Branching rules for sp(2N) algebra reduction on the chain sp(2N − 2)×sp(2)

Marcin Cerkaski
Department of Theoretical Physics, Institute of Nuclear Physics, Radzikowskiego 152, 31-342 Kraków 23, Poland and Joint Institute for Nuclear Research, Dubna, USSR

(Received 24 June 1986; accepted for publication 29 October 1986)

In this paper branching rules for the reduction sp(2N) ↘ sp(2N − 2)×sp(2) are found, and a new pattern for labeling vectors belonging to the base for unitary irreducible representations BUIR's is found.

I. INTRODUCTION

A few years ago the so-called missing label problem in the algebra reductions

$$\mathrm{sp}(2N) \searrow \mathrm{sp}(2N-2)\times A^N_N \tag{1a}$$

and

$$\mathrm{sp}(2N) \searrow \mathrm{sp}(2N-2)\times \mathrm{sp}(2) \tag{1b}$$

was investigated.[1] Here A^N_N is an operator from the Cartan subalgebra sp(2N). Two different solutions on the set of the missing label operators were obtained, but only one solution, referred to here as (b) [see (1a) and (1b)], contains the Casimir of the sp(2) subalgebra

$$C^2_{(N)} = 2(A^N_N)^2 + A^{-N}_N A^N_{-N} + A^{-N}_N A^N_{-N}, \tag{2}$$

where A^a_b are the algebra generators.

The problem of the branching rules for the reduction (1a) was solved many years ago by Zhelobenko.[2] Denoting shortly by $|(\Omega)\rangle$ the vectors belonging to the base for unitary irreducible representations (BUIR's) we have

$$|(\Omega)\rangle = \left|\begin{array}{ccccccccccccccccccc} \Omega^N_1 & & \Omega^N_2 & & \Omega^N_3 & & \cdots & & & & & & & & \Omega^N_N \\ & \overline{\Gamma}^N_1 & & \overline{\Gamma}^N_2 & & \overline{\Gamma}^N_3 & & \cdots & & \overline{\Gamma}^N_{N-1} & & & & & \overline{\Gamma}^N_N \\ & & \Omega^{N-1}_1 & & \Omega^{N-1}_2 & & \Omega^{N-1}_3 & & \cdots & & \Omega^{N-1}_{N-1} & & & & \\ \cdots & & & & & & & & & & & & & & \\ & & & & & & & & & & & & \Omega^1_1 & & \\ & & & & & & & & & & & & & \overline{\Gamma}^1_1 & \end{array}\right\rangle, \tag{3a}$$

where the first row labels the UIR of the sp(2N) algebra, the third row labels the UIR of the sp(2N − 2) algebra, and so on. The eigenvalue for operator A^k_k is denoted by h_k and we have

$$h_k = \sum_{p=1}^{k} \Omega^k_p + \sum_{p=1}^{k-1} \Omega^{k-1}_p - 2\sum_{p=1}^{k} \overline{\Gamma}^k_p . \tag{3b}$$

All numbers entering to the pattern (3a) are positive integers and the branching rules for the reduction (1a) are contained in the systems of inequalities

$$\Omega^k_1 \geqslant \overline{\Gamma}^k_1 \geqslant \Omega^k_2 \geqslant \overline{\Gamma}^k_2 \geqslant \cdots \geqslant \Omega^k_k \geqslant \overline{\Gamma}^k_k \geqslant 0 ,$$
$$\text{for } k = N, N-1, \ldots, 1 , \tag{3c}$$
$$\overline{\Gamma}^k_1 \geqslant \Omega^{k-1}_1 \geqslant \overline{\Gamma}^k_2 \geqslant \Omega^{k-1}_2 \geqslant \cdots \geqslant \Omega^{k-1}_{k-1} \geqslant \overline{\Gamma}^k_k ,$$
$$\text{for } k = N, N-1, \ldots, 2 . \tag{3d}$$

Hence we see that the same representation of the subalgebra sp(2k − 2)×A^k_k may be found more than once in the $(\Omega^k_1, \Omega^k_2, \ldots, \Omega^k_k)$ representation of the sp(2k) algebra and we use $k-1$ missing label numbers $\overline{\Gamma}^k_1, \overline{\Gamma}^k_2, \ldots, \overline{\Gamma}^k_{k-1}$ to distinquish them [here $\overline{\Gamma}^k_k$ is dependent on the h_k, see (3b)]. If we take into the consideration the results of the Bincer paper[1] we see that the pattern (3a)–(3d) is appropriate for labeling the states in the orthogonal base reduced on the chain (1a). In Sec. II we will find branching rules and an appropriate pattern for the reduction (1b).

II. BRANCHING RULES FOR THE REDUCTION sp(2N) ↘ sp(2N − 2)×sp(2)

The following pattern may be used, intead of (3a), for labeling the vectors in the BUIR's of the sp(2N) algebra

$$|(\Omega)\rangle = \left|\begin{array}{ccccccccccccccccccc} \Omega^N_1 & & \Omega^N_2 & & \Omega^N_3 & & \cdots & & & & & & & & \Omega^N_N \\ & \Gamma^N_1 & & \Gamma^N_2 & & \Gamma^N_3 & & \cdots & & \Gamma^N_{N-1} & & & & & h_N \\ & & \Omega^{N-1}_1 & & \Omega^{N-1}_2 & & \Omega^{N-1}_3 & & \cdots & & \Omega^{N-1}_{N-1} & & & & \\ \cdots & & & & & & & & & & & & & & \\ & & & & & & & & & & & & \Omega^1_1 & & \\ & & & & & & & & & & & & & h_1 & \end{array}\right\rangle, \tag{4a}$$

 0022-2488/87/050989-02$02.50

where the meaning of the Ω's numbers is the same as in (3a). The eigenvalues of the $C^2_{(k)}$ operators (2) are equal to $2\sigma_k(\sigma_k+2)$ and we have

$$\sigma_k = \sum_{p=1}^{k} \Omega_p^k + \sum_{p=1}^{k-1} \Omega_p^{k-1} - 2\sum_{p=1}^{k-1} \Gamma_p^k . \tag{4b}$$

The branching rules for the reduction (1b) are contained in the system of inequalities ($k = N, N-1,...,2$)

$$\Omega_1^k \geqslant \Gamma_1^k \geqslant \Omega_2^k \geqslant \Gamma_2^k \geqslant \cdots \geqslant \Gamma_{k-1}^k \geqslant \Omega_k^k , \tag{4c}$$

$$\Gamma_1^k \geqslant \Omega_1^{k-1} \geqslant \Gamma_2^k \geqslant \Omega_2^{k-1} \geqslant \cdots \geqslant \Gamma_{k-1}^k \geqslant \Omega_{k-1}^{k-1} , \tag{4d}$$

$$\sum_{p=a}^{k-1} (\Omega_{p+1}^k + \Omega_p^{k-1}) \geqslant \Gamma_a^k + 2\sum_{p=a+1}^{k-1} \Gamma_p^k ,$$

$$\text{for } a = 1,2,...,k-1 , \tag{4e}$$

$$h_k = \sigma_k, \sigma_k - 2, ..., -\sigma_k . \tag{4f}$$

Proof: Let us consider two sets $\Delta^k(A_1,A_2,...,A_k;M)$ and $\overline{\Delta}^k(A_1,A_2,...,A_k;M)$. The first one contains points $X = (x_1,...,x_k)$, where the x_i are integer or half-integer numbers bounded by the system inequalities (5) and M is given by (6):

$$A_i \geqslant x_i \geqslant -A_i, \quad \text{for } i = 1,2,...,k , \tag{5}$$

$$M = \sum_{i=1}^{k} x_i . \tag{6}$$

The second one contains points $Y = (y_1,...,y_k)$, where coordinates y_i are integer of half-integer numbers satisfying relations

$$A_{p+1} + y_{p-1} \geqslant y_p \geqslant |A_{p+1} - y_{p-1}|, \quad \text{for } p = 1,2,...,k , \tag{7a,b,c}$$

$$y_{k-1} \geqslant y_k \geqslant -y_{k-1} . \tag{7d}$$

Here $y_0 = A_1$ and $y_k \equiv M$.

The dimensions for the sets Δ^k and $\overline{\Delta}^k$ are the same. The above result follows from the relation

$$D[\overline{\Delta}^k(A_1,A_2,...,A_k;M)]$$

$$= \sum_m D[\overline{\Delta}^{k-1}(A_1,A_2,...,A_{k-1};M-m)]$$

$$\times D[\Delta^1(A_k,m)] , \tag{8}$$

where we denote by $D[R]$ the dimension of the set R. The proof of (8) is rather easy by induction for k.

It is obvious from (3b)–(3d) and (10) that the representation of the algebra $\mathrm{sp}(2k-2)\times A_k^k$: $(\Omega_1^{k-1}, \Omega_2^{k-1},..., \Omega_{k-1}^{k-1})\times h_k$ enters $D[\Delta^k(A_1,A_2,...,A_k;\frac{1}{2}h_k)]$ times into the representation $(\Omega_1^k, \Omega_2^k,...,\Omega_k^k)$, where

$$A_i = \tfrac{1}{2}(B_i - C_i), \quad \text{for } i = 1,...,k , \tag{9a}$$

$$B_k = \Omega_1^k , \tag{9b}$$

$$C_1 = 0 , \tag{9c}$$

$$B_p = \mathrm{Min}(\Omega_{k+1-p}^k, \Omega_{k-p}^{k-1}), \quad \text{for } p = 1,...,k-1 , \tag{9d}$$

$$C_p = \mathrm{Max}(\Omega_{k+2-p}^k, \Omega_{k+1-p}^{k-1}), \quad \text{for } p = 2,3,...,k , \tag{9e}$$

$$x_p = -\overline{\Gamma}_{k+1-p}^k + \tfrac{1}{2}(B_p + C_p) . \tag{10}$$

Now if we assume a simple relation between the missing label numbers occuring in the pattern (4a) and coordinates y_a,

$$y_a = \frac{1}{2}\left\{B_1 + \sum_{p=2}^{a+1} (B_p + C_p - 2\Gamma_{k-p+1}^k)\right\}, \tag{11}$$

for $a = 1,2,...,k-1$, we immediately obtain from (7a,b,c) the system of inequalities (4c)–(4e) and because σ_k is equal to $2y_{k-1}$ we also get (4f) from (7d), which is what we wanted to show.

[1] A. M. Bincer, "Missing label operators in the reduction $\mathrm{Sp}(2N) \downarrow \mathrm{Sp}(2N-2)$," J. Math. Phys. **21**, 671 (1980).

[2] D. P. Zhelobenko, "Klassiceskije grupy. Spektralnyj analiz konecnomiernyh predstavlenij," Usp. Mat. Nauk. **17** (1), 27 (1962).

Enlargeable graded Lie algebras of supersymmetry

Paolo Teofilatto
Department of Mathematics, King's College London, Strand, London WC2R 2LS, England

(Received 20 May 1986; accepted for publication 14 January 1987)

A criterion to enlarge infinite-dimensional Lie algebras to analytic Lie groups is used here for the extension of graded Lie algebras of physical interest to super-Lie groups.

I. INTRODUCTION

Supersymmetry[1] required new and rich geometrical structures able to treat commuting and anticommuting fields of supersymmetric field theories on the same ground. These structures were built with the usual techniques of differential geometry starting from a graded space called a superspace.

In particular, super-Lie groups[2] were defined for the exponentiation of graded Lie algebras (GLA's), which are the algebras of infinitesimal transformations in supersymmetric theories.

The study of the extension of graded Lie algebras to globally defined graded Lie groups appeared definitively interesting as a basic ingredient for the investigation of topological properties of superspace field theories.

This extension, which generalized to super-Lie groups the classical Lie third theorem,[3] was first proved by Rogers[2] for the case when the mathematical representation of superspace was a finite-dimensional Banach space. The same result was given by Bruzzo–Cianci in the case of infinite-dimensional superspace with a countable basis. These results show that any graded Lie algebra extends to a super-Lie group provided that the super-Lie group is modeled on a Banach space with a countable basis. Nevertheless, the case of a noncountable basis, hence nonseparable Banach spaces, is interesting from the mathematical and from the physical point of view, because it is still unknown what topological properties should have the mathematical representation of the superspace in order to give a precise formalization of supersymmetric theories.

In the Appendix, a space is shown that could be a good model of superspace and it is an example of an infinite-dimensional Grassman algebra which is a nonseparable Banach space.

So it is interesting to look for conditions that ensure the extension of graded Lie algebras to super-Lie groups modeled on general (i.e., also nonseparable) Banach spaces. This leads to the theory of infinite-dimensional (normed) Lie algebras where algebras exist that are not enlargeable to a group (in the sense of the Lie third theorem). Therefore conditions on such Lie algebras are needed; namely we use a theorem of Swierczkowsky[4] on solvable normed Lie algebras to prove the following proposition (see Proposition 5.1).

Proposition: Let A be a graded Lie algebra whose commutative sector (which is a Lie algebra) is solvable, then A is enlargeable to a super-Lie group *for any (representation of) superspace.*

Then, generalizing the result to any superspace, we paid the price of reducing ourselves to the set of solvable Lie algebras. But this is not a too great restriction: as we shall see, this set is big enough to allow the extension of a wide class of graded Lie algebras of physical interest (namely the supersymmetry and the superconformal algebras). Moreover, apart from the topological structure of superspace, our theorem works also in the context of real graded Lie algebras with an infinite number of generators.

This paper is organized as follows. In Sec. II, after the definition of normed Lie algebras and their local groups, we describe a sufficient condition to solve the problem of the extension of a normed Lie algebra to a Lie group. Sections III and IV are devoted to properly defining the same problem for graded Lie algebras and super-Lie groups. Section V contains our main result: we give a sufficient condition for the extension of graded Lie algebras. From the demonstration, an alternative definition of super-Lie groups arises, more suitable in the infinite-dimensional case. In Sec. VI we apply our condition for the graded Lie algebras of physical interest.

II. LIE GROUPS AND NORMED LIE ALGEBRAS

In order to describe the relation between Lie groups and Lie algebras, we consider the Lie group $GL(n,\mathbb{R})$ of linear invertible transformations on $\mathbb{R}^n$ as an example. A map, called an exponential map, is defined on the (Lie) algebra of the $n\times n$ real matrix $\mathcal{M}^{n\times n}(\mathbb{R})$ into $GL(n,R)$ by the series $\exp(A)=\Sigma_{k=0}^{\infty}(A^k/k!)$, where $A=I$ is the identity matrix and each matrix element converges absolutely.

Since $\exp(0)=I$, from the inverse function theorem it follows that any matrix in some neighborhood U of the identity in $GL(n,\mathbb{R})$ can be expressed as $\exp(A)$ for some A of an open neighborhood B of the origin of $\mathcal{M}^{n\times n}(\mathbb{R})$. The algebra $\mathcal{M}^{n\times n}(\mathbb{R})$ is isomorphic to the tangent space at the identity of $GL(n,\mathbb{R})$, and it is called the Lie algebra of the Lie group $GL(n,\mathbb{R})$. The group structure of $U\subseteq GL(n,\mathbb{R})$ is uniquely determinated by the Lie algebra through the "morphism" exp,

$$\exp(x)\cdot\exp(y)=\exp(x*y), \tag{2.1}$$

where the product on the lhs of the formula (2.1) is the product of the Lie group and the product on the rhs is the Campbell–Hausdorff product on the Lie algebra,[3]

$$x*y=x+y+\frac{1}{2}[x,y]+\sum_{n\geq 3}^{\infty}P_n(x,y), \tag{2.2}$$

with $P_n(x,y)$ the homogeneous Lie polynomial of degree n.

Then the local structure of a Lie group, the structure in a sufficiently small neighborhood, is completely determinated by its infinitesimal group, that is, by its Lie algebra. This is of great importance in applications for when properties of local

 0022-2488/87/050991-06$02.50

nature are being studied, and one needs only to consider the Lie algebra.

But it is also important to make sure that the study of infinitesimal transformations (i.e., of the action of the Lie algebra) is sufficient in describing the finite action of the (Lie) group. This is related to the question (problem of extension of a Lie algebra): given a Lie algebra L, does there exist a Lie group G that admits L as its Lie algebra?

The answer is always yes for finite-dimensional Lie algebras.[5] But in the infinite-dimensional case the answer can be no, and then the problem of extension is not trivial in that case.

In order to give sufficient conditions for the extension of an infinite-dimensional Lie algebra to a Lie group, one needs a more rigid structure called a normed-Lie algebra.[6]

Definition 2.1: A normed-Lie algebra L is a Lie algebra which is also a normed space with a norm $\|\cdot\|$ such that, for each $x, y \in L$,

$$\|[x,y]\| \leqslant M\|x\|\,\|y\|, \quad M>0. \tag{2.3}$$

Example 2.1: If A is a Banach space, the Lie algebra ΛA, given by the operation

$[a,b] = ab - ba$, $a,b \in A$, is a normed-Lie algebra.

Definition 2.1 tells us that the Lie product is continuous with respect to the topology induced by the norm;one could ask if the infinite series of the Campbell–Hausdorff formula (2.2) is convergent in this topology. Actually,[7] the Campbell–Hausdorff formula is absolutely convergent in the ball of L $\|x\| < \rho = \frac{1}{2}\log 2$ and there exists a positive number c such that if $\|x\|$, $\|y\| < c$, then $\|x*y\| < \rho$. The number c enables us to give an example of the local group of L (Ref. 6).

Definition 2.2: An open set B of a normed-Lie algebra L is a local group of L if and only if (i) the operation $*$: $B \times B \to L$ is continuous and well defined by the Campbell–Hausdorff formula, (ii) for each x,y,z of B we have $x*(y*z) = (x*y)*z$, and (iii) if x and nx belong to $B (n \in \mathbb{N})$ then $x^n = x*\cdots*x = nx$.

Example 2.2: The open ball $B = \{x \in L$ such that $\|x\| < c\}$ is a local group of L. Now we give a definition of a Lie group due to Hoffmann,[7] which is more suitable than the ordinary one in dealing with infinite-dimensional objects. In Proposition 2.2 we show that his definition is equivalent to the ordinary one.

Definition 2.3: A Lie group G is a topological group such that there exists a normed-Lie algebra L and a function exp: $L \to G$ with the following properties: (i) L has a local group B and G has an open set U such that (s.t.) $\exp B = U$ and exp: $B \to U$ is a homeomorphism with $\exp(x*y) = \exp(x)\cdot\exp(y)$, and (ii) if $x \in L$ and $r,s \in \mathbb{R}$ then $\exp(r+s)x = \exp(rx)\cdot\exp(sx)$.

The normed-Lie algebra L is the Lie algebra of the Lie group G and we write $L = \text{Lie } G$. The copy (B,U) is called the linearization of G.

Definition 2.4: A normed-Lie algebra L is enlargeable if there exists a Lie group G such that Lie $G = L$.

The extension problem is now to look for enlargeable normed-Lie algebras. Since there exist nonenlargeable normed-Lie algebras,[8] we are interested in sufficient conditions to have enlargeable normed-Lie algebras.

Example 2.3: The normed Lie algebras ΛA of example 2.1 are global with Lie group $G = \{$invertible elements of $A\}$. We recall that a Lie algebra L is solvable if the chain of ideals of L,

$$L^1 = [L,L],$$

$L^{n+1} = [L^n, L^n]$ is such that there exists $m \in \mathbb{N}$ s.t.

$$L^m = L^{m+1} = L^\infty = \{0\}. \tag{2.4}$$

In Ref. 4 Swierczkowski proved the following proposition.

Proposition 2.1: Any solvable normed-Lie algebra is enlargeable.

In the next section it will be shown that a class of normed-Lie algebras of great interest in supersymmetric theories is solvable, so it is enlargeable.

We end this section by proving[7] that Definition 2.3 of the Lie group is equivalent (in any dimension) to the classic one.

Proposition 2.2: The Lie group of Definition 2.3 is an analytic manifold with an analytic group operation.

Proof: Consider two linearizations $(B,U),(C,U)$ of the Lie group G s.t. $C*C \subseteq B$ (this is always possible, see Refs. 7 and 9). Charts on G are given by (U_g, f_g), where $g \in G$ and the maps f_g: $U_g \to C \subseteq \text{Lie}(G)$ are defined by $f_g(p) = \log g^{-1}p$ ($\log = \exp^{-1}$). If the intersection between two charts is not empty, $U_g \cap U_h \neq \phi$, one has $g^{-1}h \in U$, then $t = \log g^{-1}h$ belongs to C.

Compute the transition functions $f_{gh} = f_g \circ f_h$: $f_h(U_g \cap U_h) \to f_g(U_g \cap U_h)$: $f_{gh}(x) = f_g(h \exp x) = \log g^{-1}(h \exp x) = \log(g^{-1}h \exp x)$ and by Definition 2.3 (i) $f_{gh}(x) = \log(\exp(\log g^{-1}h * x)) = \log g^{-1}h * x = t * x$.

Therefore the transition functions are bijective maps given by an absolutely convergent series of terms multilinear and continuous on x, that is they are analytic. Since the left and right translations are clearly isomorphisms of the analytic structure, and the group operations are analytic in a neighborhood of the identity, it follows that multiplication and inversion are analytic.

III. GRADED-LIE ALGEBRAS

In looking for a solution of the problem of extension of a Lie algebra in the supersymmetric case, we have to note that in supersymmetric theories the algebra of infinitesimal transformations is not a Lie algebra, but a graded-Lie algebra (GLA).

Definition 3.1: A real graded algebra[10] A is a real vector space such that[11] (i) A is the direct sum of two subspaces, $A = A_0 \oplus A_1$; (ii) A is an algebra such that $A_k A_h \subseteq A_{h+k}$, $k,h = 0,1$ (the sum is mod 2); and (iii) for each homogeneous $a,b \in A$ it is $ab = (-1)^{|a||b|} ba$, where the degree of a homogeneous element $a \in A_k$ is $|a| = k$ $(k = 0,1)$.

Example 3.1: The Grassmann algebra over $\mathbb{R}$, B_L, is a real graded algebra. If we write the elements of B_L as $\xi = \Sigma_{\mu \in \mu_L} a_\mu e_\mu$ (where M_L is a suitable set of indices and e_μ is a basis of B_L), a norm can be defined on B_L such that it becomes a Banach algebra: $\|\xi\| = \Sigma_{\mu \in \mu_L} |a_\mu|$ (see Ref. 12). The generalization to infinite-dimensional Banach algebras of "Grassman-type" (that is, graded algebras), is given by the definition of a Banach–Grassmann algebra.[13]

Definition 3.2: A real graded Banach algebra Q is a Banach–Grassmann algebra if (i) (self-duality) for each continuous Q_0 linear map f: $Q_h \to Q_k$ $(h,k=0,1)$, there exists a unique element $u \in Q_{h+k}$ such that $\|u\| = \|f\|$ and $f(a) = ua$ for all $a \in Q_h$: (ii) the commutative sector of Q, Q_0, is $Q_0 = \mathbb{R} \oplus Q_0'$, with $\|\lambda + s\| = |\lambda| + \|s\|$ for $\lambda \in \mathbb{R}$ and $s \in Q_0'$, where Q_0' denotes the Banach subalgebra of Q generated by even powers of elements of the anticommutative sector of Q, Q_1.

An example of a Banach–Grassmann algebra will be given in the Appendix. Now we recall the definition of graded Lie algebra.[10]

Definition 3.3: A real graded algebra A whose product is denoted by an angular bracket $\langle \cdot, \cdot \rangle$ is a real graded Lie algebra (GLA) if, for each $a,b,c \in A$, then

$$(-1)^{|a||c|}\langle a,\langle b,c\rangle\rangle + (-1)^{|c||b|}\langle c,\langle a,b\rangle\rangle + (-1)^{|b||a|}\langle b,\langle c,a\rangle\rangle = 0. \tag{3.1}$$

Actually, we are interested in GLA's of a particular shape; as we shall see in the next section, the linearized structures related to a super-Lie group are graded Lie algebras of the form $Q \otimes_{\mathbb{R}} \mathcal{G}$, where $\mathcal{G}$ is a generic real GLA, and with Q we denote the graded algebra of Example 3.1 or Definition 3.2. In order to give $Q \otimes_{\mathbb{R}} \mathcal{G}$ a graded Lie algebra structure, we need two definitions.[10]

Definition 3.4: Given two graded algebras A, B, the graded tensor product $A \otimes B$ is a graded algebra through the product

$$(a \otimes b)\cdot(a' \otimes b') = (-1)^{|b||a'|} aa' \otimes bb'. \tag{3.2}$$

With respect to the product (3.2) the commutative and the anticommutative sectors of $A \otimes B$ are

$$\begin{aligned}(A \otimes B)_0 &= A_0 \otimes B_0 \oplus A_1 \otimes B_1,\\ (A \otimes B)_1 &= A_0 \otimes B_1 \oplus A_1 \otimes B_0.\end{aligned} \tag{3.3}$$

Definition 3.5: Given a real algebra $A = A_0 \oplus A_1$, a graded Lie algebra is defined by the product

$$\langle X,Y\rangle = XY - (-1)^{|x||y|} YX. \tag{3.4}$$

The product (3.4) is compatible with the grading, so A_0 is a Lie algebra. Applying Definitions 3.4 and 3.3 to the real graded algebras Q and $\mathcal{G}$, we define $Q \otimes \mathcal{G}$ as a graded Lie algebra. The commutative sector $(Q \otimes \mathcal{G})_0 = Q_0 \otimes \mathcal{G}_0 + Q_1 \otimes \mathcal{G}_1$ is a Lie algebra with Lie product

$$\begin{aligned}[(a \otimes g),(b \otimes h)] &= ab \otimes \langle g,h\rangle - ba \otimes \langle h,g\rangle\\ &= 2ab \otimes \langle g,h\rangle.\end{aligned} \tag{3.5}$$

If the algebra Q is infinite dimensional with respect to the reals (as in Definition 3.2) then $(Q \otimes \mathcal{G})_0$ is an infinite-dimensional (real) Lie algebra.

We ask if the infinite-dimensional Lie algebra $(Q \otimes \mathcal{G})_0$ is enlargeable in the sense of Definition 2.4. To use the sufficient condition of Proposition 2.1, we give $(Q \otimes \mathcal{G})_0$ a structure of normed algebra.

Proposition 3.1: It is possible to give a norm to $(Q \otimes \mathcal{G})_0$ such that it is a normed-Lie algebra.

Proof: The norm is defined as follows[14]: let $\{D_A\}$ be the Q-module basis of $Q \otimes \mathcal{G}$ and let C^E_{AB} be the structure constants with respect to such a base. If $X = X^A D_A$ is an element of $Q \otimes \mathcal{G}$, we set $\|X\| = \Sigma_A |X_A|$, where $|\cdot|$ is the norm defined on Q. Then

$$\|[X,Y]\| = \sum_E \|Y^B X^A C^E_{AB}\| \leqslant M \|X\| \, \|Y\|.$$

The normed algebra $Q \otimes \mathcal{G}$ is complete because $Q^N = Q \times \overset{N \text{ times}}{\cdots} \times Q$ is complete (N = real dimension of $\mathcal{G}$). Then from Propositions 2.1 and 3.1 the following is true.

Proposition 3.2: Let Q be the graded Banach algebra of Example 3.1 or Definition 3.2, and let $\mathcal{G}$ be any real graded Lie algebra, if the Lie algebra $(Q \otimes \mathcal{G})_0$ is solvable, there exists a Lie group G such that Lie $G = (Q \otimes \mathcal{G})_0$.

In Sec. V we will prove that if such a G exists, then G is in fact a super Lie group and $Q \otimes \mathcal{G}$ is its related graded Lie algebra. Before doing that, we shall give the definitions of these objects.

IV. SUPERDIFFERENTIAL CALCULUS AND SUPER-LIE GROUPS

This section contains a brief account of superdifferential calculus needed in the following. For a complete description of this subject we rely on the original works[12,13] and Refs. 15–17.

Roughly speaking, superdifferential calculus is a Fréchet differential calculus on the Banach algebra Q (see Example 3.1, $Q = B_L$, and Definition 3.2), which takes suitable account of the graded structure of the algebra Q.

To be a bit more concrete, recall[18] that on a real Banach space H, which is an $\mathbb{R}$ module, the Fréchet differential f' of a differentiable function on H is an $\mathbb{R}$-linear operator such that $f(p+h) = f(p) + f'(p)\cdot h + o(\|h\|)$ where the product $(\cdot)$ is the product on $\mathbb{R}$.

Because one can regard Q as a module on its commutative sector Q_0, the Fréchet differential will be defined as a Q_0-linear operator. If Q is a Banach–Grassmann algebra (see Definition 3.2), we have that the space of Q_0-linear functions defined on Q into Q, $\mathcal{L}_{Q_0}(Q,Q)$ is isomorphic to Q (self-duality). Therefore one has, for φ defined on Q into Q,

$$\varphi(p+q) = \varphi(p) + \varphi'(p)\cdot q + o(\|q\|), \tag{4.1}$$

where the product and the norm are those on Q.

A superdifferentiable function is an infinite-time differentiable function with Q_0-linear differential operator the Fréchet differential operator. Note that the Q_0-linearity implies $\mathbb{R}$ linearity, so a superdifferential function is also a C^∞-differential function (regarding Q as real Banach space).

If Q is finite dimensional ($Q = B_L$ of Example 3.1), $\mathcal{L}_{Q_0}(Q,Q)$ does not exhaust Q and formula (4.1) must be explicitly required. The superdifferential structure defined on Q obviously extends to the Q_0 module $Q^{m,n} = (Q_0)^m \times (Q_1)^n$.

In the definition of (m,n) super-Lie groups, one needs the definition of a superanalytic function on $Q^{m,n}$ and of superanalytic supermanifolds.[12,13]

Definition 4.2: Let f: $U \to Q$ be a function defined on an open set U of $Q^{m,n}$. The function f is superanalytic on U if,

for each p on U, it can be written as an absolutely convergent power series of the form

$$f(p) = \sum_{k_1,\cdots k_{m+n}=0}^{\infty} a_{k_1\cdots k_{m+n}} p_1^{k_1} p_2^{k_2} \cdots p_{m+n}^{k_{m+n}}$$
$$\text{with } a_{k_1\cdots k_{m+n}} \in Q. \tag{4.2}$$

It is easy to see that a function which is analytic, with respect to the underlying real Banach structure of $Q^{m,n}$ and with the Q_0 linear first Fréchet differential, is superanalytic.

Definition 4.3: An (m,n) superanalytic supermanifold M is a topological manifold endowed with an atlas $\mathscr{A} = \{(U_\alpha, \varphi_\alpha); \varphi_\alpha: U_\alpha \to Q^{m,n}\}$, whose transition functions are superanalytic functions.

A superversion of the concept of fiber bundle was also given.[19,20]

A standard example is the tangent bundle of a supermanifold M. This is the bundle obtained by gluing together the tangent spaces at each point p of M, T_pM. It is important to note that points of the tangent space T_pM have to be considered equivalent classes of paths on M and not derivations. In fact, the space of derivations on p, $\tilde{T}_pM$, is broader than the space of tangent vectors, because, in the graded case, the derivations are "Leibnitz-type" operators on germs of functions valued not only in the set of the "scalars" Q_0 (Ref. 21), but also in the broader Q. Actually functions into the scalar are not the only interesting functions; coordinate functions themselves are not always Q_0 valued and it is important to define the derivations on those functions.

Therefore in the graded case, one has two different spaces[12,13]: the space of derivations on p, $\tilde{T}M$, which is a free Q module on the basis $\{\partial/\partial x^1,\ldots,\partial/\partial x^m,\partial/\partial\vartheta^1,\ldots,\partial/\partial\vartheta^n\}$ and the space of tangent vectors on p generated by a "graded linear span" of the previous basis, that is,

$$T_pM = Q_0 \otimes \left\{\frac{\partial}{\partial x^1},\ldots,\frac{\partial}{\partial x^m}\right\} + Q_1\left\{\frac{\partial}{\partial\vartheta^1},\ldots,\frac{\partial}{\partial\vartheta^n}\right\}.$$

Vector fields and derivative fields are superdifferentiable sections of the superfiber bundles TM and $\tilde{T}M$ (the latter is modeled on Q^{m+n}). The definition of the super-Lie group[2] is an obvious extension of the Lie group one.

Definition 4.4: A topological group G is a super-Lie group iff G is an (m,n) superanalytic supermanifold with superanalytic group operations.

The space of left-invariant derivative fields of G is a graded Lie algebra isomorphic to $\tilde{T}_eG$, where e is the identity of G; this is called the graded Lie algebra of G (graded Lie module in Ref. 2). The space of left-invariant vector fields of G is isomorphic to T_eG and it is the Lie algebra of G.

In this context, the superanalog of the extension problem reads "given a graded Lie algebra $\Gamma = Q \otimes \mathscr{G}$, does there exist a super-Lie group G that admits Γ as its graded Lie algebra?" A way to answer this is to consider the commutative sector of $Q \otimes \mathscr{G}$, the Lie algebra $(Q \otimes \mathscr{G})_0$, and to try to extend it to a super-Lie group to show that $\tilde{T}_eG \cong \Gamma$. Following this procedure and using the classical Ado theorem, Rogers[2] proved that $Q \otimes \mathscr{G}$, with $Q = B_L$ and $\mathscr{G}$ any real GLA, extends to a super-Lie group; that is obviously generalizing Definition 2.4 to the super case, where any finite-dimensional graded Lie algebra is enlargeable.

For Q = Banach–Grassmann algebra (Definition 3.2), Bruzzo–Cianci,[14] using the universal enveloping algebras, proved that if Q has a countable basis then any graded Lie algebra $Q \otimes \mathscr{G}$ is enlargeable.

The case of noncountable basis, hence nonseparable Banach–Grassmann Q, must also be considered. Moreover, the example in the Appendix shows that there is some merit in doing that.

In the next section, we use the theorem of Swierczkowski[4] (Proposition 2.1) to show that the graded Lie algebra $Q \otimes \mathscr{G}$ is enlargeable if the Lie algebra $(Q \otimes \mathscr{G})_0$ is solvable, for a *generic* Banach–Grassmann algebra Q.

V. ENLARGEABLE GRADED LIE ALGEBRAS

From Propositions 3.1 and 2.1 we have that if a graded Lie algebra $Q \otimes \mathscr{G}$ is such that its commutative sector $(Q \otimes \mathscr{G})_0$ is solvable, then there exists an analytic Lie group G such that Lie $G = (Q \otimes \mathscr{G})_0$. Now we prove that G is actually a super-Lie group.

Proposition 5.1: Given a graded Lie algebra $Q \otimes \mathscr{G}$ whose Lie algebra $(Q \otimes \mathscr{G})_0$ is solvable, there exists a super-Lie group G that admits $(Q \otimes \mathscr{G})_0$ as its Lie algebra and $(Q \otimes \mathscr{G})$ as its graded algebra.

Proof: For the choice of the norm on $L = (Q \otimes \mathscr{G})_0$, there exists a homeomorphism $\varphi: L \to Q^{m,n}$ defined by $\varphi(x_aD^a + \vartheta_\alpha D^\alpha) = (x_a, \vartheta_\alpha)$, where $D^A (A = 1,\ldots,m+n)$ is the Q-module basis of L, $a = 1,\ldots,m = \dim_{\mathbb{R}} \mathscr{G}_0$, $\alpha = 1,\ldots,n = \dim_{\mathbb{R}} \mathscr{G}_1$.

Recall the analytic atlas of the group G, given by Proposition 2.1, s.t. Lie $G = L$, $\mathscr{A} = \{(U_g, f_g), f_g: U_g \to L\}_{g\in G}$; this induces an atlas modeled on $Q^{m,n}$, $\bar{\mathscr{A}} = \{(U_g, \varphi_g), \varphi_g: U_g \to Q^{m,n}\}_{g\in G}$ with $\varphi_g(p) = \varphi \circ f_g(p) = \varphi(\log g^{-1}p)$. The transition functions are $\varphi_{gh} = \varphi_g \circ \varphi_h^{-1}$: $\varphi_h(U_gU_h) \to \varphi_g(U_gU_h)$ then

$$\varphi_{gh}(p) = \varphi_g(h \exp \varphi^{-1}(p)) = \varphi \log(g^{-1}h \exp \varphi^{-1}(p))$$
$$= \varphi(\log \exp(\log g^{-1}h * \varphi^{-1}(p))) = \varphi(t * \varphi^{-1}(p)).$$

That is the transition functions are superanalytic by the very definition of the Campbell–Hausdorff product on L. For the same reason the group operations are superanalytic in a neighborhood of the identity. Since the translations are isomorphisms of the superanalytic structure, it follows that the group operations are superanalytic, therefore G is a super-Lie group.

By construction $(Q \otimes \mathscr{G})_0$ is the tangent space at the identity of G, that is,

$$(Q \otimes \mathscr{G})_0 = T_eG = Q_0\left\{\frac{\partial}{\partial x_1},\ldots,\frac{\partial}{\partial x_m}\right\} + Q_1\left\{\frac{\partial}{\partial\vartheta_1},\ldots,\frac{\partial}{\partial\vartheta_n}\right\}.$$

The Lie module of G is the free Q module generated by the

basis of the Lie algebra of G: $Q\{\partial/\partial x_1,\ldots,\partial/\partial\vartheta_n\}$, that is, $Q\otimes\mathcal{G}$. This completes the proof.

We call an enlargeable graded Lie algebra $Q\otimes\mathcal{G}$ s.t. there exists a super-Lie group with the properties of Proposition 5.1. Then we have proved the following: any GLA $Q\otimes\mathcal{G}$ with $(Q\otimes\mathcal{G})_0$ solvable is enlargeable.

VI. EXAMPLES: THE SUPERSYMMETRIC AND SUPERCONFORMAL ALGEBRAS

Now we want to apply the condition of Proposition 5.1 in the context of graded Lie algebras of physical interest. Actually, only some graded Lie algebras are of physical interest. In relativistic quantum field theory there exist several restrictions in defining a group of symmetry. These restrictions were proved by Coleman–Mandula[22] in a series of "no-go" theorems within the framework of Lie algebras.

The introduction of fermionic generators of (super) symmetry circumvented these no-go theorems. In fact, supersymmetry requires bosonic (commutative) and fermionic (anticommutative) generators to define a graded Lie algebra of infinitesimal transformations. Nevertheless, the Coleman–Mandula results still provide strong limitations for bosonic and fermionic generators as well. Haag–Lopuszansky–Sohnius[23] proved that the most general real graded Lie algebra in supersymmetric field theory is $A=A_0\oplus A_1$, where A_0 is given by the generators of the Poincaré translations P_μ and the generators B_r of a compact Lie algebra $\mathcal{B}$ of the internal symmetry group, A_1 is given by N spin-$\frac{1}{2}$ generators Q_α^K, $\overline{Q}_{\dot\alpha}^K$ (here the Weyl representation of spinors is used), $\alpha=1,2,3,4$. The commutation–anticommutation rules are[1]

$$
\begin{aligned}
&[P_\mu,P_\nu]=[P_\mu,Q_\alpha^k]=[P_\mu,B_r]=0,\\
&[B_r,B_m]=d_{rm}^{\ s}B_s,\\
&\{Q_\alpha^k,Q_{\dot\alpha}^\mu\}=-2\delta^{k\mu}\gamma_{\alpha\dot\alpha}^\mu P_\mu,\\
&\{\overline{Q}_{\dot\alpha}^k,\overline{Q}_{\dot\beta}^\mu\}=\{Q_{\dot\alpha}^k,Q_{\dot\beta}^\mu\}\\
&\qquad=\begin{cases}0, & \text{without central charges,}\\ \epsilon_{\alpha\rho}X^{kM}, & \text{with central charges,}\end{cases}
\end{aligned}
\tag{6.1}
$$

where the generators X^{KM} belong to an Abelian subalgebra of the algebra $\mathcal{B}$ and they commute with any other generator.

Applying our condition to the GLA $Q\otimes\tilde{\mathcal{G}}$, where $\tilde{\mathcal{G}}$ is given by (6.1), we obtain the following.

Lemma 5.1: If the algebra $\mathcal{B}$ of the internal group is solvable, then the GLA $Q\otimes\tilde{\mathcal{G}}$ is enlargeable.

Proof: Working out $L^1=[(Q\otimes\tilde{\mathcal{G}})_0,(Q\otimes\tilde{\mathcal{G}})_0]$, $L^2=[L^1,L^1]$ etc., by the rules (6.1) one has for sufficiently high n, $L^n=Q_0\times\mathcal{B}^i$, where $\mathcal{B}^{i+1}=[\mathcal{B}^i,\mathcal{B}^i]$ and $\mathcal{B}^1=[\mathcal{B},\mathcal{B}]$. Then $(Q\otimes\tilde{\mathcal{G}})$ is solvable if the algebra $\mathcal{B}=(B_r)$ is solvable.

In the following interesting cases $\mathcal{B}$ is actually solvable and one can apply Proposition 5.1.

(1) If in the algebra (6.1) it is $N=1$ then[24] $\mathcal{B}=$ algebra of U(1). This is the case of simple ($N=1$) supersymmetry with chiral U(1); the group G is a (4,4) super-Lie group. For global supersymmetry, the superanalytic supermanifold is trivial, then $G=(Q^{4,4},\varphi)$ with group operation ψ defined by the Campbell–Hausdorff formula that breaks itself to the second order: $L_2=[L,L^1]=0$. Then for X,Y in the algebra $X*Y=X+Y+\frac{1}{2}[X,Y]$ and by rules (6.1), we have

$$
\begin{aligned}
&\psi((\underline{x},\underline{\vartheta});(\underline{y},\underline{\epsilon}))\\
&\quad=\varphi(*)\circ(\varphi^{-1}\times\varphi^{-1})((\underline{x},\underline{\vartheta});(\underline{y},\underline{\epsilon}))\\
&\quad=\varphi*((x_\mu P^\mu+\vartheta_\alpha Q^\alpha);(y_\mu P^\mu+\epsilon_\alpha Q^\alpha))=\varphi((x_\mu+y_\mu)P^\mu+(\vartheta_\alpha+\epsilon_\alpha)Q^\alpha+\tfrac{1}{2}[x_\mu P^\mu+\vartheta_\alpha Q^\alpha,y_\nu P^\nu+\epsilon_\beta Q^\beta])\\
&\quad=\varphi((x_\mu+y_\mu)P^\mu+(\vartheta_\alpha+\epsilon_\alpha)Q^\alpha+\tfrac{1}{2}[\vartheta_\alpha Q^\alpha,\epsilon_\beta Q^\beta])=\varphi((x_\mu+y_\mu)P^\mu+(\vartheta_\alpha+\epsilon_\alpha)Q^\alpha+\vartheta_\alpha\epsilon_\beta\{Q^\alpha,Q^\beta\})\\
&\quad=\varphi((x_\mu+y_\mu+\vartheta_\alpha\gamma_\mu\epsilon_\beta)P^\mu+(\vartheta_\alpha+\epsilon_\alpha)Q^\alpha)=(\underline{x}+\underline{y}+\underline{\vartheta}\underline{\gamma}\underline{\epsilon};\underline{\vartheta}+\underline{\epsilon}).
\end{aligned}
$$

(2) If $\mathcal{B}=0$ the algebra (6.1) is the supersymmetry algebra of N extended pure supersymmetry. The group G is a (4,4N) super-Lie group.

(3) In addition to the various super-Poincaré algebras there are both simple and extended superversions of the conformal algebra:

$$\frac{\text{conformal algebra}+2N\text{ spinoral algebra}}{\text{Lorentz rotation}}$$

(for the commutation–anticommutation rules see Ref. 24). The superconformal algebra is obviously enlargeable and extends to a (10,2N) super-Lie group.

ACKNOWLEDGMENTS

R. Cianci, R. Cirelli, and M. Martellini are warmly thanked for several discussions.

APPENDIX: FROM SPACE TO SUPERSPACE

Originally superspaces were locally defined through four real and N anticommuting coordinates that are spinors, that is points of a vector space V in which the group SL(2,C) is represented.

The most immediate idea that occurs to obtain anticommuting spinor coordinates is to consider the spinors $\vartheta\in V$ as elements of degree 1 of the Grassmann algebra on V, $\Lambda(V)$. Then the product $\Lambda(V)_0\times\cdots\times\Lambda(V)_0\times\cdots\Lambda(V)_1$ could locally represent the superspace [here $\Lambda(V)_0$ is the set of commutative elements and $\Lambda(V)_1$ is the set of anticommutative elements of the Grassmann algebra $\Lambda(V)$]. But two problems arise: (1) superspace would depend on a particular choice of spinor spaces, and (2) as the spinor spaces are of real dimension 4, the superspace would be finite dimensional ($d=2^{3(4+N)}$).

Because of problem (2), undesirable restrictions follow. The Green's functions of fields valued on the algebra (V) will vanish if one takes a sufficiently high number of fermionic fields.[20] Then an infinite-dimensional Grassmann algebra is preferred. Here an infinite-dimensional Grassmann algebra whose elements are spinors is constructed. This algebra result is a Banach–Grassmann algebra (see Definition 3.2) *without a countable basis.*

Given a general space-time, one can associate to each orthonormal frame g a spinor space V_g, provided that the space-time admits a spin structure.[25] Changing the orthonormal frame g, one obtains many equivalent spinor spaces. Two orthonormal frames a,b lead to the same spinor space if and only if it results in $a = b\cdot\tau(\gamma)$, where τ is the double covering map τ: SL(2,C) $\to$ SO(3,1) [or SO(4)] and γ belongs to a subgroup of SL(2,C) $\mathscr{C}$ called the Crumeyrolle group.[26] Then if $\mathscr{H}$ is the quotient $SO(3,1)/\tau(\mathscr{C})$, the inequivalent spaces of spinors are parametrized by[27] $\mathscr{H}$: $h\in\mathscr{H}\to V_h$. Now let V be the direct integral of all inequivalent spinor spaces $V = \oplus_{h\in\mathscr{H}} V_h$, where the direct integral is the vector space of elements $(\Sigma_{h\in\mathscr{H}} x_h)$ in which the family $(x_h)_{h\in\mathscr{H}}$ belongs to the Cartesian product $\Pi_h V_h$, x_h belongs to V_h for each h, and $x_h = 0$ almost everywhere. Note that the set of indices $\mathscr{H}$ is not countable (in fact $\mathscr{H}$ is locally isomorphic to R^4).

We give to the space V a l^1-norm defined by

$$\|p\| = \left\| \sum_{h\in\mathscr{H}} x_h \right\| = \left\| \sum_{h\in\mathscr{H}} \mathbf{r}(h)\mathbf{b}(h) \right\| = \sum_{h\in\mathscr{H}} |\mathbf{r}(h)|,$$

here $\mathbf{b}(h)$ is the basis of V_h and $r(h)$ are the coefficients of $x_h\in V_h$. The sum is finite, because x_h vanishes almost everywhere. Sl(2,C) is still represented on the space $V = \oplus_{h\in\mathscr{H}} V_h$ by the direct integral of the representations ρ_h: SL(2,C) $\to V_h$. The Grassmann algebra on V (Ref. 28), $Q = \Lambda(V)$, is a free module on the infinite-dimensional base $\{b(h), b(h)\wedge b(k),\ldots\}_{h,k\in\mathscr{H}}$.

It is easy to see that Q with the previous l^1-norm, is a graded Banach algebra such that the property (ii) of Definition 3.2 holds. In order to prove the self-duality of Q [property (i)] of Definition (3.2), we note the following.

(1) $\mathrm{Hom}(\Lambda(\oplus_{h\in\mathscr{H}} V_h), \Lambda(V))$

$$= \mathrm{Hom}(\oplus_{h\in\mathscr{H}} \Lambda(V_h), \Lambda(V))$$

$$= \oplus_{h\in\mathscr{H}} \mathrm{Hom}(\Lambda(V_h), \Lambda(V)) \quad \text{(Ref. 28)}.$$

(2) If $X\in\Lambda(V)$, I is a finite subset of $\mathscr{H}$, and $Xb_i = 0$ for each $i\in I$, there exists a $v\in\Lambda(V)$ such that $X = vb_I$ ($b_I = b_i b_j\cdots b_k, i,j,k,I$) and $\|X\| = \|v\|$.

Following Ref. 29 one can see that the space of continuous Q_0 linear maps f: $Q_r \to Q_s$ $(r,s = 0,1)$ is isomorphic to Q.

To see the above, it is enough to prove that if f: $\Lambda(V_h)_i \to Q$ is a $\Lambda(V_h)_0$ linear and continuous map, $e_1\cdots e_N$ is the basis of $\Lambda(V_h)$ and $f_i = f(e_i)$, then there exists $q\in Q$ such that $f_i = qe_i$, $i = (1,\ldots,N)$. In fact $f_1e_1 = 0 \overset{(2)}{\Rightarrow}$ there is $q_1\in Q$ s.t. $f_1 = q_1e_2$. Now, $f_2e_1 = -f_1e_2 = qe_1e_2$ then $(f_2 - q_1e_2)e_1 = (f_2 - q_1e_2)e_2 = 0$ so it exists $v\in\Lambda(V)\neq Q$ s.t. $f_2 - q_1e_2 = ve_1e_2$.

Let q_2 be $q_2 = q_1 + ve_1$, then $f_2 = q_2e_2$. If $q = q_1 \neq q_2 + \cdots + q_N$, one has $f_i = qe_i$, $i = 1,\ldots,N$. Therefore $Q = \Lambda(V)$ is self-dual, and is a Banach–Grassmann algebra without a countable basis.

[1] J. Wess and J. Bagger, *Supersymmetry and Supergravity* (Princeton U.P., Princeton, NJ, 1980).
[2] A. Rogers, Super Lie groups," J. Math. Phys. **22**, 938 (1981).
[3] V. Varadarajan, *Lie Groups, Lie Algebras, and Their Representations* (Prentice–Hall, Englewood Cliffs, NJ, 1974).
[4] S. Swierczkowsky, "Embedding theorems for local analytic groups," Acta Math. **114**, 207 (1965).
[5] P. M. Cohn, *Lie Groups* (Cambridge U.P., Cambridge, 1957).
[6] P. de la Harpe, "Classical Banach Lie algebras and Banach Lie groups of operators in Hilbert spaces," Lect. Notes Math. **285**, 2 (1972).
[7] H. K. Hoffmann, "Algebra," from Seminaire Dubreul, 1973–1974.
[8] W. van Est and T. Korthagen, "Nonenlargible Lie algebras," Indag. Math. **26**, 15 (1964).
[9] L. Pontrjaghin, *Topological Groups* (Princeton U.P., Princeton, NJ, 1958).
[10] R.Scheunert, "The theory of Lie superalgebras," Lect. Notes Math. **716**, 2 (1979).
[11] Such an algebra is called $\mathbb{Z}_2$ graded algebra in the current literature; since all the graded algebras in this paper are $\mathbb{Z}_2$, we will omit the suffix $\mathbb{Z}_2$.
[12] A. Rogers, "A global theory of supermanifolds," J. Math. Phys. **21**, 1355 (1980).
[13] A. Jadczyk and K. Pinch, "Superspaces and supersymmetries," Commun. Math. Phys. **78**, 373 (1981).
[14] U. Bruzzo and R. Cianci, "An existence result for Super-Lie groups," Lett. Math. Phys. **8**, 279 (1984).
[15] A. Rogers, "The theory of supermanifolds," Commun. Math. Phys. **105**, 375 (1986).
[16] J. Rabin and L. Crane, "Global properties of supermanifolds," Commun. Math. Phys. 100, (1985).
[17] C. Bojer and S. Gitler, "The theory of G supermanifolds," Trans. Am. Math. Soc. **285**, 241 (1984).
[18] S. Lang, *Differential Manifolds* (Addison–Wesley, Reading, MA, 1972).
[19] U. Bruzzo and R. Cianci, "Mathematical theory of super fiber bundle," Class. Quantum Gravit. **1**, 213 (1984).
[20] J. Hoyos, M. Quirós, J. Ramírez Mittlebrunn, and F. de Urries, "Generalized supermanifolds," J. Math. Phys. **25**, 833 (1984).
[21] Note that $Q^{m,n}$ is a Q_0 module.
[22] S. Coleman and J. Mandula, Phys. Rev. **159**, 1251 (1967).
[23] R. Haag, J. Lopuszanski, and M. Sohnius, Nucl. Phys. B **88**, 257 (1975).
[24] S. Gates, M. Grisaru, M. Rocek, and W. Sigel, *Superspace* (Benjamin, New York, 1983).
[25] J. Milnorn, "Spin structures on manifolds," L'Enseignement Mathematique **9**, 198 (1963).
[26] A. Crumeyrolle, "Structures spinorielles," Ann. Inst. H. Poincaré **1**, 19 (1969).
[27] K. Bugajska, "Spinor structure implies superspace," from *Proceedings of the Second M. Grossmann Meeting on General Relativity*, 1982, pp. 229–235.
[28] C. Chevalley, *Algebra* (Feltrinelli, Milan, 1975).
[29] A. Jadczyk and K. Pilch, "Classical limit of CAR and self-duality of the infinite dimensional Grassmann algebra," in *Quantum Theory of Particles and Fields*, edited by B. Jancewicz and J. Lukierski (World Scientific, Singapore, 1983).

Generalized Burgers equations and Euler–Painlevé transcendents. II

P. L. Sachdev and K. R. C. Nair
Department of Applied Mathematics, Indian Institute of Science, Bangalore-560012, India

(Received 25 April 1986; accepted for publication 18 November 1986)

It was proposed earlier [P. L. Sachdev, K. R. C. Nair, and V. G. Tikekar, J. Math. Phys. **27**, 1506 (1986)] that the Euler–Painlevé equation $yy'' + ay'^2 + f(x)yy' + g(x)y^2 + by' + c = 0$ represents the generalized Burgers equations (GBE's) in the same manner as Painlevé equations do the KdV type. The GBE was treated with a damping term in some detail. In this paper another GBE $u_t + u^a u_x + Ju/2t = (\delta/2)u_{xx}$ (the nonplanar Burgers equation) is considered. It is found that its self-similar form is again governed by the Euler–Painlevé equation. The ranges of the parameter α for which solutions of the connection problem to the self-similar equation exist are obtained numerically and confirmed via some integral relations derived from the ODE's. Special exact analytic solutions for the nonplanar Burgers equation are also obtained. These generalize the well-known single hump solutions for the Burgers equation to other geometries $J = 1,2$; the nonlinear convection term, however, is not quadratic in these cases. This study fortifies the conjecture regarding the importance of the Euler–Painlevé equation with respect to GBE's.

I. INTRODUCTION

One of the best-known model equations in mathematical physics is the Burgers equation

$$u_t + uu_x = (\delta/2)u_{xx}. \tag{1.1}$$

This equation describes the conflict between cumulative nonlinear distortion due to convection and the competing linear diffusive processes that this distortion evokes. Equation (1.1) is essentially a mathematical model, and was written out in an intuitive manner (see Benton and Platzman[1] for the history and review of this equation and its solutions; see also a forthcoming book by Sachdev[2]). Equation (1.1) has a beautiful structure and has the distinctive feature that it can be exactly linearized to the heat equation by the celebrated Hopf–Cole transformation. Otherwise, its utility as a descriptor of physical phenomenon is rather limited. The model equations, which have been derived from the Navier–Stokes equations by suitable perturbation methods, are invariably more complicated than (1.1). An excellent review of these model equations in nonlinear acoustics has been given by Crighton.[3] Here we cite a few of these model equations. We may mention that the large number of these equations is due to the simple isotropic nondispersive scalar nature of the acoustic wave field subjected to nonlinear effects. The nonplanar Burgers equation

$$u_t + uu_x + Ju/2t = (\delta/2)u_{xx}, \tag{1.2}$$

$J = 1,2$, for cylindrical and spherical symmetry, was derived by Leibovich and Seebass[4] from the Navier–Stokes equations, using the method of multiple scales. It describes the N waves from a sonic boom or an explosion. The equation

$$u_t + 2uu_x - Uu = \delta u_{xx}, \tag{1.3}$$

where U is a constant, was treated by Murray[5] as a simple turbulence model. Lardner and Arya[6] discussed two generalized Burgers equations:

$$u_t - uu_x + \lambda u = (\delta/2)u_{xx}, \tag{1.4}$$

$$u_t - [\mu u + \nu u^2 + \nu C(t)]u_x = (\delta/2)u_{xx}. \tag{1.5}$$

Equation (1.4) describes the plane motion of a continuous medium for which the constitutive relation for the stress contains a large linear term proportional to the strain, a small term quadratic in the strain, a small dissipative term proportional to the strain rate, and a small viscous damping term proportional to the velocity. Equation (1.5) describes the motion of a continuous medium when the stress–strain relation contains a term cubic in the strain in addition to the terms described above.

A rather more complicated model is

$$u_t + \frac{1}{a_0}u_x - \frac{\gamma+1}{2a_0^2}uu_x = \frac{\delta}{2a_0^3}u_{xx} + \frac{\beta Q}{2\rho_0 c_p}, \tag{1.6}$$

which was suggested by Karabutov and Rudenko.[7] Here $u(x,t)$ is the velocity of a thermoviscous gas in one-dimensional flow, $Q(x,t)$ is the rate of heat addition prescribed by some external agency, β is the coefficient of thermal expansion, and c_p is the constant-pressure specific heat.

One might take a different viewpoint about these equations. To quote Crighton,[3] we may consider them "simply as model equations, having a certain nominal accuracy, but being regarded effectively as the exact equations governing weakly nonlinear wave propagation in various media and geometrical circumstances. That is the point of view taken here; the structure of the different model equations is of fundamental (and some practical) interest, as are the solutions of initial and boundary value problems for the model equations for shock waves, N waves, and harmonic waves."

In a previous paper,[8] hereafter to be referred to as Paper I, we had proposed that there is a class of nonlinear ordinary differential equations (ODE's)

$$yy'' + ay'^2 + f(x)yy' + g(x)y^2 + by' + c = 0, \tag{1.7}$$

which characterize generalized Burgers equations (GBE's) in the same manner as the Painlevé equation does the KdV-type equations. Equation (1.7) extends the class of nonlinear ODE's studied by Euler and Painlevé (see Kamke[9]) for which $b = c = 0$ and a is a constant, and so we referred to the

 0022-2488/87/050997-08$02.50

solutions of (1.7) as Euler–Painlevé transcendents. In particular, we studied the damped Burgers equation

$$u_t + u^\beta u_x + \lambda u^\alpha = (\delta/2)u_{xx}, \tag{1.8}$$

where $\lambda > 0$ or $\lambda < 0$, and α and β are real constants. The self-similar solutions of (1.8),

$$u = t^{1/(1-\alpha)} f(\eta), \quad \eta = x(2\delta t)^{-1/2}, \tag{1.9}$$

are governed by the nonlinear ODE

$$f'' + 2\eta f' - [4/(1-\alpha)]f - 4(2\delta)^{-1/2} f^{(\alpha-1)/2} f' - 4\lambda f^\alpha = 0, \tag{1.10}$$

provided $\beta = (\alpha-1)/2$. Equation (1.10) can be transformed into

$$HH'' - 2(1+\alpha_1)H'^2 + 2\eta HH' - 2H^2 - 2^{3/2}H' - 2\lambda_1 = 0 \tag{1.11}$$

via

$$H = \delta^{1/2} f^{(1-\alpha)/2},$$

where $\alpha_1 = \frac{1}{2}(3-\alpha)/(\alpha-1)$ and $\lambda_1 = \lambda\delta(1-\alpha)$.

Equation (1.11) is a special case of (1.7). We studied a connection problem for (1.10) which has appropriate (linear) asymptotic behavior at $\eta = +\infty$ and $\eta = -\infty$. Equation (1.11) was also studied in great detail, particularly through its series solution. Apart from the solutions that vanish at $\eta = \pm\infty$, we discovered solutions, in some ranges of the amplitude parameter, which either go to (nonzero) constant value [the exact singular solutions of (1.10)] at $-\infty$ or grow to become unbounded there. We also studied numerically the transition of several initial conditions for (1.8) to the self-similar form governed by (1.10) with vanishing asymptotic conditions at $\eta = \pm\infty$.

We continue our study of generalized Burgers equations to support our claim regarding the role of the class of Euler–Painlevé equations (1.7). We study the nonplanar Burgers equation

$$u_t + u^\alpha u_x + Ju/2t = (\delta/2)u_{xx}, \tag{1.12}$$

wherein we allow the nonlinearity to be general and characterized by the parameter α, the Burgers equations corresponding to $\alpha = 1$. Our study of (1.12) turns out to be very rewarding. In the present case we again seek self-similar solutions of (1.12) in the form $u = t^a f(\eta), a = -1/2\alpha$. Following the same steps as for (1.8) we obtain equations similar to (1.10) and (1.11) (see Sec. II). We study the connection problem for the corresponding equation in f. Equation (1.12) permits considerable analysis. We summarize some of the results. Unlike for (1.8), we are able to get explicit one parameter family of solutions in terms of exponential and error functions for spherically and cylindrically symmetric equations, namely,

$$u_t + u^{1/2}u_x + u/2t = (\delta/2)u_{xx}, \quad J = 1, \tag{1.13}$$

$$u_t + u^{1/3}u_x + u/t = (\delta/2)u_{xx}, \quad J = 2. \tag{1.14}$$

These solutions correspond exactly to the single hump solutions of the Burgers equation. Again the requirement that the single hump solutions of (1.8) vanishing at $\eta = \pm\infty$ exist leads to the conditions that $\alpha = \frac{1}{2}$ for $J = 1$ and $\frac{1}{4} < \alpha < \frac{1}{2}$ for $J = 2$. These conditions correspond to the condition $\alpha > \frac{1}{2}$ for $J = 0$. It is interesting that there is just one value of $\alpha = \frac{1}{2}$ for $J = 1$ for which such solutions exist; this is the value for which an exact explicit solution exists. For $J = 0$ and 2, we have, respectively, an infinite and a finite interval of α, for which the single hump solutions exist. We also have nonzero (singular) constant solutions for $\alpha J = 1$, $J \neq 0$, to which the solutions starting with appropriate asymptotic conditions at $\eta = \infty$ tend as $\eta \to -\infty$. This happens for $\alpha_1 = 1$, $J = 1$, and for $\alpha_2 = \frac{1}{2}$ for $J = 2$. These α values form, in fact, bifurcation points in the sense that the solutions starting at $\eta = \infty$ vanish at a finite value of η if $1/(J+2) < \alpha < 1/(J+1)$, $J = 0,2$; they tend to zero at $\eta = -\infty$ if $1/(J+1) \leqslant \alpha < 1/J$ and to the constant solutions for $\alpha = \alpha_J$. The solutions diverge to become unbounded at $\eta = -\infty$ when $\alpha > \alpha_J$. Thus α becomes the determining parameter for the behavior of the solution at $\eta = -\infty$. We are also able to get an equality involving integrals of F^2 and F'^2 over $-\infty < \eta < \infty$, with $F = f^\alpha$, which helps us decide when the solutions over the whole real line exist.

We note that the nonlinear ODE's for (1.12) corresponding to (1.10) and (1.11) are

$$f'' - 2^{3/2}\delta^{-1/2} f^\alpha f' + 2\eta f' + 2[(1-\alpha J)/\alpha]f = 0 \tag{1.15}$$

and

$$HH'' - [(\alpha+1)/\alpha]H'^2 + 2\eta HH' - 2(1-\alpha J)H^2 - 2^{3/2}H' = 0. \tag{1.16}$$

Equation (1.16) is a special case of (1.7) with $a = -(\alpha+1)/2$, $f(x) = 2x$, $g(x) = -2(1-\alpha J)$, $b = -2^{3/2}$, and $c = 0$. Besides finding the series solution for (1.16), and solving the connection problem for (1.15) numerically we study the transition of the several initial value problems for (1.12) to the self-similar form governed by (1.15) for the appropriate values of parameters for which the latter exist.

Thus the present study fortifies our claim that the class (1.7) does indeed represent the GBE's. Just as there are special cases of Painlevé transcendents that now can be explicitly solved,[10] there are special cases of (1.7), as we have noted above, that can be solved in terms of exponential and error functions. These functions seem to appear prominently as building blocks for (1.7).

The scheme of the present paper is as follows. Section II transforms (1.12) into (1.15) and (1.16) and poses the connection problem. All the analyses for (1.15) and (1.16) is carried out in this section. This includes the exact explicit solutions, the series solutions, and the conditions for the existence of various types of solutions. Section III deals with the numerical study of (1.15) while Sec. IV pertains to that of (1.12). Transition of solutions of the initial value problems for (1.12) to their self-similar form is treated in Sec. V. The conclusions of the study are contained in Sec. VI.

II. ANALYSIS OF SELF-SIMILAR SOLUTIONS—EULER–PAINLEVÉ TRANSCENDENTS

As in Paper I, we find self-similar form of solutions of (1.12) and determine the values of the parameter α for

which these solutions exist satisfying certain asymptotic conditions. Therefore, we write

$$u = t^{a_1} f((2\delta)^{-1/2} t^{b_1} x), \tag{2.1}$$

where a_1 and b_1 are real constants. Substitution of (2.1) into (1.12) shows that, for the similarity form, $a_1 = -1/2\alpha$ and $b_1 = -\frac{1}{2}$ so that (2.1) becomes

$$u = t^{-1/2\alpha} f(\eta), \quad \eta = x(2\delta t)^{-1/2}. \tag{2.2}$$

Equation (1.12) then reduces to

$$f'' - 2^{3/2}\delta^{-1/2} f^{\alpha} f' + 2\eta f' + [2(1-\alpha J)/\alpha] f = 0, \tag{1.15}$$

where a prime denotes differentiation with respect to η. Another change of variable

$$H = \delta^{1/2} f^{-\alpha} \tag{2.3}$$

transforms (1.15) into

$$HH'' - [(\alpha+1)/\alpha] H'^2 + 2\eta HH' - 2(1-\alpha J)H^2 - 2^{3/2} H' = 0. \tag{1.16}$$

As noted in the Introduction, Eq. (1.16) is a special case of (1.7) with $a = -(\alpha+1)/\alpha$, $f(x) = 2x$, $g(x) = -2(1-\alpha J)$, $b = -2^{3/2}$, and $c = 0$. First, we seek a Taylor series solution for H:

$$H(\eta) = \sum_{n=0}^{\infty} a_n \eta^n. \tag{2.4}$$

The coefficients a_k, $k = 2,3,\ldots,n$, are found by substituting (2.4) into (1.16):

$$\begin{aligned} a_2 &= \frac{1}{a_0}\left\{\frac{\alpha+1}{2\alpha} a_1^2 + (1-\alpha J) a_0^2 + 2^{1/2} a_1\right\}, \\ a_{k+2} &= \frac{2}{(k+1)(k+2)a_0}\left[\frac{\alpha+1}{2\alpha}(k+1) a_1 a_{k+1} + (1-\alpha J-k) a_0 a_k + 2^{1/2}(k+1) a_{k+1}\right. \\ &\quad + \sum_{i=1}^{k}\left\{-\frac{1}{2}(k+2-i)(k+1-i) a_i a_{k+2-i} + \frac{\alpha+1}{2\alpha}(i+1)(k+1-i) a_{i+1} a_{k+1-i}\right. \\ &\quad \left.\left. + (1-\alpha J) a_i a_{k-i} - (k-i) a_i a_{k-i}\right\}\right], \quad k = 1,2,\ldots,n. \end{aligned} \tag{2.5}$$

Thus, we have a two-parameter a_0, a_1 family of solutions. For $\alpha = 1/(J+1)$, $J = 0,1,2$, the parameter a_1 is uniquely fixed as $a_1 = -2^{3/2}\alpha/(\alpha+1)$. This special choice corresponds to the exact (explicit) solution we give below [see Eq. (2.13)]. The free parameter a_0 gives a single-parameter family of solutions. This could either be the amplitude parameter or the Reynolds number

$$R = \frac{1}{\delta}\int_{-\infty}^{\infty} u\, dx,$$

which is the ratio of the area under the profile to the coefficient of diffusivity of sound.

We find the asymptotic solution of Eq. (1.15) for large $|\eta|$ under the condition that $f \to 0$ as $\eta \to \pm\infty$. The linearized form of Eq. (1.15), namely,

$$f'' + 2\eta f' + [2(1-\alpha J)/\alpha] f = 0, \tag{2.6}$$

has the solution

$$f(\eta) = A e^{-\eta^2} H_\nu(\eta), \quad \text{for } \eta > 0, \tag{2.7a}$$

$$\sim [B\pi^{1/2}/(-\nu)^{1/2}]\, |\eta|^{J-1/\alpha}, \quad \text{for large negative } \eta, \tag{2.7b}$$

provided $\alpha J < 1$. Here $\nu = 1/\alpha - (J+1)$ and $H_\nu(\eta)$ is the Hermite function of order ν and A and B are the amplitude parameters. Thus, the linear solution decays exponentially as $\eta \to \infty$ and algebraically as $\eta \to -\infty$.

We now pose the boundary value or connection problem for (1.15), namely,

$$f'' - 2^{3/2}\delta^{-1/2} f^{\alpha} f' + 2\eta f' + [2(1-\alpha J)/\alpha] f = 0, \tag{2.8a}$$

$$f \sim A \exp(-\eta^2) H_\nu(\eta) \quad (\eta \uparrow \infty), \tag{2.8b}$$

$$f \to 0 \quad (\eta \downarrow -\infty), \tag{2.8c}$$

and

$$|f| < \infty, \quad -\infty < \eta < \infty.$$

We postpone the discussion of the numerical solution of the connection problem (2.8) to Sec. III. Here we give some exact solutions of Eq. (1.15) for certain special values of α.

For $\alpha = 1/(J+1)$, Eq. (1.15) reduces to

$$f + \eta f' + \tfrac{1}{2} f'' = (2/\delta)^{1/2} f^{\alpha} f'. \tag{2.9}$$

Integrating (2.9) once, we get

$$\eta f + \tfrac{1}{2} f' = [1/(\alpha+1)](2/\delta)^{1/2} f^{\alpha+1} + c_0. \tag{2.10}$$

The constant of integration c_0, however, is zero since f and $f' \to 0$ at $\eta = \infty$, according to (2.8).

Using the transformation $G = f^{-\alpha}$, Eq. (2.10) can be put into the form

$$G' - 2\alpha\eta G = -[2\alpha/(\alpha+1)](2/\delta)^{1/2}. \tag{2.11}$$

Integrating (2.11), we get

$$G = \left(c - \frac{2}{\alpha+1}\left(\frac{2\alpha}{\delta}\right)^{1/2} \int_0^{\alpha^{1/2}\eta} e^{-t^2}\, dt\right) e^{\alpha\eta^2}, \tag{2.12}$$

where c is the constant of integration. Thus

$$f(\eta) = \exp(-\eta^2)\left\{c - \frac{2}{\alpha+1}\left(\frac{2\alpha}{\delta}\right)^{1/2} \times \int_0^{\alpha^{1/2}\eta} e^{-t^2}\, dt\right\}^{-1/\alpha}, \tag{2.13}$$

where $c = f^{-\alpha}(0)$. The solution $u = t^{-1/2\alpha} f(\eta)$ of (1.12) with f as in (2.13), we believe, is new for $J = 1$, $\alpha = \frac{1}{2}$ and for $J = 2$, $\alpha = \frac{1}{3}$, and corresponds to the exact single hump solution for the standard Burgers equation, $J = 0$, $\alpha = 1$.

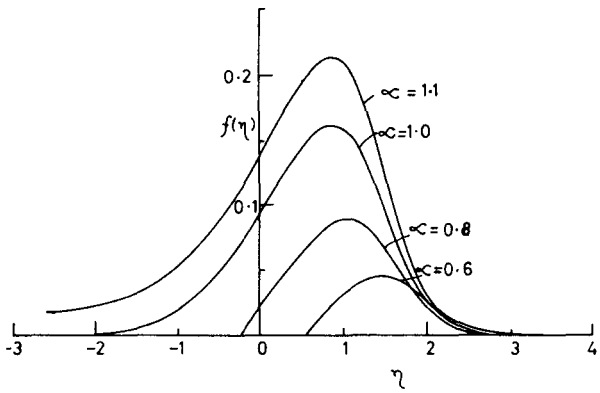


FIG. 1. The solution of the connection problem (2.8), for $J=0$, $\alpha=0.6$, 0.8, 1.0, 1.1, $A=1$.

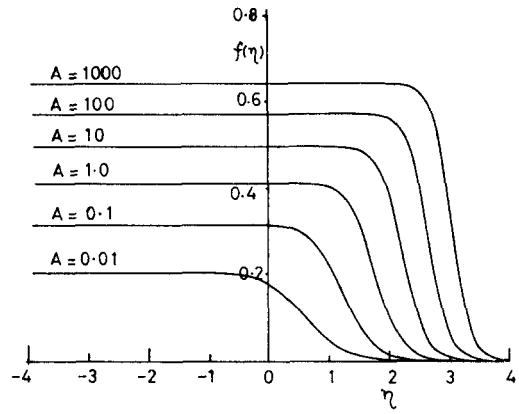


FIG. 2. The solution of the connection problem (2.8), for $J=1$, $\alpha=1$, $A=0.01$, 0.1, 1, 10, 100, 1000.

For $\alpha=1/J, J\neq 0$, f equal to an arbitrary constant ($\neq 0$) is another solution of (2.8).

We now find an equation involving integrals of F^2 and F'^2, which helps determine the parameters for which single hump solutions exist. We substitute $F=f^\alpha$ in (1.15). The resulting ODE in F is

$$\tfrac{1}{2}FF'' - [(\alpha-1)/2\alpha]F'^2 + (1-\alpha J)F^2 + \eta FF' - (2/\delta)^{1/2}F^2F' = 0. \tag{2.14}$$

Integrating (2.14) with respect to η from $-\infty$ to $+\infty$, and assuming that F and F' tend to zero as η tends to $\pm\infty$, we get

$$(2\alpha J-1)\int_{-\infty}^{\infty} F^2\,d\eta = \left(\frac{1-2\alpha}{\alpha}\right)\int_{-\infty}^{\infty} F'^2\,d\eta. \tag{2.15}$$

Equation (2.15) yields the following.

(i) For $J=0$, the ratio

$$r=\frac{\int_{-\infty}^{\infty}F^2\,d\eta}{\int_{-\infty}^{\infty}F'^2\,d\eta} = -(1-2\alpha)/\alpha > 0, \quad \text{if } \alpha > \tfrac{1}{2}.$$

Therefore, the single hump solutions exist if $\alpha>\frac{1}{2}$.

(ii) For $J=1$, the only valid choice is $\alpha=\frac{1}{2}$. This corresponds to the exact solution (2.13).

(iii) For $J=2$, the ratio of the integrals $r=(1-2\alpha)/(\alpha(4\alpha-1))$ is positive if $\frac{1}{4}<\alpha<\frac{1}{2}$. This is the range of α for which single hump solutions exist.

The numerical solution of (2.8) shows that for $\alpha<1/(J+1), J=0,2$, the solution f goes to zero at a finite point, say $\eta=\eta_0$, where $f'>0$ (see Figs. 1 and 2). This is not evident from (2.15). However, integrating (1.15) from η_0 to ∞, we get

$$\frac{\alpha(J+1)-1}{\alpha}\int_{\eta_0}^{\infty} f\,d\eta = -\frac{1}{2}f'(\eta_0) < 0. \tag{2.16}$$

Since $f>0$ for $\eta_0<\eta<\infty$, (2.16) implies that $\alpha<1/(J+1)$. Thus, single hump solutions of (1.15) vanishing at $\eta=+\infty$ and at $\eta=\eta_0$, a finite point on the left, exist only if $\alpha<1/(J+1)$. Combining this result with those in (i)–(iii), we find that single hump solutions vanishing at $\eta=+\infty$ and at $\eta=\eta_0$ exist if $1/(J+2)<\alpha<1/(J+1)$, $J=0,2$.

We note that the function f has a maximum where $f'=0, f''<0$ if $\alpha J<1$ [see Eq. (2.8)]. However, this condition is too lax. The consideration of (2.15) and (2.16) leads to very precise ranges of α for which the single hump solutions exist, as we have delineated in (i)–(iii) above. These conclusions are confirmed numerically in Sec. III.

III. NUMERICAL SOLUTION OF CONNECTION PROBLEM (2.8)

We solved Eq. (2.8) numerically starting from $\eta\sim 4$ [when f and f' are very small $O(10^{-5})$] and carried the solution to $\eta\to-\infty$ for all α for which the single hump self-similar solution exists. For a given α and $J=1,2$, the values of f and f' at $\eta=0$ were obtained and those of H and H' were computed from (2.3). The series (2.5) was then summed up. The series solution so obtained agreed closely with the numerical solution of the connection problem (2.8). It was found to be accurate to seven decimal places in single precision arithmetic over the entire range from $\eta=-\infty$ to $\eta=+\infty$. We used analytic continuation of the series solution as the convergence of the series slowed down. In particular the numerical and series solutions compared very well with exact solution (2.13) for $J=1$, $\alpha=\frac{1}{2}$ and $J=2$, $\alpha=\frac{1}{3}$ [see Table I for the solution of (1.15) for

TABLE I. Exact analytic solution (2.13), numerical solution of ODE (1.15), and series solution (2.4) of (2.9) for $J=1$, $\alpha=\frac{1}{2}$.

	Analytic	Numerical	Series	
η	f	f	H	f
−3.0	0.000 0004	0.000 0004	224.697 9	0.000 0004
−2.5	0.000 0062	0.000 0062	56.660 64	0.000 0062
−2.0	0.000 0610	0.000 0610	18.108 80	0.000 0610
−1.5	0.000 3828	0.000 3828	7.228 297	0.000 3828
−1.0	0.001 6222	0.001 6222	3.511 169	0.001 6223
−0.5	0.004 9414	0.004 9414	2.011 814	0.004 9414
0.0	0.011 4275	0.011 4274	1.322 943	0.011 4274
0.5	0.020 5568	0.020 5566	0.986 3665	0.020 5566
1.0	0.027 6067	0.027 6063	0.851 1571	0.027 6064
1.5	0.023 5493	0.023 5487	0.921 5727	0.023 5488
2.0	0.009 6266	0.009 6262	1.441 398	0.009 6264
2.5	0.001 5869	0.001 5868	3.550 143	0.001 5869
3.0	0.000 1182	0.000 1182	13.007 45	0.000 1182
3.5	0.000 0048	0.000 0048	64.873 94	0.000 0048

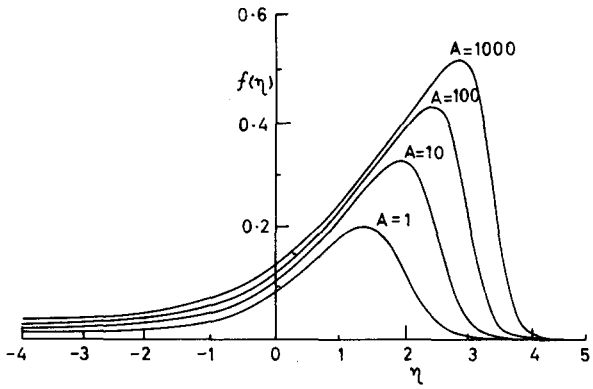


FIG. 3. The solution of the connection problem (2.8), for $J=0$, $\alpha=1.1$, $A=1$, 10, 100, 1000.

TABLE II. Single hump, monotonic, and diverging solutions of (1.15).

Behavior at left boundary	$J=0$	$J=1$	$J=2$
Solutions vanishing at $\eta=-\infty$	$\alpha=1$	$\alpha=\frac{1}{2}$	$\frac{1}{3}\leqslant\alpha<\frac{1}{2}$
Solutions vanishing at $\eta=\eta_0$	$\frac{1}{2}<\alpha<1$	⋯	$\frac{1}{4}<\alpha<\frac{1}{3}$
Solutions monotonically approaching a constant at $\eta=-\infty$	⋯	$\alpha=1$	$\alpha=\frac{1}{2}$
Solutions diverging to infinity at $\eta=-\infty$	⋯	$\alpha>1$	$\alpha>\frac{1}{2}$

$J=1$, $\alpha=\frac{1}{2}$]. For $J=0$, and $\frac{1}{2}<\alpha<1$, $f(\eta)$ vanishes at a finite point $\eta=\eta_0$ while for $\alpha\geqslant 1$, $f(\eta)\to 0$ as $n\to-\infty$ (see Figs. 1 and 3). Within the permissible (similarity) range of α, single hump solutions vanishing at $\pm\infty$ (or at $+\infty$ and a finite point η_0) exist, independent of the amplitude parameter A (see Fig. 3). For cylindrical symmetry, $J=1$, $f\to 0$ as $\eta\to-\infty$ if $\alpha=\frac{1}{2}$, $f\to\text{const}\neq 0$ if $\alpha=1$ and $f\to\infty$ if $\alpha>1$. For spherical symmetry, $J=2$, $f\to 0$ at a finite point $\eta=\eta_0$ if $\frac{1}{4}<\alpha<\frac{1}{3}$, $f\to 0$ as $\eta\to-\infty$ if $\frac{1}{3}\leqslant\alpha\leqslant\frac{1}{2}$, $f\to\text{const}\neq 0$ if $\alpha=\frac{1}{2}$ and $f\to\infty$ if $\alpha>\frac{1}{2}$ [see Table II for a summary of the nature of the solution for different J and α values and Fig. 4 for the solution of Eq. (2.8) for $J=2$]. Thus for $\alpha=1/J$, $J=1,2$, $f\to f_c$, a constant $\neq 0$ as $\eta\to-\infty$ for finite values of A (see Figs. 2 and 5). For $J=1$ and $A=1$, $f_c=0.411\,87$ and for $J=2$ and $A=1$, $f_c=0.101\,97$. For $\alpha=1/(J+1)$ and $J=0,1,2$, (2.15) was satisfied very accurately up to six decimals.

IV. NUMERICAL SOLUTION OF THE GENERALIZED BURGERS EQUATION (1.12)

We solve Eq. (1.12) subject to the initial conditions

$$u(x,t_i)=\begin{cases}0, & x<x_0,\\ g(x), & x_0<x<x_1,\\ 0, & x>x_1,\end{cases}\tag{4.1}$$

where the function $g(x)$ has the typical forms shown in Fig. 6. Since the numerical schemes for nonlinear parabolic equations of Burgers type have been discussed in detail in paper I and in Sachdev, Nair, and Tikekar,[11] we restrict ourselves to the specific discussion of Eq. (1.12). We use the pseudospectral scheme when the initial profile is discontinuous and implicit predictor–corrector[12] scheme when it is continuous. The difference analog of Eq. (1.12) is

$$u_{i+1,j+1/2}-2\left(1+\frac{2h^2}{k\delta}\right)u_{i,j+1/2}+u_{i-1,j+1/2}=\frac{Jh^2}{\delta t_{j+1/2}}u_{i,j}-\frac{4h^2}{\delta k}u_{i,j}+\frac{h}{\delta}u^{\alpha}_{i,j}(u_{i+1,j}-u_{i-1,j})\quad\text{(predictor)},\tag{4.2}$$

$$\left(1-\frac{h}{\delta}u^{\alpha}_{i,j+1/2}\right)u_{i+1,j+1}-2\left(1+\frac{2h^2}{\delta k}\right)u_{i,j+1}+\left(1+\frac{h}{\delta}u^{\alpha}_{i,j+1/2}\right)u_{i-1,j+1}=\left(\frac{h}{\delta}u^{\alpha}_{i,j+1/2}-1\right)u_{i+1,j}+2\left(1-\frac{2h^2}{\delta k}\right)u_{i,j}-\left(1+\frac{h}{\delta}u^{\alpha}_{i,j+1/2}\right)u_{i-1,j}+\frac{2Jh^2}{\delta t_{j+1/2}}u_{i,j+1/2}\quad\text{(corrector)}.\tag{4.3}$$

Here, $u_{i,j}=u(i\Delta x,j\Delta t)$ and $h=\Delta x$ and $k=\Delta t$ are spatial and time mesh sizes, respectively. The difference scheme has a truncation error $O(\Delta x^2+\Delta t^2)$. However, this scheme is not adequate to solve (1.12) if the initial profile is discontinuous and we wish to visualize the evolution of the shock wave through its embryonic shock region. The reason is that the accuracy of the solution of (1.12) with an initial discontinuous profile is severely affected by the implicit scheme (4.2) and (4.3). Therefore, we take recourse to the pseudospectral scheme. The essence of the pseudospectral scheme is that the spatial derivatives u_x, u_{xx} of the distribution $u(x,t)$ are computed very accurately by the finite Fourier transform. The finite Fourier transform of $u(x,t)$ is

$$\bar{u}(k_j,t)=\frac{1}{K}\sum_{m=0}^{K-1}u(m\Delta x,t)\exp(-ik_jm\Delta x)\tag{4.4}$$

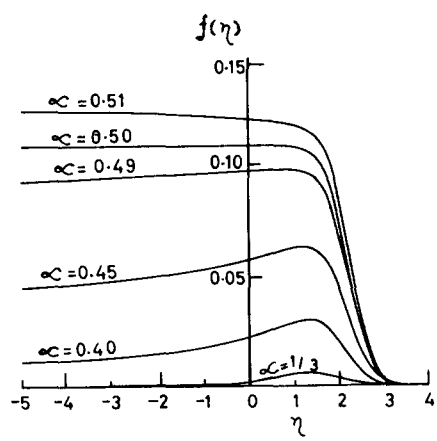


FIG. 4. The solution of the connection problem (2.8), for $J=2$, $\alpha=0.4$, 0.45, 0.49, 0.5, 0.51, $A=1$.

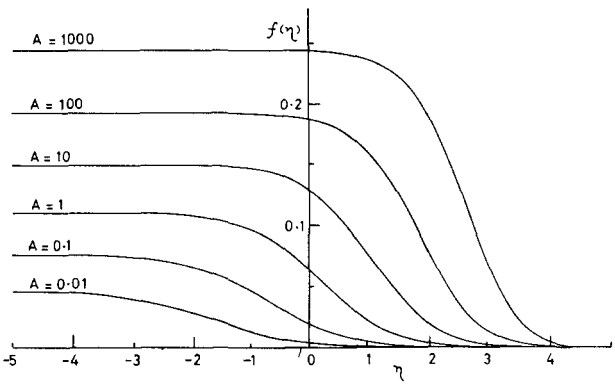


FIG. 5. The solution of the connection problem (2.8), for $J=2$, $\alpha=\frac{1}{2}$, $A=0.01, 0.1, 1, 10, 100, 1000$.

holding over the interval $(0,2\pi)$ of x. Here $\Delta x = 2\pi/K$ is the spatial mesh size, K denotes the number of mesh points, and the k_j are the wave numbers varying between 0 and $K-1$. The inverse Fourier transform is defined as

$$u(m\Delta x,t) = \sum_{|k_j|<K/2} \bar{u}(k_j,t)\exp(ik_j m\Delta x). \tag{4.5}$$

The spatial derivatives at the mesh points are

$$u_x(m\Delta x,t) = \sum_{|k_j|<K/2} (ik_j)\bar{u}(k_j,t)\exp(ik_j m\Delta x), \tag{4.6}$$

$$u_{xx}(m\Delta x,t) = \sum_{|k_j|<K/2} (ik_j)^2\bar{u}(k_j,t)\exp(ik_j m\Delta x). \tag{4.7}$$

The solution $u(x,t+\Delta t)$ at the next time level $t+\Delta t$ is obtained from truncated Taylor series

$$u(x,t+\Delta t) = u(x,t) + \Delta t u_t + \frac{\Delta t^2}{2!}u_{tt} + \frac{\Delta t^3}{3!}u_{ttt} + \cdots, \tag{4.8}$$

wherein the time derivatives u_t, u_{tt}, etc. are substituted from (1.12) and its time derivatives as follows:

$$\begin{aligned} u_t &= -u^\alpha u_x - \frac{Ju}{2t} + \frac{\delta}{2}u_{xx}, \\ u_{tt} &= -u^\alpha u_{tx} - \alpha u^{\alpha-1}u_t u_x - \frac{Ju_t}{2t} + \frac{Ju}{2t^2} + \frac{\delta}{2}u_{txx}, \\ u_{ttt} &= -u^\alpha u_{ttx} - 2\alpha u^{\alpha-1}u_t u_{tx} - \alpha(\alpha-1)u^{\alpha-2}u_t^2 u_x \\ &\quad - \alpha u^{\alpha-1}u_{tt}u_x - \frac{Ju_{tt}}{2t} + \frac{Ju_t}{t^2} - \frac{Ju}{t^3} + \frac{\delta}{2}u_{ttxx}. \end{aligned} \tag{4.9}$$

As the computation commenced we noticed a tail of negative amplitude immediately after the nonzero part of the profile. By choosing a time mesh as small as 0.0001, the magnitude of the tail was brought down to less than 0.000 01. Being spurious and negligible, the tail was artificially cut off. The tail in the subsequent steps was much smaller in magnitude and as the profile became smoother, due to diffusion and decay, the time mesh was increased in steps. We switched over to the implicit scheme when the profile became very smooth; the time mesh was increased to 0.01.

V. TRANSITION OF INITIAL VALUE PROBLEMS TO SELF-SIMILAR FORM OR INTERMEDIATE ASYMPTOTICS

We solved Eq. (1.12) with both continuous and discontinuous initial profiles (see Fig. 6), for $J=0,1,2$ and for the

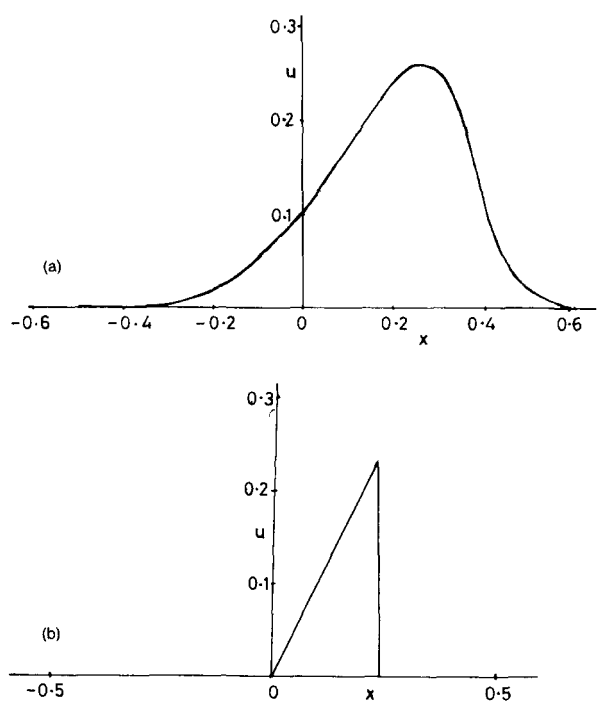


FIG. 6. The initial profiles for solving Eq. (1.12). (a) Continuous (single hump), (b) discontinuous.

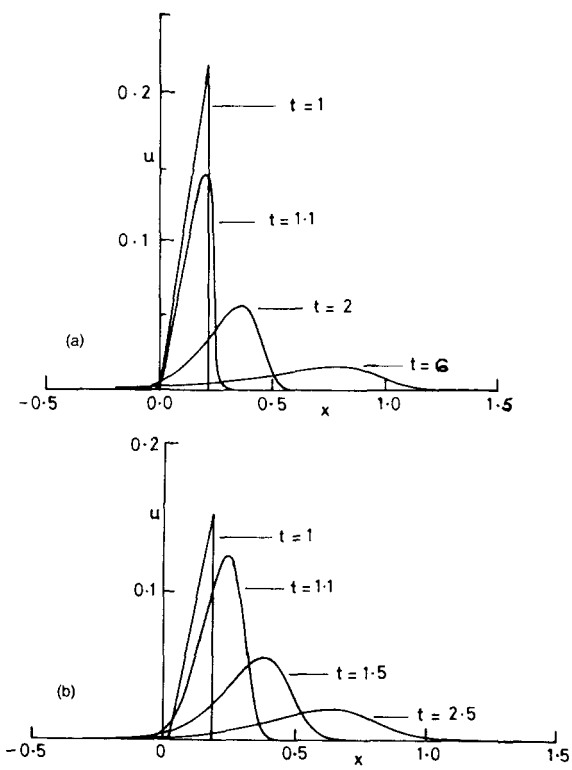


FIG. 7. The solution of Eq. (1.12). (a) $J=1$, $\alpha=\frac{1}{2}$, $\delta=0.01$. (b) $J=2$, $\alpha=\frac{1}{3}$, $\delta=0.02$.

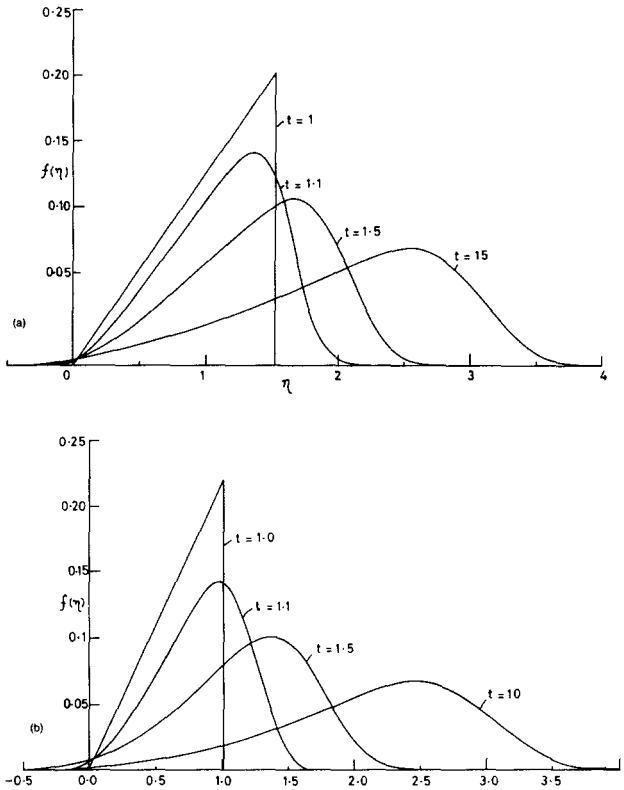


FIG. 8. The evolution of the self-similar form. (a) $J = 1$, $\alpha = \frac{1}{2}$, $\delta = 0.01$. (b) $J = 2$, $\alpha = \frac{1}{3}$, $\delta = 0.02$.

parameter α in the self-similar regime. The initial profile soon evolves into its self-similar form. Figures 7(a) and 7(b) depict the evolution and decay of single hump discontinuous initial profiles, for $J = 1,2$. Figures 8(a) and 8(b) show the evolution of these profiles into their self-similar forms. Since it is not possible to describe too precisely the evolution of these profiles into self-similar forms graphically, we have presented in Table III the maximum of u and its location at various times as well as the values of $\eta_{\max} = x_{\max}(2\delta t)^{-1/2}$. The approach of $\eta_{\max}$ and $f_{\max} = f(\eta_{\max})$ to constant values as the self-similar regime sets in are also manifest in the table.

TABLE III. Onset of self-similar form (2.2) of (1.12).

t	$x_{\max}$	$u_{\max}$	$\eta_{\max}$	$f_{\max}$
(a) $J = 1$, $\alpha = \frac{1}{2}$, $\delta = 0.01$				
1.00	0.22	0.218	1.54	0.218
1.01	0.20	0.193	1.41	0.195
1.10	0.21	0.147	1.38	0.162
1.50	0.29	0.085	1.65	0.127
2.00	0.38	0.057	1.88	0.113
3.00	0.52	0.034	2.12	0.102
4.00	0.63	0.024	2.23	0.098
5.00	0.73	0.019	2.31	0.095
6.00	0.82	0.016	2.37	0.093
7.00	0.90	0.013	2.41	0.092
8.00	0.97	0.011	2.42	0.091
9.00	1.04	0.010	2.45	0.090
10.00	1.11	0.009	2.48	0.090
11.00	1.17	0.008	2.49	0.089
12.00	1.23	0.007	2.51	0.089
13.00	1.29	0.007	2.53	0.089
14.00	1.34	0.006	2.53	0.088
15.00	1.39	0.006	2.54	0.088
(b) $J = 2$, $\alpha = \frac{1}{3}$, $\delta = 0.02$				
1.00	0.22	0.218	1.09	0.218
1.01	0.20	0.183	0.97	0.186
1.10	0.21	0.124	0.99	0.143
1.50	0.34	0.055	1.37	0.101
2.00	0.47	0.031	1.64	0.087
2.50	0.58	0.020	1.82	0.081
3.00	0.68	0.015	1.95	0.077
4.00	0.85	0.009	2.11	0.073
5.00	1.00	0.006	2.23	0.071
6.00	1.14	0.005	2.32	0.069
7.00	1.26	0.004	2.37	0.068
8.00	1.37	0.003	2.41	0.068
10.0	1.58	0.002	2.49	0.067
12.0	1.76	0.002	2.53	0.067
14.0	1.93	0.001	2.57	0.065

VI. CONCLUSIONS

We have studied the nonplanar GBE (1.12) with general nonlinearity and found its self-similar solutions. The form (1.16) again falls in the class of ODE's (1.7) whose solutions we have referred to in Paper I as Euler–Painlevé transcendents. Thus, two GBE's have representations via group theoretic methods or similarity transformations in terms of the Euler–Painlevé equations. This fortifies our conjecture regarding the connection between Euler–Painlevé transcendents and GBE's. However, we would like to caution that this has been confirmed only for two equations, unlike for the case of the Painlevé equations, which have been shown to represent a very large number of (model) nonlinear dispersive equations. More GBE's would have to be analyzed if and when they arise in applications to see that (1.7) does indeed represent GBE's.

Equation (1.12) permits much more analysis than (1.8), which we discussed in paper I. Here we have a single-parameter family of exact explicit solutions, for all geometries $J = 0,1,2$, which extend the well-known ones for the plane Burgers equation [see Eq. (2.13)]. The integral equalities (2.15) and (2.16) clearly restrict the ranges of the parameter α for which the single hump solutions vanishing at $\eta = +\infty$, and $\eta = -\infty$ or $\eta = \eta_0$, exist. We have confirmed these conclusions by numerically solving (1.15) subject to (2.8b). We have also solved the original PDE (1.12) directly to visualize the transition of initial conditions to self-similar forms in the range of α, for which the latter exist. Thus the ODE's (1.15) and (1.16) have been shown to be of considerable importance from both mathematical and physical points of view.

[1]E. R. Benton and G. W. Platzman, Q. Appl. Math. **30**, 195 (1972).
[2]P. L. Sachdev, *Nonlinear Diffusive Waves* (Cambridge U.P., Cambridge, 1987).
[3]D. G. Crighton, Ann. Rev. Fluid Mech. **11**, 11 (1979).

[4]*Nonlinear Waves*, edited by S. Leibovich and A. R. Seebas (Cornell U.P., New York, 1974).
[5]J. D. Murray, J. Fluid Mech. **59**, 263 (1973).
[6]R. W. Lardner and J. C. Arya, Acta Mech. **37**, 197 (1980).
[7]A. A. Karabutov and O. V. Rudenko, Sov. Phys. Tech. Phys. **20**, 920 (1975).
[8]P. L. Sachdev, K. R. C. Nair, and V. G. Tikekar, J. Math. Phys. **27**, 1506 (1986).
[9]E. Kamke, *Differential gleichungen: Losungs-methoden und Losungen* (Akademische, Leipzig, 1943).
[10]V. I. Gromak and N. A. Lukashevich, Diff. Eq. **8**, 317 (1982).
[11]P. L. Sachdev, K. R. C. Nair, and V. G. Tikekar, J. Fluid Mech. **172**, 347 (1986).
[12]J. Douglas and B. F. Jones, J. Soc. Ind. Appl. Math. **11**, 195 (1963).

R-separation of variables for the time-dependent Hamilton–Jacobi and Schrödinger equations

E. G. Kalnins
Mathematics Department, University of Waikato, Hamilton, New Zealand

Willard Miller, Jr.
School of Mathematics, University of Minnesota, Minneapolis, Minnesota 55455

(Received 3 November 1986; accepted for publication 7 January 1987)

The theory of R-separation of variables is developed for the time-dependent Hamilton–Jacobi and Schrödinger equations on a Riemannian manifold V^n where time-dependent vector and scalar potentials are permitted. As an application it is shown how to obtain all R-separable coordinates for the n-sphere and Euclidean n-space.

I. INTRODUCTION AND TECHNICAL CONSIDERATIONS

In the study of additive separation or R-separation of variables for Hamilton–Jacobi equations on pseudo-Riemannian manifolds one naturally distinguishes three types of equations:

$$\text{(I)}\quad \sum_{l,m} g^{lm} W_{x^l} W_{x^m} = E,$$

$$\text{(II)}\quad \sum_{l,m} g^{lm} W_{x^l} W_{x^m} + 2\lambda W_t = 0,$$

$$\text{(III)}\quad \sum_{l,m} g^{lm} W_{x^l} W_{x^m} = 0.$$

Here (g^{lm}) is the contravariant metric tensor with respect to the coordinate system $\{x^i\}$ on a Riemannian or pseudo-Riemannian manifold and E, λ are nonzero parameters. (We can also add vector and scalar potentials to the left-hand sides of each of these equations, since this is only a minor complication from the viewpoint of variable separation.) See Refs. 1 and 2 for discussions of the relevance of these equations to classical mechanics. Although (I) can be considered as a special case of (II), and (II) as a special case of (III) (in a space of two more dimensions), the three types of equations exhibit distinct forms of behavior. In particular, (II) has proved much more difficult to analyze from the viewpoint of variable separation than have (I) and (III).

For a given Hamilton–Jacobi equation, variable separation research has typically divided into three categories: (a) explicit determination of separable systems and application of these results to derive explicit solutions of the equation; (b) intrinsic, i.e., coordinate-free, characterizations of separable coordinate systems and their relation to completely integrable Hamiltonian systems; and (c) studies of the "quantization problem," the relationship between additively separable solutions of the Hamilton–Jacobi equation and multiplicatively R-separable solutions of the associated Schrödinger equation,

$$\text{(I')}\quad \Delta\psi = E\psi,$$

$$\text{(II')}\quad \Delta\psi + 2\lambda i\psi_t = 0,$$

$$\text{(III')}\quad \Delta\psi = 0,$$

where Δ is the Laplace–Beltrami operator on the pseudo-Riemannian manifold.

For equations of types (I) and (III) considerable recent progress has been made in all three of the preceding categories. (See Refs. 3 and 4 for reviews of this work.) The present paper is a contribution to category (a) for equations of type (II). Shapovalov has already announced the solution to the category (b) problem for all these equations, see Ref. 3.

In the latter half of this section we point out the sense in which (II) (with added time-dependent vector and scalar potentials) is a special case of (III) and use this connection to work out the technical conditions for a coordinate system to be R separable for (II). We show that corresponding to each R-separable coordinate system $\{y^i\}$ for (II) on a Riemannian manifold V^n there is associated a unique "time" coordinate y^1 and that the transformed equation in these new coordinates is again in Hamilton–Jacobi form (II) on the same manifold V^n. The transformed Hamiltonian and potential may, however, depend on the new time coordinate y^1. If there is no dependence of the Hamiltonian and potential on y^1 (the *regular* case) then we can use Lie theoretic methods to analyze such coordinates.[5–7]

In Sec. II we turn to the principal topic of this paper, the case where the transformed Hamiltonian $\mathscr{H}(y^1)$ is strictly y^1 dependent. We determine all such time-dependent Hamiltonians for the n-sphere S^n and Euclidean n-space E^n, and in Sec. III we show how to compute all of the associated R-separable coordinate systems for II (with added time-dependent potentials) on these manifolds.

The solution of the regular case for S^n and E^n is taken up in Sec. IV. In Sec. V it is shown that all our results extend to the time-dependent Schrödinger equations on S^n and E^n. Finally, in Sec. VI we give an intrinsic characterization of those equations of type (III) for which coordinates $\{t,x^i\}$ can be chosen such that (III) restricts to (II). All functions appearing in this paper are assumed to be locally analytic.

Technically our task is to analyze the possible R-separable solutions for the time-dependent Hamilton–Jacobi equation

$$2\lambda W_t + \sum_{l,m=1}^{n} g^{lm}(\mathbf{x}) W_{x^l} W_{x^m} + 2\lambda \sum_{l=1}^{n} A^l(\mathbf{x},t) W_{x^l} + \lambda^2 V(\mathbf{x},t) = 0, \quad g^{lm} = g^{ml}. \tag{1.1}$$

Here λ is a parameter, $\{x^i\}$ is a local coordinate system, and $g^{lm}(\mathbf{x})$ the contravariant metric tensor on the Riemannian

 0022-2488/87/051005-11$02.50

manifold V^n. In particular, the matrix $\{g^{lm}\}$ is positive definite. A solution of (1.1) is a function $W = W(t,x)$ that satisfies this equation.

We must first state precisely which transformations will be permitted in the search for separable solutions. We will do this by considering (1.1) as a special case of the (conformal) Hamilton–Jacobi equation

$$\sum_{u,v=1}^{n+2} K^{uv}(z) Z_{z^u} Z_{z^v} = 0, \tag{1.2}$$

where

$$z^i = x^i, \quad i = 1,...,n, \quad z^{n+1} = t, \quad z^{n+2} = \tau,$$
$$Z = \lambda\tau + W, \quad K^{i,n+2} = K^{n+2,i} = A^i,$$
$$K^{ij} = g^{ij}, \quad 1 \leqslant i,j \leqslant n,$$
$$K^{n+1,n+2} = K^{n+2,n+1} = 1, \quad K^{n+2,n+2} = V,$$

and all other matrix elements of K^{uv} vanish. Thus, the solutions of (1.1) can be identified with those solutions Z of (1.2) for which $Z_\tau = \lambda$ (after which we set $\tau = 0$).

The general theory of variable separation for the Hamilton–Jacobi equation (1.2) (and its relation to Lie symmetries) is well understood[8,9] and we need only modify this theory to the special requirement $Z_\tau = \lambda$. In the following paragraphs we present the modification.

We pass to separable coordinates $y^1,...,y^{n+1}, \mu$, where

$$x^k = x^k(\mathbf{y}), \quad k = 1,...,n,$$
$$t = t(\mathbf{y}), \quad \tau = \mu - R(\mathbf{y}), \tag{1.3}$$

and R is a function to be determined. Then (1.2) transforms to

$$2Z_\mu([\xi^i + \mathscr{A}^i]Z_i + G^{ij}R_iZ_j) + G^{ij}Z_iZ_j + (2[\xi^i + \mathscr{A}^i]R_i + V + G^{ij}R_iR_j)Z_\mu Z_\mu = 0, \tag{1.4}$$

where we observe the Einstein summation convention, the variables i,j take the values $1,2,...,n+1$; $Z_i = Z_{y_i}, R_i = R_{y_i}$, and $A^l(\partial y^i/\partial x^l) = \mathscr{A}^i$. Note that

$$\sum_{l,m=1}^{n} g^{lm}(\mathbf{x}) W_{x^l} W_{x^m} = G^{ij}(\mathbf{y}) Z_i Z_j, \quad Z_\mu = \lambda. \tag{1.5}$$

The separable coordinates $\mathbf{y},\mu$ are of three types: there are n_1 first kind variables y^a, n_2 second kind variables y^r, and n_3 ignorable variables y^α and μ; $n_1 + n_2 + n_3 = n + 2$. The contravariant metric tensor for Eq. (1.4), expressed in these coordinates, must be of the form

$$\begin{array}{c} \\ n_1 \\ n_2 \\ n_3 - 1 \\ \mu \end{array} \begin{array}{c} \begin{array}{cccc} n_1 & n_2 & n_3 - 1 & \mu \end{array} \\ \begin{pmatrix} QH_a^{-2}\delta^{ab} & 0 & 0 & 0 \\ 0 & 0 & Qf_r^\beta(y^r)H_r^{-2} & Qk_r(y^r)H_r^{-2} \\ 0 & Qf_s^\alpha(y^s)H_s^{-2} & Q\sum_i K_i^{\alpha\beta}(y^i)H_i^{-2} & Q\sum_i F_i^\alpha(y^i)H_i^{-2} \\ 0 & Qk_s(y^s)H_s^{-2} & Q\sum_i F_i^\beta(y^i)H_i^{-2} & Q\sum_i \mathscr{F}_i(y^i)H_i^{-2} \end{pmatrix} \end{array}. \tag{1.6}$$

Here there is no summation on n repeated indices unless explicitly indicated, and the indicated sums are $i = 1,...,n_1 + n_2$. (This expression follows immediately from Theorem 5 of Ref. 9.) The metric

$$ds'^2 = \sum_{a=1}^{n_1} H_a^2(y^b,y^s)(dy^a)^2 + \sum_{r=n_1+1}^{n_1+n_2} H_r^2(y^b,y^s)(dy^r)^2 \tag{1.7}$$

must be in Stäckel form. The matrix elements are independent of the ignorable variables y^α, μ.

Comparing (1.4) with (1.6) we have the following conditions:

$$G^{ab} = QH_a^{-2}\delta^{ab}, \quad G^{rs} = G^{ar} = 0,$$
$$G^{\alpha\beta} = Q\sum_i K_i^{\alpha\beta}H_i^{-2}, \quad G^{\alpha\alpha} = 0, \tag{1.8}$$
$$G^{r\alpha} = Qf_r^\alpha H_r^{-2},$$

and

$$\xi^a + \mathscr{A}^a + QH_a^{-2}R_a = 0,$$
$$\xi^r + \mathscr{A}^r + QH_r^{-2}\sum_\beta f_r^\beta R_\beta = Qk_rH_r^{-2},$$
$$\xi^\alpha + \mathscr{A}^\alpha + Q\sum_{\beta,i} K_i^{\alpha\beta}H_i^{-2}R_\beta + Q\sum_s f_s^\alpha H_s^{-2}R_s = Q\sum_i F_i^\alpha H_i^{-2}, \tag{1.9}$$
$$2\sum_i [\xi^i + \mathscr{A}^i]R_i + V + \sum_{i,j} G^{ij}R_iR_j = Q\sum_i \mathscr{F}_iH_i^{-2}.$$

Here $Q = Q(\mathbf{y}) \neq 0$ and each of f_r^β, k_r, $K_i^{\alpha\beta}$, F_i^α, $\mathscr{F}_i$ depend only on the variable denoted by the subscript. Finally, we have that $\xi^i p_i$ is a Killing vector for the Hamiltonian $G^{ij}p_ip_j$, i.e.,

$$\{\xi^i p_i, G^{ij}p_ip_j\} = 0, \tag{1.10}$$

where $\{\cdot,\cdot\}$ is the Poisson bracket in the canonical coordinates $y^i, \mu; p_i, p_\mu$, and there is a closed one-form $dt = df = f_i\,dy^i$ such that

$$f_i\xi^i = 1, \quad f_iG^{ij} = 0, \quad j = 1,...,n+1. \tag{1.11}$$

Conditions (1.7)–(1.11) are necessary and sufficient for R-separation of (1.1) in the coordinates y^i. The R-separable

solutions take the form

$$W = +\lambda R(\mathbf{y}) + \sum_{i=1}^{n+1} W^{(i)}(y^i), \tag{1.12}$$

where R satisfies conditions (1.9). (It is the presence of a possibly nontrivial R which leads to the term *R-separable*; if $\partial_{y^i y^j} R = 0$, $i \neq j$, the system $\{y^k\}$ is *separable*.)

We can simplify our problem somewhat by noting from (1.5) and (1.8) and the requirement (g^{lm}) positive definite that $n_2 \leqslant 1$.

Theorem 1: If the time-dependent Hamilton–Jacobi equation (1.1) is R-separable in the coordinates $\{y^i\}$ then the transformed equation (1.4) can be put in time-dependent Hamilton–Jacobi form (with potential)

$$2\lambda Z_f + \sum_{i,j \neq f} G^{ij} Z_i Z_j + \lambda \sum_{\alpha \neq f} U^\alpha Z_\alpha + \lambda^2 U = 0, \tag{1.13}$$

where $df = dt$. There are two possibilities:

Case 1: $n_2 = 0$. Then $f = y^\delta$ is ignorable and $\partial_f G^{ij} = \partial_f U^\alpha = \partial_f U = 0$. The metric (G^{ij}), $i,j \neq \delta$ determines the *same* Riemannian space V^n as does (g^{lm}).

Case 2: $n_2 = 1$. Then $f = y^s$ is a second kind coordinate and at least one of G^{ij}, U^α, U has nontrivial f dependence. For each fixed value of y^s the metric $(G^{ij}(y^s))$, $i,j \neq s$, determines the *same* Riemannian space V^n as does (g^{lm}).

Proof: Suppose (1.1) is R-separable in the coordinates $\{y^i\}$. Then from (1.11) there is a function f such that $df = dt = f_i\, dy^i$, where $f_i \xi^i = 1$ and

$$f_i G^{ij} = 0, \quad j = 1,\ldots,n+1. \tag{1.14}$$

It follows immediately from (1.8) and (1.14) that $f_a = 0$ for $a = 1,\ldots,n_1$. Furthermore, since $\mathrm{Rank}(G^{ij}) = n$ and $\partial_{y^\beta} G^{ij} = 0$ for each ignorable variable y^β, we must have $\partial_{y^\beta}(f_i/f_j) = 0$ whenever $f_j \neq 0$. It follows that f must be of the form $f = h(u,v)$ where

$$u = \sum_\alpha C_\alpha(y^r) y^\alpha, \quad v = y^r.$$

(If $n_2 = 1$ then f, C_α may depend on the single second kind coordinate y^r; if $n_2 = 0$ then the C_α must be constants.)

Now suppose $n_2 = 1$ and $\partial_u h = 0$. Then $f = f(y^r)$ and from the requirement $f_i \xi^i = f_r \xi^r = 1$ we see that $\partial_{y^i} \xi^r = 0$ for $i \neq r$. Thus, by a change of second kind coordinate $y^r = k(y^r)$ if necessary (which preserves separation), we can assume $\xi^r \equiv 1$ and $\mathscr{A}^r = 0$. Further, the condition $f_r G^{rj} = 0$ implies $G^{rj} = 0$ for $j = 1,\ldots,n+1$, so the $n \times n$ matrix (G^{ij}), $i,j \neq r$ is nonsingular. Thus Eq. (1.4) takes the form (1.13) and since $f = y^r$ is not ignorable, at least one of G^{ij}, U^α, U has nontrivial f dependence.

Next, suppose $n_2 = 1$ and $\partial_u h \neq 0$ where at least one of the C_α is nonzero. Without loss of generality we can assume $C_\gamma = 1$ for fixed ignorable variable y^γ. Then

$$\partial_{y^\beta}\left(\frac{f_r}{f_\gamma}\right) = \partial_{y^\beta}\left(\frac{h_u C'_\delta y^\delta + h_v}{h_u}\right)$$

$$= C'_\beta + C_\beta\, \partial_u\left(\frac{h_v}{h_u}\right) = 0$$

since f_r/f_γ is independent of each ignorable variable y^β. Setting $\beta = \gamma$ we have $\partial_u(h_v/h_u) = 0$. Thus $C'_\beta \equiv dC_\beta/dy^r = 0$ for each β and the C_β are constants. Further, we have $f = h(u,v) = H(u + K(v))$ for some function K. We can now pass to a new set $\{y^{r'}\}$ of equivalent R-separable coordinates such that $y^{\delta'} = C_\beta y^\beta + K(y^r)$. (See Ref. 10 for a discussion of the pseudogroup of transformations taking separable coordinates into equivalent systems of separable coordinates.) Dropping the primes, we have $f = f(y^\delta)$ and, using (1.11), $\xi^\delta = \xi^\delta(y^\delta) \neq 0$, $G^{\delta i} = 0$, $i = 1,\ldots,n+1$ and $\mathscr{A}^\delta = 0$. From the third equation of (1.9) we have

$$\xi^\delta = Q \sum_{i=1}^{n_1+n_2} F_i^\delta(y^i) H_i^{-2} = QF \neq 0.$$

Since F is a Stäckel multiplier (see Ref. 9) we can pass to an equivalent Stäckel form $\tilde{H}_i^2 = FH_i^2$ so that $\tilde{Q} = \xi^\delta$. Dividing (1.4) by the common factor $\xi^\delta = \tilde{Q}$ [see (1.8) and (1.9)] we obtain (1.13), where each term is independent of $f = y^\delta$.

Finally, suppose $n_2 = 0$, so $f = h(u)$ with $\delta_u h \neq 0$. Then a simplification of the argument in the preceding paragraph shows that we can take $f = y^\delta$ and obtain (1.13), where each term is independent of y^δ. Q.E.D.

We have shown that corresponding to each R-separable coordinate system $\{y^j\}$ for the time-dependent Hamilton–Jacobi equation on a Riemannian manifold V^n there is associated a unique time coordinate $f = y^\delta$ or $f = y^r$. The transformed equation in the $\{y^i\}$ coordinates is again in time-dependent Hamilton–Jacobi form for a Hamiltonian on V^n. The transformed Hamiltonian is strictly time dependent if and only if $f = y^r$.

In the following we will regard the problem of finding all *R-separable* solutions of a given time-dependent equation (1.1) as solved once we reduce it to the problem of finding all *separable* solutions of explicit time-independent Hamilton–Jacobi equations of the form

$$\sum_{i,j=1}^{n} G^{ij} Z_i Z_j + \lambda \sum_{i=1}^{n} U^i Z_i + \lambda^2 U = E, \tag{1.15}$$

where (G^{ij}) is the metric on V^n.

For $f = y^\delta$ this problem was solved in Ref. 7. There we studied all mappings of the form

$$t = T(f,\mathbf{y}), \quad \mathbf{x} = X(f,\mathbf{y}), \quad W = Z + \lambda h(f,\mathbf{y}) \tag{1.16}$$

that take (1.1) into another evolution equation (1.13), a "related" evolution equation. It was shown that there is a one-to-one correspondence between (equivalence classes of) related Hamilton–Jacobi equations and conformal symmetries for (1.1) of the form $\mathscr{L} = q(t,\mathbf{x})p_t + \gamma^j(t,\mathbf{x})p_{x^j} + \lambda k(t,\mathbf{x})$, where $q \neq 0$; alternatively,

$$L = q\,\partial_t + \gamma^j \partial_{x^j} + \lambda k\, \partial_W. \tag{1.17}$$

If L is a conformal symmetry with $q \neq 0$ then one can show that $\partial_{x^j} q = 0$ and that we can introduce new coordinates $f,\mathbf{y}$ and a new dependent variable z such that

$$\begin{aligned} &\partial_f = q\,\partial_t + \gamma^j \partial_{x^j}, \\ &t = T(f), \quad x^i = X^i(f,\mathbf{y}), \\ &W = Z + \lambda h(f,\mathbf{y}), \end{aligned} \tag{1.18}$$

and (1.1) transforms to the related Hamilton–Jacobi equa-

tion (1.13) with f-independent Hamiltonian.

Conversely, if case 1 of Theorem 1 occurs for (1.1) then the variable $f = y^\delta$ is ignorable. This means that the operator ∂_{y^δ} is a conformal symmetry for (1.13) which in turn transforms to a conformal symmetry for (1.1) of the form $q\partial_t + \gamma^j \partial_{x^j} + \lambda k \partial_w$ with $q \neq 0$.

Note that when f is ignorable we can require $Z_f = -E/2\lambda$ and reduce (1.13) to the time-independent equation (1.15). Thus, to find all R-separable coordinate systems $\{f,\mathbf{y}\}$ for (1.1) corresponding to case 1 we first enumerate the conjugacy classes of symmetry operators in the conformal symmetry algebra $\mathcal{G}$ for (1.1).[11] Choosing a representative operator L in each conjugacy class we make the transformation (1.18) of (1.1) into a related evolution equation and then find all R-separable systems $\{y^j\}$ for the reduced equation (1.15). (Of course, for some choices of L the reduced equation is not R-separable in any coordinate system.) See Ref. 7 for more information concerning this procedure, and Ref. 4 for a more general point of view.

We can thus regard case 1 of the preceding theorem as well understood from the viewpoint of Lie symmetries.

II. TIME-DEPENDENT HAMILTONIANS

We now turn our attention to case 2 of Theorem 1, the case where $f = y^r$ and the transformed equation (1.13) has a time-dependent Hamiltonian. Our aim, not entirely achieved, will be to enumerate the instances where this type of R-separation occurs for the Hamilton–Jacobi equation (1.1).

For simplicity we will limit ourselves to coordinate systems that are orthogonal on V^n. In other words, the metric tensor (G^{ij}), $i,j \neq f$, in (1.13) should be diagonal. (For many spaces, such as Euclidean spaces or spaces of constant curvature, only orthogonal separation can occur, so this is no restriction at all.[12]) Since orthogonal ignorable coordinates can always be considered as special cases of type 1 coordinates, without loss of generality we can assume that the separable coordinates are labeled y^a $(a = 1,...,n), y^r, \mu$; $n_1 = n, n_2 = n_3 = 1$. [This is true so long as we restrict attention to (G^{ij}) and ignore the vector and scalar potentials.]

With the above assumptions our problem simplifies substantially. The second equation in (1.9) becomes $1 = Qk_r H_r^{-2}$. Replacing the Stäckel form H_r^{-2}, H_a^{-2} by the new Stäckel form $H'^{-2}_r = k_r H_r^{-2}, H'^{-2}_a = H_a^{-2}$, we can assume $k_r = 1$. Furthermore since H'^{-2}_r is a Stäckel multiplier[5] we can pass to a new Stäckel form with $H''^{-2}_r = 1$, $H''^{-2}_a / H'^{-2}_r, Q'' = QH'^{-2}_r = 1$. Thus in terms of the coordinates y^a, y^r, μ we have (dropping the primes)

$$H_r^{-2} = 1, \quad H_a^{-2} = H_a^{-2}(y^b, y^r), \quad Q = 1. \tag{2.1}$$

The transformation from "standard" to separable coordinates becomes

$$\begin{aligned} &x^k = x^k(y^a, y^r), \quad k = 1,...,n, \\ &t = y^r, \quad \tau = \mu - R(y^a, y^r), \end{aligned} \tag{2.2}$$

and the metric becomes

$$\sum_{i,j=1}^{n} g^{ij} p_{x^i} p_{x^j} = \sum_{a=1}^{n} H_a^{-2}(y^b, t) p_{y^a}^2, \tag{2.3}$$

so for each fixed t the right-hand side of (2.3) defines a Stäckel-form metric on V^n.

To analyze this one-parameter metric we recall a few facts about Stäckel form metrics. An $n \times n$ nonsingular matrix $S(\mathbf{y})$ is said to be in Stäckel form if $S_{ij} = S_{ij}(y^i)$ and each of the elements $(S^{-1})^{1l}$, $l = 1,...,n$, is nonzero. Set

$$\mathcal{H}_i(\mathbf{y},\mathbf{p}) = \sum_{l=1}^{n} (S^{-1})^{il} p_l^2. \tag{2.4}$$

Then

$$\{\mathcal{H}_i \mathcal{H}_j\} = 0, \quad i,j = 1,...,n, \tag{2.5}$$

where $\{\cdot,\cdot\}$ is the Poisson bracket on the $2(n+1)$-dimensional symplectic manifold with canonical coordinates (y^i, p_i, t, p_r). Here $\mathcal{H}_1$ is the Hamiltonian associated with the Stäckel matrix S.

Theorem 2: Let

$$\mathcal{H}' = H_r^{-2} p_r^2 + \sum_{a=1}^{n} H_a^{-2}(\mathbf{y},t) p_a^2$$

be a Stäckel form Hamiltonian with $H_r^{-2} = 1$ and $n \geqslant 2$, and suppose t_0 is in the domain of t. Let

$$\mathcal{H}(t) = \sum_{a=1}^{n} H_a^{-2}(\mathbf{y},t) p_a^2.$$

Then there exists an $n \times n$ Stäckel matrix $S(\mathbf{y})$ such that

$$\mathcal{H}(t) = \sum_{k=1}^{n} g_k(t) \mathcal{H}_k, \tag{2.6}$$

where the $\mathcal{H}_k$ are defined by (2.4) and the g_k are scalar-valued functions with $g_k(t_0) = \delta_{1k}$.

Proof: Since $\mathcal{H}'$ is a Stäckel form Hamiltonian there exists an $(n+1) \times (n+1)$ Stäckel matrix $T'(\mathbf{y},t)$ such that

$$T' = \begin{pmatrix} T_{00}(t) & T_{01}(t) \cdots T_{0n}(t) \\ T_{10}(y^1) & T_{11}(y^1) \cdots T_{1n}(y^1) \\ \vdots & \\ T_{n0}(y^n) & T_{n1}(y^n) \cdots T_{nn}(y^n) \end{pmatrix} \tag{2.7}$$

and $(T'^{-1})^{00} = H_r^{-1} = 1$, $(T'^{-1})^{0a} = H_a^{-2}$, $a = 1,...,n$. It follows that

$$T'' = \begin{pmatrix} 1 & T_{01} & \cdots & T_{0n} \\ 0 & T_{11} & & T_{1n} \\ \vdots & \vdots & & \vdots \\ 0 & T_{n1} & & T_{nn} \end{pmatrix}$$

is also a Stäckel matrix for $\mathcal{H}'$, since $(T'^{-1})^{0i} = (T''^{-1})^{0i}$, $i = 0,1,...,n$. We can multiply column i of T'' by a nonzero constant c and column j by c^{-1} where $i \neq j$, i,j $i,j > 0$, and obtain another Stäckel matrix for $\mathcal{H}'$. Furthermore, the interchange of two such columns $\mathbf{T}_i, \mathbf{T}_j$ or the replacement of $\mathbf{T}_i$ by $\mathbf{T}_i + c'\mathbf{T}_j$ again leads to a Stäckel matrix for $\mathcal{H}'$. It follows that there is a Stäckel matrix for $\mathcal{H}'$ of the form

$$T = \begin{pmatrix} 1 & -g_1(t) & \cdots & -g_n(t) \\ 0 & S_{11}(y^1) & \cdots & S_{1n}(y^1) \\ \vdots & \vdots & & \vdots \\ 0 & S_{n1}(y^n) & \cdots & S_{nn}(y^n) \end{pmatrix}, \tag{2.8}$$

where $g_a(t_0) = \delta_{a1}$. Thus $\mathcal{H}(t)$ is given by (2.6) where the

$\mathscr{H}_k$ are computed from the $n\times n$ Stäckel matrix $(S_{ab}(y^a))$. Q.E.D.

Corollary: $\{\mathscr{H}(t_1), \mathscr{H}(t_2)\} = 0$.

Since $\mathscr{H}(t_0) = \mathscr{H}_1$ and

$$\sum_{i,j=1}^{n} g^{ij} p_{x^i} p_{x^j} = \mathscr{H}(t) \tag{2.9}$$

we see that $\mathscr{H}_1$ is a Stäckel form Hamiltonian for V^n and that $\mathscr{H}(t)$ is a one-parameter family of such Hamiltonians. The requirements that $\mathscr{H}(t)$ corresponds to V^n is a very strong condition on the functions $g_i(t)$ for any choice of separable Hamiltonian $\mathscr{H}_1$.

To show how restrictive these conditions are we consider the "generic" separable coordinates in Euclidean space E^n and on the unit sphere S^n,

$$\mathscr{H}_1 = \sum_{i=1}^{n} \frac{f(y^i)}{\pi_{i\neq j}(y^i - y^j)} p_i^2 = \sum_{i=1}^{n} H_i^{-2} p_i^2, \tag{2.10}$$

where f is a polynomial with distinct real roots. This is a separable Hamiltonian on S^n iff $\deg f = n+1$ (Jacobi elliptic coordinates), and on E^n iff $\deg f = n$ (ellipsoidal coordinates) or $\deg f = n-1$ (paraboloidal coordinates). The related $n\times n$ Stäckel matrix is[12]

$$S_{ij} = (y^i)^{n-j}/f(y^i), \quad i,j = 1,...,n. \tag{2.11}$$

In general, the orthogonal coordinates $\{x^i\}$ are separable on E^n provided the metric $ds^2 = \Sigma_{a=1}^n H_a^2(\mathbf{x})\,(dx^i)^2$ is in Stäckel form, i.e.,

$$\partial_{jk}\log H_i^2 - \partial_j \log H_i^2 \partial_k \log H_i^2 + \partial_j \log H_i^2 \partial_k \log H_j^2 + \partial_k \log H_i^2 \partial_j \log H_k^2 = 0, \quad j\neq k, \tag{2.12}$$

and $R_{hijk} = 0$ where R is the Riemann curvature tensor. For S^n (of constant curvature -1) this last condition is replaced by

$$\begin{aligned} &R_{ijji} = -H_i^2 H_j^2, \quad i\neq j, \\ &R_{hiik} = 0, \quad h,i,k \text{ distinct.} \end{aligned} \tag{2.13}$$

Eisenhart[13] (p. 269) has shown that for both E^n and S^n these conditions imply

$$\partial_{jk}\log H_i^{-2} = 0, \quad i,j,k \text{ distinct.} \tag{2.14}$$

Suppose $n\geqslant 3$. Then condition (2.4) applied to $\mathscr{H}(t)$, (2.6), where $\mathscr{H}_1$ is given by (2.10), becomes

$$\partial_{jk}\log\left(\sum_{l=1}^{n} g_l(t) \prod_{\substack{i_1<i_2<\cdots<i_{l-1}\\ i_h\neq i}} y^{i_1}\cdots y^{i_{l-1}}\right) = 0, \quad i,j,k \text{ distinct.} \tag{2.15}$$

The solution is

$$H_i^{-2}(\mathbf{y},t) = \frac{f(y^i)}{\pi_{i\neq j}(y^i - y^j)} g_1(t) \prod_{i\neq j}(1 + h(t)y^j), \quad i = 1,...,n.$$

Clearly, $g_1\neq 0$. Suppose $h\neq 0$. Under the change of coordinates $x^i = y^i/(1+hy^i)$ the metric transforms to

$$\tilde{H}_i^{-2}(\mathbf{x}) = \frac{f(x^i/(1-hx^i))}{\pi_{i\neq j}(x^i - x^j)} (1-hx^i)^{n+3} g_1(t),$$

which is again of the form (2.10), except that the polynomial in the numerator is of order $n+3$, which does not correspond to E^n or S^n. Thus $h = 0$ and

$$H_i^{-2}(\mathbf{y},t) = \frac{f(y^i) g_1(t)}{\pi_{i\neq j}(y^i - y^j)}.$$

It is well known that for $g_1(t)\neq 1$ and $V^n = S^n$ the factor g_1 changes the curvature, so that the transformed metric is not one on S^n. We conclude that for Jacobi elliptic coordinates on S^n the only possibility is $\mathscr{H}(t) = \mathscr{H}_1$ whereas for ellipsoidal or paraboloidal coordinates on E^n the only possibilities are dilatations $\mathscr{H}(t) = g_1(t)\mathscr{H}_1$. For $n = 2$ a similar but simpler argument than the preceding one yields the same result.

We can now treat the most general separable coordinate system on S^n. In Ref. 12 it is shown that the most general separable system can be constructed by "nesting" collections of the generic Jacobi elliptic coordinates. The infinitesimal distance on S^n, expressed in a separable system, can always be written in the form

$$d\omega^2 = \sum_{I=1}^{p} d\omega_I^2 \left[\frac{\pi_{l=1}^{n_1}(y^l - e_I)}{\pi_{m\neq I}(e_m - e_I)}\right] - \frac{1}{4}\sum_{i=1}^{n_1} \frac{\pi_{j\neq i}(y^i - y^j)}{\pi_{j=1}^{n_1+1}(y^i - e_j)} (dy^i)^2. \tag{2.16}$$

Here the $\{y^i\}$ are Jacobi elliptic coordinates on S^{n_1} and each $d\omega_I^2$ is the infinitesimal distance of a S^{P_I} where $\Sigma_{I=1}^p P_I + n_1 = n$, and $p\leqslant n_1 + 1$. The coordinates on each S^{P_I} are again separable and the metrics $d\omega_I^2$ can be expressed in terms of separable coordinates by using (2.16) recursively. [The case of Jacobi elliptic coordinates on S^n corresponds to $p = 0, n_1 = n$ in (2.16).] In Ref. 12 a graphical procedure is presented to elucidate that construction, and the separation equations for the Hamilton–Jacobi equation are written explicitly. Thus for every separable system on S^n it is straightforward to compute the Stäckel matrix and to construct the quadratic forms $\mathscr{H}_k$, (2.6). The y coordinates in (2.6), since they are separable and $\mathscr{H}(t)$ is analytic in t, must be of the same type (2.16) as the coordinates of $\mathscr{H}_1 = \mathscr{H}(t_0)$. Though the details are somewhat tedious, it is not difficult to use the argument of (2.15), and its following paragraphs, recursively in (2.16). The results of this argument, followed by imposition of the curvature conditions (2.13), is the following.

Theorem 3: Let S be a Stäckel matrix corresponding to a separable coordinate system on S^n, $n\geqslant 2$, and define the corresponding Hamiltonian $\mathscr{H}_1$ and constants of the motion $\mathscr{H}_i$, $i = 2,...,n$, by (2.4). Then $\mathscr{H}(t) = \Sigma_{i=1}^n g_i(t)\mathscr{H}_i$ is an S^n Hamiltonian with $\mathscr{H}(t_0) = \mathscr{H}_1$ iff $g_i(t) = \delta_{1i}$.

The most general separable coordinate system on E^n is also determined in Ref. 12. It is shown there that in the coordinates of such a system the metric ds^2 on E^n can be expressed as

$$ds^2 = \sum_{I=1}^{Q} ds_I^2, \tag{2.17}$$

where each ds_I^2 is an Euclidean space metric itself. In turn we

have

$$ds_I^2 = \sum_{i=1}^{n_I} \frac{\pi_{l=1}^{N_I}(y^l - e_i^I)}{\pi_{j\neq i}(e_j^I - e_i^I)} d\omega_i^2 + d\sigma_I^2, \tag{2.18}$$

where $d\sigma_I^2$ is the infinitesimal distance corresponding to ellipsoidal or paraboloidal coordinates $\{y^l\}$ for E^{N_I}, and each $d\omega_i^2$ is the infinitesimal distance corresponding to a separable system on the sphere S^{P_I}. Here $n = \Sigma_{I=1}^{Q}(N_I + p_I)$ and $n_I \leqslant N_I$ for $d\sigma_I^2$ ellipsoidal, $n_I < N_I$ for $d\sigma_I^2$ paraboloidal. Thus the most general separable system on E^n is constructed by first decomposing E^n as a direct sum of Q mutually orthogonal Euclidean subspaces and then in each subspace nesting collections of Jacobi elliptic coordinates into either an ellipsoidal or a paraboloidal system. The generic ellipsoidal or paraboloidal systems for E^n correspond to the case $Q = 1, n_1 = 0, N_1 = n$.

From the results of Ref. 12 it is straightforward to compute the Stäckel matrix corresponding to each separable system for E^n and to construct the quadratic forms $\mathscr{H}_k$. Again the y coordinates (2.6) must be of the same type (2.18) as the coordinates of $\mathscr{H}_1 = \mathscr{H}(t_0)$. It is tedious, though not difficult, to use the argument of (2.15) and its following paragraphs recursively in (2.18) and (2.16), followed by imposition of the Euclidean space curvature conditions to obtain the following.

Theorem 4: Let S be a Stäckel matrix corresponding to a separable coordinate system on E^n, $n \geqslant 2$, and let

$$ds^2 = \sum_{I=1}^{Q} ds_I^2$$

be the associated decomposition of the infinitesimal distance on E^n into distances ds_I^2 on Q mutually orthogonal Euclidean subspaces. Let $\mathscr{H}_1^{(I)}$ be the Hamiltonian on the I th subspace so that $\mathscr{H}_1 = \mathscr{H}(t_0) = \Sigma_{I=1}^{Q}\mathscr{H}_1^{(I)}$. Then $\mathscr{H}(t)$ is an E^n Hamiltonian for all t iff it can be expressed in the form

$$\mathscr{H}(t) = \sum_{I=1}^{Q} h_I(t)\mathscr{H}_1^{(I)},$$

with $h_I(t_0) = 1, I = 1,...,Q$.

To date we have been unsuccessful in proving Theorems 3 and 4 without using the explicit list of all separable coordinate systems on S^n and E^n.

III. COORDINATES ON S^n AND E^n

Continuing our study of case 2 of Theorem 1, let $\{y^a, y^\alpha\}$ be an orthogonal separable coordinate system on the Riemannian space V^n such that the associated infinitesimal distance

$$ds^2 = \sum_{i=1}^{n} H_i^2 (dy^i)^2 = \sum_a H_a^2 (dy^a)^2 + \sum_\alpha H_\alpha^2 (dy^\alpha)^2$$

is in Stäckel form [with Stäckel matrix (2.5)]. Here $H_\alpha^{-2} = \Sigma_{a=1}^{n_1} K_a^{\alpha\alpha}(y^a) H_a^{-2}$. Let $\mathscr{H}_1$ be the Hamiltonian in these coordinates and suppose we have determined a one-parameter family $\mathscr{H}(t) = \Sigma_{i=1}^{n} g_i(t)\mathscr{H}_i$ of Hamiltonians on V^n such that $\mathscr{H}(t_0) = \mathscr{H}_1$. We now study the remaining conditions on $g_i(t)$, $\{y^a, y^\alpha\}$ so that $\{y^r, y^a, y^\alpha\}$ will lead to R-separation of the time-dependent Hamilton–Jacobi equation

$$2\lambda W_t + \sum_{l,m=1}^{n} g^{lm}(\mathbf{x}) W_{x^l} W_{x^m} + 2\lambda \sum_{l=1}^{n} A^l(\mathbf{x}) W_{x^l} + \lambda^2 V(\mathbf{x}) = 0. \tag{3.1}$$

Here $\{x^i\}$ is a given coordinate system on V^n and we must have

$$x^i = x^i(y^a, y^r, y^\alpha), \quad i = 1,...,n, \quad t = y^r. \tag{3.2}$$

The remaining conditions to be satisfied are

$$\frac{\partial y^a}{\partial t} + \mathscr{A}^a = -H_a^{-2} R_a, \quad a = 1,...,n_1, \tag{3.3a}$$

$$\frac{\partial y^\alpha}{\partial t} + \mathscr{A}^a = -H_\alpha^{-2} R_\alpha + \sum_{a=1}^{n_1} \mathscr{F}_a^\alpha(y^a) H_a^{-2} + \mathscr{F}_r^\alpha(y^r) = -H_\alpha^{-2} R_\alpha + \mathscr{B}^\alpha, \quad \alpha = n_1,...,n, \tag{3.3b}$$

$$2R_r + V - \sum_{i=1}^{n} H_i^{-2} R_i^2 + 2\sum_\alpha R_\alpha \mathscr{B}^\alpha = \sum_a \mathscr{F}_a(y^a) H_a^{-2} + \mathscr{F}_r(y^r), \tag{3.3c}$$

where $R_a = \partial_{y^a} R(y^r, y^a)$ and R_r, R_α are defined similarly. Each $\mathscr{F}_i, \mathscr{F}_i^\alpha$ is a function of a single variable y^i. We will discuss the solution of these equations with special emphasis on the important examples S^n and E^n.

First note that (3.3a) and (3.3b) can be written in covariant form,

$$\frac{\partial z^i}{\partial t} = -G^{ij}(\mathbf{z}, t) R_{z^j} - \mathscr{A}^i + \mathscr{B}^i, \quad i = 1,...,n, \tag{3.4}$$

where (G^{ij}) is the metric for V^n in the coordinates z^i. Here, $z^i = z^i(y^a, y^\alpha)$. We can choose the initial coordinates $\{x^i\}$ and the $\{z^i\}$ to be in a convenient standard form and such that $z^i = Z^i(x^j, t)$, $i = 1,...,n$, with $Z^i(x^j, t_0) = x^i$. We will use the integrability conditions for (3.4) to determine the possible forms of the functions Z^i, and then express the $\{z^i\}$ in terms of separable $\{y^a, y^\alpha\}$ coordinates.

Consider first the space S^n. It is convenient to identify S^n with the unit sphere $\mathbf{x}\cdot\mathbf{x} = \Sigma_{j=1}^{n+1}(x^j)^2 = 1$ in E^{n+1} where $\{x^1,...,x^{n+1}\}$ are standard Cartesian coordinates. We choose $\{z^1,...,z^{n+1}\}$ to be Cartesian coordinates of the same type. Since the motion group of S^n is $O[n+1]$ (see Ref. 13, p. 23) it is clear that

$$\mathbf{z}(t) = O(t)\mathbf{x}, \quad O(t) \in O[n+1], \quad O(t_0) = I, \tag{3.5}$$

where I is the identity matrix. Equations (3.4) become

$$\sum_{l,j=1}^{n+1} \dot{O}_{ij} O_{lj} z^l = -R_{z^i} - \mathscr{A}^i + \mathscr{B}^i, \quad i = 1,...,n+1, \quad \mathbf{z}\cdot\mathbf{z} = 1. \tag{3.6}$$

The integrability conditions for (3.6) imply

$$\mathscr{A}^i - \mathscr{B}^i = -R_{z^i} + \sum_{s=1}^{n+1} X_{si}(t) z^s, \quad X = \dot{O}O^{-1}. \tag{3.7}$$

This can be regarded as a necessary condition on the vector potential in order that it permit variable separation. The second term on the right-hand side of (3.7) is "trivial" in the sense that it can always be removed by transformation to an

appropriate rotating frame $\hat{\mathbf{x}}(t) = \hat{O}(t)\mathbf{x}$, $\hat{O} \in O[n+1]$. [Thus every separable potential is via the foregoing transformation equivalent to a vector potential which is a gradient $\mathscr{A}^i - \mathscr{B}^i = -\partial_{x^i} R(\mathbf{x},t)$ in Cartesian coordinates.] Assuming the equivalent vector potential is a gradient we can remove it by an appropriate R transformation and then have $z^i \equiv x^i$, $t = y^r$, $\mathscr{A}^i = \mathscr{B}^i$, $R = R(t)$.

Thus Eq. (3.1) transforms to

$$2\lambda Z_t + \sum_{i=1}^{n} H_i^{-2}(y^a) Z_i^2 + 2\lambda \sum_{\alpha,a} F_a^{\alpha}(y^a) H_a^{-2} Z_\alpha + \sum_{\alpha} \mathscr{F}_a(y^a) H_a^{-2} = 0, \tag{3.8}$$

where $\{y^a, y^\alpha\}$ is a separable system for the time-independent Hamilton–Jacobi equation on S^n.[14]

Theorem 5: Nontrivial R-separation corresponding to case 2 does not occur for the time-dependent Hamilton–Jacobi equation on S^n, i.e., every such separable system arises from a separable system for the time-independent equation.

Now we examine the same problem for E^n. As standard coordinates $\{x^1,...,x^n\}$ we choose Cartesian coordinates. By Theorem 4, corresponding to a set of separable coordinates for E^n, there is a decomposition of this space into Q mutually orthogonal Euclidean subspaces $E_J^{N_J}$. Let $\{z_J^1,...,z_J^{N_J}\}$ be Cartesian coordinates on the Jth such subspace (of dimension N_J). The Hamiltonian on $E_J^{N_J}$ is $\mathscr{H}_1^{(J)} = \sum_{i=1}^{N_J} (p_{z_J^i})^2$ and we have

$$\mathscr{H}_1^{(J)} = \sum_{J=1}^{Q} h_J^2(t) \mathscr{H}_1^{(J)}, \tag{3.9}$$

where $h_J(t_0) = 1$. Since the motion group for E^n is the Euclidean group[13] (p. 23), it follows that the two coordinate systems are related by a t-dependent Euclidean transformation,

$$D^{-1}(t)\mathbf{z}(t) = O(t)\mathbf{x} + \mathbf{c}(t), \tag{3.10}$$

where O is an $n \times n$ orthogonal matrix, $\mathbf{c}$ is an n vector, and D is an $n \times n$ diagonal matrix whose diagonal term corresponding to z_J^i is $h_J(t)$. Equations (3.4) take the form

$$(\dot{D}D^{-1} + D\dot{O}O^{-1}D^{-1})\mathbf{z} - D\dot{O}O^{-1}\mathbf{c} + D\dot{\mathbf{c}} = -D^2 R_{\mathbf{z}} - \mathscr{A}^{\mathbf{z}} + \mathscr{B}^{\mathbf{z}}. \tag{3.11}$$

The integrability conditions for (3.11) imply that there exists a function $q(\mathbf{z},t)$ such that

$$(\mathscr{A}^i - \mathscr{B}^i) D_i^{-2} = q_{z^i} - D_i^{-1} \sum_{l,k=1}^{n} \dot{O}_{il} O_{kl} D_k^{-1} z^k, \tag{3.12}$$

where

$$-D_i^2(R_{z^i} + q_{z^i}) = \dot{D}_i D_i^{-1} z^i + D_i \dot{c}_i + D_i \sum_{l,k=1}^{n} \dot{O}_{il} O_{kl} c_k. \tag{3.13}$$

Here $D_{ij}(t) = D_i(t)\delta_{ij}$. This implies that in the standard Cartesian coordinates the vector potential has the form

$$A^j(\mathbf{x},t) - B^j(\mathbf{x},t) = -\sum_{i,s=1}^{n} O_{ij}(t)\dot{O}_{is}(t)x^s + H_{x^j}(\mathbf{x},t). \tag{3.14}$$

Just as in the S^n case we can remove the first term on the right-hand side of (3.14) through a coordinate transformation $\hat{x} = \hat{O}(t)x$, where $\hat{O}^{-1}\dot{\hat{O}} = -O^{-1}\dot{O}$. Thus any separable vector potential is equivalent to a potential in gradient form ($A^j - B^j = H_{x^j}$), so that $(\mathscr{A}^i - \mathscr{B}^i) D_i^{-2} = q_{z^i}$ and we can assume $O(t) \equiv \mathrm{I}$ in (3.12) and (3.13). Then by means of an R transformation (which does not affect separability) we can take $q = 0$ and $\mathscr{A}^i = \mathscr{B}^i$.

Thus

$$h_j^{-1}(t) z_J^i(t) = x_J^i + c_J^i(t), \quad i = 1,...,N_J, \quad J = 1,...,Q, \tag{3.15}$$

where we have adopted the same notation for the vectors $\mathbf{x}$,$\mathbf{c}$ as for $\mathbf{z}$. Substituting (3.15) back into (3.11) we obtain

$$(\dot{h}_J/h_J^3) z_J^i + \dot{c}_J^i/h_J = -R_{z_J^i},$$

so that

$$R = -\sum_{J=1}^{Q} \sum_{i=1}^{N_J} \left[\frac{1}{2} \frac{\dot{h}_J}{h_J^3} (z_J^i)^2 + \frac{\dot{c}_J^i}{h_J} z_J^i \right] + f(t). \tag{3.16}$$

We consider first the special case where the original vector and scalar potentials vanish: $\mathscr{A}^i = V = 0$. Then $\mathscr{B}^i = 0$ and we can assume $n_3 = 1$. Substituting (3.16) and (3.9) into the remaining condition (3.3c) we find

$$\sum_{J=1}^{Q} \sum_{i=1}^{N_J} \left[\frac{1}{2} (\ddot{h}_J^{-2}) (z_J^i)^2 - 2 \frac{d}{dy^r} \left(\frac{\dot{c}_J^i}{h_J} \right) z_J^i - h_J^2 \left(\frac{\dot{h}_J}{h_J^3} z_J^i + \frac{\dot{c}_J^i}{h_J} \right)^2 \right] = \sum_{J,a} h_J^2 (\mathscr{F}_{a,J}(y_J^a) H_{a,J}^{-2}(y_J^b) + \mathscr{F}_J(y^r)). \tag{3.17}$$

Since, for each fixed J, the N_J coordinates z_J^i are functions of the N_J coordinates y_J^b, a necessary condition for (3.17) to hold is that the coefficients of $(z_J^i)^2$ and z_J^i on the left-hand side are constants times h_J^2,

$$h_J^{-2}(\ddot{h}_J^{-2}) - 2\dot{h}_J^2 h_J^{-6} = \alpha_J, \tag{3.18a}$$

$$\ddot{c}_J^i/h_J^3 = \beta_J^i. \tag{3.18b}$$

It follows that under the R transformation the original Hamilton–Jacobi equation

$$2\lambda W_t + \sum_{l=1}^{n} W_{x^l}^2 = 0 \tag{3.19}$$

maps to

$$2\lambda Z_r + \sum_{J=1}^{Q} \sum_{i=1}^{N_J} \left(Z_{z_J^i}^2 + \frac{\alpha_J}{2} (z_J^i)^2 - 2\beta_J^i z_J^i \right) h_J^2 = 0, \tag{3.20}$$

where the α_J, β_J^i are constants. Note that the original Hamiltonian "decouples" into Q Hamiltonians

$$\mathscr{H}_1^{(J)} = \sum_{i=1}^{N_J} \left[p_{z_J^i}^2 + \frac{\alpha_J}{2} (z_J^i)^2 - 2\beta_J^i z_J^i \right].$$

The separable coordinates $\{y_J^b\}$ are just those that separate the time-independent Hamilton–Jacobi equations $\mathscr{H}_1^{(J)} = E$, $J = 1,...,Q$.

Equation (3.18a) is equivalent to $(\dddot{h}_J^{-2}) = 0$ and has the general solution

$$h_J^2(t) = (b_1^J t^2 + b_2^J t + b_3^J)^{-1}, \quad 2b_1^J b_3^J - \tfrac{1}{2}(b_2^J)^2 = \alpha_J. \tag{3.21}$$

We can simplify the ensuing argument by identifying coordinate systems that are equivalent under the action of the Galilean (and dilatation) symmetries of (3.19), see Ref. 2, Chap. 2. Thus, in addition to the Euclidean symmetries already employed, we identify systems related by dilatations $t\to\alpha^2 t$, $\mathbf{x}\to\alpha\mathbf{x}$, time translations $t\to t+\beta$ and velocity transformations $\mathbf{x}\to\mathbf{x}+\mathbf{c}t$. Then, in case $\alpha_J\neq 0$ for fixed J we can perform a translation of the $\{z_J^i\}$ coordinates to achieve $\beta_J^i=0$, $c_J^i=0$, $i=1,\ldots,N_J$. Thus the decoupled Hamiltonian is

$$\mathscr{H}_1^{(J)}=\sum_{i=1}^{N_J}\left[p_{z_J^i}^2+\frac{\alpha_J}{2}(z_J^i)^2\right], \tag{3.22}$$

the "harmonic oscillator" for $\alpha_J>0$ and the "repulsive oscillator" for $\alpha_J<0$. Here

$$h_J^{-1}(t)z_J^i(y_J^a)=x_J^i$$

and the possible orthogonal separable coordinates y_J^a are just those that separate $\mathscr{H}_1^{(J)}=E$.

In case $\alpha_J=0$ for fixed J we can perform a rotation of coordinates $\{z_J^i\}$ to achieve $\beta_J^i=0$ for $i=2,\ldots,N_J$. The corresponding decoupled Hamiltonian is

$$\mathscr{H}_1^{(J)}=\sum_{i=1}^{N_J}\left[p_{z_J^i}^2-2\beta_J^1 z_J^1\right], \tag{3.23}$$

the "free fall" Hamiltonian. Here

$$h_J^{-1}(t)z_J^i(y_J^a)=x_J^i+\delta^{li}c_J^1(t),$$

where

$$h_J(t)=((b_2^J/2)t+1)^{-1},$$
$$c_J^1(t)=\begin{cases}\beta_J^1(t^2/2), & b_2^J=0,\\ [2\beta_J^1/(b_2^J)^2]((b_2^J/2)t+1)^{-1}, & b_2^J\neq 0.\end{cases} \tag{3.24}$$

We have used the property that (by a suitable time translation if necessary) we can assume $t_0=0$, $h_J(0)=1$. The possible orthogonal separable coordinates y_J^a are those that separate $\mathscr{H}_1^{(J)}=E$.

Although we will not state our results as a theorem, we have reduced the problem of finding all R-separable coordinate systems for the time-dependent Euclidean equation (3.19) (with zero potential) to the problem of finding all separable coordinate systems for the time-independent Hamiltonians (3.22) and (3.23). The answer to this last set of problems is known.[15]

In the general case where the vector and scalar potentials do not both vanish we have $\mathscr{A}^i=\mathscr{B}^i$ so that in the coordinates $\{y^a,y^\alpha,y^r\}$, $\mathscr{A}^a=\mathscr{A}^r=0$,

$$\mathscr{A}^\alpha=\mathscr{B}^\alpha=\sum_{a=1}^{n_1}\mathscr{F}_a^\alpha(y^a)H_a^{-2}+\mathscr{F}_r^\alpha(y^r).$$

Then Eq. (3.1) transforms to

$$2\lambda Z_t+\sum_{J,i}h_J^2(t)H_{J,i}^{-2}(y_J^b)Z_{J,i}^2+2\lambda\sum_\alpha\left(\sum_{J,a}\mathscr{F}_{J,a}^\alpha(y_J^a)h_J^2(t)H_{J,a}^{-2}(y_J^b)+\mathscr{F}_r^\alpha(t)\right)Z_\alpha+\sum_{J,a}\mathscr{F}_{J,a}(y_J^a)h_J^2(t)H_{J,a}^{-2}=0, \tag{3.25}$$

where

$$V=-2\sum_\alpha\left(\sum_{J,a}\mathscr{F}_{J,a}^\alpha(y_J^a)h_J^2(t)H_{J,a}^{-2}(y_J^b)+\mathscr{F}_r^\alpha(t)\right)\partial_\alpha R-\sum_{J=1}^{Q}\sum_{i=1}^{N_J}\left[\frac{\alpha_J}{2}h_J^2(t)(z_J^i)^2-2h_J^2(t)\beta_J^i z_J^i-(\dot{c}_J^i)^2\right]+\sum_{J,a}h_J^2(t)\mathscr{F}_{a,J}(y_J^a)H_{a,J}^{-2}(y_J^b), \tag{3.26}$$

and α_J,β_J^i are defined by (3.18). Since, for each fixed J, the N_J coordinates z_J^i are functions of the N_J coordinates y_J^b, condition (3.26) is a strong restriction on V. Clearly, given any separable system $\{y_J^b,y_J^a\}$ for E^n there always exist potentials V for which (3.26) is satisfied. The $h_J(t)$ must be determined from this functional equation.

IV. COMMENTS AND EXAMPLES

We can also use the results of the preceding section to find all R-separable coordinate systems for the time-dependent Hamilton–Jacobi equations on S^n and E^n that correspond to case 1 of Theorem 1, i.e., such that the new time coordinate f, (1.13), is ignorable: $f=y^\delta$. (Here we will treat only the zero potential equations. The nonzero potential treatment is similar.) Since the time-independent Hamilton–Jacobi equations on S^n and E^n separate only in orthogonal coordinates,[12] we can assume $n_1=n$, $n_2=0$, $n_3=2$ so that the transformed equation (1.13) takes the form

$$2\lambda Z_{y^\delta}+\sum_{a=1}^{n}H_a^{-2}(y^b)Z_{y^a}^2+\lambda^2U(y^b)=0. \tag{4.1}$$

For S^n, $\xi^\delta=\tilde{Q}=1$ so $y^\delta=t$. Furthermore, the argument leading up to Theorem 5 shows that the separable $\{y^a\}$ coordinates are expressible entirely in terms of the $\{x^i\}$, i.e., $\partial_t y^a=0$. Combining this fact with Theorem 5 we have the following.

Theorem 6: Every R-separable coordinate system for the (zero potential) time-dependent Hamilton–Jacobi equation on S^n is purely separable and of the form $\{t,y^a\}$ where $\{y^a\}$ is an orthogonal separable system for the time-independent Hamilton–Jacobi equation on S^n.

For E^n and case 1 the results are a bit more complicated. It is easy to see that case 1 for E^n corresponds to $Q=1$ in (3.9) where now we must allow for the possibility that $h^2(t)\equiv h_1^2(t)=1$. Thus expressions (3.15)–(3.24) are correct with $Q=J=1$, $N_J=n$, $\mathscr{F}_J(y^r)\equiv 0$. The relation between the time coordinates is

$$\frac{dy^\delta}{dt}=h^2(t) \tag{4.2}$$

and the original equation

$$2\lambda W_t+\sum_{l=1}^{n}W_{x^l}^2=0 \tag{4.3}$$

maps to

$$2\lambda Z_{y^\delta}+\sum_{i=1}^{n}\left(Z_{z^i}^2+\frac{\alpha}{2}(z^i)^2-2\beta z^1\right)=0. \tag{4.4}$$

Using time translation and dilation invariance for simplifica-

tion, we obtain the following distinct possibilities:

(a) $\alpha>0,\ \beta=0,$
$h^2(t)=(1+t^2)^{-1},\quad h^{-1}(t)z^i(y^a)=x^i;$

(b) $\alpha<0,\ \beta=0,$
$h^2(t)=(1-t^2)^{-1},\quad h^{-1}(t)z^i(y^a)=x^i;$

(c) $\alpha<0,\ \beta=0,$
$h^2(t)=t^{-1},\quad \sqrt{|t|}\,z^i(y^a)=x^i;$

(d) $\alpha=0,\ \beta\neq 0,$
$h^2(t)=t^{-2},\quad tz^i(y^a)=x^i+(\delta^{i1}/2t)\beta;$

(e) $\alpha=0,\ \beta\neq 0,$
$h^2(t)=1,\quad z^i(y^a)=x^i+\delta^{i1}(\beta/2)t^2;$

(f) $\alpha=\beta=0,$
$h^2(t)=t^{-2},\quad tz^i(y^a)=x^i;$

(g) $\alpha=\beta=0,$
$h^2(t)=1,\quad z^i(y^a)=x^i.$

In each case the y^a are orthogonal separable coordinates for the Hamilton–Jacobi equation

$$\sum_{i=1}^{n}\left(Z_{z^i}^2+\frac{\alpha}{2}(z^i)^2-2\beta z^1\right)=E \tag{4.5}$$

and

$$R=\sum_{i=1}^{n}\left[\frac{1}{4}(\dot h^{-2})(z^i)^2\right]-\frac{\dot c^1}{h}z^1. \tag{4.6}$$

Basically, all case 1 R-separable systems originate from separable systems for the zero-potential equation (4.5), $\alpha=\beta=0$. For each of the types (a)–(g) one need merely determine which of the zero-potential separable systems remains separable for an added linear or quadratic equation.

For example, if $n=2$ there are four separable systems in the zero-potential types (f) and (g): Cartesian, polar, parabolic, and elliptic. For types (a)–(c), Cartesian, polar, and elliptic coordinates remain separable. Thus there are a total of $2(4)+3(3)+2(2)=21$ R-separable systems corresponding to case 1. See Ref. 11, Chap. 2, for more details.

A classification of case 2 coordinates for E^n with general n has recently been worked out by Reid.[15] Reid shows that the Hamiltonian can always be written in the form (3.10) where the functions $h_J^2(t)$ can be selected from

$$h_J^2(t)=\begin{cases}[(t+A_J)^2+B_J^2]^{-1},\\ |(t+A_J)^2-B_J^2|^{-1},\\ (t+A_J)^{-2},\\ (t+A_J)^{-1}.\end{cases}$$

He has also worked out the case 1 separable systems for general n.

V. THE TIME-DEPENDENT SCHRÖDINGER EQUATION

Our results extend rather easily to the time-dependent Schrödinger (or heat) equation

$$2\lambda\psi_t+\Delta_n\psi+2\lambda\sum_{l=1}^{n}A^l(\mathbf{x},t)\psi_{x^l}+\lambda^2V(\mathbf{x},t)\psi=0, \tag{5.1}$$

where Δ_n is the Laplace–Beltrami operator on the Riemannian manifold V^n,

$$\Delta_n=\frac{1}{\sqrt{g}}\sum_{l,m=1}^{n}\partial_{x^l}(\sqrt{g}\,g^{lm}\partial_{x^m}),\quad g^{-1}=\det(g^{lm}).$$

In analogy with Sec. I, one writes

$$\hat\psi(x,t,\tau)=e^{\lambda\tau}\psi(x,t),$$

where $\hat\psi$ satisfies the Laplace equation

$$\Delta_{n+2}\hat\psi=0,$$

$$\Delta_{n+1}=\frac{1}{\sqrt{K}}\sum_{i,j=1}^{n+2}\partial_{z^i}(\sqrt{K}\,K^{ij}\partial_{z^j}), \tag{5.2}$$

$$K^{-1}=\det(K^{ij})=-g^{-1},$$

and (K^{ij}), z^i are defined by (1.2). Equations (1.3)–(1.11) continue to hold, but, from the theory of R-separation for Laplace equations,[4,16] the R-separable solutions for (5.1) take the form

$$\psi(x,t)=\exp[-\lambda R(\mathbf{y})+S(\mathbf{y})]\prod_{i=1}^{n+1}\psi^{(i)}(y^i), \tag{5.3}$$

where R and S do not depend on the separation parameters. Theorem 1 holds with only minor modifications. The analysis of Secs. I–III for the Hamilton–Jacobi equation applies without change for the second-derivative terms in Eq. (5.2). The only complication is that the Laplace–Beltrami operator also contains first-derivative terms and, if the coefficients of these terms do not have the proper form, they could invalidate the variable separation. We can use our freedom in choosing S to partially offset this difficulty.

To be more specific we consider the case $n_2=n_3=1$ and adopt the notation of Eq. (3.3). Then (5.2) in the separable coordinates $\{y^r,y^a,\mu\}$ takes the form

$$2\hat\psi_{\mu y^r}+\partial_{y^r}(\ln h)\,\hat\psi_\mu+\frac{1}{h}\sum_a\partial_{y^a}(hH_a^{-2}\partial_{y^a})\hat\psi+\left(\sum_l\mathscr{F}_l(y^l)H_l^{-2}\right)\hat\psi_{\mu\mu}=0,\quad h=\prod_a H_a, \tag{5.4}$$

where the conditions for R-separation of the second-derivative terms are as in (2.2). Now set $\hat\psi=e^{S(y^a,y^r)}\Phi$ in (5.4) to obtain

$$2\Phi_{\mu y^r}+\partial_{y^r}(2S+\ln h)\Phi_\mu+\frac{1}{h}\sum_a\partial_{y^a}(hH_a^{-2}\partial_{y^a})\Phi+2\sum_a H_a^{-2}S_a\Phi_a+\left(\sum_l\mathscr{F}_lH_l^{-2}\right)\Phi_{\mu\mu}+\sum_a H_a^{-2}(S_{aa}+S_a^2+S_a\,\partial_{y^a}\ln(hH_a^{-2}))\Phi=0.$$

(Here $S_a=\partial_{y^a}S$, $\Phi_\mu=\partial_\mu\Phi$ but $\mathscr{F}_l$, H_a^{-2} are merely subscripted.) The coefficients of Φ_a, Φ_μ, and Φ, respectively, will be compatible with R-separation in the coordinates $\{y^a,y^r,\mu\}$ if and only if there exist functions $g_i(y^i)$, $k_i(y^i)$, each depending on the variable y^i alone, such that

$$\partial_{y^ay^b}[2S+\ln(hH_a^{-2})]=0,\quad 1\leqslant a<b\leqslant n, \tag{5.5a}$$

$$\partial_{y^r}[2S+\ln h]=\sum_{c=1}^{n_1}g_c(y^c)H_c^{-2}+g_r(y^r), \tag{5.5b}$$

$$\sum_{a=1}^{n_1}H_a^{-2}(S_{aa}-2S_a^2)=\sum_{c=1}^{n_1}k_c(y^c)H_c^{-2}+k_r(y^r). \tag{5.5c}$$

[Note that (5.5b) and (5.5c) imply that the left-hand sides of these expressions must be Stäckel multipliers.] The case

where the new time variable is ignorable ($n_2 = 0$, $n_3 = 2$) leads to conditions (5.5a) and (5.5c) with $k_r \equiv 0$. However, (5.5b) is omitted in this case.

For $V^n = S^n$ we know from Theorem 3 that $\partial_{y^a y^b} \ln(hH_a^{-2}) = 0, a \neq b$ and $\partial_{y^r} h = 0$ so $S = 0$ satisfies the above equations. For $V^n = E^n$, Theorem 4 implies that

$$\partial_{y^a y^b} \ln(hH_a^{-2}) = 0, \quad a \neq b, \quad \partial_{y^r} h = g_r(y^r),$$

so again $S = 0$ satisfies Eqs. (5.5). Thus we have the following.

Theorem 7: For any potential $(A^i(\mathbf{x},t), V(\mathbf{x},t))$ on the spaces S^n or E^n the time-dependent Hamilton–Jacobi equation additively R-separates in a given coordinate system if and only if the corresponding time-dependent Schrödinger equation multiplicatively R-separates in the same coordinates.

In general, R-separation of the Schrödinger equation implies R-separation of the Hamilton–Jacobi equation. However, it is not difficult to find examples where Eq. (5.5) cannot be satisfied, so the converse is false.

See Ref. 17 and references contained therein for applications of R-separation to time-dependent Schrödinger equations.

VI. INTRINSIC CHARACTERIZATION OF THE EQUATIONS

As was pointed out in Sec. I, the time-dependent Hamilton–Jacobi equation (1.1) can be considered as a special case of the conformal Hamilton–Jacobi equation (1.2). This suggests the interest in characterizing those pseudo-Riemannian spaces V^{n+2} for which the infinitesimal distance can be written in the form

$$ds^2 = Q\left(2\,dt\,d\tau + \sum_{a,b=1}^{n} g_{ab}\,dx^a\,dx^b\right), \tag{6.1}$$

where

$$\partial_t g_{ab} = \partial_\tau g_{ab} = 0.$$

Here Q is a nonzero function on V^{n+2}. We will employ the root structure of conformal Killing tensors to provide this characterization.

Let V^m be a pseudo-Riemannian manifold with metric $ds^2 = \Sigma G_{ij}\,dz^i\,dz^j$ in local coordinates $\{z^i\}$, and let V^m be its associated $2m$-dimensional symplectic manifold (with local canonical coordinates $\{z^i, p_j\}$). The Hamiltonian on V^m is $\mathcal{H} = \Sigma G^{ij} p_i p_j$. A (conformal) Killing tensor $\mathcal{P}(z,p)$ on V^m is a function on V^m, a polynomial in the p's with z-dependent coefficients, such that $\{\mathcal{H}, \mathcal{P}\} = \mathcal{R}\mathcal{H}$, where $\mathcal{R}$ is a function on V^m which is also a polynomial in the p's and $\{\cdot,\cdot\}$ is the Poisson bracket. If $\mathcal{R} \equiv 0$ then $\mathcal{P}$ is a Killing tensor. If $\mathcal{P}$ is linear in the p's it is a conformal Killing vector, a Killing vector if $\mathcal{R} \equiv 0$.

Let $\mathcal{A} = \Sigma A^{ij}(z) p_i p_j, A^{ij} = A^{ji}$, be a second-order Killing tensor on V^m. A root $\rho(z)$ of $\mathcal{A}$ is an analytic solution of the characteristic equation

$$\det(A^{ij} - \rho G^{ij}) = 0 \tag{6.2}$$

and an eigenform $\omega = \Sigma q_k dz^k$ corresponding to ρ is a nonzero one-form such that

$$\sum_{j=1}^{m} (A^{ij} - \rho G^{ij}) q_j = 0, \quad i = 1,\ldots,m. \tag{6.3}$$

We denote by W^ρ the vector space (over the reals) generated by the eigenforms corresponding to ρ. Roots and eigenforms are defined independent of local coordinates.

Theorem 8: Necessary and sufficient conditions that the infinitesimal distance $ds^2 = \Sigma G_{ij}\,dz^i\,dz^j$ on the pseudo-Riemannian space V^{n+2} can be expressed in the form

$$ds^2 = Q\left(2\,dt\,d\tau + \sum_{a,b=1}^{n} g_{ab}(x) dx^a\,dx^b\right) \tag{6.4}$$

are the following.

(1) There is a second-order conformal Killing tensor $\mathcal{A} = \Sigma A^{ij} p_{z^i} p_{z^j}$ $(A^{ij} = A^{ji})$ on V^{n+2} with roots 0 (multiplicity n) and $\rho \neq 0$ (multiplicity 2); $\dim W^0 = n$, $\dim W^\rho = 2$.

(2) There are two conformal Killing vectors

$$\mathcal{L}_\alpha = \sum \xi^i_\alpha\, p_{z^i}, \alpha = 1,2, \tag{6.5}$$

on V^{n+2} such that $\mathcal{A} = 2\mathcal{L}_1\mathcal{L}_2$. Furthermore $\mathcal{L}_1$, $\mathcal{L}_2$ are in involution: $\{\mathcal{L}_1, \mathcal{L}_2\} = 0$.

(3) The first covariant derivatives of $\mathcal{A}$ vanish: $A^{ij}{}_{,k} = 0$, $1 \leqslant i,j,k \leqslant n+2$. Here the covariant derivatives are taken with respect to the metric $d\bar{s}^2 = \rho^{-1}\,ds^2$.

Proof: Suppose conditions (1)–(3) are satisfied. It follows immediately from conditions (1) and (3), and the principal result of Eisenhart's paper on symmetric second-order tensor whose covariant derivatives are zero[18] (p. 303), that there is a coordinate system $\{y^1, y^2, x^1, \ldots, x^n\}$ on V^{n+2} with respect to which

$$\mathcal{A} = \sum_{c,d=1}^{2} \varphi^{c,d}(y) p_{y^c} p_{y^d}, \tag{6.6a}$$

$$\mathcal{H} = \sum_{c,d=1}^{2} \varphi^{c,d}(y) p_{y^c} p_{y^d} + \sum_{a,b=1}^{n} \gamma^{a,b}(x) p_{x^a} p_{x^b}. \tag{6.6b}$$

(Although Eisenhart's result is stated only for Riemannian spaces, his proof remains valid for pseudo-Riemannian spaces.) Condition (2) and (6.6a) imply that $\mathcal{L}_\alpha = \Sigma_{c=1}^2 \xi^c_\alpha p_{y^c}$, $\alpha = 1,2$. Since obviously $\{\mathcal{A}, \rho, \mathcal{H}\} = \{2\mathcal{L}_1\mathcal{L}_2, \rho\mathcal{H}\} = 0$ and the $\mathcal{L}_\alpha$ are conformal Killing vectors for $\rho\mathcal{H}$, it follows that $\{\mathcal{L}_\alpha, \rho\,\mathcal{H}\} = 0$. Thus $\Sigma_b \gamma^{a,b} \partial_{x^b} \xi^c_\alpha = 0$ and by the nondegeneracy of $\mathcal{H}$ we have $\partial_{x^b} \xi^c_\alpha = 0$. It follows that there is a coordinate system $\{t, \tau, x^1, \ldots, x^n\}$ on V^{n+2}, such that $t = t(y^1, y^2)$, $\tau = \tau(y^1, y^2)$ and

$$\mathcal{A} = 2p_t\,p_\tau, \quad \rho\mathcal{H} = 2p_t\,p_\tau + \sum_{a,b=1}^{n} \gamma^{a,b}(x) p_{x^a}\,p_{x^b}.$$

Setting $\rho = Q$ we obtain (6.4), where $\Sigma_l\, g_{al}\gamma^{lb} = \delta^b_a$.

Conversely, if the metric on V^{n+2} can be expressed in the form (6.4), it is straightforward to verify that conditions (1)–(3) are satisfied where $\mathcal{L}_1 = p_t$, $\mathcal{L}_2 = p_\tau$. Q.E.D.

With this result one can use existing classifications of separable coordinate systems for Hamilton–Jacobi equations $\Sigma g^{ij} p_i\, p_j = 0$ to classify separable coordinates for the time-dependent equation (1.1), e.g., Ref. 19.

ACKNOWLEDGMENT

This paper was partially supported by NSF Grant No. MCS 82-19847.

[1]V. I. Arnold, *Mathematical Methods of Classical Mechanics* (Springer, New York, 1978).
[2]W. Thirring, *A Course in Mathematical Physics I, Classical Dynamical Systems* (Springer, New York, 1978).
[3]V. N. Shapovalov, "Separation of variables in second-order linear differential equations," Differ. Eqs. **16**, 1212 (1981).
[4]W. Miller, *The Technique of Variable Separation for Partial Differential Equations*, in *Lecture Notes in Physics*, Vol. 189 (Springer, New York, 1983).
[5]R. Kuwabara, "On the symmetry algebra of the Schrödinger wave equation," Math. Japon. **22**, 243 (1977).
[6]L. V. Ovsjannikov, *Group Properties of Differential Equations* (Academic, New York, 1982).
[7]E. G. Kalnins and W. Miller, "Related evolution equations and Lie symmetries," SIAM J. Math. Anal. **16**, 221 (1985).
[8]V. N. Shapovolov, "Stäckel spaces," Siberian Math. J. **20**, 790 (1980).
[9]E. G. Kalnins and W. Miller, "Conformal Killing tensors and variable separation for Hamilton–Jacobi equations," SIAM J. Math. Anal. **14**, 126 (1983).
[10]C. P. Boyer, E. G. Kalnins, and W. Miller, "Separation of variables in Einstein spaces I," J. Phys. A: Math Gen. **14**, 1675 (1981).
[11]W. Miller, *Symmetry and Separation of Variables* (Addison–Wesley, Reading, MA, 1977).
[12]E. G. Kalnins and W. Miller, "Separation of variables on n-dimensional Riemannian manifolds 1. The n-sphere and Euclidean n-space," J. Math. Phys. **27**, 1721 (1986).
[13]L. P. Eisenhart, *Riemannian Geometry* (Princeton U. P., Princeton, 1949).
[14]A. C. T. Wu, "Separability of Schrödinger and Klein–Gordon equations with a vector potential," Nuovo Cimento A **50**, 333 (1967).
[15]G. J. Reid, "R-separation for heat and Schrödinger equations I," SIAM J. Math. Anal. **17**, 646 (1986).
[16]V. N. Shapovalov, "Separation of variables in the nonstationary Schrödinger equation," Sov. Phys. J. **17**, 1718 (1976).
[17]D. R. Truax, "Symmetry of time-dependent Schrödinger equations. I," J. Math. Phys. **22**, 1959 (1981); "Symmetry of time-dependent Schrödinger equations. II," **23**, 43 (1982).
[18]L. P. Eisenhart, "Symmetric tensors of the second order whose first covariant derivatives are zero," Trans. Am. Math. Soc. **25**, 297 (1923).
[19]E. G. Kalnins and W. Miller, "Separation of variables on n-dimensional Riemannian manifolds 3. Conformally Euclidean spaces" (to appear).

Maximum entropy summation of divergent perturbation series

Carl M. Bender, Lawrence R. Mead,[a] and N. Papanicolaou[b]
Department of Physics, Washington University, St. Louis, Missouri 63130

(Received 30 May 1986; accepted for publication 18 November 1986)

In this paper the principle of maximum entropy is used to predict the sum of a divergent perturbation series from the first few expansion coefficients. The perturbation expansion for the ground-state energy $E(g)$ of the octic oscillator defined by $H = p^2/2 + x^2/2 + gx^8$ is a series of the form $E(g) \sim \frac{1}{2} + \Sigma(-1)^{n+1} A_n g^n$. This series is terribly divergent because for large n the perturbation coefficients A_n grow like $(3n)!$. This growth is so rapid that the solution to the moment problem is not unique and ordinary Padé summation of the divergent series fails. A completely different kind of procedure based on the principle of maximum entropy for reconstructing the function $E(g)$ from its perturbation coefficients is presented. Very good numerical results are obtained.

I. INTRODUCTION

There are many processes in nature in which initial information is lost and becomes permanently irretrievable. For example, the diffusion of heat, as described by the equation $\partial u/\partial t = \sigma \nabla^2 u$, is a smoothing process which only proceeds forward in time: given $u(t_1,\mathbf{x})$ we can predict $u(t_2,\mathbf{x})$ $(t_2 > t_1)$, but given $u(t_2,\mathbf{x})$ we cannot postdict $u(t_1,\mathbf{x})$. This is because there are an infinite number of possible initial conditions $u(t_1,\mathbf{x})$ all of which evolve into $u(t_2,\mathbf{x})$. Similarly, there is no unique answer to the question of what is the exact description of an object of which we have a blurred photograph.

In cases such as these there is an interesting approach one can take based on the principle of maximum entropy. We ask a different kind of question which does have a unique answer: what is the most *likely* initial temperature distribution of all those that could have evolved into $u(t_2,\mathbf{x})$? In actual image reconstruction problems the statistical distribution is so sharply peaked that there is no question about the solution to the unblurring problem.

In this paper we examine a similar kind of question in the context of quantum-mechanical perturbation theory. We consider a Stieltjes perturbation series which is so divergent that there is no unique solution to the moment problem. Thus there are an infinite number of possible Stieltjes functions all of which have the same perturbation coefficients. As a result, no ordinary summation procedure, such as Padé summation, can give a unique result for the sum of the series. That is, information about the original function has been blurred or lost in the process of expanding it into its asymptotic expansion. To "sum" the series we use the principle of maximum entropy to determine the most likely function having the asymptotic series coefficients.

II. THEORY

We illustrate this idea by investigating the Rayleigh–Schrödinger series for the ground-state energy of the octic oscillator. The octic oscillator is defined by the Hamiltonian

$$H = p^2/2 + x^2/2 + gx^8 , \tag{1}$$

where $[x,p] = i$. It is fairly easy to compute the Rayleigh–Schrödinger perturbation expansion of the ground-state energy $E(g)$ for this Hamiltonian. This expansion takes the form of a power series in g:

$$E(g) \sim \frac{1}{2} + \sum_{n=1}^{\infty} (-1)^{n+1} A_n g^n \quad (g \to 0) . \tag{2}$$

The perturbation coefficients are obtained from a two-index recursion formula[1]

$$2jC_{n,j} = (j+1)(2j+1)C_{n,j+1} + C_{n-1,j-4} - \sum_{p=1}^{n-1} C_{p,1} C_{n-p,j} , \tag{3}$$

$$A_n = C_{n,1} ,$$

where $C_{n,j}$ satisfies the boundary conditions

$$C_{0,0} = 1, \quad C_{n,j} = 0 \quad (n \geqslant 1,\ j < 1),$$

$$C_{n,j} = 0 \quad (j > 2n+2) .$$

The numerical values of the first few perturbation coefficients are given in Table I.

The growth of A_n for large n is very rapid[2]:

$$A_n \sim \frac{3}{\pi^{3/2} 2^n} \Gamma\left(3n + \frac{1}{2}\right) \left[\frac{\Gamma(\frac{8}{3})}{\Gamma^2(\frac{4}{3})}\right]^{3n+1/2} \quad (n \to \infty) ,$$

or

$$A_n \sim \alpha n^{-1/2} \beta^n (3n)! , \tag{4}$$

where α and β are constants.

The function $E(g)$ is a generalized Stieltjes function.[3]

TABLE I. Values of the first six perturbation coefficients A_n.

n	A_n
1	6.562 5
2	$2.109\,843\,75 \times 10^3$
3	$3.137\,100\,585\,937\,5 \times 10^6$
4	$1.241\,410\,979\,868\,896\,5 \times 10^{10}$
5	$1.031\,871\,179\,414\,999\,2 \times 10^{14}$
6	$1.560\,089\,634\,068\,099\,8 \times 10^{18}$

[a] On leave from Department of Physics, University of Southern Mississippi, Hattiesburg, Mississippi 39401.

[b] Current address: Department of Physics, University of Crete, Iraklion, Crete, Greece.

 0022-2488/87/051016-03$02.50

TABLE II. The diagonal and off-diagonal Padé sequences for $g = \frac{1}{10}, \frac{1}{2}$, and 1. Observe that these sequences are monotone but do not converge to $E(g)$. Clearly, Padé summation is not useful for the octic oscillator. The values of $E_{\text{exact}}(g)$ are taken from Ref. 5.

N	$P^N_{N+1}(\frac{1}{10})$	$P^N_N(\frac{1}{10})$	$P^N_{N+1}(\frac{1}{2})$	$P^N_N(\frac{1}{2})$	$P^N_{N+1}(1)$	$P^N_N(1)$
1	0.526 80	1.015 30	0.527 72	3.0727	0.527 84	5.6445
2	0.530 54	0.965 91	0.531 77	2.8224	0.531 93	5.1429
3	0.532 94	0.939 86	0.534 39	2.6896	0.534 58	4.8768
4	0.534 64	0.923 42	0.536 27	2.6056	0.536 48	4.7082
5	0.535 92	0.911 96	0.537 68	2.5467	0.537 92	4.5902
6	0.536 92	0.903 39	0.538 81	2.5028	0.539 06	4.5020
7	0.537 74	0.896 72	0.539 71	2.4685	0.539 99	4.4332
8	0.538 43	0.891 35	0.540 49	2.4408	0.540 77	4.3776
9	0.539 01	0.886 90	0.541 15	2.4179	0.541 43	4.3316
	$E_{\text{exact}}(\frac{1}{10}) = 0.620\ 51$		$E_{\text{exact}}(\frac{1}{2}) = 0.745\ 51$		$E_{\text{exact}}(1) = 0.820\ 69$	

Therefore, the perturbation coefficients A_n are all positive and can be represented as moments of a non-negative function $p(x)$:

$$A_n = \int_0^\infty p(x) x^{n-1}\, dx \quad (n = 1,2,\ldots)\,. \tag{5}$$

However, because the moments grow more rapidly than $(2n)!$, the function $p(x)$ is not uniquely determined by the numbers A_n (see Ref. 4). Thus the Borel sum $E_B(g)$ of the series (2),

$$E_B(g) = \frac{1}{2} + g\int_0^\infty \frac{dx\, p(x)}{1+gx}, \tag{6}$$

is not uniquely determined.

The rapid growth of A_n also interferes with Padé summation of the perturbation series (2). We know that for a function of Stieltjes the diagonal Padé sequence P^N_N and the off-diagonal Padé sequence $P^N_{N+1}(g)$ converge as $N\to\infty$ for fixed g in the cut g plane. Moreover, for $g>0$, $P^N_N(g)$ forms a monotone decreasing sequence and $P^M_{M+1}(g)$ forms a monotone increasing sequence with $P^N_N(g) > P^M_{M+1}(g)$ for all M and N. However, when A_n grows more rapidly than $(2n)!$ as $n\to\infty$ the Padé sequences are not guaranteed to converge to the correct answer $E(g)$. Indeed, Table II shows that the correct ground-state energy[5] lies well between the limits of the diagonal and off-diagonal Padé sequences. Thus Padé summation is an ineffective numerical tool for summing the divergent series (2).

Although the series in (2) really does not determine $p(x)$ uniquely, we can still try to find the *most likely* function $p(x)$ of all those non-negative functions that solve the moment problem (5). To do so we invoke the principle of maximum entropy[6] (maxent). That is, we seek the distribution $p_N(x)$ that maximizes the Nth entropy functional[7]

$$S_N = -\int_0^\infty dx\,[p_N(x)\ln p_N(x) - p_N(x)] + \sum_{n=1}^N \lambda_n\left[A_n - \int_0^\infty dx\, x^{n-1} p_N(x)\right], \tag{7}$$

where the λ_n are Lagrange multipliers that enforce the moment condition in (5) for $n = 1,2,\ldots,N$.

Varying (7) with respect to λ_n and $p_N(x)$ gives the system of equations

$$A_n = \int_0^\infty dx\, x^{n-1} p_N(x) \quad (n = 1,2,\ldots,N)\,,$$
$$p_N(x) = \exp\left(-\sum_{n=1}^N \lambda_n x^{n-1}\right). \tag{8}$$

A multidimensional Newton's method algorithm for solving (8) is described in Ref. 7. Once $p_N(x)$ is known it is inserted into the integral in (6) to give the maxent value of $E_N(g)$ obtained from N moments. We have solved (8) for the case $N = 2$, 4, and 6. The values of λ_n are given in Table III and the maxent predictions for $E_N(g)$ are given in Table IV. Observe that the numerical results are quite good considering the small number of moments used. Thus the principle of maximum entropy, which has been used in such diverse areas as economics, photographic image reconstruction, and time series analysis, also appears to be very effective at extracting maximal information about the sum of a very divergent series from a small number of perturbation coefficients.

Our results for the energy suggest that the maximum entropy sequence approximating the integral in (6) is con-

TABLE III. The values of the Lagrange multipliers for the $N = 2$, 4, and 6 maxent problems in (8). The input used to determine these values comes from the perturbation coefficients given in Table I.

n	λ_n for $N=2$	λ_n for $N=4$	λ_n for $N=6$
1	$-1.134\ 757\ 74$	$-1.478\ 838\ 47$	$-1.605\ 982\ 10$
2	$3.110\ 419\ 91\times10^{-3}$	$4.986\ 297\ 77\times10^{-3}$	$-5.975\ 387\ 00\times10^{-3}$
3	...	$-1.111\ 269\ 77\times10^{-6}$	$-2.081\ 860\ 33\times10^{-6}$
4	...	$8.095\ 878\ 10\times10^{-11}$	$3.189\ 806\ 41\times10^{-10}$
5	...	...	$-1.833\ 039\ 26\times10^{-14}$
6	...	...	$3.466\ 001\ 94\times10^{-19}$

TABLE IV. $E_N(g)$ for various values of g. Observe that the numerical accuracy is far better than the predictions of Padé theory in Table II, even though the Padé predictions use three times as many perturbation coefficients.

g	$E_2(g)$	$E_4(g)$	$E_6(g)$	$E_{\text{exact}}(g)$
0.0001	0.500 636	0.500 638	0.500 638	0.500 64
0.001	0.505 194	0.505 357	0.505 391	0.505 43
0.01	0.524 490	0.527 740	0.528 811	0.532 10
0.1	0.561 572	0.576 203	0.581 994	0.620 51
0.5	0.592 609	0.618 986	0.630 054	0.745 51
1.0	0.606 451	0.638 322	0.651 910	0.820 69

vergent. On the other hand, little is known about the behavior of the sequence $p_N(x)$ as $N \to \infty$, even for less divergent series for which the Carleman condition is satisfied.[7] A rigorous analysis of these questions should shed light on the true potential of the method.

[1]A derivation of this recursion formula is given in C. M. Bender and T. T. Wu, Phys. Rev. D **7**, 1620 (1973).

[2]C. M. Bender and T. T. Wu, Phys. Rev. Lett. **27**, 461 (1971).

[3]$E(g)$ is a generalized Stieltjes function because the once-subtracted function $F(g) \equiv [E(g) - \frac{1}{2}]/g$ is a function of Stieltjes. To show the $F(g)$ is Stieltjes we recall that (i) $F(g)$ is analytic in the cut g plane [see J. J. Loeffel and A. Martin, CERN Report No. CERN-TH-1167, 1971 (unpublished)]; (ii) $E(g) \sim cg^{1/5}$ ($|g| \to \infty$) in the cut g plane (this is known as Symmanzik scaling); and (iii) the series in (2) is asymptotic to $E(g)$ as $g \to 0$ in the cut g plane [see J. J. Loeffel, A. Martin, B. Simon, and A. S. Wightman, Phys. Lett. B **30**, 656 (1969)].

[4]The Carleman condition, $\Sigma |A_n|^{-1/(2n)} = \infty$, if satisfied, implies that the moment problem has a unique solution. For the quartic oscillator $A_n \sim n!$ and for the sextic oscillator $A_n \sim (2n)!$ so in both of these cases the moment problem does have a unique solution.

[5]Numerical values for $E(g)$ are given in F. T. Hioe, D. MacMillen, and E. W. Montroll, J. Math. Phys. **17**, 1320 (1976).

[6]For a good review of the maximum entropy principle see E. T. Jaynes, Proc. IEEE **70**, 939 (1982) and *The Maximum Entropy Formalism*, edited by R. D. Levine and M. Tribus (MIT, Cambridge, MA, 1979).

[7]For a full discussion of the maximum entropy principle applied to the solution of moment problems for which the Carleman condition is satisfied see L. R. Mead and N. Papanicolaou, J. Math. Phys. **25**, 2404 (1984).

Symmetries of static, spherically symmetric space-times

Ashfaque H. Bokhari and Asghar Qadir[a)]
Department of Physics, University of Texas, Austin, Texas 78712

(Received 21 February 1986; accepted for publication 10 December 1986)

In this paper it is shown that reduction from maximal to minimal static, spherical symmetry of a space-time occurs in only one step reducing the number of independent Killing vector fields from 10 to 4. Maximal symmetry corresponds only to the de Sitter, anti-de Sitter, and Minkowski metrics, without reference to the Einstein field equations.

I. INTRODUCTION

By Noether's theorem[1] the symmetries of a Lagrangian imply the existence of conserved quantities. These symmetries have been used[2] to obtain the constants of motion for the trajectories of freely falling particles in the field of a gravitating source, e.g., in the Schwarzschild, Reissner–Nordstrom, and Kerr–Newmann geometries. In general relativity, symmetries are expressed in terms of Killing vector fields (or Killing tensor fields, as in the case of the Kerr–Newmann geometry[3]). The number of independent Killing vector fields (KV's) is related to the number of generators of the corresponding symmetry group. Rather than trying to work out the symmetries of some particular space-time by group theoretic methods, we work out all possible KV's for a static, spherically symmetric space-time by the process of elimination.

A Killing vector field is a vector field $\mathbf{k}$ relative to which the Lie derivative of the metric tensor g is zero, i.e.,

$$\mathcal{L}_{\mathbf{k}} g = 0. \tag{1}$$

In a torsion-free space, in a coordinate basis, the Killing equation reduces to[4]

$$g_{ab,c}k^{c} + g_{ac}k^{c}{}_{,b} + g_{bc}k^{c}{}_{,a} = 0 \quad (a,b,c = 0,...,3). \tag{2}$$

The number of KV's for the de Sitter, anti-de Sitter, and Minkowski geometries are known to be maximal (10) and for the Schwarzchild geometry to be minimal (4). A point that needs to be determined is whether the gaps in the number of KV's from the maximal to the minimal symmetry for a static, spherically symmetric space-time can be filled or not. In this paper we examine this point. We start by considering the most general static, spherically symmetric line element,

$$ds^2 = e^{\nu(r)}\,dt^2 - e^{\lambda(r)}\,dr^2 - r^2\,d\theta^2 - r^2\sin^2\theta\,d\phi^2. \tag{3}$$

The Killing equations are solved for all possible cases. It is found that there can be either ten or four KV's for the metric given by Eq. (3), in general.

The authors have not found any work in recent literature exactly along the lines followed here. However, there are two major lines followed that are fairly close to the approach taken in this paper. One follows the standard work of Petrov,[5] where he considers Einstein spaces, and the other is the work on exact solutions of Einstein's field equations, given by Kramer, Stephani, MacCallum, and Herlt,[6] for example.

Since we are not dealing with Einstein spaces only, the work on Einstein spaces does not apply to our considerations. We have replaced the requirement by the conditions of spherical symmetry and staticity. Thus ours is, in many ways, a more restrictive assumption. Nevertheless, there are many examples of spherically symmetric, static metrics that do not belong to Einstein spaces.

Of course, all cases considered by us are exact solutions of *some* Einstein field equations. However, the procedure generally adopted is to deal with given Einstein equations and determine the symmetry of their exact solutions. We have reversed the order to deal with a given symmetry and determine, where possible, the stress-energy tensor for such a symmetry. This procedure may seem to provide a pointless approach at first sight. However, our point of view was to look only at the symmetries obtaining in a space-time, provided that it is static and spherically symmetric.

It is instructive to put the work in group theoretic terms. What we show in our paper is that the maximal symmetry group of a spherically symmetric static four-dimensional space-time is one of the three: (a) SO(1,4), (b) SO(2,3), or (c) $SO(1,3)\otimes\mathbb{R}^4$. Here the $\mathbb{R}^4$ give the four space-time translations. Thus the groups are either the de Sitter, anti-de Sitter, or Poincaré groups. The minimal allowed symmetry group is $SO(3)\otimes\mathbb{R}$, where the $\mathbb{R}$ gives time translation and SO(3) the spatial rotations only. The remarkable result is that there does not exist any group properly containing the minimal group and properly contained in one of the minimal groups.

In the next section we explain the procedure adopted for finding KV's. This procedure is applied, in full, to one case in Sec. III while mentioning the results for all other cases without giving details. Finally, we state our main result in the form of a theorem in the concluding section.

II. PROCEDURE ADOPTED

To find the KV's for the metric given by Eq. (3) we write the complete set of first-order coupled partial differential equations obtained by inserting Eq. (3) into Eq. (2). Now, by differentiating these equations, we can obtain identities between pairs of equations, leading to first- or second-order partial differential equations that are decoupled. We then solve these differential equations by using the separation of variables. The separation and integration constants

[a)] Also the Centre of Basic Science, UGC, Islamabad, Pakistan.

 0022-2488/87/051019-04$02.50

are then allowed to take all possible values, i.e., positive, zero, or negative. In some cases the positivity of $e^{\nu(r)}$ and $e^{\lambda(r)}$ imposes a constraint on the choice of the integration constant.

Having obtained some partial solution, the expressions are inserted back into the original set of ten Killing equations. Consistency places further constraints on the integration and separation constants. This procedure is used iteratively till the general solution to the coupled differential equations is obtained. In general the solution will depend on the choice of $\nu(r)$ and $\lambda(r)$. However, for the solution of some of the equations to exist, these functions will have to satisfy some differential equations. In these cases the differential equations are solved to yield the metric coefficients for zero–zero and one–one components. It should be stressed that the Einstein field equations have *not* been appealed to.

III. APPLICATION OF THE PROCEDURE

The Killing equations for the metric given by Eq. (3) are

$$\nu'(r)k^1 + 2k^0_{,0} = 0, \tag{4}$$

$$e^{\nu(r)}k^0_{,1} - e^{\lambda(r)}k^1_{,0} = 0, \tag{5}$$

$$e^{\nu(r)}k^0_{,2} - r^2k^2_{,0} = 0, \tag{6}$$

$$e^{\nu(r)}k^0_{,3} - r^2\sin^2\theta k^3_{,0} = 0, \tag{7}$$

$$\lambda'(r)k' + 2k^1_{,1} = 0, \tag{8}$$

$$e^{\lambda(r)}k^1_{,2} + r^2k^2_{,1} = 0, \tag{9}$$

$$e^{\lambda(r)}k^1_{,3} + r^2\sin^2\theta k^3_{,1} = 0, \tag{10}$$

$$k^1 + rk^2_{,2} = 0, \tag{11}$$

$$k^2_{,3} + \sin^2\theta\, k^3_{,2} = 0, \tag{12}$$

$$k^1 + r\cot\theta k^2 + rk^3_{,3} = 0, \tag{13}$$

where a prime denotes differentiation with respect to r. Equation (8) is a differential equation involving k and its derivative with respect to r. Thus it can be integrated with respect to r to yield

$$k^1 = B(t,\theta,\phi)e^{-\lambda(r)/2}, \tag{14}$$

where $B(t,\theta,\phi)$ is the "constant" of integration. Now there are two cases: (I) $B \neq 0$, and (II) $B = 0$. We first consider case (I).

Differentiating Eqs. (9) and (11) with respect to θ and r and comparing gives (as $B \neq 0$)

$$B(t,\theta,\phi)_{\theta\theta}/B(t,\theta,\phi) = -(1 + r\lambda'(r)/2)e^{-\lambda(r)} = -\alpha, \tag{15}$$

where $B_{\theta\theta} = \partial^2 B/\partial\theta^2$. Since the left-hand side of Eq. (15) is not a function of r whereas the right-hand side is, α is a separation constant. Now there are three possibilities: (1) $\alpha > 0$, (2) $\alpha < 0$, or (3) $\alpha = 0$. First consider case (1). Here Eq. (15) can be easily solved to give

$$B(t,\theta,\phi) = B_1(t,\phi)\cos\sqrt{\alpha}\theta + B_2(t,\phi)\sin\sqrt{\alpha}\theta, \tag{16}$$

$$e^{-\lambda(r)} = (\alpha + \beta r^2), \tag{17}$$

where $B_1(t,\theta)$, $B_2(t,\theta)$, and β are "constants" of integration. Again there are three possible cases: (a) $\beta < 0$, (b) $\beta > 0$, or (c) $\beta = 0$. Consider case (a) first.

Differentiating Eqs. (4) and (5) with respect to r and t, respectively, and comparing, using Eqs. (14), (16), and (17), gives

$$\frac{B_1(t,\phi)_{tt}\cos\sqrt{\alpha}\theta + B_2(t,\phi)_{tt}\sin\sqrt{\alpha}\theta}{B_1(t,\phi)\cos\sqrt{\alpha}\theta + B_2(t,\phi)\sin\sqrt{\alpha}\theta} = \tfrac{1}{2}\{\nu''(\alpha + \beta r^2) + \beta\nu' r\} = \gamma, \tag{18}$$

where γ is the separation constant. Once again there are three possibilities: (i) $\gamma > 0$, (ii) $\gamma < 0$, or (iii) $\gamma = 0$. We first consider case (i). Equations (18) can be solved for both sides to yield

$$B_1(t,\phi) = B_{11}(\phi)\cosh\sqrt{\gamma}t + B_{12}(\phi)\sinh\sqrt{\gamma}t, \tag{19}$$

$$B_2(t,\phi) = B_{21}(\phi)\cosh\sqrt{\gamma}t + B_{22}(\phi)\sinh\sqrt{\gamma}t,$$

$$e^{\nu(r)} = -(\gamma/\alpha\beta)(\alpha + \beta r^2) = -(\gamma/\alpha\beta)e^{-\lambda(r)}. \tag{20}$$

Notice from Eq. (20) that for $e^{\nu(r)}$ to be positive γ and β should have opposite signs and α be nonzero. Using the value of k^1 in Eq. (11) with Eqs. (19) and (20) and integrating with respect to θ gives

$$\begin{aligned} k^2 = &-[(\alpha + \beta r^2)/\sqrt{\alpha}r][\{B_{11}(\phi)\cosh\sqrt{\gamma}t \\ &+ B_{12}(\phi)\sinh\sqrt{\gamma}t\}\sin\sqrt{\alpha}\theta \\ &- \{B_{21}(\phi)\cosh\sqrt{\gamma}t + B_{22}(\phi)\sinh\sqrt{\gamma}t\}\cos\sqrt{\alpha}\theta] \\ &+ A_1(t,r,\phi). \end{aligned} \tag{21}$$

Differentiating this equation with respect to r and comparing with Eq. (9) it is easily seen that A is a function of t and ϕ only. Integrating θ in Eq. (6) using Eq. (21), yields

$$\begin{aligned} k^0 = &[\beta r/(\gamma(\alpha + \beta r^2))]^{1/2}[\{B_{11}(\phi)\sinh\sqrt{\gamma}t \\ &+ B_{12}(\phi)\cosh\sqrt{\gamma}t\}\cos\sqrt{\alpha}\theta \\ &+ \{B_{21}(\phi)\sinh\sqrt{\gamma}t + B_{22}(\phi)\cosh\sqrt{\gamma}t\}\sin\sqrt{\alpha}\theta] \\ &- \alpha\beta\theta A_1(t,\phi)_t/\gamma(\alpha + \beta r^2) + A_2(t,r,\phi). \end{aligned} \tag{22}$$

Differentiating this equation with respect to r and comparing with Eq. (5) it is found that A_1 and A_2 are functions of ϕ only and of t and ϕ, respectively. Equation (10) can be integrated with respect to r. Using k^1 and Eqs. (19) and (20), one obtains

$$k^3 = [(\alpha+\beta r^2)^{1/2}/\alpha r \sin^2\theta]\,[\{B_{11}(\phi)_\phi \cosh\sqrt{\gamma}t + B_{12}(\phi)_\phi \sinh\sqrt{\gamma}t\}\cos\sqrt{\alpha}\theta + \{B_{21}(\phi)_\phi \cosh\sqrt{\gamma}t + B_{22}(\phi)_\phi \sinh\sqrt{\gamma}t\}\sin\sqrt{\alpha}\theta] + A_3(t,\theta,\phi). \quad (23)$$

Using Eqs. (22) and (23) in Eq. (7) it is easily checked that A_2 and A_3 are functions of t only and of θ and ϕ, respectively. To check consistency use Eqs. (21) and (23) in Eq. (12), which implies that it is satisfied only if

$$-2[(\alpha+\beta r^2)/\alpha r]^{1/2}[(B_{11}(\phi)_\phi \cosh\sqrt{\gamma}t + B_{12}(\phi)_\phi \sinh\sqrt{\gamma}t)(\sqrt{\alpha}\sin\theta\sin\sqrt{\alpha}\theta + \cos\theta\cos\sqrt{\alpha}\theta) - (B_{21}(\phi)_\phi \cosh\sqrt{\gamma}t + B_{22}(\phi)_\phi \sinh\sqrt{\gamma}t)\times(\sqrt{\alpha}\sin\theta\cos\sqrt{\alpha}\theta - \cos\theta\sin\sqrt{\alpha}\theta)] + [A_3(\theta,\phi)_\theta \sin^2\theta + A_1(\phi)_\phi]\sin\theta = 0.$$

This equation is satisfied if the coefficients of r and $\sin\theta$ are separately zero. Thus

$$(B_{11}(\phi)_\phi \cosh\sqrt{\gamma}t + B_{12}(\phi)_\phi \sinh\sqrt{\gamma}t)(\sqrt{\alpha}\sin\theta\sin\sqrt{\alpha}\theta + \cos\theta\cos\sqrt{\alpha}\theta) - (B_{21}(\phi)_\phi \cosh\sqrt{\gamma}t + B_{22}(\phi)_\phi \sinh\sqrt{\gamma}t)(\sqrt{\alpha}\sin\theta\cos\sqrt{\alpha}\theta - \cos\theta\sin\sqrt{\alpha}\theta) = 0, \quad (24)$$

$$A_3(\theta,\phi) = \cot\theta A_1(\phi)_\phi + A_4(\phi). \quad (25)$$

There are two possibilities for Eq. (24) to be satisfied: $(*)\alpha = 1$, $(\dagger)\alpha \neq 1$. In the first case we have the *de Sitter metric* with $\beta = -1/R^2$. From Eq. (24) we see that B_{11} and B_{12} are constants, say C_1 and C_2, respectively. Differentiating and using Eq. (23) in Eq. (13), remembering Eq. (25), we obtain

$$\begin{aligned} B_{21} &= C_3\cos\phi + C_4\sin\phi, \\ B_{22} &= C_5\cos\phi + C_6\sin\phi, \\ A_1 &= C_7\cos\phi + C_8\sin\phi, \\ A_4 &= C_9. \end{aligned} \quad (26)$$

Now from Eq. (5) it can be easily checked that A_2 is an integration constant, say C_{10}. Thus we obtain ten KV's for the de Sitter metric:

$$\begin{aligned} k^0 &= [r/R^2/(\gamma(1-r^2/R^2))^{1/2}]\,[(C_1\sinh\sqrt{r}t + C_2\cosh\sqrt{\gamma}t)\times\cos\theta + \{(C_3\cos\phi + C_4\sin\phi)\sinh\sqrt{\gamma}t + (C_s\cos\phi + C_6\sin\phi)\cosh\sqrt{r}t\}\sin\theta] + C_7, \\ k^1 &= (1-r^2/R^2)^{1/2}[(C_1\cosh\sqrt{\gamma}t + C_2\sinh\sqrt{\gamma}t)\cos\theta + \{(C_3\cos\phi + C_4\sin\phi)\cosh\sqrt{\gamma}t + (C_5\cos\phi + C_6\sin\phi)\sinh\sqrt{\gamma}t\}\sin\theta], \quad (27) \\ k^2 &= -[(1-r^2/R^2)^{1/2}/r]\,[(C_1\cosh\sqrt{\gamma}t + C_2\sinh\sqrt{\gamma}t)\times\sin\theta - \{(C_3\cos\phi C_4\sin\phi)\cosh\sqrt{\gamma}t + (C_5\cos\phi + C_6\sin\phi)\sinh\sqrt{\gamma}t\}\cos\theta] + (C_8\cos\phi + C_9\sin\phi), \\ k^3 &= [(1-r^2/R^2)^{1/2}/r\sin\theta]\,[(-C_3\sin\phi + C_4\cos\phi)\times\cosh\sqrt{\gamma}t + (-C_5\sin\phi + C_6\cos\phi)\sinh\sqrt{\gamma}t] + \cot\theta(-C_8\sin\phi + C_9\cos\phi) + C_{10}. \end{aligned}$$

Anti-de Sitter metric: Another possibility is the case (1.b. ii.*). Following the same procedure as in the first case (replacing γ by $-\gamma$ and β by $1/R^2$), we can obtain the independent Killing vector fields for the anti-de Sitter metric. These Killing vector fields are again 10, with $\sinh\sqrt{\gamma}t(\cosh\sqrt{\gamma}t)$ replaced by $\sin\sqrt{\gamma}t$ $(\cos\sqrt{\gamma}t)$ in Eqs. (27).

The Minkowski metric: Now consider the case (1.C.iii.*). Equation (18) yields

$$e^{\nu(r)} = e^{a+br}. \quad (28)$$

Now there are two possibilities: $(\$)b = 0$, $(\#)b = 0$. The first case gives the Minkowski metric (modulo a constant zero–zero metric coefficient which could be taken to be unity). Again we have ten KV's:

$$\begin{aligned} k^0 &= r[C_1\cos\theta + C_2\cos(\phi+C_3)\sin\theta] + C_4, \quad (29a) \\ k^1 &= t[C_1\cos\theta - C_2\cos(\phi+C_3)\sin\theta] + C_5\cos\theta + C_6\cos(\phi+C_7)\sin\theta, \quad (29b) \\ k^2 &= -(t/r)[C_1\sinh\theta - C_2\cos(\phi+C_3)\cos\theta] - (1/r)[C_5\sin\theta - C_6\cos(\phi+C_7)\cos\theta] + C_8\cos(\phi+C_9), \quad (29c) \\ k^3 &= -(1/r\sin\theta)[tC_2\sin(\phi+C_3) + C_6\sin(\phi+C_7) - C_8\sin(\phi+C_9)\cot\theta] + C_{10}. \quad (29d) \end{aligned}$$

We now consider the case $(\dagger)$ in which it is easy to see the reduction of KV's from 10 to 4 only. In case $(\dagger)$ Eq. (24) is satisfied if B_{11}, B_{12}, B_{21}, and B_{22} are all constants. To check consistancy we use Eqs. (14), (16), (17), and (21) in Eq. (13). It turns out that all the above constants are identically zero. In this case A_1 and A_4 are given by Eq. (26). The KV's are

$$\begin{aligned} k^0 &= C_1 \\ k^1 &= 0, \\ k^2 &= C_2\cos\phi + C_3\sin\phi, \\ k^3 &= \cot\theta(-C_2\sin\phi + C_3\cos\phi) + C_4. \end{aligned} \quad (30)$$

Notice that these are the usual KV's for the Schwarzschild metric. Here, however, the metric tensor has one–one and zero–zero components given by Eqs. (17) and (20).

We now write $e^{\nu(r)}$ for those remaining subcases of case I that are permissible within the requirement of positivity of $e^{\nu(r)}$ and $e^{\lambda(r)}$:

Cases	$e^{\nu(r)}$
(1.a.iii)	$a+(b/\sqrt{\beta})\sinh^{-1}\sqrt{\beta/\alpha}\,r$,
(1.b.iii)	$a+(b/\sqrt{\beta})\sinh^{-1}\sqrt{\beta/\alpha}\,r$,
(1.c.i)	$-2\gamma/\alpha\nu''$,
(1.c.ii)	$2\gamma/\alpha\nu''$,
(1.c.iii$\neq$)	$a+br$,
(2.b.ii)	$-(\gamma/\alpha\beta)(\alpha+\beta r^2)$,
(2.b.iii)	$a+(b/\sqrt{\beta})\sinh^{-1}\sqrt{\beta/\alpha}\,r$,
(3.b.ii)	$-2\gamma/\beta r(\nu' r)'$,
(3.b.iii)	ar.

There are only four KV's in each of the above cases. These vector fields are given by Eqs. (30).

Now consider case (II). In this case Eq. (14) gives

$$k' = 0.$$

Using this value of k in Eqs. (4) and (5) it is easily seen that k^0 is a function of θ and ϕ only. Also Eqs. (9) and (11) imply that k^2 can depend only on t and ϕ. Differentiating Eq. (7) with respect to t we obtain

$$k^3 = A_1(\theta,\phi) + A_2(\theta,\phi)t. \tag{31}$$

Differentiating Eq. (6) with respect to θ and solving gives

$$k^0 = A_3(\phi) + A_6(\phi)\theta. \tag{32}$$

Also Eq. (6) can be differentiated first with respect to t and then integrated with respect to t to yield

$$k^2 = A_5(\phi) + A_6(\phi)t. \tag{33}$$

Solving Eq. (12) with Eqs. (31) and (33) we obtain

$$\begin{aligned} A_1 &= \cot\theta A_5(\phi)_\phi + A_7(\phi), \\ A_2 &= \cot\theta A_6(\phi)_\phi + A_8(\phi). \end{aligned} \tag{34}$$

For consistency using values of k^0 and k^2 in Eq. (6) it turns out that

$$A_4(\phi)e^{\nu(r)} = r^2 A_6(\phi). \tag{35}$$

Equation (35) can be separated in r and ϕ with the separation constant γ and solved to yield

$$e^{\nu(r)} = \gamma r^2, \quad A_6(\phi) = \gamma A_4(\phi). \tag{36}$$

Notice that the separation constant can be greater than zero here [as $\gamma \leqslant 0$ in Eq. (36) is not permissible]. Using the above results in Eq. (13) we get

$$\begin{aligned} A_4 &= C_3\cos\phi + C_4\sin\phi, \\ A_5 &= C_1\cos\phi + C_2\sin\phi, \\ A_7 &= C_5, \quad A_8 = C_6. \end{aligned} \tag{37}$$

To check consistency from Eq. (16) it is easy to see that C_3, C_4, and C_6 are zero and A_3 is a constant. Using the values from Eqs. (34) and (37) in Eqs. (31)–(33) we obtain the same four KV's given by Eq. (30).

Notice that in this case since $A_4 = 0$, Eq. (35) implies that A is zero. Thus $e^{\nu(r)}$ or $e^{\lambda(r)}$ in this case have no constraints. This leads to the fact that while maintaining spherical symmetry and staticity for arbitrary $\nu(r)$ and $\lambda(r)$, the form of KV's is given by Eqs. (30).

IV. SUMMARY AND CONCLUSION

It is found that in the most general static, spherically symmetric space-time there can either be *ten* or *four* KV's without reference to the Einstein field equations. The first case includes the de Sitter (positive cosmological constant), anti-de Sitter (negative cosmological constant), and Minkowski metrics with the requirement that $\alpha = 1$ in these cases. In the case $\alpha \neq 1$, Eq. (24) is satisfied if B_{11} to B_{22} are zero. The requirement reduces the number of KV's from 10 to 4, not *allowing any number in between*.

The separation constants in the above cases have been allowed to take all possible values. It turns out that in each case different metrics are obtained. All these metrics admit of only four KV's, given by Eqs. (30).

In the case II ($B = 0$) there are four KV's given by Eqs. (30). However, the metric has no constraints on the functional form of $e^{\nu(r)}$ and $e^{\lambda(r)}$. Hence the other cases having four KV's can be incorporated into the case II. The constraints would now have to be obtained from the Einstein field equations. Thus we have the following.

Theorem: (i) Static, spherically symmetric space-times admit either ten or four KV's.

(ii) In the case of four KV's there is no restriction on the metric but in the case of ten KV's the metric is either de Sitter, anti-de Sitter, or Minkowski.

Notice that the maximal symmetry corresponds to the Lie algebra of the group O(1,4) or the Poincaré group. The minimal symmetry corresponds to one translation (along the time axis) and three "rotational" parameters, i.e., $O(3) \otimes \mathbb{R}(1)$. It would be interesting to look at the reduction of symmetries in more general cases, e.g., without assuming that the metric is static, or taking static and axially symmetric metrics, etc. In these cases the present procedure would be too complicated and group theoretic methods would have to be used.

ACKNOWLEDGMENT

We are gratefully indebted to Dr. Hasan Azad for useful discussions during the completion of this work.

[1] R. Abraham and J. E. Marsden, *Foundations of Mechanics* (Benjamin, New York, 1978).

[2] C. W. Misner, K. S. Thorne, and J. A. Wheeler, *Gravitation* (Freeman, San Francisco, 1973); S. W. Hawking and G. F. R. Ellis, *The Large Scale Structure of Space-Time* (Cambridge U.P., Cambridge, 1968).

[3] M. Walker and R. Penrose, Commun. Math. Phys. **18**, 265 (1970); M. Walker, Ph. D. thesis, Birbeck College, London University, 1969.

[4] H. Stephani, *General Relativity—An Introduction to the Theory of the Gravitational Field* (Cambridge, U.P., Cambridge, 1982).

[5] A. Z. Petrov, *Einstein Space* (Pergamon, New York, 1969).

[6] D. Kramer, H. Stephani, M. MacCallum, and E. Herlt, *Exact Solutions of Einstein's Field Equations* (Cambridge U.P., Cambridge, 1980).

Propagators for massive vector fields in anti-de Sitter space-time using Stueckelberg's Lagrangian

H. Janssen and C. Dullemond
Institute for Theoretical Physics, University of Nijmegen, Nijmegen, The Netherlands

(Received 6 June 1986; accepted for publication 10 December 1986)

Expressions are found for homogeneous and inhomogeneous propagators for vector fields of arbitrary mass in anti-de Sitter space-time using a generalization of Stueckelberg's Lagrangian for a massive vector field. The massless case (quantum electrodynamics) is also considered by taking the appropriate zero-mass limit.

I. INTRODUCTION

Recently there has been great interest in field theory in anti-de Sitter (ADS) space-time because this space occurs as a natural space-time background in extended supergravity and Kaluza–Klein theories.[1] Another place of interest is in models for hadrons in which confinement has been built in by means of an ADS bag in which quarks and gluons move along geodesics.[2] In order to take quantum effects into account, like gluon exchange, one needs the use of propagators.

Early work on propagators has been done by Fronsdal[3] for homogeneous ADS scalar propagators and Fronsdal and Haugen[4] for spinor fields. The massless case for arbitrary spin has been studied by Fronsdal[5] and Fang and Fronsdal.[6] Anti-de Sitter quantum electrodynamics (QED) was recently developed by Binegar *et al.*[7] for a particular gauge fixing choice $c = \frac{1}{3}$, and by Gazeau[8] for the general case in the framework of the representation theory of the ADS SO(3,2) symmetry group. Vector propagators in maximally symmetric spaces have recently been studied by Allen and Jacobson.[9]

In previous papers, expressions for SO(3,2) symmetric massive scalar and spinor propagators, homogeneous as well as inhomogeneous, were found using configuration space methods.[10,11] The same method will be applied in this article.

Anti-de Sitter space-time is not simply connected; therefore we need the introduction of a covering space.[3,10] Furthermore, implicit boundary conditions at infinity have to be imposed in order to get a well-posed Cauchy problem and to make the propagators unique.[10,12] This can be done by requiring the propagators to approach zero "sufficiently" fast when a certain invariant quantity approaches minus infinity.[10]

In Sec. II we give a review of the Stueckelberg method for obtaining massive vector propagators in Minkowski space. In the massless limit one obtains QED with Gupta–Bleuler quantization. In Sec. III we obtain the appropriate vector field equations with arbitrary mass in ADS space-time. In Sec. IV we construct homogeneous and inhomogeneous vector propagators and in Sec. V we discuss the normalization of the propagators using the quantum conditions. In Sec. VI we discuss the massless case (QED with Gupta–Bleuler quantization) and compare our results with those obtained by Gazeau.[8]

II. MASSIVE VECTOR FIELDS IN MINKOWSKI SPACE

Consider Stueckelberg's Lagrangian for a massive vector field in Minkowski space with metric $\eta_{\mu\nu} = \mathrm{diag}(1,-1,-1,-1)$[13]:

$$\mathscr{L} = -\tfrac{1}{4}F_{\mu\nu}F^{\mu\nu} + \tfrac{1}{2}\mu^2 A_\mu A^\mu - \tfrac{1}{2}c(\partial_\mu A^\mu)^2, \tag{2.1}$$

with field equations (for $c \neq 0$)

$$(\Box + \mu^2)A^\mu - (1-c)\partial^\mu\partial_\nu A^\nu = 0 \tag{2.2}$$

and

$$[\Box + \mu^2/c]\partial_\mu A^\mu = 0. \tag{2.3}$$

The Feynman propagator in k space is given by

$$G_{\mu\nu}(k,\mu^2) = -i\left\{\frac{\eta_{\mu\nu} - k_\mu k_\nu/\mu^2}{k^2-\mu^2+i\epsilon} + \frac{k_\mu k_\nu/\mu^2}{k^2-\mu^2/c+i\epsilon}\right\}, \tag{2.4}$$

where the first part is transverse and the second part longitudinal (pure gauge). The massless limit is given by

$$G_{\mu\nu}(k) = -i\left\{\frac{\eta_{\mu\nu}}{k^2+i\epsilon} + \left(\frac{1-c}{c}\right)\frac{k_\mu k_\nu}{(k^2+i\epsilon)^2}\right\}, \tag{2.5}$$

or in x space,

$$G_{\mu\nu}(x) = \frac{1}{4\pi^2 i}\left\{\left(1+\frac{1}{c}\right)\frac{\eta_{\mu\nu}}{2(x^2-i\epsilon)} + \left(1-\frac{1}{c}\right)\frac{x_\mu x_\nu}{(x^2-i\epsilon)^2}\right\}. \tag{2.6}$$

After quantization the commutator of two fields reads

$$[A_\mu(x), A_\nu(0)] = iD_{\mu\nu}(x), \tag{2.7}$$

where

$$D_{\mu\nu}(x) = D^{\mathrm{tr}}_{\mu\nu}(x) + D^{\mathrm{long}}_{\mu\nu}(x), \tag{2.8}$$

with $D^{\mathrm{tr}}_{\mu\nu}(x)$ and $D^{\mathrm{long}}_{\mu\nu}(x)$ given by

$$D^{\mathrm{tr}}_{\mu\nu}(x) = i\int \frac{d^4k}{(2\pi)^3} e^{-ik\cdot x} \times \epsilon(k_0)\left[\eta_{\mu\nu} - \frac{k_\mu k_\nu}{\mu^2}\right]\delta(k^2-\mu^2) \tag{2.9}$$

and

$$D^{\mathrm{long}}_{\mu\nu}(x) = i\int \frac{d^4k}{(2\pi)^3} e^{-ik\cdot x}\epsilon(k_0)\frac{k_\mu k_\nu}{\mu^2}\delta(k^2-m^2), \tag{2.10}$$

 0022-2488/87/051023-07$02.50

with $m^2=\mu^2/c$. This homogeneous propagator is normalized, such that the quantum conditions are fulfilled, to

$$[\dot{A}_\mu(x),A_\nu(0)]|_{t=0}=i\frac{\partial D_{\mu\nu}(x)}{\partial t}\bigg|_{t=0}$$
$$=i\eta_{\mu\nu}\left(1+\frac{1-c}{c}\eta_{\mu 0}\right)\delta^3(\bar{x}). \tag{2.11}$$

In terms of the conjugate momenta

$$\pi_\mu=\frac{\partial\mathcal{L}}{\partial(\dot{A}^\mu)} \tag{2.12}$$

and fields, we obtain, of course,

$$[A_\mu(x),\pi_\nu(0)]|_{t=0}=i\eta_{\mu\nu}\delta^3(\bar{x}). \tag{2.13}$$

Performing the integrals in (2.9) and (2.10) we obtain, with

$$\lambda=x_\mu x^\mu,\quad \lambda_{-\epsilon}=\lambda-i\epsilon t-\epsilon^2/4,\quad x_0=t-i\epsilon/2, \tag{2.14}$$

$$D_{\mu\nu}(x)=\frac{1}{2\pi^2}\operatorname{Im}\left\{\eta_{\mu\nu}\left[\frac{1}{\lambda_{-\epsilon}}-\frac{1}{\lambda_{-\epsilon}}\left[\frac{J_1(\mu\sqrt{\lambda})}{\mu\sqrt{\lambda}}-\frac{1}{c}\frac{J_1(m\sqrt{\lambda})}{m\sqrt{\lambda}}\right]\right.\right.$$
$$-\frac{\mu J_1(\mu\sqrt{\lambda})}{2\sqrt{\lambda}}\ln(-\lambda_{-\epsilon})+\frac{1}{2\lambda_{-\epsilon}}\left[J_2(\mu\sqrt{\lambda})-\frac{1}{c}J_2(m\sqrt{\lambda})\right]\ln(-\lambda_{-\epsilon})\Big]$$
$$+x_\mu x_\nu\left[\frac{2}{(\lambda_{-\epsilon})^2}\left[\frac{J_1(\mu\sqrt{\lambda})}{\mu\sqrt{\lambda}}+J_2(\mu\sqrt{\lambda})-\frac{1}{c}\frac{J_1(m\sqrt{\lambda})}{m\sqrt{\lambda}}-\frac{1}{c}J_2(m\sqrt{\lambda})\right]\right.$$
$$\left.\left.-\frac{1}{2\lambda_{-\epsilon}}\left[\frac{\mu J_3(\mu\sqrt{\lambda})}{\sqrt{\lambda}}-\frac{m}{c}\frac{J_3(m\sqrt{\lambda})}{\sqrt{\lambda}}\right]\ln(-\lambda_{-\epsilon})\right]\right\}. \tag{2.15}$$

In the massless limit (QED with Gupta–Bleuler quantization), we obtain

$$D^0_{\mu\nu}(x)=\frac{1}{2\pi^2}\operatorname{Im}\left\{\eta_{\mu\nu}\frac{1}{\lambda_{-\epsilon}}+\left(\frac{1-c}{c}\right)\left[\frac{\eta_{\mu\nu}}{2\lambda_{-\epsilon}}-\frac{x_\mu x_\nu}{(\lambda_{-\epsilon})^2}\right]\right\}. \tag{2.16}$$

We see that the gauge fixing choice $c=1$ (Feynman gauge) gives the simplest expressions for (2.15):

$$D_{\mu\nu}(x)=\frac{1}{2\pi^2}\operatorname{Im}\left\{\eta_{\mu\nu}\left[\frac{1}{\lambda_{-\epsilon}}-\frac{\mu J_1(\mu\sqrt{\lambda})}{2\sqrt{\lambda}}\ln(-\lambda_{-\epsilon})\right]\right\}, \tag{2.17}$$

where the last term vanishes for the massless case.

III. MASSIVE VECTOR FIELDS IN ANTI-DE SITTER SPACE

Consider a massive vector field A_μ in a curved spacetime with coordinates x^μ ($\mu=1,\ldots,4$) and metric $g_{\mu\nu}$ with signature $(+,-,-,-)$. The Lagrangian is[14]

$$\mathcal{L}_A=-\tfrac{1}{4}\sqrt{-g}F_{\mu\nu}F^{\mu\nu}+\tfrac{1}{2}\mu^2\sqrt{-g}A_\mu A^\mu, \tag{3.1}$$

where $g=\det g_{\mu\nu}$. When $\mu^2\neq 0$ the field equations are

$$\partial_\mu(\sqrt{-g}F^{\mu\nu})+\mu^2\sqrt{-g}A^\nu=0. \tag{3.2}$$

Taking the divergence, we obtain the generalized Lorentz condition

$$\partial_\mu(\sqrt{-g}A^\mu)=\sqrt{-g}A^\mu{}_{;\mu}=0, \tag{3.3}$$

where the semicolon denotes the covariant derivative:

$$A^\nu{}_{;\mu}=A^\nu{}_{,\mu}+\Gamma^\nu{}_{\rho\mu}A^\rho, \tag{3.4}$$

with $\Gamma^\nu{}_{\rho\mu}$ the affine connection. Now consider a five-dimensional space with coordinates ξ^M ($M=1,\ldots,5$) and metric

$$\eta_{MN}=\operatorname{diag}(-1,-1,-1,1,1). \tag{3.5}$$

Anti-de Sitter space can be visualized as (the covering space of) the hyperboloid

$$\xi_M\xi^M=-\bar{\xi}^2+\xi_4^2+\xi_5^2=R^2=1/\alpha=\text{const}>0. \tag{3.6}$$

The time variable t is introduced by

$$\xi_4=\sqrt{R^2+\bar{\xi}^2}\sin t/R,\quad \xi_5=\sqrt{R^2+\bar{\xi}^2}\cos t/R, \tag{3.7}$$

and is many valued in ξ^M space.

In the following we restrict ourselves to the first sheet only. We introduce a five-dimensional vector field B_M, which satisfies the transversality condition (no component perpendicular to the hyperboloid)

$$\xi^M B_M=0. \tag{3.8}$$

The Lagrangian is given by[15]

$$\mathcal{L}_B=-\tfrac{1}{4}\tilde{G}_{MN}\tilde{G}^{MN}+\tfrac{1}{2}(m+1)(m+2)B_MB^M. \tag{3.9}$$

Here, $\tilde{G}_{MN}$ is defined by

$$\tilde{G}_{MN}=(\mathring{\partial}_M-\alpha\xi_M)B_N-(\mathring{\partial}_N-\alpha\xi_N)B_M, \tag{3.10}$$

with

$$\mathring{\partial}_M=\partial_M-\alpha\xi_M\xi^N\partial_N \tag{3.11}$$

the tangential derivative. The equation of motion which follows from (3.9) is[7]

$$(M^2+2)B_M+\alpha^{-1}\mathring{\partial}_M\mathring{\partial}_NB^N-2\xi_M\mathring{\partial}_NB^N$$
$$=(m+1)(m+2)B_M. \tag{3.12}$$

This is equivalent to (3.2) when

$$B_M = \frac{\partial x^\mu}{\partial \xi^M} A_\mu \tag{3.13}$$

and $g_{\mu\nu}$ is the ADS metric. Here,

$$M^2 = \tfrac{1}{2} M_{MN} M^{MN}, \tag{3.14}$$

with

$$M_{MN} = i(\xi_M \partial_N - \xi_N \partial_M). \tag{3.15}$$

Taking the divergence of Eq. (3.12) we obtain for $m \neq -1, -2$,

$$\mathring{\partial}_M B^M = 0, \tag{3.16}$$

which is equivalent to (3.3). Then Eq. (3.12) reduces to

$$[M^2 - m(m+3)]B_M = 0, \tag{3.17}$$

which is the equation for a scalar field.

Now consider the Lagrangian which is the generalization of Stueckelberg's Lagrangian (2.1) for ADS space:

$$\mathscr{L}_B = -\tfrac{1}{4}\widetilde{G}_{MN}\widetilde{G}^{MN} + \tfrac{1}{2}(m+1)(m+2)B_M B^M - \tfrac{1}{2}c(\mathring{\partial}_M B^M)^2. \tag{3.18}$$

When $m = -1, -2$, we obtain the Lagrangian for ADS QED with a gauge fixing term. The field equation for B_M now becomes

$$[M^2 - m(m+3)]B_M - (c-1)\alpha^{-1}\mathring{\partial}_M \mathring{\partial}_N B^N - 2\xi_M \mathring{\partial}_N B^N = 0. \tag{3.19}$$

Taking the divergence we obtain the scalar equation

$$\{M^2 - [(m+1)(m+2)]/c\}\mathring{\partial}_M B^M = 0, \tag{3.20}$$

which is the analog of (2.3).

IV. CONSTRUCTION OF PROPAGATORS IN ADS SPACE

Take as the reference point ξ_0^M on the hypersurface $\xi_{0M}\xi_0^M = R^2$. The propagator matrix element $G_{MN}(\xi,\xi_0)$ is an invariant function of

$$\lambda = 1 - (\alpha\gamma)^2 \equiv 1 - z^2, \tag{4.1}$$

where

$$\gamma = \xi_0^M \xi_M.$$

The tensor structure is given by two basic transverse (with respect to ξ and ξ_0) tensors. Introduce the transverse projector

$$P_{MN} = \eta_{MN} - \alpha\xi_M\xi_N \tag{4.2}$$

(i.e., $\mathring{\partial}_M = P_{MN}\partial^N$). The two basic tensors are[8]

$$P_{MR}P_0{}^R{}_N = P_M \cdot P_{0N} = \eta_{MN} - \alpha\xi_M\xi_N - \alpha\xi_{0M}\xi_{0N} + \alpha z\xi_M\xi_{0N} \tag{4.3}$$

and

$$\xi^S P_{0SN}\xi_0^T P_{TM} = \xi\cdot P_{0N}\xi_0\cdot P_M = (\xi_N - z\xi_{0N})(\xi_{0M} - z\xi_M). \tag{4.4}$$

Writing

$$G_{MN}(z) = P_M\cdot P_{0N} f(z) + \alpha\xi\cdot P_{0N}\xi_0\cdot P_M g(z) \tag{4.5}$$

and substituting this into the homogeneous equation

$$[M^2 - m(m+3)]G_{MN} - (c-1)\alpha^{-1}\mathring{\partial}_M \mathring{\partial}^P G_{PN} - 2\xi_M \mathring{\partial}^P G_{PN} = 0, \tag{4.6}$$

we obtain for f and g the following coupled equations:

$$[M^2 - (m+1)(m+2) + 4c]f - (1-c)zf' + (-3+5c)zg + (1-c)(1-z^2)g' = 0 \tag{4.7}$$

and

$$[M^2 - (m+1)(m+2) + 3 + 5c]g + (-3+7c)zg' + (1-c)(1-z^2)g'' + (-3+5c)f' - (1-c)zf'' = 0, \tag{4.8}$$

where a prime denotes a derivative with respect to z and

$$M^2 = -(1-z^2)\frac{d^2}{dz^2} + 4z\frac{d}{dz}. \tag{4.9}$$

First consider solutions which are transverse:

$$\mathring{\partial}^M G^{\mathrm{tr}}_{MN} = 0; \tag{4.10}$$

then (4.7) and (4.8) reduce to

$$[M^2 - m(m+3) + 2]f + 2zg = 0, \tag{4.11}$$

$$[M^2 - m(m+3) + 6]g + 4zg' + 2f' = 0, \tag{4.12}$$

with two independent solutions for $z^2 \neq 1$:

$$f_1 = \frac{1}{m+3}\left\{F\left(-\frac{m+2}{2}, \frac{m+3}{2}; \frac{1}{2}; z^2\right) + (m+2)F\left(-\frac{m}{2}, \frac{m+3}{2}; \frac{1}{2}; z^2\right)\right\}, \tag{4.13}$$

$$g_1 = z\left\{(2-m)F\left(-\frac{m+4}{2}, \frac{m+5}{2}; \frac{3}{2}; z^2\right) + (m+2)F\left(-\frac{m+2}{2}, \frac{m+5}{2}; \frac{3}{2}; z^2\right)\right\}, \tag{4.14}$$

$$f_2 = z\left\{(1-m)F\left(-\frac{m+3}{2}, \frac{m+4}{2}; \frac{3}{2}; z^2\right) - m(m+2)F\left(-\frac{m+1}{2}, \frac{m+4}{2}; \frac{3}{2}; z^2\right)\right\}, \tag{4.15}$$

$$g_2 = \left\{(1-m)F\left(-\frac{m+3}{2}, \frac{m+4}{2}; \frac{1}{2}; z^2\right) + (m+2)F\left(-\frac{m+1}{2}, \frac{m+4}{2}; \frac{1}{2}; z^2\right)\right\}, \tag{4.16}$$

where $F(a,b;c;x)$ is a hypergeometric function. For $m = -1, -2$, they reduce for $\lambda \neq 0$ to

$$\begin{cases} f_1 = 1/(1-z^2)^2 = 1/\lambda^2, \\ g_1 = 4z/(1-z^2)^3 = 4\sqrt{1-\lambda}\,/\lambda^3, \end{cases} \tag{4.17}$$

and

$$f_2 = \frac{3z - z^3}{(1-z^2)^2} = \sqrt{1-\lambda}\left[\frac{2}{\lambda^2} + \frac{1}{\lambda}\right], \qquad g_2 = \frac{3 + 6z - z^4}{(1-z^2)^3} = \frac{8 - 4\lambda - \lambda^2}{\lambda^3}, \tag{4.18}$$

which correspond in the flat space limit $\alpha \to 0$ to the pure

gauge propagator

$$-i\frac{k_\mu k_\nu}{k^2}=\lim_{\mu^2\to 0}\mu^2 D^{\rm tr}_{\mu\nu}(k,\mu^2),\quad k^2\neq 0, \tag{4.19}$$

where

$$D^{\rm tr}_{\mu\nu}(k,\mu^2)=i[(\eta_{\mu\nu}-k_\mu k_\nu/\mu^2)/(k^2-\mu^2)]. \tag{4.20}$$

Now consider longitudinal solutions of the form

$$G^{\rm long}_{MN}=\alpha^{-1}\overset{\circ}{\partial}_M\overset{\circ}{\partial}_{0N}G(z)$$
$$=P_M\cdot P_{0N}G'(z)+\alpha\xi\cdot P_{0N}\xi_0\cdot P_M G''(z); \tag{4.21}$$

then $G(z)$ has to satisfy

$$\{M^2-[(m+1)(m+2)]/c\}G(z)=0. \tag{4.22}$$

Taking appropriate linear combinations of the solutions (4.13)–(4.16), we can obtain solutions which converge to zero when $\lambda\to-\infty$. Because m and $-(m+3)$ are interchangeable, we can limit ourselves to $m>-\frac{3}{2}$ and use the solutions (4.25) and (4.26) when we demand that the function approaches zero faster than $(-\lambda)^{-3/4}$. The case $m=-\frac{3}{2}$ will not be considered. For $m<0$ we find, using the same notation (we limit ourselves to the first sheet),

$$f_1=(-\lambda)^{m/2}\left[3(m+3)F\left(-\frac{m}{2},-\frac{m-2}{2};-m-\frac{1}{2};\lambda^{-1}\right)\right.$$
$$\left.+(m-1)F\left(-\frac{m+2}{2},-\frac{m-4}{2};-m-\frac{1}{2};\lambda^{-1}\right)\right], \tag{4.23}$$

$$g_1=-\frac{2}{3}(-\lambda)^{(m-1)/2}\left[((m+1)(m+2)+12)F\left(-\frac{m+1}{2},-\frac{m+3}{2};-m-\frac{1}{2};\lambda^{-1}\right)\right.$$
$$\left.-(m-1)(m-2)F\left(-\frac{m+3}{2},-\frac{m-5}{2};-m-\frac{1}{2};\lambda^{-1}\right)\right], \tag{4.24}$$

and, for $m>-3$,

$$f_2=(-\lambda)^{-(m+3)/2}\left[3mF\left(\frac{m+3}{2},\frac{m+1}{2};m+\frac{5}{2};\lambda^{-1}\right)+(m+4)F\left(\frac{m+5}{2},\frac{m-1}{2};m+\frac{5}{2};\lambda^{-1}\right)\right], \tag{4.25}$$

$$g_2=\frac{2}{3}(-\lambda)^{-(m+4)/2}\left[((m+1)(m+2)+12)F\left(\frac{m+4}{2},\frac{m}{2};-m+\frac{5}{2};\lambda^{-1}\right)\right.$$
$$\left.-(m+4)(m+5)F\left(\frac{m+6}{2},\frac{m-2}{2};m+\frac{5}{2};\lambda^{-1}\right)\right]. \tag{4.26}$$

The longitudinal solutions with the appropriate convergence properties are obtained from

$$G_1=(-\lambda)^{m'/2}F\left(-\frac{m'+2}{2},-\frac{m'}{2};-m'-\frac{1}{2};\lambda^{-1}\right) \tag{4.27}$$

and

$$G_2=(-\lambda)^{-(m'+3)/2}F\left(\frac{m'+1}{2},\frac{m'+3}{2};m'+\frac{5}{2};\lambda^{-1}\right), \tag{4.28}$$

where m' is related to m and c through

$$m'(m'+3)=(m+1)(m+2)/c. \tag{4.29}$$

We show that the transverse solutions are not normalizable in the sense of condition (5.6) because they contain λ^{-2} and λ^{-3} singularities. Expanding the solutions around $\lambda=0$ we find for both solutions,

$$f(m,\lambda)=A\left\{\frac{4}{\lambda^2}-\frac{(m+1)(m+2)}{\lambda}+\frac{3}{8}m(m+1)(m+2)(m+3)F\left(-\frac{m}{2},\frac{m+3}{2};2;\lambda\right)\ln(-\lambda)\right.$$
$$\left.-\frac{1}{192}(m-1)m(m+1)(m+2)(m+3)(m+4)\lambda F\left(-\frac{m+2}{2},\frac{m+5}{2};4;\lambda\right)\ln(-\lambda)+a(m,\lambda)\right\}, \tag{4.30}$$

$$g(m,\lambda)=A\left\{\frac{16}{\lambda^3}+2[(m+1)(m+2)-4]\frac{1}{\lambda^2}+\frac{1}{4}m(m+3)[(m+1)(m+2)+4]\frac{1}{\lambda}\right.$$
$$-\frac{1}{48}[(m+1)(m+2)+12]m(m+1)(m+2)(m+3)F\left(-\frac{m+1}{2},\frac{m+2}{2};3;\lambda\right)\ln(-\lambda)$$
$$\left.+b(m,\lambda)\right\}, \tag{4.31}$$

where A is an arbitrary constant and $a(m,\lambda)$ and $b(m,\lambda)$ are regular functions of λ in the domain $|\lambda|<1$, which are different for the two solutions. But we see that the singularity structure is the same for both solutions. For the longitudinal solutions we find

$$G'(z) = -2\sqrt{1-\lambda}\,\frac{d}{d\lambda}G = D\left\{-\frac{2}{\lambda^2} - \frac{m'(m'+3)}{2\lambda} + \frac{1}{16}m'(m'+1)(m'+2)(m'+3)\right.$$
$$\left.\times\sqrt{1-\lambda}\,F\left(-\frac{m'+2}{2},\frac{m'+5}{2};3;\lambda\right)\ln(-\lambda) + d(m',\lambda)\right\}, \tag{4.32}$$

$$G''(z) = 4(1-\lambda)\frac{d^2}{d\lambda^2}G - 2\frac{dG}{d\lambda} = D\left\{-\frac{8}{\lambda^3} - \frac{(m'-1)(m'+4)}{\lambda^2} - \frac{(m'-1)m'(m'+3)(m'+4)}{8\lambda}\right.$$
$$+\frac{1}{96}(m'-2)m'(m'+1)(m'+2)(m'+3)(m'+5)(1-\lambda)F\left(-\frac{m'+4}{2},\frac{m'+7}{2};4;\lambda\right)$$
$$\left.\times\ln(-\lambda) + \frac{1}{16}m'(m'+1)(m'+2)(m'+3)F\left(-\frac{m'+2}{2},\frac{m'+5}{2};3;\lambda\right)\ln(-\lambda) + e(m',\lambda)\right\}, \tag{4.33}$$

where $d(m',\lambda)$ and $e(m',\lambda)$ are regular functions of λ in the domain $|\lambda| < 1$, which are different for both solutions.

In order to obtain normalizable solutions in the sense of Sec. V we take appropriate combinations of transverse and longitudinal solutions. With arbitrary gauge fixing parameter c, we obtain

$$G_{MN}(\lambda) = F\left\{P_M\cdot P_{0N}\left[-\left(1+\frac{1}{c}\right)\frac{1}{\lambda} + \frac{3}{8}m(m+3)F\left(-\frac{m}{2},\frac{m+3}{2};2;\lambda\right)\ln(-\lambda)\right.\right.$$
$$-\frac{1}{192}(m-1)m(m+3)(m+4)\lambda F\left(-\frac{m+2}{2},\frac{m+5}{2};4;\lambda\right)\ln(-\lambda)$$
$$+\frac{1}{8c}(m'+1)(m'+2)\sqrt{1-\lambda}\,F\left(-\frac{m'+2}{2},\frac{m'+5}{2};4;\lambda\right)\ln(-\lambda)$$
$$\left.+f(m,c,\lambda)\right] + \alpha\xi\cdot P_{0N}\xi_0\cdot P_M\left[\left(1-\frac{1}{c}\right)\frac{2}{\lambda^2} + \frac{1}{4\lambda}2\left(1+\frac{1}{c}\right) + (m+1)(m+2)\left(1-\frac{1}{c^2}\right)\right.$$
$$-\frac{1}{48}m(m+3)((m+1)(m+2)+12)F\left(-\frac{m+1}{2},\frac{m+4}{2};3;\lambda\right)\ln(-\lambda)$$
$$-\frac{1}{2304}(m-2)(m-1)m(m+3)(m+4)(m+5)\lambda F\left(-\frac{m+3}{2},\frac{m+6}{2};5;\lambda\right)$$
$$\times\ln(-\lambda) + \frac{1}{48c}(m'-2)(m'+1)(m'+2)(m'+5)(1-\lambda)F\left(-\frac{m'+4}{2},\frac{m'+7}{2};4;\lambda\right)\ln(-\lambda)$$
$$\left.+\frac{1}{8c}(m'+1)(m'+2)F\left(-\frac{m'+2}{2},\frac{m'+5}{2};3;\lambda\right)\ln(-\lambda) + g(m,c,\lambda)\right]\right\}, \tag{4.34}$$

where $f(m,c,\lambda)$ and $g(m,c,\lambda)$ are regular functions of λ in the domain $|\lambda| < 1$.

To obtain the Feynman propagator $G^F_{MN}(\bar{\xi},t;\bar{\xi}_0,t_0)$ from Eq. (4.34) we make the following replacement (we restrict ourselves to the first sheet, i.e., $n = 0$; see Ref. 10):

$$\lambda \to \lambda - i\epsilon, \tag{4.35}$$

but leave ξ^M the same. In order to normalize this propagator, we want it to satisfy Eq. (4.6), where the right-hand side is replaced by the inhomogeneous term

$$-R^2 P_{0MN}\delta^3(\bar{\xi}-\bar{\xi}_0)\delta(t-t_0). \tag{4.36}$$

For the normalization constant F we find by direct substitution of the solution in the inhomogeneous equation

$$F = -1/8\pi^2 iR^2 \tag{4.37}$$

[cf. Eq. (2.6)]. Here we made use of representations for the δ^4 function such as

$$\lim_{\epsilon\to 0}\frac{1}{\pi^2}\frac{2\epsilon}{(\lambda - i\epsilon)^3} = \delta^4(x^\mu), \tag{4.38}$$

where $\lambda = x_\mu x^\mu$.

V. NORMALIZATION OF THE SOLUTIONS

Consider the Lagrangian (3.1) with a gauge fixing term

$$\mathcal{L}_{GF} = -(c/2)\sqrt{-g}(A^\mu{}_{;\mu})^2. \tag{5.1}$$

This total Lagrangian is equivalent to (3.18). Define the conjugate momenta

$$\pi^\rho = \frac{\partial}{\partial(\partial_0 A_\rho)} = \sqrt{-g}F^{\rho 0} - \sqrt{-g}g^{\rho 0}(A^\mu{}_{;\mu}). \tag{5.2}$$

This prescription is independent of the metric. The quantum conditions are (independent of the metric)

$$[\pi^\mu(\bar{x},t), A_\nu(\bar{y},t)] = -i\delta^\mu{}_\nu\delta^3(\bar{x}-\bar{y}), \tag{5.3}$$

from which we obtain

$$[\pi_\mu(\bar{x},t), A_\nu(\bar{y},t)] = -ig_{\mu\nu}\delta^3(\bar{x}-\bar{y}). \tag{5.4}$$

We consider the case $c = 1$ the simplest case to obtain the normalization. Using (3.13) to express the $A_\mu(\bar{x},t)$ fields in terms of $B_M(\bar{\xi},t)$ we find from (5.4),

$$[\dot{B}_M(\bar{\xi},t), B_N(\bar{\xi}_0,t)] = i(1+\alpha\bar{\xi}_0^2)P_{0MN}\delta^3(\bar{\xi}-\bar{\xi}_0). \tag{5.5}$$

Consider the solution (4.34) and perform the replacement

$$\xi^4 \to \xi^4 \cosh\left(\frac{\epsilon}{2}\right) - i\xi^5 \sinh\left(\frac{\epsilon}{2}\right) \simeq \xi^4\left(1+\frac{\epsilon^2}{8}\right) - i\xi^5\left(\frac{\epsilon}{2}\right),$$

$$\xi^5 \to i\xi^5 \sinh\left(\frac{\epsilon}{2}\right) + \xi^5 \cosh\left(\frac{\epsilon}{2}\right) \simeq i\xi^4\left(\frac{\epsilon}{2}\right) + \xi^5\left(1+\frac{\epsilon^2}{8}\right), \tag{5.6}$$

which results in

$$\lambda \to \lambda_{-\epsilon} \simeq \lambda - i\epsilon\alpha\xi_4\xi_5 - \tfrac{1}{4}\alpha(\xi_5^2 - \xi_4^2)\epsilon^2. \tag{5.7}$$

This corresponds to a complex time translation, analogous to the flat space definition of $\lambda_{-\epsilon}$ in (2.14). Then the real and imaginary parts of this solution correspond to the homogeneous propagators. Identifying the imaginary part with the commutator $-i[B_M(\bar{\xi},t), B_N(\bar{\xi}_0,t_0)]$, we find

$$\frac{\partial}{\partial t} \operatorname{Im} G_{MN}(\bar{\xi},t;\bar{\xi}_0,t_0)\Big|_{t=t_0} = -4\pi^2 F R^2 (1+\alpha\bar{\xi}_0^2) P_{0MN}\delta^3(\bar{\xi}-\bar{\xi}_0) \tag{5.8}$$

and comparing with (5.5) we find for the normalization constant

$$F = -1/4\pi^2 R^2. \tag{5.9}$$

Of course, one can do the same calculations for $c \neq 1$ with the same result, but they are more tedious. When we take the flat space limit of this solution ($\alpha \to 0$) we obtain exactly $D_{\mu\nu}(x-y)$ [Eq. (2.15)]. The Feynman propagator G^F_{MN} can also be obtained from the homogeneous propagators Re G_{MN} and Im G_{MN} in the following way analogous to the scalar case[10]:

$$G^F_{MN} = (1/2i)\{\operatorname{Re} G_{MN} + i\epsilon(t-t_0)\operatorname{Im} G_{MN}\}. \tag{5.10}$$

VI. ANTI-DE SITTER QED[7]

Consider the solutions (4.23)–(4.26). Taking the appropriate combinations of transverse and longitudinal solutions and taking the limit $m \to -1$, we obtain with the proper normalization the following solutions:

$$\begin{aligned} G^1_{MN}(m=-1) = \frac{1}{4\pi^2R^2}\Bigg\{ & P_M\cdot P_{0N}\Bigg[\left(1+\frac{1}{3c}\right)\frac{2}{\lambda} + \left(1-\frac{1}{3c}\right)\Bigg[\frac{2}{\lambda^2}\ln\left(-\frac{\lambda}{4}\right) \\ & - \left(1+\frac{2}{\lambda}\right)\frac{\sqrt{1-\lambda}}{\lambda}\ln\left|\frac{1-\sqrt{1-\lambda}}{1+\sqrt{1-\lambda}}\right|\Bigg]\Bigg] + \alpha\xi\cdot P_{0N}\xi_0\cdot P_M\Bigg[\left(1+\frac{1}{3c}\right)\frac{2\sqrt{1-\lambda}}{\lambda} \\ & + \left(1-\frac{1}{3c}\right)\Bigg[\frac{8\sqrt{1-\lambda}}{\lambda^3}\ln\left(-\frac{\lambda}{4}\right) + \frac{1}{\lambda^2}\left(\lambda+4-\frac{8}{\lambda}\right)\ln\left|\frac{1-\sqrt{1-\lambda}}{1+\sqrt{1-\lambda}}\right|\Bigg]\Bigg]\Bigg\}, \end{aligned} \tag{6.1}$$

$$\begin{aligned} G^2_{MN}(m=-1) = \frac{1}{4\pi^2R^2}\Bigg\{ & P_M\cdot P_{0N}\Bigg[\frac{4}{3c}\frac{\sqrt{1-\lambda}}{\lambda} - \left(1-\frac{1}{3c}\right)\Bigg[\frac{\sqrt{1-\lambda}}{\lambda}\left(1+\frac{2}{\lambda}\right) \\ & \times\ln\left(-\frac{\lambda}{4}\right) - \frac{2}{\lambda^2}\ln\left|\frac{1-\sqrt{1-\lambda}}{1+\sqrt{1-\lambda}}\right|\Bigg]\Bigg] + \alpha\xi\cdot P_{0N}\xi_0\cdot P_M\Bigg[-\frac{6}{\lambda^2} - \frac{2}{3c\lambda} + \frac{10}{3c\lambda^2} \\ & + \left(1-\frac{1}{3c}\right)\Bigg[\frac{1}{\lambda}\left(1+\frac{4}{\lambda}-\frac{8}{\lambda^2}\right)\ln\left(-\frac{\lambda}{4}\right) + \frac{8\sqrt{1-\lambda}}{\lambda^3}\ln\left|\frac{1-\sqrt{1-\lambda}}{1+\sqrt{1-\lambda}}\right|\Bigg]\Bigg]\Bigg\}. \end{aligned} \tag{6.2}$$

Performing the transformation $\lambda \to \lambda_{-\epsilon}$ and taking the imaginary part, we obtain the appropriate homogeneous propagators, which reduce in the flat space limit to $D^0_{\mu\nu}(x-y)$ [Eq. (2.16)].

We see that the logarithmic contributions to the propagators vanish when $c = \frac{1}{3}$, which corresponds to $c = \frac{2}{3}$ of Ref. 8. We obtain in this case, writing the solutions as

$$P_M\cdot P_{0N} d(\lambda) + \alpha^{-1}\mathring{\partial}_M\mathring{\partial}'_N\phi(\lambda) \tag{6.3}$$

that

$$d_1 = \frac{1}{2\pi^2R^2}\frac{1}{\lambda} \text{ and } \phi_1 = -\frac{1}{4\pi^2R^2}\ln\left|\frac{1-\sqrt{1-\lambda}}{1+\sqrt{1-\lambda}}\right|. \tag{6.4}$$

Adding a combination of pure gauge propagators, we obtain expressions without logarithms:

$$d_1 = \frac{1}{2\pi^2R^2}\frac{1}{\lambda}, \quad \phi_1 = -\frac{1}{2\pi^2R^2}\frac{\sqrt{1-\lambda}}{\lambda}. \tag{6.5}$$

Furthermore,

$$d_2 = \frac{1}{2\pi^2R^2}\frac{\sqrt{1-\lambda}}{\lambda}, \quad \phi_2 = -\frac{1}{2\pi^2R^2}\frac{1}{\lambda}, \tag{6.6}$$

which are the same solutions as the ones obtained by Gazeau[8] for $c = \frac{2}{3}$.

ACKNOWLEDGMENTS

We wish to thank Dr. E. van Beveren and Dr. K. Metzger for useful discussions.

Part of this work was included in the research program of the Stichting voor Fundamenteel Onderzoek der Materie (F.O.M.) with financial support from the Nederlandse Organisatie voor Zuiver-Wetenschappelijk Onderzoek (Z.W.O.).

[1] M. J. Duff, B. E. W. Nilsson, and C. N. Pope, Phys. Rep. **130**, 1 (1986), and references therein.

[2] C. Dullemond and E. van Beveren, Phys. Rev. D **28**, 1028 (1983); C. Dullemond, J. Math. Phys. **25**, 2638 (1984); in "New particle produc-

tion," *Proceedings of the XIX Rencontre de Moriond*, edited by J. Tran Thanh Van (Editions Frontières, Gif sur Yvette, 1984), p. 847; E. van Beveren, T. A. Rijken, C. Dullemond, and G. Rupp, in *Proceedings of the Bielefeld Workshop on Resonances*, edited by S. Albeverio, L. S. Ferreira, and L. Streit (Springer, New York, 1984), p. 331; E. van Beveren, T. A. Rijken, K. Metzger, C. Dullemond, G. Rupp, and J. E. Ribeiro, Z. Phys. C **30**, 615 (1986).
[3]C. Fronsdal, Phys. Rev. D **10**, 589 (1974).
[4]C. Fronsdal and R. B. Haugen, Phys. Rev. D **12**, 3810 (1975).
[5]C. Fronsdal, Phys. Rev. D **20**, 848 (1979).
[6]J. Fang and C. Fronsdal, Phys. Rev. D **22**, 1361 (1980).
[7]B. Binegar, C. Fronsdal, and W. Heidenreich, Ann. Phys. (NY) **149**, 254 (1983); J. Math. Phys. **24**, 2828 (1983).
[8]J. P. Gazeau, J. Math. Phys. **26**, 1847 (1985).
[9]B. Allen and T. Jacobson, Commun. Math. Phys. **103**, 669 (1986).
[10]C. Dullemond and E. van Beveren, J. Math. Phys. **26**, 2050 (1985).
[11]H. Janssen and C. Dullemond, J. Math. Phys. **27**, 2786 (1986).
[12]C. Fronsdal, Phys. Rev. D **12**, 3819 (1975); S. J. Avis, C. J. Isham, and D. Storey, *ibid.* **18**, 3565 (1978).
[13]N. N. Bogoliubov and D. V. Shirkov, *Quantum Fields* (Benjamin Cummings, New York, 1983); C. Itzykson and J.-B. Zuber, *Quantum Field Theory* (McGraw-Hill, New York, 1980).
[14]N. D. Birrell and P. C. W. Davies, *Quantum Fields in Curved Space* (Cambridge U.P., London, 1982).
[15]E. van Beveren, T. A. Rijken, and C. Dullemond, J. Math. Phys. **27**, 1411 (1986).

A limit theorem for basic states of disordered structures

Andrzej Korzeniowski
Department of Mathematics, University of Texas, Arlington, Texas 76019

(Received 14 October 1985; accepted for publication 14 January 1987)

Let $a(t,\omega)$ be a stationary process such that $0<E[1/a(t,\omega)]<\infty$. It is shown that the random boundary-value problem $Hy=-(d/dt)a(t,\omega)(dy/dt)=\lambda y$, $y(0)=y'(L)=0$, has a unique solution $(\lambda_i(\omega,L), y_i(t,\omega,L))$ for $i\geqslant 0$ and $\lambda_i(\omega,L)/\lambda_{oi}(L)\to 1$ almost surely as $L\to\infty$, where $\lambda_{oi}(L)$ is the ith eigenvalue of the averaged Hamiltonian $H_o y=-[1/E(1/a(t,\omega))](d^2y/dt^2)=\lambda y$, $y(0)=y'(L)=0$.

I. INTRODUCTION

Motivation for this paper comes from an article by Grenkova, Molcanov, and Sudarev[1] where, among other things, a one-dimensional random Schrödinger operator $H=-(d/dt)a(t,\omega)(d/dt)$ describing the quantum-mechanical behavior of an electron in a random medium is discussed. The result stated in Theorem 1 of Ref. 1, however, is not mathematically rigorous. Namely, the arguments used in the proof do not show that the random boundary-value problem does have a solution on $[0,L]$. It has been proved that for sufficiently large $L=L(\omega)$, there is a unique solution except on a set of arbitrarily small probability and therefore neither the solution nor the corresponding eigenvalue is a stochastic process and a random variable, respectively, as being not defined on the whole probability space. Consequently, correlation between energy levels cannot be defined. Moreover, all estimates on pp. 108 and 109 (like I.16–I.18) of Ref. 1 given for a fixed ω from a set $A_L(\epsilon)$ such that $\mathcal{P}(A_L(\epsilon))<\epsilon$ cannot be controlled on $\Omega-A_L(\epsilon)$ for fixed L which is crucial in stating that considered "random variables" are asymptotically Gaussian. Another thing that needs clarification is the following. What is meant by a solution of the random boundary-value problem as well as measurability of the corresponding eigenvalues and eigenfunctions?

Our aim here is to answer these questions as well as to prove a new result, i.e., almost surely convergence of the random eigenvalues as opposed to convergence in distributions of Ref. 1.

II. RESULT

Following Ref. 1, the random boundary-value problem is equivalent to

$$Hy=-\frac{d}{dx}a(xL,\omega)\frac{d}{dx}y=\mu y,\quad x\in[0,1],\qquad y(0)=y'(1)=0, \tag{1}$$

where $\mu_i(\omega,L)=\lambda_i(\omega,L)L^2$, $i\geqslant 0$.

At this point we explain the meaning of the solution of (1). Note that for most processes the trajectories $a(xL,\omega)$ for fixed ω are not differentiable and thus the first attempt is to assume that $a(xL,\omega)$ is absolutely continuous. Unfortunately for many process it does not work (e.g., for pure jump processes $(d/dx)(xL,\omega)=0$ for almost all $x\in[0,1]$). Therefore we come to the following.

Definition: A pair $(\mu(\omega,L),y(x,\omega,L))$ is called a solution of (1) if the following conditions are satisfied: (i) $\mu(\omega,L)$ is a random variable; (ii) $y(x,\omega,L)$ is a stochastic process whose trajectories are absolutely continuous; (iii) $(\mu(\omega,L),y(x,\omega,L))$ solves the integral equation

$$\begin{aligned}y(x,\omega,L)&=\mu(\omega,L)\int_0^1 G(x,z,\omega,L)y(z,\omega,L)dz\\&=\mu(\omega,L)\left[\int_0^x y(z,\omega,L)dz\int_0^z\frac{du}{a(uL,\omega)}\right.\\&\quad\left.+\int_0^x\frac{du}{a(uL,\omega)}\int_x^1 y(z,\omega,L)dz\right],\end{aligned}\tag{2}$$

where

$$G(x,z,\omega,L)=\begin{cases}\displaystyle\int_0^x\frac{du}{a(uL,\omega)}, & x\leqslant z,\\ \displaystyle\int_0^z\frac{du}{a(uL,\omega)}, & x>z,\end{cases}$$

is a random Green's function.

One assumes that all equalities as well as properties of considered processes hold almost surely. It is also assumed that for a given ω for almost all $u\in[0,1]$, $a(uL,\omega)\neq 0$ and $1/a(uL,\omega)$ is integrable on $[0,1]$. Obviously y in (2) is absolutely continuous and $y'=(\mu/a)\int_x^1 y$ for almost all $x\in[0,1]$. Note that $y(0)=0$; however, to guarantee the second boundary condition $y'(1)=0$ an additional assumption on the process $a(t,\omega)$ must be imposed.

It is easy to see that $a(t,\omega)\in D[0,\infty)$—class of processes which are right continuous and have left limits—is sufficient. Combining the above we assume that $a(t,\omega)$ is a stationary process in $D[0,\infty)$ such that $0<\alpha=E[1/a(0,\omega)]<\infty$. Note that by the Fubini theorem it implies $\int_0^t[1/a(u,\omega)]du<\infty$ for $t\in[0,\infty)$. In addition by the ergodic theorem

$$\sup_{0\leqslant x\leqslant 1}\left|\int_0^x\frac{du}{a(uL,\omega)}\right|\leqslant\sup_{0\leqslant t\leqslant 1}t\frac{1}{Lt}\int_0^{Lt}\left|\frac{du}{a(u,\omega)}\right|<\infty,\ \text{a.s.}$$

(almost surely) and therefore the operator

$$\mathcal{G}=\mathcal{G}(\omega,L):\ C[0,1]\to C[0,1],$$

$$\mathcal{G}y(x)=\int_0^1 G(x,z,\omega,L)y(z)dz$$

is compact, having a bounded kernel G. Consequently there is a sequence $(\mu_i(\omega,L), y_i(x,\omega,L))$, $i\geqslant 0$ of the reciprocal of

 0022-2488/87/051030-02$02.50

eigenvalues and eigenfunctions of $\mathcal{G}$ solving (2). The only thing to prove is measurability of $\mu_i(\omega)$ and $y_i(x,\omega)$ for fixed $x\in[0,1]$.

To this end note that

$$\mu_i^{-1}=\begin{cases}\sup\limits_{\|u\|_{C_i[0,1]}=1}(\mathcal{G}u,u), & \sup\limits_{\|u\|_{C_i[0,1]}=1}(\mathcal{G}u,u)\geqslant -\inf\limits_{\|u\|_{C_i[0,1]}=1}(\mathcal{G}u,u),\\ -\inf\limits_{\|u\|_{C_i[0,1]}=1}(\mathcal{G}u,u), & \text{otherwise,}\end{cases} \tag{3}$$

where $C_i[0,1]=\{y_0,\ldots,y_{i-1}\}^{\perp}$—the orthogonal complement in $C[0,1]$ with the norm induced by $(u,v)=\int_0^1 u(x)v(x)dx$ and $y_{-1}\equiv 0$.

The above implies measurability of μ_i, because $(\mathcal{G}u,u)$: $\Omega\to R$ is measurable (the kernel G of $\mathcal{G}$ is measurable). To show measurability of y_i we will approximate $\mathcal{G}$ by step operators $\mathcal{G}^n$ as follows. Take the kernel $G^n(\cdot,\cdot,\omega,L)$ such that

$$\|G^n(\cdot,\cdot,\omega,L)-G(\cdot,\cdot,\omega,L)\|\underset{C[0,1]\times[0,1]}{\to}0 \quad \text{as} \quad n\to\infty \tag{4}$$

and $G^n(\cdot,\cdot,\omega,L)\in C[0,1]^2$ does not change for $\omega\in\Omega_k$, $k=1,\ldots,n$, where Ω_k form a disjoint partition of Ω.

Pick an arbitrary $\omega_k\in\Omega_k$, $k=1,\ldots,n$, find $(\mu_i^n(\omega_k,L),y_i^n(x,\omega_k,L))$ such that

$$y_i^n=\mu_i^n\int_0^1 G^n y_i^n$$

and set

$$y_i^n(x,\omega,L)=\sum_{k=1}^{n}\mu_i^n(\omega_k,L)y_i^n(x,\omega_k,L)\mathbb{1}_{\Omega_k}(\omega)\,.$$

Then y_i^n satisfies

$$y_i^n=\mu_i^n\int_0^1 G^n y_i^n\,.$$

Since

$$\sup_n\sup_{(x,z)}|G^n(x,z,\omega,L)|<\infty,\ \text{a.s.},$$

therefore $\{y_i^n\}$ is relatively compact in $C[0,1]$ and thus has subsequence converging to $\tilde{y}_i(\cdot,\omega,L)$.

Now for fixed ω

$$\|G^n y_i^n-G\tilde{y}_i\|\leqslant\|G^n(y_i^n-\tilde{y}_i)\|+\|(G^n-G)\tilde{y}_i\|\to 0$$

and by (3) and (4), $\mu_i^n\to\mu_i$.

Consequently, $G^n y_i^n\to G\tilde{y}_i$ and

$$\tilde{y}_i=\mu_i\int_0^1 G\tilde{y}_i\,.$$

Next by uniqueness $\tilde{y}_i=y_i$, whence the sequence $\{y_i^n\}$ itself is convergent and thus y_i is measurable.

To complete our analysis consider the following[1]:

$$\mathcal{G}y=\mu(\mathcal{G}_0+\mathcal{G}_1)y=y\,, \tag{5}$$

where

$$\mathcal{G}_0 y=\int_0^1 G_0(x,z)y(z)dz,\quad G_0=\begin{cases}x\alpha, & x\leqslant z,\\ z\alpha, & x>z,\end{cases}$$

$$\mathcal{G}_1 y=\int_0^1 G_1(x,z,\omega,L)y(z)dz,$$

$$G_1=\begin{cases}\int_0^x\dfrac{du}{a(uL,\omega)}-x\alpha, & x\leqslant z,\\ \int_0^z\dfrac{du}{a(uL,\omega)}-z\alpha, & x>z,\end{cases}$$

with G_0 the Green's function of the averaged Hamiltonian

$$H_0 y=-\frac{1}{\alpha}\frac{d^2y}{dx^2}=\mu_0 y,\quad y(0)=y'(1)=0 \tag{6}$$

and G_1 the corresponding random counterpart which by the ergodic theorem converges to 0 almost surely in $L^2[0,1]^2$.

Denoting the solution of (6) by (μ_{oi},y_{oi}) and using (5) one has

$$(y_i,y_{oi})=\mu_i(\mathcal{G}_0 y_i,y_{oi})+\mu_i(\mathcal{G}_1 y_i,y_{oi}),$$

$$\|y_{oi}\|=\|y_i\|=1\,,$$

whence by self-adjointness of $\mathcal{G}_0$

$$\frac{\mu_i}{\mu_{oi}}=\frac{(y_i,y_{oi})}{(y_i,y_{oi})+\mu_{oi}(\mathcal{G}_1 y_i,y_{oi})}\to 1,\quad \text{a.s.},$$

since $|(\mathcal{G}_1 y_i,y_{oi})|\leqslant\|\mathcal{G}_1\|\to 0$ and y_i tends to y_{oi} in $L^2[0,1]$ as $L\to\infty$. Equivalently, applying (1) one obtains

$$\lambda_i(\omega,L)/y_{oi}(L)\to 1,\ \text{a.s.}\,,$$

as claimed.

Remark: To make theorem 1 of Ref. 1 work, besides measurability, one needs to show two things: (I) there is a unique solution on $[0,L]$, a.s., (II) $\mu_i(x,L)=\mu_{oi}+b_i(\omega,L)$, where $b_i(\omega,L)$ is asymptotically Gaussian for large L.

To this end apply Skorokhod's result (cf. Ref. 2, p. 281) to $\zeta_L(t)\to$ weakly Wiener process, i.e., replace it by $\eta_L(t)\overset{\text{a.s.}}{\to}$ Wiener process. Then use existence arguments similar to ours to ensure (I). Next repeating the analysis of Ref. 1 (pp. 107 and 108) we have

$$\mu_i(\omega,L)=\mu_{oi}-\left(\frac{\mu_{oi}^2}{\sqrt{L}}(\mathcal{G}_1(\omega,L)y_{oi},y_{oi})+r_{i,L}(\omega)\right)\times\mathbb{1}_{A_L}(\omega)+\mu_i(\omega,L)\mathbb{1}_{A_L^c}(\omega),$$

where $\mu_i(\omega,L)$ is the eigenvalue found in (I), $\sup_{\omega\in A_L}|r_{i,L}(\omega)|\leqslant c(i,\epsilon)/L$ and $\mathcal{P}(A_L^c)\leqslant\epsilon$ which proves (II) because $(\mathcal{G}_1(\omega,L)y_{oi},y_{oi})$ is asymptotically Gaussian for large L.

[1]L. N. Grenkova, S. A. Molcanov, and Yu. N. Sudarev, "On the basic states of one-dimensional disordered structures," Commun. Math. Phys. **90**, 101 (1983).

[2]A. V. Skorokhod, "Limit theorems for stochastic processes," in Theor. Probab. Appl. **1** (3), 261 (1956).

Hypervirial theorem and parameter differentiation: Closed formulation for harmonic oscillator integrals

J. Morales, J. Lopez-Bonilla, and A. Palma
Instituto Mexicano Del Petroleo, Investigacion Basica de Procesos, Apartado Postal 14-805 07730 Mexico, D.F. Mexico

(Received 19 August 1986; accepted for publication 7 January 1987)

A simple method has been developed to generate a closed formula for the calculation of matrix elements of arbitrary functions $f(x)$ in the representation of the harmonic oscillator. The proposed algebraic procedure is based on the combined use of the hypervirial theorem with and without the second quantization formalism along with the parameter differentiation technique. The closed formula thus obtained is given in terms of a sum involving the jth derivative of $f(x)$ evaluated at zero.

I. INTRODUCTION

Recently, an algebraic procedure based on the hypervirial theorem[1] and second quantization formalism has been developed to derive generalized recurrence relations for the calculation of arbitrary function integrals in the one-dimensional harmonic oscillator (HO) representation.[2,3] Despite the fact that recursion formulas allow us to obtain matrix elements, a closed-form expression is always desirable. In this respect, Wilcox[4] has shown, by an analytical procedure, that the one-center m,n matrix element of an arbitrary function $f(x)$ in the HO representation, is given by a closed formula in terms of an integral that involves the Fourier transform of $f(x)$; the analytical procedure to evaluate such integral is, however, sometimes long and cumbersome since j integrations by parts are necessary. Alternatively, Palma *et al.*[5] have shown, by an algebraic procedure, that the combined use of Cauchy's integral and the Baker–Campbell–Hausdorff theorem allows us to obtain a closed-form equation for the calculation of $f(x)$ HO integrals. On the other hand, heretofore, one of the main uses of hypervirial theorems has been as a means of deriving relationships between quantum mechanical matrix elements.[6] As far as we know, the power of the hypervirial methods has not been exploited to obtain closed-form expressions for the calculation of matrix elements. With this purpose, in the present work we point out a procedure that proposes the use of the hypervirial theorem, with and without the second quantization formalism, along with the parameter differentiation technique[7,8] in order to obtain closed formulas for the evaluation of HO integrals. The present scope is restricted to the one-center matrix elements case, but the proposed method can be extended to the calculation of two-center HO integrals as well as other potentials.

To schematize this approach, the next section is devoted to the determination of closed formulas for matrix elements of exponential operators. The results thus obtained are used in the subsequent paragraphs in order to derive the corresponding closed formulation of power and Gaussian function integrals. A closed-form expression for integrals of arbitrary functions is then obtained from the above results.

II. EXPONENTIAL INTEGRALS

Consider a one-dimensional problem for which the eigenfunction satisfies the time-independent Schrödinger equation

$$H|n\rangle = \left(-\alpha\frac{d^2}{dx^2} + V(x)\right)|n\rangle = E_n|n\rangle\,, \tag{2.1}$$

where $\alpha = \hbar^2/2\mu$ and the eigenenergies E_n, mass μ, and potential constants implicit in $V(x)$ are assumed to be known. In the most general case of a $f(x)$ function such that

$$[V(x), f(x)] = 0\,, \tag{2.2}$$

the exact generalized recurrence relation for the calculation of $f(x)$ matrix elements as a function of eigenenergies for any $V(x)$ is given, from the second hypervirial theorem, by[2,9,10]

$$\begin{aligned}(E_m - E_n)^2\langle m|f(x)|n\rangle &= -\alpha^2\langle m|\frac{d^4f(x)}{dx^4}|n\rangle - 2\alpha(E_m + E_n)\\ &\times\langle m|\frac{d^2f(x)}{dx^2}|n\rangle + 4\alpha\langle m|\frac{d^2f(x)}{dx^2}V(x)|n\rangle\\ &+ 2\alpha\langle m|\frac{df(x)}{dx}\frac{dV(x)}{dx}|n\rangle\,.\end{aligned} \tag{2.3}$$

In the particular case of $f(x) = \exp(-\beta x)$ and HO potential, the corresponding recurrence relation becomes

$$\begin{aligned}&((E_m - E_n)^2 + \alpha^2\beta^4 + 2\alpha\beta^2(E_m + E_n))\\ &\times\langle m|\exp(-\beta x)|n\rangle\\ &= 2K\alpha\beta^2\langle m|x^2\exp(-\beta x)|n\rangle\\ &- 2K\alpha\beta\langle m|x\exp(-\beta x)|n\rangle\,,\end{aligned} \tag{2.4}$$

where K is the force constant. Although it seems at first glance that the above recursion equation cannot be put to practical use, we will see its usefulness at once. To obtain closed formulas for $\exp(-\beta x)$ matrix elements, we assume that the HO eigenfunctions are independent of the β parameter. Then, the above recursion relation is transformed by parameter differentiation[8] into

 0022-2488/87/051032-04$02.50

$$\frac{d^2y(\beta)_{m,n}}{d\beta^2}+\frac{1}{\beta}\frac{dy(\beta)_{m,n}}{d\beta}$$
$$-\frac{1}{4\beta^2}(4A+4T\beta^2+\beta^4)y(\beta)_{m,n}=0\,, \tag{2.5}$$

where

$$y(\beta)_{m,n}=\langle m|\exp(-\beta x)|n\rangle\,,$$

$A=(m-n)^2$, and $T=m+n+1$; we use natural units $\hbar=\mu=\omega=1$.

At this point, in order to clarify the proposed method, instead of solving Eq. (2.5) directly, we shall consider some useful particular cases.

A. $m=n=0$ generator matrix element

Straightforwardly, the corresponding differential equation is given by

$$\frac{d^2y(\beta)_{0,0}}{d\beta^2}+\frac{1}{\beta}\frac{dy(\beta)_{0,0}}{d\beta}-\frac{1}{4}(\beta^2+4)y(\beta)_{0,0}=0 \tag{2.6}$$

with solution

$$y(\beta)_{0,0}=C_{0,0}\exp(\beta^2/4)=\exp(\beta^2/4)\,. \tag{2.7}$$

B. Diagonal matrix elements

Similarly to the previous case, the differential equation

$$z\frac{d^2f(\beta)}{dz^2}+(1-z)\frac{df(\beta)}{dz}+nf(\beta)=0\,, \tag{2.8}$$

where the independent variable is $z=-\beta^2/2$ and

$$y(\beta)_{m,m}=f(\beta)\exp(\beta^2/4)\,,$$

has as its solution

$$y(\beta)_{m,m}=C_{m,m}\exp(\beta^2/4)L_m(-\beta^2/2)$$
$$=\exp(\beta^2/4)L_m(-\beta^2/2)\,. \tag{2.9}$$

It is important to notice that $y(\beta)_{0,0}$ as well as $y(\beta)_{m,m}$ have been obtained without the explicit use of wave functions; the coefficient $C_{0,0}=C_{m,m}=1$ comes from the initial condition $\beta=0$ and the orthogonality requirement

$$\langle m|n\rangle=\delta_{m,n}\,. \tag{2.10}$$

C. Off-diagonal matrix elements

The (m,n)th matrix elements of the operator $\exp(-\beta x)$ come from

$$\frac{d^2Q(z)}{dz^2}+\left(\frac{m+n+1}{2z}+\frac{1-r^2}{4z^2}-\frac{1}{4}\right)Q(z)=0\,, \tag{2.11}$$

where $r^2=A$ and

$$f(z)=Q(z)P(z)\,,$$

where

$$P(z)=(-z)^{-1/2}\exp(z/2),\quad z\leqslant 0\,.$$

Thus, in the case $n>m$,

$$r=n-m$$

and Eq. (2.11) transforms to

$$\frac{d^2Q(z)}{dz^2}+\left(\frac{2m+r+1}{2z}+\frac{1-r^2}{4z^2}-\frac{1}{4}\right)Q(z)=0\,, \tag{2.12}$$

with solution

$$Q(z)=C_{m,n}\exp(-z/2)z^{(r+1)/2}L_m^r(z)\,, \tag{2.13}$$

i.e., for $n>m$,

$$y(\beta)_{m,n}=C_{m,n}(-\beta/\sqrt{2})^{n-m}$$
$$\times\exp(\beta^2/4)L_m^{n-m}(-\beta^2/2) \tag{2.14}$$

and similarly for $m>n$,

$$y(\beta)_{m,n}=C_{m,n}(-\beta/\sqrt{2})^{m-n}$$
$$\times\exp(\beta^2/4)L_n^{m-n}(-\beta^2/2)\,. \tag{2.15}$$

Here, unfortunately the $C_{m,n}$ coefficient cannot be determined from $\beta=0$ and Eq. (2.10) as before. However, Morales *et al.*[2] have proposed a simple method, based on the hypervirial theorem and second quantization formalism, for the determination of proper recurrence relations for matrix elements in the HO representation. Thus from their corresponding recurrence relation

$$(m-n)y(\beta)_{m,n}$$
$$=\beta\sqrt{n/2}y(\beta)_{m,n-1}-\beta\sqrt{m/2}y(\beta)_{m-1,n}\,, \tag{2.16}$$

along with the recursion formulas for the generalized Laguerre polynomial[11] for $m>n$,

$$zL_{n-1}^{m-n+1}=mL_{n-1}^{m-n}-nL_n^{m-n} \tag{2.17}$$

and

$$L_n^{m-n-1}=L_n^{m-n}-L_{n-1}^{m-n}\,, \tag{2.18}$$

one gets

$$(m\sqrt{n}C_{m,n-1}-\sqrt{m}C_{m-1,n})L_{n-1}^{m-n}$$
$$=\{(m-n)C_{m,n}+n\sqrt{n}C_{m,n-1}$$
$$-\sqrt{m}C_{m-1,n}\}L_n^{m-n}\,. \tag{2.19}$$

Finally, by using the fact that L_{n-1}^{m-n} and L_n^{m-n} are linearly independent,

$$m\sqrt{n}C_{m,n-1}-\sqrt{m}C_{m-1,n}=0\,,$$
$$(m-n)C_{m,n}+n\sqrt{n}C_{m,n-1}-\sqrt{m}C_{m-1,n}=0\,,$$

it follows immediately that

$$C_{m,n}=\sqrt{n!/m!},\quad m>n\,, \tag{2.20}$$

and similarly

$$C_{m,n}=\sqrt{m!/n!},\quad n>m\,. \tag{2.21}$$

The closed formulas for the evaluation of the $y(\beta)_{m,n}$ matrix elements are given by Eq. (2.14) and Eq. (2.15) with the $C_{m,n}$ corresponding coefficients specified in Eq. (2.20) and Eq. (2.21).

III. x^k AND GAUSSIAN INTEGRALS

In this section the results obtained in the precedent paragraph are used to obtain closed formulas for the evaluation of power and Gaussian matrix elements as well as some particularly useful recurrence relations.

It is enough to consider $n>m$ since m,n represent exchangeable wave functions. Then, in the first case, from the identity

$$y(\beta)_{m,n}=\sum_k \frac{(-\beta)^k}{k!}x^k_{m,n}\,, \tag{3.1}$$

where

$$x^k_{m,n}=\langle m|x^k|n\rangle$$

along with the condition $n-m=2\lambda$ in Eq. (2.14), one gets

$$x^{2l}_{m,n}=\begin{cases}\sqrt{m!n!}\displaystyle\sum_{r=0}^{[m,l-\lambda]}\frac{(2l)!}{2^{2l-r-\lambda}(l-\lambda-r)!(m-r)!(2\lambda+r)!r!}, & l\geqslant\lambda, \quad (3.2a)\\ 0, & l<\lambda, \quad (3.2b)\end{cases}$$

where $[m,\rho]$ denotes the smaller of m and ρ. Similarly, the case of $n-m=2\lambda+1$ is given by

$$x^{2l+1}_{m,n}=\begin{cases}\sqrt{m!n!}\displaystyle\sum_{r=0}^{[m,l-\lambda]}\frac{(2l+1)!}{2^{2l-\lambda-r+1/2}(l-\lambda-r)!(m-r)!(2\lambda+1+r)!r!}, & l>\lambda, \quad (3.3a)\\ 0, & l<\lambda. \quad (3.3b)\end{cases}$$

Equations (3.2b) and (3.3b) give the well-known selection rule[4] that $\langle m|x^j|n\rangle$ vanishes unless $j\geqslant n-m$. Furthermore, all the possible cases to be considered are covered by means of Eqs. (3.2) and (3.3) along with

$$X^r_{m,n}=\begin{cases}0, & r=2l+1 \text{ and } (n-m) \text{ even}, \quad (3.4a)\\ 0, & r=2l \text{ and } (n-m) \text{ odd}. \quad (3.4b)\end{cases}$$

Additionally to the above equations, some useful recurrence relations for practical calculations are

$$X^{2l+2}_{0,n}=\frac{(l+1)(2l+1)}{2(l+1-\lambda)}X^{2l}_{0,n},\quad n=2\lambda,\quad l\geqslant\lambda \tag{3.5}$$

and

$$X^{2l+3}_{0,n}=\frac{(2l+3)(l+1)}{2(l+1-\lambda)}X^{2l+1}_{0,n},\quad n=2\lambda+1,\quad l\geqslant\lambda\,, \tag{3.6}$$

where $X^{2l+1}_{0,n}=0$ for $n=2\lambda+1$, $l<\lambda$, and $X^{2l}_{0,n}=0$ for $n=2\lambda$, $l<\lambda$. In a similar way, the previous results are used to obtain the corresponding closed formula for the calculation of Gaussian-type integrals; i.e., from the definition of the Gaussian function, it is directly recognized that

$$\langle m|\exp(-\beta x^2)|n\rangle=0 \quad\Leftrightarrow\quad n-m=\text{odd}\,.$$

Consequently, when $n-m=2\lambda$ it leads to

$$\langle m|\exp(-\beta x^2)|n\rangle=\sum_{j=\lambda}^{\infty}\frac{(-\beta)^j}{j!}\sqrt{m!n!}\sum_{r=0}^{[m,j-\lambda]}\frac{(2j)!}{2^{2j-r-\lambda}(j-\lambda-r)!}\times\frac{1}{(m-r)!(2\lambda+r)!r!}\,. \tag{3.7}$$

This equation can be simplified by using the identity

$$\sum_{j=\lambda}^{\infty}=\sum_{j=\lambda}^{\lambda+m}+\sum_{j=\lambda+m+1}^{\infty} \tag{3.8}$$

to

$$\langle m|f(x)|n\rangle=\sqrt{m!n!}\sum_{j=0}^{\infty}\sum_{r=0}^{[m,j]}\frac{f^{(2j+2\lambda)}(0)}{2^{(2j+\lambda-r)}(j-r)!(m-r)!(2\lambda+r)!r!}, \tag{4.1}$$

as well as for $n-m=2\lambda+1$

$$\langle m|f(x)|n\rangle=\sqrt{m!n!}\sum_{j=0}^{\infty}\sum_{r=0}^{[m,j]}\frac{f^{(2j+2\lambda+1)}(0)}{2^{2j+\lambda-r+1/2}(j-r)!(m-r)!(n-m+r)!r!}, \tag{4.2}$$

where

$$\langle m|\exp(-\beta x^2)|n\rangle=2^{\lambda}\sqrt{m!n!}\sum_{r=0}^{m}\frac{2^r}{(m-r)!(2\lambda+r)!r!}\times\sum_{j=r+\lambda}^{\infty}\frac{(-\beta)^j(2j)!}{2^{2j}j!(j-\lambda-r)!}\,. \tag{3.9}$$

Finally, the identification of the last sum in the above equation transforms it to

$$\langle m|\exp(-\beta x^2)|n\rangle=\left(\frac{m!n!}{1+\beta}\right)^{1/2}\sum_{r=0}^{m}\left(-\frac{\beta}{2(1+\beta)}\right)^{(n-m+2r)/2}\times\frac{(n-m+2r)!}{(m-r)!r![(n-m+r)!](r+(n-m)/2)!}\,. \tag{3.10}$$

Some useful particular cases are the generator[12]

$$\langle 0|\exp(-\beta x^2)|0\rangle=(1+\beta)^{-1/2}, \tag{3.11}$$

the up/down[12]

$$\langle 0|\exp(-\beta x^2)|n\rangle=\frac{\sqrt{n!}(-\beta/2)^{n/2}}{(1+\beta)^{(n+1)/2}(n/2)!}, \tag{3.12}$$

and the diagonal

$$\langle n|\exp(-\beta x^2)|n\rangle=\frac{n!}{\sqrt{1+\beta}}\sum_{r=0}^{n}\left(-\frac{\beta}{2(1+\beta)}\right)^r\frac{(2r)!}{(r!)^3(n-r)!} \tag{3.13}$$

matrix elements.

IV. INTEGRALS OF ARBITRARY FUNCTIONS

The above results can be used in order to obtain a generalized closed formula for the calculation of integrals of arbitrary functions within the HO representation. In fact, in the case of a $f(x)$ function, such that it can be expanded in a Taylor series, it is immediate to show for $n-m=2\lambda$ that

Similar to the Gaussian integrals case, the use of the identity specified in Eq. (3.8) in the above two equations leads independently in the general case of any m and n, to

$$\langle m|f(x)|n\rangle = \left(\frac{m!n!}{2^{n-m}}\right)^{1/2} \sum_{r=0}^{m} \sum_{j=0}^{\infty} \frac{f^{(2j+n-m-2r)}(0)}{2^{2j+r}(m-r)!(n-m+r)!r!j!}, \quad n \geqslant m . \tag{4.3}$$

This equation is an exact closed-form expression for the calculation of $f(x)$ matrix elements in the HO representation. Its application to a given $f(x)$ needs exclusively the identification of the last sum over the $(2j+n-m-2r)$th derivative of the function $f(x)$ evaluated at zero. We want to point out, however, that Eq. (4.3) has also been derived recently by Palma *et al.*[5] by means of an alternative algebraic method that proposes the combined use of the Cauchy's integral formula and the Baker–Campbell–Hausdorff theorem.

V. DISCUSSION

As far as we know, in the present work the hypervirial theorem has been used for the first time, not as a means, as usually, of deriving relationships between quantum mechanical integrals, but as a useful media to obtain closed-form expressions for the calculation of matrix elements. This has been made possible with the help of the second quantization formalism along with the parameter differentiation technique. By using the HO representation, as an example, the proposed method has permitted us to rederive the well-known exact closed formulas for the calculation of exponential, power, and Gaussian functions integrals. Additionally, a generalized closed formula for integrals of arbitrary functions has been also obtained. Our proposed exact general closed formula [Eq. (4.3)] seems to be easier to handle than the one given by Wilcox.[4] Finally we want to add that the algebraic procedure shown here can be extended to the determination of the corresponding mathematical formulas for the calculation of two-center HO integrals as well as other potential matrix elements.

[1] J. O. Hirschfelder, J. Chem. Phys. **33**, 1462 (1960).
[2] J. Morales, A. Palma, and L. Sandoval, Int. J. Quantum Chem. **29**, 211 (1986).
[3] J. Morales, L. Sandoval, and A. Palma, J. Math. Phys. **27**, 2966 (1986).
[4] R. M. Wilcox, J. Chem. Phys. **45**, 3312 (1966).
[5] A. Palma, L. Sandoval, and J. Morales, to be published.
[6] See, for example J. H. Epstein and S. T. Epstein, Am. J. Phys. **30**, 266 (1962); P. D. Robinson, J. Chem. Phys. **47**, 2319 (1967); R. J. Swenson, and S. H. Danforth, *ibid.* **57**, 1734 (1972); J. C. Nash, J. Phys. B **6**, 393 (1973); R. H. Tipping, J. Mol. Spectrosc. **59**, 8 (1976); *ibid.* **73**, 400 (1978); A. Requena, P. Pena, and A. Serna, Int. J. Quantum Chem. **17**, 931 (1980); R. H. Tipping and J. F. Ogilvie, Phys. Rev. A **27**, 95 (1983); F. M. Fernandez and E. A. Castro, Int. J. Quantum Chem. **23**, 915 (1983); A. Requena, P. Pena, and J. Zuniga, J. Chem. Phys. **78**, 4792 (1983), and references therein.
[7] K. Aizu, J. Math. Phys. **4**, 762 (1963).
[8] R. M. Wilcox, J. Math. Phys. **8**, 962 (1967).
[9] This equation corresponds in the case $R_{ij} = 0$ to Eq. (16) in F. M. Fernandez and E. A. Castro, Kinam **4**, 193 (1982).
[10] For the corresponding equation in spherical coordinates see D. E. Hughes, J. Phys. B **10**, 3167 (1977).
[11] U. W. Hochstrasser, in *Handbook of Mathematical Functions*, edited by M. Abramowitz and I. A. Stegun (Dover, New York, 1970), p. 771.
[12] S. I. Chan and D. Stelman, J. Chem. Phys. **39**, 545 (1963).

Linear stability of symplectic maps

J. E. Howard[a)]
University of California, Berkeley, California 94720

R. S. MacKay
Mathematics Institute, University of Warwick, Coventry CV4 7AL, England

(Received 16 September 1986; accepted for publication 17 December 1986)

A general method is presented for analytically calculating linear stability limits for symplectic maps of arbitrary dimension in terms of the coefficients of the characteristic polynomial and the Krein signatures. Explicit results are given for dimensions 4, 6, and 8. The codimension and unfolding are calculated for all cases having a double eigenvalue on the unit circle. The results are applicable to many physical problems, including the restricted three-body problem and orbital stability in particle accelerators.

I. INTRODUCTION

Symplectic maps occur in many physical problems, including orbital stability in particle accelerators,[1] plasma wave heating,[2,3] and the restricted three-body problem.[4] Any time-periodic Hamiltonian system of n degrees of freedom generates a $2n$-dimensional symplectic map, by following the flow for one period. Similarly, an autonomous Hamiltonian system of $n+1$ degrees of freedom induces a family of $2n$-dimensional symplectic maps parametrized by the value of the Hamiltonian, by considering the first return to a surface of section. Return maps provide much useful information on the behavior of continuous time systems, including the existence or nonexistence of invariant tori, and the location and stability of resonances and periodic orbits.[5]

While 2-D maps have found wide applications to physical problems and have therefore been extensively studied,[6] higher-dimensional maps also occur in many problems of current interest and there is a need to study their properties as well. For example, four- and six-dimensional symplectic maps arise in the venerable three-body problem,[4] orbits in particle accelerators,[1] and electron-cyclotron resonance heating using two wave frequencies.[7] In addition to these practical applications, there are also interesting theoretical questions concerning higher-dimensional mappings. Arnol'd diffusion has been investigated using a model 4-D symplectic map.[8] Froeschlé and Scheideker have studied the ergodic properties of four- and six-dimensional symplectic maps.[9] Mao, Satija, and Hu have recently discovered period-doubling sequences in 4-D maps, analogous to Feigenbaum sequences in one- and two-dimensional maps.[10]

In all these investigations it is useful to have analytic formulas for the linear stability limits of the fixed points (and periodic orbits) of the mappings. For example, for 2-D area-preserving maps it is well known that a necessary condition for linear stability of a periodic orbit is $|\mathrm{Tr}\, L| \leqslant 2$, where L is the tangent map round the orbit (definition to be recalled shortly). Analogous stability criteria for 4-D symplectic maps were derived by Broucke,[4] who studied periodic orbits in the restricted three-body problem, and later by Dragt.[1] (See also Refs. 11 and 12.) However, these analyses did not take into account the crucial role of the Krein signatures.[13,14] Moreover, the important case of 6-D symplectic maps has apparently not yet received attention. In this paper we derive explicit expressions for the stability boundaries for symplectic maps of arbitrary dimension, including the effects of the Krein signatures, and treat the cases of dimension up to 8 in detail.

It is also useful to know the generic ways that linear stability can be lost. For a symplectic map to lose linear stability it is necessary to have at least one multiple eigenvalue on the unit circle (S^1). In this paper, we analyze the typical behavior of families of symplectic maps of arbitrary dimension near all cases possessing a double eigenvalue on S^1. Particular attention is given to the subclass of reversible maps, which occur frequently in physical problems.

We shall find it useful to employ two closely related notions of stability of periodic orbits, both depending only on the linearization of the map about the orbit. We define the *tangent map* to a periodic orbit of period q to be the derivative of the qth iterate of the map at one of the periodic points. For a Hamiltonian system it can be represented by a symplectic matrix (the product of the Jacobian matrices round the orbit).

Definitions[15]: (i) A periodic orbit is said to be *linearly stable* if, given $\epsilon > 0$, there is a $\delta(\epsilon) > 0$ such that all orbits of the tangent map initially within δ of 0 remain within ϵ of 0 for all forward time.

(ii) A periodic orbit is said to be *spectrally stable* if all eigenvalues of the tangent map have modulus less than or equal to 1.

A periodic orbit is linearly stable iff it is spectrally stable and all Jordan blocks corresponding to eigenvalues on the unit circle are one dimensional.[16] Since, as will be shown, the boundaries of linear and spectral stability are identical for symplectic maps, the concept of spectral stability allows us to describe stability limits without continually excluding the case of multiple eigenvalues.

The paper is organized as follows. We begin by collecting together in Sec. II some useful properties of symplectic maps. We then examine the characteristic polynomial for a general symplectic matrix of degree $2n$ and show how it may

[a)] Present address: Laboratoire PMI, Ecole Polytechnique, 91128 Palaiseau, France.

 0022-2488/87/051036-16$02.50

be reduced to a polynomial of degree n. General expressions for the stability boundaries are then given in terms of the properties of this "reduced characteristic polynomial." This is accomplished in the general case by means of Sturm's method,[17] which gives the number of real roots lying in a given interval for an arbitrary polynomial with real coefficients. Maps through dimension 6 are more simply treated by an alternative method involving the discriminant of the reduced characteristic polynomial and absolute bounds on the stability region. In Sec. III we review the important role of the Krein signatures in determining the behavior of symplectic maps of dimension greater than 2. These general results are then applied to the specific cases of dimension 2, 4, 6, and 8. The 2-D case is of course well known; the 4-D problem has been treated previously,[4] but the effects of the Krein signatures were not included. These two cases are therefore described in Secs. IV and V for completeness and unity of treatment. The central result of the paper (Sec. VI) is a complete description of the stability boundaries for 6-D symplectic maps, which arise in time-periodic, three-degree-of-freedom Hamiltonian systems. In Sec. VII we also obtain explicit stability boundaries for 8-D symplectic maps, which requires Sturm's method for a complete solution.

Section VIII addresses the question of the likelihood of encountering a double eigenvalue on S^1 and determines the typical behavior of a family of maps containing a member with a double eigenvalue on S^1 as one or more control parameters are varied. The behavior depends on the signature of the eigenvalues and the Jordan normal form. We evaluate the codimension of the various possible cases having a double eigenvalue on S^1 and unfold them. We consider both the general and the reversible cases.

II. SYMPLECTIC MAPS

In this section we review some useful properties of symplectic maps and derive some general results about the subset of stable ones.

A. Symplectic maps and canonical transformations

A mapping M of a $2n$-dimensional manifold is called *symplectic*[18] if its tangent map $L = DM$, the derivative of M, preserves a nondegenerate antisymmetric bilinear form, called the skew-scalar product,

$$[L\xi, L\eta] = [\xi, \eta] \tag{2.1}$$

for all $\xi, \eta \in \mathbb{R}^{2n}$. By Darboux's theorem,[18] local coordinates can always be found such that the skew-scalar product takes the standard form

$$[\xi, \eta] = \xi^T J \eta, \tag{2.2}$$

where

$$J = \begin{bmatrix} 0 & I_n \\ -I_n & 0 \end{bmatrix} \tag{2.3}$$

and I_n is the $n \times n$ identity. Equivalently (in such a coordinate system) a mapping M is symplectic if its Jacobian matrix L satisfies

$$L^T J L = J \tag{2.4}$$

for all $\mathbf{x}$. It can be shown that $\det L = +1$, so that symplectic maps are volume preserving.

Hamiltonian systems possess a natural symplectic structure. Indeed, Hamilton's equations in n degrees of freedom can be written

$$\frac{d\mathbf{x}}{dt} = J \cdot DH(\mathbf{x}, t), \tag{2.5}$$

where $\mathbf{x} = (\mathbf{q}, \mathbf{p})$ is the $2n$-dimensional phase space point, $H(\mathbf{x}, t)$ is the Hamiltonian, and DH is its derivative with respect to $\mathbf{x}$. Canonical transformations, those which preserve the form of Hamilton's equations and the value of the Hamiltonian, may then be recognized as symplectic maps in the standard basis (2.3). Invariance of the skew-scalar product (2.2) corresponds to conservation of the Lagrange bracket. It is also well known that the time evolution of a Hamiltonian system may be viewed as a symplectic map from arbitrary initial to final states. Hence the study of Hamiltonian systems can often be reduced to that of symplectic maps, as mentioned in the Introduction.

B. Eigenvalues of symplectic maps

Let L be a symplectic matrix operating on $\mathbb{R}^{2n}$. It may be shown from (2.4) that the characteristic polynomial

$$P(\lambda) = \det(L - \lambda I) \tag{2.6}$$

is reflexive, i.e.,

$$P(\lambda^{-1}) = \lambda^{-2n} P(\lambda) \tag{2.7}$$

so that the coefficients of $P(\lambda)$ form a palindrome[18]:

$$P(\lambda) = \lambda^{2n} - A_1 \lambda^{2n-1} + A_2 \lambda^{2n-2} - \cdots + A_2 \lambda^2 - A_1 \lambda + 1. \tag{2.8}$$

The coefficients of P are easily expressed as functions of the matrix elements.

Since L is real, it follows that complex eigenvalues occur in quadruplets $(\lambda, \lambda^{-1}, \lambda^*, \lambda^{*-1})$ unless $|\lambda| = 1$, when they occur in complex conjugate pairs, while real eigenvalues come in pairs λ, λ^{-1}. Furthermore, $\lambda, \lambda^*, \lambda^{-1}$, and λ^{*-1} have the same multiplicity and Jordan block structure, and eigenvalues ± 1 have even multiplicity. It also follows that L is spectrally stable iff $|\lambda| = 1$ for all eigenvalues λ.

Following Broucke,[4] we associate with each eigenvalue λ a *stability index*

$$\rho = \lambda + \lambda^{-1}, \tag{2.9}$$

which plays a central role in our analysis, as illustrated by the following.

Lemma: A real symplectic matrix is spectrally stable iff all stability indices ρ are real with $|\rho| \leqslant 2$.

Proof: From (2.9),

$$\lambda^2 - \rho\lambda + 1 = 0 \tag{2.10}$$

so that

$$\lambda = \tfrac{1}{2}(\rho \pm i(4 - \rho^2)^{1/2}). \tag{2.11}$$

If $0 \leqslant \rho^2 \leqslant 4$, it follows from (2.11) that $|\lambda| = 1$. Conversely, if $|\lambda| = 1$, then

$$\rho = |\lambda| e^{i\varphi} + |\lambda|^{-1} e^{-i\varphi} = 2\cos\varphi \tag{2.12}$$

with φ real. Q.E.D.

C. The reduced characteristic polynomial

Using the reflexive property (2.7), dividing the characteristic polynomial (degree $2n$) by λ^n yields a polynomial in ρ of degree n:

$$Q(\rho) = \lambda^{-n} P(\lambda)$$
$$= \rho^n - A_1' \rho^{n-1} + \cdots + (-)^n A_n', \tag{2.13}$$

which we call the *reduced characteristic polynomial*. The coefficients of Q are easily derived from those of P. In fact the transformation between them is affine and invertible.

For a given root ρ, the corresponding eigenvalues (λ, λ^{-1}) are then given by (2.11). Thus the calculation of the eigenvalues of a symplectic matrix has been reduced from solving a polynomial of degree $2n$ to one of degree n, plus the quadratic (2.10). This is not only a considerable reduction in effort, but also allows explicit expressions for the eigenvalues of the otherwise intractable 6-D and 8-D cases. More importantly, it will enable us to find necessary and sufficient conditions for spectral stability in arbitrary dimensions.

D. Spectrally stable region

We can restate the previous lemma as follows.

Proposition: The symplectic matrix $L: \mathbb{R}^{2n} \to \mathbb{R}^{2n}$ is spectrally stable iff all roots of $Q(\rho)$ are real and lie in $[-2, +2]$.

Corollary: The set of reduced characteristic polynomials $Q(\rho)$ for spectrally stable symplectic matrices is homeomorphic to a closed ball in $\mathbb{R}^n$, the interior corresponding to polynomials with all roots real and distinct and in $(-2, +2)$.

Proof: The map from $\{(\rho_1,\ldots,\rho_n): -2 \leqslant \rho_1 \leqslant \cdots \leqslant \rho_n \leqslant +2\}$ to monic polynomials of degree n with all roots real and in $[-2, +2]$, defined by

$$(\rho_1,\ldots,\rho_n) \to \prod_i (\rho - \rho_i), \tag{2.14}$$

is a homeomorphism. The proof is analogous for the interiors. Q.E.D.

In particular the stable region is simply connected. The same is true in the space of coefficients of the characteristic polynomial $P(\lambda)$, since the transformation between coefficients of P and Q is a diffeomorphism.

Also it follows from the form of the characteristic polynomial that the stable region is invariant under simultaneous reflection of all the odd coefficients of $P(\lambda)$. The same is true for $Q(\rho)$.

We now derive conditions on the coefficients of Q for all its roots to be real and in $[-2, +2]$. For a polynomial $Q(\rho)$ of degree n with roots $\rho_1,\ldots,\rho_n$, counting multiplicity, the *discriminant* is defined to be

$$\Delta(Q) = \prod_{i<j} (\rho_i - \rho_j)^2. \tag{2.15}$$

It is zero iff there is a multiple root. If all the roots are distinct it is positive or negative according as the number of complex conjugate pairs is even or odd. Thus $\Delta(Q)$ undergoes a change of sign every time a quadruplet leaves S^1. The discriminant can be computed directly from the coefficients of Q without having to find the roots (Appendix A).

Define Σ_n to be the set of polynomials Q of degree n satisfying

$$\begin{aligned} &\text{(i)}\ \ Q(+2) \geqslant 0, \\ &\text{(ii)}\ \ (-)^n Q(-2) \geqslant 0, \\ &\text{(iii)}\ \ \Delta(Q) \geqslant 0. \end{aligned} \tag{2.16}$$

Theorem: If L is spectrally stable then $Q \in \Sigma_n$.

So if any of these conditions is violated then L is spectrally unstable. However, except in 2-D, they are not sufficient for spectral stability. For example, in the 4-D case there are "wedge regions" where there are two real positive pairs or two real negative pairs of eigenvalues (look ahead to Fig. 3) contained in Σ_2. Similarly, in 8-D there are configurations with two quadruplets belonging to Σ_4. In general, Σ_n can be decomposed into several regions with different configurations of eigenvalues, of which the stable region is but one.

Proposition: The set of reduced characteristic polynomials $Q(\rho)$ for spectrally stable symplectic maps is the closure of one component of the interior of Σ_n.

E. Absolute bounds

What we require now is some method to decide whether a given $Q \in \Sigma$ is in the stable region. We present first a partial answer which will turn out to be sufficient for $n \leqslant 3$. The coefficients A_k' of Q are related to the roots ρ_i by

$$\begin{aligned} A_1' &= \sum_i \rho_i, \\ A_2' &= \sum_{i<j} \rho_i \rho_j, \\ &\vdots \\ A_n' &= \rho_1 \cdots \rho_n. \end{aligned} \tag{2.17}$$

The condition on the roots for spectral stability allows one to calculate absolute bounds on the coefficients for the stable region (Appendix B), for example,

$$\begin{aligned} &\left.\begin{matrix} -2n, & n \text{ even} \\ -2(n-1), & n \text{ odd} \end{matrix}\right\} \leqslant A_2' \leqslant 2n(n-1), \\ &\qquad\qquad |A_1'| \leqslant 2n, \\ &\qquad\qquad |A_3'| \leqslant 4n(n-1)(n-2)/3, \\ &2n(n-2) - (4n - \tfrac{16}{3})\Delta n^2 + \tfrac{2}{3}\Delta^4 \leqslant A_4' \leqslant \tfrac{2}{3} n(n-1)(n-2)(n-3), \\ &\qquad\qquad \vdots \\ &\qquad\qquad |A_n'| \leqslant 2^n, \end{aligned} \tag{2.18}$$

where Δn is the closest odd (even) integer to $(3n-4)^{1/2}$ for n odd (even). The motion is unstable if any of these conditions is violated. These bounds are optimal in the sense that for each bound there is a spectrally stable map for which equality is attained.

F. Sturm's method

A complete answer to the question of how many real roots a polynomial $Q(\rho)$ has in a given interval $[a,b]$ is provided by *Sturm's method*,[17] which we now outline. Define

$$F_0(\rho) = Q(\rho),$$
$$F_1(\rho) = Q'(\rho), \tag{2.19}$$

and F_k, $k \geqslant 2$, G_k, $k \geqslant 1$, inductively by division:

$$F_{k-2}(\rho) = G_{k-1}(\rho)F_{k-1}(\rho) - F_k(\rho),$$
$$\deg(F_k) < \deg(F_{k-1}), \tag{2.20}$$

until a constant polynomial F_t is obtained ($t \leqslant \deg(Q)$). Let $V(\rho)$ be the number of changes of sign of the sequence $F_0(\rho), F_1(\rho), \ldots, F_t$, ignoring any zeros. Then the number of distinct real roots in $(a,b]$ is

$$V(a) - V(b). \tag{2.21}$$

If Q has any multiple roots then $F_t = 0$ and conversely. In that case $F_{t-1}(\rho)$ is a greatest common divisor of $Q(\rho)$ and $Q'(\rho)$. Then the number of distinct roots in the whole complex plane is

$$\deg(Q) - \deg(F_{t-1}). \tag{2.22}$$

Thus all the roots of Q are real and in $[a,b]$ iff

$$V(a-) - V(b+) + d = \deg(Q), \tag{2.23}$$

where

$$V(a-) = \lim_{a' \nearrow a}(a'), \quad V(b+) = \lim_{b' \searrow b}(b'), \tag{2.24}$$

and

$$d = \begin{cases} 0, & \text{if } F_t \neq 0, \\ \deg(F_{t-1}), & \text{otherwise.} \end{cases} \tag{2.25}$$

This is quite an easy algorithm to implement. Typically,

$$\deg(F_k) = \deg(F_{k-1}) - 1 \tag{2.26}$$

and $t = \deg(Q)$. In this case the test reduces to

$$\begin{aligned} (-)^n F_0(-2) &\geqslant 0, & F_0(+2) &\geqslant 0, \\ (-)^n F_1(-2) &\leqslant 0, & F_1(+2) &\geqslant 0, \\ &\vdots & &\vdots \\ F_{n-1}(-2) &\leqslant 0, & F_{n-1}(+2) &\geqslant 0, \\ & F_n \geqslant 0, \end{aligned} \tag{2.27}$$

because the only way to achieve $V(-2) - V(+2) = n$ is for the signs to alternate at -2 and be constant at $+2$. The case with all the above inequalities reversed is impossible since $Q(\rho)$ begins with $+\rho^n$.

The top pair of conditions are just the first two conditions defining Σ_n (2.16). We believe that the last condition is equivalent to the third condition defining Σ_n, because $F_n = 0$ iff there is a multiple root and we suspect that F_n and Δ always have the same sign (see Appendix A). The other inequalities select the correct "component" of Σ_n.

G. Stability boundary

From the results of Sec. II B limiting the possible eigenvalue configurations, it follows that there are just three basic ways for a symplectic map to lose spectral stability, as depicted in Fig. 1.

(i) *Tangent bifurcation:* Two eigenvalues coalesce at $\lambda = +1$ and split along the positive real axis (a stability index increases through $+2$).

(ii) *Period-doubling bifurcation:* Two eigenvalues coalesce at $\lambda = -1$ and split along the negative real axis (a stability index decreases through -2).

(iii) *Krein collision:* Two complex conjugate pairs of eigenvalues collide and split off S^1 at a point where $\lambda^2 \neq 1$ (two stability indices in $[-2, +2]$ merge and become complex).

Combinations of the above cases or the occurrence of eigenvalues of multiplicity greater than 2 are clearly possible, and are included in the above statement.

The first two cases are so named because, typically, periodic orbits of the full nonlinear map suffer tangent (saddle-node) and period-doubling bifurcations, respectively, in these cases. Krein collisions are named after M. G. Krein,[19] who proved that additional invariants exist for symplectic maps with eigenvalues on S^1, which may prevent colliding eigenvalues from leaving S^1. These invariants, called the Krein signatures, will be described in Sec. III.

Thus most of the information on the boundary of spectral stability is contained in the reduced characteristic polynomial.

Theorem: The boundary of the set of reduced characteristic polynomials $Q(\rho)$ corresponding to spectrally stable symplectic maps of dimension $2n$ is the union of the sets of spectrally stable maps on three "transition boundaries" given as functions of the coefficients of the reduced characteristic polynomial $Q(\rho)$ as follows:

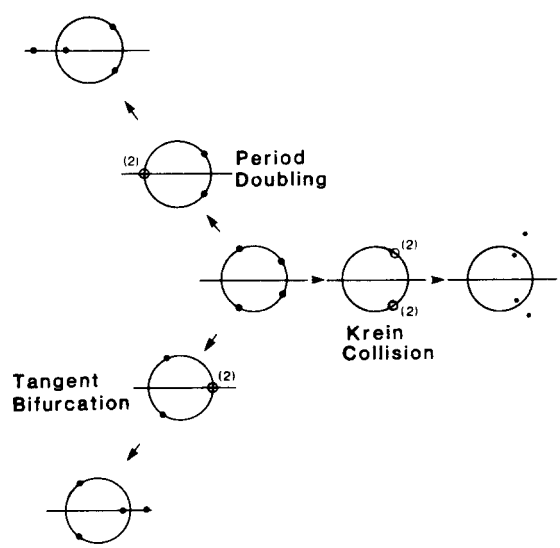


FIG. 1. The three destabilization routes for symplectic maps.

$$\begin{array}{lll} \text{(i)} & \text{Tangent bifurcation:} & Q(+2)=0 \quad (P(+1)=0); \\ \text{(ii)} & \text{Period-doubling bifurcation:} & Q(-2)=0 \quad (P(-1)=0); \\ \text{(iii)} & \text{Krein collision:} & \Delta(Q)=0. \end{array} \tag{2.28}$$

Note that not every system on one of these boundaries is spectrally stable. Absolute bounds or Sturm's method can be used to find the spectrally stable part. In particular, the transition boundary $\Delta(Q)=0$ contains other possibilities than Krein collisions, e.g., merging of two pairs of real eigenvalues into one, merging of two quadruplets into one, and merging of a quadruplet into a real pair. But the only part of this transition boundary which adjoins the stable region is the part where there is a Krein collision.

While it is easily proved that all symplectic maps satisfying (i) or (ii) are either already spectrally unstable or on the boundary of both spectral and linear stability, this is not true for maps satisfying (iii). The outcome depends on the Krein signatures, which we shall now describe.

III. KREIN'S THEOREM

We present here the basic ideas and results about Krein signatures.[13,14,19]

Consider a real linear symplectic map with a pair of eigenvalues λ, λ^* on S^1, with $\lambda^2 \neq 1$. Let V_λ be the corresponding *real invariant subspace*, that is, the space of tangent vectors of the form $\mathbf{x}=\xi+\xi^*$, where ξ is a generalized eigenvector[16] for λ, i.e., there is an integer k such that

$$(L-\lambda I)^k \xi = 0. \tag{3.1}$$

Then

$$q(\mathbf{x}) = [\mathbf{x}, L\mathbf{x}], \quad \mathbf{x}\in V_\lambda \tag{3.2}$$

is a real quadratic form which can be shown to be nondegenerate on V_λ. As is well known, a nondegenerate quadratic form can be diagonalized to a unique canonical form

$$q'(\mathbf{y}) = \sum_{i=1}^{2m} \epsilon_i y_i^2, \tag{3.3}$$

where $\epsilon_i = \pm 1$ and $2m = \dim V_\lambda$. The numbers m_+ of positive terms and m_- of negative terms are independent of how the quadratic form is diagonalized. We define the *signature* of the eigenvalue pair (λ, λ^*) to be the ordered pair (m_+, m_-). It can further be shown that $m_\pm$ are both even (for λ on $S^1 \setminus \{\pm 1\}$), so we shall often write the signature symbolically as a string of pluses and minuses, e.g. $(+\cdots+-\cdots-)$, with $m_+/2$ pluses and $m_-/2$ minuses. The signature may be calculated by any standard technique, such as Lagrange's method.[20]

The signature of an eigenvalue pair (λ,λ^*) is preserved under continuous perturbation of L so long as λ does not reach ± 1, collide with another eigenvalue, or split into two or more eigenvalues. In the second and third cases, the total signature is still conserved, provided that the eigenvalues avoid ± 1.

The signatures are significant because they may restrict the motion of the eigenvalues as parameters vary, according to the following.

Krein's Theorem[19]**:** Consider a symplectic map L with a pair of multiple eigenvalues λ, λ^* on S^1 such that $\lambda \neq \pm 1$. If the Krein signature is (positive or negative) definite (i.e., $m_- = 0$ or $m_+ = 0$), then the map has diagonal Jordan normal form on the corresponding generalized eigenspaces and the eigenvalues cannot leave S^1 under small perturbation of L.

This is easy to see because $q(\mathbf{x}) = [\mathbf{x}, L\mathbf{x}]$ is invariant under L. If it is definite then all orbits must remain bounded. Also it will remain definite for all nearby systems as they will be stable too.

This was proved independently by Moser,[12] who also proved a converse.

Moser's Theorem: If the signature is mixed ($m_+ \neq 0$ and $m_- \neq 0$) then (i) either the Jordan normal form is nontrivial or it can be made so by arbitrarily small perturbation, and (ii) the eigenvalues can be split to form (at least) a complex quadruplet by arbitrarily small perturbation of L.

Putting together these two theorems, we obtain the following.

Theorem: The boundary of spectral stability in the space of linear symplectic maps of dimension $2n$ is the subset of spectrally stable maps with an eigenvalue ± 1 or a multiple eigenvalue on S^1 of mixed signature. It is also the boundary of linear stability.

This still leaves open, however, the question of whether typical perturbations of maps with eigenvalues of mixed signature will lead to instability. In Sec. VIII we shall determine the typical behavior of systems with double eigenvalues on S^1 as parameters vary, for both mixed and definite signatures. We will also treat the cases of double eigenvalues ± 1.

The Krein signature can be defined similarly for real pairs $(\lambda, 1/\lambda)$, with $\lambda \neq \pm 1$ and for complex quadruplets $(\lambda, 1/\lambda, \lambda^*, 1/\lambda^*)$, but in both these cases there is only one possibility for the signature, viz., $m_- = m_+ = m$. If $\lambda = \pm 1$ then $q(\mathbf{x})$ must be degenerate, the degree of degeneracy depending on the Jordan normal form.

The total signature of $q(\mathbf{x}) = [\mathbf{x}, L\mathbf{x}]$, $\mathbf{x}\in\mathbb{R}^{2n}$, is the sum of the signatures of the real invariant subspaces corresponding to pairs of eigenvalues on the unit circle, complex quadruplets and eigenvalues ± 1. Thus these observations, plus the remark already made that the $m_\pm$ are even for eigenvalues on the unit circle ($\neq \pm 1$), gives one quite a lot of information about the configuration of the eigenvalues of L from $q(\mathbf{x})$ without calculating the characteristic polynomial of L. For example, if q is definite then L is linearly stable. If m_+ or m_- is odd then L is linearly unstable. But if $m_\pm$ are nonzero and even then L may be stable or unstable.

Analogous results for linear stability of equilibria of Hamiltonian systems are given in Ref. 21.

IV. TWO-DIMENSIONAL MAPS

Although the stability properties of 2-D maps are well known, we include a brief treatment of this case for com-

pleteness and to illustrate the general method. In two dimensions, symplectic maps are just those which preserve oriented area. The characteristic polynomial is

$$P(\lambda) = \lambda^2 - A\lambda + 1, \tag{4.1}$$

with $A = \mathrm{Tr}\, L$. Hence the reduced characteristic polynomial is simply

$$Q(\rho) = \rho - A, \tag{4.2}$$

so that L is spectrally stable iff $|A| \leqslant 2$. The stability boundary consists of the points $A = 2$ (tangent bifurcation), and $A = -2$ (period-doubling bifurcation). Since $Q(\rho)$ is of degree 1, the Krein collision cannot occur. Nevertheless, the Krein signature is still well defined and it is instructive to evaluate it.

Write

$$L = \begin{bmatrix} a & b \\ c & d \end{bmatrix}. \tag{4.3}$$

Then, with $\mathbf{x}^T = (x_1, x_2)$, Eq. (3.2) becomes

$$q(\mathbf{x}) = cx_1^2 + (d-a)x_1x_2 - bx_2^2, \tag{4.4}$$

which may be diagonalized by completing the square on x_1 to obtain

$$q'(\mathbf{y}) = cy_1^2 + [4 - (\mathrm{Tr}\, L)^2]y_2^2/4c \tag{4.5}$$

(if $c \neq 0$), where we have used the fact that $\det L = +1$. Thus for 2-D area-preserving maps the Krein signature is definite for elliptic orbits and mixed for hyperbolic orbits, as the results of the previous section imply. In the elliptic case the signature is the sign of c (it is easy to show then that $c \neq 0$, in fact $bc < 0$).

V. FOUR-DIMENSIONAL MAPS

Four-dimensional symplectic maps arise from time-periodic Hamiltonians of two degrees of freedom, or from autonomous three-degree-of-freedom Hamiltonian systems. The characteristic polynomial has the form

$$P(\lambda) = \lambda^4 - A\lambda^3 + B\lambda^2 - A\lambda + 1, \tag{5.1}$$

where, by the method of Leverrier,[18]

$$A = \mathrm{Tr}\, L, \quad 2B = (\mathrm{Tr}\, L)^2 - \mathrm{Tr}(L^2). \tag{5.2}$$

Alternatively, the standard expression

$$B = \sum_{i<j} \begin{vmatrix} L_{ii} & L_{ij} \\ L_{ji} & L_{jj} \end{vmatrix} \tag{5.3}$$

is somewhat more efficient computationally.

Dividing $P(\lambda)$ by λ^2 and collecting terms gives the reduced characteristic polynomial

$$Q(\rho) = \rho^2 - A\rho + B - 2. \tag{5.4}$$

The stability indices are therefore

$$\rho = \tfrac{1}{2}(A \pm (A^2 - 4B + 8)^{1/2}) \tag{5.5}$$

and (2.11) yields the four eigenvalues.

The transition boundaries for bifurcations are given by substituting $\rho = \pm 2$ in (5.4) or $\lambda = \pm 1$ in (5.1), which yields the lines

$$\lambda = +1: \quad B = +2A - 2, \tag{5.6}$$

$$\lambda = -1: \quad B = -2A - 2, \tag{5.7}$$

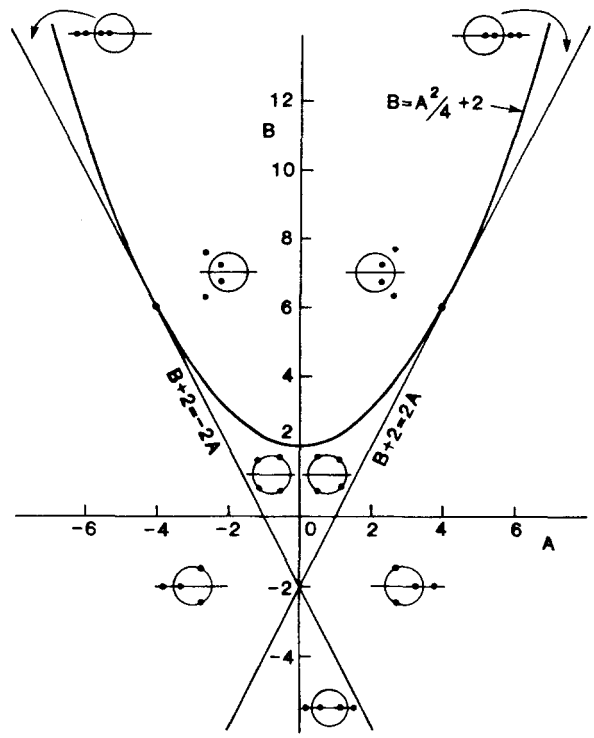


FIG. 2. Stability diagram for 4-D symplectic maps; A and B are the two coefficients of the characteristic polynomial.

which intersect transversely at $A = 0, B = -2$, as shown in Fig. 2. The motion is unstable to the right of (5.6) and to the left of (5.7). The transition boundary for Krein collisions is given by $\Delta(Q) = 0$, which is the parabola

$$B = A^2/4 + 2 \tag{5.8}$$

on which the double root is $\rho = A/2$. The motion is unstable about this curve, where the stability indices become complex. The lines (5.6) and (5.7) are tangent to the parabola at the points $A = \pm 4, B = 6$. The regions wedged between the parabola and the tangent lines to the right of $A = +4$ and left of $A = -4$ are also unstable, as shown in the inserts. The stable region is thus the arrowhead-shaped region enclosed by the three curves as shown. These results were derived by Broucke[4] in connection with the restricted three-body problem (see also Refs. 1, 11, and 12).

In summary, a 4-D symplectic map is spectrally stable iff the following conditions are simultaneously satisfied:

$$\begin{aligned} &B \geqslant 2A - 2, \quad B \geqslant -2A - 2, \\ &B \leqslant A^2/4 + 2, \quad B \leqslant 6. \end{aligned} \tag{5.9}$$

Equivalent conditions may also be obtained very efficiently by applying Sturm's method to (5.1).

Now, by Krein's theorem, not all maps reaching the parabola (5.8) can actually destabilize, only those with mixed signature. This requires calculating the total signature of $[\mathbf{x}, L\mathbf{x}]$, with $\mathbf{x} \in \mathbb{R}^4$, as explained in Sec. III. If the signature is definite then it is impossible to cross the Krein collision boundary. Indeed, we will see in Sec. VIII that it is quite unlikely to even reach it. If the signature is mixed, destabilization is possible, and in fact typical, as we shall also demonstrate in Sec. VIII.

VI. SIX-DIMENSIONAL MAPS

We now take up the main business of this paper, which is to calculate the stability boundaries for the space of 6-D symplectic maps. Such maps arise in time-periodic three degree of freedom Hamiltonian systems, and are of current interest in orbital stability in particle accelerators. The characteristic polynomial is

$$P(\lambda) = \lambda^6 - A\lambda^5 + B\lambda^4 - C\lambda^3 + B\lambda^2 - A\lambda + 1. \tag{6.1}$$

Again using the method of Leverrier, we find

$$A = \mathrm{Tr}\, L, \quad 2B = (\mathrm{Tr}\, L)^2 - \mathrm{Tr}(L^2), \quad 3C = \mathrm{Tr}(L^3) - A\,\mathrm{Tr}(L^2) + B\,\mathrm{Tr}\, L. \tag{6.2}$$

Alternatively, B and C are more economically computed using the standard expansions (5.3) and

$$C = \sum_{i<j<k} \begin{vmatrix} L_{ii} & L_{ij} & L_{ik} \\ L_{ji} & L_{jj} & L_{jk} \\ L_{ki} & L_{kj} & L_{kk} \end{vmatrix}. \tag{6.3}$$

Dividing $P(\lambda)$ by λ^3 gives the reduced characteristic polynomial

$$Q(\rho) = \rho^3 - A\rho^2 + D\rho - E, \tag{6.4}$$

where

$$D = B - 3, \quad E = C - 2A. \tag{6.5}$$

The eigenvalues are then given by (2.11). Thus the calculation of eigenvalues has been reduced from numerically solving the generally intractable sextic (6.1), for which algebraic solutions do not exist, to the solvable cubic (6.4) and quadratic (2.10).

The transition boundaries for bifurcations are again given by substituting $\rho = \pm 2$ in (6.4), or $\lambda = \pm 1$ in (6.1), which yields the planes

$$\lambda = +1: \quad A - B + C/2 = +1, \tag{6.6}$$

$$\lambda = -1: \quad A + B + C/2 = -1, \tag{6.7}$$

which intersect at an angle $\cos^{-1}(\frac{1}{9}) = 83.62°$ along the line $C = -2A, B = -1$. In viewing the stability region it therefore seems natural to orient the B axis vertically, with the planes (6.6) and (6.7) forming the lower boundary (Fig. 3).

The upper boundary is formed by the surface defined by the vanishing of the discriminant of the cubic (6.4), i.e.,

$$\Delta = -4p^3 - 27q^2 = 0, \tag{6.8}$$

where

$$p = D - A^2/3, \tag{6.9a}$$

$$q = -E + AD/3 - 2A^3/27 \tag{6.9b}$$

are the coefficients of the reduced cubic.[17] Substituting in (6.8) then yields the two-sheeted quartic surface

$$A^2D^2 + 18ADE = 4A^3E + 4D^3 + 27E^2 \tag{6.10a}$$

or

$$(AD - 9E)^2 = 4(A^2 - 3D)(D^2 - 3AE) \tag{6.10b}$$

on which the corresponding double zero of (6.4) is

$$\rho_1 = (AD - 9E)/2(A^2 - 3D) = 2(D^2 - 3AE)/(AD - 9E). \tag{6.11}$$

The two smooth sheets join at a cusped ridge where $p = q = 0$, along which $\rho_1 = \rho_2 = \rho_3 = A/3$, so that

$$A^2 = 3D, \quad AD = 9E, \quad D^2 = 3AE. \tag{6.12}$$

Figure 3 is a perspective view of the striking stability region bounded by the intersecting planes (6.6) and (6.7)

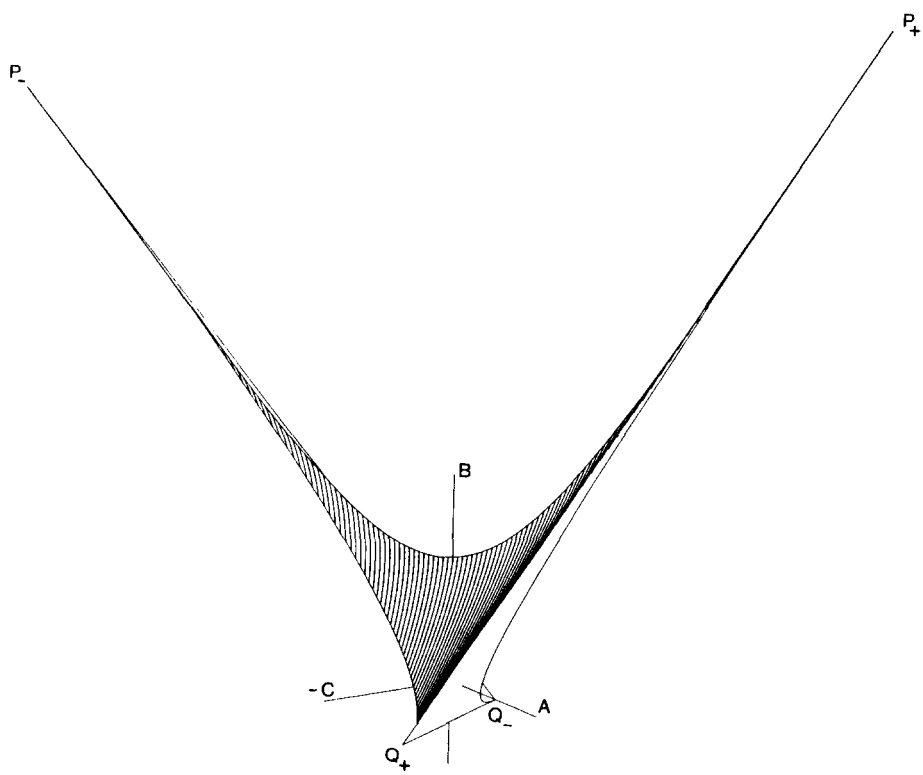


FIG. 3. Perspective view of stability region for 6-D symplectic maps in the space of characteristic polynomial coefficients (A,B,C). The planes $P_+Q_+Q_-$ and $P_-Q_-Q_+$ (the latter mostly out of view) are the transition boundaries for tangent and period-doubling bifurcations, respectively, and the quartic surface (one sheet clearly visible shaded) is the transition boundary for Krein collisions. All $\rho_i = \pm 2$ at $P_\pm$, respectively, while $\rho_1 = +2$, $\rho_2 = -2$ on Q_+Q_- with $\rho_3 = \pm 2$ at $Q_\pm$, respectively. All ρ_i are equal on the cusped ridge P_+P_-.

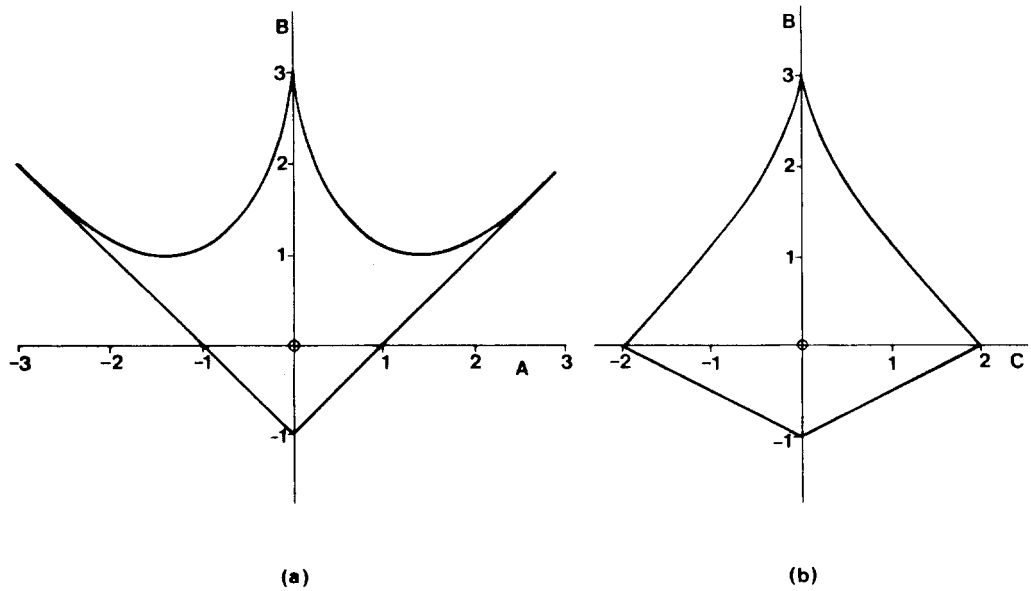


FIG. 4. Two cross sections of the 6-D stability region: (a) $A=0$ and (b) $C=0$, showing bilateral symmetry, the cusped ridge, and transverse and tangential intersections of the quartic surface with the bifurcation planes.

and the quartic surface (6.10). The stability boundary is the rather complicated surface formed by the three intersecting transition boundaries. Each sheet of the quartic surface intersects one plane tangentially and the other transversely. Together with the cusped ridge, these intersections form two roughly triangular spires terminating at the points P_+ and P_-. The ridge curls over and becomes tangent to the $\rho=\pm 2$ plane as each peak is reached. It is easily seen from the form of the characteristic equation that the stability boundary has reflection symmetry, i.e., the surface is invariant under the transformation $A\leftrightarrow -A$, $C\leftrightarrow -C$. The figure therefore has bilateral symmetry when viewed from any direction perpendicular to the B axis. Figure 4 depicts the intersections of the stability region with the AB and CB coordinate planes, clearly showing the cusped ridge, bilateral symmetry, and the transverse and tangential intersections of the quartic surface with the bifurcation planes. We now give explicit expressions for the various intersections of the transition boundaries; detailed derivations may be found in Appendix C.

The $\rho=2$ plane (6.6) is tangent to one sheet of the quartic surface along the line

$$(A-2)/1=(B+1)/4=(C+4)/6, \tag{6.13}$$

where $\rho_1=\rho_2=2$, intersecting the second sheet transversely along the curve given by

$$A^2+4=+2C, \quad A-B+C/2=1, \tag{6.14}$$

where $\rho_1=\rho_2$ and $\rho_3=+2$.

Similarly, the $\rho=-2$ plane (6.7) is tangent to the second sheet along the line

$$(A+2)/1=(B+1)/(-4)=(C-4)/6, \tag{6.15}$$

where $\rho_1=\rho_2=-2$, intersecting the first sheet transversely along the curve

$$A^2+4=-2C, \quad A+B+C/2=-1, \tag{6.16}$$

where $\rho_1=\rho_2$ and $\rho_3=-2$.

These intersections bound the line formed by the intersecting tangent planes at the points $Q_-=(-2,-1,4)$, where $\rho_1=\rho_3=-2$ and $\rho_2=2$, and $Q_+=(2,-1,-4)$, where $\rho_2=\rho_3=2$ and $\rho_1=-2$, as shown in Fig. 3. The line (6.13) and the curve (6.16), lying in the tangent plane (6.6), intersect the cusped ridge (6.12) tangentially at the point $P_+=(6,15,20)$, where $\rho_1=\rho_2=\rho_3=+2$. Similarly, the line (6.14) and curve (6.13) join with the cusped ridge at the opposite point $P_-=(-6,15,-20)$, where $\rho_1=\rho_2=\rho_3=-2$.

The simply connected stability region thus formed is shown in Appendix B to be bounded absolutely by $-1\leqslant B\leqslant 15$, $|A|\leqslant 6$, and $|C|\leqslant 20$. Of these, the single condition $B\leqslant 15$ suffices to eliminate unstable parts of Σ_3, because only one component of its interior is left.

In summary, a general 6-D symplectic map is spectrally stable iff the following conditions are simultaneously satisfied:

$$B\geqslant A+C/2-1, \quad B\geqslant -A-C/2-1,$$
$$(AD-9E)^2\leqslant 4(A^2-3D)(D^2-3AE), \quad B\leqslant 15. \tag{6.17}$$

Equivalent conditions may be obtained via Sturm's method, but they are somewhat more complicated. Note that the Krein collision condition [(6.10) and the third condition above] is quartic in A, cubic in B, but only quadratic in C, which greatly facilitates practical calculation of stability boundaries.

As in the 4-D case, whether a map satisfying the Krein collision condition is actually on the boundary of stability depends on the Krein signatures. It is straightforward to calculate the total signature by diagonalizing $[\mathbf{x},L\mathbf{x}]$, with $\mathbf{x}\in\mathbb{R}^6$. If it is definite, then the eigenvalues cannot leave S^1. However, if the signature is mixed ($++-$ or $+--$), then, except in the case of triple collisions, one must determine which eigenvalues have which signatures. This question does not arise in the 4-D case where there is only one possible collision to consider. We have not found a simple way to identify signatures with eigenvalues except by actually calculating the eigenspaces and evaluating $[\mathbf{x}, L\mathbf{x}]$ on each of the corresponding subspaces V_λ.

VII. EIGHT-DIMENSIONAL MAPS

We shall limit the treatment of the 8-D case to a derivation of the transition boundaries, without identifying tangent lines, etc.

The characteristic polynomial is

$$P(\lambda) = \lambda^8 - A\lambda^7 + B\lambda^6 - C\lambda^5 + D\lambda^4 - C\lambda^3 + B\lambda^2 - A\lambda + 1, \tag{7.1}$$

where the new coefficient D is given by

$$4D = \mathrm{Tr}(L^4) - A\,\mathrm{Tr}(L^3) + B\,\mathrm{Tr}(L^2) - C\,\mathrm{Tr}\,L \tag{7.2}$$

or

$$D = \sum_{i<j<k<l} \begin{vmatrix} L_{ii} & L_{ij} & L_{ik} & L_{il} \\ L_{ji} & L_{jj} & L_{jk} & L_{jl} \\ L_{ki} & L_{kj} & L_{kk} & L_{kl} \\ L_{li} & L_{lj} & L_{lk} & L_{ll} \end{vmatrix}, \tag{7.3}$$

with the other coefficients as in the 6-D case.

The reduced characteristic polynomial is

$$Q(\rho) = \rho^4 - A'\rho^3 + B'\rho^2 - C'\rho + D', \tag{7.4}$$

where

$$A' = A, \quad B' = B - 4, \quad C' = C - 3A, \quad D' = D - 2B + 2. \tag{7.5}$$

The transition boundaries for tangent and period-doubling bifurcations are given by setting $\rho = \pm 2$ in (7.3) or $\lambda = \pm 1$ in (7.1):

$$\lambda = +1: \quad A - B + C - D/2 = +1, \qquad \lambda = -1: \quad A + B + C + D/2 = -1. \tag{7.6}$$

These hyperplanes intersect in the two-plane

$$2B + D + 2 = 0, \quad A + C = 0. \tag{7.7}$$

The transition boundary for Krein collisions is given by setting the discriminant $\Delta(Q) = 0$. This may be accomplished by means of a theorem from the classical theory of equations. The "resolvent cubic" $R(y)$ is defined by[17]

$$R(y) = y^3 - B'y^2 + (A'C' - 4D')y - (A'^2D' - 4B'D' + C'^2). \tag{7.8}$$

It may then be shown that

$$\Delta(R) = \Delta(Q). \tag{7.9}$$

The discriminant of (7.8) is

$$\Delta(R) = -(4p^3 + 27q^2), \tag{7.10}$$

where

$$p = A'C' - 4D' - B'^2/3 \tag{7.11}$$

and

$$q = D'(8B'/3 - A'^2) - C'^2 + A'B'C'/3 - 2B'^3/27 \tag{7.12}$$

are the coefficients of the reduced cubic. Setting $\Delta(Q) = 0$ then yields the desired result, in terms of primed variables,

$$4[B'^2/3 + 4D' - A'C']^3 = 27[D'(8B'/3 - A'^2) + A'B'C'/3 - C'^2 - 2B'^3/27]^2, \tag{7.13}$$

which may alternatively be obtained by the method outlined in Appendix C.

Thus the transition boundary for Krein collisions is a sixth-degree, three-dimensional hypersurface embedded in the four-dimensional space of coefficients (A,B,C,D). Whether or not a spectrally stable system satisfying (7.13) is actually on the boundary of stability depends again on the Krein signatures.

A new feature of the eight-dimensional case is that stability can be lost by Krein collision without the discriminant (7.10) becoming negative. This is because (7.10) is the equation for a double stability index, which includes more situations than Krein collisions when $2n \geqslant 8$. In particular, it also includes collisions of complex quadruplets off S^1, that is $\rho_1 = \rho_2, \rho_3 = \rho_4 = \rho_1^*$. The condition for two double stability indices $\rho_1 = \rho_2, \rho_3 = \rho_4$ is

$$C'^2 = A'^2D', \quad A'^3 = 4A'B' - 8C'. \tag{7.14}$$

This forms a two-dimensional surface embedded in the three-dimensional hypersurface (7.13). It has two parts, corresponding to double quadruplets and to two double pairs. They meet in the special case of a pair of quadruple eigenvalues on S^1 with $\rho_1 = \rho_2 = \rho_3 = \rho_4$ (real) given by the one-dimensional curve:

$$D' = (A'/4)^4, \quad C' = A'^3/16, \quad B' = 3A'^2/8, \tag{7.15}$$

which lies on the surface (7.14).

Thus if the signature has $m_+ = m_-$, it is possible to make a transition from spectral stability to instability via (7.15), with the discriminant (7.13) remaining zero after the transition case if the system remains on (7.14) or becoming positive again by forming two complex quadruplets.

It turns out that a map on (7.14) has double quadruplets iff

$$D' > (A'/4)^4, \tag{7.16}$$

but we still require a test to determine whether a map crossing (7.13) via a point on (7.15) loses stability or not.

The stable region is invariant under reflection about the $B'D'$ plane: $A' \leftrightarrow -A', C' \leftrightarrow -C'$. The stable region may be shown (Appendix B) to be bounded by

$$|A'| \leqslant 8, \quad -8 \leqslant B' \leqslant 24, \quad |C'| \leqslant 32, \quad |D'| \leqslant 16, \tag{7.17}$$

and each of these bounds is optimal. But they do not eliminate all unstable polynomials in Σ_4; some cases with two quadruplets remain. We must use Sturm's method to obtain a complete answer.

For (7.4) Sturm's sequence is

$$\begin{aligned} F_0(\rho) &= \rho^4 - A'\rho^3 + B'\rho^2 - C'\rho + D', \\ F_1(\rho) &= 4\rho^3 - 3A'\rho^2 + 2B'\rho - C', \\ F_2(\rho) &= -F\rho^2 - E\rho - P, \\ F_3(\rho) &= H\rho + I, \\ F_4(\rho) &= P - IJ, \\ G_1(\rho) &= \rho/4 - A'/16, \\ G_2(\rho) &= -4\rho/F - G, \\ G_3(\rho) &= -F\rho/H + J, \end{aligned} \tag{7.18}$$

where

$$F = B'/2 - 3A'^2/16, \qquad E = A'B'/8 - 3C'/4,$$
$$P = D' - A'C'/16, \qquad G = -(3A' - 4E/F)/F,$$
$$H = GE - 2B' + 4P/F, \qquad I = GP + C', \tag{7.19}$$
$$J = (-E + IF/H)/H.$$

Applying the results of Sec. II F we find that the map is spectrally stable iff

$$16 + 4B' + D' \geqslant |8A' + 2C'| \quad (Q(-2), Q(+2) \geqslant 0),$$
$$|12A' + C'| \leqslant 32 + 4B',$$
$$2|E| \leqslant -P - 4F, \tag{7.20}$$
$$|I| \leqslant 2H,$$
$$P \geqslant IJ \quad (\Delta(Q) \geqslant 0).$$

VIII. CODIMENSION AND UNFOLDING OF SYMPLECTIC MAPS HAVING DOUBLE EIGENVALUES ON THE UNIT CIRCLE

It was shown in Sec. III that a Krein collision, $\rho_1 = \rho_2 \in (-2, +2)$, can never precipitate unstable motion if the corresponding Krein signature is positive or negative definite. If the signature is mixed, then Moser showed that destabilization is possible under appropriate perturbation. However, this leaves open the question of whether destabilization is likely. A preliminary question is whether collision itself is even likely.

In this section we employ some concepts of singularity theory[22,23] to quantify the likelihood that eigenvalues moving on S^1 will collide and to analyze how the eigenvalues will move for maps near a collision case. We begin by recalling a few definitions.

Definitions: Given a submanifold N embedded in a manifold M, the *codimension* of N in M is codim N = dim M − dim N. It is the dimension of any complementary space to the tangent space at any point of N.

Two submanifolds N_1 and N_2 of a manifold M are *transverse* at a point L of intersection if the sum of the tangent spaces to N_1 and N_2 at L together span the tangent space to M at L.

An *unfolding* of a point L_0 on a submanifold N of a manifold M is a differentiable family $L(\mu)$, with $\mu \in \mathbb{R}^m$, for some integer m, and $L(0) = L_0$.

The unfolding $L(\mu)$ is *transverse* to N at L_0 if the sum of the range of the derivative $\partial L/\partial \mu$ and the tangent space to N span the tangent space to M at L_0.

The unfolding $L(\mu)$ is *minitransversal*[22] if it is transverse to N at L_0 and m = codim N.

Some authors (e.g., Lu[23]) use the term "universal" for our "minitransversal," but we follow Arnol'd[22] who reserves "universal" for an unfolding $L(\mu)$ for which, given any other unfolding $L'(\nu)$, there exists a unique differentiable parameter map $\nu \to \mu$ such that the $L'(\nu)$ is equivalent to $L(\mu(\nu))$ under the equivalence relation of interest. Universal unfoldings in this sense do not exist for our problems.

A. Summary of results

We shall prove the following results in Sec. VIII B.

(1) The subsets of symplectic maps (in arbitrary dimension $2n \geqslant 4$) having a double eigenvalue $\lambda \neq \pm 1$ on S^1 and diagonal Jordan normal form (JNF) and positive or negative definite or mixed signature each have codimension 3. Thus one is unlikely to encounter such cases in one- and two-parameter families. In particular, from Krein's theorem the JNF is always diagonal for eigenvalues of definite signature. For example, one- and two-parameter families of 4-D symplectic maps with definite signature are unlikely to contain any cases with (A,B) lying on the parabola in Fig. 2.

(2) The case of a double eigenvalue $\lambda \neq \pm 1$ on S^1 with nontrivial JNF, however, has codimension 1, and can therefore occur robustly in one-parameter families.

(3) The case of a double eigenvalue ± 1 has codimension 3 for diagonal JNF and codimension 1 for nontrivial JNF.

(4) For minitransversal unfoldings of these cases there is a smooth reparametrization μ such that the relevant stability indices evolve as follows.

(a) Definite signature:

$$(\rho_2 - \rho_1)^2 = \mu_1^2 + \mu_2^2 + \mu_3^2. \tag{8.1}$$

(b) Mixed signature, nontrivial JNF:

$$(\rho_2 - \rho_1)^2 = \mu_1. \tag{8.2}$$

(c) Mixed signature, diagonal JNF:

$$(\rho_2 - \rho_1)^2 = \mu_1^2 - \mu_2^2 - \mu_3^2. \tag{8.3}$$

(d) Double eigenvalue ± 1, diagonal JNF:

$$\rho = \pm(2 + \mu_1^2 - \mu_2^2 + \mu_3^2). \tag{8.4}$$

(e) Double eigenvalue ± 1, nontrivial JNF:

$$\rho = \pm(2 + \mu_1). \tag{8.5}$$

Consequently, typical one-parameter families approaching a case with a double eigenvalue of definite signature have an "avoided collision" $(\rho_2 - \rho_1)^2 = \mu^2 + \delta^2$ for some real $\delta \neq 0$. The eigenvalues approach each other on S^1, but reach a minimum separation and then move away. However, typical one-parameter families containing a case with a double eigenvalue of mixed signature with nontrivial JNF have a parabolic collision of eigenvalues on S^1, separating parabolically as a quadruplet. One-parameter families passing near a case with a double eigenvalue of mixed signature and diagonal JNF can do various things. One possibility is a "bubble of instability," in which the eigenvalues collide on S^1, split off as a quadruplet, but then recombine on S^1, again splitting as two pairs on S^1. Conversely, a quadruplet can have a "bubble of stability," combining momentarily on S^1 where they split off as two pairs and subsequently recombine and split off S^1 as a quadruplet. Avoided collisions are also possible in this case. We leave it to the reader to make the analogous predictions for the cases of eigenvalues ± 1.

One can also deduce the form of the stability diagrams for typical two-parameter families passing near case (c). The set with a double eigenvalue is the cone $\mu_1^2 = \mu_2^2 + \mu_3^2$, so that typical 2-D sections are as sketched in Fig. 5(a), but Fig. 5(b) is not stable to perturbation.

Many symplectic maps L occurring in physical applications are *reversible*, i.e., there exists a map R with the property that $R^2 = (RL)^2 = I$, and which reverses the symplectic form. We show in Appendix D that coordinates can be cho-

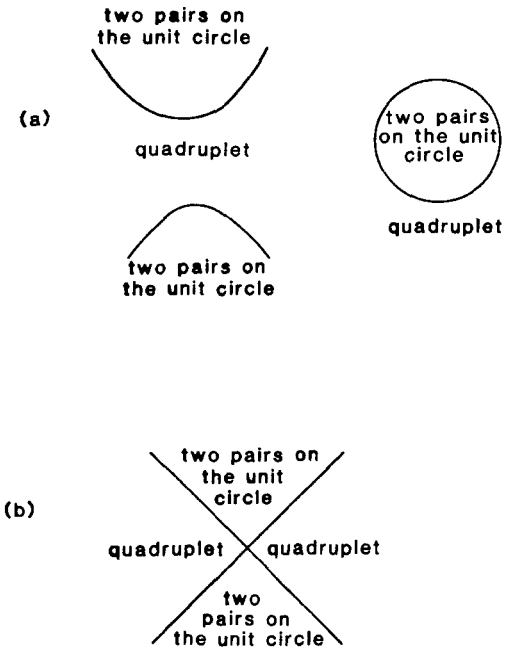


FIG. 5. (a) Two robust possibilities for stability diagrams in two-parameter families of symplectic maps near a case with a double eigenvalue of mixed signature with diagonal Jordan normal form. (b) A nonrobust stability diagram in general, but robust for two-parameter families of reversible maps.

sen such that $R(\mathbf{q},\mathbf{p}) = (\mathbf{q}, -\mathbf{p})$. Then, in the space of such matrices, the codimension of all the above cases having codimension 3 drops to 2, so that they are a little less unlikely to occur. Minitransversal unfoldings are now 2-D; one obtains the same results as before, but with $\mu_3 = 0$ in each case. The main change is that Fig. 5(b) becomes a robust stability diagram for two-parameter families with mixed signature. These results directly parallel those for the eigenvalues of equilibria of Hamiltonian systems.[18,21]

B. Calculation of codimension and unfolding

In this subsection we provide proofs for the results summarized in the previous subsection. It will prove useful to represent canonical transformations by "Poincaré generating functions of the second kind," $S(\mathbf{x}-\mathbf{x}')$, for which[24,25]

$$\mathbf{x}+\mathbf{x}' = -J\cdot DS(\mathbf{x}-\mathbf{x}'), \tag{8.6}$$

where $\mathbf{x} = (\mathbf{q},\mathbf{p})$ and derivatives are taken with respect to the difference variables $\mathbf{x}-\mathbf{x}'$. Provided

$$\det(J\cdot D^2S - I) \neq 0 \tag{8.7}$$

[which is in fact equivalent to $\det(J\cdot D^2S + I)\neq 0$] these relations generate a symplectic map $\mathbf{x}' = M(\mathbf{x})$, locally. Every symplectic map with a fixed point with no eigenvalue $\lambda = +1$ has such a generating function locally. For the linearization $L = DM$ at a fixed point, taken to be at the origin, (8.6) becomes

$$\mathbf{x}+\mathbf{x}' = -J\cdot D^2S\cdot(\mathbf{x}-\mathbf{x}'), \tag{8.8}$$

where D^2S is the $2n\times 2n$ Hessian matrix

$$D^2S = \begin{bmatrix} S_{qq} & S_{qp} \\ S_{pq} & S_{pp} \end{bmatrix} \tag{8.9}$$

evaluated at zero.

The generating function $S(\mathbf{x}-\mathbf{x}')$ is particularly well suited for treating reversible maps, as will be seen.

The eigenvalues λ_i of L are given by the characteristic equation

$$\det(J\cdot D^2S + kI) = 0, \tag{8.10}$$

where

$$k_i = (1+\lambda_i)/(1-\lambda_i). \tag{8.11}$$

Note that since $\pm k$ correspond to λ and λ^{-1}, the characteristic polynomial will be a function of k^2 alone.

The total Krein signature is easy to read off from the generating function, as follows.

Lemma: For the linear map $\mathbf{x}' = L\mathbf{x}$ generated by (8.8),

$$S(\mathbf{x}-\mathbf{x}') - S(0) = [\mathbf{x},L\mathbf{x}]. \tag{8.12}$$

Proof:

$$S(\mathbf{x}-\mathbf{x}') - S(0) = \tfrac{1}{2}(\mathbf{x}-\mathbf{x}')^T\cdot D^2S\cdot(\mathbf{x}-\mathbf{x}'). \tag{8.13}$$

Premultiplying (8.8) by $(\mathbf{x}-\mathbf{x}')^T\cdot J$ then yields

$$\begin{aligned}(\mathbf{x}-\mathbf{x}')^T\cdot D^2S\cdot(\mathbf{x}-\mathbf{x}') &= (\mathbf{x}-\mathbf{x}')^T\cdot J\cdot(\mathbf{x}+\mathbf{x}') \\ &= 2\mathbf{x}^TJ\mathbf{x}',\end{aligned} \tag{8.14}$$

which establishes (8.12). Q.E.D.

Thus since $\mathbf{x}\to\mathbf{x}-\mathbf{x}'$ is invertible (L has no eigenvalue $+1$) the quadratic form $[\mathbf{x},L\mathbf{x}]$ is equivalent to the Poincaré generating function $S(\mathbf{x}-\mathbf{x}')$ for the tangent map, so that in particular the Krein signature is the signature of S.

Next we use this Poincaré generating function representation to calculate the codimension of the subset of symplectic maps having a double eigenvalue on S^1, with definite or mixed signature, diagonal or nontrivial JNF.

1. Definite signature

Williamson[26] showed that if a symplectic map L has a double eigenvalue λ on S^1 with definite signature (without loss of generality $++$), then coordinates can be found on the subspace V_λ to cast L into a certain normal form, corresponding to the generating function

$$S(\mathbf{q},\mathbf{p}) = k^2(q_1^2+p_1^2+q_2^2+p_2^2)/2 \tag{8.15}$$

with $k = (1+\lambda)/(1-\lambda)$. Let us restrict our attention to $L_\lambda = L\,|\,V_\lambda$; the other eigenspaces play no role. We also define the *orbit* of L_λ under the group Sp(4) of 4-D symplectic matrices,

$$O(L_\lambda) = \{TL_\lambda T^{-1}:\ T\in \mathrm{Sp}(4)\}, \tag{8.16}$$

that is, the equivalence class of L_λ under symplectic coordinate changes.

The generating function for TLT^{-1} is $S\circ T^{-1}$. The orbit $O(L_\lambda)$ is a submanifold because (i) it is an algebraic variety, as it is given by polynomial equations in terms of generating functions, and (ii) it is homogeneous, because it is a group orbit. The desired quantity is then the codimension of the union over $\lambda\in S^1\setminus\{\pm 1\}$ of the orbits $O(L_\lambda)$:

$$\chi^{++} = \bigcup_\lambda O(L_\lambda). \tag{8.17}$$

It suffices to evaluate the codimension of the tangent space to χ^{++} at L_λ. All symplectic coordinate changes T: $\mathbf{x}\to\mathbf{x}'$ near the identity can be generated by the "Poincaré

generating function of the first kind," $\tau(\mathbf{x}+\mathbf{x}')$, in terms of which

$$\mathbf{x}-\mathbf{x}' = J\cdot D\tau(\mathbf{x}+\mathbf{x}'). \tag{8.18}$$

For the linearization at a fixed point, again taken at the origin,

$$\mathbf{x}-\mathbf{x}' = -J\cdot D^2\tau\cdot(\mathbf{x}+\mathbf{x}'), \tag{8.19}$$

where $D^2\tau$ is evaluated at the origin. Conversely, any quadratic function $\tau(\mathbf{x}+\mathbf{x}')$ satisfying (8.7) yields a linear symplectic map, in this case avoiding eigenvalues $\lambda=-1$ rather than $+1$. It is also well suited for dealing with reversible maps.

To first order in τ,

$$\mathbf{x} = T^{-1}\mathbf{x}' = (I-2J\cdot D^2\tau)\cdot\mathbf{x}'. \tag{8.20}$$

Using (8.20) in (8.15) and multiplying out, with $\mathbf{x}'=(\mathbf{Q},\mathbf{P})$, we find, to first order in τ,

$$\begin{aligned} S\circ T^{-1}(\mathbf{Q},\mathbf{P}) &= \tfrac{1}{2}k^2[Q_1^2+P_1^2+Q_2^2+P_2^2 \\ &\quad +4\tau_{p_1q_1}(P_1^2-Q_1^2)+4\tau_{p_2q_2}(P_2^2-Q_2^2) \\ &\quad +4(\tau_{q_1q_1}-\tau_{p_1p_1})Q_1P_1+4(\tau_{q_2q_2}-\tau_{p_2p_2})Q_2P_2 \\ &\quad +4(\tau_{p_1q_2}+\tau_{p_2q_1})(P_1P_2-Q_1Q_2) \\ &\quad +4(\tau_{q_1q_2}-\tau_{p_1p_2})(Q_1P_2+Q_2P_1)]. \end{aligned} \tag{8.21}$$

As τ varies, this yields a six-dimensional space of generating functions embedded in the ten-dimensional space of all 4-D generating functions, so that the orbit $O(L_\lambda)$ has codimension 4. Taking the union over λ (or equivalently, over k) shows that χ^{++} has codimension 3, as claimed in Sec. VIII A above.

This result implies that in one- or two-parameter families, two eigenvalues on S^1 with the same signature are unlikely to collide. To find out what they are likely to do instead, let us examine the effect of adding a general perturbation $s(\mathbf{x})$ to (8.15). Since $\pm k$ are both roots, we know that the characteristic Eq. (8.10) has the form

$$k^4+Bk^2+C=0, \tag{8.22}$$

where $B=\det D^2S$ and

$$C=\sum_{i<j}\begin{vmatrix} J\cdot D^2S_{ii} & J\cdot D^2S_{ij} \\ J\cdot D^2S_{ji} & J\cdot D^2S_{jj}\end{vmatrix}. \tag{8.23}$$

We find that the discriminant $\Delta=B^2-4C$ vanishes to first order in s. Retaining second-order terms and diagonalizing the resultant quadratic form gives

$$\begin{aligned}\Delta = k^4[&(s_{q_1q_1}+s_{p_1p_1}-s_{q_2q_2}-s_{p_2p_2})^2 \\ &+4(s_{q_1q_2}+s_{p_1p_2})^2+4(s_{q_1p_2}-s_{q_2p_1})^2]. \end{aligned}\tag{8.24}$$

Examining the formula (8.21) for the tangent space to $O(L_\lambda)$ shows that an unfolding $S+s_\mu$ is minitransversal iff it depends on three parameters, say μ_1,μ_2,μ_3, and

$$\begin{aligned}(\mu_1,\mu_2,\mu_3)\to(&s_{q_1q_1}+s_{p_1p_1}-s_{q_2q_2}-s_{p_2p_2}, \\ &2(s_{q_1q_2}+s_{p_1p_2}),2(s_{q_1p_2}-s_{q_2p_1}))\end{aligned}\tag{8.25}$$

is a diffeomorphism at 0. Since Δ is to second order in μ a nondegenerate quadratic form, it follows by the Morse lemma[23] that there is a diffeomorphic reparametrization such that

$$\Delta=\mu_1^2+\mu_2^2+\mu_3^2. \tag{8.26}$$

This establishes the claims made in Sec. A about the behavior of eigenvalues, for the case of definite signature.

If one wants to restrict attention to reversible maps, we show in Appendix D that we can always take the reversor to be $R(\mathbf{q},\mathbf{p})=(\mathbf{q},-\mathbf{p})$. It is readily shown that for the two Poincaré generating functions employed above, the resultant maps are reversible with respect to this reversor iff $S(\mathbf{q},\mathbf{p})=S(\mathbf{q},-\mathbf{p})$ and $\tau(\mathbf{q},\mathbf{p})=-\tau(\mathbf{q},-\mathbf{p})$, respectively. It follows that the space of generating functions S reversible with respect to R is only six dimensional, and from (8.21) that the orbit of L_λ under symplectic coordinate changes preserving R has codimension 3. Thus, taking the union over λ, we see that the existence of a double eigenvalue on S^1 with definite signature is only codimension 2 in the space of reversible symplectic maps. Furthermore, the third term in (8.24) is zero for reversible unfoldings, and the behavior of eigenvalues in a minitransversal family is given by $\Delta=\mu_1^2+\mu_2^2$.

2. Mixed signature

In the case of a double eigenvalue on $S^1\setminus\{\pm1\}$ having mixed signature and nontrivial Jordan normal form, Williamson obtained the normal form with generating function (actually, an equivalent one):

$$S(\mathbf{q},\mathbf{p})=k^2(q_2p_1-q_1p_2)+ap_1^2, \quad k,a\neq0. \tag{8.27}$$

In this case we can evaluate the codimension and unfolding in a simpler fashion. Add a general perturbation $s(\mathbf{q},\mathbf{p})$. One then finds that the discriminant of the characteristic polynomial (8.10) is, to first order in s,

$$\Delta=4k^4as_{q_1q_1}. \tag{8.28}$$

Since $k^4a\neq0$, the implicit function theorem guarantees that the set of maps with $\Delta=0$ is locally a codimension 1 surface.

Exceptionally, $a=0$; this is the case of diagonal Jordan normal form. We then have to use the same method as for the case of definite signature. We could use the normal form (2.27) with $a=0$ but we prefer an alternative normal form for this case:

$$S(\mathbf{q},\mathbf{p})=k^2(q_1^2+p_1^2-q_2^2-p_2^2)/2. \tag{8.29}$$

As in the definite case, one finds that the orbit of L_λ (the map generated by 8.29) has codimension 4, thus

$$\chi^{+-}=\bigcup_{\lambda\in S^1\setminus\{\pm1\}}O(L_\lambda) \tag{8.30}$$

has codimension 3. For a general perturbation $s(\mathbf{q},\mathbf{p})$ of (8.29), one finds to second order in s that the discriminant of the characteristic polynomial is

$$\begin{aligned}\Delta = k^4[&(s_{q_1q_1}+s_{p_1p_1}+s_{q_2q_2}+s_{p_2p_2})^2 \\ &-4(s_{q_1q_2}-s_{p_1p_2})^2-4(s_{q_1p_2}+s_{q_2p_1})^2]. \end{aligned}\tag{8.31}$$

From the formula for $O(L_\lambda)$ it follows that an unfolding $S+s_\mu$ is minitransversal to χ^{+-} iff it contains three parameters and

$$(\mu_1,\mu_2,\mu_3)\to(s_{q_1q_1}+s_{p_1p_1}+s_{q_2q_2}+s_{p_2p_2},$$
$$2(s_{q_1q_2}-s_{p_1p_2}),\,2(s_{q_1p_2}+s_{p_1q_2}))\qquad(8.32)$$

is a diffeomorphism at 0. Applying the Morse lemma[23] once again, we see that a diffeomorphic change of parameters can be found such that

$$\Delta=\mu_1^2-\mu_2^2-\mu_3^2.\qquad(8.33)$$

Again, in the reversible case the codimension drops to 2 and minitransversal families can be reparametrized such that

$$\Delta=\mu_1^2-\mu_2^2.\qquad(8.34)$$

3. Double eigenvalue ± 1

This subsubsection was inspired by work of Chillingworth and Afsharnejad[27] on the Mathieu equation. Double eigenvalues ± 1 are essentially a 2-D phenomenon, so we need only consider matrices

$$\begin{bmatrix}A & B\\ C & D\end{bmatrix}$$

with $AD-BC=1$. We could use generating functions, but the 2-D case is simple enough that we can manage without them. Diagonalizing the quadratic form $AD-BC$, we see that this set forms a hyperboloid in the space of A,B,C,D:

$$(A+D)^2-(A-D)^2-(B+C)^2+(B-C)^2=4.\qquad(8.35)$$

The eigenvalues are determined by the trace (stability index) $\rho=A+D$. The subset of matrices with a double eigenvalue $+1$ is given by

$$A+D=2,\qquad(A-D)^2+(B+C)^2=(B-C)^2,\qquad(8.36)$$

which is a two-dimensional cone. The vertex is the identity matrix, all others having nontrivial Jordan normal form. Thus existence of a double eigenvalue $+1$ with diagonal Jordan normal form is codimension 3 in the space of symplectic maps.

From (8.35), minitransversal families have a reparametrization (μ_1,μ_2,μ_3) such that the stability index

$$\rho=2+\mu_1^2-\mu_2^2+\mu_3^2.\qquad(8.37)$$

The case of nontrivial Jordan normal form is codimension 1 and minitransversal families have a reparametrization μ_1 such that

$$\rho=2+\mu_1.\qquad(8.38)$$

Reversibility with respect to

$$R=\begin{bmatrix}1 & 0\\ 0 & -1\end{bmatrix}$$

is obtained iff $A=D$ (Ref. 28). Thus, existence of a double eigenvalue $+1$ with diagonal Jordan normal form is only codimension 2 in the space of reversible symplectic maps, and minitransversal families have a reparametrization such that

$$\rho=2+\mu_1^2-\mu_2^2.\qquad(8.39)$$

The analysis for a double eigenvalue -1 is analogous, but with the sign of ρ changed throughout.

C. Further remarks

There are other cases whose codimension and unfolding might be worth finding, for example existence of a quadruple eigenvalue $+1$ or -1, and existence of triple eigenvalues λ, λ^* on S^1, with $\lambda\neq\pm 1$, with the various possibilities for Jordan normal form and signature. But the ones we have treated are, we believe, the major cases of interest.

Finally, it is worth remembering that whereas families of linear symplectic maps may cross the "saddle-node" boundary quite freely, a typical unfolding of a nonlinear map with a periodic orbit having an eigenvalue $+1$ has a saddle-node bifurcation there. For one sign of the parameter there are two periodic orbits with stability on opposite sides of the saddle-node boundary. The periodic orbits collide parabolically and annihilate each other.[29] In the reversible families, however, symmetry breaking is a persistent alternative.[30]

IX. DISCUSSION

A general method has been presented which reduces the calculation of the eigenvalues of a symplectic matrix of dimension $2n$ to solving a "reduced characteristic equation" of degree n. This not only greatly simplifies the calculation of eigenvalues, but also makes it possible to obtain explicit spectral stability limits in terms of the coefficients of the characteristic polynomial for symplectic maps of arbitrary dimension, using Sturm's method. This has been done here in detail for dimensions 2, 4, 6, and 8.

Stable symplectic maps have all their eigenvalues on the unit circle. In order to lose stability some of them must collide. According to Krein's theorem, however, colliding eigenvalues with definite signature and $\lambda\neq\pm 1$ are constrained from leaving S^1, because the definite conserved quadratic form (3.2) is incompatible with unstable motion. Calculation of the signature turns out to be fairly straightforward for 4-D maps, but requires calculating the invariant subspaces for each eigenvalue pair for 6-D and 8-D maps. In the case of mixed signature stability may or may not be lost. However, examination of the possible unfoldings show that in the definite case, one- and two-parameter families are unlikely to even reach the boundary for Krein collisions. Collision in the case of mixed signature, on the other hand, is codimension 1 for nontrivial Jordan normal form, and destabilization is typical.

The stability boundaries derived in this paper are of interest both in physical problems and "pure" mathematical studies. We expect their greatest utility will lie in orbital calculations for particle accelerators, which entail computing the eigenvalues of a great many 4-D or 6-D matrices. However, the potential user of these results is cautioned to calculate the Krein signatures as well as the eigenvalues, in order to fully determine the stability properties of the maps. In this connection it may be useful to observe that signatures may be calculated *inside* as well as on the stability boundary, since they are conserved under continuous perturbations, provided that the eigenvalues avoid ± 1. Another applica-

tion is the search for period-doubling sequences in six- and eight-dimensional maps, similar to those recently discovered in four-dimensional maps.[10] The stability boundaries derived in Secs. VI and VII should prove useful in locating appropriate initial conditions for such sequences.

ACKNOWLEDGMENTS

It is a pleasure to thank Paul Weiss for several useful suggestions, Allan Lichtenberg and John Greene for their comments on earlier drafts of this paper, and David Chillingworth for useful discussions.

APPENDIX A: FORMULA FOR THE DISCRIMINANT

The discriminant of a polynomial Q can be evaluated directly from the coefficients of Q without having to find the roots ρ_i.[17] Define the *Newton coefficients*

$$s_k = \sum_i \rho_i^k. \tag{A1}$$

These can be evaluated in terms of the coefficients A'_k of Q (2.13) by *Newton's identities*:

$$\begin{aligned} &s_0 = n, \\ &s_k - A'_1 s_{k-1} + \cdots + (-)^{k-1} A'_{k-1} s_1 \\ &\quad + (-)^k k A'_k = 0, \quad 1 \leqslant k \leqslant n, \\ &s_k - A'_1 s_{k-1} + \cdots + (-)^n A'_n s_{k-n} = 0, \quad k > n, \end{aligned} \tag{A2}$$

where n is the degree of Q. Then the discriminant $\Delta(Q)$ is given by the determinant

$$d_n = \begin{vmatrix} s_0 & s_1 & \cdots & s_{n-1} \\ s_1 & s_2 & \cdots & s_n \\ \vdots & & & \\ s_{n-1} & & \cdots & s_{2n-2} \end{vmatrix}. \tag{A3}$$

We conjecture that the leading coefficient of $F_i(\rho)$ in Sturm's method is given in terms of the Newton coefficients by

$$d_i \Big(\prod_{j<i} d_j^2 \Big)^{-1}, \tag{A4}$$

but have not proved it in general. In particular, it would follow that F_n is the discriminant divided by the square of a polynomial in the coefficients, hence they would have the same sign.

APPENDIX B: ABSOLUTE BOUNDS ON THE STABILITY REGION

In this appendix we derive the absolute bounds (2.18) on the stability region in the space of reduced characteristic polynomials.

To find the extrema of

$$A'_m = \sum_{1 \leqslant i_1 < \cdots < i_m \leqslant n} \rho_{i_1} \cdots \rho_{i_m} \tag{B1}$$

over the hypercube $|\rho_i| \leqslant 2$ real, it suffices to look at its values at the vertices, $\rho_i = \pm 2$, since A'_m is affine in each ρ_i separately:

$$\begin{aligned} A'_m = \rho_n &\sum_{1 \leqslant i_1 < \cdots < i_{m-1} \leqslant n-1} \rho_{i_1} \cdots \rho_{i_{m-1}} \\ &+ \sum_{1 \leqslant i_1 < \cdots < i_m \leqslant n-1} \rho_{i_1} \cdots \rho_{i_m}. \end{aligned} \tag{B2}$$

The upper bounds on A'_m in (2.18) come from taking all $\rho_i = +2$. When m is odd, the lower bounds come from taking all $\rho_i = -2$.

For m even, the lower bounds can be obtained as follows. If there are n_- roots $\rho = -2$ and n_+ roots $\rho = +2$ ($n_- + n_+ = n$) then counting arguments show that A'_m is 2^m times a sum of products of binomial coefficients which one can recognize as the coefficient of x^m in

$$f(x) = (1-x)^{n_-}(1+x)^{n_+}. \tag{B3}$$

Thus

$$A'_m = 2^m f^{(m)}(0)/m!. \tag{B4}$$

For example,

$$A'_2 = 2((\Delta n)^2 - n), \tag{B5}$$

$$A'_4 = \tfrac{2}{3}[3n(n-2) - (6n-8)(\Delta n)^2 + (\Delta n)^4], \tag{B6}$$

where

$$\Delta n = n_+ - n_-. \tag{B7}$$

The minimum of A'_m over the stability region is then given by minimizing (B4) over Δn.

An alternative set of absolute bounds on the stable region is provided by the following optimal bounds on the Newton coefficients (A1):

$$\left.\begin{matrix} 0, & k \text{ even} \\ -n\,2^k, & k \text{ odd} \end{matrix}\right\} \leqslant s_k \leqslant n\,2^k. \tag{B8}$$

APPENDIX C: GEOMETRY OF 6-D STABILITY REGION

We present here additional details on the structure of the 6-D stability region illustrated in Fig. 3. Explicit formulas have been derived in Sec. VI for the quartic surface and the two planes that bound the stability region, defined by existence of a double stability index $\rho_1 = \rho_2$, and stability indices ± 2. The edges and corners may be obtained from the coincidences of these conditions, which are most easily found by writing

$$Q(\rho) = (\rho - \rho_1)(\rho - \rho_2)(\rho - \rho_3) \tag{C1}$$

and matching coefficients with those in (6.4) to obtain

$$\begin{aligned} A &= \rho_1 + \rho_2 + \rho_3, \\ D &= \rho_1\rho_2 + \rho_1\rho_3 + \rho_2\rho_3, \\ E &= \rho_1\rho_2\rho_3. \end{aligned} \tag{C2}$$

These equations can also be used to find the quartic surface and the values of the double root $\rho_1 = \rho_2$ and the third root ρ_3 on it. A less direct method is to solve $Q(\rho) = 0$ and $Q'(\rho) = 0$ simultaneously.

We now use Eqs. (C2) to derive explicit expressions for all the edges and corners of the stability region.

$\rho_1 = \rho_2 = \rho_3$: This yields the cusped ridge (6.12). Equations (C2) show immediately that $A = 3\rho_3$, $D = 3\rho_3^2$, and $E = \rho_3^3$ serve to define the ridge parametrically and show that $p = q = 0$ there (6.9).

$\rho_1 = \rho_2 = \pm 2$: The quartic surface and the $\rho = \pm 2$ plane are tangent along the lines (6.13) and (6.15), respectively, on which $\rho_3 = A \mp 4$. The equations for the lines follow directly from Eqs. (C2) by eliminating ρ_3. Regarding tangency, we prove the somewhat more general result that for all ρ the quartic surface is tangent to the plane where there is a root ρ at the line of two roots $= \rho$.

Proof: The plane where there is a root ρ_1 has equation $Q(\rho_1) = 0$, so if (A_0, D_0, E_0) is one point on it, then it is given by

$$\rho^2 \Delta A - \rho \Delta D + \Delta E = 0, \tag{C3}$$

where $\Delta A = A - A_0$, etc. For the tangent plane to the quartic surface, set $\rho_1 = \rho_2$ in (C2) and differentiate with respect to ρ_1 and ρ_3. One finds

$$\begin{aligned} \Delta A &= 2\Delta\rho_1 + \Delta\rho_3, \\ \Delta D &= 2(\rho_1 + \rho_3)\Delta\rho_1 + 2\rho_1 \Delta\rho_3, \\ \Delta E &= 2\rho_1 \rho_3 \Delta\rho_1 + \rho_1^2 \Delta\rho_3, \end{aligned} \tag{C4}$$

to first order. One easily checks then that (C3) is satisfied for all $\Delta\rho_1$, $\Delta\rho_3$. Q.E.D.

$\rho_1 = \rho_2$, $\rho_3 = \pm 2$: The quartic surface intersects the $\rho = \pm 2$ planes transversely along the space curves

$$A^2 \pm 4 = 2C, \tag{C5}$$

along which $\rho_1 = A/2 \mp 1$, respectively. The proof follows directly from Eqs. (C2).

$\rho_1 = 2$, $\rho_2 = -2$: The two tangent planes intersect at the line $C = -2A$, $B = -1$, along which $\rho_3 = A$. Again, the proof is straightforward.

Triple coincidences are also easily found from (C2). We consider the form of the stability region near the tops of the spires.

$\rho_1 = \rho_2 = \rho_3 = \pm 2$: The tangent line $\rho_1 = \rho_2 = \pm 2$, intersection curve $\rho_1 = \rho_2, \rho_3 = \pm 2$, and the cusped ridge all intersect tangentially at the point $(A,B,C) = (\pm 6, 15, \pm 20)$. We give the proof for $\rho = +2$. From (6.13) tangent vectors to the tangent line have $(\Delta A, \Delta B, \Delta C) \propto (1,4,6)$. From (6.14), tangent vectors to the intersection curve have

$$\begin{aligned} \Delta C &= A \Delta A, \\ \Delta B &= \Delta A + \Delta C/2 = (1 + A/2)\Delta A; \end{aligned} \tag{C6}$$

setting $A = 6$ then gives $(\Delta A, \Delta B, \Delta C) \propto (1,4,6)$. Finally, tangent vectors to the cusped ridge have $(\Delta A, \Delta D, \Delta E) \propto (3, 6\rho_1, 3\rho_1^2) \propto (1,4,4)$ for $\rho_1 = +2$, so $(\Delta A, \Delta B, \Delta C) \propto (1,4,6)$ again. Q.E.D.

As we advance upwards in B, the cusped ridge curls over like a breaking wave, as shown in Fig. 5. We now show that, in the limit as $B \to 15$, the tangent plane to the surface at the ridge becomes parallel to the $\rho = 2$ plane.

Proof: Expanding the quartic surface (6.10) to second order about the ridge, we obtain to second order,

$$[\Delta A \ \ \Delta B \ \ \Delta C] \begin{bmatrix} AE & -3E & D \\ -3E & D & -A \\ D & -A & 3 \end{bmatrix} \begin{bmatrix} \Delta A \\ \Delta B \\ \Delta C \end{bmatrix} = 0. \tag{C7}$$

Diagonalizing this quadratic form, we find, to second order,

$$AE(\Delta A - 3\Delta D/A + D\Delta E/AE)^2 = 0. \tag{C8}$$

The tangent plane at the ridge is therefore

$$\Delta A - 3\Delta D/A + D\Delta E/AE = 0. \tag{C9}$$

At the peak, $A = 6$, $D = 12$, and $E = 8$, so that (C9) becomes

$$\Delta A - \Delta D/2 + \Delta E/4 = 0, \tag{C10}$$

which is easily seen to coincide with the $\rho = 2$ plane (6.6). Q.E.D.

APPENDIX D: NORMAL FORM FOR REVERSORS

A symplectic map T is called *reversible* if there exists a map R (called a *reversor*) such that $R^2 = I$, $(RT)^2 = I$ (so that $R^{-1} T R = T^{-1}$), and R is *antisymplectic*, that is

$$[DR\,\xi, DR\eta] = -[\xi, \eta] \tag{D1}$$

for all tangent vectors ξ, η, where DR is the derivative of R. We believe that by a symplectic coordinate change, R can be put into the standard form

$$R(\mathbf{q}, \mathbf{p}) = (\mathbf{q}, -\mathbf{p}) \tag{D2}$$

in a neighborhood of any fixed point of R. We will prove this here for the linearization only, as that is all we need in the present context.

Proof: $R^2 = I$ implies that the space can be decomposed as $V_+ \oplus V_-$, with

$$\begin{aligned} Rv_+ &= +v_+ \quad \forall\ v_+ \in V_+, \\ Rv_- &= -v_- \quad \forall\ v_- \in V_-. \end{aligned} \tag{D3}$$

Then R antisymplectic implies that $[v_+, w_+] = [v_-, w_-] = 0$, $\forall v_+, w_+ \in V_+$, $v_-, w_- \in V_-$. Choose a basis $(v_+^i)_{i=1\cdots i_+}$ for V_+ and $(v_-^i)_{i=1\cdots i_-}$ for V_-. The symplectic form is nondegenerate, so $\exists\ v_-^i$ such that $[v_+^1, v_-^i] \neq 0$. Now permute the basis for V_- so that $[v_+^1, v_-^1] \neq 0$ and subtract $[v_+^1, v_-^i] v_-^1 / [v_+^1, v_-^1]$ from v_-^i to achieve $[v_+^1, v_-^i] = 0$, $i > 1$. Similarly, subtract $[v_+^i, v_-^1] v_+^1 / [v_+^1, v_-^1]$ from v_+^i to obtain $[v_+^i, v_-^1] = 0$, $i > 1$. Proceed by induction. We cannot run out of v_+^i before v_-^i or vice versa because then the symplectic form would be degenerate. Thus $i_+ = i_-$ and the new basis puts the symplectic form into the cannonical form $[v_+^i, v_-^j] = \delta_{ij}$ and R into the desired form. Q.E.D.

[1] A. J. Dragt, *Lectures on Nonlinear Orbit Dynamics, AIP Conference Proceedings*, Vol. 87 (AIP, New York, 1982).
[2] M. A. Lieberman and A. J. Lichtenberg, Plasma Phys. **15**, 125 (1973).
[3] J. E. Howard, M. A. Lieberman, A. J. Lichtenberg, and R. H. Cohen, in *Proceedings of the 2nd Workshop on Hot Electron Ring Physics*, edited by N. A. Uckan, Conf. 811203, June, 1982, p. 561.
[4] R. Broucke, AIAA J. 7, 1003 (1969).
[5] G. Contopoulos, P. Magnenat, and L. Martinet, Physica D **6**, 126 (1982).
[6] A. J. Lichtenberg and M. A. Lieberman, *Regular and Stochastic Motion* (Springer, New York, 1983).
[7] J. E. Howard, A. J. Lichtenberg, M. A. Lieberman, and R. H. Cohen, Physica D **20**, 259 (1986).
[8] M. A. Lieberman, Ann. N.Y. Acad. Sci. **357**, 119 (1980).
[9] C. Froeschlé and J.-P. Scheideker, Astron. Astrophys. **16**, 172 (1972).
[10] J.-M. Mao, I. I. Satija, and B. Hu, Phys. Rev. A **32**, 1927 (1985).
[11] Danby, Astron. J. **69**, 165 (1964).

[12]R. Courant and Snyder, Ann. Phys. **3**, 1 (1958).
[13]V. I. Arnol'd and A. Avez, *Ergodic Problems of Classical Mechanics* (Benjamin, New York, 1968), Appendix 29.
[14]J. K. Moser, Commun. Pure Appl. Math. **9**, 673 (1958).
[15]D. D. Holm, J. E. Marsden, T. Ratiu, and A. Weinstein, Phys. Rep. **123**, 1 (1985).
[16]M. W. Hirsch and S. Smale, *Differential Equations, Dynamical Systems and Linear Algebra* (Academic, New York, 1974).
[17]L. E. Dickson, *New First Course in The Theory of Equations* (Wiley, New York, 1939), p. 43 *et seq.*
[18]V. I. Arnol'd, *Mathematical Methods of Classical Mechanics* (Springer, New York, 1978), Appendix 6.
[19]M. G. Krein, Pamyati, A. A. Andronova, Izv. Akad. Nauk SSSR **1955**, 413; translation in M. G. Krein, *Topics in Differential and Integral Equations and Operator Theory* (Birkhäuser, Boston, 1983), pp. 1–105.
[20]F. R. Gantmacher, *Matrix Theory* (Chelsea, New York, 1959), Vol. I, p. 87.
[21]R. S. MacKay, "Stability of equilibria of Hamiltonian systems," in *Nonlinear Phenomena and Chaos*, edited by S. Sarkar (Hilger, Bristol, 1986), pp. 254–270.
[22]V. I. Arnol'd, *Geometrical Methods in the Theory of Ordinary Differential Equations* (Springer, New York, 1983), Chap. 6.
[23]Y.-C. Lu, *Singularity Theory and an Introduction to Catastrophe Theory* (Springer, New York, 1976).
[24]H. Poincaré, *Les Méthodes Nouvelles de la Mecanique Céleste*, Tome 3 (Gauthier–Villars, Paris, 1899).
[25]See Ref. 1, Appendix 6.
[26]J. Williamson, Am. J. Math. **59**, 599 (1937).
[27]Z. Afsharnejad, "Perturbation geometry for the Mathieu equation," Indian J. Appl. Math., to appear.
[28]R. S. Mackay, thesis, Princeton University, 1982 (Univ. Microfilms Int., Ann Arbor, MI), Sec. 1.2.3.2.
[29]K. R. Meyer, Trans. Am. Math. Soc. **149**, 95 (1970).
[30]R. Rimmer, J. Diff. Eqs. **28**, 329 (1978).

Gauge and dual symmetries and linearization of Hirota's bilinear equations

S. Saito
Department of Physics, Tokyo Metropolitan University, Setagaya-ku, Tokyo, Japan 158

N. Saitoh
Department of Applied Mathematics, Yokohama National University, Hodogaya-ku, Yokohama, Japan 240

(Received 7 July 1986; accepted for publication 14 January 1987)

In Hirota's [Hiroshima University Technical Report Nos. A 6, A 9, 1981; J. Phys. Soc. Jpn. **50**, 3785 (1981)] bilinear difference equation which is satisfied by solutions to the Kadomtsev–Petviashvili (KP) hierarchy, gauge and dual symmetries are found, which enable one to reduce the problem of solving the nonlinear equation to solving a single linear equation.

I. INTRODUCTION

In studying nonlinear integrable systems it is useful to find their similarities and investigate them from a unified point of view. Along this line one of the present authors studied[1] a nonlinear equation which was derived from the Toda lattice through a transformation of independent variables. It reduces to the KdV equation and to the Toda lattice itself in certain respective limits of a parameter. This generalized Toda lattice itself was shown to be integrable in all range of the parameter.

Based on the same viewpoint Hirota proposed[2] an equation which reduces into various types of soliton equations in some limits of appropriate combination of independent variables. Among them are the KdV equation, Kadomtsev–Petviashvili (KP) equation, modified KdV equation, sine–Gordon equation, Toda lattice, two-dimensional Toda lattice, etc. He also gave three-soliton solutions and conjectured the integrability of the equation.

A further interesting observation was made by Miwa[3] who found a transformation of independent variables which connects Hirota's equation and the hierarchy of the KP equation. The latter equation has been studied intensively by many authors.[4] It contains an infinite number of soliton equations whose solutions have been classified completely by mathematical terms. This offers a typical example which shows the importance of the view described above.

Besides these soliton equations there have been known equations which are characterized by the gauge symmetry and an infinite number of solutions have been given under certain conditions. The Yang–Mills theory and the Einstein equation of gravity belong to this category.[5] The gauge symmetry plays an essential role in these theories. It is, therefore, desirable to formulate the soliton equations by means of the gauge theory.

For the purpose to establish such a formalism we like to study Hirota's equation within the framework of the gauge theory in this paper. In our formalism Hirota's equation emerges as a compatibility relation of a pair of equations for a field which are covariant under gauge transformations. Therefore the gauge symmetry appears as a hidden symmetry of the nonlinear equation. Another combination of this pair of equations yields a linear equation for the field. The pair of equations was called the duality equations in our previous article[6] in which a special limit of Hirota's equation, i.e., the two-dimensional Toda lattice, was studied along the same line of the present paper.

The most remarkable feature of the duality equations is that they are symmetric under the exchange of the roles of the gauge field and the amplitude field. This dual symmetry implies that the amplitude field itself satisfies Hirota's equation. An important consequence of this property is that this scheme provides a kind of Bäcklund transformation which enables us to generate a new solution of the nonlinear equation from a given solution of it, just by solving the single linear equation derived from the duality equations. In other words the problem of solving the nonlinear equation is reduced to solve the linear equation associated to the nonlinear equation.

It has been already shown[7] how this scheme works in the case of the two-dimensional Toda lattice which can be obtained from Hirota's equation in particular limits of parameters. There are some examples of solutions, which satisfy both the nonlinear and the linear equations simultaneously, that were given,[7] although no attention was paid to their gauge symmetric nature. Our present work claims that this scheme can be generalized to include all of the nonlinear equations described by Hirota's equation.

II. GAUGE AND DUAL SYMMETRIES

Hirota's equation is given by[2]

$$\begin{aligned}&\alpha f(\lambda+1,\mu,\nu)f(\lambda-1,\mu,\nu)\\&\quad+\beta f(\lambda,\mu+1,\nu)f(\lambda,\mu-1,\nu)\\&\quad+\gamma f(\lambda,\mu,\nu+1)f(\lambda,\mu,\nu-1)=0.\end{aligned}\tag{1}$$

As proved by Miwa,[3] this equation is satisfied by τ functions of the KP hierarchy.

To investigate gauge properties of Eq. (1) we first define covariant difference operators by

$$\begin{aligned}\nabla_+ f_n(l,m)&=\exp(-A^{\,l}_{\,n}(l,m))f_n(l+1,m)\\&\quad-f_n(l,m),\end{aligned}\tag{2a}$$

$$\begin{aligned}\nabla_- f_n(l,m)&=\exp(-A^{\,m}_{\,n-1}(l,m))f_n(l,m+1)\\&\quad-f_n(l,m).\end{aligned}\tag{2b}$$

The gauge transformation should be defined by

$$f_n(l,m)\to\exp(V_n(l,m))f_n(l,m),\tag{3a}$$

$$A^{\,l}_{\,n}(l,m)\to A^{\,l}_{\,n}(l,m)+V_n(l+1,m)-V_n(l,m),\tag{3b}$$

 0022-2488/87/051052-04$02.50

$$A_n^m(l,m) \to A_n^m(l,m) + V_n(l,m+1) - V_n(l,m). \quad (3c)$$

Then $\nabla_\pm f_n(l,m)$ transform into $\exp(V_n(l,m))\nabla_\pm f_n(l,m)$ provided $V_n(l,m)$ satisfies

$$V_n(l,m+1) - V_n(l,m) = V_{n-1}(l,m+1) - V_{n-1}(l,m). \quad (4)$$

We restrict our discussion to the case in which the gauge field can be represented by the following expressions:

$$A_n^l(l,m) = \ln(g_n(l+1,m)/g_n(l,m)), \quad (5a)$$

$$A_n^m(l,m) = \ln(g_n(l,m+1)/g_n(l,m)). \quad (5b)$$

Now we introduce the duality equations by

$$\nabla_+ f_n(l,m) = c_+(g_{n-1}(l,m)/g_n(l+1,m)) \times f_{n+1}(l+1,m), \quad (6a)$$

$$\nabla_- f_n(l,m) = c_-(g_n(l,m)/g_{n-1}(l,m+1)) \times f_{n-1}(l,m+1), \quad (6b)$$

where $c_\pm$ are arbitrary constants. These equations will be compatible if the commutator $[\nabla_+,\nabla_-]$, calculated in two ways agree with each other. First from definitions (2a) and (2b) we obtain

$$\nabla_+\nabla_- f_n(l,m) = \frac{g_n(l,m)g_{n-1}(l+1,m)}{g_n(l+1,m)g_{n-1}(l+1,m+1)} f_n(l+1,m+1) - \frac{g_n(l,m)}{g_n(l+1,m)} f_n(l+1,m) - \frac{g_{n-1}(l,m)}{g_{n-1}(l,m+1)} f_n(l,m+1) + f_n(l,m) \quad (7a)$$

and

$$\nabla_-\nabla_+ f_n(l,m) = \frac{g_{n-1}(l,m)g_n(l,m+1)}{g_{n-1}(l,m+1)g_n(l+1,m+1)} f_n(l+1,m+1) - \frac{g_n(l,m)}{g_n(l+1,m)} f_n(l+1,m) - \frac{g_{n-1}(l,m)}{g_{n-1}(l,m+1)} f_n(l,m+1) + f_n(l,m), \quad (7b)$$

which yield

$$[\nabla_+,\nabla_-] f_n(l,m) = \left(\frac{g_n(l,m)g_{n-1}(l+1,m)}{g_n(l+1,m)g_{n-1}(l+1,m+1)} - \frac{g_{n-1}(l,m)g_n(l,m+1)}{g_{n-1}(l,m+1)g_n(l+1,m+1)}\right) \times f_n(l+1,m+1). \quad (8)$$

Another expression of the commutator can be derived from (6a) and (6b) as follows; using Eq. (6b), we can write

$$\nabla_+\nabla_- f_n(l,m) = c_-\nabla_+\left(\frac{g_n(l,m)}{g_{n-1}(l,m+1)} f_{n-1}(l,m+1)\right),$$

which can be rewritten, by using the definition of ∇_+, as follows:

$$\nabla_+\nabla_- f_n(l,m) = c_-\left[\frac{g_n(l,m)}{g_n(l+1,m)}\left(\frac{g_n(l+1,m)}{g_{n-1}(l+1,m+1)} f_{n-1}(l+1,m+1)\right) - \frac{g_n(l,m)}{g_{n-1}(l,m+1)} f_{n-1}(l,m+1)\right]$$
$$= c_-\frac{g_n(l,m)}{g_{n-1}(l,m+1)}\left(\frac{g_{n-1}(l,m+1)}{g_{n-1}(l+1,m+1)} f_{n-1}(l+1,m+1) - f_{n-1}(l,m+1)\right)$$
$$= c_-\frac{g_n(l,m)}{g_{n-1}(l,m+1)}\nabla_+ f_{n-1}(l,m+1).$$

In this form of the expression, Eq. (6a) can be used to obtain

$$\nabla_+\nabla_- f_n(l,m) = c_+c_-\frac{g_n(l,m)g_{n-2}(l,m+1)}{g_{n-1}(l,m+1)g_{n-1}(l+1,m+1)} \times f_n(l+1,m+1). \quad (9a)$$

Similarly, we obtain

$$\nabla_-\nabla_+ f_n(l,m) = c_+c_-\frac{g_{n-1}(l,m)g_{n+1}(l+1,m)}{g_n(l+1,m)g_n(l+1,m+1)} \times f_n(l+1,m+1), \quad (9b)$$

which leads, together with (9a), to

$$[\nabla_+,\nabla_-] f_n(l,m) = c_+c_-\left(\frac{g_n(l,m)g_{n-2}(l,m+1)}{g_{n-1}(l,m+1)g_{n-1}(l+1,m+1)} - \frac{g_{n-1}(l,m)g_{n+1}(l+1,m)}{g_n(l+1,m)g_n(l+1,m+1)}\right) f_n(l+1,m+1). \quad (10)$$

Putting (8) and (10) equal we see that the quantity

$$\frac{g_n(l+1,m)g_n(l,m+1)}{g_n(l,m)g_n(l+1,m+1)} - c_+c_-\frac{g_{n+1}(l+1,m)g_{n-1}(l,m+1)}{g_n(l,m)g_n(l+1,m+1)}$$

does not depend on n. Denoting this constant by $-\beta/\alpha$ we obtain

$$\alpha g_n(l+1,m)g_n(l,m+1)+\beta g_n(l+1,m+1)g_n(l,m)$$
$$+\gamma g_{n+1}(l+1,m)g_{n-1}(l,m+1)=0, \tag{11}$$

where $\gamma=-c_+c_-\alpha$. It will not be difficult to convince oneself that this expression is equivalent to Hirota's equation (1) if one substitutes

$$l=(\lambda+\mu+\nu)/2, \quad m=(\mu-\lambda-\nu)/2, \quad n=\nu$$

into (11). Hence we have obtained Hirota's equation (1) as a compatibility condition of the duality equations (6).

If we form the symmetric combinations $\{\nabla_+,\nabla_-\}$ from (7a) and (7b) and also from (9a) and (9b) and put them equal we obtain the following linear equation of $f_n(l,m)$:

$$\frac{\beta}{\alpha}\frac{g_n(l,m)g_{n-1}(l,m)}{g_n(l+1,m)g_{n-1}(l,m+1)}f_n(l+1,m+1)$$
$$+\frac{g_n(l,m)}{g_n(l+1,m)}f_n(l+1,m)$$
$$+\frac{g_{n-1}(l,m)}{g_{n-1}(l,m+1)}f_n(l,m+1)-f_n(l,m)=0, \tag{12}$$

when the coefficients of this equation are determined by a solution of Eq. (11). The pair of equations (11) and (12) is now equivalent to the pair of the duality equations (6a) and (6b).

The gauge symmetry will be restricted to those satisfying

$$V_{n+1}(l+1,m)-V_n(l+1,m)=V_n(l,m)-V_{n-1}(l,m), \tag{13}$$

if one requires covariance of the duality equations (6a) and (6b).

The most striking feature of the duality equations (6a) and (6b) is that we can rewrite them by exchanging the roles of the fields f and g as follows:

$$\bar{\nabla}_+g_n(l,m)=-c_-(f_n(l,m)/f_{n+1}(l,m-1))$$
$$\times g_{n+1}(l,m-1), \tag{14a}$$
$$\bar{\nabla}_-g_n(l,m)=-c_+(f_{n+1}(l,m)/f_n(l-1,m))$$
$$\times g_{n-1}(l-1,m), \tag{14b}$$

where the new covariant difference operators are defined by

$$\bar{\nabla}_+g_n(l,m)=g_n(l,m)$$
$$-\exp(B^m_{n+1}(l,m-1))g_n(l,m-1), \tag{15a}$$
$$\bar{\nabla}_-g_n(l,m)=g_n(l,m)$$
$$-\exp(B^l_n(l-1,m))g_n(l-1,m), \tag{15b}$$

and

$$B^l_n(l,m)=\ln(f_n(l+1,m)/f_n(l,m)),$$
$$B^m_n(l,m)=\ln(f_n(l,m+1)/f_n(l,m)).$$

There is no need to repeat the same argument again. We see that the field $f_n(l,m)$ in Eqs. (6a) and (6b) itself satisfies Hirota's equation (1) as a compatibility condition of Eqs. (14a) and (14b). Therefore we call Eqs. (6a) and (6b) or equivalently Eqs. (14a) and (14b), the duality equations in the sense that they represent the duality relation between the amplitude field and the gauge field.

III. LINEAR BÄCKLUND TRANSFORMATION

Now an important result emerges from this property. If g_n is a solution to the nonlinear equation (11) and f_n satisfies the linear equation (12), then the duality equations (6a) and (6b) are fulfilled, whereas the compatibility of (14a) and (14b), which are nothing but (6a) and (6b) themselves, requires for f_n to satisfy the same equation as (11) with g_n replaced by f_n. Namely, as we solve the linear equation (12) for $f_n(l,m)$ whose coefficients are given by any solution of Hirota's equation (1), the solution itself satisfies Hirota's equation (1). In this way our duality equations enable us to obtain a tower of solutions of the nonlinear equation (1), just by solving the linear equation (12). This remarkable feature owes much to the characteristic nature of the gauge symmetric equations (6a) and (6b).

The situation might sound somewhat similar to the Bäcklund transformation of the Liouville equation.[8] The complete set of solutions have been known to be given by the solutions of a linear equation in the d'Alembert form in both ordinary Liouville equation[8] and its difference analog.[9] Therefore the Bäcklund transformation of this model connects different equations. In general, however, it has been believed[8] that soliton equations will not be reduced to linear equations, in contrast to our present analysis.

There have been papers[10] about transformations which relate one soliton equation to another. These are also called the gauge transformations, but differ from one of ours. In our procedure the gauge transformation described by Eq. (3) connects one solution to another, thus offering a series of solutions of Eq. (1) by the gauge transformations.

We should recall that Hirota's equation (1) not only contains many known integrable equations in their corresponding limits, but all solutions to the KP hierarchy can be shown to satisfy it by the transformation invented by Miwa.[3] Therefore the results shown in this paper must be quite general in integrable systems and the gauge symmetry should play an essential role there. In this connection it is worthwhile to recognize the role played by the duality, or antiduality, relation between the electric and magnetic fields in the study of non-Abelian gauge theories.[5]

Finally we remark that equations similar to (6a) and (6b) were already discussed by Hirota[2] as Bäcklund transformations. He, however, did not respect the gauge symmetric nature of the equations and also did not derive the linear equation (12). Therefore in his treatment the two equations corresponding to Eqs. (6a) and (6b) must be solved simultaneously in order to obtain a new solution through the transformations.

The set of equations (6a) and (6b) was also used in Ref. 7 in the continuum limits of the variables l and m. There the equations were regarded as recurrence formulas which generate differential equations of second order as well as the two-dimensional Toda lattice.

[1]N. Saitoh, J. Phys. Soc. Jpn. **49**, 409 (1980).
[2]R. Hirota, Hiroshima University Technical Report Nos. A 6, A 9, 1981; *Non-linear Integrable Systems*, edited by M. Jimbo and T. Miwa (World Scientific, Singapore, 1983), p. 17; J. Phys. Soc. Jpn. **50**, 3785 (1981).
[3]T. Miwa, Proc. Jpn. Acad. A **58**, 9 (1982); E. Date, M. Jimbo, and T. Miwa, J. Phys. Soc. Jpn. **51**, 4116, 4125 (1982).
[4]M. Kashiwara and T. Miwa, Proc. Jpn. Acad. A **57**, 342 (1981); E. Date, M. Kashiwara, and T. Miwa, *ibid.* A **57**, 387 (1981); E. Date, M. Jimbo, M. Kashiwara, and T. Miwa, J. Phys. Soc. Jpn. **50**, 3806, 3813 (1981), and RIMS Report No. 358, 1981; M. Sato and Y. Sato, RIMS Report No. 388, 1980, p. 183; No. 414, 1981, p. 181; M. Sato, a talk delivered at the Symposium on Random Systems and Dynamical Systems held at RIMS, Kyoto University, 1981.
[5]G. 'tHooft, Nucl. Phys. B **79**, 276 (1974); A. A. Belavin, A. M. Polyakov, A. S. Schwarz, and Y. Tyupkin, Phys. Lett. B **59**, 85 (1975); M. F. Atiyah and R. S. Ward, Commun. Math. Phys. **55**, 117 (1977); A. N. Leznov and M. V. Saveliev, *ibid.* **74**, 111 (1980); N. Sanchez, *Dynamical Problems in Soliton Systems*, edited by S. Takeno (Springer, Berlin, 1985), p. 134.
[6]N. Saitoh and S. Saito, Phys. Lett. A **19**, 287 (1986).
[7]N. Saitoh, É. I. Takizawa, and S. Takeno, J. Phys. Soc. Jpn. **54**, 4524 (1985); N. Saitoh, *ibid.* **54**, 435 (1986).
[8]See, for example, G. L. Lamb, Jr., *Elements of Soliton Theory* (Wiley, New York, 1980).
[9]R. Hirota, J. Phys. Soc. Jpn. **46**, 312 (1979).
[10]V. E. Zakharov and L. A. Takhtadzhyan, Theor. Math. Phys. **38**, 17 (1979); J. Honerkamp, J. Math. Phys. **22**, 277 (1981); D. V. Chudnovsky and G. V. Chudnovsky, *ibid.* **22**, 2518 (1981); Y. Ishimori, J. Phys. Soc. Jpn. **51**, 3036 (1982); M. Wadati and K. Sogo, *ibid.* **52**, 394 (1983).

Noether-type conservation laws for perfect fluid motions

G. Caviglia
Department of Mathematics, University of Genova, 16132 Genova, Italy

A. Morro
Department of Biophysical and Electronic Engineering, University of Genova, 16145 Genova, Italy

(Received 7 July 1986; accepted for publication 26 November 1986)

A general approach is developed for the derivation of conservation laws in continuum physics. A Noether-type theorem is applied in connection with transformations which leave the action functional invariant to within the integral of a divergence. Specific results are derived in the case of fluid dynamics: the pertinent equations are considered within the Lagrangian (material) description and are associated with a genuine variation formulation. The physical meaning of the conservation laws is emphasized and the greater generality of the approach is commented upon.

I. INTRODUCTION

Conservation laws constitute a basic tool in the analysis of systems of partial differential equations. They can be exploited to gain information on the properties of the solutions, to determine physical quantities which are constant in time, and to make such theoretical constructions as, for example, the derivation of the geodesic law of motion from the Einstein field equations. In continuum physics the structure of the governing differential equations and the associated procedure for determining conservation laws markedly depends on whether the Lagrangian (material) or the Eulerian (spatial) description is adopted. It turns out that the Lagrangian description is especially fit for the derivation of conservation laws. In this paper this feature will be emphasized just in the case of fluid dynamics where the Lagrangian description is by far less usual than the Eulerian one.

Our approach rests on the fundamental idea of Noether's theorem whereby conservation laws are related to invariance properties of the given functional. Specifically, the explicit determination of the infinitesimal symmetry transformations that leave the action functional invariant depends crucially on the fact that the Lagrangian (density) is defined up to a divergence. Moreover, the invariance is required to hold on the solutions to the field equations. On the basis of these arguments, which trace back to Bessel-Hagen,[1] we construct a very efficient algorithm leading to the generation of rather general conservation laws (Sec. II).

To illustrate the procedure through a specific application, this paper deals with conservation laws for three-dimensional isoentropic fluid flows. As a matter of fact, conservation laws for such fluid motions have been extensively investigated within the framework of the Eulerian description, namely by having recourse to Hamiltonian formulations[2] or to the concept of symmetry transformation and its connection with the generation of conserved currents.[3-5] In addition, one-dimensional flows have also been examined by using either the geometric methods of Estabrook and Wahlquist or other approaches (cf. Refs. 6 and 7). In our investigation we benefit from a recent variational formulation for isoentropic fluid motions[8] which involves the Lagrangian description (Sec. III).

Although the Lagrangian description might seem quite unnatural in fluid dynamics, it is possible to find an infinite set of independent conservation laws (Sec. IV) thus obtaining a proper extension of known results about the corresponding Eulerian description. In particular, it turns out that the conservation laws involve arbitrary functions of the Lagrangian coordinates. Another unexpected result is that the infinitesimal generators of the invariance transformations so obtained—referred to as divergence symmetries by Olver[9]—exhaust the class of symmetry generators for the field equations (Sec. V).

II. NOETHER'S THEOREM AND CONSERVATION LAWS

Consider a physical system described by a Lagrangian (density)

$$L = L(X_\Lambda, \phi_\alpha, \phi_{\alpha,\Lambda}),$$

where X_Λ stands for the independent variables describing a region V in space-time, ϕ_α denotes the unknown fields while $\phi_{\alpha,\Lambda} = \partial\phi_\alpha/\partial X_\Lambda$. Capital (lowercase) Greek letters run over the set of independent variables (unknown fields). Noether-type theorems, and the associated conservation laws, are based on the invariance of the action functional

$$f(\phi_\alpha) = \int_V L(X_\Lambda, \phi_\alpha, \phi_{\alpha,\Lambda})\, dV$$

under suitable transformations. By analogy with the usual procedure, the transformations are taken as

$$\begin{aligned} \overline{X_\Lambda} &= X_\Lambda + \epsilon h_\Lambda, \\ \overline{\phi_\alpha} &= \phi_\alpha + \epsilon \xi_\alpha, \\ \overline{\phi_{\alpha,\Lambda}} &= \phi_{\alpha,\Lambda} + \epsilon(D_\Lambda \xi_\alpha - \phi_{\alpha,\Sigma} D_\Lambda h_\Sigma), \end{aligned} \tag{2.1}$$

with h_Λ and ξ_α as functions of X_Λ and ϕ_α while ϵ is a parameter; here the summation convention is in force and

$$D_\Lambda = \frac{\partial}{\partial X_\Lambda} + \phi_{\alpha,\Lambda} \frac{\partial}{\partial \phi_\alpha} + \phi_{\alpha,\Lambda\Sigma} \frac{\partial}{\partial \phi_{\alpha,\Sigma}}.$$

The last term is inserted in the definition of D_Λ for later convenience.

As is well known, an expression of divergence type, $D_\Sigma J_\Sigma$, can be added to a Lagrangian without affecting the

 0022-2488/87/051056-05$02.50

Euler-Lagrange equations.[10] Based on this observation, we consider the invariance property of f to within the integral of a divergence field.[1,9] It is convenient to express the invariance in the form

$$\int_V L(\overline{X_\Lambda}, \overline{\phi_\alpha}, \overline{\phi_{\alpha,\Lambda}})(1+\epsilon D_\Sigma h_\Sigma)dV - \int_V L(X_\Lambda,\phi_\alpha,\phi_{\alpha,\Lambda})dV = \epsilon\int_V D_\Sigma J_\Sigma\, dV, \tag{2.2}$$

where J_Σ denotes a set of functions which are allowed to depend on X_Λ and ϕ_α. To within first-order terms in the parameter ϵ, the condition (2.2) yields

$$LD_\Lambda h_\Lambda + \frac{\partial L}{\partial X_\Lambda} h_\Lambda + \frac{\partial L}{\partial \phi_\alpha}\xi_\alpha + \frac{\partial L}{\partial \phi_{\alpha,\Lambda}}(D_\Lambda \xi_\alpha - \phi_{\alpha,\Sigma} D_\Lambda h_\Sigma) = D_\Sigma J_\Sigma. \tag{2.3}$$

Letting $\eta_\alpha = \xi_\alpha - \phi_{\alpha,\Sigma} h_\Sigma$, the relation (2.3) can be given the form

$$\eta_\alpha \frac{\partial L}{\partial \phi_\alpha} + (D_\Lambda \eta_\alpha)\frac{\partial L}{\partial \phi_{\alpha,\Lambda}} + D_\Lambda(Lh_\Lambda) = D_\Lambda J_\Lambda, \tag{2.4}$$

whence

$$\eta_\alpha\left(\frac{\partial L}{\partial \phi_\alpha} - D_\Lambda \frac{\partial L}{\partial \phi_{\alpha,\Lambda}}\right) + D_\Lambda\left(h_\Lambda L + \eta_\alpha \frac{\partial L}{\partial \phi_{\alpha,\Lambda}} - J_\Lambda\right) = 0.$$

The unknown fields ϕ_α are taken to satisfy the Euler–Lagrange equations

$$\frac{\partial L}{\partial \phi_\alpha} - D_\Lambda \frac{\partial L}{\partial \phi_{\alpha,\Lambda}} = 0. \tag{2.5}$$

Accordingly, along the solutions ϕ_α to (2.5),

$$D_\Lambda\left(h_\Lambda L + \eta_\alpha \frac{\partial L}{\partial \phi_{\alpha,\Lambda}} - J_\Lambda\right) = 0, \tag{2.6}$$

and then the vector components

$$I_\Lambda = h_\Lambda L + \eta_\alpha \frac{\partial L}{\partial \phi_{\alpha,\Lambda}} - J_\Lambda \tag{2.7}$$

are divergence-free.

In conclusion, once we know the transformation (2.1) leaving the functional f invariant in the sense of (2.2), the quantities I_Λ enter a conservation law. That is why the set of functions h_Λ, ξ_α are said to represent a divergence symmetry.[9] Operatively, the functions h_Λ and ξ_α are not given *a priori* and then the determination of conserved quantities ultimately results in the determination of the functions h_Λ, ξ_α (and then η_α), and J_Λ which satisfy (2.3). Accordingly, we may regard h_Λ, η_α, and J_Λ as functions satisfying Eq. (2.4), and otherwise arbitrary, while the fields ϕ_α are solutions to Eq. (2.5). The search for h_Λ, η_α, and J_Λ is made (relatively) simple by the circumstance that Eq. (2.4) is linear in h_Λ and η_α and by the occurrence of the arbitrary functions J_Λ. Of course, if the J_Λ's are taken to vanish we get standard formulations of Noether's theorem.[11]

Incidentally, a strictly analogous procedure holds when the functions h_Λ, ξ_α, and J_Λ are allowed to depend also on derivatives of the fields ϕ_α; this case needs only a suitable generalization of D_Λ. In this regard we mention the general property that[9] each conservation law is equivalent to one satisfying (2.6) and hence yields a corresponding symmetry.

III. A VARIATIONAL PRINCIPLE FOR FLUID DYNAMICS

Consider a perfect fluid whose motion is described in terms of the time t and the Lagrangian (Cartesian) coordinates $\mathbf{X} = (X_1,X_2,X_3)\in\mathscr{R}$, $\mathscr{R}$ being a suitable reference configuration.[12] Letting $\mathbf{x} = (x_1,x_2,x_3)$, the one-parameter family of diffeomorphisms $\mathbf{x} = \mathbf{x}(\mathbf{X},t)$ gives the position vector of the particle $\mathbf{X}$ at time t. Throughout we adopt the Lagrangian description whereby the fields under consideration are taken to depend on $\mathbf{X}$ and t. In terms of the function $\mathbf{x}(\mathbf{X},t)$ we may define the matrix $x_{aA} = \partial x_a/\partial X_A$ $(a,A = 1,2,3)$, its determinant $J = \det(x_{aA})$, and its inverse $X_{Aa} = \partial X_A/\partial x_a$. Let $\rho(\rho_0)$ be the mass density in the present (reference) configuration and p the pressure. The equation of motion and the continuity equation in the Lagrangian description are written as

$$\rho x_{a,tt} + X_{Ma}\, p_{,M} = 0, \tag{3.1}$$

$$J\rho - \rho_0 = 0. \tag{3.2}$$

As before a comma denotes partial differentiation, namely $f_{,t} = \partial f(\mathbf{X},t)/\partial t$, $f_{,M} = \partial f(\mathbf{X},t)/\partial X_M$. To Eqs. (3.1) and (3.2) we should add the energy equation. Since we are dealing with perfect fluids the energy equation may be expressed through the conservation of the specific entropy S, namely $S(\mathbf{X},t) = S_0(\mathbf{X})$.

Concerning the system (3.1) and (3.2) we observe that $\mathbf{x}(\mathbf{X},t)$ and $\rho(\mathbf{X},t)$ are the unknown functions, whereas $\rho_0(\mathbf{X})$ and $S_0(\mathbf{X})$ are given initial data. The pressure p is taken to be related to the internal energy $E(\rho)$ by

$$p = \rho^2 \frac{dE}{d\rho}; \tag{3.3}$$

the possible dependence of E on S is disregarded because S is assumed to be constant.

In Ref. 8 an application of the techniques pertaining to the inverse problem of the calculus of variations led to some variational formulations for fluid dynamics. For the purpose we have in mind we search for a formulation with $\mathbf{x}$ and ρ as unknown functions. In such a case the pertinent equations are just (3.1) and (3.2), and they are shown to follow from the Euler–Lagrange equations associated with the Lagrangian (density)

$$L(\mathbf{x},\rho) = \tfrac{1}{2}\rho_0 x_{a,t}x_{a,t} - \rho_0 E(\rho) + (J-\rho_0/\rho)\,p(\rho). \tag{3.4}$$

IV. NOETHER-TYPE CONSERVATION LAWS FOR PERFECT FLUIDS

The general method exhibited in Sec. II for the derivation of conserved quantities and the existence of a variational formulation, as described in Sec. III, allows us to determine conserved quantities for perfect fluids. In this regard we identify the independent variables X_Λ with the Cartesian coordinates X_1, X_2, X_3, and the time t. Moreover we let ϕ_α represent the Cartesian coordinates x_1, x_2, x_3, and the mass density ρ. Now, in connection with the Lagrangian (3.4) we have

$$\frac{\partial L}{\partial X_A} = \rho_{0,A}\left(\frac{1}{2}x_{a,t}x_{a,t} - E - \frac{p}{\rho}\right),$$

$$\frac{\partial L}{\partial \rho} \doteq 0, \quad \frac{\partial L}{\partial x_{aA}} = pJX_{Aa}, \quad \frac{\partial L}{\partial x_{a,t}} = \rho_0 x_{a,t},$$

the symbol $\doteq$ denoting the equality along the solution. Then the condition (2.3) becomes

$$\begin{aligned} &h_A\rho_{0,A}(\tfrac{1}{2}x_{a,t}x_{a,t} - E - p/\rho) \\ &\quad + pJX_{Aa}(D_A\xi_a - \phi_{a,\Sigma}D_A h_\Sigma) + \rho_0 x_{a,t}(D_t\xi_a \\ &\quad - \phi_{a,\Sigma}D_t h_\Sigma) + (\tfrac{1}{2}\rho_0 x_{a,t}x_{a,t} - \rho_0 E)(D_t h_t + D_A h_A) \\ &\quad - D_t J_t - D_A J_A \doteq 0. \end{aligned} \tag{4.1}$$

As a starting point choose h_Σ, ξ_α, and J_Λ as

$$\begin{aligned} h_\Sigma &= h_\Sigma(\mathbf{X},t,\mathbf{x}(\mathbf{X},t),\rho(\mathbf{X},t)), \\ \xi_\alpha &= \xi_\alpha(\mathbf{X},t,\mathbf{x}(\mathbf{X},t),\rho(\mathbf{X},t)), \\ J_\Lambda &= J_\Lambda(\mathbf{X},t,\mathbf{x}(\mathbf{X},t),\rho(\mathbf{X},t)). \end{aligned} \tag{4.2}$$

Then, for example,

$$D_t h_\Sigma = h_{\Sigma,t} + \frac{\partial h_\Sigma}{\partial x_a}x_{a,t} + \frac{\partial h_\Sigma}{\partial \rho}\rho_{,t}.$$

Upon substitution, the identical validity with respect to the derivatives $\rho_{,A}, \rho_{,t}, x_{aA}, x_{a,t}$ implies that Eqs. (4.2) reduce to

$$\begin{aligned} &h_A = h_A(\mathbf{X}), \quad h_t = \text{const}, \\ &\xi_a = \xi_a(t,\mathbf{x}), \quad \xi_t = \xi_t(\mathbf{X},t,\mathbf{x},\rho), \\ &J_A = J_A(\mathbf{X},t), \quad J_t = J_t(\mathbf{X},t,\mathbf{x}), \end{aligned}$$

while

$$\rho_0\xi_{a,t} - \frac{\partial J_t}{\partial x_a} = 0, \tag{4.3}$$

$$J_{A,A} + J_{t,t} = 0, \tag{4.4}$$

$$(\rho_0 h_A)_{,A} = 0, \tag{4.5}$$

$$\frac{\partial \xi_a}{\partial x_b} + \frac{\partial \xi_b}{\partial x_a} = 0. \tag{4.6}$$

Further information on h_Σ, ξ_α, and J_Λ arises from the integration of the differential equations (4.3)–(4.6). Equation (4.6) is a Killing-type equation in a flat space; its general solution is (cf. Ref. 12, § 84),

$$\xi_a = \omega_{ab}x_b + \zeta_a, \tag{4.7}$$

where ζ_a and ω_{ab} depend only upon t and the matrix ω_{ab} is skew symmetric. Equation (4.5) implies that there exists a triple w_C, dependent on $\mathbf{X}$, such that

$$\rho_0 h_A = \epsilon_{ABC}w_{C,B}, \tag{4.8}$$

with ϵ_{ABC} the alternating tensor. In view of (4.7) it follows from (4.3) that ω_{ab} is in fact independent of t. Then the application of (4.4) shows that ζ_a must depend linearly on t, namely $\zeta_a = c_a t + d_a$. Hence to within an inessential additive function of $\mathbf{X}$ and t, we have

$$J_t = \rho_0 c_a x_a. \tag{4.9}$$

Meanwhile Eq. (4.4) shows that J_A vanishes to within an inessential function of $\mathbf{X}$ and t. In conclusion, specific conservation laws follow from (2.6) and (2.7) by specializing the parameters ω_{ab}, c_a, d_a, and h_t. This can be seen as follows.

Consider first the transformation with $h_t \neq 0$ while the other parameters vanish. It is

$$\xi_a = 0, \quad \eta_\alpha = -\phi_{\alpha,t}h_t. \tag{4.10}$$

Accordingly, (3.4) yields

$$I_t = -h_t(\tfrac{1}{2}\rho_0 x_{a,t}x_{a,t} + \rho_0 E), \quad I_A = -h_t(x_{a,t}X_{Aa}Jp). \tag{4.11}$$

The transformation law for surface elements, from the reference configuration to the present configuration (see, e.g., Ref. 13), shows that $x_{a,t}X_{Aa}Jp$ corresponds to pv_a in the Eulerian description. Thus the conservation law

$$I_{t,t} + I_{A,A} = 0$$

is in fact the balance of energy.

Second, assume that $d_a \neq 0$ while the other parameters vanish. Since $\eta_a = d_a$ we get

$$I_t = d_a\rho_0 x_{a,t}, \quad I_A = d_a JX_{Aa}p.$$

The conservation law, the identity $(JX_{Aa})_{,A} = 0$ and the arbitrariness of d_a, $a = 1,2,3$, yield the usual balance law for linear momentum.

Third, let $c_a \neq 0$ while the other parameters vanish. Because of (4.7) and (4.9) we have

$$\eta_a = \xi_a = c_a t, \quad J_t = \rho_0 c_a x_a.$$

Then the definition (2.7) gives

$$I_t = c_a\rho_0(tx_{a,t} - x_a), \quad I_A = c_a tJX_{Aa}p.$$

The arbitrariness of c_a leads to the three conservation laws

$$\rho_0(tx_{a,t} - x_a)_{,t} + (tJx_{Aa}p)_{,A} = 0, \quad a = 1,2,3. \tag{4.12}$$

Hence

$$t(\rho_0 x_{a,tt} + JX_{Aa}p_{,A}) = 0,$$

namely t times the components of the equation of motion.

Fourth, let $\omega_{ab} \neq 0$ while the other parameters vanish. It is

$$\eta_a = \xi_a = \omega_{ab}x_b.$$

Then we have

$$I_t = \omega_{ab}\rho_0 x_b x_{a,t}, \quad I_A = \omega_{ab}x_b JX_{Aa}p,$$

and hence the conservation law represents the balance of angular momentum.

Fifth, consider the h_A's as the only nonvanishing parameters. Accordingly, it is

$$\eta_a = -x_{aA}h_A.$$

Thus

$$I_t = -h_B\rho_0 x_{aB}x_{a,t}, \quad I_A = h_A(L - pJ). \tag{4.13}$$

Hence it follows the conservation law

$$\rho_0 h_B(x_{aB}x_{a,t})_{,t} - D_A I_A = 0. \tag{4.14}$$

V. REMARKS ABOUT THE NEW CONSERVATION LAWS

The conservation laws (4.12) and (4.14) are formally and conceptually new in fluid dynamics. It is then worth emphasizing the mathematical origin and the physical significance of these laws.[14] Mathematically, the derivation of the law (4.12) is strictly related to the occurrence of the

divergence term $D_\Sigma J_\Sigma$ in the invariance condition (2.2). Indeed, it involves the time component J_t of J_Σ and hence it cannot be obtained if the presence of the term $D_\Sigma J_\Sigma$ is not allowed.

The derivation of the law (4.14) depends crucially on the arbitrary functions h_A expressing the variations of the spatial domain (material coordinates X_A). Moreover, $\rho_0 h_A$ is divergence-free [cf. (4.8)], which corresponds to the constraint of mass conservation for the coordinate transformation in $\mathscr{R}$. This geometrical property preserves the nature of the continuum; it is neatly formulated in the Lagrangian description but cannot arise out in the Eulerian description because the coordinates x_a are affected by the motion of the continuum. That is why the analog of the law (4.14) cannot appear in the Eulerian description.

The physical significance becomes more suggestive by considering the actual configuration. Let V be any region in $\mathscr{R}$ and $\mathscr{V}$ its image in the actual configuration. Upon use of the transformation law for surface elements[13] we can write the integral form of (4.12) as

$$\frac{d}{dt}\int_{\mathscr{V}} \rho_0(t,\mathbf{x}_{,t}-\mathbf{x})\,d\mathscr{v} = -t\int_{\partial\mathscr{V}} p\mathbf{n}\,d\mathscr{a},$$

$\mathbf{n}$ being the outward unit normal to $\partial\mathscr{V}$. Letting $\bar{\mathbf{x}}$ be the center of mass and m the mass of the region $\mathscr{V}$, upon a trivial integration and a comparison with the balance law for linear momentum we arrive at

$$\bar{\mathbf{x}}(t) = \bar{\mathbf{x}}(0) + t\bar{\mathbf{x}}_{,t}(0) - \frac{1}{m}\int_0^t\int_0^\tau \left(\int_{\partial\mathscr{V}} p\mathbf{n}\,d\mathscr{a}\right)(\theta)\,d\theta\,d\tau; \tag{5.1}$$

as we expect, in case the net force due to the pressure field vanishes, (5.1) makes the center of mass undergo a uniform motion. It is therefore appropriate to view the result (5.1), and hence (4.12), as the center-of-mass theorem. Incidentally, the conservation laws (4.12) constitute the continuum counterpart of a result derived by Hill[15] for a system of N particles.

Denote now by $\mathbf{y} = \mathbf{x}_{,B}h_B$ ($= -\eta$) the displacement induced by h_B in the actual configuration through the motion of the continuum ($\mathbf{x}_{,B}$). The density in (4.14) can be expressed in the integral form as

$$\int_{\mathscr{V}} \rho\mathbf{y}\cdot\mathbf{x}_{,t}\,d\mathscr{v}. \tag{5.2}$$

The condition (5.2) with $\mathbf{y}$ taken as constant constitutes the projection of the balance law for linear momentum in the direction induced by h_B. Usually, however, $\mathbf{y}$ is not constant and then (5.2) may be viewed as a weighted balance law for linear momentum, the weight being just the vector field $\mathbf{y}$.

VI. COMMENTS AND CONCLUSIONS

The present derivation of conservation laws for perfect fluid motions is based on the determination of divergence symmetries through integration of Eq. (4.1) and subsequent application of Noether's theorem. In this regard we think that the present approach can be very fruitful, especially because each conservation law can be related to a divergence symmetry, possibly dependent on the derivatives of the field functions.[9] Differently from the current applications of divergence symmetry transformations, we have regarded the quantities J_Λ as unknown functions on the same footing as the symmetry generators, which means that the integration of the equations arising in the analysis of (4.1) has to be performed without any reference to the original meaning of the variables involved. To our minds, our approach constitutes an improvement of the procedure set up by Olver[9] in that it does fully exploit the facilities associated with the existence of a variational formulation. In particular, our approach does not require the knowledge of the symmetries of the field equations, whose determination involves a lot of hard manipulations on a complicated system of partial differential equations.[3–5,9]

Further—perhaps independent—conservation laws could be generated by looking for solutions to (4.1) which depend on higher-order derivatives. Of course, the algorithm becomes more and more laborious, and the derivatives of (2.1) and (2.2) are also to be taken into account. In this regard an alternative simpler procedure exists which is based on the observation that every divergence symmetry is a symmetry transformation for the field equations[9] (2.1) and (2.2). We can construct conservation laws involving arbitrary functions and higher-order derivatives of the x_a's, e.g., by "deforming"[3,4] the field (4.13) along the "direction" (4.11). In addition, the local formulation (2.2) of the principle of conservation of mass is easily recovered by deformation of the momentum density, so that all general laws of continuum physics are embodied in the present formulation.

In principle, new conservation laws could also be determined by deformation of a given one along the direction of a symmetry, of the equations of motion, different from a divergence symmetry. However, rather long and involved calculations show that when h_Σ and ξ_α are allowed to depend on $\mathbf{X}$, t, $\mathbf{x}$, and ρ the generators of symmetry transformations for the system (2.1) and (2.2) coincide with the set of divergence symmetries described in Sec. IV. Therefore the analysis based on invariance properties of the field equations does not add anything new.

The algorithm based on deformation procedures is not exhaustive. Specifically, it does not give rise, e.g., to the so-called "total helicity integral," which is connected to a conservation law holding in the Eulerian formulation.[2,16] This result, however, is recovered in the present scheme as follows. A direct substitution shows that (4.1) is identically satisfied provided we set

$$\xi_a = h_t = J_t = J_A = 0, \quad h_A = \omega_c X_{Ac}/\rho,$$

where ω_c is simply the c component of curl $\mathbf{x}_{,t}$, that is

$$\omega_c = \epsilon_{cab}x_{b,tH}X_{Ha}.$$

In view of (2.7), the corresponding conserved density turns out to be

$$I_t = -Jx_{c,t}\omega_c,$$

which gives rise to the total helicity integral. Thus, in view of the previous results, we conclude that the conservation laws associated with the Eulerian description of perfect fluid motion[2,3,16] constitute a subset of those associated with the Lagrangian description.

With a view to the application of the present approach to other contexts, it is worth emphasizing the main aspects of our derivation of conservation laws for fluid dynamics. Although it might appear that the Eulerian description is more natural than the Lagrangian one in fluid dynamics, the results of this paper give evidence of the importance of the Lagrangian description. Indeed, in comparison with analogous investigations concerning the Eulerian description,[2,3,16] the Lagrangian description allows the elaboration of a more systematic approach and leads to a wider set of conservation laws. Essentially, this is due to the fact that the Lagrangian description provides a clear distinction between unknown fields and independent variables.

In this conjunction, we mention that the Eulerian description has been proved not to allow the existence of arbitrary functions (like h_A), thus restricting the set of conservation laws for both compressible and incompressible fluids. This constitutes a further advantage of the Lagrangian description.

The existence of a variational formulation is not crucial in determining conservation laws; the same results could be attained by considering the symmetry transformations of the given system of differential equations and using the algorithms described in Refs. 3 and 4. However, the knowledge of the Lagrangian density makes the derivation mathematically simpler and the results more suggestive as to the immediate physical meaning of the conserved quantities.

ACKNOWLEDGMENTS

The research leading to this work has been supported by the Italian Ministry of Education through the 40% Project "Problemi di evoluzione nei fluidi e nei solidi."

[1] E. Bessel-Hagen, Math. Ann. **84**, 258 (1921).
[2] P. J. Olver, J. Math. Anal. Appl. **89**, 233 (1982).
[3] G. Caviglia, "Composite variational principles and the determination of conservation laws," submitted for publication.
[4] G. Caviglia, J. Math. Phys. **27**, 972 (1986).
[5] M. Benati and G. Caviglia, "Conservation laws for 3-dimensional compressible Euler equations," Int. J. Eng. Sci. (in press).
[6] E. S. Suhubi, Int. J. Eng. Sci. **22**, 119 (1984).
[7] J. Veroski, J. Math. Phys. **25**, 884 (1984).
[8] F. Bampi and A. Morro, J. Math. Phys. **23**, 2312 (1982); **25**, 2418 (1984).
[9] P. J. Olver, Arch. Rational Mech. Anal. **85**, 112 (1984); P. J. Olver in *Systems of Nonlinear Partial Differential Equations*, edited by J. M. Ball (Reidel, Dordrecht, 1983).
[10] L. V. Ovsiannikov, *Group Analysis of Differential Equations* (Academic, New York, 1982).
[11] I. M. Gelfand and S. V. Fomin, *Calculus of Variations* (Prentice–Hall, Englewood Cliffs, NJ, 1963); D. Lovelock and H. Rund, *Tensors, Differential Forms and Variational Principles* (Wiley, New York, 1975); A. Trautman, Commun. Math. Phys. **6**, 248 (1967).
[12] C. Truesdell and R. A. Toupin, in *Encyclopedia of Physics*, edited by S. Flügge (Springer, Berlin, 1960), Vol. III/1.
[13] A. C. Eringen, *Mechanics of Continua* (Krieger, Huntington, NY, 1980).
[14] We are indebted to the anonymous referee for drawing our attention to this point.
[15] E. L. Hill, Rev. Mod. Phys. **23**, 253 (1951).
[16] H. K. Moffatt, J. Fluid Mech. **35**, 117 (1969); D. Serre, C. R. Acad. Sci. Paris **289**, 267 (1979).

Factorization of the wave equation in higher dimensions

Vaughan H. Weston
Department of Mathematics, Purdue University, West Lafayette, Indiana 47907

(Received 15 July 1986; accepted for publication 10 December 1986)

The factorization of the wave equation into a coupled system involving up-and down-going wave components is obtained for the case where the field quantities are multivariate functions of spatial variables, but the velocity c is a function of the z variable only. The form of the reflection operator is derived and the quadratic differential-integral equation satisfied by its kernel is obtained.

I. INTRODUCTION

One of the techniques that has been used in time-dependent direct and inverse scattering problems associated with the wave equation (and similar hyperbolic systems) for a nonhomogeneous medium is based upon the method of wave splitting.[1,2]

The splitting of the linear one-dimensional wave equation into up- and down-going waves reduces the wave equation to a coupled first-order system in the up- and down-going components of the fields. Exact local splittings yield uncoupled systems whenever the medium does not vary in the preferred direction. As an example, the splitting for the one-dimensional wave equation

$$\frac{\partial^2}{\partial z^2}u(z,t)=\frac{1}{c^2(z)}\frac{\partial^2}{\partial t^2}u(z,t) \tag{1}$$

is obtained by first rewriting Eq. (1) in the form[2]

$$\frac{\partial}{\partial z}\begin{bmatrix}u\\ u_z\end{bmatrix}=\begin{bmatrix}0 & 1\\ ((1/c)\partial_t)^2 & 0\end{bmatrix}\begin{bmatrix}u\\ u_z\end{bmatrix}, \tag{2}$$

then defining the components

$$\begin{bmatrix}\varphi^+\\ \varphi^-\end{bmatrix}=\frac{1}{2}c^{-1/2}\begin{bmatrix}1 & -((1/c)\partial_t)^{-1}\\ 1 & ((1/c)\partial_t)^{-1}\end{bmatrix}\begin{bmatrix}u\\ u_z\end{bmatrix}, \tag{3}$$

where

$$\partial_t^{-1}v=\int_0^t v(z,\tau)d\tau.$$

As expressed in terms of the components φ^+ and φ^-, system (2) takes the form

$$\frac{\partial}{\partial z}\begin{bmatrix}\varphi^+\\ \varphi^-\end{bmatrix}=\begin{bmatrix}-\frac{1}{c}\partial_t & -\frac{c_z}{2c}\\ -\frac{c_z}{2c} & \frac{1}{c}\partial_t\end{bmatrix}\begin{bmatrix}\varphi^+\\ \varphi^-\end{bmatrix}. \tag{4}$$

As is seen, in the case $c=$ const the system decouples and φ^+ and φ^- take the form $\varphi(z,t)=f(z\mp ct)$ of up- and down-going waves.

The analogous splitting in the frequency domain can be used to get the Bremmer series[3] and the parabolic approximation of Leontovich–Fock.

The importance of such splittings, in general, is that they lead to the use of invariant imbedding techniques.[4–7] Given a slab of inhomogeneous medium and a splitting one can define an associated scattering matrix. Invariant imbedding techniques then allow one to write a complex system of differential equations for the operator entries of the scattering matrix whose differentiation is with respect to the location of one of the planes of the slab. One can then deduce the behavior of the reflection operators for small time which provides a connection between up- and down-going wave fields and the properties of the medium on the edge of the slab.[1,2] The reflection operator can then be used in both direct and inverse scattering problems.

For the particular splitting given above, the reflection operator is given by

$$\varphi^-(z,t)=\mathscr{R}\varphi^+=\int_0^t \mathsf{R}(z,t-s)\varphi^+(z,s)ds, \tag{5}$$

with the kernel satisfying the system[2]

$$\frac{\partial}{\partial z}\mathsf{R}(z,t)-\frac{2}{c}\frac{\partial}{\partial t}\mathsf{R}(z,t)-\frac{c_z}{2c}\mathsf{R}*\mathsf{R}=0,$$
$$\mathsf{R}(z,0^+)=\tfrac{1}{4}c_z, \tag{6}$$

where $\mathsf{R}*\mathsf{R}$ is a convolution.

The wave splitting concept and the associated reflection operator have been extensively used for a variety of one-dimensional inverse problems, among these the electromagnetic inverse problem for dispersive media,[8,9] the inverse problem for viscoelastic media,[10] and the inverse problem for elastic media with oblique incidence.[11]

A formal attempt to apply the wave splitting to the multidimensional wave equation is given by Corones and Krueger[12] and Davison.[2] Here the difficulty is to diagonalize a matrix with operator entries and/or variable coefficients, and to correctly identify the appropriate plus and minus components (up- and down-going waves).

However, Fishman and McCoy[13] were successful in factorizing the reduced Helmholtz equation for the transversely inhomogeneous half-space, but their method gave rise to pseudodifferential equations. From this, though, they were able to develop a systematic derivation of the approximate extended parabolic wave theories. Another attempt to enlarge the well-known factorization of the one-dimensional case to three dimensions is given by Yagle and Levy.[14]

In this paper, we examine the up- and down-going wave condition (determination of the plus and minus quantities) for the three-dimensional case. In Sec. II, Huygen's principle is essentially used to define up- and down-going waves in a homogeneous medium. This leads to the introduction of an operator $\mathscr{K}$, and a linear relationship between u and $\partial u/\partial z$

 0022-2488/87/051061-08$02.50

on a plane $z = \text{const}$ involving the operator $\mathscr{K}$. With appropriate sign taken, this linear relation specifies whether the wave is up or down going on the plane $z = \text{const}$. Using this, the wave in a homogeneous slab can be decomposed into up- and down-going wave components.

In Sec. III the up- and down-going wave correlation and the associated operator $\mathscr{K}$ are generalized to the case where c is a continuous function of z. This is used in Sec. IV to generalize the wave decomposition and system given by Eq. (4) to the case where $u = u(x,y,z,t)$. The formal generalization of the reflection operator $\mathscr{R}$ is given in Sec. V together with the corresponding integral-differential equation it must satisfy.

II. CONDITION FOR UP-GOING AND DOWN-GOING WAVES (c CONSTANT)

With $u(x,y,z,t)$ being a solution of the wave equation with constant velocity $c = c_0$, a relationship will be established here between u and u_z on a plane $z = \text{const}$, to indicate whether the wave is up-going (propagating in the positive z direction) or down-going (propagating in the negative z direction). For simplicity, the plane will be taken to the coordinate plane $z = 0$. To be precise, up-going waves at $z = 0$ will be defined as those generated by sources in the region $z < 0$, with the half space $z \geqslant 0$, being source-free. Correspondingly, down-going waves at $z = 0$ are generated by sources in the region $z > 0$. A relationship between u and u_z on the plane $z = 0$ for up-going and down-going waves can now be established by considering the appropriate mixed (initial and boundary-value) problem. The first of these is given by the following lemma.

Note that the notation $U(x,y;\sigma)$ is used to represent the disk with center (x,y) and radius σ.

Lemma 1: The solution of the mixed problem:

$$\text{(i)}\ \frac{1}{c_0^2}\frac{\partial^2 u}{\partial t^2} - \nabla^2 u = 0, \quad t > 0, \quad z > 0,$$

$$\text{(ii)}\ u(x,y,z;0) = u_t(x,y,z;0) = 0, \quad z \geqslant 0,$$

$$\text{(iii)}\ \frac{\partial u}{\partial z}(x,y,z;t)\big|_{z=0} = v(x,y,t), \quad t \geqslant 0,$$

where v is Holder continuous in $\mathbb{R}^2 \times [0,\infty)$ with $v(x,y,0) = v_t(x,y,0) = 0$, is given by

$$u(x,y,z,t) = -\frac{1}{2\pi}\iint_{U(x,y;\sigma)} \frac{v(x',y',t - r/c_0)}{r}\,dx'\,dy', \tag{7}$$

where

$$\sigma = (c_0^2 t^2 - z^2)^{1/2} \tag{8}$$

and

$$r^2 = (x - x')^2 + (y - y')^2 + z^2. \tag{9}$$

For $z > c_0 t$ the solution vanishes.

Proof: This representation in the time-dependent formulation of the single-layer potential formulation is easily verified. Since for fixed x' and y', $v(x',y';t - r/c_0)/r$ satisifies the wave equation at all points (x,y,z), where $r \neq 0$, it follows that expression (7) satisfies the wave equation. The initial conditions are easily established using the condition that $v = 0$ at $t = 0$. The boundary condition is established by using the jump condition of the normal derivative of the single layer potential.[15] ■

The condition for up-going waves on the plane $z = 0$ can be easily obtained by replacing $v(x,y,t)$ in Lemma 1 by $u_z(x,y,0,t)$, and taking the limit of expression (7) as $z \to 0+$. This gives the condition for up-going waves

$$u(x,y,0,t) = -\frac{1}{2\pi}\iint_{U(x,y;c_0 t)} \frac{1}{R}\, u_z\left(x',y',0;t - \frac{R}{c_0}\right) dx'\,dy', \tag{10}$$

where

$$R^2 = (x - x')^2 + (y - y')^2. \tag{11}$$

The condition for down-going waves can be obtained in a similar manner by employing the corresponding mixed initial-value, boundary-value problem for the half-space $z \leqslant 0$. The resulting condition is same as that given by Eq. (10) except for a difference of sign of the term of the right-hand side.

The results can be summarized in the following Lemma.

Lemma 2: The up-going and down-going wave condition on the plane $z = 0$ is given by

$$u = \pm \mathscr{K}_0 u_z, \tag{12}$$

where the operator $\mathscr{K}_0$ is defined by

$$\mathscr{K}_0 v = -\frac{1}{2\pi}\iint_{U(x,y;c_0 t)} \frac{v(x',y',t - R/c_0)}{R}\,dx'\,dy', \tag{13}$$

and the plus and minus signs in Eq. (13) refer to up-going and down-going waves, respectively.

We want to show next that condition (12) can be expressed in the form $\mathscr{K}_0^{-1} u = \pm u_z$. To show this the following lemma is needed.

Lemma 3: The solution of the mixed problems,

$$\text{(i)}\ \frac{1}{c_0^2}\frac{\partial^2 u}{\partial t^2} - \nabla^2 u = 0, \quad t > 0, \quad z > 0,$$

$$\text{(ii)}\ u(x,y,z;0) = u_t(x,y,z;0) = 0, \quad z \geqslant 0,$$

$$\text{(iii)}\ u(x,y,0,t) = w(x,y,t), \quad t \geqslant 0,$$

where w and w_t are Holder continuous in $\mathbb{R}^2 \times [0,\infty)$ with $w(x,y,0) = w_t(x,y,0) = w_{tt}(x,y,0) = 0$, is given by

$$u(x,y,z,t) = \frac{1}{2\pi}\iint_{U(x,y;\sigma)} \left\{ w\left(x',y',t - \frac{r}{c_0}\right) + \frac{r}{c_0}\frac{\partial}{\partial t} w\left(x',y',t - \frac{r}{c_0}\right)\right\}\frac{z}{r^3}\,dx'\,dy', \tag{14}$$

with the solution vanishing for $z > c_0 t$, and σ given by Eq. (8).

Proof: This representation in terms of the time-dependent form of the double layer potential is easily verified. Since for fixed x' and y', $(\partial/\partial z)\{w(x',y',t - r/c_0)/r\}$ satisfies the wave equation for $r \neq 0$, the integrand in the expression satisfies the wave equation for $z > 0$. The conditions on v at $t = 0$ ensure that no contribution comes from the limits of the integral when expression (1.4) is inserted in the wave equation. At the same time these conditions ensure that the

initial conditions are satisfied. The boundary condition is established using the jump condition for the double layer potential.[15] ■

Relation (14) can now be used to get the alternative form for the up-going wave. First, $w(x,y,t)$ is replaced by the C^2 function $u(x,y,0,t)$, and then in the resulting expression for $u(x,y,z,t)$, the order of differentiation and integration is changed to yield for $z>0$,

$$u(x,y,z,t) = -\frac{1}{2\pi}\frac{\partial}{\partial z}\iint_{U(x,y;\sigma)} \frac{u(x',y',0;t-r/c_0)}{r}\,dx'\,dy' \tag{15}$$

with $\sigma = (c_0^2 t^2 - z^2)^{1/2}$.

Since it is easily verified that the integral expression on the right-hand side of Eq. (15) is a solution of the wave equation for $z>0$ [using conditions $u(x,y,0,0)=0$, $u_t(x,y,0,0)=0$], it can be shown that for $z>0$,

$$\frac{\partial u}{\partial z}(x,y,z,t) = -\Box_0 \frac{1}{2\pi}\iint_{U(x,y;\sigma)} \frac{u(x',y',0;t-r/c_0)}{r}\,dx'\,dy',$$

where

$$\Box_0 = \left(\frac{1}{c_0^2}\frac{\partial^2}{\partial t^2} - \frac{\partial^2}{\partial x^2} - \frac{\partial^2}{\partial y^2}\right). \tag{16}$$

Hence taking the limit as $z\to 0^+$, one obtains

$$\frac{\partial}{\partial z}u(x,y,0,t) = \Box_0 \mathscr{K}_0 u \tag{17}$$

as an alternative condition for up-going waves at $z=0$. It follows from Eqs. (12) and (17) that $\mathscr{K}_0^{-1}$ exists and is given by

$$\mathscr{K}_0^{-1} = \Box_0 \mathscr{K}_0. \tag{18}$$

A similar result to expression (17) for down-going waves can be obtained. The results are summarized as follows.

Lemma 4: The up-going and down-going wave condition on the plane $z=0$ is given by

$$\mathscr{K}_0^{-1} u = \pm u_z, \tag{19}$$

where $\mathscr{K}_0^{-1}$ is given by Eq. (18).

Applying the up-going and down-going wave conditions (12) at any plane $z=$ const other than just the plane $z=0$, we can decompose the solution $u(x,y,z,t)$ of the wave equation in terms of up- and down-going components. Using the following identity:

$$u = \frac{1}{2}\left(u + \mathscr{K}_0\frac{\partial u}{\partial z}\right) + \frac{1}{2}\left(u - \mathscr{K}_0\frac{\partial u}{\partial z}\right),$$

it follows that u can be decomposed into the two components,

$$u(x,y,z,t) = u^+(x,y,z,t) + u^-(x,y,z,t), \tag{20}$$

where

$$u^\pm(x,y,z,t) = \frac{1}{2}\left(u \pm \mathscr{K}_0\frac{\partial u}{\partial z}\right). \tag{21}$$

It is easily seen that $u^+(x,y,z,t)$ represents an up-going wave since

$$\left(u^+ - \mathscr{K}_0\frac{\partial u^+}{\partial z}\right) = \frac{1}{2}\left(u - \mathscr{K}_0^2\frac{\partial^2}{\partial z^2}u\right) = \tfrac{1}{2}\{u - \mathscr{K}_0^2\Box_0 u\} = \tfrac{1}{2}\{u-u\} = 0,$$

which implies that u^+ satisfies the up-going condition (12). Similarly, it can be shown that u^- satisfies the down-going condition (12).

III. DECOMPOSITION INTO UP- AND DOWN-GOING WAVES WHEN $c(z)$ IS A FUNCTION OF z

Here we want to extend the decomposition of u into up- and down-going waves [as given by Eq. (20)] to the case where the velocity c is a piecewise differentiable function of z. The extension will be based upon the physical idea of slicing the medium into a set of infinitesimal slabs of width Δz, in each of which c is constant, then imposing the decomposition given by Eqs. (20) and (21) in each slab. Thus by first modifying the operator $\mathscr{K}_0$ to take into account the variation of $c(z)$ with z by defining

$$\mathscr{K}w = -\frac{1}{2\pi}\iint_{U(x,y;c(z)t)} \frac{w(x',y',t-R/c(z))}{R}\,dx'\,dy', \tag{22}$$

the upward and downward wave components [given by (21)] will now be defined as follows:

$$u^\pm = \frac{1}{2}\left(u \pm \mathscr{K}\frac{\partial u}{\partial z}\right). \tag{23}$$

It is shown in Appendix A that as expected $\mathscr{K}$ has essentially the same properties as $\mathscr{K}_0$, namely, that

$$\mathscr{K}^2 w = \frac{c(z)}{2\pi}\int_0^t \iint_{U(x,y;c(t-s))} \frac{w(x',y',s)}{[c^2(z)(t-s)^2 - (x-x')^2 - (y-y')^2]^{1/2}}\,dx'\,dy'\,ds,$$

which is Poisson's formula[15] for the solution of the wave equation with zero initial conditions. Hence we have

$$\Box \mathscr{K}^2 w = w, \tag{24}$$

where

$$\Box = \frac{1}{c^2(z)}\frac{\partial^2}{\partial t^2} - \frac{\partial^2}{\partial x^2} - \frac{\partial^2}{\partial y^2}. \tag{25}$$

Thus if $w\in C^2(\mathbb{R}^2)\times C^2[0,\infty)$ (corresponding to the classical solution), then Eq. (24) states that $\Box\mathscr{K}$ is the left inverse of $\mathscr{K}$. In order for $\mathscr{K}$ to be a right inverse also, i.e., $\mathscr{K}(\Box\mathscr{K}) = (\Box\mathscr{K})\mathscr{K} = I$, an additional condition on w has to be imposed. It can be shown (by putting the integral expression for $\mathscr{K}w$ in local polar form), that $(\partial^2/\partial t^2)\mathscr{K}(\mathscr{K}w) = \mathscr{K}(\partial^2/\partial t^2)(\mathscr{K}w)$ provided that

$\mathscr{K}w = (\partial/\partial t)\mathscr{K}w = 0$ at time $t = 0$, and this in turn holds provided that $w = 0$ at time $t = 0$. One can then verify that the right-inverse condition holds provided that the domain is restricted to functions that vanish at $t = 0$. Additional results are given in Eq. (27) below. The results are summarized in the following Lemma.

Lemma 5: Let $w \in C^2(\mathbb{R}^2) \times C^2[0,\infty)$, then (i) $\Box\mathscr{K}$ is the left inverse of $\mathscr{K}$,

$$(\Box\mathscr{K})\mathscr{K}w = w, \tag{26}$$

(ii) if $w = 0$ at $t = 0$, then $\Box\mathscr{K}$ is the right inverse of $\mathscr{K}$

$$\mathscr{K}(\Box\mathscr{K})w = w, \tag{26'}$$

(iii) furthermore if $w = w_t = 0$ at $t = 0$, then

$$\mathscr{K}^2\Box w = \Box\mathscr{K}^2 w = w. \tag{27}$$

■

In the one-dimensional case, the operator $\mathscr{K}$ takes the very simple form

$$\mathscr{K}w = -c(z)\int_0^t w(s)ds. \tag{28}$$

An alternative form for $\mathscr{K}^{-1}$ is developed in Appendix B. There it is shown that if $w \in C^2(\mathbb{R}^2) \times C^2[0,\infty)$, then

$$\mathscr{K}^{-1}w = -\frac{1}{c}\frac{\partial}{\partial t}w(x,y,t) + \mathscr{L}w, \tag{29}$$

where the operator $\mathscr{L}$ is given by

$$\mathscr{L}w = \frac{1}{2\pi}\iint_{U(x,y;ct)} \left\{ w_x\left(x',y';t-\frac{R}{c}\right)\frac{(x'-x)}{R^3} + w_y\left(x',y';t-\frac{R}{c}\right)\frac{(y'-y)}{R^3}\right\} dx'\,dy', \tag{30}$$

where R is defined by Eq. (11).

IV. FACTORIZATION OF THE WAVE EQUATION IN A STRATIFIED MEDIUM

The upward and downward wave decomposition will be applied to the factorization of the wave equation for a stratified medium. Using the same initial procedure that was done for the one-dimensional (spatial) case,[1] where u is a function of z and t only, the wave equation will be written in the form

$$\frac{\partial}{\partial z}\begin{bmatrix} u \\ u_z \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ \Box & 0 \end{bmatrix}\begin{bmatrix} u \\ u_z \end{bmatrix}. \tag{31}$$

The upward and downward wave down decomposition as defined by Eq. (23) will be expressed in vector form

$$\begin{bmatrix} u^+ \\ u^- \end{bmatrix} = \mathscr{T}\begin{bmatrix} u \\ u_z \end{bmatrix}, \tag{32}$$

where $\mathscr{T}$ is the matrix operator

$$\mathscr{T} = \frac{1}{2}\begin{bmatrix} 1 & \mathscr{K} \\ 1 & -\mathscr{K} \end{bmatrix} \tag{33}$$

whose inverse is given by

$$\mathscr{T}^{-1} = \begin{bmatrix} 1 & 1 \\ \mathscr{K}^{-1} & -\mathscr{K}^{-1} \end{bmatrix}. \tag{34}$$

With the insertion of the inverse relation

$$\begin{bmatrix} u \\ u_z \end{bmatrix} = \mathscr{T}^{-1}\begin{bmatrix} u^+ \\ u^- \end{bmatrix} \tag{35}$$

into Eq. (31) and the premultiplication of the resulting equation by the matrix operator $\mathscr{T}$, the following is obtained:

$$\frac{\partial}{\partial z}\begin{bmatrix} u^+ \\ u^- \end{bmatrix} = \mathscr{T}\begin{bmatrix} 0 & 1 \\ \Box & 0 \end{bmatrix}\mathscr{T}^{-1}\begin{bmatrix} u^+ \\ u^- \end{bmatrix} - \mathscr{T}\frac{\partial\mathscr{T}^{-1}}{\partial z}\begin{bmatrix} u^+ \\ u^- \end{bmatrix}. \tag{36}$$

Using relations (33) and (34) it can be shown that

$$\mathscr{T}\begin{bmatrix} 0 & 1 \\ \Box & 0 \end{bmatrix}\mathscr{T}^{-1} = \begin{bmatrix} \mathscr{K}^{-1} & 0 \\ 0 & -\mathscr{K}^{-1} \end{bmatrix}, \tag{37}$$

$$\mathscr{T}\frac{\partial\mathscr{T}^{-1}}{\partial z} = \frac{1}{2}\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}\mathscr{K}\frac{\partial\mathscr{K}^{-1}}{\partial z}. \tag{38}$$

Using the identity $\mathscr{K}\mathscr{K}^{-1} = I$ it follows that if w is such that w, w_t vanish at $t = 0$,

$$\mathscr{K}\left(\frac{\partial\mathscr{K}^{-1}}{\partial z}\right)w = -\left(\frac{\partial\mathscr{K}}{\partial z}\right)\mathscr{K}^{-1}w = -\left(\frac{\partial\mathscr{K}}{\partial z}\right)\mathscr{K}\Box w.$$

From Eq. (A9) in the Appendix, this becomes

$$\begin{aligned}\mathscr{K}\left(\frac{\partial\mathscr{K}^{-1}}{\partial z}\right)w &= -\frac{1}{2c}\frac{\partial c}{\partial z}\left\{t\mathscr{K}^2\left(\frac{\partial}{\partial t}\Box w\right) - \mathscr{K}^2\left(t\frac{\partial}{\partial t}\Box w\right)\right\} \\ &= -\frac{1}{2c}\frac{\partial c}{\partial z}\left\{t\mathscr{K}^2\Box\frac{\partial w}{\partial t} - \mathscr{K}^2\left[\Box\left(t\frac{\partial w}{\partial t}\right) - 2\frac{1}{c^2}\frac{\partial^2}{\partial t^2}w\right]\right\} \\ &= -\frac{1}{c}\frac{\partial c}{\partial z}\mathscr{K}^2\left(\frac{1}{c^2}\frac{\partial^2}{\partial t^2}w\right).\end{aligned} \tag{39}$$

But on using the relation

$$\begin{aligned}\mathscr{K}^2\left(\frac{1}{c^2}\frac{\partial^2}{\partial t^2}w\right) &= \mathscr{K}^2\left[\Box w + \left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}\right)w\right] \\ &= w + \left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}\right)\mathscr{K}^2 w,\end{aligned}$$

the resulting expression is obtained

$$\mathscr{K}\left(\frac{\partial\mathscr{K}^{-1}}{\partial z}\right)w = -\frac{1}{c}\frac{\partial c}{\partial z}\left[w + \left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}\right)\mathscr{K}^2 w\right]. \tag{40}$$

System (36) can now be expressed in the explicit form

$$\frac{\partial}{\partial z}\begin{bmatrix} u^+ \\ u^- \end{bmatrix} = \begin{bmatrix} \mathscr{K}^{-1} & 0 \\ 0 & -\mathscr{K}^{-1} \end{bmatrix}\begin{bmatrix} u^+ \\ u^- \end{bmatrix} + \frac{1}{2c}c_z\begin{bmatrix} (1+\Lambda) & -(1+\Lambda) \\ -(1+\Lambda) & (1+\Lambda) \end{bmatrix}\begin{bmatrix} u^+ \\ u^- \end{bmatrix}, \tag{41}$$

where Λ is the operator

$$\Lambda w = \left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}\right)\mathscr{K}^2 w. \tag{42}$$

Some simplification is achieved if one sets

$$u^{\pm} = c^{1/2}\phi^{\pm}, \tag{43}$$

in which case system (41) reduces to

$$\frac{\partial}{\partial z}\begin{bmatrix}\phi^+\\ \phi^-\end{bmatrix}=\begin{bmatrix}\mathscr{K}^{-1} & 0\\ 0 & -\mathscr{K}^{-1}\end{bmatrix}\begin{bmatrix}\phi^+\\ \phi^-\end{bmatrix}$$

$$+\frac{1}{2c}c_z\begin{bmatrix}\Lambda & -(1+\Lambda)\\ -(1+\Lambda) & \Lambda\end{bmatrix}\begin{bmatrix}\phi^+\\ \phi^-\end{bmatrix}. \tag{44}$$

System (44) can now be used to derive the reflection operator as is shown in the next section.

V. REFLECTION OPERATOR

In this section, the reflection operator (relating the down-going wave to the up-going wave) given by Eq. (5) for the one-dimensional case, will be generalized to the multidimensional case. In particular, the equation and initial conditions that are satisfied by the kernel of the reflection operator will be sought. For the multidimensional case, the reflection operator will take the form (the reason for it, will be apparent later)

$$\phi^-(x,y,z,t)=\mathscr{R}\phi^+=(I+\Lambda)\mathsf{R}*\phi^+, \tag{45}$$

where Λ is the operator given by Eq. (42) and $\mathsf{R}*\phi^+$ is the convolution

$$\mathsf{R}*\phi^+=\int_0^t\int_{\mathbb{R}^2}\mathsf{R}(x-x',y-y',z,t-s)\times\phi^+(x',y',z,s)dx'\,dy'\,ds \tag{46}$$

involving the reflection operator kernel $\mathsf{R}(x,y,z,t)$.

As a preliminary a number of identities need to be deduced, among them the following:

$$\frac{\partial}{\partial z}\{(I+\Lambda)\psi\}=(I+\Lambda)\left[\frac{\partial\psi}{\partial z}+\frac{2c_z}{c}\Lambda\psi\right],\quad t>0, \tag{47}$$

where $\psi(x,y,z,t)$ is a sufficiently smooth function such that $\psi=\psi_t=0$ at $t=0$.

To obtain expression (47) the relation [obtained from (27)]

$$\Box\mathscr{K}^2\chi=\chi(x,y,t),\quad t>0,$$

valid for $\chi(x,y,t)$ such that $\chi=\chi_t=0$ at $t=0$, is differentiated with respect to z, giving

$$-2\frac{c_z}{c}\left(\frac{1}{c^2}\frac{\partial^2}{\partial t^2}\mathscr{K}^2\chi\right)+\Box\left(\frac{\partial}{\partial z}\mathscr{K}^2\right)\chi=0.$$

From relations (27) and (42) this reduces to

$$\left(\frac{\partial}{\partial z}\mathscr{K}^2\right)\chi=2\frac{c_z}{c}\mathscr{K}^2(1+\Lambda)\chi.$$

It can now be seen using this relation that

$$\frac{\partial}{\partial z}(\Lambda\psi)=\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}\right)\left(\frac{\partial}{\partial z}\mathscr{K}^2\right)\psi+\Lambda\frac{\partial\psi}{\partial z}$$

$$=\Lambda\left\{\frac{2c_z}{c}(I+\Lambda)\psi+\frac{\partial\psi}{\partial z}\right\}.$$

Hence identity (47) immediately follows.

The next result that is needed is the following lemma.

Lemma 6: The only $C^2(\mathbb{R}^3)\times C^2[0,\infty)$ solution $\psi(x,y,z,t)$ of the system

$$(I+\Lambda)\psi=0,\quad t>0, \tag{48}$$

is the trivial solution $\psi\equiv0$.

Proof: Equation (48) can be written in the form

$$\frac{\partial^2}{\partial t^2}(\mathscr{K}^2\psi)=0.$$

Since $\mathscr{K}^2\psi$, $(\partial/\partial t)\mathscr{K}^2\psi$ vanish at $t=0$, it follows that $\mathscr{K}^2\psi=0$, for $t\geqslant0$. Operating on this with $\Box$, it follows from Eq. (26) that $\psi\equiv0$, $t\geqslant0$.

We will now proceed to get the equation for the reflection operator. System (44), written out explicitly, takes the form

$$\frac{\partial\phi^{+1}}{\partial z}=\left\{\mathscr{K}^{-1}+\frac{c_z}{2c}\Lambda\right\}\phi^+-\frac{c_z}{2c}(I+\Lambda)\phi^-, \tag{49}$$

$$\frac{\partial\phi^-}{\partial z}=\left\{-\mathscr{K}^{-1}+\frac{c_z}{2c}\Lambda\right\}\phi^--\frac{c_z}{2c}(I+\Lambda)\phi^+. \tag{50}$$

It will be assumed that ϕ^+ is a sufficiently smooth function and that $\phi^+=\phi_t^+=0$ at $t=0$. It can then be shown that both $\mathsf{R}*\phi^+$ and $(\partial/\partial t)(\mathsf{R}*\phi^+)$ vanish at $t=0$. Hence it follows that

$$\mathscr{K}^{-1}(I+\Lambda)(\mathsf{R}*\phi^+)$$

$$=\mathscr{K}^{-1}\frac{1}{c^2}\frac{\partial^2}{\partial t^2}\mathscr{K}^2(\mathsf{R}*\phi^+)=\mathscr{K}\frac{1}{c^2}\frac{\partial^2}{\partial t^2}(\mathsf{R}*\phi^+)$$

$$=\frac{1}{c^2}\frac{\partial^2}{\partial t^2}\mathscr{K}(\mathsf{R}*\phi^+)=(I+\Lambda)\mathscr{K}^{-1}(\mathsf{R}*\phi^+).$$

Insert expression (45) for ϕ^- into Eq. (50), and employ relation (47), and the above interchange of operators to obtain

$$(I+\Lambda)\left\{\frac{\partial}{\partial z}(\mathsf{R}*\phi^+)+\left(\mathscr{K}^{-1}+\frac{3}{2}\frac{c_z}{c}\Lambda\right)(\mathsf{R}*\phi^+)+\frac{c_z}{2c}\phi^+\right\}=0. \tag{51}$$

Because of the assumptions on ϕ^+, Lemma 6 can be applied to Eq. (51) to yield

$$\mathsf{R}_z*\phi^++\mathsf{R}*\phi_z^+$$

$$+\left(\mathscr{K}^{-1}+\frac{3}{2}\frac{c_z}{c}\Lambda\right)(\mathsf{R}*\phi^+)+\frac{c_z}{2c}\phi^+=0. \tag{52}$$

Replace ϕ_z^+ in Eq. (52) by the right-hand side of expression (49) to give

$$\mathsf{R}_z*\phi^++\mathsf{R}*\left(\mathscr{K}^{-1}\phi^++\frac{c_z}{2c}\Lambda\phi^+\right)$$

$$-\frac{c_z}{2c}\mathsf{R}*(I+\Lambda)^2(\mathsf{R}*\phi^+)$$

$$+\left(\mathscr{K}^{-1}+\frac{3}{2}\frac{c_z}{c}\Lambda\right)(\mathsf{R}*\phi^+)+\frac{c_z}{2c}\phi^+=0. \tag{53}$$

In order to reduce expression (53) the following results involving the convolution of $f(x,y,t)$ and $g(x,y,t)$ [as defined by Eq. (46)]:

$$f*(\mathscr{K}^pg)=(\mathscr{K}^pf)*g=\mathscr{K}^p(f*g),\quad p=1,2,$$

$$f*\left(\frac{\partial^2g}{\partial x^2}\right)=\left(\frac{\partial^2f}{\partial x^2}\right)*g=\frac{\partial^2}{\partial x^2}(f*g)$$

(with a similar result holding for the derivative with respect

to y), will be employed. In addition to these relationships, the relationships

$$\frac{\partial^2}{\partial t^2}(f*g) - \left(\frac{\partial^2 f}{\partial t^2}\right)*g$$
$$= f*\left(\frac{\partial^2 g}{\partial t^2}\right) - \left(\frac{\partial^2 f}{\partial t^2}\right)*g$$
$$= \int_{\mathbb{R}^2}\int f_t(x-x',y-y',0)g(x',y',t)dx'\,dy'$$

valid for f and g such that at time $t=0$, $f=g=\partial g/\partial t=0$, will be used.

Using the relation $\mathscr{K}^{-1}=\Box\mathscr{K}$ and the fact that $\mathscr{K}\mathsf{R}=0$, and $(\partial/\partial t)(\mathscr{K}\mathsf{R}) = -c\mathsf{R}(x,y,z,0)$ at $t=0$, it can thus be shown that

$$\mathsf{R}*(\mathscr{K}^{-1}\phi^+)$$
$$= \mathscr{K}^{-1}(\mathsf{R}*\phi^+)$$
$$= (\mathscr{K}^{-1}\mathsf{R})*\phi^+ - \frac{1}{c}\int_{\mathbb{R}^2}\int \mathsf{R}(x-x',y-y',z,0)$$
$$\times\phi^+(x',y',z,t)dx'\,dy'. \tag{54}$$

From the relation

$$(I+\Lambda) = \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\mathscr{K}^2$$

and the fact that $\mathscr{K}^2\mathsf{R} = (\partial/\partial t)(\mathscr{K}^2\mathsf{R}) = 0$ at $t=0$, it also follows that

$$\mathsf{R}*(I+\Lambda)\phi^+ = [(I+\Lambda)\mathsf{R}]*\phi^+ = (I+\Lambda)(\mathsf{R}*\phi^+). \tag{55}$$

From the results of Eqs. (54) and (55), Eq. (53) can now be reduced to the form

$$(\Gamma\mathsf{R})*\phi^+ + \frac{c_z}{2c}\phi^+ - \frac{2}{c}\int_{\mathbb{R}^2}\int \mathsf{R}(x-x',y-y',z,0)$$
$$\times\phi^+(x',y',z,t)dx'\,dy' = 0, \tag{56}$$

where the operator Γ is given by

$$\Gamma\mathsf{R} = \mathsf{R}_z + 2\mathscr{K}^{-1}\mathsf{R} + 2(c_z/c)\Lambda\mathsf{R}$$
$$- (c_z/2c)(I+\Lambda)\mathsf{R}*(I+\Lambda)\mathsf{R}. \tag{57}$$

Now impose the following initial condition on R the reflection kernel:

$$\mathsf{R}(x,y,z,0) = (c_z/4)\delta(x)\delta(y). \tag{58}$$

In this case Eq. (56) reduces to $(\Gamma\mathsf{R})*\phi^+ = 0$. Since ϕ^+ may be treated as an arbitrary smooth test function such that $\phi^+ = \phi_t^+ = 0$, the following equation for the reflection kernel is immediately derived:

$$\mathsf{R}_z + 2(\mathscr{K}^{-1} + (c_z/c)\Lambda)\mathsf{R}$$
$$- (c_z/2c)(I+\Lambda)\mathsf{R}*(I+\Lambda)\mathsf{R} = 0. \tag{59}$$

Equation (59) and initial condition (58) constitutes the required system for the reflection kernel. These are seen to be an immediate generalization of the one-dimensional system (6).

The existence of the solution of equations (58) and (59) needs to be examined. Because of the delta function in the initial condition one would have to look for solutions belonging to the space of tempered distributions. The quadratic term should pose no problem since it is a convolution and the operator Λ is convolution like in its support.

Equation (59) can be used both in the direct and inverse scattering problem. In the inverse problem, $\mathsf{R}(x,y,z,t)$ is a known function on the plane $z=0$. Equation (59) is used to numerically construct $\mathsf{R}(x,y,z,t)$ in a step by step basis (peeling off layer by layer) in the region $z>0$, and the value of $c(z)$ is recovered from condition (58). The process described here needs to be examined in detail.

VI. COMMENTS

The results derived here represent an important intermediate step in extending the concept of wave splitting with its associated reflection operator, from the one-dimensional case to the full three-dimensional case where $c=c(x,y,z)$. Present investigation indicates that the concepts and analysis developed here can be extended to wave splitting for smooth nonplanar geometry.

ACKNOWLEDGMENTS

This research was supported by O.N.R. under Contract No. N00014-83-K-6038.

APPENDIX A: EVALUATION OF $\mathscr{K}^2$ AND $((\partial/\partial z)\mathscr{K})\mathscr{K}$

From Eq. (22) if follows that

$$\mathscr{K}^2 w = \frac{1}{(2\pi)^2}\int_{\mathbb{R}^2\times\mathbb{R}^2}\frac{H(ct-R'-R'')}{R'R''}$$
$$\times w\left(x'',y'',t-\frac{R'}{c}-\frac{R''}{c}\right)dx'\,dy'\,dx''\,dy'', \tag{A1}$$

where $H(\eta)$ is the Heaviside step function, $c=c(z)$, and

$$R' = [(x-x')^2+(y-y')^2]^{1/2},$$
$$R'' = [(x'-x'')^2+(y'-y'')^2]^{1/2}.$$

This can be written in the form

$$\mathscr{K}^2 w = \frac{c}{(2\pi)^2}\int_{\mathbb{R}^2\times\mathbb{R}^2}\int_0^\infty \frac{w(x'',y'',s)\delta(c(t-s)-R'-R'')}{R'R''}ds\,dx'\,dy'\,dx''\,dy''$$
$$= \frac{c}{(2\pi)^2}\int_{\mathbb{R}^2}\int_0^\infty w(x'',y'',s)I_1 dx''\,dy'', \tag{A2}$$

where

$$I_1 = \int_{\mathbb{R}^2}\frac{\delta(c(t-s)-R'-R'')}{R'R''}dx'\,dy'. \tag{A3}$$

With (x',y'), the variables of integration of I_1, free to be chosen so that the x' axis lies along the line joining points (x,y) and (x'',y'') and the origin at the midpoint, elliptic coordinates (μ,θ) given by[16]

$$x' = \tfrac{1}{2} R \cosh\mu \cos\theta, \quad y' = \tfrac{1}{2} R \sinh\mu \sin\theta \tag{A4}$$

are introduced where R is the distance between (x,y) and (x'',y'') and $0 \leqslant \theta \leqslant 2\pi$, $0 \leqslant \mu < \infty$. Using the fact that

$$(1/R'R'')dx'\,dy' = d\mu\,d\theta, \quad R' + R'' = R\cosh\mu,$$

we have

$$\begin{aligned} I_1 &= \int_0^{2\pi}\int_0^{\infty} \delta(c(t-s) - R\cosh\mu)d\mu\,d\theta, \\ &= 2\pi \frac{H(t-s)H(c(t-s)-R)}{[c^2(t-s)^2 - R^2]^{1/2}}, \end{aligned} \tag{A5}$$

which yields the following:

$$\mathscr{K}^2 w = \frac{c}{2\pi}\int_0^t \iint_{U(x,y,c(t-s))} \frac{w(x'',y'',s)dx''\,dy''\,ds}{[c^2(t-s)^2 - (x-x'')^2 - (y-y'')^2]^{1/2}}. \tag{A6}$$

The evaluation of $((\partial/\partial z)\mathscr{K})\mathscr{K}w$ proceeds in a similar fashion. First noting that

$$\left(\frac{\partial}{\partial z}\mathscr{K}\right) w = -\frac{1}{c^2}\frac{\partial c}{\partial z}\frac{1}{2\pi} \times \iint_{U(x,y;ct)} w_t\left(x',y',t-\frac{R'}{c}\right)dx'\,dy'$$

with

$$R' = [(x-x')^2 + (y-y')^2]^{1/2},$$

it follows in a similar fashion that

$$\left(\frac{\partial}{\partial z}\mathscr{K}\right)\mathscr{K}w = \frac{1}{c}\frac{\partial c}{\partial z}\frac{1}{(2\pi)^2} \times \int_{\mathbb{R}^2}\int_0^{\infty} w_s(x'',y'',s)I_2\,dx''\,dy'' \tag{A7}$$

where

$$I_2 = \int_{\mathbb{R}^2} \frac{\delta(c(t-s) - R' - R'')}{R''}dx'\,dy'.$$

Using elliptic coordinates this reduces to

$$\begin{aligned} I_2 &= \int_0^{2\pi}\int_0^{\infty} \delta(c(t-s) - R\cosh\mu) \\ &\quad \times \frac{R}{2}(\cosh\mu - \cos\theta)d\mu\,d\theta \\ &= \frac{\pi H(t-s)H(c(t-s)-R)c(t-s)}{[c^2(t-s)^2 - R^2]^{1/2}}. \end{aligned} \tag{A8}$$

Hence we have from (A7) and (A8),

$$\begin{aligned} &\left(\frac{\partial\mathscr{K}}{\partial z}\right)\mathscr{K}w \\ &= \frac{1}{4\pi}\frac{\partial c}{\partial z} \\ &\quad \times \int_0^t \int_{U(x,y;c(t-s))} \frac{w_s(x'',y'',s)(t-s)}{[c^2(t-s)^2 - R^2]^{1/2}}dx''\,dy'' \\ &= \frac{1}{2c}\frac{\partial c}{\partial z}\{t\mathscr{K}^2(w_t) - \mathscr{K}^2(tw_t)\}. \end{aligned} \tag{A9}$$

APPENDIX B: ALTERNATIVE FORM FOR $\mathscr{K}^{-1}$

In expression (22) for $\mathscr{K}w$, change the variables of integration (x',y') to local polar coordinates (R,θ) centered at (x,y), $x' = x + R\cos\theta, y' = y + R\sin\theta$, then replace R by s through the substitution $R = c(t-s)$, giving

$$\begin{aligned} \mathscr{K}w = -\frac{c}{2\pi}\int_0^t\int_0^{2\pi} w(x + c(t-s)\cos\theta, y \\ + c(t-s)\sin\theta, s)d\theta\,ds. \end{aligned} \tag{B1}$$

If $w(x,y,t) \in C^2(\mathbb{R}^2) \times C^2[0,\infty)$, then it can be shown that

$$\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right)\mathscr{K}w = -\frac{c}{2\pi}\int_0^t\int_0^{2\pi}\{w_{xx} + w_{yy}\}d\theta\,ds, \tag{B2}$$

$$\begin{aligned} \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\mathscr{K}w = -\frac{c}{2\pi}\int_0^t\int_0^{2\pi}[w_{xx}\cos^2\theta + 2w_{xy} \\ \times\cos\theta\sin\theta + w_{yy}\sin^2\theta]d\theta\,ds \\ -\frac{1}{c}\frac{\partial}{\partial t}w(x,y,t). \end{aligned} \tag{B3}$$

Hence it follows that

$$\begin{aligned} \Box\mathscr{K}w = \frac{c}{2\pi}\lim_{\epsilon\to 0}\int_0^{t-\epsilon}\int_0^{2\pi}\frac{1}{c^2(t-s)^2}\frac{\partial^2 w}{\partial\theta^2}d\theta\,ds \\ + \frac{c}{2\pi}\lim_{\epsilon\to 0}\int_0^{t-\epsilon}\int_0^{2\pi}\frac{1}{c(t-s)} \\ \times[w_x\cos\theta + w_y\sin\theta]d\theta\,ds \\ -\frac{1}{c}\frac{\partial}{\partial t}w(x,y,t). \end{aligned} \tag{B4}$$

Noting that

$$\lim_{s\to t}\int_0^{2\pi}(w_x\cos\theta + w_y\sin\theta)d\theta = 0,$$

it follows from Eqs. (B3) and (24) that

$$\mathscr{K}^{-1}w$$

$$= \frac{1}{2\pi}\int_0^t \frac{1}{(t-s)}\int_0^{2\pi}(w_x\cos\theta + w_y\sin\theta)d\theta\,ds$$

$$-\frac{1}{c}\frac{\partial}{\partial t}w(x,y,t). \tag{B5}$$

On transformation back to the variables x',y', this takes the form

$$\mathscr{K}^{-1}w = -\frac{1}{c}\frac{\partial}{\partial t}w(x,y,t) + \mathscr{L}w, \tag{B6}$$

where

$$\mathscr{L}w = \frac{1}{2\pi}\iint_{U(x,y;ct)}\left\{w_x\left(x',y';t-\frac{R}{c}\right)\frac{(x'-x)}{R^3}\right.$$

$$\left. + w_y\left(x',y';t-\frac{R}{c}\right)\frac{(y'-y)}{R^3}\right\}dx'\,dy'. \tag{B7}$$

[1]J. P. Corones, M. E. Davison, and R. J. Krueger, "Wave splittings, invariant imbedding and inverse scattering," in *Inverse Optics*, Proc. SPIE 413, edited by A. J. Devaney (SPIE, Bellingham, WA, 1983), pp. 102–106.

[2]M. Davison, "A general approach to splitting and invariant imbedding techniques for linear wave equations," Ames Laboratory preprint, to appear in J. Math. Anal. Appl.

[3]H. Bremmer, "The W.K.B. approximation as the first term of a geometric-optimal series," Commun. Pure Appl. Math. **4**, 105 (1951).

[4]R. Bellman and G. N. Wing, *An Introduction to Invariant Imbeddding* (Wiley, New York, 1975).

[5]R. Redheffer, "On the relation of transmission-line theory to scattering and transfer," J. Math.Phys. (Cambridge, MA) **41**, 1 (1962).

[6]J. Corones and R. J. Krueger, "Obtaining scattering kernels using invariant imbedding," J. Math. Anal. Appl. **95**, 393 (1983).

[7]J. P. Corones, M. E. Davison, and R. J. Krueger, "Direct and inverse scattering in the time domain by invariant imbedding techniques," J. Acoust. Soc. Am. **74**, 1535 (1983).

[8]R. A. Beezley and R. J. Krueger, "An electromagnetic inverse problem for dispersive media," J. Math. Phys. **26**, 317 (1985).

[9]J. P. Corones, M. E. Davison, and R. J. Krueger, "Dissipative inverse problems in the time domain," in *Inverse Methods in Electromagnetic Imaging*, NATO ASI series, Series C, edited by W. M. Boerner (Reidel, Dordrecht, 1985), Vol. 143, pp. 123–130.

[10]E. Ammicht, J. P. Corones, and R. J. Krueger, "Direct and inverse scattering for viscoelastic media," preprint.

[11]Robert Dougherty, Ph.D dissertation, Iowa State University, 1986.

[12]J. P. Corones and R. J. Krueger, "Higher order parabolic approximations to time-independent wave equations," J. Math. Phys. **24**, 2301 (1983).

[13]L. Fishman and J. J. McCoy, "Derivation and application of extended wave parabolic wave theories, I. The factorized Helmholtz equation," J. Math. Phys. **25**, 285 (1984).

[14]A. E. Yagle and B. L. Levy, "Layer stripping solutions of multidimensional inverse scattering problems," J. Math. Phys. **27**, 1701 (1986).

[15]V. S. Vladimirov, *Equations of Mathematical Physics* (Dekker, New York, 1971).

[16]P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill, New York, 1953), Vol. 2.

Confinement and redistribution of charges and currents on a surface by external fields

Henk F. Arnoldus, Daniel Jelski, and Thomas F. George
Department of Physics and Astronomy and of Chemistry, 239 Fronczak Hall, State University of New York at Buffalo, Buffalo, New York 14260

(Received 6 October 1986; accepted for publication 7 January 1987)

The old problem of light scattering from a perfectly conducting surface is addressed. An electromagnetic field is incident upon the boundary, where it induces a charge and current distribution. These charges and currents emit the reflected fields. A set of equations for the charges and currents on the surface is derived by eliminating the **E** and **B** fields from Maxwell's equations with the aid of the appropriate boundary conditions. An explicit and general solution is achieved, which reveals the confinement and redistribution of the charge and the current on the surface by the external field. Expressions are obtained for the surface resolvents, or the redistribution matrices, which represent the surface geometry. Action of a surface resolvent on the incident field, evaluated at the surface, then yields the charge and current distributions. The Faraday induction appears as an additional contribution to the charge density. Subsequently, the reflected fields are expanded in spherical waves, which have the surface-multipole moments as a source. Explicit expressions are presented for the surface-multipole moments, and it is pointed out that charge conservation on the surface sets constraints on these moments. The results apply to arbitrarily shaped surfaces and to any incident field. For a specific choice of the surface structure and the external field, the solutions for the charge, the current, and the reflected fields are amenable to numerical evaluation.

I. INTRODUCTION

The study of chemistry and physics near a surface has developed rapidly during the last decade. Investigations range from classical processes like periodic deposition,[1] image formation,[2–4] and dispersion of plasmon waves[5–22] to quantum mechanical issues as Raman scattering of intense laser light,[23,24] atomic fluorescence near a rough surface,[25,26] the coupling of an atomic dipole to surface polaritons,[27] and cooperative emission processes near a conductor.[28] It appears, however, that besides these well-established theories, even the simplest problem—light scattering from an arbitrarily shaped surface—is not yet completely tractable. Early approximations like the Rayleigh–Fano expansion (neglection of multiple reflections) or the small-roughness limit provide sufficient understanding of the induced effects on a boundary by incident fields, but exact solutions in the form of general expressions for the scattered fields and the surface waves are not available at present. Contemporary closed-form solutions pertain only to polarized plane waves, incident upon gratings with well-defined geometries, like square or sinusoidal wells. The results always rely on the periodicity of the surface roughness, which implies the applicability of Fourier-series expansions, or a numerical solution of the extinction theorems, as they exist in many phrasings.[7,8,11] In this paper we consider a metallic surface, which is illuminated by an externally applied electromagnetic field with an arbitrary time dependence and spatial distribution. The surface is not assumed to be periodic, and our results apply equally well to a closed surface or to assemblies of surfaces, as for example a sphere near a grating. We achieve closed-form solutions of Maxwell's equations for the charge and current distributions on the surface and for the reflected fields, although at the expense of the assumption that the metal has a perfect conductivity.

II. THE FIELD EQUATIONS

The time development and the spatial distribution of the charge density $\rho(\mathbf{r},t)$, the current density $\mathbf{j}(\mathbf{r},t)$, the electric field $\mathbf{E}(\mathbf{r},t)$, and the magnetic field $\mathbf{B}(\mathbf{r},t)$ are governed by Maxwell's equations. If we adopt a Fourier transform of the real-valued fields

$$\mathbf{E}(\mathbf{r},t)=\frac{1}{\pi}\,\mathrm{Re}\int_0^\infty d\omega\,\hat{\mathbf{E}}(\mathbf{r},\omega)e^{-i\omega t}, \tag{2.1}$$

and similarly for the other three fields, then the field equations read

$$\nabla\cdot[\epsilon(\mathbf{r})\mathbf{E}(\mathbf{r})]=\rho(\mathbf{r}), \tag{2.2}$$

$$\nabla\cdot\mathbf{B}(\mathbf{r})=0, \tag{2.3}$$

$$\nabla\times\mathbf{E}(\mathbf{r})-i\omega\mathbf{B}(\mathbf{r})=0, \tag{2.4}$$

$$\nabla\times[\mu(\mathbf{r})^{-1}\mathbf{B}(\mathbf{r})]+i\omega\epsilon(\mathbf{r})\mathbf{E}(\mathbf{r})=\mathbf{j}(\mathbf{r}), \tag{2.5}$$

where we have simplified the notation by writing $\mathbf{E}(\mathbf{r})$ rather than $\hat{\mathbf{E}}(\mathbf{r},\omega)$. The frequency dependence of the fields and of $\epsilon(\mathbf{r})$ and $\mu(\mathbf{r})$ will be suppressed throughout this paper.

We shall suppose that the entire space is occupied by two kinds of media, perfect conductors and perfect insulators, which are separated by boundaries. The set of all boundaries will then be referred to as the surface. Within each medium the dielectric constant $\epsilon(\mathbf{r})$ and the permeability $\mu(\mathbf{r})$ will be assumed to be $\mathbf{r}$ independent, but across the surface $\epsilon(\mathbf{r})$ and $\mu(\mathbf{r})$ are discontinuous. Conductors are specified by a relation like $\mathbf{j}(\mathbf{r})=\gamma\mathbf{E}(\mathbf{r})$, $\gamma>0$, and the assumption of perfect conductivity implies the limit $\gamma\to\infty$.

 0022-2488/87/051069-06$02.50

Since the current density $\mathbf{j}(\mathbf{r})$ should remain finite, we obtain $\mathbf{E}(\mathbf{r})=\mathbf{0}$ everywhere in the conductor. From Eq. (2.2) we then find $\rho(\mathbf{r})=0$, and Eq. (2.4) yields $\mathbf{B}(\mathbf{r})=\mathbf{0}$, under the restriction $\omega\neq 0$. In this paper we will exclude the trivial static case $\omega=0$. Finally, Eq. (2.5) gives $\mathbf{j}(\mathbf{r})=\mathbf{0}$, and hence Maxwell's equations in the conductor reduce to

$$\mathbf{E}(\mathbf{r})=\mathbf{0},\quad \mathbf{B}(\mathbf{r})=\mathbf{0},\quad \rho(\mathbf{r})=0,\quad \mathbf{j}(\mathbf{r})=0. \tag{2.6}$$

Around a point $\mathbf{r}$ on the surface the fields are discontinuous. Application of Gauss' theorem on (2.2) and (2.3) and of Stokes' theorem on (2.4) and (2.5) enables us to rewrite the equations in the vicinity of the surface as

$$\mathbf{E}(\mathbf{r}^+)=\epsilon^{-1}\sigma(\mathbf{r})\mathbf{n}(\mathbf{r}), \tag{2.7}$$

$$\mathbf{B}(\mathbf{r}^+)=\mu\mathbf{i}(\mathbf{r})\times\mathbf{n}(\mathbf{r}). \tag{2.8}$$

Here $\sigma(\mathbf{r})$ and $\mathbf{i}(\mathbf{r})$ are the surface charge and current density, respectively, and $\mathbf{n}(\mathbf{r})$ represents the unit normal vector in $\mathbf{r}$ on the surface, with the convention that it points from the conductor to the dielectric. We have introduced the notation $\mathbf{r}^+$ to indicate a point in the dielectric and close to $\mathbf{r}$. Explicitly, we write

$$\mathbf{r}^+=\mathbf{r}+\mathbf{n}(\mathbf{r})\delta \quad \text{with } \delta\downarrow 0. \tag{2.9}$$

We note that Eqs. (2.7) and (2.8) combine the four Maxwell equations in $\mathbf{r}$ on the surface, and that they contain four unknown fields.

The dielectric is presumed to exhibit no conductivity at all, so it can be specified by $\mathbf{j}=\gamma\mathbf{E}$ with $\gamma\to 0$. This implies $\mathbf{j}=\mathbf{0}$, and from charge conservation ($\nabla\cdot\mathbf{j}=i\omega\rho$) we find $\rho=0$, since we required $\omega\neq 0$. Hence, all charges and currents, if any, are situated on the surface as $\sigma(\mathbf{r})$ and $\mathbf{i}(\mathbf{r})$. The electric and magnetic fields in the dielectric are generated by $\sigma(\mathbf{r})$ and $\mathbf{i}(\mathbf{r})$, and they contain the incident fields. This notion allows us to write Maxwell's equations for a point $\mathbf{r}$ in the dielectric as

$$\rho(\mathbf{r})=0, \tag{2.10}$$

$$\mathbf{j}(\mathbf{r})=0, \tag{2.11}$$

$$\mathbf{E}(\mathbf{r})-\mathbf{E}(\mathbf{r})^{\text{inc}}=\frac{-1}{4\pi\epsilon}\int dA'\,\sigma(\mathbf{r}')\nabla G(\mathbf{r},\mathbf{r}') + \frac{i\omega\mu}{4\pi}\int dA'\,\mathbf{i}(\mathbf{r}')G(\mathbf{r},\mathbf{r}'), \tag{2.12}$$

$$\mathbf{B}(\mathbf{r})-\mathbf{B}(\mathbf{r})^{\text{inc}}=\frac{-\mu}{4\pi}\int dA'\,\mathbf{i}(\mathbf{r}')\times\nabla G(\mathbf{r},\mathbf{r}'), \tag{2.13}$$

where the integrals run over the complete surface. This representation involves the Green's function of the wave equation,

$$G(\mathbf{r},\mathbf{r}')=|\mathbf{r}-\mathbf{r}'|^{-1}\exp(ik|\mathbf{r}-\mathbf{r}'|), \tag{2.14}$$

and its gradient

$$\nabla G(\mathbf{r},\mathbf{r}')=(\mathbf{r}-\mathbf{r}')|\mathbf{r}-\mathbf{r}'|^{-3}(ik|\mathbf{r}-\mathbf{r}'|-1)\times\exp(ik|\mathbf{r}-\mathbf{r}'|), \tag{2.15}$$

which contain the wave number $k=(\epsilon\mu)^{1/2}\omega$. We have to solve the set (2.12) and (2.13) for $\sigma(\mathbf{r})$, $\mathbf{i}(\mathbf{r})$, $\mathbf{E}(\mathbf{r})$, and $\mathbf{B}(\mathbf{r})$, and Maxwell's equations (2.7) and (2.8) on the surface can be considered as the boundary conditions.

III. ELIMINATION OF THE FIELDS

Maxwell's equations in the dielectric medium are basically two equations with four unknown fields, but we can eliminate the radiation fields $\mathbf{E}(\mathbf{r})$ and $\mathbf{B}(\mathbf{r})$ with the boundary conditions (2.7) and (2.8). To this end we take $\mathbf{r}$ in (2.12) and (2.13) as $\mathbf{r}^+$ from (2.9), and then substitute the boundary values for $\mathbf{E}(\mathbf{r}^+)$ and $\mathbf{B}(\mathbf{r}^+)$. This procedure leaves us with a set of two equations for $\sigma(\mathbf{r})$ and $\mathbf{i}(\mathbf{r})$. The appearance of $G(\mathbf{r}^+,\mathbf{r}')$ and $\nabla_{\mathbf{r}^+}G(\mathbf{r}^+,\mathbf{r}')$ in the integrands of (2.12) and (2.13) is not convenient since it involves points $\mathbf{r}^+$, which are not situated on the surface. It will turn out to be more practical to have equations in which the Green's function connects only points of the surface, rather than a point on the surface to a point in the dielectric. However, care should be exercised in replacing $\mathbf{r}^+$ by $\mathbf{r}$, because the integrals are discontinuous across the surface. If we take the limit $\mathbf{r}^+\to\mathbf{r}$ properly (see Appendix), we obtain

$$\int dA'\,\sigma(\mathbf{r}')\nabla G(\mathbf{r}^+,\mathbf{r}') = -2\pi\sigma(\mathbf{r})\mathbf{n}(\mathbf{r})+\int dA'\,\sigma(\mathbf{r}')\nabla G(\mathbf{r},\mathbf{r}'), \tag{3.1}$$

$$\int dA'\,\mathbf{i}(\mathbf{r}')G(\mathbf{r}^+,\mathbf{r}')=\int dA'\,\mathbf{i}(\mathbf{r}')G(\mathbf{r},\mathbf{r}'), \tag{3.2}$$

$$\int dA'\,\mathbf{i}(\mathbf{r}')\times\nabla G(\mathbf{r}^+,\mathbf{r}') = -2\pi\mathbf{i}(\mathbf{r})\times\mathbf{n}(\mathbf{r})+\int dA'\,\mathbf{i}(\mathbf{r}')\times\nabla G(\mathbf{r},\mathbf{r}'), \tag{3.3}$$

and we observe that replacing $\mathbf{r}^+$ by $\mathbf{r}$ requires that we should add the terms $-2\pi\sigma(\mathbf{r})\mathbf{n}(\mathbf{r})$ and $-2\pi\mathbf{i}(\mathbf{r})\times\mathbf{n}(\mathbf{r})$ in Eqs. (3.1) and (3.3). It was already pointed out by Maradudin[29] that integrals of this kind appear to have a finite contribution from a single point. This feature can, however, also be regarded as resulting from the discontinuity of the fields across the surface. Critical comments on this issue have also been made by Agarwal[12] in a slightly different context. Combining everything then yields the set of equations

$$\sigma(\mathbf{r})\mathbf{n}(\mathbf{r})=\frac{-1}{2\pi}\int dA'\,\sigma(\mathbf{r}')\nabla G(\mathbf{r},\mathbf{r}') + \frac{i\omega\epsilon\mu}{2\pi}\int dA'\,\mathbf{i}(\mathbf{r}')G(\mathbf{r},\mathbf{r}')+2\epsilon\mathbf{E}(\mathbf{r})^{\text{inc}}, \tag{3.4}$$

$$\mathbf{i}(\mathbf{r})\times\mathbf{n}(\mathbf{r})=\frac{-1}{2\pi}\int dA'\,\mathbf{i}(\mathbf{r}')\times\nabla G(\mathbf{r},\mathbf{r}')+2\mu^{-1}\mathbf{B}(\mathbf{r})^{\text{inc}}, \tag{3.5}$$

for $\sigma(\mathbf{r})$ and $\mathbf{i}(\mathbf{r})$. We can write $\sigma(\mathbf{r})$ and $\mathbf{i}(\mathbf{r})$ in the integrands as

$$\sigma(\mathbf{r})=\mathbf{n}(\mathbf{r})\cdot(\sigma(\mathbf{r})\mathbf{n}(\mathbf{r})), \tag{3.6}$$

$$\mathbf{i}(\mathbf{r})=\mathbf{n}(\mathbf{r})\times(\mathbf{i}(\mathbf{r})\times\mathbf{n}(\mathbf{r})), \tag{3.7}$$

since $\mathbf{i}(\mathbf{r})$ is parallel to the surface, which shows that Eqs. (3.4) and (3.5) are essentially a set of equations for the vector fields $\sigma(\mathbf{r})\mathbf{n}(\mathbf{r})$ and $\mathbf{i}(\mathbf{r})\times\mathbf{n}(\mathbf{r})$ on the surface.

Equation (3.5) for $\mathbf{i}(\mathbf{r})\times\mathbf{n}(\mathbf{r})$ has the form of an inhomogeneous Fredholm equation of the second kind, where the external field $2\mu^{-1}\mathbf{B}(\mathbf{r})^{\text{inc}}$ is the inhomogeneity. In the

same fashion, Eq. (3.4) has $2\epsilon\mathbf{E}(\mathbf{r})^{\text{inc}}$ (and the current term) as an inhomogeneous part. Hence the incident fields can be regarded as the source terms of these equations. In this sense $\sigma(\mathbf{r})\neq 0$ and $\mathbf{i}(\mathbf{r})\neq\mathbf{0}$ are a result of the presence of the driving field, so the charges and the currents are confined on the surface by the field. If there is a net charge on the surface, this mechanism might also be conceived as a redistribution process. Equations (3.4) and (3.5) resemble the extinction theorem for the analogous problem of scattering of an incident field from a dielectric grating. The extinction theorem is, however, a homogeneous equation, and its solvability condition is equivalent to the dispersion relation for surface polaritons.

IV. REPRESENTATION OF THE SURFACE

Ordinary Fredholm equations are single-variable equations for a function on the complex plane, and they can be solved by an expansion of the function onto a suitable complete set. Our equations for $\sigma(\mathbf{r})\mathbf{n}(\mathbf{r})$ and $\mathbf{i}(\mathbf{r})\times\mathbf{n}(\mathbf{r})$ are three-dimensional and surface-related equations for a vector field, so we have to modify the standard technique slightly. In order to accomplish this, we introduce spherical coordinates (r,θ,ϕ) with respect to an arbitrary origin, and we will abbreviate the direction θ,ϕ by the single variable Ω. Then the assembly of all points $\mathbf{r}$, which constitute the surface, can be represented by a set of functions $\xi(\Omega)_\lambda$. The $\xi(\Omega)_\lambda$ will indicate the distance from the origin to a point $\mathbf{r}$ on the surface, in the direction Ω, while the subscript λ accounts for the multiplicity (see Fig. 1). In this fashion, the surface is divided in regions, numbered by λ, where its shape is defined by a function $\xi(\Omega)_\lambda$, which determines uniquely the spherical coordinates $(r,\theta,\phi)=(\xi(\theta,\phi)_\lambda,\theta,\phi)$ of a point $\mathbf{r}$ in this region. The shape functions $\xi(\Omega)_\lambda$ will be assumed to be given, and therefore we can represent a point on the surface by its surface coordinates (λ,Ω) rather than by its spherical coordinates (r,Ω). We will use λ as a subscript and Ω as a variable.

The measure $dA(\Omega)_\lambda$ and the direction $\mathbf{n}(\Omega)_\lambda$ of the surface at a given point (λ,Ω) are fixed by its shape $\xi(\Omega)_\lambda$. For instance, the infinitesimal surface area at (λ,Ω) is given by

$$dA(\Omega)_\lambda = f(\Omega)_\lambda\, d\Omega, \tag{4.1}$$

with

$$f(\Omega)_\lambda = \xi(\Omega)_\lambda\left\{\xi(\Omega)_\lambda^2+\left(\frac{\partial}{\partial\theta}\xi(\Omega)_\lambda\right)^2 + \frac{1}{\sin^2\theta}\left(\frac{\partial}{\partial\phi}\xi(\Omega)_\lambda\right)^2\right\}^{1/2}, \tag{4.2}$$

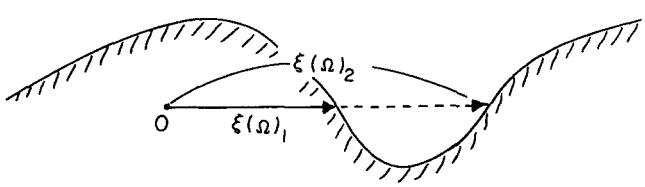


FIG. 1. Illustration of the surface multiplicity. From the origin O in the direction Ω, we find points on the surface which have a distance $\xi(\Omega)_1,\xi(\Omega)_2,\ldots$ to 0. Therefore, a description of the surface in spherical coordinates requires a set of functions $\xi(\Omega)_\lambda$.

in terms of the infinitesimal surface area $d\Omega=\sin\theta\, d\theta\, d\phi$ of the unit sphere. Hence the function $f(\Omega)_\lambda$ accounts for the deviation of the surface curvature from the curvature of a sphere, and with the aid of (4.1) we can transform a surface integral over the region λ into an integration over a part of the unit sphere. We note that not every direction Ω for a given λ corresponds to a point on the surface. It will turn out to be convenient to extend the definition (4.1) of $f(\Omega)_\lambda$ as

$$f(\Omega)_\lambda = 0, \quad \text{if } \Omega \text{ does not correspond to a point on the surface in region } \lambda. \tag{4.3}$$

Then we can write the surface integrals as

$$\int dA\cdots = \sum_\lambda \int d\Omega\, f(\Omega)_\lambda \cdots, \tag{4.4}$$

where the integrals now run over the complete unit sphere for every λ. This construction will enable us to apply the general theory of expanding vector fields on a sphere.

V. EXPANSION OF THE FIELDS

Since we are using spherical coordinates, the spherical harmonics $Y(\Omega)_{lm}$ supply a suitable complete set on the unit sphere for an expansion of the magnitude of a vector field. The direction of a vector will be expanded onto a space-fixed set of three unit vectors, denoted by $\mathbf{e}_\tau$, which is, for instance, the Cartesian set $\mathbf{e}_x,\mathbf{e}_y,\mathbf{e}_z$ or the spherical set $\mathbf{e}_{+1},\mathbf{e}_0,\mathbf{e}_{-1}$. Then the vector fields $Y(\Omega)_{lm}\mathbf{e}_\tau$ constitute a complete set on the unit sphere for an expansion of an arbitrary vector field.

It is our aim to solve Eqs. (3.4) and (3.5) for $\sigma(\mathbf{r})\mathbf{n}(\mathbf{r})$ and $\mathbf{i}(\mathbf{r})\times\mathbf{n}(\mathbf{r})$. We thus start with an expansion of these fields,

$$f(\Omega)_\lambda\sigma(\Omega)_\lambda\mathbf{n}(\Omega)_\lambda = \sum_{lm\tau} S_{lm\tau\lambda}\, Y_{lm}(\Omega)\mathbf{e}_\tau, \tag{5.1}$$

$$f(\Omega)_\lambda\mathbf{i}(\Omega)_\lambda\times\mathbf{n}(\Omega)_\lambda = \sum_{lm\tau} I_{lm\tau\lambda}\, Y_{lm}(\Omega)\mathbf{e}_\tau, \tag{5.2}$$

and note that we have included a factor $f(\Omega)_\lambda$ on the left-hand side. This is necessary, since otherwise the left-hand side of Eqs. (5.1) and (5.2) would not be properly defined for every Ω. The driving, incident fields $\mathbf{E}(\mathbf{r})^{\text{inc}},\mathbf{B}(\mathbf{r})^{\text{inc}}$ in Eqs. (3.4) and (3.5) enter only through their value on the surface, so that we can expand them on the surface set according to

$$f(\Omega)_\lambda\mathbf{E}(\Omega)_\lambda^{\text{inc}} = \sum_{lm\tau} E_{lm\tau\lambda}\, Y(\Omega)_{lm}\mathbf{e}_\tau, \tag{5.3}$$

$$f(\Omega)_\lambda\mathbf{B}(\Omega)_\lambda^{\text{inc}} = \sum_{lm\tau} B_{lm\tau\lambda}\, Y(\Omega)_{lm}\mathbf{e}_\tau. \tag{5.4}$$

The expansion coefficients for the incident fields then follow from the inverse relation

$$E_{lm\tau\lambda} = \int d\Omega\, f(\Omega)_\lambda\mathbf{E}(\Omega)_\lambda^{\text{inc}}\cdot\mathbf{e}_\tau^* Y(\Omega)_{lm}^*, \tag{5.5}$$

$$B_{lm\tau\lambda} = \int d\Omega\, f(\Omega)_\lambda\mathbf{B}(\Omega)_\lambda^{\text{inc}}\cdot\mathbf{e}_\tau^* Y(\Omega)_{lm}^*, \tag{5.6}$$

and the appearance of $f(\Omega)_\lambda$ in the integrands reflects that we actually have integrals over region λ of the surface. This

illustrates that $f(\Omega)_\lambda Y(\Omega)^*_{lm}\mathbf{e}^*_\tau$ can be considered as a complete surface set for the expansion of a vector field on the surface. Note that we allow $\mathbf{e}_\tau$ to be complex, which is the case for a spherical set.

VI. THE CHARGE AND THE CURRENT DISTRIBUTIONS

It is straightfoward to rewrite Eqs. (3.4) and (3.5) for $\sigma(\mathbf{r})\mathbf{n}(\mathbf{r})$ and $\mathbf{i}(\mathbf{r})\times\mathbf{n}(\mathbf{r})$ in terms of their expansion coefficients. We obtain

$$\sum_{l'm'\tau'\lambda'}(R^{(2)}_{lm\tau\lambda,l'm'\tau'\lambda'}-\delta_{ll'}\delta_{mm'}\delta_{\tau\tau'}\delta_{\lambda\lambda'})S_{l'm'\tau'\lambda'}$$
$$=i\omega\epsilon\mu\sum_{l'm'\tau'\lambda'}R^{(3)}_{lm\tau\lambda,l'm'\tau'\lambda'}I_{l'm'\tau'\lambda'}-2\epsilon E_{lm\tau\lambda}, \quad (6.1)$$

$$\sum_{l'm'\tau'\lambda'}(R^{(1)}_{lm\tau\lambda,l'm'\tau'\lambda'}-\delta_{ll'}\delta_{mm'}\delta_{\tau\tau'}\delta_{\lambda\lambda'})I_{l'm'\tau'\lambda'}$$
$$=-2\mu^{-1}B_{lm\tau\lambda}, \quad (6.2)$$

which are two coupled inhomogeneous linear equations for the surface charge $S_{lm\tau\lambda}$ and the surface current $I_{lm\tau\lambda}$. The expansion coefficients $E_{lm\tau\lambda}$ and $B_{lm\tau\lambda}$ for the external fields are supposed to be given. The set (6.1) and (6.2) also involves three R-matrices, with matrix elements

$$R^{(1)}_{lm\tau\lambda,l'm'\tau'\lambda'}=\frac{-1}{2\pi}\int d\Omega\int d\Omega' f(\Omega)_\lambda Y(\Omega)^*_{lm}Y(\Omega')_{l'm'}$$
$$\times\{\mathbf{e}^*_\tau\times(\mathbf{n}(\Omega')_{\lambda'}\times\mathbf{e}_{\tau'})\}\cdot\nabla G(\Omega,\Omega')_{\lambda\lambda'}, \quad (6.3)$$

$$R^{(2)}_{lm\tau\lambda,l'm'\tau'\lambda'}=\frac{-1}{2\pi}\int d\Omega\int d\Omega' f(\Omega)_\lambda Y(\Omega)^*_{lm}Y(\Omega')_{l'm'}$$
$$\times(\mathbf{n}(\Omega')_{\lambda'}\cdot\mathbf{e}^*_\tau)\mathbf{e}^*_\tau\cdot\nabla G(\Omega,\Omega')_{\lambda\lambda'}, \quad (6.4)$$

$$R^{(3)}_{lm\tau\lambda,l'm'\tau'\lambda'}=\frac{-1}{2\pi}\int d\Omega\int d\Omega' f(\Omega)_\lambda Y(\Omega)^*_{lm}Y(\Omega')_{l'm'}$$
$$\times\mathbf{e}^*_\tau\cdot(\mathbf{n}(\Omega')_{\lambda'}\times\mathbf{e}_{\tau'})G(\Omega,\Omega')_{\lambda\lambda'}, \quad (6.5)$$

where we have written $G(\Omega,\Omega')_{\lambda\lambda'}$ for the Green's function, which connects the points (λ,Ω) and (λ',Ω') of the surface. We emphasize that these R-matrices depend only on the geometry of the surface, and not on the external fields. Prescription of the shape of the surface determines the R-matrices. Recall, however, that the R-matrices depend on the frequency ω through the Green's function, but this is merely a parametric dependence and independent of the external field.

The expansion coefficients $S_{lm\tau\lambda}$ can always be arranged in a one-dimensional array, considered as a vector, and similarly $R^{(1)}$, $R^{(2)}$, and $R^{(3)}$ can be regarded as two-dimensional matrices. Then we can write (6.1) and (6.2) as

$$(R^{(2)}-1)S=i\omega\epsilon\mu R^{(3)}I-2\epsilon E, \quad (6.6)$$

$$(R^{(1)}-1)I=-2\mu^{-1}B, \quad (6.7)$$

where we have also adopted a vector representation for the driving fields. The solution of (6.6) and (6.7) is immediately found to be

$$S=\frac{2\epsilon}{1-R^{(2)}}\left\{E-i\omega R^{(3)}\frac{1}{1-R^{(1)}}B\right\}, \quad (6.8)$$

$$I=\frac{2\mu^{-1}}{1-R^{(1)}}B, \quad (6.9)$$

which expresses the charge density S and the current density I explicitly in the externally applied fields E and B and the surface-shape matrices $R^{(1)}$, $R^{(2)}$, and $R^{(3)}$.

For vanishing external fields, e.g., $E=0$ and $B=0$, the charge and current distributions also vanish, as can be seen explicitly from Eqs. (6.8) and (6.9). Hence the charges and currents are indeed confined to the surface by the external fields. Remember that we have excluded the static case $\omega=0$, for which we can have charges on a surface without external fields. Furthermore, we can identify the resolvents $(1-R^{(2)})^{-1}$ and $(1-R^{(1)})^{-1}$ as the operators that account for the redistribution of the charges and currents, respectively, as resulting from the Lorentz force between charges and between currents. The coupling of charges and currents, which is the Faraday induction, is incorporated in the $R^{(3)}$-matrix.

VII. THE REFLECTED FIELDS

The incident field induces charges and currents on the surface, and these oscillating charges and currents emit radiation, which are the reflected fields. In this section we express these fields in terms of the expansion coefficients $S_{lm\tau\lambda}$ and $I_{lm\tau\lambda}$, as they are given explicitly in the previous section.

In Eqs. (2.12) and (2.13) we expressed the reflected electric field $\mathbf{E}(\mathbf{r})-\mathbf{E}(\mathbf{r})^{\text{inc}}$ in terms of $\sigma(\mathbf{r})$ and $\mathbf{i}(\mathbf{r})$, and similarly $\mathbf{B}(\mathbf{r})-\mathbf{B}(\mathbf{r})^{\text{inc}}$ in terms of $\mathbf{i}(\mathbf{r})$. With (3.6) and (3.7) we can rewrite these equations in a way that $\sigma(\mathbf{r})\mathbf{n}(\mathbf{r})$ and $\mathbf{i}(\mathbf{r})\times\mathbf{n}(\mathbf{r})$ are the source fields, and then we can apply (5.1) and (5.2) in order to find an expansion on the spherical set. However, the resulting expressions are not transparent, since they will involve the Green's function and its gradient. In order to achieve a more comprehensible result, we expand the Green's function on the spherical set. We write[30]

$$G(\mathbf{r},\mathbf{r}'_\lambda)$$
$$=4\pi ik\sum_{lm}h^{(1)}(k\xi(\Omega')_\lambda)_l Y(\Omega')_{lm}j(kr)_l Y(\Omega)^*_{lm}, \quad (7.1)$$

where $h^{(1)}_l$ and j_l are spherical Bessel functions. Here the convention is that we choose the origin of our coordinate system in the dielectric, and in such a way that the inequality

$$\xi(\Omega)_\lambda > r \quad (7.2)$$

holds for every (λ,Ω). The vector $\mathbf{r}$ is the position in the dielectric, where we wish to evaluate the reflected fields. The expansion coefficients $S_{lm\tau\lambda}$ and $I_{lm\tau\lambda}$ depend on the position of the origin, so both the charge and current distributions and the reflected fields must be evaluated with respect to the same coordinate system. Futhermore, restriction (7.2) must hold in order to apply the series expansion (7.1) of the Green's function. For a given $\mathbf{r}$, this can always be arranged.

The solution for the fields can be cast in an appealing form by the introduction of the source-term vectors

$$\mathbf{S}^{(\lambda)}_{lm}=\sum_\tau S_{lm\tau\lambda}\mathbf{e}_\tau, \quad (7.3)$$

$$\mathbf{I}_{lm}^{(\lambda)} = \sum_{\tau} I_{lm\tau\lambda}\mathbf{e}_{\tau}. \tag{7.4}$$

In view of (5.1) and (5.2), these $\mathbf{S}_{lm}^{(\lambda)}$ and $\mathbf{I}_{lm}^{(\lambda)}$ are just the expansion coefficients of $f(\Omega)_{\lambda}\sigma(\Omega)_{\lambda}\mathbf{n}(\Omega)_{\lambda}$ and $f(\Omega)_{\lambda}\mathbf{i}(\Omega)_{\lambda}\times\mathbf{n}(\Omega)_{\lambda}$ after an expansion of these fields onto the set of spherical harmonics, but without a decomposition along the basis vectors $\mathbf{e}_{\tau}$. Futhermore, we define the vector

$$\mathbf{p}_{lm,l'm'}^{(\lambda)} = -i\int d\Omega f(\Omega)_{\lambda} Y(\Omega)_{lm} Y(\Omega)_{l'm'} \times h^{(1)}(k\xi(\Omega))_{l'}\mathbf{n}(\Omega)_{\lambda}, \tag{7.5}$$

which is a surface integral over the region λ. It is the integrated normal vector $\mathbf{n}(\Omega)_{\lambda}$ times the appropriate weight functions. This vector $\mathbf{p}_{lm,l'm'}^{(\lambda)}$ depends only on the shape of the surface. After these preliminary definitions, we can write for the reflected fields

$$\mathbf{E}(\mathbf{r}) - \mathbf{E}(\mathbf{r})^{\text{inc}} = \frac{k}{\epsilon}\sum_{lml'm'\lambda}\mathbf{p}_{lm,l'm'}^{(\lambda)}\cdot\mathbf{S}_{lm}^{(\lambda)}\nabla j(kr)_{l'}Y(\Omega)_{l'm'}^{*} - i\omega\mu k\sum_{lml'm'\lambda}\mathbf{p}_{lm,l'm'}^{(\lambda)}\times\mathbf{I}_{lm}^{(\lambda)}j(kr)_{l'}Y(\Omega)_{l'm'}^{*}, \tag{7.6}$$

$$\mathbf{B}(\mathbf{r}) - \mathbf{B}(\mathbf{r})^{\text{inc}} = \mu k\sum_{lml'm'\lambda}(\mathbf{p}_{lm,l'm'}^{(\lambda)}\times\mathbf{I}_{lm}^{(\lambda)})\nabla j(kr)_{l'}Y(\Omega)_{l'm'}^{*}. \tag{7.7}$$

These explicit expressions for the fields that are emitted by the surface charge and current distributions exhibit a clear separation between the source terms $\mathbf{S}_{lm}^{(\lambda)}$ and $\mathbf{I}_{lm}^{(\lambda)}$ and the redistribution, due to the surface geometry, which is accounted for by the vector $\mathbf{p}_{lm,l'm'}^{(\lambda)}$. The spatial distribution is represented as an expansion in the spherical waves $j(kr)_l Y(\Omega)_{lm}^{*}$ and $\nabla j(kr)_l Y_{lm}(\Omega)^{*}$.

VIII. SURFACE MULTIPOLES

We can elucidate the significance of the expansions (7.6) and (7.7) for the reflected fields by the introduction of the surface multipoles. To this end we define the multipolar moments of the charge and the current distributions as

$$C_{lm} = \frac{k}{\epsilon}\sum_{l'm'\lambda}\mathbf{p}_{l'm',lm}^{(\lambda)}\cdot\mathbf{S}_{l'm'}^{(\lambda)}, \tag{8.1}$$

$$\mathbf{J}_{lm} = \mu k\sum_{l'm'\lambda}\mathbf{p}_{l'm',lm}^{(\lambda)}\times\mathbf{I}_{l'm'}^{(\lambda)}, \tag{8.2}$$

where C_{lm} is a scalar and $\mathbf{J}_{lm}$ is a vector. These multipolar moments represent the charge and current distribution of the complete surface, not just in one region λ. The emitted fields now attain the form

$$\mathbf{E}(\mathbf{r}) - \mathbf{E}(\mathbf{r})^{\text{inc}} = \sum_{lm} C_{lm}\nabla j(kr)_l Y(\Omega)_{lm}^{*} - i\omega\sum_{lm}\mathbf{J}_{lm}j(kr)_l Y(\Omega)_{lm}^{*}, \tag{8.3}$$

$$\mathbf{B}(\mathbf{r}) - \mathbf{B}(\mathbf{r})^{\text{inc}} = \sum_{lm}\mathbf{J}_{lm}\times\nabla j(kr)_l Y(\Omega)_{lm}^{*}, \tag{8.4}$$

which greatly resembles the multipole expansion of the fields emitted by a charge and current distribution in a restricted region of space. The distinction is of course that the source terms C_{lm} and $\mathbf{J}_{lm}$ here gain contributions from everywhere in space, rather than from a localized area. This results effectively in an exchange of the spherical Bessel function $h^{(1)}(kr)_l$ with $j(kr)_l$ in the expansion of the Green's function.

The surface multipolar moments C_{lm} and $\mathbf{J}_{lm}$ are not independent. From the fact that the fields obey Maxwell's equations, as they do by construction, it follows that they are subject to some constraints. From $\nabla\cdot(\mathbf{E}(\mathbf{r}) - \mathbf{E}(\mathbf{r})^{\text{inc}}) = 0$ we readily derive the relation

$$i\sqrt{\epsilon\mu}\,C_{lm} = \frac{\sqrt{l}}{\sqrt{2l-1}}\sum_{\mu=-(l-1)}^{l-1}\sum_{\tau}(lm1\tau|l-1\mu)\mathbf{J}_{l-1,\mu}\cdot\mathbf{e}_{\tau}^{*} + \frac{\sqrt{l+1}}{\sqrt{2l+3}}\sum_{\mu=-(l+1)}^{l+1}\sum_{\tau}(lm1\tau|l+1\mu)\mathbf{J}_{l+1,\mu}\cdot\mathbf{e}_{\tau}^{*}, \tag{8.5}$$

for a spherical basis set $\mathbf{e}_{\tau}$. Here $(lm1\tau|l\pm1\mu)$ denotes a Clebsch–Gordan coefficient. The constraint (8.5) can be considered as the surface-integrated form of charge conservation ($\nabla\cdot\mathbf{j} = i\omega\rho$) for the surface charge density $\sigma(\mathbf{r})$.

IX. CONCLUSIONS

We have studied the charge and current distributions on the boundary of a perfect conductor with a dielectric, as they are confined and redistributed there by an externally applied electromagnetic field. The surface was allowed to have an arbitrary shape, and we did not impose any periodicity condition. We obtained closed-form and exact expressions for $\sigma(\mathbf{r})$ and $\mathbf{i}(\mathbf{r})$ everywhere on the surface. This was accomplished by deriving a set of inhomogeneous Fredholm equations of the second kind for $\sigma(\mathbf{r})\mathbf{n}(\mathbf{r})$ and $\mathbf{i}(\mathbf{r})\times\mathbf{n}(\mathbf{r})$ from Maxwell's equations, and subsequently solving these equations by an expansion on a discrete spherical set of basis vector functions. The solution involves surface-structure matrices, the R-matrices, which are independent of the incident field. It appears that an operation of a resolvent $(1-R)^{-1}$ on the vector representation of the impinging field on the surface yields the charge and current distributions. The Faraday induction between the $\mathbf{E}$ and the $\mathbf{B}$ fields gives rise to a coupling between the equations for $\sigma(\mathbf{r})\mathbf{n}(\mathbf{r})$ and $\mathbf{i}(\mathbf{r})\times\mathbf{n}(\mathbf{r})$, and it was accounted for by the matrix R [3].

Next, the structure of the fields, which are emitted by the oscillating charges and currents, was examined. The solution was cast in the form of a spherical multipolar expansion, and the multipolar moments were identified explicitly in terms of the solutions for $\sigma(\mathbf{r})$ and $\mathbf{i}(\mathbf{r})$. The effect of the surface geometry could be incorporated entirely by the application of a surface-integrated normal-direction matrix $\mathbf{p}_{lm,l'm'}^{(\lambda)}$. In addition, it was shown that the multipolar moments for the charge and current distributions are related, which reflects the charge conservation on the surface.

ACKNOWLEDGMENTS

This research was supported by the Office of Naval Research and the Air Force Office of Scientific Research

(AFSC), United States Air Force, under Contract No. F49620-86-C-0009.

APPENDIX: DISCONTINUOUS INTEGRALS

The integrals on the right-hand sides of Eqs. (3.1) and (3.3) are discontinuous if we pass $\mathbf{r}^+$ across the surface. Therefore, care should be taken in the evaluation of the limit $\mathbf{r}^+ \to \mathbf{r}$. In this appendix we give the details of the derivation of Eq. (3.1). Then the results (3.2) and (3.3) are obtained along similar lines. The limit to be found is

$$\text{Int} = \int dA'\, \sigma(\mathbf{r}') e^{ik|\mathbf{r}^+ - \mathbf{r}'|}(\mathbf{r}^+ - \mathbf{r}') \times \left\{ \frac{ik}{|\mathbf{r}^+ - \mathbf{r}'|^2} - \frac{1}{|\mathbf{r}^+ - \mathbf{r}'|^3} \right\}, \tag{A1}$$

with $\mathbf{r}^+ = \mathbf{r} + \mathbf{n}(\mathbf{r})\delta$ and $\delta \downarrow 0$. To this end we divide the surface into a small circle with radius R and around $\mathbf{r}$ and the remainder of the surface. This is illustrated in Fig. 2. For the integration over the region outside the circle, the integrand has no singularities, and we can replace $\mathbf{r}^+$ by $\mathbf{r}$. Inside the circle, however, the factor in curly brackets is singular for $\mathbf{r}^+ \to \mathbf{r}'$. This implies that we have to carry out the integration before we take the limit $\delta \downarrow 0$. This can be done as follows. First, for $\mathbf{r}'$ inside the circle we can write

$$\sigma(\mathbf{r}') \simeq \sigma(\mathbf{r}), \tag{A2}$$

$$e^{ik|\mathbf{r}^+ - \mathbf{r}'|} \simeq 1, \tag{A3}$$

since these functions vary negligibly over the singularity. Next, we write $\mathbf{r}^+ - \mathbf{r}' = (\mathbf{r} - \mathbf{r}') + \mathbf{n}(\mathbf{r})\delta$ for the vector in front of the brackets. Then we notice that the integral with $\mathbf{r} - \mathbf{r}'$ vanishes because of the cancellation of contributions from $\mathbf{b}$ and $-\mathbf{b}$ (see Fig. 2). This component disappears for every δ, and therefore also in the limit $\delta \downarrow 0$, which leaves us with

$$\text{Int} = \int dA'\, \sigma(\mathbf{r}') \nabla G(\mathbf{r},\mathbf{r}') + \sigma(\mathbf{r})\mathbf{n}(\mathbf{r})\delta \int_{\substack{\text{inside} \\ \text{circle}}} dA' \left\{ \frac{ik}{|\mathbf{r}^+ - \mathbf{r}'|^2} - \frac{1}{|\mathbf{r}^+ - \mathbf{r}'|^3} \right\}. \tag{A4}$$

From Fig. 2 we see that $|\mathbf{r}^+ - \mathbf{r}'|^2 = |\mathbf{r} - \mathbf{r}'|^2 + \delta^2$. After substitution into the integrand, the integration is most easily carried out in polar coordinates, which yields

$$\delta \int_{\substack{\text{inside} \\ \text{circle}}} dA' \left\{ \frac{ik}{|\mathbf{r}^+ - \mathbf{r}'|^2} - \frac{1}{|\mathbf{r}^+ - \mathbf{r}'|^3} \right\} = 2\pi \{ \tfrac{1}{2} ik\delta \log(R^2 + \delta^2) - ik\delta \log \delta + \delta/(R^2 + \delta^2)^{1/2} - 1 \}. \tag{A5}$$

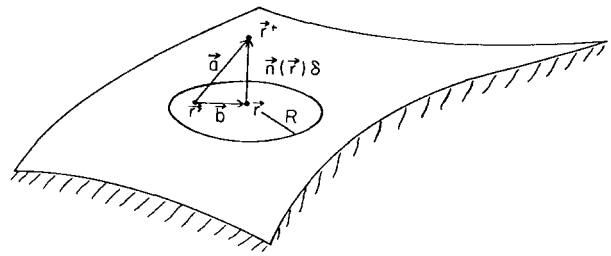


FIG. 2. Geometry for the evaluation of the limit $\mathbf{r}^+ \to \mathbf{r}$. Around $\mathbf{r}$ on the surface, we divide the surface into an infinitesimal circle of radius R, and the rest of the surface. Then the integrals are split up accordingly. The normal vector points from the surface into the dielectric and is multiplied by $\delta > 0$. We denoted $\mathbf{r}^+ - \mathbf{r}'$ by $\mathbf{a}$ and $\mathbf{r} - \mathbf{r}'$ by $\mathbf{b}$. The limit $\mathbf{r}^+ \to \mathbf{r}$ implies $R \gg \delta > 0$ and $R \to 0$. It appears that an integral over the small circle remains finite whenever the gradient of the Green's function occurs in the integrand.

In the limit $R \gg \delta > 0$ and $R \to 0$, this integral acquires the finite value of -2π, and its combination with Eq. (A4) gives expression (3.1), which was to be proved.

[1] S. R. J. Brueck and D. J. Ehrlich, Phys. Rev. Lett. **48**, 1678 (1982).
[2] A. G. Ramm and M. A. Fiddy, Opt. Commun. **56**, 8 (1985).
[3] P. J. Feibelman, Phys. Rev. B **12**, 1319 (1975).
[4] P. J. Feibelman, Phys. Rev. B **12**, 4282 (1975).
[5] H. J. Juranek, Z. Phys. **233**, 324 (1970).
[6] J. P. Marton and J. R. Lemon, Phys. Rev. B **4**, 271 (1971).
[7] J. J. Sein, Opt. Commun. **14**, 157 (1975).
[8] G. S. Agarwal, Opt. Commun. **14**, 161 (1975).
[9] A. A. Maradudin and D. L. Mills, Phys. Rev. B **11**, 1392 (1975).
[10] V. Celli, A. Marvin, and F. Toigo, Phys. Rev. B **11**, 1779 (1975).
[11] A. Marvin, F. Toigo, and V. Celli, Phys. Rev. B **11**, 2777 (1975).
[12] G. S. Agarwal, Phys. Rev. B **14**, 846 (1976).
[13] F. Toigo, A. Marvin, V. Celli, and N. R. Hill, Phys. Rev. B **15**, 5618 (1977).
[14] N. Garcia and N. Cabrera, Phys. Rev. B **18**, 576 (1978).
[15] B. Laks, D. L. Mills, and A. A. Maradudin, Phys. Rev. B **23**, 4965 (1981).
[16] N. E. Glass and A. A. Maradudin, Phys. Rev. B **24**, 595 (1981).
[17] P. Sheng, R. S. Stepleman, and P. N. Sanda, Phys. Rev. B **26**, 2907 (1982).
[18] N. Garcia and A. A. Maradudin, Opt. Commun. **45**, 301 (1983).
[19] N. Garcia, Opt. Commun. **45**, 307 (1983).
[20] M. Weber and D. L. Mills, Phys Rev. B **27**, 2698 (1983).
[21] N. E. Glass, M. Weber, and D. L. Mills, Phys. Rev. B **29**, 6548 (1984).
[22] K. T. Lee and T. F. George, Phys. Rev. B **31**, 5106 (1985).
[23] S. S. Jha, J. R. Kirtley, and J. C. Tsang, Phys. Rev. B **22**, 3973 (1980).
[24] M. Nevière and R. Reinisch, Phys. Rev. B **26**, 5403 (1982).
[25] X. Y. Huang, K. T. Lee, and T. F. George, J. Chem. Phys. **85**, 567 (1986).
[26] D. Agassi, Phys. Rev. B **33**, 2393 (1986).
[27] G. S. Agarwal and C. V. Kunasz, Phys. Rev. B **26**, 5832 (1982).
[28] K. C. Liu and T. F. George, Phys. Rev. B **32**, 3622 (1985).
[29] A. A. Maradudin, in *Surface Polaritons*, edited by V. M. Agranovich and A. A. Maradudin (North-Holland, Amsterdam, 1982), pp. 405.
[30] P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill, New York, 1953), pp. 887.

On the electromagnetic field and the Teukolsky–Press relations in arbitrary space-times

Bartolomé Coll
Chaire de Physique Mathématique, Collège de France, Paris, France
Francesc Fayos
Departament de Física de E.T.S.A.B., Universitat Politècnica de Catalunya, Barcelona, Spain
Joan Josep Ferrando
Departament de Física Teòrica, Facultat de Físiques, Burjassot (València), Spain

(Received 23 July 1986; accepted for publication 22 October 1986)

The relations on the electromagnetic field obtained by Teukolsky and Press for type D vacuum space-times are considered; these are four second-order equations in two complex components of the field with respect to a principal null tetrad. A rigorous geometric interpretation of these relations is given, showing the essential role played by the Maxwellian character of the basic null tetrad. It appears that, generically, the Teukolsky–Press relations are incomplete. Once completed, their generalizations to the general Maxwell equations (with source term) with respect to non-necessarily Maxwellian tetrads on arbitrary space-times are given.

I. INTRODUCTION

(a) Let ϕ_0, ϕ_1, ϕ_2 be the components of the electromagnetic field F with respect to a principal null tetrad in a type D vacuum space-time. Here we call *Teukolsky–Press relations*[1] the following set of four second-order partial differential equations in the two components ϕ_0, ϕ_2:

$$\begin{aligned} &T_0\colon\ \tau_0\phi_0 = 0, \quad T_2\colon\ \tilde{\tau}_0\phi_2 = 0, \\ &\underline{T}_0\colon\ \underline{\tau}_0\phi_0 + \underline{\tau}_2\phi_2 = 0, \quad \underline{T}_2\colon\ \tilde{\underline{\tau}}_0\phi_2 + \tilde{\underline{\tau}}_2\phi_0 = 0, \end{aligned} \tag{1}$$

where the τ's are given, in the Newman–Penrose notation, by

$$\begin{aligned} \tau_0 &\equiv (D - \epsilon + \bar{\epsilon} - 2\rho - \bar{\rho})(\Delta + \mu - 2\gamma) \\ &\quad - (\delta - \beta - \bar{\alpha} - 2\tau + \bar{\pi})(\bar{\delta} + \pi - 2\alpha), \\ \underline{\tau}_0 &\equiv (\bar{\delta} + 3\pi - \bar{\beta} - \alpha)(\bar{\delta} + \pi - 2\alpha), \\ \underline{\tau}_2 &\equiv (D + 3\rho + \epsilon - \bar{\epsilon})(D - \rho + 2\epsilon), \end{aligned} \tag{2}$$

and $\sim$ is the operator which permutes separately the real and complex vectors of the null tetrad. The two uncoupled equations in (1), T_0 and T_2, were first given by Teukolsky[2]; the remaining two, $\underline{T}_0$ and $\underline{T}_2$, by Teukolsky and Press.[3]

The Teukolsky–Press relations were the starting point to show[3] that the Maxwell equations can be integrated by separation of variables in perturbed Kerr geometries. For this reason, they play an important role in many problems related to the Kerr space-times. It is the case, in particular, in the problem of the perturbations of a Kerr black hole by incident electromagnetic waves, first considered by Starobinsky and Churilov,[4] which could be studied in detail (see Chandrasekhar[5]).

But, in spite of their simple derivation, the Teukolsky–Press relations are not easy to interpret: derived from the Maxwell equations, one does not know, conversely, to what extent the Maxwell equations are implied by them.

On the other hand, some authors[6,7] have given Teukolsky-Press-like relations in the Kerr-Newman space-times, but the precise conditions under which the Teukolsky-Press relations may be generalized to other space-times have not yet been found.

This paper answers both problems: we find a rigorous geometric interpretation of the Teukolsky–Press relations and their connection with the Maxwell equations, and we give their generalizations to arbitrary null tetrads and arbitrary space-times.

(b) For this task, we need two important notions: those of *Maxwellian structure* and of *conditional system* associated to a given differential system.

It is weli known that a electromagnetic field (arbitrary two-form) selects algebraically, at every point of the space-time, a pair of orthogonal two-planes which, in the regular case, define a 2 + 2 almost-product structure.[8–10] The *Maxwellian structures* are the 2 + 2 almost-product structures defined by the regular solutions to the vacuum Maxwell equations.

On the other hand, let $D_1(\phi_0,\phi_1,\phi_2)$ and $D_2(\phi_0,\phi_2)$ be two differential systems in the ϕ's. We shall say that D_2 is a *conditional system* for D_1 if all their solutions (ϕ_0,ϕ_2) may be completed to solutions (ϕ_0,ϕ_1,ϕ_2) of D_1 and if, conversely, all the solutions (ϕ_0,ϕ_1,ϕ_2) of D_1 are such that (ϕ_0,ϕ_2) are solutions of D_2.

We shall see here that *the Maxwell equations always admit a conditional system in* (ϕ_0,ϕ_2) that is, generically, of third order. Moreover, *this system degenerates to a second-order system if, and only if, the basic null tetrad is associated naturally to a Maxwellian structure.*

The principal null tetrads of the type D vacuum space-time are associated to a Maxwellian structure. Consequently, the conditional system admitted there by the Maxwell equations is a second-order one. Then, its comparison with Eqs. (1) and (2) shows that, up to a missing equation, *the Teukolsky–Press relations on the type D vacuum space-times are nothing but the conditional system in* (ϕ_0,ϕ_2) *admitted by the Maxwell equations.*

The missing equation in the Teukolsky–Press relations

 0022-2488/87/051075-05$02.50

is identically verified on the solutions of the Maxwell equations which are invariant under the isometry group of the Kerr metric. This is, perhaps, the reason why this equation has been omitted up to now: the Maxwell solutions usually considered in this context belong essentially to this class.

On the basis of our preceding results, the generalization of the Teukolsky–Press relations to any null tetrad in any space-time must be considered as given by the conditional system in (ϕ_0,ϕ_2) associated there to the Maxwell equations.

It will be then easy to characterize the type D nonvacuum space-times in which the first two Teukolsky–Press equations remain uncoupled.

(c) The paper is organized as follows: in Sec. II we introduce "*à la* Rainich" the notion of Maxwellian structure and then, give its version in the complex formalism. Section III is devoted to finding the conditional systems in (ϕ_0,ϕ_2) admitted by the Maxwell equations, and Sec. IV gives its explicit expression in terms of the spin coefficients. Finally, in Sec. V, we compare them with the Teukolsky–Press relations and discuss the remainder of the results stated in the precedent paragraph (b).

This paper contains some results published elsewhere,[11] but here we consider the general Maxwell equations (with source term), obtain the explicit form of the third-order conditional system, and give detailed proofs of our statements.

II. MAXWELLIAN STRUCTURES

(a) Let Ω be a domain of the space-time (V_4,g), g being a Lorentzian metric of signature -2. To every two-form F is associated the *Minkowski stress-energy tensor* T, given by $2T\equiv F^2+(*F)^2$ with $F^2\equiv F\times F$, $\times$ being the cross product,[12] and $*$ denoting the Hodge dual operator.[13] The tensor T verifies $T^2=\chi^2 g$, where χ is nonzero if, and only if, F is *regular*.[14] In this section we shall consider only regular two-forms, so that the tensor $P\equiv\chi^{-1}T$ defines a $2+2$ almost-product structure. Let G be the simple unit two-form characterizing the field of timelike two-planes of the structure

$$\tfrac{1}{2}\mathrm{tr}\,G^2=1,\quad \mathrm{tr}\,{*G}\times G=0,\quad P=G^2+(*G)^2, \tag{3}$$

tr being the trace operator; the field of spacelike two-planes is then characterized by $*G$, and one has[15]

$$F=e^{\phi+*\psi}G=e^{\phi}(\cos\psi G+\sin\psi *G), \tag{4}$$

where $2\phi=\ln 2\chi$. Every regular two-form F is thus biunivocally characterized by its *components* $\{G,\phi,\psi\}$. Note that, given the *geometric component* G, the *energetic component* ϕ determines the norm of the eigenvalues of T, and both, G and ϕ, characterize T itself. Finally, among all the two-forms associated with a given T, the *Rainich component* ψ selects, by a duality rotation, the particular two-form F.

(b) In terms of these components, the vacuum Maxwell equations for F,

$$\delta F=0,\quad \delta *F=0, \tag{5}$$

may be written[9]

$$d\phi=\Phi,\quad d\psi=\Psi, \tag{6}$$

where the one-forms Φ and Ψ are functionals of the sole geometric component G:

$$\begin{aligned}\Phi&\equiv *(\delta G\wedge *G+\delta *G\wedge G),\\ \Psi&\equiv *(\delta G\wedge G-\delta *G\wedge *G),\end{aligned} \tag{7}$$

δ being the codifferentiation operator and "$\wedge$" denoting the exterior product.[16]

From (6), the *Rainich Theorem*[8] follows: A simple and unitary two-form G is the geometric component of a (local) solution to the vacuum Maxwell equations if, and only if, it verifies the equations

$$d\Phi=0,\quad d\Psi=0. \tag{8}$$

The almost-product structures defined by such simple and unitary solutions to Eq. (8) will be called *Maxwellian structures*.

To every Maxwellian structure, say G, the first of relations (6), $d\phi=\Phi$, associates a (one-parameter, additive) family of energetic components ϕ, characterizing a (homothetic) family of energy tensors T. In fact, it may be shown that this relation is strictly equivalent to the conservation equation $\delta T=0$ (Ref. 17). In a similar way, the second of the relations (6), $d\psi=\Psi$, associates with G a (one-parameter, additive) family of Rainich components ψ, characterizing (up to a homothecy) a family of two-forms F related by a constant duality rotation. In fact, it may be shown that this relation is strictly equivalent to the Rainich's complexion equation. The set $\{G,\phi,\psi\}$ then defines the two-parameter family of solutions to the Maxwell equations having the same almost-product structure.

(c) The form (6) of the Maxwell equations may be easily obtained in the complex vectorial formalism.[18] Let us consider the complex two-forms Z^I $(I=0,1,2)$ given by

$$Z^0=\bar m\wedge n,\quad Z^1=n\wedge l-\bar m\wedge m,\quad Z^2=l\wedge m,$$

where $\{l,n,m,\bar m\}$ is a complex null tetrad; since the Z^I's are self-duals, $*Z^I=iZ^I$, the basis $\{Z^I,\bar Z^I\}$ of the complex two-forms separates invariantly the self-dual and anti-self-dual parts of every two-form W: $W=W_IZ^I+\underline{W}_I\bar Z^I$. In particular, for every real two-form F, the complex two-form $\hat F\equiv F-i*F$ is self-dual and its components in the basis $\{Z^I\}$ will be designed by ϕ_I: $\hat F=\phi_IZ^I$ $(I=1,2,3)$.

The *general Maxwell equations* $\delta F=J$, $\delta *F=0$, now may be written in the form

$$J=\delta\hat F=\delta\{\phi_IZ^I\}=\phi_1\delta Z^1-i(d\phi_1)Z^1+\delta H.$$

Contracting by Z^1 and taking into account that $Z^1\times Z^1=g$, one finds that *the general Maxwell equations are equivalent to the system*

$$d\phi_1=\phi_1h+\omega, \tag{9}$$

where

$$h\equiv i(\delta Z^1)Z^1,\quad \omega\equiv i(\delta H-J)Z^1,\quad H\equiv\phi_0Z^0+\phi_2Z^2. \tag{10}$$

In order to formulate the Rainich Theorem in this formalism, let us consider the almost-product structure associated to a null tetrad, defined by the element Z^1 of the corresponding self-dual basis, $Z^1=G-i*G$. The two-form G is the geometric component of every two-form F_0 having the expression (4), and one has $\hat F_0=\phi_1^0Z^1$ with $\phi_1^0=e^{\phi+i\psi}$, that is $H=0$: The vacuum Maxwell equations for F_0 are then

$$d \ln \phi_1^0 = h, \tag{11}$$

and their (local) integrability condition is

$$dh = 0. \tag{12}$$

Expressing Z^1 in terms of G in the definition (10) of h, and taking into account that $*(v \wedge A) = -(-1)^p i(v) *A$ for every one-form v and every p-form A, one finds $h = \Phi + i*\Psi$, where Φ and Ψ are the functionals of G given by (7). We then have the following.

Proposition 1 (Rainich Theorem): An almost-product structure Z^1 is (locally) Maxwellian iff the one-form $h \equiv i(\delta Z^1) Z^1$ is closed: $dh = 0$.

Let us consider the component $(dh)_1$ of dh in the basis $\{Z^I, \bar{Z}^I\}$. From the identity $(A, dv) = \delta i(v) A + i(v) \delta A$ and the orthogonal properties of the Z^I's, one has

$$-2(dh)_1 = (Z^1, dh) = \delta i(h) Z^1 + i(h) \delta Z^1$$
$$= \delta^2 Z^1 + i^2(\delta Z^1) Z^1 = 0,$$

and thus we have the following.

Proposition 2: The differential system $dh = 0$ characterizing the Maxwellian structures consists of five second-order complex equations in Z^1.

Considered as equations on the spin coefficients of a null complex tetrad compatible with the Maxwellian structure, they are first-order equations; their explicit expression may be found elsewhere.[19]

III. CONDITIONAL SYSTEMS FOR THE MAXWELL EQUATIONS

(a) The differential system (8) defining the Maxwellian structures is satisfied by the component G of all the solutions (ϕ, ψ, G) to the vacuum Maxwell equations (6) and, conversely, all his solutions G may be completed to solutions (ϕ, ψ, G) to the Maxwell system. In other words, in order that the Maxwell equations, considered as an (overdetermined) system in the two unknowns ϕ and ψ, be compatible, it is necessary and sufficient that the system (8) in G holds. We give the following definition.

Definition: Let $D_1(x, y)$ and $D_2(y)$ be two differential systems in p unknowns x and q unknowns y; let $S_1 \subset F^{p+q}$ and $S_2 \subset F^q$ be their corresponding spaces of solutions, and let $\pi: F^{p+q} \to F^q$, $(x, y) \mapsto (y)$ be the natural projection. We shall say that D_2 is a conditional system in the y's for D_1 if $\pi(S_1) = S_2$.

Thus, the Rainich Theorem may be equivalently enounced by saying that *the Maxwell equations admit a second-order conditional system in G.*

(b) Let us now consider the general Maxwell equations (9) in the unknowns ϕ_I. By differentiation, one has

$$0 = d\phi_1 \wedge h + \phi_1 \, dh + d\omega,$$

and, taking into account (9), it follows that

$$\Omega + \phi_1 \, dh = 0, \tag{13}$$

where

$$\Omega \equiv d\omega + \omega \wedge h. \tag{14}$$

Thus when dh does not vanish, a necessary condition for the existence of ϕ_1 is that the two-forms Ω and dh be proportional:

$$\Omega \otimes dh = dh \otimes \Omega. \tag{15}$$

In such a case, the sufficient conditions are obtained by imposing that the proportionality factor between both two-forms be effectively a solution of Eq. (9). These conditions are first-order equations in Ω or, from (14), third-order equations in (ϕ_0, ϕ_2). When dh vanishes, we have (locally) $h = d \ln \phi_1^0$ and Eqs. (9) may be written $d(\phi_1/\phi_1^0) = \omega/\phi_1^0$, whose integrability conditions are $\Omega = 0$. We have shown the following.

Theorem 1: The Maxwell equation always admit a conditional system in (ϕ_0, ϕ_2). It is, generically, a third-order system, and it reduces to a second-order one if, and only if, the almost-product structure associated to the self-dual basis is Maxwellian: $dh = 0$. In such a case, the system is given by

$$\Omega(\phi_0, \phi_2) = 0. \tag{16}$$

For every solution (ϕ_0, ϕ_2) to (16), there exists a family of functions which complete it to solutions to the Maxwell equations. If ϕ_1 is such a function, all the others are of the form $\phi_1 + \phi_1^0$ where ϕ_1^0 is the general solution for the electromagnetic fields admitting Z^1 as the complex geometric component.

(c) Now consider a non-Maxwellian geometry and let X be any two-form such that $(X, dh) \neq 0$. If Eqs. (15) are verified, then, according to (13), we have

$$\phi_1 = -(\Omega, X)/(dh, X), \tag{17}$$

and Eqs. (9) impose

$$(\Omega, X) d(dh, X) - (dh, X) d(\Omega, X)$$
$$= -(\Omega, X)(dh, X) h + (dh, X)^2 \omega.$$

After rearranging terms, these equations may be written in the form

$$i(X) i'(X) \{\Omega \otimes \nabla \, dh - dh \otimes \nabla \Omega$$
$$+ \Omega \otimes h \otimes dh - dh \otimes \omega \otimes dh\}$$
$$+ \{i(\Omega) i'(dh) - i(dh) i'(\Omega)\}(X \otimes \nabla X) = 0, \tag{18}$$

where $i(\,)$ and $i'(\,)$ denote, respectively, contraction over the first and last two-form elements of the tensorial basis. From Eqs. (15) and their covariant derivatives, it follows, respectively, that the term in $X \otimes \nabla X$ vanishes and that the tensor between brackets, which is in $\Lambda^2 \otimes T^* \otimes \Lambda^2$, is symmetric in their antisymmetric components. Thus, as (18) must be verified for any two-form X, we have the following.

Theorem 2: The third-order conditional system in (ϕ_0, ϕ_2) for the Maxwell equations is given by

$$\Omega \otimes \nabla dh - dh \oplus \{\nabla \Omega - h \otimes \Omega + \omega \otimes dh\} = 0. \tag{19}$$

To every solution (ϕ_0, ϕ_2) to this system, corresponds a unique solution (ϕ_0, ϕ_1, ϕ_2) to the Maxwell equations, the ϕ_1 being given by (17).

IV. THE SECOND-ORDER CONDITIONAL SYSTEM IN THE SPIN COEFFICIENTS' FORMALISM

(a) Let $\{\Omega_I, \underline{\Omega}_I\}$ be the components of the two-form Ω with respect to the chosen self-dual basis $\{Z^I, \bar{Z}^I\}$. From the orthogonality properties $(Z^0, Z^2) = 1, (Z^1, Z^1) = -2$, and the definition (14) of Ω, we have, for the component Ω_1,

$$-2\Omega_1 = (Z^1,\Omega) = (Z^1,\omega\wedge h) + (Z^1,d\omega)$$
$$= -i(\omega)i(h)Z^1 + \delta i(\omega)Z^1 + i(\omega)\delta Z^1,$$

where the adjoint character of $i(\ )$ (resp. δ) with respect to $\wedge$ (resp. d) has been taken into account. But, from $Z^1\times Z^1 = g$ and the definition (10) of h and ω, it follows $i(h)Z^1 = \delta Z^1$ and $i(\omega)Z^1 = \delta H - J$ so that we have

$$-2\Omega_1 = -i(\omega)\delta Z^1 + \delta(\delta H - J) + i(\omega)\delta Z^1 = 0.$$

Consider now the component Ω_0 and Ω_2: the second-order terms in ϕ_0 and ϕ_2 come from $d\omega$ or, according to the definitions (10) of ω and H and the orthogonality properties of the basis Z^I, from the antisymmetrization of $\nabla d\phi_0\times Z^0 + \nabla d\phi_2\times Z^2$; but $Z^0\times Z^0 = Z^2\times Z^2 = 0$, $\Omega_0 = (\Omega,Z^2)$, and $\Omega_2 = (\Omega,Z^0)$ so that Ω_0 (resp. Ω_2) does not depend on the second-order derivatives of ϕ_2 (resp. ϕ_0). On the other hand, it is clear that Ω depends at most on the first derivatives of J, and, finally, denoting by $\sim$ the operator which permutes separately the real and complex vectors of the null tetrad, $\sim^2 = \mathrm{Id}$, $\tilde{Z}^1 = -Z^1$, $\tilde{Z}^0 = -Z^2$, one has $\tilde{J} = J$, $\tilde{\phi}_0 = -\phi_2$ so that, from the definitions (10) of h and ω, it follows that $\tilde{h} = h$ and $\tilde{\omega} = -\omega$ and, consequently, $\tilde{\Omega} = -\Omega$. Taking into account all these results, we have the following.

Proposition 3: The second-order conditional system in (ϕ_0,ϕ_2) for the general Maxwell equations is of the form

$$\begin{aligned}
&-\Omega_0 \equiv D_0\phi_0 + D_2\phi_2 + \mathscr{J}_0 = 0,\\
&-\underline{\Omega}_0 \equiv \underline{D}_0\phi_0 + \underline{D}_2\phi_2 + \mathscr{J}_2 = 0,\\
&\Omega_2 \equiv \tilde{D}_0\phi_2 + \tilde{D}_2\phi_0 - \tilde{\mathscr{J}}_0 = 0,\\
&\underline{\Omega}_2 \equiv \tilde{\underline{D}}_0\phi_2 + \tilde{\underline{D}}_2\phi_0 - \tilde{\mathscr{J}}_2 = 0,\\
&\tfrac{1}{2}(\Omega_1 + \underline{\Omega}_1) \equiv D_1\phi_2 - \tilde{D}_1\phi_0 - \mathscr{J}_1 = 0,
\end{aligned} \tag{20}$$

where D_2 is a first-order derivation operator and the $\mathscr{J}_I$'s are functions of J and its first derivatives.

(b) In order to obtain the explicit expression for the components (20) of the two-form Ω, in terms of the spin coefficients and the directional derivatives associated to the null tetrad, we need of some intermediate expressions. The evaluation of the codifferentials of the Z^I's, which may be easily performed using Ref. 18, gives

$$\begin{aligned}
\delta Z^0 &= 2i(\sigma_1)Z^0 + i(\sigma_2)Z^1,\\
\delta Z^1 &= -2i(\sigma_0)Z^0 + 2i(\sigma_2)Z^2,\\
\delta Z^2 &= -i(\sigma_0)Z^1 - 2i(\sigma_1)Z^2,
\end{aligned}$$

where the σ_I's denote the following one-forms[18]:

$$\begin{aligned}
\sigma_0 &= \tau l + \kappa n - \rho m - \sigma\bar{m},\\
\sigma_1 &= \gamma l + \epsilon n - \alpha m - \beta\bar{m},\\
\sigma_2 &= \nu l + \pi n - \lambda m - \mu\bar{m}.
\end{aligned}$$

The codifferentials of the tetrad one-form are

$$\delta l = -(\epsilon+\bar{\epsilon}) + (\rho+\bar{\rho}),\quad \delta n = (\gamma+\bar{\gamma}) - (\mu+\bar{\mu}),$$
$$\delta m = -\bar{\pi} + \tau + \bar{\alpha} - \beta,\quad \delta\bar{m} = -\pi + \bar{\tau} + \alpha - \bar{\beta},$$

and the action of the operator $\sim$ on the σ_I's is

$$\tilde{\sigma}_1 = -\sigma_1,\quad \tilde{\sigma}_0 = -\sigma_2.$$

Following Crossman and Fackerell,[6] we write

$$D^{rs}_{pq} = D + (p-1)\epsilon - (q+1)\rho + (r-1)\bar{\epsilon} - s\bar{\rho},$$
$$\delta^{rs}_{pq} = \delta + (p-1)\beta - (q+1)\tau - (r-1)\bar{\alpha} + s\bar{\pi},$$

and denote by Δ^{rs}_{pq} and $\bar{\delta}^{rs}_{pq}$, respectively, the transforms of D^{rs}_{pq} and δ^{rs}_{pq} by the operator $\sim$.

Taking into account the above expressions, the computation of relations (20) is not a difficult task; denoting by

$$\tau_0 = \delta^{21}_{01}\bar{\delta}^{10}_{30} - D^{21}_{01}\Delta^{10}_{30},\quad \underline{\tau}_2 = -D^{00}_{22}D^{10}_{30},$$
$$\underline{\tau}_0 = \bar{\delta}^{00}_{22}\bar{\delta}^{10}_{30},\quad \tau_1 = D^{20}_{21}\delta^{10}_{30} - (\tau+\bar{\pi})D^{10}_{30},$$

the second-order operators on the ϕ_I's, the result is the following.

Theorem 3: In any space-time, the second-order conditional system in (ϕ_0,ϕ_2) of the general Maxwell equations, $\Omega = 0$, is of the form (20), where

$$\begin{aligned}
D_0 &= \tau_0 - \kappa\nu + \sigma\lambda,\\
D_2 &= -2\kappa\delta^{3/2\,1/2}_{3/2\,1/2} + 2\sigma D^{3/2\,1/2}_{3/2\,1/2} - \delta\kappa + D\sigma,\\
\underline{D}_0 &= \underline{\tau}_0 - \lambda D^{00}_{22} - \bar{\sigma}\Delta^{10}_{30} - \bar{\kappa}\nu - D\lambda,\\
\underline{D}_2 &= \underline{\tau}_2 - \kappa\bar{\delta}^{00}_{22} - \bar{\kappa}\delta^{10}_{30} - \sigma\bar{\sigma} - \bar{\delta}\kappa,\\
D_1 &= \tau_1 + \kappa\Delta^{20}_{21} + \sigma(\pi+\bar{\tau}) + \Delta\kappa,\\
\mathscr{J}_0 &= \kappa J^1 + \delta^{21}_{01}J^2 + \sigma J^3 + D^{21}_{01}J^4,\\
\mathscr{J}_2 &= \bar{\kappa}J^1 + \bar{\sigma}^{00}_{22}J^2 - D^{00}_{22}J^3 - \bar{\sigma}J^4,\\
\mathscr{J}_1 &= D^{20}_{21}J^1 + \Delta^{20}_{21}J^2 + (\bar{\pi}+\tau)J^3 - (\pi+\bar{\tau})J^4.
\end{aligned} \tag{21}$$

V. THE TEUKOLSKY–PRESS RELATIONS AND THEIR GENERALIZATIONS

(a) Let us consider, on any type D vacuum space-time, the null tetrads associated to the Bel directions[20] (*principal null tretrads*). In the Newman–Penrose formalism,[21] we have

$$\begin{aligned}
&\Psi_0 = \Psi_1 = \Psi_3 = \Psi_4 = 0,\\
&\kappa = \nu = \sigma = \lambda = 0,
\end{aligned} \tag{22}$$

and the Bianchi identities become

$$d\Psi_2 = 3\Psi_2\cdot h. \tag{23}$$

Then one has $dh = 0$ and thus, according to Proposition 1, the almost-product structure associated to the null tetrads is Maxwellian.

For such space-times, the four Teukolsky–Press relations may be written in the form (1) with the values (2) of the τ's. On the other hand, the evaluation of Eqs. (20) under the hypothesis (22), leads, *in the source free case* $J = 0$, to the equations

$$\Omega_A = T_A,\quad \underline{\Omega}_A = \underline{T}_A, \tag{24}$$

for $A = 0,2$ and

$$\tfrac{1}{2}(\Omega_1 + \underline{\Omega}_1) = \tau_1\phi_2 - \tilde{\tau}_1\phi_0 = 0, \tag{25}$$

for $A = 1$. Thus we have the following.

Theorem 4: On type D vacuum space-times, the conditional system in (ϕ_0,ϕ_2) for the source-free Maxwell equations, associated with the principal null tetrads, consists of the Teukolsky–Press relations (1) completed with the relation (25).

From (20) and (21) it is easy to see that (24) holds iff relations (22) hold. For *any* type D space-time, we have the following.

Proposition 4: The first two Teukolsky–Press relations

T_0 and T_2 decouple if, and only if, the Bel directions of the space-time are geodesic, are shear-free, and define a Maxwellian structure.

(b) In the particular case of the Kerr metric, the first two equations (1), in addition to being uncoupled in ϕ_0 and ϕ_2, may be separated into radial and angular parts relative to the Boyer–Linquist coordinates for the electromagnetic fields which are invariant[22] under the action of the two-dimensional isometry group. For these fields, the fifth equation (25) is identically satisfied when the first four equations (24) hold. This is perhaps the reason why Eq. (25) has not been (apparently) considered up to now. But if, in the same geometric context, one wishes to consider, for example, nonperiodic time-dependent electromagnetic fields, then Eq. (25) must be necessarily added to the usual Teukolsky–Press relations (24) in order to insure the existence of ϕ_1.

(c) Theorem 4 shows that, *once completed, the natural geometric generalization of the Teukolsky–Press relations is our conditional system in* (ϕ_0,ϕ_2). This is a manifold generalization: the second-order conditional system (20) extends the validity of the Teukolsky–Press relations, step by step, to noninvariant fields, to nonprincipal tetrads, to non-source-free Maxwell equations, and to arbitrary space-times. Finally, when the chosen null tetrads do not define a Maxwellian structure, the third-order conditional system (19) must be used instead of the second-order one.

[1]In preceding papers (see Ref. 11) we called them *Teukolsky relations*. The present appellation seems more correct.

[2]S. A. Teukolsky, J. Appl. Phys. **185**, 635 (1973).

[3]S. A. Teukolsky and W. H. Press, J. Appl. Phys. **193**, 443 (1974).

[4]A. A. Starobinsky and S. M. Churilov, Sov. Phys. JETP **38**, (1973).

[5]S. Chandrasekhar, *The Mathematical Theory of Black Holes* (Oxford U. P., London, 1983).

[6]R. G. Crossman and E. D. Fackerell, *Proceedings of the Summer School on Gravitational Radiation*, edited by C. Edwards (Springer, Berlin, 1980).

[7]V. Bellezza and V. Ferrari, J. Math. Phys. **25**, 1985 (1984).

[8]G. Y. Rainich, Trans. Am. Math. Soc. **27**, 106 (1925).

[9]R. Debever, *Colloquium Théorie de la Relativité* (C.B.R.M., Bruxelles, 1959).

[10]L. T. Phong, Ann. Inst. Fourier, Grenoble **14**, 269 (1964).

[11]B. Coll, F. Fayos, and J. J. Ferrando, C. R. Acad. Sci. Paris Ser. I, **300**, 699 (1985).

[12]The cross product is here defined as the contraction over the inner base elements of the tensorial product. In local charts: $(A \times B)_{\alpha\beta} = A_{\alpha\rho} B^{\rho}_{\beta}$.

[13]Our convention is $*\equiv(\eta,\cdot)$ where η is the volume element and (B,A) denotes the inner product, given in local charts by $p!(B,A)_{\alpha_1\cdots\alpha_q} = B_{\alpha_1\cdots\alpha_q\beta_1\cdots\beta_p} A^{\beta_1\cdots\beta_p}$.

[14]A. Lichnerowicz, Ann. Mat. Pura. Appl. **50**, 1 (1960).

[15]The exponential notation in (4) is due to C. W. Misner and J. A. Wheeler, Ann. Phys. (NY) **2**, 525 (1957).

[16]In local charts $(\delta P)_{\beta_1\cdots\beta_p} = -\nabla_\rho P^\rho{}_{\beta_1\cdots\beta_p}$.

[17]The regular two-form F verifying $\delta T = 0$ are called *pre-Maxwellian forms* [see R. Debever, Bull. Cl. Sci. Acad. R. Belg. **62**, 662 (1976)].

[18]M. Cahen, R. Debever, and L. Defrise, J. Math. Mech. **16**, 761 (1967).

[19]R. Debever and R. G. McLenaghan, J. Math. Phys. **22**, 1711 (1981).

[20]L. Bel, Thèse d'état, Universite de Paris, 1959.

[21]E. Newman and R. Penrose, J. Math. Phys. **3**, 566 (1962).

[22]More generally, the electromagnetic field may be conformally invariant with conformal factor of the form $\exp\{i(\sigma^+ t + m\phi)\}$; see Ref. 5.

Existence and analyticity of many-body scattering amplitudes at low energies

Jan Dereziński
Division of Mathematical Methods in Physics, Warsaw University, Hoża 74, 00-682 Warsaw, Poland

(Received 28 February 1986; accepted for publication 8 October 1986)

Two-cluster–two-cluster scattering amplitudes for N-body quantum systems are studied. Our attention is restricted to energies below the lowest three-cluster threshold. For potentials falling off like $r^{-1-\delta}$ it is proved that in this energy range these amplitudes exist, are continuous, and that the asymptotic completeness holds. Moreover, if the potentials fall off exponentially it is proved that these amplitudes can be meromorphically continued in the energy, with square root or logarithmic branch points at the two-cluster thresholds.

I. INTRODUCTION

Many-body scattering is less complicated at energies below the lowest three-cluster threshold. Only two-cluster scattering is possible in this energy range, which makes it possible to use effectively two-body techniques. This paper is devoted to proving various results on scattering in this energy range by stationary methods. The main tool in our approach is Eq. (1) of Sec. III, which is a kind of a resolvent equation suited to the study of the many-body resolvent below the lowest three-cluster threshold. It is closely related to the Weinberg–Van Winter equation (see Refs. 1–3) and to equations used to prove the asymptotic completeness for three- and four-body systems by stationary methods in Refs. 4–6.

Probably the most interesting result of our paper is the proof that threshold singularities of scattering amplitudes below the lowest three-cluster threshold are of the square root or logarithmic type. Sections III–V are chiefly devoted to this proof. We assume there that the two-body potentials fall off exponentially, which is the same assumption that in the two-body case guarantees the existence of a meromorphic continuation of the resolvent for all energies. In fact, in our method we express the N-body resolvent in terms of essentially two-body objects whose analytic properties we understand better. We prove that many-body scattering amplitudes can also be meromorphically continued across the real axis onto the nonphysical sheet. Our proof is valid only for the energy range below the lowest three-cluster threshold except for the thresholds where square root or logarithmic singularities may occur (the former for odd dimensions and the latter for even dimensions). By a square root (resp. logarithmic) singularity we mean that the function can be locally continued analytically onto the Riemann surface of the square root (resp. of the logarithm). Section V explains how to extend our results to potentials with singularities characteristic of the form boundedness condition.

It should be noted that results similar to ours have been obtained by Balslev.[7] They are restricted to the case $N = 3$ whereas we can handle an arbitrary finite number of particles.

Various authors have studied analytic properties of two-cluster–two-cluster amplitudes before. This was done by Balslev,[7,8] Hagedorn,[9] Hunziker and Sigal,[10] and Sigal.[11] Their methods though, with the exception of Ref. 7, went in a different direction. They assumed both the dilation analyticity and an exponential decay of the potentials and did not obtain the information on threshold singularities that we have. On the other hand, the dilation analyticity assumption allowed them to study the scattering amplitude for the whole energy range. In our next paper we will exploit the dilation analyticity to study threshold singularities above the lowest three-cluster threshold (Ref. 12). There are also some interesting though less complete results on analytic properties of other kinds of many-body scattering amplitudes—see Refs. 7, 8, 10, 11, 13, and 14.

Equation (1) of Sec. III can also be applied to a study of quantum scattering below the lowest three-cluster threshold for potentials that decay like $r^{-1-\delta}$. Those applications require almost no additional work and are given in Sec. VI. They include a construction of a generalized eigenfunction expansion outside a closed set $\mathbb{E}$ of measure zero, a proof of the asymptotic completeness and a proof of the existence and continuity of scattering amplitudes outside $\mathbb{E}$ (all those results are proved below the lowest three-cluster threshold and assume that the potentials decay like $r^{-1-\delta}$).

Most of the results of Sec. VI are not new but our proofs are entirely different from those contained in the literature. A proof of the asymptotic completeness below the lowest three-cluster threshold was found already by Combes[15] and Simon[16,17] for some restricted classes of potentials. A simple time-dependent proof of this fact for potentials that decay like $r^{-1-\delta}$ was given by Enss.[18,19] Recently Sigal and Soffer proved[20] the asymptotic completeness for the whole energy range.

The existence and continuity of two-cluster–two-cluster scattering amplitudes for the whole energy range outside of the thresholds and bound states can be obtained by the so-called commutator methods due to Mourre,[21] Perry, Sigal, and Simon,[22] and Yafaev.[23] Related results are also contained in Refs. 24–27.

The stationary technique in the many-body scattering that we apply has been used and developed by Faddeev,[4] Ginibre and Moulin,[5] Howland,[28] Sigal,[29] and Hagedorn.[6] Using this method one obtains explicit formulas that in principle can be useful for calculating scattering amplitudes, which can be regarded as an advantage of the stationary

 0022-2488/87/051080-09$02.50

method over the time-dependent and commutator methods. Another advantage of our approach is a theorem on a generalized eigenfunction expansion below the lowest three-cluster threshold outside an exceptional set $\mathbb{E}$, a result that to our knowledge has not been obtained by the other methods and is not contained in the literature.

II. NOTATION

We study a many-body Schrödinger operator acting on $L^2(\mathbb{R}^{dN})$ defined by

$$\tilde{H} = -\sum_{i=1}^{N} \frac{\tilde{\Delta}_i}{2m_i} + \sum_{\substack{i,j=1 \\ i<j}}^{N} V_{ij}(x_i - x_j) = -\sum_{i=1}^{N} \frac{\tilde{\Delta}_i}{2m_i} + V,$$

where x_i is a d-dimensional vector pointing at the position of the ith particle with mass m_i and $\tilde{\Delta}_i$ is the Laplacian in x_i.

Throughout the paper we will assume the potentials to be form bounded with respect to the free Hamiltonian with an arbitrarily small bound, which implies that the Hamiltonian is self-adjoint.[30]

Now we have to introduce some concepts that belong to the standard folklore of many-body Schrödinger operators. They are discussed in more detail in numerous papers on the scattering theory, notably in Refs. 6, 29, and 31.

First we remove the center-of-mass motion to obtain the Hamiltonian

$$H = -\sum_{i=1}^{N-1} \frac{\Delta_i}{2\mu_i} + \sum_{\substack{i,j=1 \\ i<j}}^{N} V_{ij}(x_i - x_j) = H_0 + V,$$

where Δ_i is the Laplacian corresponding to the ith coordinate in some system of Jacobi coordinates and μ_i is the corresponding reduced mass for $i = 1,...,N-1$. This Hamiltonian acts on $L^2(X)$, where X denotes the space isomorphic to $\mathbb{R}^{d(N-1)}$ that describes the relative motion of N particles.

We let $R(z) = (z - H)^{-1}$ and $R_0(z) = (z - H_0)^{-1}$.

A cluster decomposition is a partition of the set $\{1,2,...,N\}$ into nonempty disjoint subsets called clusters. A cluster decomposition will be denoted by capital letters such as D and B. A subscript on a cluster decomposition denotes the number of clusters in a given partition. $D_i \subset D_j$ means that D_i refines D_j, i.e., the clusters of D_i are obtained by further partitioning the clusters of D_j. Note that $D_i \subset D_j$ implies $i \geqslant j$ and that for any D_i we have $D_i \subset D_i$. A cluster decomposition with $N-1$ clusters may be also called a pair and denoted by a Greek letter such as σ or by (ij).

For each cluster decomposition we define

$$H_D = -\sum_{i=1}^{N-1} \frac{\Delta_i}{2\mu_i} + \sum_{(i,j)\subset D} V_{ij} = -\sum_{i=1}^{N-1} \frac{\Delta_i}{2\mu_i} + V_D,$$

$$R_D(z) = (z - H_D)^{-1}.$$

The Hamiltonian obtained from H_{D_i} by separating the cluster center of motion variables will be denoted by H^{D_i}. This separation results in a decomposition of the space $L^2(X)$ into $L^2(X^{D_i}) \otimes L^2(X_{D_i})$, where X^{D_i} is isomorphic to $\mathbb{R}^{d(N-i)}$ and X_{D_i} is isomorphic to $\mathbb{R}^{d(i-1)}$. Here the first variables, denoted by x^{D_i}, stand for intracluster degrees of freedom and the latter, denoted by x_{D_i}, for intercluster degrees of freedom. If we represent the original Hilbert space as the above tensor product we can write our cluster Hamiltonian as

$$H_{D_i} = H^{D_i} \otimes 1 + 1 \otimes T_{D_i},$$

where T_{D_i} is the kinetic energy of the c.m. motion of the clusters.

Eigenvalues of H^{D_i} for $1 < i < N$ are called i-cluster thresholds. The point zero is the only N-cluster threshold. We denote the lowest i-cluster threshold for $i > 2$ by ξ. The set of two-cluster thresholds will be denoted by Ω, and ω_1 will mean the lowest two-cluster threshold, which is at the same time the bottom of the continuous spectrum. Elements of $L^2(X^{D_i})$ that are the eigenvectors of H^{D_i} we denote by ϕ_α and call channels. We denote the threshold corresponding to the channel ϕ_α by ω_α and the corresponding cluster decomposition by $D(\alpha)$. If $D(\alpha)$ is a two-cluster decomposition then the corresponding reduced mass of intercluster motion we denote by μ_α and $\nu_\alpha(z)$ will stand for $(2\mu_\alpha(z - \omega_\alpha))^{1/2}$. The generalized eigenvector of $H_{D(\alpha)}$ corresponding to the channel ϕ_α with the intercluster momentum k we denote by $\Phi_\alpha(k)$, explicitly

$$\Phi_\alpha(k)(x) = \phi_\alpha(x^{D(\alpha)}) \exp(ikx_{D(\alpha)}).$$

We define also

$$T_\alpha = \omega_\alpha + T_{D(\alpha)} = \omega_\alpha - \Delta_{D(\alpha)}/2\mu_\alpha.$$

The scattering amplitude for the $\alpha - \beta$ scattering at the energy λ is given by the formula

$$t_{\alpha\beta}(k_1,k_2) = (\Phi_\alpha(k_1),(V - V_{D(\alpha)})\Phi_\beta(k_2)) + \lim_{\epsilon\to 0^+} (\Phi_\alpha(k_1),(V - V_{D(\alpha)})R(\lambda + i\epsilon) \times (V - V_{D(\beta)})\Phi_\beta(k_2)),$$

where

$$T_\alpha \Phi_\alpha(k_1) = \lambda \Phi_\alpha(k_1), \quad T_\beta \Phi_\beta(k_2) = \lambda \Phi_\beta(k_2),$$

and $(\ ,\ )$ denotes the scalar product. (See Ref. 32; for a rigorous derivation of this formula see Ref. 10.)

We denote by $P^{\epsilon}_{D_2}$ the orthogonal projection onto the part of spectrum of H^{D_2} with energies below $\xi - \epsilon$. We need to use the family of projections $P^{\epsilon}_{D_2}$ instead of the projection $P^{0}_{D_2}$ because of a possible occurrence of the Efimov effect. If this effect occurs in a more-than-three-particle system then there may exist infinitely many two-cluster thresholds below ξ (see Refs. 33–35).

Here $|x|$ means some fixed Euclidean norm of the vector x. The symbols ρ^b, $\rho^{D,b}$, and $\rho_D{}^b$ will denote the operators of multiplication by

$$\exp(-b(|x|^2 + 1)^{1/2}), \quad \exp(-b(|x^D|^2 + 1)^{1/2}),$$

and

$$\exp(-b(|x_D|^2 + 1)^{1/2}),$$

respectively.

We also intoduce the Sobolev spaces $\mathbb{H}_m(\mathbb{R}^k)$

$= (1-\Delta_k)^{-m/2} L^2(\mathbb{R}^k)$, where Δ_k is the k-dimensional Laplacian.

III. MAIN RESULTS

In Secs. III–V we will require the potentials to decay exponentially, which is expressed in the following assumption.

Assumption 3.1: We assume that for any i and j, $|V_{ij}|^{1/2}(\rho^{ij,c})^{-1}$ is compact from $\mathbb{H}_1(X^{ij})$ to $L^2(X^{ij})$ for some $c>0$.

The main result of our paper is contained in the following two theorems.

Theorem 3.2: For any $b>0$, $\rho^b R(z)\rho^b$ can be continued meromorphically across the real line below ξ outside of Ω. If $\omega\in\Omega$, $\omega<\xi$, and the dimension is odd, then $\rho^b R(z)\rho^b$ can be continued meromorphically onto a neighborhood of ω on the Riemann surface of $(z-\omega)^{1/2}$. If the dimension is even then the same is true with $\log(z-\omega)$ replacing $(z-\omega)^{1/2}$.

Theorem 3.3: Fix two unit vectors $\hat{e}_1$ and $\hat{e}_2$. Fix two channels ϕ_α and ϕ_β. Then the scattering amplitude $t_{\alpha\beta}[\nu_\alpha(z)\hat{e}_1, \nu_\beta(z)\hat{e}_2]$ can be continued meromorphically in z across the real line below ξ outside of Ω. If $\omega\in\Omega$, $\omega<\xi$, and the dimension is odd, then the scattering amplitude can be continued meromorphically onto a neighborhood of ω on the Riemann surface of $(z-\omega)^{1/2}$. If the dimension is even then the same is true with $\log(z-\omega)$ replacing $(z-\omega)^{1/2}$. This means that at each two-cluster threshold that lies below the lowest at least three-cluster threshold the scattering amplitude has at worst a square root branch point singularity for odd dimensions and a logarithmic singularity for even dimensions.

Most of our paper will be devoted to proving these results. Throughout this section though we will not give full proofs. We will assume in this section that $V_{ij}(\rho^{ij,c})^{-1}\in L^\infty$. The case of singular potentials will be studied in Sec. V. Moreover, we defer some technical lemmas to Sec. IV.

Lemma 3.4: We can find $a>0$ such that $(\rho^{D_2,a})^{-1} P^\epsilon_{D_2} (\rho^{D_2,a})^{-1}$ is bounded.

Proof: By the Hunziker–Van Winter–Zhislin (HVZ) theorem, the part of the spectrum of H^{D_2} associated with the range of $P^\epsilon_{D_2}$ is pure point and lies at least a distance ϵ below the bottom of the continuous spectrum. It is well known (see Ref. 3, Theorems XIII.39 and 40) that the eigenvectors with such energies belong to the domain of the multiplication by $\exp(a|x^{D_2}|)$ for some $a>0$ (a may depend on ϵ). Q.E.D.

From now on we will usually omit a in $\rho^{D_2,a}$ and we will assume that it has a value determined for a given ϵ by the above lemma. We will also usually omit z in $R(z)$ and $R_D(z)$ and ϵ in $P^\epsilon_{D_2}$.

Lemma 3.5: Let ϕ_α be a two-cluster channel, ω_α the corresponding threshold, and $\sigma\not\subset D(\alpha)$. Then we can find $b>0$ such that for any unit vector $\hat{e}$, the function

$$z\to V_\sigma(\rho^b)^{-1}\Phi_\alpha[\nu_\alpha(z)\hat{e}]$$

defined for real z greater than ω_α can be continued analytically onto a neighborhood of the real line outside of ω_α. It can also be continued onto a neighborhood of ω_α on the Riemann surface of $(z-\omega_\alpha)^{1/2}$.

Proof: We easily see that ϕ_α and V_σ with $\sigma\not\subset D_2(\alpha)$ have enough falloff to make up for the growth of $\exp[iv_\alpha(z)\hat{e}x_{D_2(\alpha)}]$ and $(\rho^b)^{-1}$. Q.E.D.

Lemma 3.6: (a) If $i\geqslant 3$, then

$$(V_\sigma)^{1/2} R_{E_i} |V_{\sigma'}|^{1/2}, \quad (V_\sigma)^{1/2} R_{E_i} P_{E_2} |V_{\sigma'}|^{1/2},$$

and

$$(V_\sigma)^{1/2} R_{E_i} P_{E_2} (\rho^{E_2})^{-1} |V_{\sigma'}|^{1/2}$$

are analytic on $\mathbb{C}-[\xi,\infty)$.

(b) $(V_\sigma)^{1/2} R_{E_2}(1-P_{E_2})|V_{\sigma'}|^{1/2}$ is analytic on $\mathbb{C}-[\xi-\epsilon,\infty)$.

Proof: (a) By the HVZ theorem the spectrum of an at least three-cluster Hamiltonian belongs to $[\xi,\infty)$.

(b) The proof is similarly obvious. Q.E.D.

Lemma 3.7: The following expressions can be continued analytically onto a neighborhood of the real line outside of the eigenvalues of H^{F_2} [if ω is an eigenvalue of H^{F_2} then they can be continued onto a neighborhood of ω on the Riemann surface of $(z-\omega)^{1/2}$ for an odd d and of $\log(z-\omega)$ for an even d]:

(a) $(V_\sigma)^{1/2} R_{F_2} P_{F_2} (\rho^{F_2})^{-1} |V_{\sigma'}|^{1/2}$,
where $\sigma, \sigma' \not\subset F_2$,

(b) $\rho^{E_2}(V_\sigma)^{1/2} R_{F_2} P_{F_2} (\rho^{F_2})^{-1} |V_{\sigma'}|^{1/2}$,
where $\sigma'\not\subset F_2$, $E_2\neq F_2$,

and

(c) $(V_\sigma)^{1/2} P_{E_2} R_{F_2} P_{F_2} (\rho^{F_2})^{-1} |V_{\sigma'}|^{1/2}$,
where $E_2\neq F_2$, $\sigma'\not\subset F_2$.

Proof: The lemma follows easily from the proof of Lemma 1 of the Appendix to §XI.6 of Ref. 31, which says the following: $\rho^b(z+\Delta)^{-1}\rho^b$ can be analytically continued across the positive real axis onto the nonphysical sheet as long as $\mathrm{Im}(z^{1/2}) > -b$. See also Ref. 36. Q.E.D.

Now we want to introduce our basic equation for the resolvent. If $k>l$ and $D_k\subset D_l$ we define

$$L_{D_k,D_l} = \sum_{D_k\subset D_{k-1}\subset\cdots\subset D_l} R_{D_k}(V_{D_{k-1}} - V_{D_k}) R_{D_{k-1}} \cdots R_{D_{l+1}} \times (V_{D_l} - V_{D_{l+1}}) R_{D_l}.$$

Statement 3.8: Suppose that z is sufficiently large and negative. Then the full resolvent is equal to the following convergent series:

$$R = \sum_{m=1}^{\infty} \sum_{k=2}^{N-1} \sum_{D^1_{N-1}\subset D^1_2 \not\supset D^2_{N-1}\subset\cdots\subset D^{m-1}_2 \not\supset D^m_{N-1}\subset D^m_k} R_0 \times V_{D^1_{N-1}} L_{D^1_{N-1},D^1_2} V_{D^2_{N-1}} \times L_{D^2_{N-1},D^2_2}\cdots L_{D^{m-1}_{N-1},D^{m-1}_2} V_{D^m_{N-1}} L_{D^m_{N-1},D^m_k} + R_0.$$

Proof: Expand both sides of the above equation by using

$$R = \sum_{n=0}^{\infty} R_0(VR_0)^n$$

and

$$R_D = \sum_{n=0}^{\infty} R_0(V_D R_0)^n.$$

Then compare both sides term by term. Q.E.D.

Statement 3.8 is also an immediate consequence of Eq. IV.4 of Ref. 6 and we refer to Ref. 6 for a more detailed derivation and discussion of similar resolvent identities. Moreover, it is related to the Weinberg–Van Winter equation (see Lemma 4.2 and Chap. XII5 of Ref. 3).

A sequence $D_k \subset D_{k-1} \subset \cdots \subset D_l$ will be called a string.

Now we transform our expression for R in the following way: for $i>1$ each $V_{D^i_{N-1}}$ that appears between $R_{D^{i-1}_2}$ and $R_{D^i_{N-1}}$ we replace with

$$\left[(1-P_{D^{i-1}_2})\left|V_{D^i_{N-1}}\right|^{1/2}\cdot\left(V_{D^i_{N-1}}\right)^{1/2}\right.$$
$$\left.+P_{D^{i-1}_2}(\rho^{D^{i-1}_2})^{-1}\left|V_{D^i_{N-1}}\right|^{1/2}\times\rho^{D^{i-1}_2}\left(V_{D^i_{N-1}}\right)^{1/2}\right]$$

where both "$\cdot$" and "$\times$" denote multiplication. We expand all the square brackets. Each summand of our series we factor by "cutting" it in the following places: (i) at each "$\times$" and (ii) at each "$\cdot$" that "belongs" to $V_{D^i_{N-1}}$ unless there is a "$\times$" at $V_{D^{i+1}_{N-1}}$.

Our aim is to convert the series into some matrix formula. We introduce a square matrix $M(z)$, a row vector $A(z)$, a column vector $B(z)$, and a scalar $C(z)$ with entries from $B(L^2(X))$ and with indices of the form (D_2,E_{N-1}) or (E_{N-1}):

$$M_{(D_2,E_{N-1})(F_2,G_{N-1})}=\sum_{D_2\supset E_{N-1}\subset E_2\supset F_{N-1}\subset F_2\supset G_{N-1}}\rho^{D_2}(V_{E_{N-1}})^{1/2}L_{E_{N-1},E_2}(1-P_{E_2})V_{F_{N-1}}L_{F_{N-1},F_2}P_{F_2}(\rho^{F_2})^{-1}\left|V_{G_{N-1}}\right|^{1/2}$$
$$+\sum_{D_2\supset E_{N-1}\subset F_2\supset G_{N-1}}\rho^{D_2}(V_{E_{N-1}})^{1/2}L_{E_{N-1},F_2}P_{F_2}(\rho^{F_2})^{-1}\left|V_{G_{N-1}}\right|^{1/2},$$

$$M_{(E_{N-1})(F_2,G_{N-1})}=\sum_{E_{N-1}\subset E_2\supset F_{N-1}\subset F_2\supset G_{N-1}}(V_{E_{N-1}})^{1/2}L_{E_{N-1},E_2}(1-P_{E_2})V_{F_{N-1}}L_{F_{N-1},F_2}P_{F_2}(\rho^{F_2})^{-1}\left|V_{G_{N-1}}\right|^{1/2},$$

$$M_{(D_2,E_{N-1})(G_{N-1})}=\sum_{D_2\supset E_{N-1}\subset E_2\supset G_{N-1}}\rho^{D_2}(V_{E_{N-1}})^{1/2}L_{E_{N-1},E_2}(1-P_{E_2})\left|V_{G_{N-1}}\right|^{1/2},$$

$$M_{(E_{N-1})(G_{N-1})}=\sum_{E_{N-1}\subset E_2\supset G_{N-1}}(V_{E_{N-1}})^{1/2}L_{E_{N-1},E_2}(1-P_{E_2})\left|V_{G_{N-1}}\right|^{1/2},$$

$$A_{(F_2,G_{N-1})}=\sum_{E_{N-1}\subset E_2\supset F_{N-1}\subset F_2\supset G_{N-1}}R_0V_{E_{N-1}}L_{E_{N-1},E_2}(1-P_{E_2})V_{F_{N-1}}L_{F_{N-1},F_2}P_{F_2}(\rho^{F_2})^{-1}\left|V_{G_{N-1}}\right|^{1/2}$$
$$+\sum_{F_{N-1}\subset F_2\supset G_{N-1}}R_0V_{F_{N-1}}L_{F_{N-1},F_2}P_{F_2}(\rho^{F_2})^{-1}\left|V_{G_{N-1}}\right|^{1/2},$$

$$A_{(G_{N-1})}=\sum_{E_{N-1}\subset E_2\supset G_{N-1}}R_0V_{E_{N-1}}L_{E_{N-1},E_2}(1-P_{E_2})\left|V_{G_{N-1}}\right|^{1/2},\quad B_{(D_2,E_{N-1})}=\sum_{i=2}^{N-1}\sum_{D_2\supset E_{N-1}\subset E_i}\rho^{D_2}(V_{E_{N-1}})^{1/2}L_{E_{N-1},E_i},$$

$$B_{(E_{N-1})}=\sum_{i=2}^{N-1}\sum_{E_{N-1}\subset E_i}(V_{E_{N-1}})^{1/2}L_{E_{N-1},E_i},\quad C=R_0+\sum_{i=2}^{N-1}\sum_{D_{N-1}\subset D_i}R_0V_{D_{N-1}}L_{D_{N-1},D_i}.$$

Using geometric series to resume the expression for R we obtain

$$R(z)=A(z)(1-M(z))^{-1}B(z)+C(z). \tag{1}$$

The above equation strongly resembles formula IV.10 from Ref. 6, which was used by Hagedorn to prove the asymptotic completeness for the cases of three and four bodies. Nevertheless, our equation is not a generalization of Hagedorn's. Below we list the most important differences.

(1) Hagedorn was interested in the whole spectrum of the Hamiltonian whereas we deal only with its part below the lowest three-cluster threshold. Thus the only projections that appear in our equation are P_{D_2}'s while Hagedorn had to take into account also P_{D_3}'s. This fact explains partly a relative simplicity of our equation in comparison with Hagedorn's.

(2) Every entry of our $M(z)$ is built out of what we call one- and two-string expressions. This implies compactness of our $M(z)$. Some of the entries in Hagedorn's $M(z)$ are not compact; it is the square of his $M(z)$ that is compact.

(3) Our $M(z)$ is indexed by cluster decompositions—Hagedorn's by strings. Because of this difference we need not use the so-called symmetrization that Hagedorn had to use (see Eq. IV.13 of Ref. 6).

(4) The hardest problem about resolvent equations used in the scattering theory is how to control the singularity of the free resolvents that appear in those equations. Here we briefly describe some facts that are commonly used in this context.

(a) The function

$$z\mapsto \exp(-b|x|)R_0(z)\exp(-b|x|)$$

can be continued analytically across the positive real line outside zero (see Lemma 3.7).

(b) The function

$$z\mapsto (1+|x|)^{-1/2-\epsilon}R_0(z)(1+|x|)^{-1/2-\epsilon}$$

is norm continuous up to the positive real line outside zero (see Lemma 6.3).

(c) Suppose that x_1 and x_2 are some coordinates and the dimension is bigger than 2. Then

$$z \mapsto (1+|x_1|)^{-1-\epsilon} R_0(z) (1+|x_2|)^{-1-\epsilon}$$

is uniformly bounded and strongly continuous up to the positive real axis including zero. In some cases we can replace "strongly continuous" by "norm continuous." (See Lemma II3 of Ref. 6 and the proof of Theorem XIII.27 of Ref. 3.)

Our equation is adapted to the use of (a) or (b) whereas Hagedorn's is adapted to the use of (c).

Now we are ready for the proofs of main results of our paper.

Proof of Theorem 3.2: Fix $\epsilon > 0$. (Recall that ϵ enters in our definition of P_{D_2}.) First we prove that, for any $b > 0$, $M(z)$, $\rho^b A(z)$, $B(z)\rho^b$, and $\rho^b C(z)\rho^b$ can be continued analytically across the real line below $\xi - \epsilon$ outside of Ω and if $\omega \in \Omega$ and $\omega < \xi - \epsilon$ then they can be continued analytically onto a neighborhood of ω on the Riemann surface of $(z-\omega)^{1/2}$ or $\log(z-\omega)$ (depending on the dimension). A direct application of Lemma 3.6 proves this fact for the following terms: $M_{(D_2,E_{N-1})(G_{N-1})}$, $M_{(E_{N-1})(G_{N-1})}$, and $\rho^b A(G_{N-1})$. All of the remaining terms contain $R_{D_2} P_{D_2}$ and have a complicated structure involving one or two strings. To prove their analyticity we have to apply also Lemma 3.7. But before this we have to do some algebraic manipulations on such expressions. These manipulations involve a repeated use of the resolvent identity and some combinatorics—proofs are given in the next section (see Consequences 4.3 and 4.7).

Next we need to show that $M(z)$ is compact and goes to zero as $z \to -\infty$. It is enough to show this for the following expressions:

$$(V_{E_{N-1}})^{1/2} L_{E_{N-1},E_2}(z)(1-P_{E_2}) \left| V_{G_{N-1}} \right|^{1/2}$$

and

$$(V_{E_{N-1}})^{1/2} L_{E_{N-1},E_2}(z) P_{E_2} (\rho^{E_2})^{-1} \left| V_{G_{N-1}} \right|^{1/2},$$

where $E_2 \not\supset G_{N-1}$. Suppose that we take a sufficiently large negative z. Then we can expand all the R_{E_k}'s that appear in the above expressions in convergent perturbation expansions. Every term of these expansions can be proved to be compact and go to zero as $z \to -\infty$ by mimicking the proof of Lemma 3A of §5, Chap. XIII, Ref. 3. (All these terms correspond to the so-called connected diagrams in the terminology of Ref. 3.)

Thus $M(z)$ is compact for large negative z. But since it is an analtyic function on a connected domain, its values have to be compact on its whole domain.

By the analytic Fredholm theorem $(1-M(z))^{-1}$ is meromorphic on the domain of analyticity of $M(z)$. Finally we apply Eq. (1), which, since we can take ϵ arbitrarily small, implies the desired analytic properties of $\rho R(z) \rho$. Q.E.D.

Proof of Theorem 3.3: We apply Theorem 3.2, Lemma 3.5, and the definition of the scattering amplitudes. Q.E.D.

IV. ONE- AND TWO-STRING EXPRESSIONS

We begin with an essentially combinatoric lemma.

Lemma 4.1: (a) Fix D_k and D_j. Then

$$\sum_{D_k \subset D_{k-1} \subset D_j} (V_{D_{k-1}} - V_{D_k}) = V_{D_j} - V_{D_k}.$$

(b) Fix D_a, b, and c. Then for some numbers $A(a,b,c)$ we have

$$\sum_{D_a \subset D_b \subset D_c} (V_{D_b} - V_{D_a}) = A(a,b,c)(V - V_{D_a}).$$

Proof: (a) If $\sigma \subset D_k$ or $\sigma \not\subset D_j$ then V_σ does not belong to $V_{D_{k-1}} - V_{D_k}$. Assume it is not the case. Then σ belongs to exactly one D_{k-1} such that $D_k \subset D_{k-1} \subset D_j$.

(b) If $\sigma \subset D_a$ then $V_{D_b} - V_{D_a}$ does not contain V_σ. Let $\sigma \not\subset D_a$. The number of D_b's such that $D_a \subset D_b$ and $\sigma \subset D_b$ is equal to the number of partitions of an $(a-1)$-element set into b nonempty subsets. The number of D_c's such that $D_b \subset D_c$ is equal to the number of partitions of a b-element set into c nonempty subsets. Here $A(a,b,c)$ equals the product of both these numbers. Q.E.D.

The terms in $\rho^b A$, M, $B\rho^b$, and $\rho^b C \rho^b$ that we study fall into two categories: in the first one there are only products involving one long string, in the second one there are sums of products involving two strings. First we study a typical expression involving one string.

Lemma 4.2: Fix D_k and D_{k-m}. Then

$$L_{D_k,D_{k-m}} = \sum_{D_k \subset D_j \subset D_{k-m}} C(k,j,k-m) R_{D_j},$$

where $C(\ )$ are some numerical coefficients.

Proof: We prove our lemmma by induction on m. For $m = 0$ the lemma is obvious. Assume it to be true for some m. Then

$$\begin{aligned} L_{D_k,D_{k-m-1}} &= \sum_{D_k \subset D_{k-1} \subset D_{k-m-1}} R_{D_k} (V_{D_{k-1}} - V_{D_k}) L_{D_{k-1},D_{k-m-1}} \\ &= \sum_{D_k \subset D_{k-1} \subset D_j \subset D_{k-m-1}} C(k-1,j,k-m-1) R_{D_k} \\ &\quad \times (V_{D_{k-1}} - V_{D_k}) R_{D_j} \\ &= \sum_{D_k \not\subset D_j \subset D_{k-m-1}} C(k-1,j,k-m-1) R_{D_k} \\ &\quad \times (V_{D_j} - V_{D_k}) R_{D_j} \\ &= \sum_{D_k \not\subset D_j \subset D_{k-m-1}} C(k-1,j,k-m-1)(R_{D_j} - R_{D_k}). \end{aligned}$$

We used in the following order: the induction step, Lemma 4.1 (a), and the resolvent identity. Q.E.D.

Consequence 4.3: Let $\epsilon > 0$. Then for any $b > 0$ the first term in $M_{(D_2,E_{N-1})(F_2,G_{N-1})}$, the first term in $\rho^b A_{(F_2,G_{N-1})}$, $B\rho^b$, and $\rho^b C \rho^b$ can be continued analytically across the real line below $\xi - \epsilon$ outside of Ω. Moreover, if $\omega \in \Omega$ and $\omega < \xi - \epsilon$ then they can be continued analytically onto a neighborhood of ω on the Riemann surface of $(z-\omega)^{1/2}$ for an odd d and of $\log(z-\omega)$ for an even d.

Now we look closer at two-string expressions. Cluster decompositions that belong to the left-hand side string will be denoted by B_i and those that belong to the right-hand side string will be denoted by D_i. We break up our study into a series of lemmas.

Lemma 4.4: Fix B_m and D_2. Then

$$\sum_{B_m\subset B_3\subset B_2} L_{B_m,B_3}(V_{B_2}-V_{B_3})R_{D_2}$$

$$=\sum_{\sigma\not\subset D_2} Q_\sigma V_\sigma R_{D_2}+Y+ZR_{D_2},$$

where Q_σ and Y are sums of R_{D_j}'s with j greater than 2 and Z is some number.

Proof: Consider the expression on the left-hand side of the equation in the lemma. By Lemma 4.2 it is equal to

$$\sum_{B_m\subset B_j\subset B_3\subset B_2} C(m,j,3)R_{B_j}(V_{B_2}-V_{B_3})R_{D_2}.$$

By Lemma 4.1(b) it can be rewritten as

$$\sum_{B_m\subset B_j} C(m,j,3)A(j,3,2)R_{B_j}(V-V_{B_j})R_{D_2}.$$

Now

$$R_{B_j}(V-V_{B_j})R_{D_2}$$

$$=R_{B_j}(V-V_{D_2})R_{D_2}+R_{D_2}-R_{B_j}. \qquad \text{Q.E.D.}$$

Lemma 4.5: Fix B_{N-1} and D_2. Then

$$\sum_{B_{N-1}\subset B_3\subset B_2\neq D_2} V_{B_{N-1}}^{1/2} L_{B_{N-1},B_3}(V_{B_2}-V_{B_3})R_{D_2}$$

$$=\sum_{\sigma\not\subset D_2} C_\sigma V_\sigma^{1/2} R_{D_2}+C,$$

where C_σ's and C are sums of products of V's, $V^{1/2}$'s, and R_{D_l}'s with l greater than 2.

Proof: Assume first that $B_{N-1}\subset D_2$. The expression can be rewritten as

$$\sum_{m=2}^{N-2} \sum_{\substack{B_{N-1}\subset B_{m+1}\subset B_m\subset B_3\subset B_2\\ B_{m+1}\subset D_2;\ B_m\not\subset D_2}} V_{B_{N-1}}^{1/2} L_{B_{N-1},B_{m+1}}(V_{B_m}-V_{B_{m+1}})L_{B_m,B_3}(V_{B_2}-V_{B_3})R_{D_2}.$$

Now we apply Lemma 4.4 and notice that $V_{B_m}-V_{B_{m+1}}$ consists of V_σ's with $\sigma\not\subset D_2$.

Now let $B_{N-1}\not\subset D_2$. Then we can apply Lemma 4.4 immediately. Q.E.D.

Lemma 4.6: Fix B_{N-1} and D_2. Then

$$\sum_{B_{N-1}\subset B_2\not\supset D_{N-1}\subset D_2} V_{B_{N-1}}^{1/2} L_{B_{N-1},B_2}(1-P_{B_2})V_{D_{N-1}}L_{D_{N-1},D_2}=\sum_{\sigma\not\subset D_2} Q_\sigma V_\sigma^{1/2}R_{D_2}+Y+\sum_{B_2\neq D_2} Z_{B_2}P_{B_2}R_{D_2},$$

where the Q_σ's, Y, and Z_{B_2}'s are sums of products of the V's, $V^{1/2}$ s, $(1-P_{B_2})R_{B_2}$'s, and RD_j's with j greater than 2.

Proof: First we apply Lemma 4.2 to L_{D_{N-1},D_2}. We get a sum that includes only terms R_{D_j} with j greater than 2 and R_{D_2}. The former we will include in Y, what is left has the form

$$\sum_{B_{N-1}\subset B_3\subset B_2\not\supset D_{N-1}\subset D_2} V_{B_{N-1}}^{1/2} C(N-1,2,2)L_{B_{N-1},B_3}(V_{B_2}-V_{B_3})R_{B_2}(1-P_{B_2})V_{D_{N-1}}R_{D_2}$$

$$=\sum_{B_{N-1}\subset B_3\subset B_2\neq D_2} V_{B_{N-1}}^{1/2} C(N-1,2,2)L_{B_{N-1},B_3}(V_{B_2}-V_{B_3})$$

$$\times\left[R_{D_2}-P_{B_2}R_{D_2}-(1-P_{B_2})R_{B_2}+(1-P_{B_2})R_{B_2}\sum_{B_2\supset\sigma\not\subset D_2} V_\sigma R_{D_2}\right].$$

After expanding the square bracket we can include the third term in Y and the fourth one in Q_σ. The second one will constitute Z. Up to a constant we are left with

$$\sum_{B_{N-1}\subset B_3\subset B_2\neq D_2} V_{B_{N-1}}^{1/2} L_{B_{N-1},B_3}(V_{B_2}-V_{B_3})R_{D_2}.$$

The above expression is taken care of by Lemma 4.5. Q.E.D.

Consequence 4.7: Let $\epsilon>0$. Then for any $b>0$ the second term in $M_{(D_2,E_{N-1})(F_2,G_{N-1})}$, $M_{(E_{N-1})(F_2,G_{N-1})}$ and the second term in $\rho^b A_{(F_2,G_{N-1})}$ can be continued analytically across the real line below $\xi-\epsilon$ outside of Ω. Moreover, if $\omega\in\Omega$ and $\omega<\xi-\epsilon$ then they can be continued analytically onto a neighborhood of ω on the Riemann surface of $(z-\omega)^{1/2}$ for an odd d and of $\log(z-\omega)$ for an even d.

V. SINGULAR POTENTIALS

This chapter shows how to modify the proofs of Theorems 3.2 and 3.3 if the potentials are singular and satisfy only Assumption 3.1.

Lemma 5.1: Suppose the potentials are form bounded with respect to the Laplacian with a zero bound. Let ϕ be an eigenvector of H^D with the energy E below the continuous spectrum. Then for some $a>0$ we have $(\rho^{D,a})^{-1}\phi\in\mathbb{H}_1(X^D)$.

Proof: It is well known that for some $a>0$ we have $(\rho^{D,a})^{-1}\phi\in L^2$ (see Ref. 3). We will drop all the reference to D in our computations. We denote $(\rho^a)^{-1}$ by $\exp(F)$ and $\exp(F)\phi$ by ϕ_F. We easily compute the following formula[37]:

$$\Delta\phi_F=[\nabla(\nabla F)+(\nabla F)\nabla]\phi_F-(\nabla F)^2\phi_F+\exp(F)\Delta\phi.$$

We apply it to our Hamiltonian:

$$(\phi_F, H\phi_F) = E(\phi_F,\phi_F) + (\phi_F,(\nabla F)^2\phi_F) < \infty .$$

Since $(\nabla F)^2 = (a\nabla|x|)^2 = a^2$ is bounded we can see that ϕ_F belongs to $\mathbb{H}_1$. Q.E.D.

Lemma 5.2: For any $b>0$, $\rho^b(z+\Delta)^{-1}\rho^b$ can be extended to an analytic function on the part of the Riemann surface of the square root or logarithm (depending on the dimension) defined by $\mathrm{Im}(z^{1/2}) > -b$, with values in bounded operators from $\mathbb{H}_{-1}$ to $\mathbb{H}_1$.

Proof: First we see that

$$\begin{aligned}(-\Delta+1)&\rho^b(z+\Delta)^{-1}\rho^b \\ &= [(-\Delta\rho^b) - 2(\nabla\rho^b)\nabla + \rho^b(-\Delta-1)] \\ &\quad\times(z+\Delta)^{-1}\rho^b \\ &= [(-\Delta\rho^b) + (z+1)\rho^b](z+\Delta)^{-1}\rho^b \\ &\quad + (-2\nabla\rho^b)\nabla(z+\Delta)^{-1}\rho^b - (\rho^b)^2.\end{aligned}$$

By mimicking the proof of Lemma 1 of the Appendix to §XI.6 of Ref. 31 we show that the above expression extends to an analytic family of bounded operators from L^2 to L^2 on the desired complex domain. This means that $\rho^b(z+\Delta)^{-1}\rho^b$ is analytic on the same domain as a function with values in bounded operators from L^2 to $\mathbb{H}_2$ and, by an analogous argument, as a function with values in bounded operators from $\mathbb{H}_{-2}$ to L^2. Now we apply interpolation. Q.E.D.

Equipped with these two lemmas, we can easily modify the proofs from Sec. III to include the singular potentials. For instance, to prove Lemma 3.7(a) we write

$$\begin{aligned}(V_\sigma)^{1/2}&R_{F_2}P_{F_2}(\rho^{F_2,a})^{-1}|V_{\sigma'}|^{1/2} \\ &= [(V_\sigma)^{1/2}P_{F_2}(\rho^b_{F_2})^{-1}][\rho^b_{F_2}P_{F_2}R_{F_2}\rho^b_{F_2}] \\ &\quad\times[(\rho^b_{F_2})^{-1}P_{F_2}(\rho^{F_2,a})^{-1}|V_{\sigma'}|^{1/2}] \\ &= QY(z)Z .\end{aligned}$$

Now for some a, $b>0$ the term Z maps L^2 into $\mathbb{H}_{-1}$, $Y(z)$ maps $\mathbb{H}_{-1}$ into $\mathbb{H}_1$ and has the desired analytic properties, and Q maps $\mathbb{H}_1$ into L^2, which proves Lemma 3.7(a).

VI. ASYMPTOTIC COMPLETENESS AND EXISTENCE OF SCATTERING AMPLITUDES

In this section instead of being interested in an analytic continuation of the resolvent onto the nonphysical sheet we are studying here just its continuous limit up to the real axis. That allows us to weaken assumptions on potentials, which instead of decaying exponentially have to decay only as $r^{-1-\epsilon}$. The main link of this chapter with the preceding ones is Theorem 6.4, which makes an extensive use of methods developed there. It deals with continuing the resolvent up to the real axis. Related properties of the resolvent of the Schrödinger equation were sometimes called a "limiting absorbtion principle" or a "limiting similarity principle" and were studied in the context of the scattering theory or of the absolute continuity of the spectrum (see, for instance, Refs. 6, 28, 29, 38, and 39). By using standard methods of the stationary scattering theory and Theorem 6.4 we are able to obtain various kinds of information about the scattering below the lowest three-cluster threshold, such as the existence of a generalized eigenfunctiion expansion (Theorem 6.6), asymptotic completeness (Theorem 6.7), and the existence and the continuity of the scattering amplitude outside an exceptional set (Theorem 6.8).

We need some additional definitions. If ϕ_α is a channel then we define the imbedding J_α of $L^2(X_{D(\alpha)})$ in $L^2(X)$ by the following formula:

$$J_\alpha(f) = f\otimes\phi_\alpha .$$

Then we define the channel wave operators:

$$W^\pm_\alpha = \underset{t\to\pm\infty}{\text{s-lim}} \exp(-iHt)\exp(iH_{D(\alpha)}t)\,J_\alpha .$$

Here $E_\alpha(B)$ will denote the spectral projection of T_α corresponding to a measurable set $B\subset\mathbb{R}$, and $E_{\mathrm{ac}}(B)$ will denote the spectral projection onto the absolutely continuous part of the spectrum of H belonging to B.

Throughout this section δ will be a fixed number greater than $\frac{1}{2}$. Let γ, γ_D, and γ^D be multiplication operators by $[1+|x|^2]^{-(1/2)\delta}$, $[1+|x_D|^2]^{-(1/2)\delta}$, and $[1+|x^D|^2]^{-(1/2)\delta}$, respectively.

We need some formalism enabling us to restrict Fourier transforms to $(d-1)$-spheres. Let $\mathscr{S}(\mathbb{R}^d)$ denote the space of Schwartz functions on $\mathbb{R}^d$. For $\nu>0$ we define

$$\pi(\nu): \mathscr{S}(\mathbb{R}^d)\to L^2(S^{d-1},d\hat{e})$$

by

$$(\pi(\nu)f)(\hat{e}) = \nu^{(d-1)/2}\hat{f}(\nu\hat{e}) ,$$

where $\hat{e}$ belongs to the unit sphere S^{d-1}, $\hat{f}$ is the Fourier transform of f, and $d\hat{e}$ is the invariant surface measure on $S^{d-1}\subset\mathbb{R}^d$. If $\mathbb{R}^d$ in the above definition is equal to X_D for some two-cluster decomposition D then such a $\pi(\nu)$ will be denoted by $\pi_D(\nu)$.

Here $\mathbb{H}_{m,n}$ denotes the space

$$(1+|x|^2)^{-n/2}(1-\Delta)^{-m/2}L^2.$$

Assumption 6.1: $|V_\sigma|^{1/2}(\gamma^\sigma)^{-1}$ is compact from H_1 to L^2.

Proposition 6.2: Suppose $\delta>\frac{1}{2}$. Then $\pi(\nu)$ extends to a bounded mapping of $H_{0,\delta}(\mathbb{R}^d)$ into $L^2(S^{d-1},d\hat{e})$. Moreover $\nu\to\pi(\nu)$ is norm Hölder continuous.

Proof: See Ref. 5. Q.E.D.

We note also that if $f\in L^2$ then $\pi(\nu)f$ makes sense for almost all ν as an $L^2(S^{d-1},d\hat{e})$-valued measurable function.

Using the above proposition we can easily prove the following lemma (see Ref. 5).

Lemma 6.3: If λ is not equal to 0 and $\epsilon\to 0^+$ then $(\lambda+i\epsilon+\Delta_d)^{-1}$ has a weak limit as an operator from $\mathbb{H}_{-1,\delta}$ into $\mathbb{H}_{1,-\delta}$.

Now we state the main technical theorem of this section.

Theorem 6.4: Suppose that Assumption 6.1 holds. Then there exists a closed set $\mathbb{E}\subset[\omega_1,\xi]$ of measure zero such that

$$R(z) = Z(z) + \sum_{D_2} R_{D_2}(z)P_{D_2}Z_{D_2}(z) ,$$

where $Z(z)$ and $Z_{D_2}(z)$ are analytic functions with values in bounded operators from $\mathbb{H}_{-1,\delta}(X)$ into $\mathbb{H}_{1,0}(X)$ and $\mathbb{H}_{-1,\delta}(X)$, respectively, that can be extended continuously up to $(\omega_1,\xi)\setminus\mathbb{E}$.

Proof: The proof is basically parallel to that of Theorem 3.2. First we change the definitions of M, A, B, and C by replacing ρ and ρ^D by γ and γ^D, respectively. Next we study their properties using Lemma 6.3 in those places where previously we used Lemma 3.7. In this way we are able to prove that the analogs of the expressions which in previous chapters could be continued analytically on an appropriate Riemannian manifold are now norm continuous up to $[\omega_1,\xi]\setminus\Omega$. For instance, $B(z)$ is norm continuous up to $[\omega_1,\xi]\setminus\Omega$ as an operator from $\mathbb{H}_{-1,\delta}(X)$ into $L^2(X)$.

Instead of the analytic Fredholm theorem we have to use its modification from Ref. 31, Chap. XI.6. By virtue of this theorem $(1-M(z))^{-1}$ is norm continuous up to $(\omega_1,\xi)\setminus\mathbb{E}$ as an operator from $L^2(X)$ to $L^2(X)$, where $\mathbb{E}\subset(\omega_1,\xi)$ is some closed set of measure zero.

The last thing to do is to split

$$R=A(1-M)^{-1}B+C$$

into Z and $R_{D_2}P_{D_2}Z_{D_2}$'s. By using Lemma 4.2 we have

$$C=\sum_D c_D R_D\,,$$

where the c_D are some numerical coefficients. We define

$$\tilde{C}=\sum_{D_k,\ k>2} c_{D_k}R_{D_k}+\sum_{D_2} c_{D_2}(1-P_{D_2})R_{D_2}\,.$$

A similar but more complicated analysis using the techniques from Sec. IV leads to the following formula:

$$A=\tilde{A}+\sum_{D_2\supset E_{N-1}}\alpha_{D_2,E_{N-1}}R_{D_2}P_{D_2}(\gamma^{D_2})^{-1}|V_{E_{N-1}}|^{1/2}\,,$$

where $\tilde{A}$ is norm continuous up to $(\omega_1,\xi)\setminus\Omega$ as an operator from $L^2(X)$ to $\mathbb{H}_{1,0}(X)$ and the $\alpha_{D_2,E_{N-1}}$'s are some row vectors with numerical entries. Eventually, we define

$$Z=\tilde{A}(1-M)^{-1}B+\tilde{C}$$

and

$$Z_{D_2}=\sum_{D_2\supset E_{N-1}}\alpha_{D_2,E_{N-1}}P_{D_2}(\gamma^{D_2})^{-1}\left|V_{E_{N-1}}\right|^{1/2}(1-M)^{-1}B+c_{D_2}P_{D_2}\,.$$ Q.E.D.

The above theorem is an important step in proving asymptotic completeness of wave operators; as is well known their existence can be shown much more easily and in greater generality.

Theorem 6.5: Let $\epsilon>0$. Suppose that $V_\sigma=U_\sigma^{(1)}U_\sigma^{(2)}$, where $U_\sigma^{(k)}(1-\Delta)^{-1/2}$ are bounded for $k=1,2$ and $U_\sigma^{(1)}\in\mathbb{H}_{0,1/2+\epsilon}$ for $d=1$, $U_\sigma^{(2)}\in\mathbb{H}_{0,\epsilon}$ for $d=2$ and $U_\sigma^{(2)}\in L^{d-\epsilon}$ for $d>2$. Then all the channel wave operators $W_\alpha^{\pm}$ exist and their ranges are orthogonal.

Proof: The proof is essentially contained in Ref. 31. (See the proofs of Theorems XI.6, XI.16, XI.26, XI.34, and XIII.27 of Ref. 31.) Q.E.D.

Next we will state the main results of this section. We do not give their proofs since very similar ones are contained in the literature and belong to the standard technique of the stationary scattering theory (see Ref. 6, 28, or 29). Our approach is closest to that of Sec. III of Ref. 6. The property of the full resolvent, which was proved in our Theorem 6.4, is a minor modification of the multiparticle limiting absorption principle of Ref. 6. Our Theorems 6.6 and 6.7 are close analogs of Proposition III.6 and Theorem III.1 of Ref. 6, respectively. They can be proved using Theorem 6.4 by mimicking methods of Ref. 6.

Theorem 6.6: Suppose that V satisfies Assumption 6.1, ϕ_α is a two-cluster channel, $\lambda\in(\omega_1,\xi)\setminus\mathbb{E}$, and $\psi\in\mathbb{H}_{0,\delta}$. Then $\widehat{(W_\alpha^{\pm *}\psi)}$ can be restricted to the sphere of radius $\nu_\alpha(\lambda)$ and

$$\pi_{D(\alpha)}(\nu_\alpha(\lambda))W_\alpha^{\pm *}\psi$$
$$=\text{w-}\lim_{\epsilon\to 0^+}\pi_{D(\alpha)}(\nu_\alpha(\lambda))\,J_\alpha^{*}Z_{D(\alpha)}(\lambda+i\epsilon)\psi\,.$$

Remark: Suppose that $f\in L^2(S^{d-1})$. Then (at least formally)

$$HW_\alpha^{\pm}\pi_{D(\alpha)}(\nu_\alpha(\lambda))^{*}f=W_\alpha^{\pm}H_\alpha\pi_{D(\alpha)}(\nu_\alpha(\lambda))^{*}f$$
$$=\lambda W_\alpha^{\pm}\pi_{D(\alpha)}(\nu_\alpha(\lambda))^{*}f\,.$$

Thus the above theorem implies that for $\lambda\in(\omega_1,\xi)\setminus\mathbb{E}$ the following limit exists in $\mathbb{H}_{0,-\delta}$:

$$\text{w-}\lim_{\epsilon\to 0^+}Z_{D(\alpha)}(\lambda+i\epsilon)^{*}J_\alpha\pi_{D(\alpha)}(\nu_\alpha(\lambda))^{*}f$$

and this limit can be interpreted as a generalized eigenfunction of H with an eigenvalue λ.

Theorem 6.7: Suppose that Assumption 6.1 holds. Then

$$E_{\rm ac}(\omega_1,\xi)=\sum_\alpha W_\alpha^{\pm}E_\alpha(\omega_1,\xi)W_\alpha^{\pm *}\,.$$

Remark: We say that the asymptotic completeness holds in the energy range $[a,b]$ if and only if

$$E_{\rm ac}(a,b)=\bigoplus_\alpha \text{Ran}\,W_\alpha^{\pm}E_\alpha(\alpha,b)\,.$$

It is easy to see that the above theorem and the orthogonality of ranges of the $W_\alpha^{\pm}$'s imply the asymptotic completeness below ξ.

The following theorem may be regarded as an analog of Theorem 3.3 in the case when the potentials fall off like $r^{-1-\epsilon}$. It is an easy consequence of Theorem 6.4.

Theorem 6.8: Let ϕ_α and ϕ_β be two-cluster channels. Define the T matrix for the $\alpha-\beta$ scattering by the following formula (see Ref. 10):

$$T_{\alpha\beta}(\lambda)$$
$$=(\phi_\alpha,\pi_{D(\alpha)}(\nu_\alpha(\lambda))(V-V_{D(\alpha)})\pi_{D(\beta)}(\nu_\beta(\lambda))^{*}\phi_\beta)$$
$$+\text{w-}\lim_{\epsilon\to 0^+}(\phi_\alpha,\pi_{D(\alpha)}(\nu_\alpha(\lambda))(V-V_{D(\alpha)})$$
$$\times R(\lambda+i\epsilon)(V-V_{D(\beta)})\pi_{D(\beta)}(\nu_\beta(\lambda))^{*}\phi_\beta)\,.$$

Suppose also that Assumption 6.1 holds. Then for $\lambda\in(\omega_1,\xi)\setminus\mathbb{E}$, $\lambda\mapsto T_{\alpha\beta}(\lambda)$ is a continuous function with values in bounded operators from $L^2(S^{d-1},d\hat{e})$ into itself.

Remark: The above result is formulated in terms of the T matrix $T_{\alpha\beta}(\lambda)$ and not, as in the previous sections, in terms of the scattering amplitudes $t_{\alpha\beta}(k_1,k_2)$. This is because the $r^{-1-\epsilon}$ decay of the potentials is not enough to guarantee the existence of the scattering amplitudes. Below we give an equation that shows the relationship between $T_{\alpha\beta}(\lambda)$ and $t_{\alpha\beta}(k_1,k_2)$.

Suppose that $f_1,f_2\in L^2(S^{d-1})$. Then we have

$$(f_1,T_{\alpha\beta}(\lambda)f_2) = (\nu_\alpha(\lambda)\nu_\beta(\lambda))^{(d-1)/2} \times \int d\hat{e}_1\,d\hat{e}_2\bar{f}(\hat{e}_1)f(\hat{e}_2)t_{\alpha\beta}(\nu_\alpha(\lambda)\hat{e}_1,\nu_\beta(\lambda)\hat{e}_2)\,.$$

ACKNOWLEDGMENTS

It is a pleasure to be able to acknowledge my enormous debt to G. Hagedorn for numerous helpful and friendly discussions. I also appreciate a discussion with I. Herbst to whom I am indebted for the proof of Lemma 5.1 and I am grateful to I. M. Sigal for his valuable remarks. Moreover, I owe my gratitude to M. Klaus for pointing out a mistake in an early version of my paper.

[1]S. Weinberg, "Systematic solution of multiparticle scattering problems," Phys. Rev. **133**, 232 (1964).
[2]C. Van Winter, "Theory of finite systems of particles. I," Mat. Fys. Skr. Danske Vid. Selsk. **1**, 1 (1964).
[3]M. Reed and B. Simon, *Methods of Modern Mathematical Physics* (Academic, New York, 1978), Vol. IV.
[4]L. D. Faddeev, *Mathematical Aspects of the Three Body Problem in the Quantum Scattering Theory* (Israel Program for Scientific Translations, Jerusalem, 1965).
[5]J. Ginibre and M. Moulin, "Hilbert space approach to the quantum mechanical three-body problem," Ann. Inst. H. Poincaré A **21**, 97 (1974).
[6]G. Hagedorn, "Asymptotic completeness for classes of two, three, and four particle Schrödinger operators," Trans. Am. Math. Soc. **258**, 1 (1980).
[7]E. Balslev, "Resonances in three-body scattering theory," Adv. Appl. Math. **5**, 260 (1984).
[8]E. Balslev, "Analytic scattering theory of quantum mechanical three-body systems," Ann. Inst. H. Poincaré A **32**, 125 (1980).
[9]G. Hagedorn, "A link between scattering resonances and dilation analytic resonances in few body quantum mechanics," Commun. Math. Phys. **65**, 181 (1979).
[10]W. Hunziker and I. M. Sigal, "Scattering theory," preprint, 1985.
[11]I. M. Sigal, "Analytic properties of the scattering matrix of many particle systems," preprint, 1985.
[12]J. Dereziński, "Threshold singularities of 2-cluster–2-cluster amplitudes for dilation analytic potentials," to appear.
[13]I. M. Sigal, "Scattering theory in many-body quantum systems. Analyticity of the scattering matrix," in *Quantum Mechanics in Mathematics, Chemistry and Physics*, edited by Gustafson and Reinhardt (Plenum, New York, 1981), p. 307.
[14]I. M. Sigal, "Mathematical theory of single channel systems. Analyticity of scattering matrix," Trans. Am. Math. Soc. **270**, N2, 517 (1982).
[15]J. M. Combes, Nuovo Cimento A **64**, 111 (1969).
[16]B. Simon, "Geometric methods in multiparticle quantum systems," Commun. Math Phys. **55**(3), 259 (1977).
[17]B. Simon, "N-body scattering in the two-cluster region," Commun. Math. Phys. **58**(2), 205 (1978).
[18]V. Enss, "Two-cluster scattering of N-charged particles," Commun. Math. Phys. **65**(2), 151 (1979).
[19]P. Perry, "Scattering theory by the Enss method," Math. Rep. **1**, part 1 (1983).
[20]I. M. Sigal and A. Soffer, "N-Particle scattering problem: Asymptotic completeness for short-range systems," preprint, 1985.
[21]E. Mourre, "Absence of singular continuous spectrum for certain self-adjoint operators," Commun. Math. Phys. **78**, 391 (1981).
[22]P. Perry, I. M. Sigal, and B. Simon, "Spectral analysis of N-body Schrödinger operators," Commun. Math. Phys. **55**, 259 (1981).
[23]D. R. Yafaev, "Remarks on the spectral theory for the Schrödinger operator of multiparticle type," Notes Sci. Seminars LOMI **133**, 277 (1984).
[24]F. Gesztesy and G. Karner, "On three-body scattering near thresholds," preprint, 1985.
[25]G. Hagedorn, "Born series for (2-cluster) → (2-cluster) scattering of two, three, and four particle Schrödinger operators," Commun. Math. Phys. **66**, 77 (1979).
[26]J. Nuttall and S. R. Singh, "Existence of partial-wave two-cluster atomic scattering amplitudes," preprint.
[27]J. Nuttall and S. R. Singh, "Continuation of partial-wave two-cluster atomic scattering amplitudes," preprint.
[28]J. S. Howland, "Abstract stationary theory of multichannel scattering," J. Funct. Anal. **22**, 250 (1976).
[29]I. M. Sigal, *Scattering Theory for Many-Body Quantum Mechanical Systems* (Springer, New York, 1983).
[30]M. Reed and B. Simon, *Methods of Modern Mathematical Physics* (Academic, New York, 1979), Vol. II.
[31]M. Reed and B. Simon, *Methods of Modern Mathematical Physics* (Academic, New York, 1979), Vol. III.
[32]R. Newton, *Scattering Theory of Waves and Particles* (McGraw-Hill, New York, 1982).
[33]V. Efimov, Phys. Lett. B **33**, 563 (1970).
[34]D. R. Yafaev, Math. Sb. **94**, 567 (1974).
[35]Y. N. Ovchinnikov and I. M. Sigal, "Number of bound states of three body systems and Efimov's effect," Ann. Phys. (NY) **123**, 274 (1978).
[36]M. Klaus and B. Simon, "Coupling constant thresholds in nonrelativistic quantum mechanics. I. Short range two body case," Ann. Phys. (NY) **130**, 251 (1980).
[37]R. Froese and I. Herbst, "Exponential bounds and absence of positive eigenvalues for N-body Schrödinger operators," Commun. Math. Phys. **87**, 429 (1982).
[38]S. Agmon, "Spectral properties of Schrödinger operators and scattering theory," Ann. Sc. Norm. Sup. Pisa **2**, 151 (1975).
[39]I. M. Sigal, "Asymptotic completeness of many body short range systems," Lett. Math. Phys. **8**, 181 (1984).

A new class of lattice identities

F. E. Low
Center for Theoretical Physics, Laboratory for Nuclear Science and Department of Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

(Received 16 October 1986; accepted for publication 17 December 1986)

Some identities involving two-dimensional lattice sums of a class of integrals are derived. A simple application to theta functions is given.

I. INTRODUCTION

In this paper we derive a new class of mathematical identities. These identities may be of interest in physical applications involving wave packets with lattice structures in coordinate and wave number space; that is, involving functions like

$$u_{n,m}(x) = f(x-n)\exp\{2\pi imx\}. \tag{1.1}$$

Here, $f(x)$ is a sufficiently smooth and bounded function and n and m are integers.

Functions like $u_{n,m}$ were first introduced into quantum theory by von Neumann[1] in his famous paper on the Ergodic Theorem. In that work the functions were taken to be Gaussians (coherent states). The completeness properties of the coherent states were studied by Perelomov,[2] who derived a special case of the identities discussed here. Functions other than Gaussians were discussed by Bacry *et al.*[3] A very complete list of references is given by Janssen.[4]

The identities we consider involve the inner products of two such functions,

$$h(n,m) = \int_{-\infty}^{+\infty} dx f^*(x)g(x-n)\exp\{2\pi imx\}, \tag{1.2}$$

and state that the sum

$$S(k,q) = \sum_{n,m=-\infty}^{n,m=+\infty} h(n,m)\exp\{i(kn-qn)\} = 0 \tag{1.3}$$

for at least one pair of values $k = k_0, q = q_0$ between zero and 2π. In particular, if f and/or g is an even function of x, $k_0 = q_0 = \pi$; if f and/or g is odd then $k_0 = q_0 = 0$.

We prove the result by noting that

$$S(k,q) = \sum_{n,m} \exp\{i(kn-qm)\} \times \int f^*(x)g(x-n)\exp\{2\pi imx\}dx \tag{1.4}$$

and that the sum over m can be done using the Poisson inversion formula to yield

$$S(k,q) = \sum_n \exp\{ikn\} \int dx f^*(x)g(x-n) \times \sum_p \delta\left(x - \frac{q}{2\pi} + p\right) \tag{1.5}$$

$$= \sum_{n,p} \exp\{ikn\} f^*\left(\frac{q}{2\pi} - p\right) g\left(\frac{q}{2\pi} - p - n\right) \tag{1.6}$$

$$= \left[\sum_p \exp\{ipk\} f\left(\frac{q}{2\pi} - p\right)\right]^* \times \left[\sum_{n'} \exp\{ikn'\} g\left(\frac{q}{2\pi} - n'\right)\right]. \tag{1.7}$$

It has been shown that the expression

$$v(k,q) = \sum_p \exp\{ikp\} f\left(\frac{q}{2\pi} - p\right)$$

has at least one zero for k and q between zero and 2π. Note that for even f, $k = q = \pi$ yields $v = 0$, whereas for odd f the zero occurs for $k = q = 0$. It follows that the sum $S(k,q)$ is zero at $k = q = \pi$ for f and/or g even, and is zero at $k = q = 0$ for f and/or g odd. The resulting identities are obvious for the case of one of the functions odd and one of them even. For both odd or both even, however, the resulting identities are not obvious. For functions that are neither even nor odd, there still must be at least one zero in the range 0 to 2π. This result was first obtained by Balian,[5] and independently by Morgan.[6] The work of Balian is concerned with the dispersion properties in x and k (the variable conjugate to x) of an orthonormal set of functions of the type (1.1). He showed that the product of the uncertainties in x and k for any member of the set diverges at least logarithmically. A similar result was obtained by the author.[7]

II. AN EXAMPLE

We consider the functions

$$f(x) = \exp - \{a^*x^2/2 + t^*x\}$$

and (2.1)

$$g(x) = \exp\{-bx^2/2 + sx\},$$

where $a = a_1 + ia_2$, $b = b_1 + b_2$ with $a_1, b_1 > 0$ and s and t arbitrary complex numbers. Here $h(n,m)$ is easily calculated:

$$h(n,m) = \left(\frac{2\pi}{a+b}\right)^{1/2} \exp\left\{\frac{(na-2\pi im)(nb+2\pi im)}{2(a+b)} + i\pi nm + \frac{(s+t)^2}{2(a+b)} + \frac{t(na-2\pi im)}{(a+b)} - \frac{s(nb+2\pi im)}{(a+b)}\right\}. \tag{2.2}$$

It is convenient to replace s by $s(a+b)^{1/2}$ and t by $t(a+b)^{1/2}$ and to define

$$z_1(n,m) = (an - 2\pi im)(a+b)^{-1/2}$$

 0022-2488/87/051089-02$02.50

and

$$z_2(n,m) = (bn + 2\pi im)(a+b)^{-1/2}.$$

This has the advantage that in the special case considered by Perelomov,[2] where z_1 and z_2 are complex conjugates, the sum over n and m in (1.3) becomes a sum over lattice points in the complex z plane, with a unit cell having the area π. The general expression for h (with normalizing factors left out) is

$$h(n,m)\alpha \exp\{-\tfrac{1}{2} z_1 z_2 + i\pi mn + st + tz_1 - sz_2\}. \tag{2.3}$$

The functions that enter into the identities (1.3) are then obtained by taking the even and odd parts of (2.3) with respect to s and t.

Before indicating how that is done, we note that setting $t = 0$ in (2.3) projects an even function of t, so that both the even and odd parts in s of (2.3) will satisfy the identity (1.3) with $q_0 = k_0 = \pi$. If we further make z_1 and z_2 complex conjugates, we recover the sum rule noted by Perelomov.[2]

Returning to the general case, our identities read

$$\sum (-1)^{m+n} P_+(s) P_+(t) h(m,n) = 0 \tag{2.4}$$

and

$$\sum P_-(s) P_-(t) h(n,m) = 0, \tag{2.5}$$

where the P's project even and odd parts with respect to s and t.

In the special case of a and b real and equal, the z lattice consists of rectangles, and the sum (2.4) and (2.5) can be expressed as relations between theta functions[8]

$$\theta_3(u,q) = \sum_{-\infty}^{+\infty} q^{n^2} \exp\{2inu\} \tag{2.6}$$

and

$$\theta_4(u,q) = \sum_{-\infty}^{+\infty} q^{n^2} (-1)^n \exp\{2inu\}. \tag{2.7}$$

Setting $q_1 = \exp\{-a/4\}$, $u_1 = i/2(a/2)^{1/2}(t-s)$, $q_2 = \exp\{-\pi^2/a\}$, and $u_2 = \tfrac{1}{2}(2/a)^{1/2}\pi(s+t)$ we find

$$\begin{aligned} 0 = \exp ts\{&\theta_4(q_1,u_1)\theta_3(q_2,u_2) + \theta_3(q_1,u_1)\theta_4(q_2,u_2) \\ &\pm \theta_4(q_1,u_1)\theta_4(q_2,u_2) \mp \theta_3(q_1,u_1)\theta_3(q_2,u_2)\} \\ &\pm \exp -ts\{q_1 \to q_1, q_2 \to q_2, u_1 \to (ia/2\pi)u_2, u_2 \\ &\to (2\pi/ia)u_1\}. \end{aligned} \tag{2.8}$$

ACKNOWLEDGMENTS

I would like to thank Professor Kenneth Johnson for several helpful conversations.

This work is supported in part by funds provided by the U. S. Department of Energy (D.O.E) under Contract No. DE-AC02-76ER03069.

[1] J. von Neumann, Z. Phys. **57**, 30 (1929).
[2] A. M. Perelomov, Theor. Math. Phys. **6**, 156 (1971).
[3] H. Bacry, A Grossman, and J. Zak, Phys. Rev. B **12**, 1118 (1975).
[4] J. E. M. Janssen, J. Math. Phys. **23**, 720 (1983).
[5] R. Balian, C. R. Acad. Sci. Paris, **292**, 1375 (1981).
[6] F. Morgan (private communication).
[7] F. E. Low, *A Passion for Physics: Essays in Honor of Geoffrey Chew* (World Scientific, Singapore, 1985), p. 17.
[8] P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill, New York, 1953), p. 430.

Two comments on nonlinear Schrödinger equations

K. P. Hadeler
Lehrstuhl für Biomathematik, Universität Tübingen, Auf der Morgenstelle 10, D-7400 Tübingen, West Germany

H. Lange
Mathematisches Institut, Universität Köln, Weyertal 86-90, D-5000 Köln 41, West Germany

(Received 30 July 1986; accepted for publication 31 December 1986)

In a recent paper, Brüll and Lange [Expos. Math. **4**, 279 (1986); Math. Meth. Appl. Sci. **8**, 559 (1986)] have discussed a class of nonlinear Schrödinger equations with rather general nonlinearities which comprises various cases occurring in the literature. Although the "potentials" in these equations are quite complicated, the equations admit various invariance properties. The present paper has two aims. First several local and global conservation laws related to conservation of mass, impulse, and energy are exhibited. One of these laws seems to be new, though not surprising. Then it is shown that the equation defined by Brüll and Lange is just suitable to apply some transformations which reduce the problem of solitary waves to a relatively simple Hamiltonian system in the plane. This method of transforming the phase plane problem into normal form is, in some respects, similar to the transformations introduced by Hadeler [Proc. Math. Soc. Edinburgh, to be published; *Free Boundary Problems: Theory and Applications*, Montecatini Conference, 1981, edited by A.Fasano and M. Primicerio (Pitman, New York, 1983), Vol. II, pp. 664–671] for parabolic and hyperbolic reaction diffusion equations.

I. CONSERVATION LAWS

The equation studied in Ref. 1 has the form

$$iu_t = -u_{xx} + W\cdot u, \tag{1}$$

where the "potential" W is given by

$$W = f(s) + 2k\cdot h'(s)\cdot[h(s)]_{xx} \tag{2}$$

and

$$s = u\bar{u} = |u|^2. \tag{3}$$

Here f and h are real smooth functions (three times continuously differentiable, say) on $[0,\infty)$, and k is a real constant. The solution u is a scalar complex-valued function.

The "potential" W can be expressed in various ways, e.g.,

$$W = f(s) + 2kh'(s)[h''(s)s_x^2 + h'(s)s_{xx}]. \tag{4}$$

If one introduces

$$v = \sqrt{s}, \tag{5}$$

and

$$\hat{f}(v) = f(s), \tag{6}$$

$$\hat{h}(v) = h(s), \tag{7}$$

then

$$W = \hat{f}(v) + k[\hat{h}'(v)/v][\hat{h}''(v)v_x^2 + \hat{h}'(v)v_{xx}]. \tag{8}$$

For a solution u the following expressions are of some interest:

$$e_0 = |u|^2 = s, \tag{9}$$

$$e_1 = \operatorname{Im} u\bar{u}_x = (1/2i)(u\bar{u}_x - \bar{u}u_x), \tag{10}$$

$$e_2 = u_x\bar{u}_x + g(s) - k[(h(s))_x]^2, \tag{11}$$

where

$$g(s) = \int_0^s f(\tau)d\tau. \tag{12}$$

The quantities e_0, e_1, e_2 are called the localized mass, impulse, and energy, respectively. We shall also consider

$$e_3 = \operatorname{Re} u\bar{u}_x. \tag{13}$$

We show the following conservation laws.

Proposition 1: Define

$$\Delta = u_x\bar{u} - u\bar{u}_x = -2ie_1, \tag{14}$$

$$G_0 = i\Delta, \quad G_3 = (i/2)\Delta_x, \tag{15}$$

$$G_1 = |u_x|^2 - \operatorname{Re} u\bar{u}_{xx} + f(s)s - g(s) + 2k\{h'(s)s[h(s)]_{xx} - \tfrac{1}{2}([h(s)]_x)^2\}, \tag{16}$$

$$G_2 = i\{u_{xx}\bar{u}_x - u_x\bar{u}_{xx} + f(s)\Delta + 2k[(h'(s)h''(s)s_x^2 + h'^2(s)s_{xx})\Delta - h'^2(s)s_x\Delta_x]\}. \tag{17}$$

Then

$$\frac{\partial e_j}{\partial t} = \frac{\partial G_j}{\partial x}, \quad j = 0,1,2,3. \tag{18}$$

The next proposition is an immediate consequence.

Proposition 2: Assume u is a solution of Eq. (1) converging so fast to zero for $|x|\to\infty$, that the integrals

$$E_i(t) = \int_{-\infty}^{+\infty} e_i(t,x)dx, \quad i = 0,1,2,3, \tag{19}$$

exist and that integration can be exchanged with differentiation with respect to t. Then

$$\frac{d}{dt}E_i(t) \equiv 0, \quad i = 0,1,2,3. \tag{20}$$

Also

$$E_3(t) \equiv 0. \tag{21}$$

Proof of Proposition 1:

$$\frac{\partial}{\partial t}e_0 = s_t = i\Delta_x, \tag{22}$$

 0022-2488/87/051091-03$02.50

$$\frac{\partial}{\partial t}e_1=\frac{\partial}{\partial t}(\operatorname{Im} u\bar{u}_x)=\frac{-1}{2i}\frac{\partial}{\partial t}\Delta$$
$$=\tfrac{1}{2}\left[u_{xx}\bar{u}_x+\bar{u}_{xx}u_x-(u\bar{u}_{xxx}+\bar{u}u_{xxx})\right]+W_x s. \tag{23}$$

On the other hand, from (2)

$$W_x=f'(s)s_x+2k\left(h'(s)h''(s)s_x^2+h'^2(s)s_{xx}\right)_x, \tag{24}$$

hence

$$\frac{\partial}{\partial t}e_1=(u_x\bar{u}_x)_x-\tfrac{1}{2}(u\bar{u}_{xx}+\bar{u}u_{xx})_x+f'(s)ss_x$$
$$+2k\left[h''^2(s)ss_x^3+h'(s)h'''(s)ss_x^3\right.$$
$$\left.+4h'(s)h''(s)ss_xs_{xx}+h'^2(s)ss_{xxx}\right]$$
$$=\frac{\partial G_1}{\partial x} \tag{25}$$

by direct calculation. From (14),

$$\Delta_x=u_{xx}\bar{u}-\bar{u}_{xx}u, \tag{26}$$
$$\Delta_{xx}=u_{xxx}\bar{u}-u\bar{u}_{xxx}+u_{xx}\bar{u}_x-u_x\bar{u}_{xx}. \tag{27}$$

From (22),

$$S_{tx}=i\Delta_{xx}. \tag{28}$$

Furthermore,

$$u_{tx}\bar{u}_x+\bar{u}_{tx}u_x=i(u_{xxx}\bar{u}_x-\bar{u}_{xxx}u_x)+iW_x(u_x\bar{u}-u\bar{u}_x)$$
$$=i(u_{xx}\bar{u}_x-u_x\bar{u}_{xx})_x+iW_x\Delta. \tag{29}$$

Hence

$$-i\frac{\partial}{\partial t}e_2=(u_{xx}\bar{u}_x-u_x\bar{u}_{xx})_x+W_x\Delta+f(s)\Delta_x$$
$$-2kh'(s)s_x(h''(s)s_x\Delta_x+h'(s)\Delta_{xx})$$
$$=(u_{xx}\bar{u}_x-u_x\bar{u}_{xx})_x+f(s)\Delta_x+f'(s)s_x\Delta$$
$$+2k\left(h'(s)h''(s)s_x^2+h'^2(s)s_{xx}\right)_x\Delta$$
$$-2kh'(s)s_x(h''(s)s_x\Delta_x+h'(s)\Delta_{xx})$$
$$=(u_{xx}\bar{u}_x-u_x\bar{u}_{xx})_x+(f(s)\Delta)_x$$
$$+2k\left[\left(h'(s)h''(s)s_x^2+h'^2(s)s_{xx}\right)\Delta\right]_x$$
$$-2k\left(h'^2(s)s_x\Delta_x\right)_x. \tag{30}$$

II. SOLITARY WAVES

A solitary wave is a solution

$$u(t,x)=v(x-ct)e^{i\varphi(x-dt)}, \tag{31}$$

where v is a real function and c,d,φ are real constants. Hence the solution consists of the absolute value which moves like a wave with speed c, and a rotational factor (or phase factor) moving with speed d, not necessarily equal to c.

In Ref. 1 it has been shown that for a solution of the form $u(t,x)=v(x-ct)\exp[i\phi(x-dt)]$, where ϕ is a real function, ϕ is necessarily linear, in fact $\phi(y)=cy/2$ (see below).

This solution (31) satisfies

$$u_t=-cv'e^{\div}-i\varphi\,dv\,e^{\div},\quad u_x=v'e^{\div}+vi\varphi e^{\div},$$
$$u_{xx}=v''e^{\div}+2i\varphi v'e^{\div}-v\varphi^2e^{\div}. \tag{32}$$

We introduce these quantities into the differential equations (1) and (8) and compare real and imaginary parts. Then

$$\varphi=c/2 \tag{33}$$

and

$$0=-v''+(c^2/4)v-(c/2)dv+\hat{f}(v)v$$
$$+k(\hat{h}'(v)\hat{h}''(v)v'^2+(\hat{h}'(v))^2v''). \tag{34}$$

Define the constant

$$\kappa=(c/2)d-c^2/4. \tag{35}$$

Then the equation reads

$$[1-k(\hat{h}'(v))^2]v''=-\kappa v+\hat{f}(v)v+k\hat{h}'(v)\hat{h}''(v)v'^2. \tag{36}$$

Let

$$\xi=v,\quad \eta=v'. \tag{37}$$

As long as the leading coefficient does not vanish, Eq. (36) is equivalent with the planar system

$$\xi'=\eta, \tag{38}$$
$$\eta'=\frac{-\kappa\xi+\hat{f}(\xi)\xi+k\hat{h}'(\xi)\hat{h}''(\xi)\eta^2}{1-k(\hat{h}'(\xi))^2}.$$

Introduce the function

$$H(\xi)=1-k(\hat{h}'(\xi))^2. \tag{39}$$

Then

$$H'(\xi)=-2k\hat{h}'(\xi)\hat{h}''(\xi) \tag{40}$$

and the system reads

$$\xi'=\eta, \tag{41}$$
$$\eta'=[-\kappa\xi+\hat{f}(\xi)\xi-\tfrac{1}{2}H'(\xi)\eta^2]/H(\xi).$$

After the transformation

$$\xi=\xi,\quad \zeta=H^{1/2}(\xi)\eta, \tag{42}$$

the system (41) assumes the form

$$\xi'=H^{-1/2}(\xi)\zeta, \tag{43}$$
$$\zeta'=H^{-1/2}(\xi)[-\kappa\xi+\hat{f}(\xi)\xi].$$

Now introduce a new time variable by

$$\tau=\int_0^t H^{-1/2}(\xi(\rho))d\rho. \tag{44}$$

This transformation (as long as it exists) does not change the trajectories of the system

$$\dot{\xi}=\zeta,\quad \dot{\zeta}=-\kappa\xi+\hat{f}(\xi)\xi. \tag{45}$$

Hence the existence problem for solitary waves is to some extent (as long as H stays positive) independent of the function h.

Equations (45) represent a Hamiltonian system with the Hamiltonian function

$$\mathscr{H}(\xi,\zeta)=\tfrac{1}{2}\zeta^2-(\kappa/2)\xi^2+\tfrac{1}{2}g(\xi^2). \tag{46}$$

Hence the system (41) is also a Hamiltonian system with the Hamiltonian function

$$\hat{\mathscr{H}}(\xi,\eta)=\tfrac{1}{2}H(\xi)\eta^2-(\kappa/2)\xi^2+\tfrac{1}{2}g(\xi^2). \tag{47}$$

One can assume that the function H does not vanish, otherwise the differential equation (36) for v is singular. In order to have a definite notion at hand, we call a solitary wave *regular* if $k(\hat{h}'(v))^2<1$. Then we can express the result as follows.

Proposition 3: There is a one-to-one correspondence of

the regular solitary waves of Eq. (1) and the solutions of the Hamiltonian system (45). Solitary waves vanishing at infinity correspond to homoclinic orbits connecting (0,0) to itself.

According to Ref. 1 several types of functions occur, $f(s)=\lambda s^p$ or

$$f(s)=\lambda \log s, \quad h(s)=s^q, \quad h(s)=(1-s)^{1/2}.$$

As an example we give a complete discussion of the case [the "classical case" $f(s)=\lambda s^p$, $h(s)\equiv 0$ has been treated in Ref. 2],

$$f(s)=\lambda s^p, \quad p>0, \quad \lambda\in\mathbb{R}, \quad h(s)=s^q, \quad q>\tfrac{1}{2}. \tag{48}$$

Then the system (45) reads

$$\dot{\xi}=\zeta, \quad \dot{\zeta}=-\kappa\xi+\lambda\xi^{2p+1}. \tag{49}$$

Always (0,0) is a stationary point. There is a second stationary point $(\bar{\xi},0)$ iff $\kappa\lambda>0$. Then

$$\bar{\xi}=(\kappa/\lambda)^{1/2p}.$$

The determinants of the Jacobians at (0,0) and $(\bar{\xi},0)$ are κ and $-2p\kappa$, respectively. If $\lambda>0$ and $\kappa<0$ then (0,0) is a saddle point and all nonconstant solutions are unbounded.

If $\lambda>0$ and $\kappa>0$ then (0,0) is a center, $(\bar{\xi},0)$ is a saddle point. Then there are periodic orbits around (0,0). These correspond to solitary waves in the form of spatially periodic wave trains.

If $\lambda<0$ and $\kappa>0$ then all nonconstant solutions are unbounded. On the other hand, if $\lambda<0$, $\kappa<0$ then the system has a saddle point at (0,0), a center at $(\bar{\xi},0)$, and a homoclinic orbit connecting (0,0) to itself. Along this orbit the Hamiltonian is constant and has the same value as at (0,0), hence $\mathcal{H}\equiv 0$ along this orbit.

After the situation for $h\equiv 0$ has been clarified one has to determine the maximal amplitude $\hat{\xi}$ of the homoclinic orbit. From $\mathcal{H}(\xi,\zeta)=0$, $\zeta=0$ one finds

$$\hat{\xi}=((\kappa/\lambda)(p+1))^{1/2p}.$$

If we choose $h(s)=s^q$, $q>\frac{1}{2}$, then $\hat{h}(v)=v^{2q}$, and $H(v)=1-k\cdot 4q^2\cdot v^{4q-2}$. Hence the homoclinic orbit of the system (45) corresponds to a homoclinic orbit of the original system (38) if either $k\leqslant 0$ or $k>0$, but $k\cdot 4q^2\cdot\hat{\xi}^{4q-2}<1$, i.e.,

$$k\cdot 4q^2((\kappa/\lambda)(p+1))^{(2q-1)/p}<1.$$

For this example we can collect the result as follows. Assume f and h are given by (48), where $p>0$, $q>\frac{1}{2}$, and $\lambda<0$,

$$\sigma=\begin{cases}+\infty & \text{if } k\leqslant 0,\\ (|\lambda|/(p+1))(4q^2k)^{-p/(2q-1)} & \text{if } k>0.\end{cases} \tag{50}$$

Then for each pair (c,d) with

$$(c/2)d-c^2/4\in(-\sigma,0), \tag{51}$$

there is a solitary wave with parameters c and d which vanishes at infinity.

[1] L. Brüll and H. Lange, "Solitary waves for quasilinear Schrödinger equations," Expo. Math. **4**, 279 (1986).
[2] W. A. Strauss, "The nonlinear Schrödinger equation," in *Contempory Developments in Continuum Mechanics and Partial Differential Equations*, edited by G. M. de La Penha and L. A. Medeiros (North-Holland, Amsterdam, 1978).

The Hamiltonian structures of the nonlinear Schrödinger equation in the classical limit

John Verosky
School of Mathematics, University of Minnesota, Minneapolis, Minnesota 55455

(Received 22 July 1986; accepted for publication 5 November 1986)

Using Madelung's hydrodynamical variables, it is shown that the bi-Hamiltonian structure of the nonlinear Schrödinger equation goes over to a bi-Hamiltonian structure of the shallow water wave equations in the classical limit.

I. INTRODUCTION

In 1971 Gardner[1] showed that the Korteweg–de Vries equation can be viewed as a completely integrable Hamiltonian system, and in 1978 Magri[2] published a theory of integrable Hamiltonian systems of partial differential equations. The main feature of Magri's theory is the concept of a bi-Hamiltonian structure, that is, the ability to put a system into Hamiltonian form in two different ways

$$u_t = J_1 E H_1 = J_2 E H_2 ,$$

where the H's are the Hamiltonian functions (functions of u and its x derivatives), the J's are Hamiltonian operators (skew-adjoint partial differential operators giving rise to a Poisson bracket satisfying the Jacobi identity), and E is the Euler operator (variational derivative). A recursion operator[3] $R = J_2 J_1^{-1}$ can be constructed from the Hamiltonian operators and can be used to get an infinite sequence of symmetries and conserved densities of the system. Such sequences are a hallmark of integrable partial differential equations (PDE's), and Magri showed why. Later Kupershmidt and Wilson[4] used the idea of a second Hamiltonian structure to study modified Lax equations and Kupershmidt[5] even found a tri-Hamiltonian system of dispersive water-wave equations containing the usual Korteweg–de Vries equation as a special subsystem. Most recently Nutku[6] has shown that the equations of finite amplitude waves also contain a tri-Hamiltonian structure. The equations of isentropic gas dynamics and the shallow water wave equations are examples of these.

One of Magri's original examples was the nonlinear Schrödinger equation (NLSE). It has two Hamiltonian structures. If the NLSE is written with $\hbar$ (Planck's constant divided by 2π) in the appropriate places, it describes a nonlinear quantum mechanical situation. In 1927, in an effort to give Schrödinger's equation a hydrodynamical interpretation, Madelung[7] used a change of variables to write Schrödinger's equation in fluid form. Purcell[8] used Madelung's transformation to study the higher-order symmetries of the resulting quantum fluid equations for both the linear and nonlinear Schrödinger equations, and has indicated applications to liquid helium. Writing the NLSE in fluid form results in the shallow water wave equations (SWWE) with another term containing $\hbar^2$. Hence in the classical limit $\hbar \to 0$, we have NLSE $\to$ SWWE. The point of this paper is that the two Hamiltonian structures of the NLSE $\to$ two of the Hamiltonian structures of the SWWE. Furthermore, the recursion operator for the NLSE must pass over to one for the SWWE, so there is a correspondence between their symmetries and their conserved densities.

II. PRELIMINARY CALCULATIONS WITH THE NLSE

The NLSE with $\hbar$ is

$$\psi_t = i(\hbar\psi_{xx} + (1/2\hbar)\psi^2\bar{\psi}) .$$

Madelung's change of variables is accomplished in two stages. First let

$$\psi = Re^{i\theta/\hbar} ,$$

where R and θ are functions of x and t. Substituting this into the NLSE results in two equations, the real and imaginary parts, respectively,

$$R_t + 2R_x\theta_x + R\theta_{xx} = 0 ,$$
$$\theta_t + \theta_x^2 + R^2/2 - \hbar^2 R_{xx}/R = 0 .$$

Letting $\rho = R^2$ and $u = 2\theta_x$ results in

$$\rho_t + \rho u_x + u\rho_x = 0 ,$$
$$u_t + uu_x + \rho_x - 2\hbar^2(R_{xx}/R)_x = 0 ,$$

which are the SWWE except for the $\hbar^2$ term, which vanishes when $\hbar \to 0$ in the classical limit.

The theory of Hamiltonian PDE's (see Ref. 9) concerns systems of real equations, so to avoid confusion it is desirable to write the NLSE as a system of two real equations by setting $\psi = v + iw$. Then we get

$$\begin{pmatrix} v \\ w \end{pmatrix}_t = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \left(\hbar \begin{pmatrix} v_{xx} \\ w_{xx} \end{pmatrix} + \frac{1}{2\hbar}(v^2 + w^2) \begin{pmatrix} v \\ w \end{pmatrix} \right) ,$$

which can be put in explicit Hamiltonian form

$$\begin{pmatrix} v \\ w \end{pmatrix}_t = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} E_v \\ E_w \end{pmatrix} \times \left[-\frac{\hbar}{2}(v_x^2 + w_x^2) + \frac{1}{8\hbar}(v^2 + w^2)^2 \right] ,$$

where E_v and E_w are the variational derivatives with respect to v and w, respectively, and the Hamiltonian function is the usual energy density. The second Hamiltonian structure is given by

$$\begin{pmatrix} v \\ w \end{pmatrix}_t = \left(\hbar D_x + \frac{1}{\hbar} \begin{pmatrix} 0 & -w \\ 0 & v \end{pmatrix} D_x^{-1} \begin{pmatrix} 0 & 0 \\ -w & v \end{pmatrix} \right) \begin{pmatrix} E_v \\ E_w \end{pmatrix} \times \left[\frac{1}{2}(wv_x - vw_x) \right] .$$

 0022-2488/87/051094-03$02.50

Here D_x is the total x derivative and D_x^{-1} is its formal inverse. The Hamiltonian function is the usual momentum density. Both of these are obtained from Magri's original example in Ref. 2 simply by changing from ψ to v and w and inserting the $\hbar$'s.

The Madelung change of variable from v and w to ρ and u is expressed by

$$\rho = v^2 + w^2,$$
$$u = 2\hbar(vw_x - wv_x)/(v^2 + w^2)\,.$$

Using the fact that the variational derivative or Euler operator of a total x derivative is zero, it can be shown that the first Hamiltonian function, the energy, becomes

$$-\frac{\hbar}{2}(\rho^{1/2})_x - \frac{1}{8\hbar}\rho u^2 + \frac{1}{8\hbar}\rho^2\,,$$

which is

$$-\frac{1}{4\hbar}\left(\frac{1}{2}\rho u^2 - \frac{1}{2}\rho^2 + 2\hbar^2(\rho^{1/2})_x\right).$$

The factor $-1/4\hbar$ in front will cancel out later. In the classical limit $\hbar \to 0$ only a classical energy

$$\tfrac{1}{2}\rho u^2 - \tfrac{1}{2}\rho^2$$

would remain. The second Hamiltonian function is obviously

$$-(1/4\hbar)\rho u\,,$$

which is a classical momentum with the same factor $-1/4\hbar$ in front.

III. CHANGE OF VARIABLES FOR THE HAMILTONIAN OPERATORS

If a Hamiltonian system

$$U_t = JE_U H$$

is rewritten in terms of new variables V, then we get

$$V_t = V_U J(V_U)^* E_V H\,,$$

where V_U is the differential of V with respect to U defined by

$$\int V_U \eta\, dx = \frac{d}{d\epsilon}\bigg|_{\epsilon=0} \int V(U + \epsilon\eta)dx\,,$$

and the star denotes the formal adjoint. The new Hamiltonian operator is

$$V_U J(V_U)^*\,.$$

The simple proof of this involves the chain rule and the definition of the Euler operator

$$\int E_u[L(V)]\eta\, dx = \frac{d}{d\epsilon}\bigg|_{\epsilon=0} \int L(V + \epsilon\eta)dx\,.$$

In our case we want to pass from vw to ρu variables. Hence J will go to

$$\begin{pmatrix} \rho_v & \rho_w \\ u_v & u_w \end{pmatrix} J \begin{pmatrix} \rho_v & \rho_w \\ u_v & u_w \end{pmatrix}^*.$$

Using chain rules $u_v = u_\theta \theta_v$ and $u_w = u_\theta \theta_w$ it is easy to check that

$$\begin{pmatrix} \rho_v & \rho_w \\ u_v & u_w \end{pmatrix} = 2\begin{pmatrix} v & w \\ -\hbar D_x w/\rho & \hbar D_x v/\rho \end{pmatrix}.$$

Note that an operator $D_x f$ means "multiply by f and then take the x derivative." The first Hamiltonian operator is thus

$$4\begin{pmatrix} 0 & \hbar D_x \\ -\hbar D_x & 0 \end{pmatrix},$$

which is the usual Hamiltonian operator for the SWWE or the gas dynamics equations.[10]

Noting that

$$\begin{pmatrix} v & w \\ -\hbar D_x w/\rho & \hbar D_x v/\rho \end{pmatrix} \frac{1}{\hbar}\begin{pmatrix} 0 & -w \\ 0 & v \end{pmatrix} \times D_x^{-1}\begin{pmatrix} 0 & 0 \\ -w & v \end{pmatrix}\begin{pmatrix} v & \hbar(w/\rho)D_x \\ w & -\hbar(v/\rho)D_x \end{pmatrix}$$

reduces simply to

$$\begin{pmatrix} 0 & 0 \\ 0 & \hbar D_x \end{pmatrix},$$

it is easy to see that the second Hamiltonian operator

$$\begin{pmatrix} v & w \\ -\hbar D_x w/\rho & \hbar D_x v/\rho \end{pmatrix}(\hbar D_x + \cdots)\begin{pmatrix} v & -\hbar D_x w/\rho \\ w & \hbar D_x v/\rho \end{pmatrix}$$

is a third-order operator with no terms of order smaller than first order. In the limit $\hbar \to 0$, only the first-order terms would be kept.

IV. SUMMARY

The first Hamiltonian structure for the fluid version of the NLSE is

$$\begin{pmatrix} \rho \\ u \end{pmatrix}_t = -\begin{pmatrix} 0 & D_x \\ D_x & 0 \end{pmatrix}\begin{pmatrix} E_\rho \\ E_u \end{pmatrix}\left[\frac{1}{2}\rho u^2 - \frac{1}{2}\rho^2 + 2\hbar^2(\rho^{1/2})_x\right]$$

and the second is

$$\begin{pmatrix} \rho \\ u \end{pmatrix}_t = (\hbar^2 C + \hbar B + A)\begin{pmatrix} E_\rho \\ E_u \end{pmatrix}[\rho u]\,,$$

where A, B, and C are first, second, and third order in D_x, respectively. Note how the extra $\hbar$ that D_x carries around canceled with the extra $1/\hbar$ of the Hamiltonians. In the classical limit any term with an $\hbar$ disappears and two Hamiltonian structures for the classical SWWE remain. Since the second Hamiltonian operator for the NLSE is Hamiltonian for any value of the parameter $\hbar$, the operator $\hbar^2 C + \hbar B + A$ (which comes from a change in variables) must also be Hamiltonian for any value of $\hbar$. Thus each of A, B, and C must be Hamiltonian operators, hence there is no worry that the limiting operator A as $\hbar \to 0$ will be Hamiltonian. Obviously a correspondence of higher-order symmetries (and their Noether generators, the conservation laws) is in effect because there is a correspondence of Magri-type recursion operators:

$$\left(\hbar D_x + \frac{1}{\hbar}\begin{pmatrix} 0 & -w \\ 0 & v \end{pmatrix} D_x^{-1}\begin{pmatrix} 0 & 0 \\ -w & v \end{pmatrix}\right)\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix},$$
$$\downarrow$$
$$(\hbar^2 C + \hbar B + A)\begin{pmatrix} 0 & D_x^{-1} \\ D_x^{-1} & 0 \end{pmatrix},$$
$$\downarrow$$
$$A\begin{pmatrix} 0 & D_x^{-1} \\ D_x^{-1} & 0 \end{pmatrix}.$$

Note that the recursion operator for the NLSE raises the differential order upon each application, whereas A being first order in D leads to a recursion operator for the SWWE that will maintain the same differential order but will change the degrees of u and ρ, depending on what explicit functions are in A. See Ref. 6 for a more detailed account of the Hamiltonian structures, recursion operators, and conserved densities of the SWWE.

ACKNOWLEDGMENTS

I am indebted to P. J. Olver and Y. Nutku for several conversations and discussions without which this work could not have been written.

This work was partially supported by NSF Grant No. DMS 86-02004.

[1]C. S. Gardner, J. Math. Phys. **12**, 1548 (1971).
[2]F. Magri, J. Math. Phys. **19**, 1156 (1978).
[3]P. J. Olver, J. Math. Phys. **18**, 1212 (1977).
[4]B. A. Kupershmidt and G. Wilson, Invent. Math. **62**, 403 (1981).
[5]B. A. Kupershmidt, Commun. Math. Phys. **99**, 51 (1985).
[6]Y. Nutku, J. Math. Phys. (to appear).
[7]E. Madelung, Z. Phys. **40**, 322 (1927).
[8]A. J. Purcell, Phys. Rev. D **30**, 2128 (1984).
[9]P. J. Olver, Math. Proc. Cambridge Philos. Soc. **88**, 71 (1980).
[10]J. Verosky, J. Math. Phys. **25**, 884 (1984).

Semiclassical treatment of spin system by means of coherent states

Hernán G. Solari[a)]
Department of Physics and Atmospheric Science, Drexel University, Philadelphia, Pennsylvania 19104

(Received 4 November 1985; accepted for publication 7 January 1987)

The semiclassical time-dependent propagator is studied in terms of the SU(2) coherent states for spin systems. The first- and second-order terms are obtained by means of a detailed calculation. While the first-order term was established in the earlier days of coherent states the second-order one is a subject of contradiction. The present approach is developed through a polygonal expansion of the discontinuous paths that enter the path integral. The results here presented are in agreement with only one of the previous approaches, i.e., the one developed on Glauber's coherent states by means of a direct WKB approximation. It is shown that the present approach gives the exact result in a simple case where it is also possible to observe differences with previous works.

I. INTRODUCTION

The formulation of semiclassical approaches to quantum problems has received a renewed interest following the popularity of the so-called coherent states (CS's),[1,2] which are in some sense the most classical states. Since the pioneering work by Klauder,[3] where the path integral formulation was established, it was apparent that a so-called complexification of the (real) classical variables (essentially position and impulse) was necessary.[3] This complexification obscured the derivation of the semiclassical time propagator (SP) especially when the second-order term, i.e., the reduced propagator (RP), ought to be considered. Attempts to avoid this procedure[4,5] have resulted in a heuristic formulation of the SP in a *P*-form[6] (*P*- and *Q*-forms associated with CS's were introduced from the beginning by Glauber[7]). Later work on Glauber's coherent states has shown[8] that the RP in the *P*-form has a more complicated structure than the one first assumed.

The reduced propagator has also been a subject of controversy in the path integral approach. The imposition of continuity conditions[9–11] to the paths considered has forced the application of the path expansion procedure[12] for the evaluation of the RP, and as a result of this procedure, the second-order term has been formally expressed in terms of the eigenvalues of a Sturm–Liouville problem.[9,10] This procedure has received some criticism[13] because it produces incorrect behavior of the SP at the starting time, among other problems.[13] Other attempts to find the semiclassical propagator present problems in the identification of the correct classical Hamiltonian.[14,15]

In the present work, we study the semiclassical propagator for spin and quasispin systems using the following steps.

(i) We decompose the propagator by means of the Trotter product formulas and slip-in identities between the product terms as done in Refs. 3–6, 9–11, and 16, but, taking advantage of the coherent states' overcompleteness,[1,2,7] the identities are taken in a more general form than in the previous works. This generality is not really a necessary tool, but it makes the following steps clearer.

(ii) We evaluate all the integrals by the Laplace method. This method requires the complexification of the variables, but, by virtue of the generality introduced in (i), this reduces to fixing the free complex parameter (which labels the equivalent identities) to a different value for each time.

(iii) The second-order term is evaluated directly from the Laplace method working out a second-order differential equation for the reduced propagator.

(iv) Finally, the equation for the reduced propagator is integrated.

The steps (i) and (ii) are developed in Sec. II; Sec. III is reserved for a detailed calculation of the reduced propagator [steps (iii) and (iv)] while Sec. IV is devoted to an almost trivial example which already shows the differences between this work and the previous ones. The conclusions and perspectives are presented in Sec. V.

II. SU(2) PATH-INTEGRAL-LIKE FORMULATION

A. The formulation

The path integral formulation can be easily found using the slip-in identities decomposed as addition of coherent states[3–5,9–11] between the terms in the Trotter product formulas

$$U=\exp(-iHt)=\lim_{N\to\infty}(1-iHt/N)^N. \tag{2.1}$$

The standard identity in terms of CS's is written[1,2] as

$$I=\frac{2J+1}{i2\pi}\int_C |z\rangle\langle z|\,\frac{dz\wedge dz}{(1+zz^*)^2}, \tag{2.2}$$

where

$$|z\rangle=\exp(z'J_+-z'^*J_-)|J,-J\rangle, \tag{2.3a}$$

$$z'=e^{-i\phi}\theta/2,\quad z=e^{-i\phi}\tan(\theta/2), \tag{2.3b}$$

(J_+,J_-,J_z) are the three generators of the SU(2) group, and $|J,-J\rangle$ is the extremal state ($J_z|J,-J\rangle=-J|J,-J\rangle$) of the J-irreducible representation of SU(2), while the domain of integration C is the complex plane.

The coherent state (2.3) may also be written taking advantage of the BCH theorems[1] in the form

[a)] Fellow of the Consejo Nacional de Investigaciones Cientificas y Tecnicas, Argentina.

 0022-2488/87/051097-06$02.50

$$|z\rangle = \exp(zJ_+)\exp(\ln(1+zz^*)J_z)\exp(-z^*J_-)|J,-J\rangle, \tag{2.4a}$$

$$|z\rangle = \exp(zJ_+)|J,-J\rangle(1+zz^*)^{-J}, \tag{2.4b}$$

$$|z\rangle = |z)(1+zz^*)^{-J}. \tag{2.4c}$$

This last formula [(2.4c)] defines the unnormalized coherent state (curved brackets) which allows us to rewrite the identity (2.2) in the form

$$I = \frac{2J+1}{2\pi i}\int_C |z)(z|[(1+zz^*)^2(z|z)]^{-1}\,dz\wedge dz^*. \tag{2.5}$$

As the CS's form an overcomplete set of states there are many different ways of writing the identity; for instance, multiplying (2.5) by

$$I = \exp(\delta J_+)\exp(-\delta J_+),$$

we obtain

$$I = II = \exp(\delta J_+)\exp(-\delta J_+)I$$
$$= \exp(\delta J_+)I\exp(-\delta J_+),$$
$$I = \frac{2J+1}{2\pi i}\int_C \exp(\delta J_+)|z)(z|\exp(-\delta J_+)$$
$$\times[(z|z)(1+zz^*)^2]^{-1}\,dz\wedge dz^*,$$

which may be put in the form (applying again the disentangling theorems[1])

$$I = \frac{2J+1}{2\pi i}\int_D |y)(x|[(x|y)(1+yx^*)^2]^{-1}\,dy\wedge dx^*, \tag{2.6a}$$

where x^* and y depend on z and z^* in the following specific form:

$$y = z+\delta, \tag{2.6b}$$

$$x^* = z^*/(1-z^*\delta). \tag{2.6c}$$

The domain of integration C in (2.5) transforms into

$$D = \{(y,x^*) \text{ such that}$$
$$(y-\delta)^* = x^*/(1+x^*y)\} \text{ in (2.6)}.$$

The identity (2.6) is valid for any arbitrary complex number δ, just because it does not depend on it.

Following the standard procedure[4,5,9,10,16] we obtain the following expression for the matrix elements of the propagator between CS's:

$$(\phi|U|\psi) = \lim_{N\to\infty}\int_{D_0}\cdots\int_{D_N}\prod_{n=0}^{N} dy_n\wedge dx_n^*$$
$$\times\{(2J+1)[2\pi i(1+y_nx_n^*)^2]^{-1}\}\exp(F), \tag{2.7}$$

where F has the form

$$F = \sum_{n=1}^{N}\ln\left[\frac{(x_n|y_{n-1})}{(x_n|y_n)}\right] - \frac{it}{N}\mathscr{H}(y_{n-1},x_n^*)$$
$$+\ln[(\phi|y_N)(x_0|\psi)/(x_0|y_0)] \tag{2.8}$$

and where the classical Hamiltonian $\mathscr{H}(y_{n-1},x_n^*)$ reads

$$\mathscr{H}(y_{n-1},x_n^*) = (x_n|H|y_{n-1})/(x_n|y_{n-1}). \tag{2.9}$$

At this point we note that there is no reason for requiring continuity of the paths just because we are dealing with nonorthogonal states. In this respect we recall that in earlier works on the subject[3,9,11] only almost-everywhere continuous paths were considered. The contribution of the discontinuous paths can be determined by the following argument: considering the evaluation of the matrix elements of the identity (2.6),

$$(\phi|\psi) = \frac{2J+1}{2\pi i}\int_D \frac{(\phi|y)(x|\psi)}{(x|y)(1+x^*y)^2}\,dy\wedge dx^*, \tag{2.10}$$

we observe that as long as the integrand is a nonsingular c-number all the allowed values of (y,x^*) contribute to the integral and not only just $y=\psi$, $x^*=\phi^*$ (it is even not necessarily in the domain of integration for an arbitrary δ!). Further, the integrand can never become singular, as is easily seen by inspection of (2.2).

The evolution operator has been decomposed, in our case, in an infinite product of infinitesimal steps (21). Each of the terms in the product is very like the identity but because of the argument concerning the matrix elements of the identity [cf. (2.10)] no notion of continuity of the paths follows from this observation. In fact the opposite is true. On the other hand, if the identities inserted between the terms of (2.1) were expressed in terms of δ-orthogonal states, an intuitive notion of continuity of the paths would follow.

In the following, we shall include discontinuous paths (as suggested in Ref. 13), with the understanding that the state $|y_n\rangle$ in (2.7) is not supposed to be $|y_n\rangle = |y_{n-1} + O(1/N)\rangle$. At this time we will not formulate a formal path integral which would call for the time derivatives of discontinuous paths. A discussion of the subject may be found in Ref. 16. In the semiclassical evaluation of (2.7) we follow a method which closely resembles the polygonal formulation of the path integral.[13] We left the large N limit as the last step to be taken.

B. Classical evaluation of the integrals

The evaluation of integrals, which depends on complex arguments (but real variables) by the Laplace or saddle point methods, requires that the integration path be extended to the complex plane[17]; this procedure was called complexification by Klauder.[3] In the present situation such a deformation of the integration path has already been done in (2.6) and the extremal points are identified by maximizing F [(2.8)] in all the variables, leading to the following set of equations:

$$\frac{\partial F}{\partial x_n^*} = 0 = \frac{\partial\{\ln[(x_n|y_{n-1})/(x_n|y_n)] - it/N\mathscr{H}(y_{n-1},x_n^*)\}}{\partial x_n^*}, \quad n=1,\ldots,N, \tag{2.11a}$$

$$\frac{\partial F}{\partial x_0^*} = 0 = \frac{\partial\ln\{(x_0|\psi)/(x_0|y_0)\}}{\partial x_0^*}, \quad \text{i.e., } \psi = y_0, \tag{2.11b}$$

$$\frac{\partial F}{\partial y_n}=0=\frac{\partial\{\ln[(x_{n+1}|y_n)/(x_n|y_n)]-it/N\mathscr{H}(y_n,x^*_{n+1})\}}{\partial y_n},\quad n=0,\ldots,N-1, \tag{2.11c}$$

$$\frac{\partial F}{\partial y_N}=0=\frac{\partial\ln[(\phi|y_N)/(x_N|y_N)]}{\partial y_N},\quad \text{i.e., } \phi^*=x^*_N. \tag{2.11d}$$

These equations fix both y_n and x^*_n in (2.7) [or z_n and δ_n, see (2.6)].

We recall here that the large parameter in which the asymptotic expansion is carried out is just $2J$ and that

$$(x|y)=(1+yx^*)^{2J}.$$

It is easy to realize that the classical equations (2.11) point out the continuous path as the most important one in the saddle point approximation. They also include the natural boundary conditions.

In what follows we are going to consider that only isolated classical paths contribute to the SP. If there is more than one path, a sum over classical paths will be understood.

Developing the integrands in (2.7) up to second order around the classical path $(\bar y_n,\bar x^*_n)$ we obtain the following expression for the SP:

$$(\phi|U|\psi)=\lim_{N\to\infty}\exp(iS(\psi,\phi^*,t,N))\times\int\prod_{n=0}^{N}\left\{d\xi_n\wedge d\xi^*_n\left[\frac{2J+1}{(2\pi i)(1+y_nx^*_n)^2}\right]\right\}\times\exp[.5(\xi^T\mathbf{S}_2\xi)], \tag{2.12}$$

with $x^*_n=\bar x^*_n+\xi^*_n$, $y_n=\bar y_n+\zeta_n$. We here define the classical (discrete) action, $S(\psi,\phi^*,t,N)$, in terms of the classical path expressed by (2.11), as follows:

$$S(\psi,\phi^*,t,N)=\sum_{n=0}^{N}(-i)\ln\left[\frac{(\bar x_n|\bar y_{n-1})}{(\bar x_n|\bar y_n)}\right]-\mathscr{H}(\bar y_n,\bar x^*_n)\frac{t}{N}-i\ln(\bar x_N|\bar y_N). \tag{2.13}$$

The secondary action appearing throughout the tridiagonal matrix $\mathbf{S}_2$ and the vector ξ are defined, respectively, by

$$\mathbf{S}_2=\begin{vmatrix} A_n & C_n & & & & \\ C_n & B_n & D_{n-1} & & & \\ & D_{n-1} & A_{n-1} & C_{n-1} & & \\ & & & & A_0 & C_0 \\ & & & & C_0 & O(1/2J) \end{vmatrix} \tag{2.14a}$$

and

$$\xi^T=(\zeta_0,\xi^*_0,\zeta_1,\xi^*_1;\ldots;\zeta_N,\xi^*_N). \tag{2.14b}$$

The matrix elements A_n, B_n, C_n, and D_n are the various second derivatives of F evaluated at the extreme points, i.e.,

$$A_n=\left.\frac{\partial^2F}{\partial y_n^2}\right|_{\bar y,\bar x^*}, \tag{2.15a}$$

$$B_n=\left.\frac{\partial^2F}{\partial x_n^{*2}}\right|_{\bar y,\bar x^*}, \tag{2.15b}$$

$$C_n=\left.\frac{\partial^2F}{\partial y_n\,\partial x^*_n}\right|_{\bar y,\bar x^*}, \tag{2.15c}$$

$$D_n=\left.\frac{\partial^2F}{\partial y_{n-1}\,\partial x^*_n}\right|_{\bar y,\bar x^*}, \tag{2.15d}$$

and the $O(1/2J)$ symbol means order of $1/(2J)$ as J goes to ∞.

III. EVALUATION OF THE REDUCED PROPAGATOR

The evaluation of the SP represented by (2.12) simply involves the Gaussian integrations; we obtain

$$(\phi|U|\psi)=\exp(iS(\psi,\phi^*,t))\lim_{N\to\infty}\left\{((-)^N\det(\mathbf{S}_2))^{-1/2}\times\prod_{n=0}^{N}[(1+y_nx^*_n)^{-2}(2J+1)]\right\}, \tag{3.1}$$

where

$$S(\psi,\phi^*,t)=\lim_{N\to\infty}S(\psi,\phi^*,t,N)=\int_0^t\left[i\frac{\partial\ln(x|y)}{\partial y}\dot y-\mathscr{H}(y,x^*)\right]ds-i\ln(\phi|y(t)) \tag{3.2}$$

and

$$\bar x^*_n\to\bar x^*(s),\quad \bar y_n\to\bar y(s)$$

is the (continuous) classical path. [The dot in (3.2) means time derivative.]

We are going to follow a number of steps in the evaluation of the second-order term. First we factorize out of $\mathbf{S}_2$ the C_n elements and a factor $-i$ of each row, defining $\mathbb{M}_N$ as

$$\mathbf{S}_2=\begin{vmatrix} iC_n & & & \\ & iC_n & & \\ & & iC_{n-1} & \\ & & & iC_{n-1} \end{vmatrix}\times\mathbb{M}_N \tag{3.3}$$

and taking into account that

$$\det\mathbf{S}_2=\det\mathbb{M}_N(-)^N\prod_{n=0}^{N}C_n^2 \tag{3.4}$$

the reduced propagator, K, turns out to be

$$K=\lim_{N\to\infty}\left((-)^N\det(\mathbf{S}_2)^{-1/2}\times\prod_{n=0}^{N}[(1+y_nx^*_n)^{-2}(2J+1)]\right),$$

$$K = \lim_{N\to\infty} (\det \mathbb{M}_N)^{-1/2} \times \prod_{n=0}^{N} [(1+y_n x_n^*)^{-2} C_n^{-1}(2J+1)] . \quad (3.5)$$

An explicit evaluation of C_n [(2.15c)] shows that

$$K = \lim_{N\to\infty} (\det \mathbb{M}_N)^{-1/2} \prod_{n=0}^{N} (1+(2J)^{-2}). \quad (3.6)$$

The term $(1+1/2J)^N$ has to be taken as unity as long as it is unity plus the error in the evaluation of the integrals by the Laplace or saddle point method—this procedure lets us write the expression

$$K = \lim_{N\to\infty} \det(\mathbb{M}_N)^{-1/2}. \quad (3.7)$$

The matrix $\mathbb{M}_N$ is a tridiagonal one and has the following explicit form:

$$\mathbb{M}_N = \left|\begin{array}{ccccc} iA_n/C_n & i & & & \\ i & iB_n/C_n & iD_{n-1}/C_n & & \\ & iD_{n-1}/C_{n-1} & iA_{n-1}/C_{n-1} & & \end{array}\right| \begin{array}{l} - \mathbb{M}_n \\ - \mathbb{M}'_n . \end{array} \quad (3.8)$$

The $\det(\mathbb{M}_N)$ can be evaluated by recursion, using the submatrices $\mathbb{M}_n$ and $\mathbb{M}'_n$; the relations are

$$M_n = \det \mathbb{M}_n, \quad M'_n = \det \mathbb{M}'_n, \quad (3.9a)$$

$$M_n = iA_n/C_n M'_n + M_{n-1}, \quad M_0 = 1, \quad (3.9b)$$

$$M'_n = iB_n/C_n M_{n-1} + (D_{n-1})^2/(C_n C_{n-1}) M'_{n-1}, \quad M'_0 = 0. \quad (3.9c)$$

In the limit $N\to\infty$ it is easy to realize from (2.15) and (2.11) that the different coefficients behave in the following way:

$$\frac{A_n}{C_n} = -\frac{\partial \dot{\bar{x}}^*}{\partial \bar{y}}\bigg|_{\bar{x}^*} \frac{t}{N} + O\left(\left(\frac{t}{N}\right)^2\right), \quad (3.10a)$$

$$\frac{B_n}{C_n} = \frac{\partial \dot{\bar{y}}}{\partial \bar{x}^*}\bigg|_{\bar{y}} \frac{t}{N} + O\left(\left(\frac{t}{N}\right)^2\right), \quad (3.10b)$$

$$\frac{D_{n-1}}{C_{n-1}} = 1 - \frac{\partial \dot{\bar{y}}}{\partial \bar{y}}\bigg|_{\bar{x}^*} \frac{t}{N} + O\left(\left(\frac{t}{N}\right)^2\right), \quad (3.10c)$$

$$\frac{D_{n-1}}{C_n} = 1 + \frac{\partial \dot{\bar{x}}^*}{\partial \bar{x}^*}\bigg|_{\bar{y}} \frac{t}{N} + O\left(\left(\frac{t}{N}\right)^2\right). \quad (3.10d)$$

This behavior allows us to transform the recursion formulas (3.9) into a set of coupled first-order differential equations

$$\dot{M} = -i \frac{\partial \dot{\bar{x}}^*}{\partial \bar{y}}\bigg|_{\bar{x}^*} M', \quad (3.11a)$$

$$\dot{M}' = i \frac{\partial \dot{\bar{y}}}{\partial \bar{x}^*}\bigg|_{\bar{y}} M + \left(\frac{\partial \dot{\bar{y}}}{\partial \bar{y}}\bigg|_{\bar{x}^*} - \frac{\partial \dot{\bar{x}}^*}{\partial \bar{x}^*}\bigg|_{\bar{y}}\right) M', \quad (3.11b)$$

with the boundary conditions

$$M(0) = 1, \quad (3.11c)$$

$$M'(0) = 0. \quad (3.11d)$$

The final step in the evaluation of the RP is to integrate Eq. (3.11). It may be checked that the solution we are looking for is the following one:

$$M(t) = \left[\frac{\partial \bar{x}^*(t)}{\partial \bar{x}^*(0)}\bigg|_{\bar{y}(0)} \frac{\partial \bar{y}(t)}{\partial \bar{y}(0)}\bigg|_{\bar{x}^*(t)}\right]^{1/2} \times \exp\left\{\frac{1}{2}\int_0^t \left(\frac{\partial \dot{\bar{y}}}{\partial \bar{y}}\bigg|_{\bar{x}^*} - \frac{\partial \dot{\bar{x}}^*}{\partial \bar{x}^*}\bigg|_{\bar{y}}\right) ds\right\}, \quad (3.12a)$$

$$M'(t) = iM(t) \frac{\partial \bar{y}(t)}{\partial \bar{x}^*(t)}\bigg|_{\bar{y}(0)} . \quad (3.12b)$$

These expressions can in turn be put in terms of the second derivative of the action S [(3.2)], taking into account that

$$i\frac{\partial S(\bar{y}(0),\bar{x}^*(t),t)}{\partial \bar{y}} = (2J)\frac{\bar{x}^*(0)}{1+\bar{y}(0)\bar{x}^*(0)}, \quad (3.13a)$$

$$i\frac{\partial S(\bar{y}(0),\bar{x}^*(t),t)}{\partial \bar{x}^*(t)} = (2J)\frac{\bar{y}(t)}{1+\bar{y}(t)\bar{x}^*(t)}, \quad (3.13b)$$

and

$$\frac{\partial S}{\partial t} = -\mathscr{H}(\bar{y}(t),\bar{x}^*(t)), \quad (3.13c)$$

the determinant M [(3.12)] then equals

$$M(t) = (1+\bar{y}(0)\bar{x}^*(0))^2(1+\bar{y}(t)\bar{x}^*(t))^2 \times \left(i\frac{\partial^2 S}{\partial \bar{x}(t)^* \partial \bar{y}(0)}\right)^{-1} \times \exp\left\{\frac{1}{2}\int_0^t \left(\frac{\partial \dot{\bar{y}}}{\partial \bar{y}} - \frac{\partial \dot{\bar{x}}^*}{\partial \bar{x}^*}\right) ds\right\}. \quad (3.14)$$

The final expression for the matrix elements of the semiclassical propagator (2.12) reads

$$(\phi|U|\psi) = \exp\{iS(\psi,\phi^*,t)\}\left[i\frac{\partial^2 S}{\partial\psi\,\partial\phi^*}\right]^{1/2} \times [(1+\psi\bar{x}^*(0))(1+\bar{y}(t)\phi^*)/(2J)] \times \exp\left\{-\frac{1}{4}\int_0^t \left(\frac{\partial \dot{\bar{y}}}{\partial \bar{y}} - \frac{\partial \dot{\bar{x}}^*}{\partial \bar{x}^*}\right) ds\right\}, \quad (3.15)$$

where $\bar{y}$ and $\bar{x}^*$ are the classical (complex) coordinate and impulse, which start at $\bar{y}(0) = \psi$ and end at $\bar{x}^*(t) = \phi^*$ following the equation of motion (2.11) in the $N\to\infty$ limit

$$i\frac{\partial^2 \ln(\bar{x}|\bar{y})}{\partial \bar{x}^* \partial \bar{y}} \dot{\bar{y}} = \frac{\partial \mathscr{H}(\bar{y},\bar{x}^*)}{\partial \bar{x}^*}, \quad (3.16a)$$

$$-i\frac{\partial^2 \ln(\bar{x}|\bar{y})}{\partial \bar{x}^* \partial \bar{y}} \dot{\bar{x}}^* = \frac{\partial \mathscr{H}(\bar{y},\bar{x}^*)}{\partial \bar{y}}. \quad (3.16b)$$

The semiclassical expression (3.15) can be interpreted as the contribution of several factors: first of all, the classical contribution, i.e., the exponential of the action S; and second, the square root of the term

$$(2J)^{-1}(1+\psi\bar{x}^*(0))(1+\bar{y}(t)\phi^*)\, i\frac{\partial^2 S(\psi,\phi^*,t)}{\partial\psi\,\partial\phi^*} = \frac{1+\bar{y}(t)\phi^*}{1+\psi\bar{x}^*(0)} \frac{\partial \bar{y}(t)}{\partial \phi^*}\bigg|_{\psi} = \frac{1+\psi\bar{x}^*(0)}{1+\bar{y}(t)\phi^*} \frac{\partial \bar{x}(0)}{\partial \psi}\bigg|_{\phi^*}, \quad (3.17)$$

which accounts for the change in the density of the paths due both to the Hamiltonian flow and the curvature of the phase space.

We call the last factor

$$\exp\left\{-\frac{1}{4}\int_0^t \left(\frac{\partial \dot{\bar{y}}}{\partial \bar{y}} - \frac{\partial \dot{\bar{x}}^*}{\partial \bar{x}^*}\right) ds\right\} \quad (3.18)$$

the "extra-phase" term because it provides an extra phase term in the simplest examples, although it is not necessarily of modulo one in general. We do not have at present a physical interpretation for this term, but we observe that it is like a signature of the coherent states, in the sense that the expression (3.15) looks like the expression of the semiclassical propagator in term of space coordinates[13] except for the presence of this term, (3.18), and the ratio of the metric at the initial and final points in (3.17). An equivalent term to (3.18) is also present in the P-propagator of Ref. 8.

IV. EXAMPLE

The simplest example we may look for is the J_z Hamiltonian:

$$H = J_z. \tag{4.1}$$

The classical Hamiltonian (2.9) is now

$$\mathscr{H}(y,x^*) = (x|J_z|y)/(x|y) = -J(1-yx^*)/(1+yx^*) \tag{4.2}$$

and the classical motion (3.16) is

$$i\dot{y} = y, \quad y(0) = \psi, \tag{4.3a}$$

$$i\dot{x}^* = -x^*, \quad x^*(t) = \phi^*. \tag{4.3b}$$

These equations have straightforward integrals that allow the evaluation of the action in the form given by (3.2):

$$S(\psi,\phi^*,t) = Jt - i2J\ln(1+\psi\phi^* e^{-it}). \tag{4.4}$$

Taking the second derivative of S (4.4) we obtain

$$\frac{\partial^2 S}{\partial\psi\,\partial\phi^*} = -\frac{i2Je^{-it}}{(1+\psi\phi^* e^{-it})^2}. \tag{4.5}$$

The evaluation of the SP (3.15) from (4.3)–(4.5) gives

$$(\phi|U|\psi) = e^{iJt}(1+\psi\phi^* e^{-it})^{2J}. \tag{4.6}$$

This last formula [(4.6)] is in fact exact for the matrix elements of the SP. The result (4.6) in this example is more accurate than the previous results where the same matrix elements were calculated in the form [Eqs. (3.7) and (3.49), Ref. 9]

$$(\phi|U|\psi) = \exp(iS(\psi,\phi^*,t))(2i\pi\sin(t))^{-1/2}. \tag{4.7}$$

Before attempting a comparison between the present results and the corresponding ones of Refs. 8, 14, and 15, we have to go from SU(2) coherent states to $Np4$ coherent states (Gaussian wave packets in the space or momentum basis). The procedure consists of contracting the SU(2) algebra into the $Np4$ algebra and simultaneously mapping the coherent states. The details of the method are explained in detail in Refs. 1 and 18. We recall here the main results.

Under the limit $J\to\infty$ the operators and coherent states associated with SU(2) go into operators and coherent states of $Np4$ in the form

$$\mathrm{SU}(2)\to Np4,$$

$$J_z + J \to \hat{N} = a^+ a \quad \text{(number operator)},$$

$$J_+/(2J)^{1/2} \to a^+ \quad \text{(creation operator)},$$

$$J_-/(2J)^{1/2} \to a \quad \text{(destruction operator)},$$

$$|J,-J\rangle \to |0\rangle \quad (\hat{N}|0\rangle = 0|0\rangle),$$

$$(2J)^{1/2} y \to y \quad \text{(coherent states map)}. \tag{4.8}$$

The contraction can be seen in essence as the linear expansion of the phase space $\{y,x^*\}$ around the point $\{0,0\}$.

With these identifications, we obtain, from (4.6), the matrix elements of $F = \exp(-it\hat{N})$ expressed as

$$(\phi|F|\psi) = \lim_{J\to\infty}(\phi|U|\psi)e^{-iJt} = \exp(\psi\phi^* e^{-it}), \tag{4.9}$$

where the limit will be understood as the contraction procedure previously outlined.

This latest expression (4.9) is exact and is the one obtained in Ref. 8, while it appears in Refs. 14 and 15 multiplied by $e^{it/2}$. This factor is irrelevant in the present trivial example but it accounts for a missing term in the general semiclassical expression of Ref. 14. (In Ref. 15 the factor was compensated by an *ad hoc* identification of the classical Hamiltonian.)

The contraction procedure applied here is not limited to the Hamiltonian of the example and is valid in general.

V. CONCLUSIONS

We have developed the semiclassical propagator in terms of SU(2) coherent states in an almost closed form. The resulting formula is well behaved for short times and in addition it matches the exact result for Hamiltonians which are linear combinations of the SU(2) generators. It also agrees with the results obtained in Ref. 8 using Glauber's coherent states in a direct WKB approximation to the SP in the P-form.

Looking for possible generalizations, we recall here that the present approach is fully based on the existence of an algebraic classical limit,[19] expressed by the large $(2J)$ approximation. *While it does not appear that it could be difficult to generalize these results to other systems from a technical point of view, it is worth keeping in mind that the existence of an algebraic limit is a requisite from both physical and mathematical points of view.* (It may express the existence of a large number of particles or quasiparticles or to have other meaning depending upon the problem.)

In order to make sense the evaluation of the integrals by the Laplace or saddle point methods it is required that the overlap between two unnormalized coherent states behaves as C^λ, where C is a complex number and λ is the order parameter that is expected to be linked with physical situations. The nonexistence of a parameter in which the asymptotic expansion is carried out makes the application of the Laplace method uncertain and does not make room for necessary operations like the one performed while going from (3.6) to (3.7). As a major mention of the importance of this fact we recall that the standard time-dependent Hartree–Fock equations, which have been formally derived in the classical limit,[4,11] do not have an identified large parameter associated.[20] This fact raises important questions about the justification of these derivations.

As a physical situation that may be treated by the present approach we may mention the Coulomb excitation of a

nucleus as the result of scattering if the nucleus is described by an IBM model.[20] Another point of possible physical interest is the requantification of the solutions applying Gutzwiller's method[22] adapted to CS's. In this context, the lowest lying state in this approximation is the one predicted by the random phase approximation as it may be easily realized shifting the real time to an imaginary one ($i\beta$) and looking for the $\beta \to \infty$ limit (i.e., the zero temperature limit).

Our last point about the present approach is that it does respect dynamical symmetries if they can be expressed by the exponential of a linear combination of the SU(2) generators.[8] This point brings up several questions as to the correct way of taking mean values of operators in the semiclassical approximation because SU(2)–TDHF expressions are symmetry breaking (see, for example, Refs. 23 and 24). Further work on this subject is in progress.

ACKNOWLEDGMENTS

I fully acknowledge support of the Consejo Nacional de Investigaciones Cientificas y Tecnicas of Argentina and hospitality of Drexel University. I am kindly indebted to Dr. Michel Vallieres for a careful reading of previous versions of this work.

This work has been partially supported by the National Science Foundation (NSF) under Grant No. PHY-84-41891.

[1]R. Gilmore, *Lie Groups and Lie Algebras and Some of Their Applications* (Wiley, New York, 1974).
[2]A. Perelomov, Commun. Math. Phys. **26**, 222 (1972).
[3]J. Klauder, Phys. Rev. D **19**, 2349 (1979).
[4]H. Kuratsuji and T. Suzuki, Phys. Lett. B **92**, 19 (1980).
[5]H. Kuratsuji, Phys. Lett. B **103**, 79 (1981).
[6]T. Suzuki, Nucl. Phys. A **398**, 557 (1983).
[7]R. J. Glauber, Phys. Rev. **131**, 2766 (1963).
[8]H. G. Solari, J. Math. Phys. **27**, 1351 (1986).
[9]H. Kuratsuji and Y. Mizobuchi, J. Math. Phys. **22**, 757 (1981).
[10]H. Kuratsuji and Y. Mizobuchi, Phys. Lett. A **82**, 2798 (1981).
[11]J. P. Blaizot and H. Orland, Phys. Rev. C **24**, 1740 (1981).
[12]S. Levit, Ann. Phys. (NY) **103**, 198 (1977).
[13]L. S. Schulman, *Techniques and Applications of Path Integration* (Wiley, New York, 1981).
[14]Y. Weissman, J. Chem. Phys. **76**, 4067 (1982).
[15]Y. Wiessman, J. Phys. A: Math. Gen. **16**, 2693 (1983).
[16]J. R.Klauder and B.-S. Skagerstan, *Coherent States* (World Scientific, Singapore, 1985).
[17]E. T. Copson, *Asymptotic Expansions* (Cambridge U.P., Cambridge, 1965).
[18]F. T. Arecchi, E. Courtens, R. Gilmore, and H. Thomas, Phys. Rev. A **6**, 2211 (1972).
[19]R. Gilmore, J. Math Phys. **20**, 891 (1979).
[20]S. Levit, Phys. Rev. C **21**, 1594 (1980).
[21]R. Gilmore and D. H. Feng, *Interacting Bose–Fermi Systems in Nuclei*, edited by F. Iachello (Plenum, New York, 1981).
[22]M. C. Gutzwiller, J. Math. Phys. **12**, 343 (1971).
[23]H. G. Solari and E. S. Hernandez, Phys. Rev. C **26**, 2310 (1982).
[24]H. G. Solari and E. S. Hernandez, Phys. Rev. C **28**, 2472 (1983).

Generalization of Levinson's theorem to particle–matter interactions

S. G. Chung
Department of Physics, Western Michigan University, Kalamazoo, Michigan 49008

Thomas F. George
Departments of Chemistry and of Physics and Astronomy, 239 Fronczak Hall, State University of New York at Buffalo, Buffalo, New York 14260

(Received 17 June 1986; accepted for publication 21 January 1987)

It is shown that Levinson's theorem in static potential scattering can be generalized to a particle dynamically interacting with one-dimensional matter systems (liquids or solids). A restriction on a particle–matter interaction is that it decays faster than an inverse quadratic of the particle–matter separation.

I. INTRODUCTION

Levinson's theorem is one of the classic theorems in scattering theory. For s-wave motion of a particle in a spherically symmetric potential $V(r)$ in three dimensions, Levinson showed that the scattering phase shift $\delta(k)$ as a function of incident wave number k is related to the number of s-wave bound states N as

$$N=\delta(+0)/\pi \tag{1}$$

under certain conditions on the potential $V(r)$ (Refs. 1 and 2). Jauch[3] and then Kazes[4] and Ida[5] later developed the method of scattering operator algebra, and succeeded to generalize the theorem to cases of nonlocal potentials. In this paper, we shall point out that the theorem can be generalized to the case of dynamical particle–matter interactions in one dimension (1-D).

A desire for this generalization arose in the course of our recent study of low-temperature adsorption of atoms on a material surface.[6] Consider a scattering eigenstate characterized by two wave numbers k_x and k_z of the incident particle as shown in Fig. 1(a) (the particle motion is in the xz plane). The scattering wave function takes an asymptotic form at $z\to\infty$ of

$$|k^+\rangle \sim \phi_0 e^{ik_x x}(e^{-ik_z z}-S(k_x,k_z)e^{ik_z z}), \tag{2}$$

where ϕ_0 represents the matter ground state ($T=0$ K for simplicity, and we assume that the ground state is nondegenerate), and the S-matrix element $S(k_x,k_z)$ is in general a function of both k_x and k_z. Sometimes $S(k_x,k_z)$ has a weak k_x dependence, whereby the problem becomes essentially one dimensional. One such example is found in recent experiments for ^{4}He atom scattering from a liquid ^{4}He surface, reporting a weak k_x dependence for the reflectance coefficient as a function of k_x and k_z (Ref. 7). Indeed, previously people mainly considered a simplified 1-D model of particle–matter interactions to study low-temperature adsorption [cf. Fig. 1(b)]. We note that a 1-D model must be of finite size, because otherwise the matter does not have a well-defined boundary at finite temperatures, and the question of calculating, for example, the adsorption probability of a particle becomes meaningless.[8]

A long-standing controversy in low-temperature adsorption based on a finite 1-D model concerns the importance of correlated motions of a particle near a material surface.[9,10] This is essentially a question on the importance of many-body effects. We thus encounter an interesting question: is it possible to dynamically generalize Levinson's theorem? In this paper, we shall show that there indeed exists a dynamical version of Levinson's theorem. The only restriction in our arguments is that the potential created by a matter system and seen by a particle must decay faster than an inverse quadratic of the particle–matter separation. We also assume that the ground state of the matter system is nondegenerate, which in fact is very likely the case for a finite system without a special symmetry.

We have organized the present paper as follows. In the next section, as a natural generalization of the static case,[1,11] we describe a scattering eigenstate of a finite 1-D model, particularly a Jost function and its general aspects. In Sec. III, we discuss analytic properties of the Jost solution and Jost function. To do this, again as a natural generalization of the static case,[11,12] we consider an integral Schrödinger equation for the Jost solution, and its formal solution in terms of the Fredholm series. A dynamical generalization of Levinson's theorem is then straightforward (Sec. IV). Finally in Sec. V, our conclusion is given.

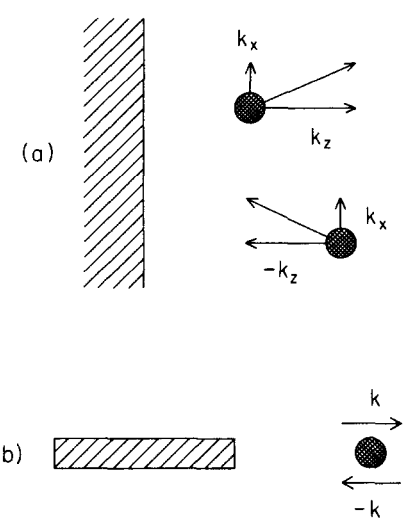


FIG. 1. (a) Three-dimensional geometry for the scattering eigenstate characterized by the parallel and perpendicular wave numbers, k_x and k_z. (b) Its one-dimensional simplification when the parallel and perpendicular motions are approximately separable.

 0022-2488/87/051103-04$02.50

II. SCATTERING EIGENSTATES

The Hamiltonian for a particle interacting with a matter system is written in general as

$$H_{\rm tot} = H(\tilde{X},\tilde{P}) + V(\tilde{X},x) + K(p), \tag{3}$$

where $(\tilde{X},\tilde{P})$ are vector operators describing the positions and momenta of the matter atoms, and (x,p) describe the position and momentum of the particle. Here $H(\tilde{X},\tilde{P})$ is a 1-D matter Hamiltonian, $K(p)$ is the kinetic energy of the particle, and $V(\tilde{X},x)$ describes the interaction between the particle and the 1-D matter system. Let us use the notations $\tilde{r} \equiv (\tilde{X},x)$, m = mass of particle, and $\phi_0(\tilde{X})$ and E_0, respectively, for the ground state of $H(\tilde{X},\tilde{P})$ ($T = 0$ K) and its energy. For a given total energy $E(k) = E_0 + \hbar^2 k^2/2m$, the Schrödinger equation

$$H_{\rm tot}\psi(\tilde{r},k) = E(k)\psi(\tilde{r},k) \tag{4}$$

has two independent solutions $F(\tilde{r}, \pm k)$ with the asymptotic properties at $x \to \infty$,

$$F(\tilde{r}, \pm k) \to \phi_0(\tilde{X})e^{\mp ikx}. \tag{5}$$

The scattering state $\psi(\tilde{r},k)$ is then given as a linear combination of the Jost solutions $F(\tilde{r}, \pm k)$. Noting that (4) is real and $\psi(\tilde{r},k)$ is an even function of k, we can write, in general,

$$\psi(\tilde{r},k) = (i/2k) \times [f(-k)F(\tilde{r},k) - f(k)F(\tilde{r},-k)]. \tag{6}$$

To determine the Jost function $f(k)$ in the static case, one imposes the condition

$$\psi(x = x_0,k) = 0, \tag{7}$$

which is a requirement that the particle cannot reach the point $x = x_0$, where the potential energy is large. The corresponding physical condition in our dynamic case is that

$$\psi(\tilde{r} = \tilde{r}_c,k) = 0, \tag{8}$$

where $\tilde{r}_c$ is a constant vector independent of k. With a suitable choice of normalization, one can then take the Jost function $f(k)$ as

$$f(k) = F(\tilde{r}_c,k). \tag{9}$$

A remark here is that the vectors $\tilde{r}_c$ which satisfy condition (8) generally form a hypersurface. A consistent situation, therefore, is that by choosing the Jost function as (9) for a special point $\tilde{r} = \tilde{r}_c$ on the hypersurface, condition (8) must be automatically satisfied for all the other points on the hypersurface. In other words, the Jost solutions $F(\tilde{r}, \pm k)$ must be strongly correlated.

In the next section, we shall examine an analytic property of the Jost solution $F(\tilde{r},k)$ in the complex k plane, which leads to the same analytic property of the Jost function $f(k)$ due to (9). Before doing so, let us mention some general properties of $f(k)$. First, it is seen from (5) and (6) that the zeros of the Jost function $f(k)$ on the negative imaginary axis in the complex k plane describe the bound states of $H_{\rm tot}$. In this paper, we restrict ourselves to those particle–matter potentials which decay faster than an inverse quadratic of the particle–matter separation. For such potentials, one can readily see that the number of bound states is finite, and therefore, except for physically uninteresting accidental situations,

$$f(0) \neq 0. \tag{10}$$

Second, the reality of $H_{\rm tot}$ means that

$$H_{\rm tot}F^*(\tilde{r}, -k^*) = E(k)F^*(\tilde{r}, -k^*), \tag{11}$$

but since $F^*(\tilde{r}, -k^*) \sim \phi_0(\tilde{X})e^{-ikx}$ as $x \to \infty$, we have

$$F^*(\tilde{r}, -k^*) = F(\tilde{r},k). \tag{12}$$

Now since $\psi(\tilde{r},k)$ as given by (6) is a real, even function of k,

$$\psi^*(\tilde{r},k^*) = \psi(\tilde{r},k). \tag{13}$$

From (6), (12), and (13), we obtain the well-known relationship

$$f^*(-k^*) = f(k). \tag{14}$$

For real k, in particular, upon writing the Jost function as

$$f(k) = |f(k)|e^{i\delta(k)}, \tag{15}$$

where $\delta(k)$ is a scattering phase shift, (10) and (14) give

$$-\delta(-k) = \delta(k), \tag{16}$$

under the convention that $\delta(\pm\infty) = 0$. A note on (16) is that $\delta(\pm\infty)$ need not be the same, so that they are not necessarily zero.

III. ANALYTICITY OF THE JOST FUNCTION

We now discuss an analytic property of the Jost solution in the complex k plane, leading to the same analytic property of the Jost function due to (9). Let us consider the following integral Schrödinger equation for $F(\tilde{r},k)$:

$$F(\tilde{r},k) = F_0(\tilde{r},k) + \int d\tilde{r}'\, K(\tilde{r},\tilde{r}';k)F(\tilde{r}',k), \tag{17}$$

where the integral kernel is

$$K(\tilde{r},\tilde{r}';k) \equiv -G(\tilde{r},\tilde{r}';k)V(\tilde{r}'), \tag{18}$$

$V(\tilde{r}) \equiv V(\tilde{X},x)$, $F_0(\tilde{r},k) \equiv \phi_0(\tilde{X})e^{-ikx}$, and the Green's function $G(\tilde{r},\tilde{r}';k)$ is defined by

$$[H(\tilde{X},\tilde{P}) + K(p) - E(k)]G(\tilde{r},\tilde{r}';k) = \delta(\tilde{r} - \tilde{r}'). \tag{19}$$

Introducing an orthonormal complete basis set $\{\phi_i(\tilde{X})\}$ for the matter Hamiltonian $H(\tilde{X},\tilde{P})$, we can write the Green's function G as

$$G(\tilde{r},\tilde{r}';k) = \sum_i \int \frac{dk'}{2\pi} \frac{e^{ik'(x-x')}}{E(i) + k'^2 - k^2 + i\epsilon} \times \phi_i(\tilde{X})\phi_i^*(\tilde{X}'), \tag{20}$$

where $E(i)$ is the energy difference between the states $\phi_i(\tilde{X})$ and $\phi_0(\tilde{X})$, and we have put $\hbar^2/2m = 1$. In (20) we have added the term $i\epsilon$ (ϵ = infinitesimal positive number) in the denominator to describe an outgoing wave.

In carrying out the k' integration in (20), as will become clear below, we need only consider k in the region D surrounded by the contour C: $[-k_0,k_0]$, $[k_0,k_0 - i\infty]$, $[k_0 - i\infty, -k_0 - i\infty]$, and $[-k_0 - i\infty, -k_0]$, where k_0 is an infinitesimal positive number. On the other hand, our 1-D matter is finite, and thus the excitation above the ground state has a gap, that is, $E(i) > 0$. Therefore, for k in the region D, it is always realized that

$$E(i,k) \equiv [E(i) - k^2]^{1/2} > 0. \tag{21}$$

With (21) in mind, we perform a contour integral over k' to obtain

$$G(\tilde{r},\tilde{r}';k)=\sum_i \frac{e^{-E(i,k)|x-x'|}}{2E(i,k)}\phi_i(\tilde{X})\phi_i^*(\tilde{X}'). \tag{22}$$

The integral equation (17) can be solved formally by the Fredholm method[13]:

$$F(\tilde{r},k)=F_0(\tilde{r},k)+\frac{1}{\Delta}\int d\tilde{r}'\,\Delta(\tilde{r},\tilde{r}')F_0(\tilde{r}',k), \tag{23}$$

where

$$\Delta=1+\sum_{n=1}^{\infty}\frac{(-)^n}{n!}\int d\tilde{r}_1\cdots\int d\tilde{r}_n \times\begin{vmatrix} K_{11} & \cdots & K_{1n}\\ \vdots & & \vdots\\ K_{n1} & \cdots & K_{nn}\end{vmatrix}, \tag{24}$$

where $K(\tilde{r}_i,\tilde{r}_j)$ is abbreviated as K_{ij} and

$$\Delta(\tilde{r},\tilde{r}')=K(\tilde{r},\tilde{r}')+\sum_{n=1}\frac{(-)^n}{n!}\int d\tilde{r}_1\cdots\int d\tilde{r}_n \times\begin{vmatrix} K_{rr'} & K_{r1} & \cdots & K_{rn}\\ K_{1r'} & K_{11} & \cdots & K_{1n}\\ \vdots & \vdots & & \vdots\\ K_{nr'} & K_{n1} & \cdots & K_{nn}\end{vmatrix}. \tag{25}$$

We note that both $F_0(\tilde{r},k)$ and the kernel $K(\tilde{r},\tilde{r}')$, as given by (18) and (22), are analytic in region D. Therefore if the Fredholm series in (24) and (25) converge, we reach the conclusion that the Jost solution $F(\tilde{r},k)$ as given by (23) is also analytic in region D.

We now show the convergence of Δ. In a similar way, we can show the convergence of $\Delta(\tilde{r},\tilde{r}')$. We first note that from (18) and Hadamard's inequality[14] we can write

$$\begin{aligned}&\int d\tilde{r}_1\cdots\int d\tilde{r}_n \det_{1\le i,j\le n}\|k_{ij}\|\\ &\le\int dx_1\cdots\int dx_n\int d\tilde{X}_1\cdots\int d\tilde{X}_n\\ &\quad\times|V(\tilde{r}_1)\cdots V(\tilde{r}_n)|\cdot\|g_1\|\cdots\|g_n\|,\end{aligned} \tag{26}$$

where $\|g_i\|$ is the norm of the ith column vector of the matrix G_{ij}. Next, since our k is in the low-energy region D, the excitation of the matter from its ground state $\phi_0(\tilde{X})$ is limited to a finite number of low-lying excited states, i.e., with some integer I, (22) gives

$$|G_{rr'}|\le\sum_{i<I}\frac{|\phi_i(\tilde{X})\phi_i(\tilde{X}')|}{2E(i,k)}. \tag{27}$$

The wave functions of low-lying excited states are well localized in the $\tilde{X}$ space, and therefore, when carrying out the integrations $\int d\tilde{X}_1\cdots\int d\tilde{X}_n$ in (26), one can apply the average-value theorem. This means that there exist a certain constant vector $\tilde{X}_0$ and finite constants A and B such that

$$\begin{aligned}&\int d\tilde{X}_1\cdots\int d\tilde{X}_n|V(\tilde{r}_1)\cdots V(\tilde{r}_n)|\cdot\|g_1\|\cdots\|g_n\|\\ &=|V(\tilde{X}_0,x_1)|\cdots|V(\tilde{X}_0,x_n)|\\ &\quad\times\int d\tilde{X}_1\cdots\int d\tilde{X}_n\|g_1\|\cdots\|g_n\|\end{aligned}$$

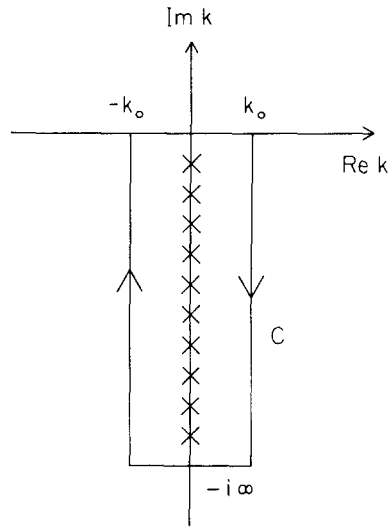


FIG. 2. The contour C in the integral of (31). The crosses on the negative imaginary axis denote the zeroes of the Jost function $f(k)$.

$$\le|V(\tilde{X}_0,x_1)|\cdots|V(\tilde{X}_0,x_n)|A^n(Bn^{1/2})^n. \tag{28}$$

Physically, $\tilde{X}_0$ describes a most probable configuration of the matter atoms at low temperatures. We finally note that since our $|V(\tilde{X}_0,x)|$ decays faster than x^{-2} at $x\to\infty$ by assumption,

$$\int dx|V(\tilde{X}_0,x)|\le M<\infty. \tag{29}$$

From (26), (28), and (29) we obtain

$$\int d\tilde{r}_1\cdots\int d\tilde{r}_n \det_{1\le i,j\le n}\|K_{ij}\|\le(MAB)^n n^{n/2}, \tag{30}$$

which assures the convergence of Δ.

IV. DYNAMICAL LEVINSON'S THEOREM

In the preceding sections, we have discussed some general properties of the Jost function $f(k)$ and its analytic property in the complex k plane. We are now ready to claim the existence of dynamical Levinson's theorem in a similar manner as in the static potential scattering,

$$\begin{aligned}N_b&=-\frac{1}{2\pi i}\int_C dk\frac{f'(k)}{f(k)}=-\frac{1}{2\pi i}\int_C d[\ln f(k)]\\ &=-(1/2\pi)[\delta(-0)-\delta(+0)]=\delta(+0)/\pi,\end{aligned} \tag{31}$$

where N_b is the number of bound states of H_{tot}, and the contour C is as given in Fig. 2. This can be seen as follows: since the Jost function $f(k)$ is analytic in the region D surrounded by the contour C, the integrand $f'(k)/f(k)$ has simple poles of unit strength at zeroes of $f(k)$, each of which corresponds to a bound state. For q degenerate bound states, the strength of the corresponding pole is q. This is the first equality in (31). The remaining equalities in (31) are trivial from (10), (15), (16), and the analyticity of $f(k)$ in region D.

V. CONCLUSION

In this paper, we have considered a 1-D model which describes a particle dynamically interacting with a finite 1-D

matter. We have shown that if the matter has a nondegenerate ground state and is well localized in space, and hence the collision of the particle with the matter is well defined, and if the particle–matter potential decays faster than an inverse quadratic of the distance, there exists a dynamical version of Levinson's theorem, connecting the zero-energy phase shift $\delta(+0)$ to the number of bound states of the total system. This dynamical Levinson's theorem has recently played an essential role in the study of low-temperature adsorption.[6] Furthermore, in light of its general, many-body character, we expect its fruitful applications in other physical problems.

ACKNOWLEDGMENTS

This research was supported by the Office of Naval Research, the Air Force Office of Scientific Research (AFSC), United States Air Force, under Contract No. F49620-86-C-0009, and the National Science Foundation under Grant No. CHE-8519053.

[1]L. I. Schiff, *Quantum Mechanics* (McGraw-Hill, New York, 1968), p. 344.
[2]R. G. Newton, *Scattering Theory of Waves and Particles* (McGraw-Hill, New York, 1966).
[3]J. M. Jauch, Helv. Phys. Acta. **30**, 143 (1957).
[4]E. Kazes, Nuovo Cimento **13**, 983 (1959).
[5]M. Ida, Prog. Theor. Phys. **21**, 625 (1959).
[6]S. G. Chung and T. F. George, to be published.
[7]D. O. Edwards, Physica **109, 110**B, 1531 (1982).
[8]S. G. Chung and T. F. George, to be published.
[9]T. R. Knowles and H. Suhl, Phys. Rev. Lett. **39**, 1417 (1977).
[10]G. Doyen, Phys. Rev. B **22**, 497 (1980).
[11]R. Jost and A. Pais, Phys. Rev. **82**, 840 (1951).
[12]T. Y. Wu and T. Ohmura, *Quantum Theory of Scattering* (Prentice–Hall, Englewood Cliffs, NJ, 1962).
[13]A. B. Whittaker and C. D. Watson, *Modern Analysis* (Cambridge U. P, London, 1940), 4th ed., Chap. 11.
[14]See Ref. 13, p. 212.

Double Kerr–Schild equivalence and hyperheavens

Krzysztof Rózga
Institute of Mathematics, The Pedagogical University, 25-406 Kielce, ul. Konopnickiej 21, Poland

(Received 18 March 1985; accepted for publication 8 October 1986)

Double Kerr–Schild equivalence of hyperheavens is studied within a spinorial formalism based on the fact of the existence of a congruence of self-dual null strings. In particular, it is shown that a given hyperheavenly metric structure generates all members of the corresponding equivalence class in terms of a certain spinor field and the generalized key function subject to the generalized hyperheavenly equation. The treatment is manifestly covariant: no coordinates are employed. Finally, a link with the classical results on hyperheavens (Plebański and Robinson [J. F. Plebański and I. Robinson, Phys. Rev. Lett. **37**, 493 (1976)] and Finley and Plebański [J. D. Finley, III and J. F. Plebański, J. Math. Phys. **17**, 2207 (1976)]) is established.

I. INTRODUCTION

A theory of algebraically degenerate space-times is one of those areas of general relativity that has been explored successfully since 1960.[1] That success is founded, in principal, on the Goldberg–Sachs theorem.[2] In fact, it is the existence of a shear-free congruence of null geodesic lines that is employed to construct geometric coordinate systems, in which Einstein equations can be reduced essentially (see Ref. 1 and, for instance, Ref. 3). A generalized formulation of the Goldberg–Sachs theorem in terms of bivectors was provided by Robinson and Schild in Ref. 4.

A field of Debever–Penrose (DP) null directions can be viewed as an intersection of two distributions, of two-dimensional, complex subspaces that are totally null (any two vectors of a subspace are orthogonal) and self-dual or anti-self-dual, respectively. With each of them there is associated a simple bivector. Now, in Einstein spaces,[5] the Goldberg–Sachs theorem states that a DP direction is algebraically degenerate if and only if those distributions are formally integrable, i.e., a collection of all complex vector fields contained pointwise in subspaces of one of these distributions is closed with respect to the Lie bracket operation. That fact can be restated conveniently in terms of the corresponding two-forms. The formal integrability does not yet mean integrability in the usual sense (i.e., the existence of integral submanifolds) since the space-time is real and the subspaces are complex. However, if the space-time can be extended analytically, one gets integrability of the corresponding distributions. Their integral submanifolds are called left or self-dual null strings and right or anti-self-dual null strings, respectively. That fact has been one of the main reasons that brought the idea of a complex space-time into general relativity. The complex version of the Goldberg–Sachs theorem was discussed extensively by Plebański *et al.* in Refs. 6–8.

It turns out that for a general complex Weyl tensor, algebraic types of its self-dual and anti-self-dual counterparts are independent. This is unlike the situation in a real case, where both of them are identical.

The first attempt to integrate Einstein equations was made for complex space-times with a trivial one of the irreducible parts of the Weyl conformal tensor.[9–11] These spaces are referred to as heavens. (There has been an independent interest in them due to Newman[12] and Penrose.[13]) Soon it was realized that similar results can be obtained for hyperheavens, i.e., complex space-times with one-side-degenerate conformal curvature tensor.[14–15] The metric structure of a hyperheaven is determined entirely by one scalar function, fulfilling a differential constraint of the second order.

It turned out that the complex line element of a hyperheaven exhibits an interesting algebraic structure. It is double-Kerr–Schild (dKS)-equivalent to another metric which is flat. To put it in other words, the difference between these two metrics is spanned at each point by two null and orthogonal vectors. Moreover, those vectors are tangent to the null strings associated wtih the algebraically degenerate part of the conformal curvature tensor. Thus

$$ds^2 = \eta + Ak\otimes k + Bm\otimes m + C(k\otimes m + m\otimes k), \quad (1.1)$$

where η is flat, and k and m are null and orthogonal. Those properties of k and m hold with respect to both metrics, ds^2 and η. Moreover,

$$[k,m]\wedge k\wedge m = 0. \quad (1.2)$$

We want to point out that the idea of dKS equivalence is not new. The concept was discovered by Plebański in Ref. 16. (See also Refs. 17 and 18.) In a paper by Plebański and Schild, the authors discussed dKS-equivalent metrics (called there dKS-conjugated) using the null-tetrad formalism.[19]

To describe hyperheavens one constructs a coordinate system $\{q^{\dot A},p^{\dot B}\}$, $\dot A,\dot B = \dot 1,\dot 2$, such that two-dimensional submanifolds $q^{\dot A} = \text{const}$ form the congruence of self-dual null strings and $\{p^{\dot B}\}$ are some parameters along them. Then K and m are spanned by $\{dq^{\dot A}\}$ only. Working in that coordinate system, one obtains from Einstein equations the existence of the key function and the corresponding hyperheavenly equation.[15]

Although this result reflects geometric properties of a complex space-time it has been derived in specific coordinates. In this paper we reformulate it in a covariant way. We get an even more general result. To arrive at it one considers instead of (1.1), the relation

$$ds'^2 = ds^2 + Ak\otimes k + Bm\otimes m + C(k\otimes m + m\otimes k), \quad (1.3)$$

where ds^2 is this time a given hyperheavenly, otherwise arbi-

 0022-2488/87/051107-07$02.50

trary metric structure. For ds'^2 to be a hyperheaven one obtains some conditions on A, B, and C, that can be reduced to the existence of the so-called generalized key function and the corresponding generalized hyperheavenly equation. Obviously, the generalized key function depends on a choice of ds^2. However, the result is manifestly coordinate independent.

We should like to mention here that a problem of a covariant formulation of a theory of hyperheavens has been already discussed in Refs. 20–22. As far as our treatment of this problem is concerned the main emphasis is put into dKS-equivalence aspects of that theory.

In Sec. II the basic concepts, facts, and important formulas concerning complex, one-side degenerate spaces are listed. Most of them were discussed extensively in Ref. 8.

Section III contains the main results of this paper: a brief outline of a proof of the existence of the generalized key function and a derivation of the generalized hyperheavenly equation in its covariant form. In Sec. IV the correspondence between our results and those obtained by Plebański and Robinson[14] and by Finley and Plebański[15] is established. In all that it seems to be very convenient to operate on spinorial objects. In this respect the technique developed by Plebański[8] is appropriate.

II. COMPLEX SPACE-TIMES AND CONGRUENCES OF NULL STRINGS

Throughout this paper by a complex space-time is meant a complex, four-dimensional manifold with a holomorphic metric structure ds^2, that satisfies Einstein equations

$$C_{AB\dot{C}\dot{D}} = 0 = R. \tag{2.1}$$

Here $C_{AB\dot{C}\dot{D}}$ stands for the spinorial image of the trace-free Ricci tensor and R for the Ricci scalar curvature.[8]

A congruence of self-dual null strings is determined by a nowhere vanishing spinor field k_A, such that

$$k^A k^B \nabla_{A\dot{C}} k_B = 0. \tag{2.2}$$

This condition is invariant with respect to arbitrary rescalings of k_A, $k_A \to \psi k_A$, $\psi \neq 0$. However, one finds it[8] more advantageous perhaps, to normalize k_A in such a way that

$$\nabla_{A\dot{B}} k_C = 3Z_{A\dot{B}} k_C + 2\epsilon_{AC} k^M Z_{M\dot{B}}, \tag{2.3}$$

where $Z_{A\dot{B}}$ is another spinor field, called Sommers vector. Now, the rescalings of k_A are reduced to those with ψ being constrained by

$$k^A \nabla_{A\dot{B}} \psi = 0. \tag{2.4}$$

That special normalization of k_A, referred to as canonical, is equivalent to the requirement that the self-dual two-form

$$\Sigma := k_A k_B S^{AB} \tag{2.5}$$

is closed.[8]

A careful examination of integrability conditions for Eqs. (2.3) shows (Ref. 8) that

$$\nabla^A{}_{\dot{C}} Z^B{}_{\dot{D}} = 2k^{(A}\chi^{B)}{}_{\dot{C}\dot{D}} + 2Z^{A}{}_{(\dot{C}} Z^{B}{}_{\dot{D})} - C^{AB}{}_{\dot{C}\dot{D}} + \epsilon_{\dot{C}\dot{D}}\rho^{AB} + \epsilon^{AB}\rho_{\dot{C}\dot{D}} - \epsilon^{AB}\epsilon_{\dot{C}\dot{D}}\eta, \tag{2.6}$$

where

$$\eta = \tfrac{1}{2} Z_{A\dot{B}} Z^{A\dot{B}} - R/24, \tag{2.7}$$

$$2\rho_{AB} = k^S \psi_{SAB}, \quad k_{(A}\psi_{BCD)} = -\tfrac{1}{2} C_{ABCD}, \tag{2.8}$$

$$2\rho_{A\dot{B}} = k^S \chi_{SA\dot{B}}, \quad \chi_{AB\dot{C}} = \chi_{A(B\dot{C})}. \tag{2.9}$$

The deviation one-form θ (also known as an expansion)[8] is

$$\theta := -2k_A k^S Z_{S\dot{B}} g^{A\dot{B}}. \tag{2.10}$$

Here $\{g^{A\dot{B}}\}$ stands for one-forms (further on referred to as null-tetrad one-forms) in terms of which the expression for the metric tensor ds^2 reads

$$ds^2 = -\tfrac{1}{2} g_{A\dot{B}} \underset{s}{\otimes} g^{A\dot{B}}. \tag{2.11}$$

The Goldberg–Sachs theorem for complex space-times states that the conformal curvature spinor C_{ABCD} is algebraically degenerate, with k_A being its multiple DP spinor, i.e.,

$$C_{ABCD} = 6k_{(A}k_B\psi_{CD)}, \tag{2.12}$$

if and only if k_A determines a congruence of null strings (Ref. 8).

By a hyperheaven is meant, in this paper, a complex space-time that satisfies condition (2.12).

From now on all discussion of this section concerns hyperheavens. We remark, however, that similar results can be obtained under much weaker conditions (consult Ref. 8). And so, the Sommers vector can be put into a form of

$$Z_{A\dot{B}} = k_A \alpha_{\dot{B}} + \tfrac{1}{2} \nabla_{A\dot{B}} \ln \Phi. \tag{2.13}$$

Next, let $\nabla_{\dot{A}}$ be an operator of a covariant derivative along null strings defined by

$$\nabla_{\dot{A}} := \Phi^{-1/2} k^B \nabla_{B\dot{A}}, \tag{2.14}$$

and let $\Omega_{\dot{C}\dot{D}\cdots\dot{E}}$ be an arbitrary spinor field with dotted indices only. It is not difficult to show that

$$\nabla^{\dot{A}} \nabla_{\dot{A}} \Omega_{\dot{C}\dot{D}\cdots\dot{E}} = 0. \tag{2.15}$$

The computation is straightforward. It involves Ricci identities,[8] Eqs. (2.3) and (2.13). Equation (2.15) show that the comutator of $\nabla_{\dot{A}}$ and $\nabla_{\dot{B}}$ is zero, when considered as an operator acting on dotted spinor fields. This feature of $\nabla_{\dot{A}}$ is very useful in computations. As an application of that property one gets a lemma.

Lemma: The equation $\nabla_{\dot{C}} \rho_{\dot{D}\cdots\dot{E}} = \Omega_{\dot{C}\dot{D}\cdots\dot{E}}$ is completely integrable (i.e., its integrability conditions are satisfied identically) if and only if $\nabla^{\dot{C}} \Omega_{\dot{C}\dot{D}\cdots\dot{E}} = 0$.

Complete integrability implies that the solution exists for arbitrary "initial data," where "initial data" means a two-dimensional surface (transversal at each of its points to the corresponding null string) together with a spinor field $\overset{0}{\rho}{}_{\dot{D}\cdots\dot{E}}$ along it. (That type of a Cauchy–Kowalewski-like problem has been discussed in a similar context in Ref. 23.)

We notice for future references that, in particular, the equations of the form

$$\nabla_{\dot{A}} \rho_{\dot{B}} = 0 \tag{2.16}$$

and

$$\nabla_{\dot{B}} P^{\dot{A}} = \delta^{\dot{A}}{}_{\dot{B}} \tag{2.17}$$

are completely integrable.

Now, let $\theta_{\dot{A}}$ be defined by

$$\theta^{\dot{A}} := k^{M} Z_{M}{}^{\dot{A}}. \tag{2.18}$$

It is not difficult to show employing Eqs. (2.14), (2.3), (2.6), (2.1), and (2.12) that

$$\Phi^{1/2} \nabla^{\dot{A}} \theta^{\dot{B}} = 3\theta^{\dot{A}} \theta^{\dot{B}}. \tag{2.19}$$

We remark here that in spaces with a congruence of null strings such that the corresponding k^{A} is a multiple DP spinor field, Eqs. (2.19) is equivalent to the condition

$$k^{A} k^{B} C_{ABC\dot{D}} = 0. \tag{2.20}$$

The same remark applies to the validity of Eqs. (2.15). Now, notice that because of Eqs. (2.14), we have

$$\theta^{\dot{A}} = \tfrac{1}{2} \Phi^{1/2} \nabla^{\dot{A}} \ln \Phi. \tag{2.21}$$

This being substituted back into Eqs. (2.19) provides a constraint on Φ,

$$\nabla^{\dot{A}} \nabla^{\dot{B}} \Phi^{-1} = 0. \tag{2.22}$$

Let $J^{\dot{A}}$ be a spinor field defined according to

$$J^{\dot{A}} := \nabla^{\dot{A}} \Phi^{-1}, \tag{2.23}$$

and let $P^{\dot{A}}$ be a particular solution of Eqs. (2.17). Then

$$\nabla^{\dot{A}} J^{\dot{B}} = 0, \tag{2.24}$$

and

$$\Phi^{-1} = J_{\dot{A}} P^{\dot{A}} + K, \tag{2.25}$$

where the function K is constant along null strings, i.e.,

$$\nabla_{\dot{A}} K = 0. \tag{2.26}$$

From now on, any spinor field annihilated by the operator $\nabla_{\dot{A}}$ will be referred to as a *spinor field covariantly constant along null strings.*

At the conclusion of this section we remark, that $J^{\dot{A}} \neq 0$ if and only if the deviation one-form θ is nontrivial ($\theta \neq 0$). That fact is a simple consequence of Eqs. (2.10), (2.18), (2.21), and (2.23).

III. dKS EQUIVALENCE OF HYPERHEAVENS

(i) dKS-*equivalent complex metrics*: Let $\{g^{A\dot{B}}\}$ be null-tetrad one-forms for the metric tensor ds^2 [Eqs. (2.11)]. It is not difficult to prove that two metrics ds'^2 and ds^2 are dKS-equivalent in the sense of (1.3), iff there exist spinor fields k_A ($k_A \neq 0$) and $\omega_{\dot{A}\dot{B}}$,

$$\omega_{\dot{A}\dot{B}} = \omega_{(\dot{A}\dot{B})}, \tag{3.1}$$

such that the one-forms

$$g'^{A\dot{B}} = g^{A\dot{B}} + k^{A} \omega^{\dot{B}}{}_{\dot{N}} k_{M} g^{M\dot{N}} \tag{3.2}$$

represent null-tetrad one-forms for the metric tensor ds'^2. Then

$$ds'^2 := -\tfrac{1}{2} g'_{A\dot{B}} \underset{s}{\otimes} g'^{A\dot{B}} = ds^2 - \omega_{B\dot{N}} k_{A} g^{A\dot{B}} \underset{s}{\otimes} k_{M} g^{M\dot{N}}. \tag{3.3}$$

Further on, we shall assume ds^2 to be a hyperheavenly metric structure and k_A to be the corresponding multiple DP spinor field constrained, due to the Goldberg-Sachs theorem (Sec. II), by the condition (2.3). Our ultimate goal is to obtain and then to integrate the conditions on $\omega_{\dot{A}\dot{B}}$ for ds'^2 to be another hyperheavenly line element. We adopt the convention that all objects related to ds'^2 are endowed with "prime."

At the first stage one finds relations between the connection one-forms Γ'_{AB}, $\Gamma'_{\dot{A}\dot{B}}$ and Γ_{AB}, $\Gamma_{\dot{A}\dot{B}}$. (For definitions of Γ's consult Ref. 8.) To this end one applies to both sides of formula (3.2) an operator of the external covariant differentiation D (Ref. 8). Then one employs the first structural equations

$$Dg^{A\dot{B}} = 0, \tag{3.4}$$

and the like for primed objects. One gets as the result

$$H^{A}{}_{M} \wedge g'^{M\dot{B}} + H^{\dot{B}}{}_{\dot{M}} \wedge g'^{A\dot{M}} = D(k^{A} \omega^{\dot{B}}{}_{\dot{N}} k_{M}) \wedge g^{M\dot{N}}, \tag{3.5}$$

where

$$H_{AB} := \Gamma_{AB} - \Gamma'_{AB} \tag{3.6}$$

and

$$H_{\dot{A}\dot{B}} := \Gamma_{\dot{A}\dot{B}} - \Gamma'_{\dot{A}\dot{B}}. \tag{3.7}$$

Condition (3.5) can be solved for the components of H_{AB} and $H_{\dot{A}\dot{B}}$, i.e., $H_{ABC\dot{D}}$ and $H_{\dot{A}\dot{B}C\dot{D}}$ ($H_{AB} = -\tfrac{1}{2} H_{ABC\dot{D}} g^{C\dot{D}}$ and $H_{\dot{A}\dot{B}} = -\tfrac{1}{2} H_{\dot{A}\dot{B}C\dot{D}} g^{C\dot{D}}$) by simple algebraic manipulations involving spinorial techniques. The final formulas are

$$\begin{aligned} H_{ABC\dot{D}} = {} & k_{(A} k_{B} \nabla_{C)}{}^{\dot{N}} \omega_{\dot{N}\dot{D}} + 6 k_{(A} k_{B} Z_{C)}{}^{\dot{N}} \omega_{\dot{N}\dot{D}} \\ & - 2\omega k_{A} k_{B} k_{C} k^{N} Z_{N\dot{D}} - k_{A} k_{B} k_{C} \omega^{\dot{M}}{}_{\dot{N}} k_{M} \nabla^{M\dot{N}} \omega_{\dot{M}\dot{D}} \\ & + \tfrac{1}{3} k_{(A} \epsilon_{B)C} k_{N} \nabla^{N\dot{N}} \omega_{\dot{N}\dot{D}} \end{aligned} \tag{3.8}$$

and

$$H_{\dot{A}\dot{B}C\dot{D}} = k_{C} k_{M} \{ \nabla^{M}{}_{\dot{D}} \omega_{\dot{A}\dot{B}} - \epsilon_{\dot{C}(\dot{A}} \nabla^{M\dot{N}} \omega_{\dot{B})\dot{N}} \}, \tag{3.9}$$

where

$$\omega := -\tfrac{1}{2} \omega_{\dot{A}\dot{B}} \omega^{\dot{A}\dot{B}}, \tag{3.10}$$

and covariant derivatives of k_A have been already eliminated by means of Eqs. (2.3).

At the next stage one finds the correspondence between S'^{AB},$S'^{\dot{A}\dot{B}}$, and S^{AB},$S^{\dot{A}\dot{B}}$. (Their definitions can be found in Ref. 8.) The computations are straightforward and they amount to

$$\begin{aligned} S'^{AB} &:= \tfrac{1}{2} \epsilon_{\dot{R}\dot{S}} g'^{A\dot{R}} \wedge g'^{B\dot{S}} \\ &= S^{AB} - \omega k^{A} k^{B} k_{M} k_{N} S^{MN} - k^{A} k^{B} \omega_{\dot{C}\dot{D}} S^{\dot{C}\dot{D}}, \end{aligned} \tag{3.11}$$

and

$$S'^{\dot{A}\dot{B}} := \tfrac{1}{2} \epsilon_{RS} g'^{R\dot{A}} \wedge g'^{S\dot{B}} = S^{\dot{A}\dot{B}} - k_{C} k_{D} \omega^{\dot{A}\dot{B}} S^{CD}. \tag{3.12}$$

Now, one is prepared to establish relations between $C'_{AB\dot{C}\dot{D}}$, R' and $C_{AB\dot{C}\dot{D}}$, R. To this end one employs the second structural equations (Ref. 8).

One obtains from them

$$\begin{aligned} C'_{ABCD} = {} & C_{ABCD} - k_{(A} k_{B} \{ u_{CD)\dot{C}\dot{D}} + C_{CD)\dot{C}\dot{D}} \} \omega^{\dot{C}\dot{D}} \\ & - \omega k_{(A} k_{B} \{ C_{CD)MN} + v_{CD)MN} \} k^{M} k^{N} \\ & + \tfrac{1}{12} R \omega k_{A} k_{B} k_{C} k_{D} + v_{(ABCD)}, \end{aligned} \tag{3.13}$$

$$\begin{aligned} C'_{AB\dot{C}\dot{D}} = {} & C_{AB\dot{C}\dot{D}} + u_{AB\dot{C}\dot{D}} - (C_{ABCD} k^{C} k^{D} \\ & - \tfrac{1}{12} R k_{A} k_{B} + v_{ABCD} k^{C} k^{D}) \omega_{\dot{C}\dot{D}}, \end{aligned} \tag{3.14}$$

and

$$R' = R - 4v^{CD}{}_{CD} + 4\omega k^{A} k^{B} k^{C} k^{D}(C_{ABCD} + v_{ABCD}) + 4k^{C} k^{D}(u_{CD\dot{C}\dot{D}} + C_{CD\dot{C}\dot{D}})\omega^{\dot{C}\dot{D}}, \tag{3.15}$$

where

$$u_{AB\dot{C}\dot{D}} = -\tfrac{1}{2}\{\nabla^{M}{}_{(\dot{C}} H_{|ABM|\dot{D})} + H_{AMN(\dot{D}} H^{M}{}_{|B|}{}^{N}{}_{\dot{C})}\} \tag{3.16}$$

and

$$v_{ABCD} = \tfrac{1}{2}\{\nabla_{(D}{}^{\dot{N}} H_{|AB|C)\dot{N}} + H_{AM(C|\dot{N}|} H^{M}{}_{|B|D)}{}^{\dot{N}}\}. \tag{3.17}$$

(Explicit formulas for v_{ABCD} and $u_{AB\dot{C}\dot{D}}$ are listed in Appendix A.)

(ii) *The generalized key function:* The conditions $R' = 0$ and $k^{A} C'_{AB\dot{C}\dot{D}} = 0$ can be expressed in terms of $\omega_{\dot{A}\dot{B}}$ by means of Eqs. (3.15), (3.14), (2.12), (A1), and (A5). They read

$$\nabla^{A}\nabla^{B}\omega_{\dot{A}\dot{B}} = 0, \tag{3.18}$$

and

$$\nabla^{\dot{N}}\Phi^{2}\nabla_{(\dot{C}}\omega_{\dot{D})\dot{N}} = 0. \tag{3.19}$$

We remark here, that k_A is a multiple DP spinor for C'_{ABCD} [Eqs. (3.13)], i.e., $C'_{ABCD} k^{B} k^{C} k^{D} = 0$. Indeed, it is true for C_{ABCD} [Eqs. (2.12)] and one can verify that $v_{(ABCD)} k^{B} k^{C} k^{D} = 0$ [Eqs. (A8)]. This fact shows the consistence of requirements that both ds^2 and ds'^2 are to be hyperheavenly metric tensors. An integration of Eqs. (3.18) and (3.19) can be carried out as it has been done in similar circumstances in Ref. 15 (see also Refs. 20–22). Indeed, although in our approach ∇_A is an operator of a covariant derivative defined in a coordinate independent way, it can be treated formally as "$\partial/\partial p^{A}$" due to its property expressed by Eqs. (2.15).

Again, there are two cases to be considered.

Case 1: $J_{\dot{A}} \neq 0$ (expanding case): It is not difficult to show that the general solution of Eqs. (3.18) and (3.19) can be represented by

$$\omega_{\dot{A}\dot{B}} = \alpha_{\dot{A}\dot{B}} + \nabla_{(\dot{A}}\Phi^{-4}\nabla_{\dot{B})} W, \tag{3.20}$$

where $\alpha_{\dot{A}\dot{B}}$ is a symmetric covariantly constant along null strings spinor field and W is a function called from now on the generalized key function.

Case 2: $J_{\dot{A}} = 0$ (nonexpanding case): In this case $\Phi^{-1} = K$ [Eqs. (2.25)] and consequently $K \neq 0$. The general solution of Eqs. (3.18) and (3.19) can be represented by

$$\omega_{\dot{A}\dot{B}} = -\tfrac{1}{3}K^{-1} L_{(\dot{A}} P_{\dot{B})} + K^{4}\nabla_{\dot{A}}\nabla_{\dot{B}} W, \tag{3.21}$$

where $L_{\dot{A}}$ is covariantly constant along null strings spinor field and $P_{\dot{A}}$ is a particular solution of Eqs. (2.17). We remark that the nonexpanding case can be thought of as a limit of the expanding one[15] and therefore a further discussion of this paper concerns the general case only.

It is to be pointed out that there are alternative forms of $\omega_{\dot{A}\dot{B}}$. Our representation of $\omega_{\dot{A}\dot{B}}$ [Eqs. (3.20)] seems to be very natural, however, it is not the one that has been employed in Ref. 15. That point is to be explained later in Sec. IV. Also in our representation of $\omega_{\dot{A}\dot{B}}$ there is some ambiguity for $\alpha_{\dot{A}\dot{B}}$ and W. It is not difficult to figure out. It turns out that the transformations, which preserve $\omega_{\dot{A}\dot{B}}$, are of the form

$$W \to W + \Delta W, \tag{3.22}$$

and

$$\alpha_{\dot{A}\dot{B}} \to \alpha_{\dot{A}\dot{B}} + \Delta\alpha_{\dot{A}\dot{B}}, \tag{3.23}$$

where

$$\Delta W = \Delta_{0,0} + \Phi^{3}\{\Delta_{-3,0} + (P^{\dot{M}} J_{\dot{M}})\Delta_{-2,0} + (P^{\dot{M}} K_{\dot{M}})\Delta_{-3,1}\}, \tag{3.24}$$

$$\Delta\alpha_{\dot{A}\dot{B}} = 2J_{\dot{A}} J_{\dot{B}}\Delta_{-2,0} + 2J_{(\dot{A}} K_{\dot{B})}\Delta_{-3,1}. \tag{3.25}$$

The coefficients $\Delta_{m,n}$'s are covariantly constant along null strings, and $K^{\dot{A}}$ is a covariantly constant along null strings spinor field such that

$$K^{\dot{A}} J_{\dot{A}} = 1. \tag{3.26}$$

Consequently, one can always arrange $\Delta\alpha_{\dot{A}\dot{B}}$ in such a way that the new $\alpha_{\dot{A}\dot{B}}$ fulfills the following condition:

$$\alpha_{\dot{A}\dot{B}} K^{\dot{B}} = 0. \tag{3.27}$$

(iii) *The generalized hyperheavenly equation:* It turns out that a pattern of an integration of Einstein equations discovered in Refs. 14 and 15 (see also Refs. 20–22) manifests itself in our approach as well. To be more specific, let l_A be a spinor field such that

$$k^{A} l_{A} = 1. \tag{3.28}$$

The field equations $C'_{AB\dot{C}\dot{D}} = 0$ for the metric tensor ds'^2 (one assumes that $R' = 0$ has been already satisfied) can be split into a triplet of equations

$$C'_{AB\dot{C}\dot{D}} k^{A} k^{B} = 0, \tag{3.29}$$

$$C'_{AB\dot{C}\dot{D}} k^{A} l^{B} = 0, \tag{3.30}$$

and

$$C'_{AB\dot{C}\dot{D}} l^{A} l^{B} = 0. \tag{3.31}$$

It turns out that with Eqs. (3.29) and (3.30) being satisfied, Eq. (3.31) can be put into a form of

$$\nabla_{\dot{C}}\nabla_{\dot{D}}\Lambda = 0. \tag{3.32}$$

To find the corresponding Λ one has to pull out the operator $\nabla_{\dot{C}}\nabla_{\dot{D}}$ in front of the left-hand side member of Eqs. (3.31), expressed already in terms of $W, \alpha_{\dot{A}\dot{B}}$ and their derivatives. That requires a repeated application of the Leibnitz rule for differentiation of a product of spinor fields, Ricci identities, and Eqs. (2.3), (2.6), (2.13), and (3.28). During that process one gets some compensating terms, which are not of the form $\nabla_{\dot{C}}\nabla_{\dot{D}}\nu$. It turns out, however, that the total contribution from those terms is equal to zero. (See also Appendix B.) All of that procedure is manifestly covariant; no coordinates are employed.

Before an expression for the corresponding Λ is provided, we would like to make a remark. We observe that from the Bianchi identities

$$\nabla_{M}{}^{A} C_{A\dot{B}\dot{C}\dot{D}} = 0, \tag{3.33}$$

it follows that

$$\nabla^{A} C_{A\dot{B}\dot{C}\dot{D}} = 0. \tag{3.34}$$

Equations (3.34) in turn, imply that there exists a function C, such that

$$C_{A\dot{B}C\dot{D}} = \nabla_{\dot{A}}\nabla_{\dot{B}}\nabla_{\dot{C}}\nabla_{\dot{D}}C. \tag{3.35}$$

This result is a simple consequence of the lemma from Sec. II.

Now an expression for L can be written down. It reads

$$\begin{aligned}\Lambda = & -\tfrac{1}{2}\Phi^{-5}(\nabla_{\dot{A}}\Phi^2\nabla_{\dot{B}}\Phi^{-3}W)(\nabla^{\dot{A}}\Phi^2\nabla^{\dot{B}}\Phi^{-3}W)\\ & -\alpha^{\dot{A}\dot{B}}\nabla_{\dot{A}}\nabla_{\dot{B}}(\Phi^{-3}W) - \Phi^{-4}\Box W\\ & +(12\Phi^{-4}Z^{\dot{A}\dot{B}} + 5\Phi^{-3}\nabla^{\dot{A}\dot{B}}\Phi^{-1})\nabla_{\dot{A}\dot{B}}W\\ & +3\Phi^{-3}\{\Box\Phi^{-1} + \Phi(\nabla^{\dot{A}\dot{B}}\Phi^{-1})(\nabla_{\dot{A}\dot{B}}\Phi^{-1})\}\\ & \times W - 2C_{\dot{A}\dot{B}}\alpha^{\dot{A}\dot{B}},\end{aligned} \tag{3.36}$$

where

$$\Box := -\tfrac{1}{2}\nabla_{\dot{A}\dot{B}}\nabla^{\dot{A}\dot{B}}, \tag{3.37}$$

$$C_{\dot{A}\dot{B}} = \nabla_{\dot{A}}\nabla_{\dot{B}}C. \tag{3.38}$$

Equation (3.32), after being integrated twice, takes the form

$$\Lambda = N_{\dot{A}}P^{\dot{A}} + \Gamma, \tag{3.39}$$

where Λ is determined by (3.36) and $N_{\dot{A}}$ and Γ are covariantly constant along null strings. This equation is called from now on the generalized hyperheavenly equation. Notice that Λ does not depend on l_A, as one could expect it, since l_A has been an auxiliary spinor field only [Eqs. (3.28)].

There is still some freedom in the structural elements that constitute an expression for Λ, i.e., Φ, k^A, W, $\alpha_{\dot{A}\dot{B}}$, and $C_{\dot{A}\dot{B}}$. We now list the corresponding transformations and their effect on Λ.

(a) $C_{\dot{A}\dot{B}} \to C_{\dot{A}\dot{B}} + \Delta C_{\dot{A}\dot{B}}$,

where $\Delta C_{\dot{A}\dot{B}}$ is such that

$$\Delta C_{\dot{A}\dot{B}} = \nabla_{\dot{A}}\nabla_{\dot{B}}\Delta C$$

and

$$\nabla_{\dot{A}}\nabla_{\dot{B}}\nabla_{\dot{C}}\nabla_{\dot{D}}\Delta C = 0.$$

Then $\Lambda \to \Lambda - 2\alpha^{\dot{A}\dot{B}}\Delta C_{\dot{A}\dot{B}}$.

(b) $\Phi \to \chi\Phi$, $W \to \chi^5 W$, $C_{\dot{A}\dot{B}} \to \chi C_{\dot{A}\dot{B}}$,

where χ is such that $\chi \neq 0$ and $\nabla_{\dot{A}}\chi = 0$. Then $\Lambda \to \chi\Lambda$.

(c) $k^A \to \psi k^A$, $W \to \psi^{-4}W$,

$\alpha_{\dot{A}\dot{B}} \to \psi^{-2}\alpha_{\dot{A}\dot{B}}$, $C_{\dot{A}\dot{B}} \to \psi^{-2}C_{\dot{A}\dot{B}}$,

where ψ is such that $\psi \neq 0$ and $\nabla_{\dot{A}}\psi = 0$. Then $\Lambda \to \psi^{-4}\Lambda$.

(d) $W \to W + \Delta W$, $\alpha_{\dot{A}\dot{B}} \to \alpha_{\dot{A}\dot{B}} + \Delta\alpha_{\dot{A}\dot{B}}$,

where ΔW and $\Delta\alpha_{\dot{A}\dot{B}}$ are given by Eqs. (3.22)–(3.25). Then $\Lambda \to \Lambda + \Delta\Lambda$, where $\Delta\Lambda$ is such that $\nabla_{\dot{C}}\nabla_{\dot{D}}\Delta\Lambda = 0$.

An explicit form of $\Delta\Lambda$ is this time more complicated and it will be presented elsewhere.

IV. A STANDARD APPROACH TO HYPERHEAVENS

In this section we show how the original results on hyperheavens [14–15] can be obtained within our formalism by certain specifications of coordinates and the metric structure ds^2.

And so, a coordinate system $\{q^{\dot{A}}, p^{\dot{B}}\}$ is the one associated with a congruence of null strings (null strings are determined by equations $q^{\dot{A}} = \text{const}$). The metric tensor ds^2 is flat and it is of the form

$$ds^2 = 2\Phi^{-2}\,dq^{\dot{A}}dp_{\dot{A}}, \tag{4.1}$$

where

$$\Phi = j_{\dot{A}}p^{\dot{A}} + \kappa \tag{4.2}$$

and $j_{\dot{A}}$ and κ are numerically constant.

We choose null-tetrad one-forms $\{g^{A\dot{B}}\}$ to be

$$g^{1\dot{A}} = -\sqrt{2}\,dp^{\dot{A}}, \tag{4.3}$$

and

$$g^{2\dot{A}} = -\sqrt{2}\Phi^{-2}\,dq^{\dot{A}}. \tag{4.4}$$

A spinor field k^A in its canonical normalization [Eqs. (2.3)] is given by

$$k^A = (\Phi^2/\sqrt{2})\delta^A_1. \tag{4.5}$$

Next, we study spinor fields covariantly constant along null strings, i.e., solutions of the equation

$$\nabla_{\dot{A}}u_{\dot{B}} = 0. \tag{4.6}$$

An explicit formula for the operator $\nabla_{\dot{A}}$ can be worked out. Then Eqs. (4.6) take the form of

$$\frac{\partial u_{\dot{B}}}{\partial p^{\dot{A}}} + \tfrac{1}{2}\Phi^{-1}(j_{\dot{B}}u_{\dot{A}} + \epsilon_{\dot{A}\dot{B}}u^{\dot{M}}j_{\dot{M}}) = 0. \tag{4.7}$$

It is not difficult to find its general solution. It turns out that

$$u_{\dot{A}} = \Phi^{-1/2}\{uj_{\dot{A}} + v(p_{\dot{A}} + (\kappa/\tau)k_{\dot{A}})\}, \tag{4.8}$$

where $k_{\dot{A}}$ is a fixed, numerically constant spinor, such that

$$k^{\dot{A}}j_{\dot{A}} = \tau \neq 0, \tag{4.9}$$

and u and v are arbitrary functions of $q^{\dot{A}}$'s only, i.e., constant along null strings. In particular, $J^{\dot{A}}$ of Eqs. (2.23) is of that form. Indeed, one can verify that

$$J_{\dot{A}} = \nabla_{\dot{A}}\Phi^{-1} = -\Phi^{-1/2}j_{\dot{A}}, \tag{4.10}$$

which corresponds to $(u,v) = (-1,0)$ in formula (4.8).

As for a spinor field $K_{\dot{A}}$ [Eqs. (3.26)], it can be taken in the form of

$$K_{\dot{A}} = -\Phi^{-1/2}(p_{\dot{A}} + (\kappa/\tau)k_{\dot{A}}). \tag{4.11}$$

Finally, we observe that

$$P_{\dot{A}} = -(\Phi^{-1/2}/\tau)k_{\dot{A}} + \Phi^{-1/2}K(p_{\dot{A}} + (\kappa/\tau)k_{\dot{A}}) \tag{4.12}$$

is a particular solution of Eqs. (2.17). With that choice of $P_{\dot{A}}$ formulas (2.25) and (4.2) are consistent.

We now assume that the generalized key function and the corresponding spinor field $\alpha_{\dot{A}\dot{B}}$ have been arranged already in such a way that $\alpha_{\dot{A}\dot{B}}K^{\dot{B}} = 0$ [Eqs. (3.27)]. Consequently,

$$\alpha_{\dot{A}\dot{B}} = -\mu K_{\dot{A}}K_{\dot{B}}, \tag{4.13}$$

where μ is a function constant along null strings.

Let the function W_s be defined by

$$W_s = \tfrac{1}{4}\mu\Phi^4(P^{\dot{M}}K_{\dot{M}})^2 + \tfrac{1}{2}W. \tag{4.14}$$

It turns out that W_s is the key function of the standard ap-

proach to hyperheavens.[15] That fact can be verified by a substitution of W from Eqs. (4.14) into Eqs. (3.39). Additionally one has to represent all covariant derivatives involved in Eqs. (3.39) as well as $Z_{A\dot{B}}$ in the coordinate system $\{q^{\dot{A}}, p^{B}\}$. The computations are straightforward.

V. DISCUSSION

(i) The results of this paper reveal a mechanism according to which any hyperheavenly metric tensor ds^2 can be used to generate new hyperheavens. To this end the generalized hyperheavenly equation has to be solved for the generalized key function [Eqs. (3.2), (3.20), (3.36), and (3.39)]. These results generalize those of Refs. 14 and 15, which from the point of view of our formalism can be obtained by an assumption that an initially given metric tensor ds^2 is flat (see Sec. IV).

(ii) An assumption that the metric tensor ds^2 represents a hyperheaven does not seem to be very essential. Indeed, some facts, for instance, Eqs. (2.13) and (2.15), do not require it. Therefore of particular interest is a question concerning the weakest conditions on ds^2 under which the whole procedure of an integration of the field equations for ds'^2 [Eqs. (3.3)] works again. A partial answer to that question is already known; ds^2 can be conformally flat (for details see Refs. 20–22).

(iii) In this paper one does not touch at all a problem of hyperheavens with a cosmological constant and an electromagnetic field. They were discussed within a standard formalism of Refs. 14 and 15 in Refs. 24 and 25, respectively (see also Refs. 20–22). One expects, therefore, the corresponding generalization to exist.

(iv) A formalism of this paper is manifestly covariant: no coordinates are employed. A slightly different approach to a problem of a covariant formulation of a theory of hyperheavens (with a cosmological constant and an electromagnetic field) has been proposed in Refs. 20–22. The author of those papers employs canonical coordinates $\{p^{\dot{A}}, q^{\dot{B}}\}$ and to arrive at a covariant expression for the hyperheavenly equation, directional derivatives are translated into the corresponding covariant ones. A discussion is confined to a hyperheavenly metric tensor ds'^2 in a form of Eqs. (1.3), where ds^2 is conformally flat.

ACKNOWLEDGMENT

The work for this paper was partially supported by the U. S. National Science Foundation under Grant No. PHY-8306104.

APPENDIX A: EXPRESSIONS FOR $u_{AB\dot{C}\dot{D}}$ AND v_{ABCD}

Here we list expressions for $u_{AB\dot{C}\dot{D}}$ [Eqs. (3.16)] and v_{ABCD} [Eqs. (3.17)]:

$$u_{AB\dot{C}\dot{D}} = k_A k_B u_{\dot{C}\dot{D}} + k_{(A} u_{B)\dot{C}\dot{D}}, \tag{A1}$$

where

$$u_{\dot{C}\dot{D}} = 4\theta^{\dot{N}}\theta^{\dot{M}}\omega_{\dot{N}\dot{C}}\omega_{\dot{M}\dot{D}} - \omega\theta_{\dot{C}}\theta_{\dot{D}} - 18Z_M{}^{\dot{N}}Z^{M}{}_{(\dot{C}}\omega_{\dot{D})\dot{N}} + \theta^{\dot{N}}\omega_{\dot{N}(\dot{C}}\tilde{\nabla}^{\dot{M}}\omega_{\dot{D})\dot{M}} + 3Z_{M(\dot{C}}\nabla^{M\dot{N}}\omega_{\dot{D})\dot{N}} + 3\nabla_{M(\dot{C}}Z^{M\dot{N}}\omega_{\dot{D})\dot{N}}$$
$$- \tilde{\nabla}_{(\dot{C}}\omega\theta_{\dot{D})} + \tfrac{1}{2}\omega^{M}{}_{\dot{N}}\theta_{(\dot{C}}\tilde{\nabla}^{\dot{N}}\omega_{\dot{D})M} - \tfrac{1}{2}\tilde{\nabla}_{(\dot{C}}\omega^{M\dot{N}}\tilde{\nabla}_{|\dot{N}|}\omega_{\dot{D})M} + \tfrac{1}{2}\tilde{\nabla}_{M(\dot{C}}\nabla^{M\dot{N}}\omega_{\dot{D})\dot{N}}, \tag{A2}$$

$$u_{B\dot{C}\dot{D}} = 6Z_{B(\dot{C}}\omega_{\dot{D})\dot{N}}\theta^{\dot{N}} - 12Z_B{}^{\dot{N}}\omega_{\dot{N}(\dot{C}}\theta_{\dot{D})} + 2\nabla_{B(\dot{C}}\omega_{\dot{D})\dot{N}}\theta^{\dot{N}} + \tfrac{3}{2}Z_{B(\dot{C}}\tilde{\nabla}^{\dot{N}}\omega_{\dot{D})\dot{N}} - 2\theta_{(\dot{C}}\nabla_{|B|}{}^{\dot{N}}\omega_{\dot{D})\dot{N}} + \tfrac{1}{2}\nabla_{B(\dot{C}}\tilde{\nabla}^{\dot{N}}\omega_{\dot{D})\dot{N}}, \tag{A3}$$

and

$$\tilde{\nabla}_{\dot{A}} := k^{M}\nabla_{M\dot{A}}. \tag{A4}$$

Next,

$$v_{ABCD} = \tfrac{1}{2}\nabla_{(C}{}^{\dot{N}}H_{|AB|D)\dot{N}} + \tfrac{1}{2}H_{AM(C|\dot{N}|}H^{M}{}_{|B|D)}{}^{\dot{N}}, \tag{A5}$$

where

$$\nabla_{(C}{}^{\dot{N}}H_{|AB|D)\dot{N}} = k_A k_B\{\nabla_{(C}{}^{\dot{M}}\nabla_{D)}{}^{\dot{N}}\omega_{\dot{M}\dot{N}} + 12Z_{(C}{}^{\dot{N}}\nabla_{D)}{}^{\dot{M}}\omega_{\dot{M}\dot{N}} + 6\omega_{\dot{M}\dot{N}}\nabla_{(C}{}^{\dot{N}}Z_{D)}{}^{\dot{M}} + k_{(C}\nabla_{D)}{}^{\dot{N}}(\omega^{\dot{M}}{}_{\dot{L}}\tilde{\nabla}^{\dot{L}}\omega_{\dot{M}\dot{N}} - 2\omega\theta_{\dot{N}})$$
$$+ 9k_{(C}Z_{D)}{}^{\dot{N}}(\omega^{\dot{M}}{}_{\dot{L}}\tilde{\nabla}^{\dot{L}}\omega_{\dot{M}\dot{N}} - 2\omega\theta_{\dot{N}}) + 36Z_{(C}{}^{\dot{N}}Z_{D)}{}^{\dot{M}}\omega_{\dot{M}\dot{N}} - k_{(A}\epsilon_{B)(C}\{4\theta^{\dot{N}}\nabla_{D)}{}^{\dot{M}}\omega_{\dot{M}\dot{N}}$$
$$+ \nabla_{D)}{}^{\dot{N}}\tilde{\nabla}^{\dot{M}}\omega_{\dot{M}\dot{N}} + 4\nabla_{D)}{}^{\dot{N}}\theta^{\dot{M}}\omega_{\dot{M}\dot{N}} + 4k_{D)}\theta^{\dot{N}}\omega^{\dot{M}}{}_{\dot{L}}\tilde{\nabla}^{\dot{L}}\omega_{\dot{M}\dot{N}} + 3Z_{D)}{}^{\dot{N}}\tilde{\nabla}^{\dot{M}}\omega_{\dot{M}\dot{N}} + 36Z_{D)}{}^{\dot{N}}\theta^{\dot{M}}\omega_{\dot{M}\dot{N}}\}$$
$$- 2\epsilon_{A(C}\epsilon_{D)B}\{\theta^{\dot{N}}\tilde{\nabla}^{\dot{M}}\omega_{\dot{M}\dot{N}} + 4\theta^{\dot{N}}\theta^{\dot{M}}\omega_{\dot{M}\dot{N}}\} \tag{A6}$$

and

$$H_{AM(C|\dot{N}|}H^{M}{}_{|B|D)}{}^{\dot{N}} = 6k_{(C}Z_{D)}{}^{\dot{M}}\omega^{\dot{N}}{}_{\dot{M}}\tilde{\nabla}^{\dot{R}}\omega_{\dot{R}\dot{N}} + k_A k_B k_C k_D\{\omega^{\dot{L}}{}_{\dot{M}}(\tilde{\nabla}^{\dot{R}}\omega_{\dot{R}\dot{N}})\tilde{\nabla}_{\dot{L}}\omega^{\dot{M}\dot{N}} - 4\theta_{\dot{R}}\omega^{\dot{N}\dot{R}}\omega^{\dot{M}\dot{L}}\tilde{\nabla}_{\dot{L}}\omega_{\dot{M}\dot{N}}$$
$$- 2\omega\theta^{\dot{M}}\tilde{\nabla}^{\dot{N}}\omega_{\dot{M}\dot{N}} - 8\omega\theta^{\dot{M}}\theta^{\dot{N}}\omega_{\dot{M}\dot{N}} - 12\omega Z_{M\dot{N}}Z^{M\dot{N}}\} - k_A k_B k_{(C}(\nabla_{D)}{}^{\dot{M}}\omega_{\dot{M}\dot{N}})\{\tilde{\nabla}^{\dot{R}}\omega^{\dot{N}}{}_{\dot{R}} + 4\theta^{\dot{R}}\omega^{\dot{N}}{}_{\dot{R}}\}. \tag{A7}$$

In particular one obtains from Eqs. (A5)–(A7) that

$$v_{(ABCD)}k^{B}k^{C}k^{D} = 0. \tag{A8}$$

APPENDIX B: THE STRUCTURE OF AN EXPRESSION FOR $l^A l_B C'_{ABCD}$

An expression for $l^A l^B C'_{ABCD}$ (with $C_{ABCD} = 0$) can be worked out from Eqs. (3.14), (2.12), and (A1)–(A7). After some slight changes due to the Ricci identities, it reads

$$l^A l^B C'_{ABCD} = N_{CD} + L_{CD}, \tag{B1}$$

where

$$\begin{aligned} N_{CD} = & -\omega_{CD}\tilde{\nabla}^{M}\theta^{N}\omega_{NM} + \theta^{N}\omega_{N(C}\tilde{\nabla}^{M}\omega_{D)M} \\ & + 3\theta^{M}\theta^{N}\omega_{MC}\omega_{ND} + \tfrac{1}{2}\omega^{M}{}_{N}\theta_{(C}\tilde{\nabla}^{N}\omega_{D)M} \\ & - \tfrac{1}{2}\tilde{\nabla}_{(C}\{2\omega\theta_{D)} + \omega^{MN}\tilde{\nabla}_{|N|}\omega_{D)M}\} \end{aligned} \tag{B2}$$

and

$$\begin{aligned} L_{CD} = & \omega_{CD}\{-\psi_{MN}k^{M}k^{N} + 6Z_{MN}Z^{MN} + 3\theta^{N}\nabla_{MN}l^{M} \\ & - l^{M}\nabla_{M}{}^{N}\theta_{N}\} + \omega_{N(C}\{9\theta_{D)}l^{M}Z_{M}{}^{N} \\ & + l^{M}\nabla_{|M|}{}^{N}\theta_{D)} + 3\theta_{D)}\nabla_{M}{}^{N}l^{M}\} \\ & + 3\tilde{\nabla}_{(C}\omega_{D)N}l_{M}Z^{MN} + \theta^{N}l^{M}\nabla_{M(C}\omega_{D)N} \\ & + \tfrac{1}{2}l_{M}\tilde{\nabla}_{(C}\nabla^{MN}\omega_{D)N}. \end{aligned} \tag{B3}$$

We remark that indices C and D are in positions corresponding to operators ∇_C and ∇_D. To see this fact it suffices to recall an expression for ω_{AB} [Eqs. (3.20)], Eqs. (2.21), (A4), and to confront them with Eqs. (B2) and (B3).

[1] I. Robinson and A. Trautman, Phys. Rev. Lett. **4**, 431 (1960); Proc. R. Soc. London Ser. A **265**, 463 (1962).
[2] J. N. Goldberg and R. K. Sachs, Acta Phys. Pol. **22**, 13 (1962).
[3] G. C. Debney, R. P. Kerr, and A. Schild, J. Math. Phys. **10**, 1842 (1969).
[4] I. Robinson and A. Schild, J. Math. Phys., **4**, 484 (1963).
[5] Einstein space is an empty space-time with a cosmological constant. For a discussion of Goldberg–Sachs theorem in that case we refer the reader to the next three references.
[6] J. F. Plebański and S. Hacyan, J. Math. Phys. **16**, 2403 (1975).
[7] M. Przanowski and J. F. Plebański, Acta Phys. Pol. B **10**, 485, 573 (1979).
[8] J. F. Plebański and K. Rózga, J. Math. Phys. **25**, 1930 (1984).
[9] J. F. Plebański, J. Math. Phys. **16**, 2395 (1975).
[10] J. D. Finley, III and J. F. Plebański, J. Math. Phys. **17**, 585 (1976).
[11] C. P. Boyer and J. F. Plebański, J. Math. Phys. **18**, 1022 (1977).
[12] E. T. Newman, Gen. Relativ. Gravit. **7**, 107 (1976).
[13] R. Penrose, Gen. Relativ. Gravit. **7**, 31 (1976).
[14] J. F. Plebański and I. Robinson, Phys. Rev. Lett. **37**, 493 (1976). See also in the *Proceedings of the Symposium on Asymptotic Structure of Space-Time*, University of Cincinnati, Ohio, June, 1976, edited by F. P. Esposito and L. Witten (Plenum, New York, 1977).
[15] J. D. Finley, III and J. F. Plebański, J. Math. Phys. **17**, 2207 (1976).
[16] J. F. Plebański, Ann. Phys. (NY) **90**, 196 (1975).
[17] J. F. Plebański, Ann. NY Acad. Sci. **262**, 247 (1975).
[18] J. F. Plebański and M. Demiański, Cal. Tech. preprint, OAP-401, April, 1975.
[19] J. F. Plebański and A. Schild, Nuovo Cimento B **35**, 35 (1976).
[20] G. F. Torres del Castillo, J. Math. Phys. **24**, 590 (1983).
[21] G. F. Torres del Castillo, J. Math. Phys. **25**, 342 (1984).
[22] G. F. Torres del Castillo, J. Math. Phys. **26**, 152 (1985).
[23] I. Robinson and K. Rózga, J. Math. Phys. **25**, 1941 (1984).
[24] A. García, J. F. Plebański, and I. Robinson, Gen. Relativ. Gravit. **8**, 841 (1977).
[25] J. D. Finley, III and J. F. Plebański, J. Math. Phys. **18**, 1662 (1977).

New analytical models for anisotropic spheres in general relativity

J. Ponce de León[a)]

Universidad Simon Bolivar, Division de Fisica y Matematicas, Departamento de Fisica, Apartado 80659, Caracas 1081-A, Venezuela and Departamento de Fisica, Facultad de Ciencias, Universidad Central de Venezuela, Caracas 1051, Venezuela

(Received 8 July 1986; accepted for publication 10 December 1986)

Two new exact analytical solutions to Einstein's field equations representing static fluid spheres with anisotropic pressures are presented. One solution has a maximum value of mass of about 0.42 times the radius of the fluid sphere, dictated by causality and the corresponding values for the surface red shift and the central red shift are 1.58 and 9.03, respectively. The other solution has a maximum mass of about 0.435 times the radius of the fluid sphere and the corresponding red shifts from the surface and from the center are 1.77 and 16.28, respectively. In the low mass limit both solutions reduce to the constant density Schwarzschild interior solution.

I. INTRODUCTION

In investigations concerning massive objects in general relativity the matter distribution is usually assumed to be locally isotropic. However, in the last few years theoretical studies on realistic stellar models indicate that some massive objects may be locally anisotropic.[1–5] In the literature there are a number of interesting solutions that have provided insight into the effects of anisotropy on star parameters.[6–8] Nevertheless, many of these solutions have a limited applicability to astrophysical situations since they do not satisfy certain physical restrictions usually imposed upon density and pressure, viz., that the pressures should not exceed the energy density (dominant energy condition), and that the (adiabatic) derivatives of the pressure with respect to the density should be less than or equal to unity[9] (macrocausality condition).

In this paper we propose two new analytical models for anisotropic, static fluid spheres. These models are exact solutions to the Einstein's field equations and have reasonable physical properties. The metric coefficients are sufficiently simple so that all quantities of interest (such as mass, radius, surface, and central red shifts) can be computed exactly. One of the assumptions made for obtaining the solutions is that the space-time be conformally flat. This assumption has been widely used in the literature,[10–20] and it is well known that the Schwarzschild interior solution is the unique conformally flat, static perfect fluid space-time. Therefore in the low mass limit ($\ll$the fluid radius) the solutions obtained become identical (to first order in mass/radius) to the Schwarzschild interior solution.

II. FIELD EQUATIONS

The line element for a static, spherically symmetric distribution of matter may be written as

$$ds^2 = e^{\nu(r)}\,dt^2 - e^{\lambda(r)}\,dr^2 - r^2(d\theta^2 + \sin^2\theta\,d\phi^2). \quad (1)$$

With this choice of coordinates the Einstein's field equations for an anisotropic fluid read

$$8\pi\rho = -e^{-\lambda}(1/r^2 - \lambda'/r) + 1/r^2, \quad (2)$$

$$8\pi p_r = e^{-\lambda}(1/r^2 + \nu'/r) - 1/r^2, \quad (3)$$

$$8\pi p_\perp = \frac{e^{-\lambda}}{2}\left(\nu'' + \frac{\nu'^2}{2} + \frac{\nu' - \lambda'}{r} - \frac{\nu'\lambda'}{2}\right), \quad (4)$$

where the primes indicate differentiation with respect to r, ρ is the energy density, and p_r and $p_\perp$ are the radial and tangential "pressure," respectively.

Now we assume that the space-time is conformally flat. As is known, the space-time is conformally flat if all the components of Weyl tensor vanish.[21] For the metric (1) our assumption leads to the equation

$$\frac{e^{\lambda}}{r^2} - \frac{1}{r^2} - \frac{\nu'^2}{4} + \frac{\nu'\lambda'}{4} - \frac{\nu''}{2} - \frac{(\lambda' - \nu')}{2r} = 0. \quad (5)$$

Equations (3)–(5) may be combined to obtain

$$xZ_{,x} + 1 - Z - x\Delta = 0, \quad (6)$$

$$(4Zx^2)y_{,xx} + (2x^2Z_{,x})y_{,x} - (xZ_{,x} - Z + 1)y = 0, \quad (7)$$

where $\Delta \equiv 4\pi(p_\perp - p_r)$, $x \equiv r^2$, $Z \equiv e^{-\lambda}$, $y^2 \equiv e^{\nu}$, and the subscript x following a comma denotes differentiation with respect to x.

Equations (6) and (7) can be formally integrated to obtain the metric functions as follows:

$$e^{-\lambda} \equiv Z = 1 + Cr^2 + 2r^2\int_0^r \frac{\Delta(r)}{r}\,dr, \quad (8)$$

$$e^{\nu} \equiv y^2 = r^2[Ae^{u/2} + Be^{-u/2}]^2, \quad (9)$$

where u is a function of r defined as

$$u(r) = 2\int \frac{e^{\lambda/2}}{r}\,dr, \quad (10)$$

and the constants of integration A, B, and C are specified by matching the metric functions (8) and (9) to the exterior Schwarzschild solution for a mass M, at radius r_0. The result is

$$A = \frac{e^{-u(r_0)/2}}{2r_0}\left[\frac{3M}{r_0} - 1 + \left(1 - \frac{2M}{r_0}\right)^{1/2}\right], \quad (11)$$

[a)] Postal address: Apartado 2816, Caracas 1010-A, Venezuela.

 0022-2488/87/051114-04$02.50

$$B=\frac{e^{u(r_0)/2}}{2r_0}\left[1-\frac{3M}{r_0}+\left(1-\frac{2M}{r_0}\right)^{1/2}\right], \tag{12}$$

$$e^{-\lambda(r_0)}=(1-2M/r_0). \tag{13}$$

III. SOLUTIONS TO THE FIELD EQUATIONS

Examination of Eqs. (2)–(4) and (8)–(10) shows that we have three equations in four unknowns, namely, $\rho, p_r, p_\perp$, and $u(r)$. The problem becomes determinate by choosing any of these unknowns as a function of r or by specifying an equation of state for the stresses. In particular, for the equation of state $p_r=p_\perp$ we recover the well-known Schwarzschild interior solution.

Solution 1: For the first solution we choose the energy density as follows:

$$8\pi\rho=(3C/2)[(3+Cr^2)/(1+Cr^2)^2], \tag{14}$$

where C is a constant to be determined from the boundary conditions. We notice at this point that the perfect fluid solution corresponding to this distribution has been discussed in detail by Durgapal and Bannerji.[22]

In the present case substituting (14) into Eq. (2) we find

$$e^{\lambda}=2(1+\alpha v)/(2-\alpha v), \tag{15}$$

with

$$\alpha=(4M/r_0)/(3-4M/r_0), \tag{16}$$

where M and r_0 are, respectively, the mass and radius of the sphere, and $v\equiv r^2/r_0^2$.

The function $u(r)$ is given by

$$e^{-u}r_0^2=\frac{4+\alpha v+2\sqrt{2}(2+\alpha v-\alpha^2v^2)^{1/2}}{v}\times\exp\left[\sqrt{2}\sin^{-1}\left(\frac{1-2\alpha v}{3}\right)\right]. \tag{17}$$

The metric function e^{ν} can be obtained by substituting this expression into Eq. (9) and using A and B from (11) and (12). The density and radial pressure are as follows:

$$8\pi\rho r_0^2=\tfrac{3}{2}\alpha[(3+\alpha v)/(1+\alpha v)^2], \tag{18}$$

$$8\pi p_r r_0^2=[G(4-5\alpha v-2\sqrt{2}(2+\alpha v-\alpha^2v^2)^{1/2}+\alpha v(4-5\alpha v+2\sqrt{2}(2+\alpha v-\alpha^2v^2)^{1/2})]\times[2v(1+\alpha v)(\alpha v+G)]^{-1}, \tag{19}$$

with

$$G=(B/A)(4+\alpha v+2\sqrt{2}(2+\alpha v-\alpha^2v^2)^{1/2})\times\exp[\sqrt{2}\sin^{-1}((1-2\alpha v)/3)].$$

The tangential pressure is given by

$$8\pi p_\perp r_0^2=8\pi p_r r_0^2+3\alpha^2 v/(1+\alpha v)^2. \tag{20}$$

At the center of the distribution $p_r=p_\perp$, and the ratio of central pressure p_c to central density ρ_c is

$$\frac{p_c}{\rho_c}\cong\frac{1}{3}\left[\frac{A}{4.851B}-2\right]. \tag{21}$$

Solution 2: For the second solution we choose the metric function e^{λ} as follows:

$$e^{-\lambda}\equiv Z=(1+Cr^2)^2/(1-3Cr^2)^2, \tag{22}$$

where C is a constant. Substituting (22) into the field equations and using (9)–(13) we find the second solution in the following form:

$$y^2\equiv e^{\nu}=\frac{[16\beta(1-\beta)^4v+4(1-3\beta)(1+\beta)(1-\beta v)^4]^2}{16(3\beta+1)^4(1-\beta)^4(1-\beta v)^4}, \tag{23}$$

$$e^{\lambda}=(1+3\beta v)^2/(1-\beta v)^2, \tag{24}$$

$$8\pi\rho r_0^2=\frac{24\beta+16\beta^2v+24\beta^3v^2}{(1+3\beta v)^3}, \tag{25}$$

$$8\pi p_r r_0^2=\frac{2\beta(3n+2)(4-8\beta v-12\beta^2v^2)-16\beta(1-\beta v)^4}{[2\beta(3n+2)v+(1-\beta v)^4](1+3\beta v)^2}, \tag{26}$$

$$8\pi p_\perp r_0^2=8\pi p_r r_0^2+\frac{80\beta^2v+48\beta^3v^2}{(1+3\beta v)^3}, \tag{27}$$

where the parameter β was defined as

$$\beta=(1-(1-2M/r_0)^{1/2})/(1+3(1-2M/r_0)^{1/2}), \tag{28}$$

and M and r_0 are, respectively, the mass and radius of the sphere, $v\equiv r^2/r_0^2$, and n is the ratio of central pressure to density,

$$n\equiv p_c/\rho_c=\tfrac{2}{3}[(1-\beta)^4/(1-3\beta)(\beta+1)-1]. \tag{29}$$

IV. PROPERTIES OF THE SOLUTIONS

Solutions 1 and 2 are free of singularities and describe anisotropic spheres whose densities drop continuously from their maximum values at the center to values which are positive at the boundary. From Eqs. (20) and (27) we see that for the two models $p_r=p_\perp$ at the center, whereas $p_\perp>p_r$ for $r>0$. Examination of Eqs. (18)–(21) and (28) and (29) reveals that for the two solutions the central pressure become infinite when the mass reaches $4r_0/9$. It should be noted that this value is exactly the same as is found in the perfect fluid Schwarzschild interior case.

It is easy to prove that in the low mass limit, $M\ll r_0$,

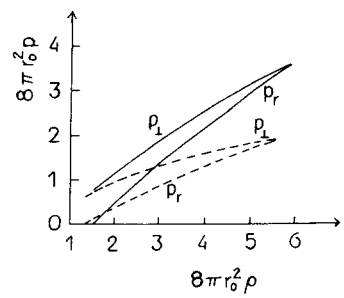


FIG. 1. Equations of state $p_r=p_r(\rho)$ and $p_\perp=p_\perp(\rho)$ for solution 1. Dotted curves correspond to $\rho_c=3p_c$ and $M=0.417r_0$. Heavy curves correspond to $p_c=0.6\rho_c$ and $M=0.425r_0$. For masses greater than $0.425r_0$ the fluid becomes noncausal.

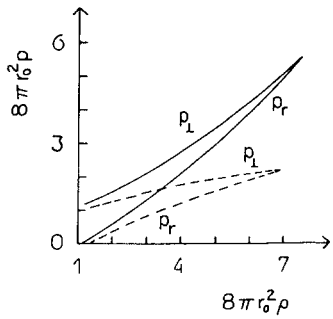


FIG. 2. Equations of state $p_r = p_r(\rho)$ and $p_\perp = p_\perp(\rho)$ for solution 2. Dotted curves correspond to $\rho_c = 3p_c$ and $M = 0.427r_0$. Heavy curves correspond to $p_c = 0.74\rho_c$ and $M = 0.435r_0$. For masses greater than $0.435r_0$ the fluid becomes noncausal.

solutions 1 and 2 and the constant density solution of Schwarzschild have (to first order in M/r_0) the same common limit, namely, the following:

$$e^{-\lambda} = 1 - (2M/r_0)v, \tag{30}$$

$$e^{\nu} = 1 - (M/r_0)(3 - v), \tag{31}$$

where

$$v \equiv r^2/r_0^2.$$

In addition, solution 1 has the following properties.

(i) The dominant energy condition ($\rho > p_r$, $\rho > p_\perp$) holds everywhere within the sphere when the mass is less than about $0.431\, r_0$.

(ii) For $M \lesssim 0.425r_0$, $dp_r/d\rho \leqslant 1$, and $dp_\perp/d\rho \leqslant 1$ throughout the configuration. For $M > 0.425\, r_0$ these derivatives exceed unity in parts of the fluid.

For solution 2 we have found that the dominant energy condition and the macrocausality condition hold at all points within the sphere for $M \lesssim 0.436r_0$ and $M \lesssim 0.435r_0$, respectively. For masses greater than $0.435r_0$ the fluid becomes noncausal.

V. APPLICATIONS

To illustrate the astrophysical applications of the solutions we have calculated the surface gravitational potential M/r_0, the red shift at the center Z_c, and the surface red shift Z_s under different central conditions. To obtain some numerical values for M and r_0 we have taken a particular surface density, viz., $\rho_s = 2\times 10^{14}$ g/cm^3. The results for the first and the second solution are given in Tables I and II, respectively. Figures 1 and 2 show the equations of state p_r vs ρ and $p_\perp$ vs ρ corresponding to solutions 1 and 2 for different values of p_c/ρ_c.

TABLE I. Neutron star parameters as calculated from solution 1.

p_c/ρ_c	M/r_0	Z_c	Z_s	$r_0(Km)$	$M(M_\odot)$
0.1	0.401	3.94	1.247	20.41	5.46
0.333	0.417	6.41	1.454	20.58	5.72
0.6	0.425	9.03	1.581	20.61	5.84
1.000	0.431	12.95	1.691	20.66	5.94
infinite	0.444	infinite	2.000	20.74	6.14

TABLE II. Neutron star parameters as calculated from solution 2.

p_c/ρ_c	M/r_0	Z_c	Z_s	$r_0(Km)$	$M(M_\odot)$
0.1	0.417	5.87	1.45	19.18	5.33
0.33	0.427	9.03	1.61	19.11	5.44
0.74	0.435	16.28	1.77	19.05	5.52
1.00	0.436	18.57	1.79	19.02	5.33
infinite	0.444	infinite	2.00	18.92	5.60

VI. CONCLUSIONS

We have obtained two exact analytical solutions for static spheres with anisotropic pressures. We have seen that the first and second solutions are free of singularities and have reasonable equations of state for masses less than about 0.42 and 0.435 times the radius of the fluid sphere (in geometric units), respectively. The solutions may be used in describing ultracompact objects.[23] An interesting feature of the solutions is the large value for both the central and the surface red shift. For example, from Table II we see that for a neutrino emitted from the center of a star of $M = 0.435r_0$ the red shift at infinity is about 16.28. Tables I and II indicate that for a surface density of 2×10^{14} g/cm^3 the maximum masses dictated by causality, are about of $5.84M_\odot$ and $5.52M_\odot$ for the first and the second model, respectively. Consequently the maximum values for the masses and for the red shifts, in our models, are greater than that obtained from perfect fluid models computed under similar conditions.[22–30] However, it must be emphasized that for each fixed value of p_c/ρ_c an increase (decrease) in the surface density ρ_s correspond to a decrease (increase) in $M/M_\odot$ and also in r_0, so that "neutron star" models smaller (bigger) in mass and size are obtainable from our solutions. The tables also show that the masses and radii of our models are relatively insensitive to the ratio p_c/ρ_c (i.e., to the limiting form of the equation of state at the center). It is interesting to note that a similar feature has the perfect fluid model of Misner and Zapolsky.[26]

[1] M. Ruderman, Ann. Rev. Astron. Astrophys. **10**, 427 (1972).
[2] V. Canuto, *Neutron Stars: General Review*, Solvay Conference on Astrophysics and Gravitation (Brussels, Belgium, 1973).
[3] R. L. Bowers and E. P. T. Liang, Astrophys. J. **188**, 657 (1974).
[4] P. S. Letelier, Phys. Rev. D **22**, 807 (1980).
[5] S. S. Bayin, Phys. Rev. D **26**, 1262 (1982).
[6] M. Cosenza, L. Herrera, M. Esculpi, and L. Witten, J. Math. Phys. **22**, 118 (1981).
[7] K. D. Krori, P. Borgohain, and R. Devi, Can. J. Phys. **62**, 239 (1983).
[8] L. Herrera and J. Ponce de León, J. Math. Phys. **26**, 2018 (1985).
[9] S. W. Hawking and G. F. R. Ellis, *The Large Scale Structure of Space-Time* (Cambridge U. P., Cambridge, 1973), p. 91.
[10] D. R. K. Reddy, J. Math. Phys. **20**, 23 (1979).
[11] A. Banerjee and N. O. Santos, J. Math. Phys. **22**, 824 (1981).
[12] A. Banerjee and N. O. Santos, J. Math. Phys. **22**, 1075 (1981).
[13] B. W. Steward, J. Phys. A **15**, 2419 (1982).
[14] S. R. Maiti, Phys. Rev. D **25**, 2518 (1982).
[15] Z. Shi-Chang, Gen. Relativ. Gravit. **15**, 293 (1983).
[16] E. N. Glass, J. Math. Phys. **20**, 1508 (1979).
[17] H. Knutsen, J. Math. Phys. **24**, 2188 (1983).
[18] H. Knutsen, Phys. Scr. **28**, 357 (1983).

[19]S. B. Kalyanshetti and B. B. Waghmode, Phys. Rev. D **27**, 2835 (1983).
[20]J. Beckers, S. Sinzinkayo, and J. Demaret, Phys. Rev. D **30**, 1846 (1984).
[21]J. L. Singe, *Relativity: The General Theory* (North-Holland, Amsterdam, 1964), p. 341.
[22]M. C. Durgapal and R. Bannerji, Phys. Rev. D **27**, 328 (1983).
[23]B. R. Iyer, C. V. Vishveshwara, and S. V. Dhurandhar, Class. Quantum Gravit. **2**, 219 (1985).
[24]D. N. Pant and A. Sah, Phys. Rev. D **32**, 1358 (1985).
[25]R. J. Adler, J. Math. Phys. **15**, 727 (1974).
[26]C. W. Misner and H. S. Zapolsky, Phys. Rev. Lett. **12**, 635 (1964).
[27]R. C. Adams and J. M. Cohen, Astrophys. J. **198**, 507 (1975).
[28]P. G. Whitman, J. Math. Phys. **18**, 868 (1977).
[29]P. G. Whitman and R. W. Redding, Astrophys. J. **224**, 993 (1978).
[30]P. G. Whitman, J. Math. Phys. **26**, 792 (1984).

On spherically symmetric shear-free perfect fluid configurations (neutral and charged). I

Roberto A. Sussman
School of Mathematical Sciences, Queen Mary College, Mile End Road, London E1 4NS, England

(Received 12 February 1986; accepted for publication 31 December 1986)

A class of solutions describing a wide variety of nonstatic, spherically symmetric, charged, shear-free perfect fluid configurations is derived. It is presented in the form of Jacobian elliptic functions characterized by seven free parameters: five constants and two arbitrary functions of time. This class of solutions is the most general charged version of the class derived by Kustaanheimo and Qvist [Comment. Phys. Math. Helsingf. **13**, 12 (1948); *Exact Solutions of Einstein's Field Equations* (Cambridge U. P., Cambridge, 1980), Chap. 12, Sec. 2]. It is found that many of the charged particular solutions expressible by elementary functions are new. Particular solutions, including neutral and uniform density solutions, are classified in detail. The physical interpretation of these solutions, including the study of their singularity structure, will be presented in a subsequent paper (Part II).

I. INTRODUCTION

A physically motivated approach in the process of finding exact solutions in general relativity would start with supplying a realistic equation of state and a given set of initial conditions; then solving the set of Einstein (or Einstein–Maxwell) equations would determine the dynamical evolution of the configuration. Proceeding with this physical strategy, even assuming spherical symmetry and that matter can be described by a perfect fluid, one usually finds intractable equations requiring numerical integration. Therefore in order to find analytical solutions which (one hopes) would be physically relevant, many authors turn out to a strategy of "mathematical simplicity" which inverts the priorities of the physically motivated strategy.

Spherically symmetric shear-free perfect fluid solutions (which will be denoted hereafter as "SSSF solutions") are a typical example illustrating the strategy of mathematical simplicity. That is, instead of supplying an equation of state and then finding out how the fluid evolves, a given simple fluid motion (i.e., *shear-free* motion) is imposed on the fluid from the outset so that Einstein (or Einstein–Maxwell) equations simplify considerably to the point where they can be integrated analytically. Then, the state variables that make the fluid evolve in this simple manner are found through the same Einstein (or Einstein–Maxwell) equations and the Bianchi identities. Following this procedure, an equation of state need not be imposed, and the state variables are treated as some sort of "agents" ensuring that the fluid moves in the prescribed shear-free manner. However, because of the lack of an equation of state, the Einstein (or Einstein–Maxwell) system of equations remains undetermined, so that the solutions obtained contain arbitrary parameters that cannot be fixed with the information contained in the field equations.

Ideally, the strategy of mathematical simplicity would be successful (from the point of view of physics) if the above mentioned free parameters could be fixed in such a way that the state variables ended up obeying a reasonable equation of state. This is the case in the Friedman–Robertson–Walker (FRW) solutions, though one can argue that these solutions can be derived following a physically motivated strategy plus imposing extra requirements on the symmetry of space-time. Unfortunately, for neutral SSSF solutions other than FRW, this ideal alternative seems to have been ruled out by Mansouri,[1] Glass,[2] Mashhoon and Partovi,[3,4] Collins and Wainwright,[5] Srivastava and Prasad,[6] and Collins.[7,8] These authors have verified conclusively that the Wyman solution[9] is the only SSSF solution (apart from the FRW solutions) compatible with a barotropic equation of state, which turns to be unphysical.[4,7,8] A generalization of this result to charged SSSF solutions was carried on by Mashhoon and Partovi,[3,4] quoted above, and Srivastava and Prasad.[10] As in the neutral case, the charged version of the Wyman solution also admits an unphysical barotropic equation of state. However, Mashhoon and Partovi[3] found another new charged SSSF solution (see Appendix D) compatible with an unphysical barotropic equation of state. Therefore, unless SSSF solutions could be found compatible with physically realistic nonbarotropic equations of state, the applicability of these solutions as models of physically realistic objects seems to be limited. These solutions, though, could still have a theoretical interest as exact "test" solutions surveying the effects of pressure gradient and (if the fluid is charged) electric forces in certain extreme conditions, such as the late stages of a gravitational collapse, or near a space–time singularity, where the lack of a universal equation of state is not a serious handicap.

Most papers on SSSF solutions can be grouped in three categories: (a) authors who discuss the mathematical derivation of the solutions and offer a (sometimes comprehensive) catalog of metrics, (b) authors who have found a given particular solution (or a reduced class of solutions) and examine its physical and geometric properties (sometimes in detail), and (c), authors (the ones mentioned above) who have concentrated on the question of the compatibility of these solutions with a barotropic equation of state, and thus,

 0022-2488/87/051118-19$02.50

one could say, have (to a certain degree) demonstrated the limitations of the strategy of mathematical simplicity.

This paper (Part I) and its continuation (Part II) try to close the gap between these almost mutually exclusive categories mentioned above. In Part I, which fits the category (a), a large class of solutions is derived and classified. This class of solutions will be referred to thereafter in Parts I and II as the "charged Kustaanheimo–Qvist (ChKQ) class," as it is the most general charged version of the "NKQ" class of neutral solutions derived by Kustaanheimo and Qvist.[11,12] Therefore Part I extends and complements the work of many authors in (a), who either have considered only neutral configurations or have derived classes of charged solutions contained in the ChKQ class. Part II examines in detail the general physical and geometric properties common to all or to a large number of ChKQ solutions. Since the solutions dealt with in most papers in category (b) are contained in the ChKQ class, these papers are thus linked and extended in Part II by discussing common important features (conformal structure, singularities, global view, etc.) which have been omitted or marginally studied by these authors. In particular, it will prove very useful to compare the free parameters of the ChKQ solutions with those of the FRW solutions. The work of authors in the category (c) can be appreciated then as a search for the specific choices of free parameters under which a barotropic equation of state is obeyed. Part II will extend and complement the work of some authors in this category, such as Collins and Wainwright,[5] Collins,[7,8] and Mashhoon and Partovi,[4] who have examined some of the above mentioned aspects but only for the case of the Wyman solution.

Going back to the strategy of mathematical simplicity, such a strategy allows one to obtain analytical solutions of the field equations because these equations simplify considerably by reducing the number of independent field variables. This fact can be appreciated from the work of Barnes[13] and Collins and White,[14] who also discussed in detail geometric invariant aspects of shear-free perfect fluid configurations not necessarily restricted to be spherically symmetric. For spherically symmetric neutral configurations using comoving spatial coordinates, the three originally independent metric coefficients are reduced to one independent quantity,[2,12] while for charged configurations one ends up with only two independent quantities: a metric coefficient and an electric potential term.[3,15,16] In either case, analytical expressions for the independent quantities can be obtained by integrating a single nonlinear second-order partial differential equation known in the literature as the equation of "pressure isotropy."[12] This equation, obtained by combining two of the Einstein (or Einstein–Maxwell) equations, has one of two variable coefficients for neutral or charged configurations, respectively.

For neutral fluids, there are different procedures to find solutions of the equation of pressure isotropy. A pioneering work was that of Kustaanheimo and Qvist[11,12] who introduced a suitable ansatz fixing the form of the variable coefficient, and then found the most general solution (the NKQ class) under this restriction. More recent papers by Wyman,[17,18] McVittie,[19] Stephani,[20] and Srivastava[21] have further systematized the methods of integration of the equation of pressure isotropy, investigating which conditions the variable coefficient must satisfy in order to lead to analytical solutions. While Wyman, McVittie, and Srivastava worked within the framework of the theory of differential equations, Stephani applied to the equation of pressure isotropy a more elegant method of obtaining first integrals developed by Lie. These authors have obtained solutions not contained in the NKQ class, and from their work it can be appreciated how the ansatz proposed by Kustaanheimo and Qvist follows as a sufficient condition transforming the equation of pressure isotropy into an equation whose solutions are expressible in terms of elliptic functions.

For charged fluids, the equation of pressure isotropy involves two arbitrary variable coefficients,[3,15] and it should be possible to generalize the algorithms used for solving this equation in the neutral case. In a recent paper, an algorithm proposed by Wyman[17] was extended to the charged case by Chatterjee,[22] obtaining a class of solutions expressible in terms of elliptic functions. However, as will be shown in Sec. V, Chatterjee's expressions are not the most general solutions of the equation of pressure isotropy that one can obtain following this procedure. Such a general class of solutions is the ChKQ class which will be obtained here. However, except for the solutions discussed in Appendix D, charged solutions outside the ChKQ class are still unknown.

The contents of this paper can be summarized as follows: Sec. II shows how the assumption of shear-free motion simplifies the metric associated with spherically symmetric perfect fluid configurations. In Sec. III, the equation of pressure isotropy is integrated leading to a generalization of the ansatz proposed by Kustaanheimo and Qvist for solving this equation in the neutral case (see Kramer *et al.*[12]). Section IV presents the ChKQ solutions obtained in Sec. III expressed in the form of Jacobian elliptic functions. Table I classifies these solutions in terms of the parameters appearing in these functions. Section V presents and classifies particular cases of the ChKQ class. These cases include the class of neutral solutions NKQ that are classified in Table II, the Wyman solution and its electric version mentioned earlier, Chatterjee's solutions,[22] uniform density solutions and solutions in which one of the arbitrary functions of time (time dependent free parameters) appearing in the integration of Eq. (18) is set to a constant. Charged solutions expressible in terms of elementary functions are presented in Sec. VI and classified in Tables III and VIII. With the exception of the charged solution presented in Appendix D, all charged SSSF solutions previously examined in the literature can be identified as particular cases of the ChKQ class. This fact is shown in Table IV, so that one can identify those charged solutions which (as far as I am aware) have never been mentioned or studied before. Section VII presents the subclass of conformally flat uniform density solutions. Neutral solutions and uniform density solutions (including the conformally flat cases) expressible in terms of elementary functions are similarly classified and identified in Tables V–VII.

Appendix A presents the set of Einstein–Maxwell field equations, the nonzero components of the Weyl tensor, and the conformal scalar invariant $\Psi_{(2)}$ calculated for the metric

tensor given by Eq. (11) in Sec. II. Appendix B shows how each solution classified in the tables subdivides into a triplet of solutions, i.e., one solution for each of the three possible values of k in Eqs. (14) and (17). This aspect of the solutions is a sort of generalization of that found in the FRW solutions, in which these values of k distinguish between three different solutions whose surfaces of constant cosmic time have positive, negative, and zero curvature. Appendix C presents the transformation relating the radial coordinate used in this paper with that used by authors working with "isotropic coordinates." Appendix D discusses some simple solutions not belonging to the ChKQ class.

II. SPHERICALLY SYMMETRIC SHEAR-FREE CONFIGURATIONS

Spherically symmetric, nonstatic perfect fluid configurations (whether neutral or electrically charged) can be described by the following metric tensor given in spatially comoving coordinates[12]:

$$ds^2 = g_{\alpha\beta}\, dx^\alpha\, dx^\beta$$
$$= -G^2(t,r)dt^2 + H^2(t,r)dr^2 + R^2(t,r)d\Omega^2, \qquad (1)$$

where

$$d\Omega^2 \equiv d\theta^2 + \sin^2\theta\, d\phi^2$$

and

$$u^\alpha = \delta^\alpha{}_t\,(1/G). \qquad (2)$$

Geometric units have been chosen in (1) and (2), so that the speed of light and Newton's gravitational constant are set to unity. The four-velocity associated with the fluid is u^α, while $\delta^\alpha{}_\beta$ is the Kronecker delta tensor. Greek indices run from zero to three (coordinates t,r,θ,ϕ), while Latin indices run from one to three (coordinates r,θ,ϕ). As usual, semicolons and commas indicate covariant and partial derivatives, respectively.

The kinematical parameters associated with this type of fluid in this particular coordinate representation are the following[2]:

expansion,

$$\Theta \equiv u^\alpha{}_{;\alpha} = (1/G)\,[H/H + 2R/R\,], \qquad (3)$$

acceleration,

$$a_\alpha \equiv u_{\alpha;\beta}u^\beta = \delta_\alpha{}^r G'/G, \qquad (4)$$

shear,

$$\sigma_{\alpha\beta} \equiv u_{(\alpha;\beta)} - a_{(\alpha}u_{\beta)} - (\theta/3)(g_{\alpha\beta} - u_\alpha u_\beta), \qquad (5a)$$

$$\sigma^t{}_t = 0,$$

$$\sigma^r{}_r = \sigma^\theta{}_\theta = -\tfrac{1}{2}\sigma^\phi{}_\phi = (G/3)\,[\dot{R}/R - \dot{H}/H\,], \qquad (5b)$$

where a "dot" and a "prime" refer to partial derivatives with respect to t and r, respectively, and the bracket in the first and second terms of the left-hand side of (5a) denotes symmetrization of the indices α and β.

From Eq. (5b), if the motion of the fluid is to be shear-free, then the metric coefficients H and R are related by

$$\dot{R}/R = \dot{H}/H \;\Rightarrow\; R(t,r) = f(r)H(t,r), \qquad (6)$$

where $f(r)$ is an arbitrary function. As one has the freedom of performing arbitrary coordinate transformations of the form $r^* = r^*(r)$, one can in particular "absorb" $f(r)$ into a new radial coordinate defined as $r^* = f(r)$ (Refs. 12 and 19). However, $f(r)$ will be left unspecified for the time being and will be fixed in Sec. III by a specific coordinate choice.

If the fluid is charged, it is always possible to introduce a gauge[15,16] in which the electromagnetic vector potential has the form $A_\alpha = \delta_\alpha{}^t V(t,r)$. If $q(t,r)$ is the electric charge density and the fluid is nonconducting, the set of Maxwell equations $F^{\alpha\beta}{}_{;\beta} = 4\pi q u^\alpha$ in the comoving representation given by (1) and (2) reduce to

$$(RV'/GH)^{\cdot} = 0 \;\Rightarrow\; V' = (GH/R)E(r), \qquad (7a)$$

$$E' = 4\pi q H R^2. \qquad (7b)$$

A neutral fluid in the presence of a source-free electric field corresponds to the particular case of Eqs. (7) with $E = \text{const}$. The only nonzero components of the stress-energy tensor are

$$T^t{}_t = -\rho - E^2/8\pi R^4, \qquad (8a)$$

$$T^r{}_r = p - E^2/8\pi R^4, \qquad (8b)$$

$$T^\theta{}_\theta = T^\phi{}_\phi = p + E^2/8\pi R^4, \qquad (8c)$$

where $\rho(t,r)$ and $p(t,r)$, respectively, denote the matter density and pressure associated with the fluid. Using the shear-free condition (6), the Einstein–Maxwell equation $G^t{}_r = 0$ for (1) and (8) can be integrated obtaining[12]

$$G(t,r) = g(t)\dot{H}/H, \qquad (9)$$

where $g(t)$ is an arbitrary function. Although one also has the freedom to relabel the time coordinate arbitrarily as $t^* = t^*(t)$, before computing $\partial H/\partial t$ explicitly one cannot "absorb" $g(t)$ into a new time coordinate t^* defined as, say, $t^* = \int g(t)dt$. This is so, because this function would reappear in the transformed metric coefficient $g_{t^*t^*}$ due to the factor $\partial H(t,r)/\partial t = g(t^*)\partial H(t^*,r)/\partial t^*$. Instead, inserting (6) and (3) and comparing with (9), one can identify $g(t)$ as

$$g(t) = 3/\Theta \;\Rightarrow\; \Theta = \Theta(t). \qquad (10)$$

It will be shown in Part II that Θ cannot be determined from the information contained in the field equations, but by physically motivated boundary conditions on the fluid. Different prescriptions of time coordinate will give different relations between Θ and the time derivative of H. For the rest of Part I, the choice of time coordinate will be kept unspecified and Θ will be thought as of an arbitrary function of time appearing in the metric coefficient G.

The metric (1), now specialized for shear-free configurations, is given by

$$ds^2 = -((\dot{H}/H)/(\Theta/3))^2\, dt^2 + H^2[dr^2 + f^2(r)d\Omega^2]. \qquad (11)$$

This simplification of the metric (1), which follows merely by imposing shear-free motion, is the backbone of the strategy of mathematical simplicity mentioned in the Introduction. The set of Einstein–Maxwell field equations for (11) and (8) is presented in Appendix A, these equations are considerably simpler than the equivalent field equations for the more general metric (1). In the next section, following this strategy, analytical expressions will be obtained for the field variables H and E.

III. INTEGRATION OF THE EQUATION OF PRESSURE ISOTROPY

Analytical expressions for H and E can be obtained by solving a nonlinear partial differential equation known in the literature[12] as the "equation of pressure isotropy." This equation, a *consequence* more than a definition of pressure isotropy, is obtained by eliminating p as the Einstein–Maxwell equations are combined in the form

$$G^{\theta}{}_{\theta} - G^{r}{}_{r} = 8\pi(T^{\theta}{}_{\theta} - T^{r}{}_{r}) = 2E^2/(fH)^4. \quad (12)$$

From the form of Einstein–Maxwell equations given in Appendix A, Eq. (12) explicitly reads

$$\frac{1}{H^2}\left[\frac{f'H'}{fH} - \frac{2H'^2}{H^2} + (\dot{H})^{-1}\left(\frac{4\dot{H}'H'}{H} - \dot{H}''\right) - f^{-2}(ff'' - f'^2 + 1)\right] = \frac{2E^4}{(fH)^4}. \quad (13)$$

Looking at Eq. (13) suggests fixing the radial comoving coordinate r by choosing $f(r)$ to satisfy[19,23,24]

$$ff' - f'^2 + 1 = 0 \Rightarrow f'^2 = [1 - kf^2],$$

that is

$$f(r) = \begin{cases} r & (k=0), \quad (14a) \\ \sin r & (k=1), \quad (14b) \\ \sinh r & (k=-1), \quad (14c) \end{cases}$$

so that surfaces of constant t will be manifestly conformal to surfaces of constant curvature. Most authors, though, prefer the so-called "isotropic coordinates" in which these surfaces are conformally flat. The transformation relating isotropic coordinates with those given by (22) is shown in Appendix C.

Using the coordinate choice (14), and multiplying both members of (13) by $H^2\partial H/\partial t$, that equation is easily integrated once with respect to t yielding

$$\left(\frac{H'}{H}\right)' - \frac{H'f'}{Hf} - \left(\frac{H'}{H}\right)^2 + \frac{3j(r)}{H} - \frac{2E^2}{f^4H^2} = 0, \quad (15)$$

where $j(r)$ is an arbitrary function of integration, which, as shown by Eqs. (A4) and (A5) (see Appendix A), is related to the conformal scalar invariant $\Psi_{(2)}$ (Refs. 2 and 16). In order to simplify this equation, it is convenient to introduce the following quantities[23,24]:

$$Y = 1/H, \quad (16a)$$

$$y(r) = \int_0^r f(\bar{r})d\bar{r}, \quad (16b)$$

so that f, now as a function of y, is given by

$$f^2(y) = y(2 - ky) \quad (17)$$

and Eq. (15) transforms into

$$\frac{\partial^2 Y}{\partial y^2} + \frac{3j(y)}{f^2}Y^2 - \frac{2E^2(y)}{f^6}Y^3 = 0, \quad (18)$$

where

$$\frac{\partial Y}{\partial y} = \frac{1}{f}Y'.$$

Equation (18) is a second-order partial differential equation in y, therefore its integration determines completely how H and E depend on r. The time dependence of these variables will be contained in (at most) two arbitrary functions appearing as "constants" of integration. The state variables p, q, and ρ can be expressed in terms of H and E from (7b) and the field equations (A1) and (A2) or (A7) (see Appendix A). However, for each expression $H(t,y)$ and $E(y)$ obtained by solving Eq. (18) one has three different solutions, each one corresponding to one of the values of k in (14), (16), and (17). This is demonstrated rigorously in Appendix B, and is an analogous situation as in the FRW solutions. In the latter, k distinguishes between solutions where surfaces $t = \text{const}$ have constant positive, negative, and zero curvature. For more general shear-free solutions, where these surfaces do not have constant curvature, the nature of the difference between solutions with different k is more complicated and will be discussed in detail in part II.

The most general spherically symmetric shear-free perfect fluid solution would be the most general solution of Eq. (18), that is the solution with arbitrary variable coefficients $j(y)$ and $E(y)$ and two arbitrary functions of time. Since this general solution is intractable, one usually fixes a specific functional dependence of the coefficients j and E on y so that Eq. (18) transforms into a differential equation whose first integral is known. The solutions obtained in this form will then be the most general solutions under these restrictions.

The procedure to be followed in this section consists of finding sufficient conditions that $j(y)$ and $E(y)$ must satisfy so that Eq. (18) becomes a differential equation whose solutions can be given as elliptic functions, that is a differential equation whose first integral has the form

$$\left(\frac{\partial W}{\partial \omega}\right)^2 = \sum_{i=0}^{4}(b_i W^i), \quad (19)$$

where the coefficients b_i are arbitrary constants. This procedure is a particular case of a general method which consists of introducing two transformations: $Y \to Y(W,h)$ and $y \to y(\omega)$, involving arbitrary functions $h(y)$ and $\omega(y)$ to be determined by demanding that Eq. (18) has a first integral expressible in closed analytical form as a quadrature. This method of obtaining first integrals was developed by Lie and has been systematically applied to the neutral case ($E = 0$) of Eq. (18) by Stephani,[20] where it is shown that the first integral of (18) with $E = 0$ will have the form (19) if

$$W(t,y) = h(y)Y(t,y), \quad (20a)$$

$$\frac{\partial}{\partial y} = \frac{d\omega}{dy}\frac{\partial}{\partial \omega}, \quad (20b)$$

provided that $h(y)$ and $\omega(y)$ have the forms specified by Kustaanheimo and Qvist[11,12] and Wyman.[17] Besides more general transformations, Stephani considers transformations like those given by (20) but with forms of $h(y)$ and $\omega(y)$ leading to first integrals of (18) (with $E = 0$) different from (and much more complicated than) (19). In this paper, only the transformation of the type (20) leading to a first integral like (19) will be applied to the charged case ($E \neq 0$) of Eq. (18).

If W and Y are related by (20a) and $\omega(y)$ is defined by (20b), then Eq. (18) has a first integral of the form (19) given specifically by

$$\left(\frac{\partial W}{\partial \omega}\right)^2 = Q(w) \tag{21a}$$

with

$$Q(W) \equiv \epsilon^2 W^4 - 2\mu W^3 + \Delta W^2 + L(t), \tag{21b}$$

where

$$\begin{aligned} &\epsilon \text{ and } \mu \text{ are arbitrary constants,} \\ &L(t) \text{ is an arbitrary function of integration,} \\ &\Delta = b^2 - 4ac, \end{aligned} \tag{21c}$$

provided that $j(y)$, $E(y)$, and $\omega(y)$ satisfy

$$3j(y)/f^2 \equiv 3\mu[h(y)]^5, \tag{22a}$$

$$2E^2(y)/f^6 \equiv 2\epsilon^2[h(y)]^6, \tag{22b}$$

$$\frac{d\omega}{dy} = \frac{1}{h^2}, \tag{22c}$$

where

$$\begin{aligned} &h(y) = [ay^2 + 2by + c]^{-1/2}, \\ &a, b, \text{ and } c \text{ arbitrary constants.} \end{aligned} \tag{23}$$

The form of Eqs. (22a), (22c), and (23) coincide with analogous expressions obtained by Kustaanheimo and Qvist[11,12] (the ansatz that they proposed, as mentioned in the Introduction) and Wyman[17] dealing with the neutral case ($E = 0$). Chatterjee[22] obtained also these expressions plus (22b) for the charged case, but his expression analogous to $Q(W)$ is less general than the one given by (21a). Chatterjee's solutions will be discussed in Sec. V.

A second integration transforms (21a) into an elliptic integral of the first kind:

$$\xi F[\Psi,\eta] = \int_{Z_i}^{Z} \frac{dW}{[Q(W)]^{1/2}} = T(t) + X(y), \tag{24a}$$

where

$$X(y) \equiv \int h^2(y)dy \tag{24b}$$

and $T(t)$ is a second arbitrary function of integration. For the different possible combinations of constant parameters a, b, c characterizing $h(y)$ in (23), the integral of (24b) can take any one of the following five forms:

$$\Delta > 0, \quad a \neq 0, \quad X_{(1)} = \frac{1}{2\sqrt{\Delta}} \ln\left|\frac{ay + b - \sqrt{\Delta}}{ay + b + \sqrt{\Delta}}\right|, \tag{25a}$$

$$\Delta > 0, \quad a = 0, \quad X_{(2)} = (1/2b)\ln|2by + c|, \tag{25b}$$

$$\Delta < 0, \quad X_{(3)} = (1/\sqrt{-\Delta})\tan^{-1}(ay + b)/\sqrt{-\Delta}, \tag{25c}$$

$$\Delta = 0, \quad b^2 = ac, \quad X_{(4)} = -1/\sqrt{a}(\sqrt{ay} + b), \tag{25d}$$

$$\Delta = 0, \quad b = a = 0, \quad X_{(5)} = y/c. \tag{25e}$$

The specific values of the integration limits Z, Z_i in (24a) depend on the form of $Q(W)$ (see the next section and Fig. 1). The spatial dependence of the solutions is governed by the functions $h(y)$ and $X(y)$ as given by Eqs. (23) and (25), different types of particular solutions will arise from different forms of these functions.

The argument $\Psi(W,\eta) = am^{-1}[(T + X)/\xi,\eta]$, modulus η, and a proportionality factor ξ, are defined by the following relation:

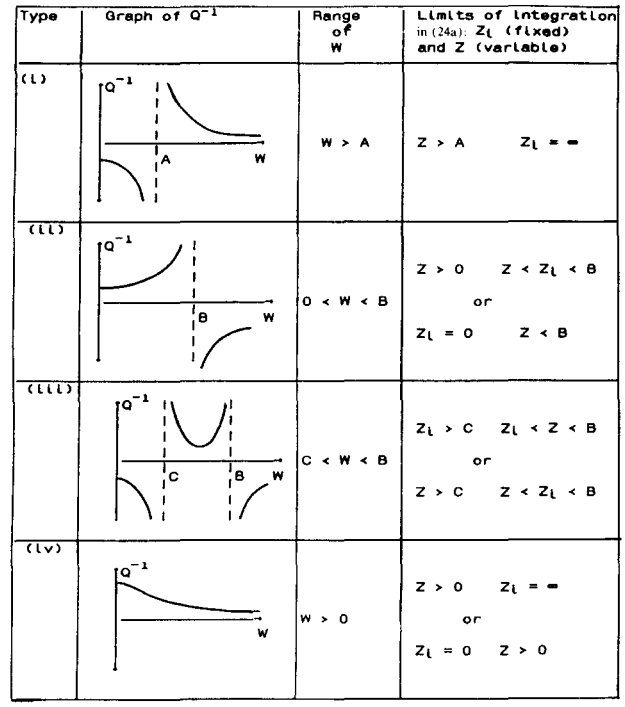


FIG. 1. The integral (24a) must be evaluated for those values $W > 0$ for which $Q(W) > 0$. Thus for different forms of Q, one has four types denoting four different possibilities for range of W and limits of integration in (24a). These types are illustrated in this figure for solutions given by Eqs. (27) and (29). As the range of W is usually contained between consecutive roots of the quartic Q, this classification also characterizes charged solutions expressible by elementary functions where Q is a quartic with repeated roots, and neutral solutions where Q is a cubic. The range of W and the limits of integration of (24a) will be discussed in Part II in connection with the physical and geometric properties of the solutions.

$$\frac{dW}{[Q(W)]^{1/2}} = \frac{\xi\, d\Psi}{[1 - \eta^2 \sin^2\Psi]^{1/2}}, \tag{26}$$

so that the specific form of these quantities depends on the roots of the quartic $Q(W)$. Information about parameters associated to elliptic integrals and functions can be found in any standard text on the subject.[25] Notice that, in general, the modulus and the proportionality factor will depend on t through the function $L(t)$.

Equation (24a) is the general form representing the ChKQ class mentioned in the Introduction, and is the most general solution under the restrictions (22) and (23). It is characterized by seven free parameters: five constants ϵ, μ, a, b, c; and two arbitrary functions of time $L(t)$, $T(t)$. The electric field term $E(y)$ is given by Eq. (22b), so that the electric charge density q can be calculated from (7b). Particular cases will be obtained in Secs. V–VII by imposing restrictions on the free parameters.

IV. THE "CHARGED KUSTAANHEIMO–QVIST" CLASS OF SOLUTIONS

In order to classify the ChKQ class of solutions derived in the previous section, it is convenient to examine qualitatively the nature of the roots of a quartic like $Q(W)$. Since the coefficient ϵ^2 is always non-negative, $Q(W)$ has at least a minimum and $Q \to \infty$ as $W \to \pm\infty$, and also $Q(W)$ has an extremum at $W = 0$. The other parameters, μ and Δ, can be

positive, zero, or negative, so that applying Descartes' rule of signs, $Q(W)$ will have at most three positive real roots ($\mu>0$, $\Delta<0$, $L<0$) or at most three negative real roots ($\mu<0$, $\Delta>0$, $L>0$). In general, a quartic like $Q(W)$ can be written as one of the following three forms, according to the number of real roots:

$$Q_{(4)} = \epsilon^2[\alpha_1 W-\beta_1 A\,][\alpha_2 W-\beta_2 B\,] \times[\alpha_3 W-\beta_3 C\,][\alpha_4 W-\beta_4 D\,], \quad \alpha_i=\pm 1,\ \beta_i=\pm 1,\ i=1,2,3,4, \quad D<C<B<A,\ \alpha_1\alpha_2\alpha_3\alpha_4=1, \tag{27a}$$

$$Q_{(2)} = \epsilon^2[\alpha_1 W-\beta_1 A\,][\alpha_2 W-\beta_2 B\,][(W-C)^2+D^2] \quad \alpha_i=\pm 1,\ \beta_i=\pm 1,\ i=1,2, \quad B<A,\ \alpha_1\alpha_2=1, \tag{27b}$$

$$Q_{(0)} = \epsilon^2[(W-A)^2+B^2][(W-C)^2+D^2]. \tag{27c}$$

However, by definition $W=h/H=fh/R$, and H and R are non-negative, therefore once the positive (negative) sign is chosen in (23) one can disregard negative (positive) roots of $Q(W)$. Another restriction on the range of W arises when $Q(W)<0$, which happens, for example, if $Q(W)$ has a minimum between two consecutive real roots A_i and A_{i+1}. Thus the range of W will correspond to those values for which $Q(W)\geqslant 0$, which together with $W\geqslant 0$, sets for each case in (27) the limits of integration in (24a). Figure 1 shows the four different types of range of W and integration limits of (24a) which can arise for a quartic Q given by Eqs. (27). As will be shown in Part II, the range of W, the roots of $Q(W)$, and the limits of integration of (24a) bear a strong relation to regularity conditions and existence of singularities in these solutions.

Depending on which form one can write $Q(W)$, Eq. (24a) can be inverted so that H is cast in terms of Jacobian elliptic functions[25] as

$$Q_{(4)}:\ \sin\Psi=\mathrm{sn}[(T+X)/\xi,\eta], \quad H=h\frac{1-\alpha^2\,\mathrm{sn}^2[(T+X)/\xi,\eta]}{A_i-A_j\alpha^2\,\mathrm{sn}^2[(T+X)/\xi,\eta]}, \tag{28a}$$

$$Q_{(2)}:\ \cos\Psi=\mathrm{cn}[(T+X)/\xi,\eta], \quad H=h\frac{\gamma_1-\gamma_2+(\gamma_1+\gamma_2)\,\mathrm{cn}[(T+X)/\xi,\eta]}{C\gamma_1-D\gamma_2+(C\gamma_1+D\gamma_2)\mathrm{cn}[(T+X)/\xi,\eta]}, \tag{28b}$$

$$Q_{(0)}:\ \tan\Psi=\mathrm{tn}[(T+X)/\xi,\eta], \quad H=h\frac{1+\delta_1\,\mathrm{tn}[(T+X)/\xi,\eta]}{(C^2-D^2\delta_1)\{1+\delta_2\,\mathrm{tn}[(T+X)/\xi,\eta]\}}, \tag{28c}$$

where the specific form of the parameters α, γ_1, γ_2, δ_1, δ_2 which depend on the roots of Q, is given in Table I.

An alternative form to Eqs. (28), which will be used in Part II, can be obtained by choosing

$$T(t)=-X_0+F[\Psi_0,\eta], \tag{29}$$

where $\Psi_0\equiv\Psi(t,r_0)$, $r_0\geqslant 0$ being a fixed arbitrary value of the radial comoving coordinate. With this choice of T, H in Eqs. (28) becomes a function of H_0, L, and r. Inserting (29) into (28), these expressions become

$$Q_{(4)}:\ H=h(\Pi_1-\alpha^2\Pi_2)/(A_i\Pi_1-\alpha^2 A_j\Pi_2),$$

$$\Pi_1=[1-\eta^2\sin^2\Psi_0\,\mathrm{sn}\,\chi]^2,$$

$$\Pi_2=[\cos\Psi_0(1-\eta^2\sin\Psi_0)^{1/2}\,\mathrm{sn}\,\chi+\sin\Psi_0\,\mathrm{cn}\,\chi\,dn\,\chi]^2,$$

$$\alpha^2\sin^2\Psi_0=(h_0-A_iH_0)/(h_0-A_jH_0), \tag{30a}$$

$$Q_{(2)}:\ H=h\frac{(\gamma_1-\gamma_2)\Pi_1+(\gamma_1+\gamma_2)\Pi_2}{(C\gamma_1-D\gamma_2)\Pi_1+(C\gamma_1+D\gamma_2)\Pi_2},$$

$$\Pi_1=1-\eta^2(1-\eta^2\cos^2\Psi_0),$$

$$\Pi_2=\cos\Psi_0\,\mathrm{cn}\,\chi-\sin\Psi_0(1-\eta^2(1+\cos^2\Psi_0))^{1/2}\,\mathrm{sn}\,\chi\,dn\,\chi,$$

$$\cos\Psi_0\frac{(\gamma_1-\gamma_2)h_0-(C\gamma_1-D\gamma_2)H_0}{(\gamma_1+\gamma_2)h_0-(C\gamma_1+D\gamma_2)H_0}, \tag{30b}$$

$$Q_{(0)}:\ H=h\frac{\Pi_1+\delta_1\Pi_2}{(B-D\delta_1)[\Pi_1+\delta_2\Pi_2]},$$

$$\Pi_1=1-(1-\eta^2\sin^2\Psi_0)^{1/2}\tan\Psi_0\,\mathrm{tn}\,\chi\,dn\,\chi,$$

$$\Pi_2=(1-\eta^2\sin^2\Psi_0)^{1/2}\mathrm{tn}\,\chi+\tan\Psi_0\,dn\,\chi,$$

$$\tan\Psi_0=\frac{(B-D\delta_1)\delta_2H_0-\delta_1h_0}{h_0-(B-D\delta_1)H_0}, \tag{30c}$$

where the argument of the Jacobian elliptic functions sn χ, cn χ, and tn χ is given by $\chi\equiv(X-X_0)/\xi$. As for Eqs. (28), the quantities η, ξ, and the parameters α^2, γ_i, and δ_i are defined in Table I. The forms of the solutions given by (28) and (30) are entirely equivalent, though one form could be more suited than the other for understanding a given aspect of the solutions. In Part II, both forms will be used interchangeably as needed. Table I provides a classification of the general forms characterizing the ChKQ class of solutions.

V. PARTICULAR SOLUTIONS

A. The NKQ class of neutral solutions

If $\epsilon=0$, from Eqs. (7b) and (22b): $q=E=0$, and the solutions describe electrically neutral fluids. Under this restriction, Eqs. (24) reduce to Eq. (14.35), p. 168, of Kramer *et al.*,[12] and one has the NKQ class of solutions discovered by Kustaanheimo and Qvist in 1948. Expressions for H, analogous to (28) and (30), can be obtained for different forms of Q, which is now a cubic in W. If $\epsilon=0$ but $\mu\neq 0$ (the case $\mu=0$ will be considered separately, see Sec. VII), Q can have the following forms depending on the number of real roots:

$$Q_{(3)}=2|\mu|[\alpha_1 W-\beta_1 A\,][\alpha_2 W-\beta_2 B\,][\alpha_3 W-\beta_3 C\,], \tag{31a}$$

$$Q_{(1)}=2|\mu|[\alpha_1 W-\beta_1 A\,][(W-\beta)^2+c^2], \tag{31b}$$

from which one obtains the following expression analogous to (28):

$$Q_{(3)}:\ H=h\frac{\gamma_1+\gamma_2\,\mathrm{cn}^2[(T+X)/\xi,\eta]}{\alpha_1-\alpha_2\,\mathrm{sn}^2[(T+X)/\xi,\eta]}, \quad \text{where } \gamma_1,\gamma_2=0,1, \tag{32a}$$

$$Q_{(1)}:\ H=h\frac{1+\nkappa\,\mathrm{cn}[(T+X)/\xi,\eta]}{(A-\nkappa\delta)+(\nkappa A+\delta)\mathrm{cn}[(T+X)/\xi,\eta]}, \quad \nkappa=\pm 1. \tag{32b}$$

TABLE I. This table provides the values of all unspecified parameters appearing in Eqs. (27), (28), and (30). These parameters are all given in terms of the roots of the quartic Q, and can be computed right away if these roots are known. As the spatial dependence of the solutions is given in terms of y defined by Eq. (16b), one has three different solutions corresponding to each value of k in Eqs. (14) and (17) (see Appendix B). The classification of types (i)–(iv) for the range of W is illustrated in Fig. 1.

$Q_{(\ )}$	η^2	ξ^2	$\Psi(Z,\eta)$	Range of W	Form of Q in (27)	Parameters in (28) and (30)
$Q_{(4)}$	$\dfrac{(B-C)(A-D)}{(A-C)(B-D)}$	$\dfrac{4}{(A-C)(B-D)}$	$\sin^{-1}\left(\dfrac{(B-C)(Z-A)}{(A-C)(Z-B)}\right)^{1/2}$	type (i) $W>A$	$\alpha_1=\alpha_2=\alpha_3=\alpha_4=1$, $\beta_1=1,\quad A>0$.	$\alpha^2=\dfrac{A-D}{B-D}>1$, $A_i=A,\quad A_j=B$.
			$\sin^{-1}\left(\dfrac{(A-C)(B-Z)}{(B-C)(A-Z)}\right)^{1/2}$	type (ii) $0<W<B$	$\alpha_1=\alpha_2=\beta_1=\beta_2=-1$, $\alpha_3=\alpha_4=\beta_3=\beta_4=1$, $D<C<0$.	$\alpha^2=\dfrac{B-C}{A-C}<\eta^2$, $A_i=B,\quad A_j=A$.
				type (iii) $C<W<B$	As in type (ii) with $A>B>C>0$, $D<0$.	
$Q_{(2)}$	$\dfrac{(\gamma_1+\gamma_2)^2-(A-B)^2}{4\gamma_1\gamma_2}$	$\dfrac{1}{\gamma_1\gamma_2}$	$\cos^{-1}\left(\dfrac{(\gamma_2-\gamma_1)Z+\gamma_1B-\gamma_2A}{(\gamma_2+\gamma_1)Z-\gamma_1B-\gamma_2A}\right)$	type (i) $W>A$	$\alpha_1=\alpha_2=\beta_1=1$, $A>0$.	
				type (ii) $0<W<B$	$\alpha_1=\alpha_2=\beta_1=\beta_2=-1$, $0<B<A$.	$\gamma_1^2=(\beta_2B-C)^2+D^2$.
				type (iv) $W>0$	As in type (i) with $B<A<0$.	$\gamma_2^2=(\beta_1A-C)^2+D^2$.
$Q_{(0)}$	$\dfrac{4\sigma_1\sigma_2}{(\sigma_1+\sigma_2)^2}$, $\sigma_1^2=(A-C)^2+(B+D)^2$, $\sigma^2=(A-C)^2+(B-D)^2$.	$\dfrac{4}{(\sigma_1+\sigma_2)^2}$	$\tan^{-1}\left(\dfrac{Z-A+\delta_1B}{B+\delta_1A-\delta_2Z}\right)$	type (iv) $W>0$		$\delta_1^2=\dfrac{4B^2-(\sigma_1-\sigma_2)^2}{(\sigma_1+\sigma_2)^2-4B^2}$, $\delta_2^2=\dfrac{(B+\delta_1A)}{(A-\delta_1B)}$.

The argument of the elliptic functions, $(T+X)/\xi$, has been defined in the same manner as for the ChKQ solutions. The parameters $\alpha_1, \alpha_2, \gamma_1, \gamma_2, \not{\alpha}, \delta$ and the modulus and argument η and Ψ are given in Table II for $Q_{(3)}$, $Q_{(1)}$, and for different ranges of W. These solutions can be cast in the form (30) by inserting (29) in Eqs. (32). Solutions belonging to the NKQ class, expressible in terms of elementary functions, are discussed and classified in Sec. VI. It is worth mentioning that some charged solutions in the class ChKQ do not reduce to any neutral solution in NKQ as $\epsilon \to 0$.

B. Chatterjee's solutions

The expressions presented in (24), (28), and (30) give the most general form for H under the restrictions (22). As it was mentioned in the Introduction, Chatterjee[22] followed the same method of integration of Eq. (18) and obtained a similar class of charged solutions. However, these solutions are contained in the ChKQ class presented in Sec. IV. Chatterjee's solution can be obtained from (21) and (24a) by setting

$$\epsilon^2 = 1, \quad \Delta = \mu^2/\epsilon^2, \quad W = \tilde{W} + \mu/2\epsilon^2, \tag{33}$$

so that Eq. (21a) transforms into

$$\left(\frac{\partial \tilde{W}}{\partial \omega}\right)^2 = \tilde{W}^4 - \frac{\Delta}{2}\tilde{W}^2 + \left(L + \frac{\Delta^2}{16}\right), \tag{34}$$

which coincides with Chatterjee's Eq. (9) if one identifies "A" (in Chatterjee's notation) with $L + \Delta^2/16$. Chatterjee's solutions can be cast in the form (28) and (30) just by specializing the parameters ϵ, μ, Δ as in (33), these solutions correspond to $Q_{(4)}$ type (iii) and $Q_{(0)}$ type (iv) of Table I for $L<0$ and $L>0$, respectively. For each case, the roots of the quartic Q, as it appears in Eqs. (27), are given by

$$Q_{(4)}: \ A = \Delta^{1/2}/2 + [\Delta/4 + |L|^{1/2}]^{1/2} = \Delta^{1/2} - D,$$

$$B = \Delta^{1/2}/2 + [\Delta/4 - |L|^{1/2}]^{1/2} = \Delta^{1/2} - C, \tag{35a}$$

$$\begin{aligned} Q_{(0)}: \ \sqrt{2}A &= \Delta^{1/2}/2 + ([\Delta^2/16 + L]^{1/2} + \Delta/4)^{1/2} \\ &= \Delta^{1/2} - \sqrt{2}D, \\ \sqrt{2}B &= \Delta^{1/2}/2 + ([\Delta^2/16 + L]^{1/2} - \Delta/4)^{1/2} \\ &= \Delta^{1/2} - \sqrt{2}C. \end{aligned} \tag{35b}$$

Particular solutions of (34) expressible in terms of elementary functions were obtained by Chatterjee, see Tables III and IV.

C. MacVittie-type solutions

If $L = 0$ one has a large subclass of charged ($\epsilon \neq 0$) and neutral ($\epsilon = 0$, $\mu \neq 0$) solutions which contain the famous pioneering solution obtained by MacVittie[31] in 1933. (See Tables V and VI.) This subclass will be called "McVittie-type solutions."

Neutral McVittie-type solutions must not be confused with the "McVittie metrics" examined by McVittie in a recent paper.[19] The latter are also a class of neutral shear-free solutions with $T(t)$ as the only arbitrary function of time, however, these solutions form a larger class than the "McVittie-type" solution presented here, as they include cases with $L \neq 0$ (Wyman-type solutions discussed below) and solutions not belonging to the class NKQ, that is, solutions with $j(y)$ different from (22a) (see Appendix D).

Particular cases of charged and neutral MacVittie-type solutions, usually describing bounded fluid spheres, have been examined by authors in the category (b). (See the Introduction.) In particular, Chatterjee and Chakravorty[30] derived, classified, and identified all charged McVittie-type solutions, see Table IV. Since Q has always multiple roots if $L = 0$, all McVittie-type solutions are expressible in terms of elementary functions. They will be classified in Sec. VI.

TABLE II. This table is the neutral analog of Table I, corresponding to solutions given by Eqs. (31) and (32). See Sec. V.

$Q_{(\ )}$	ξ^2	η^2	$\Psi(Z,\eta)$	Range of W	Parameters in (31)	Parameters in (32)
$Q_{(3)}$				type (i)	$\alpha_i = \beta_i = 1$,	$\gamma_1 = 0$,
		$\frac{B-C}{A-C}$	$\sin^{-1}\left(\frac{Z-A}{Z-B}\right)^{1/2}$	$W>A>0$	with: $i = 1,2,3$	$\gamma_2 = 1$,
				type (iv)	$\mu<0$.	$\alpha_1 = A$,
				$W>A \quad A<0$		$\alpha_2 = B$.
				type (iii)	$\alpha_3 = \beta_3 = 1$,	$\gamma_2 = 0$,
	$\frac{4}{A-C}$		$\sin^{-1}\left(\frac{Z-C}{B-C}\right)^{1/2}$	$C<W<B$	all other $\alpha_i, \beta_i = -1$,	$\gamma_1 = 1$,
					$\mu<0, \quad A>B>C>0$.	$\alpha_1 = C$,
						$\alpha_2 = C - B$.
			$\sin^{-1}\left(\frac{A-Z}{A-B}\right)^{1/2}$	type (ii)	$\alpha_3 = \beta_3 = -1$,	
				$B<W<A$	all other $\alpha_i, \beta_i, = 1$,	$\gamma_1 = 1, \quad \alpha_1 = A$,
		$\frac{A-B}{A-C}$			$\mu>0 \quad C<0$.	$\gamma_2 = 0, \quad \alpha_2 = A - B$
			$\sin^{-1}\left(\frac{C-Z}{B-Z}\right)^{1/2}$	type (ii)	$\alpha_i = \beta_i = -1$,	$\gamma_1 = 0, \quad \alpha_1 = C$,
				$0<W<C$	$i = 1,2,3, \quad \mu>0$.	$\gamma_2 = 1, \quad \alpha_2 = B$.
$Q_{(1)}$				type (i)	$\alpha_1 = \beta_1 = 1$,	$\not{\alpha} = -1$
	$1/\delta$	$\frac{\delta+B-A}{2\sigma}$	$\cos^{-1}\left(\frac{Z-A-\delta}{Z-A+\delta}\right)$	$W>A$	$\mu<0, \quad A>0$.	
				type (iv)	As type (i) above	
				$W>0$	with $A<0$.	
	$\delta^2 = (B-A)^2 + C^2$	$\frac{\delta-B+A}{2\sigma}$	$\cos^{-1}\left(\frac{\delta-A+Z}{\delta+A-Z}\right)$	type (ii)	$\alpha_1 = \beta_1 = -1$	$\not{\alpha} = 1$
				$0<W<A$	$\mu>0, \quad A>0$.	

TABLE III. The simplified form of H has been obtained by substituting in Eqs. (45) the explicit values of the parameters given in Table VIII. The classification scheme is explained in the text (see Sec. VI). Notice that each entry of this table represents three different solutions, one for each value of k in Eqs. (14) and (17) (see Appendix B).

Classification scheme	Simplified form of H	
ChMcV($r4$)($X4$) ChMcV($r4$)($X5$)	$\epsilon h[T \pm X]$	h and X as in (25a) h and X as in (25e)
ChMcV($r3$)($X4$) ChMcV($r3$)($X5$)	$\frac{\mu h}{2}\left((T \pm X)^2 - \frac{\epsilon^2}{\mu^2}\right)$	h and X as in (25d) h and X as in (25e)
ChWy($r3$)($X1$) ChWy($r3$)($X2$)	$\frac{12\epsilon h}{\mu}\frac{4\epsilon^2 - \mu^2(T \pm X)^2}{12\epsilon^2 + \mu^2(T \pm X)^2}$	h and X as in (25a) h and X as in (25b)
ChMcV($r2,r2$)($X1$) ChMcV($r2,r2$)($X2$)	$\frac{1}{\Delta v(y)}[\tilde{T}(t) \pm \sqrt{\Delta}\epsilon u(y)]$	$\tilde{T}(t) = \mu e^{\sqrt{\Delta}T}\quad v(y) = \frac{1}{\sqrt{a}}(ay + b + \sqrt{\Delta}),\quad u(y) = \left(\frac{ay + b + \sqrt{\Delta}}{ay + b - \sqrt{\Delta}}\right)^{1/2}$ $\tilde{T}(t) = \mu e^{bT},\quad v(y) = b^2,\quad u(y) = [2by + c]^{-1/2}$
ChWy($r2,r2$)($X1$) ChWy($r2,r2$)($X2$) ChWy($r2,r2$)($X3$)	$\frac{\sqrt{\lvert\Delta\rvert}h}{\sqrt{2}\epsilon}\frac{1 \pm \lambda\tilde{T}(t)u(y)}{\tilde{T}(t) \pm u(y)}$	$\tilde{T}(t) = \tan\frac{\sqrt{\Delta}T}{\sqrt{2}\epsilon},\quad u(y) = \tan\frac{\sqrt{\Delta}}{\sqrt{2}\epsilon}X_{(1)},\quad \lambda = 1$ as above with $\Delta = b^2$ and $X = X_{(2)}$ $\Delta < 0,\ \tilde{T}(t) = \tanh\frac{\sqrt{-\Delta}T}{\sqrt{2}\epsilon},\quad u(y) = \tanh\frac{\sqrt{-\Delta}}{\sqrt{2}\epsilon}X_{(3)},\ \lambda = -1$
ChMcV($r2$)($X1$) ChMcV($r2$)($X2$)	$\frac{1}{\Delta v(y)\tilde{T}(t)}\left[\left(\tilde{T}(t) \pm \frac{\mu}{2}u(y)\right)^2 - \frac{1}{4}\Delta\epsilon^2 u^2(y)\right]$	$\tilde{T}$, u, and v as in ChMcV($r2,r2$)($X1$) with $4a = \Delta\epsilon^2 - \mu^2$ u and v as in ChMcV($r2,r2$)($X2$), $\tilde{T} = n_0 + n(t),\quad bn(t) = LnT(t),\quad 4\exp(2bn_0) = b^2\epsilon^2 - \mu^2$
ChMcV($r2$)($X3$)	$\frac{1}{\lvert\Delta\rvert}[\mu h + (\mu^2 - \Delta\epsilon^2)\lvert\Delta\rvert^{1/2} \times(\sin\tilde{T}(t) \pm u(y)\cos\tilde{T}(t))]$	$\tilde{T}(t) = \sqrt{-\Delta}T,\quad u(y) = (-\Delta)^{-1/2}(ay + b)$
ChWy($r2$)($X1$) ChWy($r2$)($X2$)	$\frac{9\mu h}{2\Delta}\frac{(T \pm X)^2 - 4u_0}{(T \pm X)^2 - u_0}$	$u_0 = \frac{27\mu^2}{16\epsilon\Delta^2}$, h and X as in (25a) u_0 as above. h and X as in (25b)
ChWy($r2$)($X4$) ChWy($r2$)($X5$)	$\frac{2\epsilon^2 h}{3\mu}\frac{\tilde{T}^2(t) + 2\sqrt{2}u(y)\tilde{T}(t) - u^2(y)}{\tilde{T}^2(t) + 2u(y)\tilde{T}(t) - u^2(y)}$	$\tilde{T}(t) = \exp\left(\frac{3\mu}{\sqrt{2}\epsilon}T\right),\quad u(y) = \exp\left(\frac{3\mu}{\sqrt{2}\epsilon}X_{(4)}\right)$ as above with $X = X_{(5)}$
ChWy($r2$)I($X1$) ChWy($r2$)I($X2$)	$\frac{\sqrt{2}\epsilon h}{\sqrt{\Delta}}[1 + \sec[(\Delta/2)^{1/2}(T \pm X)]]^{-1}$	h and X as in (25a) h and X as in (25b)
ChWy($r2$)$I\delta_+$($X1$) ChWy($r2$)$I\delta_+$($X2$)	$\frac{h}{A}\frac{\pm\tilde{T}^2(t) \mp \alpha_1 u(y)\tilde{T}(t) \mp u^2(y)}{\pm\tilde{T}^2(t) \mp \alpha_2 u(y)\tilde{T}(t) \mp u^2(y)}$	$\tilde{T}$ and u as in ChMcV($r2,r2$)($X1$) with: $\epsilon^2\delta = \Delta > 0$, $\alpha_1 = 2(A + \nu)(\zeta - \nu^2)^{-1/2},\quad \alpha_2 = 2(\zeta + \nu A)(\zeta - \nu^2)^{-1/2}$ as above with $X = X_{(2)}$
ChWy($r2$)$I\delta_-$($X3$)	$\frac{h}{A}\frac{\sin\tilde{T}(t) + u(y)\cos\tilde{T}(t) + \alpha_1 h^{-1}}{\sin\tilde{T}(t) + u(y)\cos\tilde{T}(t) + \alpha_2 h^{-1}}$	$\tilde{T}$ and u as in ChMcV($r2$)($X3$) with: $\epsilon^2\delta = \Delta < 0$, $\alpha_1 = \frac{A + \nu}{\epsilon[a\delta(\zeta - \nu^2)]^{1/2}},\quad \alpha_2 = \frac{\delta + A(A + \nu)}{\epsilon A[a\delta(\zeta - \nu^2)]^{1/2}}$

D. Wyman-type solutions

If L is a constant different from zero, one has another large subclass of charged ($\epsilon \neq 0$) and neutral ($\mu \neq 0$) solutions containing the famous solution obtained by Wyman[9] in 1946. The time dependence of H for Wyman-type solutions is simpler than in the more general case with $dL/dt \neq 0$, since the only time-dependent function characterizing the solutions is T in Eqs. (28) and (32) [or H_0 in Eqs. (30)].

Wyman's solution is the particular Wyman-type solution corresponding to[4,5] $\epsilon = 0$, $T(t) = t$, $L = \text{const}$, $\Theta/3 = [6Lt + L_0]^{1/2}$ with L_0 constant, $\Delta = 0$ with X as in (25e) with $c = 1$ and y given by (16b) with $k = 0$. It can be given in the form of (24a), that is, as an elliptic integral:

$$t + \frac{r^2}{2} = \int \frac{dW}{[-2\mu W^3 + L]^{1/2}}, \tag{36}$$

which, save for constant factors and changes of notation, coincides with the usual expressions provided by authors dealing with this important solution (see Collins and Wainwright[5]). However, the Wyman solution can also be cast in the form (32b) by inverting (36), as it corresponds to $Q_{(1)}$ in Eqs. (31) with $A = (L/2\mu)^{1/3}$, $B = -A/2$, $C^2 = 3A^2/4$. Thus, with the help of Table II, the parameters appearing in (32b) and ranges of W in Fig. (1) can be computed explicitly leading to the following three forms of H:

TABLE IV. The authors in the left-hand column have been listed chronologically. Comparing the list of solutions examined by these authors with those presented in Tables III and VIII, it is evident that the following charged solutions expressible by elementary functions are new: ChWy($r3$)($X1$), ChWy($r3$)($X2$), ChWy($r2,r2$)($X1$), ChWy($r2,r2$)($X2$), ChWy($r2,r2$)($X3$), ChWy($r2$)($X1$), ChWy($r2$)($X2$), ChWy($r2$)($X4$), ChWy($r2$)$I\delta_+$, and ChWy($r2$)$I\delta_-$ with $X = X1$, $X2$, and $X3$. See Sec. VI.

Authors	Solutions examined
P. C. Vaidya and Y. Shah,[a] A. Banerjee, N. Chakravorty, and S. B. Duttachoudhury[b]	The class of solutions discovered by Vaidya and Shah include ChMcV($r4$)($X4,5$) and ChMcV($r2,r2$)($X1,2$) as particular cases. Other Valdya–Shah solutions do not correspond to the choice of $j(y)$ and $E(y)$ given by (22). See Appendix D.
P. C. Valdya and Y. Shah[c]	They derived the charged version of McVittie solution. This solution is ChMcV($r2$)($X2$) with $c = 0$.
M. C. Faulkes[d]	He found the most general ChKQ solution with $L =$ const and with $j(y)$ and $E(y)$ proportional to f^2 and f^3, that is those solutions with $X = X_{(5)}$. The following solutions are contained as particular cases: ChMcV($r4$)($X5$), ChMcV($r3$)($X5$), ChWy($r2$)I($X5$).
P. Vickers[e]	He found ChMcV($r2$)($X2$) which contains the charged version of McVittie's ($c = 0$) and Narial's solutions ($c \neq 0$). See Table VII.
A. Banerjee, N. Chakravorty, and S. B. Duttachoudhury[f]	They found the ChWy($r3$)($X1$) and ($X2$) solutions.
N. Chakravorty and S. Chatterjee[g]	They found and identified all charged McVittie-type solutions. This is equivalent to integrating (24a) with $L = 0$. They identified the charged version of Narial's solution as ChMcV($r2$)($X2$) with $c \neq 0$.
B. Mashhoon and H. Partovi[h]	They obtained the case $k = 0$ of ChMcV($r2$)($X1$) and ChMcV($r2$)($X2$) in a single expression. Their particular case "$\nu = 0$" (their notation) corresponds to $\mu^2 = \Delta\epsilon^2$ (my notation) and so to ChMcV($r2,r2$)($X1$) and ($X2$). They also derive the corresponding solutions with uniform density (see Table VIII).
B. Mashhoon and H. Partovi[i,j] D. C. Srivastava and S. Prasad[k]	They proved that the only charged shear-free solutions satisfying a barometric equation of state are the charged version of the Wyman solution (see Sec. IV) and a new solution belonging to the class discovered by P. C. Vaidya and Y. Shah quoted above. The latter was discovered by Mashhoon and Partovi. See Appendix D.
S. Chatterjee[l]	He integrated Eqs. (24a) with the restrictions given by (32) (see Sec. IV). He rediscovered ChMcV($r2,r2$)($X1$). ChMcV($r2,r2$)($X2$). And discovered ChWy($r2$)I($X1$) and ChWy($r2$)I($X2$) which follow as particular cases from his solution.

[a] See Ref. 30.
[b] See Ref. 31.
[c] See Ref. 32.
[d] See Ref. 15.
[e] See Ref. 24.
[f] See Ref. 34.
[g] See Ref. 27.
[h] See Ref. 16.
[i] See Ref. 5.
[j] See Ref. 6.
[k] See Ref. 12.
[l] See Ref. 22.

TABLE V. This table is the neutral analog of Table VIII, with "N" in the classification scheme denoting "neutral" instead of "Ch" for charged solutions. See Sec. VI.

$Q_{(r)}$	Parameters in (46)	Parameters in (21b)	Range of W (see Fig. 1)	Classification scheme
$Q_{(r3)}$		$\Delta = L = 0$	type (iv) $W > 0$	NMcV($r3$)($X4$) NMcV($r3$)($X5$)
$Q_{(r2)}$	$A = 0$, $\zeta = D = \Delta$	$L = 0$, $\Delta > 0$	type (i) $\mu > 0$, $W > 2\mu$ type (ii) $0 < W < 2\mu$ type (iv) $\mu < 0$, $W > 0$	NMcV($r2$)($X1$) NMcV($r2$)($X2$)
		$L = 0$, $\mu < 0$, $\Delta < 0$	type (i) $W > 2\mu$	NMcV($r2$)($X3$)
	$A = -\frac{\Delta}{3\mu}$, $\zeta = \mu A$	$\Delta > 0$, $\mu < 0$ $0 > L = \Delta^3/27\mu^2$	type (i) $W > \Delta/6\mu$	NWy($r2$)(X1) NWy($r2$)($X2$)
		$\Delta < 0$, $L > 0$ $\lvert L \rvert$ as above	type (ii) $0 < W < \lvert\Delta\rvert/3\mu$	NWy($r2$)($X3$)

TABLE VI. The simplified form of H has been obtained by inserting the values of the parameters of Table V into Eqs. (47). See Sec. VI. Two neutral solutions frequently mentioned in the literature are those discovered by McVittie[26] and Nariai,[33] these solutions are NMcV$(r2)(X2)$ with $c=0$ and $c\neq 0$, respectively.

Classification scheme	Simplified form of H		Previous occurrences in the literature
NMcV$(r3)(X4)$	$\lvert\mu\rvert^2 h[T\pm X]^2$	h and X as in (25d)	The case $k=0$ was discovered by Banerjee and Banerji.[a] Reviewed by Banerjee *et al.*,[b] Glass,[c] and McVittie.[d]
NMcV$(r3)(X5)$		h and X as in (25e)	The case $k=0$ was discovered by Faulkes.[e] Reviewed by the same authors as above.
NMcV$(r2)(X1)$	$\dfrac{1}{\Delta v(y)T(t)}\left(\tilde{T}(t)\pm\dfrac{\mu}{2}u(y)\right)^2$	$\tilde{T}(t)=e^{\sqrt{\Delta}T}$, $v(y)=\dfrac{1}{\sqrt{a}}(ay+b+\sqrt{\Delta})$, $u(y)=\left(\dfrac{ay+b+\sqrt{\Delta}}{ay+b-\sqrt{\Delta}}\right)^{1/2}$	The case $k=0$ was studied by Glass and Mashhoon.[f] Reviewed by Kramer *et al.*,[g] Banerjee *et al.*,[b] and McVittie.[d]
NMcV$(r2)(X2)$		$\tilde{T}(t)=e^{bT}$, $v(y)=b$, $u(y)=(2by+c)^{-1/2}$	Studied and reviewed by the same authors as above. Contains Nariai's solution $(c\neq 0)$[b] and if $c=0$ one has the McVittie solution.[i]
NMcV$(r2)(X3)$	$\dfrac{\lvert\mu\rvert h}{\lvert\Delta\rvert}\left(1+\dfrac{\sqrt{-\Delta}}{\sqrt{a}}h\,[\cos\tilde{T}(t)+u(y)\sin\tilde{T}(t)]\right)$	$\tilde{T}(t)=\sqrt{-\Delta}T$, $u(y)=\dfrac{1}{\sqrt{-\Delta}}(ay+b)$	Discovered by Banerjee *et al.*[b] and Wyman.[j] Particular case of solutions presented by McVittie.[d]
NWy$(r2)(X1)$	$\dfrac{6\lvert\mu\rvert h}{\Delta}\left(1+3\tan^2\left[\dfrac{\sqrt{\Delta}}{2}(T+X)\right]\right)^{-1}$	h and X as in (25a)	As above.
NWy$(r2)(X2)$		h and X as in (25b)	
NWy$(r2)(X3)$	$\dfrac{6\lvert\mu\rvert h}{\lvert\Delta\rvert}\left(1+3\tanh^2\left[\dfrac{\sqrt{-\Delta}}{2}(T+X)\right]\right)^{-1}$	h and X as in (25c)	As above.

[a] See Ref. 35.
[b] See Ref. 36.
[c] See Ref. 4.
[d] See. Ref. 19.
[e] See Ref. 37.
[f] See Ref. 38.
[g] See Ref. 2.
[h] See Ref. 33.
[i] See Ref. 26.
[j] See Ref. 17.

$\mu>0, \quad L>0, \quad$ type(ii)

$$H=A^{-1}\frac{1+\operatorname{cn} v}{(\sqrt{3}+1)\operatorname{cn} v-(\sqrt{3}-1)}, \tag{37a}$$

$\mu<0, \quad L<0, \quad$ type(i) $\quad H=A^{-1}[1-\operatorname{cn} v]$, (37b)

$\mu<0, \quad L>0, \quad$ type(iv) $\quad H=A^{-1}(1-\operatorname{cn} v)/\operatorname{cn} v$, (37c)

with

$$v\equiv t+r^2/2, \quad A=(L/2\mu)^{1/3} \tag{37d}$$

and cn v being a Jacobian elliptic function[25] whose modulus η, argument Ψ, and other parameters can be found from Table II. The importance of the Wyman solution lies in the fact that it is the only NKQ solution satisfying a barotropic equation of state, and thus its physical and geometric properties have been extensively discussed (See Collins[7] and Mashhoon–Partovi.[4]) This solution will be further examined in Part II, in connection with the study of the properties of NKQ and ChKQ solutions in general (i.e., not admitting a barotropic equation of state), and offering a parallel comparison with the work of Collins and Mashhoon–Partovi.

The charged version of Wyman's solution[4,10] follows also from (24a) under the same restrictions as the neutral case with $\epsilon\neq 0$. In this case, the form of Q in Eqs. (27) can be as $Q_{(2)}$ if $L<27\mu^4/16\epsilon^6$, or $Q_{(0)}$ if $L>27\mu^4/16\epsilon^6$, and it can be cast as in (28b), (28c), (30b), or (30c) depending on the values of L, μ, and ϵ. It can be of either one of the following types in Table I and Fig. 1: type (i) if Q has one positive real root, type (ii) if it has two positive roots, and type (iv) if it has two negative real root or no real roots. Particular Wyman-type solutions expressible in terms of elementary functions will be classified in Sec. VI.

E. Solutions with $dT/dt=0$ but $dL/dt\neq 0$

As far as I am aware, this class of solutions has not been considered in the literature. Since all the time dependence of the solutions is contained in the modulus of the elliptic functions, these solutions are as complicated to handle as the more general case with both dL/dt and dT/dt different from zero. In Eqs. (30), $dT/dt=0$ does not imply $dH_0/dt\neq 0$, so that the time dependent of these expressions is contained in both H_0 and L. If $\epsilon=0$ with $\mu\neq 0$, one has neutral solutions of this type. This subclass (whether charged or neutral) does not contain particular solutions expressible in terms of elementary functions.

F. Static solutions

If $dL/dt=dT/dt=0$, there is no time dependence in H, and one has a large class of static solutions. These solutions will be examined in Part II.

G. Solutions with uniform matter density

The necessary and sufficient conditions for matter density to be independent of r imply a restriction on the form of h in (23), this fact can be appreciated from Eqs. (A7) and (A8) in Appendix A. From (A8)

$$\rho'=0 \Rightarrow J'=E'=0 \Rightarrow fh=c_0=\text{const}. \tag{38}$$

From Eqs. (17) and (23), (38) implies

$$\Delta c_0^4=1, \tag{39a}$$

$$c=0, \quad a=-kc_0^{-2}, \quad b=c_0^{-2}. \tag{39b}$$

In order to verify that conditions (39) imply $\rho'=0$, it is necessary to evaluate ρ from Eq. (A7) by eliminating the term $[fR'/R]$ in terms of R and E using Eqs. (16a), (20a), (21a), and (21b). This term is

$$fR'/R=(fh)'/h-fhR\,[Q(W)]^{1/2}, \tag{40}$$

where

$$W=fh/R=h/H$$

and Q is given by (21b). Inserting Eqs. (39) and (40) in Eq. (A7) the latter equation becomes

$$\tfrac{8}{3}\pi\rho(t)=\Theta^2/9-c_0^2L(t)\,. \tag{41}$$

Uniform density solutions can be either McVittie- or Wyman-type solutions if $dL/dt=0$, or neither of these types as there is no contradiction between the conditions for uniform density and $dL/dt\neq 0$. For uniform density solutions, X takes the following form:

$$X_{(UD)}=c_0^{-2}\ln|y/(2-ky)|^{1/2}\,. \tag{42}$$

The form of H for uniform density solutions is that of (28), (30), or (32) under the restrictions on h and X given by (39b) and (42). The neutral case ($\epsilon=0$) of these homogeneous density solutions has been reviewed by Barnes[13] and Kramer *et al.*,[12] and it has been examined by Eisenstaedt[37] in a cosmological context.

Since Eq. (A8) holds in general for all charged spherically symmetric shear-free perfect fluid solutions, it follows from Eq. (7b) that uniform matter density implies $q=0$. Though, if $\epsilon\neq 0$ so that $E=\text{const}\neq 0$, one has a neutral fluid in the presence of a source-free electric field. For the ChKQ class of solutions given by (28) and (30), if the electric charge density vanishes the fluid must have uniform density. This is true because J and E are related by $J=\mu(E/\epsilon)^{5/3}$. However, for shear-free solutions for which restrictions (22) do not hold, one could have neutral fluids in the presence of an electric field with nonuniform density. Notice that the restrictions (39) do not imply $\mu=0$, although, for neutral solutions, $\mu=0$ does imply $\rho=\rho(t)$ because of Eq. (A8). The subclass of conformally flat, homogeneous density solutions corresponding to Eq. (24a) with $\epsilon=\mu=0$ will be treated in Sec. VII. Particular homogeneous density solutions expressible in terms of elementary functions will be classified in Table VII.

VI. SOLUTIONS EXPRESSIBLE IN TERMS OF ELEMENTARY FUNCTIONS

The solutions given by (28), (30), and (32) with $\mu\neq 0$ reduce to elementary functions if the modulus η of the elliptic functions is zero or unity. Since η, as given in Tables I and II, is a function of time only because it depends on $L(t)$, thus if $\eta=0,1$, L must be a constant. Therefore, if $\epsilon\neq 0$ in (28) and (30) and $\mu\neq 0$ in (32), nonstatic solutions expressible in terms of elementary functions cannot belong to the class presented in Sec. V E, and must be either McVittie-type solutions or particular cases of Wyman-type solutions. This fact can also be appreciated if one recalls that $\eta=0,1$ is

TABLE VII. The classification scheme appearing in the first column is analogous to that introduced for solutions with nonuniform density. See Sec. VI. The form of H for the "ChUD" and "NUD" solutions is that of the corresponding $X1$ solution (indicated in the middle column) with X given by Eq. (25a). Conformally flat solutions introduced in Sec. VII are denoted by "CF."

Classification scheme	Form of H (see Tables IV and VII)	Previous occurrences in the literature
ChWy($r3$)(UD)	As ChWy($r3$)	None
ChMcV($r2,r2$)(UD)	As ChMcV($r2,r2$)	B. Mashhoon and H. Partovi[a]
ChWy($r2,r2$)(UD)	As ChWy($r2,r2$)	None
ChMcV($r2$)(UD)	As ChMcV($r2$)	B. Mashhoon and H. Partovi[a]
ChWy($r2$)(UD)	As ChWy($r2$)	None
ChWy($r2$)I(UD)	As ChWy($r2$)I	None
ChWy($r2$)$I\delta_+$(UD)	As ChWy($r2$)$I\delta_+$	None
ChWy($r2$)$I\delta_-$(UD)	As ChWy ($r2$)$I\delta_-$	None
NMcV($r2$)(UD)	As NMcV($r2$)	H. Knutsen[b]
NWy($r2$)(UD)	As NWy($r2$)	None
CF(L,T)	As. Eq. (48)	H. Bondi,[c] I. H. Thompson and G. J. Whitrow,[d] H. Nariai and K. Tomita[c]
CF(T)	As above with $L = \text{const}$	W. B. Bonnor and M. C. Faulkes[f]
CF(FRW)	As above with $kc_0^4 L = -\exp[2T/c_0^2]$	The FRW-type solutons[g]

[a] See Ref. 16.
[b] See Ref. 39.
[c] See Ref. 40.
[d] See. Ref. 41.
[e] See Ref. 42.
[f] See Ref. 43.
[g] See Ref. 44.

equivalent to Q having multiple roots. From the form of Q given in Eq. (21b), the necessary and sufficient condition for $Q=0$ to have multiple roots is the vanishing of the discriminant[29]:

$$D = L\,[\epsilon^2 I_1^{\,2} + \Delta I_1 I_2 - 3L I_2^{\,2}] = 0\,, \tag{43}$$

with

$$I_1 \equiv L\epsilon^2 + \Delta^2/12, \quad I_2 \equiv 2\Delta\epsilon^2/3 - 3\mu^2/4\,.$$

From Eq. (43) it is clear that Q has trivial multiple roots at $W=0$ corresponding to $L=0$, that is, to McVittie-type solutions. If $L\neq 0$, finding multiple roots of Q is not so simple, and this fact explains why Wyman-type solutions have been barely discussed in the literature (see Tables III–VIII). In this case, the condition for the occurrence of a multiple root at $W\neq 0$ can be expressed as a quadratic equation on L from the term in brackets in Eq. (43), therefore L will be a constant different from zero in this case. If $\epsilon=0$ but $\mu\neq 0$ in (43), the same result holds: multiple roots in the cubic Q are incompatible with $dL/dt\neq 0$. However, in neutral solutions, if $\mu=0$ and Δ and H satisfy restrictions (39), then it is possible to have solutions expressible in terms of elementary functions and keeping both T and L as functions of time. These uniform density solutions are also conformally flat and will be considered in the next section.

If the quartic Q given by (21b) has multiple roots, it can be expressed as one of the following forms according to the number of repeated roots (r):

$$Q_{(r4)} = \epsilon^2 W^4\,, \tag{44a}$$

$$Q_{(r3)} = \epsilon^2[W-A\,]^3[W-B\,]\,, \tag{44b}$$

$$Q_{(r2,r2)} = \epsilon^2[W-A\,]^2[W-B\,]^2\,, \tag{44c}$$

$$Q_{(r2)} = \epsilon^2[W-A\,]^2[W^2+2\nu W+\zeta\,]\,. \tag{44d}$$

For each form given above, Eq. (24a) can be integrated and inverted yielding the following expressions analogous to (28):

$$Q_{(r4)}\colon\ H = \epsilon h\,|T\pm X\,|\,, \tag{45a}$$

$$Q_{(r3)}\colon\ H = h\,\frac{\epsilon^2(A-B)^2(T+X)^2-4}{|A\epsilon^2(A-B)^2(T+X)^2-4B\,|}\,, \tag{45b}$$

$$Q_{(r2,r2)}\colon\ H = h\,\frac{1-\exp(\pm\epsilon(A-B)(T+X))}{B-A\exp(\pm(A-B)(T+X))}\,, \tag{45c}$$

$Q_{(r2)}$:

$$H = h\,\frac{|\zeta-\nu^2|^{1/2}S(2\epsilon|\delta|^{1/2}(T+X))-A+\nu}{A\,|\zeta-\nu^2|^{1/2}S(2\epsilon|\delta|^{1/2}(T+X))+\nu A+\zeta}\,, \tag{45d}$$

$$H = h\,\frac{\epsilon^2(A+\nu)^2(T+X)^2-1}{A\,[\epsilon^2(A+\nu)^2(T+X)^2+1]+2\nu}\,,\quad \delta=0\,, \tag{45e}$$

where

$$\delta \equiv A^2+2\nu A+\zeta\,,$$

$$S\equiv\sin \quad \text{if } \delta<0,\quad \zeta-\nu^2<0\,,$$

$$S\equiv\sinh \quad \text{if } \delta>0,\quad \zeta-\nu^2>0\,.$$

Different particular solutions arise for different forms of h and X given by Eqs. (25). These solutions are classified in Table VIII, which provides the values of the repeated roots A and B, and the parameters ν and ζ appearing in Eqs. (45). A classification scheme has been introduced in Table VIII, so that charged Wyman-type and charged McVittie-type solutions are labeled as "ChWy(rN_1)$I\delta_{\pm}$(XN_2)" and "ChMcV(rN_1)(XN_2)," where the number N_1 denotes N_1 repeated roots in Eqs. (44), the number N_2 corresponds to the number labeling X in Eqs. (25), the symbol I in some Wyman-type solutions indicates that I_1 and I_2 in Eqs. (43) are different from zero, while "$\delta_{\pm}$" indicates that the quantity defined as δ (positive or negative) in Eqs. (45d) and (45e) can take its most general value. In order to compare these solutions with solutions previously discussed in the literature, a simplified form of H is provided in Table III. This form of H is obtained in each case by inserting the corresponding values of the roots A and B and the parameters $\epsilon, \mu, \Delta, L, \nu, \zeta$, and δ from Table VIII into Eqs. (45). Forms of H analogous to those of Eqs. (30) can also be obtained by choosing $N(t)$ as in Eq. (29).

Previous occurrences of particular shear-free charged solutions in the literature are listed in Table IV roughly following a chronological order. As most authors use "isotropic" coordinates, in order to facilitate the comparison of the solutions studied by these authors with those presented in Tables III and VIII, the transformation relating these coordinates with the ones used in this paper is given in Appendix B. Since the metric coefficient g_{tt} in (11) can be obtained from H, it will be enough to compare the forms of H in each case with those presented in Table VIII. All authors listed in Table IV have made a choice of time coordinate by demanding $\Theta(t)$ to be proportional to dT/dt, other possible choices of time coordinate will be discussed in Part II. From Tables III, IV, and VIII, one can appreciate the fact that most particular solutions previously found in the literature are McVittie-type solutions in which Q has multiple roots at $W=0$. Most Wyman-type solutions classified in Tables II and III are new, and some of these new solutions, like

ChWy($r2,r2$)($X1$), ChWy($r2,r2$)($X2$),

ChWy($r2,r2$)($X3$), ChWy($r2$)($X1$),

and

ChWy($r2$)($X2$),

have mathematically complicated forms and it is unlikely that they will be studied in detail. However, the other new solutions,

ChMcV($r2$)($X3$), ChWy($r2$)I($X4$),

ChWy($r2$)$I\delta_+$($X1$), ChWy($r2$)$I\delta_+$($X2$),

and

ChWy($r2$)$I\delta_-$($X3$),

are more appealing and the examination of their physical and geometric properties might be worthwhile.

For neutral solutions with $\mu\neq 0$, the cubic Q having multiple roots can be cast in any of the following forms analogous to Eqs. (44):

$$Q_{(r3)} = |2\mu|W^3\,, \tag{46a}$$

$$Q_{(r2)} = [W-A\,]^2[\,\zeta-2\mu W\,]\,, \tag{46b}$$

TABLE VIII. This table provides the values of all unspecified parameters appearing in Eqs. (44) and (45). The classification scheme displayed in the last column is explained in the text. See Sec. VI.

$Q_{(r)}$	Parameters in (44)	Parameters in (21b)	Range of W (see Fig. 1)	Classification scheme
$Q_{(r4)}$		$\mu=\Delta=L=0$	type (iv) $W>0$	ChMcV($r4$)($X4$) ChMcV($r4$)($X5$)
$Q_{(r3)}$	$A=0,\quad B=\frac{2\mu}{\epsilon^2}.$	$\mu\neq 0,\quad \Delta=L=0.$	type (i) $\mu>0,\quad W>2\mu/\epsilon^2,$ type (ii) $\mu<0,\quad 0<W<2\mu/\epsilon^2.$	ChMcV($r3$)($X4$) ChMcV($r3$)($X5$)
	$A=\frac{3\mu}{4\epsilon^2},\quad B=-\frac{\mu}{4\epsilon^2}.$	$L<0,\quad \Delta=2\sqrt{3}\lvert L\rvert^{1/2}=\frac{9\mu}{8\epsilon^2}>0.$	type (i) $W>3\mu/4\epsilon^2$	ChWy($r3$)($X1$) ChWy($r3$)($X2$)
$Q_{(r2,r2)}$	$A=0,\quad B=\frac{\mu}{\epsilon^2}.$	$L=0,\quad \Delta=\frac{\mu^2}{\epsilon^2}>0.$	type (i) $\mu>0,\quad W>\mu/\epsilon^2,$ type (ii) $0<W<\mu/\epsilon^2,$ type (iv) $\mu<0,\quad W>0.$	ChMcV($r2,r2$)($X1$) ChMcV($r2,r2$)($X1$)
	$A=\frac{\sqrt{\Delta}}{\sqrt{2}\epsilon},\quad B=-A.$	$0<L=\frac{\Delta^2}{4\epsilon^2},\quad \mu=0,\quad \Delta>0.$	type (i) $W>\sqrt{\Delta}/\sqrt{2}\epsilon,$ type (ii) $0<W<\sqrt{\Delta}/\sqrt{2}\epsilon.$	ChWy($r2,r2$)($X1$) ChWy($r2,r2$)($X2$)
	$A=\frac{i\sqrt{\Delta}}{\sqrt{2}\epsilon},\quad B=-A.$	As above with: $\Delta<0.$	type (iv) $W>0$	ChWy($r2,r2$)($X3$)
$Q_{(r2)}$	$A=0,\quad \nu=-\frac{\mu}{\epsilon^2},\quad \zeta=\frac{\Delta}{\epsilon^2}.$	$L=0,\quad \mu^2-\Delta\epsilon^2<0,\quad \Delta>0.$	type (i) $\mu>0,\quad W>W_+,$ type (ii) $0<W<W_-,$ $W_\pm=\mu\epsilon^{-2}[1\pm\sqrt{\mu^2-\Delta\epsilon^2}],$ type (iv) $\mu<0,\quad W>0.$	ChMcV($r2$)($X1$) ChMcV($r2$)($X2$)
		$L=0,\quad \mu^2-\Delta\epsilon^2>0,\quad \Delta<0.$	type (iv) $W>0$	ChMcV($r2$)($X3$)
	$A\neq 0,\quad 1_1=1_2=0,$ $A=\frac{2\Delta}{3\mu},\quad \nu=-\frac{9\Delta}{2\mu},\quad \zeta=-\frac{4\Delta^2}{27\mu^2}.$	$0>L=\frac{\Delta^2}{12};\quad \Delta=\frac{9\mu^2}{8\epsilon^2}>0,$ $\delta=0.$	type (i) $\mu>0,\quad W>9\Delta W_1/2\mu,$ $W_1=1+\sqrt{1-16/2187},$ type (iv) $\mu<0,\quad W>0.$	ChWy($r2$)($X1$) ChWy($r2$)($X2$)
	$A\neq 0,\quad 1_1,1_2\neq 0,$ $A=\frac{1}{3\mu}\left[\Delta-2\epsilon^2\frac{12L\epsilon^2+\Delta^2}{8\Delta\epsilon^2-9\mu^2}\right],$ $\nu=A-\frac{\mu}{\epsilon^2},$ $\zeta=3A^2-\frac{4\mu}{\epsilon^2}A+\frac{\Delta}{\epsilon^2},$ $\delta=A^2+2\nu A+\zeta.$	$\Delta=0,\quad 0<L=\frac{27\mu^4}{16\epsilon^6},$ $A=\frac{3\mu}{2\epsilon^2},\quad \nu=\frac{A}{3},\quad \zeta=\frac{A^2}{3},\quad \delta>0.$	type (i) $\mu>0,\quad W>3\mu/2\epsilon^2,$ type (ii) $0<W<3\mu/2\epsilon^2,$ type (iv) $\mu<0,\quad W>0,$ $\delta=9\mu^2/2\epsilon^4.$	ChWy($r2$)I($X4$) ChWy($r2$)I($X5$)
		$\Delta=\frac{\mu^2}{\epsilon^2}>0,\quad 0<L=\frac{\Delta^2}{16},$ $A=\frac{\sqrt{\Delta}}{2\epsilon},\quad \nu=-A,\quad \zeta=4A^2,\quad \delta>0.$	type (i) $W>\sqrt{\Delta}/2\epsilon,$ type (ii) $0<W<\sqrt{\Delta}/2\epsilon,$ $\delta=3\Delta/4\epsilon^2.$	ChWy($r2$)I($X1$) ChWy($r2$)I($X2$)
		$\delta>0$	type (i) $A>0.\quad W-<0<W<W_+.\quad W>W_+,$ $W_-<W_+<0,\quad W>A,$ $A<0.\quad W_+>W_->0,\quad W>W_+,$ type (ii) $A>0,\quad W_-<W_+<0,\quad 0<W<A,$ $A<0,\quad W_+>W_->0,\quad 0<W<W_-,$ type (iii) $0<W_-<W_+<A,\quad W_-<W<W_+,$ $W_\pm=-\nu\pm\sqrt{\nu^2-\zeta}.$	ChWy($r2$)$I\delta_+$($X1$) ChWy($r2$)$I\delta_+$($X2$) ChWy(r2)$I\delta_+$($X3$)
		$\delta<0$	type (i) $A>0,\quad W>A,$ type (ii) $0<W<A,$ type (iv) $A<0,\quad W>0.$	As above with δ_-

where, as for the charged case, Eq. (24a) can be integrated and inverted yielding

$$Q_{(r3)}:\ H=(\mu h/2)[T\pm X]^2\,, \tag{47a}$$

$$Q_{(r2)}:\ H=h\frac{2\mu C^2(|\zeta-2\mu A|^{1/2}[T+X])}{\lambda\zeta-\lambda AS^2(|\zeta-2\mu A|^{1/2}[T+X])}\,, \tag{47b}$$

where

$$S\equiv\sin,\quad C\equiv\cos,\quad \lambda=-1,\quad \text{if}\quad 2\mu A<\zeta\,,$$
$$S\equiv\sinh,\quad C\equiv\cosh,\quad \lambda=1,\quad \text{if}\quad 2\mu A>\zeta\,,$$

and, as with the case of charged solutions, different particular solutions correspond to different choices of h and X in Eqs. (25). Particular solutions of (32) are classified and identified in Tables V and VI, which are analogous to Tables III, IV, and VIII. All these solutions have already been discovered and classified by Wyman,[17] McVittie,[19] and (some of them) in Kramer *et al.*[12] Unlike the charged case, all Wyman-type neutral solutions expressible by elementary functions have mathematically unappealing forms and it is probably not worthwhile studying them in detail. A classification scheme analogous to that used in Tables III and VIII has been introduced in Tables V and VI.

Some of the expressions presented in (45) and (47) contain uniform density solutions expressible in terms of elementary functions, this happens when h and X satisfy (39) and (42). The form of H for these solutions can be obtained by specializing h and X in (45) and (47) into (39) and (42). All types of uniform density solutions expressible in terms of elementary functions are classified and identified in Table III. A classification scheme analogous to those used in Tables III–VI and VIII has been introduced for Table VII, so that solutions with $\epsilon\neq0$ are denoted as "Ch...(UD)", those with $\epsilon=0$ but $\mu\neq0$ as "N...(UD)".

All solutions expressible as elementary functions reduce to static solutions if $T(t)$ becomes a constant. However, these static limits will be discussed in Part II.

VII. CONFORMALLY FLAT SOLUTIONS

If $\epsilon=\mu=0$, the functions j and E in (22a) and (22b) vanish and so does the Weyl tensor, as is shown in Appendix A [see Eqs. (A4), (A5), and (A6)]. Thus solutions with these two parameters set to zero are the only conformally flat, spherically symmetric, nonstatic, shear-free solutions. These solutions have also homogeneous density and their metric coefficient H is obtained from integrating (24a) with Q restricted by $\epsilon=\mu=0$. However, conditions (38) and (39) must be satisfied so that $\rho=\rho(t)$ is given by (41) and Eq. (18) with $\epsilon=\mu=0$ holds. These solutions are the only particular solutions of the ChKQ class expressible in terms of elementary functions which have two arbitrary functions of time. The form of H for these solutions will be

$$H=2e^{T/c_0^2}/[ye^{2T/c_0^2}-(2-ky)c_0{}^4L\,]\,. \tag{48}$$

Conformally flat solutions are also classified in Table VII where they are denoted as "CF." The Friedman–Robertson–Walker (FRW) family of solutions belongs to this class (see Table VII).

ACKNOWLEDGMENTS

I would like to thank Dr. M. A. H. MacCallum for encouragement and illuminating discussion. I would also like to thank Dr. B. Mashhoon for his hospitality and useful suggestions while I visited him at Koln.

I am indebted to the National University of Mexico (UNAM) for giving me financial support.

APPENDIX A: FIELD EQUATIONS AND CERTAIN GEOMETRIC QUANTITIES ASSOCIATED WITH THE METRIC TENSOR GIVEN BY EQ. (11)

Having already used the equation $G^t{}_r=0$ (Ref. 2) for the metric (1) and stress-energy tensor (8) in order to obtain Eq. (9), the remaining independent field equations for (8) and (11) are

$$G^t{}_t=8\pi T^t{}_t\,,$$
$$-8\pi\rho-\frac{E^2}{(fH)^4}=\frac{\Theta^2}{3}+\frac{1}{H^2}\left[\left(\frac{H'}{H}\right)^2-\frac{4f'H'}{fH}-\frac{2H''}{H}-\frac{2ff''+f'^2-1}{f^2}\right], \tag{A1}$$

$$G^r{}_r=8\pi T^r{}_r\,,$$
$$8\pi p-\frac{E^2}{(fH)^4}=(\dot H/H)^{-1}\frac{\partial}{\partial t}(\Theta^2/9)+\frac{\Theta^2}{3}+\frac{1}{H^2}\left[-\frac{2\dot H'}{\dot H}\left(\frac{H'}{H}+\frac{f'}{f}\right)+\frac{(H')^2}{H^2}+\frac{1-(f')^2}{f^2}\right], \tag{A2}$$

$$G^\theta{}_\theta=G^\phi{}_\phi=8\pi T^\theta{}_\theta=8\pi T^\phi{}_\phi\,,$$
$$8\pi p+\frac{E^2}{(fH)^4}=\left(\frac{\dot H}{H}\right)^{-1}\frac{\partial}{\partial t}\left(\frac{\Theta^2}{9}\right)+\frac{\Theta^2}{3}+\frac{1}{H^2}\left[\frac{\dot H'}{\dot H}\left(\frac{2H'}{H}-\frac{f'}{f}\right)-\frac{(H')^2}{H^2}-\frac{f''}{f}-\frac{\dot H''}{H}\right]. \tag{A3}$$

The conformal scalar invariant $\Psi_{(2)}$ is found to satisfy the simple relation

$$R^3\Psi_{(2)}=-J+E^2/R\,, \tag{A4}$$

$$J(y)\equiv jf^3\,, \tag{A5}$$

where $R=fH$ by Eq. (6), and $j(y)$ is the same function appearing in Eq. (18). This simple relation was discovered by Glass[2] for the neutral case ($E=0$) of Eq. (18). For the charged case, see Mashhoon and Partovi.[16] The rest of the null tetrad components of the Weyl tensor are zero, which confirms the known fact[13,14] that these solutions have Petrov types D or (if $\Psi_{(2)}=0$) type O.

The nonzero components of the Weyl tensor in the coordinate system (t,r,θ,ϕ) can be given in terms of the invariant $\Psi_{(2)}$ as

$$C_{trtr}=C_{\theta\phi\theta\phi}=2\Psi_{(2)}\,,$$
$$C_{r\theta r\theta}=C_{r\phi r\phi}=-C_{t\theta t\theta}=-C_{t\phi t\phi}=\Psi_{(2)}\,, \tag{A6}$$

so that the necessary condition for solutions characterized by the metric (11) to be conformally flat (i.e., the vanishing

of the Weyl tensor) is given by $\Psi_{(2)}=0$. There is a conformally flat subclass[2] of neutral nonstatic solutions contained in the ChkQ class. These conformally flat solutions are presented in Sec. VII, they will be labeled as "CF" in the classification scheme [see Eq. (46) and Table VII]. From (A4),(A5), conformally flat charged ChKQ solutions must be static.

Using Eq. (15), the field equations given by (A1), (A2), and (A3) can be simplified by eliminating second derivatives of the form H''. For certain calculations, it is useful to transform (11) into a metric where y and Y, as defined by Eqs. (16), replace r and H, respectively. The advantage of using this representation is that the elimination of second derivatives like $\partial^2 Y/\partial y^2$ is easier using Eq. (18) than Eq. (15). In particular, Eq. (A1) can be brought to the form

$$\frac{8}{3}\pi\rho=\frac{\Theta^2}{9}+\frac{1}{R^2}\left[1-\frac{2J}{R}+\frac{E^2}{R^2}-\left(\frac{fR'}{R}\right)^2\right], \tag{A7}$$

where J is the same function appearing in (A5). Differentiating Eq. (A7) with respect to r, and again eliminating derivatives like R'' with the help of Eqs. (15) or (18), one obtains

$$\tfrac{8}{3}\pi\rho' R^3=-2J'+(E^2)'/R\,. \tag{A8}$$

Equations (A7) and (A8) will be needed in Sec. V as solutions with uniform density are discussed. All quantities calculated in this appendix have been obtained using the algebraic computing language SHEEP.[45]

APPENDIX B: NONEQUIVALENCE OF SOLUTIONS WITH DIFFERENT VALUES OF k IN EQS. (14) AND (17)

The form of the metric (11) with $f(r)$ given by Eq. (14) suggests a comparison with the FRW solutions in which H is a function of time only (see Table VIII). In this particular case, the constant $k=0,\pm 1$ indicates the sign of the (constant) curvature of surfaces of constant proper (cosmic) time. For any solution (other than FRW) presented in the tables and characterized by $H(t,y)$ satisfying Eq. (15) [or (18)], it is also desirable to find out whether the three different values of k $(0,\pm 1)$ denote three different solutions. It is shown in this appendix that this is actually the case, so that each entry appearing in the tables is really a triplet of solutions. The demonstration will be restricted to neutral solutions excluding solutions with uniform density. The generalization to charged solutions is straightforward, and the case of uniform density solutions (including the conformally flat subclass) will be examined in Part II. The interpretation of the different choices of k in terms of geometric properties of the solutions is not as simple as with the FRW solutions, and thus will also be left for Part II.

The problem of testing whether two metrics are equivalent, in the sense that they correspond to the same space-time manifold, is a nontrivial problem known in the literature as the "equivalence problem." Broadly speaking one can say that two metrics g and $\bar{g}$ are equivalent if there is a nonsingular coordinate transformation $x^a=x^a(\bar{x}^b)$ relating them. In practice it is rather difficult to guess if such a coordinate transformation exists, and testing the equivalence of metrics in general requires elaborate mathematical techniques (see Karlhede[46]). Though there are no standard prescriptions, one could proceed by assuming that the desired coordinate transformation exists, and then testing the consistency (or inconsistency) of a system of simultaneous equations of the form

$$\begin{aligned} I_1(x^a)&=\bar{I}_1(\bar{x}^b)\,,\\ I_2(x^a)&=\bar{I}_2(\bar{x}^b)\,,\\ &\vdots\\ I_n(x^a)&=I_n(\bar{x}^b)\,, \end{aligned} \tag{B1}$$

where I_i are a set of suitable invariants. In general, one would take the I_i to be the components of the Riemann tensor and its covariant derivatives in a canonical null tetrad. However, in some cases it might be possible to find other mathematically simpler invariants.

Fortunately, for the solutions under consideration in this paper the equivalence of metrics with different k in Eqs. (14) and (17) is relatively easy to test. Because of spherical symmetry and the mathematical simplification of the metric coefficients, one can find simple invariant quantities which can be defined without reference to any coordinate system. With these quantities one can construct a system of equations similar to (B1), and thus reduce the test of equivalence of metrics with different k to a test of consistency of such a system. Consider the following quantities as the above mentioned invariants.

(i) Proper surface of two-spheres generated by the world lines of comoving observers,

$$A(t,y)=4\pi R^2=4\pi(fH)^2\,.$$

(ii) Change of $R=(A/4\pi)^{1/2}$ with respect to the proper time of comoving observers,

$$\frac{dR}{d\tau}=(-g^{rr})^{1/2}\frac{\partial R}{\partial t}=\frac{1}{3}\theta R\,.$$

(iii) The conformal scalar invariant $\Psi^{(2)}$ given by Eq. (A5). The functions h and f are given by Eqs. (17) and (23),

$$\Psi_{(2)}=-\mu(fh)^5/R^3\,.$$

(iv) Matter-energy density ρ, the eigenvalue of the timelike eigenvector of the momentum-energy tensor ρ given by (A7) and (38).

Consider two metrics corresponding to the same slot in the classification scheme (i.e., H is the same function of y, the constants μ, a, b, c are equal) but with y corresponding to two different values of k in Eqs. (17). Call (t,y) and $(\bar{t},\bar{y})$ the coordinates characterizing each metric. Let us assume that these metrics are equivalent, that is, there exists a nonsingular coordinate transformation of the type $t=t(\bar{t},\bar{y})$ and $y=y(\bar{t},\bar{y})$ such that the following system of simultaneous equations holds:

$$A(t,y)=\bar{A}(\bar{t},\bar{y})\Rightarrow R(t,y)=\bar{R}(\bar{t},\bar{y})\,, \tag{B2}$$

$$\frac{dR}{d\tau}=\frac{d\bar{R}}{d\bar{\tau}}\Rightarrow\Theta(t)R(t,y)=\bar{\Theta}(\bar{t})\bar{R}(\bar{t},\bar{y})\,, \tag{B3}$$

$$\Psi_{(2)}=\bar{\Psi}_{(2)}\Rightarrow\frac{[f(y)h(y)]^5}{R^3(t,y)}=\frac{[\bar{f}(\bar{y})\bar{h}(\bar{y})]^5}{\bar{R}^3(\bar{t},\bar{y})}\,, \tag{B4}$$

$$\tfrac{8}{3}\pi\rho(\Theta,R,f,h)=\tfrac{8}{3}\pi\bar{\rho}(\bar{\Theta},\bar{R},\bar{f},\bar{h})\,. \tag{B5}$$

The functions $f(y)$ and $h(y)$ in (B4) are given by (17) and (23), $\bar{f}(\bar{y})$ and $\bar{h}(\bar{y})$ are

$$\bar{h}(\bar{y}) = [a\bar{y}^2 + 2b\bar{y} + c]^{-1/2}, \tag{B6a}$$

$$\bar{f}(\bar{y}) = [\bar{y}(2 - \bar{k}\bar{y})]^{1/2}, \tag{B6b}$$

where the bar on top of k indicates that this constant is to take any of the other two possible values it can have different from the value of k without the bar. That is if $k = 0$, $\bar{k}$ can be 1 or -1, if $k = 1$, $\bar{k} = 0, -1$, etc.

If the system (B2) to (B5) is consistent, metrics with different k in (14) and (17) are equivalent. If (B2), (B3), and (B4) hold simultaneously, then one has

$$R = \bar{R}, \quad \Theta = \bar{\Theta}, \quad fh = \bar{f}\bar{h}. \tag{B7}$$

From the third equation in (B7), Eqs. (17), (23), and (B6), it follows that

$$\frac{2y - ky^2}{ay^2 + 2by + c} = \frac{2\bar{y} - \bar{k}\bar{y}^2}{a\bar{y}^2 + 2b\bar{y} + c}, \tag{B8}$$

which relates y with $\bar{y}$. Inserting (B7) in (B5) with ρ and $\bar{\rho}$ written explicitly with the help of (38) and (A7), one has

$$a(y - \bar{y})(fh)^2 = ky - \bar{k}\bar{y}, \tag{B9}$$

which can be shown to be in contradiction with (B8) unless $k = \bar{k}$ and $y = \bar{y}$. Therefore the metrics corresponding to different values of k are not equivalent.

This demonstration can be extended to include charged solutions by adding to the list of invariants the Maxwell field invariant $F_{ab}F^{ab}$ and proceeding in exactly the same manner. This equivalence of metrics for uniform density solutions will be considered in Part II.

APPENDIX C: ISOTROPIC COORDINATES

Most authors dealing with particular shear-free solutions use spatial comoving coordinates with the radial coordinate x defined in such a way that the metric of surfaces of constant t is given by

$$dl^2 = H^2(x,t)[dx^2 + x^2\, d\Omega^2]. \tag{C1}$$

These coordinates are called "isotropic" and relate to the coordinates introduced in Eqs. (14) through the following transformation[19,26]:

$$x(r) = \begin{cases} r, & k = 0 \\ 2\tan(r/2), & k = 1, \\ 2\tanh(r/2), & k = -1. \end{cases} \tag{C2}$$

In order to compare the forms of H given in Tables III and VI with those obtained by authors using isotropic coordinates, the following relations are helpful:

$$H(r,t) = (1 + kx^2/4)H(x,t), \tag{C3}$$

$$y(x) = x^2/2(1 + kx^2/4)^{-1}, \tag{C4}$$

$$f(x) = x(1 + kx^2/4)^{-1}. \tag{C5}$$

APPENDIX D: SOME SIMPLE SOLUTIONS NOT BELONGING TO THE ChKQ CLASS

The ChKQ solutions presented in Sec. IV follow from integrating Eq. (18) under the restrictions given by Eqs. (22). The simplest charged solutions which do not satisfy these restrictions are those discovered by Vaidya and Shah[30] and further examined by Banerjee, Chakravorty, and Duttachoudhury[31] (see Table IV). For these solutions, H has the form

$$H = [1/(\alpha y + \beta)][T(t) \pm u(y)], \tag{D1}$$

where α and β are arbitrary constants and $T(t)$ is an arbitrary function. The function $u(y)$ is related to $j(y)$ and $E(y)$ by Eq. (18). Inserting (D1) in Eq. (18), remembering that $Y \equiv H^{-1}$, one obtains

$$\frac{3j(y)}{f^2} = 2\alpha\frac{du}{dy} - (\alpha y + \beta)\frac{d^2u}{dy^2}, \tag{D2}$$

$$\frac{2E^2(y)}{f^6} = (\alpha y + \beta)\left(\frac{du}{dy}\right)^2. \tag{D3}$$

Some of the solutions classified in Tables III and VIII are particular cases of the Vaidya–Shah family of solutions. If α, β, and $u(y)$ are chosen so that (D1) becomes equal to H for solutions ChMcV($r4$)($X4$), ChMcV($r4$)($X5$), ChMcV($r2,r2$)($X1$), and ChMcV($r2,r2$)($X2$) of Table III, then $j(y)$ and $E(y)$ defined by Eqs. (D2) and (D3) will have the forms specified by Eqs. (22). However, H given by (D1) satisfies Eq. (18) for any function $u(y)$ related to $j(y)$ and $E(y)$ by (D2) and (D3), therefore one can choose in general functions $u(y)$ for which $j(y)$ and $E(y)$ will be different from (22). An example of a charged solution belonging to the Vaidya–Shah class but not contained in the class presented in Sec. IV is the solution discovered by Mashhoon and Partovi.[5] This solution [their Eq. (115)] satisfies a barotropic equation of state, it has the form (D1) with $\alpha = 0, \beta = \text{const}$ and, with obvious changes of notation, the function $u(y)$ satisfies

$$\frac{du}{dr} = -\frac{3}{4}\frac{\sinh 2r - 2(r + \rho_0)}{\sinh^2 r}, \tag{D4}$$

where ρ_0 is a constant and r is related to y by Eqs. (14c) and (16b).

It is quite likely that neutral solutions that do not satisfy the restriction (22a) will have charged versions, though it is also likely that charged shear-free solutions without neutral counterparts exist. Obtaining these types of solutions can be a topic of further research.

For neutral solutions, Wyman,[18] McVittie,[19] Stephani,[20] and Srivastava[21] have discussed the conditions to be imposed on $j(y)$ in order to be able to integrate Eq. (18) with $E = 0$. Solutions not belonging to the NKQ class presented here can be found in their papers and in references quoted therein. In particular, for some of these solutions the variable coefficient in (18) takes the following odd form:

$$3j/f^2 = [\nu y + \zeta]^{-n}, \quad n = \tfrac{15}{7}, \tfrac{20}{7}, \quad \nu, \zeta \text{ constants}. \tag{D5}$$

Stabell and Knutsen[47] and Knutsen[48,49] seem to have examined some solutions corresponding to (D5). The latter author has recently studied[50] simple solutions which have the form

$$H = u^{1/3n}[(1/n)T(t) + \alpha u^{1/3n}]^2, \tag{D6}$$

where α and n are arbitrary constants. This solution reduces to NMcV($r3$)($X4$) (see Table VII) if $n = 1$. For $n \neq 1$, it seems to correspond to $j(y)$ different from (22a) and (D5).

[1] R. Mansouri, Ann. Inst. H. Poincaré 27, 175 (1977).
[2] E. Glass, J. Math. Phys. 20, 1508 (1979).

[3]B. Mashhoon and H. Partovi, Ann. Phys. (NY) **130**, 1, 99 (1982).
[4]B. Mashhoon and H. Partovi, Phys. Rev. D **30**, 1839 (1984).
[5]C. B. Collins and J. Wainwright, Phys. Rev. D **27**, 1209 (1983).
[6]D. C. Srivastava and S. Prasad, Gen. Relativ. Gravit. **15**, 65 (1983).
[7]C. B. Collins, J. Math. Phys. **26**, 2009 (1985).
[8]C. B. Collins, Can. J. Phys. **64**, 191 (1986).
[9]M. Wyman, Phys. Rev. **70**, 396 (1946).
[10]D. C. Srivastava and S. Prassad, "On shear-free motion of charged perfect fluids obeying an equation of state," preprint Univ. of Gorakhpur, India, May, 1984, to appear in Gen. Relativ. Gravit.
[11]P. Kustaanheimo and B. Qvist, Comment. Phys. Math. Helsingf. **13**, 12 (1948).
[12]D. Kramer, H. Stephani, M. A. H. MacCallum, and E. Herlt, *Exact Solutions of Einstein's Field Equations* (Cambridge U. P., Cambridge, 1980), Chap. 12, Sec. 2.
[13]A. Barnes, Gen. Relativ. Gravit. **4**, 2, 129 (1973).
[14]A. J. White and C. B. Collins, J. Math. Phys. **25**, 332, 1460 (1984).
[15]M. C. Faulkes, Can. J. Phys. **47**, 1989 (1969).
[16]B. Mashhoon and H. Partovi, Phys. Rev. D **20**, 2455 (1979).
[17]M. Wyman, Aust. J. Phys. **31**, 111 (1978).
[18]M. Wyman, Can. Math. Bull. **19**, 343 (1976).
[19]G. C. McVittie, Ann. Inst. H. Poincaré **40**, 3, 235 (1984).
[20]H. Stephani, J. Phys. A **16**, 3529 (1983).
[21]D. C. Srivastava, "Exact solutions for shear-free motion of spherically symmetric perfect fluid distributions...," preprint submitted for publication.
[22]S. Chatterjee, Gen. Relativ. Gravit. **16**, 381 (1984).
[23]H. Nariai, Prog. Theor. Phys. **40**, 1013 (1968).
[24]P. Vickers, Ph. D. thesis, University of London, 1973.
[25]P. F. Byrd and M. D. Friedman, *Handbook of Elliptic Integrals for Engineers and Scientists* (Springer, Berlin, 1971), 2nd ed.
[26]P. C. Vaidya and Y. Shah, Ann. Inst. H. Poincaré **6**, 219 (1967).
[27]A. Banerjee, N. Chakravorty, and S. B. Duttachoudhury, Nuovo Cimento B **29**, 357 (1975).
[28]P. C. Vaidya and Y. Shah, Tensor (N. S.) **19**, 191 (1968).
[29]A. Banerjee, N. Chakravorty, and S. B. Duttachoudhury, Acta Phys. Pol. B **7**, 675 (1976).
[30]N. Chakravorty and S. Chatterjee, Acta Phys. Pol. B **9**, 777 (1978).
[31]G. C. McVittie, Mon. Not. R. Astron. Soc. **93**, 325 (1933); see also Astrophys. J. **143**, 682 (1966).
[32]A. Banerjee and S. Banerji, Acta Phys. Pol. B **7**, 389 (1976).
[33]A. Banerjee, N. Chakravorty, and S. B. Duttachoudhury, Aust. J. Phys. **29**, 113 (1976).
[34]M. C. Faulkes, Prog. Theor. Phys. **42**, 1139 (1969).
[35]E. N. Glass and B. Mashhoon, Astrophys. J. **205**, 570 (1976).
[36]H. Nariai, Prog. Theor. Phys. **38**, 92 (1967).
[37]J. Eisenstaedt, Phys. Rev. D **11**, 2021 (1975); **12**, 1573 (1975).
[38]H. Knutsen, Phys. Scr. **30**, 289 (1984).
[39]H. Bondi, Mon. Not. R. Soc. Phys. **142**, 333 (1969).
[40]I. H. Thompson and G. J. Whitrow, Mon. Not. R. Astron. Soc. **136**, 207 (1967), **139**, 499 (1968).
[41]H. Nariai and K. Tomita, Progr. Theor. Phys. **40**, 3, 679 (1968).
[42]W. B. Bonnor and M. C. Faulkes, Mon. Not. R. Astron. Soc. **137**, 239 (1967).
[43]A. Friedman, Z. Phys. **10**, 377 (1922); **21**, 326 (1924).
[44]W. S. Burnside and A. W. Panton, *The Theory of Equations* (Dublin U. P., Dublin, and Longmans, London, 1935), 10th edition.
[45]I. Frick and R. A. d'Inverno, Gen. Relativ. Gravit. **12**, 693 (1980).
[46]A. Karlhede, Gen. Relativ. Gravit. **14**, 835 (1982).
[47]H. Knutsen and R. Stabell, Ann. Inst. H. Poincaré **31**, 4, 339 (1979).
[48]H. Knutsen, Ann. Inst. H. Poincaré **39**, 2, 101 (1983).
[49]H. Knutsen, Gen. Relativ. Gravit. **16**, 8, 777 (1984).
[50]H. Knutsen, Phys. Scr. **31**, 305 (1985).

Linear independence of renormalization counterterms in curved space-times of arbitrary dimensionality

Ian Jack[a] and Leonard Parker
Department of Physics, University of Wisconsin-Milwaukee, Milwaukee, Wisconsin 53201

(Received 29 September 1986; accepted for publication 3 December 1986)

The counterterms in the Lagrangian of a renormalizable quantum field theory that involve the Riemann curvature tensor are considered. It is proved that for six or fewer dimensions the counterterms not containing derivatives of the Riemann tensor are linearly independent. It is shown that there appears to be a maximum space-time dimension for which identities can exist among invariants involving the product of n Riemann tensors (without derivatives acting on them.) Space-times that would require these products as counterterms for a renormalizable theory have a dimensionality which is one higher than this maximum dimension. This makes it plausible that the required set of counterterms not involving derivatives of the Riemann tensor is linearly independent in arbitrary dimensions. In the Appendix, a relation cubic in the Weyl tensor, which relates invariants cubic in the Riemann tensor in five dimensions, is proved.

I. INTRODUCTION

It is well known that in a three-dimensional space-time, the Weyl curvature tensor vanishes, which implies that[1]

$$R_{\lambda\mu\nu\kappa} = g_{\lambda\nu}R_{\mu\kappa} - g_{\lambda\kappa}R_{\mu\nu} - g_{\mu\nu}R_{\lambda\kappa} + g_{\mu\kappa}R_{\lambda\mu} - \tfrac{1}{2}(g_{\lambda\nu}g_{\mu\kappa} - g_{\lambda\kappa}g_{\mu\nu})R. \tag{1.1}$$

This yields the following relation quadratic in the Riemann tensor:

$$R^{\lambda\mu\nu\kappa}R_{\lambda\mu\nu\kappa} - 4R^{\mu\kappa}R_{\mu\kappa} + R^2 = 0. \tag{1.2}$$

In a four-dimensional curved space-time background, one requires, among the counterterms in the Lagrangian of a renormalizable quantum field theory, terms proportional to $R^{\lambda\mu\nu\kappa}R_{\lambda\mu\nu\kappa}$, $R^{\mu\nu}R_{\mu\nu}$, and R^2. If a linear relation like Eq. (1.2) were to hold in four dimensions, then the set of counterterms would be reducible. (Of course, we assume that the symmetries of the Riemann tensor have already been used to reduce the set of counterterms as much as possible.) Our aim in this paper is to assist in the determination of a minimal set of counterterms in arbitrary dimensions by precluding linear relations among invariants formed from products of Riemann tensors with no derivatives. We will not be concerned with linear relations among such invariants which hold only to within a total derivative. An example of such a relation is the Gauss–Bonnet theorem, which implies that the left-hand side of Eq. (1.2) is a total derivative in four dimensions.

In six dimensions, the counterterms not containing derivatives of the Riemann tensor are cubic in that tensor,[2–4] and one must consider the possibility of linear relations existing among such terms. Indeed, such relations among cubic invariants do exist in four[5] and in five dimensions.[6] In the Appendix, we prove a relation cubic in the Weyl tensor (and hence in the Riemann tensor) which holds in five dimensions. We will prove below that no such identities exist in six or more dimensions. We will also present arguments suggesting that there is no linear relation among the invariants of order n in the Riemann tensor (without derivatives acting on it) in a space-time of dimension $2n$ or greater. In a space-time of dimension $2n$, the counterterms for a renormalizable theory which do not involve derivatives of the Riemann tensor are of order n in that tensor. Hence one cannot eliminate such a counterterm by means of a linear combination of similar terms of order n. (Counterterms of the type under consideration here do not appear in odd-dimensional space-times.) By way of introduction, we will first outline a proof that no linear relation among $R^{\lambda\mu\nu\kappa}R_{\lambda\mu\nu\kappa}$, $R^{\mu\nu}R_{\mu\nu}$, and R^2 exists in four dimensions.

II. QUADRATIC INVARIANTS IN FOUR DIMENSIONS

Let $J_1 = R^{\lambda\mu\nu\kappa}R_{\lambda\mu\nu\kappa}$, $J_2 = R^{\mu\nu}R_{\mu\nu}$, and $J_3 = R^2$. Suppose a relation exists for some given dimension $d > 3$ of the form

$$\sum_{j=1}^{3} c_j J_j = 0. \tag{2.1}$$

Consider a d-dimensional space-time $Q_{d,p}$ in a coordinate system in which it is manifestly the direct product of a p-dimensional de Sitter space-time and a $(d-p)$-dimensional flat Euclidean space. The curvature tensor is then given by

$$R_{\alpha\beta\lambda\delta} = \begin{cases} K(g_{\alpha\gamma}g_{\beta\delta} - g_{\alpha\delta}g_{\beta\gamma}), & \alpha,\beta,\gamma,\delta = 0,1,\ldots,p-1, \\ 0, & \text{otherwise.} \end{cases} \tag{2.2}$$

It is evident that the invariants are the same as for a p-dimensional de Sitter space

$$J_1 = 2p(p-1), \quad J_2 = p(p-1)^2, \quad J_3 = p^2(p-1)^2. \tag{2.3}$$

Then Eq. (2.1) is

$$p(p-1)[c_3p^2 + (c_2 - c_3)p + 2c_1 - c_2] = 0. \tag{2.4}$$

By assumption, Eq. (2.1) holds for any space-time in d dimensions. In particular, Eq. (2.1) must hold for all $Q_{d,p}$ with $1 \leqslant p \leqslant d$. Hence Eq. (2.4) must have roots $p = 1,2,\ldots,d$. Clearly, $p = 0$ is also a root. Therefore Eq. (2.4) has at least $(d+1)$ roots. Thus, if $d > 3$, then Eq. (2.4) is a polynomial of degree 4 with more than four roots. Hence it must be trivial, which implies that

$$c_1 = c_2 = c_3 = 0. \tag{2.5}$$

Therefore no nontrivial relation of the form (2.1) can exist in a dimension d greater than 3.

[a] Present address: Department of Physics, University of Southampton, Southampton SO9 5NH, England.

 0022-2488/87/051137-03$02.50

III. CUBIC INVARIANTS IN SIX DIMENSIONS

The symmetries of $R_{\alpha\beta\gamma\delta}$ imply that there are eight independent invariants cubic in the Riemann tensor and not involving covariant derivatives.[7] In this section, we will prove in six or more dimensions that no linear combination of these invariants is zero.

The eight invariants are

$$\begin{aligned} &I_1 = R_{\alpha\beta}{}^{\gamma\sigma} R^{\alpha\lambda}{}_{\gamma\tau} R_{\lambda}{}^{\beta\tau}{}_{\sigma}, \\ &I_2 = R_{\alpha\beta}{}^{\gamma\sigma} R_{\gamma\sigma}{}^{\lambda\tau} R_{\lambda\tau}{}^{\alpha\beta}, \\ &I_3 = R^{\lambda}{}_{\tau} R_{\lambda\alpha\beta\gamma} R^{\tau\alpha\beta\gamma}, \quad I_4 = R R_{\alpha\beta\gamma\lambda} R^{\alpha\beta\gamma\lambda}, \\ &I_5 = R^{\alpha\gamma} R^{\beta\lambda} R_{\alpha\beta\gamma\lambda}, \quad I_6 = R_{\alpha}{}^{\beta} R_{\beta}{}^{\gamma} R_{\gamma}{}^{\alpha}, \\ &I_7 = R R_{\alpha\beta} R^{\alpha\beta}, \quad I_8 = R^3. \end{aligned} \tag{3.1}$$

Suppose that a relation of the form

$$\sum_{i=1}^{8} b_i I_i = 0 \tag{3.2}$$

exists in a dimension d greater than or equal to 6. For a space-time $Q_{d,p}$ defined in Sec. II, the invariants have the values

$$\begin{aligned} &I_1 = p(p-1)(p-2), \quad I_2 = 4p(p-1) \\ &I_3 = 2(p-1)^2 p, \quad I_4 = 2(p-1)^2 p^2, \\ &I_5 = p(p-1)^3, \quad I_6 = p(p-1)^3, \\ &I_7 = p^2(p-1)^3, \quad I_8 = p^3(p-1)^3. \end{aligned} \tag{3.3}$$

The relation in Eq. (3.2) thus becomes an equation of degree 6 in p, namely

$$\begin{aligned} p(p-1)\{&b_8 p^4 + b_7 p^3 + (2b_4 + b_5 + b_6) p^2 \\ &+ (b_1 + 2b_3 - 2b_4 - 2b_5 - 2b_6) p - 2b_1 \\ &+ 4b_2 - 2b_3 + b_5 + b_6\} = 0. \end{aligned} \tag{3.4}$$

By assumption Eq. (3.2) holds for any space-time in d dimensions, in particular for each $Q_{d,p}$ with $1 \leqslant p \leqslant d$. Hence Eq. (3.4) must have roots $p = 1,2,...,d$. Moreover we see by inspection that $p = 0$ is also a root of Eq. (3.4). Hence Eq. (3.4) has at least $(d+1)$ roots. Thus if $d > 5$ then Eq. (3.4) is a polynomial equation of degree 6 with more than six roots, and hence must be trivial. We then have

$$b_7 = b_8 = 0, \tag{3.5a}$$

$$2b_4 + b_5 + b_6 = 0, \tag{3.5b}$$

$$b_1 + 2b_3 - 2b_4 - 2b_5 - 2b_6 = 0, \tag{3.5c}$$

$$-2b_1 + 4b_2 - 2b_3 + b_5 + b_6 = 0. \tag{3.5d}$$

By considering further specific space-times we may show that all the b_i must vanish. First consider a space-time with metric

$$\begin{aligned} ds^2 = dt^2 - (&t^{4/3} dx_1^2 + t^{4/3} dx_2^2 \\ &+ t^{-2/3} dx_3^2 + dx_4^2 + \cdots + dx_d^2). \end{aligned} \tag{3.6}$$

This is the direct product of a Kasner space-time in four dimensions with a $(d-4)$-dimensional Euclidean space-time. We find for this space-time

$$R_{\alpha\beta} = 0, \tag{3.7}$$

$$I_1 = -\tfrac{128}{243}(1/t^6), \quad I_2 = -\tfrac{256}{243}(1/t^6). \tag{3.8}$$

Equation (3.7) implies that $I_3 = I_4 = \cdots = I_8 = 0$, and hence on substituting (3.8) into (3.2) we find

$$b_1 = -2b_2. \tag{3.9}$$

Equations (3.5a)–(3.5d) together with (3.9) then yield

$$b_1 = b_2 = 0, \tag{3.10}$$

$$b_3 + b_4 = 0. \tag{3.11}$$

Next we consider a space-time with metric

$$ds^2 = dt^2 - t(dx_1^2 + dx_2^2) - (dx_3^2 + \cdots + dx_d^2). \tag{3.12}$$

We find after some calculation that for this space-time (3.2) becomes

$$b_3 + 3b_4 + b_6 = 0. \tag{3.13}$$

From (3.5b), (3.11), and (3.13) we obtain

$$b_5 = 0. \tag{3.14}$$

Finally we consider a space-time with metric

$$ds^2 = dt^2 - t(dx_1^2 + dx_2^2 + dx_3^2) - (dx_4^2 + \cdots + dx_d^2). \tag{3.15}$$

We find that in view of the relations already found between the b_i, this space-time will only satisfy (3.2) provided

$$b_3 = b_4 = b_6 = 0. \tag{3.16}$$

To sum up, if the relation (3.2) is to be valid for a general space-time in d dimensions with $d > 5$, we must have

$$b_i = 0, \quad i = 1,...,8. \tag{3.17}$$

In other words there is no nontrivial relationship among $I_1, I_2, ..., I_8$ in more than five dimensions.

IV. CONCLUSIONS

We have proved that a relation of the form (2.1) does not exist for a dimension greater than or equal to 4, and a relation of the form (3.2) does not exist for a dimension greater than or equal to 6.

Although these proofs have dealt with invariants quadratic and cubic in the Riemann tensor, it seems clear that one could construct a similar proof for invariants of any order n in the Riemann tensor (without derivatives), showing that relationships among them would only be possible up to some maximum dimension of space-time. In fact, assuming such a proof exists, we can immediately determine this maximum dimension. It is clear that upon substitution of the curvature tensor for $Q_{d,p}$ into the invariants of order n in the Riemann tensor, the highest power of p which occurs in the equation analogous to (3.4) is p^{2n} (from the invariant R^n). One of the roots of that equation is guaranteed to be $p = 0$ because each term which contributes to an invariant, after substituting Eq. (2.2), contains at least one factor of p given by a contraction over the metric of the p-dimensional maximally symmetric subspace in $Q_{d,p}$. Since each $Q_{d,p}$ with $1 \leqslant p \leqslant d$ is a particular example of a d-dimensional space-time, the equation of order $2n$ in p must have the $d+1$ roots $p = 0,1,...,d$. Therefore, if $d + 1 > 2n$, then all the coefficients of the polynomial must be zero. It seems likely that one can then prove, as in Sec. III, that there can be no nontrivial linear relationship among the invariants of order n in the Riemann tensor. Therefore it is plausible that the maximum

dimension of space-times in which linear relations can exist among invariants of order n in the Riemann tensor (without derivatives) is given by

$$d_{\max} = 2n - 1. \tag{4.1}$$

For a renormalizable quantum field theory in a curved space-time of $2n$ dimensions, the counterterms which depend only on the Riemann tensor (and not its derivatives) must be of order n in that tensor, so that the action remains dimensionless. From Eq. (4.1), these counterterms are linearly independent. Although our rigorous proof extends only up to six dimensions, if as we have argued Eq. (4.1) is valid for all n, then these counterterms are always linearly independent.

Our conclusions concerning the linear independence of the counterterms not containing derivatives of the Riemann tensor should be useful in arriving at the minimal set of counterterms for a quantum theory in a curved space-time of arbitrary dimension.

ACKNOWLEDGMENTS

We thank John L. Friedman for helpful conversations.

We also thank the National Science Foundation for support under Grant No. Phy-8603173.

APPENDIX: INVARIANT CUBIC RELATION IN FIVE DIMENSIONS

In this Appendix, we prove that the relation

$$C_{\alpha\beta}{}^{\gamma\alpha} C^{\alpha\lambda}{}_{\gamma\tau} C_{\lambda}{}^{\beta\tau}{}_{\sigma} = \tfrac{1}{2} C_{\alpha\beta}{}^{\gamma\sigma} C_{\gamma\sigma}{}^{\lambda\tau} C_{\lambda\tau}{}^{\alpha\beta} \tag{A1}$$

holds in space-times of dimension 5 or less, where $C^{\alpha}{}_{\beta\gamma\delta}$ denotes the Weyl tensor appropriate to the given dimension. In Ref. 5, Xu gives a proof of this relation for four-dimensional space-times, using spinor methods special to four dimensions. Here, we give a proof valid in five (or fewer) dimensions.

In five or less dimensions, one clearly has the identity

$$C^{[\alpha\beta}{}_{\kappa\lambda} C^{\gamma\delta}{}_{\mu\nu} C^{\epsilon\eta]}{}_{\sigma\tau} = 0, \tag{A2}$$

where $[\cdots]$ denotes antisymmetrization with respect to the six contravariant indices only. Contracting between the lower and upper indices, keeping track of the various permutations, and using the tracelessness of the Weyl tensor, yields

$$C^{\alpha\beta}{}_{\gamma\delta} C^{\gamma\delta}{}_{\epsilon\eta} C^{\epsilon\eta}{}_{\alpha\beta} = 4 C^{\alpha\beta}{}_{\gamma\epsilon} C^{\gamma\delta}{}_{\alpha\eta} C^{\epsilon\eta}{}_{\beta\delta}. \tag{A3}$$

(This can also be obtained by realizing that only two essentially different contractions appear, and finding the numerical constant, 4, by evaluating each side of the relation for a simple metric.) This already gives a relation cubic in the Riemann tensor valid in five or less dimensions. By the same method, one can clearly obtain similar relations among contracted products of n Riemann tensors valid in $2n - 1$ or less dimensions.

To show that Eq. (A1) is equivalent to Eq. (A3), we show that

$$C^{\alpha\beta}{}_{\delta\eta} C_{\alpha}{}^{\gamma\delta}{}_{\epsilon} C^{\epsilon\eta}{}_{\gamma\beta} = 2 C^{\alpha\beta}{}_{\gamma\epsilon} C^{\gamma\delta}{}_{\alpha\eta} C^{\epsilon\eta}{}_{\beta\delta}. \tag{A4}$$

First use the cyclic identity

$$C^{\alpha}{}_{\beta\gamma\delta} = -C^{\alpha}{}_{\delta\beta\gamma} - C^{\alpha}{}_{\gamma\delta\beta}, \tag{A5}$$

on the last tensor in the product on the left side of Eq. (A6), to obtain

$$\begin{aligned} C^{\alpha\beta}{}_{\delta\eta} C_{\alpha}{}^{\gamma\delta}{}_{\epsilon} C^{\epsilon\eta}{}_{\gamma\beta} = & -C^{\epsilon\eta}{}_{\gamma\beta} C^{\beta\alpha}{}_{\eta\delta} C^{\delta}{}_{\alpha}{}^{\gamma}{}_{\epsilon} \\ & + C^{\alpha\beta}{}_{\gamma\epsilon} C^{\gamma\delta}{}_{\alpha\eta} C^{\epsilon\eta}{}_{\beta\delta}. \end{aligned} \tag{A6}$$

Then use the cyclic identity on all three Weyl tensors of that same product, obtaining a sum of eight terms cubic in that tensor. After a long calculation using the symmetry properties of the Weyl tensor (including the cyclic identity), one finds that the eight terms can be reduced to three:

$$\begin{aligned} C^{\alpha\beta}{}_{\delta\eta} C_{\alpha}{}^{\gamma\delta}{}_{\epsilon} C^{\epsilon\eta}{}_{\gamma\beta} = & 3 C^{\epsilon\eta}{}_{\gamma\beta} C^{\beta\alpha}{}_{\eta\delta} C^{\delta}{}_{\alpha}{}^{\gamma}{}_{\epsilon} \\ & + C^{\alpha\beta}{}_{\gamma\epsilon} C^{\gamma\delta}{}_{\alpha\eta} C^{\epsilon\eta}{}_{\beta\delta} \\ & + C^{\alpha\beta}{}_{\gamma\delta} C^{\gamma\delta}{}_{\epsilon\eta} C^{\epsilon\eta}{}_{\alpha\beta}. \end{aligned} \tag{A7}$$

As we have already proved Eq. (A3), we can write this in the form

$$\begin{aligned} C^{\alpha\beta}{}_{\delta\eta} C_{\alpha}{}^{\gamma\delta}{}_{\epsilon} C^{\epsilon\eta}{}_{\gamma\beta} = & 3 C^{\epsilon\eta}{}_{\gamma\beta} C^{\beta\alpha}{}_{\eta\delta} C^{\delta}{}_{\alpha}{}^{\gamma}{}_{\epsilon} \\ & + 5 C^{\alpha\beta}{}_{\gamma\epsilon} C^{\gamma\delta}{}_{\alpha\eta} C^{\epsilon\eta}{}_{\beta\delta}. \end{aligned} \tag{A8}$$

From Eqs. (A6) and (A8), one immediately obtains Eq. (A4). The latter, together with Eq. (A3), implies that

$$C^{\alpha\beta}{}_{\delta\eta} C_{\alpha}{}^{\gamma\delta}{}_{\epsilon} C^{\epsilon\eta}{}_{\gamma\beta} = \tfrac{1}{2} C^{\alpha\beta}{}_{\gamma\delta} C^{\gamma\delta}{}_{\epsilon\eta} C^{\epsilon\eta}{}_{\alpha\beta}, \tag{A9}$$

which is equivalent to Eq. (A1), as was to be proved. In Ref. 6, we give explicitly the relation involving the Riemann tensor that this yields in five dimensions.

[1] S. Weinberg, *Gravitation and Cosmology* (Wiley, New York, 1972), p. 144.
[2] D. J. Toms, Phys. Rev D **26**, 2713 (1982).
[3] J. Kodaira, Phys. Rev. D **33**, 2882 (1986).
[4] I. Jack, Nucl. Phys. B **274**, 139 (1986).
[5] D. Xu, Phys. Rev. D **35**, 769 (1987).
[6] I. Jack and L. Parker, Phys. Rev. D **35**, 771 (1987).
[7] P. B. Gilkey, J. Differential Geom. **10**, 601 (1975); Compositio Math. **38**, 201 (1979).

The nonlinear Boltzmann equation with partially absorbing boundary conditions. Global existence and uniqueness results

G. Toscani
Dipartimento di Matematica, Università di Ferrara, 44100 Ferrara, Italy

V. Protopopescu[a)]
Istituto di Matematica Applicata "G. Sansone," Università di Firenze-50139 Firenze, Italy

(Received 13 June 1986; accepted for publication 31 December 1986)

The method applied by Bellomo and Toscani [J. Math. Phys. **26**, 334 (1985)] for the Boltzmann equation in an infinite medium to establish global results for bounded media with partially absorbing boundary conditions is generalized. The method does not require that the equilibrium solution be the vacuum state and, accordingly, does not rely on positivity/monotonicity arguments. The growth produced by the nonlinearity is compensated by the combined effect of streaming and (partial) absorption (leakage) at the boundary.

I. INTRODUCTION

The nonlinear Boltzmann equation describes the evolution of a moderately dense gas whose state is supposed to be completely described by the one-particle distribution function $f=f(\mathbf{x},\mathbf{v},t)$ depending on position $\mathbf{x}\in\Omega\subset\mathbb{R}^3$, velocity $\mathbf{v}\in\mathbb{R}^3$, and time $t\in[0,T]$. In the absence of an external field the nonlinear Boltzmann equation reads

$$\frac{\partial f}{\partial t}+\mathbf{v}\cdot\nabla_x f=J(f,f)\,. \tag{1}$$

The flow (streaming) term, $\mathbf{v}\cdot\nabla_x f$, describes the change in the distribution caused by free translational movement between collisions, while the nonlinear operator J accounts for changes produced by two-particle collisions.

For interparticle potentials with cutoff,[1,2] J can be actually separated into a difference of two terms, the gain and loss operators, respectively:

$$J(f,f)=Q(f,f)-fR(f)\,, \tag{2}$$

$$Q(f,g)(\mathbf{x},\mathbf{v},t)=\int_{\mathbb{R}^3}d\mathbf{w}\int_0^{\pi/2}d\theta\int_0^{2\pi}d\epsilon\,B(\theta,q)\times f(\mathbf{x},\mathbf{v}',t)g(\mathbf{x},\mathbf{w}',t)\,, \tag{3}$$

$$R(f)(\mathbf{x},\mathbf{v},t)=\int_{\mathbb{R}^3}d\mathbf{w}\int_0^{\pi/2}d\theta\int_0^{2\pi}d\epsilon\,B(\theta,q)f(\mathbf{x},\mathbf{w},t)\,. \tag{4}$$

In the expressions above, $(\mathbf{v},\mathbf{w})$ are the precollisional velocities, $(\mathbf{v}',\mathbf{w}')$ are the postcollisional velocities, and the angles ϵ and θ are, respectively, the polar and azimuthal angles of $\mathbf{v}'$ in a spherical coordinate system with z axis in the direction of $\mathbf{q}=\mathbf{w}-\mathbf{v}$. Namely,

$$\mathbf{v}'=\mathbf{v}+(\mathbf{n}\cdot\mathbf{q})\mathbf{n}\,,\quad \mathbf{w}'=\mathbf{w}-(\mathbf{n}\cdot\mathbf{q})\mathbf{n}\,, \tag{5}$$

where $\mathbf{n}$ is a vector in the plane of the collision which bisects the angle formed by $\mathbf{v}-\mathbf{w}$ and $\mathbf{w}'-\mathbf{v}'$. Taking into account the definition of θ, it follows that $\mathbf{n}\cdot\mathbf{q}=q\cos\theta$. The collision kernel $B(\theta,q)$ is determined by the interparticle potential; in what follows, we shall suppose that B satisfies the inequality[1]:

$$\left|\frac{B(\theta,q)}{\sin\theta\cos\theta}\right|\leqslant F\frac{1+q}{q^{\delta}}\,,\quad 0\leqslant\delta<1\,. \tag{6}$$

Condition (6) is fulfilled for both soft and hard interactions including the hard sphere model; in the latter case, $\delta=0$. Supplemented with an initial (or initial and boundary) condition(s), the evolution equation (1) generates an initial-value (or initial-boundary value) problem to be solved—or, at least, to be proven solvable—in a suitable functional setting.

More than a hundred years after the proposal of the nonlinear Boltzmann equation, the global unique solvability question is still unanswered in full generality.

For simplified versions (mainly homogeneous[3] and discrete[4,5] models) global existence proofs have been provided in various settings at a satisfactory degree of generality.[5–8] Lately, global existence proofs for the full nonhomogeneous Boltzmann equation have been essayed in the framework of nonstandard analysis,[9,10] but their relevance for the classical solutions is still under investigation.

The global results for the full equation in classical settings could be obtained only under certain restrictive conditions on the scattering kernel and/on the data. The state of the art has been surveyed in Refs. 11 and 12. The major difficulty in getting global existence results for the nonhomogeneous Boltzmann equation comes from the absence of an adequate controlling mechanism for the possible growth induced by the nonlinear gain term. Several compensating effects have been proposed, such as mollification,[13] rarefaction,[14,15] discretization,[4,6–8] or closeness to equilibrium. Because of the difference in techniques, in the last category we distinguish between the equilibrium represented by the nontrivial Maxwellian state and the vacuum (zero) state. In the latter case, one takes essential advantage of positivity and point estimates. In principle, the compensation of the nonlinear gain term must come from the loss term. Unfortunately, the presently available formalisms are not able to take best advantage of the potential cancellation of the two terms, since the minus sign is lost in the estimation of the norms. It is not clear whether this is due to the lack of refinement of the used techniques or to an intrinsic difficulty of the equation itself. Some comments on this point can be found in Ref. 16.

Another possible compensation may arrive from the flow term $-\mathbf{v}\cdot\nabla_x f$, whose net eventual effect (at least in an infinite medium) is a local rarefaction of the gas.

[a)] Present address: Engineering Physics and Mathematics Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831.

 0022-2488/87/051140-06$02.50

The idea to use the flow term as a compensating term for $Q(f,f)$ seems to have been introduced by Tartar.[17] Different versions have been used lately by Illner and Shinbrot,[14] Bellomo and Toscani,[18] and Hamdache.[19,20]

The aim of this paper is to extend these results to bounded media. In this case, the compensation will arrive from the combined effect of the free streaming and the leakage through the surfaces.

Notwithstanding technical differences, all approaches used so far have applied a unique strategy: (1) choose the functional setting such that it be closed to linear operations and to the quadratic operation described by $Q(f,f)$; (2) find a convenient norm estimate of the nonlinear term which includes the possibly compensating factors; and (3) apply the contraction principle and the fixed point theorems by appropriately varying the compensating factors.

So far, the compensating factors have been essentially the norm of the departure of the initial state from the equilibrium state. When the equilibrium state is the vacuum, this norm depends on the initial state of rarefaction of the gas (expressible in terms of density, mean free path, etc.) and on the aggregating properties of the intermolecular potential. In this paper two new parameters will control the norm and its time evolution: the spatial dimension of the body and the absorption coefficient at the boundary. The smallness condition on the initial data is propagated at subsequent times by using the method of the characteristics. The results have been announced in Ref. 21.

II. STATEMENT OF THE PROBLEM AND NOTATION

We shall consider the initial-boundary value problem for the nonlinear equation (1) in three situations.

Problem A: The spatial domain Ω is the parallelepiped $\{2a_1,2a_2,2a_3\}$ and the boundary conditions are partially periodic:

$$f(\mp a_i,\mathbf{v},t)=\alpha f(\pm a_i,\mathbf{v},t),\quad v_i\gtrless 0,\qquad i=1,2,3,\quad 0\leqslant\alpha<1. \tag{7}$$

The accommodation coefficient α is related to the absorptive properties of the boundaries. The case $\alpha=0$ describes perfect absorption. The case $\alpha=1$ (perfect periodicity) is not included in the present treatment.

Problem B: The spatial domain Ω is the parallelepiped $\{2a_1,2a_2,2a_3\}$ and the boundary conditions are partially specularly reflecting:

$$f(\pm a_i,\mathbf{v},t)=f(\pm a_i,\mathbf{v}-2(\mathbf{v}\cdot\mathbf{n})\mathbf{n},t),\qquad i=1,2,3,\quad 0\leqslant\alpha<1, \tag{8}$$

where $\mathbf{n}$ is a generic notation for the exterior unit normal on the corresponding faces of Ω.

Problem C: The spatial domain Ω is the semi-infinite medium $\{\mathbf{x}\in\mathbb{R}^3/x_1<0\}$ and the boundary conditions are partially or perfectly specularly reflecting:

$$f(0,\mathbf{v},t)=\alpha f(0,\mathbf{v}-2(\mathbf{v}\cdot\mathbf{e}_1)\mathbf{e}_1,t),\quad 0\leqslant\alpha\leqslant 1, \tag{9}$$

where $\mathbf{e}_1$ is the exterior unit normal on the plane $x_1=0$.

The free-streaming operator, $-\mathbf{v}\cdot\nabla_{\mathbf{x}}$, with any of the boundary conditions (7)–(9) in the corresponding geometries generates a C_0 semigroup of positive contractions, $U(t)$. (For recent proofs in general settings see Refs. 22–24.)

The common feature of the problems A–C is that the action of the semigroup $U(t)$ can be explicitly computed for each of them as follows:

$$\text{(A)}\ (U(t)f)(\mathbf{x},\mathbf{v})=\sum_{\mathbf{j}\in\mathbb{Z}^3}\alpha^{|\mathbf{j}|}f(\mathbf{x}+\mathbf{ja}-\mathbf{v}t,\mathbf{v}), \tag{10}$$

$$\mathbf{j}=(j_1,j_2,j_3)\in\mathbb{Z}^3, \tag{11}$$

$$\mathbf{ja}=(2a_1j_1,2a_2j_2,2a_3j_3), \tag{12}$$

$$|\mathbf{j}|=|j_1|+|j_2|+|j_3|. \tag{13}$$

$$\text{(B)}\ (U(t)f)(\mathbf{x},\mathbf{v})=\sum_{\mathbf{j}\in\mathbb{Z}^3}\alpha^{|\mathbf{j}|}f(\mathbf{x}_{\mathbf{j}}+\mathbf{ja}-\mathbf{v}_{\mathbf{j}}t,\mathbf{v}_{\mathbf{j}}), \tag{14}$$

$$\mathbf{x}_{\mathbf{j}}=((-1)^{j_1}x_1,(-1)^{j_2}x_2,(-1)^{j_3}x_3). \tag{15}$$

$$\text{(C)}\ (U(t)f)(\mathbf{x},\mathbf{v})=f(\mathbf{x}-\mathbf{v}t,\mathbf{v})+\alpha f((\mathbf{x}-\mathbf{v}t)-2((\mathbf{x}-\mathbf{v}t)\cdot\mathbf{e}_1)\mathbf{e}_1,\mathbf{v}-2(\mathbf{v}\cdot\mathbf{e}_1)\mathbf{e}_1). \tag{16}$$

The functions appearing in formulas (10), (14), and (16) are always to be understood as functions with spatial support contained in Ω. In other words, $f=0$ if the spatial argument does not take values in Ω.

Using the semigroup $U(t)$ the solution of the original problem (1) with initial condition

$$f(\mathbf{x},\mathbf{v},t=0)=f_0(\mathbf{x},\mathbf{v}) \tag{17}$$

can be formally written in mild form as

$$f(\mathbf{x},\mathbf{v},t)=(U(t)f_0)(\mathbf{x},\mathbf{v})+\int_0^t ds(U(t-s)J(f(s),f(s)))(\mathbf{x},\mathbf{v}). \tag{18}$$

The integral form (18) is much better suited for the kind of arguments we shall develop; however, it will provide only mild solutions to the original problem.

We shall present in detail problem A. Problem B can be worked out along the same lines. At the end, we shall sketch the proof for problem C.

Let $d=\max(a_1,a_2,a_3)$ and, for any set E, let $C_b(E)$ be the space of bounded continuous functions defined in E.

The functional setting for problems A and B will be the Banach space:

$$\mathbb{B}_r=\{f\in C_b(\Omega\times\mathbb{R}^3\times\mathbb{R}_+)\times[f(\mathbf{x},\mathbf{v},t)(1+(1/d^2)|\mathbf{x}-\mathbf{v}t|^2)e^{r^2v^2}]^{-1}\in C_b(\Omega\times\mathbb{R}^3\times\mathbb{R}_+)\}, \tag{19}$$

with the norm defined by

$$\|f\|_{\mathbb{B}_r}=\sup_{(\mathbf{x},\mathbf{v},t)}|f(\mathbf{x},\mathbf{v},t)|(1+(1/d^2)|\mathbf{x}-\mathbf{v}t|^2)\exp(r^2v^2). \tag{20}$$

For problem C we shall use the space

$$\mathbb{B}_{h,r}=\{f\in C_b(\Omega\times\mathbb{R}^3\times\mathbb{R}_+)\times[|f(\mathbf{x},\mathbf{v},t)|e^{r^2v^2}]^{-1}\leqslant ch(|\mathbf{x}-\mathbf{v}t)\}, \tag{21}$$

where h is a given strictly positive function in $C_b(\mathbb{R}_+)$ and the constant c may depend on f.

In order to apply the contraction mapping principle, we need some technical results which we shall summarize in the next section.

III. ESTIMATES FOR THE FREE-STREAMING SEMIGROUP AND FOR THE COLLISION OPERATOR

Lemma 1: $U(t)$ maps $\mathbb{B}_r$ into $\mathbb{B}_r$.

Proof: For any $\varphi\in\mathbb{B}_r$, there is a constant K such that

$$(U(t)\varphi)(\mathbf{x},\mathbf{v})\leqslant K\sum_{\mathbf{j}\in\mathbf{Z}^3}\alpha^{|\mathbf{j}|}e^{-r^2v^2}\chi_\Omega(\mathbf{x}+\mathbf{ja}-\mathbf{v}t)\,, \quad (22)$$

where $\chi_\Omega(\mathbf{x})$ is the characteristic function of the domain Ω:

$$\chi_\Omega(\mathbf{x})=\begin{cases}1, & \mathbf{x}\in\Omega\,,\\ 0, & \mathbf{x}\notin\Omega\,.\end{cases} \quad (23)$$

Now consider the function

$$f_\alpha(x_1)=\sum_{l\in\mathbf{Z}}\alpha^{|l|}\chi_{](2l-1)a_1,(2l+1)a_1[}(x_1)\,, \quad (24)$$

and let us define

$$M_\alpha=\max_{k>0}\alpha^k(1+(2k+1)^2)\,. \quad (25)$$

Then we have

$$f_\alpha(x_1)\leqslant M_\alpha(1+(1/a_1^2)x_1^2)^{-1}\,. \quad (26)$$

Proceeding in a similar way for x_2 and x_3 and collecting the results, we get

$$(U(t)\varphi)(\mathbf{x},\mathbf{v})\leqslant Ke^{-r^2v^2}M_\alpha^3(1+(1/d^2)x^2)^{-1}\,, \quad (27)$$

which concludes the proof. □

Lemma 2: Let $\mathbf{u}$ and $\mathbf{w}$ be two orthogonal vectors, $|\mathbf{u}|>0$, $|\mathbf{v}|>0$. Then for any $a>0$, any $\mathbf{x}\in\mathbb{R}^3$, and any $t\in\mathbb{R}_+$, the following inequality holds:

$$\int_0^t(1+a^2|\mathbf{x}+\mathbf{u}s|^2)^{-1}(1+a^2|\mathbf{x}+\mathbf{w}s|^2)^{-1}\,ds \leqslant\frac{\pi}{1+a^2x^2}\frac{1}{a}\left\{\frac{1}{|\mathbf{u}|}+\frac{1}{|\mathbf{w}|}+\frac{1}{|\mathbf{u}+\mathbf{w}|}\right\}. \quad (28)$$

Proof: This Lemma is a particular case of Lemma 1 in Ref. 20, where the proof can also be found. □

Lemma 3: Let $f,g\in\mathbb{B}_r$. Then, under the hypothesis (6)

$$\int_0^t ds(U(t-s)[Q(f(s),g(s))-f(s)R(g(s))])(\mathbf{x},\mathbf{v})\in\mathbb{B}_r\,. \quad (29)$$

Moreover,

$$\left\|\int_0^t ds(U(t-s)Q(f(s),g(s)))(\mathbf{x},\mathbf{v})\right\| \leqslant 12Fd\pi^3S_{\delta,r}N_\alpha\|f\|\,\|g\|\,,$$

$$\left\|\int_0^t ds(U(t-s)[f(s)R(g(s))])(\mathbf{x},\mathbf{v})\right\| \leqslant 4Fd\pi^3S_{\delta,r}N_\alpha\|f\|\,\|g\|\,, \quad (30)$$

where

$$N_\alpha=\max_{x>0}\alpha^{x/2}(1+x^2)^2\,, \quad (31)$$

$$S_{\delta,r}=\sup_{\mathbf{v}\in\mathbf{R}^3}\int_{\mathbf{R}^3}\frac{1+q}{q^{\delta+1}}e^{-r^2w^2}\,d\mathbf{w}\,. \quad (32)$$

Proof: Let $f,g\in\mathbb{B}_r$. Then

$$\begin{aligned}&|Q(f,g)(\mathbf{x}+\mathbf{ja}-\mathbf{v}(t-s),\mathbf{v},s)|\\ &=\left|\int_{\mathbf{R}^3}d\mathbf{w}\int_0^{\pi/2}d\theta\int_0^{2\pi}d\epsilon\,B(\theta,q)\right.\\ &\quad\left.\times f(\mathbf{x}+\mathbf{ja}-\mathbf{v}(t-s),\mathbf{v}',s)g(\mathbf{x}+\mathbf{ja}-\mathbf{v}(t-s),\mathbf{w}',s)\right|\\ &\leqslant\int_{\mathbf{R}^3}d\mathbf{w}\int_0^{\pi/2}d\theta\int_0^{2\pi}d\epsilon\,B(\theta,q)\cdot\|f\|\,\|g\|\\ &\quad\cdot\left(1+\frac{1}{d^2}|\mathbf{x}+\mathbf{ja}-\mathbf{v}t+(\mathbf{v}-\mathbf{v}')s|^2\right)^{-1}\cdot e^{-r^2v'^2}\\ &\quad\times\left(1+\frac{1}{d^2}|\mathbf{x}+\mathbf{ja}-\mathbf{v}t+(\mathbf{v}-\mathbf{w}')s|^2\right)^{-1}\cdot e^{-r^2w'^2}\,.\end{aligned} \quad (33)$$

From (6), (10), and (33) it follows that $U(t-s)Q(f(s),g(s))(\mathbf{x},\mathbf{v})$ is bounded by

$$|U(t-s)Q(f,g)|\leqslant\left(\frac{1+\alpha}{1-\alpha}\right)^3F\pi^2\|f\|\cdot\|g\|e^{-r^2v^2} \times\int_{\mathbf{R}^3}d\mathbf{w}\,\frac{1+|\mathbf{v}|+|\mathbf{w}|}{q^\delta}\,e^{-r^2w^2}\,. \quad (34)$$

We can now evaluate

$$\begin{aligned}&\left|\int_0^t ds(U(t-s)Q(f(s),g(s))(\mathbf{x},\mathbf{v}))\right|\\ &\leqslant\sum_{\mathbf{j}\in\mathbf{Z}^3}\alpha^{|\mathbf{j}|}\int_0^t ds|Q(f,g)(\mathbf{x}+\mathbf{ja}-\mathbf{v}(t-s),\mathbf{v},s)|\\ &\quad\times\chi_\Omega(\mathbf{x}+\mathbf{ja}-\mathbf{v}(t-s))\\ &\leqslant 4N_\alpha\int_0^t ds\int_{\mathbf{R}^3}d\mathbf{w}\int_0^{\pi/2}d\theta\\ &\quad\times\int_0^{2\pi}d\epsilon\,B(\theta,q)\|f\|\,\|g\|e^{-r^2v^2}\cdot e^{-r^2w^2}\\ &\quad\times(1+(1/d^2)|\mathbf{x}-\mathbf{v}t+(\mathbf{v}-\mathbf{v}')s|^2)^{-1}\\ &\quad\cdot(1+(1/d^2)|\mathbf{x}-\mathbf{v}t+(\mathbf{v}-\mathbf{w}')s|^2)^{-1}\,,\end{aligned} \quad (35)$$

where we have used that $|\mathbf{j}|\geqslant|\mathbf{ja}|/2d$,

$$\alpha^{|\mathbf{ja}|/2d}\leqslant N_\alpha\cdot(1+(1/d^2)|\mathbf{ja}|^2)^{-2}$$

and

$$(1+(1/d^2)|\mathbf{ja}|^2)(1+(1/d^2)|\mathbf{x}-\mathbf{v}t+\mathbf{ja}+(\mathbf{v}-\mathbf{v}')s|^2) \geqslant\tfrac{1}{2}(1+(1/d^2)|\mathbf{x}-\mathbf{v}t+(\mathbf{v}-\mathbf{v}')s|^2)\,. \quad (36)$$

Due to the geometry of the binary collisions, the vectors $\mathbf{v}-\mathbf{v}'$ and $\mathbf{v}-\mathbf{w}'$ appearing in (35) are orthogonal. Recalling that $|\mathbf{v}-\mathbf{v}'|=q\cos\theta$, $|\mathbf{v}-\mathbf{w}'|=q\sin\theta$, and $|\mathbf{v}-\mathbf{v}'+\mathbf{v}-\mathbf{w}'|=q$, we have

$$\frac{1}{|\mathbf{v}-\mathbf{v}'|}+\frac{1}{|\mathbf{v}-\mathbf{w}'|}+\frac{1}{|\mathbf{v}-\mathbf{v}'+\mathbf{v}-\mathbf{w}'|} \leqslant\frac{3}{q\cos\theta\sin\theta}\,. \quad (37)$$

Taking into account (36) and (37), exchanging the order of

integration and applying Lemma 2 in (35), we obtain

$$\int_{\mathbf{R}^3} d\mathbf{w} \int_0^{\pi/2} d\theta \int_0^{2\pi} d\epsilon\, B(\theta,q) e^{-r^2w^2} \times \int_0^t (1+(1/d^2)|\mathbf{x}-\mathbf{v}t+(\mathbf{v}-\mathbf{v}')s|^2)^{-1} \cdot(1+(1/d^2)|\mathbf{x}-\mathbf{v}t+(\mathbf{v}-\mathbf{w}')s|^2)^{-1}$$

$$\leqslant 3dF\pi^3 \cdot \frac{1}{1+(1/d^2)|\mathbf{x}-\mathbf{v}t|^2} \cdot \int_{\mathbf{R}^3} \frac{1+q}{q^{1+\delta}} e^{-r^2w^2} d\mathbf{w} \leqslant \frac{3Fd\pi^3 S_{\delta,r}}{1+(1/d^2)|\mathbf{x}-\mathbf{v}t|^2} . \tag{38}$$

Analogously, let $f,g \in \mathbb{B}_r$. Then $|U(t-s)[f(s)R(g(s))]|$ is bounded as in (34).

Therefore proceeding as in the first part of the Lemma:

$$\left| \int_0^t ds (U(t-s)[f(s)R(g(s))])(\mathbf{x},\mathbf{v}) \right| \leqslant \sum_{\mathbf{j}\in\mathbf{Z}^3} \alpha^{|\mathbf{j}|} \int_0^t ds |fR(g)|(\mathbf{x}+\mathbf{j}a-\mathbf{v}(t-s),\mathbf{v},s) \times \chi_\Omega(\mathbf{x}+\mathbf{j}a-\mathbf{v}(t-s))$$

$$\leqslant 4N_\alpha \|f\| \, \|g\| e^{-r^2v^2}(1+(1/d^2)|\mathbf{x}-\mathbf{v}t|^2)^{-1} \cdot \int_0^t ds \int_{\mathbf{R}^3} d\mathbf{w} \int_0^{\pi/2} d\theta \int_0^{2\pi} d\epsilon\, B(\theta,q) \times (1+(1/d^2)|\mathbf{x}-\mathbf{v}t-\mathbf{q}s|^2)^{-1} e^{-r^2w^2}$$

and

$$\int_{\mathbf{R}^3} d\mathbf{w} \int_0^{\pi/2} d\theta \int_0^{2\pi} d\epsilon\, B(\theta,q) e^{-r^2w^2} \times \int_0^t ds (1+(1/d^2)|\mathbf{x}-\mathbf{v}t-\mathbf{q}s|^2)^{-1} \leqslant Fd\pi^3 S_{\delta,r} . \quad \square$$

IV. THE CONTRACTION MAPPING PRINCIPLE

The estimates calculated in the preceding section will allow us to show that for certain initial conditions one has a contraction property for a nonlinear mapping related to Eq. (18).

Let us consider the evolution equation (1) without the loss term, i.e.,

$$\frac{\partial f}{\partial t} + \mathbf{v}\cdot\nabla_{\mathbf{x}} f = Q(f,f) \tag{39}$$

with initial condition f_0. The (formal) mild solution of (39) is

$$f(\mathbf{x},\mathbf{v},t) = (U(t)f_0)(\mathbf{x},\mathbf{v}) + \int_0^t ((U(t-s)Q(f(s),f(s)))(\mathbf{x},\mathbf{v}))ds . \tag{40}$$

For any $\varphi \in \mathbb{B}_r$, we define the operator A by the formula (40), i.e.,

$$(A\varphi)(\mathbf{x},\mathbf{v},t) = (U(t)\varphi_0)(\mathbf{x},\mathbf{v}) + \int_0^t ds (U(t-s)Q(\varphi(s),\varphi(s)))(\mathbf{x},\mathbf{v}) , \tag{41}$$

where $\varphi_0(\mathbf{x},\mathbf{v}) = \varphi(\mathbf{x},\mathbf{v},0)$. From the previous estimates, it is easy to see that A is a well-defined operator from $\mathbb{B}_r$ into $\mathbb{B}_r$.

Proposition 1: If $0 \leqslant \varphi_0 \in C_b(\Omega \times \mathbb{R}^3)$ and $\varphi_0(\mathbf{x},\mathbf{v}) \leqslant Ke^{-r^2v^2}$, where

$$K < (48\pi^3 FdS_{\delta,r} N_\alpha M_\alpha^3)^{-1} , \tag{42}$$

then A satisfies the contraction mapping principle on $\mathbb{D}_K$, where

$$\mathbb{D}_K = \{0 \leqslant \varphi \in \mathbb{B}_r / \|\varphi\| \leqslant 2\|U(t)\varphi_0\|\} . \tag{43}$$

Proof: Since $\varphi_0 \leqslant Ke^{-r^2v^2}$, we have, by Lemma 1, that

$$(U(t)\varphi_0)(\mathbf{x},\mathbf{v}) \leqslant KM_\alpha^3 e^{-r^2v^2}(1+(1/d^2)|\mathbf{x}-\mathbf{v}t|^2)^{-1} ,$$

therefore

$$\|U(t)\varphi_0\| \leqslant KM_\alpha^3 . \tag{44}$$

Moreover, by using Lemma 3, inequality (44) and condition (42), we get for $\varphi \in \mathbb{D}_K$,

$$\left| \left| \int_0^t ds (U(t-s)Q(\varphi(s),\varphi(s)))(\mathbf{x},\mathbf{v}) \right| \right| \leqslant 12Fd\pi^3 S_{\delta,r} N_\alpha \cdot 4\|U(t)\varphi_0\|^2 \leqslant 48Fd\pi^3 S_{\delta,r} N_\alpha KM_\alpha^3 \|U(t)\varphi_0\| \leqslant \|U(t)\varphi_0\| .$$

Therefore

$$\|A\varphi\| \leqslant 2\|U(t)\varphi_0\| , \tag{45}$$

implying that $A\varphi \in \mathbb{D}_K$.

We shall now prove that A is a contraction mapping on $\mathbb{D}_K$ with respect to the norm (20), provided (42) is satisfied. Let $\varphi,\psi \in \mathbb{D}_K$. Then

$$|A\varphi - A\psi|(\mathbf{x},\mathbf{v},t) \leqslant \int_0^t ds |U(t-s)Q(\varphi(s),\varphi(s)) - U(t-s)Q(\psi(s),\psi(s))|(\mathbf{x},\mathbf{v})$$

$$\leqslant \int_0^t ds \sum_{\mathbf{j}\in\mathbf{Z}^3} \alpha^{|\mathbf{j}|} \int_{\mathbf{R}^3} d\mathbf{w} \int_0^{\pi/2} d\theta \int_0^{2\pi} d\epsilon\, B(\theta,q) \{\varphi(\mathbf{x}+\mathbf{j}a-\mathbf{v}(t-s),\mathbf{v}',s) \cdot|\varphi-\psi|(\mathbf{x}+\mathbf{j}a-\mathbf{v}(t-s),\mathbf{w}',s) + \psi(\mathbf{x}+\mathbf{j}a-\mathbf{v}(t-s),\mathbf{w}'s) \cdot|\varphi-\psi|(\mathbf{x}+\mathbf{j}a-\mathbf{v}(t-s),\mathbf{v}',s)\} \chi_\Omega(\mathbf{x}+\mathbf{j}a-\mathbf{v}(t-s))$$

$$= \int_0^t ds [|(U(t-s)Q(\varphi(s),|\varphi-\psi|(s)))(\mathbf{x},\mathbf{v})| + |(U(t-s)Q(|\varphi-\psi|(s),\psi(s)))(\mathbf{x},\mathbf{v})|] .$$

From this point on, taking into account the first of (30)

$$
\begin{aligned}
&|A\varphi - A\psi|(\mathbf{x},\mathbf{v},t) \\
&\quad \leqslant 4\|U(t)\varphi_0\|\|\varphi - \psi\|(1 + (1/d^2)|\mathbf{x} - \mathbf{v}t|^2)^{-1} \\
&\qquad \cdot e^{-r^2v^2} \cdot 12Fd\pi^3 S_{\delta,r} \\
&\quad \leqslant 48Fd\pi^3 S_{\delta,r} N_\alpha M_\alpha^3 \cdot K\|\varphi - \psi\| , \qquad (46)
\end{aligned}
$$

which proves the proposition.

V. THE ITERATION SCHEME

The contraction principle derived in the preceding section solves the "auxiliary" Boltzmann equation (39) which is obtained from the original Eq. (1) by removing the loss term $-fR(f)$. It is natural to expect that restoring the original equation by reintroducing the loss term will not affect the existence result.

The method was first introduced by Kaniel and Shinbrot[25] and soon became a standard procedure.[14,26] We remark that the method takes essential advantage of positivity/monotonicity arguments and, as a consequence, could be applied so far only to situations in which the equilibrium solution is the vacuum state.

Let $0 < T < \infty$ and let l_0 and u_0 be two functions in $\mathbb{D}_K$ such that $0 \leqslant l_0(\mathbf{x},\mathbf{v},t) \leqslant u_0(\mathbf{x},\mathbf{v},t)$ pointwise.

We define recursively two sequences, l_k and u_k, as solutions of the equations

$$
\frac{\partial l_{k+1}}{\partial t} + \mathbf{v}\cdot\nabla_{\mathbf{x}} l_{k+1} + l_{k+1} R(u_k) = Q(l_k,l_k) , \qquad (47)
$$

$$
\frac{\partial u_{k+1}}{\partial t} + \mathbf{v}\cdot\nabla_{\mathbf{x}} u_{k+1} + u_{k+1} R(l_k) = Q(u_k,u_k) , \qquad (48)
$$

$$
l_{k+1}(\mathbf{x},\mathbf{v},0) = u_{k+1}(\mathbf{x},\mathbf{v},0) = f_0(\mathbf{x},\mathbf{v}); \quad k = 1,2,..., \qquad (49)
$$

where $f_0(\mathbf{x},\mathbf{v})$ is the initial value distribution of the problem to be actually solved. At any step, l_k and u_k are solutions of simple linear systems for which the existence and uniqueness theory is a settled issue.[22–25] Moreover, if the pair (l_0,u_0) satisfies the beginning condition,[29] viz.

$$
0 \leqslant l_0(t) \leqslant l_1(t) \leqslant u_1(t) \leqslant u_0(t), \quad t \in [0,T] , \qquad (50)
$$

then the solutions of the systems (47)–(49) are unique, belong to $\mathbb{D}_k$, and satisfy the inequalities

$$
0 \leqslant l_0(t) \leqslant l_1(t) \leqslant l_2(t) \leqslant \cdots \leqslant u_2(t) \leqslant u_1(t) \leqslant u_0(t) . \qquad (51)
$$

The argument in Ref. 29 shows that, when the two sequences converge to the same limit $f = \lim u_k(t) = \lim l_k(t)$, this limit is the mild solution of the nonlinear Boltzmann equation (1) with initial condition (17). Summarizing, we can propose the following.

Theorem 1: Let $0 \leqslant f_0 \in C_b(\Omega \times \mathbb{R}^3)$ and $f_0(\mathbf{x},\mathbf{v}) \leqslant Ke^{-r^2v^2}$, with

$$
k \leqslant (56Fd\pi^3 S_{\delta,r} N_\alpha M_\alpha^3)^{-1} . \qquad (52)
$$

Then Problem A has a unique, non-negative global mild solution.

Proof: Take $l_0 = 0$. The beginning condition reduces to

$$
0 \leqslant u_1(t) \leqslant u_0(t) . \qquad (53)
$$

Since $l_0 = 0$, Eq. (48) gives

$$
\begin{aligned}
u_1(\mathbf{x},\mathbf{v},t) &= (U(t)u_0(0))(\mathbf{x},\mathbf{v}) \\
&\quad + \int_0^t ds(U(t-s)Q(u_0(s),u_0(s)))(\mathbf{x},\mathbf{v}) \\
&= (Au_0)(\mathbf{x},\mathbf{v},t) .
\end{aligned}
$$

Choose

$$
u_0(x,v,t) = 2\|U(t)f_0\|e^{-r^2v^2}(1 + (1/d^2)|\mathbf{x} - \mathbf{v}t|^2)^{-1}
$$

and recall that K satisfies inequality (52); then $u_0 \in \mathbb{D}_K$, and, since A maps $\mathbb{D}_K$ into $\mathbb{D}_K$,

$$
\begin{aligned}
&0 \leqslant u_1(\mathbf{x},\mathbf{v},t)e^{+r^2v^2}(1 + (1/d^2)|\mathbf{x} - \mathbf{v}t|^2) \\
&\quad \leqslant \|u_1\| = \|Au_0\| \leqslant 2\|U(t)f_0\| ,
\end{aligned}
$$

which proves the beginning condition (53).

Now we prove that the limits of the monotone sequences u_k and l_k are equal.

Owing to Eqs. (47) and (48), these limits satisfy

$$
\begin{aligned}
l(\mathbf{x},\mathbf{v},t) &= (U(t)f_0)(\mathbf{x},\mathbf{v}) + \int_0^t ds(U(t-s)[Q(l(s),l(s)) \\
&\quad - u(s)R(l(s))])(\mathbf{x},\mathbf{v}) , \qquad (54)
\end{aligned}
$$

$$
\begin{aligned}
u(\mathbf{x},\mathbf{v},t) &= (U(t)f_0)(\mathbf{x},\mathbf{v}) + \int_0^t ds(U(t-s)[Q(u(s),u(s)) \\
&\quad - l(s)R(u(s))])(\mathbf{x},\mathbf{v}) . \qquad (55)
\end{aligned}
$$

Therefore since $l \leqslant u$:

$$
\begin{aligned}
(u-l)(\mathbf{x},\mathbf{v},t) &\leqslant \int_0^t ds(U(t-s)[Q(u(s),(u-l)(s)) \\
&\quad + Q((u-l)(s),l(s)) \\
&\quad + l(s)R((u-l)(s))])(\mathbf{x},\mathbf{v}) . \qquad (56)
\end{aligned}
$$

Now, apply inequalities (30) to (56), recalling that $\|l\| \leqslant \|u\| \leqslant 2\|U(t)f_0\|$,

$$
\begin{aligned}
\|u-l\| &\leqslant 4\|U(t)f_0\| \cdot 12Fd\pi^3 S_{\delta,r} N_\alpha \cdot \|u-l\| \\
&\quad + 2\|U(t)f_0\| \cdot 4Fd\pi^3 S_{\delta,r} N_\alpha \|u-l\| \\
&\leqslant 56Fd\pi^3 S_{\delta,r} N_\alpha K M_\alpha^3 \|u-l\| . \qquad (57)
\end{aligned}
$$

When (52) holds, (57) implies $u = l$.

VI. CONCLUDING REMARKS

We have proved a global existence theorem for the nonlinear nonhomogeneous Boltzmann equation in 3-D-parallelepipedic geometry with partially absorbing-partially periodic boundary conditions (problem A).

From a technical point of view the result falls into the area of point-estimate results, based on positivity/monotonicity arguments. While implementing these arguments, we have used the explicit form of the free-streaming evolution semigroup and the fact that the equilibrium state of the system is the vacuum (zero) state. This explains the choice of both the geometries and the accommodation coefficients which do not include the conservative case ($\alpha = 1$). For the case $\alpha = 1$, the constant K in Proposition 1 becomes zero.

Conservative boundary conditions ($\alpha = 1$) have been studied in different functional settings and with different

techniques[27–29]; the two types of results can be considered complementary.

The constant K appearing in Proposition 1 and in Theorem 1 controls the norm of the initial data for which the global existence results can be proved. It is essentially a smallness condition on the initial data (closeness to equilibrium) of the same type like those imposed in Refs. 8, 14, 15, 18, and 20. The noticeable difference in our condition (42) is that two new parameters control the value of the constant K: the typical dimension of the body in which the gas evolves, d, and the accommodation coefficient at the surface of the body α. The dependence of K on d and α supports the physical intuition. The smaller is d and the higher the absorption (i.e., the smaller is α), the larger is the expected compensating effect and, therefore, the larger is the class of allowed initial conditions.

The method used throughout this paper can also be applied to any spatial domain with partially absorbing–partially backward reflecting walls including the purely absorptive walls ($\alpha = 0$) (see Ref. 30).

The technical reason for this is that the free-streaming semigroup for partial backward reflection maintains the functional dependence on the spatial variable[31] which is essential for our estimates. Certainly, this does not apply to specular reflection in domains other than parallelepipedic and to any diffuse reflection.

Obviously, one can solve the exterior problem with partially absorbing–partially specularly reflecting boundary conditions for rather arbitrary domains.[15]

With minimal changes, the proof applies to problem B and to variants of problems A and B in one or two dimensions (infinite slab, infinite prism) with boundary conditions analogous to (7) and (8).

A slight modification is required when dealing with problem C ($\Omega = \{\mathbf{x}\in\mathbb{R}^3/x_1<0\}$). For any $h\in L_1(\mathbb{R}_+)\cap C_b(\mathbb{R}_+)$, $h>0$, one defines then the space

$$\mathbb{B}_{h,r} = \{f\in C_b(\Omega\times\mathbb{R}^3\times\mathbb{R}_+)/|f(\mathbf{x},\mathbf{v},t)| \leqslant ch(|\mathbf{x}-\mathbf{v}t|)e^{-r^2v^2}\}$$

endowed with the norm

$$\|f\|_{\mathbb{B}_{h,r}} = \sup_{(\mathbf{x},\mathbf{v},t)} |f(\mathbf{x},\mathbf{v},t)|e^{-r^2v^2}h^{-1}(|\mathbf{x}-\mathbf{v}t|)\ .$$

From this point on, the estimates in Lemma 1 and Lemma 3 follow as in Ref. 26, with the action of the free-streaming semigroup replaced by (16). In this case, due to the form of the semigroup, perfect reflection can be included.

The results can be generalized as to allow a much slower decay of the data at large velocities. Instead of the Maxwellian $e^{-r^2v^2}$, one may use powerlike decays of the form $(1+v^2)^{-k}$, $k>(3-\delta)/2$ (see Refs. 20 and 26). The estimates become more elaborate but not less straightforward. Also, depending on the functional space chosen to solve the problem in, slightly weaker cutoff conditions may replace (6) (see Ref. 24).

ACKNOWLEDGMENTS

One of the authors (V.P.) acknowledges the supportive attitude of the Engineering Physics and Mathematics Division at ORNL during the final revision of the paper.

He also acknowledges the financial support of C.N.R. and the kind hospitality of the Institute of Applied Mathematics "G. Sansone," where the first draft of this work was written.

[1]H. Grad, "Principles of the kinetic theory of gases," in *Handbuch der Physik*, edited by S. Flügge (Springer, Berlin, 1958), Vol. 12.
[2]C. Cercignani, *Theory and Application of the Boltzmann Equation* (Scottish Academic, Edinburgh, 1975).
[3]T. Carleman, *Problèmes mathématiques dans la théorie cinétique des gaz* (Almquist and Wiksells, Uppsala, 1957).
[4]R. Gatignol, "Contribution à la théorie cinétique des gaz à repartition discrète de vitesses," in *Lecture Notes in Physics*, Vol. 36 (Springer, Berlin, 1975).
[5]L. Arkeryd, Arch. Rat. Mech. Anal. **45**, 1, 16 (1972).
[6]S. Kaniel and M. Shinbrot, J. Mec. **19**, 581 (1980).
[7]G. Toscani, Ann. Mat. Pura Appl. **138**, 4 (1984).
[8]Y. Shizuta and S. Kawashima, Hokkaido Math. J. **14**, 249 (1985).
[9]L. Arkeryd, Arch. Rat. Mech. Anal. **77**, 1 (1981).
[10]L. Arkeryd, Arch. Rat. Mech. Anal. **86**, 85 (1984).
[11]W. Greenberg, J. Polewczak, and P. F. Zweifel, "Global existence proofs for the Boltzmann equation," in *Nonequilibrium Phenomena I, The Boltzmann Equation*, edited by J. L. Lebowitz and E. W. Montroll (North-Holland, Amsterdam, 1983).
[12]W. Fiszdom, H. Lachowicz, and A. Palczewski, "Existence problems of the nonlinear Boltzmann equation," in *Trends and Applications of Pure Mathematics to Mechanics*, edited by P. G. Ciarlet and M. Roseau (Springer, Berlin, 1984).
[13]A. Ya. Povzner, Am. Math. Soc. Transl. (2) **47**, 193 (1962).
[14]R. Illner and M. Shinbrot, Commun. Math. Phys. **95**, 117 (1984).
[15]M. Shinbrot, Transp. Theory Stat. Phys. **15**, (1986).
[16]G. Frosali and V. Protopopescu, "Can the mixed norm be used in kinetic theory?," preprint, Seminari dell'Istituto di Matematica Applicata "G. Sansone" (1985).
[17]L. Tartar, *Some Existence Theorems for Semilinear Hyperbolic Systems in One Space Variable* (University of Wisconsin, Madison, WI, 1980).
[18]N. Bellomo and G. Toscani, J. Math. Phys. **26**, 334 (1985).
[19]K. Hamdache, C. R. Acad. Sci. Paris **297**, Série I 619 (1983).
[20]K. Hamdache, Jpn. J. Appl. Math. **2**, (1984).
[21]G. Toscani and V. Protopopescu, C. R. Acad. Sci. Paris **302**, Série I 255 (1986).
[22]Y. Shizuta, Commun. Pure Appl. Math. **36**, 705 (1983).
[23]R. Beals and V. Protopopescu, J. Math. Anal. Appl. (1986).
[24]S. Ukay, preprint, 1985.
[25]S. Kaniel and M. Shinbrot, Commun. Math. Phys. **58**, 65 (1978).
[26]G. Toscani, Arch. Rat. Mech. Anal. **95**, 37 (1986).
[27]S. Ukai, Proc. Jpn. Acad., Ser. A **50**, 179 (1974).
[28]Y. Shizuta and K. Asano, Proc. Jpn. Acad. Ser. A **53**, 3 (1977).
[29]S. Ukai, N. Point, and H. Ghidouche, J. Math. Pures Appl. **57**, 203 (1978).
[30]H. Babowsky, Transp. Theory Stat. Phys. **13**, 475 (1984).
[31]G. Frosali, C. V. M. VanderMee, and V. Protopopescu, Math. Meth. Appl. Sci. (1986).

Four-dimensional boson field theory. III. Nontriviality

George A. Baker, Jr. and J. D. Johnson
Theoretical Division, Los Alamos National Laboratory, University of California, Los Alamos, New Mexico 87545

(Received 22 May 1986; accepted for publication 31 December 1986)

The results of numerical investigations based on series analysis indicate clearly that the method of phantom fields constructs nontrivial, self-interacting scalar Euclidean boson field theories. It is found that these continuum theories arise as the scaling limit to normal critical points of lattice statistical mechanical models. The character of these theories is numerically indistinguishable from that of a classical theory on the lambda line near the tricritical point.

I. INTRODUCTION AND SUMMARY

Recently Baker and Johnson[1] announced a procedure for the construction of a nontrivial, self-interacting, Euclidean boson field theory in four dimensions. In a subsequent paper[2] it was proven their method, the method of "phantom fields," constructs a theory which has all the usual properties of a field theory. The exception is rotational invariance which was not proved because a lattice based ultraviolet cutoff was used, although current information is consistent with rotational invariance.

Two important properties have not yet been treated in detail. The first is the question of nontriviality and the second is the question of in principle computability by series methods of the resulting field theory. It is the purpose of this paper to study numerically these two questions. The second question, if answered in the affirmative (in a more detailed manner than we can treat it), implies unique limits to the limiting processes described previously[2] but does not address the question of, for example, lattice independence. We do answer numerically the computability question to a sufficient extent to allow us to treat numerically the question of nontriviality. De Carvalho *et al.*[3] have suggested, but not proved for our case, that for a four-dimensional self-interacting Euclidean field theory to be nontrivial, the critical indices of the corresponding statistical-mechanical, critical-point theory must be classical without logarithmic corrections. Our results are in fact in accord with their suggestions and the resulting theory we obtain is nontrivial and has, within fairly small numerical error, the aforementioned properties.

We concentrate our numerical work on the Blume[4]–Capel[5] model in four dimensions, which is a special case of the phantom field method. In order to calculate its properties we rely on the analysis of high-temperature series.[6–8] We consider the region along the lambda line near where it runs into the tricritical point. Put otherwise, we look at a range of parameters where, in the continuum limit, the renormalized four-line coupling constant is positive (and goes to zero at the tricritical point).

The main results of this analysis are that there are values, of the parameter S, which characterizes the Blume–Capel model, for which the high-temperature series expansions determine a continuum limit which represents a *nontrivial* self-interacting Euclidean boson field theory in four dimensions. The characteristics of this theory seem to agree with the classical exponents (mean-field and Ornstein–Zernike theories).

In the second section of this paper we introduce the phantom field model and attendant notation. We show how the Blume–Capel, or hyper-strong-coupling limit fits in the phantom field picture. We also discuss briefly the nature of the high-temperature series and their employment in the analysis of the continuum limit.

In the third section we analyze the magnetic susceptibility χ and the correlation length ξ^2. In terms of the high-temperature expansion variable K in the limit as the critical point is approached,

$$\chi \propto (K_c - K)^{-\gamma}, \quad \xi \propto (K_c - K)^{-\nu} \tag{1.1}$$

define the critical indices γ and ν. We find that as a function of K, χ may have a confluent singularity which diverges like $(K_c - K)^{-\gamma+\hat{\Delta}}$, where $\hat{\Delta} \simeq 0.25 \pm$, but it is very weak and may or may not be present. We estimate that in the range of field theoretic interest for the hyper-body-centered-cubic lattice $\gamma \simeq 1.00 \pm 0.02$, where 1.00 is the classical value. We find for ξ^2 that there is not much evidence for a confluent singularity. If one exists, it must be weak. We estimate $2\nu = 1.00 \pm 0.02$, where 1.00 is the classical value. In addition we have used these analyses to determine $K_c(S)$. As a further study, in preparation for latter sections, we have computed $\chi(\xi^2)$. This study has the advantage that the singular point is known to be $\xi^2 = \infty$. For convenience[6] we use the argument x, $\xi^2 = 0.1x/(1-x)$ so $\xi^2 = \infty$ corresponds to $x = 1$. Here we estimate $\gamma/2\nu = 1.00 \pm 0.01$ where 1.00 is the classical value. No real evidence for a confluent singularity was found. If one exists, it is rather weak. Corresponding results have been obtained for the hyper-simple-cubic lattice, but with lower precision.

In Sec. IV we analyze in various ways $\partial^2\chi/\partial H^2$, where H is the magnetic field. As the critical point is approached,

$$\frac{\partial^2 \chi}{\partial H^2} \propto (K_c - K)^{-\gamma-2\Delta}, \tag{1.2}$$

which defines the critical exponent Δ. The quantity of most interest is the four-line, renormalized coupling constant g,

 0022-2488/87/051146-13$02.50

$$g = -v\frac{\partial^2\chi}{\partial H^2}(a^4\chi^2\xi^4)^{-1}$$

$$= -100v(1-x)^2\frac{\partial^2\chi}{\partial H^2}(\xi^2(x))[a^4x^2\chi^2(\xi^2(x))]^{-1}, \tag{1.3}$$

where v is the specific volume per lattice site and a is the lattice spacing. We find as a function of the parameter S that g traces out a gently curving line which crosses zero going from positive to negative as S increases. The portion where g is positive corresponds to a nontrivial field theory. The portion where g is negative, in the context of the mean-field analysis,[9] represents a spinodal point which is the analytic continuation through a first-order phase transition. This theory differs by a change of limits from that proved[2] to exist and its significance to constructive field theory is not yet clear to us, although field theoretic perturbation theory in low order appears able to construct such a case. The numerical example of Baker and Johnson[1] is of this class.

In Sec. V we review the implications of convexity and how they apply to whether or not the singularities seen in Secs. III and IV represent actual second-order phase transitions or an analytic continuation to a spinodal point. It is shown that a resolution of this question in the neighborhood of $H=0$ involves the consideration of $\partial^4\chi/\partial H^4$ and in some cases, of course, even higher magnetic field derivatives. Mean-field theory gives a specific prediction for the critical point limit of $\partial^4\chi/\partial H^4$. From the global point of view a scan for the region $0\leqslant K\leqslant K_c$, $0\leqslant H\leqslant\infty$ for singularities in χ is required, although this search can be restricted in H by use of the high-field Griffiths–Hurst–Sherman (GHS) inequalities of Ellis *et al.*[10] They also prove that for $1\leqslant S\leqslant\sqrt{3}$ enough inequalities hold to assure that a normal critical point occurs. Using results of Newman,[11] we have slightly extended this range.

In Sec. VI we give an analysis of the behavior of $\partial^4\chi/\partial H^4$ and we present evidence, making due allowance for numerical errors of estimation, that a mean-field prediction for $\partial^4\chi/\partial H^4$ holds. Thus the general mean-field picture[9] that the variation of S leads to a lambda line of ordinary critical points which end in a tricritical point and continue with a line of first-order phase transitions appears valid. A surprising detail of mean-field theory, which also appears to be valid, is the analytic continuation through the phase transition along the $H=0$ line to a spinodal point. This behavior if correct, is contrary to what is known about the up–down magnetization phase boundary in the two-dimensional Ising model.[12,13]

In the seventh section we report a global scan over K and H in the region of S where $g>0$. We find that the results are in accord with expectations and that this region does in fact correspond to a normal critical point so the continuum limit exists and is in principal computable from the high-temperature series expansion. This result (numerical) answers the second of the two questions we set ourselves and completes our numerical study with good evidence that the method of phantom fields can construct nontrivial, self-interacting scalar boson Euclidean field theories in four dimensions.

II. PHANTOM FIELD MODEL

As the purpose of this paper is to examine numerically the questions of trivality and computability of the models proposed by Baker and Johnson[1] and proved to exist by Baker,[2] it will suffice to consider in detail a particular model. Specifically we start with the lattice cutoff version of the structure

$$Z=\int\mathscr{D}\phi(x)\exp\Big\{-\int d^4x[(\nabla\phi)^2 + m_0^2\phi^2+\lambda_0:(\phi^6+b\phi^4):]\Big\} \tag{2.1}$$

of a scalar, Euclidean, boson field theory. After we introduce the lattice cutoff, (2.1) becomes

$$Z=M^{-1}\int_{-\infty}^{+\infty}\cdots\int\prod_{\mathbf{r}}d\phi_{\mathbf{r}} \times\exp\Big\{-\sum_{\mathbf{r}}v\Big[\frac{8}{q}\sum_{\{\delta\}}\frac{(\phi_{\mathbf{r}}-\phi_{\mathbf{r}+\delta})^2}{a^2} + m_0^2\phi_{\mathbf{r}}^2+\lambda_0:(\phi_{\mathbf{r}}^6+b\phi_{\mathbf{r}}^4):\Big]\Big\}, \tag{2.2}$$

where M is a formal normalization constant, $\mathbf{r}$ ranges over a finite portion of the space lattice, $\{\delta\}$ is one-half the set of nearest neighbor sites on the lattice, v is the specific volume per lattice site, e.g., a^4 for the hyper-simple-cubic lattice, a is the lattice spacing, q is the lattice coordination number, and $:\phi^n:$ is the normal-ordered product. The normal order product[14] on a lattice ($a>0$) is

$$:\phi^{2p}:=\sum_{j=0}^{p}\frac{(2p)!(-1)^j}{(2p-2j)!\,j!}2^{-j}C^j\phi^{2p-2j}, \tag{2.3}$$

where C is the commutator $[\phi^-,\phi^+]$ and, in four dimensions is proportional to a^{-2}. It is convenient to reexpress (2.2) as

$$Z=M^{-1}\int_{-\infty}^{+\infty}\cdots\int\prod_{\mathbf{i}}d\sigma_{\mathbf{i}}\exp\Big\{K\sum_{\mathbf{i}}\sum_{\{\delta\}}\sigma_{\mathbf{i}}\sigma_{\mathbf{i}+\delta} - \sum_{\mathbf{i}}[\tilde{A}\sigma_{\mathbf{i}}^2+\tilde{g}_0\sigma_{\mathbf{i}}^4+\tilde{\lambda}_0\sigma_{\mathbf{i}}^6-H_{\mathbf{i}}\sigma_{\mathbf{i}}]\Big\}, \tag{2.4}$$

where a magnetic field has been added at each site,

$$\sigma_{\mathbf{i}}=(16v/qa^2K)^{1/2}\phi_{\mathbf{i}},\quad \tilde{\lambda}_0=\lambda_0q^3a^6K^3/(4096v^2), \tag{2.5}$$

and the value of $\tilde{A}$ is determined by

$$\langle\sigma_{\mathbf{i}}^2\rangle=1=\frac{\int_{-\infty}^{+\infty}x^2\exp(-\tilde{A}x^2-\tilde{g}_0x^4-\tilde{\lambda}_0x^6)dx}{\int_{-\infty}^{+\infty}\exp(-\tilde{A}x^2-\tilde{g}_0x^4-\tilde{\lambda}_0x^6)dx}, \tag{2.6}$$

for $H_{\mathbf{i}}=K=0$. The parameter K plays the role of the inverse temperature in the continuous spin Ising model. The limit $K\to K_c^-$, the critical point, corresponds to the continuum limit, $a\to0$, because the correlation length in units of the lattice spacing ξ and the Euclidean mass m satisfy[6] the relation $ma\xi=1$, and $\xi\to\infty$ as $K\to K_c$ at an ordinary second-order phase transition. We will discuss the possibility that a triple point might intervene to prevent the computation of this limit by series methods. For the cases we study in this report that problem does not arise.

The case of (2.4) that we have selected for detailed numerical investigation is the hyperstrong, bare coupling-constant limit. That is to say, $\lambda_0 \to \infty$ before $a \to 0$. By taking the normalization condition (2.6) into account, we find that the limiting single site spin distribution reduces to[5,14]

$$\tfrac{1}{2}[\delta(s-S)+\delta(s+S)]S^{-2}+(1-S^{-2})\delta(s), \quad S \geqslant 1, \tag{2.7}$$

in the absence of a magnetic field, where $\delta(x)$ is a Dirac delta function. The moments of (2.7) are, for zero magnetic field,

$$\begin{aligned} &I_{2n+1} \equiv \langle s^{2n+1} \rangle = 0, \quad I_0 \equiv \langle 1 \rangle = 1, \\ &I_{2n} \equiv \langle s^{2n} \rangle = S^{2(n-1)}, \quad n \geqslant 1. \end{aligned} \tag{2.8}$$

The various quantities of interest which we will study are the magnetic susceptibility χ, and the second moment definition of the correlation length ξ^2. They are defined by

$$\chi = \sum_i \langle \sigma_0 \sigma_i \rangle, \quad \xi^2 = \left[\sum_i i^2 \langle \sigma_0 \sigma_i \rangle\right](8\chi)^{-1}. \tag{2.9}$$

We may deduce from the divergence of these quantities the location of the critical point, which as we have remarked, corresponds to the continuum limit. Further quantities of interest are the renormalized four-line coupling constant g and the renormalized six-line coupling constant λ. They are defined by

$$\begin{aligned} g &= -\left(\frac{v}{a^4}\right)\frac{\partial^2 \chi}{\partial H^2}(\chi^2 \xi^4)^{-1}, \\ \lambda &= -\left(\frac{v}{a^4}\right)^2 \frac{\partial^4 \chi}{\partial H^4}(\chi^3 \xi^8)^{-1}. \end{aligned} \tag{2.10}$$

In subsequent sections we will analyze them by the method of high-temperature expansions. These expansions are known for general single-site spin distributions through tenth order in K. The information for the series for χ, ξ^2, and $\partial^2\chi/\partial H^2$ is given by Baker and Kincaid.[6] That for $\partial^4\chi/\partial H^2$ is given by Johnson and Baker.[8] We will also need χ for general H and that series is directly derivable from Kincaid *et al.*[7]

As a function of the parameter S, see (2.7), there are a number of significant values for the model. For $S=1$, Eq. (2.7) reduces to the spin-$\frac{1}{2}$ Ising model as there are only two peaks in the distribution for that case. For $S \leqslant \sqrt{2}$, the Yang–Lee theorem holds (Lieb and Sokal[15]). It is useful to write out explicitly the first few terms of

$$\begin{aligned} \frac{\partial^2 \chi}{\partial H^2} &= (I_4 - 3I_2^2) + 4qK(I_4 - 3I_2^2)I_2 + O(K^2), \\ \frac{\partial^4 \chi}{\partial H^4} &= (I_6 - 15I_2I_4 + 30I_2^3) + qK(6I_2I_6 \\ &\quad + 10I_4^2 - 150I_2^2I_4 + 270I_2^4) + O(K^2), \end{aligned} \tag{2.11}$$

which become, on substituting in the moments,

$$\begin{aligned} \frac{\partial^2 \chi}{\partial H^2} &= (S^2 - 3) + 4qK(S^2 - 3) + O(K^2), \\ \frac{\partial^4 \chi}{\partial H^4} &= (S^4 - 15S^2 + 30) + qK(6S^4 \\ &\quad - 140S^2 + 270) + O(K^2), \end{aligned} \tag{2.12}$$

where q is, as at (2.2), the lattice coordination number. At $S = [\frac{1}{2}(15 - \sqrt{105})]^{1/2} \simeq 1.541\,598\,07...$, the $\partial^4\chi/\partial H^4$ changes from positive to negative for $K=0$. At $S = [\frac{1}{2}(15 + \sqrt{105})]^{1/2} \simeq 3.552\,953\,05...$, it changes back to positive at $K=0$ and remains so for all larger values of S. At $S = \sqrt{3} \simeq 1.732\,050\,8...$, $\partial^2\chi/\partial H^2$ changes from negative to positive for $K=0$ and remains so for all larger values of S. This zero is a double zero in K. These signs of these derivatives are important in determining the signs of the renormalized coupling constants (2.10) as χ and ξ^2 are necessarily non-negative by Griffiths inequalities.[16] Of course, we need to know the signs and values of the limit as $K \to K_c$ for the determination of the corresponding field theory.

III. ANALYSIS OF THE SERIES FOR THE SUSCEPTIBILITY AND THE CORRELATION LENGTH

We here begin the discussion of our numerical analysis with an analysis of the series for the magnetic susceptibility χ and the second moment definition of the correlation length ξ^2 [Eq. (2.9)]. From an analysis of these series we determine estimates of the critical temperature K_C, the divergence exponent γ for χ, and the divergence exponent 2ν for ξ^2. In a separate analysis, we also estimate the ratio $\frac{1}{2}\gamma/\nu$. The main method of analysis in this section is to compute the Padé approximants to the logarithmic derivative of χ and of ξ^2. The rationale behind this procedure is that if, for example, $\chi \propto (K_c - K)^{-\gamma}$ for K near K_c, then $d \ln \chi/dK = -\gamma/(K - K_c) + o[(K - K_c)^{-1}]$ for K near K_c. This leading order can be well represented by Padé approximants.

Sample of the series we consider are, for example, $S = 1.77$ on the hyper-body-centered cubic (HBCC) lattice

$$\begin{aligned} \chi(K) = 1 &+ 16K + 257.0632K^2 + 4130.069\,500K^3 + 66\,210.948\,37K^4 + 1\,061\,448.254K^5 + 1.699\,821\,922 \times 10^7 K^6 \\ &+ 2.722\,135\,043 \times 10^8 K^7 + 4.356\,595\,244 \times 10^9 K^8 + 6.972\,493\,233 \times 10^{10} K^9 \\ &+ 1.115\,469\,449 \times 10^{12} K^{10} + \cdots, \end{aligned} \tag{3.1}$$

$$\begin{aligned} \xi^2(K) = 2K &+ 32K^2 + 514.132\,287\,5K^3 + 8260.233\,199K^4 + 132\,420.5991K^5 + 2\,122\,878.690K^6 + 3.399\,634\,065K^7 \\ &+ 5.444\,240\,577 \times 10^8 K^8 + 8.713\,255\,847 \times 10^9 K^9 + 1.394\,507\,671 \times 10^{11} K^{10} + \cdots. \end{aligned} \tag{3.2}$$

For the analysis of the ratio $\frac{1}{2}\gamma/\nu$ it is convenient to consider the series $\chi(x)$ where we first revert (3.2) to give $K(\xi^2)$, next substitute that series in $\chi(K)$ to give $\chi(\xi^2)$, and finally use the transformation [as used in Eq. (1.3)], $\xi^2 = 0.1x/(1-x)$, to map the critical point $\xi^2 = \infty$ into the point $x = 1$,

$$\chi^{-1}(x) = 1 - 0.8x - 0.162\,658x^2 - 3.093\,68\times10^{-2}x^3 - 4.527\,533\,372\times10^{-3}x^4 - 8.5472\times10^{-4}x^5$$
$$- 2.046\,432\,231\times10^{-4}x^6 - 1.907\,577\,470\times10^{-4}x^7 - 1.248\,786\,254\times10^{-4}x^8$$
$$- 1.034\,627\,788\times10^{-4}x^9 - 7.233\,145\,199\times10^{-5}x^{10} - \cdots . \tag{3.3}$$

The use of the Padé method and the assessment of its apparent errors is discussed in more detail in the next section.

We present in Table I a general survey at $H = 0$ of the more common critical point parameters of the Blume–Capel model,[4,5] together with their apparent error. This error is plus or minus the number in parentheses and is in the last digit quoted. For the hyper-simple-cubic (HSC) and HBCC lattices, we examine the Blume–Capel model for $1.7 \leqslant S \leqslant 2.1$.

Anticipating results from subsequent sections, we will find that there is a tricritical point at about $S_t = 1.829$ for the HBCC lattice and at about $S_t = 1.939$ for the HSC lattice. Therefore those results in Table I for S larger than these values are likely to be for a spinodal point in a metastable region. Those data for S smaller are for a true critical point while, of course, $S = S_t$ is for the tricritical point itself. As we will also see of additional results throughout this paper, the results in Table I are consistent with the critical properties of the mean field analysis of Blume *et al.*[9]

The K_c estimates of Table I are determined through Padé analyses. We computed Padé approximants to $d \ln \chi/dK$ and $d \ln \xi^2/dK$ to obtain estimates of the K_c. Both analyses were examined to determine a best K_c for each lattice and value of S. Additional analyses beyond what are reported in Table I have been made, but we have just reported a representative sample. The reader should take note that the precision falls at the ends of this range. As will be seen in later sections this result is a general feature. Since logarithmic corrections to, for example, $\chi \propto (1 - K/K_c)^{-1}$ are expected[3,6] for $S \simeq 1.0$, and none are observed in this range, it is logical to suppose that the behavior changes for some intermediate value of S. More particularly, in the course of the analysis of the next section we see some evidence that this value is near 1.73.

TABLE I. Critical temperatures and exponents for the magnetic susceptibility and correlation length on the HBCC and HSC lattices.

S	K_c	$2\nu/\gamma$	γ	2ν
HBCC				
1.7	0.063 31(3)	1.00(1)	1.03(3)[a]	1.0(1)
1.8	0.062 309(8)	1.00(1)	1.02(2)[a]	1.01(2)
1.829	0.062 015(5)	1.000(3)	1.01(2)[a]	1.01(1)
1.9	0.061 29(5)	1.00(2)	1.00(1)	1.00(1)
2.0	0.060 23(2)	1.00(1)	0.99(5)[a]	1.0(4)
2.1	0.059 17(4)	1.00(1)	0.96(6)	1.0(5)
HSC				
1.7	0.129 (1)	1.01(2)	1.05(8)	1.1(2)
1.8	0.125 8(2)	1.005(6)	1.04(2)	1.0(1)
1.9	0.122 5(1)	1.00(3)	1.03(1)	1.04(9)
1.9393	0.121 2(9)	1.00(3)	1.03(7)	1.03(4)
1.95	0.120 9(6)	1.00(4)	1.02(5)	1.03(5)
2.0	0.119 (5)	1.00(5)	1.0(4)	1.0(2)
2.1	0.116 0(8)	1.00(8)	1.00(7)	1.0(3)

[a] Indicates error estimates increased as implied by the Baker–Hunter analysis.

A further possible problem in the analysis of these series is the possibility of confluent singularities. This situation is discussed in more detail in Sec. IV. Here we have employed the Baker–Hunter[17] method of analysis which, while depending on an estimate of K_c, takes account of this possibility. Here χ may have a confluent singularity which diverges like $(K_c - K)^{-\gamma+\hat{\Delta}}$, where $\hat{\Delta} \simeq 0.25 \pm$, but it is very weak and may or may not be present. For the function ξ^2, there is not much evidence of a confluent singularity. If one exists it must be a very weak one. The values of γ and 2ν were determined by the aforementioned ($d \log/dK$) analyses. The error estimates were increased where necessary as indicated by the Baker–Hunter analyses. The results are all consistent with $\gamma = 2\nu = 1$, the mean-field prediction.

To estimate directly the ratio $\gamma/2\nu$, we analyze the series of the type Eq. (3.3). A quick look at Eq. (3.3) shows that $\chi^{-1}(x)$ is apparently converging at the critical point, $x = 1$. Direct analyses of this series reveals a zero at $x = 1$ with an error of less than 1×10^{-4} for the highest orders (using nine or ten terms) of Padé approximants. The Baker–Hunter analysis about $x \equiv 1$ results are given in Table I. Note that in the current case the x analysis yields a smaller apparent error than the $d \log/dK$.

IV. ANALYSIS OF THE SERIES FOR THE RENORMALIZED COUPLING CONSTANT

The first step in the analysis of any series is to identify its salient features. As pointed out in Sec. I, Eq. (1.3), the series we are interested in is the ratio $-\partial^2\chi/\partial H^2/\chi^2\xi^4$. As further pointed out, this ratio changes sign at high temperatures from positive for $S < \sqrt{3}$ to negative for $S > \sqrt{3}$. In fact,[18] $\partial^2\chi/\partial H^2$ possesses a double zero in K at $S = \sqrt{3}$. Numerical studies show that for $S < \sqrt{3}$ these two zeros occur for complex values of K roughly in the directions of plus and minus i. They collide at $K = 0$ for $S = \sqrt{3}$ and then one moves, as S increases, out the positive real axis and the other along the negative real axis. The one of physical interest is the one which moves along the positive real axis. We follow again the transformation of Eq. (1.3) to express g in terms of a variable x such that the critical point, when it exists, always corresponds to $x = 1$. It is convenient to use

$$G = 0.01\left(\frac{a^4}{v}\right)x^2 g$$
$$= -(1-x)^2\frac{\partial^2\chi}{\partial H^2}(\xi^2(x))[\chi^2(\xi^2(x))]^{-1} \tag{4.1}$$

as an object for study since it is finite at $x = 0$.

We expect, if the field theory is to be nontrivial, that g and hence G will be finite for $x = 1$. One such sample series is for $S = 1.77$ on the HBCC lattice. The reason for this choice will be clear later.

$$G(x) = -0.1329 + 5.316\times10^{-2}x + 0.137\,899\,503\,6x^2 + 1.275\,148\,92\times10^{-3}x^3 + 9.248\,526\,797\times10^{-3}x^4$$
$$+ 2.796\,826\,869\times10^{-3}x^5 + 3.133\,267\,087\times10^{-3}x^6 + 1.812\,932\,607\times10^{-3}x^7$$
$$+ 1.449\,688\,782\times10^{-3}x^8 + 9.731\,249\,025\times10^{-4}x^9 + 7.453\,208\,147\times10^{-4}x^{10} + \cdots . \quad (4.2)$$

We have analyzed this series by the method of Padé approximants.[19] The results are displayed in Fig. 1 as well as those for $S = 1.80$ on the HSC lattice. The location of the nearest singularity of $G(x)$ can be estimated by the poles of the approximants. This method suggests a singularity in the vicinity of $x = 1.3$–1.4, and so a radius of convergence of the Maclaurin series for $G(x)$ is estimated to be noticeably greater than unity. It will be noticed from the figure that the zero on the positive real axis which started from $x = 0$ at $S \simeq 1.732$ has now moved out to about $x \simeq 0.78$. It is this scale of movement which reflects the sensitivity of the model to the value of S.

The Padé approximant $[L/M]$ to a function $f(x)$ is defined by the equations

$$[L/M] = P_L(x)/Q_M(x),$$
$$Q_M(x)f(x) - P_L(x) = O(x^{L+M+1}), \quad (4.3)$$
$$Q_M(0) = 1.0,$$

where P_L and Q_M are polynomials of degrees L and M, respectively. In Table II we list the values of various approximants for $S = 1.77$ on HBCC lattice at $x = 1$. The standard procedure[20] for analyzing the apparent error is as follows. First eliminate approximants with defects ([5/5], [6/4], and [7/3] in this case). Then examine the sequence of near diagonal approximants where $L + M = 9$, 10 for the largest and smallest values. Here for $L + M = 10$ they are [3/7] and [4/6] and for $L + M = 9$ they are [3/6] and [5/4]. Referring to a table of values (not herein reported) we find for this case that these extremal properties continue to be valid for all values of x between zero and unity. The largest difference is between the [3/6] and the [3/7]. These approximants give the upper and lower limits shown in Fig. 1. If we examine the difference between these two approximants, we find [by (4.3)] that it is proportional to x^{10}. Next the standard procedure is to compare the values of the approximants $0 < x \leqslant 1$ to see what the largest coefficient of x^{10} inferred is. Here we find that the largest such coefficient occurs at $x = 1$ and is (times 100) 6.2×10^{-2}. Thus we estimate that $G(1) = 8.14\times10^{-2} \pm 6\times10^{-4}$ for this case, or $g = 4.07 \pm 0.03$. A similar analysis has been performed for a number of values of S on both the HSC and the HBCC lattice. We do not always estimate a symmetrical error bound but sometimes there appears to be more error on one side of the central value than the other. These results are summarized in Fig. 2. Here $G(1)$ vs S is plotted together with the apparent errors and an interpolating line for the central values for each lattice. Note that $S = 1.77$ for HBCC and $S = 1.80$ for the HSC seem to be roughly optimal in terms of the ratio of $G(1)$ to the apparent error. At $S = 1.80$ for the HSC lattice we estimate $G(1) = 0.116 \pm 8\times10^{-3}$ or $g = 11.6 \pm 0.8$. The point $G(1) = 0$ will be considered in more detail in the next section and its relation to tricritical phenomena discussed.

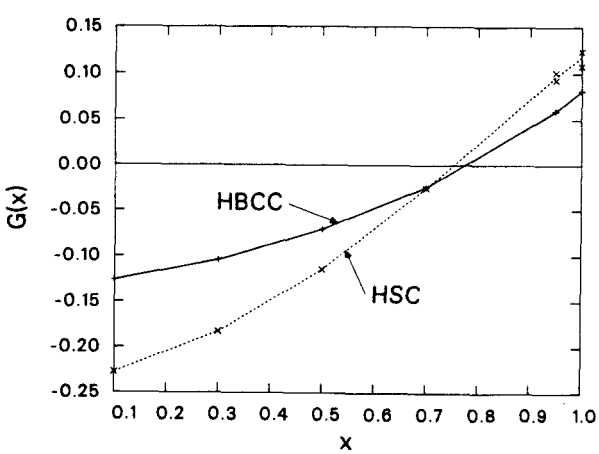


FIG. 1. $G(x)$ as a function of x. The solid curve is for $S = 1.77$ on the HBCC lattice and the dotted curve is for $S = 1.80$ on the HSC lattice.

There is a further possible source of error. It might possibly be that the point $x = 1$ is not a regular point but some kind of singular point of $G(x)$. To check this possibility, we have analyzed the series (4.2) (and of course the other cases as well) by the confluent singularity method of Baker and Hunter.[17] In brief, this method assumes

$$g(x) = \sum_{i=1}^{n} A_i(1-x)^{-\gamma_i}, \quad (4.4)$$

where here the singular point is taken as $x = 1$, and transforms the series for $G(x)$ to a series with simple poles at $1/\gamma_i$ and residues $-A_i/\gamma_i$, which can be analyzed by Padé approximants. The results of this analysis are not definitive, but are consistent with the following two possibilities. First, $x = 1$ is a regular point of $G(x)$. Second, there is some indication that the approach to $G(1)$ might be like $(1-x)^{\phi}$, $0.5 < \phi < 0.8$. In this case, however, the coefficient is negative so the curve is hooking upwards. Under these circumstances the values quoted for $G(1)$ may be too low, but the conclusion $G(1) > 0$ holds *a fortiori*. Hence we conclude that there

TABLE II. Values of the Padé approximants to $G(1.0)$ for $S = 1.77$ on the HBCC lattice times 100.

D \ N	2	3	4	5	6	7
2	6.926 59	7.396 57	7.959 47	8.094 60	8.133 52	8.131 64
3	7.256 02	11.054 1	8.100 86	8.148 45	8.131 72	8.133 35[b]
4	7.565 06	8.009 89	8.136 04	8.132 37	8.140 23[c]	
5	7.751 89	8.196 78	8.132 08	8.135 33[a]		
6	7.884 67	8.115 82	8.146 33			
7	7.970 07	8.177 95				

[a] Pole at $-0.208\,287$ with a residue of 5×10^{-12}.
[b] Pole at $-0.098\,890$ with a residue of 1×10^{-14}.
[c] Pole at $-0.558\,796$ with a residue of -4×10^{-8}.

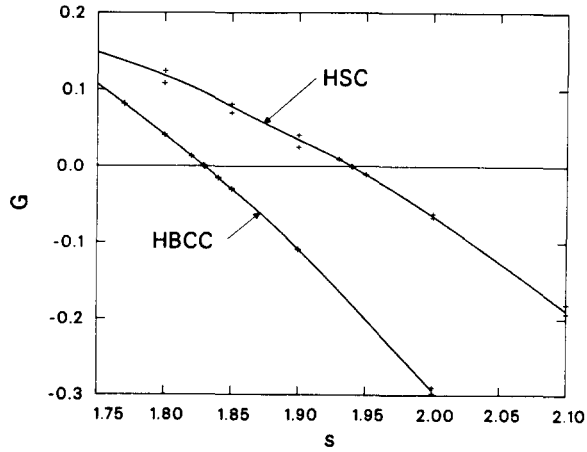


FIG. 2. $G(1)$ as a function of S. The solid curves are the central values for the HBCC and HSC lattices and the apparent errors are indicated by + signs.

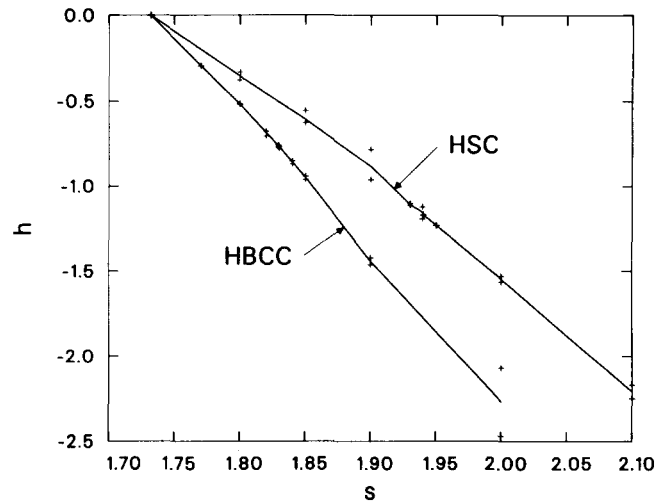


FIG. 4. $h(1)$ as a function of S. The solid curves are the central values for the HBCC and HSC lattices and the apparent errors as indicated by + signs.

exist ranges of S for which the renormalized coupling differs from zero and hence the theory is nontrivial.

It is also of interest to consider the behavior of the positive real zero, x_0, as a function of S. This behavior can also be deduced numerically from the direct Padé analysis of $G(x)$. We display, along with the apparent errors, the results of our analysis of this quantity in Fig. 3. The point at which the line of zeros crosses $x = 1$ corresponds to $G(1) = 0$, which we discuss in the next section. We estimate that the crossing occurs at about $S = 1.829$ for the HBCC and at about $S = 1.939$ for the HSC lattice. It is interesting to see that the line of zeros passes smoothly to $x > 1$ which corresponds to $\xi^2 < 0$. In order to look further at the structure of $G(x)$ we have computed

$$h(x) = G(x)/(1 - x/x_0) \tag{4.5}$$

and plotted $h = h(1)$ in Fig. 4. Although the results are not definitive, they are least consistent with the idea that $h(x)$ is a smooth function so that $(1 - x/x_0)$ factors out of $G(x)$ as a simple zero. Note that since $x_0 = 0$ for $S = \sqrt{3}$, $h(1) = 0$ automatically for that S.

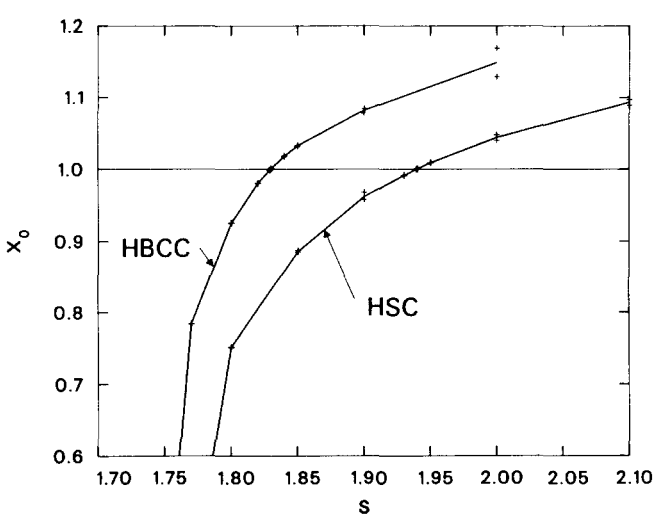


FIG. 3. The positive real zero, x_0, of $G(x)$ as a function of S. The central values are connected by solid curves and the apparent errors indicated by + signs.

V. CONVEXITY

The hyperstrong coupling model discussed in this paper is mathematically the same as the Blume[4]–Capel[5] model. It is of interest to see how our results compare with the Blume *et al.*[9] mean-field analysis. Briefly, their results, insofar as we require them, are as follows. The Helmholtz free energy is, in mean-field approximation, given by

$$A(M) - A(0) = aM^2 + bM^4 + cM^6 + \cdots, \tag{5.1}$$

where

$$\begin{aligned} a &= S^2/2\beta - \tfrac{1}{2}J, \\ b &= (1/8\beta)(S^4 - \tfrac{1}{3}S^6), \\ c &= (1/6\beta)(\tfrac{1}{2}S^6 - \tfrac{3}{8}S^8 + \tfrac{3}{40}S^{10}), \end{aligned} \tag{5.2}$$

with $\beta = 1/kT$, k is Boltzmann's constant and T is the absolute temperature, and $J = qS^2K/\beta$. The critical temperature is identified by $a = 0$, i.e., $K_c = q^{-1}$, and it is a normal critical point if $b > 0$. This situation prevails if $S < \sqrt{3}$, independent of temperature. When we note that $c > 0$ for all S, $1 \leqslant S < \infty$, we find a tricritical point for $S = \sqrt{3}$ and a triple point for $S > \sqrt{3}$, $b < 0$. In this latter case, the triple point occurs at a temperature determined by $4ac = b^2$ when the phases $M = 0$, $M = \pm(-\frac{1}{2}b/c)^{1/2}$ are in equilibrium with each other. For the same value of S at a higher temperature determined by $5ac = 3b^2$ there appear at $M = \pm(-\frac{1}{5}b/c)^{1/2}$ two critical end points at nonzero values of the magnetic field H which are joined by first-order transition lines (in the H-T plane) to the above mentioned triple point. [For this discussion we require $M \ll 1$, in order that expansion (5.1) be valid.]

Put more simply, as far as our present work is concerned, when b changes sign the critical point at $H = 0$ passes through the tricritical point and for b negative lies behind a triple point and hence becomes a spinodal point on the other side of a first-order phase transition from the point

about which our series expansions are constructed. Thus the continuum limit would not be accessible by our numerical methods. From the computational point of view, as S increases from $\sqrt{3}$, there is a divergent susceptibility at $H=0$, $T=T_t$, the tricritical point, which bifurcates into two critical end points ($H\neq 0$) at each of which there is a divergent magnetic susceptibility. In mean-field theory, the triple point which remains at $H=0$ is transparent and a simple analysis of $\chi(T)$ for $H=0$ will show only the spinodal point with no indication of the passage through the first-order phase transition to the, in principle, inaccessible region. Of course, as long as b remains positive no such problem arises and an ordinary critical point occurs.

The actual situation is considerably different. It is elementary to show, in the one phase region, that the Helmholtz free energy, which from thermodynamic considerations is convex as a function of the magnetization M, is given by

$$A(M)-A(0)=\frac{1}{2}\chi\left(\frac{M}{\chi}\right)^2-\frac{1}{24}\left(\frac{M}{\chi}\right)^4\frac{\partial^2\chi}{\partial H^2}-\left(\frac{M}{\chi}\right)^6\times\left[\frac{1}{720}\frac{\partial^4\chi}{\partial H^4}-\frac{1}{72}\frac{1}{\chi}\left(\frac{\partial^2\chi}{\partial H^2}\right)^2\right]+O(M^8)\,, \tag{5.3}$$

for small M. At an actual ordinary critical point, we expect from the ideas of thermodynamic scaling that

$$\frac{\partial^{2n}\chi}{\partial H^{2n}}\propto(1-K/K_c)^{-\gamma-2n\Delta}\,, \tag{5.4}$$

for $H=0$ as $K\to K_c$ from smaller values of K. The critical indices γ and Δ are susceptibility and gap indices, respectively, as we have remarked before. To compare (5.3) with (5.1) we see that $\chi\to\infty$ as $K\to K_c$ corresponds to $a\to 0$. The next term, b, is generally of order unity so that $3\gamma-2\Delta\simeq 0$. This relation is satisfied by the mean-field values $\gamma=1$ and $\Delta=\frac{3}{2}$ in four dimensions. The renormalization group suggests logarithmic corrections, but our discussion here is not a precise one so we will not go into this point now. The third term c is also of order unity. However, each of the two terms which form the coefficient of M^6 in (5.3) are proportional to $(1-K/K_c)^{2(3\gamma-2\Delta)-\gamma}$ which diverges roughly like $(1-K/K_c)^{-\gamma}$. Thus in order for the mean-field picture of the tricritical point to hold we must have

$$\lim_{K\to K_c}\chi\frac{\partial^4\chi}{\partial H^4}\left(\frac{\partial^2\chi}{\partial H^2}\right)^{-2}=10\,, \tag{5.5}$$

plus sufficient cancellations in the higher-order terms (in M) to make a local expansion analysis about $M=0$ a valid guide to the tricritical behavior. Without such cancellations thermodynamic scaling suggests the structure of (5.3) to be

$$A(M)-A(0)=(1-K/K_c)^{2\Delta-\gamma}yB(y)\,, \tag{5.6}$$

where

$$y=M^2(1-K/K_c)^{2(\gamma-\Delta)}\,. \tag{5.7}$$

In addition to the mean-field analysis of the Blume–Capel model described above, a renormalization group analysis has been made, and is reviewed by Lawrie and Sarbach.[21] If we follow their computations (Chap. 5) but replace the coefficient v of ψ^6 by $v(1-K/K_c)$, then the behavior parallels that for λ_0 implied by the phantom field constraint $\tilde{\lambda}_0=\text{const}$ in (2.5) when we use the common value $\nu=\frac{1}{2}$ [as defined by (1.1)]. In their Eq. (5.147) this behavior means that $\tilde{p}|g|^{-\phi_{p,t}}$ is not an additional scaling variable. The resulting theory gives the critical and tricritical exponents exactly the same as in the mean-field analysis. The logarithmic corrections to the critical exponents are clearly absent both along the "lambda line" [S less than the S for which $\partial^2\chi/\partial H^2(K_c)$ changes sign] and at the tricritical point. Thus this aspect also agrees with mean-field theory. This point is important as theoretical evidence[3,22,23] links the occurrence of logarithmic corrections to the triviality of the theory.

Related results have been obtained by Hara *et al.*[24] They show that for spatial dimension $d>4$, which is of course not our current case, that when [Eq. (2.4)] $\tilde{A},\tilde{g}_0,\tilde{\lambda}_0$ are positive and when $\tilde{g}_0/\tilde{A}^2$ and $\tilde{\lambda}_0/\tilde{A}^3$ are small enough, that mean-field theory correctly gives the thermodynamic critical indices. For example $\gamma=1$, $\alpha=0$, and $\Delta_4=\frac{3}{2}$, where α is the critical index for the specific heat and Δ_4 is the gap index determined by the ratio $\partial^2\chi/\partial H^2/\chi$. Their results give a further indication of the relevance of mean-field theory to field theories and are in agreement in this respect with the Ginsburg criterion.[21]

If we use, in addition, Sokal's[25] inequality, $2\nu\geqslant\gamma$, which follows (has been proved on hyper-simple-cubic lattices) by reflection positivity for these models,[2] then we find that (2.10) becomes

$$g=-\left(\frac{v}{a^d}\right)\frac{\partial^2\chi}{\partial H^2}(\chi^2\xi^d)^{-1}\leqslant\text{const}(K_c-K)^{(1/2)(d-4)}\,. \tag{5.8}$$

Thus the field theory, constructed by the phantom field method for these "soft" systems, has no two-particle zero energy scattering for $d>4$ and hence is trivial[26] since the Lebowitz inequalities also hold[24] here.

To investigate the validity of the mean-field picture locally, we compute in the next section the behavior of $\partial^4\chi/\partial H^4$ to see if (5.5) does in fact hold. It is, of course, only necessary and not sufficient but if it fails in a significant way then the mean-field analysis is inadequate to describe the actual behavior.

Now consider the scaling forms (5.6). For small values of S, $1\leqslant S\leqslant\sqrt{2}$, we know by the Yang–Lee theorem[25] that we have a normal critical point at $M=0, K=K_c$. By the results of Ellis *et al.*,[10] even though the Yang–Lee theorems fails for $S>\sqrt{2}$, the Griffiths–Hurst–Sherman inequalities and the Lebowitz inequalities continue to hold for $1\leqslant S\leqslant\sqrt{3}$. These results assure that a normal critical point continues to exist at $M=0, K=K_c$ and that it is characterized by a divergent correlation length and magnetic susceptibility. In other words the mass gap goes continuously to zero as $K\to K_c^-$ and so the continuum limit is accessible to high temperature series methods. Ellis *et al.*[10] in fact prove more; they show that as S increases that those same inequalities continue to hold for

$$\cosh SH\geqslant(S^4-2S^2-1)/(S^2-1)\,, \tag{5.9}$$

where H is the magnetic field variable. Further improvements are possible. Using results of Newman,[11] we have been able to show that those inequalities again hold for

$$2\cosh K \geqslant 2 - 3S^2 + S^4, \tag{5.10}$$

which in the region $K < K_c$ gives a slight improvement over $S = \sqrt{3}$ in the range over which we can rigorously prove the existence of a normal critical point and hence a continuum field theory. Newman's method should be capable of further extension.

We increase S and look for a possible conversion of this critical point to a tricritical point. The most sensitive criterion available is a violation of convexity, i.e., a failure of $\partial^2 A / \partial M^2 \geqslant 0$. If this failure of convexity should occur for the region $y = O(1)$ where our expansion (5.3) is presumed to be valid, then we find by direct computation the necessary conditions (based on the first three terms only) are

$$\frac{\partial^2 \chi}{\partial H^2} \geqslant 0, \quad \chi \frac{\partial^4 \chi}{\partial H^4} \geqslant \frac{17}{2}\left(\frac{\partial^2 \chi}{\partial H^2}\right)^2. \tag{5.11}$$

The first condition corresponds to $b \leqslant 0$ in (5.1) and the second one is necessary that there be a real solution for y. If the first condition fails, then A is automatically convex for the region $y = o(1)$ without further consideration. Note that the mean-field case (5.5) satisfies (5.11). So long as either of the conditions (5.11) fail and the coefficients of M^2, M^4, and M^6 do not simultaneously vanish [we do not expect this as a stronger cancellation than (5.5) would be required], $A(M)$ is locally convex in the sector $y = o(1)$. This result does not preclude trouble for $y = O(1)$, which values still scale to $M = 0$ as $K \to K_c$.

In order to complete our check of the identification of K_c as a normal critical point, we note that, by standard thermodynamics

$$\left.\frac{\partial^2 A}{\partial M^2}\right|_T = \frac{1}{\chi(K,H)}, \tag{5.12}$$

where χ is the magnetic susceptibility in any set of variables we choose as convenient. In order to assure that there are no other intruding singularities, we must scan the region $0 \leqslant K \leqslant K_c$, $-\infty \leqslant H \leqslant +\infty$ for values of $\chi = \infty$. This project is carried out in Sec. VII by use of the series through tenth order of Kincaid *et al.*[7] for $M(K,H)$. The coefficients are polynomials in the moments of the single-site spin distribution. We first differentiate the magnetization M to get χ. Then we note that if we weight (2.7) by e^{Hs}, the moments are

$$\begin{aligned} I_0 &= 1, \quad I_{2n} = \frac{S^{2(n-1)}(1+\tau^2)}{1 + (2S^{-2} - 1)\tau^2}, \\ I_{2n-1} &= \frac{2S^{2n-3}\tau}{1 + (2S^{-2} - 1)\tau^2}, \end{aligned} \tag{5.13}$$

where we have used the notation,

$$\tau = \tanh \tfrac{1}{2} HS. \tag{5.14}$$

The substitution of these moments directly into the series gives the required results.

VI. ANALYSIS OF THE SERIES FOR $\partial^4\chi/\partial H^4$

Thermodynamic scaling predicts that the quantity $\partial^4\chi/\partial H^4/(\chi^3\xi^8)$, which is related to the six-line coupling constant λ of Eq. (2.10), should be finite in the limit $\xi \to \infty$, if this limit exists. A convenient quantity to study here is

$$\begin{aligned} \Lambda(x) &= 10^{-4}\left(\frac{a^4}{v}\right)^2 x^4 \lambda \\ &= -(1-x)^4 \frac{\partial^4 \chi}{\partial H^4}(\xi^2(x))[\chi^3(\xi^2(x))]^{-1}, \end{aligned} \tag{6.1}$$

where x is given as in (1.3) and $\Lambda(0)$ is finite. As explained in Sec. II for $S < [\frac{1}{2}(15 - \sqrt{105})]^{1/2}$, $\Lambda(0) < 0$, for $[\frac{1}{2}(15 - \sqrt{105})]^{1/2} < S < [\frac{1}{2}(15 + \sqrt{105})]^{1/2}$, $\Lambda(0) > 0$, and for S larger $\Lambda(0) < 0$ again. By (5.1), (5.3), (4.1), and (6.1) the coefficient of M^6 in the expansion of the free energy is

$$c = \frac{1}{720\chi^3}\left(\frac{1}{1-x}\right)^4 [\Lambda(x) + 10G^2(x)]. \tag{6.2}$$

If c is positive finite or infinite, or negative and finite, when $x \to 1$ and $G(1) > 0$, then locally (i.e., near $H = 0$) the free energy has the structure of a normal critical point.

Let us begin by studying $\Lambda(x)$. For $S = 1.77$ on the HBCC lattice, the series[8] is

$$\begin{aligned} \Lambda(x) = {} & 7.178\,437\,59 - 11.626\,799\,424x + 3.517\,844\,534x^2 + 1.253\,104\,024x^3 - 0.477\,791\,0471x^4 \\ & + 0.122\,437\,001\,0x^5 - 9.434\,831\,223\times 10^{-2}x^6 + 3.368\,411\,731\times 10^{-2}x^7 - 1.093\,361\,292\times 10^{-2}x^8 \\ & + 1.214\,038\,453\times 10^{-2}x^9 - 9.115\,136\,829\times 10^{-4}x^{10} + \cdots. \end{aligned} \tag{6.3}$$

We analyze this series in the same general way as we did in Sec. IV. In Fig. 5 we show, together with the apparent errors, the Padé approximant values as a function of x for (6.3). We also show in this figure the corresponding results for $\Lambda(x)$ on the HSC lattice for $S = 1.80$. It is to be noted that the zero of $\Lambda(x)$ which occurred at $x = 0$ for $S \simeq 1.5416$ has moved to about $x = 0.965$ for the HBCC case and to about $x = 0.925$ for the HSC case. In Table III we list the Padé approximants to $\Lambda(1)$ based on the series (6.3). We estimate $\Lambda(1) = -0.093 \pm 0.02$. That the relative error is so large is mainly because, as can be seen from the figure, the value is quite small for $\Lambda(1)$. The corresponding result for $S = 1.80$ on the HSC lattice is $\Lambda(1) = -0.17 \pm 0.03$. We have performed corresponding analyses for a number of values of S. These results for $\Lambda(1)$ are summarized in Fig. 6 for the HBCC lattice and Fig. 7 for the HSC lattice.

A further feature of $\Lambda(x)$ which is of interest is the location of the zero as a function of S which enters the region of physical interest at $S \simeq 1.5416$. We display our Padé estimates of the location of this zero y_0, in Fig. 8 for the HBCC

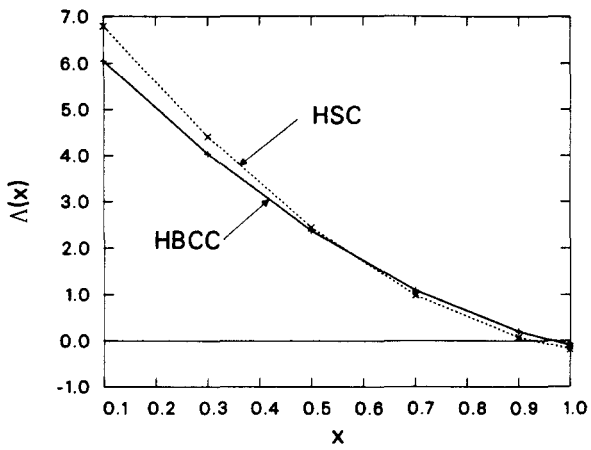


FIG. 5. $\Lambda(x)$ as a function of x. The solid curve is for $S = 1.77$ on the HBCC lattice. The apparent errors are indicated by + signs. The dotted curve is for $S = 1.80$ on the HSC lattice. The apparent errors are indicated by × signs.

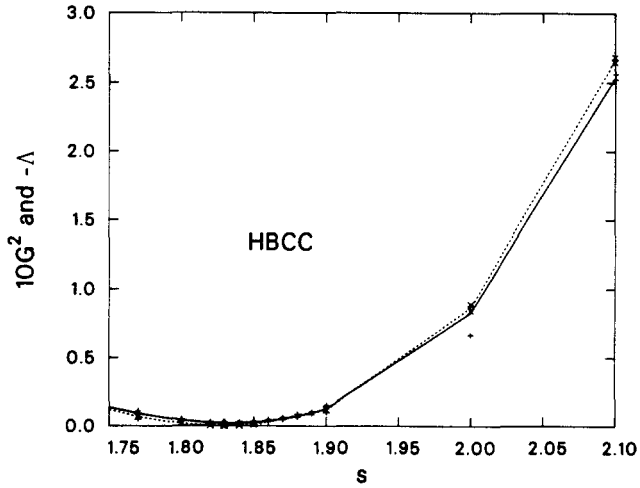


FIG. 6. The solid curve displays $-\Lambda(1)$ vs S for the HBCC lattice. The apparent errors are indicated by + signs. The dotted curve shows $10G(1)^2$, where $G(1)$ is as in Fig. 2 with the apparent error indicated by × signs.

lattice and Fig. 9 for the HSC lattice again with the apparent errors shown. As these figures show, the zero contour runs up to near $x = 1$ and then runs back to smaller x and, of course, leaves the physical region at $x = 0$, $S \simeq 3.5530$. The contour Fig. 9 for the HSC is much less well determined than that for the HBCC, but is easily consistent with the idea that the zero contour is tangent to the $x = 1$ line for the same value of S as that for which $G(1) = 0$. The contour Fig. 8 for the HBCC reaches a maximum for a value of S which agrees within error with that for which $G(1) = 0$. The value at this maximum falls a little short of the $x = 1$ line, but we are inclined to think, since the discrepancy is only a small multiple of the apparent error, that these results are not inconsistent with the idea that $y_0(S)$ is tangent to $x = 1$ at the S for which $G(1) = 0$.

We turn now to a study of c of Eq. (6.2). In Figs. 6 and 7 we show both $-\Lambda(1)$ and $10G(1)^2$ vs S for the HBCC and the HSC lattices. The correlation between these two curves is very strong. As pointed out in the previous section, mean-field theory predicts that $[\Lambda(x) + 10G^2(x)]$ vanishes to leading order. In order to investigate this question we have computed

$$c_1(x) = -[0.1\Lambda(x) + G^2(x)]/(1-x), \qquad (6.4)$$

which we have displayed in Fig. 10 as a function of x, together with the apparent errors for $S = 1.77$ on the HBCC lattice and $S = 1.80$ on the HSC lattice. The method of Padé analysis is as described above. A Baker–Hunter confluent singularity analysis[17] for $x = 1$ (HBCC) was also performed and, although not definitive, was consistent with the idea that at $x = 1$, $c_1(1)$ is negative and finite while giving some indication that $c_1(1)$ is approached like $-(1-x)^{0.3-1.0}$; in such a case $c_1(x)$ would hook upward and hence $c_1(1)$ is larger (c smaller) than estimated. This type of error would not change the local nature of the free energy and hence locally it would look like a critical point so long as $c_1(1) < \infty$. We show in Fig. 11 the behavior of $c_1(1)$ as a function of S. We conclude that as $c_1(1)$ is finite and as from a previous section $\chi \propto (1-x)^{-1}$ that c [Eq. (6.2)] is finite or at least we cannot, by our numerical analysis, reject the mean-field hypothesis.

It is somewhat surprising, but correct so far as we can tell numerically, that perfectly reasonable covergent results are obtained along the $x = 1$ ($\xi^2 = \infty$) line in the x-S plane both for $G(1) > 0$ and $G(1) < 0$. Since we have numerically verified that locally mean field theory predictions are valid, it appears that in addition as predicted by mean-field theory the first-order phase transition characterized by a line of triple points can be penetrated by analytic continuation to the spinodal curve, $\xi^2 = \infty$, lying behind it. In other words we see numerically no indications of corrections to leading-order mean-field behavior such as logarithms. This is not to say, of course, that there may not perhaps be nonanalytic corrections terms subdominate to the leading order behav-

TABLE III. Values of the Padé approximants to $\Lambda(1.0)$ for $S = 1.77$ on the HBCC lattice.

D \ N	2	3	4	5	6	7
2	−0.218 140	−0.055 599	0.193 108	−0.108 265	−0.101 721	−0.102 380[a]
3	0.087 791	−0.085 803	−0.105 176	−0.099 265	−0.088 513	−0.092 734
4	−0.135 023	−0.098 467	−0.100 577	−0.070 650	−0.092 447	
5	−0.065 031	−0.100 191	−0.099 197[a]	−0.093 200		
6	−0.109 939	−0.097 994	−0.107 859[a]			
7	−0.818 88	−0.094 291				

[a] Defective approximant.

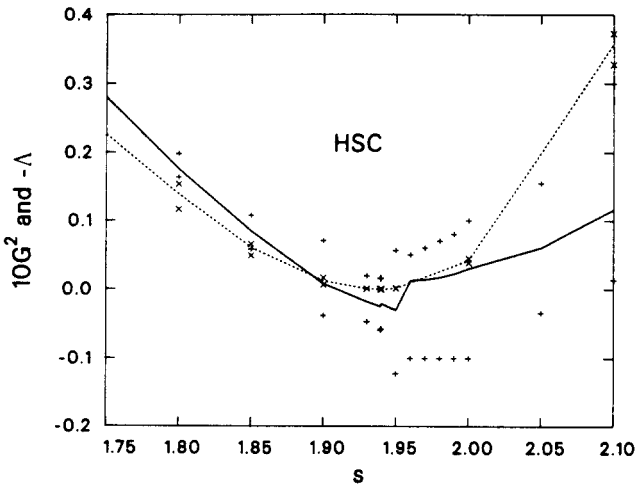


FIG. 7. The solid curve displays $-\Lambda(1)$ vs S for the HSC lattice. The apparent errors are indicated by + signs. The dotted curve shows $10G(1)^2$, where $G(1)$ is as in Fig. 2 with the apparent error indicated by × signs.

ior. These results fill in extra detail with regards to the mean field picture of the behavior of the "phantom field" model in the vicinity of the tricritical point, and we think add versimilitude to our picture.

VII. ANALYSIS OF $\chi(K,H)$, GLOBAL CONVEXITY

As pointed out in Sec. V the global convexity, and hence the nature of $K = K_c$, $H = 0$ as a critical, rather than a spinodal, point can be assured by showing that $\chi^{-1} > 0$ for $0 \leqslant K \leqslant K_c$, $0 < H \leqslant \infty$. In fact, by the use of a theorem of Ellis *et al.*,[10] we need only check H smaller than that given by (5.9). In this section we report our numerical calculations on this subject. Following the methods discussed in Sec. V, we have produced by computer manipulation from the results of Kincaid *et al.*[7] the coefficients of $\chi(K,H)$ as series in K whose coefficients are polynomials in the cumulants of the single-site spin distributions. For example the coefficient of K^{10} has 755 terms. The cumulants are expressed in terms of

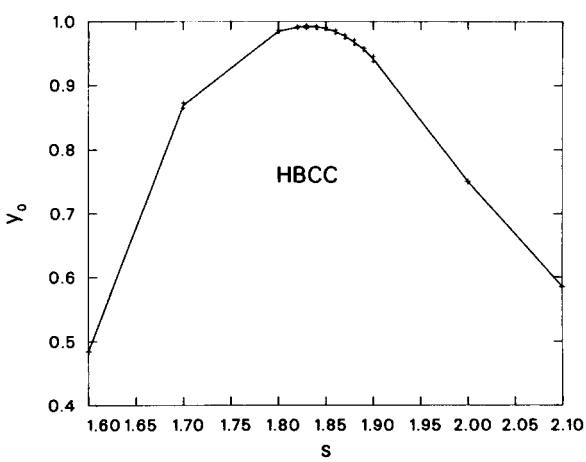


FIG. 8. The zero contour $y_0(S)$, for $\Lambda(x)$ in the x-S plane for the HBCC lattice. The apparent errors are indicated by + signs.

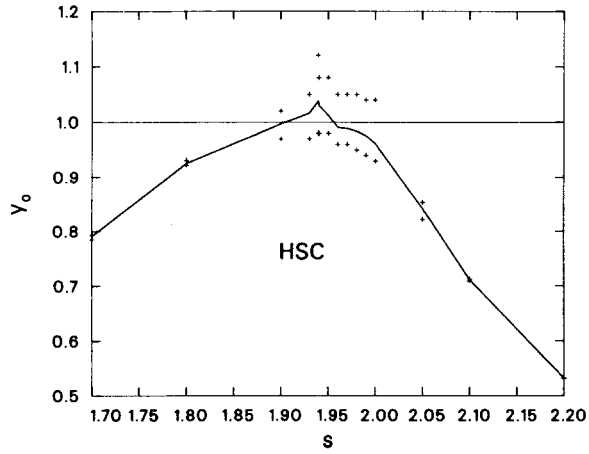


FIG. 9. The zero contour, $y_0(S)$, for $\Lambda(x)$ in the x-S plane for the HSC lattice. The apparent errors are indicated by + signs.

the moments in the usual way. If we use the spin distribution as given by Eq. (2.7) multipled by e^{sH} and the parameter τ, (5.14), which is convenient to express

$$\begin{aligned} \cosh Hs &= (1+\tau^2)/(1-\tau^2), \\ \sinh Hs &= 2\tau/(1-\tau^2), \end{aligned} \tag{7.1}$$

then the moments [Eq. (2.8)] are given by (5.13). The bound, (5.9), above which χ decreases monotonically with τ for all K becomes

$$\tau_0^2 = S^2(S^2-3)/(S^2+1)(S^2-2) \tag{7.2}$$

for $S^2 \geqslant 3$ and $\tau_0 = 0$ for $1 \leqslant S^2 \leqslant 3$. A sample of the HBCC series is

$$\begin{aligned} \chi(K,H) = \kappa_2 &+ 16(\kappa_2^2 + \kappa_1\kappa_3)K + 8(96\kappa_1\kappa_2\kappa_3 + 16\kappa_1^2\kappa_4 \\ &+ 32\kappa_2^3 + \kappa_3^2 + \kappa_2\kappa_4)K^2 + \cdots, \end{aligned} \tag{7.3}$$

where as usual the cumulants are

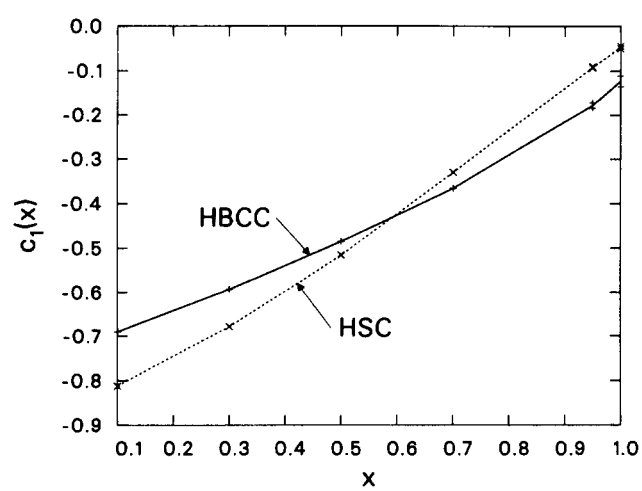


FIG. 10. The factor $c_1(x)$, Eq. (6.4), of the coefficient, c, of M^6 in the expansion (5.3) of the free energy is displayed as a function of x on the HBCC lattice for $S = 1.77$ and on the HSC lattice for $S = 1.80$. The apparent errors are indicated by + signs for the HBCC and × signs for the HSC.

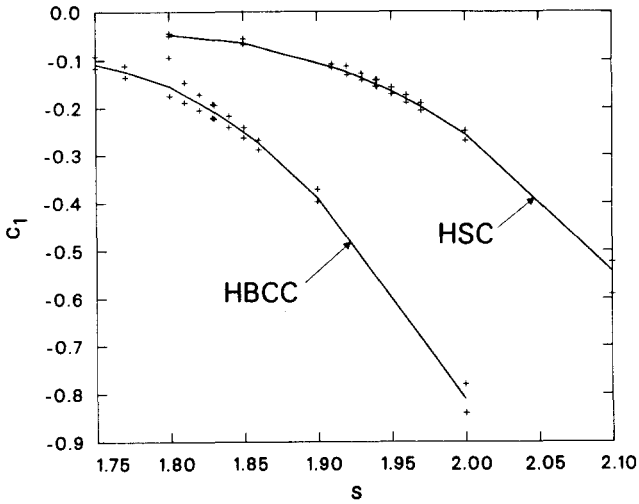


FIG. 11. The factor $c_1(1)$, Eq. (6.4), of the coefficient, c, of M^6 in the expansion (5.3) of the free energy is displayed as a function of S for the HBCC and HSC lattice. The apparent errors are indicated by + signs.

$$\kappa_1 = I_1, \quad \kappa_2 = I_2 - I_1^2, \quad \kappa_3 = I_3 - 3I_2I_1 + 2I_1^3, \qquad (7.4)$$
$$\kappa_4 = I_4 - 4I_3I_1 - 3I_2^2 + 12I_2I_1^2 - 6I_1^4, \ldots .$$

By the substitution of (5.13) into (7.4), we obtain the cumulants exactly for any desired S and τ. Then we substitute them in (7.3) and so obtain the K series for χ, given S and τ.

By way of orientation, we illustrate in Fig. 12 various expected contour plots of χ to show what we expect in various circumstances. In Fig. 12(a) we show the contour map of χ for $1 \leqslant S \leqslant \sqrt{3}$. In this region the GHS inequalities hold and χ is monotonically increasing for $\tau < 0$ and monotonically decreasing for $\tau > 0$. These conditions are considered normal and are unshaded. This picture holds for all dimensions greater than $d = 1$. In the case $d = 1$ there is, of course, no critical point nor phase boundary. The ridge line in the τ direction (dotted line) continues up the whole $\tau = 0$ line. In the region $\sqrt{3} < S$ we can show directly from the first term of $\partial\chi(K=0)/\partial H$ that there is a region along $v = 0$ line near $\tau = 0$, where χ increases for $\tau > 0$ and decreases for $\tau < 0$. This region is shaded in Fig. 12(b). The ridge line (in the τ direction) has a bifurcation point as shown and we expect, and can even prove for S an extremely small distance above $\sqrt{3}$, that there continues to be a normal critical point. The bifurcation point corresponds to the zero discussed previously in $\partial^2\chi(K,H=0)/\partial H^2$. When the zero moves out as S increases to coincide, at $S = S_t$, in location with the critical point, we expect a tricritical point to result and have shown it in Fig. 12(c). As S increases above S_t, Fig. 12(d), we expect a triple point with two-wing phase transition lines to occur. These two new phase transition lines should end in critical end points with infinite susceptibility. It is these critical end points that we seek in our global scan; because in the case that they occur the critical point which we had been following at $H = \tau = 0$ would lie behind the triple point and become a spinodal point for which the necessary properties to construct a field theory have not been demonstrated. Finally we illustrate, Fig. 12(e), the behavior for $d = 1$. We note that numerically a small discontinuity in the magnetization

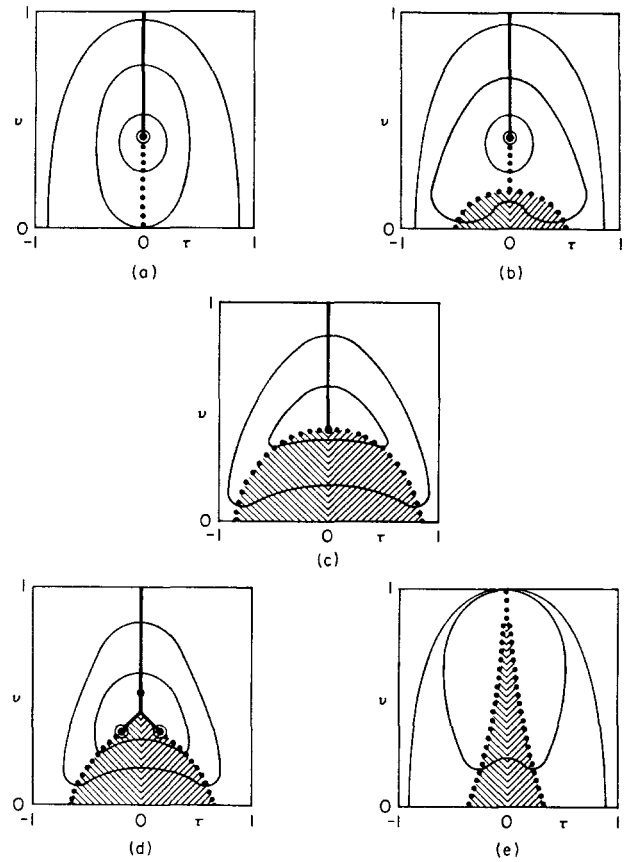


FIG. 12. Sketches of possible contour maps of the magnetic susceptibility χ for the Blume–Capel model. For $\tau < 0$ the unshaded portion is monotonically increasing and for $\tau > 0$ it is monotonically decreasing. The contours are light solid lines. First-order phase boundaries are heavy solid lines. Circled points are critical points. Dotted lines are the ridge lines in the τ direction, $v = \tanh K$. (a) This case corresponds to $1 \leqslant S \leqslant \sqrt{3}$. (b) This case corresponds to $\sqrt{3} < S < S_t$, where S_t is the tricritical point value. (c) $S = S_t$ the large dot is the tricritical point. (d) This case corresponds to $S > S_t$. The intersection of the three first-order phase boundaries is the triple point. The large dot is the spinodal point, which in mean field theory is reached by analytic continuation along the line $\tau = 0$ through the triple point into a metastable region. (e) The one-dimensional case.

M and a very high ridge in χ are rather hard to distinguish and so cases (b)–(d), and even (e), are hard to differentiate when S is near to S_t. Fortunately for the purposes of this paper such an effort is not required.

Our principle analysis of the behavior of $\chi(K,H)$ in this section is to consider the K series and to transform the K series (7.3) to a series in x as was done in Eq. (1.3). We mention explicitly that the correlation length used in the transformation is that for $H = 0$ so that $x = 1$ corresponds to $K = K_c$ (when appropriate) for all $-1 \leqslant \tau \leqslant 1$. We analyze both the x and K series by the method of Padé approximants. The results of these analyses are displayed as a contour map of χ in Fig. 13 for the HBCC lattice with $S = 1.77$ and in Fig. 14 for the HSC lattice with $S = 1.80$. The regions of uncertainty are shown by dashed lines. The curves in this region are approximated by the formula

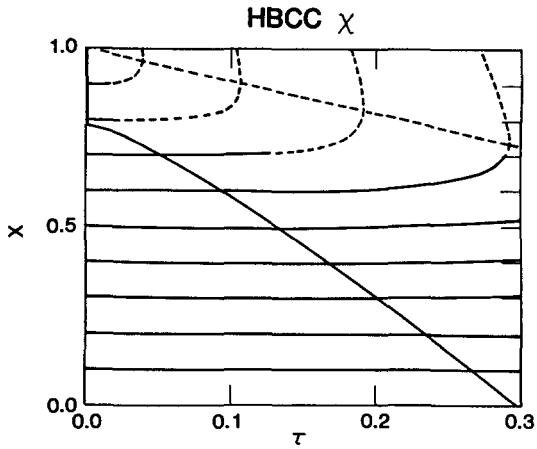


FIG. 13. A contour map of χ for $S = 1.77$ on the HBCC lattice. The solid contours represent our numerical results. The τ direction and K direction ridge lines are displayed. The dashed portions of these curves are included as an aid to the eye.

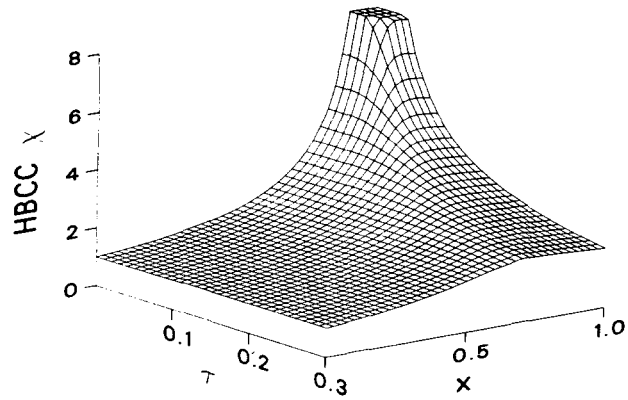


FIG. 15. The function $\chi(x,\tau)$ for $S = 1.77$ on the HBCC lattice is displayed as a three-dimensional plot. The large values near $\tau = 0, x = 1$ are truncated and displayed falsely as a flat top.

$$\chi(x,\tau) = \frac{a/(1-x)+b}{[1+c(x-d)[\tau^2/(1-x)^3]+f[\tau^4/(1-x)^6]]^{1/6}}, \tag{7.5}$$

for appropriate values of the constants $a,\ldots,f$. We are able to obtain sufficient precision to see that the contour plots are of type 12(b), which corresponds to a normal critical point and hence verifies that for those cases at least the continuum limit ($\xi^2 = \infty$) occurs, and so a field theory, in the sense discussed previously, can be constructed and is nontrivial (Sec. IV). A perhaps more revealing display of our results is the three-dimensional representation given in Fig. 15 of the same case as Fig. 13.

We have performed some additional analyses. In particular, a number of Padé analysis of the logarithm derivative of χ were performed, searching for possible singularities reflecting critical end points. We mention only a couple of these results. First for $S = 2.1$ on the HBCC we found a contour map like that of Fig. 12(d). The two wing phase transition lines were represented by a line of singularities that end somewhere between $\tau = 0.1$ and 0.2. For larger τ these singularities split and no longer appear for real K but move into the complex K plane in the vicinity of the ridge in the K direction. We have also examined $S = 1.85$ on the HBCC lattice, but it is too near to S_t for the contour plot to clearly resolve between the contour map types of Fig. 12(b), 12(c), or 12(d). The singularities bifurcating into the complex K plane near the K-ridge line is a general effect. It is well known in complex variable theory that at regular points analytic functions do not process maxima or minima. Consequently, when, for real values of K, $\chi(K,H)$ passes over a maximum, it must be a minimum in the imaginary direction at that point. It is not unusual for the function to then increase to a singular point at some nearby complex point. In the case of χ, since by the theory of functions of several complex variables the critical point singularity cannot simply disappear as τ changes, the above described bifurcation phenomena is the expected behavior for the cases of Figs. 12(a)–12(c). However, the representation of these complex singularities limits the useful convergence of Padé approximants to our ten-term series to about the distance of the K-ridge line. This limit is a "short series effect" and not an intrinsic one to the method (except in special cases).[19]

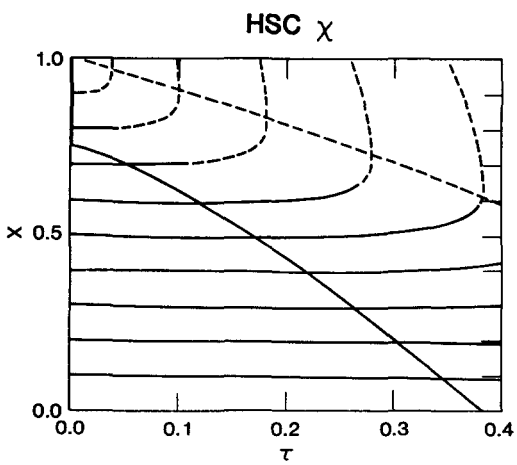


FIG. 14. A contour map of χ for $S = 1.80$ on the HSC lattice. The solid contours represent our numerical results. The τ direction and the K direction ridge lines are displayed. The dashed portions of these curves are included as an aid to the eye.

ACKNOWLEDGMENTS

The authors are grateful to J. Bricmont and C. M. Newman for helpful discussions.

Work for this paper was performed under the auspices of the U. S. Department of Energy.

[1] G. A. Baker, Jr. and J. D. Johnson, J. Phys. A **18**, L261 (1985).
[2] G. A. Baker, Jr., J. Math. Phys. **27**, 2379 (1986).
[3] C. A. de Carvalho, S. Caracciolo, and J. Frölich, Nucl. Phys. B **215** [F57], 209 (1983).
[4] M. Blume, Phys. Rev. **141**, 517 (1966).
[5] H. W. Capel, Physica **32**, 966 (1966).
[6] G. A. Baker, Jr. and J. M. Kincaid, J. Stat. Phys. **24**, 469 (1981).

[7]J. M. Kincaid, G. A. Baker, Jr., and L. W. Fullerton, "High-Temperature series expansions of the continuous spin Ising model," Los Alamos report LA-UR-79-1575, 1979.
[8]J. D. Johnson and G. A. Baker, Jr., "High-temperature series expansion of $\partial^4\chi/\partial H^4$ for a general continuous-spin Ising model," Los Alamos report LA-10275-MS, 1984.
[9]M. Blume, V. J. Emery, and R. B. Griffiths, Phys. Rev. A **4**, 1071 (1971).
[10]R. S. Ellis, C. M. Newman, and M. R. O'Connell, J. Stat. Phys. **26**, 37 (1981).
[11]C. M. Newman, J. Stat. Phys. **27**, 836 (1982); (private communication).
[12]G. A. Baker, Jr. and D. Kim, J. Phys. A **13**, L103 (1980).
[13]S. N. Isakov, Commun. Math. Phys. **95**, 427 (1984).
[14]G. A. Baker, Jr., *Phase Transitions and Critical Phenomena*, edited by C. Domb and J. L. Lebowitz (Academic, London, 1984), Vol. 9, pp. 233–311.
[15]E. H. Lieb and A. D. Sokal, Commun. Math. Phys. **80**, 153 (1981).
[16]J. Ginibre, Commun. Math. Phys. **16**, 310 (1970).
[17]G. A. Baker, Jr. and D. L. Hunter, Phys. Rev. B **7**, 3377 (1973).
[18]G. A. Baker, Jr., J. Phys. A **17**, L621 (1984).
[19]G. A. Baker, Jr. and P. R. Graves-Morris, *Padé Approximants Parts I and II, Encyclopedia of Mathematics and Its Applications*, edited by G. C. Rota (Addison–Wesley, Reading, MA, 1981), Vols. 13 and 14.
[20]D. L. Hunter and G. A. Baker, Jr., Phys. Rev. B **7**, 3346 (1973).
[21]I. D. Lawrie and S. Sarbach, in *Phase Transitions and Critical Phenomena*, edited by C. Domb and J. L. Lebowitz (Academic, London, 1984), Vol. 9, pp. 1–161.
[22]D. Brydges, J. Frölich, and T. Spencer, Commun. Math. Phys. **83**, 123 (1982).
[23]J. Frölich, Nucl. Phys. B **200** [FS4], 281 (1982).
[24]T. Hara, T. Hattori, and H. Tasaki, J. Math. Phys. **26**, 2922 (1985).
[25]A. D. Sokal, Ann. Inst. H. Poincaré **37**, 317 (1982).
[26]J. Glimm and A. Jaffee, *Recent Developments in Gauge Theories*, Cargese, 1979, edited by G. 't Hooft, C. Itzykson, A. Jaffe, H. Lehmann, P. K. Mitter, I. M. Singer, and R. Stora (Plenum, New York, 1980). Their arguments are stated for ϕ^4 theory only but are equally valid for the present case.

A few remarks on the paper "Necessary versus sufficient conditions for exact solubility of statistical models on lattices" [J. Math. Phys. 27, 593 (1986)]

Eugene Gutkin
Department of Mathematics, University of Southern California, Los Angeles, California 90089-1113

(Received 6 June 1986; accepted for publication 10 December 1986)

Lochak and Maillard [J. Math. Phys. **27**, 593 (1986)] claim that the Baxter condition, which was known to be sufficient for the commutativity of transfer matrices, is also necessary (under some additional technical assumptions). Although the claim is correct, the proof in that paper is false. In this paper the errors of Lochak and Maillard are pointed out and correct proofs are outlined.

I. INTRODUCTION

The paper of Lochak and Maillard[1] addresses the question of solubility of lattice models of statistical mechanics (cf. Ref. 2). For the benefit of the reader, we recall the basic definitions here. The paper[1] discusses two kinds of such models: the vertex and the spin models with the nearest neighbor interactions. Both types live on square lattices with periodic boundary conditions. Let $L(M,N)$ be such a lattice of width M and height N.

II. VERTEX MODELS

Let m and n be positive integers. A vertex model is determined by m^2n^2 numbers $u(i,j|k,l)$, $1\leqslant i,j\leqslant m$, $1\leqslant k,l\leqslant n$, which are called Boltzmann weights. The Boltzmann weight $u(i,j|k,l)$ corresponds to an elementary configuration of spins in the vertex model [see Fig. 1(a)]. To any configuration ω of spins on the edges of $L(M,N)$ we assign the number

$$u(\omega)=\prod_{i,j,k,l} u(i,j|k,l),$$

where the product is taken over all elementary configurations contained in ω. The partition function is given by

$$F_{M,N}(u)=\sum_{\omega} u(\omega),$$

where the summation is over all possible configurations ω and the argument u of F stands for the m^2n^2 vector $u(i,j|k,l)$ of Boltzmann weights.

It is standard to express the partition function $F_{M,N}(u)$ in terms of the (row-to-row) transfer matrix $T_M(u)$ of the model which acts on the space $\otimes^M \mathbb{C}^n$. The matrix elements

$$T_M(u)^{l_1,\ldots,l_n}_{k_1,\ldots,k_n}$$

of $T_M(u)$ are expressed via Boltzmann matrices $U(k,l)$ acting on $\mathbb{C}^m$, where k and l run from 1 to n. The matrix elements $U(k,l)^j_i$ are given by $U(k,l)^j_i=u(i,j|k,l)$ and omitting M and u from $T_M(u)$ we have

$$T^{l_1,\ldots,l_M}_{k_1,\ldots,k_M}=\operatorname{tr}[U(k_1,l_1)\cdots U(k_M,l_M)]. \tag{1}$$

Then

$$F_{M,N}(u)=\operatorname{tr} T_M(u)^N. \tag{2}$$

We summarize the construction above. We are given a set $u(i,j|k,l)$, $1\leqslant i,j\leqslant m$, $1\leqslant k,l\leqslant n$, of Boltzmann weights denoted for brevity by $u\in\mathbb{R}^{m^2n^2}$. We arrange them as n^2 Boltz-

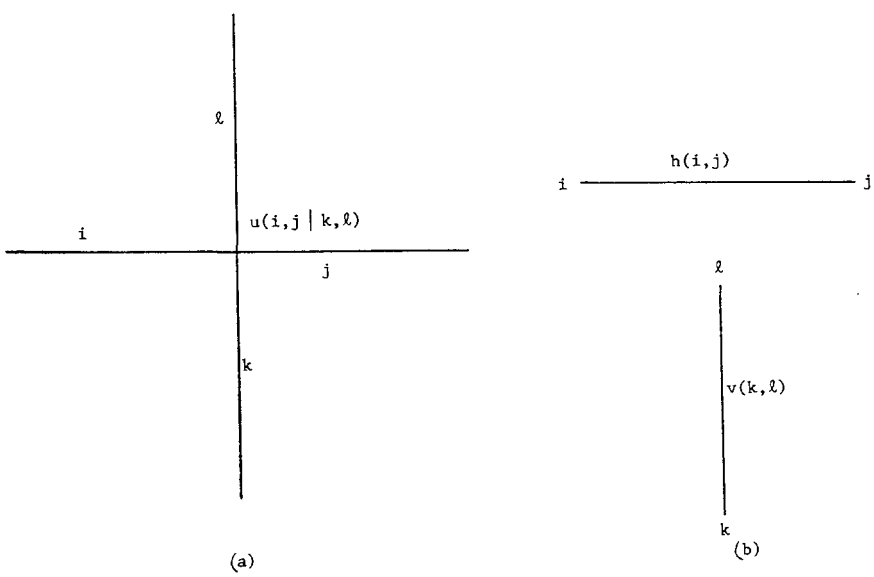


FIG. 1. Boltzmann weights of elementary configurations in (a) vertex model and (b) spin model.

 0022-2488/87/051159-05$02.50

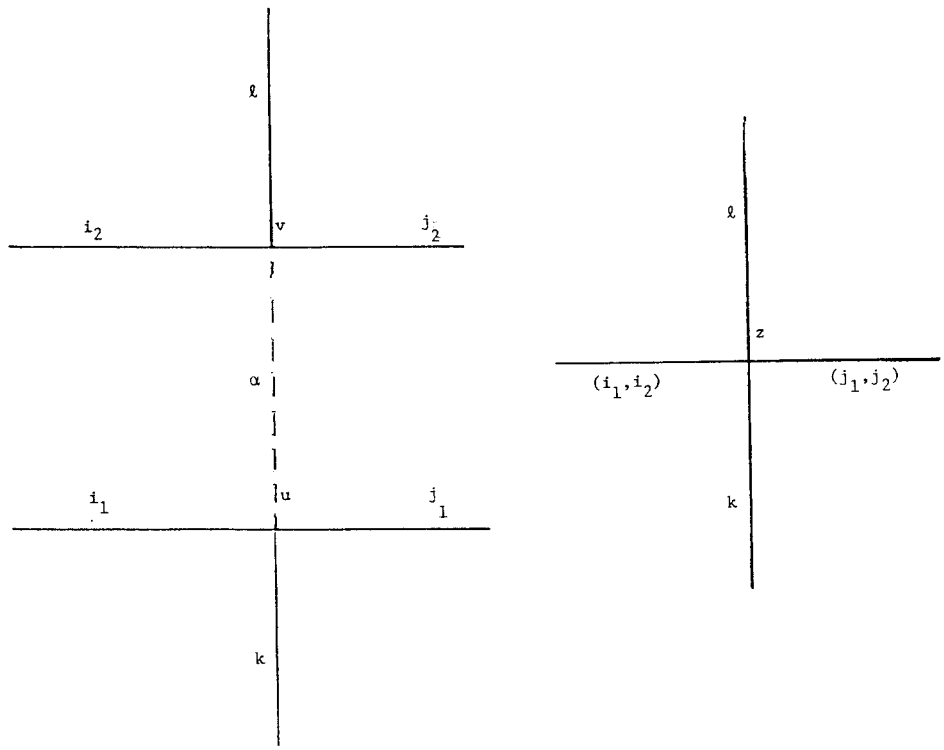


FIG. 2. The Boltzmann weights u,v yield, after summation over α, the Boltzmann weight z.

mann matrices $U(k,l)$ of order $m\times m$. From those we construct by (1) the transfer matrix $T=T_M(u)$ acting on $\otimes^M\mathbb{C}^n$, which determines, by (2), the partition function $F_{M,N}(u)$.

Solubility of lattice models is known to be related to the question whether the transfer matrices $T_{M,N}(u)$ and $T_{M,N}(v)$ corresponding to two sets of Boltzmann weights $u,v\in\mathbb{R}^{m^2n^2}$ commute for all M and N (see Ref. 2). Reference 1 discusses this latter question. We fix M and N and omit them for the moment from the notation $T_{M,N}(u)$. It is known (cf. Ref. 3) that $T(u)T(v)$ is itself a transfer matrix $T(z)$ corresponding to the set $z\in\mathbb{R}^{m^4n^2}$ of Boltzmann weights obtained from $u,v\in\mathbb{R}^{m^2n^2}$ by

$$z(i_1,i_2;j_1,j_2|k,l)=\sum_{\alpha=1}^{n}u(i_1,j_1|k,\alpha)v(i_2,j_2|\alpha,l) \tag{3}$$

(see Fig. 2 for an illustration). Analogously, $T(v)T(u)=T(w)$, where the coordinates of the Boltzmann vector $w\in\mathbb{R}^{m^4n^2}$ are given by

$$w(i_1,i_2;j_1,j_2|k,l)=\sum_{\alpha=1}^{n}v(i_1,j_1|k,\alpha)u(i_2,j_2|\alpha,l). \tag{4}$$

We organize the vectors z and w into n^2 Boltzmann matrices $Z(k,l)$ and $W(k,l)$, respectively, acting on the space $\mathbb{C}^{m^2}$. In view of (1), the equation $T(u)T(v)=T(v)T(u)$ will be satisfied if there exists a nondegenerate matrix R on $\mathbb{C}^{m^2}$ such that for all k and l,

$$RZ(k,l)=W(k,l)kR. \tag{5}$$

Thus (5), which I call Baxter's condition,[3] because it is closely related to the star-triangle relation of Baxter's,[2] is sufficient for the commutativity of $T_M(u)$ and $T_M(v)$ for all M.

The first goal of Ref. 1 is to show that (under some technical assumption) (5) is also necessary. Although the assertion is correct, its proof in Ref. 1 is erroneous and I explain the error here.

Denote for brevity m^2 by d. The equation

$$T_M(u)T_M(v)=T_M(v)T_M(u) \tag{6}$$

implies, by (1), that the two sets of matrices $Z(k,l)$ and $W(k,l)$ on $\mathbb{C}^d$, $1\leqslant k,l\leqslant n$, satisfy

$$\mathrm{tr}[Z(k_1,l_1)\cdots Z(k_M,l_M)]=\mathrm{tr}[W(k_1,l_1)\cdots W(k_M,l_M)] \tag{7}$$

for any set $k_1,\ldots,k_M,l_1,\ldots,l_M$ of indices. To show the necessity of Baxter's condition we need to prove that (7) implies (5). The authors of Ref. 1 understand it. After stating this, they say that the proof that (7) implies (5) "is easily seen to be reduced" to the following assertion (see Ref. 1, Theorem 1).

Assertion 1: Let A and A' be two subalgebras of the full matrix algebra $L_d(\mathbb{C})$ of operators on $\mathbb{C}^d$. Suppose that there is no nontrivial subspace of $\mathbb{C}^d$ invariant under A (this is the technical assumption I mentioned before). Suppose further that there is an algebra homomorphism $\varphi: A\to A'$ which is onto and satisfies

$$\mathrm{tr}(a_1\cdots a_k)=\mathrm{tr}(\varphi(a_1\cdots a_k)) \tag{8}$$

for any $a_1,\ldots,a_k\in A$. Then there is a nondegenerate matrix $R\in L_d(\mathbb{C})$ such that for all $a\in A$

$$\varphi(a)=RaR^{-1}. \tag{9}$$

Then Ref. 1 proceeds to prove this assertion. The first point is that Assertion 1 is true even without (8), which should be a part of the conclusion and not of the assumption. The proof is as follows. Since elements of A have no common invariant

subspace, by Burnside's Theorem (cf. Ref. 4, p. 182), A is the full matrix algebra $L_d(\mathbb{C})$. By Skolem's Theorem (cf. Ref. 5, p. 99), any nonzero homomorphism φ of the full matrix algebra to itself is inner, i.e., there is an invertible matrix R such that $\varphi(a) = RaR^{-1}$. In particular, the assumption of Ref. 1 that $\varphi: A \to A'$ is onto is also erroneous. It should be replaced by $\varphi \neq 0$. Since φ is inner, for any $a \in A$ (in particular for $a = a_1 \cdots a_k$) we have $\mathrm{tr}(\varphi(a)) = \mathrm{tr}(RaR^{-1}) = \mathrm{tr}\, a$. This proves the following proposition, which is the correct version of Assertion 1.

Proposition 1: Let $A \subset L_d(\mathbb{C})$ be an irreducible subalgebra and let φ by a nonzero homomorphism $\varphi: A \to L_d(\mathbb{C})$. Then there exists an invertible matrix R such that $\varphi(a) = RaR^{-1}$ and therefore $\mathrm{tr}\,\varphi(a) = \mathrm{tr}\, a$.

To summarize, Proposition 1 which is the corrected version of Theorem 1 of Ref. 1 is trivial. The second point is that even after this correction, the necessity of Baxter's condition, i.e., the implication (7) → (5), is not proved in Ref. 1 because its reduction to Proposition 1 is not there. Actually, for a good reason, because the implication (7) → (5) does not reduce to Proposition 1. The implication is proved in Ref. 3, Theorem 1, where it is the main part of the proof. For the convenience of the reader I state here the assertion that is missing in Ref. 1 and outline the crucial point in the proof.

Proposition 2: Let a_i and a_i', $i \in I$, a finite set of indices, be two sets of operators on $\mathbb{C}_d$. Let for any $i_1, \ldots, i_n \in I$,

$$\mathrm{tr}(a_{i_1} \cdots a_{i_n}) = \mathrm{tr}(a'_{i_1} \cdots a'_{i_n}). \tag{10}$$

Denote by A and A' the matrix algebras generated by a_i and a_i', $i \in I$, respectively.

Assume that the operators $\{a_i, i \in I\}$ (or $\{a_i', i \in I\}$) have no nontrivial common invariant subspace in $\mathbb{C}^d$. Then the correspondence $a_i' = \varphi(a_i)$, $i \in I$, continues to the algebra homomorphism $\varphi: A \to A'$.

Outline of Proof: Denote by $\underline{A}$ and $\underline{A}'$ the abstract free algebras generated by the symbols a_i and a_i', $i \in I$, respectively. The correspondence $a_i' = \varphi(a_i)$ defines the isomorphism $\underline{\varphi}: \underline{A} \to \underline{A}'$. The matrix algebras A and A' are the quotients of $\underline{A}$ and $\underline{A}'$ by their ideals, i.e., $A = \underline{A}/J$, $A' = \underline{A}'/J'$. The isomorphism $\underline{\varphi}$ descends to a homomorphism $\varphi: A \to A'$ of matrix algebras if and only if $\underline{\varphi}(J) \subset J'$. The proof of this inclusion is rather intricate (see the proof of Theorem 1 in Ref. 3) and this is where Eq. (10) and the invariant subspace assumption have to be used. The assertion of Proposition 2 fails if we omit either one of these assumptions (examples are easy to construct).

For the reader's convenience we state here the theorem on Baxter's commutativity condition for vertex models whose proof was the subject of the preceding discussion.

Theorem 1: Consider two vertex models on the square lattice given by their Boltzmann weights $u(i,j|k,l)$ and $v(i,j|k,l)$, where $1 \leqslant i,j \leqslant m$, $1 \leqslant k,l \leqslant n$. Define n^2 matrices $Z(k,l)$ and $W(k,l)$, $1 \leqslant k,l \leqslant n$, on $\mathbb{C}^{m^2}$ by (3) and (4), respectively.

Assume that the n^2 operators $Z(k,l)$ [or $W(k,l)$] have no nontrivial common invariant subspace. Then the Baxter condition (5) is *necessary and sufficient* for the transfer matrices $T(u)$ and $T(v)$ to commute [see Eq. (6)].

The invariant subspace assumption, although generically satisfied, is very hard to check. It can be replaced by a certain symmetry assumption on the Boltzmann weights.[6]

Theorem 1 has a close analog for spin models which we discuss in the next section.

III. SPIN MODELS

A spin model on $L(M,N)$ with n spin states is determined by $2n^2$ Boltzmann weights $h(i,j)$, $v(k,l)$, $1 \leqslant i,j,k,l \leqslant n$, where $h(i,j)$ and $v(k,l)$ correspond to elementary configurations of spins in the model [see Fig. 1(b)]. Denoting by $u \in \mathbb{R}^{2n^2}$ the vector determined by the Boltzmann weights we define the partition function $F_{M,N}(u)$ of a spin model analogously to that of a vertex model.

One can define (in many ways) a transfer matrix $T(u)$ of a spin model so that $F_{M,N}(u)$ is obtained from $T(u)$ by a formula similar to (2).

Solubility of a spin model is again related to the question when $T(u)$ and $T(u')$ commute for all M,N (see Ref. 2). Depending on a particular choice of the transfer matrix T one can define sufficient conditions for commutativity of $T(u)$ and $T(u')$ similar to (5) (we call them Baxter conditions) and ask whether they are necessary.

We remark first that for any spin model S one can canonically define a vertex model V such that S and V have the same partition function (we say that S and V are equivalent, see Ref. 7). This observation makes a separate consideration of spin models in some sense redundant.

Using the general correspondence $S \to V$ of Ref. 7 may not be practical but if $N = M = 2K$ there is a simple way to transform a spin model into a vertex model (which is presumably known to the experts). Namely, we rotate the lattice $L(2K,2K)$ by 45 deg so that the two diagonals of the original lattice become horizontal and vertical, respectively, in the new lattice L'. Consider separately the even and the odd numbered horizontal rows of the new lattice (see Fig. 3 and Ref. 2, Fig. 7.1). Figure 3 shows the transformation of the spin model S on L' into a vertex model V on the lattice $L(2K,K)$. The number n of spin states of V and S is the same and the Boltzmann weights of V are given by $u(i,j|k,l) = v(i,l)h(l,j)v(k,j)h(i,k)$.

Let us nevertheless consider the row-of-spins to row-of-spins transfer matrix T for a spin model. It acts on $\otimes^M \mathbb{C}^n$ and its matrix elements are given by

$$T^{j_1,\ldots,j_M}_{i_1,\ldots,i_M} = v(i_1,j_1)\cdots v(i_M,j_M) \times [h(i_1,i_2)h(j_1,j_2)\cdots h(i_M,i_1)h(j_M,j_1)]^{1/2}. \tag{11}$$

Then the partition function is given by $F = \mathrm{tr}\, T^N$. Given another pair (h',v') of Boltzmann weights we have two transfer matrices $T(h,v) = T$ and $T(h',v') = T'$.

We define n^4 matrices $a(\alpha,\beta|\gamma,\delta)$ on $\mathbb{C}^n$, $1 \leqslant \alpha,\beta,\gamma,\delta \leqslant n$ by

$$a(\alpha,\beta|\gamma,\delta) = \sum_{p,q=1}^{n} h(\alpha,\beta)^{1/2} v(\alpha,p)v(\beta,q)h(p,q)^{1/2} \times h'(p,q)^{1/2} v'(p,\gamma)v'(q,\delta)h'(\gamma,\delta)^{1/2}. \tag{12}$$

Then

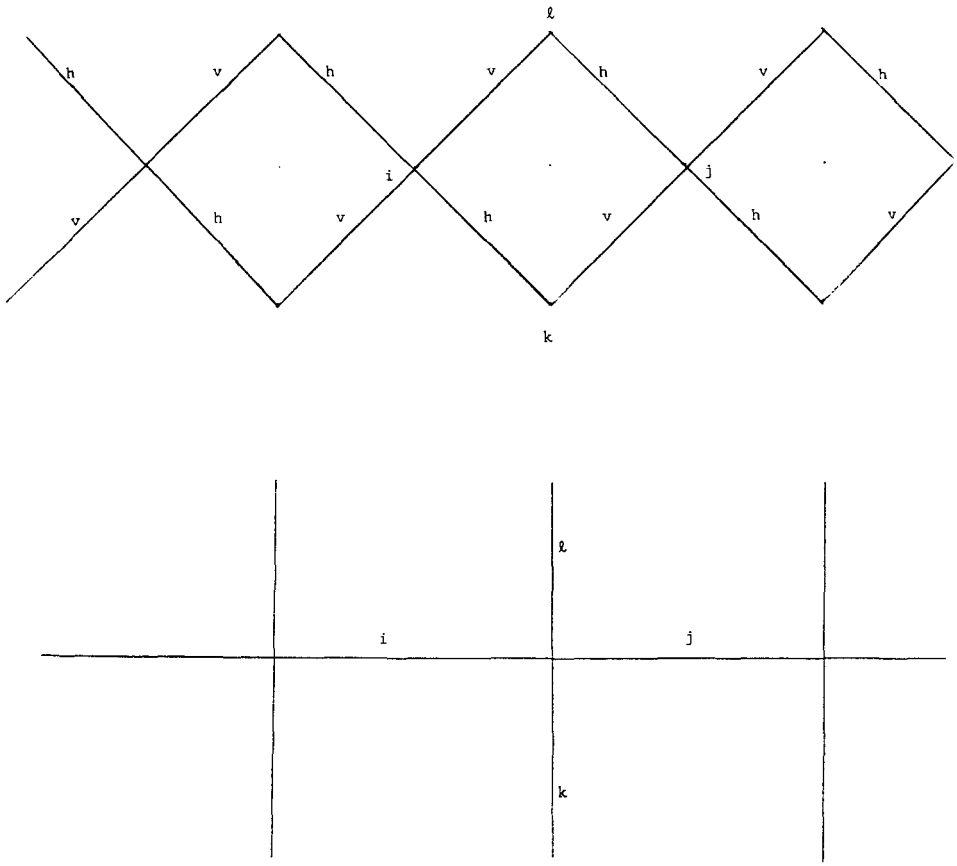


FIG. 3. Transformation of a spin model (above) into an equivalent vertex model (below).

$$(TT')^{j_1,\dots,j_M}_{i_1,\dots,i_M} = \mathrm{tr}[a(i_1,j_1|i_2,j_2) \times a(i_2,j_2|i_3,j_3)\cdots a(i_M,j_M|i_1,j_1)]. \quad (13)$$

Switching T' and T around we obtain another n^4 matrices $b(\alpha,\beta|\gamma,\delta)$ such that

$$(T'T)^{j_1,\dots,j_M}_{i_1,\dots,i_M} = \mathrm{tr}[b(i_1,j_1|i_2,j_2)\cdots b(i_M,j_M|i_1,j_1)]. \quad (14)$$

Denote for brevity (α,β) by i, $1\leqslant i\leqslant n^2$. The argument above shows the following.

Proposition 3: Let S and S' be two spin models on the square lattice L with n spin states given by Boltzmann weights (h,v) and (h',v'), respectively. Let T and T' be their row-to-row transfer matrices. Equation (12) defines two sets $a(i,j)$ and $b(i,j)$ of matrices on $\mathbb{C}^n$, $1\leqslant i,j\leqslant n^2$, whose elements depend on h,v,h',v' such that T and T' commute for any size of L if and only if for any indices $i_1,\dots,i_M$,

$$\mathrm{tr}[a(i_1,i_2)a(i_2,i_3)\cdots a(i_M,i_1)] = \mathrm{tr}[b(i_1,i_2)b(i_2,i_3)\cdots b(i_M,i_1)]. \quad (15)$$

Now let I be any finite set of indices and let $a(i,j)$ and $b(i,j)$, $i,j\in I$, be two sets of matrices on $\mathbb{C}^n$. Assume that there are nondegenerate operators $R_i\in L_n(\mathbb{C})$, such that for any $i,j\in I$,

$$b(i,j) = R_i a(i,j) R_j^{-1}. \quad (16)$$

Let us call (16) the Baxter condition (in this setting). The obvious implication (16) $\rightarrow$ (15) means that Baxter condition is sufficient for the commutativity of T and T' discussed earlier. The following is a close analog of Theorem 1 in the context of spin models (we use the notation of Proposition 3).

Theorem 2: Assume that for some $k\in I$ either the set $\{a(k,i_1)\cdots a(i_{M-1},i_M)a(i_M,k)\colon i_1,\dots,i_M\in I\}$ or the set $\{b(k,i_1)\cdots b(i_M,k)\colon i_1,\dots,i_M\in I\}$ do not have a nontrivial invariant subspace in $\mathbb{C}^n$ and that the matrices $a(k,i)$, $a(i,k)$, $b(k,i)$, $b(i,k)$ are invertible for all $i\in I$. Then the transfer matrices T and T' commute if and only if the Baxter condition (16) holds.

In view of the discussion above, it suffices to show that (15) implies (16). The argument below deduces the implication from Propositions 1 and 2. This argument is similar to the argument of Theorem 2 in Ref. 1. Our argument is simpler and our statement is more general. Essentially, Theorem 2 of Ref. 1 is the special case of our theorem corresponding to $n = 2$.

Proof of Theorem 2: Assume for concreteness that $k = 1$. Applying Propositions 1 and 2 to the sets $\{a(1,i_1)\cdots a(i_M,1)\colon i_1,\dots,i_M\in I\}$ and $\{b(1,i_1)\cdots b(i_M,1)\colon i_1,\dots,i_M\in I\}$, we obtain that there is an operator $R_1\in GL_n(\mathbb{C})$ such that for any $i_1,\dots,i_M\in I$,

$$b(1,i_1)\cdots b(i_M,1) = R_1 a(1,i_1)\cdots a(i_M,1) R_1^{-1}. \quad (17)$$

For any $i\in I$ set

$$R_i = b(1,i)^{-1} R_1 a(1,i). \quad (18)$$

As a special case of (17) we have for any $j\in I$

$$b(1,j)b(j,1) = R_1 a(1,j)a(j,1) R_1^{-1}, \quad (19)$$

which implies

$$b(j,1)R_1a(j,1)^{-1} = b(1,j)^{-1}R_1a(1,j) = R_j. \quad (20)$$

now for any $i,j \in I$,

$$\begin{aligned} b(i,j) &= b(1,i)^{-1}[b(1,i)b(i,j)b(j,1)]b(j,1)^{-1} \\ &= [b(1,i)^{-1}R_1a(1,i)]a(i,j)[a(j,1)R_1^{-1}b(j,1)^{-1}] \\ &= R_ia(i,j)R_j^{-1}. \end{aligned}$$

The theorem is proved.

ACKNOWLEDGMENT

The author was partially supported by NSF Grant No. DMS 84-03238.

[1] P. Lochak and J. M. Maillard, "Necessary versus sufficient conditions for exact solubility of statistical models on lattices," J. Math. Phys. **27**, 593 (1986).

[2] R. J. Baxter, *Exactly Solved Models in Statistical Mechanics* (Academic, New York, 1982).

[3] E. Gutkin, "A comment on Baxter condition for commutativity of transfer matrices," J. Stat. Phys. **44**, 193 (1986).

[4] C. W. Curtis and I. Reiner, *Representation Theory of Finite Groups and Associative Algebras* (Wiley, New York, 1962).

[5] I. N. Herstein, "Noncommutative rings," in *Carus Mathematical Monographs* (Wiley, New York, 1968).

[6] E. Gutkin, "Another comment on the Baxter condition for commutativity of transfer matrices," submitted to Phys. Lett. A.

[7] E. Gutkin, "Equivalence of lattice models of statistical mechanics," submitted to Physica A.

On the motion of the first Lee–Yang zero

Hal Tasaki[a)]
Department of Physics, Faculty of Science, University of Tokyo, Hongo, Tokyo 113, Japan

(Received 6 August 1986; accepted for publication 31 December 1986)

In the d-dimensional ferromagnetic Ising and φ^4 systems, the motion of the first Lee–Yang zero when the lattice size tends to infinity is studied. In particular, a power law behavior at the critical point, which is distinct from those in noncritical points, is elucidated.

I. INTRODUCTION

The present paper is devoted to a problem concerning the Lee–Yang zeros of the partition function of the Ising and φ^4 ferromagnets.

Since the pioneering works of Lee and Yang,[1] which include the proof of the Lee–Yang circle theorem, there have been published considerably many works related to the Lee–Yang zeros. Among them are extensions of the circle theorem to other systems,[2,3] studies of the basic structure of the zeros,[4] and its applications in deriving correlation inequalities,[5] normal fluctuation theorems,[6] or phenomenological scaling arguments.[7] It is, however, surprising to find that our knowledge on the actual behavior (or motion) of the Lee–Yang zeros in specific models is still not so rich.

We concentrate here on the systems in the finite d-dimensional hypercubic lattice, and study *how the location of the first Lee–Yang zero behaves* when the size of the lattice tends to infinity. We find, at the critical point, that the motion of the zero is governed by a power law, distinct from that in high and low temperature phases. (See Corollaries 4 and 5.) Our proof is based on the various correlation inequalities.[8–14]

II. RESULTS AND PROOFS

Consider a finite d-dimensional hypercubic lattice $\Lambda(L)=\{-L/2,-L/2+1,\ldots,L/2\}^d$ with a *free boundary condition*. The size L, and the number of the sites L^d of the lattice play central roles throughout the present paper. The Ising (and φ^4) models on $\Lambda(L)$, with inverse temperature J and vanishing external field, are described by the following thermal expectation:

$$\langle\cdots\rangle_J^L = Z(J,L)^{-1}\int\prod_{x\in\Lambda(L)}d\nu(\varphi_x)\,e^{-J\mathscr{H}(L)}\quad \langle 1\rangle_J^L=1,$$
$$\mathscr{H}(L) = -\frac{1}{2}\sum_{\substack{|x-y|=1\\ x,y\in\Lambda(L)}}\varphi_x\varphi_y,\quad J\geqslant 0,\quad \varphi_x\in\mathbb{R}, \tag{1}$$

where $d\nu(\varphi)$ is a single site measure. Here we mainly consider the Ising model with $d\nu(\varphi)=\delta(\varphi^2-1)d\varphi$. The φ^4 model is defined by $d\nu(\varphi)=\exp(-\mu\varphi^2-\lambda\varphi^4)d\varphi$, with $\mu\in\mathbb{R}$ and $\lambda>0$.

Let the *partition function* for the system under complex external field z be defined by

$$f(z;L,J)=\left\langle\exp\left(z\sum_{x\in\Lambda(L)}\varphi_x\right)\right\rangle_J^L,\quad z\in\mathbb{C}. \tag{2}$$

By the *Lee–Yang zeros*, we denote $z\in\mathbb{C}$ which satisfies

$$f(z;L,J)=0. \tag{3}$$

The *Lee–Yang theorem*[1,3] states that these zeros are contained in the region $\mathrm{Re}(z)=0$. Let $z=\pm i\alpha_1(L,J)$ be the *first Lee–Yang zeros*, i.e., the zeros of $f(z;L,J)$ nearest to the origin $z=0$. To clarify the behavior of $\alpha_1(L,J)$ when L tends to infinity, for various values of J, is the main interest of the present paper.

Let us define the quantities χ, s_2, and u_4 as the following[5]:

$$\chi(L,J)=\sum_{x\in\Lambda(L)}\langle\varphi_0\varphi_x\rangle_J^L, \tag{4}$$

$$s_2(L,J)=\sum_{x,y\in\Lambda(L)}\langle\varphi_x\varphi_y\rangle_J^L, \tag{5}$$

where $0=(0,0,\ldots,0)$, and

$$u_4(L,J)=\sum_{x,y,z,w\in\Lambda(L)}U_4(x,y,z,w). \tag{6}$$

Here the four-point Ursell function (or cumulant) $U_4(x,y,z,w)$ is defined by

$$U_4(x,y,z,w)=\langle\varphi_x\varphi_y\varphi_z\varphi_w\rangle-\langle\varphi_x\varphi_y\rangle\langle\varphi_z\varphi_w\rangle - \langle\varphi_x\varphi_z\rangle\langle\varphi_y\varphi_w\rangle-\langle\varphi_x\varphi_w\rangle\langle\varphi_y\varphi_z\rangle.$$

Note that, by Griffith's inequalities,[12] χ and s_2 are related by the inequalities

$$(L/2)^d\chi(L/2,J)\leqslant s_2(L,J)\leqslant L^d\cdot\chi(2L,J). \tag{7}$$

The following Lemma is among the main tools in the present investigation.

Lemma 1: For arbitrary real k, $J\geqslant 0$, and L, we have the following inequality:

$$f(ik;L,J)-\exp(-s_2(L,J)\cdot k^2)\geqslant -\tfrac{1}{16}|u_4(L,J)|\cdot k^4\cdot\cosh([s_2(L,J)/2]k^2). \tag{8}$$

Proof: Let us expand $f(ik)$ as

$$f(ik)=\sum_{n=0}^{\infty}\frac{(-k^2)^n}{(2n)!}s_{2n},$$

where

$$s_{2n}(L,J)=\sum_{\substack{x_i\in\Lambda(L)\\ (i=1,\ldots,2n)}}\langle\varphi_{x_1}\cdots\varphi_{x_{2n}}\rangle.$$

Applying the Gaussian inequality[13] and Aizenman's inequality (Proposition 12.1 of Ref. 8) to s_{2n}, we find

[a)] New address: Department of Physics, Princeton University, Princeton, New Jersey 08544.

 0022-2488/87/051164-03$02.50

$$0 \geqslant s_{2n} - (2n-1)!!(s_2)^n \geqslant -\frac{2n!(2n-5)!!3}{4!(2n-4)!2}|u_4|(s_2)^{n-2}.$$

Summing up the inequality, we get the desired inequality (8). ■

The above lemma and the result by Newman[5] enable us to bound the location of the first Lee–Yang zero α_1 by simple quantities of the systems under *vanishing* external field.

Proposition 2: For arbitrary L and $J \geqslant 0$, the location of the first Lee–Yang zero $\alpha_1(L,J)$ satisfies

$$\frac{1}{s_2(L,J)} \leqslant \alpha_1(L,J)^2 \leqslant \frac{s_2(L,J)}{|u_4(L,J)|}\left(\leqslant \frac{2^{d-1}\chi(2L,J)}{\chi(L/2)^3}\right). \quad (9)$$

Here the final bound in the bracket is valid only for the Ising systems.

Proof: The upper bound $(\alpha_1)^2 \leqslant s_2/|u_4|$ is due to Newman.[5] The bound in the bracket is easily derived by a version of the Griffiths, Hurst, and Sherman (GHS) inequality[14] $U_4(x,y,z,w) \leqslant -2\langle\varphi_x\varphi_y\rangle\langle\varphi_x\varphi_z\rangle\langle\varphi_x\varphi_z\rangle$ valid only for the Ising models. To prove the lower bound, note that the inequality in Lemma 1 with bound $|u_4| \leqslant 2(s_2)^2$ [Theorem 5 of Ref. 13 (b), Eq. (5.3) of Ref. 8] imply $f(ik) \geqslant \exp(-X) - (X^2/4)\cosh X$, where $X = s_2 k^2/2$. Then we find that $f(ik)$ is strictly positive in the region $X \leqslant 0.9874...$, i.e., $k^2 \leqslant 1.9748.../s_2$. ■

It is well known that the models in consideration, in their infinite volume limit $L \to \infty$, undergo phase transition provided that $d \geqslant 2$. If $J < J_c$, the system is in the high temperature phase where the correlation functions cluster exponentially, and if $J > J_c$, it is in the low temperature phase where the spontaneous magnetization appears.[15]

Our first results on the motion of the first Lee–Yang zero are concerned with the systems not at the critical point J_c.

Corollary 3: For the Ising models with $d \geqslant 2$, we have

$$L^{-d/2} \lesssim \alpha_1(L,J) \lesssim \text{const}, \quad J < J_c, \quad (10)$$

$$\alpha_1(L,J) \sim \text{const}, \quad J < J_0 \leqslant J_c, \quad (11)$$

$$\alpha_1,(L,J) \sim L^{-d}, \quad J > J_c, \quad (12)$$

as $L \to \infty$. Here J_0 is a certain constant depending on d. If $d = 1$, Eq. (11) is valid for all $J < \infty$, and Eq. (12) holds for $J = \infty$. [Here the relation $g(L,J) \lesssim h(L,J)$ implies that there is a continuous function $c(J)$ independent of L, and $g(L,J) \leqslant c(J)h(L,J)$ holds for all L; $g \sim h$ implies that $g \lesssim h$ and $g \gtrsim h$ hold simultaneously.]

Proof: Note that we have $\chi(L,J) \sim \chi(L = \infty, J) < \infty$ for $J < J_c$, and $\chi(L,J) \sim M_s(J)^2 L^d$ for $J > J_c$, where $M_s(J)$ is the spontaneous magnetization. These with Eq. (7) and Proposition 2 immediately imply (10) and (12). Equation (11) is a consequence of an independent argument on the analyticity of the free energy.[16] The result for $d = 1$ follows from an explicit calculation of Lee and Yang.[1] ■

It is strongly expected that the behavior (11) is valid for all $J < J_c$.

Our Proposition 2 also provides information on the behavior of α_1 for the systems strictly at the critical point J_c. First assuming the decay property at the critical point as

$$\langle\varphi_x\varphi_y\rangle_{J_c}^L \sim |x-y|^{-(d-2+\eta)}, \quad (13)$$

we find that the behavior of α_1 and J_c is distinct from those in the noncritical phases.

Corollary 4: For the Ising models with $d \geqslant 2$ at the critical point, we have

$$L^{-(d+2-\eta)/2} \lesssim \alpha_1(L,J_c) \lesssim L^{\eta-2}, \quad \text{as } L \to \infty \quad (14)$$

under the assumption (13).

We can also state the following rigorous version of (14) without any assumptions.

Corollary 5: For the Ising models at the critical point, we have

$$L^{-15/8} \lesssim \alpha_1(L,J_c) \lesssim L^{-14/8}, \quad d = 2, \quad (15)$$

$$L^{-1-d/2} \lesssim \alpha_1(L,J_c) \lesssim L^{-1}, \quad d \geqslant 3, \quad (16)$$

as $L \to \infty$.

Proofs: Equations (14) and (15) are the direct consequences of Proposition 2 and Eq. (13), where in the latter we have used the exact result $\eta = \frac{1}{4}$ for $d = 2$. Equation (16) is proved using Proposition 2 and Lemma 7 in the Appendix. ■

It seems that, at the critical point, our lower bound in Proposition 2 is more strict in lower dimensions, and becomes weaker in higher dimensions.

For the φ^4 systems, we can show Eqs. (10), (11), and $\alpha_1(L,J) \gtrsim L^{-d}$ $(J > J_c)$ instead of (12). Similarly, we can only prove the lower bounds of Eqs. (14) and (16).

Finally, let us state a specific result for the "soft" φ^4 models in high dimensions.

Corollary 6: For the φ^4 models in $d > 4$ dimensions with sufficiently small λ, we have

$$L^{-1-d/2} \lesssim \alpha_1(L,J_c) \lesssim L^{-3}, \quad \text{as } L \to \infty. \quad (17)$$

Proof: The lower bound is the same as that in Eq. (16). For the proof of the upper bound, we use the facts $|u_4| \sim \lambda L^d \chi^4$, and $\langle\varphi_x\varphi_y\rangle \sim |x-y|^{-d+2}$ and Proposition 2. The former is derived (as in Sec. III of Ref. 17) from the first- and second-order skeleton inequalities,[18] and the latter is a consequence of the rigorous renormalization group analysis of Gawedzki and Kupiainen.[19] ■

From the representation $f(ik) = \exp(\Sigma(-1)^n k^{2n} u_{2n})$ and a perturbative estimate $u_{2n} \sim (\text{combinatoric factor})(-\lambda)^{n-1} L^d \chi^{3n-2}$ for φ^4 models, it is conjectured that

$$\alpha_1(L,J_c) \sim \chi(L,J_c)^{-3/2} \sim L^{-3} \quad (18)$$

is the correct behavior of $\alpha_1(L,J_c)$ for $d > 4$.

In dimensions $d > 4$, it has been proved[8,9] that the scaling limits of the Ising and φ^4 models are Gaussian. Let us describe a heuristic relation between this fact and the motion of the first Lee–Yang zero at J_c.

Consider an averaged spin variable $\Phi = L^{-(d/2+1)}\Sigma_{x\in\Lambda(L)}\varphi_x$, where we used the normalization factor of the critical (i.e., massless) Gaussian theory. If we denote by $Z = \pm iA_1$ the Lee–Yang zero of the partition function $F(Z) = \langle\exp(Z\Phi)\rangle$, we obviously have $A_1 = L^{d/2+1}\alpha_1$. Now substituting the conjectured behavior of α_1 (18), we observe $A_1 \sim L^{(d-4)/2} \to \infty$ as $L \to \infty$ if $d > 4$. Therefore the first zero A_1, along with all the other zeros, is driven away to infinity in the thermodyamic limit. Since a

distribution without Lee–Yang zeros is Gaussian,[5] this suggests that Φ becomes a Gaussian random variable in this limit.

APPENDIX: BOUNDS FOR s_2

Here we prove a technical lemma based on the infrared bounds[10] and Simon's argument[11] on the critical decay.

Lemma 7: At the critical point $J = J_c$, s_2 satisfies the following bounds

$$\text{const}\, J_c^{-1} L^{d+1} \leqslant s_2(L,J_c) \leqslant \text{const}\, J_c^{-1} L^{d+2}. \tag{19}$$

Proof: To prove the *upper bound*, recall that Griffiths II inequality[12] and Sokal's argument[20] on the infrared bounds[10] imply

$$s_2(L,J) \leqslant \sum_{x,y\in\Lambda(L)} \langle \varphi_x \varphi_y \rangle_J^P \leqslant \text{const}\, J^{-1} L^{d+2} + M_s(J)^2,$$

where $\langle \cdots \rangle_J^P$ is an infinite volume expectation with the periodic boundary condition, and $M_s(J)$ is the corresponding spontaneous magnetization. For $J < J_c$, we have $M_s(J) = 0$. Since the expectation in a finite box is continuous in J, we have

$$s_2(L,J_c) \leqslant \text{const}\, J_c^{-1} L^{d+2}.$$

To prove the *lower bound*, we recall that Simon's argument[11] combined with the Simon–Lieb inequality[21] imply that for arbitrary finite sublattice Λ of Z^d,

$$\sum_{x\in\partial\Lambda} \langle \varphi_0 \varphi_x \rangle_\Lambda \geqslant \text{const}\, J_c^{-1}$$

holds at the critical point. Thus we have

$$\left(\frac{L}{2}\right)^{-d} s_2(L,J_c) \geqslant \sum_{L'=1}^{L/4} \sum_{x\in\partial\Lambda(L')} \langle \varphi_0 \varphi_x \rangle_{J_c}^{L'} \geqslant \text{const}\, J_c^{-1} L,$$

which, with (7), leads us to the desired lower bound. ■

ACKNOWLEDGMENTS

I wish to thank Professor M. Suzuki for continual encouragements, invaluable discussions, and a careful reading of the manuscript, and Dr. H. Nishimori and T. Hara for useful discussions.

[1] C. N. Yang and T. D. Lee, Phys. Rev. **87**, 404 (1952); T. D. Lee and C. N. Yang, Phys. Rev. **87**, 410 (1952).
[2] T. Asano, J. Phys. Soc. Jpn. **29**, 350 (1970); M. Suzuki and T. Fisher, J. Math. Phys. **12**, 235 (1971).
[3] B. Simon and R. B. Griffiths, Commun. Math. Phys. **33**, 145 (1973); C. M. Newman, *Constructive Field Theory* (Springer, Berlin, 1973); E. H. Lieb and A. D. Sokal, Commun. Math. Phys. **80**, 153 (1981).
[4] H. Nishimori and R. B. Griffiths, J. Math. Phys. **24**, 2637 (1983).
[5] C. M. Newman, Commun. Math. Phys. **41**, 1 (1975).
[6] D. Iagolnitzer and B. Souillard, Phys. Rev. B **19**, 1515 (1979).
[7] M. Suzuki, Prog. Theor. Phys. **38**, 289, 1225 (1967); **39**, 349 (1968).
[8] M. Aizenman, Commun. Math. Phys. **86**, 1 (1982).
[9] J. Fröhlich, Nucl. Phys. B **200** [FS4], 281 (1982).
[10] J. Fröhlich, B. Simon, and T. Spencer, Commun. Math. Phys. **50**, 79 (1976).
[11] B. Simon, Commun. Math. Phys. **77**, 111 (1980).
[12] R. B. Griffiths, J. Math. Phys. **8**, 478, 484 (1967).
[13] (a) J. L. Lebowitz, Commun. Math. Phys. **35**, 87 (1974); (b) C. M. Newman, Z. Wahrscheinlichkeitstheor. Verw. Geb. **33**, 75 (1975).
[14] R. B. Griffiths, C. A. Hurst, and S. Sherman, J. Math. Phys. **11**, 790 (1970).
[15] M. Aizenman, Phys. Rev. Lett. **54**, 839 (1985); M. Aizenman, D. J. Barsky, and R. Fernandez, preprint.
[16] See, for example, J. L. Lebowitz, in *Lecture Notes in Physics*, Vol. 39 (Springer, Berlin, 1975).
[17] T. Hara, T. Hattori, and H. Tasaki, J. Math. Phys. **26**, 2922 (1985).
[18] D. Brydges, J. Fröhlich, and A. D. Sokal, Commun. Math. Phys. **91**, 117 (1983); A. Bovier and G. Felder, Commun. Math. Phys. **93**, 259 (1984).
[19] K. Gawedzki and A. Kupiainen, Commun. Math. Phys. **99**, 197 (1985).
[20] Lemma A3 and its proof in A. D. Sokal, Ann. Inst. H. Poincaré A **37**, 317 (1982).
[21] E. H. Lieb, Commun. Math. Phys. **77**, 127 (1980); D. Brydges, J. Fröhlich, and T. Spencer, Commun. Math. Phys. **83**, 123 (1982).

Generalized Riccati equations for self-dual Yang–Mills fields

L.-L. Chau and H. C. Yen
Physics Department, Brookhaven National Laboratory, Upton, New York 11973

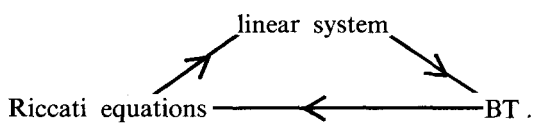
(Received 20 June 1986; accepted for publication 17 September 1986)

It is shown that although no Riccati equations in the strict sense are likely to exist for the self-dual Yang–Mills fields, certain "generalized Riccati equations" derivable from the Bäcklund transformation do exist, and are capable of reproducing the linear system when a certain constraint is imposed.

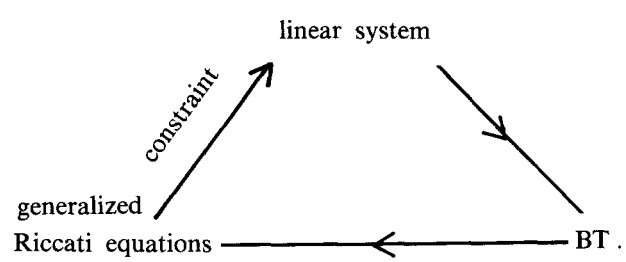
I. INTRODUCTION

By now it is well established that the two-dimensional principal chiral model and the four-dimensional self-dual Yang–Mills (SDYM) fields are both examples of integrable systems[1] with all the characteristic integrability properties, such as the existence of linear systems, an infinite number of nonlocal conservation laws, Riemann–Hilbert transformations, Kac–Moody algebra, Bianchi–Bäcklund transformations, etc. In fact, these two systems are very similar to each other in their integrability structures, except in one important area: the Riccati equations.

For the chiral model, it is possible to derive a pair of Riccati equations from the Bäcklund transformation (BT) equations generated from the linear system. These Riccati equations can be used, on the other hand, to establish the existence of an infinite set of local conservation laws, and, on the other hand, to reconstruct the linear system, thus completing the following logic cycle:



No corresponding Riccati equations have ever been found for the SDYM fields. Indeed, we will show in this paper that no such Riccati equations are likely to exist for SDYM. Nevertheless, we will also show that the BT equation for SDYM can be transformed into a certain set of equations, which we will call "generalized Riccati equations," and which, together with the constraint equation contained in the BT, are sufficient to reconstruct the linear system. The situation can be expressed schematically as



II. A REVIEW ON THE RICCATI EQUATIONS OF THE CHIRAL MODEL

In order to make clear the problem at hand, we first review the derivation of the Riccati equations and their relation to the linear system for the chiral model. It has been shown in Ref. 1 that a linear system for the principal chiral fields is

$$\partial_\xi \psi = [\lambda/(1-\lambda)]A_\xi \psi\,, \quad \partial_\eta \psi = [-\lambda/(1+\lambda)]A_\eta \psi\,, \tag{2.1}$$

where λ is a complex parameter and $A_\xi = g^+\partial_\xi g$, $A_\eta = g^+\partial_\eta g$, with $g^+g = 1$. The integrability condition of (2.1) is the field equation

$$\partial_\xi A_\eta + \partial_\eta A_\xi = 0\,. \tag{2.2}$$

From (2.1) a specific procedure can be taken to generate a BT[2,3]

$$g'^+\partial_\xi g' - g^+\partial_\xi g = \partial_\xi(g'^+g)\,, \tag{2.3a}$$

$$g'^+\partial_\eta g' - g^+\partial_\eta g = -\partial_\eta(g'^+g)\,, \tag{2.3b}$$

together with an imposed constraint

$$g^+g' + g'^+g = -2\beta\,, \tag{2.4}$$

which is consistent with (2.3) in the sense that (2.3) in itself already implies that $g^+g' + g'^+g =$ constant matrix.

Now to derive the Riccati equations, we define $\Gamma \equiv -g^+g'$, which satisfies $\Gamma^+\Gamma = 1$, and rewrite (2.3) and (2.4) completely in terms of Γ and A_ξ, A_η:

$$(1-\Gamma)\Gamma_{,\xi} = [\Gamma, A_\xi]\,, \tag{2.5a}$$

$$(1+\Gamma)\Gamma_{,\eta} = [\Gamma, A_\eta]\,, \tag{2.5b}$$

$$\Gamma + \Gamma^{-1} = 2\beta \quad \text{or} \quad \Gamma^2 = 2\beta\Gamma - 1\,. \tag{2.6}$$

The matrix $(1 \mp \Gamma)$ is invertible, and indeed, from (2.6),

$$(1 \mp \Gamma)^{-1} = [1/2(1 \mp \beta)](1 \mp \Gamma^{-1})\,. \tag{2.7}$$

Multiplying $(1 \mp \Gamma)^{-1}$ on (2.5) from the left, we get the following Riccati equations:

$$\begin{aligned}\Gamma_{,\xi} &= [1/2(1-\beta)] \\ &\quad \times [-\Gamma A_\xi \Gamma + \Gamma A_\xi - (1-2\beta)A_\xi\Gamma - A_\xi]\,, \\ \Gamma_{,\eta} &= [1/2(1+\beta)] \\ &\quad \times [\Gamma A_\eta \Gamma + \Gamma A_\eta - (1+2\beta)A_\eta\Gamma + A_\eta]\,,\end{aligned} \tag{2.8}$$

where (2.6) has been taken into account.

The prescription to reconstruct the linear system from (2.8) has been described in Ref. 1, yielding the following equations:

$$\partial_\xi M = \frac{A_\xi}{2(1-\beta)}\begin{bmatrix} -(1-2\beta) & -1 \\ 1 & -1 \end{bmatrix} M\,, \tag{2.9a}$$

 0022-2488/87/051167-03$02.50

$$\partial_\eta M=\frac{A_\eta}{2(1+\beta)}\begin{bmatrix}-(1+2\beta) & 1\\ -1 & -1\end{bmatrix}M\,, \tag{2.9b}$$

where M can be either a two-component column vector or a 2×2 matrix. Conversely, it is also possible to derive (2.8) from (2.9) in the following way.[1] Suppose (2.9) has a solution

$$M=\begin{pmatrix}M_1 & M_2\\ M_3 & M_4\end{pmatrix},$$

where each M_i is in itself a matrix. Define

$$\Gamma\equiv(M_1C+M_2)(M_3C+M_4)^{-1}\,, \tag{2.10}$$

where C is a constant matrix. Then it is straightforward to show, without the aid of any constraint on Γ, that such Γ satisfies the Riccati equations (2.8).

Finally, when (2.9) are simultaneously diagonalized, they become

$$\partial_\xi M'=\frac{A_\xi}{2}\begin{bmatrix}-1+i\tau & 0\\ 0 & -1-i\tau\end{bmatrix}M'\,, \tag{2.11a}$$

$$\partial_\eta M'=\frac{A_\eta}{2}\begin{bmatrix}-1-i/\tau & 0\\ 0 & -1+i/\tau\end{bmatrix}M'\,, \tag{2.11b}$$

where $\tau=[(1+\beta)/(1-\beta)\,]^{1/2}$. Equation (2.11) is in fact of the same form as (2.1), with λ of (2.1) given by either root of the equation

$$\lambda^2=2\beta\lambda-1\,.$$

We note in passing that one can show from (2.8) that the following continuity equation holds:

$$\partial_\xi\,\mathrm{Tr}[2(1-\beta)\Gamma A_\eta\,]+\partial_\eta\,\mathrm{Tr}[2(1+\beta)\Gamma A_\xi\,]=0\,, \tag{2.12}$$

which, when expanded asymptotically at $\beta=\pm 1$, gives rise to an infinite set of local conservation laws.

III. IMPOSSIBILITY OF RICCATI EQUATIONS FOR THE SDYM CASE

For SDYM fields in the J-formulation,[1] the linear system is given by

$$\begin{aligned}&[\partial_y-(1/\lambda)\partial_{\bar z}\,]\chi+B_y\chi=0\,,\\ &[\partial_z+(1/\lambda)\partial_{\bar y}\,]\chi+B_z\chi=0\,,\end{aligned} \tag{3.1}$$

where λ is a complex parameter, and $B_y=J^{-1}J_{,y}$, $B_z=J^{-1}J_{,z}$, with $J^+=J$. The integrability condition of (3.1) is the equation of motion

$$\partial_{\bar y}B_y+\partial_{\bar z}B_z=0\,. \tag{3.2}$$

From (3.1), the following BT has been generated[2,3]:

$$J'^{-1}J'_{,y}-J^{-1}J_{,y}=(J'^{-1}J)_{,\bar z}\,, \tag{3.3a}$$

$$J'^{-1}J'_{,z}-J^{-1}J_{,z}=-(J'^{-1}J)_{,\bar y}\,, \tag{3.3b}$$

together with an imposed constraint

$$J^{-1}J'-J'^{-1}J=\beta\,. \tag{3.4}$$

The structures of (3.3) and (3.4) are similar to those of (2.3) and (2.4), but there is a difference in that (3.3a) and (3.4) together can be shown to contain already (3.3b) [or alternatively, (3.3b) and (3.4) contain already (3.3a)], but no such relations exist in (2.3) and (2.4).

Now let $\Gamma\equiv J^{-1}J'$. Then (3.3) and (3.4) become

$$\Gamma_{,y}-\Gamma\Gamma_{,\bar z}=[\Gamma,B_y]\,, \tag{3.5a}$$

$$\Gamma_{,z}+\Gamma\Gamma_{,\bar y}=[\Gamma,B_z]\,, \tag{3.5b}$$

$$\Gamma-\Gamma^{-1}=\beta\quad\text{or}\quad\Gamma^2=\beta\Gamma+1\,. \tag{3.6}$$

The steps that have been followed to derive (2.8) from (2.5) cannot be carried over for the present case, due to the structural difference in (3.5) and (2.5). It is apparently possible to bring the lhs of (3.5a) into a form more closely resembling the lhs of (2.5a), by multiplying (3.5a) from the left with a factor $(1+\sigma\Gamma)$, where the constant σ is such that

$$(1+\sigma\Gamma)\Gamma=(1+\sigma\Gamma)\sigma\,,$$

or equivalently, $\sigma^2=\beta\sigma+1$. Then we have

$$(1+\sigma\Gamma)(\Gamma_{,y}-\sigma\Gamma_{,\bar z})=(1+\sigma\Gamma)[\Gamma,B_y]\,. \tag{3.7}$$

However, the inverse of $(1+\sigma\Gamma)$ does not exist and we are unable to convert (3.7) into a Riccati equation.

Now we will give an argument indicating why there should be no Ricatti equations for the SDYM system. If the Riccati equations were to exist for SDYM, the most plausible form for them would be

$$(\partial_y-\alpha\,\partial_{\bar z})\Gamma=\Gamma A\Gamma+\Gamma B+C\Gamma+D\,, \tag{3.8a}$$

$$(\partial_z+\alpha\,\partial_{\bar y})\Gamma=\Gamma A'\Gamma+\Gamma B'+C'\Gamma+D'\,, \tag{3.8b}$$

where α is some constant, and A, B, C, D are matrices proportional to B_y, while A',B',C',D' are matrices proportional to B_z. This specific structure of (3.8) is suggested by the relations between (2.8), (2.9), and (2.11), and by the desired form of the linear system (3.1).

Taking the complex conjugate of (3.8a) and noting that $\Gamma^+=J\Gamma J^{-1}$, we get

$$\begin{aligned}&[\partial_z-(1/\alpha^*)\partial_{\bar y}\,]\Gamma\\ &\quad=[\Gamma,B_z]-(1/\alpha^*)[\Gamma,(J^{-1}B_y^{\,+}J)]\\ &\qquad-(1/\alpha^*)[\Gamma(J^{-1}A^{\,+}J)\Gamma+\Gamma(J^{-1}C^{\,+}J)\\ &\qquad+(J^{-1}B^{\,+}J)\Gamma+(J^{-1}D^{\,+}J)]\,.\end{aligned} \tag{3.9}$$

Equation (3.9) is independent of (3.8b), since $\alpha\neq-1/(\alpha^*)$. From (3.9) and (3.8b) we would get two separate Riccati equations,

$$\partial_z\Gamma=\cdots\,,\quad\partial_{\bar y}\Gamma=\cdots\,. \tag{3.10}$$

Similarly, from (3.8a) and the complex conjugate of (3.8b), we would obtain

$$\partial_y\Gamma=\cdots\,,\quad\partial_{\bar z}\Gamma=\cdots\,. \tag{3.11}$$

Equations (3.10) and (3.11) altogether would constitute too large a number of Riccati equations to correctly reconstruct the linear system (3.1).

To avoid the above difficulty, suppose we keep only (3.8a) and its complex conjugate,

$$[\partial_y-\alpha\partial_{\bar z}\,]\Gamma=\Gamma A\Gamma+\Gamma B+C\Gamma+D\,, \tag{3.12a}$$

$$[\partial_z-(1/\alpha^*)\partial_{\bar z}\,]\Gamma=\Gamma A''\Gamma+\Gamma B''+C''\Gamma+D''\,, \tag{3.12b}$$

while discarding (3.8b) and its complex conjugate. From (3.12) it is possible to construct a linear system following the procedure that led us from (2.8) to (2.9), but such a linear system would not produce the correct equation of motion (3.2). In summary, there seems to be no room for Riccati

equations in the SDYM case, in contrast to the chiral model. This impossibility might root in the higher dimensionality than 2 of the SDYM theory.

IV. GENERALIZED RICCATI EQUATIONS FOR SDYM

Since no Riccati equations seem possible for SDYM, we will settle for (3.5) and (3.6) as a replacement of the Riccati equations. We will call (3.5) the generalized Riccati equations, and describe a method to reconstruct the linear system from them. Assume a solution Γ of (3.5) can be factorized into a product of two undetermined matrices

$$\Gamma = XY^{-1}, \tag{4.1}$$

with one more relation between X and Y to be imposed later, so that each of them can be completely specified. Substituting (4.1) into (3.5a), we get

$$X_{,y} - \Gamma Y_{,y} - \Gamma(X_{,\bar{z}} - \Gamma Y_{,\bar{z}}) = \Gamma B_y Y - B_y X. \tag{4.2}$$

We then use (2.6) to get rid of the only Γ^2 term appearing in (3.2), and obtain

$$(X_{,y} + B_y X + Y_{,\bar{z}}) - \Gamma(X_{,\bar{z}} + Y_{,y} + B_y Y - \beta Y_{,\bar{z}}) = 0. \tag{4.3}$$

Now we supply the missing relation between X and Y by imposing

$$X_{,y} + B_y X + Y_{,\bar{z}} = 0, \tag{4.4}$$

so that we get from (3.3)

$$X_{,\bar{z}} + Y_{,y} + B_y Y - \beta Y_{,\bar{z}} = 0. \tag{4.5}$$

Equations (4.4) and (4.5) can be combined in a matrix form

$$\begin{pmatrix} \partial_y + B_y & \partial_{\bar{z}} \\ \partial_{\bar{z}} & \partial_y + B_y - \beta\partial_{\bar{z}} \end{pmatrix} \begin{pmatrix} X \\ Y \end{pmatrix} = 0. \tag{4.6a}$$

A similar maneuver on (2.5b) gives

$$\begin{pmatrix} \partial_z + B_z & -\partial_{\bar{y}} \\ -\partial_{\bar{y}} & \partial_z + B_z + \beta\partial_{\bar{y}} \end{pmatrix} \begin{pmatrix} X \\ Y \end{pmatrix} = 0. \tag{4.6b}$$

Equations (3.3a) and (3.3b) can be simultaneously diagonalized to become

$$\begin{pmatrix} \partial_y + B_y + (1/\sigma)\partial_{\bar{z}} & 0 \\ 0 & \partial_y + B_y - \sigma\partial_{\bar{z}} \end{pmatrix} \times \begin{pmatrix} X + (1/\sigma)Y \\ X - \sigma Y \end{pmatrix} = 0, \tag{4.7a}$$

$$\begin{pmatrix} \partial_z + B_z - (1/\sigma)\partial_{\bar{y}} & 0 \\ 0 & \partial_z + B_z + \sigma\partial_{\bar{y}} \end{pmatrix} \times \begin{pmatrix} X + (1/\sigma)Y \\ X - \sigma Y \end{pmatrix} = 0, \tag{4.7b}$$

where $\sigma = \frac{1}{2}[\beta + (\beta^2 + 4)^{1/2}]$. Comparing with the original linear system (3.1), we find $\lambda = \sigma$ or $-1/\sigma$, i.e., λ is either root of the equation $\lambda^2 = \beta\lambda + 1$. Notice that to derive the linear system (4.6) from (3.5), it has been necessary to use the constraint (3.7), while in going from (2.8) to (2.9), no constraint on Γ is needed.

Conversely, to derive the generalized Riccati equations (3.5) from the linear system (4.6) will also need help from the constraint (3.7), as we will show in the following. Suppose $\binom{X_1}{Y_1}$ and $\binom{X_2}{Y_2}$ are two independent solutions of (4.6). Define

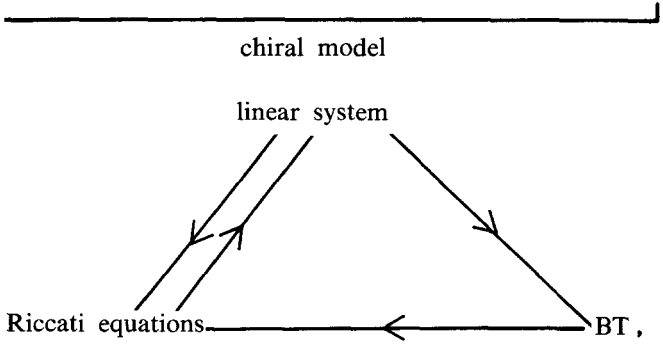
$$\Gamma \equiv (X_1 C + X_2)(Y_1 C + Y_2)^{-1}, \tag{4.8}$$

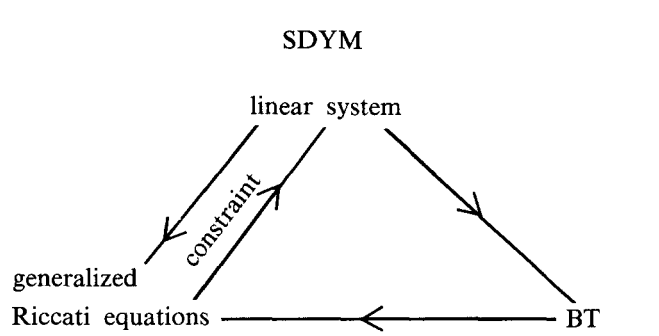
where C is a constant matrix. Then a direct computation on (4.8) yields

$$\Gamma_{,y} = \Gamma\Gamma_{,\bar{z}} + (\Gamma, B_y) + (\Gamma^2 - \beta\Gamma - 1) \times (Y_{1,\bar{z}} - C + Y_{2,\bar{z}})(Y_1 C + Y_2)^{-1}. \tag{4.9}$$

If Γ of (4.8) happens to satisfy the constraint (3.7), then (4.9) becomes identical to (3.5a). Equation (2.6b) can be similarly reproduced. Thus the constraint (3.7) is needed in addition to the linear system (4.6) to return to the generalized Riccati equations (3.5).

To conclude, we have shown that the chiral model and the SDYM theory differ from each other in regard to the Riccati equation, and the situation can be summarized by the following diagrams:





ACKNOWLEDGMENTS

One of us (H.C.Y.) wishes to thank Dr. H. H. Chen, Dr. C. R. Lee, and Dr. J. C. Shaw for many useful discussions.

[1] L.-L. Chau, in *Proceedings of Nonlinear Phenomena*, Mexico, 1982, *Lecture Notes in Physics*, Vol. 189 (Springer, New York, 1983).

[2] L.-L. Chau, in *Proceedings of Workshop on Vertex Operators in Mathematics and Physics*, Berkeley, 10–17 November 1983; in *Proceedings of the 13th International Colloquium on Group Theoretical Methods in Physics*, Maryland, 21–25 May, 1984.

[3] L.-L. Chau and H. C. Yen, "A unified derivation of Bäcklund transformations for integrable nonlinear equations," BNL preprint, 1986.

Finite-dimensional Lorentz covariant bifurcations

Mayer Humi
Department of Mathematical Sciences, Worcester Polytechnic Institute, Worcester, Massachusetts 01609

(Received 27 September 1985; accepted for publication 31 December 1986)

In this paper finite-dimensional Lorentz covariant bifurcation equations are constructed and their properties, solutions, and gradient structures are examined. The possible applications of these ideas and techniques to elementary particle physics are considered.

I. INTRODUCTION

In the last few years extensive research was done on bifurcations covariant with respect to the rotation group in three dimensions and their applications in various physical contexts.[1-5] In view of these efforts the study of Lorentz covariant bifurcations seems to be both natural and interesting from physical and mathematical considerations.

First, from a mathematical point of view, the Lorentz group is a simple noncompact group [as compared to O(3) which is compact] and hence all its nontrivial finite-dimensional representations are nonunitary, thus possibly introducing a new element in the analysis of the bifurcation equations. More important from a physical point of view is the fact that the Lorentz group is the invariance group of all local relativistic physical phenomena and hence covariant bifurcations with respect to this group should govern all bifurcations of relativistic processes. In particular we wish to point out that the production of new (elementary) particles through a collision of other particles at relativistic velocities can be viewed as a bifurcation process. Thus in this instance the original (stable) state of the system (consisting of the particles before the collision) becomes, above certain energy threshold, unstable due to the collision and the system bifurcates to new states or particles. It follows then that the detailed study of these Lorentz covariant bifurcations, which are independent of the explicit form of the interaction, might lead to better understanding of these processes, which goes beyond those consisting of spin and energy alone.

The plan of the paper is as follows: in Sec. II we summarize briefly the general setting for covariant bifurcations as discussed by Sattinger[1,2] and comment on the possible difficulties in its application to noncompact groups. In Sec. III the construction of Lorentz covariant bifurcation equations is carried out and in Sec. IV we present an explicit example of these equations and their solutions. In Sec. V we prove that Lorentz covariant bifurcations of the second order have a gradient structure even though the corresponding representations are nonunitary. Some possible implications of these techniques to the physics of elementary particles are considered in Sec. VI. Finally in the Appendix we show the need for a minor modification in the formula for the Clebsch–Gordan coefficients of the Lorentz group.

II. A SHORT REVIEW OF BIFURCATION THEORY WITH SYMMETRY[1,2]

We are considering the bifurcations of a nonlinear functional equation $G(u,\lambda)=0$. Under proper conditions on $G(\lambda,u)$ we can reduce this equation at a bifurcation point (λ_0,u_0) via the Lyapunov–Schmidt method to a finite-dimensional problem:

$$F_i(\lambda,\mathbf{v})=0, \quad i=1,...,n, \tag{1}$$

$\mathbf{v}\in R^n$ and $n=\dim\ker G_u(\lambda_0,u_0)$. Expanding $\mathbf{F}(\lambda,\mathbf{v})$ in a power series in $\mathbf{v}$ we obtain,

$$\mathbf{F}(\lambda,\mathbf{v})=A(\lambda)\mathbf{v}+\mathbf{B}_2(\lambda,\mathbf{v},\mathbf{v})+\mathbf{B}_3(\lambda,\mathbf{v},\mathbf{v},\mathbf{v})+\cdots. \tag{2}$$

On can infer then that if the original problem is covariant with respect to a representation Γ of a group G then the same holds for each term in the expansion (2). Furthermore, since $\mathbf{B}_2(\lambda,\mathbf{v},\mathbf{w})$ must be symmetric in $\mathbf{v},\mathbf{w}$ it follows that $\mathbf{B}_2$ must belong to the subspace of symmetric second-order tensors which transform as Γ under the action of G.

For the rest of this work we approximate $\mathbf{F}(\lambda,\mathbf{v})$ by the first two terms in (2) (obviously, if $\mathbf{B}_2\equiv 0$ one must consider $\mathbf{B}_3$, etc.) and denote $\mathbf{B}_2$ by B.

The construction and anlysis of second-order G-covariant bifurcations proceeds as follows: first, we identify those representations for which Γ appears as a symmetric tensor in the decomposition of $\Gamma\times\Gamma$. Then to construct B explicitly we can either use the Clebsh–Gordan coefficients of G or apply the Lie generators of G directly on some "ground state" of B. Furthermore, if Γ is irreducible it follows then from Schur's Lemma that $A(\lambda)=\lambda I$. Once the solutions of

$$B(\mathbf{v},\mathbf{v})+A(\lambda)\mathbf{v}=0 \tag{3}$$

have been found (usually there are several solutions) one can infer the stability of each bifurcating state by introducing the parametrization

$$\lambda=-\epsilon, \quad \mathbf{v}=-\epsilon\zeta \tag{4}$$

and calculating the eigenvalues of $J-\lambda I$, where J is the Jacobian of B at the solution. For $\epsilon>0$, negative and positive eigenvalues correspond then to stable and unstable subcritical branching states, respectively.

Although the results of bifurcation theory reviewed above are rather general, proper care should be exercised in their application to the Lorentz group since it is a noncompact group. Due to this fact there exist some open mathematical questions as to whether the Fredholm alternative and the Lyapunov–Schmidt procedure hold under these conditions. While these problems should be addressed formally, we would like to observe that from a physical point of view the Lorentz group is a local symmetry group. Accordingly, in our analysis of the bifurcation equations we have to consider only a proper neighborhood of the identity in this group. It is our contention then that under these restrictions

 0022-2488/87/051170-05$02.50

the bifurcation theory as developed in Refs. 1 and 2 holds for Lorentz covariant bifurcations.

Another possible source of trouble in applying the general theory to the finite-dimensional representations of the Lorentz group is that these are nonunitary.

However, a close examination of the theory developed in Ref. 1 (and Theorem 4.1 in particular) shows that the unitarity assumption is not needed and we can apply these results in our context.

III. CONSTRUCTION OF LORENTZ COVARIANT BIFURCATION EQUATIONS

We first observe that there are two ways to characterize the spinor representations of the (proper) Lorentz group (which we denote henceforth by G.) These are $(k,n)_s$ and (j_0,j_1). The first of these notations relates to the spinor contents of the representation while the second relates to its decomposition with respect to O(3).[6]

Lemma 1: Let $\Gamma = (k,n)_s$ be an irreducible representation of G and let $F(\lambda,\mathbf{v})$ be covariant with respect to Γ, then $B \equiv 0$ unless k,n are even.

Proof: For B to be different from zero it is necessary (but not sufficient) for Γ to appear in the decomposition of $\Gamma \times \Gamma$. However,

$$\Gamma \times \Gamma = \sum \oplus (k',n')_s, \tag{5}$$

where

$$k' = 0,2,\ldots,2k, \quad n' = 0,2,\ldots,2n.$$

Hence, we infer that $(k,n)_s$ appears in this decomposition only if k,n are even.

Proposition 1: If $\Gamma = (k,n)_s$, then B does not vanish if and only if k,n and $(k+n)/2$ are even integers.

Proof: In view of Lemma 1 Γ appears (once) in the decomposition of $\Gamma \times \Gamma$ under the conditions of this proposition. We must prove, however, that it appears as a symmetric tensor. To show this we denote the states of Γ by $x(k,n,j,m)$. The states of a representation $(k',n')_s$, which appear in the decomposition of $\Gamma \times \Gamma$, are then given by

$$x(k',n',j,m) = \sum H^{k',n',j,m}_{k,n,j_1,m_1;k,n,j_2,m_2} x(k,n,j_1,m_1)x(k,n,j_2,m_2), \tag{6}$$

where the H's are the Clebsch–Gordan (CG) coefficients of G. Hence Γ appears as a symmetric tensor in the decomposition if and only if

$$H^{k,n,j,m}_{k,n,j_1,m_1;k,n,j_2,m_2} = H^{k,n,j,m}_{k,n,j_2,m_2;k,n,j_1,m_1}. \tag{7}$$

However, it is well known that the CG coefficients of G are given by[6,7] (see, however, the discussion in the Appendix)

$$H^{k,n,j,m}_{k_1,n_1,j_1,m_1;k_2,n_2,j_2,m_2} = (-1)^{(k+n)/2}[(k+1)(n+1)]^{1/2}\alpha(j_1,j_2,j) \times \begin{pmatrix} j_1,j_2,j \\ m_1,m_2,m \end{pmatrix} \begin{Bmatrix} k_1/2 & n_1/2 & j_1 \\ k_2/2 & n_2/2 & j_2 \\ k/2 & n/2 & j \end{Bmatrix}, \tag{8}$$

where α is symmetric in j_1,j_2. Hence, using the symmetry properties of the $3j$ and the $9j$ symbols[8] we infer that (7) is true if and only if $(-1)^{(k+n)/2} = 1$, which proves our statement.

We now note that the explicit form of B is known if the CG coefficients of G are known. In fact for $\Gamma = (k,n)_s$ we have

$$B(j,m) = \sum H^{j,m}_{j_1,m_1;j_2,m_2} x(j_1,m_1)x(j_2,m_2) \tag{9}$$

(for brevity we dropped the designation of the representation and shall do so henceforth whenever its meaning is clear). However, since the actual calculation of the H's is tedious we describe a direct method to do so for irreducible spinor representations of the form $(n,n)_s \equiv (0,n+1)$, i.e., the ladder representations whose decomposition with respect to O(3) contains the irreducible representations $j = 0,1,\ldots,n$. (The method to be described can be applied to other spinor representations with minor modifications.)

Proposition 2: The action of the operator

$$F^2 = -(B_1^2 + B_2^2 + B_3^2) = F_-F_+ + F_3^2 - J_3 \tag{10}$$

on $x(k,n,j,m)$ (here k,n are arbitrary) is given by

$$F^2 x(k,n,j,m) = (j_1^2 - j_0^2 + j^2 + j + 1)x(k,n,j,m). \tag{11}$$

Proof: The proof of this proposition proceeds through direct (and long) computation using the results in Ref. 6 regarding the matrix elements of the operators F_+, F_-, and F_3.

At this point we would like to note that the operator F^2 seems to have an intrinsic importance from a group theoretical point of view. Thus, as is obvious from (11), any state $x(k,n,j,m)$ of $(k,n)_s$ is an eigenstate of F^2. Moreover, the corresponding eigenvalues are independent of m. However, we found no reference to this operator in the classical literature on G.

We start the construction of the quadratic form $B(j,m)$ with $B(0,0)$. To this end we observe that the representation $j = 0$ appears only in the decomposition of $j \times j$. Hence we attempt to write

$$B(0,0) = \sum_{j=0}^{n} \sum_{m=-j}^{j} a(j,m)x(j,m)x(j,-m). \tag{12}$$

To determine the coefficients $a(j,m)$ we use the fact that

$$J_+B(0,0) = 0 \tag{13}$$

[or equivalently $J_-B(0,0) = 0$]. This yields after some simple algebra

$$B(0,0) = \sum_{j=0}^{n} b(j) \sum_{m=-j}^{j} (-1)^m x(j,m)x(j,-m), \tag{14}$$

where $b(j)$ are constants which depend on j only. To determine these coefficients we now apply F^2 to $B(0,0)$ using (11) with $j_0 = 0, j_1 = n+1$,

$$F^2B(0,0) = -(n^2+2n)B(0,0). \tag{15}$$

Evaluating the left-hand side of this equation by direct application of $F_3^2 + F_-F_+ - J_3$ to (14) yields a system of linear equations for $b(j)$ (note that F^2 is not a derivation) which when solved determines $B(0,0)$. The other components of B can be obtained then by repeated aplications of F_+ and J_- or F_- and J_+.

In order to solve the second-order bifurcation equations

we shall use the following results, which are completely analogous to those for O(3).[1]

Lemma 2: The states $x(j,m)$ in an irreducible spinor representation of G can be chosen so that

$$\bar{x}(j,m) = (-1)^m x(j,-m). \tag{16}$$

Proof: An irreducible spinor representation of G can be decomposed into irreducible representations of O(3), each of which appears once. It was shown by Sattinger[1] that due to the uniqueness of $x(j,m)$ in an irreducible representation of O(3) it is possible to satisfy condition (14). This proves the lemma.

Proposition 3: Let the reduced second-order bifurcation equations for $(n,n)_s$

$$B(j,m) + \lambda x(j,m) = 0 \tag{17}$$

be restricted to the subclass of solutions with the symmetry

$$x(j,m) = (-1)^m x(j,-m), \tag{18}$$

then

$$B(j,-m) = (-1)^m B(j,m), \tag{19}$$

i.e., the bifurcation equations for $m<0$ are redundant.

Proof: From the previous lemma it follows that $B(j,m)$ can be chosen so that

$$\bar{B}(j,m) = (-1)^m B(j,-m). \tag{20}$$

However, by construction $B(j,m)$ for $(n,n)_s$ is a quadratic form with real coefficients, hence

$$\bar{B}(j,m)(x(j,m)) = B(j,m)(\bar{x}(j,m)). \tag{21}$$

But from Lemma 2 and condition (18) we obtain

$$\bar{x}(j,m) = (-1)^m x(j,-m) = x(j,m). \tag{22}$$

Thus,

$$\begin{aligned} B(j,-m)(x(j,m)) &= (-1)^m \bar{B}(j,m)(x(j,m)) \\ &= (-1)^m B(j,m)(jx(j,m)) \\ &= (-1)^m B(j,m)(x(j,m)), \end{aligned} \tag{23}$$

which is the desired result.

Finally we note that when Γ is reducible the additional considerations that are necessary to construct and solve the bifurcation equations are completely analogous to the O(3) case[1] and will not be discussed further here.

Remark: The representations $(j,j+1)$ of G are equivalent to irreducible representations of O(3). It follows then that the construction of covariant bifurcation equations which are related to these representations proceeds exactly as in the O(3) case.

IV. AN EXAMPLE

In this section we construct and solve explicitly the bifurcation equations for $(2,2)_s$. To begin with we apply F^2 to $B(0,0)$ and use (9) and (11). We obtain the following equations for $b(j)$:

$$b(0) = 2b(1),\ b(1) + b(2) = 0. \tag{24}$$

Hence, up to a multiplicative constant, $B(0,0)$ in this case is given by

$$\begin{aligned} B(0,0) = {} & 2x^2(0,0) + x^2(1,0) \\ & - 2x(1,1)x(1,-1) - 2x(2,-2)x(2,2) \\ & + 2x(2,1)x(2,-1) - x^2(2,0). \end{aligned}$$

The other components of B can now be evaluated by repeated application of F_+ and J_-,

$$\begin{aligned} B(1,1) = {} & -2\sqrt{3}x(1,-1)x(2,2) + \sqrt{6}x(1,0)x(2,1) \\ & - \sqrt{2}x(1,1)x(2,0) + 2x(0,0)x(1,1), \\ B(1,0) = {} & -\sqrt{6}x(1,-1)x(2,1) + 2\sqrt{2}x(1,0)x(2,0) \\ & - \sqrt{6}x(1,1)x(2,-1) + 2x(0,0)x(1,0), \\ B(1,-1) = {} & -2\sqrt{3}x(1,1)x(2,-2) + \sqrt{6}x(1,0)x(2,-1) \\ & - \sqrt{2}x(1,-1)x(2,0) + 2x(0,0)x(1-1), \\ B(2,2) = {} & -2x(2,2)[\sqrt{2}x(2,0) + x(0,0)] + \sqrt{3}x^2(2,1) \\ & + \sqrt{3}x^2(1,1), \\ B(2,1) = {} & -2\sqrt{3}x(2,-1)x(2,2) + \sqrt{2}x(2,0)x(2,1) \\ & - 2x(0,0)x(2,1) + \sqrt{6}x(1,0)x(1,1), \\ B(2,0) = {} & -2\sqrt{2}x(2,-2)x(2,2) - \sqrt{2}x(2,-1)x(2,1) \\ & + \sqrt{2}x^2(2,0) - 2x(0,0)x(2,0) \\ & + \sqrt{2}x(1,-1)x(1,1) + \sqrt{2}x^2(1,0), \\ B(2,-1) = {} & -2\sqrt{3}x(2,-2)x(2,1) + \sqrt{2}x(2,-1)x(2,0) \\ & - 2x(0,0)x(2,-1) + \sqrt{6}x(1,0)x(1,-1), \\ B(2,-2) = {} & -2x(2,-2)[\sqrt{2}x(2,0) + x(0,0)] \\ & + \sqrt{3}x^2(2,-1) + \sqrt{3}x^2(1,-1). \end{aligned}$$

Solving the second-order bifurcation equations $B(\mathbf{v},\mathbf{v}) + \lambda \mathbf{v} = 0$ under the restrictions of Proposition 3 we found the following four solutions for $(2,2)_s$ [note that for an irreducible representation $A(\lambda) = \lambda I$ in (21)]. The nonzero components of these solutions are

$$\begin{aligned} &(1)\ x(0,0) = -\lambda/2; \quad (2)\ x(0,0) = \lambda/6, \\ &x(2,0) = -(\sqrt{2}/3)\lambda; \quad (3,4)\ x(2,2) = \pm(\sqrt{3}/6)\lambda, \\ &x(2,0) = (\sqrt{2}/6)\lambda, \quad x(0,0) = \lambda/6. \end{aligned}$$

To determine the stability of these bifurcating solutions we let $\lambda = 1$ and evaluate the eigenvalues of $J-I$, where J is the Jacobian of $B(\mathbf{v},\mathbf{v})$ at the solution. It is interesting to note that for all the four solutions given above these eigenvalues are $(-3,-2,0)$ (the corresponding multiplicities are 3, 1, 5). Hence if we introduce the scaling $\lambda = -\epsilon$, $x(j,m) = \epsilon z(j,m)$ then for $\epsilon > 0$ we obtain subcritical branching with one neutral and two stable modes. (From a physical point of view the most natural interpretation of λ is as the total energy of the system.)

V. GRADIENT STRUCTURE OF THE BIFURCATION EQUATIONS

In this section we show that for $\gamma = (n,n)_s$ there exists a Lorentz invariant polynomial $p(x(j,m))$ so that

$$B(j,m) = \frac{\partial p}{\partial x(j,m)} \tag{25}$$

if the $B(j,m)$ satisfy the condition (19). To prove this assertion we note that p has to be a third-order polynomial in

$x(j,m)$. Taking into account the invariance of p with respect to O(3) it is easy to infer that p must be of the form

$$p=\sum_{j=0}^{n} d(j) \sum_{m=-j}^{j} (-1)^m B(j,m)x(j,-m). \tag{26}$$

To determine the coefficients $d(j)$ we use the invariance of p, which implies that $F_3(p)=0$. This immediately yields $d(j)=d(j-1), j=1,...,n$,i.e., up to a multiplicative constant,

$$p=\sum_{j=0}^{n} \sum_{m=-j}^{j} (-1)^m B(j,m)x(j,-m). \tag{27}$$

Using (9) and a little algebra then yields

$$p=\sum_{j_1j_2j_3} \sum_{m_1,m_2,m_3} (-1)^m H^{j_3,-m_3}_{j_1m_1j_2,m_2} \times x(j_1,m_1)x(j_2,m_2)x(j_3,m_3). \tag{28}$$

A careful analysis of this expression using the symmetry properties of the H's (Ref. 8) and (A7) shows that the coefficient of

$$x(j_1,m_1)x(j_2,m_2)x(j_3,m_3)$$

(for fixed j_i,m_i, $i=1,2,3$) is given by

$$6(-1)^{m_3} H^{j_3,-m_3}_{j_1m_1;j_2m_2}.$$

Hence when we differentiate p with respect to $x(j_3,m_3)$ we obtain

$$\frac{1}{3}\frac{\partial p}{\partial x(j_3,m_3)} = (-1)^{m_3} \sum_{m_1m_2} H^{j_3,-m_3}_{j_1,m_1,j_2,m_2} x(j_1,m_1)x(j_2,m_2) = (-1)^m B(j_3,-m_3) = B(j_3,m_3)$$

[where we took into account the fact that $x(j_1,m_1)x(j_2,m_2)$ appears twice in this summation].

We verified this result explicitly for $(2,2)_s$.

VI. POSSIBLE PHYSICAL APPLICATIONS

To summarize the results of this paper from a physical point of view we observe that without any reference to the explicit form of the interaction involved in the bifurcating process we were able to deduce, using Lorentz covariance alone, the number of bifurcating modes and their stability. It follows then that if the proposed application of bifurcation theory to colliding beams of particles is correct one should be able to predict *a priori* certain experimental facts which were derived so far on a phenomenological basis only.

The major obstacle for such a direct application is that particles at relativistic speeds are "dressed particles." Thus it is not *a priori* clear that the same representation which is related to a given particle at low energies is appropriate for its description at high energies. In particular, one should not rule out the use of the ladder representations of G for the description of bifurcating processes involving elementary particles.[9]

A possible objection to this point of view is that in treating elementary particles one should consider the quantized fields rather than the classical equations of motion. However, regardless of the paradigm one adopts for the study of these bifurcating processes G covariance must hold and our calculations should be applicable to it. (The same applies to any physical process which is G covariant.)

Another open question that is related to this bifurcation theory is the determination of the isotropy group for the bifurcating modes [this is open even for O(3)]. The identification of such an isotropy group should in principle lead to additional quantum numbers which characterize the bifurcating states.

APPENDIX: ON THE CG COEFFICIENTS OF G

The CG coefficients for the spinor representations of the Lorentz group appear in various references (see Refs. 6 and 7). The one due to Gel'fand *et al.*[6] is equivalent to

$$H^{k,n,j,m}_{k_1,n_1j_1,m_1;k_2,n_2j_2m_2} = \sum_{m_1',m_2'} \left(\frac{k}{2},m_1'+m_2';\frac{n}{2},m-m_1'-m_2'\middle|j,m\right)\left(\frac{k_1}{2},m_1';\frac{k^2}{2},m_2'\middle|\frac{k}{2},m_1'+m_2'\right) \times\left(\frac{n_1}{2},m_1-m_1';\frac{n_2}{2},m_2-m_2'\middle|\frac{n}{2},m-m_1'-m_2'\right)\left(\frac{k_1}{2},m_1';\frac{n_1}{2},m_1-m_1'\middle|j_1,m_1\right)\left(\frac{k_2}{2},m_2';\frac{n_2}{2},m_2-m_2'\middle|j_2,m_2\right), \tag{A1}$$

while the other[7] is given by

$$H^{k,n,j,m}_{k_1,n_1j_1,m_1;k_2,n_2j_2,m_2} = (-1)^{(k+n)/2}[(k+1)(n+1)(2j_1+1)(2j_2+1)]^{1/2}(j_1,m_1;j_2,m_2|j,m) \begin{Bmatrix} k_1/2 & n_1/2 & j_1 \\ k_2/2 & n_2/2 & j_2 \\ k/2 & n/2 & j \end{Bmatrix}. \tag{A2}$$

However, it is easy to show that these two formulas agree up to an unimportant factor of $(-1)^c$, where c depends only on k_i,n_i, $i=1,2,3$. We contend, nevertheless, that in both of these equivalent formulas a factor of

$$\alpha(j_1,j_2,j) = (-1)^{(j_1+j_2-j_3)/2} \tag{A3}$$

should be inserted on the right-hand side.

To begin with we observe that (qualitatively) both (A1) and (A2) are incompatible with the matrix elements of the Lie algebra of G as they appear in Refs. 6 and 7. In fact, according to (A1) and (A2) the CG coefficients of G are all real. Hence the states of Γ' in the decomposition of $\Gamma\times\Gamma$ should all appear with real coefficients. However, when one applies F_+ or F_3 on such a state of Γ' one obtains, in general

(i.e., when $j_0 \neq 0$), an expression with both real and complex coefficients. Such a state cannot be a multiple of the one constructed using (A1) or (A2).

More concretely we computed the Clebsch–Gordan coefficients for the ground state $B(0,0)$ of $(2,2)_s$ in the decomposition of $(2,2)_s \times (2,2)_s$ using both (A1) and (A2) and found in both cases that

$$B(0,0) = (1/2\sqrt{3})\{2x^2(0,0) + 2x(1,1)x(1,-1) - x^2(1,0) - 2x(2,2)x(2,-2) + 2x(1,1)x(1,-1) - x^2(2,0)\}. \quad \text{(A4)}$$

However, (A3) is incorrect. In fact from Refs. 6 and 7 we infer that

$$F_+ B(0,0) = (4/\sqrt{3})B(1,1). \quad \text{(A5)}$$

But if we apply F_+ to (A3) we obtain

$$B(1,1) = 6x(0,0)x(1,1). \quad \text{(A6)}$$

This expression for $B(1,1)$ is wrong since from (A1) we obtain, e.g.,

$$H^{1,1}_{2,2;1,-1} = \tfrac{1}{2},$$

i.e., a term with $x(2,2)x(1,-1)$, should appear in (A6).

Similarly if we construct the highest weight $B(2,2)$ of $(2,2)_s$ using (A1) we obtain

$$B(2,2) = (1/2\sqrt{3})\{-2x(0,0)x(2,2) + 2\sqrt{2}(2,0)x(2,2) + \sqrt{3}x^2(1,1) - \sqrt{3}x^2(2,1)\}.$$

However, the application of F_+ on this state yields

$$F_+ B(2,2) = \tfrac{4}{3}x(1,1)x(2,2),$$

rather than zero as it should.

By a little algebra one finds that the required adjustment in (A1) [or (A2)] for $(2,2)_s$ is given by (A3). We conjecture, however, that this is true for all spinor representations of G since (A3) is independent of (k,n). We in fact verified this statement directly for $(4,4)_s$. The general proof of this conjecture will require a separate publication (which is outside our main thrust in this paper). The general formula for the CG coefficients of G is given therefore by

$$H^{k,n,j,m}_{k_1,n_1,j_1,m_1;k_2,n_2,j_2,m_2} = (-1)^{(k+n+j_1+j_2-j_2)/2}[(k+1)(n+1)(2j_1+1) \times (2j_2+1)]^{1/2}(j_1 m_1; j_2, m_2 | jm) \times \begin{Bmatrix} k_1/2 & n_1/2 & j_1 \\ k_2/2 & n_2/2 & j_2 \\ k/2 & n/2 & j \end{Bmatrix}. \quad \text{(A7)}$$

We would like to note, however, that the main results of this paper are independent of the proposed adjustment in (A1).

[1] D. H. Sattinger, SIAM J. Math. Anal. **8**, 179 (1977).
[2] D. H. Sattinger, *Group Theoretic Methods in Bifurcation Theory* (Springer, Berlin, 1979); J. Math. Phys. **19**, 1720 (1978).
[3] P. C. H. Martens, Phys. Rep. **115**, 315 (1948). This paper contains an extensive list of references on the subject.
[4] G. Gaeta, J. Phys. A **16**, 1607 (1983); G. Gaeta and P. Rossi, J. Math. Phys. **25**, 1671 (1984).
[5] A. K. Bajaj, SIAM J. Appl. Math. **42**, 1078 (1982); M. Golubitsky and D. Schaeffer, Commun. Pure Appl. Math. **35**, 81 (1982).
[6] I. M. Gel'fand, R. A. Minlos, and Z. Ya. Shapiro, *Representations of the Rotation and Lorentz Groups and Their Applications* (Pergamon, Oxford, 1963). We use the notation of this reference regarding the representation of the Lorentz group.
[7] R. L. Anderson, R. Raczka, M. A. Rashid, and P. Winternitz, J. Math. Phys. **11**, 1059 (1970).
[8] C. L. Biedenharn and H. Van Dam, *Quantum Theory of Angular Momentum* (Academic, New York, 1965); D. M. Brink and G. R. Satchler, *Angular Momentum* (Oxford U. P., London, 1968).
[9] M. Humi and S. Malin, Phys. Rev. **187**, 2278 (1969).

Cylindrical group and massless particles

Y. S. Kim
Department of Physics and Astronomy, University of Maryland, College Park, Maryland 20742

E. P. Wigner
Joseph Henry Laboratories, Princeton University, Princeton, New Jersey 08544

(Received 30 September 1986; accepted for publication 31 December 1986)

It is shown that the representation of the E(2)-like little group for photons can be reduced to the coordinate transformation matrix of the cylindrical group, which describes movement of a point on a cylindrical surface. The cylindrical group is isomorphic to the two-dimensional Euclidean group. As in the case of E(2), the cylindrical group can be regarded as a contraction of the three-dimensional rotation group. It is pointed out that the E(2)-like little group is the Lorentz-boosted O(3)-like little group for massive particles in the infinite-momentum/zero-mass limit. This limiting process is shown to be identical to that of the contraction of O(3) to the cylindrical group. Gauge transformations for free massless particles can thus be regarded as Lorentz-boosted rotations.

I. INTRODUCTION

In their 1953 paper,[1] Inonu and Wigner discussed the contraction of the three-dimensional rotation group [or O(3)] to the two-dimensional Euclidean group [or E(2)]. Since the little groups governing the internal space-time symmetries of massive and massless particles are locally isomorphic to O(3) and E(2) respectively,[2] it is quite natural for us to expect that the E(2)-like little group is a limiting case of the O(3)-like little group.[3]

The kinematics of the O(3)-like little group for a massive particle is well understood. The identification of this little group with O(3) can best be achieved in the Lorentz frame in which the particle is at rest.[2] In this frame, we can rotate the direction of the spin without changing the momentum. Indeed, for a massive particle, the little group is for the description of the spin orientation in the rest frame.

The kinematics of the E(2)-like little group has been somewhat less transparent, because there is no Lorentz frame in which the particle is at rest. While the geometry of E(2) can best be understood in terms of rotations and translations in two-dimensional space, there is no physical reason to expect that the translationlike degress of freedom in the E(2)-like little group represent translations in an observable space. In fact, the translationlike degrees of freedom in the little group are the gauge degrees of freedom.[4] Therefore, in the past, the correspondence between the E(2)-like little group and the two-dimensional Euclidean group has been strictly algebraic.

In this paper, we formulate a group theory of a point moving on the surface of a circular cylinder. This group is locally isomorphic to the two-dimensional Euclidean group. We show that the transformation matrix of the little group for photons reduces to that of the coordinate transformation matrix of the cylindrical group. The cylindrical group therefore bridges the gap between E(2) and the E(2)-like little group.

As in the case of E(2), we can obtain the cylindrical group by contracting the three-dimensional rotation group. While the contraction of O(3) to E(2) is a tangent-plane approximation of a spherical surface with large radius,[1] the contraction to the cylindrical group is a tangent-cylinder approximation. Using this result, together with the fact that the representation of the E(2)-like little group reduces to that of the cylindrical group, we show that the gauge degree of freedom for massless particles comes from Lorentz-boosted rotations.

In Sec. II, we discuss the cylindrical group and its isomorphism to the two-dimensional Euclidean group. Section III deals with the E(2)-like little group for photons and its isomorphism to the cylindrical group. It is shown in Sec. IV that the cylindrical group can be regarded as an equatorial-belt approximation of the three-dimensional rotation group, while E(2) can be regarded as a north-pole approximation. In Sec. V, we combine the conclusions of Sec. III and Sec. IV to show that the gauge degrees of freedom for free massless particles are Lorentz-boosted rotational degrees of freedom.

II. TWO-DIMENSIONAL EUCLIDEAN GROUP AND CYLINDRICAL GROUP

The two-dimensional Euclidean group, often called E(2), consists of rotations and translations on a two-dimensional Euclidian plane. The coordinate transformation takes the form

$$x' = x\cos\alpha - y\sin\alpha + u, \qquad y' = x\sin\alpha + y\cos\alpha + v. \tag{2.1}$$

This transformation can be written in the matrix form as

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} \cos\alpha & -\sin\alpha & u \\ \sin\alpha & \cos\alpha & v \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}. \tag{2.2}$$

The three-by-three matrix in the above expression can be exponentiated as

$$E(u, v, \alpha) = \exp[-i(uP_1 + vP_2)]\exp(-i\alpha L_3), \tag{2.3}$$

where L_3 is the generator of rotations, and P_1 and P_2 generate translations. These generators take the form

 0022-2488/87/051175-05$02.50

$$L_3=\begin{bmatrix}0 & -i & 0\\ i & 0 & 0\\ 0 & 0 & 0\end{bmatrix},$$
$$P_1=\begin{bmatrix}0 & 0 & i\\ 0 & 0 & 0\\ 0 & 0 & 0\end{bmatrix},\quad P_2\begin{bmatrix}0 & 0 & 0\\ 0 & 0 & i\\ 0 & 0 & 0\end{bmatrix}, \tag{2.4}$$

and satisfy the commutation relations

$$[P_1,P_2]=0,\quad [L_3,P_1]=iP_2,\quad [L_3,P_2]=-iP_1, \tag{2.5}$$

which form the Lie algebra for E(2).

The above commutation relations are invariant under the sign change in P_1 and P_2. They are also invariant under Hermitian conjugation. Since L_3 is Hermitian, we can replace P_1 and P_2 by

$$Q_1=-(P_1)^\dagger,\quad Q_2=-(P_2)^\dagger, \tag{2.6}$$

respectively, to obtain

$$[Q_1,Q_2]=0,\quad [L_3,Q_1]=iQ_2,\quad [L_3,Q_2]=-iQ_1. \tag{2.7}$$

These commutation relations are identical to those for E(2) given in Eq. (2.5). However, Q_1 and Q_2 are not the generators of Euclidean translations in the two-dimensional space. Let us write their matrix forms:

$$Q_1=\begin{bmatrix}0 & 0 & 0\\ 0 & 0 & 0\\ i & 0 & 0\end{bmatrix},\quad Q_2=\begin{bmatrix}0 & 0 & 0\\ 0 & 0 & 0\\ 0 & i & 0\end{bmatrix}. \tag{2.8}$$

Here L_3 is given in Eq. (2.4). As in the case of E(2), we can consider the transformation matrix

$$C(u,v,\alpha)=C(0,0,\alpha)C(u,v,0), \tag{2.9}$$

where $C(0,0,\alpha)$ is the rotation matrix and takes the form

$$C(0,0,\alpha)=\exp(-i\alpha L_3)=\begin{bmatrix}\cos\alpha & -\sin\alpha & 0\\ \sin\alpha & \cos\alpha & 0\\ 0 & 0 & 1\end{bmatrix}, \tag{2.10}$$

$$C(u,v,0)=\exp[-i(uQ_1+vQ_2)]=\begin{bmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ u & v & 1\end{bmatrix}. \tag{2.11}$$

The multiplication of the above two matrices results in the most general form of $C(u,v,\alpha)$. If this matrix is applied to the column vector (x,y,z), the result is

$$\begin{bmatrix}\cos\alpha & -\sin\alpha & 0\\ \sin\alpha & \cos\alpha & 0\\ u & v & 1\end{bmatrix}\begin{bmatrix}x\\ y\\ z\end{bmatrix}=\begin{bmatrix}x\cos\alpha-y\sin\alpha\\ x\sin\alpha+y\sin\alpha\\ z+ux+vy\end{bmatrix}. \tag{2.12}$$

This transformation leaves (x^2+y^2) invariant, while z can vary from $-\infty$ to $+\infty$. For this reason, it is quite appropriate to call the group of the above linear transformation the *cylindrical group*. This group is locally isomorphic to E(2).

If, for convenience, we set the radius of the cylinder to be unity,

$$(x^2+y^2)=1, \tag{2.13}$$

then x and y can be written as

$$x=\cos\phi,\quad y=\sin\phi, \tag{2.14}$$

and the transformation of Eq. (2.12) takes the form

$$\begin{bmatrix}\cos\alpha & -\sin\alpha & 0\\ \sin\alpha & \cos\alpha & 0\\ u & v & 1\end{bmatrix}\begin{bmatrix}\cos\phi\\ \sin\phi\\ z\end{bmatrix}=\begin{bmatrix}\cos(\phi+\alpha)\\ \sin(\phi+\alpha)\\ z+u\cos\phi+v\sin\phi\end{bmatrix}. \tag{2.15}$$

We shall see in the following sections how this cylindrical group describes gauge transformations for massless particles.

III. E(2)-LIKE LITTLE GROUP FOR PHOTONS

Let us consider a single free photon moving along the z direction. Then we can write the four-potential as

$$A^\mu(x)=A^\mu e^{i\omega(z-t)}, \tag{3.1}$$

where

$$A^\mu=(A_1,A_2,A_3,A_0).$$

The momentum four-vector is clearly

$$p^\mu=(0,0,\omega,\omega). \tag{3.2}$$

Then, the little group applicable to the photon four-potential is generated by

$$J_3=\begin{bmatrix}0 & -i & 0 & 0\\ i & 0 & 0 & 0\\ 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0\end{bmatrix}, \tag{3.3}$$

$$N_1=\begin{bmatrix}0 & 0 & -i & i\\ 0 & 0 & 0 & 0\\ i & 0 & 0 & 0\\ i & 0 & 0 & 0\end{bmatrix},\quad N_2=\begin{bmatrix}0 & 0 & 0 & 0\\ 0 & 0 & -i & i\\ 0 & i & 0 & 0\\ 0 & i & 0 & 0\end{bmatrix}.$$

These matrices satisfy the commutation relations:

$$[J_3,N_1]=iN_2,\quad [J_3,N_2]=-iN_1,\quad [N_1,N_2]=0, \tag{3.4}$$

which are identical to those for E(2). From these generators, we can construct the transformation matrix:

$$D(u,v,\alpha)=D(0,0,\alpha)D(u,v,0), \tag{3.5}$$

where

$$D(u,v,0)=\exp[-i(uN_1+vN_2)],$$

$$D(0,0,\alpha)=R(\alpha)=\exp[-i\alpha J_3].$$

We can now expand the above formulas in power series, and the results are

$$R(\alpha)=\begin{bmatrix}\cos\alpha & -\sin\alpha & 0 & 0\\ \sin\alpha & \cos\alpha & 0 & 0\\ 0 & 0 & 1 & 0\\ 0 & 0 & 0 & 1\end{bmatrix}, \tag{3.6}$$

and

$$D(u,v,0) = \begin{bmatrix} 1 & 0 & -u & u \\ 0 & 1 & -v & v \\ u & v & 1-(u^2+v^2)/2 & (u^2+v^2)/2 \\ u & v & -(u^2+v^2)/2 & 1+(u^2+v^2)/2 \end{bmatrix}. \tag{3.7}$$

When applied to the four-potential, the above D matrix performs a gauge transformation,[4] while $R(\alpha)$ is the rotation matrix around the momentum.

The D matrices of Eq. (3.5) have the same algebraic property as that for the E matrices discussed in Sec. II. Why, then, do they look so different? In the case of the O(3)-like little group, the four-by-four matrices of the little group can be reduced to a block diagonal form consisting of the three-by-three rotation matrix and one-by-one unit matrix.[2] Is it then possible to reduce the D matrices to the form which can be directly compared with the three-by-three E or C matrices discussed in Sec. II?

One major problem in bringing the D matrix to the form of the E matrix is that the D matrix is quadratic in the u and v variables. In order to attack this problem, let us impose the Lorentz condition on the four-potential:

$$\frac{\partial}{\partial x^\mu}(A^\mu(x)) = p^\mu A_\mu(x) = 0, \tag{3.8}$$

resulting in $A_3 = A_0$. Since the third and fourth components are identical, the N_1 and N_2 matrices of Eq. (3.3) can be replaced, respectively, by

$$N_1 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ i & 0 & 0 & 0 \\ i & 0 & 0 & 0 \end{bmatrix}, \quad N_2 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & i & 0 & 0 \\ 0 & i & 0 & 0 \end{bmatrix}. \tag{3.9}$$

At the same time, the $D(u,v,0)$ of Eq. (3.7) becomes

$$D(u,v,0) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ u & v & 1 & 0 \\ u & v & 0 & 1 \end{bmatrix}. \tag{3.10}$$

This matrix has some resemblance to the representation of the cylindrical group given in Eq. (2.11).[5]

In order to make the above form identical to Eq. (2.11), we use the light cone coordinate system in which the combinations $x, y, (z+t)/\sqrt{2}$, and $(z-t)/\sqrt{2}$ are used as the coordinate variables.[6] In this system the four-potential of Eq. (3.1) is written as

$$A^\mu = (A_1, A_2, (A_3 + A_0)/\sqrt{2}, (A_3 - A_0)/\sqrt{2}). \tag{3.11}$$

The linear transformation from the four-vector of Eq. (3.1) to the above expression is straightforward. According to the Lorentz condition, the fourth component of the above expression vanishes. We are thus left with the first three components.

During the transformation into the light-cone coordinate system, J_3 remains the same. If we take into account the fact that the fourth component of A^μ vanishes, N_1 and N_2 become

$$N_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ i & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad N_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & i & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}. \tag{3.12}$$

As a consequence, $D(u,v)$ takes the form

$$D(u,v,0) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ u/\sqrt{2} & v/\sqrt{2} & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \tag{3.13}$$

and $R(\alpha)$ remains the same as before. It is now clear that the four-by-four representation of the little group is reduced to one three-by-three matrix and one trivial one-by-one matrix. If we use $\tilde{J}_3$, $\tilde{N}_1$, and $\tilde{N}_2$ for the three-by-three portion of the four-by-four J_3, N_1, and N_2 matrices, respectively, then

$$\tilde{J}_3 = L_3, \quad \tilde{N}_1 = (1/\sqrt{2})Q_1, \quad \tilde{N}_2 = (1/\sqrt{2})Q_2. \tag{3.14}$$

Now the identification of E(2)-like little group with the cylindrical group is complete.

IV. THE CYLINDRICAL GROUP AS A CONTRACTION OF O(3)

The contraction of O(3) to E(2) is well known and discussed widely in the literature.[1] The easiest way to understand this procedure is to consider a sphere with large radius, and a small area around the north pole. This area would appear like a flat surface. We can then make Euclidean transformations on this surface, consisting of translations along the x and y directions and rotations around any point within this area. Strictly speaking, however, these Euclidean transformations are SO(3) rotations around the x axis, y axis, and around the axis which makes a very small angle with the z axis.

Let us start with the generators of O(3), which satisfy the commutation relations:

$$[L_i, L_j] = i\epsilon_{ijk} L_k. \tag{4.1}$$

Here L_3 generates rotations around the north pole , and its matrix form is given in Eq. (2.4). Also, L_1 and L_2 take the form

$$L_1 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{bmatrix}, \quad L_2 = \begin{bmatrix} 0 & 0 & i \\ 0 & 0 & 0 \\ -i & 0 & 0 \end{bmatrix}. \tag{4.2}$$

For the present purpose, we can restrict ourselves to a small region near the north pole, where z is large and is equal to the radius of the sphere R, and x and y are much smaller than the radius. We can then write

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1/R \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix}. \tag{4.3}$$

The column vectors on the left- and right-hand sides are, respectively, the coordinate vectors on which the E(2) and O(3) transformations are applicable. We shall use the notation A for the three-by-three matrix on the right-hand side. In the limit of large R,

$$L_3 = AL_3A^{-1},$$
$$P_1 = (1/R)AL_2A^{-1}, \tag{4.4}$$
$$P_2 = -(1/R)AL_1A^{-1}.$$

This procedure leaves L_3 invariant. However, L_1 and L_2 become the P_1 and P_2 matrices discussed in Sec. II. Furthermore, in terms of P_1, P_2 and L_3, the commutation relations for O(3) given in Eq. (4.1) become

$$[L_3,P_1] = iP_2, \quad [L_3,P_2] = -iP_1,$$
$$[P_1,P_2] = -i(1/R)^2L_3. \tag{4.5}$$

In the large-R limit, the commutator $[P_1, P_2]$ vanishes, and the above set of commutators becomes the Lie algebra for E(2).

We have so far considered the area near the north pole where z is much larger than $(x^2+y^2)^{1/2}$. Let us next consider the opposite case, in which $(x^2+y^2)^{1/2}$ is much larger than z. This is the equatorial belt of the sphere. Around this belt, x and y can be written as

$$x = R\cos\phi, \quad y = R\sin\phi. \tag{4.6}$$

We can now write

$$\begin{bmatrix}\cos\phi\\ \sin\phi\\ z\end{bmatrix} = \begin{bmatrix}1/R & 0 & 0\\ 0 & 1/R & 0\\ 0 & 0 & 1\end{bmatrix}\begin{bmatrix}x\\ y\\ z\end{bmatrix}, \tag{4.7}$$

to obtain the vector space for the cylindrical group discussed in Sec. II. The three-by-three matrix on the right-hand side of the above expression is proportional to the inverse of the matrix A given in Eq. (4.3). Thus in the limit of large R,

$$L_3 = A^{-1}L_3A,$$
$$Q_1 = -(1/R)A^{-1}L_2A, \tag{4.8}$$
$$Q_2 = (1/R)A^{-1}L_1A.$$

In terms of L_3, Q_1, and Q_2, the commutation relations for O(3) given in Eq. (4.1) become

$$[L_3,Q_1] = iQ_2, \quad [L_3,Q_2] = -iQ_1,$$
$$[Q_1,Q_2] = -i(1/R)^2L_3, \tag{4.9}$$

which become the Lie algebra for E(2) in the large-R limit. The contraction of O(3) to E(2) and to the cylindrical group is illustrated in Fig. 1.

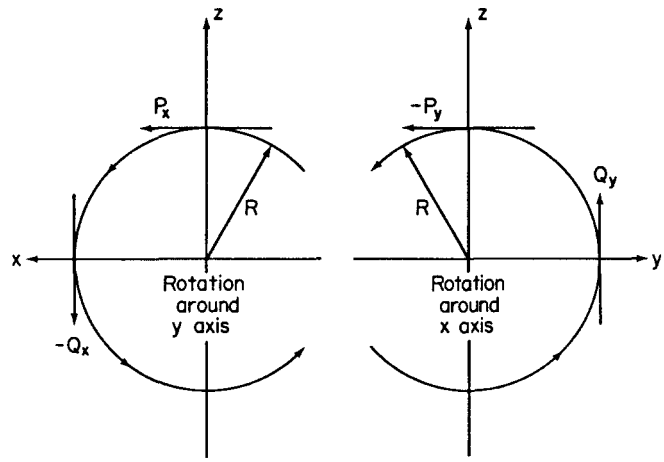


FIG. 1. Contraction of the three-dimensional rotation group to the two-dimensional Euclidean group and to the cylindrical group. The rotation around the z axis remains unchanged as the radius becomes large. In the case of E(2), rotations around the y and x axes become translations in the x and $-y$ directions, respectively, within a flat area near the north pole. In the case of the cylindrical group, the rotations around the y and x axes result in translations in the negative and positive z directions, respectively, within a cylindrical belt around the equator.

V. E(2)-LIKE LITTLE GROUP AS AN INFINITE-MOMENTUM/ZERO-MASS LIMIT OF THE O(3)-LIKE LITTLE GROUP FOR MASSIVE PARTICLES

If a massive particle is at rest, the symmetry group is generated by the angular momentum operators J_1, J_2, and J_3. If this particle moves along the z direction, J_3 remains invariant, and its eigenvalue is the helicity. However, what happens to J_1 and J_2, particularly in the infinite-momentum limit?

In order to tackle this problem, let us summarize the results of the preceding sections. The generators of the E(2)-like little group can be reduced to those of the cylindrical group. The cylindrical group can be obtained from the three-dimensional rotation group through a large-radius approximation. Therefore if the boost matrix takes a diagonal form as in the case of Eq. (4.3) or Eq. (4.7), we should be able to obtain N_1 and N_2 by boosting J_2 and J_1, respectively, along the z direction.[7]

Indeed, in the light-cone coordinate system, the boost matrix takes the form

$$B(P) = \begin{bmatrix}1 & 0 & 0 & 0\\ 0 & 1 & 0 & 0\\ 0 & 0 & R & 0\\ 0 & 0 & 0 & 1/R\end{bmatrix}, \tag{5.1}$$

with

$$R = \left(\frac{1+\beta}{1-\beta}\right)^{1/2},$$

where β is the velocity parameter of the particle. Under this boost, J_3 will remain invariant:

$$J_3' = BJ_3B^{-1} = J_3. \tag{5.2}$$

Here J_1 and J_2 in the light-cone coordinate system take the form

$$J_1 = \frac{1}{\sqrt{2}}\begin{bmatrix}0 & 0 & 0 & 0\\ 0 & 0 & -i & i\\ 0 & i & 0 & 0\\ 0 & -i & 0 & 0\end{bmatrix},$$
$$J_2 = \frac{1}{\sqrt{2}}\begin{bmatrix}0 & 0 & i & -i\\ 0 & 0 & 0 & 0\\ -i & 0 & 0 & 0\\ i & 0 & 0 & 0\end{bmatrix}. \tag{5.3}$$

If we boost this massive particle along the z direction, the boosted J_1 and J_2 become

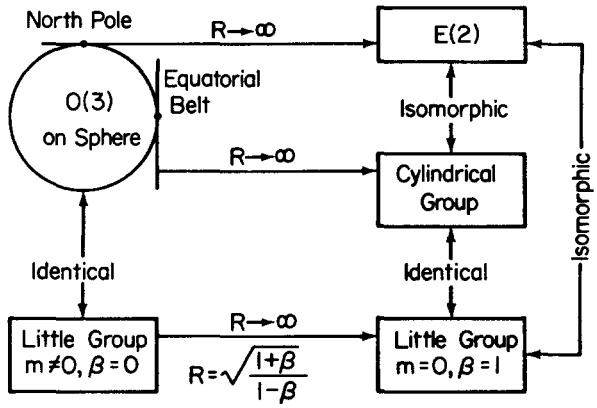


FIG. 2. Here are E(2), the E(2)-like little group for massless particles, and the cylindrical group. The correspondence between E(2) and the E(2)-like little group is isomorphic but not identical. The cylindrical group is identical to the E(2)-like little group. Both E(2) and the cylindrical group can be regarded as contractions of O(3) in the large-radius limit. The Lorentz boost of the O(3)-like little group for a massive particle at rest to the E(2)-like little group for a massless particle is exactly the same as the contraction of O(3) to the cylindrical group. The radius of the sphere in this case can be identified as $((1+\beta)/(1-\beta))^{1/2}$.

$$J_1' = BJ_1B^{-1} = \frac{1}{\sqrt{2}} \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & -i/R & iR \\ 0 & iR & 0 & 0 \\ 0 & -i/R & 0 & 0 \end{bmatrix}, \tag{5.4}$$

$$J_2' = BJ_2B^{-1} = \frac{1}{\sqrt{2}} \begin{bmatrix} 0 & 0 & i/R & -iR \\ 0 & 0 & 0 & 0 \\ -iR & 0 & 0 & 0 \\ i/R & 0 & 0 & 0 \end{bmatrix}.$$

Because of the Lorentz condition, the iR terms in the fourth column of the above matrices can be dropped. Therefore, in the large-R limit which is the limit of large momentum,

$$N_1 = -(1/R)J_2', \quad N_2 = (1/R)J_1', \tag{5.5}$$

where N_1 and N_2 are given in Eq. (3.12). This completes the proof that the gauge degrees of freedom in the E(2)-like little group for photons are Lorentz-boosted rotational degrees of freedom. The limiting process is the same as the contraction of the three-dimensional rotation group to the cylindrical group.

VI. CONCLUDING REMARKS

The isomorphism between the two-dimensional Euclidean group and the little group for massless particles is well known and well understood. However, the isomorphism in this case does not mean that they are identical. We have shown in this paper that the E(2)-like little group can be reduced to the identity group and the cylindrical group which is isomorphic to E(2). As in the case of E(2), we can obtain the cylindrical group by contracting the three-dimensional rotation group. This contraction procedure is identical to the Lorentz boost of the O(3)-like little group for a massive particle at rest to the E(2)-like little group for a massless particle. The result of the present paper is summarized in Fig. 2.

[1] E. Inonu and E. P. Wigner, Proc. Natl. Acad. Sci. USA **39**, 510 (1953); J. D. Talman, *Special Functions, A Group Theoretical Approach Based on Lectures by E. P. Wigner* (Benjamin, New York, 1968). See also R. Gilmore, *Lie Groups, Lie Algebras, and Some of Their Applications in Physics* (Wiley, New York, 1974).

[2] E. P. Wigner, Ann. Math. **40**, 149 (1939); V. Bargmann and E. P. Wigner, Proc. Natl. Acad. Sci. USA **34**, 211 (1946); E. P. Wigner, Z. Phys. **124**, 665 (1948); A. S. Wightman, in *Dispersion Relations and Elementary Particles*, edited by C. De Witt and R. Omnes (Hermann, Paris, 1960); M. Hamermesh, *Group Theory* (Addison-Wesley, Reading, MA, 1962); E. P. Wigner, in *Theoretical Physics*, edited by A. Salam (IAEA, Vienna, 1962); A. Janner and T. Jenssen, Physica **53**, 1 (1971); **60**, 292 (1972); J. L. Richard, Nuovo Cimento A **8**, 485 (1972); H. P. W. Gottlieb, Proc. R. Soc. London Ser. A **368**, 429 (1979); H. van Dam, Y. J. Ng, and L. C. Biedenharn, Phys. Lett. B **158**, 227 (1985). For a recent textbook on this subject, see Y. S. Kim and M. E. Noz, *Theory and Applications of the Poincaré Group* (Reidel, Dordrecht, Holland, 1986).

[3] E. P. Wigner, Rev. Mod. Phys. **29**, 255 (1957). See also D. W. Robinson, Helv. Phys. Acta **35**, 98 (1962); D. Korff, J. Math. Phys. **5**, 869 (1964); S. Weinberg, in *Lectures on Particles and Field Theory, Brandeis 1964*, edited by S. Deser and K. W. Ford (Prentice-Hall, Englewood Cliffs, NJ, 1965) Vol. 2; S. P. Misra and J. Maharana, Phys. Rev. D **14**, 133 (1976); D. Han, Y. S. Kim, and D. Son, J. Math. Phys. **27**, 2228 (1986).

[4] S. Weinberg, Phys. Rev. B **134**, 882 (1964); B **135**, 1049 (1964); J. Kuperzstych, Nuovo Cimento B **31**, 1 (1976); D. Han and Y. S. Kim, Am. J. Phys. **49**, 348 (1981); J. J. van der Bij, H. van Dam, and Y. J. Ng, Physica A **116**, 307 (1982); D. Han, Y. S. Kim, and D. Son, Phys. Rev. D **31**, 328 (1985).

[5] D. Han, Y. S. Kim, and D. Son, Phys. Rev. D **26**, 3717 (1982). For an earlier effort to study the E(2)-like little group in terms of the cylindrical group, see L. J. Boya and J. A. de Azcarraga, An. R. Soc. Esp. Fis. Quim. A **63**, 143 (1967). We are grateful to Professor Azcarraga for bringing this paper to our attention.

[6] P. A. M. Dirac, Rev. Mod. Phys. **21**, 392 (1949); L. P. Parker and G. M. Schmieg, Am. J. Phys. **38**, 218, 1298 (1970); Y. S. Kim and M. E. Noz, J. Math. Phys. **22**, 2289 (1981).

[7] D. Han, Y. S. Kim, and D. Son, Phys. Lett. B **131**, 327 (1983); D. Han, Y. S. Kim, M. E. Noz, and D. Son, Am. J. Phys. **52**, 1037 (1984). These authors studied the correspondence between the contraction of O(3) to E(2) and the Lorentz boost of the O(3)-like little group.

Invariant supersymmetric multilinear forms and the Casimir elements of *P*-type Lie superalgebras

Manfred Scheunert[a)]
Department of Physics, University of Freiburg, D-7800 Freiburg, Federal Republic of Germany

(Received 25 September 1986; accepted for publication 18 November 1986)

The Casimir elements of the *P*-type Lie superalgebras are investigated. Depending on the class of algebras under consideration either there do not exist any nontrivial Casimir elements at all or else the Casimir elements are highly degenerate. Basic to the investigation is a lemma about invariant supersymmetric multilinear forms on a finite-dimensional module over a Lie superalgebra. Some comments on the Cartan subalgebras of a Lie superalgebra are also included. An Appendix provides some information on multilinear algebra with ϵ-commuting scalars.

I. INTRODUCTION

The present work is one more in a series of papers[1-3] dealing with the Casimir elements of Lie superalgebras.[4,5] Once the cases of the general linear, the special linear, and the orthosymplectic Lie superalgebras have been discussed in some detail (see also Refs. 6–13), we now turn to the so-called strange classical Lie superalgebras. A special class of Q-type algebras has already been dealt with in Ref. 14, and some partial results on P-type and Q-type algebras have been obtained in Ref. 15. In an as yet unpublished investigation, the author himself has considered the simple Q-type algebras, and in Ref. 2 the P-type algebras have been touched upon.

To remind the reader of the findings in the P-type case, let us first fix our notation (see Sec. IV for more details). The "proper" P-type algebras, i.e., those whose Lie algebra is isomorphic to $\mathrm{sl}(n)$, will be denoted by $P(n-1)$. Closely related to these are the algebras which we denote by $GP(n)$ (Ref. 15) and whose Lie algebra is $\mathrm{gl}(n)$. Formally they can be obtained by adjoining a grading derivation to $P(n-1)$, geometrically they arise as the invariance algebra of a nondegenerate odd supersymmetric bilinear form. The algebra $P(n-1)$ is the commutator algebra of $GP(n)$.

The preliminary results for the P-type algebras are somewhat surprising: All attempts to construct nontrivial Casimir elements for these algebras have failed . A standard technique, which has been successful in all other cases, only produced the zero Casimir element.[2,15] Moreover, an explicit investigation showed that for $n \geqslant 3$ the $GP(n)$ and the $P(n-1)$ algebras have no nontrivial Casimir elements of order $\leqslant 4$ (for $n = 1,2$, these algebras are degenerate and hence these cases were excluded).

Thus the question arose[2] whether the P-type algebras have any nontrivial Casimir elements at all. In the present work we are going to answer this question as follows. The $GP(n)$ algebras do not have any nontrivial Casimir elements. In the $P(n-1)$ case, we characterize the Casimir elements without constant term by certain symmetric polynomials in n indeterminates and derive several necessary conditions which these Casimir elements must satisfy [for example, they are Z homogeneous of degree $-n$ and the order of the nonzero ones is at least equal to $\frac{1}{2}n(n+1)$]. Unfortunately, I do not yet know whether our conditions are also sufficient. I have only been able to settle the degenerate cases $n = 1, 2$ and to construct the lowest-order Casimir element (of order 6) for $n = 3$. To my knowledge, the latter is the first example of a nontrivial odd Casimir element.

Nevertheless, our results are sufficient to show that the Casimir elements of the $P(n-1)$ algebras are highly degenerate: The product of any two Casimir elements without constant term vanishes, and the image of any such Casimir element under a completely reducible representation is equal to zero.

Let us now briefly comment on how we are going to proceed. For any finite-dimensional Lie superalgebra L, the classification of all Casimir elements amounts to the classification of all invariant supersymmetric multilinear forms on the coadjoint module L^* (Ref. 2). We shall first look for graded subspaces U of L^* with the property that the aforementioned forms are uniquely determined by their restriction onto U. Of course, a similar problem arises for any finite-dimensional graded L-module W in place of L^*. The basic lemma of Sec. II contains a sufficient criterion for a graded subspace U of W to meet this condition. In Sec. III the lemma is applied to the adjoint and coadjoint modules of L. In this connection we also have to comment on the Cartan subalgebras of a Lie superalgebra.

The subsequent sections deal with the P-type Lie superalgebras. We recall some of their basic properties (Sec. IV), apply the lemma and derive the necessary conditions which the multilinear forms and hence the Casimir elements must satisfy (Sec. V), and discuss the examples I have studied (Sec. VI).

In the proof of the lemma we utilize a tool that is successfully employed in most applications of supersymmetry, the introduction of anticommuting scalars.[16] Essentially, we are dealing with a Lie supergroup, even though we do not explicitly rely on the theory of these groups. For the convenience of the reader, the pertinent algebraic structures have been collected in a rather extensive Appendix. This Appendix contains much more material than is needed in the proof

[a)] Part of this work has been presented at the Conference on Differential Geometric Methods in Theoretical Physics, Clausthal, West Germany, July 1986.

 0022-2488/87/051180-12$02.50

of the lemma, and we allow for arbitrary gradations and arbitrary commutation factors.[1]

We close this Introduction by the remark that throughout this work the base field K will be any commutative field of characteristic zero.

II. A BASIC LEMMA

In the present section we are going to show that an invariant supersymmetric multilinear form on a graded module W over a Lie superalgebra L is uniquely determined by its restriction onto a suitable graded subspace of W.

Lemma: Let L be a finite-dimensional Lie superalgebra and let ρ be a graded representation of L in a finite-dimensional graded vector space W. Suppose we are given a graded subspace U of W, an element $u \in U_{\bar{0}}$, and elements $A_1,...,A_r \in L_{\bar{0}}$ such that the following conditions are satisfied.

(1) The linear mappings $\rho(A_i)$ are nilpotent.

(2) The vector space $W_{\bar{0}}$ is generated by the elements $\rho(A_1)u,...,\rho(A_r)u$ and its subspace $U_{\bar{0}}$.

(3) The vector space $W_{\bar{1}}$ is generated by its subspaces $\rho(L_{\bar{1}})u$ and $U_{\bar{1}}$.

Then any L-invariant supersymmetric n-linear form ϕ on W is uniquely determined by its restriction onto U.

Proof: Without loss of generality we may assume that ϕ is homogeneous. Let V be any finite-dimensional vector space and let $S = \wedge V$ denote its Grassmann algebra, considered as a Z_2-graded algebra. Our main tool will be the extension of the domain of scalars from K to S, as described in the Appendix.

Let $\hat{\phi}$ be the graded S-multilinear extension of ϕ onto $(S \otimes W)^n$ and let $\hat{\rho}$ be the representation of $S \otimes L$ in $S \otimes W$ obtained form ρ by the extension of the domain of scalars from K to S. We know that $\hat{\phi}$ is homogeneous of the same degree as ϕ, furthermore, $\hat{\phi}$ is supersymmetric and $S \otimes L$ invariant.

Obviously, the mappings $\hat{\rho}(Q)$ with $Q \in (S \otimes L)_{\bar{0}}$ map $(S \otimes W)_{\bar{0}}$ into itself. Consequently, $\hat{\rho}$ induces a representation $\tilde{\rho}$ of the Lie algebra $(S \otimes L)_{\bar{0}}$ in the vector space $(S \otimes W)_{\bar{0}}$, and the restriction $\hat{\phi}_{\bar{0}}$ of $\hat{\phi}$ onto $(S \otimes W)_{\bar{0}}$ is symmetric and $(S \otimes L)_{\bar{0}}$ invariant.

Finally, let $\tilde{\phi}$ denote the polynomial mapping of $(S \otimes W)_{\bar{0}}$ into S defined by

$$\tilde{\phi}(x) = \hat{\phi}(x,...,x) \text{ for all } x \in (S \otimes W)_{\bar{0}}.$$

If Q is an element of $(S \otimes L)_{\bar{0}}$ such that $\tilde{\rho}(Q)$ is nilpotent, then the $(S \otimes L)_{\bar{0}}$ invariance of $\hat{\phi}_{\bar{0}}$ implies that

$$\hat{\phi}_{\bar{0}}(e^{\tilde{\rho}(Q)}x_1,...,e^{\tilde{\rho}(Q)}x_n) = \hat{\phi}_{\bar{0}}(x_1,...,x_n),$$

for all $x_1,...,x_n \in (S \otimes W)_{\bar{0}}$, and hence that

$$\tilde{\phi}(e^{\tilde{\rho}(Q)}x) = \tilde{\phi}(x) \quad \text{for all } x \in (S \otimes W)_{\bar{0}}.$$

Now choose any elements $B_1,...,B_s \in L_{\bar{1}}$ such that the vector space $W_{\bar{1}}$ is generated by the elements $\rho(B_1)u,...,\rho(B_s)u$ and its subspace $U_{\bar{1}}$ (condition 3). Consider the polynomial mapping

$$F: (S_{\bar{0}})^r \times (S_{\bar{1}})^s \times (S \otimes U)_{\bar{0}} \to (S \otimes W)_{\bar{0}}$$

defined by

$$F(a_1,...,a_r; b_1,...,b_s; y) = e^{\tilde{\rho}(a_1 \otimes A_1)} \cdots e^{\tilde{\rho}(a_r \otimes A_r)} e^{\tilde{\rho}(b_1 \otimes B_1)} \cdots e^{\tilde{\rho}(b_s \otimes B_s)} y .$$

Note that the mappings $\tilde{\rho}(a_i \otimes A_i)$ and $\tilde{\rho}(b_j \otimes B_j)$ are nilpotent, hence F is well defined (for the former, this follows from condition 1, for the latter we only have to remark that the b_j are odd).

The derivative of F at the point $(0,...,0; 0,...,0; 1 \otimes u)$ is the K-linear mapping of the vector space $(S_{\bar{0}})^r \times (S_{\bar{1}})^s \times (S \otimes U)_{\bar{0}}$ into the vector space $(S \otimes W)_{\bar{0}}$, defined by

$$(a_1,...,a_r; b_1,...,b_s; y) \to y + \sum_{i=1}^{r} a_i \otimes \rho(A_i)u + \sum_{j=1}^{s} b_j \otimes \rho(B_j)u .$$

Condition 2 and the choice of the B_j (hence condition 3) imply that this mapping is surjective. But then it is an elementary fact from algebraic geometry that any polynomial mapping of $(S \otimes W)_{\bar{0}}$ into a finite-dimensional vector space is uniquely determined by its restriction onto the image of F (see Ref. 17 for an easy proof). In view of the invariance property of $\tilde{\phi}$ this implies that $\tilde{\phi}$ is uniquely fixed by its restriction onto $(S \otimes U)_{\bar{0}}$.

Because of its symmetry, $\hat{\phi}_{\bar{0}}$ is uniquely determined by $\tilde{\phi}$, and the restriction of $\tilde{\phi}$ onto $(S \otimes U)_{\bar{0}}$ is uniquely determined by the restriction of ϕ into U. Thus all we have to show is that ϕ is uniquely determined by $\hat{\phi}_{\bar{0}}$, and this is true provided we choose V such that

$$\dim V \geqslant \dim W_{\bar{1}} .$$

The simple proof of this fact is left to the reader.

III. FIRST APPLICATIONS OF THE LEMMA

We will now apply our lemma to the adjoint and coadjoint representations of a finite-dimensional Lie superalgebra L. Our results are not necessary for the rest of the paper, but I think they are worth mentioning.

The adjoint case requires some preparatory remarks. Following Kac's basic work on Lie superalgebras[4] it has become customary to simply identify the concepts of a Cartan subalgebra of L or $L_{\bar{0}}$. On the other hand, it is easy to transcribe the classical definition to the supercase, as follows. A graded Cartan subalgebra h of L is a nilpotent graded subalgebra of L which coincides with its normalizer in L (i.e., such that, for any $A \in L$, the relation $\langle A, h \rangle \subset h$ implies $A \in h$).

One may now sit down and transcribe the classical theory of Cartan subalgebras and regular elements of a Lie algebra[18] to the supercase. I do not want to go into detail here but only mention a proposition which establishes a simple bijective correspondence between the Cartan subalgebras of $L_{\bar{0}}$ and the graded Cartan subalgebras of L.

Let us first introduce a notation. For any subset $t \subset L_{\bar{0}}$, the set of all $A \in L$ such that, for any $T \in t$,

$$(\text{ad } T)^r A = 0 \quad \text{if } r \text{ is sufficiently large,}$$

will be denoted by $L^0(t)$. It is easy to see that $L^0(t)$ is a graded subalgebra of L.

Proposition 1: Let L be a finite-dimensional Lie superalgebra.

(a) If h is a graded Cartan subalgebra of L, then $h_{\bar{0}}$ is a

Cartan subalgebra of $L_{\bar{0}}$ and $h = L^0(h_{\bar{0}})$.

(b) If k is a Cartan subalgebra of $L_{\bar{0}}$, then $h = L^0(k)$ is a graded Cartan subalgebra of L and $h_{\bar{0}} = k$.

(c) A graded subspace h of L is a graded Cartan subalgebra of L if and only if $L^0(h_{\bar{0}}) = h$.

The following proposition is a corollary to the basic lemma.

Proposition 2: Let L be a finite-dimensional Lie superalgebra and let h be a graded Cartan subalgebra of L. Then any invariant supersymmetric multilinear form on L is uniquely determined by its restriction onto h.

Proof: Let K' be any extension field of K. Then a graded subalgebra h of L is a graded Cartan subalgebra of L if and only if $K' \otimes h$ is a graded Cartan subalgebra of the Lie superalgebra $K' \otimes L$ over K'. Hence we may assume that the base field is algebraically closed.

We know that $h_{\bar{0}}$ is a Cartan subalgebra of $L_{\bar{0}}$. Consequently, we can construct the root space decomposition of L with respect to $h_{\bar{0}}$.[5] For any linear form λ on $h_{\bar{0}}$ let $L^\lambda(h_{\bar{0}})$ denote the primary component of L corresponding to λ. In particular, we have

$$L^0(h_{\bar{0}}) = h.$$

If Δ is the set of all nonzero $\lambda \in h_{\bar{0}}^*$ such that $L^\lambda(h_{\bar{0}}) \neq \{0\}$, i.e., the set of all nonzero roots of L with respect to $h_{\bar{0}}$, then

$$L = h \oplus \bigoplus_{\lambda \in \Delta} L^\lambda(h_{\bar{0}}) .$$

To apply our lemma, we set $U = h$ and choose an element $u \in h_{\bar{0}}$ such that

$$\lambda(u) \neq 0 \quad \text{for all } \lambda \in \Delta .$$

Now let $\lambda \in \Delta$ and $\alpha \in Z_2$. Then ad u induces a bijective linear mapping of $L_\alpha^\lambda(h_{\bar{0}})$ onto itself, and for any $A \in L_\alpha^\lambda(h_{\bar{0}})$, the mapping ad A is nilpotent. The rest is obvious.

Remark: Proposition 2 generalizes, in the supercase, proposition 2 of Ref. 2.

Let us next consider the case where W is the coadjoint module L^* of L (this case will be of interest for the P-type Lie superalgebras). Of course, the smaller the subspace U of W is chosen the stronger the conclusion of the lemma will be. Disregarding condition 1 as well as the requirement that u be an element of U, the strategy must be to choose $u \in W_{\bar{0}}$ such that $\rho(L)u$ has the largest possible dimension and then to take for U some graded subspace of W that is complementary to $\rho(L)u$.

In the case of the coadjoint module we are thus led to search for even linear forms u on L such that the mapping $A \to -u \circ \text{ad } A$ of L into L^* has maximal rank, i.e., such that the subspace

$$R^u = \{B \in L \mid u(\langle A, B \rangle) = 0 \text{ for all } A \in L\}$$

has minimal dimension, which is to say that the rank of the bilinear form B^u on L defined by

$$B^u(A, B) = u(\langle A, B \rangle)$$

is maximal.

For any $u \in (L^*)_{\bar{0}}$, B^u is even and super-skew-symmetric, and the radical R^u of B^u is a graded subalgebra of L. Moreover, the rank of B^u is maximal if and only if both the restrictions of B^u onto $L_{\bar{0}}$ and $L_{\bar{1}}$ have maximal rank. Generalizing a concept from classical Lie algebra theory[19] such a linear form u will be called regular on L. As in Ref. 19 one can then show that, for any regular linear form $u \in (L^*)_{\bar{0}}$, any two elements of R^u supercommute provided they are homogeneous of the same degree.

IV. *P*-TYPE ALGEBRAS

Let us now recall some properties of the P-type Lie superalgebras. It is well known[4,5] that these algebras have a natural consistent Z gradation. This fact will play a vital role in the subsequent discussion. Thus it is most convenient to carry out our analysis in the framework of ϵ Lie algebras,[20] with $\Gamma = Z$ as group of degrees and with the commutation factor ϵ defined by $\epsilon(r,s) = (-1)^{rs}$ for all $r,s \in Z$.

Let $n \geqslant 1$ be an integer. Consider $K^{2n} = K^n \oplus K^n$ as a Z-graded vector space of column vectors, where the vectors $x = (x_i)_{1 \leqslant i \leqslant 2n}$ of degree zero (resp. one) are those with $x_{n+1} = \cdots = x_{2n} = 0$ (resp. $x_1 = \cdots = x_n = 0$) while all the other homogeneous subspaces are equal to $\{0\}$. The vector space of all linear mappings of this space into itself is canonically identified with the space of all $2n \times 2n$ matrices over K. As usual, these matrices are written in a block form $\begin{pmatrix} A & B \\ C & D \end{pmatrix}$, where A, B, C, D are $n \times n$ matrices. It is well known that this space has a natural Z gradation: The matrices of the types $\begin{pmatrix} A & 0 \\ 0 & D \end{pmatrix}$, $\begin{pmatrix} 0 & B \\ 0 & 0 \end{pmatrix}$, and $\begin{pmatrix} 0 & 0 \\ C & 0 \end{pmatrix}$ form the homogeneous subspaces of degrees 0, -1, and 1, respectively.

If the ϵ commutator (which, in the present case, is nothing but the usual supercommutator) is used to introduce a multiplication in this space, the algebra that emerges is just the general linear Lie superalgebra $\mathrm{gl}(n,n)$, endowed with its natural Z gradation.[5]

Consider next the bilinear form on K^{2n} whose matrix (with respect to the canonical basis) is equal to $\begin{pmatrix} 0 & J \\ J & 0 \end{pmatrix}$, where J is the $n \times n$ unit matrix. This form is nondegenerate, ϵ symmetric, homogeneous of Z degree -1, and the corresponding ϵ-adjoint operation,[1] denoted by an asterisk, is given by

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix}^* = \begin{pmatrix} {}^tD & -{}^tB \\ {}^tC & {}^tA \end{pmatrix},$$

where, for example, ${}^tA$ is the usual transpose of the matrix A.

We define[1]

$$GP(n) = \{X \in \mathrm{gl}(n,n) \mid X^* = -X\}$$

$$= \left\{ \begin{pmatrix} A & B \\ C & -{}^tA \end{pmatrix} \middle| \, {}^tB = B, \ {}^tC = -C \right\}$$

and[21]

$$W = \{X' \in \mathrm{gl}(n,n) \mid X'^* = X'\}$$

$$= \left\{ \begin{pmatrix} A' & B' \\ C' & {}^tA' \end{pmatrix} \middle| \, {}^tB' = -B', \ {}^tC' = C' \right\}.$$

Then $GP(n)$ is the Z-graded subalgebra of $\mathrm{gl}(n,n)$ consisting of all elements that leave the mentioned bilinear form invariant. On the other hand, W is a Z-graded $GP(n)$-invariant subspace of $\mathrm{gl}(n,n)$ which is complementary to $GP(n)$ and which can be identified with the coadjoint module of $GP(n)$, the canonical pairing

$$W \times GP(n) \to K$$

being given by the supertrace (equal to the ϵ trace)

$$(X', Y) \to \mathrm{Str}(X'Y) .$$

For completeness, we give the action of $GP(n)$ on $\mathrm{gl}(n,n)$:

$$\left\langle \begin{pmatrix} A & 0 \\ 0 & -{}^tA \end{pmatrix}, \begin{pmatrix} A' & B' \\ C' & D' \end{pmatrix} \right\rangle = \begin{pmatrix} [A,A'] & AB' + B'\,{}^tA \\ -{}^tAC' - C'A & [-{}^tA, D'] \end{pmatrix},$$

$$\left\langle \begin{pmatrix} 0 & B \\ C & 0 \end{pmatrix}, \begin{pmatrix} A' & B' \\ C' & D' \end{pmatrix} \right\rangle = \begin{pmatrix} BC' + B'C & BD' - A'B \\ CA' - D'C & CB' + C'B \end{pmatrix}.$$

The first of these equations exhibits the transformation properties of the matrices A', B', C', D' under the action of the Lie algebra $GP(n)_0 \simeq \mathrm{gl}(n)$. In particular, when applying the usual tensor calculus for $\mathrm{gl}(n)$, we shall write the indices of the matrices A, B, C, D in the following positions:

$$A = (A^i{}_j), \quad B = (B^{ij}), \quad C = (C_{ij}), \quad D = (D_i{}^j) .$$

The algebra $GP(n)$ contains the element

$$R = -\tfrac{1}{2}\begin{pmatrix} J & 0 \\ 0 & -J \end{pmatrix}.$$

It has the useful property that ad R is the grading derivation of $\mathrm{gl}(n,n)$: An element $X \in \mathrm{gl}(n,n)$ is homogeneous of Z-degree r if and only if

$$\langle R, X \rangle = rX .$$

Besides the $GP(n)$ algebras we are interested in the proper P-type algebras defined by

$$P(n-1) = \langle GP(n), GP(n) \rangle = GP(n) \cap \mathrm{sl}(n,n) ,$$

where, as usual, $\langle GP(n), GP(n) \rangle$ denotes the subspace generated by all supercommutators $\langle X, Y \rangle$ with $X, Y \in GP(n)$. The $P(n-1)$ are Z-graded ideals of the $GP(n)$ and are known to be simple Lie superalgebras provided that $n \geqslant 3$.

Obviously, $GP(n)$ is the direct sum of its subspaces $P(n-1)$ and $K \cdot R$. The subspace of W orthogonal to $P(n-1)$ is equal to $K \cdot I$, where I denotes the $2n \times 2n$ unit matrix. Consequently, the coadjoint module of $P(n-1)$ can be identified with $W/K \cdot I$.

Occasionally it is useful to notice that, when considered as $GP(n)_0$ modules, $GP(n)$ is the direct sum of $K \cdot R$, $P(n-1)_{\pm 1}$, $P(n-1)_0$, and W is the direct sum of $K \cdot I$, $W_{\pm 1}$, and

$$W_{00} = \{X' \in W_0 \mid \mathrm{Str}(X'R) = 0\} = \left\{ \begin{pmatrix} A' & 0 \\ 0 & {}^tA' \end{pmatrix} \middle| \, \mathrm{Tr}(A') = 0 \right\}.$$

However, we stress that the subspace $W_{-1} \oplus W_{00} \oplus W_1$ of W is not $P(n-1)$ invariant and hence cannot be identified with the coadjoint module of $P(n-1)$.

Finally we want to comment on the weight space decompositions of $GP(n)$ and W when considered as modules over $GP(n)$ or $P(n-1)$. Let h denote the subspace of all diagonal matrices in $GP(n)$. Then h is a Cartan subalgebra of $GP(n)_0$. Define the linear forms ϵ_i, $1 \leqslant i \leqslant n$, on h by the requirement that $\epsilon_i(H)$ be the ith diagonal element of H, for all $H \in h$. Obviously, the ϵ_i form a basis of the dual h^* of h.

Let $E_{ij}, 1 \leqslant i,j \leqslant n$, be the canonical basis matrices of $\mathrm{gl}(n)$,

i.e.,

$$(E_{ij})_{kl} = \delta_{ik}\delta_{jl} \quad \text{for } 1 \leqslant i,j,k,l \leqslant n .$$

Then the block matrices $\left(\begin{smallmatrix} A & B \\ C & D \end{smallmatrix}\right)$, where, in turn, A, B, C, D are set equal to E_{ij} and the remaining three submatrices are set equal to zero are weight vectors of $\mathrm{gl}(n,n)$ with respect to h and the corresponding weights are $\epsilon_i - \epsilon_j$, $\epsilon_i + \epsilon_j$, $-\epsilon_i - \epsilon_j$, and $-\epsilon_i + \epsilon_j$, respectively. By forming the obvious linear combinations, it is then easy to construct a complete set of weight vectors for the submodules $GP(n)$ and W. We do not work out the details but only mention those block matrices which will play a role in the subsequent section, namely,

$$\tilde{E}_{ij} = \begin{pmatrix} E_{ij} & 0 \\ 0 & -E_{ji} \end{pmatrix}, \quad E'_{ij} = \begin{pmatrix} E_{ij} & 0 \\ 0 & E_{ji} \end{pmatrix},$$

$$F'_i = \begin{pmatrix} 0 & 0 \\ E_{ii} & 0 \end{pmatrix};$$

they are weight vectors of $GP(n)$, W, and W, respectively, and correspond to the weights $\epsilon_i - \epsilon_j$, $\epsilon_i - \epsilon_j$, and $-2\epsilon_i$.

For $P(n-1)$ the discussion can be repeated almost verbatim: One simply has to replace h by

$$h^s = h \cap \mathrm{sl}(n,n) ,$$

the linear forms ϵ_i by their restrictions ϵ_i^s onto h^s,

$$\epsilon_i^s = \epsilon_i | h^s ,$$

and to note that the ϵ_i^s are no longer linearly independent but satisfy a linear relation which is unique up to an overall factor,

$$\sum_{i=1}^{n} \epsilon_i^s = 0 .$$

V. CASIMIR ELEMENTS OF THE P-TYPE ALGEBRAS

We are now ready to investigate the Casimir elements of the $GP(n)$ and $P(n-1)$ algebras. According to Ref. 2 all we have to do is to determine the invariant supersymmetric (i.e., ϵ symmetric) multilinear forms on the coadjoint modules W of $GP(n)$ and $W/K \cdot I$ of $P(n-1)$, respectively. Without loss of generality we may (and will) also assume that these forms are homogeneous in the sense of the Z gradation.

We shall treat both cases simultaneously by first investigating the $P(n-1)$-invariant supersymmetric Z-homogeneous r-linear forms on W. To settle the $GP(n)$ case we then have to require in addition that these forms are annihilated by the action of the element R specified in Sec. IV, which means that the forms are Z homogeneous of degree zero; in the $P(n-1)$ case the additional condition is that the forms have to vanish whenever one of the arguments is equal to the $2n \times 2n$ unit matrix I.

Let us first apply our basic lemma. We set

$$U=\left\{\begin{pmatrix}A' & 0\\ C' & A'\end{pmatrix}\in W \,\middle|\, A',C' \text{ diagonal}\right\}.$$

Obviously, U is a Z-graded subspace of W. We choose for ρ the coadjoint representation of $P(n-1)$ in W, for the A_s the elements $\tilde{E}_{ij}$ with $i\neq j$, and for u a block matrix of the form $\left(\begin{smallmatrix}D' & 0\\ 0 & D'\end{smallmatrix}\right)$, where D' is a diagonal $n\times n$ matrix whose diagonal elements are different from each other. Then the lemma applies and shows that the forms under consideration are uniquely fixed by their restriction onto U.

Let ϕ be one of these forms. We show next that the restriction of ϕ onto U_0 vanishes. In fact, the elements of U_0 have the form $X'=\left(\begin{smallmatrix}A' & 0\\ 0 & A'\end{smallmatrix}\right)$ with diagonal $n\times n$ matrices A'. If T is the element of $P(n-1)_{-1}$ defined by $T=\left(\begin{smallmatrix}0 & J\\ 0 & 0\end{smallmatrix}\right)$ (where J is the $n\times n$ unit matrix), then

$$X'=\langle T,Y'\rangle \text{ with } Y'=\begin{pmatrix}0 & 0\\ A' & 0\end{pmatrix}\in U_1$$

and

$$\langle T,X'\rangle=0\,.$$

Let $X'_1,\ldots,X'_r$ be any elements of U_0. Write $X'_1=\langle T,Y'\rangle$ with $Y'\in U_1$ as above; then

$$\phi(X'_1,\ldots,X'_r)=\phi(\langle T,Y'\rangle,X'_2,\ldots,X'_r)$$
$$=\sum_{q=2}^{r}\phi(Y',X'_2,\ldots,\langle T,X'_q\rangle,\ldots,X'_r)=0\,.$$

This result settles the $GP(n)$ case: In that case ϕ has to be Z homogeneous of degree zero, and consequently it is already fixed by its restriction onto U_0 and this implies that $\phi=0$.

In the subsequent discussion we assume that $\phi\neq 0$. Let us investigate $\phi(X'_1,\ldots,X'_r)$ for arguments X'_q which are homogeneous elements of U. Since the elements F'_i, $1\leqslant i\leqslant n$, as defined in Sec. IV form a basis of U_1, we may assume without loss of generality that the elements X'_q lying in U_1 belong to this basis. Let λ_q be the weight of X'_q with respect to h^s (equal to 0 if $X'_q\in U_0$ and equal to $-2\epsilon^s_i$ if $X'_q=F'_i$). The h^s invariance of ϕ implies that

$$\sum_{q=1}^{r}\lambda_q(H)\phi(X'_1,\ldots,X'_r)=0 \quad \text{for all } H\in h^s\,.$$

But we know that, up to an overall factor, the ϵ^s_i satisfy the unique linear relation

$$\sum_{i=1}^{n}\epsilon^s_i=0\,.$$

Moreover, ϕ is supersymmetric and we have shown that its restriction onto U_0 is equal to zero. All this implies that $\phi(X'_1,\ldots,X'_r)$ vanishes unless each of the F'_i appears among the X'_q exactly once.

It follows that ϕ must be Z homogeneous of degree $-n$, that $r\geqslant n$, and that ϕ is uniquely determined by the $(r-n)$-linear form

$$(X'_{n+1},\ldots,X'_r)\to\phi(F'_1,\ldots,F'_n,X'_{n+1},\ldots,X'_r)$$

on U_0. Since this form is symmetric, ϕ is already fixed by the polynomial function

$$X'\to\phi(F'_1,\ldots,F'_n,X',\ldots,X')$$

on U_0.

To study this function we recall a well-known fact. Let ρ be the coadjoint representation of $P(n-1)$ in W. If Q is any element of $P(n-1)_0$ such that $\rho(Q)$ is nilpotent, the invariance of ϕ (in the Lie sense) implies that ϕ is invariant under $e^{\rho(Q)}$ (in the group sense).

We apply this remark as follows. Let i and j be two different elements of $\{1,\ldots,n\}$. Then it is well known (and easy to check) that

$$e^{\rho(\tilde{E}_{ij})}e^{-\rho(\tilde{E}_{ji})}e^{\rho(\tilde{E}_{ij})}E'_{kk}=\begin{cases}E'_{kk}, & \text{if } k\neq i,j,\\ E'_{jj}, & \text{if } k=i,\\ E'_{ii}, & \text{if } k=j,\end{cases}$$

and the formula remains valid if E'_{ll} is replaced by F'_l.

If we now write

$$X'=\sum_{i=1}^{n}x_iE'_{ii}\,,$$

it follows that $\phi(F'_1,\ldots,F'_n,X',\ldots,X')$ is a skew-symmetric polynomial function in $x_1,\ldots,x_n$. Consequently, there exists a homogeneous symmetric polynomial φ in n indeterminates such that

$$\phi(F'_1,\ldots,F'_n,X',\ldots,X')=\varphi(x_1,\ldots,x_n)\prod_{i<j}(x_i-x_j)\,,$$

and we know that ϕ is uniquely determined by φ. Visibly, the degree of φ is equal to $r-\frac{1}{2}n(n+1)$, whence it follows that $r\geqslant\frac{1}{2}n(n+1)$.

Finally, we recall that ϕ is derived from a multilinear form on $W/K\cdot I$ if and only if it vanishes whenever one of its arguments is equal to I. Let us investigate what this condition implies for φ. The following discussion applies except in the trivial case $n=r=1$.

Suppose first that this condition is satisfied. Then $\phi(F'_1,\ldots,F'_n,X'+tI,\ldots,X'+tI)$ is independent of the parameter t, for any $X'=\Sigma x_iE'_{ii}$. This in turn means that $\varphi(x_1+t,\ldots,x_n+t)$ does not depend on t or, what is the same, that

$$\sum_{i=1}^{n}\partial_i\varphi=0\,,$$

where ∂_i denotes the derivative with respect to the ith indeterminate.

Conversely, let us assume that this equation holds. Then we conclude first that

$$\phi(F'_1,\ldots,F'_n,X',\ldots,X',I)=0$$

for all $X'\in U_0$. Now consider the $(r-1)$-linear form ϕ' on W defined by

$$\phi'(X'_1,\ldots,X'_{r-1})=\phi(X'_1,\ldots,X'_{r-1},I)\,.$$

This form is $P(n-1)$ invariant [since $P(n-1)$ annihilates I], supersymmetric, and Z homogeneous (of degree $-n$). Consequently, ϕ' is uniquely determined by the polynomial function on U_0 whose value at $X'\in U_0$ is

$$\phi'(F'_1,\ldots,F'_n,X',\ldots,X')=\phi(F'_1,\ldots,F'_n,X',\ldots,X',I)=0\,.$$

Thus we have shown that $\phi'=0$ and hence that ϕ vanishes whenever one of its arguments is equal to I.

We summarize the results obtained above in the following proposition.

Proposition 3: (a) There are no nonzero $GP(n)$-invariant supersymmetric multilinear forms on the coadjoint module W of $GP(n)$.

(b) Let ϕ be any nonzero $P(n-1)$-invariant supersymmetric r-linear form on W. Then $r \geqslant \frac{1}{2}n(n+1)$ and there exists a homogeneous symmetric polynomial φ in n indeterminates such that, for all $X' = \Sigma x_i E'_{ii}$,

$$\phi(F'_1,\ldots,F'_n,X',\ldots,X') = \varphi(x_1,\ldots,x_n)\prod_{i<j}(x_i - x_j) .$$

The form ϕ is uniquely determined by φ. In particular, ϕ is Z homogeneous of degree $-n$.

(c) With the notation of part (b), the equation

$$\sum_{i=1}^{n} \partial_i \varphi = 0$$

is a necessary and sufficient condition for ϕ to vanish whenever one of its arguments is equal to I.

Corollary: (a) The Lie superalgebras $GP(n)$ have no nontrivial Casimir elements.

(b) If a nonzero Casimir element of the Lie superalgebra $P(n-1)$ has no constant term, then it is Z homogeneous of degree $-n$ and its order is at least equal to $\frac{1}{2}n(n+1)$. In particular, the product of any two such Casimir elements vanishes.

The final statement of the corollary has another strange consequence. Namely, let V be any (possibly infinite-dimensional) irreducible $P(n-1)$ module (this assumption may be understood in the Z-graded, Z_2-graded, or nongraded sense).[22] Then if C is a Casimir element of $P(n-1)$ without a constant term, the square of the corresponding Casimir operator C_V vanishes, which implies that C_V itself is equal to zero. Of course, the same holds if V is any completely reducible $P(n-1)$ module.

On the other hand, for any nonzero Casimir element C of $P(n-1)$ there exists a finite-dimensional Z-graded $P(n-1)$-module V such that $C_V \neq 0$. In fact, one can prove the following general proposition, which in the Lie algebra case is due to Harish-Chandra.

Proposition 4: Let ϵ be a commutation factor on an Abelian group Γ, let L be a finite-dimensional ϵ Lie algebra, and let $U(L)$ denote its universal enveloping algebra. For any nonzero element $X \in U(L)$ there exists a finite-dimensional Γ-graded L-module V such that the representative X_V of X in V is different from zero.

Using the results of Refs. 1, 2, and 20, the classical proof[19] can immediately be transcribed to the ϵ Lie algebra case.

Remark: Of course, Proposition 3 only solves part of our problem: One would like to know whether the conditions on φ are also sufficient. Stated differently: To any symmetric polynomial φ in n indeterminates which is homogeneous of degree d, does there exist a $P(n-1)$-invariant supersymmetric $(d + \frac{1}{2}n(n+1))$-linear form ϕ on W (necessarily Z homogeneous of degree $-n$) which corresponds to φ in the sense of Proposition 3? In view of the examples given in the subsequent section I think there is a chance that this might be correct. Be that as it may, our results clearly indicate that Casimir elements will hardly play a major role in the representation theory of $P(n-1)$.

VI. EXAMPLES

In the present section we are going to deal with the cases $n = 1,2$ and the simplest example for $n = 3$. For lack of any basic method of construction, we have to use the following tedious approach: Employing the representation theory of $P(n-1)_0 \simeq \mathrm{sl}(n)$, we determine the $P(n-1)_0$-invariant supersymmetric multilinear forms on W which satisfy the necessary conditions found in Proposition 3 and then try to fix the remaining free parameters such as to make the form $P(n-1)$ invariant.

To simplify the notation we shall identify W_0, W_{-1}, W_1 with the $\mathrm{sl}(n)$ modules of the corresponding $n \times n$ matrices $A' = (A'^i{}_j)$, $B' = (B'^{ij})$, $C' = (C'_{ij})$, which are arbitrary, skew symmetric, or symmetric, respectively.

A. The case $n=1$

This case is completely degenerate and trivial; it is only included for the sake of completeness. For any integer $r \geqslant 1$ there is, up to the normalization, just one nonzero supersymmetric Z-homogeneous r-linear form on W of degree -1. The form is $P(0)$ invariant, and it is derived from an r-linear form on $W/K \cdot I$ if and only if $r = 1$. This result is to be compared to the obvious fact that the enveloping algebra of $P(0)$ is nothing but the Grassmann algebra of the one-dimensional vector space $P(0)_{-1}$.

B. The case $n=2$

This case is more interesting but still somewhat degenerate. We will first summarize our results and then comment on how they have been obtained.

Choose any integers $s,t \geqslant 0$; we are going to define a $P(1)$-invariant supersymmetric Z-homogeneous $(2s + t + 3)$-linear form $\phi_{s,t}$ on W of degree -2. Obviously, it is sufficient to specify its restrictions onto the products $W_1^2 \times W_0^{2s+t+1}$ and $W_1^3 \times W_{-1} \times W_0^{2s+t-1}$ (if $2s + t \geqslant 1$).

Let us define the 2×2 matrix G by

$$G = (G^{ij}) = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

Then the restriction of $\phi_{s,t}$ onto $W_1^2 \times W_0^{2s+t+1}$ is given by

$$\phi_{s,t}(C'^1, C'^2, A', \ldots, A') = \mathrm{Tr}([GC'^1, GC'^2]\hat{A}'^{2s+1})(\mathrm{Tr}\,A')^t$$

and its restriction onto $W_1^3 \times W_{-1} \times W_0^{2s+t-1}$ by

$$\phi_{s,t}(C'^1, C'^2, C'^3, B', A', \ldots, A') = c_{s,t}\,\mathrm{Tr}([GC'^1, GC'^2]GC'^3) \times \mathrm{Tr}(\hat{A}'^{2s}B'G^{-1})(\mathrm{Tr}\,A')^{t-1} ,$$

where $c_{s,t}$ is some constant to be fixed below and $\hat{A}' = A' - \frac{1}{2}\mathrm{Tr}(A')$ denotes the traceless part of A'. For $t = 0$ the constant $c_{s,0}$ and the latter expression are set equal to zero. Note that for convenience the arguments from W_0 have been chosen to be equal, which is sufficient since $\phi_{s,t}$ has to be symmetric in these arguments. Also, we remind the reader that the square of any traceless 2×2 matrix (like $\hat{A}'$) is a scalar multiple of the unit matrix.

The form $\phi_{s,t}$ is $P(1)$ invariant if and only if

$$2c_{s,t}(2s+t)+t/(2s+t+1)=0,$$

and the polynomial $\varphi_{s,t}$ corresponding to $\phi_{s,t}$ in the sense of Proposition 3 is given by

$$\varphi_{s,t}(x_1,x_2)=(\tfrac{1}{2}(x_1-x_2))^{2s}(x_1+x_2)^t .$$

The polynomials $(x_1-x_2)^{2s}(x_1+x_2)^t$ form a basis of the algebra of all symmetric polynomials in x_1, x_2. According to Proposition 3 this implies that the $\phi_{s,t}$ with $2s+t+3=r$ form a basis in the space of all $P(1)$-invariant supersymmetric r-linear forms on W. Furthermore, a $P(1)$-invariant supersymmetric multilinear form on W vanishes whenever one of its arguments is equal to I if and only if it is proportional to some $\phi_{s,0}$. More generally, the $(2s+t+2)$-linear form obtained from $\phi_{s,t}$ by setting one of its arguments equal to I is just $2t/(2s+t+1)$ times $\phi_{s,t-1}$ (equal to zero if $t=0$).

To derive the aforementioned results it is advantageous to first classify those forms ϕ of the type in question, which vanish whenever one of their arguments is equal to I, and to identify these forms with their restrictions onto $W_{-1}\oplus W_{00}\oplus W_1$. According to Proposition 3, they must be $(2s+3)$ linear, with $s=0,1,2,\ldots$.

Next we recall that the mth symmetric power of the adjoint sl(2) module is isomorphic to the direct sum of the irreducible sl(2) modules of the dimensions $2m-4p+1$, where p is an integer with $0\leqslant 2p\leqslant m$. This observation shows that the restriction of ϕ onto $W_1^2\times W_0^{2s+1}$ is fixed up to its normalization and that the restriction of ϕ onto $W_1^3\times W_{-1}\times W_0^{2s-1}$ (if $s\geqslant 1$) must vanish. Moreover, it can be used to obtain a particularly simple proof of the fact that the form constructed is indeed $P(1)$ invariant.

Once this case is settled it is easy to guess the general ansatz. The results of the special case can then be used to check that this ansatz indeed works.

C. The simplest example for $n=3$

According to Proposition 3 we want to construct a $P(2)$-invariant supersymmetric six-linear form ϕ on W which is Z homogeneous of degree -3. Note that this form will automatically vanish whenever one of its arguments is equal to I, thus we could well include this property among our assumptions.

Obviously, it is sufficient to specify the restrictions of ϕ onto $W_0^3\times W_1^3$ and $W_{-1}\times W_0\times W_1^4$ (we could even replace W_0 by W_{00}).

Next we exploit the invariance of ϕ under $P(2)_0\simeq \mathrm{sl}(3)$ and begin with the restriction of ϕ onto $W_0^3\times W_1^3$. The third exterior power $\overset{3}{\wedge} W_1$ of the sl(3) module W_1 is the direct sum of a 10- and a $\overline{10}$-submodule. On the other hand, the third symmetric power $S_3(W_{00})$ of the sl(3) module W_{00} contains both a 10- and a $\overline{10}$-submodule exactly once, and the same holds true for $S_3(W_0)$. This implies that there are exactly two linearly independent sl(3)-invariant six-linear forms on $W_0^3\times W_1^3$ which have the correct symmetry properties and that these forms vanish whenever one of the arguments from W_0 is equal to I. Moreover, the foregoing analysis suggests the following construction of these forms.

Let V_0 and $\overline{V}_0$ denote the sl(3) modules of all rank three tensors with three upper or three lower indices, respectively, and let V and $\overline{V}$ be the corresponding submodules of symmetric tensors (they carry the 10- and $\overline{10}$-representations, respectively). We define the skew-symmetric sl(3)-invariant trilinear mappings

$$S\colon W_1^3\to V,\quad \overline{S}\colon W_1^3\to\overline{V},$$

by

$$S^{ijk}(C'^1,C'^2,C'^3)=C'^a_{wp}\epsilon^{pqi}C'^b_{qr}\epsilon^{rsj}C'^c_{sv}\epsilon^{vwk}\epsilon_{abc}$$

and

$$\overline{S}_{ijk}(C'^1,C'^2,C'^3)=C'^a_{ip}C'^b_{jq}C'^c_{kr}\epsilon^{pqr}\epsilon_{abc},$$

where the ϵ symbols are skew symmetric and normalized by the condition

$$\epsilon^{123}=\epsilon_{123}=1 .$$

Of course, it has to be checked that S^{ijk} and $\overline{S}_{ijk}$ really are symmetric in i,j,k.

Similarly, we define the symmetric sl(3)-invariant trilinear mappings

$$T\colon W_0^3\to V_0,\quad \overline{T}\colon W_0^3\to\overline{V}_0,$$

by

$$T^{ijk}(A',A',A')=(A'A')^i{}_pA'^j{}_q\epsilon^{pqk},$$

$$\overline{T}_{ijk}(A',A',A')=(A'A')^p{}_iA'^q{}_j\epsilon_{pqk} .$$

Since T and $\overline{T}$ are symmetric it is sufficient to specify them for coinciding arguments.

Our ansatz for the restriction of ϕ onto $W_0^3\times W_1^3$ now reads

$$\begin{aligned}\phi(A',A',A',C'^1,C'^2,C'^3)&=e\overline{T}_{ijk}(A',A',A')S^{ijk}(C'^1,C'^2,C'^3)\\&\quad+fT^{ijk}(A',A',A')\overline{S}_{ijk}(C'^1,C'^2,C'^3),\end{aligned}$$

where e and f are some constants to be fixed later.

One might hope to obtain a relation between e and f by calculating $\phi(X',X',X',F_1',F_2',F_3')$ with $X'=\Sigma x_iE'_{ii}$ and demanding that this should be proportional to

$$(x_1-x_2)(x_1-x_3)(x_2-x_3),$$

see Proposition 3. However, with these arguments both $\overline{T}S$ and $T\overline{S}$ are equal to

$$(x_1-x_2)(x_1-x_3)(x_2-x_3)$$

and hence no relation for e and f can be derived this way.

This is remarkable: The requirement for $P(2)_0$ invariance, supersymmetry, and the "correct" form of $\phi(X',X',X',F_1',F_2',F_3')$, $X'=\Sigma x_iE'_{ii}$, is not sufficient to fix the restriction of ϕ onto $W_0^3\times W_1^3$.

Let us next construct the ansatz for the restriction of ϕ onto $W_{-1}\times W_0\times W_1^4$. The fourth exterior power $\wedge W_1$ of the sl(3) module W_1 is isomorphic to the irreducible 15-dimensional sl(3) module M consisting of the traceless tensors (Q^{ij}_k) which are symmetric in the upper indices. The corresponding dual module $\overline{M}$ consists of the traceless tensors $(\overline{Q}^k_{ij})$ which are symmetric in the lower indices. Since the tensor product $W_{-1}\otimes W_0$ contains a unique sl(3) sub-

module isomorphic to $\overline{M}$, we conclude that there is, up to the normalization, just one nonzero sl(3)-invariant six-linear form on $W_{-1}\times W_0\times W_1^4$ which has the correct symmetry property, and this form vanishes if the argument from W_0 is equal to I.

To construct this form let $\overline{M}_0$ denote the reducible tensor space consisting of all tensors of the form $(\overline{Q}^{\,k}_{ij})$. Then the bilinear mapping

$$\overline{R}\colon W_{-1}\times W_0\to\overline{M}_0\,,$$

defined by

$$\overline{R}^{\,k}_{ij}(B',A') = \epsilon_{pqi}B'^{pq}A'^{k}_{\ j}\,,$$

and the skew-symmetric quadrilinear mapping

$$R\colon W_1^4\to M\,,$$

defined by

$$R^{ij}_{\ k}(C'^1,C'^2,C'^3,C'^4) = C'^{a}_{kp}C'^{b}_{sq}C'^{c}_{tr}C'^{d}_{vw}\epsilon^{pqr}\epsilon^{svi}\epsilon^{twj}\epsilon_{abcd}\,,$$

are sl(3) invariant. The ϵ symbol with four indices is skew symmetric and normalized such that $\epsilon_{1234}=1$.

Actually, $\overline{R}$ yields an sl(3)-module isomorphism of $W_{-1}\otimes W_0$ onto $\overline{M}_0$ and R yields an sl(3)-module isomorphism of $\overset{4}{\wedge} W_1$ onto M. According to our foregoing remarks, it follows from the sl(3) invariance that $(R^{ij}_{\ k})$ must be traceless and symmetric in the upper indices, but this can also be checked directly, of course.

Our ansatz for the restriction of ϕ onto $W_{-1}\times W_0\times W_1^4$ now reads

$$\phi(B',A',C'^1,C'^2,C'^3,C'^4) = g\overline{R}^{\,k}_{ij}(B',A')R^{ij}_{\ k}(C'^1,C'^2,C'^3,C'^4)\,,$$

where g is one more constant to be fixed.

The main task is then to find the conditions on e,f,g under which the form ϕ is $P(2)$ invariant. It turns out that this is the case if and only if

$$e=-12g,\quad f=-24g\,.$$

The proof is quite tedious: I have repeatedly used the sl(3) invariance as well as several tricks to make the calculations feasible.

Thus the multilinear form we have been looking for does indeed exist. Correspondingly, the algebra $P(2)$ has a Casimir element of order 6 which is Z homogeneous of degree -3. To my knowledge this is the first example of a nontrivial odd Casimir element of a Lie superalgebra. Unfortunately, it is quite degenerate and I cannot imagine how it could be useful.

More efficient techniques have to be developed before we can proceed to construct higher-order Casimir elements or to tackle the cases $n\geqslant 4$.

Our results show that for $n=1,2,3$ the Lie superalgebra $P(n-1)$ has a Casimir element of the lowest possible order $\frac{1}{2}n(n+1)$. I conjecture that this should be true for all n, even if the much more general question asked at the end of Sec. V should be answered in the negative.

APPENDIX: MULTILINEAR ALGEBRA WITH ϵ-COMMUTING SCALARS

The use of a Grassmann algebra as the domain of scalars in multilinear algebra and analysis is an important tool for dealing with Bose–Fermi symmetry (see Ref. 16 and the literature cited therein). For the convenience of the reader, I want to describe the pertinent algebraic structures in a language most suitable for our purposes and to comment on some special results which have been used in the proof of the basic lemma in Sec. II. Also, I take the opportunity to present the material in its general formal setting by allowing for arbitrary Abelian groups of degrees and arbitrary commutation factors. Here and in the sequel we use without comment the notation and results of Refs. 1 and 20.

In the following, Γ denotes an Abelian group and ϵ a commutation factor on Γ with values in K. Moreover, S is a Γ-graded associative ϵ-commutative algebra over K with a nonzero unit element (necessarily homogeneous of degree zero). Thus K can be identified with a subalgebra of S. All modules V over S are assumed to be unitary in the sense that under scalar multiplication the unit element of S acts as the identity operator on V. In particular, any S module has a canonical structure of a vector space over K. We follow the convention of Ref. 1 according to which the degree of a homogeneous element is denoted by the "corresponding" lowercase Greek letter. The gradations that are going to appear all have Γ as their group of degrees. Thus we can simplify the notation and speak about graded vector spaces, graded algebras, etc. without further specification.

1. Some basic definitions and results

Our subsequent discussion is based on the following elementary observation. If V is a left (resp. right) graded S module,[23,24] we can introduce on V a new structure of a right (resp. left) graded S module by keeping the addition and the gradation and introducing the scalar multiplication through the equation

$$xs=\epsilon(\xi,\sigma)sx\quad[\text{resp. } sx=xs\epsilon(\sigma,\xi)]$$

for all homogeneous elements $x\in V$ and $s\in S$. In both cases, the left and the right S-module structures on V are compatible in the sense that

$$(sx)s'=s(xs')$$

for all $x\in V$ and $s,s'\in S$. Note also that the transitions from left to right and from right to left graded S modules are inverse to each other in the obvious sense and that the underlying structure of a graded vector space over K remains unchanged under these transitions.

Whenever we shall speak about a graded S module V it is always understood that V is endowed with the structures of a left and of a right graded S module and that the two are related as described above.

Let $V_1,\ldots,V_n$, W be graded S modules and let $\gamma\in\Gamma$. Consider the set $Lgr_{n,S}(V_1,\ldots,V_n;W)_\gamma$ of all n-additive mappings g of $V_1\times\cdots\times V_n$ into W, which are homogeneous of degree γ and satisfy the conditions

$$g(sx_1,x_2,\ldots,x_n)=\epsilon(\gamma,\sigma)sg(x_1,\ldots,x_n)\,,$$

$$g(x_1,...,x_r,s,x_{r+1},...,x_n) = g(x_1,...,x_r,sx_{r+1},...,x_n)\,,$$
$$\text{if } 1 \leqslant r \leqslant n-1\,,$$
$$g(x_1,...,x_{n-1},x_n s) = g(x_1,...,x_n)s\,,$$

for all elements $x_i \in V_i$ and all homogeneous elements $s \in S$. (Note that each of these $n+1$ conditions is a consequence of the other n ones.) Obviously, this set is a subspace of $Lgr_n(V_1,...,V_n; W)_\gamma$. The sum of these subspaces of $Lgr_n(V_1,...,V_n; W)$ (with $\gamma \in \Gamma$) is direct and will be denoted by $Lgr_{n,S}(V_1,...,V_n; W)$. This is a graded subspace of $Lgr_n(V_1,...,V_n; W)$ whose elements are called the graded S-multilinear mappings of $V_1 \times \cdots \times V_n$ into W.

It is easy to introduce on $Lgr_{n,S}(V_1,...,V_n; W)$ a structure of a left graded S module: If $g \in Lgr_{n,S}(V_1,...,V_n; W)$ and $s \in S$, we define the mapping $sg \in Lgr_{n,S}(V_1,...,V_n; W)$ by

$$(sg)(x_1,...,x_n) = sg(x_1,...,x_n) \text{ for all } x_i \in V_i\,.$$

The corresponding structure of a right graded S module is then given by

$$(gs)(x_1,...,x_n) = g(sx_1,x_2,...,x_n)$$

(same notation as before).

Of course, the case $n = 1$ is most important and should have been considered first. In this case, we simplify the notation and write Lgr_S instead of $Lgr_{1,S}$. If V and W are two graded S modules, the elements of $Lgr_S(V,W)$ are called the graded S-linear mappings of V into W. Note that $Lgr_S(V,W)$ consists of all S-linear mappings (in the usual sense) of the right S module V into the right S module W which belong to $Lgr(V,W)$.

The elements of $Lgr_S(V,W)_0$ are called homomorphisms of the graded S module V into the graded S module W. Isomorphisms of graded S modules are defined correspondingly.

The definition of graded multilinearity has been chosen such that the classical relation between tensor products and multilinear mappings remains valid. If $V_1,...,V_n$ are graded S modules, their tensor product $V_1 \underset{S}{\otimes} \cdots \underset{S}{\otimes} V_n$ is, at first, only an Abelian group which is generated by the decomposable tensors $x_1 \otimes \cdots \otimes x_n$, with $x_i \in V_i$. By definition, these tensors satisfy the equations

$$x_1 \otimes \cdots \otimes x_r s \otimes x_{r+1} \otimes \cdots \otimes x_n = x_1 \otimes \cdots \otimes x_r \otimes s x_{r+1} \otimes \cdots \otimes x_n$$

for all $x_i \in V_i$ and all $s \in S$. Nevertheless, the tensor product has a natural gradation and natural left and right S-module structures which convert it into a graded S module. The gradation is fixed by the requirement that $x_1 \otimes \cdots \otimes x_n$ is homogeneous of degree $\xi_1 + \cdots + \xi_n$ if the elements $x_i \in V_i$ are homogeneous of degree ξ_i, and the S-module structures are given by the equations

$$s(x_1 \otimes \cdots \otimes x_n) = (sx_1) \otimes x_2 \otimes \cdots \otimes x_n\,,$$
$$(x_1 \otimes \cdots \otimes x_n)s = x_1 \otimes \cdots \otimes x_{n-1} \otimes (x_n s)\,,$$

for all $x_i \in V_i$ and all $s \in S$.

Classical results on tensor products now imply that for any mapping $g \in Lgr_{n,S}(V_1,...,V_n; W)$ there exists a unique mapping $\tilde{g} \in Lgr_S(V_1 \underset{S}{\otimes} \cdots \underset{S}{\otimes} V_n, W)$ such that for all $x_i \in V_i$

$$g(x_1,...,x_n) = \tilde{g}(x_1 \otimes \cdots \otimes x_n)\,,$$

and the assignment $g \to \tilde{g}$ defines an isomorphism of the graded S module $Lgr_{n,S}(V_1,...,V_n; W)$ onto the graded S module $Lgr_S(V_1 \underset{S}{\otimes} \cdots \underset{S}{\otimes} V_n, W)$.

The foregoing discussion immediately leads to a series of definitions concerning algebras and their modules.

(a) A graded S algebra is a graded S module A, endowed with a graded S-bilinear mapping of $A \times A$ into A (the product mapping), which is homogeneous of degree zero. By restricting the domain of scalars, any graded S algebra can be considered as a graded K algebra.

(b) Associativity and ϵ commutativity of graded S algebras are defined as usual. For any graded S module V, the graded S module $Lgr_S(V,V)$, endowed with the usual multiplication (i.e., composition) of mappings, is an associative graded S algebra.

(c) An ϵ Lie algebra over S is a graded S algebra, whose multiplication is ϵ skew-symmetric and satisfies the ϵ Jacobi identity. Any associative graded S algebra is converted into an ϵ Lie algebra over S if the original multiplication is replaced by the ϵ commutator. In particular, if V is a graded S module, we can apply this remark to $Lgr_S(V,V)$ and obtain an ϵ Lie algebra over S which will be denoted by $gl_S(V,\epsilon)$.

(d) Let A be an associative graded S algebra or an ϵ Lie algebra over S. A graded representation of A in a graded S module V is a homomorphism ρ of the graded S algebra A into the graded S algebra $Lgr_S(V,V)$ or $gl_S(V,\epsilon)$, respectively. (According to our conventions this implies that ρ is homogeneous of degree zero.) A graded S module V that is endowed with a graded representation of A is called a (left) graded A module over S. Equivalently, this definition can be formulated as follows. A graded A module V over S is at the same time a graded S module and a graded A module over K. Both of these structures are built over the same graded vector space structure of V and they are related by the requirement that the product mapping of $A \times V$ into V be graded S bilinear.

In a systematic presentation of the theory we would now have to go through the basic constructions with graded algebras and modules (over K) and to investigate how these can be generalized to the present setting. We do not want to embark on this boring exercise but rather mention some special results.

In the Lie case, we comment on a few sections of Ref. 1. Let L denote an ϵ Lie algebra over S.

Section 3 of Ref. 1 can immediately be transcribed, with the sole proviso that the canonical homomorphism π might no longer be injective. In particular, the definition of the graded tensor product of linear mappings remains valid, and the graded tensor product of graded L modules over S is defined as usual and is still associative.

In order to generalize Sec. 4 of Ref. 1, let $V_1,..,V_n$, W be graded L modules over S. Then the usual action of L makes the graded S modules $Lgr_{n,S}(V_1,...,V_n; W)$ and $Lgr_S(V_1 \underset{S}{\otimes} \cdots \underset{S}{\otimes} V_n, W)$ into graded L modules over S and the canonical mapping $g \to \tilde{g}$ considered above is an isomorphism of graded L modules over S.

In Sec. 5 of Ref. 1 the action of the symmetric group on graded tensor products is discussed. Almost all the results of this section can immediately be transcribed to the present setting. In particular, the symmetry transformations S_π and $\check{S}_\pi$ are defined as usual. A reservation has to be made only in connection with the representation theory of the symmetric group. This theory is usually formulated over a commutative field, whereas in the present case an analogous theory over S would be required.

In the associative case, we mention the following results. Let $A_1,...,A_n$ be associative graded S algebras. Then there exists a unique multiplication in $A_1 \underset{S}{\otimes} \cdots \underset{S}{\otimes} A_n$ such that

$$(a_1 \otimes \cdots \otimes a_n)(b_1 \otimes \cdots \otimes b_n)$$
$$= \prod_{i>j} \epsilon(\alpha_i, \beta_j) a_1 b_1 \otimes \cdots \otimes a_n b_n ,$$

for all homogeneous elements $a_i, b_i \in A_i$. Endowed with its structure of a graded S module and this multiplication, $A_1 \underset{S}{\otimes} \cdots \underset{S}{\otimes} A_n$ is an associative graded S algebra which is called the graded tensor product of the A_i and will be denoted by $A_1 \underset{S}{\overline{\otimes}} \cdots \underset{S}{\overline{\otimes}} A_n$.

Now let ρ_i be a graded representation of the associative graded S algebra A_i in a graded S module V_i, for $1 \leqslant i \leqslant n$. Then there exists a unique graded representation ρ of the associative graded S algebra $A_1 \underset{S}{\overline{\otimes}} \cdots \underset{S}{\overline{\otimes}} A_n$ in the graded S module $V_1 \underset{S}{\otimes} \cdots \underset{S}{\otimes} V_n$ such that

$$\rho(a_1 \otimes \cdots \otimes a_n) = \rho_1(a_1) \overline{\otimes} \cdots \overline{\otimes} \rho_n(a_n)$$

[graded tensor product of the $\rho_i(a_i)$] for all $a_i \in A_i$. The representation ρ is called the graded tensor product of the representations ρ_i.

2. Extension of the domain of scalars from K to S

In practice, most of the graded S modules and S algebras arise through an extension of the domain of scalars from K to S, a process which we are now going to describe.

Let V be a graded vector space over K. Then the graded vector space $S \otimes V$ (tensoring with respect to K) has a natural structure of a left graded S module such that

$$s'(s \otimes x) = (s's) \otimes x$$

for all $s,s' \in S$ and all $x \in V$. The corresponding structure of a right graded S module is given by

$$(s \otimes x)s' = \epsilon(\xi,\sigma')(ss') \otimes x$$

for all homogeneous elements $s,s' \in S$ and $x \in V$.

Similarly, $V \otimes S$ has a natural structure of a right graded S module such that

$$(x \otimes s)s' = x \otimes (ss')$$

for all $x \in V$ and $s,s' \in S$, and the corresponding structure of a left graded S module is given by

$$s'(x \otimes s) = \epsilon(\sigma',\xi) x \otimes (s's)$$

for all homogeneous elements $x \in V$ and $s,s' \in S$.

Now it is easy to see that there exists a unique additive mapping of $S \otimes V$ into $V \otimes S$ that maps $s \otimes x$ onto $\epsilon(\sigma,\xi) x \otimes s$, for all homogeneous elements $s \in S$ and $x \in V$. This mapping is an isomorphism of graded S modules. Thus there is no loss of generality if we restrict our attention to one of these modules. From habit I work with $S \otimes V$ although for $V \otimes S$ some of the subsequent formulas would simplify a little. We say that the S module $S \otimes V$ is obtained from V by extension of the domain of scalars from K to S.

Let $V_1,...,V_n$ be graded vector spaces (over K). Then there exists a unique canonical isomorphism of graded S modules

$$(S \otimes V_1) \underset{S}{\otimes} (S \otimes V_2) \underset{S}{\otimes} \cdots \underset{S}{\otimes} (S \otimes V_n)$$
$$\to S \otimes (V_1 \otimes \cdots \otimes V_n)$$

such that

$$(s_1 \otimes x_1) \otimes \cdots \otimes (s_n \otimes x_n)$$
$$\to \prod_{i<j} \epsilon(\xi_i,\sigma_j) s_1 \cdots s_n \otimes (x_1 \otimes \cdots \otimes x_n)$$

for all homogeneous elements $x_i \in V_i$ and $s_j \in S$.

We are now ready to discuss the extension of multilinear mappings. The linear case is trivial. If V and W are graded vector spaces and if $g \in Lgr(V,W)$, then $\mathrm{id}_S \overline{\otimes} g$ (graded tensor product) is an element of $Lgr_S(S \otimes V, S \otimes W)$ which will be called the graded S-linear extension of g onto $S \otimes V$.

Consider next the graded vector spaces $V_1,...,V_n$, W and a mapping $g \in Lgr_n(V_1,...,V_n; W)$. Let $\tilde{g}$ be the corresponding K-linear mapping of $V_1 \otimes \cdots \otimes V_n$ into W. Its extension $\mathrm{id}_S \overline{\otimes} \tilde{g}$ belongs to

$$Lgr_S(S \otimes (V_1 \otimes \cdots \otimes V_n), S \otimes W).$$

Composing it with the isomorphism constructed above, we obtain an element of

$$Lgr_S((S \otimes V_1) \underset{S}{\otimes} \cdots \underset{S}{\otimes} (S \otimes V_n), S \otimes W)$$

which canonically corresponds to an element

$$\hat{g} \in Lgr_{n,S}(S \otimes V_1,...,S \otimes V_n; S \otimes W).$$

If g is homogeneous of degree γ, then

$$\hat{g}(s_1 \otimes x_1,...,s_n \otimes x_n)$$
$$= \epsilon\Big(\gamma, \sum_k \sigma_k\Big) \prod_{i<j} \epsilon(\xi_i,\sigma_j) s_1 \cdots s_n \otimes g(x_1,...,x_n)$$

for all homogeneous elements $x_i \in V_i$ and $s_j \in S$. The mapping $\hat{g}$ is called the graded S-multilinear extension of g onto $(S \otimes V_1) \times \cdots \times (S \otimes V_n)$. Of course we have $\hat{g} = \mathrm{id} \overline{\otimes} g$ if $n = 1$.

The foregoing constructions will now be applied to graded algebras and their modules. Let A be a graded algebra over K. We make the graded S module $S \otimes A$ into a graded S algebra by demanding that the product mapping of $S \otimes A$ be the graded S-bilinear extension of the product mapping of A. Thus the multiplication in $S \otimes A$ is fixed by

$$(s \otimes a)(s' \otimes a') = \epsilon(\alpha,\sigma')(ss') \otimes (aa')$$

for all homogeneous elements $s,s' \in S$ and $a,a' \in A$. We say that the algebra $S \otimes A$ is obtained from A by extension of the domain of scalars from K to S. It is easy to verify the following statements.

(1) If A' is a second graded K algebra and if $f: A \to A'$ is a homomorphism of graded K algebras, then $\mathrm{id} \otimes f$: $S \otimes A \to S \otimes A'$ is a homomorphism of graded S algebras.

(2) If A is associative or ϵ commutative, then $S \otimes A$ is likewise.

(3) If e is a unit element of A, then $1 \otimes e$ is a unit element of $S \otimes A$.

(4) If A is an ϵ Lie algebra over K, then $S \otimes A$ is an ϵ Lie algebra over S.

Suppose next that A is an associative graded algebra or an ϵ Lie algebra over K. If V is a graded A module, we convert the graded S module $S \otimes V$ into a graded $S \otimes A$ module over S by demanding that the product mapping of $(S \otimes A) \times (S \otimes V)$ into $S \otimes V$ be the graded S-bilinear extension of the product mapping of $A \times V$ into V. Thus we have

$$(s \otimes a)(s' \otimes x) = \epsilon(a,\sigma')(ss') \otimes (ax)$$

for all homogeneous elements $s,s' \in S$, $a \in A$, and $x \in V$. Of course, it has to be checked that this prescription really defines a graded $S \otimes A$ module over S. Instead, we give a different but equivalent defintion from which this will be obvious.

We begin with a preparatory remark. Let V be a graded vector space and $g \in Lgr(V,V)$, $s \in S$. We know that $\mathrm{id} \bar{\otimes} g$, the graded S-linear extension of g, belongs to $Lgr_S(S \otimes V, S \otimes V)$. Thus the same is true for $s(\mathrm{id} \bar{\otimes} g)$; let us denote this mapping by $s \bar{\otimes} g$. Obviously, there exists a unique K-linear mapping

$$S \otimes Lgr(V,V) \to Lgr_S(S \otimes V, S \otimes V)$$

such that

$$s \otimes g \to s \bar{\otimes} g$$

for all $s \in S$ and $g \in Lgr(V,V)$, and it is easy to check that this is, in fact, a homomorphism of graded S algebras. Obviously, the same holds true if this is considered to be a mapping of $S \otimes \mathrm{gl}(V,\epsilon)$ into $\mathrm{gl}_S(S \otimes V, \epsilon)$.

Now let V be a graded A module and let ρ be the corresponding homomorphism of the graded algebra A into the graded algebra $Lgr(V,V)$ or $\mathrm{gl}(V,\epsilon)$, respectively. Then $\mathrm{id} \otimes \rho$ is a homomorphism of the graded S algebra $S \otimes A$ into the graded S algebra $S \otimes Lgr(V,V)$ or $S \otimes \mathrm{gl}(V,\epsilon)$, respectively. Composed with the homomorphism above, we thus obtain a graded representation $\hat{\rho}$ of the graded S algebra $S \otimes A$ in the graded S module $S \otimes V$. By definition, we have

$$\begin{aligned}(\hat{\rho}(s \otimes a))(s' \otimes x) &= (s \bar{\otimes} \rho(a))(s' \otimes x) \\ &= \epsilon(a,\sigma')ss' \otimes \rho(a)x\end{aligned}$$

for all homogeneous elements $s,s' \in S$, $a \in A$, and $x \in V$, which is exactly the prescription given above. We say that the representation $\hat{\rho}$ (resp. the $S \otimes A$ module $S \otimes V$) is obtained from the representation ρ (resp. from the A module V) by extension of the domain of scalars from K to S.

In the following, we restrict our attention to the Lie case. Let L be an ϵ Lie algebra (over K) and let $V_1,\ldots,V_n$ be graded L modules. We have already defined a canonical isomorphism of graded S modules

$$\begin{aligned}&(S \otimes V_1) \underset{S}{\otimes} (S \otimes V_2) \underset{S}{\otimes} \cdots \underset{S}{\otimes} (S \otimes V_n) \\ &\quad \to S \otimes (V_1 \otimes \cdots \otimes V_n)\,.\end{aligned}$$

Under the present assumptions, both of these modules have a natural structure of a graded $S \otimes L$ module over S, and it is easy to see that our mapping is even an isomorphism of these modules.

Consider one more graded L module W. For any element $g \in Lgr_n(V_1,\ldots,V_n; W)$ we have defined its graded S-multilinear extension $\hat{g}$ which belongs to $Lgr_{n,S}(S \otimes V_1,\ldots,S \otimes V_n; S \otimes W)$. Consequently, there exists a unique K-linear mapping

$$S \otimes Lgr_n(V_1,\ldots,V_n; W) \to Lgr_{n,S}(S \otimes V_1,\ldots,S \otimes V_n; S \otimes W)$$

such that

$$s \otimes g \to s\hat{g}$$

for all $s \in S$ and $g \in Lgr_n(V_1,\ldots,V_n; W)$. Under our assumptions, both spaces have a natural structure of a graded $S \otimes L$ module over S, and it is easy to see that our mapping is, in fact, a canonical homomorphism of these modules. In particular, if $g \in Lgr_n(V_1,\ldots,V_n; W)$ is L invariant, its graded S-multilinear extension $\hat{g}$ is $S \otimes L$ invariant.

Next we want to comment on the action of the symmetric group. We keep the notation above. Then for any permutation π of $\{1,\ldots,n\}$ and any $g \in Lgr_n(V_{\pi(1)},\ldots,V_{\pi(n)}; W)$ we have (with a slight abuse of notation)

$$\check{S}_\pi \hat{g} = (\check{S}_\pi g)\hat{}\,.$$

In particular, if $V_1 = \cdots = V_n = V$ and if g is ϵ symmetric, then so is $\hat{g}$. Of course, an analogous (dual) commutativity result holds for $S \otimes (V_1 \otimes \cdots \otimes V_n)$ and $(S \otimes V_1) \underset{S}{\otimes} \cdots \underset{S}{\otimes} (S \otimes V_n)$.

Finally, let us comment on how the graded S modules $S \otimes V$, with V a graded vector space, are characterized within the class of all graded S modules. The answer is simple: A graded S module $\bar{V}$ is isomorphic to some $S \otimes V$, with V a graded vector space over K, if and only if it has a homogeneous basis over S.

Some care is needed to understand this result. *A priori*, we have to distinguish between bases of the left and of the right S module $\bar{V}$. Obviously, there are examples where a graded S module has no basis whatsoever. Moreover, it is conceivable (although I have not tried to construct an example) that a basis of the left S module $\bar{V}$ is not necessarily a basis of the right S module $\bar{V}$ (and vice versa). In any case, if a family of elements of $\bar{V}$ is homogeneous (i.e., consists of homogeneous elements only), then it is a basis of the left S module $\bar{V}$ if and only if it is a basis of the right S module $\bar{V}$, and it may then be called a homogeneous basis of the graded S module $\bar{V}$ without further specification. The reader is warned that the number of basis elements which are homogeneous of a fixed degree $\gamma \in \Gamma$ may be different for different homogeneous bases, the most obvious reason being that S may contain invertible elements which are homogeneous of nonzero degree.

We could now proceed to develop a graded calculus for matrices over S; however, this would extend the Appendix beyond a reasonable size. Instead, we refer the reader to Ref. 25 where substantial results of this type have been obtained.

[1]M. Scheunert, J. Math. Phys. **24**, 2658 (1983).
[2]M. Scheunert, J. Math. Phys. **24**, 2671 (1983).

[3]M. Scheunert, J. Math. Phys. **24**, 2681 (1983).
[4]V. G. Kac, Adv. Math. **26**, 8 (1977).
[5]M. Scheunert, *The Theory of Lie Superalgebras, Lecture Notes in Mathematics*, Vol. 716 (Springer, Berlin, 1979).
[6]F. A. Berezin, Inst. Theor. Exp. Phys. ITEP-66, 75-78 (5 parts), Moscow (1977).
[7]V. G. Kac, Commun. Algebra **5**, 889 (1977).
[8]V. G. Kac, in *Differential Geometrical Methods in Mathematical Physics, Lecture Notes in Mathematics*, Vol. 676, edited by K. Bleuler, H. R. Petry, and A. Reetz (Springer, Berlin, 1978), p. 597.
[9]P. D. Jarvis and H. S. Green, J. Math. Phys. **20**, 2115 (1979).
[10]A. N. Sergeev, C. R. Acad. Bulgare Sci. **35**, 573 (1982).
[11]A. M. Bincer, J. Math. Phys. **24**, 2546 (1983).
[12]V. G. Kac, Proc. Natl. Acad. Sci. USA **81**, 645 (1984).
[13]C. O. Nwachuku and M. A. Rashid, J. Math. Phys. **26**, 1914 (1985).
[14]A. N. Sergeev, Lett. Math. Phys. **7**, 177 (1983).
[15]P. D. Jarvis and M. K. Murray, J. Math. Phys. **24**, 1705 (1983).
[16]B. DeWitt, *Supermanifolds* (Cambridge U. P., Cambridge, 1984).
[17]Séminaire "Sophus Lie," "Théorie des algèbres de Lie. Topologie des groupes de Lie," exposé 15, Paris, École Norm. Sup., 1954–55.
[18]N. Bourbaki, *Groupes et algèbres de Lie* (Hermann, Paris, 1975), Chap. 7, Secs. 2 and 3.
[19]J. Diximer, *Algèbres enveloppantes* (Gauthier–Villars, Paris, 1974).
[20]M. Scheunert, J. Math. Phys. **20**, 712 (1979).
[21]In Ref. 1, W has been denoted by W^1.
[22]M. Scheunert, in *Supersymmetry*, edited by K. Dietz, R. Flume, G. von Gehlen, and V. Rittenberg (Plenum, New York, 1985), p. 421.
[23]N. Bourbaki, *Algèbre* (Hermann, Paris, 1962), 3rd ed., Chap. II.
[24]N. Bourbaki, *Algèbre* (Hermann, Paris, 1971), new ed., Chap. III.
[25]Y. Kobayashi and S. Nagamachi, J. Math. Phys. **25**, 3367 (1984).

Generalized quasispin for supergroups

P. D. Jarvis
Department of Physics, University of Tasmania, Hobart, Tasmania, 7001, Australia

Mei Yang and B. G. Wybourne
Department of Physics, University of Canterbury, Christchurch, New Zealand

(Received 16 September 1986; accepted for publication 18 November 1986)

The embedding of the dynamical algebra $U(M/N)$ of nuclear supersymmetries in larger algebraic structures is studied. A noncompact $\mathbb{Z}_2 \oplus \mathbb{Z}_2$ graded color superalgebra $SpO(2M/1/2N/0)$ is identified as a receptacle for various chains containing boson and fermion (super) algebras. The existence of a generalized quasispin algebra is demonstrated and discussed.

I. INTRODUCTION

Recent experimental and theoretical interest in nuclear[1] (and other[2]) dynamical supersymmetries has emphasized the need for the study of the underlying algebraic structures beyond the finite-dimensional irreps of compact forms of $U(M/N)$ (see Refs. 3–6). A recent extension[7] of the ideas of supersymmetric interacting boson models to explore the role of fermion pairing (via a seniority scheme) used an intervening $OSp(M/N)$ subalgebra. Although phenomenologically reasonable it suffers[7] from unusual features of nonconservation of nucleon number and non-Hermitian interaction V_{BF}, and it has recently been shown[8,9] that the embedding in $U(M/N)$ involves indecomposable representations. Other approaches[10,11] to nuclear collective models using noncompact Lie algebras [e.g., $Sp(6,R) \supset U(3)$] also have supersymmetric enlargements in terms of noncompact $OSp(M/N)$ superalgebras[12]; recent attempts to apply supersymmetric IBM models consistently over a range of nuclei[13] also suggest larger structures. Finally, an eventual microscopic foundation of the IBM ideas will presumably involve infinite-dimensional mappings.[14]

In the present paper we identify a noncompact $\mathbb{Z}_2 \oplus \mathbb{Z}_2$ color superalgebra[15–19] $SpO(2M/1/2N/0)$ as a natural receptacle for various chains containing boson and fermion (super) algebras. The first $\mathbb{Z}_2$ corresponds to the usual Fermi–Bose sign factor, while the additional grading arises from the inclusion of generators both linear (*odd*) and bilinear (*even*) in the boson–fermion realization. The assignment of gradings and commutation factor is given in (2.10). It is shown in Sec. II below that $SpO(2M/1/2N/0)$ has as subalgebras (superalgebras) both the usual fermionic $O(2N+1)$ and bosonic $SpO(2M/1)$ (an alternative to the Heisenberg algebra). Its Fock space realization[8,12] comprises one irrep with just two constituents with respect to $SpO(2M/2N)$. The Casimir invariant is a specific linear combination of number operators and suitably defined pairing operators. Indeed the latter are identified with a generalized quasispin algebra $Sp_{\pm}(2)$ occurring in the $Sp_{\pm}(2) \times OSp_{\pm}(M/N)$ subalgebra of $SpO(2M/1/2N/0)$ (the $\pm$ correspond to equivalent choices, interchanged by Hermitian conjugation). Finally, we discuss briefly the significance of the generalized quasispin algebra.

II. THE COLOR SUPERALGEBRA

The construction of ordinary Lie algebras [e.g., $SO(2k)$, $SO(2k+1)$, $U(k)$, $Sp(2k)$, etc.] using either boson or fermion operators is well known. In recent times Lie superalgebras of various sorts [e.g., $U(M/N)$, $OSp(M/N)$, etc.] have been constructed using both boson and fermion operators. In this section we establish the existence of a $\mathbb{Z}_2 \oplus \mathbb{Z}_2$ color superalgebra, which contains generators both linear and bilinear in terms of boson and fermion operators, with a generalized quasispin subgroup.

After the gradation of the generators is established we show that the system satisfies closure under the supersymmetric commutation factor and hence forms a color superalgebra which we shall denote $SpO(2M_{(0,0)}/1_{(0,1)}/2N_{(1,0)}/0_{(1,1)})$, where $2M,1,2N,0$ refer to the dimension, and subscripts to the double grading, of the underlying vector spaces $V_{(0,0)}$, $V_{(0,1)}$, $V_{(1,0)}$, and $V_{(1,1)}$, respectively, and the symbol SpO reminds us that the maximal Lie superalgebra is $SpO(2M/2N)$ where the bosons span $Sp(2M)$ and the fermions span $SO(2N)$ irreps.

Some care must be exercised in establishing a concise and consistent notation. Boson and fermion operators will be collectively designated by the operators C_A^{ϵ}, where $\epsilon = 0$ or 1 refers to annihilation or creation operator, respectively, and $A = a$ or α where the Latin boson index $a = l_a m_a$ $(-a \equiv l_a, -m_a)$ belongs to the boson indexed set B

$$a \in B, \quad B \equiv \{a = l_a, m_a \,|\, m_a = \pm l_a, \pm(l_a - 1), \ldots, \pm 1, 0;\ l_a \text{ positive integer}\} \tag{2.1a}$$

and the Greek fermion index $\alpha = j_\alpha m_\alpha$ $(-\alpha \equiv j_\alpha, -m_\alpha)$ belongs to the fermion indexed set F

$$\alpha \in F, \quad F \equiv \{\alpha = j_\alpha, m_\alpha \,|\, m_\alpha = \pm j_\alpha, \pm(j_\alpha - 1), \ldots, \pm \tfrac{1}{2};\ j_\alpha \text{ positive half-integer}\}\,. \tag{2.1b}$$

 0022-2488/87/051192-06$02.50

We define the total index set as $I = B \cup F$ with $A = j_A, m_A \in I(-A \equiv j_A, -m_A)$. Finally, C^0_A is defined as

$$C^0_A = (-)^{j_A \mp m_A}(C^1_{-A})^\dagger , \tag{2.2}$$

i.e., C^0_A and C^1_A are tensor operators under SO(3). Note that either choice of $-$, $+$ sign in the phase factor is acceptable and we will keep track of both in the subsequent discussions.

The number of distinct boson and fermion indices is given by

$$M = \sum_l (2l+1) , \tag{2.3a}$$

$$N = \sum_j (2j+1) , \tag{2.3b}$$

and the total number of C^ϵ_A operators is $2M + 2N$. Here $\{C^\epsilon_A, A \in I, \epsilon = 0,1\}$ forms a well-known $\mathbb{Z}_2$ graded vector space with the usual ϵ independent grading defined by

$$(C^\epsilon_a) = (a) = 0, \quad (C^\epsilon_\alpha) = (\alpha) = 1 , \tag{2.4}$$

i.e., bosons and fermions belong to even and odd subspaces of the same $\mathbb{Z}_2$ graded vector space

$$V(2M_{(0)}/2N_{(1)}) = V(2M)_{(0)} \oplus V(2N)_{(1)} , \tag{2.5}$$

commutation and anticommutation relations among the boson and fermion operators may be compactly written as

$$\langle C^\epsilon_A, C^\zeta_B \rangle = G^{\epsilon\zeta}_{AB}, \quad A,B \in I, \quad \epsilon, \zeta = 0 \text{ or } 1 , \tag{2.6}$$

where the supercommutator $\langle , \rangle$ is defined as

$$\langle C^\epsilon_A, C^\zeta_B \rangle = C^\epsilon_A C^\zeta_B - (-1)^{(A)(B)} C^\zeta_B C^\epsilon_A . \tag{2.7}$$

Here (A), (B) are the grading vectors (one dimensional) for C^ϵ_A and C^ζ_B, respectively, $(-1)^{(A)(B)}$ is the commutation factor which determines whether $\langle , \rangle$ is a commutator $[,]$ or an anticommutator $\{ , \}$, and $G^{\epsilon\zeta}_{AB}$ is the $\mathbb{Z}_2$ graded metric tensor containing $0, \pm 1$ as its elements. Explicitly,

$(G^{\epsilon\zeta}_{AB}) =$

$\langle C^\epsilon_A, C^\zeta_B \rangle$ \ C^ζ_B / C^ϵ_A	C^0_b	C^1_b	C^0_β	C^1_β
C^0_a	0	g_{ab}	0	0
C^1_a	$-g_{ab}$	0	0	0
C^0_α	0	0	0	$g_{\alpha\beta}$
C^1_α	0	0	$-g_{\alpha\beta}$	0

, (2.8)

where

$$g_{ab} \equiv \langle C^0_a, C^1_b \rangle = (-1)^{l_a \mp m_a} \delta_{-ab} = g^{ab} = g_{ba} , \tag{2.9a}$$

$$g_{\alpha\beta} \equiv \langle C^0_\alpha, C^1_\beta \rangle = (-1)^{j_\alpha \mp m_\alpha} \delta_{-\alpha\beta} = -g^{\alpha\beta} = -g_{\beta\alpha} , \tag{2.9b}$$

N.B.

$$G^{\epsilon\zeta}_{AB} = 0 \text{ if } \epsilon = \zeta \quad \text{or} \quad (A+B) = (A) + (B) = 1 . \tag{2.9c}$$

The metric tensor $G^{\epsilon\zeta}_{AB}$ permits considerable compactification of supercommutation calculations.

The $\mathbb{Z}_2$ graded vector space $V(2M_{(0)}/2N_{(1)})$ may be extended to a doubly graded $\mathbb{Z}_2 \oplus \mathbb{Z}_2$ vector space as follows. The index set $I = B \cup F$ is extended to $I' = B \cup F \cup \{*\}$, and $C^0_* \equiv C^1_* \equiv e_*$, the identity operator, is introduced. With the extended grading

$$(C^\epsilon_A) = [(A),0) = \begin{cases} (0,0), & \text{for bosons,} \\ (1,0), & \text{for fermions,} \end{cases} \tag{2.10a}$$

$$(C^0_*) = (C^1_*) = (e_*) = (0,1) . \tag{2.10b}$$

The basis $\{C^\epsilon_A | A \in I', \epsilon = 0,1\}$ will now span the doubly graded $\mathbb{Z}_2 \oplus \mathbb{Z}_2$ vector space $V(2M_{(0,0)}|1_{(0,1)}|2N_{(1,0)}|0_{(1,1)})$, where $2M,1,2N,0$ indicates the number of basis vectors (dimension) in each of the doubly graded subspaces $V_{(0,0)}$, $V_{(0,1)}$, $V_{(1,0)}$, and $V_{(1,1)}$, respectively.

A supercommutator on the $\mathbb{Z}_2 \oplus \mathbb{Z}_2$ graded vector space $V(2M_{(0,0)}/1_{(0,1)}/2N_{(1,0)}/0_{(1,1)})$ may be defined as

$$\langle C^\epsilon_A, C^\zeta_B \rangle = C^\epsilon_A C^\zeta_B - (-1)^{(A)\cdot(B)} C^\zeta_B C^\epsilon_A , \tag{2.11}$$

and the $\mathbb{Z}_2 \oplus \mathbb{Z}_2$ graded metric tensor $G^{\epsilon\zeta}_{AB}$ on the $\mathbb{Z}_2 \oplus \mathbb{Z}_2$ vector space is given explicitly by

$(G^{\epsilon\zeta}_{AB}) =$

$\langle C^\epsilon_A, C^\zeta_B \rangle$ \ C^ζ_B / C^ϵ_A	C^0_b	C^1_b	C^0_β	C^1_β	$\zeta e_*/\sqrt{2}$
C^0_a	0	g_{ab}	0	0	0
C^1_a	$-g_{ab}$	0	0	0	0
C^0_α	0	0	0	$g_{\alpha\beta}$	0
C^1_α	0	0	$-g_{\alpha\beta}$	0	0
$\zeta e_*/\sqrt{2}$	0	0	0	0	1

, (2.12)

where g_{ab} and $g_{\alpha\beta}$ are defined in the same way as, e.g., (2.9). Comparing the above definition with Eqs. (2.6)–(2.9) we notice that $A,B\in I'$ instead of I and in the commutation factor $(-1)^{(A)\cdot(B)}$ we have a scalar product $(A)\cdot(B)$ instead of the ordinary multiplication $(A)(B)$.

The direct product space of a $\mathbb{Z}_2\oplus\mathbb{Z}_2$ graded vector space with itself is also a $\mathbb{Z}_2\oplus\mathbb{Z}_2$ graded vector space where a typical element in the space $C_A^{\epsilon}C_B^{\zeta}$ has an induced grading

$$(C_A^{\epsilon}C_B^{\zeta})\equiv(C_A^{\epsilon})+(C_B^{\zeta})\in\mathbb{Z}_2\oplus\mathbb{Z}_2\,. \tag{2.13}$$

In particular, consider the operator $S_{AB}^{\epsilon\zeta}$ which belongs to the direct product space, and is defined as a superanticommutator among normalized basis vectors $\{(1/\sqrt{2})C_A^{\epsilon}, A\in I', \epsilon=0,1\}$, i.e.,

$$\begin{aligned} S_{AB}^{\epsilon\zeta}&\equiv\langle(1/\sqrt{2})C_A^{\epsilon},(1/\sqrt{2})C_B^{\zeta}\rangle_+ \quad A,B\in I',\quad \epsilon,\zeta=0,1\,,\\ &=1/2(C_A^{\epsilon}C_B^{\zeta}+(-1)^{(A)\cdot(B)}C_B^{\zeta}C_A^{\epsilon})\\ &=(-1)^{(A)\cdot(B)}S_{BA}^{\zeta\epsilon}\,. \end{aligned} \tag{2.14}$$

Explicitly, we have

$(S_{AB}^{\epsilon\zeta})=$

$\frac{1}{2}\langle C_A^{\epsilon},C_B^{\zeta}\rangle_+$ \ C_B^{ζ} / C_A^{ϵ}	$(1/\sqrt{2})C_b^0$	$(1/\sqrt{2})C_b^1$	$(1/\sqrt{2})C_\beta^0$	$(1/\sqrt{2})C_\beta^1$	$(1/\sqrt{2})e_*$
$(1/\sqrt{2})C_a^0$	$C_a^0C_b^0$	$C_a^0C_b^1-\frac{1}{2}(-)^{l_a\mp m_a}\delta_{-ab}$	$C_a^0C_\beta^0$	$C_a^0C_\beta^1$	$(1/\sqrt{2})C_a^0$
$(1/\sqrt{2})C_a^1$	$C_a^1C_b^1$		$C_a^1C_\beta^0$	$C_a^1C_\beta^1$	$(1/\sqrt{2})C_a^1$
$(1/\sqrt{2})C_\alpha^0$			$C_\alpha^0C_\beta^0$	$C_\alpha^0C_\beta^1-\frac{1}{2}(-)^{l_\alpha\mp m_\alpha}\delta_{-\alpha\beta}$	$(1/\sqrt{2})C_\alpha^0$
$(1/\sqrt{2})C_\alpha^1$				$C_\alpha^1C_\beta^1$	$(1/\sqrt{2})C_\alpha^1$
$(1/\sqrt{2})e_*$					0

(2.15)

The natural induced grading for $S_{AB}^{\epsilon\zeta}$ is given as

$$(S_{AB}^{\epsilon\zeta})\equiv(A+B)=(A)+(B)\in\mathbb{Z}_2\oplus\mathbb{Z}_2\,. \tag{2.16}$$

The evaluation of the supercommutator

$$\langle S_{AB}^{\epsilon\zeta},S_{CD}^{\sigma\tau}\rangle\equiv S_{AB}^{\epsilon\zeta}S_{CD}^{\sigma\tau}-(-1)^{(A+B)\cdot(C+D)}S_{CD}^{\sigma\tau}S_{AB}^{\epsilon\zeta} \tag{2.17}$$

proceeds by using the super identities for arbitrary $\mathbb{Z}_2\oplus\mathbb{Z}_2$ graded operators P,Q,R,S,

$$\begin{aligned} \langle PQ,RS\rangle&=(-1)^{(Q)\cdot(R)}\langle P,R\rangle QS\\ &\quad+(-1)^{(P+Q)\cdot(R)+(P)\cdot(S)}R\langle P,S\rangle Q\\ &\quad+P\langle Q,R\rangle S+(-1)^{(P+Q)\cdot(R)}RP\langle Q,S\rangle\,, \end{aligned} \tag{2.18}$$

where

$$\langle P,Q\rangle\equiv PQ-(-1)^{(P)\cdot(Q)}QP$$

and

$$(QR)\equiv(Q)+(R)\in\mathbb{Z}_2\oplus\mathbb{Z}_2\,,$$

leading to

$$\begin{aligned} \langle S_{AB}^{\epsilon\zeta},S_{CD}^{\sigma\tau}\rangle&=(-1)^{(A)\cdot(B)}G_{AC}^{\epsilon\sigma}S_{BD}^{\zeta\tau}\\ &\quad+(-1)^{(A)\cdot(B)+(C)\cdot(D)}G_{AD}^{\epsilon\tau}S_{BC}^{\zeta\sigma}\\ &\quad+G_{BC}^{\zeta\sigma}S_{AD}^{\epsilon\tau}+(-1)^{(C)\cdot(D)}G_{BD}^{\zeta\tau}S_{AC}^{\epsilon\sigma}\,. \end{aligned} \tag{2.19}$$

In particular, if $D=*$,

$$\begin{aligned} \langle S_{AB}^{\epsilon\zeta},S_{C*}^{\sigma}\rangle&=\langle S_{AB}^{\epsilon\zeta},(1/\sqrt{2})C_C^{\sigma}\rangle\\ &=(-1)^{(A)\cdot(B)}G_{AC}^{\epsilon\sigma}(1/\sqrt{2})C_B^{\zeta}\\ &\quad+G_{BC}^{\zeta\sigma}(1/\sqrt{2})C_A^{\epsilon}\,, \end{aligned} \tag{2.20}$$

and if $B=D=*$,

$$\langle S_{A*}^{\epsilon},S_{C*}^{\sigma}\rangle=\langle(1/\sqrt{2})C_A^{\epsilon},(1/\sqrt{2})C_C^{\sigma}\rangle=S_{AC}^{\epsilon\sigma}\,. \tag{2.21}$$

Hence

$\{S_{AB}^{\epsilon\zeta}, A,B\in I', \epsilon,\zeta=0,1\}$ form a $\mathbb{Z}_2\oplus\mathbb{Z}_2$ color superalgebra under the supercommutation defined in (2.18), which is denoted here as

$$\mathrm{SpO}(2M_{(0,0)}/1_{(0,1)}/2N_{(1,0)}/0_{(1,1)})\,. \tag{2.22}$$

III. SUBALGEBRAS OF THE $\mathbb{Z}_2\oplus\mathbb{Z}_2$ COLOR SUPERALGEBRA

The $\mathbb{Z}_2\oplus\mathbb{Z}_2$ color superalgebra obtained in (2.22) has a rich subalgebra structure which is illustrated in Fig. 1. [In this section we assume that the summation convention for repeated upper and lower indices is adopted and, unless otherwise stated we assume $a,b\in B$, $\alpha,\beta\in F$, $A,B\in I$, $\epsilon,\zeta,\sigma,\tau=0$ or 1, where B and F are as defined in Eq. (2.1) and $I=B\cup F$.] Various subalgebras (Lie superalgebras or ordinary Lie algebras) are labeled by their conventional names and the generators for each of them is given. The various subalgebras are established by either discarding selected sets of the generators of the big algebra or by forming particular linear combinations of them and projecting the $\mathbb{Z}_2\oplus\mathbb{Z}_2$ grading vector down to $\mathbb{Z}_2$ and further to $\mathbb{Z}_1$.

There are four chains of subalgebras shown in Fig. 1.

Chain 1:

$$\begin{aligned} &\mathrm{SpO}(2M_{(0,0)}/1_{(0,1)}/2N_{(1,0)}/0_{(1,1)})\\ &\quad\supset\mathrm{SpO}(2M/2N)\supset\mathrm{Sp}(2)\oplus\mathrm{OSp}(M/N)\\ &\quad\supset\mathrm{U}(1)\oplus\mathrm{OSp}(M/N)\supset\mathrm{U}(1)\oplus\mathrm{O}(M)\oplus\mathrm{Sp}(N)\,. \end{aligned} \tag{3.1a}$$

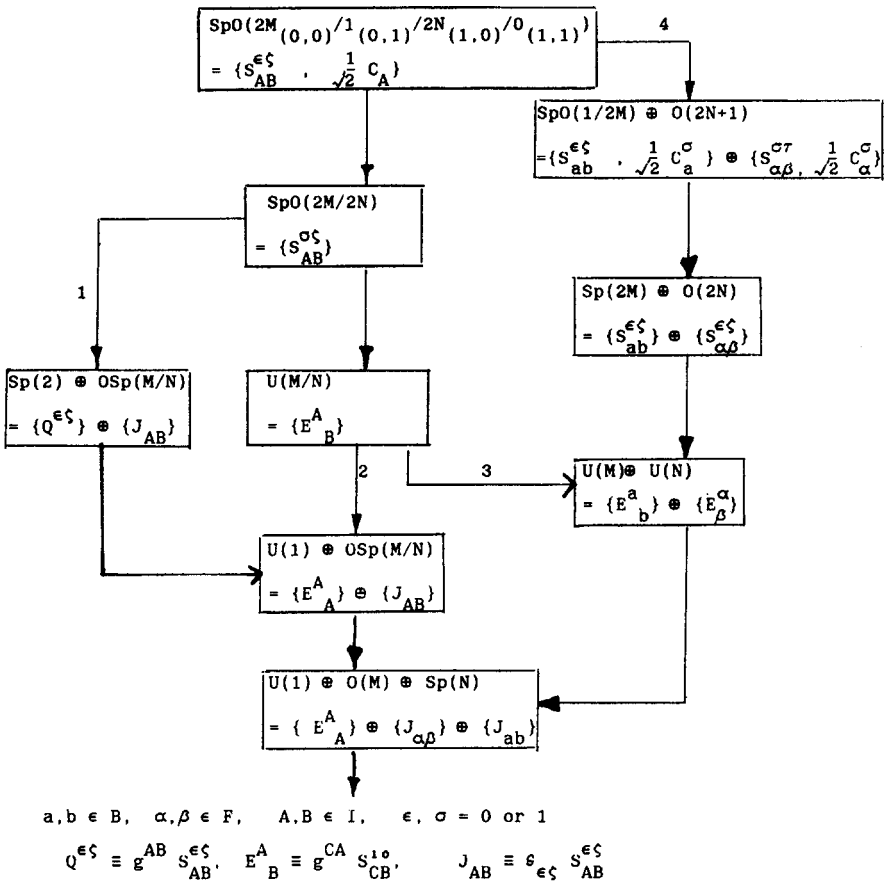


FIG. 1. Subalgebra structure of the $Z_2 \oplus Z_2$ color algebra.

Chain 2:

$$\mathrm{SpO}(2M_{(0,0)}/1_{(0,1)}/2N_{(1,0)}/0_{(1,1)}) \supset \mathrm{SpO}(2M/2N) \supset \mathrm{U}(M/N) \supset U(1) \oplus \mathrm{OSp}(M/N) \supset U(1) \oplus \mathrm{O}(M) \oplus \mathrm{Sp}(N) . \tag{3.1b}$$

Chain 3:

$$\mathrm{SpO}(2M_{(0,0)}/1_{(0,1)}/2N_{(1,0)}/0_{(1,1)}) \supset \mathrm{SpO}(2M/2N) \supset \mathrm{U}(M) \oplus \mathrm{U}(N) \supset \mathrm{U}(1) \oplus \mathrm{O}(M) \oplus \mathrm{Sp}(N) . \tag{3.1c}$$

Chain 4:

$$\mathrm{SpO}(2M_{(0,0)}/1_{(0,1)}/2N_{(1,0)}/0_{(1,1)}) \supset \mathrm{SpO}(1/2M) \oplus \mathrm{O}(2N+1) \supset \mathrm{Sp}(2M) \oplus \mathrm{O}(2N) \supset \mathrm{U}(M) \oplus \mathrm{U}(N) \supset \mathrm{U}(1) \oplus \mathrm{O}(M) \oplus \mathrm{Sp}(N) . \tag{3.1d}$$

The generalized quasispin algebra Sp(2) appears in the first chain and the generator $Q^{\epsilon\zeta}$ is defined as

$$Q^{\epsilon\zeta} \equiv \tfrac{1}{2} g^{AB} S^{\epsilon\zeta}_{AB} , \tag{3.2a}$$

and the commutation relation satisfied by $Q^{\epsilon\zeta}$ is

$$[Q^{\epsilon\zeta}, Q^{\sigma\tau}] = \tfrac{1}{2}(\theta^{\epsilon\sigma} Q^{\zeta\tau} + \theta^{\epsilon\tau} Q^{\zeta\sigma} + \theta^{\zeta\sigma} Q^{\epsilon\tau} + \theta^{\zeta\tau} Q^{\epsilon\sigma}) , \tag{3.2b}$$

where

$$\theta^{\epsilon\sigma} \equiv \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \text{ i.e., } \theta^{00} = \theta^{11} = 0, \quad \theta^{01} = -\theta^{10} = 1 .$$

Explicitly, we have

$$Q^{01} = Q^{10} = \tfrac{1}{4}(M - N) + \tfrac{1}{2}(n_b + n_f) , \tag{3.3a}$$

$$Q^{00} = \frac{1}{2} \sum_a (-)^{l_a \mp m_a} C^0_a C^0_{-a} - \frac{1}{2} \sum_\alpha (-)^{j_\alpha \mp m_\alpha} C^0_\alpha C^0_{-\alpha} , \tag{3.3b}$$

$$Q^{11} = \frac{1}{2} \sum_a (-)^{l_a \mp m_a} C^1_a C^1_{-a} - \frac{1}{2} \sum_\alpha (-)^{j_\alpha \mp m_\alpha} C^1_\alpha C^1_{-\alpha} , \tag{3.3c}$$

where n_b and n_f are the boson and fermion number operator, a and α refer to boson and fermion operators, and satisfy commutation relations

$$[Q^{01}, Q^{00}] = -Q^{00} , \tag{3.4a}$$

$$[Q^{01}, Q^{11}] = +Q^{11} , \tag{3.4b}$$

$$[Q^{00}, Q^{11}] = 2Q^{01} . \tag{3.4c}$$

If we now start from the big algebra and look along the other three chains at the generators for the subalgebras, we find $S^{\epsilon\zeta}_{AB}, E^{A}_{B}, J_{AB}$ appearing successively. The bilinear operator $S^{\epsilon\zeta}_{AB}$, $A,B\in I$, is really only $\mathbb{Z}_2$ graded since the second component of the grading vector is always zero. We define

$$(S^{\epsilon\zeta}_{AB}) = (A+B) = (A) + (B) \in \mathbb{Z}_2 , \tag{3.5a}$$

where (A) (B) is 0 or 1 as defined in (2.4). This redefinition does not affect the supercommutation

$$\begin{aligned}\langle S^{\epsilon\tau}_{AB}, S^{\sigma\zeta}_{CD}\rangle &= (-)^{(A)(B)} G^{\epsilon\sigma}_{AC} S^{\zeta\tau}_{BD} + (-)^{(A)(B)+(C)(D)} G^{\epsilon\tau}_{AD} S^{\zeta\sigma}_{BC} \\ &\quad + G^{\zeta\sigma}_{BC} S^{\epsilon\tau}_{AD} + (-)^{(C)(D)} G^{\zeta\tau}_{BD} S^{\epsilon\sigma}_{AC} ,\end{aligned} \tag{3.5b}$$

except $G^{\epsilon\tau}_{AB}$ is now $\mathbb{Z}_2$ graded as defined in (2.8) instead of being $\mathbb{Z}_2 \oplus \mathbb{Z}_2$ graded as defined in (2.12), $(A),(B)\in\mathbb{Z}_2$ $\{S^{\epsilon\zeta}_{AB}\}$ generates Lie superalgebra SpO$(2M/2N)$.

The bilinear $\mathbb{Z}_2$ graded generator E^{A}_{B} is defined in terms of the $\mathbb{Z}_2$ graded generator $S^{\epsilon\tau}_{AB}$ as follows:

$$E^{A}_{B} = g^{CA} S^{10}_{CB} , \tag{3.6a}$$

where g^{CA} is the inverse of $g_{AC} = G^{01}_{AC}$,

$$(E^{A}{}_{B}) = (A+B) = (A) + (B) \in \mathbb{Z}_2 , \tag{3.6b}$$

and under supercommutation (1.7) satisfies

$$\langle E^{A}{}_{B}, E^{C}{}_{D}\rangle = \delta_{B}{}^{C} E^{A}{}_{D} - (-)^{(A+B)(C+D)} \delta_{D}{}^{A} E^{C}{}_{B} , \tag{3.6c}$$

where $\delta_{B}{}^{C}$ is the Kroneker delta. Here $\{E^{A}{}_{B}\}$ generates the Lie superalgebra U(M/N), which is a subalgebra of SpO$(2M/2N)$, while the even operators $\{E^{a}{}_{b}, E^{\alpha}{}_{\beta}\}$ generate the direct sum of the ordinary Lie algebras U(M) $\oplus$ U(N).

Define the bilinear $\mathbb{Z}_2$ graded operators J_{AB} as

$$J_{AB} = \theta_{\epsilon\zeta} S^{\epsilon\zeta}_{AB} = S^{10}_{AB} - S^{01}_{AB} , \tag{3.7a}$$

where

$$(\theta_{\epsilon\zeta}) = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix},$$

with grading

$$(J_{AB}) = (A+B) = (A) + (B) \in \mathbb{Z}_2 . \tag{3.7b}$$

Then under supercommutation, (2.7) satisfies

$$\begin{aligned}\langle J_{AB}, J_{CD}\rangle &= -(-)^{(A)(B)} g_{AC} J_{BD} + (-)^{(A)(B)+(C)(D)} g_{AD} J_{BC} \\ &\quad + g_{BC} J_{AD} - (-)^{(C)(D)} g_{BD} J_{AC} .\end{aligned} \tag{3.7c}$$

Here $\{J_{AB}\}$ generates Lie superalgebra OSp(M/N) while its even part ($\mathbb{Z}_1$ graded) generates the direct sum of ordinary Lie algebra O(M) $\oplus$ Sp(N). Since both $E^{A}{}_{B}$ and J_{AB} are defined in terms of $S^{\epsilon\zeta}_{AB}$, the supercommutation relation for both of them can be easily evaluated by using Eq. (3.5b). The generators in different algebras of a direct sum algebra should be both disjoint and commute with each other. We see that the commutation is often trivially satisfied by noticing one algebra may contain entirely boson and the other entirely fermion operators, and it is an easy task to show also $[Q^{\epsilon\zeta}, J_{AB}] = 0$. We have thus obtained the subalgebra structure of Fig. 1. Of particular interest to us is the chain that contains the quasispin algebra Sp(2). Further discussion on quasispin is left to the next section.

IV. PROPERTIES OF THE GENERALIZED QUASISPIN ALGEBRA

The quasispin operators defined in equations (3.3) and (3.4) contain both bosonic (B) and fermionic (F) operators with

$$Q_0 = Q^{01} = B_0 + F_0 , \tag{4.1a}$$

$$Q_+ = Q^{11} = B_+ \mp F_+ , \tag{4.1b}$$

$$Q_- = Q^{00} = B_- \pm F_- . \tag{4.1c}$$

The bosonic quasispin operators

$$B_0 = \frac{1}{2} \sum_a C^{1}_{a} C^{0}_{a} + \frac{M}{4} , \tag{4.2a}$$

$$B_+ = \frac{1}{2} \sum_a (-1)^{l_a - m_a} C^{1}_{a} C^{1}_{-a} , \tag{4.2b}$$

$$B_- = \frac{1}{2} \sum_a (-1)^{l_a - m_a} C^{0}_{a} C^{0}_{-a} \tag{4.2c}$$

satisfy the commutation relations

$$[B_0, B_\pm] = \pm B_\pm , \quad [B_+, B_-] = -2B_0 \tag{4.3}$$

of the boson quasispin SU(1,1) algebra[20] while the fermionic quasispin operators

$$F_0 = \frac{1}{2} \sum_\alpha C^{1}_{\alpha} C^{0}_{\alpha} - \frac{N}{4} , \tag{4.4a}$$

$$F_+ = \frac{1}{2} \sum_\alpha (-1)^{j_\alpha - m_\alpha} C^{1}_{\alpha} C^{1}_{-\alpha} , \tag{4.4b}$$

$$F_- = -\frac{1}{2} \sum_\alpha (-1)^{j_\alpha - m_\alpha} C^{0}_{\alpha} C^{0}_{-\alpha} \tag{4.4c}$$

satisfy the commutation relations

$$[F_0, F_\pm] = \pm F_\pm , \quad [F_+, F_-] = 2F_0 \tag{4.5}$$

of the fermion quasispin SU(2) algebra.[21]

Thus the operators $(Q_0, Q_\pm)$ form a generalization of the usual fermionic and bosonic quasispin algebras with the commutation relations given in Eq. (3.4) being rewritten as

$$[Q_0, Q_\pm] = \pm Q_\pm , \quad [Q_+, Q_-] = -2Q_0 . \tag{4.6}$$

The generalized quasispin algebra (GQA) is noncompact with the result that all nontrivial irrep are infinite dimensional. The irrep of the GQA may be labeled in terms of the eigenvalues of the second-order Casimir invariant and those of the quasispin operator Q_0.

The second-order Casimir invariant for the GQA is defined as

$$Q^2 = Q_0(Q_0 - 1) - Q_+ Q_- , \tag{4.7}$$

where as usual

$$[Q^2, Q_+] = [Q^2, Q_0] = 0 . \tag{4.8}$$

Consider an arbitrary n particle state say $|n\rangle$. The action of Q_0 on $|n\rangle$ is to count the number of particles in the state (bosons and fermions). Thus

$$Q_0 |n\rangle = ((M-N)/4 + n/2)|n\rangle . \tag{4.9}$$

The operators Q_+ and Q_- constitute pairs of creation or

annihilation operators coupled to zero total angular momentum and thus create or destroy pairs of bosons or fermions of zero total angular momentum changing the number of particles in the state by ± 2 and hence change the eigenvalue of Q_0 by unity.

The fermionic part of the action of Q_+ is restricted by the Pauli exclusion principle but not for the bosonic part. Starting with an initial state Q_+ can be applied repeatedly without limit to connect an infinite set of states whereas the action of Q_- terminates when a state of v unpaired particles is reached, there being no zero coupled pairs left for Q_- to destroy, i.e.,

$$Q_-|v\rangle = 0 . \tag{4.10}$$

In this case we have

$$Q_0|v\rangle = ((M-N)/4 + v/2)|v\rangle . \tag{4.11}$$

Each GQA multiplet has a unique lowest state $|v\rangle$.

Consider the action of Q^2 on $|v\rangle$. Recalling (4.10) we have

$$\begin{aligned} Q^2|v\rangle &= [Q_0(Q_0-1) - Q_+Q_-]|v\rangle \\ &= \left(\frac{M-N}{4} + \frac{v}{2}\right)\left(\frac{M-N}{4} + \frac{v}{2} - 1\right)|v\rangle . \end{aligned} \tag{4.12}$$

Hence the eigenvalue of the Casimir invariant Q^2 may be taken as $Q(Q-1)$, where

$$Q = (M-N)/4 + v/2 , \tag{4.13}$$

where Q may be integer, half-integer, or quarter-integer. The eigenvalues of the operator Q^2 are symmetric under[22]

$$Q \to Q' = -Q + 1 , \tag{4.14}$$

with $Q(Q-1) \to Q'(Q'-1) \geqslant 0$. All states connected to $|v\rangle$ by the action Q_+ will have the same Casimir invariant and the irrep of the GQA uniquely labeled by Q with the basis vectors being labeled as

$$|QQ_0\rangle = \left|\frac{M-N}{4} + \frac{v}{2}, \frac{M-N}{4} + \frac{n}{2}\right\rangle , \tag{4.15}$$

where v is the total number of unpaired particles and

$$n = v + 2i, \quad i = 0,1,2,\ldots . \tag{4.16}$$

Within each irrep, i.e., for a fixed value of Q, Q_0 takes the values

$$Q_0 = Q, Q+1, Q+2, \ldots , \tag{4.17}$$

corresponding to the basis states

$$|QQ\rangle, |Q,Q+1\rangle, |Q,Q+2\rangle, \ldots . \tag{4.18}$$

The GQA is of importance in physical applications in providing a natural extension of the notions of pairing and seniority to mixed boson–fermion systems. Indeed in the Casimir invariant (4.7) the term

$$\begin{aligned} V_{BF} &= F_+B_- + B_+F_- \\ &= -\frac{1}{4}\sum_{a,\alpha}(-1)^{l_a \mp m_a}(-1)^{j_\alpha \mp m_\alpha} \\ &\quad \times (C^1_\alpha C^1_{-\alpha} C^0_a C^0_{-a} + C^1_a C^1_{-a} C^0_\alpha C^0_{-\alpha}) \end{aligned} \tag{4.19}$$

gives via (3.12) a boson–fermion pairing interaction whose eigenvalues are a function of M, N, and v (Ref. 21). Precisely the same V_{BF} was noted in the dynamical supersymmetry scheme based on the $U(M/N) \supset \mathrm{OSp}(M/N)$ second-order chain.[7] A direct comparison of Q^2 with the second-order $\mathrm{OSp}(M/N)$ Casimir invariant confirms that

$$Q^2 = \tfrac{1}{8} J_{AB} J^{BA} + \tfrac{1}{4}(M-N)(\tfrac{1}{4}(M-N) - 1) , \tag{4.20}$$

so that the GQA gives an alternative and more direct insight into the physics inherent in such models. Beyond this special (exactly soluble) case, the use of GQA in general permits the explicit n dependence of matrix elements of interactions to be expressed in terms of coupling coefficients via the celebrated Wigner–Eckart theorem.[21]

Finally, it should be pointed out that the sign choice in (2.2) leads to the two alternative GQA's (4.1) which are interchanged by Hermitian conjugation. In particular the above V_{BF} is anti-Hermitian,[7] even though Q has real eigenvalues. This situation is perhaps not unexpected if the algebraic models are regarded as reflecting a truncation of the true space of states.

V. CONCLUSION

Our primary purpose in this paper has been to demonstrate the existence of a genealized quasispin algebra that can arise in supersymmetric systems. In the process a noncompact $\mathbb{Z}_2 \oplus \mathbb{Z}_2$ graded color superalgebra $\mathrm{SpO}(2M/1/2N/0)$ has been established. This should allow the systematic analysis of group substructures relevant to the applications of supersymmetry concepts to nuclei and other systems.[23]

[1]See, for instance, articles in Nucl. Phys. A **421** (1984).
[2]See, for instance, articles in Physica (Utrecht) D **15** (1985).
[3]A. B. Balantekin and I. Bars, J. Math. Phys. **22**, 1810 (1981).
[4]A. B. Balantekin, and I. Bars, J. Math. Phys. **22**, 1149 (1981).
[5]P. H. Dondi and P. D. Jarvis, Z. Phys. C **4**, 201 (1980).
[6]P. H. Dondi and P. D. Jarvis, J. Phys. A: Math. Gen. **14**, 547 (1981).
[7]I. Morrison and P. D. Jarvis, Nucl. Phys. A **435**, 461 (1985).
[8]Q-Z. Han, H-Z. Sun, M. Shang, and D-H. Feng, J. Math. Phys. **26**, 1822 (1985).
[9]M. Baake, and P. Reinicke, J. Math. Phys. **27**, 1430 (1986).
[10]D. J. Rowe, Rep. Prog. Phys. **48**, 1419 (1985).
[11]M. Moshinsky and C. Quesne, J. Math. Phys. **11**, 1631 (1970); **12**, 1772 (1971).
[12]Q-Z. Han, F-s. Liu, and H-Z. Sun, Commun. Theor. Phys. (Beijing) **1**, 529 (1984).
[13]M. Baake, P. Reinicke, and A. Gelberg, Phys. Lett. B **166**, 10 (1986).
[14]I. Talmi, Comments Nucl. Part. Phys. **11**, 241 (1983).
[15]V. Rittenberg and D. Wyler, Nucl. Phys. B **139**, 189 (1978).
[16]V. Rittenberg and D. Wyler, J. Math. Phys. **19**, 2193 (1978).
[17]J. Lukierski and V. Rittenberg, Phys. Rev. D **18**, 385 (1967).
[18]M. Scheunert, J. Math. Phys. **20**, 712 (1979).
[19]H. S. Green, and P. D. Jarvis, J. Math. Phys. **24**, 1681 (1983).
[20]H. Ui, Ann. Phys. (NY) **49**, 69 (1968).
[21]B. R. Judd, *Group Theory and its Applications* (Academic, New York, 1968), Vol. 1, p. 183.
[22]A. P. Jucys, Int. J. Quantum Chem. IV 1001 (1969).
[23]P. D. Jarvis and G. E. Stedman, J. Phys. A: Math. Gen. **17**, 775 (1984).

A rigorous modified Thomas–Fermi theory for atomic systems

Jerome A. Goldstein and Gisèle Ruiz Rieder[a)]
Department of Mathematics and Quantum Theory Group, Tulane University, New Orleans, Louisiana 70118

(Received 26 August 1986; accepted for publication 17 December 1986)

Recently Parr and Ghosh [Proc. Natl. Acad. Sci. USA **83**, 3577 (1986)] proposed a variant of the classical Thomas–Fermi theory of electrons in an atom. They produced a continuous electron density by introducing the constraint that the integral $\int_{\mathbb{R}^3} e^{-2k|x|}\,\Delta\rho(x)dx$ exists, where k is determined by the nuclear cusp condition. Their results give improved calculations of ground state electron densities and energies. The present paper provides a rigorous mathematical foundation for the work of Parr and Ghosh and converts their results into theorems. Some generalizations are also obtained.

I. INTRODUCTION

Our goal is to establish rigorously results suggested by the Parr–Ghosh extension[1] of Thomas–Fermi theory.[2,3] In this theory the (approximate) electron density is finite at a nucleus and satisfies the cusp condition. Our treatment will be parallel to that of the rigorous conventional Thomas–Fermi theory, so we begin by reviewing that theory in some detail.

We shall follow the Euler–Lagrange equation approach of Bénilan and Brezis,[4–6] whose work was inspired by the pioneering work of Lieb and Simon,[7–9] who used the direct methods of the calculus of variations. The work of Bénilan and Brezis was done in the late 1970's and is outlined in two articles by Brezis.[5,6] Their full joint paper[4] still has not been completed.

In conventional Thomas–Fermi theory we seek the ground state electron density ρ (for a system of N electrons in $\mathbb{R}^3$) which minimizes the energy functional

$$E(\rho) = T(\rho) + V_{ne}(\rho) + V_{ee}(\rho) \tag{1}$$

on

$$D_N = \left\{\rho\in L^1(\mathbb{R}^3)\ \middle|\ \rho\geqslant 0,\ \int_{\mathbb{R}^3}\rho(y)dy = N,\ \rho\in\mathrm{Dom}(T)\cap\mathrm{Dom}(V_{ne})\cap\mathrm{Dom}(V_{ee})\right\}.$$

Here $T(\rho)$ represents the kinetic energy, $V_{ne}(\rho)$ represents the electron-nuclear attraction, and $V_{ee}(\rho)$ is the electron–electron repulsion term. One can argue on physical grounds that the kinetic energy term, which is horribly complicated as a functional of the density, can be approximated by an expression of the form

$$T(\rho) = c_p\int_{\mathbb{R}^3}\rho(x)^p\,dx$$

for some $p>1$, and the classical approximation of Thomas and Fermi is $p=\frac{5}{3}$ and $c_p = \frac{3}{10}(3\pi^2)^{2/3}$. We will work with the more general kinetic energy term

$$T(\rho) = \int_{\mathbb{R}^3} J(\rho(y))dy; \tag{2}$$

here J is a convex function on $[0,\infty)$ satisfying $J(0) = J'(0) = 0$, $J''(r)\geqslant 0$, and $J(r)>0$ for $r>0$. In the most general case

$$V_{ne}(\rho) = \int_{\mathbb{R}^3} V(y)\rho(y)dy, \tag{3}$$

where V is a measurable real-valued operator on $\mathbb{R}^3$. One can show[4–6,10] that a necessary condition for the existence of a solution to the minimization problem is

$$V\in L^1_{\mathrm{loc}}(\mathbb{R}^3) \text{ and } V<0 \text{ on a set of positive measure.} \tag{V}$$

In the case of an atom with Z protons fixed at the origin the electron-nuclear attraction is the Coulomb potential

$$V(x) = -Z/|x|. \tag{4}$$

We shall restrict ourselves to the atomic case in this paper. For the electron–electron repulsion term we take

$$V_{ee}(\rho) = \frac{c_{ee}}{2}\int_{\mathbb{R}^3}\int_{\mathbb{R}^3}\frac{\rho(x)\rho(y)}{|x-y|}\,dx\,dy. \tag{5}$$

Usually one takes $c_{ee}=1$, but the Fermi–Amaldi approximation[11] of $c_{ee} = (N-1)/N$ also gives interesting results. This approximation essentially agrees with $c_{ee}=1$ for N large and vanishes for $N=1$, when no electron–electron repulsion is present. If we define the operator B by $Bf = (1/4\pi)\,(1/|\cdot|)*f$, then $B = (-\Delta)^{-1}$, and V_{ee} can be expressed in the convenient form

$$V_{ee}(\rho) = 2\pi c_{ee}\int_{\mathbb{R}^3}\rho(x)\,(B\rho)(x)dx. \tag{6}$$

The minimization problem was first solved rigorously by Lieb and Simon[7,8] with $J(r) = c_{5/3}r^{5/3}$ and V in a class of potentials including (4). Benilan and Brezis[4] then solved a more general problem [with J as in (2) and a more general potential V] by deriving the Euler–Lagrange equations corresponding to the energy minimization problem, converting those equations into a form involving nonlinear elliptic partial differential equations, and solving the resulting problem.

The following problem is the Euler–Lagrange problem associated with the minimization problem. Here T is given by (2), V_{ne} by (3), and V_{ee} by (6).

Find $(\rho_0,\lambda)\in D'_N\times\mathbb{R}$ satisfying

$$\begin{aligned} &J'(\rho_0) + V + kB\rho_0 + \lambda = 0 \quad \text{a.e. on } [\rho_0>0],\\ &J'(\rho_0) + V + kB\rho_0 + \lambda\geqslant 0 \quad \text{a.e. on } [\rho_0 = 0], \end{aligned} \tag{7}$$

[a)] Current address: Department of Mathematics, Louisiana State University, Baton Rouge, Louisiana 70803.

 0022-2488/87/051198-05$02.50

where

$$D'_N = \left\{\rho \in L^1(\mathbb{R}^3) \mid \rho \geqslant 0 \text{ and } \int_{\mathbb{R}^3} \rho(y)dy = N\right\},$$

and $k = 4\pi c_{ee}$.

Thomas–Fermi theory, which is an approximate theory, possesses interesting features, some of which may be termed "flaws." The following theorem shows that the Euler–Lagrange problem is not equivalent to the minimization problem in the generality of the present context.

Theorem 1:[4–6,10] If ρ_0 is a solution of the minimization problem (1) on D_N, then there exists $\lambda \in \mathbb{R}$ such that (ρ_0, λ) solves the Euler–Lagrange problem (7) on $D'_N \times \mathbb{R}$. Conversely, suppose (ρ_0, λ) solves (7) on $D'_N \times \mathbb{R}$ and suppose there is a $K \in \mathbb{R}$ such that

$$J^*([K - V(x)]_+) \in L^1(\mathbb{R}^3); \tag{8}$$

here J^* is the convex conjugate of J, and a_+ is the positive part of the real number a. Then ρ_0 is a solution of (1) on D_N.

One can show that if (8) does not hold, no solution of the minimization problem can exist.[5,6,10] If $J(r) = cr^p$ and V is the atomic Coulomb potential (4), then (8) is equivalent to $p > \frac{3}{2}$. It is also known that the Euler–Lagrange problem has a solution for $p > \frac{4}{3}$. Thus, if $\frac{4}{3} < p \leqslant \frac{3}{2}$, there is a density ρ_0 which satisfies (7), but for such p we have $\inf E(\rho) = -\infty$.

Another feature, and indeed a fault, of Thomas–Fermi theory lies in the fact that the solution density ρ_0 diverges at an atomic nucleus. To see this with $J(r) = cr^p$ and V the atomic Coulomb potential given by (4), we observe that $V(r) \sim \text{const } r^{-1}$ as $r \to 0$ and that $B\rho_0$ is bounded as $r \to 0$. But if ρ_0 satisfies (7), then

$$\rho_0(r)^{p-1} \sim \text{const } r^{-1} \quad \text{as } r \to 0,$$

that is,

$$\rho_0(r) \sim \text{const } r^{-1/(p-1)} \quad \text{as } r \to 0. \tag{9}$$

In particular, $\rho_0(r) \to \infty$ as $r \to 0$. This is undesirable from a physical point of view; the quantum mechanical ground state density of an atom should be continuous and have a finite maximum at the origin (that is, at the nucleus). The correct behavior for the quantum mechanical electron density is known to be, to first order in r,

$$\rho(r) \sim \text{const } e^{-2Zr} \quad \text{as } r \to 0. \tag{10}$$

Using $c_{ee} = 1$ in (5), Bénilan and Brezis[4–6] prove that the density ρ_0 that solves (7) has compact support for a positive ion, i.e., for $N < Z$, and the support is $\mathbb{R}^3$ for $N = Z$. Moreover, if we use the Fermi–Amaldi approximation in (5), we know that the density which solves (7) has compact support for the case of a neutral atom as well as for a positive ion, in fact, for $N < Z + 1$ (see Ref. 10). (The support is $\mathbb{R}^3$ for $N = Z + 1$.) Thus we expect to have $\nabla\rho_0(x) \to 0$ as $|x| \to \infty$. This leads naturally to the condition that $\int_{\mathbb{R}^3} \Delta\rho_0$ should vanish. However, this condition on $\Delta\rho_0$ does not follow from conventional Thomas–Fermi theory. The above mentioned features and flaws may be remedied by imposing a continuity constraint on the domain of the energy functional (1), as was noted by Parr and Ghosh[1] in the special case $J(r) = c_{5/3}r^{5/3}$. In this paper we shall study the effect of implementing the continuity constraint in a more general and more rigorous mathematical framework. The density ρ_0 which solves this amended Thomas–Fermi problem dispenses with the previously mentioned curiosities and has several interesting properties of its own.

II. THE EULER–LAGRANGE PROBLEM

Let V be given by (4) and define

$$E(\rho) = \int J(\rho) + \int V\rho + 2\pi c_{ee} \int \rho B\rho \tag{11}$$

on the domain

$$D_{N,M} = \left\{\rho \in L^1(\mathbb{R}^3) \cap L^\infty(\mathbb{R}^3) \,\middle|\, \rho \geqslant 0,\right.$$
$$\int_{\mathbb{R}^3} \rho(y)dy = N,\ \Delta\rho \in L^1_{\text{loc}}(\mathbb{R}^3),$$
$$\text{“}\nabla\rho(\infty) = 0\text{”},\ \int e^{-2k|x|}\Delta\rho(x)dx = M \in \mathbb{R},$$
$$\left.\rho \in \text{Dom}(T) \cap \text{Dom}(V_{ne}) \cap \text{Dom}(V_{ee})\right\}.$$

We define “$g(\infty) = 0$” to mean that for all $\epsilon > 0$ there is a Borel set $A_\epsilon \subset \mathbb{R}^3$ such that $\text{meas}(A_\epsilon) = \int_{A_\epsilon} dx < \infty$ and $|g(x)| < \epsilon$ for all $x \notin A_\epsilon$. It is easy to see that $E(\rho)$ on $D_{N,M}$ is a convex functional. If ρ_0 were the minimum of E, then we would expect $E'(\rho_0) = 0$. However, this is not precise; we have three constraints to consider.

Let $\lambda_1, \lambda_2, M \in \mathbb{R}$, and define

$$E_M(\rho) = \int J(\rho) + \int V\rho + 2\pi c_{ee} \int \rho B\rho + \lambda_1\left(\int \rho - N\right)$$
$$+ \lambda_2\left(\int e^{-2k|x|}\Delta\rho(x)dx - M\right)$$

on $\mathscr{D}_M = \cup\{D_{N,M} : N > 0\}$ and with V given by (4). Applying the divergence theorem and using “$\nabla\rho(\infty) = 0$” we see that

$$\int e^{-2k|x|}\Delta\rho(x)dx = \int \Delta e^{-2k|x|}\rho(x)dx,$$

and so

$$E_M(\rho) = \int J(\rho) + \int V\rho + 2\pi c_{ee} \int \rho B\rho + \lambda_1\left(\int \rho - N\right)$$
$$+ \lambda_2\left[\int \left(4k^2 - \frac{4k}{|x|}\right)e^{-2k|x|}\rho(x)dx - M\right]. \tag{12}$$

If ρ_0 minimizes E_M on $D_{N,M}$, then, formally, $E_M(\rho) = E_M(\rho_0) + \langle E'_M(\rho_0),\ \rho - \rho_0\rangle + o(\rho - \rho_0)$. By the convexity of the functional E_M on $D_{N,M}$, we expect ρ_0 to be the unique solution of

$$0 = E'_M(\rho_0) = J'(\rho_0) + V + 4\pi c_{ee}B\rho_0$$
$$+ \lambda_1 + \lambda_2[(4k^2 - 4k/|x|)e^{-2k|x|}]. \tag{13}$$

Since E'_M is independent of M, we call it E' for short. We choose $\lambda_2 = Z/4k$ so that (13) reduces to

$$0 = J'(\rho_0) + (-Z/|x|) + 4\pi c_{ee}B\rho_0 + \lambda_1$$
$$+ (Z/|x|)e^{-2k|x|} + Zke^{-2k|x|}.$$

This gets rid of the singularity of the potential at $x=0$. However, we must also consider the constraint $\rho \geqslant 0$. Accounting for this leads us to the following Euler–Lagrange problem.

Problem: Find $(\rho_0,\lambda)\in \cup\{D'_{N,M}\times\mathbb{R}: M>0\}$ such that

$$J'(\rho_0)+\tilde{V}+4\pi c_{ee}B\rho_0+\lambda=0 \quad \text{a.e. on } [\rho_0>0],$$
$$J'(\rho_0)+\tilde{V}+4\pi c_{ee}B\rho_0+\lambda\geqslant 0 \quad \text{a.e. on } [\rho_0=0], \tag{14}$$

where

$$\tilde{V}(x)=(-Z/|x|)(1-e^{-2k|x|})+kZe^{-2k|x|} \tag{15}$$

and

$$D'_{N,M}=\left\{\rho\in L^1(\mathbb{R}^3)\cap L^\infty(\mathbb{R}^3)\,\middle|\,\rho\geqslant 0,\ \int\rho=N,\ \text{“}\nabla\rho(\infty)=0\text{”},\ \int_{\mathbb{R}^3}e^{-2k|x|}\Delta\rho_0(x)dx=M\in\mathbb{R}\right\}.$$

Note that $\tilde{V}$ also satisfies condition (V); and letting $|x|\to\infty$ in (14) shows that $\lambda\geqslant 0$.

Theorem 2: Suppose ρ_0 minimizes (11) on $D_{N,M}$. Then there exists $\lambda\in\mathbb{R}$ such that (ρ_0,λ) satisfies (14) on $\cup\{D'_{N,M}: M\in\mathbb{R}\}$. Conversely, if $(\rho_0,\lambda)\in D'_{N,M}\times\mathbb{R}$ satisfies the Euler–Lagrange problem (14) and if there is a real number K with

$$J^*[(K-\tilde{V}(x))_+]\in L^1(\mathbb{R}^3), \tag{16}$$

then ρ_0 minimizes (12) on $D_{N,M}$.

The proof of the theorem is essentially the same as the proof of Theorem 1. We omit the details.[4–6,10]

Corollary: Let $J(r)=cr^p$ for $1<p<\infty$. Then the minimization of (12) on $D_{N,M}$ is equivalent to solving the Euler–Lagrange problem (14) on $\cup\{D'_{N,M}: M\in\mathbb{R}\}$.

Proof: It suffices to show (16) holds. Now $J(r)=cr^p$ implies $J^*(r)=\tilde{c}r^q$. Choose $K<0$. Then $(K-\tilde{V}(x))_+=0$ for $|x|>R$ for some sufficiently large R, since $\tilde{V}(x)\to 0$ as $|x|\to\infty$. If we observe that $\tilde{V}\in L^\infty(\mathbb{R}^3)$, we find that

$$\int_{\mathbb{R}^3}J^*[(K-\tilde{V}(x))_+]dx=\tilde{c}\int_{B_R(0)}[(K-\tilde{V}(x))_+]^q dx$$
$$\leqslant c_1\int_0^R[\tilde{V}(r)^q+|K|^q]r^2\,dr$$
$$\leqslant c_2[\|\tilde{V}\|_\infty^q+|K|^q]R^3<\infty. \quad \square$$

The value of M is really irrelevant in Theorem 2 and its corollary. Let ρ_0 minimize $E(\rho)$ given by (11) and belong to D_{N,M_0} for some $M_0\in\mathbb{R}$. Then for some $\lambda\in\mathbb{R}$ (ρ_0,λ) satisfies the Euler–Lagrange problem. Conversely if (ρ_0,λ) satisfies the Euler–Lagrange problem, then necessarily M defined by $M=\int_{\mathbb{R}^3}e^{-2k|x|}\Delta\rho_0(x)dx$ exists as a real number, and if (16) holds, then $\rho_0\in D_{N,M}$ for this M. If ρ_j minimizes (12) on D_{N,M_j} for $j=0,1$, then (ρ_j,λ_j) satisfies (14) on $\cup\{D'_{N,M}: M\in\mathbb{R}\}$ for some $\lambda_j\in\mathbb{R}$, $j=0,1$. But in the next section we shall show that solutions ρ_j of the Euler–Lagrange problem are unique. Thus $\rho_0=\rho_1$ and $M_0=M_1$.

The functional $E(\rho)$ given by (11) is a strictly convex functional on the convex set $\cup\{D_{N,M}: M\in\mathbb{R}\}$, so it will have at most one minimum. If it does, this minimum will belong to $D_{N,M}$ for precisely one value of M. Thus M is really determined by the problem through the constraint that $\int_{\mathbb{R}^3}e^{-2k|x|}\Delta\rho_0(x)dx$ is finite. In this sense the value of M is irrelevant.

III. THE PARTIAL DIFFERENTIAL EQUATION

Let $u=-\tilde{V}-4\pi c_{ee}B\rho_0$. Rearranging terms in (14), we see $J'(\rho_0)=u-\lambda$ a.e. on $[\rho_0>0]$. Define

$$\Gamma(r)=\begin{cases}(J')^{-1}(r), & \text{for } r\in(0,\infty),\\ 0, & \text{for } r\in(-\infty,0].\end{cases} \tag{17}$$

Applying $(J')^{-1}$ to (14) yields

$$\rho_0=\Gamma(u-\lambda),$$

and applying $-\Delta$ to the definition of u gives

$$-\Delta u+4\pi c_{ee}\Gamma(u-\lambda)=\Delta\tilde{V}.$$

Thus we arrive at the following nonlinear elliptic problem associated with the Euler–Lagrange problem (4) on $D_{N,M}$.

Problem: Find $u\in\mathscr{M}^3(\mathbb{R}^3)$ and $\lambda\in\mathbb{R}$ such that

$$-\Delta u+4\pi c_{ee}\Gamma(u-\lambda)=\Delta\tilde{V},$$
$$N=\int_{\mathbb{R}^3}\Gamma(u(x)-\lambda)dx, \quad \Gamma'_+(-\lambda)=c<\infty, \tag{18}$$

where

$$\Gamma(u(\cdot)-\lambda)\in L^1(\mathbb{R}^3)\cap L^\infty(\mathbb{R}^3).$$

Here $\Gamma'_+(-\lambda)$ is the derivative of Γ from the right at $-\lambda$.

The Marcinkiewicz spaces (or weak L^p spaces) $\mathscr{M}^p(\mathbb{R}^3)$ are defined as follows:

$$\mathscr{M}^p(\mathbb{R}^3)=\{u\,|\,u\in L^1_{\text{loc}}(\mathbb{R}^3) \text{ and } \|u\|_{\mathscr{M}^p}<\infty\},$$

where

$$\|u\|_{\mathscr{M}^p}=\min\left\{c\in[0,\infty]\,\middle|\,\int_A|u(x)|\leqslant c\,\text{meas}(A)^{1/q}\right.$$
$$\left.\text{for all measurable sets } A\subset\mathbb{R}^3 \text{ of finite measure}\right\},$$
$$p^{-1}+q^{-1}=1.$$

There are several basic properties of these spaces which we record as Proposition 1 below. For a more complete treatment and for the proofs, one may consult Ref. 12.

Proposition 1: (i) The function $x\to|x|^{-\alpha}$ belongs to $\mathscr{M}^{3/\alpha}(\mathbb{R}^3)$ for $0<\alpha<3$.

(ii) If $E\in\mathscr{M}^p(\mathbb{R}^3)$ for some p, $1<p<\infty$, and if $f\in L^1(\mathbb{R}^3)$, then

$$E*f\left[\text{defined by } (E*f)(x)=\int_{\mathbb{R}^3}E(x-y)f(y)dy\right]$$

belongs to $\mathscr{M}^p(\mathbb{R}^3)$ and

$$\|E*f\|_{\mathscr{M}^p}\leqslant\|E\|_{\mathscr{M}^p}\|f\|_{L^1}.$$

(iii) Suppose

$$u\in L^1_{\text{loc}}(\mathbb{R}^3),\ \Delta u\in L^1(\mathbb{R}^3),$$

and

$$\lim_{n\to\infty}\int_{1\leqslant|x|\leqslant 2}|u(nx)|dx=0. \tag{19}$$

Then $u = B(-\Delta u)$. In particular, $u \in \mathscr{M}^3(\mathbb{R}^3)$, $\nabla u \in [\mathscr{M}^{3/2}(\mathbb{R}^3)]^3$, $\|u\|_{\mathscr{M}^3} \leqslant C_0 \|\Delta u\|_{L^1}$, and $\|\nabla u\|_{\mathscr{M}^{3/2}} \leqslant C_0 \|\Delta u\|_{L^1}$ for some constant C_0 independent of u.

(iv) Let $u \in \mathscr{M}^3(\mathbb{R}^3)$ and $\Delta u \in L^1(\mathbb{R}^3)$. Then for every ξ in $\mathscr{J}_0 = \{\xi \in C^1(\mathbb{R}) \cap L^\infty(\mathbb{R}) \mid \xi'(x) \geqslant 0$ for all x and $\xi(0) = 0\}$, we have

$$(\xi'(u))^{1/2} |\nabla u| \in L^2(\mathbb{R}^3)$$

and

$$\int \xi'(u) |\nabla u|^2 + \int (\Delta u) \xi(u) \leqslant 0.$$

Note that if $u \in L^1(\mathbb{R}^3)$ or if $u \in \mathscr{M}^3(\mathbb{R}^3)$, then u satisfies (19).

Proposition 2: Let $\lambda \in [0, \infty)$. Suppose u is a solution of (18). Then there exists a solution (ρ_0, λ) of the Euler–Lagrange problem.

Proof: Define $\rho_0 = \Gamma(u - \lambda)$ a.e. [where Γ is defined by (17)]. Applying $B = (-\Delta)^{-1}$ to the first line of (18) yields $u + 4\pi c_{ee} B \rho_0 = -\tilde{V}$ a.e., that is $u = -\tilde{V} - 4\pi c_{ee} B \rho_0$. Then

$$\begin{aligned} \rho_0 &= \Gamma(-\tilde{V} - 4\pi c_{ee} B \rho_0 - \lambda) \\ &= (J')^{-1}[(-\tilde{V} - 4\pi c_{ee} B \rho_0 - \lambda)_+]. \end{aligned}$$

We apply J' to both sides of this equation and rearrange terms to obtain (14). It follows immediately that $\rho_0 \geqslant 0$ a.e. and $\int_{\mathbb{R}^3} \rho_0(y)\,dy = N$. Since $\Delta \tilde{V} \in L^1(\mathbb{R}^3)$, we must have $\Delta u \in L^1(\mathbb{R}^3)$. By Proposition 1 (iii) we see "$\nabla u(\infty) = 0$". But $\nabla \rho_0 = \Gamma'(u - \lambda) \nabla u$, so that

$$\begin{aligned} \text{“}\nabla \rho_0(\infty) &= \Gamma'(u(\infty) - \lambda) \nabla u(\infty) \\ &= \Gamma'_+(-\lambda) \nabla u(\infty) = 0\text{”} \end{aligned}$$

since $\Gamma'_+(-\lambda) < \infty$. Finally, $\Gamma(u - \lambda) \in L^\infty(\mathbb{R}^3)$ implies $\rho_0 \in L^\infty(\mathbb{R}^3)$. Thus we have $\rho_0 \in D_{N,M}$.

Remark: In conventional Thomas–Fermi theory, i.e., $c_{ee} = 1$, the condition $\Gamma'_+(-\lambda) < \infty$ is always satisfied if $N < Z$. If we use the Fermi–Amaldi approximation of $c_{ee} = (N-1)/N$, then $\Gamma'_+(-\lambda) < \infty$ holds for $N < Z + 1$. We shall see below that $\lambda = 0$ corresponds to

$$N = Z c_{ee}^{-1} = \begin{cases} Z, & \text{if } c_{ee} = 1, \\ Z + 1, & \text{if } c_{ee} = (N-1)/N. \end{cases}$$

When $J(r) = cr^p$ for $p > 1$, the condition $\Gamma'_+(0) < \infty$ is equivalent to $p \leqslant 2$; $\Gamma'_+(-\lambda) < \infty$ always holds for $p > 1$ and $\lambda > 0$.

Freeze $\lambda \geqslant 0$ and set

$$\beta(u) = 4\pi c_{ee} \Gamma(u - \lambda), \tag{20}$$

where Γ is given by (17). A simple calculation shows

$$\Delta \tilde{V} = 4k^3 Z e^{-2kr}, \tag{21}$$

and thus $\Delta \tilde{V} \in L^1(\mathbb{R}^3) \cap L^\infty(\mathbb{R}^3)$.

Proposition 3: Let $\beta: \mathbb{R} \to \mathbb{R}$ be a continuous, nondecreasing function with $\beta(0) = 0$. Then for every non-negative radial function $f \in L^1(\mathbb{R}^3) \cap L^\infty(\mathbb{R}^3)$ there is a unique solution $u \in \mathscr{M}^3(\mathbb{R}^3)$ satisfying

$$-\Delta u + \beta(u) = f \quad \text{in } \mathbb{R}^3. \tag{22}$$

Moreover, $\beta(u) \in L^1(\mathbb{R}^3) \cap L^\infty(\mathbb{R}^3)$ and $\|\beta(u)\|_\infty \leqslant \|f\|_\infty$.

Proof: Under the above hypothesis on β, if $f \in L^1(\mathbb{R}^3)$, then a unique solution $u \in \mathscr{M}^3(\mathbb{R}^3)$ of (22) exists and $\beta(u) \in L^1(\mathbb{R}^3)$ (cf. Ref. 12). Let $f \in \mathscr{C}_0^\infty(\mathbb{R}^3)$. Then Proposition 1 (iv) implies

$$\int_{\mathbb{R}^3} \beta(u) \xi(u)\,dx \leqslant \int_{\mathbb{R}^3} f \xi(u)\,dx$$

for any $\xi \in \mathscr{J}_0$. Take $\xi(\eta) = (\operatorname{sgn} \eta) \beta(\eta)^{q-1}$ for $\eta \in \mathbb{R}$ and large q. Then

$$\begin{aligned} 0 &\leqslant \left(\int_{\mathbb{R}^3} \beta(u(y)) \xi(u(y))\,dy \right)^{1/q} \\ &= \left(\int_{\mathbb{R}^3} |\beta(u(y))|^q\,dy \right)^{1/q} \leqslant \left(\int_{\mathbb{R}^3} f \xi(u(y))\,dy \right)^{1/q} \\ &\leqslant \left(\int_{\mathbb{R}^3} |f|\, |\beta(u(y))|^{q-1}\,dy \right)^{1/q}. \end{aligned}$$

Then by Hölder's inequality

$$\|\beta(u)\|_q^q \leqslant \|f\|_q \|\beta(u)^{q-1}\|_{q'} \leqslant \|f\|_q \|\beta(u)\|_q^{q-1},$$

where $1/q + 1/q' = 1$. It follows that

$$\|\beta(u)\|_q \leqslant \|f\|_q.$$

Letting $q \to \infty$, we see that $\beta(u) \in L^\infty$ and

$$\|\beta(u)\|_\infty \leqslant \|f\|_\infty. \tag{23}$$

Since f is a non-negative radial function, we can construct a sequence $\{f_n\} \subset \mathscr{C}_0^\infty(\mathbb{R}^3)$ such that $f_n \to f$ a.e. and in $L^1(\mathbb{R}^3)$ with $\|f_n\|_\infty \leqslant \|f\|_\infty$. Let u_n be the unique solution of $-\Delta u_n + \beta(u_n) = f_n$, $\Delta u_n \in L^1(\mathbb{R}^3)$, $u_n \in \mathscr{M}^3(\mathbb{R}^3)$, and $\beta(u_n) \in L^1(\mathbb{R}^3)$. Then there exists a subsequence $\{u_{n_k}\}_k$ such that $u_{n_k} \to u$ and $\beta(u_{n_k}) \to \beta(u)$ a.e. and in $L^1_{\text{loc}}(\mathbb{R}^3)$, where u is the solution of (22) (cf. Refs. 4–6 and 10). By (23), $\|\beta(u_{n_k})\|_\infty \leqslant \|f_n\|_\infty$. Passing to the limit we get

$$\|\beta(u)\|_\infty \leqslant \|f\|_\infty.$$

Remark: As Gallouët and Morel noted in Ref. 13, if f is in addition radial nonincreasing (resp. radial decreasing), then the solution u of (22) is also radial nonincreasing (resp. radial decreasing).

In the present context we apply Proposition 3 with $\beta(r) = 4\pi c_{ee} \Gamma(r - \lambda)$ and $f = \Delta \tilde{V}$. It is clear that both β and f defined in this way satisfy the hypotheses of Proposition 3. Thus we have a unique solution $u_\lambda \in \mathscr{M}^3(\mathbb{R}^3)$ of (18) for each $\lambda \geqslant 0$.

Theorem 3: Let $J(r) = cr^p$ with $\frac{4}{3} \leqslant p \leqslant 2$ and V the atomic Coulomb potential (4). Take $N_0 = \gamma Z$ where

$$\gamma = c_{ee}^{-1} = \begin{cases} 1, & \text{if } c_{ee} = 1 \\ N/(N-1), & \text{if } c_{ee} = (N-1)/N. \end{cases}$$

Then the minimization problem [for (12)] has a unique solution ρ_0 for $0 < N \leqslant N_0$ and no solution for $N > N_0$. Moreover if $0 < N < N_0$, the solution ρ_0 has compact support (and the assumption that $p \leqslant 2$ can be omitted). Finally, for $0 < N \leqslant N_0$,

$$\rho_0(r) \sim \text{const } e^{-2Zr} \quad \text{as } r \to 0. \tag{24}$$

Proof: Under the above hypotheses, the preceding discussion shows there is a unique solution u_λ of the first equation of (18). If we let $\rho_0 = \Gamma(u_\lambda - \lambda)$, Proposition 2 shows we have a solution of the Euler–Lagrange problem, and thus

we have a solution of the minimization problem by the Corollary to Theorem 2. Next set

$$N(\lambda) = \int_{\mathbb{R}^3} \Gamma(u_\lambda(x) - \lambda)dx.$$

The existence and nonexistence parts of the theorem follow from the next lemma.

Lemma: The function N: $\mathbb{R}^+ \to \mathbb{R}^+$ is nonincreasing, continuous and

$$\lim_{\lambda \to +\infty} N(\lambda) = 0.$$

In addition $N(\lambda)$ is strictly decreasing on $\{\lambda: N(\lambda) > 0\}$ and $N_0 = N(0) > 0$.

The proof of this lemma is analogous to that of Lemma 7 in Ref. 10 (cf. also Refs. 5 and 6). If $0 < N < N_0$, then $N = N(\lambda)$ for some unique $\lambda > 0$, and so $\Gamma(u_\lambda(x) - \lambda) \to \Gamma(-\lambda) = 0$ as $|x| \to \infty$, $N = N_0$ corresponds to $\lambda = 0$.

The inequality $N_0 \geqslant \gamma Z$ comes from the observations that $\int \Delta u_0 \geqslant 0$,

$$-\Delta u_0 + 4\pi c_{ee}\rho_0 = \Delta\tilde{V}, \quad \int_{\mathbb{R}^3} \Delta\tilde{V} = 4\pi Z,$$

which follows easily from (2.1). The upper bound follows from the fact $u_0(x) \sim c/|x|$ as $|x| \to \infty$, and that $p \geqslant \frac{4}{3}$.

It remains to show that (24) holds to first order in r. From (14) near $r = 0$, we have

$$pc_p\rho_0^{p-1}(r) - (Z/r)(1 - e^{-2kr}) - kZe^{-2kr} + 4\pi c_{ee}B\rho_0 + \lambda = 0.$$

We note that $B\rho_0$ is an even function of r. Consequently if we expand it in a Taylor series about $r = 0$, the first-order term in r vanishes. Likewise there is no first-order contribution from the Lagrange multiplier term λ. From the Taylor series expansion for e^{-2kr} we see

$$(Z/r)(1 - e^{-2kr}) = 2Zk - 2Zk^2r + (\text{terms of higher order}),$$

$$Zke^{-2kr} = Zk - 2Zk^2r + (\text{terms of higher order}).$$

If $\rho_0(r)$ satisfies (24), then to first order in r (14) becomes

$$-2c_p p\rho_0(0)^{p-1}(p-1)Zr = -4Zk^2r.$$

Choosing

$$k = [(c_p p(p-1)/2)\rho_0(0)^{p-1}]^{1/2} \tag{25}$$

then leads to the desired behavior at the origin.

This concludes the proof of the Theorem. □

Note that for $p = \frac{5}{3}$, (25) reduces to

$$k = [\tfrac{5}{9}c_{5/3}\rho_0(0)^{2/3}]^{1/2},$$

which is the k obtained by Parr and Ghosh.[1]

The Parr–Ghosh constraint

$$\int_{\mathbb{R}^3} \Delta\rho(x)e^{-2k|x|}dx = \int_{\mathbb{R}^3} \Delta(e^{-2k|x|})\rho(x)dx \in \mathbb{R} \tag{26}$$

forces the ground state electron density to be a bounded continuous function. As shown above, this is equivalent to requiring that

$$\int_{\mathbb{R}^3} r^{-1}\gamma(r)e^{-2kr}\rho(x)dx \in \mathbb{R} \tag{27}$$

with

$$\gamma(r) = 4k^2r - 4k. \tag{28}$$

But the constraint (27) can be imposed with other choices of γ which will ensure that the density is bounded and continuous and satisfies the nuclear cusp condition. Thus (26) is not a "canonical constraint." It would be of interest to know what further physically motivated conditions make a choice of γ unique. And if this can be done, would the unique γ be the Parr–Ghosh γ given by (28)? This is an open question.

ACKNOWLEDGMENTS

We deeply appreciate a conversation with Professor Robert Parr in November 1985. Parr informed us of the problem of this paper and kindly provided a preprint of Ref. 1.

The first named author gratefully acknowledges the partial support of an NSF grant.

[1] R. G. Parr and S. K. Ghosh, Proc. Natl. Acad. Sci. USA **83**, 3577 (1986).
[2] L. H. Thomas, Proc. Cambridge Philos. Soc. **23**, 542 (1927).
[3] E. Fermi, Rend. Acad. Naz. **6**, 602 (1927).
[4] Ph. Bénilan and H. Brezis, "The Thomas–Fermi problem," in preparation.
[5] H. Brezis, in *Contemporary Developments in Continuum Mechanics and Partial Differential Equations*, edited by G. M. de la Penha and L. A. Medeiros (North-Holland, Amsterdam, 1978), p. 81.
[6] H. Brezis, in *Variational Inequalities and Complementarity Problems: Theory and Applications*, edited by R. W. Cottle, F. Giannessi, and J. L. Lions (Wiley, New York, 1980), p. 53.
[7] E. H. Lieb and B. Simon, Phys. Rev. Lett. **33**, 681 (1973).
[8] E. H. Lieb and B. Simon, Adv. Math. **23**, 22 (1977).
[9] E. H. Lieb, Rev. Mod. Phys. **53**, 603 (1981).
[10] G. R. Rieder, Ph. D. thesis, Tulane University, 1986.
[11] E. Fermi and E. Amaldi, Mem. Accad. Ital. **6**, 119 (1934).
[12] Ph. Bénilan, H. Brezis, and M. G. Crandall, Ann. Scuola Norm. Sup. Pisa Cl. Sci. **2**, 523 (1975).
[13] Th. Gallouët and J. M. Morel, Nonlinear Anal. **7**, 971 (1983).

Relativistic plasma dispersion functions: Series, integrals, and approximations

P. A. Robinson[a)]
School of Physics, University of Sydney, New South Wales 2006, Australia

(Received 28 August 1986; accepted for publication 31 December 1986)

A number of results are presented involving the plasma dispersion functions appropriate to waves in weakly relativistic, magnetized, thermal plasmas. These results include generating functions, series, integral forms and interrelations, and several useful approximations.

I. INTRODUCTION

The dielectric properties of magnetized plasmas are of interest in studies of cyclotron emission, absorption, dispersion, and instability in physical situations as diverse as those pertaining to tokamaks, magnetic mirrors, planetary and stellar magnetospheres, and solar flares. (A list of references to these applications was given in Ref. 1.) In the case of weakly relativistic, thermal plasmas the relevant dielectric properties can be expressed in terms of a class of relativistic plasma dispersion functions (PDF's) whose properties were reviewed and extended in Ref. 1.

The purpose of the present work is to present, as briefly as possible, a number of results involving PDF's that have been obtained since the publication of Ref. 1. These results include several series and integral relationships, a family of generating functions, and a number of useful approximations. The approximations, in particular, are of interest in studies of Bernstein waves and magnetized Langmuir waves while the exact results significantly add to the known store of calculational tools available when working with PDF's.

II. EXACT RESULTS

A number of series and integral relationships involving the PDF's are derived in this section. Results from Ref. 1 are indicated by the prefix I.

The most commonly discussed relativistic PDF's are the *Dnestrovskii functions*[2] of index q, which may be written in the form[1,3] (I.84)

$$F_q(z) = \frac{1}{\Gamma(q)}\int_0^\infty du \frac{u^{q-1}e^{-u}}{u+z}. \tag{1}$$

Generating functions for families of Dnestrovskii functions may be obtained by substituting (1) into the series

$$S(q,h) = \sum_{j=0}^{\infty} \frac{h^j \Gamma(q+j)}{j!} F_{q+j}(z). \tag{2}$$

Provided $h<1$ and $q>0$ this step yields

$$S(q,h) = [\Gamma(q)/(1-h)^{q-1}]F_q[z(1-h)]. \tag{3}$$

Together, (2) and (3) yield a generating function for the functions $F_{q+j}(z)$.

Dnestrovskii functions with half-integer index are the ones of most interest in plasma physics. In the case $q=\frac{1}{2}$, (3) may be expressed in terms of known functions

$$S(\tfrac{1}{2},h) = (\pi/z^{1/2})e^{z(1-h)}\,\mathrm{erfc}[z^{1/2}(1-h)^{1/2}] \tag{4a}$$

$$= (-i\pi^{1/2}/z^{1/2})Z[iz^{1/2}(1-h)^{1/2}], \tag{4b}$$

where Z is the PDF of Fried and Conte.[4]

A further result follows from (2) if we take the limit $h\to 1$, in which case (Ref. 5, Eq. 6.1.17)

$$S(q,1) = z^{q-1}\int_0^\infty \frac{dv\, v^{q-1}}{v+1}$$

$$= \pi z^{q-1}\csc(\pi q), \quad 0<\mathrm{Re}\, q<1. \tag{5}$$

The *Shkarofsky functions*[1,6] are defined

$$\mathscr{F}_q(z,a) = -i\int_0^\infty \frac{dt}{(1-it)^q}\exp\left[izt - \frac{at^2}{1-it}\right]. \tag{6}$$

These functions contain the Dnestrovskii functions as a special case with $F_q(z) = \mathscr{F}_q(z,0)$.

The Shkarofsky functions can be written in a form analogous to (1) by the use of the identity [Ref. 6; (I.9)]

$$\mathscr{F}_q(z,a) = e^{-a}\sum_{j=0}^{\infty}\frac{a^j}{j!}F_{q+j}(z-a),$$

which leads to

$$\mathscr{F}_q(z,a) = a^{(1-q)/2}e^{-a} \times \int_0^\infty du \frac{u^{(q-1)/2}e^{-u}I_{q-1}[2(au)^{1/2}]}{u+z-a}, \tag{7}$$

where I_{q-1} is a modified Bessel function. If z and a are real and $z<a$ the contour of integration in (7) is chosen to pass above the pole to reproduce the known result [Ref. 6; Eq. (I.43)] for $\mathrm{Im}\,\mathscr{F}_q(z,a)$.

The theory of Bernstein waves requires the introduction of a class of functions more general then $F_q(z)$ or $\mathscr{F}_q(z,a)$. These functions are defined [(I.62) and (I.63)]

$$\mathscr{R}_l(z,a,\lambda,s) = -i\int_0^\infty \frac{dt}{(1-it)^l}e^{-\Lambda} \times I_s(\Lambda)\exp[izt - at^2/(1-it)], \tag{8}$$

$$R_l(z,\lambda,s) = \mathscr{R}_l(z,0,\lambda,s), \tag{9}$$

with $\Lambda = \lambda/(1-it)$.

Substitution of the sum

$$\sum_{s=-\infty}^{\infty} I_s(\Lambda) = e^{\Lambda}$$

into (8) immediately yields the result

[a)] New address: Department of Astrophysical, Planetary and Atmospheric Sciences, Campus Box 391, University of Colorado, Boulder, Colorado 80309-0391.

 0022-2488/87/051203-03$02.50

$$\sum_{j=-\infty}^{\infty} \mathscr{R}_l(z,a,\lambda,j) = \mathscr{F}_l(z,a). \tag{10}$$

A number of integral relations between Shkarofsky and Dnestrovskii functions are known. These include the following [(I.80) and (I.86); Ref. 3]:

$$\mathscr{F}_l(z,a) = \frac{2}{\Gamma(k)}\int_0^\infty du\, u^{2k-1} \times \exp(-u^2)\mathscr{F}_{l-k}(z+u^2,a) \tag{11}$$

$$= \pi^{-1/2}\int_{-\infty}^{\infty} du \times \exp(-u^2)F_{l-1/2}(z+u^2-2a^{1/2}u). \tag{12}$$

These formulas can be extended to the functions $\mathscr{R}_l$ and R_l with the aid of the identity (I.70)

$$\mathscr{R}_l(z,a,\lambda,s) = 2\int_0^\infty dx\, x \exp(-x^2) \times J_s^2[(2\lambda)^{1/2}x]\mathscr{F}_{l-1}(z+x^2,a). \tag{13}$$

If (11) is substituted into (13) and the order of the resulting integrals is reversed we find

$$\mathscr{R}_l(z,a,\lambda,s) = \frac{2}{\Gamma(k)}\int_0^\infty du\, u^{2k-1}\exp(-u^2) \times \mathscr{R}_{l-k}(z+u^2,a,\lambda,s). \tag{14}$$

Similarly, substitution of (12) into (13) and reversal of the order of the resulting integrals yields

$$\mathscr{R}_l(z,a,\lambda,s) = \pi^{-1/2}\int_{-\infty}^{\infty} du \exp(-u^2) \times R_{l-1/2}(z+u^2-2a^{1/2}u,\lambda,s). \tag{15}$$

III. APPROXIMATIONS

Under some circumstances the expressions (8) and (13) for the functions $\mathscr{R}_l$ can be quite difficult to evaluate numerically and to work with analytically. It is therefore of interest to assemble a set of approximate forms of the functions $\mathscr{R}_l$ which avoid these difficulties. Such approximations have already been found useful in the theory of Bernstein waves.[7–10]

We generalize the work of Robinson[9] by noting that each of the asymptotic forms of $\mathscr{R}_l$ and R_l is of the form

$$\mathscr{R}_l(z,a,\lambda,s) = e^{-\lambda}I_s(\lambda)\mathscr{G}_l(z,a,\lambda,s), \tag{16a}$$

$$R_l(z,\lambda,s) = e^{-\lambda}I_s(\lambda)G_l(z,\lambda,s), \tag{16b}$$

$$G_l(z,\lambda,s) = \mathscr{G}_l(z,0,\lambda,s), \tag{16c}$$

where $\mathscr{G}_l$ and G_l are linear combinations of Shkarofsky functions and Dnestrovskii functions, respectively. Robinson[9] proposed an approximation of the form

$$\mathscr{G}_l(z,a,\lambda,s) = \sum_{j=1}^{N} c_j(\lambda,s)\mathscr{F}_{q(j)}(z,a), \tag{17a}$$

with

$$\sum_{j=1}^{N} c_j(\lambda,s) = 1. \tag{17b}$$

The function $\mathscr{G}_l$ is not strongly sensitive to the exact form of the coefficients $c_j(\lambda,s)$ since the functions $\mathscr{F}_q$ depend only weakly on q. Robinson[9,10] used the following approximation with $a=0$ in studies of Bernstein waves

$$\mathscr{G}_l(z,a,\lambda,s) = \mathscr{F}_{l+s}(z,a)\exp\{-|\lambda|/(s+1)^2\} + \mathscr{F}_l(z,a)[1-\exp\{-|\lambda|/(s+1)^2\}]. \tag{18}$$

This approximation reproduces the correct limiting forms[1,9] for large and small $|\lambda|$ with a fraction error of order l^{-1}. On the principal branch, the fractional error in (18) is proportional to $|z|^{-1}$ and $|a|^{-1/2}$ for $|z|\gg 1$ and $|a|\gg 1$, respectively.

Alternative approximations to the integral (8) may be obtained by introducing the Debye approximation for $I_s(\Lambda)$ (Ref. 5, Eq. 9.7.7),

$$I_s(\Lambda)e^{-\Lambda} \simeq H^s/(2\pi)^{1/2}(s^2+\Lambda^2)^{1/4}, \tag{19a}$$

with

$$H = \frac{\Lambda}{s}\frac{\exp[(1+\Lambda^2/s^2)^{1/2}-\Lambda/s]}{1+(1+\Lambda^2/s^2)^{1/2}}. \tag{19b}$$

Upon approximating $I_s(\Lambda)e^{-\Lambda}$ to first order in the quantity it, (8) can then be written

$$\mathscr{R}_l(z,a,\lambda,s) \simeq -i\int_0^\infty dt \frac{e^{-\lambda}I_s(\lambda)}{(1-it)^{q(\lambda,s)}} \times \exp[izt - at^2/(1-it)] = e^{-\lambda}I_s(\lambda)\mathscr{F}_{q(\lambda,s)}(z,a), \tag{20}$$

with

$$q(\lambda,s) = l + (s^2+\lambda^2)^{1/2} - \lambda - \lambda^2/2(s^2+\lambda^2). \tag{21}$$

Although (20) is somewhat difficult to evaluate numerically for arbitrary values of q, this expression reproduces the correct asymptotic behavior of $\mathscr{R}_l$ and has proved to be of use in analytic work on Bernstein waves.[11] We note that the final term in (21) arises from approximation of the factor $(s^2+\Lambda^2)^{-1/4}$ in (19a).

The accuracy of the expression (20) is better than that of (18): the appropriate limiting forms for $|\lambda|\gg 1$ and $|\lambda|\ll 1$ are reproduced exactly by (20) while, on the principal branch, the fractional error is proportional to $|z|^{-1}$, $|a|^{-1/2}$ and l^{-1} for $|z|\gg 1$, $|a|\gg 1$, and $l\gg 1$, respectively. If l is fixed, the largest errors occur when $|a|$ is small; in this case a numerical analysis yields maximum errors of $\sim 20\%$, $\sim 10\%$, and $\sim 6\%$ for $z>0$ and s small with $l=\frac{5}{2}$, $\frac{7}{2}$, and $\frac{9}{2}$, respectively. In each case the largest errors occur for $z\ll 1$.

An alternative approximate expression for $\mathscr{R}_l$ may be obtained by evaluating (8) using the Debye approximation (19a) and (19b) and the method of steepest descents. This gives

$$\mathscr{R}_l(z,a,\lambda,s) \simeq -i\int_0^\infty dt\, e^{-\lambda}I_s(\lambda)\exp(i\beta t - \alpha t^2) = -\tfrac{1}{2}\alpha^{-1/2}e^{-\lambda}I_s(\lambda)Z\left(\tfrac{1}{2}\beta\alpha^{-1/2}\right), \tag{22}$$

with

$$\alpha = a + \frac{1}{2}l - \lambda + \frac{s^2+2\lambda^2}{2(s^2+\lambda^2)^{1/2}} - \frac{\lambda^2(3s^2+2\lambda^2)}{4(s^2+\lambda^2)^2}, \tag{23a}$$

$$\beta = z + l + (s^2 + \lambda^2)^{1/2} - \lambda - \frac{\lambda^2}{2(\lambda^2 + s^2)}. \quad (23b)$$

The final term in each of (23a) and (23b) arises from approximation of the factor $(s^2 + \Lambda^2)^{-1/4}$ in (19a). Since it involves only Bessel functions and the Z function, (22) is simpler to evaluate numerically than (20), but is valid only on the principal branch of $\mathscr{R}_l$. This is not a serious disadvantage since the principal branch is the one of interest in most applications. Equation (22) generalizes a similar result for $\mathscr{F}_q(z,a)$ obtained by Maroli and Petrillo[12] [Eq. (I.38)] and has fractional errors of order $l^{-1/2}$, $|a|^{-1/2}$, and $|z|^{-1}$ for $l \gg 1$, $|a| \gg 1$, and $|z| \gg 1$, respectively. This approximation is the least accurate of those considered here but is qualitatively correct and is useful for semiquantitative work because of its relatively simple functional form.

IV. SUMMARY

We have obtained a number of exact and approximate results involving those relativistic PDF's which are appropriate to the description of cyclotron waves in weakly relativistic thermal plasmas. The exact results consist of a family of generating functions for the Dnestrovskii functions and a number of series and integral relations involving these and more general PDF's. Approximations are given for the PDF's relevant to the description of Bernstein waves and other large-wave-number cyclotron waves; some of these approximations have already found application in analyses of such waves.

[1] P. A. Robinson, J. Math. Phys. **27**, 1206 (1986).
[2] Y. N. Dnestrovskii, D. P. Kostomarov, and N. V. Skrydlov, Sov. Phys. Tech. Phys. **8**, 691 (1964).
[3] A. C. Airoldi and A. Orefice, J. Plasma Phys. **27**, 515 (1982).
[4] B. D. Fried and S. D. Conte, *The Plasma Dispersion Function* (Academic, New York, 1961).
[5] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1980).
[6] I. P. Shkarofsky, Phys. Fluids **9**, 561 (1966).
[7] A. Airoldi-Crescentini, E. Lazzaro, and A. Orefice, *Proceedings of the 2nd Joint Grenoble–Varenna Symposium on Heating in Toroidal Plasmas* (CEC, Brussels, 1980), p. 225.
[8] M. Bornatici, C. Maroli, and V. Petrillo, *Proceedings of the 3rd Joint Varenna–Grenoble Symposium on Heating in Toroidal Plasmas* (CEC, Brussels, 1982), p. 691.
[9] P. A. Robinson, "Dispersion of electron Bernstein waves including weakly relativistic and electromagnetic effects. I. Ordinary modes," J. Plasma Phys, in press.
[10] P. A. Robinson, "Dispersion of electron Bernstein waves including weakly relativistic and electromagnetic effects. II. Extraordinary modes," J. Plasma Phys, in press.
[11] P. A. Robinson, "Weakly relativistic dispersion of Bernstein waves," submitted to Phys. Fluids.
[12] C. Maroli and V. Petrillo, Phys. Scripta **24**, 955 (1981).

Critical exponent of susceptibility for a class of general ferromagnets in $d>4$ dimensions

Kei-ichi Kondo[a)]
Department of Physics, Faculty of Science, University of Tokyo, Bunkyo-ku, Tokyo 113, Japan
Takashi Hara
Institute of Physics, College of General Education, University of Tokyo, Komaba, Meguro-ku, Tokyo 153, Japan

(Received 27 August 1986; accepted for publication 7 January 1987)

A "rigorous proof" is presented that the critical exponent γ of the susceptibility χ takes its mean field value in $d>4$ dimensions for ferromagnets with single spin measure in the Brydges–Fröhlich–Spencer class, modulo a numerical calculation of a certain function $I(d)$ of dimension d. This class of ferromagnets contains, for example, the Ising spin model and lattice scalar φ^4 model.

I. INTRODUCTION AND RESULTS

In this paper we consider one-component lattice scalar field or spin models defined on the hypercubic lattice Z^d. The n-point lattice Schwinger or correlation function is defined by

$$\begin{aligned} S_n(x_1,\ldots,x_n) &\equiv \langle \varphi(x_1)\cdots\varphi(x_n)\rangle \\ &= Z^{-1}\int \prod_{x\in Z^d} d\nu(\varphi(x)) \\ &\quad\times \exp\left[\frac{1}{2}\sum_{x,y\in Z^d} J_{x,y}\varphi(x)\varphi(y)\right] \\ &\quad\times \varphi(x_1)\cdots\varphi(x_n), \end{aligned} \tag{1}$$

where $d\nu$ is the single spin measure and Z is the partition function. Here $\varphi(x)$ denotes an (unbounded) spin variable at site x. In the following, we restrict our attention to the ferromagnetic nearest-neighbor interaction.

$$J_{x,y} = J\delta_{|x-y|,1}, \quad J\geqslant 0. \tag{2}$$

In addition, we assume that the single spin measure belongs to the BFS (Brydges–Fröhlich–Spencer) class.[1] That is, the single spin measure is written in the form

$$d\nu(\varphi(x)) = \exp[-V(\varphi(x)^2)]d\varphi(x), \tag{3}$$

with the potential function $V(\varphi^2)$ satisfying the following condition:

$$V''(x)\geqslant 0 \quad \text{for } x\geqslant 0. \tag{4}$$

For example, the scalar φ^4 model, with

$$V(\varphi^2) = \lambda\varphi^4 + \sigma\varphi^2, \quad \lambda\geqslant 0, \quad \sigma\in R,$$

and the Ising model are obviously in the BFS class. [Note that $V(x)$ is not necessarily a polynomial in x.]

We define the connected four-point function by

$$\begin{aligned} &U_4(x_1,\ldots,x_4) \\ &= S_4(x_1,\ldots,x_4) - S_2(x_1,x_2)S_2(x_3,x_4) \\ &\quad - S_2(x_1,x_3)S_2(x_2,x_4) - S_2(x_1,x_4)S_2(x_2,x_3). \end{aligned} \tag{5}$$

It is expected that, in $d>4$ dimensions, the mean field theory is exact and critical exponents take their mean field values. In fact, we can prove the following statement rigorously modulo a numerical evaluation of a certain function $I(d)$ [defined by (14)] of dimension d.

Theorem[2,3]**:** The critical exponent of the susceptibility takes its mean field value, i.e., $\gamma = 1$ in $d\geqslant 5$ dimensions for one-component lattice scalar models defined above.

In order to prove $\gamma = 1$, we must show that there are constants c_1, c_2 such that

$$c_1 \leqslant -\frac{\partial\chi^{-1}}{\partial J} \leqslant c_2. \tag{6}$$

In fact, integrating out this from J ($<J_c$) and J_c and taking account of $\chi(J)^{-1}\downarrow 0$ as $J\uparrow J_c$, we obtain

$$c_1(J_c - J)\leqslant \chi^{-1}(J)\leqslant c_2(J_c - J), \tag{7}$$

which implies $\gamma = 1$. Since

$$\begin{aligned} \frac{\partial\chi}{\partial J} &= \frac{1}{2}\sum_{\substack{x_2,x_3,x_4 \\ |x_3-x_4|=1}} \langle\varphi(0)\varphi(x_2);\ \varphi(x_3)\varphi(x_4)\rangle \\ &= \frac{1}{2}\sum_{\substack{x_2,x_3,x_4 \\ |x_3-x_4|=1}} \{U_4(x_1,\ldots,x_4) \\ &\quad + \langle\varphi(0)\varphi(x_3)\rangle\langle\varphi(x_2)\varphi(x_4)\rangle \\ &\quad + \langle\varphi(0)\varphi(x_4)\rangle\langle\varphi(x_2)\varphi(x_3)\rangle\}, \end{aligned} \tag{8}$$

we obtain

$$-\frac{\partial\chi}{\partial J} = 2d + \frac{1}{2}\chi^{-2}\sum_{\substack{x_2,x_3,x_4 \\ |x_3-x_4|=1}} U_4(0,x_2,x_3,x_4). \tag{9}$$

Hence $\gamma\geqslant 1$ is derived for any d by the Lebowitz inequality[2,4] $U_4\leqslant 0$ with $c_2 = 2d$. (We omit all the details about the infinite volume limit, see, for example, Sokal.[5])

Now, in order to show the converse, we need an appropriate lower bound on U_4. The following inequality due to Aizenman and Fröhlich[2,3] gives such a lower bound:

[a)] Fellow of the Japan Society for the Promotion of Sciences. Address after 1 April 1987: Department of Physics, Faculty of Science, Nagoya University, Chikusa-ku, Nagoya 464, Japan.

 0022-2488/87/051206-03$02.50

$$U_4(x_1,\ldots,x_4)$$

$$\geqslant -\sum_{z,z',z''\in Z^d}[S_2(x_1,z)J_{z,z'}S_2(z',x_2)S_2(x_3,z)J_{z,z''}S_2(z'',x_4) + \text{two permutations}] - E_4(x_1,\ldots,x_4), \tag{10a}$$

where the sum runs over all the lattice sites and E is the "error term"

$$E_4(x_1,\ldots,x_4) = \delta_{x_3,x_4}\sum_{z'\in Z^d} S_2(x_1,x_3)J_{x_3,z'}S_2(z',x_2)S_2(x_3,x_3) + \delta_{x_1,x_2}S_2(x_1,x_1)\sum_{z''\in Z^d} S_2(x_1,x_3)J_{x_1,z''}S_2(z'',x_4) + \delta_{x_1,x_2}\delta_{x_3,x_4}\delta_{x_1,x_3}S_2(x_1,x_1)S_2(x_3,x_3) + \text{two permutations}. \tag{10b}$$

Inequalities of this type can be proved[2-4,6] for lattice scalar models with the potential function satisfying condition (4).

Note that $\gamma = 1$ has been proved for φ^4 models in Refs. 3 and 7. Its proof is based on the Griffiths–Simon representation and hence its generalization to other scalar, e.g., φ^6, models is not clear at present. On the other hand, using the inequality (10), $\gamma = 1$ can be proved for all scalar models satisfying (4), as is announced in Ref. 2. Because its complete exposition has not been found in the literature, we find it worthwhile to give its detailed proof in this paper. For the present achievement of the proof of the mean field properties for other critical exponents, see Ref. 8.

II. LOWER BOUND ON $|\partial\chi^{-1}/\partial J|$

Substituting inequality (10) into (9), we obtain

$$-\frac{\partial\chi^{-1}}{\partial J} \geqslant 2d - \frac{1}{2}J^2\chi^{-2}\sum_{\substack{x_2,x_3,x_4\\ |x_3-x_4|=1}}\Bigg(\sum_{\substack{z,z',z''\\ |z-z'|=1\\ |z-z''|=1}}\{S_2(x_1,z) \times S_2(z',x_2)S_2(x_3,z)S_2(z'',x_4) + \text{two permutations}\} + E_4(x_1,\ldots,x_4)\Bigg). \tag{11}$$

By the translation invariance and the Fourier transform,

$$\sum_{\substack{x_2,x_3,x_4\\ |x_3-x_4|=1}}\sum_{\substack{z,z',z''\\ |z-z'|=1\\ |z-z''|=1}}\{S_2(x_1,z)S_2(z',x_2)S_2(x_3,z)S_2(z'',x_4) + \text{two permutations}\}$$

$$= \chi^2\int\frac{d^dp}{(2\pi)^d}\left(2d\left|\sum_{|z|=1}e^{ip\cdot z}\right|^2 + 2\left|\sum_{|z|=1}e^{ip\cdot z}\right|^3\right)\times\{\tilde{S}_2(p)\}^2$$

$$\leqslant 6d\chi^2\int\frac{d^dp}{(2\pi)^d}\left|2\sum_{\mu=1}^d\cos p_\mu\right|^2\{\tilde{S}_2(p)\}^2 \leqslant 6dI(d)\chi^2/J^2, \tag{12}$$

where we defined

$$I(d)\equiv\int_{p\in(-\pi,\pi)^d}\frac{d^dp}{(2\pi)^d}\frac{|\Sigma_{\mu=1}^d\cos p_\mu|^2}{|\Sigma_{\mu=1}^d(1-\cos p_\mu)|^2} \tag{13}$$

and used the infrared bound[9]

$$\tilde{S}_2(p)\leqslant\left(2J\sum_{\mu=1}^d(1-\cos p_\mu)\right)^{-1}, \tag{14}$$

in the last step.

Similarly, the contribution from the "error term" is

$$\sum_{\substack{x_2,x_3,x_4\\ |x_3-x_4|=1}}E_4(x_1,\ldots,x_4) = 4J\chi\langle\varphi^2\rangle\sum_{\substack{z,z'\\ |z|=|z'|=1}}S_2(0,z+z') + J\langle\varphi^2\rangle\sum_{\substack{y,z,z'\\ |z|=|z'|=1}}S_2(0,y)S_2(0,y+z+z') \leqslant \text{const }(d)\chi. \tag{15}$$

Therefore we obtain

$$-\frac{\partial\chi^{-1}}{\partial J}\geqslant 2d\left(1-\frac{3}{2}I(d)+O\left(\frac{1}{\chi}\right)\right). \tag{16}$$

Note that, as $J\uparrow J_c$, $\chi(J)^{-1}\downarrow 0$.

Hence, in order to prove the theorem, we have only to show that $I(d)<\frac{2}{3}$ for $d\geqslant 5$, which will be shown in the next section.

TABLE I. Integrals $I(d;1)$, $I(d;2)$, and $I(d)$.

d	$I(d;1)$	$I(d;2)$	$I(d)$
3	0.505 462 019 7173(1)		
4	0.309 866 780 4621(1)		
5	0.231 261 624 9680(1)	0.077 397 657 6153(1)	0.622 325 190 7019(1)
6	0.186 160 562 2044(1)	0.042 059 662 6951(1)	0.280 221 110 5712(3)
7	0.156 272 330 7983(1)	0.027 893 595 2965(1)	0.178 973 538 3505(4)
8	0.134 830 876 5021(1)	0.020 140 668 5891(1)	0.131 708 765 6684(5)
9	0.118 638 454 0424(1)	0.015 307 074 4570(1)	0.104 380 858 2516(5)

TABLE II. Integrals $L(d;1)$ and $L(d;2)$.

d	$L(d;1)$	$L(d;2)$
3	0.516 386 059 1520(1)	
4	0.239 467 121 8485(1)	
5	0.156 308 124 8402(1)	0.934 941 440 3823(1)
6	0.116 963 373 2267(1)	0.514 147 857 0246(1)
7	0.093 906 315 5878(1)	0.366 786 169 5262(2)
8	0.078 647 012 0169(1)	0.289 002 789 7022(2)
9	0.067 746 086 3814(1)	0.239 873 031 0144(2)

III. PROPERTIES AND ESTIMATES OF THE INTEGRAL $I(d;n)$ AND $I(d)$

First, define

$$I(d;n) \equiv \int_{-\pi}^{\pi} \prod_{\mu=1}^{d} \frac{dp_\mu}{2\pi} \left(\sum_{\mu=1}^{d} (1 - \cos p_\mu) \right)^{-n}. \quad (17)$$

Applying the identity

$$\int_0^\infty dt\, t^{n-1} e^{-tX} = X^{-n} \quad (X \geqslant 0, \quad n = 1,2,\ldots),$$

with $X = \Sigma_{\mu=1}^{d} (1 - \cos p_\mu)$, we obtain

$$I(d;n) = \int_0^\infty dt\, t^{n-1} e^{-dt} f(t)^d, \quad (18a)$$

where

$$f(t) \equiv \int_{-\pi}^{\pi} \frac{d\theta}{2\pi} e^{t\cos\theta} = I_0(t). \quad (18b)$$

By the change of variable $t = x/d$,

$$d^n I(d;n) = \int_0^\infty dx\, x^{n-1} e^{-x} f\left(\frac{x}{d}\right)^d \equiv K(d;n). \quad (19)$$

Then we find

$$I(d) = K(d;2) - 2K(d;1) + 1. \quad (20)$$

Since $f(x/d)^d$ is monotone decreasing in d (see Refs. 10 and 11) and

$$f(x/d)^d \downarrow 1 \quad \text{as } d \uparrow \infty, \quad (21)$$

$K(d;n)$ is monotone decreasing in d (see Ref. 10) and

$$K(d;n) \downarrow \int_0^\infty dx\, x^{n-1} e^{-x} = (n-1)! \quad \text{as } d \uparrow \infty. \quad (22)$$

Therefore we find

$$I(d) \downarrow 0 \quad \text{as } d \uparrow \infty. \quad (23)$$

We also remark that $I(d;n)$ diverges if $d \leqslant 2(1+n)$. In particular, $I(d) = \infty$ for $d \leqslant 4$. The results of numerical calculations are listed in Tables I and II.

Proof of $I(d) < \frac{2}{3}$:

Case (i) $d = 5$: Use directly the results of the numerical calculations (Table I).

Case (ii) $d \geqslant 6$: Introducing

$$L(d;n) = K(d;n) - (n-1)!$$

$$= \int_0^\infty dx\, x^{n-1} e^{-x} \left[f\left(\frac{x}{d}\right)^d - 1 \right] \geqslant 0. \quad (24)$$

We can write

$$I(d) = L(d;2) - 2L(d;1). \quad (25)$$

As $K(d;n)$ is monotone decreasing in d, so is $L(d)$. Thus

$$I(d) \leqslant L(d;2) \leqslant L(6;2).$$

By Table III, $L(6;2) < \frac{2}{3}$.

This completes the proof of the theorem for one-component systems. ■

ACKNOWLEDGMENTS

The numerical calculation was done at the Nagoya University Computation Center.

This work was supported in part by a Grant-in-Aid for Scientific Research of the Ministry of Education, Science and Culture (60790026).

[1] D. Brydges, J. Fröhlich, and T. Spencer, Commun. Math. Phys. **83**, 123 (1982).
[2] J. Fröhlich, Nucl. Phys. B **200** (FS4), 281 (1982).
[3] M. Aizenman, Phys. Rev. Lett. **47**, 1, 886 (E) (1981); Commun. Math. Phys. **86**, 1 (1982).
[4] T. Hattori, J. Math. Phys. **24**, 2200 (1983).
[5] A. D. Sokal, J. Stat. Phys. **25**, 25 (1981).
[6] D. Brydges, "Field theory and Symanzik's polymer representation," in *Gauge Theories: Fundamental Interactions and Rigorous Results*, 1981 Brasov lectures, edited by P. Dita, V. Georgescu, and R. Purice (Birkhäuser, Boston, 1982).
[7] M. Aizenman and R. Graham, Nucl. Phys. B **225**, (FS9), 209 (1983).
[8] K.-I. Kondo, Ph. D. thesis, Nagoya University, February 1986; K.-I. Kondo, Prog. Theor. Phys. **77**, 473 (1987) T. Hara, T. Hattori, and H. Tasaki, J. Math. Phys. **26**, 2922 (1985).
[9] J. Fröhlich, B. Simon, and T. Spencer, Commun. Math. Phys. **50**, 79 (1976).
[10] F. Dyson, E. Lieb, and B. Simon, J. Stat. Phys. **18**, 335 (1978).
[11] W. Driessler, L. Laudau, and J. Fernando Perez, J. Stat. Phys. **20**, 123 (1979).

Erratum: Automorphisms of algebraic varieties and Yang–Baxter equations [J. Math. Phys. 27, 2776 (1986)]

Jean-Marie Maillard
Laboratoire de Physique Théorique et Hautes Energies, Université Pierre et Marie Curie, Tour 16-ler étage, 4, place Jussieu, 75252 Paris Cédex 05 France

(Received 18 November 1986; accepted for publication 31 December 1986)

Page 2779, left column, line 32 from the top, should read as follows:
model,[19] the elliptic parametrization of the model is giv-

The final sentence of Sec. IV A should be replaced by the following: The genus of an algebraic curve defined in general by the intersection of a quadric, a cubic, and a quartic in P_4 can be calculated from the formula of addition of the characteristic of Euler–Poincaré...leading to a rather high genus; in fact the hard hexagon parametrization corresponds to two relations between the previous constants C_i, $C_1 \cdot C_2 = 1$ and $C_1 + C_2 = C_3$, that reduce the number of equations to only two, leading to a ruled surface $(E \times P_1)$.

 0022-2488/87/051209-01$02.50