon class of field theories. The single-kink solutions to the wave equations along with small oscillations about these kinks will be described. The various quantities described in this section are collected in Table I for the sine–Gordon, ϕ^4, and double-quadratic

---

[a)] Present address: Center for Nonlinear Studies, Los Alamos National Laboratory, Los Alamos, New Mexico 87545.

 0022-2488/87/071619-13$02.50

TABLE I. Various quantities for the ϕ^4, SG, and DQ systems. Here $V(\phi)$ is the nonlinear potential, $\phi_k(x)$ is the kink (+) or antikink (−) solution, $V''[\phi_k(x)]$ is the potential which enters in the Schrödinger-like phonon equation [see Eq. (2.8)], and $f_{b,i}(x)$ and $f_k(x)$ are the bound and scattering states of $V''[\phi_k(x)]$ ($\omega_{b,1}=0$ for all three cases; $\omega_{b,2}=\sqrt{3}/2$ for the ϕ^4 potential).

	ϕ^4	SG	DQ
$V(\phi)$	$\frac{1}{8}[\phi^2-1]^2$	$1-\cos(\phi)$	$\frac{1}{2}[\lvert\phi\rvert-1]^2$
$\phi_k(x)$	$\tanh[\pm x/2]$	$4\tan^{-1}(e^{\pm x})$	$\pm\,\mathrm{sgn}(x)[1-e^{-\lvert x\rvert}]$
$V''[\phi_k(x)]$	$1-\frac{3}{2}\mathrm{sech}^2[x/2]$	$1-2\,\mathrm{sech}^2(x)$	$1-2\delta(x)$
$f_{b,n}(x)$	$f_{b,1}(x)=\sqrt{\frac{3}{8}}\,\mathrm{sech}^2[x/2]$ $f_{b,2}(x)=\sqrt{\frac{3}{4}}\,\mathrm{sech}[x/2]\tanh[x/2]$	$f_{b,1}(x)=\sqrt{\frac{1}{2}}\,\mathrm{sech}(x)$	$f_{b,1}(x)=e^{-\lvert x\rvert}$
$f_k(x)$	$\dfrac{e^{ikx}[3\tanh^2(x/2)-6ik\tanh(x/2)-(1+4k^2)]}{\sqrt{8\pi(1+k^2)(1+4k^2)}}$	$\dfrac{e^{ikx}[k+i\tanh(x)]}{\sqrt{2\pi(1+k^2)}}$	$\sin(kx)/\sqrt{2\pi},\ k<0$ $\dfrac{k\cos(kx)-\mathrm{sgn}(x)\sin(kx)}{\sqrt{2\pi(1+k^2)}},\ k>0$

potentials [this table corrects some errors in Table 1 of Ref. 15 and a similar error in Eq. (4.16b) in Ref. 1].

The general nonlinear Klein–Gordon Lagrangian we consider has the form

$$L=\int_{-\infty}^{\infty}dx\left\{\frac{1}{2}\phi_t^2-\frac{1}{2}\phi_x^2-V(\phi)\right\},\tag{2.1}$$

where x and t are dimensionless space and time variables and $V(\phi)$ is the nonlinear potential. The nonlinear wave equation satisfied by $\phi(x,t)$ is

$$\phi_{tt}-\phi_{xx}+V'(\phi)=0,\tag{2.2}$$

where the prime on $V(\phi)$ denotes a derivative with respect to ϕ. Static single-kink solutions, $\phi_k(x)$, of Eq. (2.2) may be obtained by direct integration with the boundary conditions

$$\left.\frac{d\phi_k(x)}{dx}\right|_{x=\pm\infty}=0.\tag{2.3}$$

The static kink (+) and antikink (−) solutions are given by

$$x=\pm\frac{1}{\sqrt{2}}\int_{\phi_k(0)}^{\phi_k(x)}\frac{d\phi}{\sqrt{V(\phi)}}.\tag{2.4}$$

Moving solutions can be obtained by a Lorentz boost.

The equation governing the small oscillations about the static kink waveform is obtained by substituting

$$\phi(x,t)=\phi_k(x)+\psi(x,t),\tag{2.5}$$

into Eq. (2.2) and linearizing in ψ:

$$\psi_{tt}-\psi_{xx}+V''[\phi_k(x)]\psi=0.\tag{2.6}$$

Here $V''[\phi_k(x)]$ denotes the second derivative of $V(\phi)$ with respect to ϕ evaluated for $\phi=\phi_k(x)$. Writing ψ as

$$\psi(x,t)=f(x)e^{-i\omega t},\tag{2.7}$$

leads to the following eigenvalue equation:

$$-f_{xx}+V''[\phi_k(x)]f=\omega^2 f.\tag{2.8}$$

Due to the localized nature of the kink waveform $\phi_k(x)$, the function $V''[\phi_k(x)]$ varies mainly in the region of the kink center (assumed to be at $x=0$) and approaches a constant (taken to be unity) far from the kink center:

$$V''[\phi_k(x)]\underset{|x|\to\infty}{\to}1.\tag{2.9}$$

Moreover, the function $V''[\phi_k(x)]$ has a minimum at $x=0$ such that

$$V''[\phi_k(0)]<0.\tag{2.10}$$

From these properties, we see that there exists a close analogy between Eq. (2.8) and the Schrödinger equation for a particle moving in a one-dimensional "potential well," $V''[\phi_k(x)]$. The "bound state(s)" and "continuum" states for this potential are of fundamental importance for statistical mechanics phenomenologies,[15] perturbation theories for kink dynamics,[2] and quantization procedures for kink states.[10,12,13,16–19]

Since the Lagrangian (2.1) possesses translational invariance, the spectrum of the small oscillations about the single kink must contain a zero-frequency ($\omega=0$) "translation" mode (Goldstone mode) which restores the translational invariance broken by the introduction of the kink. In addition to this translation mode there may be other discrete eigenvalues ("bound states") with frequencies between 0 and 1. These solutions correspond to "internal" oscillation modes in which the kink undergoes a harmonically varying shape change localized about the kink center. We denote these bound-state eigenfrequencies by $\omega_{b,1}\cdots\omega_{b,N}$, where N is the total number of bound states. The lowest of these is $\omega_{b,1}=0$ since all other $\omega_{b,i}^2$ must be non-negative in order for the kink to be stable against small oscillations.

In addition to the bound states, there exist continuum states ("phonons") which are labeled by a wave vector k. These states have eigenvalues ω_k^2 given by

$$\omega_k^2=1+k^2,\tag{2.11}$$

which is precisely the dispersion relation for small oscillations in the absence of kinks.

The continuum states together with the bound states form a complete set and satisfy the completeness relation,

$$\sum_{i=1}^{N}f_{b,i}^*(x)f_{b,i}(x')+\int_{-\infty}^{\infty}dk\,f_k^*(x)f_k(x')=\delta(x-x'),\tag{2.12}$$

and the following orthogonality relations:

$$\int_{-\infty}^{\infty} dx\, f_{b,n}(x) f_{b,m}(x) = \delta_{m,n}, \tag{2.13a}$$

$$\int_{-\infty}^{\infty} dx\, f_k^*(x) f_{k'}(x) = \delta(k-k'), \tag{2.13b}$$

$$\int_{-\infty}^{\infty} dx\, f_k(x) f_{b,n}(x) = 0. \tag{2.13c}$$

Table I lists the nonlinear potentials, kink waveforms, small oscillation potentials, bound and scattering states for the SG, ϕ^4, and DQ potentials.

III. ANALYTIC EVALUATION OF THE GREEN'S FUNCTIONS

For the set $\{f_{b,i}(x), f_k(x)\}$ of solutions satisfying the "phonon" equation (2.8), we define the full Green's function as

$$G(x,x',\tau) = \sum_{\text{bound states}} f_{b,i}^*(x) f_{b,i}(x') \int_{-\infty}^{\infty} \frac{d\omega\, e^{i\omega\tau}}{2\pi(\omega_{b,i}^2 - \omega^2)} + \int_{-\infty}^{\infty} dk\, f_k^*(x) f_k(x') \int_{-\infty}^{\infty} \frac{d\omega\, e^{i\omega\tau}}{2\pi(\omega_k^2 - \omega^2)}, \tag{3.1}$$

where $\tau \equiv t - t'$. Using the completeness relation (2.12), and the fact that the set $\{f_{b,i}(x), f_k(x)\}$ satisfy Eq. (2.8), one can show that the full Green's function satisfies the usual equation

$$\{\partial_{tt} - \partial_{xx} + V''[\phi_k(x)]\} G(x,x',\tau) = \delta(x-x')\delta(\tau). \tag{3.2}$$

Once a set of boundary conditions is chosen the ω integral in Eq. (3.1) may be evaluated without choosing a particular set of $\{f_{b,i}(x), f_k(x)\}$. In this paper we choose retarded boundary conditions obtained by moving both of the poles in the ω integral above the real ω axis. Carrying out the ω integral yields

$$G(x,x',\tau) = G_b(x,x',\tau) + G_p(x,x',\tau), \tag{3.3}$$

where $G_b(x,x',\tau)$ and $G_p(x,x',\tau)$ are the bound state and phonon contributions given by

$$G_b(x,x',\tau) = \theta(\tau) \left\{ \tau f_{b,1}^*(x) f_{b,1}(x') + \sum_{i=2}^{N} f_{b,i}^*(x) f_{b,i}(x') \frac{\sin(\omega_i \tau)}{\omega_{b,i}} \right\}, \tag{3.4a}$$

$$G_p(x,x',\tau) = \theta(\tau) \int_{-\infty}^{\infty} dk\, f_k^*(x) f_k(x') \frac{\sin(\omega_k \tau)}{\omega_k}, \tag{3.4b}$$

with N the number of bound states [if $n = 1$, the second term is omitted from Eq. (3.4a)] and $\theta(\tau)$ is the Heaviside step function,

$$\theta(\tau) = \begin{cases} 0, & -\infty < \tau < 0, \\ 1, & 0 \leqslant \tau < \infty. \end{cases} \tag{3.5}$$

In order to obtain explicit forms for these contributions to the Green's function, one must insert the appropriate set of linearized solutions into Eqs. (3.4a) and (3.4b). As examples, we evaluate the phonon contribution for the SG, ϕ^4, and DQ potentials.

A. The SG Potential

Since the bound state contribution (3.4a) is already expressed in terms of known functions, we turn to the evaluation of the phonon contribution given in Eq. (3.4b). Inserting the functions $f_k(x)$ from the SG column of Table I into Eq. (3.4b) we have, after collecting common terms,

$$G_p^{\mathrm{SG}}(x,x',\tau) = \theta(\tau)\{I_1 + \beta_2 I_2 + \beta_3 \operatorname{sgn}(z) I_3\}, \tag{3.6}$$

where

$$I_1 = \frac{1}{\pi} \int_0^{\infty} \frac{dk}{\sqrt{1+k^2}} \cos(|z|k) \sin(\tau\sqrt{1+k^2}), \tag{3.7a}$$

$$I_2 = \frac{1}{\pi} \int_0^{\infty} \frac{dk}{(1+k^2)^{3/2}} \cos(|z|k) \sin(\tau\sqrt{1+k^2}), \tag{3.7b}$$

$$I_3 = \frac{1}{\pi} \int_0^{\infty} \frac{dk}{(1+k^2)^{3/2}} k \sin(|z|k) \sin(\tau\sqrt{1+k^2}), \tag{3.7c}$$

with the definitions

$$\tau \equiv t - t', \quad z \equiv x - x', \quad \beta_2 \equiv \tanh(x)\tanh(x') - 1,$$
$$\beta_3 \equiv \tanh(x') - \tanh(x). \tag{3.8}$$

Since I_2 is uniformly convergent for all $|z|$ and τ, we may differentiate with respect to $|z|$ to obtain

$$I_3 = -\frac{dI_2}{d|z|}. \tag{3.9}$$

Therefore only I_1 and I_2 need to be evaluated. These integrals may be evaluated by considering the integral $I(\mu)$ given by

$$I(\mu) = \frac{1}{\pi} \int_0^{\infty} \frac{dk}{\sqrt{\mu^2+k^2}} \cos(|z|k) \sin(\tau\sqrt{\mu^2+k^2}), \tag{3.10}$$

$$= [\theta(\tau - |z|)/2] J_0(\mu\sqrt{\tau^2 - z^2}), \tag{3.11}$$

where the integral is found in the tables.[20] The special case $I(1)$ is precisely the integral I_1. Since the derivative of the integrand of Eq. (3.10) is a continuous function of both μ and k, we may differentiate $I(\mu)$ with respect to μ to obtain

$$I_2 = \lim_{\mu \to 1} \left\{ -\frac{dI(\mu)}{d\mu} + \frac{\tau}{2\pi} \int_{-\infty}^{\infty} \frac{dk}{\mu^2+k^2} \times \cos(|z|k) \cos(\tau\sqrt{\mu^2+k^2}) \right\}, \tag{3.12}$$

$$= \frac{\theta(\tau - |z|)}{2} \sqrt{\tau^2 - z^2}\, J_1(\sqrt{\tau^2 - z^2}) + \frac{\tau}{2\pi} \int_{-\infty}^{\infty} \frac{dk}{1+k^2} \cos(|z|k) \cos(\tau\sqrt{1+k^2}). \tag{3.13}$$

In the integral remaining in Eq. (3.13) we substitute $k = \sinh(u)$, which gives us

$$\frac{\tau}{2\pi} \int_{-\infty}^{\infty} \frac{dk}{1+k^2} \cos(|z|k) \cos(\tau\sqrt{1+k^2}) = \frac{\tau}{2\pi} \int_{-\infty}^{\infty} \frac{du}{\cosh(u)} \cos[|z|\sinh(u)] \cos[\tau\cosh(u)], \tag{3.14}$$

$$= \frac{\tau}{4\pi} \int_{-\infty}^{\infty} \frac{du}{\cosh(u)} \{\cos[|z|\sinh(u) + \tau\cosh(u)] + \cos[\tau\cosh(u) - |z|\sinh(u)]\}, \tag{3.15}$$

$$= \frac{\tau}{2\pi} \int_{-\infty}^{\infty} \frac{du\, e^{u}}{e^{2u}+1} \{\cos[ae^{u} + be^{-u}] + \cos[ae^{-u} + be^{u}]\}, \tag{3.16}$$

$$= \frac{\tau}{2\pi} \int_{0}^{\infty} \frac{dt}{t^2+1} \left\{\cos\left[at + \frac{b}{t}\right] + \cos\left[\frac{a}{t} + bt\right]\right\}, \tag{3.17}$$

$$= \frac{\tau}{\pi} \int_{0}^{\infty} \frac{dt}{t^2+1} \cos\left[at + \frac{b}{t}\right], \tag{3.18}$$

where in passing from Eq. (3.17) to Eq. (3.18) we have let $t \to 1/t$ in the second cosine term and in Eq. (3.16) we have introduced the quantities

$$a \equiv (\tau + |z|)/2, \tag{3.19a}$$

$$b \equiv (\tau - |z|)/2. \tag{3.19b}$$

For $b<0$ the integral in Eq. (3.18) is found in the tables[21] to be

$$\frac{1}{\pi} \int_{0}^{\infty} \frac{dt}{t^2+1} \cos\left[at - \frac{|b|}{t}\right] = \frac{1}{2} e^{-(a-b)}. \tag{3.20}$$

For $b>0$, the integral in Eq. (3.18) may be expressed in terms of "modified" Lommel functions of two variables.[14] The modified functions, namely Lommel functions in which the first argument is pure imaginary, have not been found in the literature. Hence we introduce the notation $\Lambda_n(w,s)$ and $\Xi_n(w,s)$ for the modified Lommel functions and give their series representations in terms of Bessel functions:

$$\Lambda_n(w,s) \equiv i^{-n} U_n(iw,s) = \sum_{m=0}^{\infty} \left(\frac{w}{s}\right)^{2m+n} J_{2m+n}(s), \tag{3.21a}$$

$$\Xi_n(w,s) \equiv i^{-n} V_n(iw,s) = \sum_{m=0}^{\infty} \left(\frac{w}{s}\right)^{-2m-n} J_{-2m-n}(s). \tag{3.21b}$$

With these definitions, we write for $b>0$

$$\frac{1}{\pi} \int_{0}^{\infty} \frac{dt}{t^2+1} \cos\left[at + \frac{|b|}{t}\right] = \frac{1}{2} e^{-(a-b)} - \Lambda_1(w,s), \tag{3.22}$$

where

$$s \equiv \sqrt{\tau^2 - z^2}, \tag{3.23a}$$

$$w \equiv \tau - |z|. \tag{3.23b}$$

Combining Eqs. (3.20) and (3.22) we have for I_2

$$I_2 = \tfrac{1}{2}\tau e^{-|z|} + \theta(\tau - |z|)\{sJ_1(s)/2 - \tau\Lambda_1(w,s)\}. \tag{3.24}$$

Using Eq. (B6) from Appendix B we differentiate Eq. (3.24) with respect to $|z|$ which results in

$$\frac{dI_2}{d\,|z|} = -\frac{1}{2}\tau e^{-|z|} + \frac{\theta(\tau - |z|)}{2} \times \{-(\tau + |z|)J_0(s) + 2\tau\Lambda_0(w,s)\}. \tag{3.25}$$

In Eqs. (3.24) and (3.25), I_2 and its derivative appear to have terms which grow linearly in τ which is impossible in view of the integral representations in Eqs. (3.7). Using asymptotic expressions for the modified Lommel functions, we shall show in Sec. IV that the large τ dependence is actually an inverse square root.

Writing the phonon contribution as

$$G_p^{SG}(x,x',\tau) = \theta(\tau)\left\{I_1 + \beta_2 I_2 - \beta_3 \operatorname{sgn}(z) \frac{dI_2}{d\,|z|}\right\}, \tag{3.26}$$

we notice that with I_1, I_2, and $dI_2/d\,|z|$ given by Eqs. (3.11), (3.24), and (3.25), there is a term which does not vanish outside of the "light cone" [i.e., a term which does not have $\theta(\tau - |z|)$ as a prefactor], namely

$$\theta(\tau)\tau e^{-|z|}/2\{\beta_2 + \operatorname{sgn}(z)\beta_3\}. \tag{3.27}$$

One can show that this term may be rewritten as

$$-\theta(\tau)\tau f_{b,1}^{*}(x) f_{b,1}(x'). \tag{3.28}$$

Hence when the bound state contribution is added to Eq. (3.26) to obtain the full Green's function, we are left with an expression which vanishes identically outside of the light cone,

$$G^{SG}(x,x',\tau) = [\theta(\tau - |z|)/2]\{J_0(s) + \beta_2[sJ_1(s) - 2\tau\Lambda_1(w,s)] - \beta_3 \operatorname{sgn}(z)[-(\tau + |z|)J_0(s) + 2\tau\Lambda_0(w,s)]\}, \tag{3.29}$$

explicitly demonstrating the retarded boundary conditions which have been applied.

B. The ϕ^4 potential

With a slight generalization, the techniques used to evaluate the SG Green's function may be applied to the ϕ^4 potential. Proceeding along the same lines, we write the phonon contribution as

$$G_p^{\phi^4}(x,x',\tau) = \frac{\theta(\tau)}{4}\left\{\gamma_0 I_0 - \gamma_1 \operatorname{sgn}(z) \frac{dI_0}{d\,|z|} + \gamma_2 I_2 + \gamma_3 \operatorname{sgn}(z) \frac{dI_2}{d\,|z|} + I_4\right\}, \tag{3.30}$$

where I_2 and $dI_2/d\,|z|$ are given in Eqs. (3.24), (3.25), and

$$I_0 = \frac{1}{\pi} \int_{0}^{\infty} dk \frac{\cos(|z|k)\sin(\tau\sqrt{1+k^2})}{(1+k^2)^{3/2}(1+4k^2)}, \tag{3.31a}$$

$$I_4 = \frac{1}{\pi} \int_{0}^{\infty} dk \times \frac{(1+4k^2)\cos(|z|k)\sin(\tau\sqrt{1+k^2})}{(1+k^2)^{3/2}}, \tag{3.31b}$$

$$= 2\theta(\tau - |z|)J_0(s) - 3I_2, \tag{3.32}$$

$$\begin{aligned} &\gamma_0 \equiv 9\{\tanh^2(y)\tanh^2(y') - \tanh(y)\tanh(y')\},\\ &\gamma_1 \equiv 18\{\tanh(y)\tanh^2(y') - \tanh^2(y)\tanh(y')\},\\ &\gamma_2 \equiv 9\tanh(y)\tanh(y') - 3\tanh^2(y) - 3\tanh^2(y'),\\ &\gamma_3 \equiv 6\tanh(y) - 6\tanh(y'),\\ &y \equiv x/2,\\ &y' \equiv x'/2, \end{aligned} \tag{3.33}$$

where Eq. (3.11) has been used to simplify Eq. (3.31b). The remaining integral, I_0, may be reduced by partial fractions to

$$I_0=\frac{4}{3\pi}\int_0^\infty dk\,\frac{\cos(|z|k)\sin(\tau\sqrt{1+k^2})}{\sqrt{1+k^2}(1+4k^2)}-\frac{I_2}{3}, \qquad (3.34)$$

$$=\tfrac{4}{3}I_{01}-\tfrac{1}{3}I_2, \qquad (3.35)$$

with I_{01} defined by

$$I_{01}=\frac{1}{\pi}\int_0^\infty dk\,\frac{\cos(|z|k)\sin(\tau\sqrt{1+k^2})}{\sqrt{1+k^2}(1+4k^2)}. \qquad (3.36)$$

To evaluate I_{01} we again substitute $k=\sinh(u)$ which gives us

$$I_{01}=\frac{1}{\pi}\int_0^\infty du\,\frac{\cos[|z|\sinh(u)]\sin[\tau\cosh(u)]}{1+4\sinh^2(u)}, \qquad (3.37)$$

$$=\frac{1}{2\pi}\int_0^\infty\frac{t\,dt}{t^4-t^2+1}\sin\left[at+\frac{b}{t}\right], \qquad (3.38)$$

where in going from Eq. (3.37) to Eq. (3.38) substitutions similar to those made in Eqs. (3.14)–(3.18) have been made. Factoring the denominator of Eq. (3.38), we define

$$\beta^2_\pm=-t^2_\pm=-\beta_\mp=(-1\mp i\sqrt{3})/2, \qquad (3.39)$$

where $t^2_\pm$ are the roots of t^4-t^2+1. Using partial fractions, we may write Eq. (3.38) as

$$I_{01}=\frac{1}{2\pi i\sqrt{3}}\left\{\int_0^\infty\frac{t\,dt}{t^2+\beta^2_+}\sin\left[at+\frac{b}{t}\right]-\int_0^\infty\frac{t\,dt}{t^2+\beta^2_-}\sin\left[at+\frac{b}{t}\right]\right\}, \qquad (3.40)$$

$$=\frac{-1}{2i\sqrt{3}}[J(\beta^2_-)-J^*(\beta^2_-)], \qquad (3.41)$$

$$=\frac{-1}{\sqrt{3}}\mathrm{Im}[J(\beta^2_-)], \qquad (3.42)$$

where

$$J(\beta^2)\equiv-\frac{1}{\pi}\int_0^\infty\frac{t\,dt}{t^2+\beta^2}\sin\left[at+\frac{b}{t}\right]. \qquad (3.43)$$

The integral defined in Eq. (3.43) is a slight generalization of Hardy's integrals for Lommel functions.[14,22] The evaluation of $J(\beta^2)$ follows Hardy's with a few modifications and is presented in Appendix A for completeness. From Eq. (A21) in Appendix A we have

$$J(\beta^2_-)=\tfrac{1}{2}e^{-(a\beta_--b/\beta_-)}-\theta(b)\Lambda_2[2b/\beta_-,2\sqrt{ab}\,], \qquad (3.44)$$

$$=\tfrac{1}{2}e^{-(1/2)(|z|+i\sqrt{3}\tau)}-\theta(\tau-|z|)\Lambda_2(\beta_+w,s). \qquad (3.45)$$

Therefore we have for I_{01}

$$I_{01}=(1/2\sqrt{3})e^{-|z|/2}\sin(\omega_{b,2}\tau)+[\theta(\tau-|z|)/\sqrt{3}]\mathrm{Im}[\Lambda_0(\beta_+w,s)], \qquad (3.46)$$

where

$$\omega_{b,2}\equiv\sqrt{3}/2, \qquad (3.47)$$

and we have used

$$\mathrm{Im}[\Lambda_2(\beta_+w,s)]=\mathrm{Im}[\Lambda_0(\beta_+w,s)+J_0(s)]=\mathrm{Im}[\Lambda_0(\beta_+w,s)]. \qquad (3.48)$$

From Eq. (3.30) we see that we need a derivative of I_0, and hence I_{01}, with respect to $|z|$. Using Eqs. (B6) and (B16) from Appendix B we have

$$\frac{dI_{01}}{d|z|}=\frac{-1}{4\sqrt{3}}e^{-|z|/2}\sin(\omega_{b,2}\tau)-\frac{\theta(\tau-|z|)}{2\sqrt{3}}\mathrm{Im}[\Lambda_1(\beta_+w,s)], \qquad (3.49)$$

where

$$(\beta^2_++1)/\beta_+=1, \qquad (3.50)$$

has also been used. Collecting all of the pieces, we write for the phonon contribution

$$G_p^{\phi^4}(x,x',\tau)=\frac{\theta(\tau)}{4}\left\{\frac{4}{3}\gamma_0I_{01}-\frac{4}{3}\gamma_1\,\mathrm{sgn}(z)\frac{dI_{01}}{d|z|}+\left[\gamma_2-\frac{\gamma_0}{3}-3\right]I_2+\mathrm{sgn}(z)\left[\frac{\gamma_1}{3}+\gamma_3\right]\frac{dI_2}{d|z|}+2\theta(\tau-|z|)J_0(s)\right\}. \qquad (3.51)$$

As in the sine–Gordon case one may show that when we combine the "nonretarded" pieces of the phonon contribution, we obtain exactly the negative of the bound state contribution; specifically we have

$$\frac{1}{8}\left[\gamma_2-\frac{\gamma_0}{3}-3\right]\tau e^{-|z|/2}-\frac{\mathrm{sgn}(z)}{8}\left[\frac{\gamma_1}{3}+\gamma_3\right]\tau e^{-|z|/2}=-\tau f^*_{b,1}(x)f_{b,1}(x'), \qquad (3.52a)$$

$$\frac{1}{6\sqrt{3}}e^{-|z|/2}\sin(\omega_{b,2}\tau)\gamma_0+\frac{1}{12\sqrt{3}}e^{-|z|/2}\sin(\omega_{b,2}\tau)\mathrm{sgn}(z)\gamma_1=-\frac{\sin(\omega_{b,2}\tau)}{\omega_{b,2}}f^*_{b,2}(x)f_{b,2}(x'). \qquad (3.52b)$$

With the nonretarded portion cancelled by the bound state contribution, we have for the full Green's function

$$\begin{aligned}G^{\phi^4}(x,x',\tau)&=\theta(\tau-|z|)\{(1/3\sqrt{3})\mathrm{Im}[\gamma_0\Lambda_0(\beta_+w,s)\\&\quad+\tfrac{1}{2}\gamma_1\,\mathrm{sgn}(z)\Lambda_1(\beta_+w,s)]\\&\quad+\tfrac{1}{8}[\gamma_2-\gamma_0/3-3][sJ_1(s)-2\tau\Lambda_1(w,s)]\\&\quad+[\mathrm{sgn}(z)/8][\gamma_1/3+\gamma_3][-(\tau+|z|)J_0(s)\\&\quad+2\tau\Lambda_0(w,s)]+\tfrac{1}{2}J_0(s)\}.\end{aligned} \qquad (3.53)$$

C. The DQ potential

As a final example, we evaluate the DQ Green's function. The phonon contribution in this case is

$$G_p^{\mathrm{DQ}}(x,x',\tau)=\theta(\tau-|z|)\left\{I_1-\left[I_2(z_+)-\frac{dI_2(z_+)}{dz_+}\right]\right\}, \qquad (3.54)$$

where I_1 is given in Eq. (3.11) (with $\mu=1$) and $I_2(z_+)$ is given in Eq. (3.24) with $|z|$ replaced by $z_+\equiv|x|+|x'|$. Factoring out the nonretarded piece we have

$$G^{DQ}(x,x',\tau) = [\theta(\tau-|z|)/2]\{J_0(s) - s_+J_1(s_+) + 2\tau\Lambda_1(w_+,s_+) + (\tau+z_+)J_0(s_+) + 2\tau\Lambda_0(w_+,s_+)\}, \tag{3.55}$$

with

$$z_+ \equiv |x| + |x'|, \quad w_+ \equiv \tau - z_+, \quad s_+ \equiv \sqrt{\tau^2 - z_+^2}. \tag{3.56}$$

All three of the Green's functions derived above have been checked against numerical integration. Over a large range of values for x, x', and τ, we find agreement to eight significant digits, which is presently the accuracy of our routines which compute the modified Lommel functions. In addition we have applied the small oscillation operator [see Eq. (3.2)] to each of the analytic expressions which, after some tedious algebra, yield the appropriate delta functions. To obtain a final check, we note that by using the orthogonality relation in Eq. (2.13c) and linear superposition, we see that the phonon contribution to the Green's functions must be orthogonal to the bound state(s). Numerical integrations confirm this property for all three Green's functions.

IV. ASYMPTOTIC BEHAVIOR

To obtain asymptotic expressions ($\tau \to \infty$) for the Green's functions, we must first find the appropriate limits of the modified Lommel functions. In Appendix C we examine $\Lambda_0(w,s)$ and $\Lambda_1(w,s)$ in the limit as $s \to \infty$ while $w/s \to 1$, which, when w and s are related to τ and z by Eqs. (3.23), corresponds to $\tau \gg |z|$. This limit is interesting because the expressions for the phonon contributions to the Green's functions have a term linear in τ which, in view of the integral expressions, must be cancelled by the other terms.

Since all of the Green's functions are expressible in terms of the integrals I_{01}, I_2 and their derivatives with respect to $|z|$, we consider the asymptotic expressions for these quantities first and then combine them to obtain the limits for the Green's functions.

To apply the results of Appendix C we must first recast these results in terms of the variables τ and z which are related to w and s by

$$w = \beta(\tau - |z|), \quad s = \sqrt{\tau^2 - z^2}, \tag{4.1}$$

where β is either unity or β_+. From Eqs. (C31) and (C32) of Appendix C, we have for $\beta = 1$,

$$\Lambda_0(w,s) \approx \frac{J_0(s)}{2} + \frac{e^{-|z|}}{2} + \frac{|z|}{2\tau}\sqrt{\frac{2}{\pi s}}\left\{\cos\left(s-\frac{\pi}{4}\right) + \sin\left(s-\frac{\pi}{4}\right)\frac{2R_2(1,\kappa)}{8s}\right\} + O(\tau^{-7/2}), \tag{4.2}$$

$$\Lambda_1(w,s) \approx \frac{e^{-|z|}}{2} - \frac{s}{2\tau}\sqrt{\frac{2}{\pi s}}\left\{\cos\left(s-\frac{\pi}{4}\right)\left[\frac{2[R_2(1,\kappa)-2]}{8s}\right] - \sin\left(s-\frac{\pi}{4}\right)\left[1 + \frac{2[R_4(1,\kappa)+12R_2(1,\kappa)]}{(8s)^2}\right]\right\} + O(\tau^{-9/2}), \tag{4.3}$$

where $\kappa \equiv w/s$, R_2, and R_4 are defined in Eqs. (C29) and (C30), and we have used [see Eqs. (C13)]

$$\epsilon(1,\kappa) = |z|/s, \quad \sigma_1(1,\kappa) = \tau/2s, \quad \sigma_2(1,\kappa) = \tau/2|z|,$$
$$\sigma_1(1,\kappa)/\sqrt{1+\epsilon^2(1,\kappa)} = \tfrac{1}{2}, \quad \epsilon(1,\kappa)\sigma_2(1,\kappa)/[1+\epsilon^2(1,\kappa)] = s/2\tau. \tag{4.4}$$

Inserting the expression for $\Lambda_1(w,s)$ given by Eq. (4.3) into Eq. (3.24), we see that the linear τ dependence exactly cancels [for large τ and $\tau \gg |z|$, both $\theta(\tau - |z|)$ and $\theta(\tau)$ are unity], leaving us with

$$I_2 \approx \frac{sJ_1(s)}{2} + \frac{s}{2}\sqrt{\frac{2}{\pi s}}\left\{\cos\left(s-\frac{\pi}{4}\right)\left[\frac{2[R_2(1,\kappa)-2]}{8s} - \frac{40R_4(1,\kappa)}{(8s)^3}\right] - \sin\left(s-\frac{\pi}{4}\right)\left[1 + \frac{2[R_4(1,\kappa)+12R_2(1,\kappa)]}{(8s)^2}\right]\right\} + O(\tau^{-7/2}). \tag{4.5}$$

In Eq. (4.5), I_2 now seems to have a $\sqrt{s}$ and therefore $\sqrt{\tau}$ dependence; however, this again exactly cancels when $J_1(s)$ is expanded in its asymptotic series resulting in

$$I_2 \approx \frac{1}{2}\sqrt{\frac{2}{\pi s}}\left\{\sin\left(s-\frac{\pi}{4}\right)\left[\frac{15 - 4[R_4(1,\kappa)+12R_2(1,\kappa)]}{16(8s)}\right] + \cos\left(s-\frac{\pi}{4}\right)\left[\frac{2R_2(1,\kappa)-1}{8} + \frac{5[21/16 - R_4(1,\kappa)]}{(8s)^2}\right]\right\} + O(\tau^{-7/2}). \tag{4.6}$$

Similarly, we have

$$\frac{dI_2}{d|z|} \approx \frac{|z|}{2}\sqrt{\frac{2}{\pi s}}\left\{\cos\left(s-\frac{\pi}{4}\right)\left[\frac{9+4R_2(1,\kappa)}{2(8s)^2}\right] + \sin\left(s-\frac{\pi}{4}\right)\left[\frac{2R_2(1,\kappa)-1}{(8s)}\right]\right\} + O(\tau^{-7/2}). \tag{4.7}$$

Next we turn to the I_{01} expression which involves modified Lommel functions evaluated at $\beta_+ w$ and s. With $\beta = \beta_+$, $\epsilon(\beta,\kappa)$, $\sigma_1(\beta,\kappa)$, and $\sigma_2(\beta,\kappa)$ become

$$\epsilon(\beta_+,\kappa) = (|z| + i\sqrt{3}\tau)/2s, \quad \sigma_1(\beta_+,\kappa) = (\tau + i\sqrt{3}|z|)/4\pi s, \quad \sigma_2(\beta_+,\kappa) = \frac{\kappa}{2s}\frac{(\tau+i\sqrt{3}|z|)(\tau+|z|)}{|z|+i\sqrt{3}\tau}. \tag{4.8}$$

Inserting Eqs. (4.8) into Eqs. (C31) and (C32), we have

$$\Lambda_0(\beta_+w,s)\approx\frac{1}{2}e^{-|z|/2}e^{i\omega_{b,2}t}+\frac{1}{2}\frac{1}{\sqrt{1+\epsilon^2(\beta_+,\kappa)}}\sqrt{\frac{2}{\pi s}}\left\{\cos\left(s-\frac{\pi}{4}\right)\left[1+\frac{2R_4(\beta_+,\kappa)}{(8s)^2}\right]\right.$$
$$\left.+\sin\left(s-\frac{\pi}{4}\right)\frac{2R_2(\beta_+,\kappa)}{(8s)}\right\}+O(\tau^{-7/2}),\tag{4.9}$$

$$\Lambda_1(\beta_+w,s)\approx\frac{1}{2}e^{-|z|/2}e^{-i\omega_{b,2}t}-\frac{1}{2}\frac{1}{\sqrt{1+\epsilon^2(\beta_+,\kappa)}}\sqrt{\frac{2}{\pi s}}\left\{\cos\left(s-\frac{\pi}{4}\right)\left[\frac{2[R_2(\beta_+,\kappa)-2]}{8s}\right.\right.$$
$$\left.\left.-40\frac{R_4(\beta_+,\kappa)}{(8s)^3}\right]-\sin\left(s-\frac{\pi}{4}\right)\left[1+\frac{2[R_4(\beta_+,\kappa)+12R_2(\beta_+,\kappa)]}{(8s)^2}\right]\right\}+O(\tau^{-9/2}),\tag{4.10}$$

where we have used

$$\sigma_1(\beta_+,\kappa)/\sqrt{1+\epsilon^2(\beta_+,\kappa)}=\tfrac{1}{2},\quad \epsilon(\beta_+,\kappa)\sigma_2(\beta_+,\kappa)/\sqrt{1+\epsilon^2(\beta_+,\kappa)}=\tfrac{1}{2}.\tag{4.11}$$

When Eq. (4.9) is inserted into the expression for I_{01}, the oscillatory term in τ cancels leaving

$$I_{01}\approx\frac{1}{2\sqrt{3}}\,\mathrm{Im}\left\{\frac{1}{\sqrt{1+\epsilon^2(1,\kappa)}}\sqrt{\frac{2}{\pi s}}\left[\cos\left(s-\frac{\pi}{4}\right)\left(1+\frac{2R_4(\beta_+,\kappa)}{(8s)^2}\right)\right.\right.$$
$$\left.\left.+\sin\left(s-\frac{\pi}{4}\right)\frac{2R_2(\beta_+,\kappa)}{(8s)}\right]\right\}+O(\tau^{-7/2}),\tag{4.12}$$

and

$$\frac{dI_{01}}{d|z|}\approx\frac{1}{2\sqrt{3}}\,\mathrm{Im}\left\{\frac{1}{\sqrt{1+\epsilon^2(\beta_+,\kappa)}}\sqrt{\frac{2}{\pi s}}\left[\cos\left(s-\frac{\pi}{4}\right)\left(\frac{2[R_2(\beta_+,\kappa)-2]}{8s}\right)\right.\right.$$
$$\left.\left.-\sin\left(s-\frac{\pi}{4}\right)\left(1+\frac{2[R_4(\beta_+,\kappa)+12R_2(\beta_+,\kappa)]}{(8s)^2}\right)\right]\right\}+O(\tau^{-7/2}).\tag{4.13}$$

Now all of the contributions are at hand to obtain, through $O(\tau^{-7/2})$, the asymptotic forms for the Green's functions. However, since the expressions are lengthy and not particularly illuminating, we list only the leading terms. Due to the simple analytic form of the bound state contribution, we list only the phonon portions,

$$G_p^{\mathrm{SG}}(x,x',\tau)\approx\sqrt{\frac{2}{\pi s}}\left\{\cos\left(s-\frac{\pi}{4}\right)+\frac{1}{8s}\sin\left(s-\frac{\pi}{4}\right)\right\}+O(\tau^{-5/2}),\tag{4.14}$$

$$G_p^{\phi^4}(x,x',\tau)\approx\sqrt{\frac{2}{\pi s}}\left\{\cos\left(s-\frac{\pi}{4}\right)\left[\frac{\gamma_0}{6\sqrt{3}}\,\mathrm{Im}\left(\frac{1}{\sqrt{1+\epsilon^2(\beta_+,\kappa)}}\right)+\frac{1}{8}\left(\gamma_2-\frac{\gamma_0}{3}-3\right)\left(\frac{2R_2(1,\kappa)-1}{8}\right)+2\right]\right.$$
$$\left.-\sin\left(s-\frac{\pi}{4}\right)\left[\frac{\gamma_1\,\mathrm{sgn}(z)}{12\sqrt{3}}\,\mathrm{Im}\left(\frac{1}{\sqrt{1+\epsilon^2(\beta_+,\kappa)}}\right)\right]\right\}+O(\tau^{-3/2}),\tag{4.15}$$

$$G_p^{\mathrm{DQ}}(x,x',\tau)\approx\sqrt{\frac{2}{\pi s}}\cos\left(s-\frac{\pi}{4}\right)-\frac{1}{2}\sqrt{\frac{2}{\pi s_+}}\cos\left(s_+-\frac{\pi}{4}\right)\left[\frac{2R_2(1,\kappa_+)-1}{8}\right]+O(\tau^{-3/2}),\tag{4.16}$$

where in Eq. (4.16), $\kappa_+\equiv w_+/s_+$.

One may notice that although we have shown that there is no linear τ term in the phonon contributions to the Green's functions, the full Green's functions have a linear τ term due to the first bound state, namely,

$$\theta(\tau)\tau f^*_{b,1}(x)f_{b,1}(x').\tag{4.17}$$

This term may be understood by realizing that when computing the response of a soliton to a perturbation, the effect of this term is to produce a coefficient of the translation mode $f_{b,1}(x)$ which increases with time. Therefore the soliton will move from its initial position as time progresses. Hence in this case, the linear term is associated with the translation of the soliton.

The secularity referred to in the Introduction is made evident by the linear τ behavior in the coefficient of the translation-mode contribution to the full Green's function. Indeed, the use of the full Green's function in a perturbation theory of kink dynamics in the presence of external influences is equivalent to the procedure introduced by Fogel *et al.*[2] The use of the collective-coordinate method[10–13] avoids the secularity associated with the translation mode since only the phonon part of the Green's function is employed[10–13] [together with the contribution from other bound states, if any ($N\geqslant 2$)].

Note added in proof: Recently we have been able to obtain analytic expressions for the Laplace transform of the product of the Lommel function $\Lambda_n(w,s)$ and the step function $\theta(\tau-|z|)$, with w and s related to τ and z by Eqs. (3.23). Specifically we have

$$\int_0^\infty d\tau\, e^{-\bar{s}\tau}\theta(\tau-|z|)\Lambda_n(w,s)$$
$$=\frac{1}{2s}\frac{\exp(-|z|\sqrt{\bar{s}^2+1})}{\sqrt{\bar{s}^2+1}}\left[\frac{1}{\sqrt{\bar{s}^2+1}+\bar{s}}\right]^{n-1},$$

where $\bar{s}$ is the Laplace transform variable. This leads to the remarkably simple expression for the Laplace transform of the sine–Gordon Green's function

$$G^{SG}(x,x';\bar{s})\equiv\int_0^\infty d\tau\, G^{SG}(x,x',\tau)e^{-\bar{s}\tau}$$
$$=\frac{\exp(-|z|\sqrt{\bar{s}^2+1})}{2}$$
$$\times\left\{\frac{1}{\sqrt{\bar{s}^2+1}}-\frac{\beta_2}{\bar{s}^2\sqrt{\bar{s}^2+1}}-\frac{\beta_3\,\mathrm{sgn}(z)}{\bar{s}^2}\right\},$$

with similar expressions for the ϕ^4 and the DQ Green's functions.

V. REPRESENTATIVE PLOTS

To illustrate the behavior of the Green's functions, we present several plots of the *phonon* part of the SG Green's function [plots for the other Green's functions derived look very similar] in the x-x' plane for different values of τ. The numerical values for these plots are easily obtained from the formulas in Appendix C.

Since the Green's functions depend only on $\tau=t-t'$, we are free to choose t' and let t be fixed by τ and t'. Choosing $t'=0$, Fig. 1 shows the evolution for the phonon contribution as time progresses. To interpret these plots, recall that $G(x,x',t,t')$ may be viewed as the response of the system at (x,t) due to a delta function source at (x',t'). Fixing $x'=8$ in Fig. 1(a), we move in the direction of increasing x, starting at $x=0$. Until x is on the order of 2, $G(x,x',\tau)$ is zero, meaning that the disturbance has not yet had enough time to propagate from $x=8$ to $x<2$ (or $x>14$). For $\tau=4$, time has progressed (recall we have fixed $t'=0$) and the disturbance has propagated out further. At $t=8$ the pulse reaches $x=8$. In Fig. 1(e)–1(h) the pulse has propagated off the scales, leaving behind "ripples." As τ further increases the amplitude continues to decrease in accord with the asymptotic behavior derived in Sec. IV.

If one were to follow the procedure outlined in the preceding paragraph with $x'=3$, one would note that before the pulse arrives at a particular position, the Green's function is not zero. This is because we have plotted the phonon contribution, which has a nonretarded part which exactly cancels the bound state contribution. It is this nonretarded part which gives a nonzero value for the phonon contribution to the Green's function "before the pulse arrives." We see this only near $x=x'=0$ because the bound state contribution is proportional to $e^{-|z|}$ (SG), $\mathrm{sech}(x)\mathrm{sech}(x')$ [ϕ^4], or $e^{-|x|}e^{-|x'|}$ (DQ).

ACKNOWLEDGMENTS

R. J. F. wishes to thank the Center for Nonlinear Studies at the Los Alamos National Laboratory for its hospitality and support during the completion of this work.

S. E. T. wishes to acknowledge the support of an Alfred P. Sloan Research Fellowship and also to thank the Department of Physics at the University of California, Irvine, for its hospitality during a sabbatical visit when this work was begun. The research reported here was partially supported by the U. S. Department of Energy.

APPENDIX A: EVALUATION OF THE INTEGRAL $J(\beta^2)$

The integral $J(\beta^2)$ [Eq. (A1)] differs from Hardy's integral for Lommel functions[14,22] only in that in the denominator t^2+1 is replaced by $t^2+\beta^2$. The only restriction placed on β is that $\mathrm{Re}(\beta)>0$. We first consider the case in which $b<0$ for which we have from the tables[23]

$$J(\beta^2)=\frac{1}{\pi}\int_0^\infty\frac{t\,dt}{t^2+\beta^2}\sin\left[at+\frac{b}{t}\right]$$
$$=\frac{1}{2}e^{-(a\beta-b/\beta)},\tag{A1}$$

where the restriction $\mathrm{Re}(\beta)>0$ is required.

For $b>0$ we distinguish between $b<a$ and $b>a$. The latter may be reduced to the $b<a$ case by using the relation[24]

$$\frac{1}{\pi}\int_0^\infty\frac{t\,dt}{t^2+\beta^2}\sin\left[at+\frac{b}{t}\right]$$
$$=J_0(2\sqrt{ab})-\frac{1}{\pi}\int_0^\infty\frac{t\,dt}{t^2+1/\beta^2}\sin\left[\frac{a}{t}+bt\right].\tag{A2}$$

Therefore we need only consider $b<a$. We may further restrict ourselves to $|\beta|=1$ by writing $\beta=|\beta|\,e^{i\varphi}$ which allows us to write

$$J(\beta^2)=\frac{1}{\pi}\int_0^\infty\frac{t\,dt}{|\beta|^2[t^2/|\beta|^2+e^{2i\varphi}]}\sin\left[at+\frac{b}{t}\right],\tag{A3}$$
$$=\frac{1}{\pi}\int_0^\infty\frac{t\,dt}{t^2+e^{2i\varphi}}\sin\left[a't+\frac{b'}{t}\right],\tag{A4}$$

where a' and b' are a and b scaled by $1/|\beta|$. Therefore with $b<a$ and $|\beta|=1$, we define

$$x\equiv2\sqrt{ab},\quad c\equiv(1/\beta)\sqrt{b/a},\tag{A5}$$

in terms of which we may write $J(\beta^2)$ as

$$J(\beta^2)=\frac{1}{\pi}\int_0^\infty\frac{t\,dt}{t^2+\beta^2}\sin\left[\frac{x}{2}\left(t\sqrt{\frac{a}{b}}+\frac{1}{t}\sqrt{\frac{b}{a}}\right)\right],\tag{A6}$$
$$=\frac{c}{\pi}\int_{-\infty}^\infty\frac{e^u\,du}{ce^u+1/ce^u}\sin[x\cosh(u)],\tag{A7}$$
$$=\frac{c}{\pi}\int_0^\infty du\left\{\frac{e^{-u}}{ce^{-u}+(ce^{-u})^{-1}}+\frac{e^u}{ce^u+(ce^u)^{-1}}\right\}$$
$$\times\sin[x\cosh(u)],\tag{A8}$$

$$=\frac{1}{2\pi}\int_1^\infty\frac{d\tau}{\sqrt{\tau^2-1}}\frac{c^2-1+2\tau^2}{\theta^2+\tau^2}\sin(x\tau),\tag{A9}$$

with

$$\theta\equiv\frac{1}{2}\left(c-\frac{1}{c}\right)=\frac{c'^2+1}{2c'}\left\{\frac{c'^2-1}{c'^2+1}\mathrm{Re}(\beta)-i\,\mathrm{Im}(\beta)\right\},\tag{A10}$$

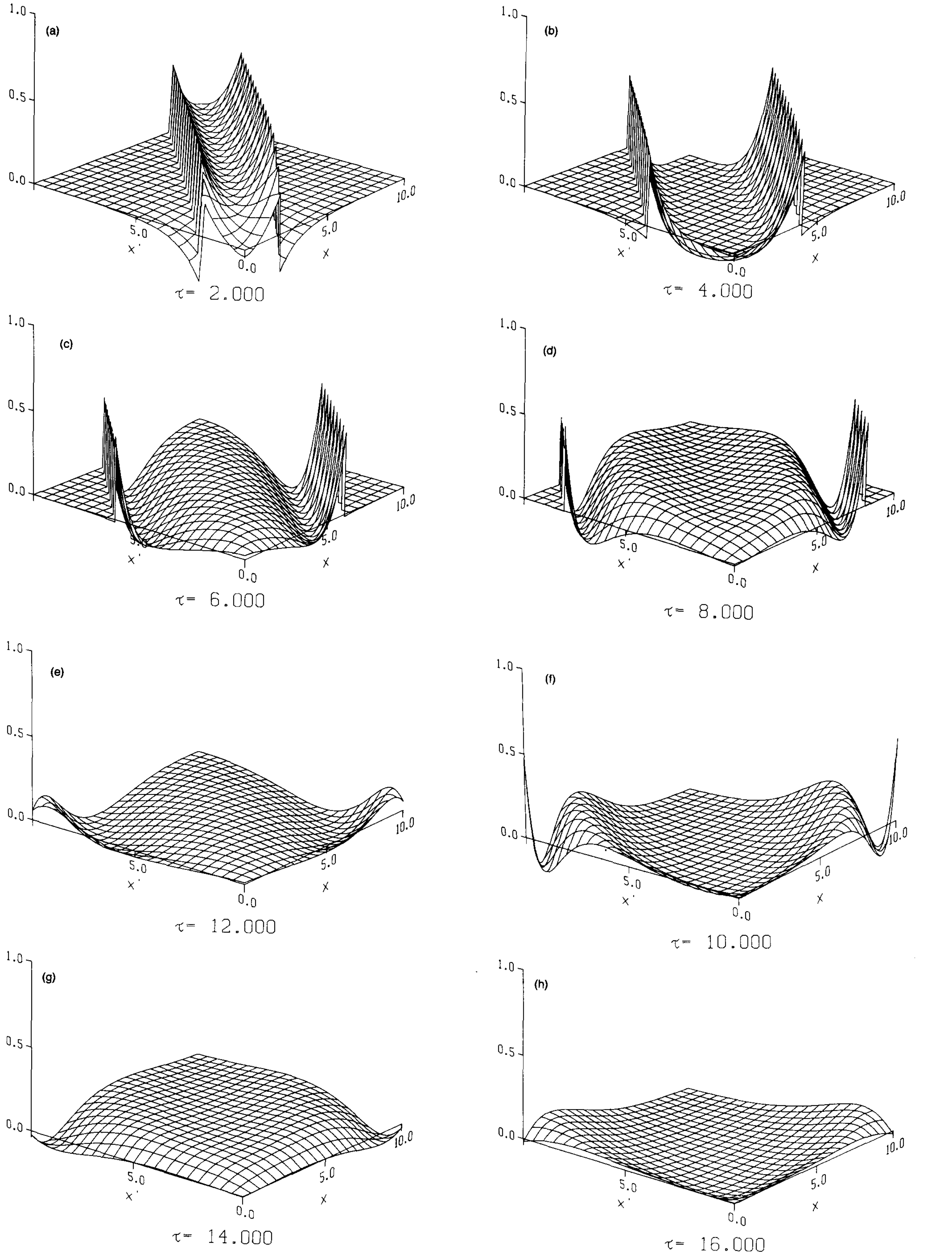

FIG. 1. The time evolution of the phonon contribution to the SG Green's function $G(x,x',t-t')$ in the xx' plane. Here we have chosen $t'=0$, therefore $\tau=t$. In Figs. 1(a)–1(d) we see a disturbance "propagating outward," 1(a) and 1(b) show the nonretarded portion near $x=x'=0$. In 1(e)–1(h) the pulse has moved off of our scales, leaving behind undulations which decrease with increasing time.

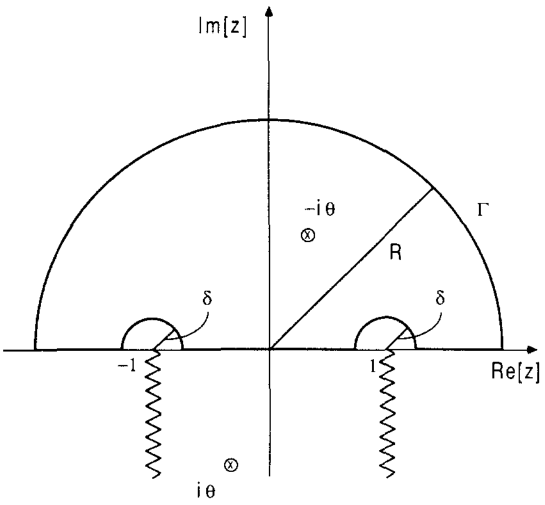

FIG. 2. The contour Γ for the evaluation of the integral $J(\beta^2)$.

$$c' \equiv \sqrt{b/a}. \tag{A11}$$

Since $\mathrm{Re}(\beta) > 0$ and $c' < 1$, θ is never pure imaginary; therefore θ^2 does not lie on the negative real axis and the only poles of the integrand in Eq. (A9) are at $\tau = \pm 1$. We evaluate Eq. (A9) by considering the contour integral $\Gamma(\beta^2)$ given by

$$\Gamma(\beta^2) \equiv \int_\Gamma \frac{dz\, e^{ixz}}{\sqrt{z^2-1}} \frac{c^2 - 1 + 2z^2}{\theta^2 + z^2}. \tag{A12}$$

With the branch cuts chosen as in Fig. 2, $\Gamma(\beta^2)$ becomes

$$\Gamma(\beta^2) = 2i \int_1^\infty \frac{d\tau \sin(x\tau)}{\sqrt{\tau^2-1}} \frac{c^2-1+2\tau^2}{\theta^2+\tau^2} - 2i \int_{-1}^{1} \frac{d\tau\, e^{ix\tau}}{\sqrt{1-\tau^2}} \frac{c^2-1+2\tau^2}{\theta^2+\tau^2}. \tag{A13}$$

Therefore we have for $J(\beta^2)$,

$$J(\beta^2) = \frac{1}{2\pi i} \frac{\Gamma(\beta^2)}{2} + \frac{1}{2\pi} \times \int_0^1 \frac{d\tau \cos(x\tau)}{\sqrt{1-\tau^2}} \frac{c^2-1+2\tau^2}{\theta^2+\tau^2}, \tag{A14}$$

$$= \frac{\mathrm{Res}[f(z); -i\theta]}{2} + \frac{1}{2\pi} \times \int_0^{\pi/2} d\varphi \cos[x\cos(\varphi)] \frac{c^2 - 1 + 2\cos^2(\varphi)}{\theta^2 + \cos^2(\varphi)}, \tag{A15}$$

where $\mathrm{Res}[f(z); -i\theta]$ is the residue of $f(z)$ evaluated at $-i\theta$ with $f(z)$ given by the integrand of Eq. (A12). In writing Eq. (A14) we have used the fact the contributions to $\Gamma(\beta^2)$ from the large and small semicircles vanish when $R \to \infty$ and $\delta \to 0$, respectively. Evaluating the residue at the simple pole $-i\theta$ we have

$$\mathrm{Res}[f(z); -i\theta] = e^{-(a\beta - b/\beta)}. \tag{A16}$$

The remaining integral in Eq. (A15) may be evaluated by noting that

$$\frac{c^2 - 1 + 2\cos^2(\varphi)}{\theta^2 + \cos^2(\varphi)} = -4 \sum_{k=1}^{\infty} (ic)^{2k} \cos(2k\varphi). \tag{A17}$$

Since $c < 1$, the sum in Eq. (A17) is uniformly convergent and we may insert it into Eq. (A15) and integrate term by term. We also make the substitution

$$\cos[x\cos(\varphi)] = J_0(x) + 2 \sum_{n=1}^{\infty} (-1)^n J_{2n}(x) \cos(2n\varphi). \tag{A18}$$

The double sum resulting from substitution of Eqs. (A17) and (A18) into Eq. (A15) is reduced to a single sum by orthogonality of the functions $\{\cos(2n\varphi)\}$ on $[0,\pi/2]$, leaving

$$\frac{1}{2\pi} \int_0^\infty \frac{d\tau \cos(x\tau)}{\sqrt{1-\tau^2}} \frac{c^2-1+2\tau^2}{\theta^2+\tau^2} = -\sum_{k=1}^{\infty} c^{2k} J_{2k}(x), \tag{A19}$$

$$= -\Lambda_2[2b/\beta, 2\sqrt{ab}\,]. \tag{A20}$$

Finally collecting Eqs. (A1), (A16), and (A20) we have

$$J(\beta^2) = \tfrac{1}{2} e^{-(a\beta - b/\beta)} - \theta(b)\Lambda_2[2b/\beta, \sqrt{2ab}\,]. \tag{A21}$$

APPENDIX B: PROPERTIES OF LOMMEL FUNCTIONS OF TWO VARIABLES

In this Appendix we review some of the properties of the Lommel functions $U_n(w,s)$ and derive additional relations and limiting forms for the special case in which the arguments are of the form

$$w = \beta(\tau - |z|), \tag{B1a}$$

$$s = \sqrt{\tau^2 - z^2}, \tag{B1b}$$

with β a complex constant independent of τ and z. Below we list some properties which we shall use to derive additional relations. We restrict ourselves to the $U_n(w,s)$ Lommel functions although similar relations exist for the $V_n(w,s)$ functions and may be found in the literature[14,25–27] along with many other properties not listed here. Using the recurrence relation for Bessel functions,[28] and the defining series for Lommel functions,

$$U_n(w,s) = \sum_{m=0}^{\infty} (-1)^m \left(\frac{w}{s}\right)^{2m+n} J_{2m+n}(s), \tag{B2}$$

one may derive the following:

$$U_n(w,s) = (w/s)^n J_n(s) - U_{n+2}(w,s), \tag{B3}$$

$$\frac{\partial U_n(w,s)}{\partial s} = -\frac{s}{w} U_{n+1}(w,s), \tag{B4}$$

$$\frac{\partial U_n(w,s)}{\partial w} = \frac{1}{2} U_{n-1}(w,s) + \frac{1}{2}\left(\frac{s}{w}\right)^2 U_{n+1}(w,s). \tag{B5}$$

For the variables (w,s) as defined in Eq. (B1) we have

$$\frac{\partial U_n(\beta w,s)}{\partial |z|} = -\frac{1}{2}\left[\beta U_{n-1}(\beta w,s) + \frac{1}{\beta} U_{n+1}(\beta w,s)\right], \tag{B6}$$

$$\frac{\partial U_n(\beta w,s)}{\partial \tau}$$

$$= -\frac{1}{2}\left[\beta U_{n-1}(\beta w,s) - \frac{1}{\beta} U_{n+1}(\beta w,s)\right], \quad \text{(B7)}$$

$$\frac{\partial^2 U_n(\beta w,s)}{\partial^2 |z|^2} = \frac{1}{4}\left[\beta^2 U_{n-2}(\beta w,s) + 2U_n(\beta w,s) + \frac{1}{\beta^2} U_{n+2}(\beta w,s)\right], \quad \text{(B8)}$$

$$\frac{\partial^2 U_n(\beta w,s)}{\partial^2 \tau^2} = \frac{1}{4}\left[\beta^2 U_{n-2}(\beta w,s) - 2U_n(\beta w,s) + \frac{1}{\beta^2} U_{n+2}(\beta w,s)\right]. \quad \text{(B9)}$$

Subtracting Eq. (B8) from Eq. (B9) we have

$$\frac{\partial^2 U_n(\beta w,s)}{\partial^2 \tau^2} - \frac{\partial^2 U_n(\beta w,s)}{\partial^2 |z|^2} = -U_n(\beta w,s). \quad \text{(B10)}$$

Therefore $U_n(\beta w,s)$ is a solution of the "massive" Klein–Gordon equation (at least in the positive half-space, since $|z|>0$).

The above properties hold for arbitrary complex w and s. We now focus on the modified functions in which w is pure imaginary ($w \to iw$). Introducing the notation

$$\Lambda_n(w,s) = i^{-n} U_n(iw,s), \quad \text{(B11)}$$

with w and s given by Eq. (B1), we consider the limit $z \to 0$ for which

$$w/s = [(\tau - |z|)/(\tau + |z|)]^{1/2} \to 1. \quad \text{(B12)}$$

For n even we have

$$\Lambda_{2n}(\tau,\tau) = -\sum_{m=1}^{n-1} J_{2m}(\tau) + \frac{1 - J_0(\tau)}{2}. \quad \text{(B13)}$$

For odd n we use an integral representation

$$\Lambda_{2n+1}(\tau,\tau) = -\sum_{m=0}^{n-1} J_{2m+1}(\tau) + \frac{1}{2}\int_0^\infty dx\, J_0(x), \quad \text{(B14)}$$

or in terms of Struve functions,[29]

$$\Lambda_{2n+1}(\tau,\tau) = -\sum_{m=0}^{n-1} J_{2m+1}(\tau) + \frac{1}{2} \times\left\{\tau J_0(\tau) + \frac{\pi\tau}{2}[J_1(\tau)\mathbf{H}_0(\tau) - J_0(\tau)\mathbf{H}_1(\tau)]\right\}. \quad \text{(B15)}$$

Finally we consider the limiting case of $\tau = |z|$, i.e., $s = w = 0$. Since for all $n \geqslant 1$ $J_n(0) = 0$, we have

$$\Lambda_0(0,0) = 1, \quad \text{(B16a)}$$

$$\Lambda_n(0,0) = 0, \quad n \geqslant 1. \quad \text{(B16b)}$$

While some of the properties (especially B10) derived above are useful for the actual derivation of the Green's functions, they are most useful when checking the analytic expressions by operating on them with the differential operator

$$\partial_{tt} - \partial_{xx} + V''[\phi_k(x)]. \quad \text{(B17)}$$

APPENDIX C: NUMERICAL EVALUATION AND ASYMPTOTIC FORMS FOR MODIFIED LOMMEL FUNCTIONS OF TWO VARIABLES

Numerical evaluation of the Green's functions derived in Sec. III requires an evaluation of the modified Lommel functions. Although Lommel functions of two real variables[30] and two purely imaginary variables[31] have been studied, to our knowledge no one has yet considered the modified functions. Below we present methods which are valid for w complex and s real (since we start by considering the modified functions and w may be complex, our methods also include the case of two real variables). Representing the first argument as βw, where $|\beta| = 1$ and w and s are real, we have for the defining series

$$\Lambda_n(\beta w,s) = \sum_{m=0}^{\infty}\left(\frac{\beta w}{s}\right)^{2m+n} J_{2m+n}(s), \quad \text{(C1)}$$

from which we deduce the symmetries

$$\Lambda_n(-\beta w,s) = (-1)^n \Lambda_n(\beta w,s), \quad \text{(C2a)}$$

$$\Lambda_n(\beta w,-s) = \Lambda_n(\beta w,s). \quad \text{(C2b)}$$

From Eqs. (C2) we see that we need only investigate the first quadrant of the s-w plane. Another relationship exists which allows us to further restrict our attention to the angular region $(0,\pi/4)$, i.e, the first octant. We obtain this property by recalling the generating function for Bessel functions[32]

$$e^{(s/2)[\beta\kappa - 1/\beta\kappa]} = \sum_{m=-\infty}^{\infty} (\beta\kappa)^m J_m(s), \quad \text{(C3)}$$

where $\kappa \equiv w/s$. Using the symmetry of the Bessel functions about the origin we have,

$$\cosh\left[\frac{s}{2}\left(\beta\kappa - \frac{1}{\beta\kappa}\right)\right] = \sum_{m=-\infty}^{\infty} (\beta\kappa)^{2m} J_{2m}(s), \quad \text{(C4a)}$$

$$\sinh\left[\frac{s}{2}\left(\beta\kappa - \frac{1}{\beta\kappa}\right)\right] = \sum_{m=-\infty}^{\infty} (\beta\kappa)^{2m+1} J_{2m+1}(s). \quad \text{(C4b)}$$

Next we note that

$$\Lambda_n\left(\frac{s^2}{\beta w},s\right) = \sum_{m=0}^{\infty}\left(\frac{s}{\beta w}\right)^{2m+n} J_{2m+n}(s), \quad \text{(C5)}$$

which leads us to

$$\sinh\left[\frac{s}{2}\left(\beta\kappa - \frac{1}{\beta\kappa}\right)\right] = \Lambda_1(\beta w,s) - \Lambda_1\left(\frac{s^2}{\beta w},s\right), \quad \text{(C6a)}$$

$$\cosh\left[\frac{s}{2}\left(\beta\kappa - \frac{1}{\beta\kappa}\right)\right] = -J_0(s) + \Lambda_0(\beta w,s) + \Lambda_0\left(\frac{s^2}{\beta w},s\right). \quad \text{(C6b)}$$

From Eqs. (C6) we see that we have a relationship which allows us to consider only the region of the first quadrant of the s-w plane in which $w/s < 1$, namely the first octant. In this region the series definition converges uniformly; however, that rate of convergence is very slow when w/s approaches 1. By comparison with the geometric series we see that since $J_n(s) < 1$ $\forall n$, we have as an error estimate for truncation after N terms

$$R_N < \kappa^{2N}/(1-\kappa^2). \quad \text{(C7)}$$

We note that the error estimate in Eq. (C7) is a very crude one as it does not take into account the decaying nature of the Bessel functions; however, it suffices for our calculations.

As $w/s \to 1$, the number of terms in the series needed to attain a given accuracy becomes unreasonably large. For values of $\kappa = w/s$ larger than some κ_0, we turn to an asymptotic expansion[33] of the modified Lommel functions. We begin by following Mayall's[34] procedure for obtaining an integral representation for the Lommel functions by substitution of an integral representation for the Bessel functions into the series and summing the series explicitly. We restrict ourselves to deriving expressions for Λ_0 and Λ_1. For small n the asymptotic expansion for Λ_n may be obtained from the recurrence relation for Lommel functions. The large n limit has not yet been examined.

Starting with the integral representation for Bessel functions

$$J_{2m}(s) = \frac{(-1)^m}{\pi}\int_0^\pi d\theta\, e^{is\cos(\theta)}\cos(2m\theta), \tag{C8}$$

we have

$$\Lambda_0(\beta w,s)$$

$$= \sum_{m=0}^{\infty} (\beta\kappa)^{2m}(-1)^m \frac{1}{\pi}\int_0^\pi d\theta\, e^{is\cos(\theta)}\cos(2m\theta), \tag{C9}$$

$$= \frac{1}{\pi}\int_0^\pi d\theta \frac{1+(\beta\kappa)^2\cos(2\theta)}{1+2(\beta\kappa)^2\cos(2\theta)+(\beta\kappa)^4} e^{is\cos(\theta)}, \tag{C10}$$

$$= \frac{J_0(s)}{2} + \frac{1-(\beta\kappa)^4}{2\pi} \times \int_0^\pi d\theta \frac{e^{is\cos(\theta)}}{1+2(\beta\kappa)^2\cos(2\theta)+(\beta\kappa)^4}, \tag{C11}$$

$$= \frac{J_0(s)}{2} + \sigma_1(\beta,\kappa)\frac{\epsilon(\beta,\kappa)}{\pi} \times \int_0^\pi d\theta \frac{e^{is\cos(\theta)}}{\epsilon^2(\beta,\kappa)+\cos^2(\theta)}, \tag{C12}$$

where

$$\epsilon(\beta,\kappa) \equiv [1-(\beta\kappa)^2]/2\beta\kappa, \tag{C13a}$$

$$\sigma_1(\beta,\kappa) \equiv [1+(\beta\kappa)^2]/4\beta\kappa, \tag{C13b}$$

and uniform convergence of the sum has been used. Similarly we may write

$$\Lambda_1(\beta w,s) = -\sigma_2(\beta,\kappa)\frac{\epsilon(\beta,\kappa)}{\pi}\frac{d}{ds} \times \int_0^\pi d\theta \frac{e^{is\cos(\theta)}}{\epsilon^2(\beta,\kappa)+\cos^2(\theta)}, \tag{C14}$$

with

$$\sigma_2(\beta,\kappa) \equiv \frac{1+(\beta\kappa)^2}{4} + \frac{\beta\kappa[1+\epsilon^2(\beta,\kappa)]}{2\epsilon(\beta,\kappa)}. \tag{C15}$$

At this point, Mayall's method no longer applies (unless $\beta = \pm i$) and we turn to an alternate derivation.

The integral

$$I(\epsilon,s) = \frac{\epsilon}{\pi}\int_0^\pi d\theta \frac{e^{is\cos(\theta)}}{\epsilon^2+\cos^2(\theta)}, \tag{C16}$$

which occurs in Eqs. (C12) and (C14), is a strong function of ϵ since in the limit as $\epsilon \to 0$ ($w/s \to 1$), we obtain a delta function. Other major contributions occur at the stationary points $\theta = 0,\pi$. To evaluate $I(\epsilon,s)$, we substitute $t = \cos(\theta)$, deform the contour and represent the integrals as a residue which captures the strong ϵ behavior, plus two integrals for which asymptotic expansions are easily derived. Substituting we have

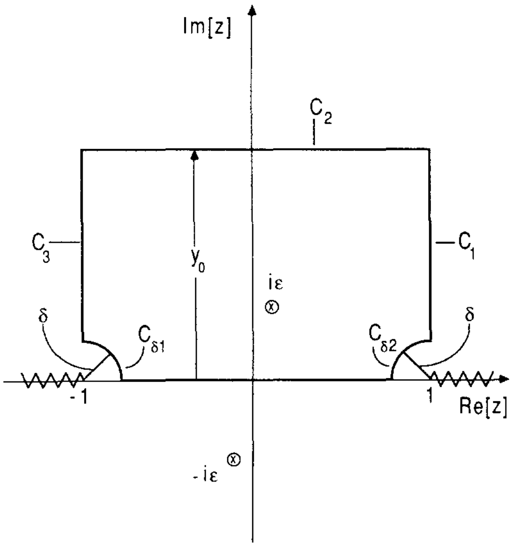

FIG. 3. The contour for the computation of the asymptotic expression for the modified Lommel function.

$$I(\epsilon,s) = \frac{\epsilon}{\pi}\int_{-1}^{1} dt \frac{e^{ist}}{(\epsilon^2+t^2)\sqrt{1-t^2}}, \tag{C17}$$

$$= \frac{\epsilon}{\pi}\left\{2\pi i \,\mathrm{Res}[f(z),i\epsilon] - \int_{c_1} dz \frac{e^{isz}}{(\epsilon^2+z^2)\sqrt{1-z^2}} - \int_{c_3} dz \frac{e^{isz}}{(\epsilon^2+z^2)\sqrt{1-z^2}}\right\}, \tag{C18}$$

where $f(z)$ is given by

$$f(z) = \frac{e^{isz}}{(\epsilon^2+z^2)\sqrt{1-z^2}}, \tag{C19}$$

and the contours are shown in Fig. 3. We have used the fact that as $\delta \to 0$ and $y_0 \to \infty$, the contributions from the contours $c_{\delta 1}$, $c_{\delta 2}$, and c_2 vanish by Jordon's lemma. Evaluating the residue and shifting the variables, we have

$$I(\epsilon,s) = \frac{e^{-\epsilon s}}{\sqrt{1+\epsilon^2}} - \frac{\epsilon}{\pi}\int_0^{i\infty} dz \frac{e^{isz}e^{is}}{[\epsilon^2+(z+1)^2]\sqrt{1-(z+1)^2}} - \frac{\epsilon}{\pi}\int_{i\infty}^{0} dz \frac{e^{isz}e^{-is}}{[\epsilon^2+(z-1)^2]\sqrt{1-(z-1)^2}}, \tag{C20}$$

$$= \frac{e^{-\epsilon s}}{\sqrt{1+\epsilon^2}} - \frac{\epsilon}{\pi}[J+J^*], \tag{C21}$$

where

$$J \equiv ie^{is}\int_0^\infty dy \frac{e^{-sy}}{[\epsilon^2+(iy+1)^2]\sqrt{1-(iy+1)^2}}, \tag{C22}$$

$$= 2ie^{is}\int_0^\infty dx \frac{e^{-sx^2}}{[\epsilon^2+(ix^2+1)^2]\sqrt{x^2-2i}}. \tag{C23}$$

As written in Eq. (C23), J is in one of Dingle's[35] standard integral forms which has as an asymptotic expansion

$$J \approx 2ie^{is}\sqrt{\frac{\pi}{2F_2}}\, e^{-F_0}\sum_{r=0}^{\infty} Q_r, \tag{C24}$$

where

$$\begin{aligned} Q_0 &= G_0,\\ Q_1 &= (-\sqrt{2}/3\sqrt{\pi}F_2^{3/2})[-3G_1F_2],\\ Q_2 &= (1/24F_2^3)[12G_2F_2^2],\\ Q_3 &= (-\sqrt{2}/135\sqrt{\pi}F_2^{9/2})[-45G_3F_2^3],\\ Q_4 &= (1/1152F_2^6)[144G_4F_2^4], \end{aligned} \tag{C25}$$

$$F_\nu = \left(\frac{d}{dx}\right)^\nu sx^2, \tag{C26}$$

$$G_\nu = \left(\frac{d}{dx}\right)^\nu \frac{1}{[\epsilon^2 + (ix^2+1)^2]\sqrt{x^2-2i}}. \tag{C27}$$

Carrying out the derivatives, we have, including up to Q_4

$$J + J^* = -\frac{2}{1+\epsilon^2}\sqrt{\frac{2\pi}{s}}\left\{\cos\left(s-\frac{\pi}{4}\right)\left[\frac{1}{2}+\frac{R_4(\beta,\kappa)}{(8s)^2}\right] + \sin\left(s-\frac{\pi}{4}\right)\left[\frac{R_2(\beta,\kappa)}{(8s)}\right]\right\} + O(s^{-7/2}), \tag{C28}$$

where

$$R_2(\beta,\kappa) = [9+\epsilon^2(\beta,\kappa)]/2[1+\epsilon^2(\beta,\kappa)], \tag{C29}$$

$$R_4(\beta,\kappa) = -\frac{9}{4} + \frac{12}{1+\epsilon^2(\beta,\kappa)} - \frac{96}{(1+\epsilon^2(\beta,\kappa))^2}. \tag{C30}$$

With Eq. (C28) we now have an asymptotic expansion for $I(\epsilon,s)$, which leads to the following expressions for $\Lambda_0(\beta w,s)$ and $\Lambda_1(\beta w,s)$:

$$\Lambda_0(\beta w,s) \approx \frac{J_0(s)}{2} + \sigma_1(\beta,\kappa)\frac{e^{-\epsilon(\beta,\kappa)s}}{\sqrt{1+\epsilon^2(\beta,\kappa)}} + \sigma_1(\beta,\kappa)\sqrt{\frac{2}{\pi s}}\frac{\epsilon(\beta,\kappa)}{1+\epsilon^2(\beta,\kappa)}\left\{\cos\left(s-\frac{\pi}{4}\right)\left[1+\frac{2R_4(\beta,\kappa)}{(8s)^2}\right] + \sin\left(s-\frac{\pi}{4}\right)\left[\frac{2R_2(\beta,\kappa)}{8s}\right]\right\} + \frac{\sigma_1(\beta,\kappa)}{\sqrt{1+\epsilon^2(\beta,\kappa)}}O(s^{-7/2}), \tag{C31}$$

$$\Lambda_1(\beta w,s) \approx \frac{\epsilon(\beta,\kappa)\sigma_2(\beta,\kappa)}{\sqrt{1+\epsilon^2(\beta,\kappa)}}\left\{e^{-\epsilon(\beta,\kappa)s} - \frac{1}{\sqrt{1+\epsilon^2(\beta,\kappa)}}\sqrt{\frac{2}{\pi s}}\left[\cos\left(s-\frac{\pi}{4}\right)\left(\frac{2[R_2(\beta,\kappa)-2]}{8s} - 40\frac{R_4(\beta,\kappa)}{(8s)^3}\right) - \sin\left(s-\frac{\pi}{4}\right)\left(1+\frac{2[R_4(\beta,\kappa)+12R_2(\beta,\kappa)]}{(8s)^2}\right)\right]\right\} + \frac{\epsilon(\beta,\kappa)\sigma_2(\beta,\kappa)}{1+\epsilon^2(\beta,\kappa)}O(s^{-9/2}). \tag{C32}$$

[1] A. R. Bishop, J. A. Krumhansl, and S. E. Trullinger, Physica D **1**, 1 (1980).
[2] M. B. Fogel, S. E. Trullinger, A. R. Bishop, and J. A. Krumhansl, Phys. Rev. Lett. **36**, 1411 (1976); Phys. Rev. B **15**, 1578 (1977).
[3] S. E. Trullinger, M. D. Miller, R. A. Guyer, A. R. Bishop, F. Palmer, and J. A. Krumhansl, Phys. Rev. Lett. **40**, 206, 1603(E) (1978).
[4] M. J. Rice, A. R. Bishop, J. A. Krumhansl, and S. E. Trullinger, Phys. Rev. Lett. **36**, 432 (1976).
[5] F. Nabarro, *Theory of Crystal Dislocations* (Oxford U. P., Oxford, 1967).
[6] J. C. Fernandez, J. M. Gambaudo, S. Gauthier, and G. Reinisch, Phys. Rev. Lett. **46**, 753 (1981); G. Reinisch and J. C. Fernandez, Phys. Rev. B **24**, 835 (1981); **25**, 7352 (1982); J. J. P. Leon, G. Reinisch, and J. C. Fernandez, *ibid.* **27**, 5817 (1983).
[7] P. C. Dash, Phys. Rev. Lett. **51**, 2155 (1983).
[8] O. H. Olsen and M. R. Samuelson, Phys. Rev. Lett. **48**, 1569 (1982); Phys. Rev. B **28**, 210 (1983).
[9] D. J. Kaup, Phys. Rev. B **29**, 1072 (1984).
[10] S. E. Trullinger, Solid State Commun. **29**, 27 (1979).
[11] R. J. Flesch, S. E. Trullinger, and B. Horovitz, in *Nonlinearity in Condensed Matter*, edited by A. R. Bishop, D. K. Cambell, P. Kumar, and S. E. Trullinger, Springer Ser. Solid-State Sci., Vol. 69 (Springer, Berlin, 1987).
[12] B. Horovitz, R. J. Flesch, and S. E. Trullinger, to be published.
[13] E. Tomboulis, Phys. Rev. D **12**, 1678 (1975); J. L. Gervais, A. Jevicki, and B. Sakita, *ibid.* **12**, 1038 (1975); Phys. Rep. C **23**, 281 (1976); J. L. Gervais and A. Jevicki, Nuclear Phys. B **110**, 93 (1976).
[14] G. N. Watson, *A Treatise on the Theory of Bessel Functions* (Cambridge U. P., Cambridge, 1980), 2nd ed., §16.5.
[15] J. F. Currie, J. A. Krumhansl, A. R. Bishop, and S. E. Trullinger, Phys. Rev. B **22**, 477 (1980).
[16] For a review, see R. Jackiw, Rev. Mod Phys. **49**, 681 (1977).
[17] N. H. Christ and T. D. Lee, Phys. Rev. D **12**, 1606 (1975).
[18] J. Goldstone and R. Jackiw, Phys. Rev. D **11**, 1486 (1975).
[19] R. Dashen, B. Hasslacher, and A. Neveu, Phys. Rev. D **11**, 3424 (1975).
[20] I. S. Gradshetyn and I. M. Rhyzhik, *Tables of Integrals, Series and Products* (Academic, New York, 1980), 4th ed., p. 472, formula #3.876.1.
[21] See Ref. 20, p. 470, formula #3.869.2.
[22] G. H. Hardy, Quart. Journal **XXXII**, 374 (1901), (The London Mathematical Society); *Collected Papers of G. H. Hardy, Volume 5* (Oxford U. P., Clarendon, 1972), pp. 514–517.
[23] See Ref. 20, p. 470, formula #3.869.1.
[24] This result is incorrectly stated in Ref. 14. Below Eq. (2) in §16.57, an error of $\frac{1}{2}$ occurs; it should read: $(1/\pi)\int_0^\infty (dt/t)\sin[at+b/t] = J_0(2\sqrt{ab})$.
[25] N. A. Shastri, Philos. Mag. (7) **XXV**, 930 (1938).
[26] J. Walker, *The Analytical Theory of Light* (Cambridge U. P., Cambridge, 1904).
[27] E. N. Dekanosidze, Vyčisl. Mat. i Vyčisl. Tehn. (Kharkov) **2**, 97 (1955) (in Russian).
[28] See Ref. 20, p. 967, formula #8.471.
[29] See Ref. 20, p. 666, formula #6.511.6.
[30] E. N. Dekanosidze, *Tables of Lommel's Functions of Two Variables* (Pergamon, London, 1960).
[31] L. S. Bark and P. I. Kuznetsov, *Tables of Lommel Functions* (Pergamon, London, 1965); E. W. Ng, Jet Propulsion Lab Report No. NASA-CR-91729; JPL-TR-32-1193.
[32] See Ref. 20, p. 973, formula #8.511.1.
[33] An error in Eq. 6 of §16.59 of Ref. 14 has been corrected by A. S. Yudina and P. I. Kuznetsov, USSR Comp. Math. Math. Phys. **11**, 258 (1971).
[34] R. H. D. Mayall, Proc. Cambridge Philos. Soc. **IX**, 259 (1898).
[35] R. B. Dingle, *Asymptotic Expansions: Their Derivation and Interpretation* (Academic, London, 1960), p. 119.

Current algebra for chiral gauge theories

M. V. Manías, M. C. von Reichenbach, F. A. Schaposnik, and M. Trobo[a)]
Departamento de Física, Facultad de Ciencias Exactas, Universidad Nacional de La Plata, 1900 La Plata, Argentina

(Received 16 September 1986; accepted for publication 4 March 1987)

Chiral gauge theories are studied with a special emphasis on the treatment of gauge degrees of freedom so as to obtain a gauge-invariant effective action from which current commutators can be evaluated. It is explicitly shown in a simple example that these commutators are those to be expected in a gauge-invariant theory.

I. INTRODUCTION

It is well known that theories where Weyl fermions interact with gauge fields can be inconsistent due to the existence of anomalies[1] which usually manifest through nonconservation of the fermion current. Quantization of such models recently received renewed interest after the proposal of Faddeev and Shatashvilli[2] concerning the construction of a sensible quantum theory by introducing a new *physical (chiral) field* with a Wess–Zumino action, this leading to a gauge-invariant theory. There has been also progress in the comprehension of how chiral theories can be consistent and unitary although non-gauge-invariant after the works of Jackiw and Rajaraman[3] and others[4] on two-dimensional chiral models.

More recently, it has been shown by Babelon, Schaposnik, and Viallet[5] that a proper treatment of the gauge degrees of freedom in chiral theories uncovers the presence of the Wess–Zumino action: *the group of gauge transformations acquires the status of a physical field and the anomaly is absorbed.* This phenomenon, which also happens with the Liouville action in the quantization of strings, as first shown by Polyakov,[6] causes the Wess–Zumino action to naturally emerge in the process of quantization without *ad hoc* introduction of additional fields. On the contrary, a group–valued field, which at the classical level was "lost" due to gauge invariance, naturally reappears after quantization. Similar ideas were independently developed by Harada and Tsutsui.[7]

Following this approach in the present work we give an analysis leading to the current algebra of chiral gauge theories. Although the current commutators we explicitly evaluate correspond to a simple (two-dimensional) model, the scheme we develop, based in the definition of an effective gauge-invariant action from which current–current correlation functions can be computed, is applicable to realistic four-dimensional theories. Some of the results in this paper have been already discussed in previous works.[3–7] We think, however, that it is worthwhile to present them in light of the framework established in Ref. 5 with the aim of clarifying some obscure points concerning quantization of chiral gauge theories, in particular, in connection with current algebra.

[a)] Postal address: Departamento de Física. UNLP, Casilla de Correos No. 67, 1900 La Plata, Argentina.

II. THE GENERAL TREATMENT

The developments in Ref. 5 start from the observation that, when quantizing a gauge theory in the path-integral approach, the generating functional should be considered as an integral over the whole space of connections rather than as an integral over the orbit space. This distinction causes no harm when Dirac fermions are present (no axial couplings) since, as first shown by Faddeev and Popov[8] for pure Yang–Mills theories, an integration over the gauge group factorizes. On the contrary, when Weyl fermions are present, the Faddeev–Popov procedure has to be revised and a trivial factorization is no more valid. Instead, a Fujikawa Jacobian[9] arises leading to an effective theory which contains the gauge group as a physical field and which is, remarkably, gauge invariant.

Let us start by briefly describing this facts by using the original Faddeev–Popov approach (a more formal presentation can be found in Ref. 5). The generating functional for a gauge theory with chiral (left handed for definiteness) fermions is

$$Z=\int \mathscr{D}A_\mu\, \mathscr{D}\bar\Psi\, \mathscr{D}\Psi\, e^{iS[A,\bar\Psi,\Psi]} \tag{1}$$

with

$$S=-\tfrac{1}{4}F_{\mu\nu}F^{\mu\nu}+\bar\Psi\, D(A)\Psi, \tag{2}$$

$$D(A)\equiv(i\partial\!\!\!/+A\!\!\!/)(1-\gamma_5)/2, \tag{3}$$

where the left-handed projector ensures that fermions have only one chirality. As stressed in Ref. 10, Z should be considered as an integral over the whole space of connections A_μ rather than as an integral over the orbit space. Usually, in the gauge fixing procedure, one passes from the former to the latter by factorizing a gauge-group integration; however, when Weyl fermions are present the standard procedure has to be reconsidered. Indeed, let us write

$$1=\Delta_{\rm FP}(A)\int \mathscr{D}g\, \delta[F[A^g]], \tag{4}$$

where the Faddeev–Popov determinant $\Delta_{\rm FP}(A)$ is obviously gauge invariant,

$$\Delta_{\rm FP}[A]=\Delta_{\rm FP}[A^g] \tag{5}$$

with

$$A_\mu^g=g^{-1}A_\mu\, g+(1/i)g^{-1}\,\partial_\mu\, g \tag{6}$$

and $\mathscr{D}g$ an appropriate measure over the gauge group. After insertion of identity (4) in (1) we get

 0022-2488/87/071632-04$02.50

$$Z = \int \mathscr{D}A_\mu \, \mathscr{D}\bar{\Psi} \, \mathscr{D}\Psi \, \mathscr{D}g \, e^{iS[A,\bar{\Psi},\Psi]} \delta[F[A^g]] \Delta_{FP}(A). \tag{7}$$

In trying to factorize the group integration one usually exploits gauge invariance of the action S, of Δ_{FP} and of $\mathscr{D}g$. Also

$$\mathscr{D}A_\mu = \mathscr{D}A^g_\mu. \tag{8}$$

Now, when writing a relation of the kind (8) for the Weyl fermion measure, one has to include a Fujikawa Jacobian[9] in the form

$$\mathscr{D}\bar{\Psi} \, \mathscr{D}\Psi = J(g,A) \mathscr{D}\bar{\Psi}^g \, \mathscr{D}\Psi^g, \tag{9}$$

since it is not possible to define in this case a gauge invariant measure (see below).[11,12] We have explicitly indicated the A_μ dependence of the Jacobian so as to stress its appearance due to the regularization procedure.

After relabeling variables one finally has

$$Z = \int \mathscr{D}A_\mu \, \delta[F[A]] \mathscr{D}\bar{\Psi} \, \mathscr{D}\Psi \, \mathscr{D}g \times J(g, A^{g^{-1}}) e^{iS[A,\bar{\Psi},\Psi]} \Delta_{FP}(A). \tag{10}$$

As explained above, the presence of the Jacobian prevents the factorization of the gauge group integration and so *the g field has acquired the status of a physical field.* Also, as it was shown in Ref. 5 the effective action defined by integration over fermions and the g field,

$$e^{iS_{\text{eff}}[A]} \equiv Z_{\Psi,\bar{\Psi},g} = \int \mathscr{D}\bar{\Psi} \, \mathscr{D}\Psi \, \mathscr{D}g \times \exp\left(i \int \bar{\Psi} D(A) \Psi \, dx\right) J(g,A) \tag{11}$$

is gauge invariant, i.e., it depends on the gauge equivalence class of A_μ and not on the particular choice of a representative on the orbit. This independence on the gauge condition choice then translates into BRS invariance.[5]

Let us now discuss with more detail the evaluation of the Jacobian defined by relation (9). The natural measure for fermions, $\mathscr{D}\bar{\Psi} \, \mathscr{D}\Psi \exp(i\int\bar{\Psi}D(A)\Psi \, dx)$, once integrated defines the fermion determinant. Hence one gets from (9):

$$J(g,A) = \det D(A)/\det D(A^g). \tag{12}$$

Now, as it is well known, the definition of Weyl fermion determinants is problematic since the corresponding Dirac operator maps negative chirality spinors into positive chirality ones and, consequently, it does not have a well-defined eigenvalue problem.[9,11,12] Precisely this problem is at the root of the anomalous behavior of chiral gauge theories. To handle this, one can define[12] an operator $\hat{D}(A)$ acting on Dirac fermions,

$$\hat{D}(A) = D(A) + i\partial\!\!\!/(1+\gamma_5)/2 = i\partial\!\!\!/ + A\!\!\!/(1-\gamma_5)/2, \tag{13}$$

which then leads to a well-defined eigenvalue problem. [As explained in Ref. 12 the doubling in the number of degrees of freedom implied by (13) affects only the overall normalization of the fermion integral since the positive chirality pieces do not couple to the gauge field.] One then defines

$$\det D(A) \equiv \det \hat{D}(A)|_{\text{Reg}}, \tag{14}$$

with the rhs appropriately regularized since the product of eigenvalues of the Dirac operator (13) grows without bound. The crucial point in this scheme is that the $\hat{D}(A)$'s eigenvalues are not gauge invariant, this being at the origin of the nontriviality of J. Concerning the regularization prescription, the guiding precept is usually gauge invariance. In the present case, however, once definition (14) has been adopted there is no more gauge invariance principle to invoke and, if one uses, for example, the heat-kernel regularization technique, more general regulators than $\hat{D}(A)$ can be used. As we shall see below, this fact can be exploited to render the resulting theory not only gauge invariant but also unitary.

It is important to note that the Jacobian (12) can be identified with a one cocycle[13] $w_1(A;g)$:

$$J(g,A) = e^{-2\pi i \omega(A;g)}, \tag{15}$$

which satisfies the condition

$$w_1(A^h; h^{-1}g) = w_1(g,A) - w_1(h,A), \tag{16}$$

this being at the origin of the explicit gauge invariance in the generating functional (11).

The Jacobian can be computed by integration of the axial anomaly. Indeed, consider an infinitesimal transformation

$$g = 1 + i\delta\theta, \tag{17}$$

whose corresponding Jacobian is[9]

$$J(\delta\theta) = \exp\left(i \operatorname{tr} \int \mathscr{A}(A) \delta\theta \, dx\right) \tag{18}$$

which $\mathscr{A}$ the consistent anomaly,

$$\langle D_\mu j^\mu \rangle_{\bar{\Psi},\Psi} = \mathscr{A}(A) = \frac{1}{i} \frac{\delta \log}{\delta\theta} J(\delta\theta, A). \tag{19}$$

The finite transformation Jacobian is gotten just by iteration of infinitesimal transformations. The simplest way to do it is to introduce a parameter t, $0 \leqslant t \leqslant 1$, to build up the finite transformation from the infinitesimal one. The answer is[14]

$$J(g,A) = \frac{\det D(A)}{\det D(A^g)} = \exp\left(i \operatorname{tr} \int \mathscr{A}(A^{g(t)}) \theta \, dt\right), \tag{20}$$

where

$$g(t) = e^{it\theta}. \tag{21}$$

Coming back to the effective action (11), note that it is analogous to that proposed by Faddeev and Shatashvili.[2] In the present formulation it is important to note that no chiral field has to be introduced *ad hoc* to recover gauge invariance. It is the gauge group which plays a physical rôle solving the anomaly problem. Of course $g(x)$ became physical after quantization since it trivially decouples the classical equations of motion.

From the effective action (11) one can compute v.e.v.'s of product of currents just by differentiation. For example the current–current correlation function is given by

$$G_{\mu\nu}(x,y) = \langle j_\mu(x), j_\nu(y) \rangle = \frac{\delta^2 S_{\text{eff}}[A]}{\delta A^\mu(x) \delta A^\nu(y)} \tag{22}$$

with $j_\mu = e\bar{\Psi}\gamma_\mu\Psi$.

Current commutators can be evaluated from (22) by means of the Bjorken–Johnson–Low method.[15] We shall now show, by studying a simple example, how these commutators are those expected in a gauge invariant theory.

III. A TWO-DIMENSIONAL EXAMPLE

We shall consider the chiral Schwinger model[3] (two-dimensional quantum electrodynamics with Weyl fermions). In order to obtain S_{eff} we have to compute the Jacobian from (12) or (20). Since the model is two dimensional, each determinant in J (and not only its ratio) can be computed exactly. As stated above, there is a regularization freedom associated to the noncovariance of the operator $\hat{D}(A)$ used to define the Weyl-fermion determinant. Following Ref. 14, the fermion determinant can be written as

$$\det D(A) \equiv \det \hat{D}(A) = \mathcal{N} \exp\left(-i \int d^2x \int_0^1 dt \, \mathrm{tr}\, \gamma_5 (\phi - \eta) \Big|_{\text{Reg}} \right) \tag{23}$$

with

$$A_\mu = -(1/e)(\epsilon_{\mu\nu} \partial^\nu \phi - \partial_\mu \eta), \tag{24}$$

and Reg meaning some regularization prescription. In two space-time dimensions determinants can be evaluated exactly since a chiral change transforms the problem to a free one (see Refs. 14 or 16). In the usual Schwinger model (Dirac fermions) this fact naturally leads to the following regularization:

$$\det \not{D} = \lim_{M^2 \to \infty} \mathcal{N} \times \exp\left(-i \int d^2x \int_0^1 dt \, \mathrm{tr}\, e^{D^2 t/M^2} \gamma_5 (\phi - \eta) \right), \tag{25}$$

with

$$\hat{Dt} = e^{(\gamma_5 \phi - i\eta)t} (i\not{\partial} + \not{A}) e^{(\gamma_5 \phi + i\eta)t}. \tag{26}$$

This choice ensures gauge invariance (see Ref. 14 for a detailed discussion). When Weyl fermions are present, a more general $\hat{Dt}$ is acceptable since, as we stated before, $\hat{D}(A)$ is not gauge covariant. Following Refs. 4 we shall choose, instead of (26)

$$\hat{Dt} = e^{(i/2)(1+\gamma_5)(\eta - \phi)t} \times [\hat{D}(A) + (a/2)\not{A}(1+\gamma_5)] e^{-(i/2)(1+\gamma_5)(\eta-\phi)t}, \tag{27}$$

with a a parameter to be determined demanding unitarity and consistency.[3] One then finds for the determinant (23),

$$\det D(A) = \exp\left(\frac{i}{8\pi} \int d^2x [(1+a)\phi\Box\phi + (1-a)\eta\Box\eta - 2\eta\Box\phi] \right). \tag{28}$$

Now, the effective action (11) can be written in the form

$$e^{iS_{\text{eff}}[A]} = \int \mathcal{D}g \, J(A_\mu^{g^{-1}}, g) \det D(A) = \int \mathcal{D}g \det D(A^{g^{-1}}) \tag{29}$$

with

$$g = e^{i\theta}.$$

Using (28) with $\eta \to \eta - \theta$ and explicitly performing the Gaussian θ integration we finally get

$$S_{\text{eff}}(A) = \frac{-i}{8\pi} \frac{a^2}{1-a} \int \phi\Box\phi \, d^2x, \tag{30}$$

which coincides with the normal Schwinger model result for $a = 2$. The important point is that (30) is gauge invariant for arbitrary a. As we shall see, the current commutators are those expected in a gauge-invariant theory provided $a > 1$ so that the Schwinger term has a coefficient with the correct sign ensuring positivity.[17] This condition for a can be also derived by noting that the effective action (30) plus the $F_{\mu\nu}^2$ term corresponds to a "photon" with a mass $-(1/4\pi)(e^2a^2/1-a)$ and hence for $a < 1$ the model would produce tachyons.[3] One can explicitly construct the current commutators from (30) and (22). One gets for $G_{\mu\nu}(x,y)$:

$$G_{\mu\nu}(x,y) = -(ie^2/4\pi)[a^2/(1-a)] \times (\delta_{\mu\nu} \partial_x^\alpha \Box_{xy}^{-1} \partial_{y\alpha} + \partial_{\mu x} \partial_{\nu y} \Box_{xy}^{-1}) \tag{31}$$

with

$$\Box_{xy}^{-1} = (1/2\pi)\log|x-y|. \tag{32}$$

Following the Bjorken–Johnson–Low method[15] one writes

$$\langle [j_\mu(x), j_\nu(y)]_{et} \rangle = \lim_{\epsilon \to 0} [G_{\mu\nu}(x_1,x_0;y_1,x_0+\epsilon) - G_{\mu\nu}(x_1,x_0;y_1,x_0-\epsilon)], \tag{33}$$

finally getting

$$[j_0, j_1]_{et} = (-ie^2/4\pi)[a^2/(a-1)]\delta'(x-y), \tag{34}$$

$$[j_0, j_0]_{et} = 0, \tag{35}$$

which are the relations expected in a gauge-invariant theory. Note that with our conventions the Schwinger term coefficient has to have a negative sign[17] this forcing $a > 1$.

To conclude, both the general analysis leading to the effective action (11) and the results for the current commutators in the simple model indicate that gauge invariance is maintained at the quantum level in chiral models if the gauge degrees of freedom are properly treated. No extra matter fields (as can be envisaged following Ref. 18) or chiral fields (as advocated in Ref. 2) have to be *ad hoc* introduced. It is the gauge group which becomes physical, solving the problem of gauge anomalies. It is interesting to note that, although gauge invariance is maintained, only using the regularization freedom (at least in the two-dimensional example) ensures a consistent unitary theory. We shall return to this point thoroughly in a future work.

ACKNOWLEDGMENTS

M.C.v.R. and F.A.S. are financially supported by CIC, Buenos Aires, Argentina.

M. V. Manias and M. Trobo are financially supported by CONICET Argentina.

[1]See, for example, R. Jackiw, in *Current Algebra and Anomalies*, edited by S. Treiman, R. Jackiw, B. Zumino, and E. Witten (World Scientific, Singapore, 1985), and references therein.

[2]L. D. Faddeev and S. L. Shatashvili, Phys. Lett. B **167**, 225 (1986).
[3]R. Jackiw and R. Rajamaran, Phys. Rev. Lett. **54**, 1219 (1985).
[4]R. Rajamaran, Phys. Lett. B **154**, 305 (1985); **162**, 148 (1985); J. G. Halliday, E. Rabinovici, A. Schwimmer, and M. Chonowitz, Nucl. Phys. B **268**, 413 (1986); B. Chanowitz, Phys. Lett. B **171**, 280 (1986); R. Banerjee, Phys. Rev. Lett. **56**, 1889 (1986); K. Harada, H. Kubota, and I. Tsutsui, Phys. Lett. B **173**, 77 (1986); J. Webb, Z. Phys. C **31**, 301 (1986); F. A. Schaposnik, LPTHE preprint 1986, unpublished; F. A. Schaposnik and J. V. Webb, Manchester Preprint 86/9.
[5]O. Babelon, F. A. Schaposnik, and C. M. Viallet, Phys. Lett. B **177**, 385 (1986).
[6]A. M. Polyakov, Phys. Lett. B **103**, 207 (1981).
[7]K. Harada and I. Tsutsui, Phys. Lett. B **183**, 311 (1987).
[8]L. D. Faddeev and V. N. Popov, Phys. Lett. B **25**, 29 (1967).
[9]K. Fujikawa, Phys. Rev. Lett. **42**, 1195 (1979); Phys. Rev. D **21**, 2848 (1980); **22**, 1499 (1980) (E).
[10]O. Babelon and C. M. Viallet, Phys. Lett. B **85**, 246 (1979).
[11]M. P. Balachandran, G. Marmo, V. P. Nair, and C. G. Trahern, Phys. Rev. D **25**, 2718 (1982); M. B. Einhorn and D. R. T. Jones, *ibid.* **29**, 331 (1984); S. K. Hu, B. L. Young, and D. W. McKay, *ibid.* **30**, 836 (1984); A. Andrianov and L. Bonora, Nucl. Phys. B **233**, 232 (1984); R. E. Gamboa Saraví, F. A. Schaposnik, and J. E. Solomin, Phys. Rev. D **28**, 856 (1983).
[12]L. Alvarez Gaumé and P. Ginsparg, Nucl. Phys. B **243**, 449 (1984).
[13]R. Jackiw, in *E. S. Fradkin Festchrift* (Adam Hilger, Bristol, 1986).
[14]R. E. Gamboa Saraví, M. A. Muschietti, F. A. Schaposnik, and J. E. Solomin, Ann. Phys. (NY) **157**, 145 (1984).
[15]J. D. Bjorken, Phys. Rev. **148**, 1467 (1966); K. Johnson and F. E. Low, Prog. Theor. Phys. (Kyoto), Suppl. 37 and 38, 74 (1966).
[16]R. Roskies and F. A. Schaposnik, Phys. Rev. D **23**, 558 (1981).
[17]J. Schwinger, Phys. Rev. Lett. **3**, 296 (1959).
[18]E. d'Hoker and E. Fahri, Nucl. Phys. B **248**, 59 (1984).

Poincaré covariant infraparticle sectors

Hans-Georg Stark
Physikalisches Institut der Universität Würzburg, Am Hubland, D-8700 Würzburg, West Germany

(Received 24 October 1984; accepted for publication 11 March 1987)

Footing on the infraparticle picture of quantum electrodynamics, a Poincaré covariant description of asymptotic sectors, containing n charged particles ($n \geqslant 1$) and the associated radiation field, is given. The analogy of this method to Wigner's description of elementary particles is discussed.

I. INTRODUCTION

The presence of massless particles in quantum field theories causes some complications due to the fact that physical states, of finite energy and momentum, may contain an infinite number of "soft" massless particles. A physically meaningful number operator cannot exist then for these particles, and besides the well-known difficulties encountered in perturbation theory this leads to the additional problem of choosing the correct state space of the theory. One might hope that some information may be obtained already from the study of the asymptotic fields, which are much easier to handle than the interacting fields of the full theory. Previous attempts to find the correct asymptotic state space for QED have led to two different directions, which also yield quite different qualitative features.

One of these proposals is the "infraparticle picture," as discussed by Fröhlich *et al.* in Refs. 1 and 2 (for a more general discussion with respect to full QED see also Ref. 3). It has the following main feature: the representation of the (asymptotic) photon field is coupled to the momenta of the charged particles, with inequivalent representations belonging to different momenta. As an "infraparticle," the charged particle is thus always surrounded by a dynamically coupled "soft photon cloud." In Refs. 1 and 2 also a definite choice for the momentum dependent photon representation is made: it is taken to be a generalized coherent representation, as suggested by a "correspondence principle" formulated in these references.

With this choice, however, the Poincaré covariance of the charged one-particle sector is broken; i.e., Lorentz boosts are not unitarily implementable in this sector. However, in view of the importance of relativistic covariance in proving analyticity properties of Wightman functions, which in turn are essential for some celebrated theorems of quantum field theory, one should perhaps look for suitable modifications of the above picture in order to reestablish relativistic symmetry in charged particle sectors. This is the aim of the present paper, which is a shortened version of Ref. 4.

For this purpose we shall use another class of so-called *symplectic* representations of the photon field (see Refs. 5 and 6), which are also useful in the alternative "infravacuum" description of asymptotic QED. (A short discussion of the latter may be found in Ref. 7). In these representations the photon field contains infinitely many soft photons of finite total energy, but in contrast to coherent representations not only space and time translations but also space rotations are unitarily implementable. We construct Poincaré covariant asymptotic sectors with a single charged particle by "boosting" a fixed symplectic representation of the photon field along the particle momentum. This construction is described in more detail in Sec. III. Crucial for the resulting Poincaré covariance is the Euclidean and time translation covariance of the symplectic representations, which excludes coherent representations from this procedure.[8] Nevertheless we arrive in this way at an "infraparticle" picture, since again the photon field is dynamically coupled to the particle momentum. This method may be generalized to asymptotic sectors containing an arbitrary number of charged particles and antiparticles with arbitrary spin.

We turn to a short summary of the contents of the sections. Section II contains preparatory material for the boosting procedure by recalling a general construction of covariant field representations as described in Refs. 9 and 10. In Sec. III we apply this construction to the free photon field. This procedure may be interpreted as the "boosting" mentioned above, and leads to a relativistically covariant sector with one charged scalar particle. Some properties of this sector are discussed. In Sec. IV the generalization to a particle with arbitrary spin is made. An appealing feature of the resulting description of infraparticles is its close analogy to Wigner's description of "ordinary" elementary particles by induced representations $[m,s]$ of the Poincaré group. Section V contains the construction of covariant charged n-particle sectors ($n \geqslant 1$) with r identical (infra)particles and $n-r$ identical antiparticles. The nonuniqueness of the Poincaré group representation in infraparticle sectors (as implied by the reducibility of the photon field) is discussed in Sec. VI. Physical arguments for choosing the "correct" representation are given. Some material concerning unitary representations of symmetry groups in reducible covariant representations of C^*-algebras is collected in the Appendix.

Two additional remarks might be appropriate.

(1) In this paper we are dealing exclusively with sectors containing charged particles. The existence of a vacuum state is therefore not touched by our considerations.

(2) It is not yet known whether the representations described here indeed appear as—say LSZ-type—limits in full QED.

We close this introduction with some technical statements, mainly in order to fix the notation.

We shall assume that there exists a C^*-algebra $\mathfrak{A}$, associated with the free asymptotic electromagnetic field, on which the Poincaré group $P^{\uparrow}_{+}$ acts as a symmetry group. The automorphism corresponding to the group element

 0022-2488/87/071636-08$02.50

$g \in P^{\uparrow}_{+}$ is denoted by τ_g. The map $g \to \tau_g$ forms a representation of the Poincaré group. If in a representation π of $\mathfrak{A}$, acting on a Hilbert space $\mathscr{H}$, these automorphisms are implemented by a unitary group representation [i.e., if $\pi(\tau_g(A)) = U(g)\pi(A)U(g)^{-1}$ for all $A \in \mathfrak{A}$], we call it a covariant representation of $\mathfrak{A}$ and denote it by $(\pi, U, \mathscr{H})$.

Using standard results of C^*-algebra representation theory (see Refs. 11 and 12), it is shown in Ref. 1 that any representation π of $\mathfrak{A}$ in the charged asymptotic one-particle sector may be decomposed into a direct integral over the particle momenta:

$$\pi = \int_{\oplus \mathbb{R}^3} \psi_{\mathbf{p}} \frac{d^3p}{\sqrt{m^2 + \mathbf{p}^2}}. \tag{1.1}$$

If the "component representations" $\psi_{\mathbf{p}}$ are mutually inequivalent for different momenta, we have the above-mentioned coupling of the photon field to the particle momentum characteristic for the infraparticle situation. Instead of using coherent representations for $\psi_{\mathbf{p}}$ as in Refs. 1 and 2, we shall use the "symplectic" representations $(\psi_{\mathbf{p}}, \mathscr{H}_{\mathbf{p}})$ described in Sec. III. These representations have been introduced in Ref. 5, where the Coulomb gauge was used; in Ref. 6 they were translated into the Gupta–Bleuler gauge. Since here we require a positive Hilbert space metric, our represenation space $\mathscr{H}_{\mathbf{p}}$ has to be identified either with the physical state space of an indefinite metric version of QED (Ref. 6), or with the state space of the Coulomb gauge.[5]

II. REDUCIBLE *G*-COVARIANT REPRESENTATIONS

Throughout this section $\mathfrak{A}$ denotes an arbitrary C^*-algebra. We shall recall a method of obtaining G-covariant representations $(\pi, U, \mathscr{H})$ from K-covariant representations $(\psi, W, \mathscr{K})$, where K is a suitable subgroup of G. (For more details see Ref. 10).

Let the topological group G act as a symmetry group of a C^*-algebra $\mathfrak{A}$ and assume the following.

(1) There exist a subgroup K and a subset T of G, such that every group element may be decomposed uniquely in the form

$$g = k \cdot t, \quad k \in K, \quad t \in T, \tag{2.1}$$

with both k and t depending continuously on g. Then for any fixed g the relation

$$t'' g = k' t' \tag{2.2}$$

defines continuous maps $\alpha_g : T \to T$ and $\beta_g : T \to K$ by

$$\beta_g(t'') = k', \quad \alpha_g(t'') = t'. \tag{2.3}$$

The uniqueness of the decomposition (2.1) also implies

$$\beta_{g_1 g_2}(t) = \beta_{g_1}(t) \beta_{g_2}(\alpha_{g_1}(t)), \tag{2.4}$$

$$\alpha_{g_1 g_2}(t) = \alpha_{g_2}(\alpha_{g_1}(t)). \tag{2.5}$$

(2) There exists a measure $d\nu$ on T which is invariant under α_g, i.e.,

$$\int_T f(t) d\nu(t) = \int_T f(\alpha_g(t)) d\nu(t) \tag{2.6}$$

for arbitrary integrable functions f on T.

Starting from a K-covariant representation $(\psi, W, \mathscr{K})$ (with continuous K-representation W) we obtain a G-covariant representation $(\pi, U, \mathscr{H})$ (with continuous G-representation U) as follows:

$$\begin{aligned} &\mathscr{H} = L^2(\mathscr{K}; T, d\nu), \\ &(\pi(A)\phi)(t) = \psi(\tau_t(A))\phi(t), \\ &(U(g)\phi)(t) = W_{\beta_g(t)} \phi(\alpha_g(t)). \end{aligned} \tag{2.7}$$

The space $\mathscr{H}$ consists of all $\mathscr{K}$-valued ν-measurable and ν-square integrable functions ϕ on T. In terms of direct integrals (see, e.g., Ref. 11), the representation $(\pi, \mathscr{H})$ of $\mathfrak{A}$ may be written as

$$\pi = \int_{\oplus T} \psi_t \, d\nu(t), \tag{2.8}$$

acting on

$$\mathscr{H} = \int_{\oplus T} \mathscr{H}_t \, d\nu(t), \tag{2.9}$$

where

$$\psi_t(A) = \psi(\tau_t(A)) \tag{2.10}$$

and

$$\mathscr{H}_t \equiv \mathscr{K} \tag{2.11}$$

for (almost) all t.

Under rather general assumptions, which are fulfilled in the case of the Poincaré group, one may show the unitary equivalence of the "group theoretic part" $(U, \mathscr{H})$ of (2.7) to the G-representations $(U^W, \mathscr{H}^W)$ induced by the representation $(W, \mathscr{K})$ of K. (For details see Ref. 13, for the theory of induced representations see Refs. 14 and 15.) One can also prove the equivalence of $(\pi, U, \mathscr{H})$ as given by (2.7) to the G-covariant representation constructed in Ref. 9 (see Ref. 13).

III. THE SCALAR ONE-PARTICLE SECTOR

In order to fix the notation, we recall some properties of the Poincaré group $P^{\uparrow}_{+}$ and the Euclidean group with time translations, which is a subgroup of $P^{\uparrow}_{+}$ ($\otimes$ denotes the semidirect product, $L^{\uparrow}_{+}$ the proper orthochronous Lorentz group, and $\mathbb{R}^4$ the real four-translations).

We have

$$P^{\uparrow}_{+} = L^{\uparrow}_{+} \otimes \mathbb{R}^4 = \{(\Lambda, d) \mid \Lambda \in L^{\uparrow}_{+}, d \in \mathbb{R}^4\}. \tag{3.1}$$

The group multiplication law in $P^{\uparrow}_{+}$ is

$$(\Lambda_1, d_1)(\Lambda_2, d_2) = (\Lambda_1 \Lambda_2, d_1 + \Lambda_1 d_2). \tag{3.2}$$

The Euclidean group together with time translations is the semidirect product of the rotations with $\mathbb{R}^4$ and will be denoted by K from now on:

$$K = O(3)^{+} \otimes \mathbb{R}^4 = \{(R, d) \mid R \in O(3)^{+}, d \in \mathbb{R}^4\}. \tag{3.3}$$

We shall now show how the above construction leads from a given (irreducible) K-covariant representation $(\psi, W, \mathscr{K})$ of the photon algebra $\mathfrak{A}$ to a (reducible) $P^{\uparrow}_{+}$-covariant representation $(\pi, U, \mathscr{H})$. In particular, we shall see that the resulting field representation is of the form (1.1), and hence is adapted to the case of a spinless infraparticle. We emphasize again that, as stated in the Introduction, an essential ingredient is the assumed K-covariance of $(\psi, W, \mathscr{K})$, which excludes coherent representations of the photon field[8] and favors symplectic representations.[5]

We first determine the subset $T \subset P^{\uparrow}_{+}$ "complementary" to K. Here the mass m of the infraparticle enters in a natural way. Its mass shell

$$\mathscr{M}^m = \{ p^\mu = (p^0,\mathbf{p}) | p^0 > 0, \quad p^2 = (p^0)^2 - \mathbf{p}^2 = m^2 \} \tag{3.4}$$

may be converted into a standard transitive G-space (G being the Poincaré group) by defining the action of $g = (\Lambda,d)$ on $\mathscr{M}^m$ by

$$p \in \mathscr{M}^m \to p_g = \Lambda^{-1} p \in \mathscr{M}^m. \tag{3.5}$$

In particular,

$$p_{g_1 g_2} = (p_{g_1})_{g_2}. \tag{3.6}$$

Introducing the reference momentum

$$p^R = (m,\mathbf{0}), \tag{3.7}$$

K may be characterized alternatively as the little group of p^R:

$$K = \{ g \in P^{\uparrow}_{+} | p^R_g = p^R \}. \tag{3.8}$$

Exploiting the transitivity of $\mathscr{M}^m$, boosts $b(p) \in L^{\uparrow}_{+}$ may be defined by the equation

$$p^R_{b(p)} = b(p)^{-1} p^R \overset{!}{=} p. \tag{3.9}$$

[For notational convenience, the symbol $b(p)$ will also be used for the element $(b(p),0)$ of $P^{\uparrow}_{+}$.]

It is possible to fix the boosts $b(p)$ uniquely, such that the map $p \to b(p)$ is (piecewise) analytic. (For several choices of boosts see, e.g., Ref. 16.) For definiteness, we take $b(p)$ to be pure Lorentz transformations. Now T may be identified with the set of all boosts:

$$T = \{ b(p) | p \in \mathscr{M}^m \}. \tag{3.10}$$

To justify (3.10), one has to verify the existence of the unique decomposition (2.1). Writing

$$g = k b(p^R_g) \tag{3.11}$$

we obtain

$$p^R_k = p^R_{g b(p^R_g)^{-1}} = (p^R_g)_{b(p^R_g)^{-1}} = p^R,$$

where we have used (3.6) and (3.9); hence $k \in K$. Assume $k' b(p') = k b(p)$. Then $b(p') = k'' b(p)$, where $k'' \in K$, and therefore $p^R_{b(p')} = p' = (p^R_{k''})_{b(p)} = p^R_{b(p)} = p$, i.e., $b(p') = b(p)$ and thus $k' = k$. Thus the decomposition (3.11) is unique. In order to determine the maps α_g and β_g, consider

$$b(p) g = b(p) g b(p_g)^{-1} b(p_g).$$

Using (3.6) and (3.9) again one has

$$p^R_{b(p) g b(p_g)^{-1}} = (p_g)_{b(b_g)^{-1}} = p^R,$$

i.e., $b(p) g b(p_g)^{-1} \in K$. With $g = (\Lambda,d)$ one thus obtains, using (3.2) and (3.5), the maps β_g and α_g defined in (2.3):

$$\begin{aligned} &\beta_g(b(p)) = (r(g,p), b(p) d), \\ &\alpha_g(b(p)) = b(\Lambda^{-1} p), \end{aligned} \tag{3.12}$$

where

$$r(g,p) = b(p) \Lambda b(\Lambda^{-1} p)^{-1} \tag{3.13}$$

is the familiar Wigner rotation. Identifying $T = \mathscr{M}^m = \mathbb{R}^3$ by the one-to-one correspondences $b(p) \leftrightarrow p = ((\mathbf{p}^2 + m^2)^{1/2}, \mathbf{p}) \leftrightarrow \mathbf{p}$, as done in the rest of this paper, an invariant measure on T is thus given by

$$d\nu(\mathbf{p}) = d^3 p / \sqrt{m^2 + \mathbf{p}^2}. \tag{3.14}$$

Hence (2.7) reads for $G = P^{\uparrow}_{+}$:

$$\begin{aligned} &\mathscr{H}^{(0,1)} = L^2(\mathscr{K};\mathbb{R}^3, d\nu(\mathbf{p})) = \int_{\oplus \mathbf{R}^3} \mathscr{K} \, d\nu(p), \\ &(\pi^{(0,1)}(A)\phi)(\mathbf{p}) = \psi(\tau_{b(p)}(A))\phi(\mathbf{p}), \\ &(U^{(0,1)}(\Lambda,d)\phi)(\mathbf{p}) = W_{(r(g,p), b(p)d)} \phi(\Lambda^{-1} p). \end{aligned} \tag{3.15}$$

(Anticipating the general notation from Sec. V, the first superscript on $\mathscr{H}^{(0,1)}$, etc. denotes the spin and the second the number of particles.) In particular, $(\pi^{(0,1)}, \mathscr{H}^{(0,1)})$ may be rewritten as

$$\pi^{(0,1)} = \int_{\oplus \mathbf{R}^3} \psi_{\mathbf{p}} \, d\nu(\mathbf{p}), \quad \psi_{\mathbf{p}}(A) = \psi(\tau_{b(p)}(A)). \tag{3.16}$$

Comparing with (1.1) we see that (3.15) corresponds to the infraparticle picture, but in contrast to Ref. 1 we now have a $P^{\uparrow}_{+}$-covariant one-particle sector $(\pi^{(0,1)}, U^{(0,1)}, \mathscr{H}^{(0,1)})$. Defining the momentum of the charged particle by

$$P^\mu_c = \int_{\oplus \mathbf{R}^3} p^\mu \, d\nu(\mathbf{p}) \quad \text{with } p^\mu = (\sqrt{m^2 + \mathbf{p}^2}, \mathbf{p}),$$

i.e.,

$$(P^\mu_c \phi)(\mathbf{p}) = p^\mu \phi(\mathbf{p}), \tag{3.17}$$

one immediately verifies its covariance

$$U(\Lambda,d) e^{iP_c d'} U(\Lambda,d)^{-1} = e^{i\Lambda^{-1} P_c d'}. \tag{3.18}$$

Denoting by P the total momentum—i.e., the generator of translations $U(1,d)$—(3.18) also proves the commutativity of P and P_c:

$$U(1,d) e^{iP_c d'} = e^{iPd} e^{iP_c d'} = e^{iP_c d'} U(1,d) = e^{iP_c d'} e^{iPd}.$$

Hence one may introduce the field momentum $P_F = P - P_c$ and the corresponding unitary group

$$e^{iP_F d} = e^{iPd} e^{-iP_c d}.$$

Then, using that the "diagonal"[17] operator $e^{iP_c d}$ commutes with all $\pi^{(0,1)}(A)$, one obtains

$$e^{iPd} \pi^{(0,1)}(A) e^{-iPd} = e^{iP_F d} \pi^{(0,1)}(A) e^{-iP_F d}.$$

Collecting these results we get the following proposition.

Proposition 3.1: The total momentum P, the particle momentum P_c and the field momentum P_F are mutually commuting operators obeying

$$P^\mu = P^\mu_F + P^\mu_c. \tag{3.19}$$

The field dynamics is determined by P_F, and the particle momentum P_c is affiliated to the commutant $\pi^{(0,1)}(\mathfrak{A})'$.

Lemma 3.2: Let

$$\pi = \int_{\oplus T} \psi_t \, d\nu(t)$$

be the direct integral representation (2.8) with $\psi_t \not\simeq \psi \equiv \psi_e$ for all $t \in T$, $t \neq e$. Then we have $\psi_t \not\simeq \psi_{t'}$ for $t \neq t'$, i.e., the component representations are pairwise inequivalent.

Proof: Suppose there exist t and $t' \neq t$ in T and a unitary W, such that $\psi_t(A) = \psi(\tau_t(A)) = W\psi_{t'}(A) W^{-1} = W\psi(\tau_{t'}(A)) W^{-1} \ \forall A \in \mathfrak{A}$. This is equivalent to $\psi(\tau_t(\tau_{t'^{-1}}(A))) = \psi(\tau_{tt'^{-1}}(A)) = W\psi(A) W^{-1} \ \forall A \in \mathfrak{A}$, i.e.,

$\tau_{tt'^{-1}}$ is implemented in ψ. Let $tt'^{-1} = k''t''$ be the unique decomposition of tt'^{-1}. Then $t \neq t'$ implies $t'' \neq e$, for otherwise we would have $t = k''t'$, i.e., a nonunique decomposition of t. But then, since ψ is K covariant and tt'^{-1} is implemented, $t'' = k''^{-1}tt'^{-1} \in T$ is implementable and $\neq e$, which is a contradiction. ∎

Choosing for $(\psi, W, \mathscr{K})$ irreducible symplectic representations of the photon field, it is known that pure Lorentz transformations are not unitarily implementable in such representations.[5] Hence in (3.16) $\psi_{\mathbf{p}} \not\simeq \psi$ for all $\mathbf{p} \neq \mathbf{0}$, and therefore $\psi_{\mathbf{p}} \not\simeq \psi_{\mathbf{p}'}$, if $\mathbf{p} \neq \mathbf{p}'$, by the above lemma. Under these conditions one may show that the infraparticle is not localizable.[18]

IV. GENERALIZATION TO ARBITRARY SPIN

In order to motivate the following construction, we recall Wigner's concept of elementary particles of mass m and spin s as irreducible representations $[m,s]$ of $P_+^\uparrow$ (see Refs. 19 and 16). In the formulas below, $r(g,p)$ and $d\nu(\mathbf{p})$ denote the Wigner rotation and the invariant measure, as defined in (3.13) and (3.14), respectively.

In the case of a spinless elementary particle $[m,0]$ one has

$$\begin{aligned} &\tilde{\mathscr{H}} = L^2(\mathbb{R}^3, d\nu(\mathbf{p})), \\ &(\tilde{U}(g)\phi)(\mathbf{p}) = e^{ipd}\phi(\Lambda^{-1}p). \end{aligned} \tag{4.1}$$

The amplitudes $\phi(0)$ for the particle at rest transform under rotations according to the trivial representation $R \to 1$ of $O(3)^+$, which means that the particle is spinless.

In the case of spin s, $(\tilde{U}^{(s)}, \tilde{\mathscr{H}}^{(s)})$ is given by

$$\begin{aligned} &\tilde{\mathscr{H}}^{(s)} = L^2(\mathbb{C}^{(2s+1)}; \mathbb{R}^3, d\nu(\mathbf{p})), \\ &(\tilde{U}^{(s)}(g)\phi)_m(\mathbf{p}) = e^{ipd} D^{(s)}_{mm'}(r(g,p))\phi_{m'}(\Lambda^{-1}p), \end{aligned} \tag{4.2}$$

where a summation convention is used. Again, the transformation properties of the amplitudes for the particle at rest under rotations define the spin: Since $r(R,p^R) \equiv R$ for rotations R and p^R as defined in (3.7), one verifies immediately that the rest amplitudes $\phi_m(0)$ transform according to $R \to D^{(s)}(R)$.

Now inspection of (3.15) shows that the amplitudes $\phi(0)$ $(\in \mathscr{K})$ of the spinless infraparticle at rest transform according to the representation

$$(R,0) \to W_{(R,0)} \tag{4.3}$$

of $O(3)^+$. This suggests the following generalization: The rest amplitudes for the infraparticle with spin s should transform under rotations according to

$$(R,0) \to W_{(R,0)} \otimes D^{(s)}(R). \tag{4.4}$$

This may be achieved by a construction analogous to the one described in the last section, with the representation $(\psi, W, \mathscr{K})$ replaced by $(\psi^{(s,1)}, W^{(s,1)}, \mathscr{K}^{(s,1)})$, where

$$\begin{aligned} &\mathscr{K}^{(s,1)} = \mathscr{K} \otimes \mathbb{C}^{(2s+1)}, \\ &\psi^{(s,1)}(A) = \psi(A) \otimes 1_{\mathbb{C}^{(2s+1)}}, \\ &W^{(s,1)}_{(R,d)} = W_{(R,d)} \otimes D^{(s)}(R). \end{aligned} \tag{4.5}$$

The representation properties and the K covariance of $(\psi^{(s,1)}, W^{(s,1)}, \mathscr{K}^{(s,1)})$ follow immediately from the corresponding properties of $(\psi, W, \mathscr{K})$. Applying (2.7) to $(\psi^{(s,1)}, W^{(s,1)}, \mathscr{K}^{(s,1)})$ now leads to the representation $(\pi^{(s,1)}, U^{(s,1)}, \mathscr{H}^{(s,1)})$, where

$$\begin{aligned} \mathscr{H}^{(s,1)} &= L^2(\mathscr{K} \otimes \mathbb{C}^{(2s+1)}; \mathbb{R}^3, d\nu(\mathbf{p})) \\ &= L^2\left(\bigoplus_{m=-s}^{+s} \mathscr{K}; \mathbb{R}^3, d\nu(\mathbf{p})\right), \\ (\pi^{(s,1)}(A)\phi)_m(\mathbf{p}) &= \psi(\tau_{b(p)}(A))\phi_m(\mathbf{p}), \\ (U^{(s,1)}(\Lambda,d)\phi)_m(\mathbf{p}) & \\ &= D^{(s)}_{mm'}(r(g,p))W_{(r(g,p),b(p)d)}\phi_{m'}(\Lambda^{-1}p). \end{aligned} \tag{4.6}$$

(The equality sign between Hilbert spaces means isomorphism.) Indeed, then, the rest amplitudes

$$\phi(0) \left(\in \bigoplus_{m=-s}^{+s} \mathscr{K}\right)$$

transform under rotations according to $(R,0) \to W_{(R,0)} \otimes D^{(s)}(R)$.

Comparing (4.1) with (4.2) and (3.15) with (4.6), we see that in both cases the transition from the spinless situation to the one with nonvanishing spin has been made by "adding Wigner's rotation in the representation $D^{(s)}$."

The particle momentum is defined as in the spinless case. Its $P_+^\uparrow$ covariance is immediately verified.

As noted for the spinless case at the end of Sec. II, the "group theoretic" part $(U^{(s,1)}, \mathscr{H}^{(s,1)})$ of (4.6) is again equivalent to the representation $(U^{W^{(s,1)}}, \mathscr{H}^{W^{(s,1)}})$ of $P_+^\uparrow$ induced by the representation $(W^{(s,1)}, \mathscr{K}^{(s,1)})$ of the Euclidean group with time translations K. Comparing this with the prescription for constructing the representations $[m,s]$ of an ordinary elementary particle (see, e.g., Ref. 16), the analogy between Wigner's concept of elementary particles and our description of infraparticles may be summarized in Table I.

V. COVARIANT n-PARTICLE SECTORS

In this section a construction is given which yields a covariant description of a system consisting of n charged particles of spin s and the associated radiation field, i.e., of n infraparticles of spin s. Starting point again is the given, irreducible, K-covariant representation $(\psi, W, \mathscr{K})$ of $\mathfrak{A}$. The method is formulated for the general case of r particles and n-r antiparticles, with $r = 0,1,...,n$.

The following requirements have to be met.

(1) The state space is a subspace, to be specified below, of

$$\begin{aligned} \mathscr{H}^{(s,n)} &= L^2(\mathbb{C}^{(2s+1)}; T, d\nu) \\ &\quad \times \otimes \cdots \otimes L^2(\mathbb{C}^{(2s+1)}; T, d\nu) \otimes \mathscr{K} \\ &= L^2((\mathbb{C}^{(2s+1)})^n \otimes \mathscr{K}; T^n, d\nu^n), \end{aligned} \tag{5.1}$$

TABLE I. Relativistic transformation laws (preliminary version).

	$P_+^\uparrow$ representation
Elementary particle	induced by the representation $(R,d) \to e^{ip^Rd} \otimes D^{(s)}(R)$ of K
Infraparticle	induced by the representation $(R,d) \to W_{(R,d)} \otimes D^{(s)}(R)$ of K

where

$$T^n = T\otimes\cdots\otimes T,\ (\mathbb{C}^{(2s+1)})^n = \mathbb{C}^{(2s+1)}\otimes\cdots\otimes\mathbb{C}^{(2s+1)}, \tag{5.2}$$
$$dv^n = dv\otimes\cdots\otimes dv.$$

(2) Assume that the indices $1,\ldots,r$ label the particle variables and $r+1,\ldots,n$ label the antiparticle variables. Then, according to the spin-statistics theorem, the state space is given by the subspace of $\mathscr{H}^{(s,n)}$ consisting of functions which are symmetric (for integer spin) or antisymmetric (for half-odd-integer spin) under separate permutations of the indices $(1,\ldots,r)$ and $(r+1,\ldots n)$. The field representation $\pi^{(s,n)}$ and the $P_+^\uparrow$-representation $U^{(s,n)}$ have to leave this subspace invariant.

(3) $(\pi^{(s,n)}, U^{(s,n)}, \mathscr{H}^{(s,n)})$ has to be a covariant representation of $\mathfrak{A}$.

Definition V.1: Let j be a continuous map from $T\times\cdots\times T$ to T with

$$(1)\ j(t_1\cdots t_n) = j(t_{P(1)}\cdots t_{P(n)}), \tag{5.3}$$

where P is an arbitrary permutation of the numbers $(1,\ldots,n)$;

$$(2)\ \alpha_g(j(t_1\cdots t_n)) = j(\alpha_g(t_1)\cdots\alpha_g(t_n)), \tag{5.4}$$

where α_g is defined in (3.12).

In the next step we modify $(\psi, W, \mathscr{K})$ in fashion analogous to (4.5): The K-covariant representation $(\psi^{(s,n)}, W^{(s,n)}, \mathscr{K}^{(s,n)})$ is defined as

$$\mathscr{K}^{(s,n)} = \mathscr{K}\otimes(\mathbb{C}^{(2s+1)})^n,$$
$$\psi^{(s,n)}(A) = \psi(A)\otimes 1_{(\mathbb{C}^{(2s+1)})^n}, \tag{5.5}$$
$$W^{(s,n)}_{(R,d)} = W_{(R,d)}\otimes(D^{(s)})^n(R)$$

[with $(D^{(s)})^n = D^{(s)}\otimes\cdots\otimes D^{(s)}$]. Then $(\pi^{(s,n)}, U^{(s,n)}, \mathscr{H}^{(s,n)})$ may be defined as

$$\mathscr{H}^{(s,n)} = L^2(\mathscr{K}^{(s,n)}; T^n, dv^n) = L^2(\mathscr{K}\otimes(\mathbb{C}^{(2s+1)})^n; T^n, dv^n),$$
$$(U^{(s,n)}(g)\phi)(t\cdots t_n) = W^{(s,n)}_{\beta_g(j(t_1\cdots t_n))}\phi(\alpha_g(t_1)\cdots\alpha_g(t_n)), \tag{5.6}$$
$$(\pi^{(s,n)}(A)\phi)(t_1\cdots t_n) = \psi^{(s,n)}(\tau_{j(t_1\cdots t_n)}(A))\phi(t_1\cdots t_n).$$

The above requirements (1) and (2) are fulfilled, as follows from the permutation invariance (5.3) of j, and the invariance of operators of the type $(D^{(s)})^n$ under permutation of the n identical factor $D^{(s)}$. It still remains to show that $\pi^{(s,n)}$ indeed is a representation of $\mathfrak{A}$, that $U^{(s,n)}$ is a unitary representation of $P_+^\uparrow$, and that τ_g is unitarily implemented by $U^{(s,n)}(g)$.

The representation property of $\pi^{(s,n)}$ follows trivially from the corresponding property of ψ. The unitarity of $U^{(s,n)}(g)$ follows from the unitarity of the operators $W_k^{(s,n)}$ $(k\in K)$ in $\mathscr{K}\otimes(\mathbb{C}^{(2s+1)})^n$ and the invariance of the measure dv^n under $(t_1\cdots t_n)\to(\alpha_g(t_1)\cdots\alpha_g(t_n))$, which is a consequence of the invariance (2.6) of the factors dv.

To check the representation property of $U^{(s,n)}$, calculate first

$$\beta_{g_1g_2}(j(t_1\cdots t_n)) = \beta_{g_1}(j(t_1\cdots t_n))\beta_{g_2}(\alpha_{g_1}(j(t_1\cdots t_n))) = \beta_{g_1}(j(t_1\cdots t_n))\beta_{g_2}(j(\alpha_{g_1}(t_1)\cdots\alpha_{g_1}(t_n))), \tag{5.7}$$

where in the first step we have used the formula (2.4) and in the second step (5.4). Hence

$$\begin{aligned}(U^{(s,n)}(g_1)U^{(s,n)}(g_2)\phi)(t_1\cdots t_n) &= W^{(s,n)}_{\beta_{g_1}(j(t_1\cdots t_n))}W^{(s,n)}_{\beta_{g_2}(j(\alpha_{g_1}(t_1)\cdots\alpha_{g_1}(t_n)))}\\ &\quad\times\phi(\alpha_{g_2}(\alpha_{g_1}(t_1))\cdots\alpha_{g_2}(\alpha_{g_1}(t_n)))\\ &= W^{(s,n)}_{\beta_{g_1g_2}(j(t_1\cdots t_n))}\phi(\alpha_{g_1g_2}(t_1)\cdots\alpha_{g_1g_2}(t_n))\\ &= (U^{(s,n)}(g_1g_2)\phi)(t_1\cdots t_n).\end{aligned}$$

In the second step formula (2.5), the representation property of $W^{(s,n)}$ and Eq. (5.7) have been used.

To verify covariance one needs the following relation, which is a consequence of (5.4), (2.2), and (2.3):

$$\beta_g(j(t_1\cdots t_n))j(\alpha_g(t_1)\cdots\alpha_g(t_n)) = \beta_g(j(t_1\cdots t_n))\alpha_g(j(t_1\cdots t_n)) = j(t_1\cdots t_n)g. \tag{5.8}$$

Thus

$$\begin{aligned}(U^{(s,n)}(g)\pi^{(s,n)}(A)\phi)(t_1\cdots t_n) &= W^{(s,n)}_{\beta_g(j(t_1\cdots t_n))}\psi^{(s,n)}(\tau_{j(\alpha_g(t_1)\cdots\alpha_g(t_n))}(A))\\ &\quad\times\phi(\alpha_g(t_1)\cdots\alpha_g(t_n))\\ &= \psi^{(s,n)}(\tau_{\beta_g(j(t_1\cdots t_n))j(\alpha_g(t_1)\cdots\alpha_g(t_n))}(A))\\ &\quad\times W^{(s,n)}_{\beta_g(j(t_1\cdots t_n))}\phi(\alpha_g(t_1)\cdots\alpha_g(t_n))\\ &= \psi^{(s,n)}(\tau_{j(t_1\cdots t_n)g}(A))(U^{(s,n)}(g)\phi)(t_1\cdots t_n)\\ &= (\pi^{(s,n)}(\tau_g(A))U^{(s,n)}(g)\phi)(t_1\cdots t_n).\end{aligned}$$

In the second and third step we used the K covariance of $(\psi^{(s,n)}, W^{(s,n)}, \mathscr{K}^{(s,n)})$ and Eq. (5.8), respectively.

One also easily verifies—identifying T with $\mathscr{M}^m$ as mentioned in Sec. III and using (5.6) and (3.12)—that the total particle momentum P_c defined by

$$(P_c^\mu\phi)(\mathbf{p}_1,\ldots,\mathbf{p}_n) = \sum_{i=1}^n p_i^\mu\phi(\mathbf{p}_1,\ldots,\mathbf{p}_n)\quad\left(p_i^\mu = (\sqrt{m^2+\mathbf{p}_i^2},\mathbf{p}_i)\right) \tag{5.9}$$

transforms covariantly under $U^{(s,n)}$:

$$U^{(s,n)}(\Lambda,d)e^{iP_cd'}U^{(s,n)}(\Lambda,d)^{-1} = e^{i\Lambda^{-1}P_cd'}. \tag{5.10}$$

Again, as in the one-particle sector, P_c is affiliated to the commutant $\pi^{(s,n)}(\mathfrak{A})'$.

Thus the construction of the covariant n-particle sector (5.6) is reduced to the problem of finding a map j with the properties of definition 5.1.

A possible choice of j is

$$j(p_1\cdots p_n) = \frac{m}{(\sum_{i,k=1}^n p_ip_k)^{1/2}}\sum_{i=1}^n p_i. \tag{5.11}$$

The map j is well defined, since p_i and p_k belong to the mass shell $\mathscr{M}^m$, and thus $p_ip_k>0$. Moreover, clearly, $j(p_1\cdots p_n)$ lies on the mass shell $\mathscr{M}^m$. The required permutation invariance (5.3) is obvious. With α_g defined in (3.12), and the invariance of the four-product under Lorentz transformations, (5.4) is easily checked. Therefore the construction (5.6) of the n-particle sector may be carried out with the map j given in (5.11).

The physical picture underlying this model might seem to be questionable, since the photon cloud is correlated only to the total particle momentum, and hence the dependence on the individual particle momenta is of a rather indirect nature only, in contrast to the case when coherent representations are used.[1] Nevertheless, the sector $(\pi^{(s,n)},U^{(s,n)},\mathscr{H}^{(s,n)})$ contains all the necessary degrees of freedom and is relativistically covariant. One might also hope that a more satisfactory physical interpretation may be achieved by choosing a map j which is more sensitive than (5.11) to the individual particle momenta.

We close this section by noticing that $\pi^{(s,n)}$ may be rewritten as a direct integral over (T^n,dv^n), but now with component representations which are not pairwise inequivalent. Rather, they are constant on the hyperplanes $j(p_1\cdots p_n) = \text{const}$, which just expresses the fact that the photon clouds are coupled to the total particle momentum only.

VI. COCYCLES AND SPECTRUM CONDITION

With the total particle momentum P_c defined in (5.9) and an arbitrary real number α, consider the unitary operator

$$V^{\alpha}_{(\Lambda,d)} = e^{i\alpha P_c d}, \tag{6.1}$$

which depends continuously on the group element $g=(\Lambda,d)$ and belongs to the commutant $\pi^{(s,n)}(\mathfrak{A})'$. Using (5.10) one easily proves the following theorem.

Theorem VI.1: For any $\alpha\in\mathbb{R}$,

$$U^{(s,n)}_{\alpha}(g) = V^{\alpha}_{g}U^{(s,n)}(g) \tag{6.2}$$

defines a unitary representation of $P^{\uparrow}_{+}$ implementing the $P^{\uparrow}_{+}$-automorphisms in $\pi^{(s,n)}$.

According to the Appendix, (6.1) defines a cocycle; i.e., the map

$$V^{\alpha}:\quad (\Lambda,d)\to V^{\alpha}_{(\Lambda,d)}$$

belongs to $Z^{\pi^{(s,n)}}_{c}(P^{\uparrow}_{+})$. (The notation is explained in the Appendix.)

The result of the following analysis is formulated as the following.

Theorem VI.2: Let the n-particle sector $(\pi^{(s,n)}, U^{(s,n)}, \mathscr{H}^{(s,n)})$ be given. Let $P^{(s,n)}$ denote the total momentum of $U^{(s,n)}$, and let P_c be given by (5.9). Then we have the following.

(1) P^{μ}_{c} and $P^{(s,n)\nu}$ form a set of commuting operators, and the field dynamics is determined by

$$P^{(s,n)}_{F} = P^{(s,n)} - P_c. \tag{6.3}$$

(2) The momentum of $U^{(s,n)}_{\alpha}$, as defined in (6.2), is

$$P^{(s,n)}_{\alpha} = \alpha P_c + P^{(s,n)}.$$

Assume, moreover, that $P^{(s,n)}$ satisfies the spectrum condition. Then we have (3).

(3) For all $\alpha\geqslant 0$, $P^{(s,n)}_{\alpha}$ also satisfies the spectrum condition. On the other hand, α may be chosen such that $P^{(s,n)}_{\alpha}$ does not satisfy the spectrum condition. Hence the (s,n) cohomology of $P^{\uparrow}_{+}$ (see the Appendix for the terminology) is nontrivial:

$$Z^{\pi^{(s,n)}}_{c}(P^{\uparrow}_{+}) \neq B^{\pi^{(s,n)}}(P^{\uparrow}_{+}).$$

Proof: Statement (1) follows from (5.10) and from

$$e^{iP_c d}\in\pi^{(s,n)}(\mathfrak{A})'$$

as for the one-particle sector in Sec. III. Statement (2) follows from

$$U^{(s,n)}_{\alpha}(1,d) = e^{iP^{(s,n)}_{\alpha}d} = V^{\alpha}_{(1,d)}\,U^{(s,n)}(1,d) = e^{i\alpha P_c d}e^{iP^{(s,n)}d}.$$

To prove statement (3), consider the Gårding domain D of P_c and $P^{(s,n)}$ (see the Appendix of Refs. 4). Since $P^{(s,n)}$ is assumed to satisfy the spectrum condition, we have

$$(\Phi,P^{(s,n)}a\Phi)\geqslant 0\quad\text{for all } a\in\overline{V}_{+}\text{ and all }\Phi\in D.$$

Then, for all $\alpha\geqslant 0$, also

$$(\Phi,P^{(s,n)}_{\alpha}a\Phi) = \alpha(\Phi,P_c a\Phi) + (\Phi,P^{(s,n)}a\Phi)\geqslant 0$$

since P_c, as a sum of particle momenta, satisfies the spectrum condition. Since $P^{(s,n)}_{\alpha}a$ is essentially self-adjoint on D (see the Appendix of Ref. 4), the above inequality is valid for all Φ in the domain of $P^{(s,n)}_{\alpha}a$. Hence $P^{(s,n)}_{\alpha}$ satisfies the spectrum condition if $\alpha\geqslant 0$.

To prove the last part of statement (3), take a fixed $\Phi\in D$ and a fixed $a\in\overline{V}_{+}$ [e.g., $a=(1,\mathbf{0})$], and choose α sufficiently negative, such that $(\Phi,P^{(s,n)}_{\alpha}a\Phi)<0$. Now, since the corresponding $P^{(s,n)}_{\alpha}$ does not satisfy the spectrum condition whereas $P^{(s,n)}$ does, the $P^{\uparrow}_{+}$-representations $U^{(s,n)}_{\alpha}$ and $U^{(s,n)}$ cannot be unitarily equivalent. Then from Theorem (A4) it follows that V^{α} cannot be a coboundary. ■

Now let P' denote the four-momentum of the K-representation W and $P'^{(s,n)}$ the momentum of $W^{(s,n)}$. We will show that the total momentum operator $P^{(s,n)}$ in an n-particle sector satisfies the spectrum condition if P' does.

Let $\phi(\mathbf{p}_1\cdots\mathbf{p}_n)$ be an element of $\mathscr{H}^{(s,n)}$. For pure translations $(1,d)$ we obtain

$$\begin{aligned}(U^{(s,n)}&(1,d)\phi)(\mathbf{p}_1\cdots\mathbf{p}_n)\\ &= (e^{iP^{(s,n)}d}\phi)(\mathbf{p}_1\cdots\mathbf{p}_n)\\ &= W^{(s,n)}_{\beta_g(j(p_1\cdots p_n))}\phi(\mathbf{p}_1\cdots\mathbf{p}_n)\\ &= e^{iP'^{(s,n)}b(j(p_1\cdots p_n))d}\phi(\mathbf{p}_1\cdots\mathbf{p}_n).\end{aligned}$$

Here we have used (3.12), (3.13), with $g=(1,d)$, and (5.6). Thus

$$P^{(s,n)} = \int_{\oplus T^n} b(j(p_1\cdots p_n))^{-1}P'^{(s,n)}\,dv^{(n)}(\mathbf{p}_1\cdots\mathbf{p}_n), \tag{6.4}$$

and hence, for all $d\in\overline{V}_{+}$,

$$\begin{aligned}(\Phi,&P^{(s,n)}d\Phi)\\ &= \int\cdots\int(\phi(\mathbf{p}_1\cdots\mathbf{p}_n), P'^{(s,n)}b(j(p_1\cdots p_n))\\ &\quad\times d\phi(\mathbf{p}_1\cdots\mathbf{p}_n))dv(\mathbf{p}_1)\cdots dv(\mathbf{p}_n)\geqslant 0,\end{aligned}$$

since $P'^{(s,n)}$ satisfies the spectrum condition if P' does, as a consequence of

$$W^{(s,n)}_{(1,d)} = W_{(1,d)}\otimes 1_{(\mathbb{C}^{(2s+1)})^n}.$$

Choosing for $(\psi, W, \mathscr{K})$ irreducible symplectic representations of the photon field, it is known that the momentum P' of W may be chosen to satisfy the spectrum condition.[5] From the above remarks then follows that $P^{(s,n)}$ satisfies the spectrum condition when symplectic representations $(\psi, W, \mathscr{K})$ are used; hence the premises of statement (3) in Theorem 6.2 are fulfilled in this case.

By Theorem 6.2 we now have uncountably many $P^{\uparrow}_{+}$-representations $U_{\alpha}^{(s,n)}$ corresponding to $\alpha \geqslant 0$, all satisfying the spectrum condition and implementing the Poincaré automorphisms in $\pi^{(s,n)}$.

We propose to choose $\alpha = 1$ for the "physical" representation. By (6.3) and (6.4), then,

$$P_{1}^{(s,n)} = P_{c} + \int_{\oplus T^{n}} b(j(p_{1}\cdots p_{n}))^{-1} P'^{(s,n)} d\nu^{n}(\mathbf{p}_{1}\cdots \mathbf{p}_{n}), \tag{6.5}$$

with

$$e^{ib(j(p_{1}\cdots p_{n}))^{-1}P'^{(s,n)}d} = W^{(s,n)}_{b(j(p_{1}\cdots p_{n}))d}$$

implementing the translation $\tau_{(1,d)}$ in the component fields

$$\psi^{(s,n)}(\tau_{b(j(p_{1},\dots,p_{n}))}(A)),$$

as easily checked. Thus the total four-momentum is just the sum of the particle momenta plus the (generalized) sum of the momenta of all component fields, and hence the total energy is bounded below by n times the particle mass.

Another argument in favor of $\alpha = 1$ is the following: Consider the one-particle sector $(s,1)$. In (4.5) substitute

$$W^{(s,1)}_{(R,d)} \to e^{ip^{R}d} \otimes W_{(R,d)} \otimes D^{(s)}(R) = \widetilde{W}^{(s,1)}_{(R,d)}.$$

Then it is easy to carry out the construction of the $(s,1)$ sector, now by applying (2.7) to $(\psi^{(s,1)}, \overline{W}^{(s,1)}, \mathscr{K}^{(s,1)})$. With β_{g} defined in (3.12), (3.13) we obtain

$$\begin{aligned}
(\overline{U}^{(s,1)}(\Lambda,d)\phi)_{m}(\mathbf{p})
&= e^{ip^{R}b(p)d} D^{(s)}_{mm'}(r(g,p)) W_{(r(g,p),b(p)d)} \phi_{m'}(\Lambda^{-1}p) \\
&= e^{ib(p)^{-1}p^{R}d} D^{(s)}_{mm'}(r(g,p)) W_{(r(g,p),b(p)d)} \phi_{m'}(\Lambda^{-1}p) \\
&= e^{ipd} D^{(s)}_{mm'}(r(g,p)) W_{(r(g,p),b(p)d)} \phi_{m'}(\Lambda^{-1}p).
\end{aligned}$$

In the last step the boost definition (3.9) was used. Thus

$$\overline{U}^{(s,1)}(g) = V_{g}^{1} U^{(s,1)}(g).$$

Hence, choosing $\alpha = 1$, the asymmetry in the description of "ordinary" elementary particles and infraparticles (see Table I) is removed, and Table I may be replaced by Table II.

TABLE II. Relativistic transformation laws.

	$P^{\uparrow}_{+}$ representation
Elementary particle	induced by the representation $(R,d) \to e^{ip^{R}d} \otimes D^{(s)}(R)$ of K
Infraparticle	induced by the representation $(R,d) \to e^{ip^{R}d} \otimes D^{(s)}(R) \otimes W_{(R,d)}$ of K

ACKNOWLEDGMENT

I thank Professor K. Kraus for numerous fruitful discussions and suggestions. His constant support made this work possible. Useful comments of Dr. G. Reents and J. Hering and the patient help of Mrs. Ch. Steinbauer and Dr. W. Petzold during the preparation of the manuscript are gratefully acknowledged.

APPENDIX: *G*-COVARIANT REPRESENTATIONS

Let $(\pi, U, \mathscr{H})$ be a given G-covariant representation π of a C^*-algebra $\mathfrak{A}$ on a Hilbert space $\mathscr{H}$, G being a topological symmetry group of $\mathfrak{A}$. Here U denotes the unitary G-representation implementing the automorphisms τ_{g} of $\mathfrak{A}$, i.e.,

$$U(g_{1})U(g_{2}) = U(g_{1}g_{2}) \tag{A1}$$

and

$$U(g)\pi(A)U(g)^{-1} = \pi(\tau_{g}(A)). \tag{A2}$$

Since we study arbitrary, in general reducible, representations π of $\mathfrak{A}$, the commutant $\pi(\mathfrak{A})'$ need not be trivial. We will first introduce unitary cocycles (see Refs. 3 and 20).

Definition A.1: A unitary cocycle is a map V: $g \to V_{g}$ from G into the unitary operators on $\mathscr{H}$ obeying the cocycle equation

$$V_{g_{1}g_{2}} = V_{g_{1}} U(g_{1}) V_{g_{2}} U(g_{1})^{-1}. \tag{A3}$$

The set of all unitary cocycles is denoted by $Z(G)$, the set of all unitary cocycles with range in $\pi(\mathfrak{A})'$ by $Z^{\pi}(G)$. We shall denote by $Z_{c}(G)$ and $Z_{c}^{\pi}(G)$, respectively, the corresponding subsets of continuous cocycles. Now any unitary operator W defines an element of $Z(G)$ by

$$V_{g}^{W} = WU(g)W^{-1}U(g)^{-1}. \tag{A4}$$

Cocycles of this form are called coboundaries.[20] They form the subsets $B(G)$, $B_{c}(G)$, $B^{\pi}(G)$, $B_{c}^{\pi}(G)$ of the above introduced sets of cocycles.

Definition A.2: Let V^{1}, V^{2} be elements of $Z(G)$. Then V^{1} is cohomologous to V^{2} ($V^{1} \sim V^{2}$), if there exists a unitary operator W such that (see Ref. 20)

$$V_{g}^{1} = WV_{g}^{2}U(g)W^{-1}U(g)^{-1}. \tag{A5}$$

As easily shown, this defines an equivalence relation, so we can form the corresponding quotient spaces $H(G)$, $H_{c}(G)$, $H^{\pi}(G)$, $H_{c}^{\pi}(G)$. The elements of these spaces are called cohomology classes.

Theorem A.3: Let $(\pi, U, \mathscr{H})$ be the given G-covariant representation of $\mathfrak{A}$, with a (continuous) unitary representation U of G implementing τ_{g}. Then U' is another unitary (continuous) G-representation implementing τ_{g} if and only if it may be written as

$$U'(g) = V_{g}U(g) \tag{A6}$$

with an element V of $Z^{\pi}(G)(Z_{c}^{\pi}(G))$.

Proof: Given an arbitrary $V \in Z^{\pi}_{(c)}(G)$, U' as defined in (A6) is easily seen to be a unitary (continuous) G-representation implementing τ_{g}. Vice versa, given such a representation U', define V by $V_{g} = U'(g)U(g)^{-1}$. Then it follows immediately that V satisfies (A3) and that $V_{g} \in \pi(\mathfrak{A})'$. ∎

Since in applications we shall identify G with the Poincaré group $P^{\uparrow}_{+}$, we will focus on continuous G-representations in order to guarantee the existence of observables like energy, momentum or angular momentum of the system under consideration. Let $\mathscr{U}$ be the set of all continuous unitary G-representations implementing τ_g in $(\pi,\mathscr{H})$, $\overline{\mathscr{U}}$ the set of (unitary) equivalence classes of elements from $\mathscr{U}$. Then we have the following theorem.

Theorem A.4: Equation (A.6) induces a one-to-one map from $\overline{\mathscr{U}}$ onto $H_c^\pi(G)$.

Proof: Let $\cong$ denote unitary equivalence. Then using Theorem A3 we get for elements U' and U'' from $\mathscr{U}$:

$$\begin{aligned} U' \cong U'' &\leftrightarrow U'(g) \\ &= V_g' U(g) = W U''(g) W^{-1} \\ &= W V_g'' U(g) W^{-1} \text{ with a unitary } W \\ &\leftrightarrow V_g' = W V_g'' U(g) W^{-1} U(g)^{-1} \\ &\leftrightarrow V' \sim V''. \end{aligned}$$ ■

If $\pi(\mathfrak{A})'$ is Abelian, $Z^\pi(G)$ is an Abelian group with the product cocycle defined by $g \to (V^1 \circ V^2)_g = V_g^1 V_g^2$ and the identity element $g \to 1$. As an example take the quasilocal algebra $\mathfrak{A}$ of a physical system (see Ref. 21) and choose the representation $\pi = \oplus_i \pi_i$ on $\mathscr{H} = \oplus_i \mathscr{H}_i$, where the indices label the superselection sectors of $\mathfrak{A}$ (i.e., each π_i is irreducible, and $\pi_i \not\cong \pi_j$ for $i \neq j$). Then $\pi(\mathfrak{A})'$ is generated by the projections P_i on the subspaces $\mathscr{H}_i$, and hence is Abelian.[21]

[1]J. Fröhlich, G. Morchio, and F. Strocchi, Ann. Phys. (NY) **119**, 241 (1979).
[2]J. Fröhlich, G. Morchio, and F. Strocchi, Phys. Lett. B **89**, 61 (1979).
[3]D. Buchholz, Commun. Math. Phys. **85**, 49 (1982).
[4]H.-G. Stark, Dissertation, Universität Würzburg, 1984.
[5]K. Kraus, L. Polley, and G. Reents, Ann. Inst. H. Poincaré **26**, 109 (1977).
[6]G. Reents, *Modelle zum Infrarotproblem in der Quantenelektrodynamik* (Habilitationsschrift, Universität Würzburg, 1977).
[7]K. Kraus, Found. Phys. **13**, 701 (1983).
[8]G. Roepstorff, Commun. Math. Phys. **19**, 301 (1970).
[9]P. Basarab-Horwath, R. F. Streater, and J. R. Wright, Commun. Math. Phys. **68**, 195 (1979).
[10]K. Kraus and H.-G. Stark, J. Phys. A: Math. Gen. **15**, 2041 (1982).
[11]J. Dixmier, *C*-Algebras* (North-Holland, Amsterdam, 1982).
[12]S. Sakai, *C*-Algebras and W*-Algebras* (Springer, Berlin, 1971).
[13]H. -G. Stark, "A Note on Induced Group Representations and Covariant Representations of C*-Algebras," to be published in Rep. Math. Phys.
[14]G. W. Mackey, Ann. Math. **55**, 101 (1952).
[15]G. W. Mackey, *Unitary Group Representations in Physics Probability and Number Theory* (Benjamin/Cummings, Reading, MA, 1978).
[16]W. Rühl, *The Lorentz Group and Harmonic Analysis* (Benjamin, New York, 1970).
[17]J. Dixmier, *Les Algèbras d'Opérateurs dans L'Espace Hilbertian (Algèbres de von Neumann)* (Gauthiers-Villars, Paris, 1957).
[18]K. Kraus (personal communication).
[19]E. P. Wigner, Ann. Math. **40**, 149 (1939).
[20]G. W. Mackey, *Induced Representations of Groups and Quantum Mechanics* (Benjamin/Boringheri, New York, 1968).
[21]R. Haag and D. Kastler, J. Math. Phys. **5**, 848 (1964).

Group theoretical aspects of the extended interacting boson model

V. K. B. Kota
Physical Research Laboratory, Navrangpura, Ahmedabad 380009, India

J. Van der Jeugt,[a] H. De Meyer,[b] and G. Vanden Berghe
Seminarie voor Wiskundige Natuurkunde, Rijksuniversiteit Gent, Krijgslaan 281-S9, B9000 Gent, Belgium

(Received 2 December 1986; accepted for publication 11 March 1987)

Group theoretical aspects of the extended interacting boson model (*g*IBM) with *s*, *d*, and *g* bosons are discussed. The basis states and branching rules are given for the seven dynamical symmetries. Casimir operators and their eigenvalues are obtained and enable one to write down energy formulas.

I. INTRODUCTION

The interacting boson model (IBM) of nuclei, introduced by Arima and Iachello,[1] is extremely successful in correlating the spectroscopic properties of low lying levels that contain quadrupole collectivity and pairing effects. This model treats pairs of valence nucleons (particles/holes) as bosons and allows the bosons to carry angular momentum $l=0$ (s bosons) or $l=2$ (d bosons). The most important aspect of IBM is that it contains three dynamical symmetries, described by the groups U(5), SU(3), and O(6). The phenomenal success of this model is well documented.[2] The quest for providing an algebraic description of hexadecupole ($l=4$) collectivity, on the one hand, and the microscopic theories[3] of IBM, on the other hand, lead one to include g bosons also in IBM. The interacting boson model with s, d, and g bosons is referred to as *g*IBM in Refs. 4 and 5. The electric hexadecupole (E4) decay properties of $L=4^+$ levels in nuclei[6] and also the rotational bands with $k=1^+, 3^+$, and 4^+ bandheads[7,8] strongly point out that one should further explore *g*IBM. In this paper we discuss some of the group theoretical aspects of *g*IBM.

The different ways hexadecupole collectivity manifests itself in the low-lying levels in nuclei is given by the various dynamical symmetries of the *g*IBM Hamiltonian. Moreover the identification and exploitation of the dynamical symmetries make *g*IBM analytically tractable and thereby allow for a rapid analysis of experimental data. In a recent letter[5] we showed that the *g*IBM possesses seven dynamical symmetries. The purpose of the present paper is to elaborate the results given there. In Sec. II the generators of the various groups in the seven dynamical symmetry group chains are given. In Sec. III the plethysm problem (reduction of a group representation into that of its subgroups) for the various group–subgroup chains in *g*IBM is solved. In Sec. IV we give the results that enable one to write down energy formulas in the symmetry limits. Some concluding remarks are made in Sec. V.

Note that throughout the paper we always deal with classical Lie algebras. Even when we use the symbol of the Lie group, we mean the Lie algebra, as is the custom in IBM literature.

II. THE *g*IBM AND THE GENERATORS OF ITS LIMITING SYMMETRIES

The symmetry group for *g*IBM is U(15), the unitary group in 15 dimensions. The 15 dimensions correspond to the 15 single particle states, one coming from the s orbit, five coming from the d orbit, and nine coming from the g orbit. The 225 generators of U(15) are

$$b^\dagger_{lm}\, b_{l'm'}\,, \tag{2.1}$$

with $l,l'\in\{0,2,4\}$ and $b_0=s$, $b_2=d$, $b_4=g$. As usual, we let

$$\tilde{b}_{lm} = (-1)^{l+m} b_{l,-m}\,. \tag{2.2}$$

Then the generators (2.1) can be linearly combined into angular momentum tensor operators by

$$(b^\dagger_l \times \tilde{b}_{l'})^{(L_0)}_\mu = \sum_{m_1 m_2} \langle l\, m_1\, l' m_2 | L_0 \mu\rangle b^\dagger_{lm_1} \tilde{b}_{l'm_2}\,. \tag{2.3}$$

The angular momentum generators of the physical O(3) subgroup are described by

$$L^{(1)}_\kappa = (d^\dagger \times \tilde{d})^{(1)}_\kappa + \sqrt{6}(g^\dagger \times \tilde{g})^{(1)}_\kappa\,. \tag{2.4}$$

In order to study the limiting symmetries of *g*IBM, we should know all the subgroups in the chain $U(15)\supset G \supset G' \cdots \supset O(3)$. This problem has been considered by several authors.[4,5,9] Kota[4] gave a classification of the U(15) subgroups, based on physical arguments, and recently it was shown[5] that this classification is complete. In this section, we shall give the generators for the seven chains in $U(15)\supset O(3)$.

A. The unitary orbit chains

Given the single boson orbits with $l=s,d,g$, we can combine them into the orbit combinations (sd,g), (sg,d), and (s,dg), leading to the subgroups $U(6)\oplus U(9)$, $U(10)\oplus U(5)$, and $U(1)\oplus U(14)$, respectively. For each of these subgroups, we shall now describe the generators and the chain to O(3).

(I) $U(6)\oplus U(9)$: The subgroup U(6) is generated by $(b^\dagger_l \times \tilde{b}_{l'})^{(L_0)}_\mu$ with $l,l'\in\{0,2\}$ and U(9) by $(b^\dagger_4 \times \tilde{b}_4)^{(L_0)}_\mu$ ($L_0=0,1,2,\ldots,8$). Hence U(6) is precisely the usual IMB group with s and d bosons, and the chains $U_{sd}(6)\supset O_d(3)$

---

[a] Senior research assistant of the National Fund for Scientific Research (Belgium).

[b] Senior research associate of the National Fund for Scientific Research (Belgium).

 0022-2488/87/071644-09$02.50

are well known.[1] The algebra of U(9) contains O(9) as a subalgebra, spanned by

$$(b_4^\dagger \times \tilde{b}_4)_\mu^{(L_0)}, \quad L_0 = 1,3,5,7\,, \tag{2.5}$$

and $(b_4^\dagger \times \tilde{b}_4)_\mu^{(1)}$ is the $O_g(3)$ subalgebra of O(9). The physical O(3) group of U(15) is then spanned by the diagonal O(3) contained in $O_d(3) \oplus O_g(3)$. Thus the present chain can be described as follows:

$$\begin{tikzcd}
 & U(9) \arrow[r, dash] & O(9) \arrow[rr, dash] & & O_g(3) \arrow[dr, dash] & \\
U(15) \arrow[ur, dash] \arrow[ddr, dash] & \times & U(5) \arrow[r, dash] & O(5) \arrow[ddr, dash] & \times & O(3) \\
 & & O(6) \arrow[ur, dash] & & & \\
 & U(6) \arrow[uur, dash] \arrow[ur, dash] \arrow[r, dash] & U(3) \arrow[rr, dash] & & O_d(3) \arrow[uur, dash] &
\end{tikzcd} \tag{2.6}$$

(II) $U(5) \oplus U(10)$: U(5) is generated by $(d^\dagger \times \tilde{d})_\mu^{(L)}$ $(L = 0,1,2,3,4)$, and is the same U(5) as in (2.6); U(10) has generators $(b_l^\dagger \times \tilde{b}_{l'})_\mu^{(L)}$ with $l,l' = 0,4$. The algebra U(9) in (2.6) is, of course, a subalgebra of U(10), but there is still another subalgebra spanned by

$$\begin{aligned}
&(g^\dagger \times \tilde{g})^{(L)}, \quad L = 1,3,5,7\,, \\
&(s^\dagger \times \tilde{g})^{(4)} + \alpha (g^\dagger \times \tilde{s})^{(4)} \quad (\alpha = \pm)\,.
\end{aligned} \tag{2.7}$$

These are the generators of O(10); we can use either sign for α. Hence the chains to O(3) are

$$\begin{tikzcd}
 & U(5) \arrow[r, dash] & O(5) \arrow[rr, dash] & & O_d(3) \arrow[dr, dash] & \\
U(15) \arrow[ur, dash] \arrow[dr, dash] & \times & O(10) \arrow[dr, dash] & & \times & O(3) \\
 & U(10) \arrow[ur, dash] \arrow[dr, dash] & & O(9) \arrow[r, dash] & O_g(3) \arrow[ur, dash] & \\
 & & U(9) \arrow[ur, dash] & & &
\end{tikzcd} \tag{2.8}$$

(III) $U(1) \oplus U(14)$: The operator $\hat{n}_s = s^\dagger s$ generates a U(1) algebra; $\hat{n}_s$ counts the number of s bosons. Then U(14) is generated by $(b_l^\dagger \times \tilde{b}_{l'})_\mu^{(L)}$ with $l, l' \in \{2,4\}$. The reduction of U(14) to O(3) goes over O(14) and O(5). The basis elements of O(14) are given by

$$\begin{aligned}
&(d^\dagger \times \tilde{d})^{(1,3)}, \quad (g^\dagger \times \tilde{g})^{(1,3,5,7)}, \\
&[(d^\dagger \times \tilde{g})^{(k)} + \alpha_0(-1)^k (g^\dagger \times \tilde{d})^{(k)}], \\
&\quad k = 2,3,4,5,6 \quad (\alpha_0 = \pm)\,.
\end{aligned} \tag{2.9}$$

There is still an O(5) subalgebra contained in (2.9), spanned by the operators $L_\kappa^{(1)}$ (2.4) and by the following rank-3 tensor operator

$$\begin{aligned}
Q^{(3)} = {}&(d^\dagger \times \tilde{d})^{(3)} - \tfrac{3}{8}\sqrt{11}\,(g^\dagger \times \tilde{g})^{(3)} \\
&- \alpha \tfrac{3}{4}\sqrt{\tfrac{5}{2}}(d^\dagger \times \tilde{g} + g^\dagger \times \tilde{d})^{(3)} \quad (\alpha = \pm)\,.
\end{aligned} \tag{2.10}$$

The $\pm$ sign in (2.9) and (2.10) may be freely chosen: the commutation relations of $\{L_\kappa^{(1)}, Q_\mu^{(3)}\}$ are independent of this sign. The present chain reads

$$U(15) \text{——} U(14) \text{——} O(14) \text{——} O(5) \text{——} O(3)\,. \tag{2.11}$$

Note that with the O(5) subgroup in (2.11), α_0 in (2.9) is constrained to be -1.

B. The SU(5) and SU(3) limits

One can think of the l values $l = 0,2,4$ to be those of a two-boson system with each "pseudo" boson carrying angular momentun $\tilde{l} = 2$. This gives rise to a subgroup $SU(2\tilde{l}+1) = SU(5)$ of U(15). On the other hand, every system with $l = 0,2,4,\ldots,k$ bosons contains Elliott's SU(3) subgroup.[10] Note that for sd IBM the pseudoboson subgroup $SU(2\tilde{l}+1)$ $(\tilde{l} = 1)$ coincides with SU(3). This is the reason why SU(5) and SU(3) are discussed in the same subsection.

(IV) SU(5): Here SU(5) contains a tensor of rank 1, 2, 3, and 4. The rank-1 tensor is $L_\kappa^{(1)}$, given by (2.4), and the rank-3 tensor is $Q^{(3)}$ of (2.10). The remaining tensor operators have the following expression:

$$\begin{aligned}
Q^{(2)} = {}&(d^\dagger \times \tilde{d})^{(2)} - \sqrt{22}(g^\dagger \times \tilde{g})^{(2)} \\
&- \alpha(4/\sqrt{5})(d^\dagger \times \tilde{g} + g^\dagger \times \tilde{d})^{(2)} \\
&- \alpha\beta(14/3\sqrt{5})(d^\dagger \times \tilde{s}^\dagger \times \tilde{d})^{(2)}\,, \\
Q^{(4)} = {}&(d^\dagger \times \tilde{d})^{(4)} + (\sqrt{143}/4\sqrt{5})(g^\dagger \times \tilde{g})^{(4)} \\
&+ \alpha(\sqrt{55}/2\sqrt{2})(d^\dagger \times \tilde{g} + g^\dagger \times \tilde{d})^{(4)} \\
&+ \beta(7/2\sqrt{5})(s^\dagger \times \tilde{g} + g^\dagger \times \tilde{s})^{(4)} \\
&\quad (\alpha = \pm, \quad \beta = \pm) \quad .
\end{aligned} \tag{2.12}$$

As before, the structure constants are independent of the choice of α and β. Obviously, the O(5) algebra of (2.11) is contained in SU(5). Therefore we are dealing with the chain

$$U(15) \text{——} SU(5) \text{——} O(5) \text{——} O(3)\,. \tag{2.13}$$

(V) SU(3): The SU(3) algebra consists of the operators $L_\kappa^{(1)}$ of (2.4) and the rank-2 tensor

$$\begin{aligned}
T^{(2)} = {}&(d^\dagger \times \tilde{d})^{(2)} + (3\sqrt{2}/\sqrt{11})(g^\dagger \times \tilde{g})^{(2)} \\
&+ \alpha(18\sqrt{2}/11\sqrt{5})(d^\dagger \times \tilde{g} + g^\dagger \times \tilde{d})^{(2)} \\
&+ \beta(14\sqrt{2}/11\sqrt{5})(s^\dagger \times \tilde{d} + d^\dagger \times \tilde{s})^{(2)} \quad (\alpha,\beta = \pm)\,.
\end{aligned} \tag{2.14}$$

The SU(3) chain is the "shortest" chain to the physical O(3) subgroup:

$$U(15) \text{——} SU(3) \text{——} O(3)\,. \tag{2.15}$$

C. The O(15) and SU(6) limits

An obvious subgroup of U(15) is the generalized seniority group O(15), just like O(6) is the seniority group of U(6) for sd IBM. Finally, one can think of the l values $l = 0,2,4$ to be those of two fermions, with each "pseudofermion" carrying angular momentum $\tilde{j} = \frac{5}{2}$, giving rise to the group chain

$$\begin{aligned}
&U(15) \supset SU(2\tilde{j}+1) \\
&\quad = SU(6) \supset Sp(2\tilde{j}+1) = Sp(6) \supset O(3)\,.
\end{aligned}$$

Note that for sd IBM $\tilde{j} = \frac{3}{2}$, and hence $U(6) \supset SU(4) \supset Sp(4) \supset O(3)$. Since $SU(4) \sim O(6)$ and $Sp(4) \sim O(5)$, this is again the seniority group chain.

(VI) O(15): The generators of O(15) are given by

$$\begin{aligned}
&(d^\dagger \times \tilde{d})^{(1,3)}, (g^\dagger \times \tilde{g})^{(1,3,5,7)}, \\
&(s^\dagger \times \tilde{d} + \alpha d^\dagger \times \tilde{s})^{(2)}, \quad (s^\dagger \times \tilde{g} + \beta g^\dagger \times \tilde{s})^{(4)}\,, \\
&[(d^\dagger \times \tilde{g})^{(k)} - \alpha\beta(-1)^k (g^\dagger \times \tilde{d})^{(k)}]_{k=2,3,4,5,6} \\
&\quad (\alpha,\beta = \pm)\,.
\end{aligned} \tag{2.16}$$

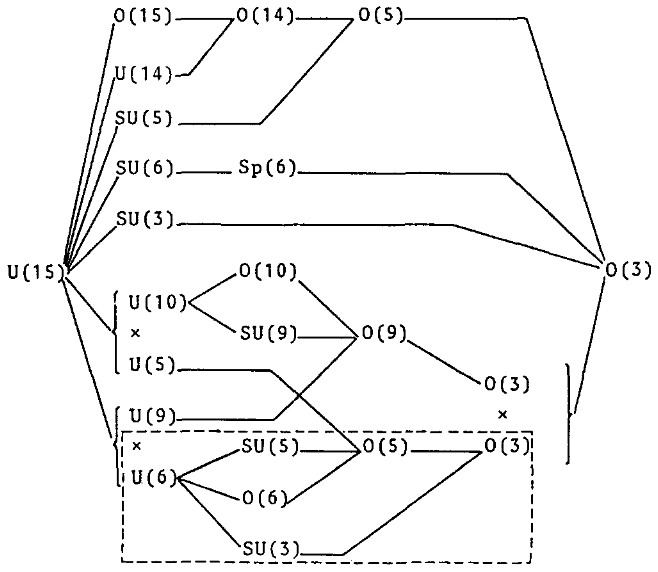

FIG. 1. Classification of the dynamical symmetry groups in the U(15) extended IBM and their consecutive decomposition into the physical O(3) group.

Clearly, O(15) contains O(14) generated by (2.9), and hence also the O(5) subgroup of (2.11):

$$\mathrm{U}(15)\text{——}\mathrm{O}(15)\text{——}\mathrm{O}(14)\text{——}\mathrm{O}(5)\text{——}\mathrm{O}(3)\,. \tag{2.17}$$

Note that the α,β in (2.16), $Q^{(3)}$ in (2.10), and α_0 in (2.9) should be chosen consistently.

(VII) SU(6): The SU(6) algebra contains the O(3) generators $P^{(1)}_{\kappa} = L^{(1)}_{\kappa}$ and the rank-k tensors $P^{(k)}$ with $k = 2,3,4,5$. An explicit expression reads

$$\begin{aligned}
P^{(2)} &= (d^{\dagger}\times\tilde{d})^{(2)} - (3\sqrt{11}/5\sqrt{2})(g^{\dagger}\times\tilde{g})^{(2)} \\
&\quad - \alpha(9\sqrt{3}/10)(d^{\dagger}\times\tilde{g} + g^{\dagger}\times\tilde{d})^{(2)} \\
&\quad - \alpha\beta\tfrac{7}{5}(s^{\dagger}\times\tilde{d} + d^{\dagger}\times\tilde{s})^{(2)}, \\
P^{(3)} &= (d^{\dagger}\times\tilde{d})^{(3)} + (\sqrt{11}/9)(g^{\dagger}\times\tilde{g})^{(3)} \\
&\quad - \alpha(5\sqrt{6}/9)(d^{\dagger}\times\tilde{g} + g^{\dagger}\times\tilde{d})^{(3)}, \\
P^{(4)} &= (d^{\dagger}\times\tilde{d})^{(4)} - (\sqrt{143}/3\sqrt{5})(g^{\dagger}\times\tilde{g})^{(4)} \\
&\quad + \alpha(\sqrt{22}/3\sqrt{3})(d^{\dagger}\times\tilde{g} + g^{\dagger}\times\tilde{d})^{(4)} \\
&\quad + \beta(14/3\sqrt{15})(s^{\dagger}\times\tilde{g} + g^{\dagger}\times\tilde{s})^{(4)}, \\
P^{(5)} &= (g^{\dagger}\times\tilde{g})^{(5)} + \alpha(\sqrt{15}/\sqrt{26})(d^{\dagger}\times\tilde{g} + g^{\dagger}\times\tilde{d})^{(5)} \\
&\quad (\alpha,\beta = \pm)\,.
\end{aligned} \tag{2.18}$$

The Sp(6) subgroup of SU(6) is generated by the odd rank tensors $P^{(k)}$, $k = 1,3,5$. Thus the last group chain is given by

$$\mathrm{U}(15)\text{——}\mathrm{SU}(6)\text{——}\mathrm{Sp}(6)\text{——}\mathrm{O}(3)\,. \tag{2.19}$$

The seven subgroup chains of U(15) are shown in Fig. 1.

III. THE BASIS STATES IN THE VARIOUS LIMITS AND BRANCHING RULES

The basis states of gIBM are classified into totally symmetric representations of U(15), labeled[11] by the Cartan–Dynkin numbers $(N,0,0,...) = [N]$, where N is the total number of nucleon pairs (or holes) near a closed shell. In this section, we shall discuss the reduction of such representations into the irreducible representations (irreps) of the various subgroups.

(I) $\mathrm{U}(6)\oplus\mathrm{U}(9)$: In the reduction $\mathrm{U}(15)\supset\mathrm{U}(6)\oplus\mathrm{U}(9)$, the irreps $[N]$ decompose into $[n_{sd}]\otimes[n_g]$ $(n_{sd},n_g \geqslant 0)$, with $n_{sd} + n_g = N$, of $\mathrm{U}(6)\oplus\mathrm{U}(9)$. The reduction of $\mathrm{U}_{sd}(6)\supset G\supset \mathrm{O}_{sd}(3)$ is the classical IBM chain[1] and is well known. For the chain $\mathrm{U}_g(9)\supset\mathrm{O}(9)\supset\mathrm{O}(3)$, the basis states are denoted by

$$\left|\begin{matrix}\mathrm{U}(9) & \supset & \mathrm{O}(9) & \supset & \mathrm{O}(3)\\ n_g & & v_g & & \alpha,L_g\end{matrix}\right\rangle. \tag{3.1}$$

Herein, v_g is the O(9) label, since for the reduction $\mathrm{U}(9)\supset\mathrm{O}(9)$ one finds

$$\begin{aligned}
&[n_g]\to\sum[v_g] = (v_g,0,0,0)\,,\\
&v_g = n_g, n_g-2,...,1 \text{ or } 0\,.
\end{aligned} \tag{3.2}$$

The reduction $\mathrm{O}(9)\supset\mathrm{O}(3)$ is not multiplicity-free: α stands for the missing labels in this chain. The multiplicity $M(L_g)$ of L_g in the decomposition of the irrep $[v_g]$ may be calculated by means of Littlewood's rules.[12] On the other hand, there are also tables[13] available for the reduction $\mathrm{O}(9)\supset\mathrm{O}(3)$. Since we are dealing with totally symmetric representations of U(9), there is still an alternative method of obtaining the L contents of a U(9) irrep $[N]$. This procedure is powerful and easy to implement for machine calculations. Consider the group $\mathrm{U}(n)$, $n = \sum_{i=1}^{k}(2l_i+1)$, and the decomposition of the irreps $[N]$ of $\mathrm{U}(n)$ into O(3) irreps (L), with the condition $[1]\to(l_1)+(l_2)+\cdots+(l_k)$. To this end, we first generate the single particle spectrum for the l_z operator: this consists of a number of l_z eigenvalues m_i, each with degeneracy d_i. Then, distributing the given number of bosons N in the l_z orbits in all possible ways, one finds the degeneracy $d(m)$ of a given total m value. Let $(n_1,...,n_k)$ be a distribution of N bosons so that $N = \sum_{i=1}^{k} n_i$ and $m = \sum_{i=1}^{k} n_i m_i$. For this distribution, the degeneracy of the m value is $\prod_{i=1}^{k}\binom{d_i+n_i-1}{n_i}$. Thus the degeneracy of the $\sum_{i=1}^{k} l_z(i)$ eigenvalue m is

$$d(m) = \sum_{(n_1,n_2,...,n_k)}\prod_{i=1}^{k}\binom{d_i+n_i-1}{n_i}, \tag{3.3}$$

where the summation is over all configurations $(n_1,...,n_k)$ such that $n_1+...+n_k = N$ and $m = \sum_{i=1}^{k} n_i m_i$. Then the simple difference formula

$$D_N(L) = d(m=L) - d(m=L+1) \tag{3.4}$$

gives the degeneracy of a given L value, and thus, according to (3.2),

$$M_{v_g}(L) = D_{n_g=v_g}(L) - D_{n_g=v_g-2}(L)\,. \tag{3.5}$$

For the reduction $\mathrm{O}(9)\supset\mathrm{O}(3)$ we shall list the decomposition of all irreps $(v_g,0,0,0)$ for $v_g \leqslant 10$ in Table I.

(II) $\mathrm{U}(5)\oplus\mathrm{U}(10)$: The U(5) algebra is the same as in sd IBM, and hence also the labeling can be maintained. The reduction $\mathrm{U}(15)\supset\mathrm{U}(5)\oplus\mathrm{U}(10)$ is given by

$$[N]\to[0]\otimes[N],[1]\otimes[N-1],...,[N]\otimes[0]\,. \tag{3.6}$$

TABLE I. Reduction of O(9) irreps $(v_g,0,0,0)$ for $v_g \leqslant 10$ into O(3) irreps (L). The multiplicity of (L) is written as an exponent.

1	4
2	2, 4, 6, 8
3	$0, 2, 3, 4, 5, 6^2, 7, 8, 9, 10, 12$
4	$0, 2^2, 3, 4^3, 5^2, 6^3, 7^2, 8^3, 9^2, 10^3, 11, 12^2, 13, 14, 16$
5	$0, 1, 2^3, 3^2, 4^4, 5^4, 6^5, 7^4, 8^6, 9^4, 10^5, 11^4, 12^4, 13^3, 14^3, 15^2, 16^2, 17, 18, 20$
6	$0^2, 1, 2^4, 3^4, 4^7, 5^5, 6^9, 7^7, 8^9, 9^8, 10^9, 11^7, 12^9, 13^6, 14^7, 15^5, 16^5, 17^3, 18^4, 19^2, 20^2, 21, 22, 24$
7	$0^2, 1^2, 2^6, 3^6, 4^{10}, 5^9, 6^{12}, 7^{12}, 8^{14}, 9^{13}, 10^{15}, 11^{13}, 12^{14}, 13^{12}, 14^{13}, 15^{10}, 16^{11}, 17^8, 18^8, 19^6, 20^5, 21^4, 22^4, 23^2, 24^2, 25, 26, 28$
8	$0^3, 1^3, 2^9, 3^8, 4^{14}, 5^{14}, 6^{18}, 7^{17}, 8^{22}, 9^{19}, 10^{23}, 11^{21}, 12^{23}, 13^{20}, 14^{22}, 15^{18}, 16^{19}, 17^{16}, 18^{16}, 19^{12}, 20^{13}, 21^9, 22^9, 23^6, 24^6, 25^4, 26^4, 27^2, 28^2, 29, 30, 32$
9	$0^4, 1^5, 2^{11}, 3^{13}, 4^{19}, 5^{19}, 6^{26}, 7^{25}, 8^{30}, 9^{30}, 10^{33}, 11^{31}, 12^{35}, 13^{32}, 14^{33}, 15^{31}, 16^{31}, 17^{27}, 18^{28}, 19^{23}, 20^{23}, 21^{19}, 22^{18}, 23^{14}, 24^{14}, 25^{10}, 26^9, 27^7, 28^6, 29^4, 30^4, 31^2, 32^2, 33, 34, 36$
10	$0^5, 1^6, 2^{16}, 3^{17}, 4^{26}, 5^{27}, 6^{35}, 7^{35}, 8^{42}, 9^{41}, 10^{48}, 11^{45}, 12^{50}, 13^{47}, 14^{50}, 15^{46}, 16^{49}, 17^{43}, 18^{44}, 19^{39}, 20^{39}, 21^{33}, 22^{33}, 23^{27}, 24^{26}, 25^{21}, 26^{20}, 27^{15}, 28^{15}, 29^{10}, 30^{10}, 31^7, 32^6, 33^4, 34^4, 35^2, 36^2, 37, 38, 40$

The remaining reductions are described by

$U(10) \supset O(10)$:

$$[n_{sg}] \to (v_{sg},0,0,0,0), \quad v_{sg} = n_{sg}, n_{sg}-2,\ldots,1 \text{ or } 0\,; \tag{3.7}$$

$U(10) \supset U(9)$:

$$[n_{sg}] \to [n_g]\,, \quad n_g = n_{sg}, n_{sg}-1,\ldots,0\,; \tag{3.8}$$

$O(10) \supset O(9)$:

$$(v_{sg},0,0,0,0) \to (v_g,0,0,0), \quad v_g = v_{sg}, v_{sg}-1,\ldots,0\,, \tag{3.9}$$

and the reductions $U(9) \supset O(9)$ and $O(9) \supset O(3)$ have already been discussed in the previous chain. Thus the states in the U(10) chain are labeled by

$$|n_{sg}, v_{sg}, v_g, \alpha, L_g\rangle \tag{3.10a}$$

or

$$|n_{sg}, n_g, v_g, \alpha, L_g\rangle\,. \tag{3.10b}$$

(III) U(14): In the U(14) chain, the basis states are given by

$$\left|\begin{matrix} U(15) \supset U(14) \supset O(14) \supset & O(5) & \supset O(3) \\ N \qquad\quad n_{dg} \qquad\quad v_{dg} & \alpha,(a_1,a_2) & \beta,L \end{matrix}\right\rangle.$$

The reductions $U(15) \supset U(14)$ and $U(14) \supset O(14)$ are classical for symmetric irreps, and are determined by

$$\begin{aligned} n_{dg} &= N, N-1,\ldots,1,0\,, \\ v_{dg} &= n_{dg}, n_{dg}-2,\ldots,1 \text{ or } 0\,. \end{aligned} \tag{3.11}$$

The reductions $O(14) \supset O(5)$ and $O(5) \supset O(3)$ are more involved. There are several ways to deal with these problems, but since we are considering the reduction of symmetric representations, we shall discuss a method similar to the one in $U(9) \supset O(3)$, i.e., a method based on weight space techniques. First, consider the decomposition of U(14) irreps $(n_{dg},0,0,\ldots)$ into irreps of O(5). It is easy to see that in this reduction the weights of the 14-dimensional standard representation $(1,0,0,\ldots)$ are projected into the following weights of O(5):

$$\begin{aligned} &(2,-2),(-2,2),(2,0),(-2,0),(1,-1),(-1,1), \\ &\quad(0,-2),(0,2),(2,2),(-2,2),(1,1), \\ &\quad(-1,-1),(0,0),(0,0)\,. \end{aligned} \tag{3.12}$$

Then, we consider all totally symmetric tensor products of $(1,0,0,\ldots)$ with itself. In their projection onto the O(5) weight space, the weights of these tensor products are generated by

$$\left[\prod_{\mu}(1-Ue^{\mu})\right]^{-1}, \tag{3.13}$$

where μ runs over all values (3.12), and e^{μ} is the formal exponential. In the expansion of (3.13), the factor F_n of U^n is then the character of the irrep $(n,0,0,\ldots)$ of U(14) projected onto the O(5) weight space. Using Weyl's character formula[14] one knows that F_n is of the following form:

$$F_n = \sum_{\lambda} \frac{\xi_\lambda}{\Delta}, \tag{3.14}$$

where the summation is over the highest weights λ of the O(5) irreps in which $(n,0,0,\ldots)$ decomposes. In (3.14), ξ_λ is the characteristic of (λ), and Δ is Weyl's denominator:

$$\Delta = \prod_{\alpha \in R^+} (e^{\alpha/2} - e^{-\alpha/2})\,, \tag{3.15}$$

where R^+ is the set of positive roots of the Lie algebra. For O(5), R^+ equals

$$\{(1,1),(2,0),(1,-1),(0,-2)\}\,. \tag{3.16}$$

Hence in order to find the reduction of $(n,0,0,\ldots)$ into O(5) irreps (a_1,a_2) we perform the following algorithm: (1) compute F_n in (3.13); (2) multiply F_n by Δ, given in (3.15); (3) preserve all the terms in the dominant Weyl sector in $F_n \cdot \Delta$, i.e., preserve all terms $e^{(\lambda_1,\lambda_2)}$ with $\lambda_2 \leqslant 0$ and $\lambda_1 \geqslant -\lambda_2$; (4) for every such highest weight (λ_1,λ_2), the corresponding Cartan–Dynkin labels are given by $(a_1,a_2) = (\lambda_1+\lambda_2-1, -\lambda_2-1)$. Note that an irrep with Cartan–Dynkin labels (a_1,a_2) has Young labels $[\tau_1,\tau_2] = [a_1/2+a_2, a_1/2]$. This algorithm is easy to program, and is a straightforward extension of the one given for $U(9) \supset O(3)$. As soon as we have the reduction $U(14) \supset O(5)$, we can use (3.11) in order to obtain the reduction $O(14) \supset O(5)$. These results are listed in Table II for $v_{dg} \leqslant 10$.

For the reduction of irreps (a_1,a_2) of O(5) into irreps (L) of O(3), several solutions have been given in the literature. Tables of McKay and Patera[13] list the reduction for $\dim(a_1,a_2) \leqslant 5000$, and for all other irreps one can easily use the branching rule generating function (23) of Gaskell *et al.*[15] Therefore we shall not give any tables in the present paper.

(IV) SU(5): This limit has been discussed in some detail by Sun *et al.*[9] The decomposition of $[N]$ of U(15) into irreps (m_1,m_2,m_3,m_4) of SU(5) is determined by (see Thrall[16] for a proof)

$$[N] \to \sum_{p,q,r,s} (2N-4p-6q-8r-10s, 2p, 2q, 2r)\,. \tag{3.17}$$

TABLE II. Reduction of O(14) irreps $[v_{dg}]$ for $v_{dg} \leqslant 10$ into O(5) irreps (a_1, a_2); the multiplicity of (a_1, a_2) is written in front. Note that $[\tau_1, \tau_2] = [a_1/2 + a_2, a_1/2]$ are the Young labels.

1	(0 2).						
2	(0 2),	(0 4),	(4 0).				
3	(0 0),	(0 2),	(0 4),	(2 2),	(4 0),	(0 6),	(4 2).
4	(0 0),	2(0 2),	2(0 4),	(2 2),	(4 0),	(4 1),	(0 6),
	(2 4),	2(4 2),	(0 8),	(4 4),	(8 0).		
5	(0 0),	2(0 2),	3(0 4),	2(2 2),	2(4 0),	(4 1),	2(0 6),
	2(2 4),	3(4 2),	(4 3),	(6 1),	(0 8),	(2 6),	2(4 4),
	(6 2),	(8 0),	(0 10),	(4 6),	(8 2).		
6	(0 0),	3(0 2),	4(0 4),	2(2 2),	3(4 0),	(2 3),	(4 1),
	4(0 6),	3(2 4),	5(4 2),	(6 0),	2(4 3),	(6 1),	2(0 8),
	2(2 6),	4(4 4),	2(6 2),	2(8 0),	(4 5),	(6 3),	(8 1),
	(0 10),	(2 8),	2(4 6),	(6 4),	2(8 2),	(0 12),	(4 8),
	(8 4),	(12 0).					
7	(0 0),	4(0 2),	5(0 4),	3(2 2),	3(4 0),	(2 3),	2(4 1),
	5(0 6),	5(2 4),	7(4 2),	(6 0),	(2 5),	3(4 3),	2(6 1),
	4(0 8),	4(2 6),	7(4 4),	4(6 2),	3(8 0),	2(4 5),	2(6 3),
	2(8 1),	2(0 10),	2(2 8),	4(4 6),	3(6 4),	4(8 2),	(4 7),
	(6 5),	(8 3),	(10 1),	(0 12),	(2 10),	2(4 8),	(6 6),
	2(8 4),	(10 2),	(12 0),	(0 14),	(4 10),	(8 6),	(12 2).
8	2(0 0),	4(0 2),	7(0 4),	4(2 2),	4(4 0),	(2 3),	3(4 1),
	.7(0 6),	7(2 4),	9(4 2),	(6 0),	2(2 5),	5(4 3),	3(6 1),
	6(0 8),	6(2 6),	9(4 4),	6(6 2),	5(8 0),	(2 7),	4(4 5),
	4(6 3),	3(8 1),	4(0 10),	4(2 8),	8(4 6),	6(6 4),	7(8 2),
	(10 0),	2(4 7),	2(6 5),	3(8 3),	2(10 1),	2(0 12),	2(2 10),
	4(4 8),	3(6 6),	5(8 4),	2(10 2),	2(12 0),	(4 9),	(6 7),
	(8 5),	(10 3),	(12 1),	(0 14),	(2 12),	2(4 10),	(6 8),
	2(8 6),	(10 4),	2(12 2),	(0 16),	(4 12),	(8 8),	(12 4),
	(16 0).						
9	2(0 0),	5(0 2),	8(0 4),	5(2 2),	5(4 0),	2(2 3),	3(4 1),
	10(0 6),	9(2 4),	12(4 2),	2(6 0),	3(2 5),	7(4 3),	4(6 1),
	8(0 8),	10(2 6),	15(4 4),	9(6 2),	6(8 0),	2(2 7),	7(4 5),
	7(6 3),	5(8 1),	6(0 10),	7(2 8),	13(4 6),	10(6 4),	11(8 2),
	2(10 0),	(2 9),	4(4 7),	5(6 5),	6(8 3),	3(10 1),	4(0 12),
	4(2 10),	8(4 8),	7(6 6),	9(8 4),	5(10 2),	4(12 0),	2(4 9),
	2(6 7),	3(8 5),	3(10 3),	2(12 1),	2(0 14),	2(2 12),	4(4 10),
	3(6 8),	(8 6),	3(10 4),	4(12 2),	(4 11),	(6 9),	(8 7),
	(10 5),	(12 3),	(14 1),	(0 16),	(2 14),	2(4 12),	(6 10),
	2(8 8),	(10 6),	2(12 4),	(14 2),	(16 0),	(0 18),	(4 14),
	(8 10),	(12 6),	(16 2).				
10	2(0 0),	6(0 2),	10(0 4),	6(2 2),	6(4 0),	(0 5),	2(2 3),
	4(4 1),	12(0 6),	12(2 4),	15(4 2),	2(6 0),	4(2 5),	9(4 3),
	6(6 1),	12(0 8),	14(2 6),	11(4 4),	12(6 2),	8(8 0),	4(2 7),
	11(4 5),	10(6 3),	7(8 1),	9(0 10),	11(2 8),	19(4 6),	16(6 4),
	16(8 2),	3(10 0),	2(2 9),	8(4 7),	9(6 5),	10(8 3),	6(10 1),
	6(0 12),	7(2 10),	14(4 8),	12(6 6),	16(8 4),	8(10 2),	6(12 0),
	(2 11),	4(4 9),	5(6 7),	7(8 5),	6(10 3),	4(12 1),	4(0 14),
	4(2 12),	8(4 10),	7(6 8),	10(8 6),	7(10 4),	8(12 2),	(14 0),
	2(4 11),	2(6 9),	3(8 7),	3(10 5),	3(12 3),	2(14 1),	2(0 16),
	2(2 14),	4(4 12),	3(6 10),	5(8 8),	3(10 6),	5(12 4),	2(14 2),
	2(16 0),	(4 13),	(6 11),	(8 9),	(10 7),	(12 5),	(14 3),
	(16 1),	(0 18),	(2 16),	2(4 14),	(6 12),	2(8 10),	(10 8),
	2(12 6),	(14 4),	2(16 2),	(0 20),	(4 16),	(8 12),	(12 8),
	(16 4),	(20 0).					

Note that the Young labels of (m_1,m_2,m_3,m_4) are $\{m_1+m_2+m_3+m_4, m_2+m_3+m_4, m_3+m_4, m_4\}$. The reduction of SU(5) irreps (m_1,m_2,m_3,m_4) into irreps (a_1,a_2) of O(5) is in general a four-missing label problem. Tables for this reduction have been listed by McKay and Patera.[13] On the other hand, there are also some analytic decomposition rules available for SU(5) irreps of the form $(m,0,0,0)$, $(m,2,0,0)$, $(m,4,0,0)$, and $(m,2,2,0)$ (see Refs. 4

TABLE III. Reduction of U(15) irreps $[N]$ for $N\leqslant 10$ into SU(3) irreps (λ,μ); the multiplicity of (λ,μ) is written in front.

1	(4 0).						
2	(8 0),	(4 2),	(0 4).				
3	(12 0),	(8 2),	(6 3),	(4 4),	(6 0),	(3 3),	(0 6),
	(2 2),	(0 0).					
4	(16 0),	(12 2),	(10 3),	2(8 4),	(10 0),	(7 3),	(8 1),
	2(4 6),	(5 4),	2(6 2),	(3 5),	(4 3),	(5 1),	(0 8),
	2(2 4),	2(4 0),	(1 3),	(0 2).			
5	(20 0),	(16 2),	(14 3),	2(12 4),	(14 0),	(10 5),	(11 3),
	(12 1),	2(8 6),	(9 4),	3(10 2),	(6 7),	2(7 5),	2(8 3),
	2(9 1),	2(4 8),	(5 6),	4(6 4),	(7 2),	3(8 0),	(3 7),
	2(4 5),	3(5 3),	(6 1),	(0 10),	(1 8),	3(2 6),	2(3 4),
	4(4 2),	(1 5),	(2 3),	(3 1),	2(0 4),	2(2 0).	
6	(24 0),	(20 2),	(18 3),	2(16 4),	(18 0),	(14 5),	(15 3),
	(16 1),	3(12 6),	(13 4),	3(14 2),	(10 7),	2(11 5),	3(12 3),
	2(13 1),	3(8 8),	2(9 6),	5(10 4),	2(11 2),	4(12 0),	(6 9),
	2(7 7),	4(8 5),	5(9 3),	2(10 1),	2(4 10),	2(5 8),	6(6 6),
	4(7 4),	7(8 2),	2(3 9),	3(4 7),	5(5 5),	5(6 3),	3(7 1),
	2(0 12),	4(2 8),	4(3 6),	7(4 4),	2(5 2),	5(6 0),	2(1 7),
	2(2 5),	4(3 3),	2(4 1),	4(0 6),	2(1 4),	4(2 2),	2(0 0).
7	(28 0),	(24 2),	(22 3),	2(20 4),	(22 0),	(18 5),	(19 3),
	(20 1),	3(16 6),	(17 4),	3(18 2),	2(14 7),	2(15 5),	3(16 3),
	2(17 1),	3(12 8),	2(13 6),	6(14 4),	2(15 2),	4(16 0),	2(10 9),
	3(11 7),	5(12 5),	6(13 3),	3(14 1),	3(8 10),	3(9 8),	8(10 6),
	6(11 4),	9(12 2),	(6 11),	3(7 9),	6(8 7),	8(9 5),	8(10 3),
	5(11 1),	3(4 12),	2(5 10),	8(6 8),	8(7 6),	13(8 4),	5(9 2),
	7(10 0),	2(3 11),	4(4 9),	8(5 7),	9(6 5),	9(7 3),	6(8 1),
	(0 14),	(1 12),	5(2 10),	5(3 8),	12(4 6),	8(5 4),	11(6 2),
	4(1 9),	4(2 7),	8(3 5),	7(4 3),	4(5 1),	5(0 8),	3(1 6),
	9(2 4),	3(3 2),	5(4 0),	3(1 3),	(2 1),	3(0 2).	
8	(32 0),	(28 2),	(26 3),	2(24 4),	(26 0),	(22 5),	(23 3),
	(24 1),	3(20 6),	(21 4),	3(22 2),	2(18 7),	2(19 5),	3(20 3),
	2(21 1),	4(16 8),	2(17 6),	6(18 4),	2(19 2),	4(20 0),	2(14 9),
	3(15 7),	5(16 5),	6(17 3),	3(18 1),	4(12 10),	4(13 8),	9(14 6),
	7(15 4),	10(16 2),	2(10 11),	4(11 9),	8(12 7),	10(13 5),	10(14 3),
	6(15 1),	4(8 12),	3(9 10),	11(10 8),	11(11 6),	17(12 4),	7(13 2),
	9(14 0),	(6 13),	4(7 11),	8(8 9),	13(9 7),	15(10 5),	14(11 3),
	9(12 1),	3(4 14),	3(5 12),	10(6 10),	11(7 8),	21(8 6),	15(9 4),
	18(10 2),	(11 0),	2(3 13),	5(4 11),	11(5 9),	14(6 7),	18(7 5),
	16(8 3),	9(9 1),	2(0 16),	(1 14),	6(2 12),	8(3 10),	18(4 8),
	14(5 6),	23(6 4),	10(7 2),	10(8 0),	4(1 11),	6(2 9),	13(3 7),
	14(4 5),	14(5 3),	7(6 1),	6(0 10),	7(1 8),	15(2 6),	9(3 4),
	14(4 2),	1(5 0),	(0 7),	7(1 5),	5(2 3),	5(3 1),	7(0 4),
	(1 2),	4(2 0).					
9	(36 0),	(32 2),	(30 3),	2(28 4),	(30 0),	(26 5),	(27 3),
	(28 1),	3(24 6),	(25 4),	3(26 2),	2(22 7),	2(23 5),	3(24 3),
	2(25 1),	4(20 8),	2(21 6),	6(22 4),	2(23 2),	4(24 0),	3(18 9),
	3(19 7),	6(20 5),	6(21 3),	3(22 1),	4(16 10),	4(17 8),	10(18 6),
	7(19 4),	10(20 2),	3(14 11),	5(15 9),	9(16 7),	11(17 5),	11(18 3),
	6(19 1),	5(12 12),	4(13 10),	13(14 8),	13(15 6),	19(16 4),	8(17 2),
	10(18 0),	2(10 13),	5(11 11),	11(12 9),	16(13 7),	19(14 5),	17(15 3),
	11(16 1),	4(8 14),	5(9 12),	14(10 10),	16(11 8),	28(12 6),	20(13 4),
	23(14 2),	2(15 0),	2(6 15),	4(7 13),	10(8 11),	18(9 9),	23(10 7),
	27(11 5),	25(12 3),	13(13 1),	3(4 16),	3(5 14),	12(6 12),	15(7 10),
	30(8 8),	26(9 6),	36(10 4),	17(11 2),	15(12 0),	3(3 15),	6(4 13),
	14(5 11),	21(6 9),	29(7 7),	31(8 5),	28(9 3),	15(10 1),	2(0 18),
	(1 16),	7(2 14),	10(3 12),	22(4 10),	24(5 8),	38(6 6),	25(7 4),
	29(8 2),	4(9 0),	5(1 13),	9(2 11),	19(3 9),	23(4 7),	28(5 5),
	23(6 3),	14(7 1),	9(0 12),	9(1 10),	23(2 8),	21(3 6),	29(4 4),
	12(5 2),	13(6 0),	3(0 9),	11(1 7),	12(2 5),	16(3 3),	7(4 1),
	11(0 6),	7(1 4),	12(2 2),	(0 3),	2(1 1),	4(0 0).	
10	(40 0),	(36 2),	(34 3),	2(32 4),	(34 0),	(30 5),	(31 3),
	(32 1),	3(28 6),	(29 4),	3(30 2),	2(26 7),	2(27 5),	3(28 3),
	2(29 1),	4(24 8),	2(25 6),	6(26 4),	2(27 2),	4(28 0),	3(22 9),

TABLE III. (*Continued.*)

3(23 7),	6(24 5),	6(25 3),	3(26 1),	5(20 10),	4(21 8),	10(22 6),
7(23 4),	10(24 2),	3(18 11),	5(19 9),	10(20 7),	11(21 5),	11(22 3),
6(23 1),	6(16 12),	5(17 10),	14(18 8),	14(19 6),	20(20 4),	8(21 2),
10(22 0),	3(14 13),	6(15 11),	13(16 9),	18(17 7),	21(18 5),	18(19 3),
12(20 1),	5(12 14),	6(13 12),	17(14 10),	19(15 8),	32(16 6),	23(17 4),
26(18 2),	2(19 0),	3(10 15),	6(11 13),	14(12 11),	23(13 9),	30(14 7),
33(15 5),	30(16 3),	16(17 1),	5(8 16),	5(9 14),	17(10 12),	22(11 10),
40(12 8),	35(13 6),	47(14 4),	23(15 2),	18(16 0),	(6 17),	5(7 15),
12(8 13),	22(9 11),	33(10 9),	43(11 7),	46(12 5),	39(13 3),	22(14 1),
4(4 18),	4(5 16),	14(6 14),	20(7 12),	39(8 10),	42(9 8),	60(10 6),
42(11 4),	44(12 2),	7(13 0),	3(3 17),	7(4 15),	17(5 13),	27(6 11),
41(7 9),	50(8 7),	53(9 5),	43(10 3),	26(11 1),	2(0 20),	(1 18),
8(2 16),	11(3 14),	29(4 12),	33(5 10),	55(6 8),	49(7 6),	60(8 4),
28(9 2),	22(10 0),	7(1 15),	11(2 13),	26(3 11),	36(4 9),	47(5 7),
46(6 5),	42(7 3),	22(8 1),	10(0 14),	13(1 12),	32(2 10),	33(3 8),
50(4 6),	36(5 4),	37(6 2),	3(7 0),	3(0 11),	19(1 9),	23(2 7),
31(3 5),	25(4 3),	16(5 1),	18(0 8),	16(1 6),	27(2 4),	12(3 2),
14(4 0),	3(0 5),	9(1 3),	4(2 1),	7(0 2).		

and 9). For example,

$$(m,0,0,0) \to (a_1,a_2) = \sum_{r=0}^{[m/2]} (m-2r,0) ,$$

$$\begin{aligned}(m-2,2,0,0) &\to (a_1,a_2)\\ &= (m,0) + \sum_{r=1}^{[(m-2)/2]} (m-2r,0)^2 \\ &\quad + (m-2[m/2],0) + \sum_{r=1}^{[(m-1)/2]} (m-2r,2) \\ &\quad + \sum_{r=0}^{[(m-2)/2]} (m-2r-2,4) .\end{aligned} \tag{3.18}$$

In (3.18) $[x]$ is integer part of x. Finally, the reduction $O(5)\supset O(3)$ has already been discussed in chain III. Thus the basis states in the present chain are labeled by

$$|N\,(m_1,m_2,m_3,m_4)\;\alpha,(a_1,a_2)\;\beta,L\,\rangle . \tag{3.19}$$

(V) SU(3): The SU(3) chain is the one that has been studied most extensively.[8,9,17–21] In order to obtain the reduction $U(15)\supset SU(3)$, we have used a similar method as previously described for $U(14)\supset O(5)$. For an alternative procedure which is based on Littlewood's theorems, see Kota.[22] Listings for the reduction of $[N]$ into SU(3) irreps (λ,μ) are given in Table III for $N\leqslant 10$. The reduction of SU(3) irreps into SO(3) multiplets is well known,[10] and thus the states in this chain are labeled by

$$|N\,\alpha,(\lambda,\mu)\;K,L\,\rangle . \tag{3.20}$$

(VI) O(15): The reductions $U(15)\supset O(15)$ and $O(15)\supset O(14)$ are classical, and determined by

$U(15)\supset O(15)$:

$$[N]\to\sum(v_{sdg},0,\ldots,0),$$

$$v_{sdg}=N,N-2,\ldots,1 \text{ or } 0 ; \tag{3.21}$$

$O(15)\supset O(14)$:

$$(v_{sdg},0,\ldots)\to\sum(v_{dg},0,\ldots,0),$$

$$v_{dg}=v_{sdg},v_{sdg}-1,\ldots,0 . \tag{3.22}$$

The decomposition of $(v_{dg},0,\ldots)$ has been discussed before. The states in this chain are labeled by

$$|N\,v_{sdg}\;v_{dg}\;\alpha,(a_1,a_2)\;\beta,L\,\rangle . \tag{3.23}$$

(VII) SU(6): Using weight space techniques, as described in chain III, one can prove that the reductions in the chain $U(15)\supset SU(6)\supset Sp(6)$ are given by the following rules:

$U(15)\supset SU(6)$:

$$[N\,]\to\sum_{j,k}(0,j,0,N-2j-3k,0) ; \tag{3.24}$$

$SU(6)\supset Sp(6)$:

$$(0,\mu,0,\nu,0)\to\sum_{p=0}^{\min(\mu,\nu)}\;\sum_{q=0}^{p}\;\sum_{r=p-q}^{\mu+\nu-p-q}(q,r,q) . \tag{3.25}$$

For SU(6), the Young labels of $(0,\mu,0,\nu,0)$ are $\{\mu+\nu,\mu+\nu,\nu,\nu,0\}$, and for Sp(6) the Young labels of (q,r,q) are given by $\langle 2q+r,q+r,q\rangle$. Finally, we have to consider the decomposition of Sp(6) irreps (q,r,q) into O(3) irreps (L). There is no analytic formula for the reduction $Sp(6)\supset O(3)$, hence we have to list the reductions in tables. Such tables have actually already been provided by McKay and Patera,[13] therefore we do not repeat them here; see also Kota.[23] The basis states in this chain are described by

$$|N\,(0,\mu,0,\nu,0)\;p(q,r,q)\;\alpha,L\,\rangle . \tag{3.26}$$

IV. CASIMIR OPERATORS AND ENERGY SPECTRA

When the gIBM-Hamiltonian H is expressible in terms of the Casimir operators of the groups appearing in a symmetry group chain, one speaks of a dynamical symmetry. In that case, we can write down the energy formula. With the generators given in Sec. II one can construct the Casimir operators of the various groups in the seven limits of gIBM. The expressions for the matrix elements of the Casimir operators of $U(N)$, $Sp(N)$, and $O(N)$ are well known.[24] If we assume that H is a $(1+2)$-body operator, we have to deal with the linear and quadratic Casimir operators only.

Rather than giving the energy formulas for each chain separately, we shall only list here the general expressions for the Casimir operators for various representations. The linear and quadratic Casimir operators of $\mathrm{U}(M)$ have the following eigenvalue when acting on states of symmetric irreps $(k,0,...,0) = [k]$:

$$\langle C_1(\mathrm{U}(M))\rangle = k, \quad \langle C_2(\mathrm{U}(M))\rangle = k(k+M) . \qquad (4.1)$$

We encouter them while dealing with $\mathrm{U}_d(5)$, $\mathrm{U}_g(9)$, $\mathrm{U}_{sd}(6)$, $\mathrm{U}_{sg}(10)$, $\mathrm{U}_{dg}(14)$, $\mathrm{U}_s(1)$, and $\mathrm{U}_{sdg}(15)$ groups, the subscripts denoting the relevant l orbits. Here the Casimir operators $C_1(\mathrm{U}(M))$, and $C_2(\mathrm{U}(M))$ are simply expressible in terms of number operators. For example, for the $\mathrm{U}_{sg}(10)$ group they are $\hat{n}_{sg} = (\hat{n}_s + \hat{n}_g)$ and $\hat{n}_{sg}(\hat{n}_{sg} + 10)$, respectively. The quadratic Casimir of $\mathrm{O}(M)$, acting on states of the irrep with Cartan–Dynkin labels $(k,0,...,0)$ has the value

$$\langle C_2(\mathrm{O}(M))\rangle = k(k+M-2) . \qquad (4.2)$$

This expression is useful for the $\mathrm{O}_d(5)$, $\mathrm{O}_g(9)$, $\mathrm{O}_{sd}(6)$, $\mathrm{O}_{sg}(10)$, $\mathrm{O}_{dg}(14)$, and $\mathrm{O}_{sdg}(15)$ groups. The explicit form of the Casimir operator $C_2(\mathrm{O}(M))$ for these groups can be given by the following general considerations. Given $l_1,l_2,...,l_k$ orbit, we can define $u^{(\lambda)}(l_il_i)$, $v^{(\lambda')}(l_il_j)$ operators as

$$u_\nu^{(\lambda)}(l_il_i) = (b_{l_i}^\dagger \times \tilde{b}_{l_i})_\nu^{(\lambda)} ,$$
$$v_{\nu'}^{(\lambda')}(l_il_j) = \left[(b_{l_i}^\dagger \times \tilde{b}_{l_j})_{\nu'}^{(\lambda')} + \alpha(l_il_j)(-1)^{\lambda'}(b_{l_j}^\dagger \times \tilde{b}_{l_i})_{\nu'}^{(\lambda')}\right] , \qquad (4.3)$$

where $\alpha(l_il_j) = \pm 1$, $\alpha(l_il_j) = \alpha(l_jl_i)$, and $\alpha(l_il_j)\,\alpha(l_jl_k) = -\alpha(l_il_k)$. Now $u_\nu^{(\lambda)}(l_il_i)$ and λ odd and $v_\nu^{(\lambda)}(l_il_j)$ with $l_i > l_j$ (or $l_i < l_j$) generate the orthogonal group $\mathrm{O}(M = \Sigma(2l_i + 1))$. Now the quadratic Casimir operator of the group $\mathrm{O}(M)$ is given by

$$C_2(\mathrm{O}(M)) = \left[2 \sum_{\substack{l_i \\ \lambda\ \mathrm{odd}}} u^{(\lambda)}(l_il_i)\cdot u^{(\lambda)}(l_il_i)\right] + \left[\sum_{\substack{l_i > l_j \\ \lambda'}} v^{(\lambda')}(l_il_j)\cdot v^{(\lambda')}(l_jl_i)\right] . \qquad (4.4)$$

Herein $u^{(\lambda)}\cdot u^{(\lambda)}$ stands for $(-1)^\lambda\sqrt{2\lambda+1}\,(u^{(\lambda)}\times u^{(\lambda)})_0^{(0)}$. The quadratic Casimir operator of the $\mathrm{SU}(5)$ group in chain IV can be expressed in terms of the operators $L^{(1)}$, $Q^{(3)}$ and $Q^{(2)}$, $Q^{(4)}$, defined in (2.4), (2.10), and (2.12), respectively, as

$$C_2(\mathrm{SU}(5)) = 10\left[\tfrac{1}{4}L^{(1)}\cdot L^{(1)} + \tfrac{9}{196}Q^{(2)}\cdot Q^{(2)} + \tfrac{16}{49}Q^{(3)}\cdot Q^{(3)} + \tfrac{4}{49}Q^{(4)}\cdot Q^{(4)}\right] . \qquad (4.5)$$

Its eigenvalue for representations (a_1,a_2,a_3,a_4) of $\mathrm{SU}(5)$ is given by

$$\langle C_2(\mathrm{SU}(5))\rangle = 2a_1^2 + 3a_2^2 + 3a_3^2 + 2a_4^2 + 3a_1a_2 + 2a_1a_3 + a_1a_4 + 4a_2a_3 + 2a_2a_4 + 3a_3a_4 + 5(2a_1 + 3a_2 + 3a_3 + 2a_4) , \qquad (4.6)$$

or, equivalently, in terms of the Young labels $\{m_1,m_2,m_3,m_4\}$:

$$\langle C_2(\mathrm{SU}(5))\rangle = 2(m_1^2 + m_2^2 + m_3^2 + m_4^2) - (m_1m_2 + m_1m_3 + m_1m_4 + m_2m_3 + m_2m_4 + m_3m_4) + 5(2m_1 + m_2 - m_4) . \qquad (4.7)$$

Similarly for the $\mathrm{O}(5)$ group in chain IV,

$$C_2(\mathrm{O}(5)) = 16\left[\tfrac{1}{4}L^{(1)}\cdot L^{(1)} + \tfrac{16}{49}Q^{(3)}\cdot Q^{(3)}\right] \qquad (4.8)$$

and for representations of $\mathrm{O}(5)$ with Dynkin labels (a_1,a_2) and Young labels $[\tau_1,\tau_2]$ one finds

$$\langle C_2(\mathrm{O}(5))\rangle = \tfrac{1}{2}(a_1^2 + 2a_1a_2 + 2a_2^2 + 4a_1 + 6a_2) \qquad (4.9)$$
$$= \tau_1(\tau_1 + 3) + \tau_2(\tau_2 + 1) . \qquad (4.10)$$

For the $\mathrm{SU}(3)$ group in chain V, the quadratic Casimir operator and its matrix elements are

$$C_2(\mathrm{SU}(3)) = \tfrac{3}{4}\left[\tfrac{242}{21}T^{(2)}\cdot T^{(2)} + 10L^{(1)}\cdot L^{(1)}\right] \qquad (4.11)$$

and

$$\langle C_2(\mathrm{SU}(3))\rangle = \lambda^2 + \mu^2 + \lambda\mu + 3(\lambda + \mu) , \qquad (4.12)$$

where $T^{(2)}$ is defined in (2.14). Finally for the $\mathrm{SU}(6)$ and $\mathrm{Sp}(6)$ groups in chain VII we have

$$C_2(\mathrm{SU}(6)) = 3\left[\tfrac{1}{7}L^{(1)}\cdot L^{(1)} + \tfrac{25}{294}P^{(2)}\cdot P^{(2)} + \tfrac{81}{392}P^{(3)}\cdot P^{(3)} + \tfrac{45}{392}P^{(4)}\cdot P^{(4)} + \tfrac{13}{28}P^{(5)}\cdot P^{(5)}\right] , \qquad (4.13)$$

$$C_2(\mathrm{Sp}(6)) = 4\left[\tfrac{1}{7}L^{(1)}\cdot L^{(1)} + \tfrac{81}{392}P^{(3)}\cdot P^{(3)} + \tfrac{13}{28}P^{(5)}\cdot P^{(5)}\right] , \qquad (4.14)$$

where the operators $P^{(r)}$ are defined in (2.18). The eigenvalues in terms of Cartan–Dynkin labels or Young labels are given by

$$C_2(\mathrm{SU}(6))_{(0\mu0\nu0)} = \mu^2 + \nu^2 + \mu\nu + 6(\mu + \nu) , \qquad (4.15)$$

$$C_2(\mathrm{SU}(6))_{\{m_1,m_2,m_3,m_4,m_5\}} = \sum_{i=1}^{5} m_i(m_i - 2i + 7) , \qquad (4.16)$$

$$C_2(\mathrm{Sp}(6))_{(qrq)} = 3q^2 + r^2 + 3qr + 9q + 5r , \qquad (4.17)$$

$$C_2(\mathrm{Sp}(6))_{\langle\lambda_1\lambda_2\lambda_3\rangle} = \lambda_1(\lambda_1 + 6) + \lambda_2(\lambda_2 + 4) + \lambda_3(\lambda_3 + 2) . \qquad (4.18)$$

With the expressions given above for the Casimir operators and their matrix elements, it is easy to construct gIBM Hamiltonians having any of the seven dynamical symmetries and also we can write down the corresponding energy formulas. For example, in chain II with states (3.10a) one has, up to an N-dependent term,

$$E_{\mathrm{II}} = A_1n_d + A_2n_d(n_d + 5) + A_3v(v + 3) + A_4L_d(L_d + 1) + B_1n_{sg} + B_2n_{sg}(n_{sg} + 10) + B_3v_{sg}(v_{sg} + 8) + B_4v_g(v_g + 7) + B_5L_g(L_g + 1) + C_1L(L + 1) . \qquad (4.19)$$

With the representations of the groups in the seven limits, given in Sec. III and the Casimir operator expressions given above one can easily construct the typical spectra that appear in gIBM.

V. CONCLUDING REMARKS

The first task in the study of the dynamical symmetries of gIBM is completed in this paper, namely the identification of the symmetry group chains, the determination of their generators, the construction of the Casimir operators in terms of the generators, and finally the solution of the plethysm problem for the various group–subgroup chains. Now one is in a position to construct the energy spectra in all the symmetry limits. Moreover, by diagonalizing a linear combination of the Casimir operators of the groups in a given chain, in a convenient basis like $|s^{n_s};d^{n_d}\alpha L_d;g^{n_g}\beta L_g;LM\rangle$, one can construct symmetry defined basis states.

In order to further investigate which symmetry limits have any physical relevance, it is necessary to calculate the electromagnetic transition strengths and particle transfer strengths, since these calculations would give us a clear insight into the band structure of energy spectra. This program, however, requires the derivation of analytic expressions for reduced matrix elements of tensor operators in the various group chains.

[1]A. Arima and F. Iachello, Ann. Phys. (NY) **99**, 253 (1976); **111**, 201 (1978); **123**, 468 (1979).

[2]A. Arima and F. Iachello, Adv. Nucl. Phys. **13**, 139 (1983); J. P. Elliott, Rep. Prog. Phys. **48**, 171 (1985); A. E. L. Dieperink and G. Wenes, Annu. Rev. Nucl. Part. Sci. **35**, 77 (1985).

[3]T. Otsuka, Nucl. Phys. A **368**, 244 (1981). N. Yoshinaga, A. Arima, and T. Otsuka, Phys. Lett. B **143**, 5 (1984); A. Van Egmond and K. Allaart, Nucl. Phys. A **425**, 275 (1984); W. Pannert, P. Ring, and Y. K. Gambhir, *ibid.* **443**, 189 (1985).

[4]V. K. B. Kota, in *Interacting Boson–Boson and Boson–Fermion Systems*, edited by O. Scholten (World Scientific, Singapore, 1984), p. 83.

[5]H. De Meyer, J. Van der Jeugt, G. Vanden Berghe, and V. K. B. Kota, J. Phys. A: Math. Gen. **19**, L565 (1986).

[6]F. Todd Baker, Phys. Rev. C **32**, 1430 (1985); F. Todd Baker, A. Sethi, V. Penumetcha, G. T. Emery, W. P. Jones, M. A. Grimm, and M. L. Whiten, Phys. Rev. C **32**, 2212 (1985); P. B. Goldhoorn, M. N. Harakeh, Y. Iwasaki, L. W. Put, F. Zwarts, and P. Van Isacker, Phys. Lett. B **103**, 291 (1981); P. Van Isacker, K. Heyde, M. Waroquier, and G. Wenes, Phys. Lett. B **104**, 5 (1981); K. Heyde, P. Van Isacker, M. Waroquier, G. Wenes, Y. Gigase, and J. Stachel, Nucl. Phys. A **398**, 235 (1983).

[7]S. Pittel, J. Dukelsky, R. P. J. Perazzo, and H. M. Sofia, Phys. Lett. B **144**, 145 (1984).

[8]N. Yoshinaga, Y. Akiyama, and A. Arima, Phys. Rev. Lett. **56**, 1116 (1986).

[9]H. Z. Sun, M. Moshinsky, A. Frank, and P. Van Isacker, Kinam Rev. Fis. **5**, 135 (1983).

[10]J. P. Elliott, Proc. R. Soc. London Ser. A **245**, 128, 562 (1958).

[11]From now on a labeling with parentheses $(a_1,a_2,...)$ always refers to the Cartan–Dynkin labels of an irrep of a Lie algebra. However, for all irreps considered we also give the corresponding notation in Young labels in the text, such as $[N]$ for symmetric irreps of $\mathrm{U}(n)$, $[v]$ for symmetric irreps of $\mathrm{O}(n)$, $[\tau_1,\tau_2]$ for irreps of $\mathrm{O}(5)$, $\langle\lambda_1,\lambda_2,\lambda_3\rangle$ for irreps of $\mathrm{Sp}(6)$. This is done for the convenience of nuclear physicists.

[12]B. G. Wybourne, *Symmetry Principles and Atomic Spectroscopy* (Wiley, New York, 1970).

[13]W. G. McKay and J. Patera, "Tables of dimensions, indices and branching rules for representations of simple Lie algebras," in *Lecture Notes in Pure and Applied Mathematics*, Vol. 69 (Dekker, New York, 1981).

[14]H. Weyl, Math. Z. **23**, 271 (1925) [reprinted in H. Weyl, *Selecta* (Birkhauser, Basel, 1956)].

[15]R. Gaskell, A. Peccia, and R. T. Sharp, J. Math. Phys. **19**, 727 (1978).

[16]R. M. Thrall, Am. J. Math. **64**, 371 (1942).

[17]R. D. Ratna Raju, Phys. Rev. C **23**, 518 (1981).

[18]H. C. Wu, Phys. Lett. B **110**, 1 (1982).

[19]H. C. Wu and X-Q. Zhou, Nucl. Phys. A **417**, 67 (1984).

[20]Y. Akiyama, Nucl. Phys. A **433**, 369 (1985).

[21]N. Yoshinaga, Nucl. Phys. A **456**, 21 (1986).

[22]V. K. B. Kota, J. Phys. A **10**, L39 (1977).

[23]V. K. B. Kota, "Tables of group representations for the six limiting symmetries in gIBM," Physical Research Laboratory (India), Technical Report, PRL-TN-86-54, 1986.

[24]B. G. Wybourne, *Classical Groups for Physicists* (Wiley, New York, 1974).

Theta series and magic numbers for diamond and certain ionic crystal structures

N. J. A. Sloane
AT&T Bell Laboratories, Murray Hill, New Jersey 07974

(Received 26 September 1986; accepted for publication 25 February 1987)

Two earlier papers by Teo and the author [J. Chem. Phys. **83**, 6520 (1985); Inorg. Chem. **25**, 2315 (1986)] studied circular and spherical clusters in the simplest close-packed structures in two and three dimensions. The present work considers clusters in other fundamental structures (the hexagonal net, diamond), and applies the results to study clusters in related structures (Lonsdaleite, graphite) and in binary arrays with the structure of the idealized ionic crystals NaCl, CsCl, ZnS, CaF_2, TiO_2, O_3Bi_2.

I. INTRODUCTION

In Refs. 1 and 2 Teo and the author investigated circular clusters in the square and hexagonal lattices in two dimensions (2D), and spherical clusters in the simple, face-centered, and body-centered cubic lattices and the hexagonal close-packing in three dimensions (3D), for various choices for the center of the cluster. In these papers we gave the theta series for each cluster, table of the coefficients and their partial sums, which are the nuclearities or magic numbers of the clusters, and coordinates for the atoms in the first few shells of each cluster.

The present paper has two goals: (i) to give the theta series for certain fundamental structures not considered in Refs. 1 and 2 (e.g., the hexagonal net in 2D, the diamond net in 3D); and (ii) to show how these fundamental theta series may be used to study clusters in more complicated structures, including binary and higher-order compounds, and to enumerate the atoms of the individual elements in the clusters. This technique is illustrated by considering the hexagonal diamond (or Lonsdaleite) net, various 3-D nets related to graphite, and binary compounds having the structure of the ionic crystals NaCl, CsCl, ZnS (zinc blende and wurtzite), CaF_2, TiO_2, and O_3Bi_2. For applications of these results see Refs. 1 and 2.

Theta series have been used for almost 100 years in the calculation of *numerical* sums (such as Madelung's constant) associated with lattices: see, for example, Tosi,[3] Glasser and Zucker,[4] and Borwein *et al.*[5] However, the present series of papers appears to represent the first application of theta series (at any rate in recent years) to the *enumerative* or *combinatorial* study of clusters. For example, the especially simple expressions (17) and (19) for clusters in the diamond structure do not appear to have been published before. On the other hand, it would not be surprising if they were to be found somewhere in the older literature. The author would appreciate hearing of any references that have been overlooked. Computer programs (such as MACSYMA[6,7]) that are capable of performing algebraic computations now make it particularly easy to manipulate theta series.

Since the methods are similar to those used in Refs. 1 and 2, the treatment here will be brief. Except for diamond we give just an analytic expression for each theta series, and the number of atoms in the first few shells of the clusters. The nuclearities of the clusters are the partial sums of the latter numbers, and can easily be derived from the information provided, as is illustrated for diamond in Table I.

The same methods may be applied to lattices in spaces of higher dimension; these results are described elsewhere.[8,9] However, it seems worth giving two particularly appealing examples. The theta series of the E_8 or Gosset lattice in eight dimensions (with respect to a lattice point) is

$$\Theta_{E_8}(X) = \tfrac{1}{2}(\theta_2(X)^8 + \theta_3(X)^8 + \theta_4(X)^8). \qquad (1)$$

Comparison of this expression with Eq. (17) shows that E_8 may be regarded as an eight-dimensional diamond lattice. The theta series of the notorious Leech lattice in 24 dimensions (with respect to a lattice point) is

TABLE I. Clusters in diamond structure, centered at (a) an atom, (b) the midpoint of two neighboring atoms, and (c) the center of a tetrahedral hole. Here S_n is the number of points in the spherical shell of radius $\sqrt{n}$, and G_n is the magic number of the cluster.

(a) Diamond, atom			(b) Diamond, edge			(c) Diamond, hole		
n	S_n	G_n	$n-\frac{3}{16}$	S_n	G_n	n	S_n	G_n
0	1	1	0	2	2	$\frac{3}{4}$	4	4
$\frac{3}{4}$	4	5	1	6	8	1	6	10
2	12	17	2	12	20	$2\frac{3}{4}$	12	22
$2\frac{3}{4}$	12	29	3	12	32	3	8	30
4	6	35	4	6	38	$4\frac{3}{4}$	12	42
$4\frac{3}{4}$	12	47	5	18	56	5	24	66
6	24	71	6	18	74	$6\frac{3}{4}$	16	82
$6\frac{3}{4}$	16	87	7	12	86	$8\frac{3}{4}$	24	106
8	12	99	8	30	116	9	30	136
$8\frac{3}{4}$	24	123	9	14	130	$10\frac{3}{4}$	12	148
10	24	147	10	6	136	11	24	172
$10\frac{3}{4}$	12	159	11	30	166	$12\frac{3}{4}$	24	196
12	8	167	12	24	190	13	24	220
$12\frac{3}{4}$	24	191	13	18	208	$14\frac{3}{4}$	36	256
14	48	239	14	30	238	$16\frac{3}{4}$	12	268
$14\frac{3}{4}$	36	275	15	26	264	17	48	316
16	6	281	16	24	288	$18\frac{3}{4}$	28	344
$16\frac{3}{4}$	12	293	17	30	318	19	24	368
18	36	329	18	24	342	$20\frac{3}{4}$	36	404
$18\frac{3}{4}$	28	357	19	18	360	21	48	452
20	24	381	20	24	384	$22\frac{3}{4}$	24	476
$20\frac{3}{4}$	36	417	21	36	420	$24\frac{3}{4}$	36	512
22	24	441	22	24	444	25	30	542
$22\frac{3}{4}$	24	465	23	48	492	$26\frac{3}{4}$	36	578
24	24	489	24	30	522	27	32	610

 0022-2488/87/071653-05$02.50

$$\Theta_{\text{Leech}}(X) = \tfrac{1}{2}(\theta_2(X)^{24} + \theta_3(X)^{24} + \theta_4(X)^{24}) - \tfrac{69}{16}(\theta_2(X)\theta_3(X)\theta_4(X))^8. \tag{2}$$

There is no 3-D analog of this lattice.

II. THETA SERIES

Let T be an array of points (or *atoms*) in Euclidean space (of any numbers of dimensions). The *norm* $s \cdot s$ of a vector s is its squared length. The *theta series* of T with respect to an arbitrary point P is the formal power series

$$\Theta_{T,P}(X) = \sum_{t \in T} X^{(t-P)\cdot(t-P)}. \tag{3}$$

The subscript P may be omitted if it is clear from the context. The coefficient of X^n in $\Theta_{T,P}(X)$, S_n say, is therefore the number of atoms in T at squared distance n from P, i.e., the number of atoms on the spherical shell of radius $\sqrt{n}$ around P. The partial sum

$$G_n = \sum_{m \le n} S_m \tag{4}$$

is the total number of atoms in the *spherical cluster* of radius $\sqrt{n}$ centered at P, i.e., the *nuclearity* or *magic number* of that cluster.

If T is a *binary* array, consisting of two types of atoms, say

$$T = T_X \cup T_Y,$$

where T_X (resp. T_Y) is the set of atoms of type X (resp. Y), then the bivariate theta series of T with respect to an arbitrary point P is defined to be

$$\Theta_{T,P}(X,Y) = \Theta_{T_X,P}(X) + \Theta_{T_Y,P}(Y). \tag{5}$$

For greater emphasis we sometimes replace X and Y by the symbols for the corresponding elements, as is illustrated in Eqs. (27)–(29).

The symbols Z, Z^2, Z^3 denote the *integer points* along a line, the *square lattice* of points with integer coordinates in 2D, and the *simple cubic lattice* in 3D, respectively. The *hexagonal lattice* in 2D (with coordination number 6, and scaled so that neighboring atoms are at unit distance apart) is denoted by $\mathbf{A}_2$.

The theta series of the structures considered in this paper may be conveniently expressed in terms of the seven fundamental functions given in Eqs. (6)–(12). These are

$$\eta_a(X) = \sum_{m=-\infty}^{\infty} X^{(m+a)^2} = \Theta_{Z+a}(X), \tag{6}$$

for any real number a, which is the theta series with respect to the origin of the 1-D array

$\ldots, a-1, a, a+1, a+2, a+3, \ldots$;

$$\theta_2(X) = \eta_{1/2}(X) = \Theta_{Z+1/2}(X) = 2X^{1/4} + 2X^{9/4} + 2X^{25/4} + 2X^{49/4} + \cdots, \tag{7}$$

$$\theta_3(X) = \eta_0(X) = \Theta_Z(X) = 1 + 2X + 2X^4 + 2X^9 + 2X^{16} + 2X^{25} + \cdots, \tag{8}$$

$$\theta_4(X) = \theta_3(-X) = 1 - 2X + 2X^4 - 2X^9 + 2X^{16} - 2X^{25} + \cdots \tag{9}$$

(θ_2, θ_3, and θ_4 are particular examples of Jacobi theta series[10]); and the theta series of the hexagonal lattice with respect to an atom, the midpoint of an edge joining two neighboring atoms, and the center of a triangular hole, respectively, which are

$$\phi_0(X) = \Theta_{\mathbf{A}_2,\text{atom}}(X) = \theta_2(X)\theta_2(X^3) + \theta_3(X)\theta_3(X^3) = 1 + 6X + 6X^3 + 6X^4 + 12X^7 + 6X^9 + 6X^{12} + \cdots, \tag{10}$$

$$\phi_1(X) = \Theta_{\mathbf{A}_2,\text{edge}}(X) = \tfrac{1}{2}\theta_2(X^{1/4})\theta_2(X^{3/4}) = 2X^{1/4} + 2X^{3/4} + 4X^{7/4} + 2X^{9/4} + 4X^{13/4} + \cdots, \tag{11}$$

$$\phi_2(X) = \Theta_{\mathbf{A}_2,\text{hole}}(X) = \theta_2(X)\eta_{1/6}(X^3) + \theta_3(X)\eta_{1/3}(X^3) = \tfrac{1}{2}(\phi_0(X^{1/3}) - \phi_0(X)) = 3X^{1/3} + 3X^{4/3} + 6X^{7/3} + 6X^{13/3} + 3X^{16/3} + \cdots. \tag{12}$$

Tables 4–6 of Ref. 1 give the first 80 terms in (10)–(12). The Jacobi theta series $\theta_2,\theta_3,\theta_4$ satisfy numerous identities[1,4,10]; these have been used to simplify later formulas whenever possible.

III. TWO-DIMENSIONAL NETS

Two-dimensional nets have been extensively studied: see, for example, Wells[11,12] and O'Keeffe and Hyde.[13] There are three regular nets, two of which (the square lattice $4^4 = Z^2$ and the hexagonal *lattice* $3^6 = \mathbf{A}_2$) we investigated in Refs. 1, 2, and 14.

We now consider the third regular net, the *hexagonal net* $6^3 = H_2$ (Fig. 1). This may be regarded as the union of $\mathbf{A}_2$ and a reflected copy of $\mathbf{A}_2$. Let neighboring atoms be at unit distance apart. Then, using the results in Ref. 1, it follows that the theta series of H_2 with respect to an atom, the midpoint of an edge joining neighboring atoms, and the center of a hexagonal hole are, respectively,

$$\Theta_{H_2,\text{atom}}(X) = \phi_0(X^3) + \phi_2(X^3) = \tfrac{1}{2}(\phi_0(X) + \phi_0(X^3)) = 1 + 3X + 6X^3 + 3X^4 + 6X^7 + 6X^9 + 6X^{12} + 6X^{13} + 3X^{16} + 6X^{19} + \cdots, \tag{13}$$

$$\Theta_{H_2,\text{edge}}(X) = \phi_1(X) - \phi_1(X^3) = \theta_2(X^{3/4})\eta_{1/6}(X^{9/4}) = 2X^{1/4} + 4X^{7/4} + 4X^{13/4} + 4X^{19/4} + 2X^{25/4} + 4X^{31/4} + 4X^{37/4} + 4X^{43/4} + 6X^{49/4} + 4X^{61/4} + \cdots, \tag{14}$$

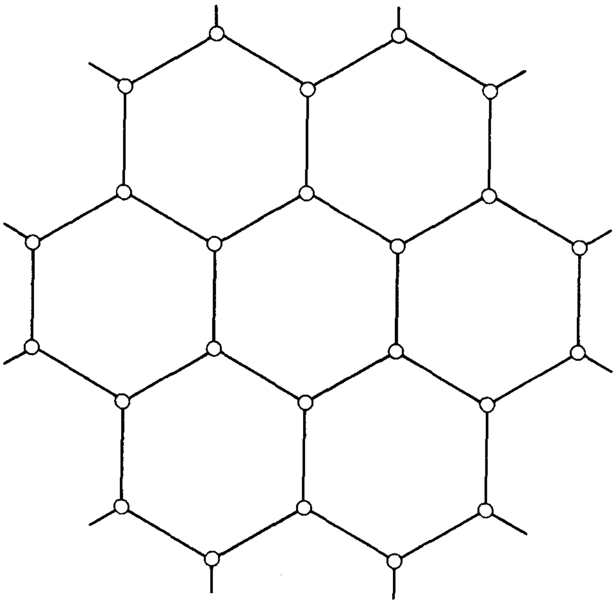

FIG. 1. Two-dimensional hexagonal net H_2.

$$\Theta_{H_2,\text{hole}}(X) = \phi_0(X) - \phi_0(X^3) = 2\phi_2(X^3) = 6X + 6X^4 + 12X^7 + 12X^{13} + 6X^{16} + 12X^{19} + 6X^{25} + 12X^{28} + 12X^{31} + 12X^{37} + \cdots. \quad (15)$$

Remarks: In principal there is no difficulty in calculating the theta series of any 2-D or 3-D structure (as illustrated in Sec. IX of Ref. 1). The interesting question is to find as simple an expression as possible. Using Refs. 1 and 2 and the above formulas simple expressions may be obtained for many other 2-D nets. For example the theta series of the *Kagomé net* (Fig. 8 of Ref. 13) with respect to an atom and the center of a hexagonal hole are

$$\phi_0(z) - \phi_1(4z) \quad \text{and} \quad \phi_0(z) - \phi_0(4z), \quad (16)$$

respectively.

IV. THE DIAMOND NET

The *diamond net* (p. 117 of Ref. 11, p. 121 of Ref. 12, p. 26 of Ref. 15) may be regarded as the union of a face-centered cubic (fcc) lattice [in which the 12 neighbors of the origin have coordinates of the form $(\pm 1, \pm 1, 0)$] and a translation of this fcc lattice by $(\frac{1}{2},\frac{1}{2},\frac{1}{2})$. The minimal distance between atoms is $\sqrt{3/2}$. Then using the results in Ref. 1 we find that the theta series of diamond with respect to an atom [e.g., the point $(0,0,0)$], the midpoint of an edge [e.g., $(\frac{1}{4},\frac{1}{4},\frac{1}{4})$], and the center of a tetrahedral hole [e.g., $(0,0,1)$] are, respectively,

$$\Theta_{\text{diamond,atom}}(X) = \tfrac{1}{2}(\theta_2(X)^3 + \theta_3(X)^3 + \theta_4(X)^3) = 1 + 4X^{3/4} + 12X^2 + 12X^{11/4} + \cdots, \quad (17)$$

$$\Theta_{\text{diamond,edge}}(X) = \theta_2(\sqrt{X})\{\theta_2(X^2)\eta_{3/8}(X^4) + \theta_3(X^2)\eta_{1/8}(X^4)\} = 2X^{3/16} + 6X^{19/16} + 12X^{35/16} + 12X^{51/16} + \cdots, \quad (18)$$

$$\Theta_{\text{diamond,hole}}(X) = \tfrac{1}{2}(\theta_2(X)^3 + \theta_3(X)^3 - \theta_4(X)^3) = 4X^{3/4} + 6X + 12X^{11/4} + 8X^3 + \cdots. \quad (19)$$

Table I displays the first 50 coefficients S_n in (17)–(19) and their partial sums G_n [see Eq. (4)]. The first column agrees with and extends a small table given by Prins and Petersen[16] (reprinted on p. 1039 of Ref. 17).

For the closely related *hexagonal diamond* or *Lonsdaleite net* (Fig. 3.35c of Ref. 12) we give just the theta series with respect to an atom, which on the same scale is

$$\{\theta_2(X^{16/3})\phi_2(X^2) + \theta_3(X^{16/3})\phi_0(X^2)\} + \{\eta_{1/8}(X^{16/3}) \times\phi_2(X^2) + (\tfrac{1}{2}\theta_2(X^{4/3}) - \eta_{1/8}(X^{16/3}))\phi_0(X^2)\} = 1 + 4X^{3/4} + 12X^2 + X^{25/12} + 9X^{11/4} + 6X^4 + 6X^{49/12} + 9X^{19/4} + 2X^{16/3} + 18X^6 + \cdots \quad (20)$$

(using Tables 16 and 22 of Ref. 1).

V. GRAPHITE AND RELATED 3-D NETS

The hexagonal net H_2 contains twice as many atoms as hexagonal holes, and many different 3-D nets may be formed by stacking appropriately displaced layers of H_2 (p. 922 of Ref. 12). By stacking identical layers, i.e., using the sequence $AAA\cdots$ we obtain the *primitive hexagonal array* (p. 596 of Ref. 13). If the distance between layers is a, and the distance between atoms in the same layer is 1, the theta series of the primitive hexagonal array with respect to an atom is, from (13),

$$\tfrac{1}{2}\theta_3(X^{a^2})(\phi_0(X^{1/2}) + \phi_0(X^{3/2})). \quad (21)$$

By stacking layers in the sequence $ABAB\cdots$ we obtain the ordinary *graphite net* (p. 922 of Ref. 12), in which there are two geometrically distinct types of atoms. With respect to an atom which is opposite *atoms* in the two adjacent layers, the theta series of graphite is also given by (21). With respect to an atom which is opposite *holes* in the two adjacent layers, the theta series of graphite is

$$\theta_3(X^{4a^2})(\phi_0(X^{3/2}) + \phi_2(X^{3/2})) + 2\theta_2(X^{4a^2})\phi_2(X^{3/2}). \quad (22)$$

By stacking layers in the order $ABCABC\cdots$ we obtain the *rhombohedral graphite net* (p. 923 of Ref. 12), in which all atoms are geometrically equivalent. The theta series with respect to any atom is

$$\tfrac{1}{4}\phi_0(X^2)(3\theta_3(X^{9a^2}) - \theta_3(X^{a^2})) + \tfrac{1}{4}\phi_0(X^{2/3})(3\theta_3(X^{a^2}) - \theta_3(X^{9a^2})). \quad (23)$$

VI. IONIC CRYSTAL STRUCTURES

To illustrate the application of theta series to binary compounds we consider seven of the most regular ionic crystal structures. The first six are pictured on p. 15 of Ref. 18. We use the most symmetrical (idealized) versions of these structures, in which the parameters of the unit cell are such as to give the highest coordination number. The same techniques may be applied to more general structures, but the resulting theta series are not as simple.

The first four examples all follow the same pattern. We begin with the simplest.

(i) The idealized *rock salt* (NaCl) structure consists of two types of atoms (Na and Cl, or more generally X and Y) placed alternately at the points of the simple cubic lattice. The theta series with respect to an X-type atom is [from (45) and (53) of Ref. 1]

$$
\begin{aligned}
&\Theta_{\text{rock salt},X}(X,Y)\\
&\quad=\Theta_{\text{fcc,atom}}(X)+\Theta_{\text{fcc,oct. hole}}(Y)\\
&\quad=\tfrac{1}{2}(\theta_3(X)^3+\theta_4(X)^3)+\tfrac{1}{2}(\theta_3(Y)^3-\theta_4(Y)^3)\\
&\quad=X^0+6Y+12X^2+8Y^3+6X^4+24Y^5\\
&\qquad+24X^6+12X^8+30Y^9+24X^{10}+\cdots. \qquad (24)
\end{aligned}
$$

The X- and Y-type atoms are geometrically equivalent, so the theta series with respect to a Y-type atom is obtained by interchanging X and Y in (24).

(ii) The *cesium chloride* (CsCl) structure similarly consists of two types of atoms placed alternately at the points of the body-centered cubic (bcc) lattice. The theta series with respect to an X-type atom is [from (37), (43) of Ref. 1]

$$
\begin{aligned}
&\Theta_{\text{CsCl},X}(X,Y)\\
&\quad=\Theta_{Z^3,\text{atom}}(X)+\Theta_{Z^3,\text{hole}}(Y)\\
&\quad=\theta_3(X)^3+\theta_2(Y)^3\\
&\quad=X^0+8Y^{3/4}+6X+12X^2+24Y^{11/4}+8X^3\\
&\qquad+6X^4+24Y^{19/4}+24X^5+24X^6+\cdots. \qquad (25)
\end{aligned}
$$

(iii) The *zinc blende* (ZnS) structure is obtained in the same way from the diamond net. The theta series with respect to an X-type atom is [compare Eq. (17)]

$$
\begin{aligned}
&\Theta_{\text{zinc blende},X}(X,Y)\\
&\quad=\tfrac{1}{2}(\theta_2(Y)^3+\theta_3(X)^3+\theta_4(X)^3)\\
&\quad=X^0+4Y^{3/4}+12X^2+12Y^{11/4}+6X^4+12Y^{19/4}\\
&\qquad+24X^6+16Y^{27/4}+12X^8+24Y^{35/4}+\cdots. \qquad (26)
\end{aligned}
$$

(iv) The *wurtzite* (ZnS) structure[19] is similarly obtained from hexagonal diamond. The theta series with respect to an X-type atom is obtained by replacing X by Y in the second bracketed expression in (20).

(v) In the idealized *fluorite* (CaF_2) structure the two types of atoms are not geometrically equivalent, and so it seems clearest to replace X and Y by the appropriate chemical symbols. With respect to a calcium atom we have [using (45) and (51) of Ref. 1]

$$
\begin{aligned}
&\Theta_{\text{CaF}_2,\text{Ca}}(\text{Ca,F})\\
&\quad=\Theta_{\text{fcc,atom}}(\text{Ca})+2\Theta_{\text{fcc,tet.hole}}(\text{F})\\
&\quad=\tfrac{1}{2}(\theta_3(\text{Ca})^3+\theta_4(\text{Ca})^3)+\theta_2(\text{F})^3\\
&\quad=\text{Ca}^0+8\,\text{F}^{3/4}+12\,\text{Ca}^2+24\,\text{F}^{11/4}\\
&\qquad+6\,\text{Ca}^4+24\,\text{F}^{19/4}+24\,\text{Ca}^6+32\,\text{F}^{27/4}\\
&\qquad+12\,\text{Ca}^8+48\,\text{F}^{35/4}+\cdots \qquad (27)
\end{aligned}
$$

[compare (26)]. The fluorite structure may also be regarded as a simple cubic lattice of F atoms with half the holes filled by Ca atoms. Therefore, with respect to a fluorine atom, we have

$$
\begin{aligned}
&\Theta_{\text{CaF}_2,\text{F}}(\text{Ca,F})\\
&\quad=\Theta_{Z^3,\text{atom}}(\text{F})+\tfrac{1}{2}\Theta_{Z^3,\text{hole}}(\text{Ca})\\
&\quad=\theta_3(\text{F})^3+\tfrac{1}{2}\theta_2(\text{Ca})^3\\
&\quad=\text{F}^0+4\,\text{Ca}^{3/4}+6\,\text{F}+12\,\text{F}^2+12\,\text{Ca}^{11/4}+8\,\text{F}^3\\
&\qquad+6\,\text{F}^4+12\,\text{Ca}^{19/4}+24\,\text{F}^5+24\,\text{F}^6+\cdots. \qquad (28)
\end{aligned}
$$

For *antifluorite* (p. 161 of Ref. 12) we interchange Ca and F in (27) and (28).

(vi) The *rutile* (TiO_2) structure illustrates how less regular structures may be handled. Rutile consists of Ti atoms at the positions of the bcc lattice and O atoms at four translates of the simple cubic lattice Z^3 by the amounts

$$\pm(u,u,0),\ \pm(\tfrac{1}{2}+u,\tfrac{1}{2}-u,0),$$

where $u\approx 0.30$. The theta series of the translate $Z^3+(u,v,w)$ is $\eta_u(X)\eta_v(X)\eta_w(X)$. It follows using (71) of Ref. 1 that

$$
\begin{aligned}
&\Theta_{\text{TiO}_2,\text{Ti}}(\text{Ti,O})\\
&\quad=\theta_2(\text{Ti})^3+\theta_3(\text{Ti})^3\\
&\quad+2\eta_u(\text{O})^2\theta_3(\text{O})+2\eta_{u+1/2}(\text{O})^2\theta_2(\text{O}). \qquad (29)
\end{aligned}
$$

(vii) The final example illustrates how close-packed tetrahedral structures (such as are described on pp. 161 and 162 of Ref. 12) may be handled. The idealized O_3Bi_2 structure consists of Bi atoms at the points of an fcc lattice in which three-quarters of the tetrahedral holes are occupied by O atoms. The theta series with respect to a Bi atom is [using (45) and (51) of Ref. 1, and counting the holes with the correct multiplicity]

$$
\begin{aligned}
&\Theta_{\text{O}_3\text{Bi}_2,\text{Bi}}(\text{O,Bi})\\
&\quad=\Theta_{\text{fcc,atom}}(\text{Bi})+\tfrac{3}{4}\cdot 2\Theta_{\text{fcc,tet. hole}}(\text{O})\\
&\quad=\tfrac{1}{2}(\theta_2(\text{Bi})^3+\theta_3(\text{Bi})^3)+\tfrac{3}{2}\theta_2(\text{O})^3\\
&\quad=\text{Bi}^0+6\,\text{O}^{3/4}+12\,\text{Bi}^2+18\,\text{O}^{11/4}+6\,\text{Bi}^4\\
&\qquad+18\,\text{O}^{19/4}+24\,\text{Bi}^6+16\,\text{O}^{27/4}+12\,\text{Bi}^8\\
&\qquad+24\,\text{O}^{35/4}+\cdots. \qquad (30)
\end{aligned}
$$

The coordination number of 6 agrees with Table 4.5 of Ref. 12.

Erratum of Ref. 1: In Eq. (10) of Ref. 1, the second + sign should be omitted.

ACKNOWLEDGMENT

This project was proposed by my former colleague Boon K. Teo, and I thank him for his suggestions. I regret that circumstances have interrupted our collaboration.

[1]N. J. A. Sloane and B. K. Teo, J. Chem. Phys. **83**, 6520 (1985).
[2]B. K. Teo and N. J. A. Sloane, Inorg. Chem. **25**, 2315 (1986).
[3]M. P. Tosi, Solid State Phys. **16**, 1 (1964).
[4]M. L. Glasser and I. J. Zucker, Theor. Chem. Adv. Perspect. **5**, 67 (1980).
[5]D. Borwein, J. M. Borwein, and K. F. Taylor, J. Math Phys. **26**, 2999 (1985).
[6]Mathlab Group, MACSYMA *Reference Manual* (M.I.T., Laboratory for

Computer Science, Cambridge, MA, 1983), Version 10.

[7]N. J. A. Sloane, Not. Am. Math. Soc. **33**, 40, 318 (1986).

[8]N. J. A. Sloane, IEEE Trans. Inf. Theory **27**, 327 (1981).

[9]J. H. Conway and N. J. A. Sloane, *Sphere-Packings, Lattices and Groups* (Springer, New York, 1987).

[10]E. T. Whittaker and G. N. Watson, *A Course of Modern Analysis* (Cambridge U. P., Cambridge, 1963), 4th ed.

[11]A. F. Wells, *Three-Dimensional Nets and Polyhedra* (Wiley, New York, 1977).

[12]A. F. Wells, *Structural Inorganic Chemistry* (Oxford U. P., Oxford, 1984), 5th ed.

[13]M. O'Keeffe and B. G. Hyde, Philos. Trans. R. Soc. Ser. A **295**, 553 (1980).

[14]B. K. Teo and N. J. A. Sloane, Inorg. Chem. **24**, 4545 (1985).

[15]R. W. G. Wyckoff, *Crystal Structures* (Wiley, New York, 1963), 2nd ed., Vol. 1.

[16]J. A. Prins and H. Petersen, Physica **3**, 147 (1936).

[17]J. O. Hirschfelder, C. F. Curtis, and R. B. Bird, *Molecular Theory of Gases and Liquids* (Wiley, New York, 1954).

[18]F. A. Cotton and G. Wilkinson, *Advanced Inorganic Chemistry* (Wiley, New York, 1980), 4th ed.

[19]M. L. Glasser, J. Phys. Chem. Solids **10**, 229 (1959).

Global qualitative study of Bianchi universes in the presence of a cosmological constant

Edmond Weber[a)]
Université Catholique de Louvain, Institut de Physique Théorique, B-1348 Louvain-la-Neuve, Belgium and Université du Burundi, B. P. 2700 Bujumbura, Burundi

(Received 22 January 1986; accepted for publication 4 March 1987)

A particular class of nonempty spatially homogeneous orthogonal cosmologies of the Bianchi classification is considered in the presence of a nonzero cosmological constant Λ, i.e., cosmologies that can be transformed into a three-dimensional autonomous dynamical system: these are the axisymmetric type II, the pseudoaxisymmetric type VI_0, and the ($n_\alpha^\alpha = 0$) types V, VI_h (including $\mathrm{III} = \mathrm{VI}_{-1}$). A qualitative method to investigate such dynamical systems as a whole is presented. The qualitative study yields for every type a set of solutions of nonzero measure and another one of zero measure becoming isotropic in an infinite cosmological time when $\Lambda > 0$.

I. INTRODUCTION

The discovery of the microwave background radiation in 1965 confirmed the isotropy assumption in the early Friedmann–Lemaître–Robertson–Walker cosmological models. However, it left open the question of why our universe should be isotropic. Several attempts to solve this problem have been done since that time. The most recent is the inflationary model of the Universe,[1] suggested by Guth.[2] In this model, a period of exponential expansion can reduce an initial anisotropy, but it has been shown by Barrow[3] that an initially high anisotropy will prevent the Universe from getting into an exponentially expanding phase. Collins and Hawking showed[4] that the set of spatially homogeneous cosmological models approaching isotropy at infinite times is of zero measure. They concluded therefore that the isotropy of the Universe cannot be explained without postulating special initial conditions.[5] However, they did not consider the possible influence of a nonzero cosmological constant Λ on the isotropy of the Universe. In two previous articles[6,7] we have shown then that for the special class of the anisotropic cosmologies of the Kantowski-Sachs type, there exist sets of solutions of nonzero measure (when they originate at certain singular points, in the context of autonomous dynamical systems, outlined in the following) as well as a set of zero measure which is not contained in the preceding ones (because the models start in this case at a different singular point), all approaching isotropy at infinite times, when a positive cosmological constant is considered. This brings us back to the idea of Misner who suggested[8] that our Universe started off in a chaotic state with inhomogeneities and anisotropies of all kinds. In order to investigate this point of view, Misner has considered[9] the effects of neutrino viscosity in damping out the anisotropy of the homogeneous cosmological model of Bianchi type I. The "dissipative process" we use here is only a nonzero cosmological constant.

The results for the Kantowski–Sachs type, however, have been obtained by studying a three-dimensional autonomous dynamical system, which gives us global qualitative solutions and quantitative asymptotic behaviors around the singularity points. Autonomous dynamical systems are well known in cosmology[10,11] but only in the plane case is the mathematical theory well established.[12,13] The generalization to three dimensions was initiated by Bihari[14] and Couper,[15] who studied the modes of approach to the simplest types[16] of singular points (i.e., where the right-hand sides disappear simultaneously). When the dimension is higher than 3, we have at our disposal a general theorem, stated by Bogoyavlensky,[17] applying to nondegenerate singular points of autonomous dynamical systems of any dimension. Because of its interest it will be given in the Appendix. Autonomous dynamical systems of (at least) three dimensions appear quite naturally in Bianchi cosmologies.[10,17,18] We present here a method to investigate globally three-dimensional autonomous dynamical systems. Let us indicate that our method differs from the one given by Bogoyavlensky.[17,19] We obtain then the evolutions of cosmological models from one singular point to another. In this paper, we will restrict ourselves to a certain class of orthogonal Bianchi models containing a perfect fluid: to the class A in the notation of Ellis and MacCallum,[20] where we analyze the axisymmetric model of Bianchi type II and the pseudoaxisymmetric one of type VI_0; to those of class B, which have $n_\alpha^\alpha = 0$: the models are then of the types III, V, VI_h where $\mathrm{III} = \mathrm{VI}_{-1}$. The other axisymmetric models of class A are either lower- or higher-dimensional autonomous systems. The special case $h = -1/9$ of class B will be considered elsewhere.[18] The technique for the asymptotic behavior around the singularity points for these models has been investigated by MacCallum.[21] The main results in our paper are, as well as the qualitative study of a class of empty Bianchi models with cosmological constant, the existence of sets of zero and especially nonzero measure of solutions becoming isotropic in an infinite cosmological time when $\Lambda > 0$ for all types studied herein. These sets whose scale factor is zero at the beginning of the evolution then asymptotically approach the de Sitter solution, indicated by Wald[22] as being the asymptotic behavior for all Bianchi models when $\Lambda > 0$. Some of the models belonging to the set of zero measure have the additional property of starting at the Einstein–de Sitter

[a)] Present address: Université du Burundi, B. P. 2700 Bujumbura, Burundi.

 0022-2488/87/071658-09$02.50

model. Let us indicate that in this paper we present global qualitative solutions of a particular class of Bianchi models with a particular matter stress-energy tensor whereas Wald gives only the asymptotic behavior but for all Bianchi models without any assumption about the nature of the matter stress-energy tensor. Furthermore our method enables us to conclude that there exists a set of solutions approaching asymptotically the de Sitter one, which is of nonzero measure.

The paper is organized as follows. In Sec. II the global qualitative method to study three-dimensional autonomous dynamical systems is described. In Sec. III we discuss qualitatively empty Bianchi models of the types II, III, V, VI, and VI_h in the presence of a nonzero cosmological constant and in Sec. IV the nonempty models for all these types are studied by means of a three-dimensional autonomous dynamical system in the presence of the same constant. Conclusions are outlined in Sec. V.

II. QUALITATIVE METHOD

Einstein's field equations can be written in the form

$$x' = x[(3\gamma-2)(1-x) - \beta'^2 + \tfrac{2}{3}\Lambda z], \tag{2.1}$$

$$\beta'' = \tfrac{1}{2}\beta'[4 - (3\gamma-2)x - \beta'^2 + (2\Lambda/3)z] - 2^{-1}C(4 - 4x - \beta'^2 - (4\Lambda/3)z), \tag{2.2}$$

$$z' = -2z[1 + \tfrac{1}{2}(3\gamma-2)x + \tfrac{1}{2}\beta'^2 - (\Lambda/3)z], \tag{2.3}$$

with the first integral

$$\beta'^2 = 4 - 4x - \tfrac{4}{3}\Lambda z - z\,e^{2\Omega}V_1, \tag{2.4}$$

for the restricted class of perfect fluid Bianchi models,[10] indicated in Sec. I, in the presence of a nonzero cosmological constant, where $' \equiv d/d\Omega$ is the derivation with respect to the time variable $\Omega = \Omega(t)$, first used by Misner[9,23–25] to study the asymptotic behavior of models of the Bianchi types I and IX. The variables $x = 3\mu/\theta^2$ and $\beta' = 2\sqrt{3}\sigma/\theta$ (where σ is the shear scalar) measure, respectively, the dynamical importance of the matter content and the rate of shear in terms of the volume expansion θ, $z = 9\theta^{-2}$, and V_1 is an effective potential for the model's anisotropy,[21] $\beta \equiv (\beta_1, \beta_2)$, where for axisymmetric models $\beta = \beta_1$ and for $(n_a^a = 0)$ models, $\sqrt{3}\beta_1 = k\beta_2$, for some constant k, β being then defined by $-\beta\sqrt{3+k^2} \equiv k\beta_1 + \sqrt{3}\beta_2$. Here C takes the values 4, -2, 0, $0 < C < 2$ for types II, VI_0, V, and VI_h, respectively, $C = 1$ for type III and $C = 2q/\sqrt{q^2 - 3h}$ for type VI_h (where q is defined for instance in Ref. 20). The equation of state has the barotropic form $p = (\gamma - 1)\mu$, where p is the pressure, μ the density of matter, and γ a constant whose values lie in the range $1 \leqslant \gamma \leqslant 2$. The upper limit $\gamma = 2$ corresponds to Zel'dovich's stiff equation of state.

In order to study the three-dimensional autonomous dynamical system (2.1)–(2.3), we first calculate the singular points at finite distance (called also critical points or equilibrium points or rest points) and examine the behavior of integral curves in their neighborhood. We obtain the singular points at infinite distance by introducing three different Poincaré transformations[6,17]

$$x = us^{-1}, \quad x = s^{-1}, \quad x = us^{-1},$$

$$\beta' = s^{-1}, \quad \beta' = us^{-1}, \quad \beta' = vs^{-1},$$

$$z = vs^{-1}, \quad z = vs^{-1}, \quad z = s^{-1},$$

alternatively in the system (2.1)–(2.3). The first integral (2.4) determines the region of physical interest: we have $\beta'^2 - 4 + 4x + \tfrac{4}{3}\Lambda z < 0$, since V_1 is purely exponential for the models dealt with in this paper. Let us remember[6] that for the Kantowski–Sachs type we had $\beta'^2 - 4 + 4x + \tfrac{4}{3}\Lambda z > 0$. In addition we have $x > 0, z > 0$. As for a plane system, we distinguish between simple (nondegenerate: all the eigenvalues have nonzero real parts) and multiple (degenerate: at least one eigenvalue is zero) singular points. In the first case, two different types appear: saddles and nodes. When two of the three eigenvalues have real parts with the same sign, we have a saddle point. If the three real parts have the same sign, we have a node, attracting (stable) if all the real parts are negative, repelling (unstable) in the other case.

We analyze first a saddle point with the autonomous dynamical system put into the canonical form

$$X' = \lambda X + f, \quad Y' = \mu Y + g, \quad z' = \nu Z + h, \tag{2.5}$$

where λ, μ, ν are (nonzero) eigenvalues and where f, g, h are the three nonlinear terms of the system.[26] The singular point is at the origin of the coordinates (X,Y,Z). Let us assume $\lambda > 0, \mu > 0, \nu < 0$ for the system (2.5). Here, θ, ϕ intervening in what follows correspond to the angles of the usual spherical coordinates ($X = r\sin\theta\cos\phi$, $Y = r\sin\theta\sin\phi$, $Z = r\cos\theta$). The sign of ν is denoted by $\sigma(\nu)$. The following theorem[14,27] gives us then the behavior of the separatrices of the saddle.

Theorem 2.1: If the autonomous system (2.5) has its characteristic roots as assumed, then there is (a) only one orbit tending to the origin with $\sigma(\nu)\Omega \to -\infty$ and $\theta_\infty \equiv \lim_{\sigma\Omega \to -\infty} \theta = 0$ and only one orbit with $\theta_\infty = \pi$; and (b) a curve γ satisfying the following properties: γ and its projection on the plane (X,Y) are homeomorphic to a circle, and every orbit starting at some point of γ is tending to the origin with $\sigma(\nu)\Omega \to +\infty$ and $\lim_{\sigma\Omega \to +\infty} \theta = \pi/2$. The angle ϕ_∞ is fixed according to the well-known theorems of nodes and foci for a plane autonomous system.[12,13,28]

We consider now a node for the same[26] autonomous dynamical system (2.5) and assume $|\nu| < |\mu| \leqslant |\lambda|$.

Theorem 2.2[14]: If the eigenvalues are all of the same sign, there is a sphere centered at the origin such that every orbit starting at its surface tends to the singular point with $\sigma\Omega \to -\infty$.

Theorem 2.3[14]: If the eigenvalues are such as assumed, then every orbit which tends to the origin does go alongside the positive Z axis or the negative one except those orbits which start at a curve γ which is homeomorphic to a circle, as well as its projection on the plane (X,Y); in this case $\theta_\infty = \pi/2$. The same remark as before applies for the angle ϕ_∞.

These two latter theorems show what is essentially new in Bihari's work about nodes in three dimensions: there are double infinities of orbits tending to the singular point (in this case alongside the Z axis), instead of only simple infinities as in the plane case. The generalization of these theorems about the topological structure of nondegenerate critical points to yet higher-dimensional autonomous dynamical systems is given in the Appendix. The case where one of the characteristic roots is zero and the other two of opposite sign, i.e., the case of a multiple or degenerate singular point, has been studied by Couper.[15] In this case the behavior of the integral curves is significantly more complex than for nondegenerate critical points.[29] All the theorems in this section and in the Appendix apply to the case of isolated singular points. When we have continuous sets, i.e., nonisolated singular points, it is sometimes possible to extend the above definitions and theorems to these cases.[30]

In order to arrive at a global picture, we have to join the different equilibrium states, at finite distance and at infinite distance (if the variables of our system extend that far). We do this by analyzing the three surfaces

$$\left\{\frac{dx}{d\Omega}=0\right\},\quad \left\{\frac{d\beta'}{d\Omega}=0\right\},\quad \left\{\frac{dz}{d\Omega}=0\right\},$$

in particular their intersections. We obtain then the sign of, for instance, dz/dx along some orbit and can so derive the global behavior of all the orbits between the singular points.

III. PLANE AUTONOMOUS SYSTEMS: THE EMPTY CASE

When the cosmological constant vanishes, we obtain plane autonomous dynamical systems in the variables x, β' for all Bianchi types under investigation. They have been studied in detail by Collins.[10] For Bianchi type V, $C=0$: we have a plane system in the variables x, z. By setting $x=0$ in (2.1), we obtain a class of empty plane autonomous systems in the presence of a cosmological constant

$$\beta''=\tfrac{1}{2}\beta'(4-\beta'^2+(2\Lambda/3)z)-\tfrac{1}{2}C(4-\beta'^2-(4\Lambda/3)z), \tag{3.1}$$

$$z'=-2z(1+\tfrac{1}{2}\beta'^2-(\Lambda/3)z). \tag{3.2}$$

These two latter cases have not yet been studied by qualitative methods and are very useful as further subcases of the three-dimensional systems in order to obtain a better idea of the global behavior of the integral curves in three dimensions.

For type II models, the critical points at finite distance are simple. The point (2,0) is a saddle point, $(-2,0)$ an improper node, as well as $(0,3/\Lambda)$, which exists only for $\Lambda>0$. When $\Lambda<0$, the Poincaré transformation[12] $(\beta=vs^{-1}, z=s^{-1})$ enables us to study the critical point at infinity ($v=0, s=0$) on the Z-axis, which is a double singular point. We find three directions of approach (in polar coordinates) $\psi=0$, arctan $(-\frac{1}{8})$, and π. For $\Lambda>0$, the typical behavior is the following one: the model starts at the critical point $(-2,0)$ which is a pancake with the one-axis distinguished, where Λ is negligible while the shear σ and the expansion θ are dominant, and tends to the singular point $(0,3/\Lambda)$ (from the point of view of an autonomous system), which we call a "divergent type," since $X\to\infty$ and $Y=Z\to\infty$ for $\Omega\to\infty$, where X, Y, Z are time-dependent functions on which the metric depends,[21] not to be confused with the same symbols used in Sec. II as coordinates. The quantities Λ and θ are dominant and σ is negligible. The critical point $(-2,0)$ is a cosmological singularity because $t\to 0_{\pm}$ and because the curvature invariant $R_{\alpha\beta\gamma\delta}R^{\alpha\beta\gamma\delta}$ diverges as one approaches this point.[31–34] The average length scale $l=e^{-\Omega}$ is proportional to $(\pm t)^{1/3}$. It is not the case for $(0,3/\Lambda)$, for which $t\to\mp\infty$ with $l\sim\exp(\mp\sqrt{\Lambda/3}t)$. When $\Lambda<0$, models start at $(-2,0)$, extending to infinitely large values of z and coming back to the same critical point. One curve is time symmetric. The global picture for $\Lambda>0$ and $\Lambda<0$ is drawn in Fig. 1. The arrows depict the entire course of evolution, but the time reverses are also possible. All the type II models considered here are locally rotationally symmetric (LRS).

The type V models divide into two classes: those in the (x,z) plane and those in the (β',z) plane. In the variables x,z we have three singular points. Two of them, (1,0) and $(0,3/\Lambda)$, are improper nodes and (0,0) is a saddle point. Here $(0,3/\Lambda)$ exists only for $\Lambda>0$. By the Poincaré transformation ($x=vs^{-1}, z=s^{-1}$) we find, for $\Lambda<0$, the double singular points ($s=0$, $0\leqslant v\leqslant-\Lambda/3$). It is a continuous line of nonisolated singular points. The direction of approach is given by $\tan\psi=[(3\gamma-2)v_0-2\Lambda/3]/3\gamma v_0$, where $0\leqslant v_0\leqslant-\Lambda/3$. When $\Lambda>0$, models start at (1,0), which is a point singularity, where matter is dynamically important, and tend to $(0,3/\Lambda)$ which is of the divergent type with the matter dynamically negligible but Λ important. The singular point (1,0) is a cosmological singularity with $l\sim(\pm t)^{2/3\gamma}$. In the case $\Lambda<0$, all the models are time symmetric. The typical behaviors are drawn in Fig. 2.

The empty case in the variables β', z contains three or four simple equilibrium states according to $\Lambda<0$ and $\Lambda>0$.

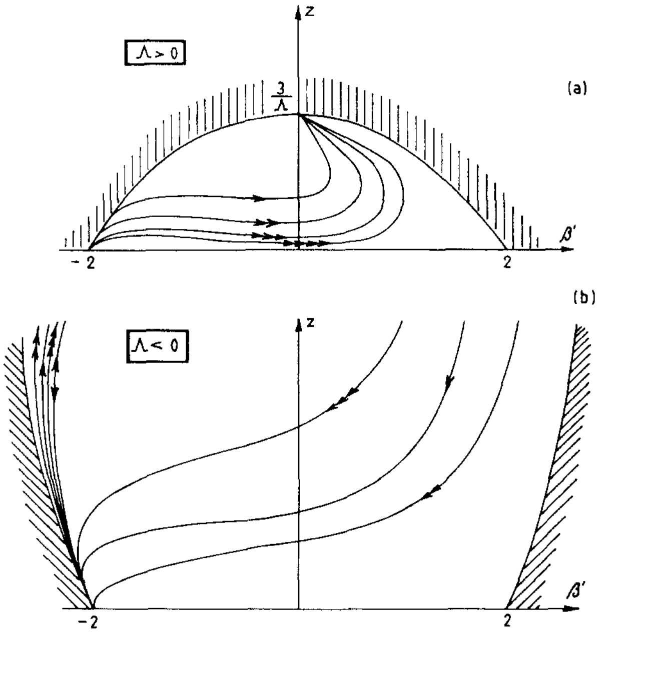

FIG. 1. The entire evolution of the empty Bianchi models with $\Lambda>0$ and $\Lambda<0$ is depicted, in terms of the variables $\beta'=2\sqrt{3}(\sigma/\theta)$ and $z=9\theta^{-2}$. Here x measures the dynamical importance of matter and β' measures the importance of shear anisotropy. When $\Lambda>0$, arrows refer to the evolution of the model with the time reverse also possible. When $\Lambda<0$, there exists one single time symmetric model. The types of arrows indicating the entire evolution are otherwise associate. The general behavior of the integral curves has been found from the qualitative method and then been used to sketch qualitative figures by hand. We describe the trajectories of the LRS($\beta_2\equiv 0$) type II model when $\Lambda>0$ [diagram (a)] and $\Lambda<0$ [diagram (b)].

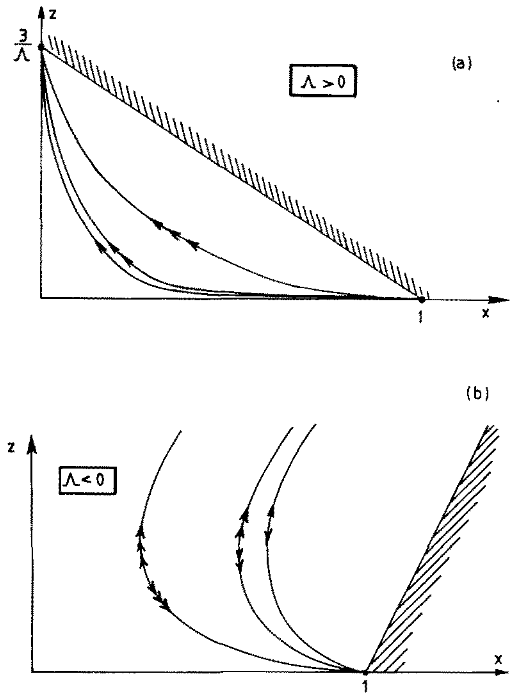

FIG. 2. The plane case in the variables (x,z) for the $(n^{\alpha}_{\alpha} = 0)$ type V model. In (b), all the models are time symmetric. See also the caption to Fig. 1.

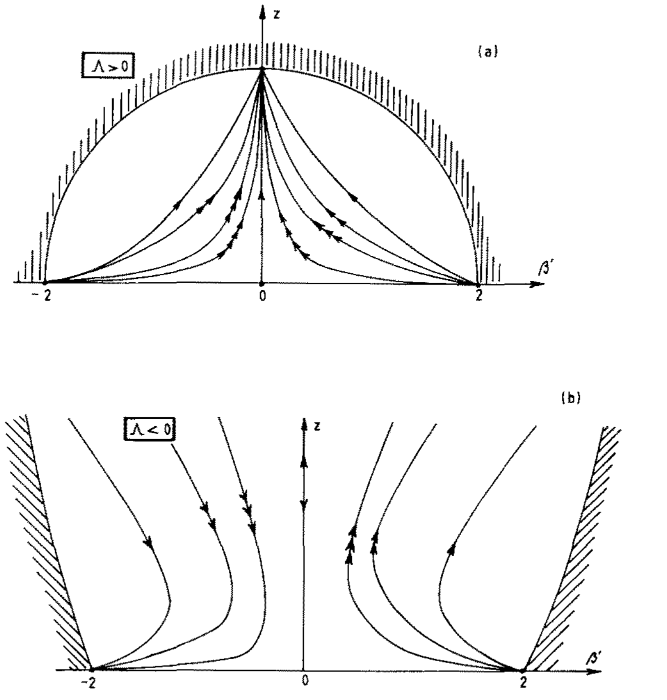

FIG. 3. The plane case in the variables (β',z). See also the caption to Fig. 1.

The singular points (2,0), $(-2,0)$, and $(0,3/\Lambda)$ are improper nodes and (0,0) is a saddle point. At infinitely large values for z, we obtain one double singular point, whose directions of approach are $\psi = 0$, $\pi/2$, and π. As shown in Fig. 3, we have a single model starting at (0,0) and tending to $(0,3/\Lambda)$ when $\Lambda > 0$. The saddle point (0,0) is then a point singularity. All other models evolve from (2,0) or $(-2,0)$, which are of the cigar type with the three- or two-axis preferred, to the divergent type $(0,3/\Lambda)$. When $\Lambda < 0$, there is one time-symmetric model; the other ones evolve from one node to the other via the singular point at infinity.

The pseudoaxisymmetric type VI_0 empty models with $\Lambda = 0$ have globally behaviors similar to those of type II (Fig. 4). These behaviors are in fact corresponding as in the case[10] of the nonempty models with a vanishing cosmological constant. The singular point (2,0) is now an improper node which is cigar shaped with the one-axis preferred, and $(-2,0)$ is a saddle point. The directions of approach at infinity are $\psi = 0$, $\arctan\frac{1}{4}$, and π.

The type VI_h models investigated herein have $0 < k < \infty$, $k \neq 3$ ($k = 3$ corresponds to the type BbII) and consequently $0 < C < 2$ ($C \neq \sqrt{3}$) where $C = 2k/\sqrt{3 + k^2}$. A particular case is Bianchi III, for which $k = C = 1$. The plane autonomous system for this latter type is the same as for the Kantowski–Sachs models[6]; the only difference resides in the physical region: we have $\beta'^2 + \frac{4}{3}\Lambda z - 4 < 0$ for type III whereas this expression is positive for the Kantowski–Sachs cosmologies. There are four simple equilibrium states. Here (2,0) and $(0,3/\Lambda)$ are always improper nodes while $(-2,0)$ is an improper node except for the value $C = 1$, for which it is a proper node. The critical point $(C,0)$ is a saddle point. The directions of approach are shown in Figs. 5 and 6. When $\Lambda < 0$, $(0,3/\Lambda)$ does not exist but there is then a double singular point at infinity whose directions of approach are $\psi = 0$, $\arctan(-1/2C)$, π. When $\Lambda > 0$, there is one single model evolving from $(C,0)$ to $(0,3/\Lambda)$, which is of the divergent type while $(C,0)$ is a point singularity when $0 < C < 1$, a barrel with the three-axis preferred for $C = 1$, and a cigar with the three-axis preferred when $1 < C < 2$. Models start from the cigar-shaped singular-

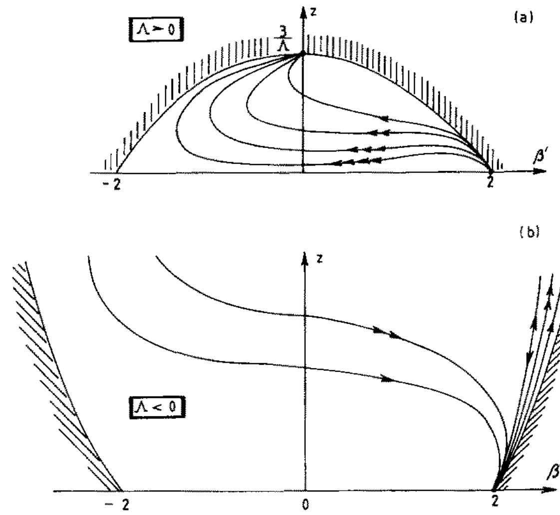

FIG. 4. The global behavior of the axisymmetric $(\beta_2 \equiv 0)$ type VI_0 model. See also the caption to Fig. 1.

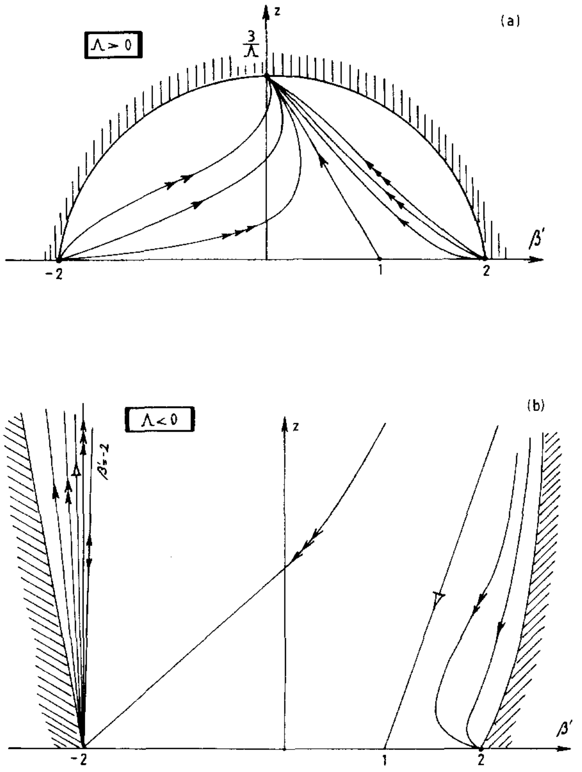

FIG. 5. The particular case $\text{III} = \text{VI}_{-1}$ for which $C = 1$. See also the caption to Fig. 1.

ity (2,0) with the three-axis preferred, as well as from ($-$ 2,0) being a cigar with the two-axis preferred, a pancake with the three-axis preferred, or a cigar with the one-axis preferred according to $k<1$, $k=1$, or $k>1$, tending all to $(0,3/\Lambda)$. When $\Lambda<0$, one single model evolves from $(C,0)$ to $(-2,0)$; there is one time-symmetric solution, $\beta' = -2$ when $C=1$. Typical behaviors are models which evolve from $(-2,0)$ to infinity and come back either to $(-2,0)$ or (2,0).

IV. NONEMPTY BIANCHI TYPES WITH A COSMOLOGICAL CONSTANT

We will follow the method outlined in Sec. II to study the three-dimensional autonomous dynamical system (2.1)–(2.3) with first integral (2.4) according to the values of C. The region of physical interest is $[x>0, z>0, \beta'^2 - 4 + 4x + (4\Lambda/3)z < 0]$. The singular points $(0, \pm 2, 0)$ and $(1,0,0)$ are simple and exist for all types, when $1 \leqslant \gamma < 2$, and for Λ positive or negative (Fig. 7). The topological type of $(x=0, \beta'=2, z=0)$ is a node except for Bianchi type II for which it is a saddle point; the characteristic roots $\lambda = 3\gamma - 6$, $\mu = -4+2C$, $\nu = -6$ are negative except for $C=4$. We have three corresponding characteristic vectors $l_1 = (1,-1,0)$, $l_2 = (0,1,0)$, and $l_3 = (0,-1,3/\Lambda)$ when $\Lambda>0$ [or $(0,1,-3/\Lambda)$ when $\Lambda<0$] in all cases, except for values of C in the range $\frac{1}{2} \leqslant C \leqslant 2$, where we have

$$1 \leqslant \gamma < (2+2C)/3: |\mu| < |\lambda| < |\nu|, \tag{4.1}$$

$$\gamma = (2+2C)/3: |\mu| = |\lambda| < |\nu|, \tag{4.2}$$

$$(2+2C)/3 < \gamma < 2: |\lambda| < |\mu| < |\nu|. \tag{4.3}$$

When $\gamma = (2+2C)/3$, we have an infinity of characteristic directions in the plane (X,Y). No physical orbits tend to (0,2,0) for Bianchi type II; there is a double infinity of orbits starting at a sphere centered at (0,2,0) (see Theorem 2.2) and tending to this point $\Omega \to \infty$, alongside the vector l_1, except for the cases (4.1) and (4.2) for which the orbits tend to (0,2,0) alongside l_2 and along the plane (X,Y), respectively (Fig. 8).

The point $(0,-2,0)$ is a node with negative eigenvalues

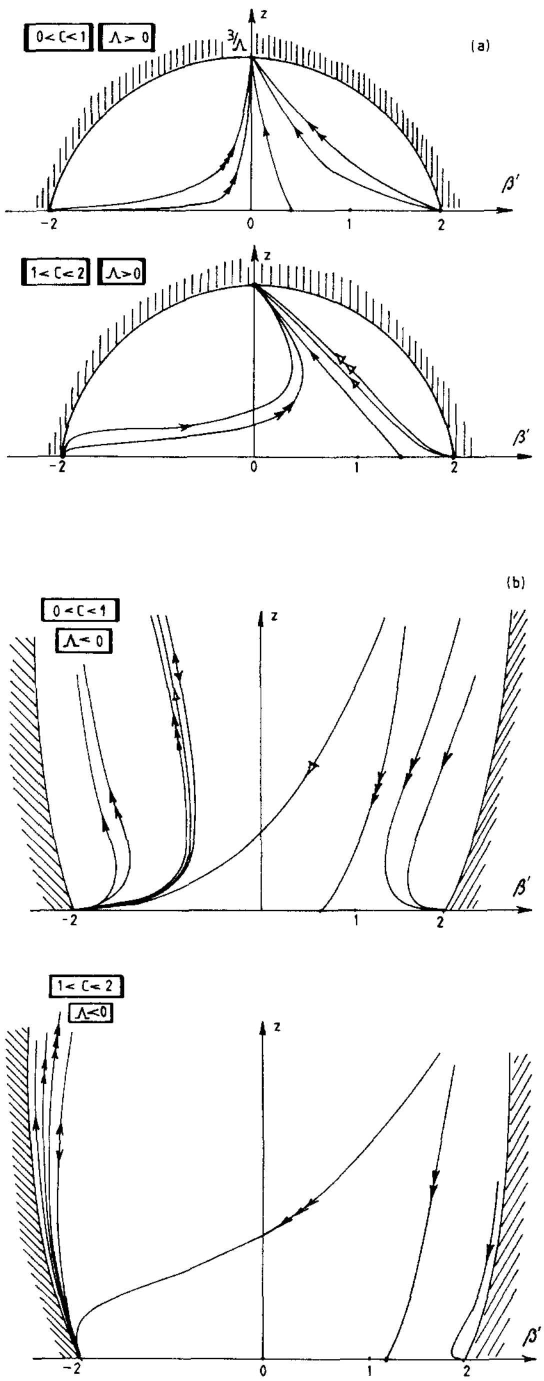

FIG. 6. The $(n^a_\alpha = 0)$ type VI_h $(h \neq 0)$ model, whose behavior differs as $0<C<1$ or $1<C<2$. See also the caption to Fig. 1.

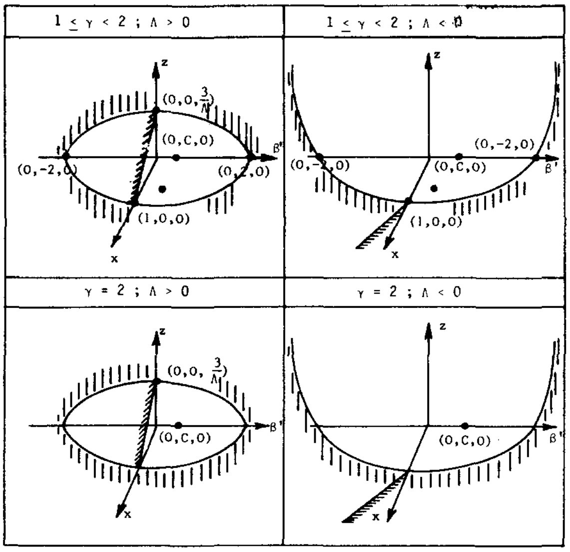

FIG. 7. The region of physical interest for the three-dimensional autonomous system, as well as the singular points at finite distance according to the value of γ and of the cosmological constant Λ are depicted. Striped parts are outside the region of physical interest.

$\lambda = 3\gamma - 6$, $\mu = -4 - 2C$, $\nu = -6$, except for the Bianchi type VI_0. In this case the eigenvalue $\mu = 0$ and $(0, -2, 0)$ behaves as a saddle point, with no physical orbits tending to it. A double infinity of orbits tends to $(0, -2, 0)$ alongside the vector l_1 for all other types.

The singular point $(1,0,0)$ is a saddle point for all types with eigenvalues $\lambda = -3\gamma + 2$, $\mu = 3 - (\frac{3}{2})\gamma$, $\nu = -3\gamma$ and with corresponding eigenvectors $l_1 = (1, -4C/(2+3\gamma), 0)$, $l_2 = (0,1,0)$, $l_3 = (\mp 1, 0, \pm 3/\Lambda)$ according to $\Lambda > 0$ or $\Lambda < 0$. We find a simple infinity of orbits tending to $(1,0,0)$ along the vector l_1 (see Theorem 2.1).

In the plane (x, β') we have two singular points $(0,C,0)$ and $((C^2 - 3\gamma + 2)/C^2, (3\gamma - 2)/C, 0)$, which do not exist for all types. The point $(0,C,0)$ is a saddle point for all types and for $\gamma \in [1,2]$ except when $C = 4$; in this case it does not exist. When $C = -2$, it coincides with the singular point $(0, -2, 0)$. The eigenvalues are $\lambda = 3\gamma - 2 - C^2$, $\mu = 2 - C^2/2$, $\nu = -2 - C^2$ and the corresponding vectors $l_1 = (1, (-3C + 3C\gamma/2)/(4 - 3\gamma + C^2/2), 0)$, $l_2 = (0,1,0)$, $l_3 = (0, \mp 1, \pm (4 + C^2/2)/\Lambda C)$ according to $\Lambda > 0$ or $\Lambda < 0$. To $C = 0$ corresponds Bianchi type V and $l_3 = (0,0,1)$. An interesting behavior occurs for $C \in]1,2[$ and $\gamma < (C^2 + 2)/3$; in this case there is a simple infinity of orbits tending to $(0,C,0)$ along the vector l_1.

The critical point $((C^2 - 3\gamma + 2)/C^2, (3\gamma - 2)/C, 0)$ exists only when $\gamma \neq 2$ and for Bianchi types II, VI_0, and for VI_h when $\gamma < (C^2 + 2)/3$. This latter inequality happens only when $1 < C < 2$. This critical point is a saddle point with two complex eigenvalues. There is one single orbit along the vector $l_3 = (0,0,1)$. The results concerning the behavior of the solutions in the neighborhood of these two singular points are valid for $\Lambda > 0$ and $\Lambda < 0$.

When $\gamma = 2$ there exists a continuous line of singular points in the plane (x, β'): $(4 - 4x - \beta'^2 = 0, z = 0)$. In this case we have a simple infinity of orbits tending to each singular point along its eigenvector in this plane.[30,35]

There remains finally in the plane (β', z) the singular point $(0,0,3/\Lambda)$ when $\Lambda > 0$ and $\gamma \in [1,2]$. It is a node with eigenvalues $\lambda = 3\gamma$, $\mu = 3$, $\nu = 2$ and corresponding eigenvectors $l_1 = (1, 0, -3/\Lambda)$, $l_2 = (0,1,0)$, $l_3 = (0, 1, -3/2\Lambda C)$ for $C \neq 0$, $l_3 = (0,0,1)$ for Bianchi type V.

This concludes the study of behaviors around singular points at finite distance (Table I). When $\Lambda < 0$, by setting $x = us^{-1}$, $\beta = vs^{-1}$, $z = s^{-1}$ in (2.1)–(2.4) we find a continuous line of double singular points at infinite distance: $(s = 0,\ v = 0,\ 0 \leqslant u \leqslant -\Lambda/3)$ which are not in the plane (x, β'). The general expression of the directions of approach, not in the plane $(s = 0)$ of the singular points $(s = 0, v = 0, u = u_0)$, is given then by

$$\{\omega_i\} = \left(\left[2(3\gamma - 2)u_0 - \frac{4\Lambda}{3}\right]^{-1}, \frac{3\gamma u_0}{2[(3\gamma - 2)u_0 - 2\Lambda/3]^2}, \frac{2C(u_0 + \Lambda/3)}{[(3\gamma - 2)u_0 - 2\Lambda/3]^2}\right).$$

One should notice that $\tan \psi = \omega_1/\omega_3 = -1/2C$, when $u_0 = 0$, in accordance with all plane autonomous systems in the variables β', z; when $u_0 = -\Lambda/3$, $\tan \psi \to \infty$.

By analyzing the three surfaces $\{dx/d\Omega = 0\}$, $\{d\beta'/d\Omega = 0\}$, $\{dz/d\Omega = 0\}$ we obtain a global picture of the orbits. For Bianchi type II models, we distinguish between four different cases: $(1 \leqslant \gamma < 2, \Lambda > 0)$; $(1 \leqslant \gamma < 2, \Lambda < 0)$; $(\gamma = 2, \Lambda > 0)$; $(\gamma = 2, \Lambda < 0)$. In the first case, we have a double infinity of orbits starting at $(0, -2, 0)$, a simple infinity starting at $(1,0,0)$, and one single orbit at $((18 - 3\gamma)/16, (3\gamma - 2)/4, 0)$ all tending toward $(0,0,3/\Lambda)$. The singular point $(0, -2, 0)$ is a pancake with the one-axis distinguished; the two equilibrium states $(1,0,0)$ and $((18 - 3\gamma)/16, (3\gamma - 2)/4, 0)$ are point singularities and $(0,0,3/\Lambda)$ is, as before, of divergent type. Around $(0, -2, 0)$ matter is dynamically negligible, becoming important during the evolution and again negligible around $(0,0,3/\Lambda)$. All these models isotropize in an infinite cosmic time as they tend to $(0,0,3/\Lambda)$ which is the de Sitter model in the Stabell–Refsdal[36] formalism. The point $(1,0,0)$ corresponds to the Ein-

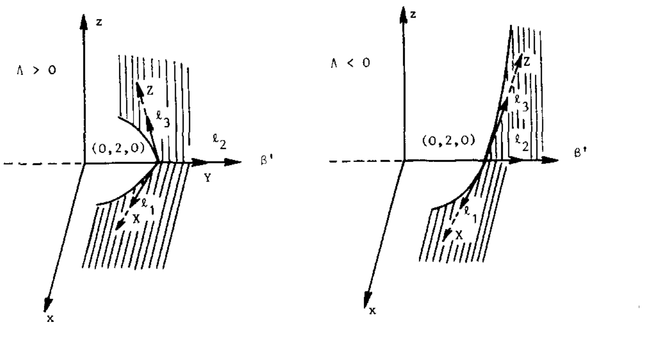

FIG. 8. We indicate the region of physical interest which is inside the striped part in the neighborhood of the critical point $(0,2,0)$ for $\Lambda > 0$ and $\Lambda < 0$. Here l_1, l_2, l_3 are the three eigenvectors.

TABLE I. In this table the different topological types of the singularity points are indicated for the Bianchi types studied herein.

Bianchi type	II	V	VI_0	VI_h
C	4	0	-2	$0<C<2$
$1\leqslant\gamma<2$ (x,β',z)				
$(0,2,0)$	saddle point	node	node	node
$(0,-2,0)$	node	node	saddle point	node
$(1,0,0)$	saddle point	saddle point	saddle point	saddle point
$((C^2-3\gamma+2/C^2),(3\gamma-2)/C,0)$	saddle point		saddle point	$1<C<2$ saddle point
$1\leqslant\gamma\leqslant2$ $(0,C,0)$		saddle point	saddle point	saddle point
$\Lambda>0$ $(0,0,3/\Lambda)$	node	node	node	node

stein–de Sitter (EdS) model. When $\gamma=2$ and $\Lambda>0$ we have a simple infinity of orbits coming from each singular point in the plane (x,β') and tending to $(0,0,3/\Lambda)$. All these models evolve from $(0\leqslant\epsilon_0^2<2,\ q_0=2,\ 0<\Sigma_0\leqslant\frac{1}{2})$ to the de Sitter model. The variables $\epsilon_0^2,q_0,\Sigma_0$ are, respectively, the relative root-mean-square deviation from isotropy, the deceleration parameter, and the density parameter of the perfect fluid at a given time $t=t_0$. For $\Lambda<0$ and $1\leqslant\gamma<2$, a time-symmetric surface of orbits approaches the singular points at infinity with $\psi=\arctan\omega_1/\omega_3$; when $u_0=0$, the orbit is identical to the time-symmetric one in the (β',z) plane case and by $(-4+4x+(4\Lambda/3)z=0)$ when $u_0=-\Lambda/3$. All the other orbits start from $(0,-2,0)$ alongside the vector l_1. There is further a double infinity of orbits starting at $(0,-2,0)$ extending to infinity with $\psi=\pi$ and coming back to the same critical point. There exists finally a simple infinity of orbits starting at $(1,0,0)$ and a single one at $((18-3\gamma)/16,\ (3\gamma-2)/4,\ 0)$ extending to infinity with $\psi=0$ and tending to $(0,-2,0)$ alongside l_1. In the last case, a simple infinity of orbits come from each singular point in the plane (x,β') and tend to the singular line at infinity with $\psi=0$ and π when $-2\leqslant\beta'\leqslant0$; there is then also a time-symmetric surface of orbits approaching the singularities at infinity with $\psi=\arctan\omega_1/\omega_2$. When $0<\beta'\leqslant2$, there is only a simple infinity tending to the singular line at infinity with $\psi=0$.

The global behavior of Bianchi type V models in three dimensions can be divided into the same four cases as for the type II. We obtain the same symmetric picture that we have already found in the two plane systems (x,β') and (β',z). From the points $(0,\pm2,0)$ starts a double infinity of orbits and from $(1,0,0)$ a simple infinity, tending all to $(0,0,3/\Lambda)$ with x,β',z being finite for the whole evolution when $\Lambda>0$ and $1\leqslant\gamma<2$. The simple infinity evolving from $(1,0,0)$ to $(0,0,3/\Lambda)$ is the same as in the plane system with variables x, z. It corresponds to a set of zero measure of EdS models evolving to the de Sitter solution. The points $(0,\pm2,0)$ are cigar shaped with the three- and two-axis distinguished, respectively. We have the same behavior as for type II when $\gamma=2$ and $\Lambda>0$. The singularity types in the plane (x,β') are the same as in Ref. 10. For $\Lambda<0$ and $1\leqslant\gamma<2$, we have a time-symmetric surface of orbits starting at $(1,0,0)$ and evolving toward the singular line at infinity with $\psi=\pi/2$: this is the plane case (x,z) with $\Lambda<0$, studied previously (in Sec. III). All the other orbits start as a double infinity at $(0,2,0)$, tend to the line at infinity with $\psi=0$, and finally toward $(0,-2,0)$. When $\gamma=2$, we have simple infinities starting from each singular point in the plane (x,β'). The behavior at infinity is the same as for $\gamma\neq2$.

For Bianchi type VI_0 models, we obtain an identical behavior with type II. The only differences reside in the cigar-shaped point $(0,2,0)$ from which starts a double infinity of orbits along l_1 and in the fact that there exists for $\gamma=2$ and $\Lambda<0$ a time-symmetric surface of orbits when $0\leqslant\beta'<2$.

The Bianchi type VI_h models divide into the same four cases as before. In the first case, we distinguish between several subcases according to the values of C and γ. When $\Lambda>0$, $1\leqslant\gamma<2$, and $0<C<\frac{1}{2}$, we have a double infinity of orbits starting at $(0,\pm2,0)$ alongside l_1 and a simple infinity starting at $(1,0,0)$, all tending to $(0,0,3/\Lambda)$. When $\frac{1}{2}\leqslant C\leqslant1$, we find the same behavior except around $(0,2,0)$. In that case for $\gamma<(2+2C)/3$, the double infinity evolves alongside l_2; for $\gamma=(2+2C)/3$, there is a simple infinity in every direction of the plane (X,Y), and for $\gamma>(2+2C)/3$, there is a double infinity alongside l_1. The Bianchi type III is included in this global behavior (for $C=1$). When $1<C<2$, we distinguish three subcases according to $\gamma<,=,>(C^2+2)/3$. In the first subcase, we find in addition to the family of orbits around $(1,0,0)$ and $(0,-2,0)$ discussed previously a double infinity of orbits starting at $(0,2,0)$ along the vector l_2, a simple infinity from $(0,C,0)$ along l_1, and a single orbit from $([C^2-(3\gamma-2)]/C^2,\ (3\gamma-2)/C,\ 0)$, all tending toward $(0,0,3/\Lambda)$. For the second and third subcases, the three typical behaviors around $(0,2,0)$ occur. When $\gamma=2$, we have an identical behavior with the previous types. With a negative cosmological constant, $1\leqslant\gamma<2$ and $0<C<\frac{1}{2}$, we find a behavior around $(0,-2,0)$ identical with the case of Bianchi type III. We have a double infinity around $(0,2,0)$ and a simple infinity around $(1,0,0)$ tending toward $(0,-2,0)$ via the singular line at infinity. For $\frac{1}{2}\leqslant C\leqslant1$, the typical behavior occurs around $(0,2,0)$. When $1<C<2$, we distinguish again between three subcases according to the value of γ. Besides typical behaviors, we have to add the simple infinity starting at $(0,C,0)$ and the single orbit at $([C^2-(3\gamma-2)]/C^2,\ (3\gamma-2)/C,\ 0)$ extending to infinity and tending to $(0,-2,0)$. For the second and third subcases, we have to mention again the triple behavior around $(0,2,0)$. For $\gamma=2$, we have an identical behavior as for Bianchi type II. The singularity types for $(0,\pm2,0)$ and $(1,0,0)$ are as mentioned in Ref. 10. The point $(0,C,0)$ is of the cigar type with the three-axis preferred in the case we are interested in.

V. CONCLUSION

We have carried out a detailed qualitative but global analysis of some orthogonal Bianchi models in the presence of a cosmological constant by means of autonomous dynamical systems with two and three dimensions.

The plane autonomous systems gave us a class of empty Bianchi cosmologies, with $\Lambda > 0$ and $\Lambda < 0$, which have not yet been analyzed by qualitative methods and which are important insofar as there do not exist any known exact solutions for all of these models.[37]

When we considered the three-dimensional autonomous systems, we found for every Bianchi type in addition to a set of zero measure, one of nonzero measure of solutions becoming isotropic in an infinite cosmic time, i.e., tending to the de Sitter model. This is an important result regarding the isotropy of the present universe because the double infinity of solutions found in this paper are global ones, although qualitative with quantitative asymptotic behavior. By this we particularize the results of Collins and Hawking[4] who showed that the set of spatially homogeneous cosmological models approaching isotropy at infinite times is of zero measure. However they did not consider a nonzero Λ, which is essential in our discussion and implies an open set of cosmologies approaching asymptotically the de Sitter solution, and without postulating any special initial conditions. The present work agrees also with the one of Wald,[22] done from a more general point of view but with different techniques, and throws further light on the qualitative but global behavior of Bianchi models with a cosmological constant.

It is clear that we have here a method to be applied to other (orthogonal) Bianchi models when Λ is nonzero, i.e., Bianchi IX, in order to obtain global behaviors, which will be reported on elsewhere.[18]

ACKNOWLEDGMENTS

We would like to thank Dr. J. Demaret (University of Liège) for helpful discussions and comments. We are grateful to Dr. A. Moussiaux and Ph. Tombal (University of Namur) for their help in computation.

We acknowledge the considerable support of the Institute of Theoretical Physics (FYMA) at Louvain-la-Neuve (Belgium) during the final elaboration of this work.

APPENDIX: THEOREMS FOR MULTIDIMENSIONAL AUTONOMOUS DYNAMICAL SYSTEMS[17]

Consider an n-dimensional autonomous dynamical system in the form

$$\frac{dX_i}{d\Omega} = f_i(X_1,...,X_n), \quad 1 \leqslant i \leqslant n. \tag{A1}$$

Definition 1: A point $(X_1^0,...,X_n^0)$ is a singular point of the system (A1) if $f_i(X_1^0,...,X_n^0) = 0$, $1 \leqslant i \leqslant n$.

Consider now the eigenvalues $\lambda_1,...,\lambda_n$ of the system (A1) at the singular point $(X_1^0,...,X_n^0)$.

Definition 2: A singular point is called nondegenerate if all the eigenvalues λ_i have nonzero real parts.

Definition 3: A nondegenerate singular point is an attracting (repelling) node, if $\operatorname{Re} \lambda_i < 0$ (>0), $1 \leqslant i \leqslant n$.

Theorem 1: If we have an attracting (repelling) node, then all trajectories in the neighborhood of the singular point approach the point for $\Omega \to {}^{+}_{(-)} \infty$.

Theorem 2: (a) If at some singular point $(X_1^0,...,X_n^0)$ the system (A1) has m eigenvalues $\lambda_1,...,\lambda_m$ with negative (positive) real parts $\alpha_1 \leqslant \cdots \leqslant \alpha_m < 0$ $(\alpha_1 \geqslant \cdots \geqslant \alpha_m > 0)$ where the eigenvalues are counted with their multiplicities, then there exists a (locally) invariant m-dimensional stable (unstable) manifold $M^m_{s(u)}$, on which all trajectories of the system (A1) approach (leave) the critical point for $\Omega \to \infty$.

(b) If there is only one eigenvalue λ_m with the maximum (minimum) negative (positive) real part α_m then the corresponding eigenvector is tangent to almost all trajectories on the invariant manifold $M^m_{s(u)}$ (where by invariant manifold we mean one such that each trajectory passing through some nonsingular point on M lies entirely in M, i.e., $-\infty < \Omega < +\infty$).

Definition 4: A nondegenerate singular point is called a saddle if at this point the system (A1) has m eigenvalues with negative real parts and $n-m$ eigenvalues with positive real parts.

Theorem 3: There exist two invariant manifolds M^m_s and M^{n-m}_u passing through a saddle and filled with separatrices approaching or leaving this singular point. All other trajectories not lying on these manifolds do not approach the saddle.

[1] A. D. Linde, Rep. Prog. Phys. **47**, 925 (1984).
[2] A. H. Guth, Phys. Rev. D **23**, 347 (1981).
[3] J. D. Barrow, Quart. J. R. Astron. Soc. **23**, 344 (1982).
[4] C. B. Collins and S. W. Hawking, Astrophys. J. **180**, 317 (1973).
[5] S. W. Hawking and J. C. Luttrell, Phys. Lett. B **143**, 83 (1984).
[6] E. Weber, J. Math. Phys. **25**, 3279 (1984).
[7] E. Weber, J. Math. Phys. **26**, 1308 (1985).
[8] C. W. Misner, in *Battelle Rencontres, 1967 Lectures in Mathematics and Physics* (Benjamin, New York, 1968).
[9] C. W. Misner, Astrophys. J. **151**, 431 (1968).
[10] C. B. Collins, Commun. Math. Phys. **23**, 137 (1971).
[11] M. A. H. MacCallum, in *General Relativity: An Einstein Centenary Survey*, edited by S. W. Hawking and W. Israël (Cambridge U. P., Cambridge, 1979), and references therein.
[12] A. A. Andronov, E. A. Leontovich, I. I. Gordon, and A. G. Maier, *Qualitative Theory of Second-Order Dynamic Systems* (Wiley, New York, 1973).
[13] V. V. Nemytskii and V. V. Stepanov, *Qualitative Theory of Differential Equations* (Princeton U. P., Princeton, NJ, 1960).
[14] I. Bihari, Studia Sci. Math. Hungar. **1**, 261 (1966).
[15] G. Couper, J. Differ. Eqs. **10**, 27 (1971).
[16] In Sec. II we will give the exact meaning of "simplest" types.
[17] O. I. Bogoyavlensky, *Methods in the Qualitative Theory of Dynamical Systems in Astrophysics and Gas Dynamics* (Springer, Berlin, 1985).
[18] E. Weber and J. Demaret, in progress.
[19] The difference lies in the manner to relate the different singular points. Our method will be presented in Sec. II.
[20] G. F. R. Ellis and M. A. H. MacCallum, Commun. Math Phys. **12**, 108 (1969).
[21] M. A. H. MacCallum, Commun. Math. Phys. **20**, 57 (1971).
[22] R. M. Wald, Phys. Rev. D **28**, 2118 (1983).
[23] C. W. Misner, Phys. Rev. Lett. **22**, 1071 (1969).
[24] C. W. Misner, Phys. Rev. **186**, 1319 (1969).
[25] C. W. Misner, in *Relativity*, edited by M. Carmeli, S. I. Fickler, and L. Witten (Plenum, New York, 1970).
[26] We restrict ourselves to this canonical form, because it is the only one of interest in the context of this paper. A discussion about other canonical

forms, for nondegenerate critical points, with eventually complex eigenvalues may be found in Ref. 14.

[27]The wording of the theorem is adapted to our needs and so is somewhat different from the one given in Ref. 14.

[28]See also Theorems 2 and 3 of the Appendix in Ref. 6.

[29]For more information about it and the generalization to n-dimensional autonomous dynamical systems we refer the reader to the Appendix of Ref. 6 and to Ref. 17.

[30]We will come back to this in Sec. IV, in the context of the three-dimensional autonomous systems investigated in this paper.

[31]We have done the computation by MACSYMA (Ref. 32) for Bianchi type II as well as for the other types V, VI_0, VI_h.

[32]*MACSYMA—Reference Manual* (Symbolics, Inc., M.I.T., Cambridge, MA).

[33]The divergence of the curvature invariant shows us that we have a curvature singularity at this tpoint (Ref. 34). This seems to us to be an adequate guide in these (more difficult) empty cases in order to detect real singulasities.

[34]G. F. R. Ellis and B. G. Schmidt, Gen. Relativ. Gravit. **8**, 915 (1977).

[35]In fact, the critical set has dimension 1 (k in general) and is nondegenerate, because at almost all critical points of this set there are two eigenvalues (n-k in general) with nonzero real parts. In this case, the definitions and theorems of the Appendix apply (Ref. 17).

[36]R. Stabell and S. Refsdal, Mon. Not. R. Astron. Soc. **132**, 379 (1966).

[37]D. Kramer, H. Stephani, R. Herlt, M. A. H. MacCallum, and E. Schmutzer, *Exact Solutions of Einstein's Field Equations* (Cambridge U. P., Cambridge, 1980).