Normal forms of elements of classical real and complex Lie and Jordan algebras[a)]

D. Ž. Djoković
Faculty of Mathematics, University of Waterloo, Waterloo, Ontario, Canada

J. Patera and P. Winternitz
Centre de recherche de mathématiques appliquées, Université de Montréal, Montréal, Québec, Canada

H. Zassenhaus
Department of Mathematics, Ohio State University, Columbus, Ohio 43210

(Received 21 April 1982; accepted for publication 1 July 1982)

Elements of the classical complex and real Lie and Jordan algebras with involutions are classified into conjugacy classes under the action of the corresponding classical Lie group. Normal forms of representatives of each conjugacy class are chosen so as to resemble the Jordan normal forms of $n \times n$ complex matrices. For completeness similar results are given for $gl(n,\mathbb{C})$, $gl(n,\mathbb{R})$, and $gl(n,\mathbb{H})$.

PACS numbers: 02.10.Sp, 02.20.Hj, 02.20.Sv

1. INTRODUCTION

The purpose of this paper is to provide a comprehensive list of normal forms of matrices of two kinds. The first are matrices belonging to the defining representations of real and complex classical algebras. The normal forms are suitably chosen representatives of the conjugacy classes of the elements under the action of the corresponding classical Lie group. The solution of this problem has been known[1] for a long time for the algebras $gl(N,D)$ and $sl(N,D)$, where D stands either for the real numbers $\mathbb{R}$, the complex numbers $\mathbb{C}$, or the quaternions $\mathbb{H}$.

The normal forms are the familiar Jordan normal forms of the corresponding $N \times N$ matrices. For the remaining classical Lie algebras, i.e., those with an involution α, the problem has been solved relatively recently by Burgoyne and Cushman,[2] although for Lie algebras of several types it has been considered earlier by a number of authors.[3,4] Our solution of the problem is independent of that of Ref. 2; the results are naturally equivalent. In addition to a "basis independent" description of the normal forms, we summarize the results in an explicit ready-to-use matrix form.

The set

$$\sigma_-(\alpha/D^{N\times N}) = \{X \mid X \in gl(N,D) \text{ and } \alpha(X) = -X\} \tag{1.1}$$

of α-skew-symmetric matrices over D is closed under commutation and thus forms the classical Lie algebra $\sigma_-(\alpha/D^{N\times N})$ with the involution α. In many applications, however, one also encounters matrices which are α-symmetric. The set

$$\sigma_+(\alpha/D^{N\times N}) = \{X \mid X \in gl(N,D) \text{ and } \alpha(X) = X\} \tag{1.2}$$

of such matrices is closed under anticommutation and thus forms a Jordan algebra. It is defined in terms of the same involution α which determines the Lie algebra (1.1). The main result of this paper is a unified method of treatment of α-symmetric and α-skew-symmetric matrices. The new results are just for the case $D = \mathbb{H}$ and $\rho = \pm 1$. For the connection to some related classification tasks see the discussion at the very end of the paper.

This article contains a classification of the elements X of $\sigma_\rho(\alpha/D^{N\times N})$, $\rho = \pm 1$ up to equivalence transformation gXg^{-1}, $g \in O(\alpha/D^{N\times N})$, and in each conjugacy class a particularly simple representative is chosen as the normal form.

In applications the suggested normal forms make it particularly easy to recognize to which orbit an arbitrarily chosen element belongs and they simplify the task of transferring results obtained for a particular Lie algebra, say $sl(n,\mathbb{C})$, to all other classical Lie algebras. Indeed, our normal forms have proven to be particularly suitable for the study of versal deformations.[5] They can be used whenever the problem of orbit analysis on Lie groups and Lie algebras or specific sets of matrices arises, which is becoming more and more frequent in theoretical physics and applied mathematics. As examples let us mention: (i) The relationship between normal forms in phase space.[4] (ii) The study of the separation of variables in partial differential equations[6] (an "ignorable variable" corresponds to each different normal form and a classification of "essential" separable variables is related to a classification of α-symmetric matrices). (iii) The problem of finding systems of ordinary nonlinear differential equations with nonlinear superposition principles.[7] Our own interest in the problem arose in connection with the problem of classifying all maximal abelian subalgebras of the classical Lie algebras[8] for which a convenient normal form of the corresponding Lie algebra elements was indispensable.

The article is organized as follows. Section 2 contains notations, definitions and some conventions. The results of the paper are found in Sec. 3. There are four classification theorems, two for the algebras of the linear type and two for the quadratic ones, as well as Table III containing the indecomposable normal forms in the quadratic case. Section 4 contains a discussion of the results, some examples, as well as the discussion of the classification of α-normal matrices.

[a)] Supported in part by the Natural Sciences and Engineering Research Council of Canada and by the Ministère de l'Education du Québec. Published with the help of an FCAC grant for support of research.

 0022-2488/83/061363-12$02.50 © 1983 American Institute of Physics

2. DEFINITIONS AND NOTATIONS

1. Let D be one of the three division algebras $\mathbb{R}$, $\mathbb{C}$, and $\mathbb{H}$ of dimension 1, 2, and 4, respectively, over the real field, and F be the center of D. Thus one has

D	$\mathbb{R}$	$\mathbb{C}$	$\mathbb{H}$
F	$\mathbb{R}$	$\mathbb{C}$	$\mathbb{R}$

(2.1)

2. The *general linear classical Lie algebra*

$$gl(N,D) = \{X \mid X \in D^{N\times N}\} \tag{2.2}$$

is the Lie algebra associated with the $N \times N$ matrices over D. The *special linear classical Lie algebra*

$$sl(N,D) = \{X \mid X \in D^{N\times N} \text{ and } \operatorname{Tr} X = 0\} \tag{2.3}$$

is the Lie algebra of traceless matrices of $D^{N\times N}$. Thus there are three Lie algebras of the general linear type: $gl(N,\mathbb{R})$, $gl(N,\mathbb{C})$, and $gl(N,\mathbb{H})$. Since $\mathbb{H}$ has a nontrivial center $F = \mathbb{R}$, the Lie algebra $gl(N,\mathbb{H})$ of quaternionic $N \times N$ matrices is a real Lie algebra of dimension $d = 4N^2$.

The general linear group $\mathrm{GL}(N,D)$ consists of all invertible matrices of $D^{N\times N}$:

$$\mathrm{GL}(N,D) = \{Y \mid Y \in D^{N\times N} \text{ and } YY^{-1} = I_N\}, \tag{2.4}$$

where I_N is the $N \times N$ identity matrix. The group $\mathrm{GL}(N,D)$ has N^2 parameters over D, so that it has $d = N^2 \dim D/\mathbb{R}$ real parameters. It contains the special linear group

$$\mathrm{SL}(N,D) = \{Y \mid Y \in D^{N\times N} \text{ and } \det Y = 1\} \tag{2.5}$$

of $\mathrm{GL}(N,D)$ elements with determinant equal to 1 as a normal subgroup.

3. The involution α is an involutory antiautomorphism $\alpha: D^{N\times N} \to D^{N\times N}$ acting on $X \in D^{N\times N}$ according to

$$\alpha(X) = KX^{\dagger}K^{-1}, \quad \det K \neq 0, \quad K^{\dagger} = \pm K = \epsilon K. \tag{2.6}$$

The symbol $X^{\dagger}$, defined by

$$X^{\dagger} \equiv (x_{ij})^{\dagger} = (\alpha_0(x_{ij}))^T = (\alpha_0(x_{ji})), \tag{2.7}$$

is the transposition of X combined with the action of an involution $\alpha_0: D \to D$ fixing the real number field.

Two involutions α, α' are equivalent if one is obtained from the other by conjugation by an automorphism of $D^{N\times N}$, i.e., $S\alpha(X)S^{-1} = \alpha'(SXS^{-1})$. Thus, if α is given by (2.6) and α' by

$$\alpha'(X) = LX^{\dagger}L^{-1}, \quad \det L \neq 0, \quad L^{\dagger} = \pm L, \tag{2.8}$$

then α and α' are equivalent iff there exists S such that

$$L = \lambda SKS^{\dagger}, \quad S \in D^{N\times N}, \quad \det S \neq 0, \quad 0 \neq \lambda \in F. \tag{2.9}$$

4. The set of matrices

$$\sigma_{\rho}(\alpha/D^{N\times N}) = \{X \mid X \in D^{N\times N} \text{ and } \alpha(X) = \rho X\}, \quad \rho = \pm 1, \tag{2.10}$$

which is either α-symmetric ($\rho = +1$) or α-skew-symmetric ($\rho = -1$) over the field of reference

$$F^{\alpha} = \{\xi \mid \xi \in F \text{ and } \alpha_0(\xi) = \xi\}, \tag{2.11}$$

is the main subject of interest in this paper. The classification of $\sigma_{\rho}(\alpha/D^{N\times N})$ hinges on the classification of nonequivalent involutions α. The results are well known.[9] They are summarized in Table I, where also the Lie algebra $\sigma_{-}(\alpha/D^{N\times N})$ corresponding to each α is shown. In particular one has

$$\alpha(K) = \epsilon K \text{ or equivalently } K^{\dagger} = \epsilon K. \tag{2.12}$$

5. The largest of the locally isomorphic Lie groups of matrices corresponding to the Lie algebra $\sigma_{-}(\alpha/D^{N\times N})$ is defined by

$$O(\alpha/D^{N\times N}) = \{X \mid X \in D^{N\times N} \text{ and } \alpha(X) = X^{-1}\}. \tag{2.13}$$

Notice that from (2.13) one has

$$XKX^{\dagger} = K. \tag{2.14}$$

6. It is often convenient to represent complex numbers and quaternions by 2×2 matrices. One has the 1-1 correspondence

$$\alpha = a + ib \leftrightarrow \begin{pmatrix} a & b \\ -b & a \end{pmatrix}, \tag{2.15}$$

$$q = a + ib + jc + kd \leftrightarrow \begin{pmatrix} \beta & \gamma \\ -\gamma^* & \beta^* \end{pmatrix} = \begin{pmatrix} a+id & b+ic \\ -b+ic & a-id \end{pmatrix}, \tag{2.16}$$

where $a,b,c,d \in \mathbb{R}$, $\alpha,\beta,\gamma \in \mathbb{C}$, and $i^2 = j^2 = k^2 = -1$, $ij = k$ (cyclic). Then, in particular, the conjugation turns out to be the transposition in the case of $\mathbb{C}$ and Hermitian conjugation for $\mathbb{H}$:

TABLE I. Involutions defining the classical quadratic Lie algebras.

$\alpha_0(x)$	ϵ	D	F	F^{α}	Lie algebras	N	K	
x	1	$\mathbb{C}$	$\mathbb{C}$	$\mathbb{C}$	$o(N,\mathbb{C})$	N	$K = K^T$	
x	−1	$\mathbb{C}$	$\mathbb{C}$	$\mathbb{C}$	$sp(2n,\mathbb{C})$	$2n$	$K = -K^T$	
x	1	$\mathbb{R}$	$\mathbb{R}$	$\mathbb{R}$	$o(p,q)$	$p+q$	$K = K^T$	$\operatorname{sgn} K = (p,q)$
x	−1	$\mathbb{R}$	$\mathbb{R}$	$\mathbb{R}$	$sp(2n,\mathbb{R})$	$2n$	$K = -K^T$	
x^*	1	$\mathbb{C}$	$\mathbb{C}$	$\mathbb{R}$	$u(p,q)$	$p+q$	$K^{\dagger} = K$	$\operatorname{sgn} K = (p,q)$
x^*	1	$\mathbb{H}$	$\mathbb{R}$	$\mathbb{R}$	$sp(p,q)$	$p+q$	$K^{\dagger} = K$	$\operatorname{sgn} K = (p,q)$
x^*	−1	$\mathbb{H}$	$\mathbb{R}$	$\mathbb{R}$	$o^*(2n)$	n	$K^{\dagger} = -K$	

$$\alpha^* \leftrightarrow \begin{pmatrix} a & b \\ -b & a \end{pmatrix}^T = \begin{pmatrix} a & -b \\ b & a \end{pmatrix}, \tag{2.17}$$

$$q^* \leftrightarrow \begin{pmatrix} \beta & \gamma \\ -\gamma^* & \beta^* \end{pmatrix}^* = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} \beta & \gamma \\ -\gamma^* & \beta^* \end{pmatrix}^T \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} \beta^* & -\gamma \\ \gamma^* & \beta \end{pmatrix}. \tag{2.18}$$

Here x^* denotes complex (quaternionic) conjugation if $x \in \mathbb{C}$ ($x \in \mathbb{H}$) and Hermitian conjugation if x is an $n \times n$ matrix, $n > 1$.

7. It is useful to distinguish two types of elements X of $gl(N,D)$: decomposable and indecomposable ones. A matrix X is *decomposable* if there exist two invariant nontrivial subspaces V_1 and V_2 of the space V on which X acts such that

$$D^{N\times 1} = V = V_1 \oplus V_2, \quad XV_1 \subseteq V_1, \quad XV_2 \subseteq V_2. \tag{2.19}$$

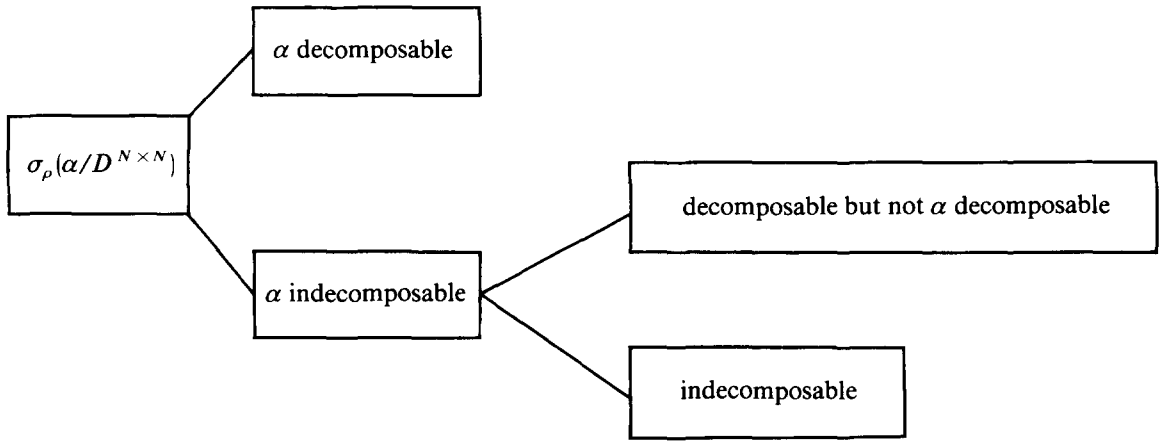
In case $D = \mathbb{H}$ the "linear" space V on which $gl(N,\mathbb{H})$ acts is the column space $\mathbb{H}^{N\times 1}$ of all N-columns over $\mathbb{H}$ with componentwise addition and multiplication with elements of $\mathbb{H}$ on the *right*. A linear subspace then is any submodule of V that is invariant under right multiplication by elements of $\mathbb{H}$.

A matrix which is not decomposable is said to be *indecomposable*. A decomposable X is GL(N,D)-conjugate to a block diagonal sum of indecomposables X_i,

$$X = \bigoplus_{i=1}^{s} X_i, \quad X_i \in gl(N_i,D), \quad \sum_{i=1}^{s} N_i = N. \tag{2.20}$$

Two elements X and X' of $gl(N,D)$ which differ only by permutation of the blocks X_i are GL(N,D)-conjugate. They transform into each other by means of a suitable permutation matrix. For the purpose of our classification X and X' should not be distinguished. By the general decomposition theorem (McLagan–Wedderburn–Remak–Krull–Schmidt–Fitting[10]) the block diagonal components are unique up to conjugacy.

8. The classification of elements into decomposable and indecomposable ones has to be refined for $X \in \sigma_\rho(\alpha/D^{N\times N})$. Indeed, a decomposition may or may not be compatible with the involution α. Consequently, one has the following hierarchy of elements:



More precisely $X \in \sigma_\rho(\alpha/D^{N\times N})$ is α-decomposable if in addition to the properties (2.19) one has also the α orthogonality:

$$V_1 \perp V_2, \text{ i.e., } \sum_{i,j=1}^{N} \alpha_0(f_i) K_{ij} f'_j = 0 \tag{2.21}$$

for any $f \in V_1, f' \in V_2$. Here α_0 and the matrix K define the involution α [cf. (2.6) and Table I].

9. The minimal polynomial $m(\lambda)$ over F of a matrix $X \in D^{N\times N}$ is the monic polynomial (i.e., its highest degree coefficient equals 1) of the lowest degree over F such that

$$m(X) = 0. \tag{2.22}$$

The minimal polynomial always divides the characteristic polynomial $\chi(\lambda)$,

$$\chi(\lambda) = \det(X - \lambda I), \tag{2.23}$$

where in the case $D = \mathbb{H}$ the matrix X is of degree $2N$ over $\mathbb{C}$. The characteristic polynomial $\chi(\lambda)$ shares all its irreducible divisors with the minimal polynomial $m(\lambda)$. The minimal polynomial for an indecomposable matrix is a power of a polynomial $m_0(\lambda)$, which is irreducible over F and called the irreducible polynomial of X.

For $D = \mathbb{R}$ or $\mathbb{H}$ the irreducible polynomials over F are of degree 1 or 2, for $D = \mathbb{C}$ they are of degree 1.

As examples, let us consider several matrices which will reappear as ingredients of normal forms. Let $a, b, \in \mathbb{R}$; then one has the following table:

X	$\chi(\lambda)$	$m(\lambda)$	$m_0(\lambda)$
$\begin{pmatrix} a & & & \\ 1 & a & & \\ & 1 & a & \\ & & 1 & a \end{pmatrix}$	$(\lambda-a)^4$	$\chi(\lambda)$	$\lambda-a$
$\begin{pmatrix} a & & & \\ 1 & a & & \\ & 1 & a & \\ & & 0 & a \end{pmatrix}$	$(\lambda-a)^4$	$(\lambda-a)^3$	$\lambda-a$
$\begin{pmatrix} a & & & \\ 1 & a & & \\ & 0 & a & \\ & & 0 & a \end{pmatrix}$	$(\lambda-a)^4$	$(\lambda-a)^2$	$\lambda-a$
$\begin{pmatrix} a & & & \\ 1 & a & & \\ & 0 & a & \\ & & 1 & a \end{pmatrix}$	$(\lambda-a)^4$	$(\lambda-a)^2$	$\lambda-a$
$\begin{pmatrix} a & b & & \\ -b & a & & \\ & & a & b \\ & & -b & a \end{pmatrix}$	$[(\lambda-a)^2+b^2]^2$	$(\lambda-a)^2+b^2$	$m(\lambda)$
$\begin{pmatrix} a & b & & \\ -b & a & & \\ 1 & 0 & a & b \\ 0 & 1 & -b & a \end{pmatrix}$	$[(\lambda-a)^2+b^2]^2$	$\chi(\lambda)$	$(\lambda-a)^2+b^2$

Similarly let $\alpha, \beta, \in \mathbb{C}$, then one has for the generic case ($\alpha \neq \alpha^*, \beta \neq 0$)

X	$\chi(\lambda)$	$m(\lambda)$	$m_0(\lambda)$
$\begin{pmatrix} \alpha & \beta \\ -\beta^* & \alpha^* \end{pmatrix}$	$(\alpha-\lambda)(\alpha^*-\lambda)+\lvert\beta\rvert^2$	$\chi(\lambda)$	$\chi(\lambda)$
$\begin{pmatrix} \alpha & \beta & & \\ -\beta^* & \alpha^* & & \\ 1 & 0 & \alpha & \beta \\ 0 & 1 & -\beta^* & \alpha^* \end{pmatrix}$	$[(\alpha-\lambda)(\alpha^*-\lambda)+\lvert\beta\rvert^2]^2$	$\chi(\lambda)$	$(\alpha-\lambda)(\alpha^*-\lambda)+\lvert\beta\rvert^2$
$\begin{pmatrix} \alpha & \beta & & \\ -\beta^* & \alpha^* & & \\ & & \alpha & \beta \\ & & -\beta^* & \alpha^* \end{pmatrix}$	$[(\alpha-\lambda)(\alpha^*-\lambda)+\lvert\beta\rvert^2]^2$	$(\alpha-\lambda)(\alpha^*-\lambda)+\lvert\beta\rvert^2$	$m(\lambda)$

10. We make use of the matrix operations

$$A \oplus B = \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix},$$
$$A \otimes B = A \otimes (b_{rs}) = (Ab_{rs}), \tag{2.24}$$

where $B = (b_{rs})$, and the $n \times n$ matrices

$$I_n = \begin{pmatrix} 1 & & \\ & \ddots & \\ & & 1 \end{pmatrix}, \quad H_n = \begin{pmatrix} & & 1 \\ & \iddots & \\ 1 & & \end{pmatrix},$$
$$F_n = \begin{pmatrix} & & & 1 \\ & & -1 & \\ & \iddots & & \\ (-1)^{n-1} & & & \end{pmatrix}, \tag{2.25}$$
$$J_n(x) = \begin{pmatrix} x & & & \\ 1 & & & \\ & \ddots & \ddots & \\ & & 1 & x \end{pmatrix},$$

where we allow x to be also a matrix. Then, for instance, one has

$$J_2\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} = \begin{pmatrix} \alpha & \beta & 0 & 0 \\ \gamma & \delta & 0 & 0 \\ 1 & 0 & \alpha & \beta \\ 0 & 1 & \gamma & \delta \end{pmatrix}. \tag{2.26}$$

11. The most often used involution α of complex matrix calculus is the involution which sends any matrix of degree N over $\mathbb{C}$ on its transpose complex conjugate. The corresponding classifying group is the group $U(N)$ of unitary matrices of degree N over $\mathbb{C}$. Symmetry is Hermitian symmetry α_0. Antisymmetry of X is equivalent to Hermitian symmetry of iX. It is well known that the principal axis theorem asserting the diagonalizability of such matrices under conjugation by unitary matrices extends precisely to the *normal matrices* which are characterized by the condition

$$\alpha(X)X = X\alpha(X) \tag{2.27}$$

which is easy to test.

Generalizing the concept of normal matrices of complex matrix calculus we define a matrix X of $D^{N\times N}$ to be α *normal* for an arbitrary involution α of $D^{N\times N}$ if it satisfies (2.27). Equivalently, the α-invariant F-algebra generated by I_N, X, $\alpha(X)$ is commutative.

Each matrix X of degree N over D is uniquely decomposed into the sum

$$X = X^+ + X^- \tag{2.28a}$$

of the α-symmetric matrix

$$X^+ = (X + \alpha(X))/2, \tag{2.28b}$$

and the α-skew-symmetric matrix

$$X^- = (X - \alpha(X))/2, \tag{2.28c}$$

and the α normality of X is tantamount to the commutativity rule

$$X^+X^- = X^-X^+. \tag{2.28d}$$

Both α-summetric and α-skew-symmetric elements are α normal. But, for example, the elements of $O(\alpha/D^{N\times N})$ are α normal, though not necessarily α-symmetric or α-skew symmetric.

Our results will be extended in Sec. 4 to a classification of the α-normal elements under $O(\alpha/D^{N\times N})$ conjugacy.

3. CLASSIFICATION THEOREMS

In this section we present four theorems. The first two deal with classical Lie algebras of linear type $gl(N,D)$, $D = \mathbb{R}$, $\mathbb{C}$, and $\mathbb{H}$. Their content is well known and they are listed here just for completeness. No proofs are given. The last two theorems apply to quadratic classical Lie algebras $\sigma_-(\alpha/D^{N\times N})$ as well as to $\sigma_+(\alpha/D^{N\times N})$. Their content is known only for $\sigma_-(\alpha/D^{N\times N})$ (cf. Ref. 2) and $\sigma_+(\alpha/\mathbb{C}^{N\times N})$ (cf. Ref. 1). For $\sigma_+(\alpha/D^{N\times N})$, where $D = \mathbb{R}$ or $\mathbb{H}$, they appear here for the first time.

Theorem 1: The normal form of an indecomposable element X of a linear classical Lie algebra $gl(N,D)$ can be chosen to be

$$X \equiv X_m = J_\mu(X_{m_0}) = X_{m_0} \otimes I_\mu + I_g \otimes J_\mu(0), \tag{3.1}$$

where X_{m_0} is an irreducible matrix of degree g over D with minimal polynomial $m_0(t)$ over F, $N = g\mu$, I_μ and J_μ are given in (2.25).

For $D = \mathbb{C}$ we have $g = 1$ and $X_{m_0} = \beta \in \mathbb{C}$; for $D = \mathbb{R}$ we have either $g = 1$ and $X_{m_0} = a \in \mathbb{R}$, or $g = 2$ and $X_{m_0} = \left(\begin{smallmatrix} a & b \\ -b & a \end{smallmatrix}\right)$, $a \in \mathbb{R}$, $b > 0$; for $D = \mathbb{H}$ we have $g = 1$ and $X_{m_0} = \left(\begin{smallmatrix} \zeta & 0 \\ 0 & \zeta^* \end{smallmatrix}\right)$, $\zeta \in \mathbb{C}$, $\mathrm{Im}\,\zeta \geqslant 0$.

Comment: For $D = \mathbb{H}$ we interpret X as an $N \times N$ matrix of 2×2 complex matrices, cf. (2.16).

The indecomposable normal forms of elements of $gl(N,D)$ are summarized in Table II.

Theorem 2: The normal form of a decomposable element of a classical linear Lie algebra $gl(N,D)$ can be chosen to be the block diagonal sum of indecomposable matrices X_i of the form (3.1);

$$X = \bigoplus_{i=1}^{s} X_i, \quad s \geqslant 2, \tag{3.2}$$

where

$$X_i \in D^{N_i \times N_i}, \quad N_i \in \mathbb{Z}^{>0}, \quad \sum_{i=1}^{s} N_i = N.$$

The sum (3.2) is defined uniquely up to permutations of the blocks X_i.

TABLE II. Indecomposable normal forms of elements of the linear classical Lie algebras. For notations see subsec. 2.10.

Algebra	Indecomposable normal form	
$gl(n,\mathbb{C})$	$J_n(\alpha)$	$\alpha \in \mathbb{C}$
$gl(2n+1,\mathbb{R})$	$J_{2n+1}(a)$	$a \in \mathbb{R}$
	$J_{2n}(a)$	$a \in \mathbb{R}$
$gl(2n,\mathbb{R})$	$J_n\left(\begin{smallmatrix} a & b \\ -b & a \end{smallmatrix}\right)$	$a,b \in \mathbb{R}, b > 0$
$gl(n,\mathbb{H})$	$J_n\left(\begin{smallmatrix} \zeta & 0 \\ 0 & \zeta^* \end{smallmatrix}\right)$	$\zeta = a + ib$, $a,b \in \mathbb{R}, b \geqslant 0$

TABLE III. List of α-indecomposable normal forms of elements of classical Lie and Jordan algebras. The parameter $\kappa = \pm 1$ indicates the existence of two nonequivalent normal forms (X,K) and $(X,-K)$.

D	†	ϵ	ρ	Normal form X	Restrictions	K	Signature $p-q$	Classification group	No.
$\mathbb{C}$	T	+	+	$J_n(\zeta)$		H_n		$O(n,\mathbb{C})$	1
			−	$J_{2n+1}(0)$		F_{2n+1}		$O(2n+1,\mathbb{C})$	2
				$J_n(\zeta)\oplus(-J_n(\zeta))^T$	n even if $\zeta=0$	$I_n\otimes\begin{pmatrix}0&1\\1&0\end{pmatrix}$		$O(2n,\mathbb{C})$	3
		−	+	$J_n(\zeta)\oplus J_n(\zeta)^T$		$I_n\otimes\begin{pmatrix}0&1\\-1&0\end{pmatrix}$		$Sp(2n,\mathbb{C})$	4
			−	$J_{2n}(0)$		F_{2n}		$Sp(2n,\mathbb{C})$	5
				$J_\eta(\zeta)\oplus(-J_n(\zeta))^T$	n odd if $\zeta=0$	$I_n\otimes\begin{pmatrix}0&1\\-1&0\end{pmatrix}$		$Sp(2n,\mathbb{C})$	6
	*	+	+	$J_n(a)$		κH_n	$\frac{\kappa}{2}(1-(-1)^n)$	$U(p,q),\ p+q=n$	7
				$J_n(\zeta)\oplus J_n(\zeta)^*$	$\mathrm{Im}\,\zeta>0$	$I_n\otimes\begin{pmatrix}0&1\\1&0\end{pmatrix}$	0	$U(n,n),\ p=q=n$	8
			−	$J_n(ib)$		$\kappa i^{n-1}F_n$	$\frac{\kappa}{2}(1-(-1)^n)$	$U(p,q),\ p+q=n$	9
				$J_n(\zeta)\oplus(-J_n(\zeta))^*$	$\mathrm{Re}\,\zeta>0$	$I_n\otimes\begin{pmatrix}0&1\\1&0\end{pmatrix}$	0	$U(n,n),\ p=q=n$	10
R	T	+	+	$J_n(a)$		κH_n	$\frac{\kappa}{2}(1-(-1)^n)$	$O(p,q),\ p+q=n$	11
				$J_n\begin{pmatrix}a&b\\-b&a\end{pmatrix}$	$b>0$	H_{2n}	0	$O(n,n),\ p=q=n$	12
			−	$J_{2n+1}(0)$		$\kappa(-1)^nF_{2n+1}$	κ	$O(p,q),\ p+q=2n+1$	13
				$J_n(a)\oplus(-J_n(a))^T$	$a\geqslant 0$ n even if $a=0$	$I_n\otimes\begin{pmatrix}0&1\\1&0\end{pmatrix}$	0	$O(n,n),\ p=q=n$	14
			−	$J_n\begin{pmatrix}0&b\\-b&0\end{pmatrix}$	$b>0$	$\kappa\begin{pmatrix}0&1\\-1&0\end{pmatrix}^{n-1}\otimes F_n$	$\kappa(1-(-1)^n)$	$O(p,q),\ p+q=2n$	15
				$J_n\begin{pmatrix}a&b\\-b&a\end{pmatrix}\oplus\left(-J_n\begin{pmatrix}a&b\\-b&a\end{pmatrix}\right)^T$	$a>0,b>0$	$I_{2n}\otimes\begin{pmatrix}0&1\\1&0\end{pmatrix}$	0	$O(2n,2n),\ p=q=2n$	16
		−	+	$J_n(a)\oplus J_n(a)^T$		$I_n\otimes\begin{pmatrix}0&1\\-1&0\end{pmatrix}$		$Sp(2n,\mathbb{R})$	17
				$J_n\begin{pmatrix}a&b\\-b&a\end{pmatrix}\oplus\left(J_n\begin{pmatrix}a&b\\-b&a\end{pmatrix}\right)^T$	$b>0$	$I_{2n}\otimes\begin{pmatrix}0&1\\-1&0\end{pmatrix}$		$Sp(4n,\mathbb{R})$	18
			−	$J_{2n}(0)$		κF_{2n}		$Sp(2n,\mathbb{R})$	19
				$J_n(a)\oplus(-J_n(a))^T$	$a\geqslant 0$ n odd if $a=0$	$I_n\otimes\begin{pmatrix}0&1\\-1&0\end{pmatrix}$		$Sp(2n,\mathbb{R})$	20
				$J_n\begin{pmatrix}0&b\\-b&0\end{pmatrix}$	$b>0$	$\kappa\begin{pmatrix}0&1\\-1&0\end{pmatrix}^n\otimes F_n$		$Sp(2n,\mathbb{R})$	21
				$J_n\begin{pmatrix}a&b\\-b&a\end{pmatrix}\oplus\left(-J_n\begin{pmatrix}a&b\\-b&a\end{pmatrix}\right)^T$	$a>0,b>0$	$I_{2n}\otimes\begin{pmatrix}0&1\\-1&0\end{pmatrix}$		$Sp(4n,\mathbb{R})$	22

TABLE III. (continued).

D	†	ϵ	ρ	Normal form X	Restrictions	K	Signature $p-q$	Classification group	No.
H	*	+	+	$J_n\begin{pmatrix} a & 0 \\ 0 & a \end{pmatrix}$		κH_{2n}	$\frac{\kappa}{2}(1-(-1)^n)$	Sp(p,q), $p+q=n$	23
				$J_n\begin{pmatrix} \zeta & 0 \\ 0 & \zeta^* \end{pmatrix} \oplus \left(J_n\begin{pmatrix} \zeta & 0 \\ 0 & \zeta^* \end{pmatrix}\right)^*$	$\operatorname{Im}\zeta > 0$	$I_{2n} \otimes \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$	0	Sp(n,n)	24
			−	$J_n\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$		$\kappa^n \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}^{n-1} \otimes F_n$	$\frac{\kappa}{2}(1-(-1)^n)$	Sp(p,q), $p+q=n$	25
				$J_n\begin{pmatrix} 0 & b \\ -b & 0 \end{pmatrix}$	$b>0$	$\kappa \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}^{n-1} \otimes F_n$	$\frac{\kappa}{2}(1-(-1)^n)$	Sp(p,q), $p+q=n$	26
				$J_n\begin{pmatrix} \zeta & 0 \\ 0 & \zeta^* \end{pmatrix} \oplus \left(-J_n\begin{pmatrix} \zeta & 0 \\ 0 & \zeta^* \end{pmatrix}\right)^*$	$\operatorname{Re}\zeta > 0$ $\operatorname{Im}\zeta \geqslant 0$	$I_{2n} \otimes \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$	0	Sp(n,n)	27
		−	+	$J_n\begin{pmatrix} \zeta & 0 \\ 0 & \zeta^* \end{pmatrix}$	$\operatorname{Im}\zeta \geqslant 0$	$\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \otimes H_n$		O*(2n)	28
			−	$J_n\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$		$\kappa^{n-1} \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}^n \otimes F_n$		O*(2n)	29
				$J_n\begin{pmatrix} 0 & b \\ -b & 0 \end{pmatrix}$	$b>0$	$\kappa \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}^n \otimes F_n$		O*(2n)	30
				$J_n\begin{pmatrix} \zeta & 0 \\ 0 & \zeta^* \end{pmatrix} \oplus \left(-J_n\begin{pmatrix} \zeta & 0 \\ 0 & \zeta^* \end{pmatrix}\right)^*$	$\operatorname{Re}\zeta > 0$ $\operatorname{Im}\zeta \geqslant 0$	$I_{2n} \otimes \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$		O*(4n)	31

Let us now turn to elements of $\sigma_\rho(\alpha/D^{N\times N})$. In order that $X \in \sigma_\rho(\alpha/D^{N\times N})$, it is necessary and sufficient to have

$$\alpha(X) = \rho X = KX^\dagger K^{-1} \tag{3.3}$$

as follows from (2.6) and (2.10). Equivalently,

$$KX^\dagger - \rho XK = 0 \tag{3.4}$$

or

$$(XK)^\dagger - \epsilon\rho XK = 0. \tag{3.5}$$

In order to fix a particular $\sigma_\rho(\alpha/D^{N\times N})$, the values of ϵ and ρ have to be fixed (cf. Table I). If in addition a particular form of the matrix K is chosen, the requirement (3.4) is invariant with respect to the transformation

$$X \to YXY^{-1}, \quad Y \in O(\alpha/D^{N\times N}). \tag{3.6}$$

It is, however, desirable for the purpose of the universal classification to let K vary within the range permitted by the equivalence relation (2.9) of the involutions. The full range should not be used because we need relations

$$(X \oplus Y,\ K \oplus L) \sim (X' \oplus Y', K' \oplus L')$$

as a consequence of

$$(X,K) \sim (X',K') \text{ and } (Y,L) \sim (Y',L').$$

The problem is then to classify simultaneously the pairs of matrices (X,K) under the action

$$S:(X,K) \to (SXS^{-1}, SKS^\dagger) \tag{3.7}$$

of $S \in \mathrm{GL}(N,D)$. It is straightforward to verify that (3.4) is invariant with respect to (3.7). We say that two pairs (X,K) and (X',K') are equivalent and write $(X,K) \sim (X',K')$ iff

$$X' = SXS^{-1}, \quad K' = SKS^\dagger \tag{3.8}$$

for some $S \in GL(N,D)$.

In Theorem 3 we treat simultaneously α-indecomposable elements that are decomposable and those that are indecomposable. A list of α-indecomposable normal forms of pairs (X,K) is given in Table III. Theorem 4 deals with α-decomposable elements.

Theorem 3: A. Let $X \in D^{N\times N}$ be an indecomposable matrix such that $\alpha(X) = \rho X, \rho = \pm 1$. The normal form of the pair (X,K) can be chosen as follows. The matrix

$$X = J_\mu(X_{m_0}) = X_{m_0} \otimes I_\mu + I_g \otimes J_\mu(0) \tag{3.9}$$

is the same as in Theorem 1 and

$$K = \begin{cases} \widetilde{K} \otimes H_\mu & \text{for } \rho = +1, \\ \widetilde{K} \otimes F_\mu & \text{for } \rho = -1, \end{cases} \tag{3.10}$$

where $\widetilde{K}$ is a nonsingular matrix of degree g over D satisfying the symmetry condition

$$\widetilde{K} = \epsilon\rho^{\mu-1}\widetilde{K}^\dagger \tag{3.11}$$

with $\epsilon = -1$ for $sp(2n,\mathbb{R})$, $sp(2n,\mathbb{C})$, $o^*(2n)$, and $\epsilon = +1$ for the remaining algebras. In addition X_{m_0} satisfies

$$\widetilde{K}X_{m_0}^\dagger - \rho X_{m_0}\widetilde{K} = 0. \tag{3.12}$$

B1. Let X be an α-indecomposable decomposable matrix such that $\alpha(X) = \rho X, \rho = \pm 1$. The normal form of the pair (X,K) can be chosen to be

$$X = X_m \oplus (\rho X_m^\dagger), \quad K = \begin{pmatrix} 0 & I_{N/2} \\ \epsilon I_{N/2} & 0 \end{pmatrix}, \tag{3.13}$$

where X_m is as in Theorem 1, i.e., $m = m_0^\mu$ and $X_m = J_\mu(X_{m_0}), N = 2g\mu$.

B2. Conversely, a pair of matrices (X,K) as given by (3.13) is α-indecomposable precisely if either (i) X_m and $\rho X_m^\dagger$ are not conjugate under $\mathrm{GL}(N/2,D)$, i.e., X_m and $\rho X_m^\dagger$ have distinct minimal polynomials over F, or, equivalently, X_m and $\rho X_m^\dagger$ have distinct eigenvalues, or (ii) X_m and $\rho X_m^\dagger$ are conjugate under $\mathrm{GL}(N/2,D)$ and one of the cases 3 (with

$\zeta = 0$), 4, 6 (with $\zeta = 0$), 14 (with $a = 0$), 17, 18, 20 (with $a = 0$), and 24 of Table III occur.

The result of the classification of α-indecomposable pairs may be summarized as follows: Let $D, \rho, \epsilon, \dagger$ be fixed (one choice from Table III). The pair (X,K) of Table III corresponding to that choice is α-indecomposable and is nonequivalent to any other choice. Every α-indecomposable pair is equivalent precisely to one pair in Table III.

Proof of Theorem 3: First the reader must work his way through Table III in order to verify that each entry (X,K) satisfies the required conditions. This work consists (1) of verifying that each case is the appropriate particular case of the generic form of the α-indecomposable pair (X,K) of Theorem 3; (2) one must verify that the generic form really meets the requirements stated in Theorem 3.

(1) Thus, for example, for line 12 of Table III one must verify that $J_n\left(\begin{smallmatrix} a & b \\ -b & a \end{smallmatrix}\right)$, with a, b real, $b > 0$, and $K = H_{2n}$, actually is an indecomposable element of $o(p,p)$, where $p = n$, $K = K^\dagger$ as claimed in the ϵ column, $KX^\dagger K^{-1} = X$ as claimed in the ρ column, and finally that all matrices $J_n\left(\begin{smallmatrix} a & b \\ -b & a \end{smallmatrix}\right)$, $a, b \in \mathbb{R}$, $b \neq 0$, are equivalent to the normal form of Table III. Because of the invariance of the minimal polynomial, it follows indeed that the pair $\left(J_n\left(\begin{smallmatrix} a & b \\ -b & a \end{smallmatrix}\right), H_{2n}\right)$ is equivalent to the pair $\left(J_n\left(\begin{smallmatrix} a & \lvert b\rvert \\ -\lvert b\rvert & a \end{smallmatrix}\right), H_{2n}\right)$.

(2) A. It follows from Theorem 1 that X defined by (3.9) is indecomposable. A simple computation shows that K defined by (3.10) meets the conditions $K^\dagger = \epsilon K$, $KX^\dagger K^{-1} = \rho X$, provided (3.11) and (3.12) are satisfied.

(3)B. It follows from Theorems 1 and 2, that X defined by (3.13) is fully decomposed so that the pair (X,K) of (3.13) is α-indecomposable.

Conversely, let us assume that (X,K_0) is an α-indecomposable pair. We have to show that it is equivalent to the normal form provided by Theorem 3.

A. If X is indecomposable then according to Theorem 1, after a suitable equivalence transformation, it is as in (3.9). Regarding K_0 we merely know that $K_0^\dagger = \epsilon K_0$, $K_0 X^\dagger K_0^{-1} = \rho X$ with stated sign of ρ and ϵ. Writing $K_0 = (K_{ik})$, $i,k = 1,\ldots,\mu$, as a matrix of matrices of degree g of D, one verifies that $K_0 X^\dagger = \rho X K_0$ implies $K_{ik} = 0$ if $i + k \leqslant \mu$. Since K of (3.10) also satisfies (2.12) and (3.4), the quotient $U_0 = KK_0^{-1}$ commutes with X. Since X is in the form (3.9), U_0 must be of the form

$$U_0 = KK_0^{-1} = U_1(I_N + M) = (I_N + M)\,U_1, \qquad (3.14)$$

where the matrix U_1 is a polynomial in the block-diagonal matrix $X_{m_0} \otimes I_\mu$ and M is a nilpotent matrix. The factorization (3.14) of U_0 is unique. A simple computation yields

$$U_0^\dagger = K^{-1} U_0 K. \qquad (3.15)$$

Using the uniqueness of (3.14) we find

$$U_1^\dagger = K^{-1} U_1 K, \quad M^\dagger = K^{-1} M K. \qquad (3.16)$$

Computing $U = (I_N + M)^{1/2}$ by means of the binomial series we obtain

$$U = (I_N + M)^{1/2} = I_N + MP(M), \qquad (3.17)$$

where P is a polynomial with rational coefficients. We have thus constructed a matrix U that is invertible, commutes with X, and satisfies

$$U^2 = I_N + M, \quad KU^\dagger K^{-1} = U. \qquad (3.18)$$

We now introduce a new matrix

$$\begin{aligned} K_1 &= UK_0U^\dagger = UU_0^{-1} KU^\dagger \\ &= UU_0^{-1}(KU^\dagger K^{-1})\,K = UU_0^{-1}UK \\ &= U^2U_0^{-1}K = U_1^{-1}K, \end{aligned}$$

so that K_1 is a side-diagonal matrix of matrices. We have thus shown the equivalence

$$(X,K_0) \sim (X,K_1) = (UXU^{-1}, UK_0U^\dagger). \qquad (3.19)$$

and it is easy to verify that the equation $(XK_1)^\dagger = \epsilon\rho XK_1$ implies that K_1 coincides with K in (3.10) up to a nonzero scalar factor.

B. The α-indecomposability of X is equivalent to the statement that the centralizer C of X in $D^{N\times N}$ contains no α-invariant idempotent other than I_N. Because of the α-symmetry or α-skew-symmetry of X it follows that C is α-invariant. Also the radical $J(C)$ of C, i.e., its maximal nilpotent ideal is α invariant.

Let us show that any α invariant idempotent coset of C modulo $J(C)$ contains an α invariant idempotent e' of C. In other words if $e \in C$, $e \notin J(C)$, $e^2 = e + u$, $\alpha(e) = e + v$, u,v in $J(C)$, then there is also an α-invariant idempotent e' of C congruent to e modulo $J(C)$. We observe that $\alpha(e + \alpha(e))/2 = \alpha(e + \alpha(e))/2 = e (\mathrm{mod}\, J(C))$.

Now let e' be an element of $e/J(C)$ for which $\alpha(e') = e'$ and the nilpotency exponent η of $u' = e'^2 - e'$ is minimal. If $\eta > 1$, then $u^\eta = 0, u^{\eta-1} \neq 0$, and now the element $e_1 = 3e'^2 - 2e'^3 = e' + u' - 2u'e' = \alpha(e_1)$ belongs to $e/J(C)$ and $e_1^2 - e_1 = u'^2(4u' - 3)$ has nilpotency exponent less than η. Hence $\eta = 1$, e' is an α invariant idempotent in $C/J(C)$. Note that the same argument also shows the existence of an α-invariant idempotent in any α-invariant non-nilpotent subalgebra of C. In other words: X is α-indecomposable precisely if the factor algebra $C/J(C)$ contains no α-invariant idempotent other than $I_N/J(C)$. Now let X be α-indecomposable. If $C/J(C)$ contains a primitive central idempotent $e/J(C) \neq I_N/J(C)$ then $\alpha(e)/J(C)$ is a primitive central idempotent $\neq e/J(C)$, $(e + \alpha(e))/J(C)$ is an α-invariant idempotent. Hence $e + \alpha(e) = I_N$ [mod $J(C)$]. In this case $C/J(C)$ is the algebraic sum of two division algebras that are mapped by α one on another.

Vice versa, if $C/J(C)$ is the algebraic sum of two α-conjugate division algebras then X is α-indecomposable, but decomposable.

On the other hand, if $C/J(C)$ contains no primitive central idempotent $\neq I_N/J(C)$ then $C/J(C)$ is isomorphic to a ring of matrices of finite degree over $\mathbb{R}$, $\mathbb{C}$, and $\mathbb{H}$. If the degree of this ring of matrices is larger than 2 then there is always an α-invariant idempotent e not equal to identity. For example, the subalgebra generated by a primitive idempotent and its α image contains an e. However, if $d = 2$ then it may or may not happen that there is an α-invariant idempotent not equal to identity. Finally, if the degree is 1 then X is indecomposable.

From this analysis it follows that the normal form of

Theorem 3B (in accordance with Theorem 1) must have precisely two Jordan components which are linked up by the involution α in such a way that the matrix X maps the first component space of the decomposition onto the second so that with appropriate choice of basis we have

$$K=\begin{pmatrix}0 & I_{N/2}\\ \epsilon I_{N/2} & 0\end{pmatrix}, \tag{3.20}$$

and correspondingly,

$$X=X_m\oplus(\rho X_m^{\dagger}), \tag{3.21}$$

with X_m as in Theorem 1.

The pairs (X,K) obtained in this way are always α-indecomposable if the two matrices $X_m,\rho X_m^{\dagger}$ are not conjugate under $\mathrm{GL}(N/2,D)$. This is because there is essentially only one decomposition of X, the factor algebra $C/J(C)$ contains only central idempotents and it is the algebraic sum of two α-conjugate division algebras.

However, if X_m and $\rho X_m^{\dagger}$ are $\mathrm{GL}(N/2,D)$ conjugate then the factor algebra $C/J(C)$ is a ring of matrices of degree 2 over a division algebra. Here it may or may not happen that there is no α-invariant idempotent.

In each of the cases to be studied we find an involution β of a ring of matrices $\Psi^{2\times 2}$, where $\Psi=\mathbb{R},\mathbb{C}$, or $\mathbb{H}$, and it must be decided whether the involution β leaves invariant an indempotent $\neq I_2$.

Using the known normal forms of the involution β the answer is as follows:

Division algebra	Action of β on Ψ	2×2 matrix representing β	idempotent fixed by β
$\mathbb{R}$	$1_{\mathbb{R}}$	$\begin{pmatrix}\pm 1 & \\ & \pm 1\end{pmatrix}$	$\begin{pmatrix}1 & \\ & 0\end{pmatrix}$
	$1_{\mathbb{R}}$	$\begin{pmatrix} & 1\\ -1 & \end{pmatrix}$	none
$\mathbb{C}$	$1_{\mathbb{C}}$	$\begin{pmatrix}1 & \\ & 1\end{pmatrix}$	$\begin{pmatrix}1 & \\ & 0\end{pmatrix}$
	$1_{\mathbb{C}}$	$\begin{pmatrix} & 1\\ -1 & \end{pmatrix}$	none
	complex conjugation	$\begin{pmatrix}\pm 1 & \\ & \pm 1\end{pmatrix}$	$\begin{pmatrix}1 & \\ & 0\end{pmatrix}$
$\mathbb{H}$	quaternionic conjugation	$\begin{pmatrix}\pm 1 & \\ & \pm 1\end{pmatrix}$	$\begin{pmatrix}1 & \\ & 0\end{pmatrix}$
	quaternionic conjugation	$\begin{pmatrix}\pm i & \\ & \pm i\end{pmatrix}$	$\begin{pmatrix}1 & \\ & 0\end{pmatrix}$

Thus the only case in doubt is the case when $\Psi=\mathbb{R}$ or $\mathbb{C}$ and β is symplectic, $\beta^T=-\beta$.

Orthogonal indecomposability is tantamount to the condition that β be symplectic. Thus verification has to be made in each case, its results are summarized in part 2 of Theorem 3B.

Suppose

$$X=X_m\oplus\rho X_m^{\dagger},\quad K=\begin{pmatrix}0 & I_{N/2}\\ \epsilon I_{N/2} & 0\end{pmatrix}$$

as stated in part 1 of Theorem 3B, and assume that

$$\rho X^{\dagger}=AX_mA^{-1}$$

for some nonsingular A of degree $N/2$ over D. In fact the matrix A can be assumed to be a side diagonal matrix of matrices of the form

$$A=\begin{pmatrix} & & A_1\\ & \iddots & \\ A_1 & & \end{pmatrix}=(\delta_{i,\mu+1-i}A_1),\quad A_1\in\mathrm{GL}(g,D),$$

such that

$$A_1X_{m_0}A_1^{-1}=\rho X_{m_0}^{\dagger}.$$

Now we want to test whether the centralizer C of X in $D^{N\times N}$ contains α-invariant idempotents. Since

$$X=B(X_m\oplus X_m)B^{-1},\quad B=I_{N/2}\otimes A$$

it follows that

$$C=BC_1^{2\times 2}B^{-1},$$

where C_1 is the centralizer of $X_m\in D^{N/2\times N/2}$ and

$$C_1^{2\times 2}=C_1\otimes_F F^{2\times 2}$$

is the centralizer of $X_m\oplus X_m$ in $D^{N\times N}$. Moreover,

$$C_1=C_{10}\otimes I_\mu+C_{10}\otimes J_\mu(0)+\cdots+C_{10}\otimes J_\mu(0)^{\mu-1},$$

where C_{10} is the centralizer of X_{m_0} in $D^{g\times g}$. This happens to be

$$C_{10}=\begin{cases}D & \text{if } g=1,\ X_{m_0}\text{ is linear},\\ FI_g\otimes FX_{m_0} & \text{if } g=2.\end{cases}$$

The radical of C is $BJ(C_1)^{2\times 2}B^{-1}$ where the radical of C_1 is

$$J(C_1)=C_{10}\otimes J_\mu(0)+\cdots+C_{10}\otimes J_\mu(0)^{\mu-1}$$

and the factor algebra $C/J(C)$ therefore is F isomorphic to $C_{10}^{2\times 2}$. It is represented by the subalgebra

$$(C_{10}^{2\times 2})=(C_{10})e_{11}+(C_{10})e_{12}+(C_{10})e_{21}+(C_{10})e_{22}$$

of C, where the "matrix units" e_{ik} are

$$e_{11}=I_{N/2}\oplus 0,\quad e_{12}=\begin{pmatrix}0 & A^{-1}\\ 0 & 0\end{pmatrix},$$

$$e_{21}=\begin{pmatrix}0 & 0\\ A & 0\end{pmatrix},\quad e_{22}=0\oplus I_{N/2},$$

subject to the standard multiplication rules

$$e_{ik}e_{rs}=\delta_{kr}e_{is}\quad(i,k,r,s=1,2).$$

The F-isomorphism θ of C_{10} into C is defined by setting

$$\theta(Y)=Y\otimes I_\mu\oplus A_1YA_1^{-1}\otimes I_\mu.$$

The centralizer C_{10} of the irreducible matrix X_{m_0} in $D^{g\times g}$ is a division algebra which is isomorphic either to $\mathbb{R}$, $\mathbb{C}$, or $\mathbb{H}$. If it is isomorphic to $\mathbb{H}$ then we know already that X is α-decomposable. Thus we may assume that m_0 is quadratic in case $D=\mathbb{H}$.

The involution β induced by α on $\theta(C_{10}^{2\times 2})$ is described by $\beta(\theta(Y))=K\theta(Y)^{\dagger}K^{-1}$ for Y in C_{10} and furthermore by

$$\beta(e_{11})=e_{22},\quad \beta(e_{22})=e_{11},$$
$$\beta(e_{12})=(\epsilon A^{\dagger-1}A\oplus 0)e_{12},$$
$$\beta(e_{21})=e_{21}(A^{-1}A^{\dagger}\oplus 0).$$

If the action of β on $\theta(C_{10})$ is not identity then we have seen already that X is α-decomposable. From now on let $K\theta(Y)^{\dagger}K^{-1}=\theta(Y)$ for $Y\in C_{10}$, or equivalently $A_1 Y A_1^{-1}=Y^{\dagger}$. For $Y=X_{m_0}$ this gives $X^{\dagger}_{m_0}=A_1 X_{-m_0} A_1^{-1}=\rho X^{\dagger}_{m_0}$. Thus if $\rho=-1$ then $X_{m_0}=(0)$, $g=1$ and $m_0(t)=t$. Here we have normalized X_{m_0} in such a way that in all cases

$$\beta(e_{12})=\beta_0 e_{12},\quad \beta(e_{21})=\beta_0 e_{21}$$

with $\beta_0=\pm 1$. If $\beta_0=1$, then β is not symplectic. If $\beta_0=-1$, then β is symplectic. ■

A classification of α-decomposable pairs (X,K) is the result of application of

Theorem 4: The normal form of the α-decomposable pair (X,K), where $X\in\sigma_p(\alpha/D^{N\times N})$, can be chosen as the block diagonal sum of α-indecomposable pairs (X_i,K_i), $1\leqslant i\leqslant s$, $s\in\mathbb{Z}^{>1}$, is given as

$$X=\bigoplus_{i=1}^{s} X_i,\quad K=\bigoplus_{i=1}^{s} K_i,\quad K_i\in D^{N_i\times N_i},$$
$$1\leqslant N_i<N,\quad \sum_{i=1}^{s} N_i=N, \tag{3.22}$$

where (X_i,K_i) have the form specified in Theorem 3 and all satisfy the coherence condition

$$\alpha(X_i)=K_i X_i^{\dagger} K_i^{-1}=\rho X_i,\quad K_i^{\dagger}=\epsilon K_i. \tag{3.23}$$

Proof of Theorem 4: The existence of the orthogonal decomposition as stated is trivial. Concerning its uniqueness, apart from the redundancy explained earlier, it follows from the following remarks. Any element X of $D^{N\times N}$ satisfying $\alpha(X)=\rho X$ generates an F-algebra $F[X]=\langle FI_N, X\rangle$ that is α invariant and contains the unital subalgebra $F[X]^{\alpha}=\{Y\mid Y\in F[X] \text{ and } \alpha(Y)=Y\}$ formed by the α-symmetric elements. The primitive idempotents of $F[X]^{\alpha}$ lead to a uniquely determined α-orthogonal decomposition of X which upon further refinement leads to the α-decomposition with only α-indecomposable components. Without loss of generality we may assume that $F[X]^{\alpha}$ contains only one primitive idempotent, namely I_N. Then all α-orthogonal components of X are nilpotent. The invariance of the decomposition occurring in Theorem 4 follows from the invariance of the decomposition occurring in Theorem 2.

4. EXAMPLES AND COMMENTS

Table III contains in a concise form all the information about α-indecomposable pairs (X,K), where X is an element of classical Lie or Jordan algebra with involution α. The two cases are distinguished by the entry in column ρ, $\rho=-1(+1)$ indicating Lie (Jordan) algebra. The type of a particular algebra is determined by the involution α, i.e., comparing the entries in the left three columns with the corresponding ones in Table I, or even more easily from the last column of Table III, where the corresponding Lie group is shown. Whenever one is dealing with an entry corresponding to algebras of the same type but different signatures [i.e., $u(p,q)$, $o(p,q)$, $sp(p,q)$] the signature can be read from the column $p-q$.

Consider a few examples. Let us pick up line 9 of Table III. From column D we see that we are dealing with an algebra of complex matrices, the operation $\dagger$ is Hermitian conjugation and the matrix K is symmetric: $K^{\dagger}=K$. Thus we have a Lie algebra $u(p,q)$. The normal form of $X\in u(p,q)$ is an irreducible Jordan block

$$J_n(ib)=\begin{pmatrix} ib & & & \\ 1 & \ddots & & \\ & \ddots & \ddots & \\ & & 1 & ib \end{pmatrix} \tag{4.1}$$

with purely imaginary eigenvalues ib. The entry in the K column $K=\kappa i^{n-1}F_n$ indicates that for each n there are two indecomposable pairs $(J_n(ib), i^{n-1}F_n)$ and $(J_n(ib), -i^{n-1}F_n)$ which are not equivalent, their signatures are opposite. The matrix $\pm i^{n-1}F_n$ varies with n as well as its signature. For the lowest values of n one has

$$n=1\quad K=\kappa,\quad p-q=\kappa,\quad p+q=1,$$
$$n=2\quad K=\kappa\begin{pmatrix} 0 & i \\ -i & 0 \end{pmatrix},\quad p=q=1, \tag{4.2}$$
$$n=3\quad K=\kappa\begin{pmatrix} 0 & 0 & -1 \\ 0 & 1 & 0 \\ -1 & 0 & 0 \end{pmatrix},$$
$$p-q=\kappa,\quad p+q=3.$$

As the second example consider the last line of Table III. From the first three columns and from the last line of Table I one deduces that the Lie algebra in this case is of the type $o^*(4n)$; moreover, the matrix X is

$$X=J_n\begin{pmatrix}\zeta & 0\\ 0 & \zeta^*\end{pmatrix}\oplus\left(-J_n\begin{pmatrix}\zeta & 0\\ 0 & \zeta^*\end{pmatrix}\right)^*=\left(\begin{array}{cccc|cccc} \zeta & 0 & & & & & & \\ 0 & \zeta^* & & & & & & \\ 1 & 0 & \zeta & 0 & & & & \\ 0 & 1 & 0 & \zeta^* & & & & \\ & & & & \ddots & & & \\ \hline & & & & -\zeta^* & 0 & -1 & 0 \\ & & & & 0 & -\zeta & 0 & -1 \\ & & & & & & -\zeta^* & 0 \\ & & & & & & 0 & -\zeta \end{array}\right), \tag{4.3}$$

where $\left(\begin{smallmatrix}\zeta & 0\\ 0 & \zeta^*\end{smallmatrix}\right)$, $\mathrm{Re}\,\zeta>0$, $\mathrm{Im}\,\zeta\geqslant 0$, stands for a quaternion represented by a 2×2 complex matrix and transformed without loss of generality to the diagonal form. One could have equally well transformed it, for instance, into the form $\left(\begin{smallmatrix}a & b\\ -b & a\end{smallmatrix}\right)$ with $a>0$ and $b\geqslant 0$.

Note that in lines 25 and 29 if n is even we have just one matrix K because $\kappa^n=1$, but for n odd we obtain two (nonequivalent) indecomposable pairs, one for $\kappa=+1$ and the other for $\kappa=-1$.

The α-symmetric and α-skew-symmetric matrices are particular cases of a more general class of α-normal matrices. A matrix X is α-normal if it commutes with its α-image (see Subsec. 11 of Sec. 2). For instance, elements of the classical Lie groups are α normal with $\alpha(X)=X^{-1}$.

Let us use Theorems 3 and 4 to establish normal forms for the pairs (Y,K) formed by an α-normal matrix Y and a matrix K establishing an involution

$$\alpha: D^{N\times N}\to D^{N\times N},\qquad \alpha(U)=KU^{\dagger}K^{-1}\quad (U\in D^{N\times N}) \tag{4.4}$$

of $D^{N\times N}$ relative to a $\dagger$ operation. We generalize Theorems 3 and 4 without detailed proof.

Theorem 5: The F-algebra

$$A=\langle FI_N, Y, \alpha(Y)\rangle \tag{4.5}$$

generated by an α-indecomposable α-normal element Y of $D^{N\times N}$ and its image $\alpha(Y)$ under the involution α with matrix K are equivalent to one of the following normal forms:

1. If A can be generated by a single element X over FI_N then we have two cases: (a) and (b): X,K are chosen as in Theorem 3A with either (a) $m_0(t)=t$ or (b) $m_0(t)=t^2+1$; the corresponding normal form of Y is obtained as

$$\text{(a) } Y=\lambda I_N+P(J_N(0))(\lambda\in F), \tag{4.6}$$

where the polynomial P of degree $[P]$ in t over F is one of the following: If $F^{\alpha}=F,\rho=1$ then $P=t$. If $F^{\alpha}=F,\rho=-1$, then $P=t+P_1(t^2)$ where $P_1\in tF(t)$, $[P_1]<N/2$. If $F=\mathbb{C}$, $F^{\alpha}=\mathbb{R},\rho=1$ then either $P=t+iP_1$ or $P=it+tP_2$ with $P_1,P_2\in t\mathbb{R}(t)$, $[P_1]<N$, $[P_2]<N-1$.

$$\text{(b)}\quad Y=aI_N+bX_{m_0}\otimes I_{\mu}+P(I_g\otimes J_{\mu}(0)),\qquad F=\mathbb{R}\quad (a\in\mathbb{R}, b\in\mathbb{R}^{>0}), \tag{4.7}$$

where the polynomial $P\in\mathbb{R}(t)$ is one of the following. If $\rho=1$ then $P=t$. If $\rho=-1$ then $P=t+P_1[t^2]$, where $P_1\in t\mathbb{R}[t]$, $[P_1]<N/2$.

2. If A cannot be generated by a single element over FI_N but if A is indecomposable then we have $D=\mathbb{R}$, N even, and

$$K=\begin{pmatrix}0 & I_{N/2}\\ I_{N/2} & 0\end{pmatrix},\qquad A=\langle FI_N, X_1, X_2, X_3\rangle,$$
$$X_1=\begin{pmatrix}0 & -1\\ 1 & 0\end{pmatrix}\otimes I_{N/2}=-\alpha(X_1),\qquad X_2=\begin{pmatrix}0 & 1\\ 0 & 0\end{pmatrix}\otimes J_{N/2}(0)=\alpha(X_2),$$
$$X_3=\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix}\otimes J_{n/2}(0)=-\alpha(X_3); \tag{4.8}$$

the corresponding normal form of Y is obtained as either

$$\text{(a) } Y=aI_N+bX_1+X_2+P_1(X_3)\quad (a\in\mathbb{R}, b\in\mathbb{R}^{>0}) \tag{4.9}$$

with $P_1\in t\mathbb{R}[t]$, $[P_1]<N/2$, or

$$\text{(b) } Y=aI_N+bX_1+X_2P_2(X_2)+X_3 \tag{4.10}$$

with $P_2\in\mathbb{R}[t]$, $[P_2]<-1+N/2$.

3. If A is decomposable then

$$K=\begin{pmatrix}0 & I_{N/2}\\ \epsilon I_{N/2} & 0\end{pmatrix},\qquad A=\langle FI_N, X, J(0)\oplus O_{N/2}\rangle, \tag{4.11}$$

where (X,K) is as in parts B1 and B2 of Theorem 3, but with the restriction that either

(a) $m_0=t$,

or

(b) $m_0=t^2+1$, $F=\mathbb{R}$.

As in 2, we present A in the form

$$A=\langle FI_N, X_1, X_2, X_3\rangle \tag{4.12}$$

with

$$X_1=X_{m_0}\otimes I_{\mu}\oplus(X_{m_0}\otimes I_{\mu})^{+},\qquad X_2=X-X_1=\rho\alpha(X_2),$$
$$X_3=2J_{\mu}(0)-X_2=-\rho\alpha(X_3);$$

the corresponding normal form of Y is obtained as

$$\text{(a) } Y=aI_N+X_2+P_1(X_3),$$

or

$$Y=aI_N+X_2P_2(X_2)+X_3\quad (a\in F), \tag{4.13}$$
$$\text{(b) } Y=aI_N+bX_1+X_2+P_1(X_3),$$

or

$$Y=aI_N+bX_1+X_2P_2(X_2)+X_3\quad (a\in\mathbb{R}, b\in\mathbb{R}^{>0}),$$

where in all cases P_1,P_2 are polynomials of $F[t]$ for which $[P_1]<\mu$, $[P_2]<\mu-1$.

Theorem 6: The normal form of the α-decomposable pair (Y,K), when Y is α normal, can be chosen as the block diagonal sum of α-indecomposable pairs (Y_i,K_i), $1\leqslant i\leqslant s$, $s\in\mathbb{Z}^{>1}$, and is given as

$$Y=\overset{s}{\underset{i=1}{\oplus}}Y_i,\quad K=\overset{s}{\underset{i=1}{\oplus}}K_i,\quad K_i\in\mathrm{GL}(N_i,D),\qquad 1\leqslant N_i,\quad \sum_{i=1}^{s}N_i=N, \tag{4.14}$$

where (Y_i,K_i) have the form specified in Theorem 5 and all satisfy the coherence condition

$$K_i=\epsilon K_i. \tag{4.15}$$

Proof: Theorem 6 just is the outcome of a full α decomposition of a given pair (Y,K). We remark that the normal form is unique up to permutation of blocks.

For the proof of Theorem 5 one makes use of Dickson's theorem by which $A=B+J(A)$ is the direct sum of a uniquely determined semisimple algebra B over F and the maximal nilpotent ideal $J(A)$ of A. Because of the α-invariance of A also B and $J(A)$ are α invariant. Because of the α-indecomposability of the commutative algebra B it follows

that B is a field, either $B = FI_N$ or $F = \mathbb{R}$, $B \simeq \mathbb{C}$.

Hence, $Y = Y_1 + X_2$ with Y_1 in B, X_2 in $J(A)$, $B = \langle I_N, Y_1 \rangle$, $J(A) = \langle FX_2, FX_3 \rangle$ with $X_3 = \alpha(X_2)$. If $J(A)$ cannot be generated by a single element over F then

$$J(A) = \langle FX_2 \rangle \oplus \langle FX_3 \rangle.$$

If $J(A)$ is one-generated over F then we have $J(A) = \langle FX_2 \rangle$ and either $J(A) = \langle FX_2^+ \rangle$ or $J(A) = \langle FX_2^- \rangle$. The remainder of the theorem follows by application of Theorems 3 and 4 to the more general situation of Theorem 5.

Finally let us point out several classification problems that are related to the one solved here.[11] With the notations of this paper let $G = \mathrm{GL}(N,D)$ and

$$\Omega(\epsilon,\rho) = \{(X,K):\ K^\dagger = \epsilon K,\ (XK)^\dagger = \epsilon\rho XK\},$$

$$\Omega'(\epsilon,\rho) = \{(Y,K):\ K^\dagger = \epsilon K,\ Y^\dagger = \epsilon\rho Y\},$$

where X, Y, K are $N \times N$ matrices over D.

Then G acts on these sets as follows:

$$A\cdot(X,K) = (AXA^{-1}, AKA^\dagger),$$

$$A\cdot(Y,K) = (AYA^\dagger, AKA^\dagger).$$

The map $\omega: \Omega(\epsilon,\rho) \to \Omega'(\epsilon,\rho)$ defined by $\omega(X,K) = (XK,K)$ is G equivariant.

Let $\Omega_0(\epsilon,\rho)$, resp. $\Omega_0'(\epsilon,\rho)$, be the subset of $\Omega(\epsilon,\rho)$, resp. $\Omega'(\epsilon,\rho)$, consisting of all pairs (X,K), resp. (Y,K), with K invertible. The restriction of ω gives a bijection $\omega_0: \Omega_0(\epsilon,\rho) \to \Omega_0'(\epsilon,\rho)$.

We have the following problems:

- (A) Classify G-orbits in $\Omega(\epsilon,\rho)$,
- (A') Classify G-orbits in $\Omega'(\epsilon,\rho)$,
- (A_0) Classify G-orbits in $\Omega_0(\epsilon,\rho)$,
- (A_0') Classify G-orbits in $\Omega_0'(\epsilon,\rho)$.

In our paper we solved the problem A_0. Using ω_0 it is clear that A_0 is equivalent to A_0'. It is also clear that A' is more general than A_0'.

The problem A', in the cases $D = \mathbb{R}$ or $\mathbb{C}$ and $\dagger = T$, has been solved by Ermolaev in Ref. 11. One obtains a solution of A_0 in two steps: 1° choose those pairs (Y_i,K_i) from Ermolaev's list for which K_i is invertible; 2° for each pair (Y_i,K_i) chosen in 1° form the pair $(Y_iK_i^{-1},K_i) = (X_i,K_i)$. These pairs (X_i,K_i) give the list of indecomposable pairs for our problem A_0. Although this is easy to do we have not done it systematically since these forms (X_i,K_i) are not as nice as our forms.

[1] A. I. Malcev, *Foundations of Linear Algebra* (Freeman, San Francisco, 1963).

[2] N. Burgoyne and R. Cushman, J. Algebra **44**, 339 (1977).

[3] H. Zassenhaus, Canad. Math. Bull. **1**, 31, 101, 193 (1958); I. K. Cikunov, Ukr. Math. Zh. **18**, 79, 122 (1966); J. Milnor, Invent. Math. **8**, 83 (1969); T. A. Springer and R. Steinberg, in *Lecture Notes in Mathematics*, Vol. 131 (Springer, New York, 1970), p. 167; J. G. F. Belinfante and P. Winternitz, J. Math. Phys. **12**, 1041 (1971).

[4] J. Williamson, Amer. J. Math. **59**, 141 (1936); N. Burgoyne and R. Cushman, Celest. Mech. **8**, 435 (1974); A. J. Laub and K. Meyer, Celest. Mech. **9**, 213 (1974); M. Moshinsky and P. Winternitz, J. Math. Phys. **21**, 1667 (1980); D. Ž. Djoković, Trans. Am. Math. Soc. **270**, 217 (1982).

[5] J. Patera and C. Rousseau, J. Math. Phys. **23**, 705 (1982); J. Patera, C. Rousseau, and D. Schlomiuk, J. Phys. A, 1063 (1982); J. Patera, C. Rousseau, and D. Schlomiuk, J. Math. Phys. **23**, 4109 (1982).

[6] W. Miller Jr., *Symmetry and Separation of Variables* (Addison-Wesley, Reading, Mass., 1977); W. Miller Jr., J. Patera, and P. Winternitz, J. Math. Phys. **22**, 251 (1981); P. Winternitz, I. Lukač, Ya. A. Smorodinskii, Yad. Fiz. **7**, 192 (1968) [Sov. J. Nucl. Phys. **7**, 139 (1968)].

[7] R. L. Anderson, Lett. Math. Phys. **4**, 1 (1980); R. L. Anderson, J. Harnad, and P. Winternitz, Lett. Math. Phys. **5**, 143 (1981); Physica D **4**, 164 (1982); P. Winternitz, Physica A **114**, 105 (1982); J. Harnad, P. Winternitz, and R. Anderson, "Superposition principles for matrix Riccati equations," J. Math. Phys. (to be published).

[8] J. Patera, P. Winternitz, and H. Zassenhaus, C. R. Math. Rep. Acad. Sci. Canada **2**, 231, 237 (1980).

[9] S. Helgason, *Differential Geometry, Lie Groups, and Symmetric Spaces* (Academic, New York, 1978); R. Gilmore, *Lie Groups, Lie Algebras and Some of Their Applications* (Wiley, New York, 1974).

[10] H. Zassenhaus, *The Theory of Groups*, 2nd ed. (Chelsea, New York, 1958), p. 114, Theorem 7.

[11] Yu. B. Ermolaev, Soviet Math. Doklady **1**, 523 (1960); R. Scharlau, Math. Z. **178**, 359 (1981).

Versal deformations of elements of classical Jordan algebras[a)]

J. Patera
Centre de recherche de mathématiques appliquées, Université de Montréal, Montréal, Québec, Canada

C. Rousseau
Département de mathématiques et de statistique, Université de Montréal, Montréal, Québec, Canada

(Received 13 April 1982; accepted for publication 22 October 1982)

Every involution α defining, by $\alpha(X) = -X$, the elements of one of the classical Lie algebras $o(n, \mathbb{C})$ and $sp(2n, \mathbb{C})$ over $\mathbb{C}$, and $o(p, q)$, $sp(2n,\mathbb{R})$, $u(p, q)$, $o^*(2n)$, and $sp(2p,2q)$ over $\mathbb{R}$, also defines α-symmetric matrices $Y = \alpha(Y)$ spanning the classical Jordan algebras. The paper contains an explicit description of all perturbations of α-symmetric matrices, considered up to equivalence under the action of the corresponding classical Lie group. Such exhaustive perturbations are called versal deformations.

PACS numbers: 02.20.Hj, 02.20.Sv

1. INTRODUCTION

Jordan algebras took their origin in physics.[1] Interest in them varied widely in the course of years.[2] Properties of individual elements, usually represented by operators acting in functional spaces, were frequently studied in operator theory[3] or in connection with differential equations. Thus, for instance, in Ref. 4 it is shown that the general solution of a matrix "symplectic" Riccati equation is given in terms of four particular solutions, one of which W_4, is α-symmetric of pseudo-orthogonal type. As long as one is interested in one Riccati equation only, it is possible to choose a diagonal solution as W_4. If, however, the equation depends on several parameters, W_4 may have unavoidable singularities of small codimension k (k is less than the number of parameters). Then the simplest possible form of the matrix W_4 depending smoothly on the parameters is the one described for this case below.

The purpose of this paper is to compute versal deformations of normal forms of elements of Jordan algebras with involution α. For classical Lie algebras the problem originated in the article of Arnold,[5] was further pursued in Refs. 6–9, and completed in Ref. 10.

The Jordan algebras considered here consist of real, complex or quaternionic matrices X which are α-symmetric, i.e.,

$$\alpha(X) = X. \tag{1.1}$$

Here α is an involution defining the classical Lie algebras of types $o(n, \mathbb{C})$, $sp(2n, \mathbb{C})$, $u(p, q)$, $o(p, q)$, $o^*(2n)$, $sp(2n, \mathbb{R})$, and $sp(2p, 2q)$. These Lie algebras consist of matrices which are α-skew-symmetric, i.e.,

$$\alpha(X) = -X. \tag{1.2}$$

Reference 11 contains a convenient description of the involutions together with a classification of α-symmetric and α-skew-symmetric matrices into conjugacy classes under the action

$$X' = gXg^{-1} \tag{1.3}$$

of elements g of the corresponding Lie group G. We summarize these normal forms in Sec. 2.

The versal deformations have interesting applications, in particular:
(i) They give, for any family of elements of a Jordan algebra depending smoothly on parameters, a normal form depending smoothly on the same parameters which is analogous to the Jordan normal form of matrices.
(ii) Any perturbation of an element is described up to equivalence (1.3) by its versal deformation.
(iii) They allow one to compute bifurcation diagrams[5–10] of families of elements in general position.
Here the description of the approach is rather brief. More details can be found in Refs. 5, 8, and 10.

Let us point out that α-symmetric matrices as well as α-skew-symmetric ones are particular cases of α-normal matrices

$$X\alpha(X) = \alpha(X)X.$$

A description of versal deformations of α-normal matrices is certainly of interest; however, their classification is far from complete. A different example of α-normal matrices are elements of the classical Lie groups. For them $\alpha(X) = X^{-1}$.

2. CONJUGACY CLASSES OF α-SYMMETRIC MATRICES

In this section we recall the classification of α-symmetric matrices in a way suitable for our problem.[11]

A nondegenerate bilinear form on a linear space V,

$$f(x, y) = xK\bar{y}, \quad x,y \in V, \tag{2.1}$$

($\bar{y} = y^T$ or y^+ and $V = \mathbb{C}^n$, $\mathbb{R}^n$ or $\mathbb{H}^n$) splits the space of linear operators on V into the direct sum of two subspaces.

(1) The subspace $\mathcal{G}$ of α-skew-symmetric operators S satisfying

$$f(Sx, y) = -f(x, Sy). \tag{2.2}$$

The classical Lie algebras $o(n, \mathbb{C})$, $sp(2n, \mathbb{C})$, $o(p, q)$, $o^*(2n)$, $sp(2n, \mathbb{R})$, $sp(2p, 2q)$, and $u(p, q)$ appear in this way. If S is represented by a matrix M then

$$\alpha(M) = -M \tag{2.3}$$

for the involution α given by

$$\alpha(M) = K^{-1}\bar{M}K, \tag{2.4}$$

[a)] Work supported in part by the Natural Science and Engineering Research Council of Canada and by the Ministère de l'Education du Québec.

 0022-2488/83/061375-06$02.50 © 1983 American Institute of Physics

where

$$\bar{M} = \begin{cases} M^T & \text{in the context of } o(n, \mathbb{C}),\ sp(2n, \mathbb{C}),\\ & o(p, q),\ sp(2n, \mathbb{R}),\\ M^+ & \text{in the context of } u(p, q),\ o^*(2n),\\ & sp(2p, 2q). \end{cases} \tag{2.5}$$

(2) The subspace $\mathcal{S}$ of α-symmetric operators S satisfying

$$f(Sx, y) = f(x, Sy)\,. \tag{2.6}$$

This gives for the matrix X representing S,

$$\alpha(X) = X\,. \tag{2.7}$$

We call these operators symmetric operators of type $o(n, \mathbb{C})$, $sp(2n, \mathbb{C})$, etc. depending on the Lie algebra defined by the involution α. They form the classical Jordan algebras.

For the classical Lie and Jordan algebras, K satisfies

$$\bar{K} = \epsilon K\,, \tag{2.8}$$

where $\bar{K}$ is given in (2.5) and

$$\begin{aligned} &\epsilon = +1 \quad \text{for } o(n, \mathbb{C}),\ o(p, q),\ u(p, q),\ sp(2p, 2q)\,,\\ &\epsilon = -1 \quad \text{for } sp(2n, \mathbb{C}),\ sp(2n, \mathbb{R}),\ o^*(2n)\,. \end{aligned} \tag{2.9}$$

Moreover the signature of K is specified to be (p, q) [respectively $(2p, 2q)$] in the cases of $o(p, q)$, $u(p, q)$ [respectively, $sp(2p, 2q)$].

In the case of quaternionic algebras $o^*(2n)$ and $sp(2p, 2q)$, the quaternions are identified with 2×2 complex matrices

$$\begin{pmatrix} \gamma & \beta \\ -\beta^* & \gamma^* \end{pmatrix}, \quad \gamma, \beta \in \mathbb{C}\,, \tag{2.10}$$

representing quaternions. Then quaternionic conjugation becomes Hermitian conjugation.

In order to specify an element of $\mathcal{S}$, the set of α-symmetric operators, one gives the matrix K of the involution α given by (2.4) and a matrix X satisfying $\alpha(X) = X$ or equivalently,

$$\bar{X}K = KX\,. \tag{2.11}$$

In a given algebra we allow the matrix K to take different forms in order to get for X the simplest possible form. This can be done through transformations

$$K' = P K \bar{P}^{\dagger}, \quad \det P \neq 0\,. \tag{2.12}$$

We have a decomposition theorem for α-symmetric matrices analogous to the Jordan decomposition theorem.

TABLE I. Indecomposable Jordan normal forms X_j of elements of the Jordan algebras of type $o(n, \mathbb{C})$, $sp(2n, \mathbb{C})$, $o(p, q)$, $sp(2n, \mathbb{R})$, $u(p, q)$, $sp(2p, 2q)$, and $o^*(2n)$. Matrices $J_n(\cdot)$ and H_n are defined in Sec. 2.

Type	ϵ	V	Specifications for K	$\bar{M}$	Case	X_j	K	Signature	
$o(n, \mathbb{C})$	$+1$	$\mathbb{C}^n$	$K = K^T$	M^T	I	$J_n(\beta)$, $\beta\in\mathbb{C}$	H_n		
$sp(2n, \mathbb{C})$	-1	$\mathbb{C}^n$	$K = -K^T$	M^T	I	$\begin{pmatrix} A & 0 \\ 0 & A^T \end{pmatrix}$, $A = J_n(\beta)$, $\beta\in\mathbb{C}$	$\begin{pmatrix} 0 & I_n \\ -I_n & 0 \end{pmatrix}$		
$o(p, q)$	$+1$	$\mathbb{R}^n$	$K = K^T$ sign $K = (p, q)$	M^T	I ($\theta = 1$) I' ($\theta = -1$)	$J_n(a)$, $a\in\mathbb{R}$	θH_n	$p - q = \pm 1$, $p = q$,	n odd n even
					II	$J_n\begin{pmatrix} a & b \\ -b & a \end{pmatrix}$, $a\in\mathbb{R}, b>0$	$\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}\otimes H_n$	$p = q$	
$sp(2n, \mathbb{R})$	-1	$\mathbb{R}^n$	$K = -K^T$	M^T	I	$\begin{pmatrix} A & 0 \\ 0 & A^T \end{pmatrix}$, $A = J_n(a)$, $a\in\mathbb{R}$	$\begin{pmatrix} 0 & I_n \\ -I_n & 0 \end{pmatrix}$		
					II	$\begin{pmatrix} A & 0 \\ 0 & A^T \end{pmatrix}$, $A = J_n\begin{pmatrix} a & b \\ -b & a \end{pmatrix}$, $a\in\mathbb{R}$, $b>0$	$\begin{pmatrix} 0 & I_{2n} \\ -I_{2n} & 0 \end{pmatrix}$		
$u(p, q)$	1	$\mathbb{C}^n$	$K = K^+$ sign $K = (p, q)$	M^+	I($\theta = 1$) I'($\theta = -1$)	$J_n(a)$, $a\in\mathbb{R}$	θH_n	$p - q = \pm 1$, $p = q$,	n odd n even
					II	$\begin{pmatrix} A & 0 \\ 0 & A^+ \end{pmatrix}$, $A = J_n(\alpha)$	$\begin{pmatrix} 0 & I_n \\ I_n & 0 \end{pmatrix}$	$p = q$	
						$\alpha\in\mathbb{C}$, $\operatorname{Im}\alpha > 0$			
$o^*(2n)$	-1	$\mathbb{H}^n$	$K = -K^+$	M^+	I	$J_n\begin{pmatrix} \alpha & 0 \\ 0 & \alpha^* \end{pmatrix}$, $\alpha\in\mathbb{C}$, $\operatorname{Im}\alpha > 0$	$\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}\otimes H_n$		
					II	$J_n\begin{pmatrix} a & 0 \\ 0 & a \end{pmatrix}$, $a\in\mathbb{R}$	$\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}\otimes H_n$		
$sp(2p, 2q)$	$+1$	$\mathbb{H}^n$	$K = K^+$	M^+	I ($\theta = 1$)	$J_n\begin{pmatrix} a & 0 \\ 0 & a \end{pmatrix}$, $a\in\mathbb{R}$	$\theta I_2\otimes H_n$	$p - q = \pm 1$,	n odd
			sign $K = (2p, 2q)$		I' ($\theta = -1$)			$p = q$,	n even
					II	$\begin{pmatrix} A & 0 \\ 0 & A^T \end{pmatrix}$, $A = J_n\begin{pmatrix} a & b \\ -b & a \end{pmatrix}$, $a\in\mathbb{R}$, $b>0$	$\begin{pmatrix} 0 & I_{2n} \\ I_{2n} & 0 \end{pmatrix}$	$p = q$	

Any such matrix X is a direct sum of blocks of two types.

(1) Irreducible "generalized Jordan blocks." For the matrix elements we allow 2×2 matrices.

(2) Blocks of the form $\left(\begin{smallmatrix} A & 0 \\ 0 & A \end{smallmatrix}\right)$ with A a generalized Jordan block. Such a block is orthogonally indecomposable with respect to the form f.

The precise form of the irreducible blocks and a summary of the information of this section appears in Table I. In each case the matrix K is given.

Notation for TABLE I

(1)

$$J_n(\alpha)=\begin{pmatrix} \alpha & 1 & & \\ & \ddots & \ddots & \\ & & & 1 \\ & & & \alpha \end{pmatrix}, \tag{2.13}$$

(2)

$$J_n\begin{pmatrix} a & b \\ -b & a \end{pmatrix}=\begin{pmatrix} a & b \\ -b & a \end{pmatrix}\otimes I_n + I_2 \otimes J_n(0)$$

$$=\begin{pmatrix} a & b & 1 & 0 & & & & \\ -b & a & 0 & 1 & & & & \\ & & \ddots & & & & & \\ & & & & 1 & 0 & & \\ & & & \ddots & 0 & 1 & & \\ & & & & a & b & & \\ & & & & -b & a & & \end{pmatrix}. \tag{2.14}$$

Similarly for $J_n\left(\begin{smallmatrix} \alpha & 0 \\ 0 & \alpha^* \end{smallmatrix}\right)$,

(3)

$$H_n=\begin{pmatrix} & & 1 \\ & \iddots & \\ 1 & & \end{pmatrix}. \tag{2.15}$$

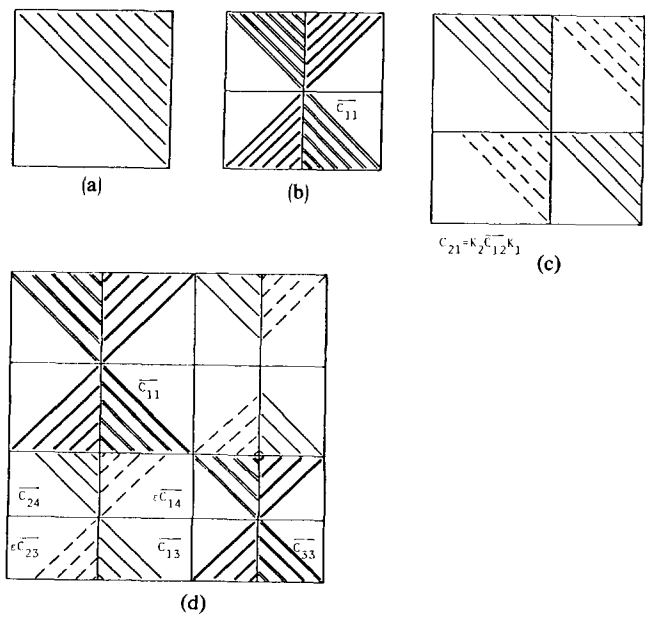


FIG. 1. Centralizers of normal forms of elements of $\mathcal{S}$. Details of notation in Sec. 3.

3. VERSAL DEFORMATIONS OF SYMMETRIC MATRICES

First we define briefly the notion of versal deformation. Details can be found in Refs. 5, 8, and 10.

In what follows we call $\mathcal{S}$ the space of α-symmetric operators on V, G and $\mathcal{G}$ the corresponding Lie group and Lie algebra.

Definition: (1) A *deformation* of an element $X_0\in\mathcal{S}$ is a family of elements of $\mathcal{S}$ indexed by $\lambda\in\Lambda$, $X:\Lambda\to\mathcal{S}$ where $\Lambda\subset\mathbb{C}^m(\mathbb{R}^m)$ is a neighborhood of 0, and where $X(0)=X_0$ and X depends holomorphically (smoothly) on the parameters. (2) A deformation $X(\lambda)=X(\lambda_1,\ldots,\lambda_m)$ of X_0 is *versal* iff for any deformation $X'(\mu_1,\ldots,\mu_k)\in\mathcal{S}$ of X_0, $X'(\mu)$ is induced by $X(\lambda)$, i.e., there exists a neighborhood V of 0 in $\mathbb{C}^k(\mathbb{R}^k)$, a map $\varphi:V\to\mathbb{C}^m(\mathbb{R}^m)$ with $\varphi(0)=0$, and a map $g:V\to G$ with $g(0)=I$ such that $\forall\mu\in V$, $X'(\mu)=g(\mu)X(\varphi(\mu))g^{-1}(\mu)$ with φ and g holomorphic (smooth).

The parameters involved in the definition are complex for the symmetric matrices of type $o(n,\mathbb{C})$ or $sp(2n,\mathbb{C})$ and real in the other cases.

It is obviously enough to consider versal deformations of elements X_J in generalized Jordan normal form, since if $X_J=S^{-1}X_0S$ is the generalized Jordan normal form of X_0, and $X(\lambda)$ is a versal deformation of X_J, then $SX(\lambda)S^{-1}$ is a versal deformation of X_0.

The technique for finding versal deformations is given by the following lemma:

Lemma 1[5]: A deformation $X(\lambda)$ of a matrix X_0 is versal iff $X(\lambda)$ is transversal to Orb X_0 at the point $\lambda=0$. The minimal number of parameters appearing in a versal deformation is therefore equal to the codimension of Orb X_0.

To find versal deformations with minimal number of parameters we need to compute a complement of the tangent space of Orb X_J of an element X_J of $\mathcal{S}$ in generalized Jordan normal form. This tangent space T_{X_J} Orb X_J is the set of matrices $[D,X_J]$, $D\in\mathcal{G}$. Among the complements of T_{X_J} Orb X_J we specify two here:

(1) the orthogonal complement under the inner product

$$(A,B)=\operatorname{tr}(AB^{+}),$$

(2) a complement with minimal number of nonzero entries.

For the first one we use the following lemma.

Lemma 2[5]: $\mathbb{C}\in\mathcal{S}$ is orthogonal to Orb X_0 *iff* $[C^{+},X_0]=0$.

We need therefore to compute centralizers in $\mathcal{S}$. We compute first centralizers of indecomposable elements, then centralizers of direct sums of pairs of indecomposable elements. Where the indecomposable elements have different eigenvalues, the centralizer is the direct sum of the two centralizers. We therefore restrict ourselves to pairs of indecomposable elements with the same eigenvalues.

The centralizers in $\mathcal{S}$ are the sets

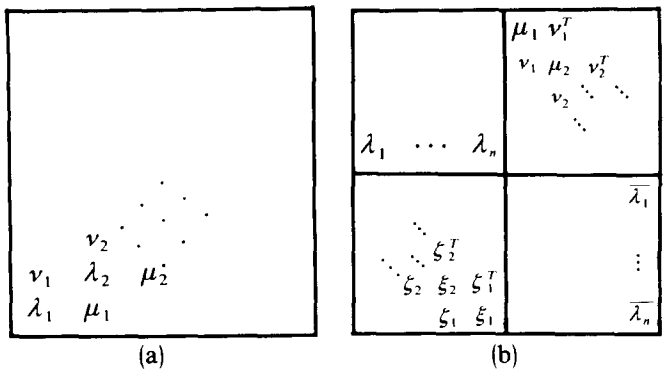
$$\operatorname{Centr}_{\mathcal{S}}X=\{X\mid[C,X]=0 \text{ and } KX=\overline{X}K\}, \tag{3.1}$$

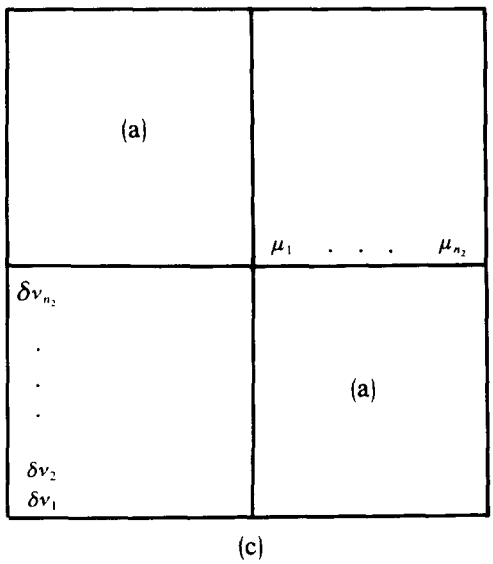
where $\overline{X}$ is given in (2.8).

The centralizers are given by Fig. 1(a)–(d) together with the explanations appearing in Table II.

TABLE II. Minimal versal deformations of elements of $\mathscr{S}$. Details of notation in Sec. 3.

Type	Case	Centralizer	$X(\lambda)$	Order of X_J	d
$o(n, \mathbb{C})$	I	Fig. 1(a)—$\gamma\in\mathbb{C}$	Fig. 2(a), $\lambda_i, \mu_i\in\mathbb{C}$, $\nu_i=\mu_i$	n	n
	I + I	Fig. 1(c)--$\gamma\in\mathbb{C}$	Fig. 2(c), $\mu_i\in\mathbb{C}$, $\delta=1$, $\nu_i=\mu_i$	$n_1\geqslant n_2$	n_1+2n_2
$sp(2n, \mathbb{C})$	I	Fig. 1(b)=$\gamma\in\mathbb{C}$—line of 0	Fig. 2(b), $\lambda_i\in\mathbb{C}$, $\mu_i=\nu_i=\xi_i=\zeta_i=0$	$2n$	n
$\epsilon=-1$	I + I	Fig. 1(a)—$\gamma\in\mathbb{C}$--$\gamma\in\mathbb{C}$	Fig. 2(d), $\varphi_i, \chi_i, \psi_i, \omega_i\in\mathbb{C}$	$2n_1\geqslant 2n_2$	n_1+5n_2
$o(p, q)$	I, I′	Fig. 1(a)—$c\in\mathbb{R}$	Fig. 2(a), $\lambda_i, \mu_i\in\mathbb{R}$, $\nu_i=\mu_i$	n	n
$\epsilon=+1$	II	Fig. 1(a)—$\begin{pmatrix} c & d \\ -d & c \end{pmatrix}$	Fig. 2(a), $\lambda_i\sim\begin{pmatrix} c & d \\ -d & 0 \end{pmatrix}$, $\mu_i=\begin{pmatrix} 0 & d_i \\ 0 & c_i \end{pmatrix}$, $\nu_i=\begin{pmatrix} 0 & 0 \\ -d_i & c_i \end{pmatrix}$	$2n$	$2n$
	I + I I′ + I′ I + I′ ($\delta=-1$)	Fig. 1(c)--$c\in\mathbb{R}$	Fig. 2(c), $\mu_i\in\mathbb{R}$, $\nu_i=\mu_i$	$n_1\geqslant n_2$	n_1+2n_2
	II + II	Fig. 1(c)--$\begin{pmatrix} c & d \\ -d & c \end{pmatrix}$	Fig. 2(c), $\mu_i=\begin{pmatrix} 0 & 0 \\ c_i & d_i \end{pmatrix}$, $\nu_i=\begin{pmatrix} 0 & -c_i \\ 0 & d_i \end{pmatrix}$, $\delta=1$	$2n_1\geqslant 2n_2$	$2n_1+4n_2$
$sp(2n, \mathbb{R})$	I	Fig. 1(b)=$c\in\mathbb{R}$—line of 0	Fig. 2(b), $\lambda_i\in\mathbb{R}$, $\mu_i=\nu_i=\xi_i=\zeta_i=0$	$2n$	n
$\epsilon=-1$	II	Fig. 1(b)=$\begin{pmatrix} c & d \\ -d & c \end{pmatrix}$—line of 0	Fig. 2(b), $\lambda_i\in\mathbb{R}$, $\mu_i=\nu_i=\xi_i=\zeta_i=0$	$4n$	$2n$
	I + I	Fig. 1(d)—$c\in\mathbb{R}$--$c\in\mathbb{R}$	Fig. 2(d), $\varphi_i, \chi_i\psi_i, \omega_i\in\mathbb{R}$	$2n_1\geqslant 2n_2$	n_1+5n_2
	II + II	Fig. 1(d)—$\begin{pmatrix} c & d \\ -d & c \end{pmatrix}$--$\begin{pmatrix} c & d \\ d & -c \end{pmatrix}$	Fig. 2(d), $\varphi_i, \chi_i, \psi_i, \omega_i\in\mathbb{R}$	$4n_1\geqslant 4n_2$	$2n_1+10n_2$
$u(p, q)$	I, I′	Fig. 1(a)—$c\in\mathbb{R}$	Fig. 2(a), $\lambda_i, \mu_i\in\mathbb{R}$, $\nu_i=\mu_i$	n	n
$\epsilon=+1$	II	Fig. 1(b)=$\gamma\in\mathbb{C}$—line of 0	Fig. 2(b), $\mu_i=\nu_i=\xi_i=\zeta_i=0$	$2n$	$2n$
	I + I I′ + I′ I + I′ ($\delta=-1$)	Fig. 1(c)--$\gamma\in\mathbb{C}$	Fig. 2(c), $\mu_i\in\mathbb{C}$, $\nu_i=\mu_i^*$	$n_1\geqslant n_2$	n_1+3n_2
	II + II	Fig. 1(d)—$\gamma\in\mathbb{C}$--line of 0	Fig. 2(d), $\phi_i, \omega_i\in\mathbb{C}$, $\chi_i=\psi_i=0$	$2n_1\geqslant 2n_2$	$2n_1+6n_2$
$o^*(2n)$	I	Fig. 1(a)—$\begin{pmatrix} \gamma & 0 \\ 0 & \gamma^* \end{pmatrix}$	Fig. 2(a), $\lambda_i, \mu_i\sim\begin{pmatrix} \gamma & 0 \\ 0 & \gamma^* \end{pmatrix}$, $\nu_i=\mu_i$	$2n$	$2n$
$\epsilon=-1$	II	Fig. 1(a)—$\begin{pmatrix} \gamma & \beta \\ -\beta^* & \gamma^* \end{pmatrix}$	Fig. 2(a), $\lambda_i, \mu_i\sim\begin{pmatrix} \gamma & \beta \\ -\beta^* & \gamma^* \end{pmatrix}$, $\nu_i=\mu_i$	$2n$	$4n$
	I + I $\delta=1$	Fig. 1(c)--$\begin{pmatrix} \gamma & 0 \\ 0 & \gamma^* \end{pmatrix}$	Fig. 2(c), $\mu_i\sim\begin{pmatrix} \gamma & 0 \\ 0 & \gamma^* \end{pmatrix}$, $\nu_i=\mu_i$	$2n_1\geqslant 2n_2$	$2n_1+4n_2$
	II + II $\delta=1$	Fig. 1(c)---$\begin{pmatrix} \gamma & \beta \\ -\beta^* & \gamma^* \end{pmatrix}$	Fig. 2(c), $\mu_i\sim\begin{pmatrix} \gamma & \beta \\ -\beta^* & \gamma^* \end{pmatrix}$, $\nu_i=\mu_i$	$2n_1\geqslant 2n_2$	$4n_1+8n_2$
$sp(2p, 2q)$	I, I′	Fig. 1(a)—$\begin{pmatrix} c & 0 \\ 0 & c \end{pmatrix}$	Fig. 2(a), $\lambda_i, \mu_i\sim\begin{pmatrix} c & 0 \\ 0 & c \end{pmatrix}$, $\nu_i=\mu_i$	$2n$	n
$\epsilon=+1$	II	Fig. 1(b)=$\begin{pmatrix} c & d \\ -d & c \end{pmatrix}$—line of 0	Fig. 2(b), $\lambda_i\sim\begin{pmatrix} c & d \\ -d & c \end{pmatrix}$, $\mu_i=\nu_i=\xi_i=\zeta_i=0$	$4n$	$2n$
	I + I I′ + I′ I + I′ ($\delta=-1$)	Fig. 1(c)--$\begin{pmatrix} c & d \\ -d & c \end{pmatrix}$	Fig. 2(c), $\mu_i\sim\begin{pmatrix} c & d \\ -d & c \end{pmatrix}$, $\nu_i=\mu_i^T$	$2n_1\geqslant 2n_2$	n_1+2n_2
	II + II	Fig. 1(d)—$\begin{pmatrix} c & d \\ -d & c \end{pmatrix}$--$\begin{pmatrix} ic & id \\ id & -ic \end{pmatrix}$	Fig. 2(d), $\phi_i, \omega_i\sim\begin{pmatrix} c & d \\ -d & c \end{pmatrix}$, $\chi_i, \psi_i\sim\begin{pmatrix} ic & id \\ id & -ic \end{pmatrix}$ (read Fig. 2(d) with $\epsilon=-1$)	$4n_1\geqslant 4n_2$	$2n_1+10n_2$





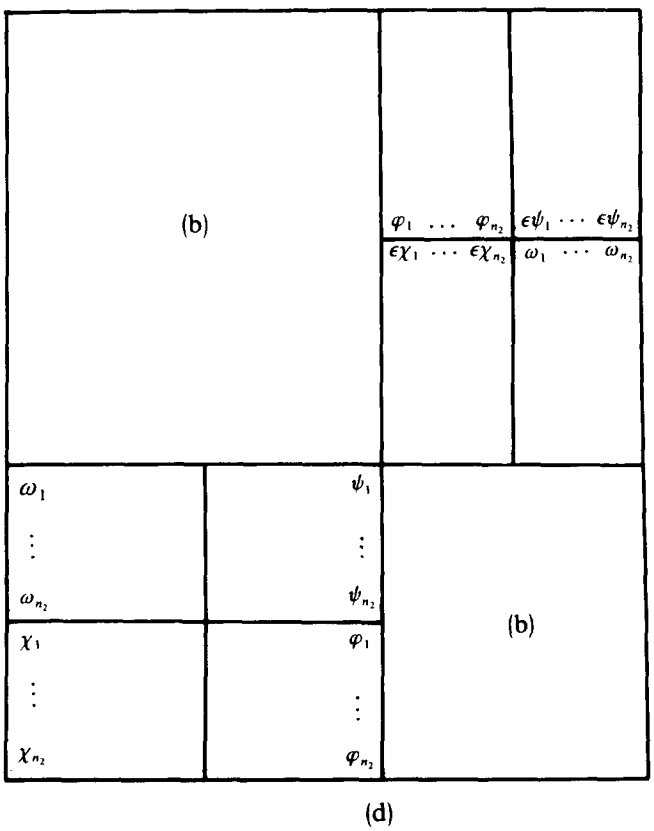


FIG. 2. Transversal subspaces to orbits of normal forms of elements of $\mathcal{S}$ with minimal number of nonzero entries. Details of notation in Sec. 3.

Notation for Figure 1

(1) A line (full or broken) denotes a repetition of the same element along the line, the element being of one of the following types: $c \in \mathbb{R}$, $id \in \mathbb{R}$, $\gamma \in \mathbb{C}$, $\left(\begin{smallmatrix} c & d \\ -d & c \end{smallmatrix}\right)$, $c, d \in \mathbb{R}$, $\left(\begin{smallmatrix} \gamma & \beta \\ -\beta^* & \gamma^* \end{smallmatrix}\right) \in \mathbb{H}$, etc.

(2) C_{ij} is defined in (2.5).

(3) ϵ is given for the different algebras in Table II.

Notation for Table II

(1) Roman (Greek) letters denote real (complex) numbers.

(2) The centralizer column gives the legend for the parts of Fig. 1. For decomposable cases, for example I + III, the legend of the diagonal blocks is found next to the simpler cases I, respectively III. In I + III one only finds the legend of nondiagonal blocks.

(3) In the column "order of X_J," for the decomposition cases we give the respective orders of the two blocks mentioned in the column case.

(4) The last column gives the minimal number of complex (real) parameters for a versal deformation.

(5) A matrix C in the centralizer is thought of as a matrix of blocks $C = (C_{ij})$ where C_{ij} is a matrix.

Theorem 1: Let $X_J \in \mathcal{S}$ be in generalized Jordan form. Then $X_A(1) = X_J + C^+(\lambda)$ is a minimal versal deformation of X_J, where $C^+(\lambda)$ is a generic element of $\mathrm{Centr}_{\mathcal{S}} X_J$. The number d of independent parameters is equal to the dimension of $\mathrm{Centr}_{\mathcal{S}} X_J$ and is computed according to the last column of Table II.

Proof: This follows immediately from Lemmas 1 and 2, the proofs of which are in Ref. 5.

From the calculation of d one has

Theorem 2: The codimension d of the orbit of an element $X \in \mathcal{S}$ can be computed from its Jordan normal form in $gl(n, \mathbb{C})$, $gl(n, \mathbb{R})$, or $gl(n, \mathbb{H})$ depending on the context.

(A corresponding theorem holds also in the α-skew-symmetric case. Cf. Ref. 9.)

We now give a minimal versal deformation with minimal number of nonzero entries.

Theorem 3: Let $X_J \in \mathcal{S}$ be in generalized Jordan form. Then $X_A(2) = X_J + X(\lambda)$ is a minimal versal deformation of X_J with minimal number of nonzero entries, where $X(\lambda)$ is given in Fig. 2 and Table II.

Proof: We need to take $X(\lambda)$ transversal to T_{X_J} Orb X_J. We therefore compute the general form of an element of T_{X_J} Orb X_J. Such an element is of the form $[D, X_J]$, $D \in \mathcal{G}$. The general form of an element $[D, X_J]$ is related to the general form of the adjoint C^+ of an element $C \in \mathrm{Centr}\, X_J$. For each line (full or broken) of C^+ filled with nonzero elements [nonzero blocks of type $\left(\begin{smallmatrix} c & d \\ -d & c \end{smallmatrix}\right)$, $\left(\begin{smallmatrix} ic & id \\ id & -ic \end{smallmatrix}\right)$, $\left(\begin{smallmatrix} \gamma & 0 \\ 0 & \gamma^* \end{smallmatrix}\right)$, $\left(\begin{smallmatrix} \gamma & \beta \\ -\beta^* & \gamma^* \end{smallmatrix}\right)$] in Fig. 1, we have that the sum of the elements (blocks) of the corresponding line is zero [of one of the forms $\left(\begin{smallmatrix} c & d \\ d & -c \end{smallmatrix}\right)$, $\left(\begin{smallmatrix} ic & id \\ -id & ic \end{smallmatrix}\right)$, $\left(\begin{smallmatrix} 0 & \beta \\ -\beta^* & 0 \end{smallmatrix}\right)$, zero]. Once we know the conditions for an element to be in T_{X_J} Orb X_J, a minimal complement to T_{X_J} Orb X_J is found in taking an element that satisfies none of the conditions. The simplest way to take an element such that for example the sum of the elements on a line is different from 0 is to have a unique

nonzero entry on that line (cf. Refs. 5 and 10 for more details).

Notation for Figure 2 and TABLE II

(1) The quantities λ_j, μ_j, etc. of Fig. 2 are of the form $c \in \mathbb{R}$, $id \in \mathbb{R}$, $\left(\begin{smallmatrix} c & d \\ -d & c \end{smallmatrix}\right)$, $\left(\begin{smallmatrix} \gamma & \beta \\ -\beta & \gamma \end{smallmatrix}\right)$, etc.

(2) $\lambda_i \sim \left(\begin{smallmatrix} c & d \\ -d & c \end{smallmatrix}\right)$ means λ_i is of the form $\left(\begin{smallmatrix} c & d \\ -d & c \end{smallmatrix}\right)$.

Example: Let us consider a symmetric operator X_J of type $o(5,5)$ having two Jordan blocks of type I and I', $J_3(a)$ and $J_2(a)$, and two generalized Jordan blocks of type II, $J_2\left(\begin{smallmatrix} c & d \\ -d & c \end{smallmatrix}\right)$ and $J_1\left(\begin{smallmatrix} c & d \\ -d & c \end{smallmatrix}\right)$. Then by Fig. 2 and Table II we get

$$X_A(\lambda,\mu) = \begin{pmatrix} a & 1 & 0 & & & \\ \lambda_2 & a+\lambda_3 & 1 & & & \\ \lambda_1 & \lambda_2 & a & \lambda_6 & \lambda_7 & \\ -\lambda_7 & & & a+\lambda_5 & 1 & \\ -\lambda_6 & & & \lambda_4 & a+\lambda_5 & \end{pmatrix} \oplus \begin{pmatrix} c & d & & & & \\ -d-\mu_3 & c+\mu_4 & & & & \\ \mu_1 & \mu_2 & c & d+\mu_3 & & \\ -\mu_2 & 0 & -d & c+\mu_4 & \mu_7 & \mu_8 \\ & -\mu_7 & & & c+\mu_5 & d+\mu_6 \\ & -\mu_8 & & & -d-\mu_6 & c \end{pmatrix} \tag{3.2}$$

The matrix K of the involution α is

$$K = H_3 \oplus (-H_1) \oplus \left(\begin{smallmatrix} 1 & \\ & -1 \end{smallmatrix}\right) \otimes H_2 \oplus \left(\begin{smallmatrix} 1 & \\ & -1 \end{smallmatrix}\right). \tag{3.3}$$

[1]P. Jordan, J. von Neumann, and E. P. Wigner, Ann. Math. **35**, 29 (1934).
[2]F. Gürsey, in *Lecture Notes in Physics*, Vol. 94 (Springer, New York, 1978).
[3]M. S. Birman and M. Z. Solomyak, *Spectral Theory of Self-adjoint Operators in Hilbert Space* (Leningrad University, Leningrad, 1980).
[4]J. Harnad, P. Winternitz, and R. Anderson, "Superposition principles for matrix Riccati equations," J. Math. Phys. **24**, 1062 (1983).
[5]V. I. Arnold, "On matrices depending on parameters," Russian Math. Surveys **26**, 29 (1971).
[6]D. M. Galin, "On real matrices depending on parameters," Usp. Math. Nauk **27**, 241 (1972) (in Russian).
[7]D. M. Galin, "Versal deformations of linear Hamiltonian systems," Tr. Sem. I. G. Petrovsky **1**, 63 (1975) (in Russian).
[8]J. Patera and C. Rousseau "Complex orthogonal and symplectic matrices depending on parameters," J. Math. Phys. **23**, 705 (1982).
[9]J. Patera, C. Rousseau, and D. Schlomiuk, "Dimensions of orbits and strata in complex and real classical Lie algebras," J. Math. Phys. **23**, 490 (1982).
[10]J. Patera, C. Rousseau, and D. Schlomiuk, "Versal deformations of elements of real classical Lie algebras," J. Phys. A **15**, 1982.
[11]D. Djoković, J. Patera, P. Winternitz, and H. Zassenhaus, "Normal forms of elements of classical real and complex Lie and Jordan algebras," J. Math. Phys. **24**, 1363 (1983).

The fundamental invariants of inhomogeneous classical groups[a)]

M. Perroud
Département de mathématiques appliquées, Ecole Polytechnique de Montréal, Montréal, H3C 3A7, Canada

(Received 29 June 1982; accepted for publication 15 October 1982)

The Casimir operators of the following groups are explicitly constructed :$\mathbb{R}^n \square Gl(n,\mathbb{R})$, $\mathbb{R}^n \square Sl(n,\mathbb{R})$, $\mathbb{R}^n \square O(p,q)$,$\mathbb{C}^n \square U(p,q)$,$\mathbb{R}^{2n} \square Sp(n,\mathbb{R})$, $\mathbb{C}^n \square Gl(n,\mathbb{C})$,$\mathbb{C}^n \square Sl(n,\mathbb{C})$, $\mathbb{C}^n \square O(n,\mathbb{C})$, $\mathbb{C}^{2n} \square Sp(n,\mathbb{C})$, $\mathbb{H}^n \square U^*(2n)$, $\mathbb{H}^n \square SU^*(2n)$, $\mathbb{H}^n \square O^*(2n)$, $\mathbb{H}^n \square Sp(p,q)$, and $\mathbb{R} \square (\mathbb{R}^{2n} \square Sp(n,\mathbb{R}))$. The method is based on a particular fiber bundle structure of the generic orbits generated by the co-adjoint representation of a semidirect product.

PACS numbers: 02.20.Nq

The Casimir operators of the classical Lie groups are well known and several articles have been devoted to them in the physical literature.[1] Curiously enough no systematic attempt has been undertaken in order to determine the Casimir operators of the inhomogeneous groups associated to these classical groups. Of course those of the Euclidean group $\mathscr{E}(3)$ and of the Poincaré group $\mathscr{E}(1,3)$ are known; nevertheless it does not seem that the quite remarkable structure of these invariants has succeeded in convincing someone to look for its origin. Some other isolated cases have been studied: the complex affine group $\mathbb{C}^6 \square Gl(6,\mathbb{C})$[2] and various inhomogeneous groups associated to the unitary group $U(n)$.[3] Some of these inhomogeneous groups appear in physics, not only as subgroups of classical groups, but also very naturally in certain models of unified gauge field theories.[4] In this context the knowledge of the Casimir operators might play a useful role as a guideline. Another field of applicability is found in nuclear physics, particularly in some recent developments on the collective model.[5] [I thank the referee for pointing out these three articles and especially the latter one (Guillemin and Sternberg) in which a very elegant analysis of the structure of the co-adjoint orbits of a semidirect product is developed.] A first systematic study in the search for the Casimir operators of the inhomogeneous classical groups can be found in a recent series of papers[6] in which the number of Casimir operators of the groups $\mathbb{R}^n \square Gl(n,\mathbb{R})$, $\mathbb{R}^n \square Sl(n,\mathbb{R})$, $\mathbb{R}^n \square O(p,q)$, $\mathbb{C}^n \square U(p,q)$, and $\mathbb{C}^n \square SU(p,q)$ have been computed.

The Casimir operators have played a central role in the early study of irreducible representations of compact Lie groups and they continue to be an essential tool in physical applications. Furthermore, after some eclipse in the representation theory of Lie groups, they found a renewal of interest with Kirillov's method of orbits[7] and the closely related geometric quantization theories. According to a result of Gel'fand,[8] which we shall recall in Sec. 1, the Casimir operators of a group are in 1-1 correspondence with the invariant polynomial functions on the dual space of the Lie algebra of the group. On the other hand, according to Kirillov's conjecture (already verified for wide classes of groups) a relation should exist between equivalence classes of irreducible representations of the group and some (or all) orbits generated in this dual space by its co-adjoint representation. In the case where these orbits are contained in the simultaneous level sets of polynomial functions, Casimir operators exist and have a nice geometric interpretation. Of course there exist Lie groups for which the orbits are contained in simultaneous level sets of functions which are not polynomials. No true Casimir operators can be associated to these functions; however, by "forcing" the correspondence, some interesting properties of these formal operators have been observed for groups of low dimension.[9] Casimir operators and their possible generalizations are thus relevant in group representation theory and its applications.

The usual recipe for finding the possible invariant polynomial functions associated to the Casimir operators of a non-semi-simple group is to solve a system of linear PDE's. Although this has been successfully used for groups of low dimension,[9] it is not effective for an arbitrary group. Since the first integrals of this system are related to the co-adjoint orbits, it is reasonable to reverse the point of view and to reconstruct the invariant functions from the knowledge of the orbits. This is the point of view adopted in this article.

In Sec. 2, we exhibit the particular form of the co-adjoint representation of the semidirect product of a Lie group with an abelian one. In Sec. 3, we study in detail the structure of the co-adjoint orbits of such a group. It is shown that these orbits are completely determined by an orbit generated by the homogeneous group in the dual space of the translation group and by a co-adjoint orbit generated by the little group of a point belonging to the first one. This is exactly the geometric analog of the result obtained by Mackey–Wigner's induction method in the search for the irreducible unitary representations of a semidirect product.[10] In Secs. 5–10 we compute in some detail the fundamental invariant polynomials of the real inhomogeneous classical groups using a procedure proposed in Sec. 4. The cases of complex and quaternionic inhomogeneous classical groups are briefly exposed in Secs. 10 and 11. An Appendix in which the classical groups and their fundamental invariants are summarized completes the text.

1. CASIMIR OPERATORS AND INVARIANT POLYNOMIAL FUNCTIONS

In this section we briefly recall an important result due to Gel'fand[8] which plays a key role in the following.

Let G be a Lie group, $\mathscr{g}$ its Lie algebra and $\mathscr{U}(\mathscr{g})$ the universal enveloping algebra of $\mathscr{g}$.[11] It is known that, with

[a)] Supported in part by the Natural Sciences and Engineering Research Council Canada.

 0022-2488/83/061381-11$02.50 © 1983 American Institute of Physics

respect to any basis $\{X_1,...,X_n\}$ of $\mathcal{g}$, the elements of $\mathcal{U}(\mathcal{g})$ can be uniquely written in the form of polynomials in the noncommuting variables $X_1,...,X_n$,

$$P=\sum a_{i_1\cdots i_l}X_{i_1}\cdots X_{i_l} \tag{1}$$

if we stipulate that the coefficients $a_{i_1\cdots i_l}$ are symmetric in all their suffixes.

Let $\mathcal{g}^*$ be the dual space of $\mathcal{g}$ and denote by $\langle \eta, X\rangle$ the scalar product with $\eta\in\mathcal{g}^*$ and $X\in\mathcal{g}$. As a vector space, $\mathcal{U}(\mathcal{g})$ is isomorphic to the vector space $P[\mathcal{g}^*]$ of all polynomial functions on $\mathcal{g}^*$. Indeed, to any element of $\mathcal{U}(\mathcal{g})$ corresponds the polynomial function

$$P(\eta)=\sum a_{i_1\cdots i_l}\eta_{i_1}\cdots\eta_{i_l}, \tag{2}$$

where $\eta_j=\langle \eta, X_j\rangle$, and conversely, to any polynomial function

$$P(\eta)=\sum b_{k_1\cdots k_n}\eta_1^{k_1}\cdots\eta_n^{k_n} \tag{3}$$

corresponds a unique element of $\mathcal{U}(\mathcal{g})$ defined by

$$P=\mathrm{Sym}\left(\sum b_{k_1\cdots k_n}X_1^{k_1}\cdots X_n^{k_n}\right), \tag{4}$$

where Sym denotes the symmetrizer in $\mathcal{U}(\mathcal{g})$. It can be shown that this isomorphism does not depend on the chosen basis of $\mathcal{g}$.

The elements of the center $Z(\mathcal{g})$ of $\mathcal{U}(\mathcal{g})$ are generally called Casimir operators of G in the physical literature. An important problem in the representation theory of Lie groups is to construct a minimal set of such operators which algebraically generate the center $Z(\mathcal{g})$. Gel'fand has given a complete characterization of $Z(\mathcal{g})$ in terms of invariant polynomial functions on $\mathcal{g}^*$. The meaning of this invariance is the following: let us consider the co-adjoint representation Ad* of G defined by

$$\langle \mathrm{Ad}^*(g)\eta, X\rangle=\langle \eta,\mathrm{Ad}(g)^{-1}X\rangle \quad \forall\eta\in\mathcal{g}^*,\ \forall X\in\mathcal{g}. \tag{5}$$

Then a function f defined on $\mathcal{g}^*$ is called invariant if $f\circ\mathrm{Ad}^*(g)=f$ for all $g\in$G. Gel'fand has proved the following[7]:

Proposition 1: An element $P\in\mathcal{U}(\mathcal{g})$ belongs to $Z(\mathcal{g})$ if and only if the corresponding polynomial function $P(\eta)$ is invariant.

This proposition asserting the existence of a vector space isomorphism between $Z(\mathcal{g})$ and the space $P_G[\mathcal{g}^*]$ of invariant polynomial functions has meanwhile been improved by Duflo[12] who has shown that it is actually an algebra isomorphism.

The invariant functions on $\mathcal{g}^*$ are not always generated by invariant polynomials: it depends on the group. Let us state a useful criterion[13]: if the derived Lie algebra $[\mathcal{g},\mathcal{g}]$ is equal to $\mathcal{g}$, then the invariant functions on $\mathcal{g}^*$ are generated by invariant polynomials. Not all of the groups under consideration satisfy this criterion but it turns out that for those which do not, no nontrivial invariant function exists at all.

The problem of finding invariant functions on $\mathcal{g}^*$ can be solved by purely algebraic methods in the case of the classical Lie groups. For groups of low dimension they have been found[9] by solving the system of partial differential equations

$$X_i^*f(\eta)=0, \quad i=1,...,n,$$

where X_i^* is the differential operator associated to the element X_i of a basis of $\mathcal{g}$:

$$X_i^*f(\eta)=\frac{d}{dt}f(\mathrm{Ad}^*(\exp\{-tX_i\})\eta)|_{t=0}.$$

This latter method is of course impractical for groups of arbitrary dimension and we had to consider another approach.

Each simultaneous level set of the invariant functions contains one or several orbits of the co-adjoint representation of G. Therefore an orbit is completely or partially determined by the corresponding values of the invariant functions on the orbit. Our approach consists of adopting the opposite point of view and of extracting the invariant polynomial functions from the knowledge of the orbits.

Let us mention that the effectiveness of such an approach depends essentially on the particular structure of the co-adjoint representation of the inhomogeneous groups which we present in the next section.

2. CO-ADJOINT REPRESENTATION OF AN INHOMOGENEOUS LIE GROUP

Let G be a Lie group, A a real vector space, and T a representation of G in the space A. We consider the inhomogeneous group

$$\mathrm{H}=\mathrm{A}\square\mathrm{G}$$

with multiplication

$$(a,g)(a',g')=(a+T(g)a',gg'). \tag{6}$$

Clearly

$$e_{\mathrm{H}}=(0,e_G), \quad (a,g)^{-1}=(-T(g)^{-1}a,g^{-1}).$$

As usual, the adjoint representation of H is defined by the derivative

$$\mathrm{Ad}(a,g)=\Delta(a,g)_*(e_H),$$

where $\Delta(a,g)$ denotes the inner automorphism of H:

$$\Delta(a,g)(b,k)=(a,g)(b,k)(a,g)^{-1}.$$

Since A is an abelian group, the Lie algebra $\mathcal{h}$ of H is of the form $\mathcal{h}=\mathrm{A}\times\mathcal{g}$ where $\mathcal{g}$ is the Lie algebra of G. Let us denote by $t=T_*(e_G)$: $\mathcal{g}\to\mathrm{End}(\mathrm{A})$ the associated representation of $\mathcal{g}$ and by Ad_G the adjoint representation of G; then we obtain easily

$$\mathrm{Ad}(a,g)(\xi,X)=(T(g)\xi-t(\mathrm{Ad}_G(g)X)a,\mathrm{Ad}_G(g)X), \tag{7}$$

where $(\xi,X)\in\mathrm{A}\times\mathcal{g}$. In the important case where $\mathrm{G}\subseteq\mathrm{Gl}(\mathrm{A})$, we have $\mathcal{g}\subseteq\mathrm{gl}(\mathrm{A})$ and $\mathrm{Ad}_G(g)X=gXg^{-1}$; if in addition we set $T(g)a=ga$ and $t(X)a=Xa$, then (7) becomes

$$\mathrm{Ad}(a,g)(\xi,X)=(g\xi-gXg^{-1}a,gXg^{-1}). \tag{8}$$

Let $\mathcal{h}^*=A^*\times\mathcal{g}^*$ be the dual space of $\mathcal{h}$ with the scalar product

$$\langle(p,M),(\xi,X)\rangle=\langle p,\xi\rangle_1+\langle M,X\rangle_2, \quad (p,M)\in A^*\times\mathcal{g}^*.$$

Since

$$\mathrm{Ad}(a,g)^{-1}(\xi, X) = (T(g)^{-1}\xi + t(\mathrm{Ad}_G(g^{-1})X) \times T(g^{-1})a, \mathrm{Ad}_G(g^{-1})X)$$

and

$$t(\mathrm{Ad}_G(g^{-1})X) = T(g)^{-1}t(X)T(g),$$

we have

$$\langle \mathrm{Ad}^*(a,g)(p,M),(\xi, X)\rangle = \langle (p,M),(T(g)^{-1}(\xi + t(X)a),\mathrm{Ad}_G(g)^{-1}X)\rangle = \langle p,T(g)^{-1}(\xi + t(X)a)\rangle_1 + \langle M,\mathrm{Ad}_G(g)^{-1}X\rangle_2.$$

By introducing the co-adjoint representation Ad^*_G of G and the contragredient representation T^* of T defined by

$$\langle T^*(g)p,\eta\rangle_1 = \langle p,T(g)^{-1}\eta\rangle_1,$$

we can write

$$\langle \mathrm{Ad}^*(a,g)(p,M),(\xi, X)\rangle = \langle T^*(g)p,\xi\rangle_1 + \langle T^*(g)p, \times t(X)a\rangle_1 + \langle \mathrm{Ad}^*_G(g)M, X\rangle_2.$$

Furthermore, since the mapping

$$X \mapsto \langle T^*(g)p,t(X)a\rangle_1$$

is linear, there exists a unique element $L(T^*(g)p,a)$ belonging to $\mathcal{g}^*$ such that

$$\langle T^*(g)p,t(X)a\rangle_1 = \langle L(T^*(g)p,a), X\rangle_2. \tag{9}$$

Therefore, the co-adjoint representation of H is given by

$$\mathrm{Ad}^*(a,g)(p,M) = (T^*(g)p,\mathrm{Ad}^*_G(g)M + L(T^*(g)p,a)). \tag{10}$$

3. STRUCTURE OF THE AD*-ORBITS

First let us fix the notations: if K is a group acting on a set E, we denote by $O_K(q)$ the orbit of the point $q \in E$ generated by K.

Directly from (10) we see that H acts on A^* through the representation T^* of G. Let us denote by $\mathrm{H}(p) = A \square G(p)$ the stabilizer of the point $p \in A^*$ where $G(p)$ is its little group. Since $T^*(g)p = p$ for any $g \in G(p)$, we see from (10) again that $H(p)$ acts on $\mathcal{g}^*$ as

$$M \mapsto \mathrm{Ad}^*_G(g)M + L(p,a). \tag{11}$$

The orbit $O_H(p_0,M_0)$ is provided with a natural fiber bundle structure

$$\Pi: O_H(p_0,M_0) \to O_H(p_0) \tag{12}$$

with

$$\Pi(p,M) = p, \quad \Pi^{-1}[p] = O_{H(p)}(M)$$

and with $O_{H(p_0)}(M_0)$ as standard fiber.

Indeed, since

$$O_H(p_0) \simeq H/H(p_0) \simeq G/G(p_0),$$

for each point $p \in O_H(p_0)$, there exist a neighborhood $\mathcal{V}_\alpha$ open in $O_H(p_0)$, a mapping $\Lambda_\alpha: \mathcal{V}_\alpha \to G$ such that

$$T^*(\Lambda_\alpha(p))p_0 = p$$

and a local trivialization

$$\varphi_\alpha: O_H(p_0,M_0)|_{\mathcal{V}_\alpha} \to \mathcal{V}_\alpha \times O_{H(p_0)}(M_0)$$

defined by

$$\varphi_\alpha(p,M) = (p,\mathrm{Ad}^*_G(\Lambda_\alpha(p))^{-1}M). \tag{13}$$

We remark that φ_α is actually defined on the whole manifold $\mathcal{V}_\alpha \times \mathcal{g}^*$.

Let us study the structure of the standard fiber $O_{H(p_0)}(M_0)$. We shall show that it is a cylinder in $\mathcal{g}^*$ whose directrix is diffeomorphic to an $\mathrm{Ad}^*_{G(p_0)}$-orbit in $\mathcal{g}(p_0)^*$ and whose generatrix is equal to the vector subspace $\mathrm{Ker}\, i(p_0)^*$, where $i(p_0)^*: \mathcal{g}^* \to \mathcal{g}(p_0)^*$ denotes the canonical surjection associated to the canonical injection $i(p_0): \mathcal{g}(p_0) \to \mathcal{g}$ of the Lie algebra of the little group $G(p_0)$ in $\mathcal{g}$.

Let us first introduce the linear mapping $L(p_0): A \to \mathcal{g}^*$ defined by

$$L(p_0)a = L(p_0,a)$$

that is, according to (9):

$$\langle L(p_0)a, X\rangle_2 = \langle p_0,t(X)a\rangle_1 = \langle -t^*(X)p_0,a\rangle_1,$$

where $t^*: \mathcal{g} \to \mathrm{End}(A^*)$ is the contragredient representation of t.

The dual mapping $L(p_0)^*: \mathcal{g} \to A^*$ is thus given by

$$L(p_0)^*X = -t^*(X)p_0 \tag{14}$$

and we have

Proposition 2:

$$\mathrm{Im}\, L(p_0) = \mathrm{Ker}\, i(p_0)^*. \tag{15}$$

Proof: The orbit $O_H(p_0)$ is generated by the T^*-action of G, hence the tangent space at the point p_0 of this orbit is, by (14), equal to $\mathrm{Im}\, L(p_0)^*$ ($\mathcal{g}^*$ being a vector space, we identify the tangent space $T_{p_0}\mathcal{g}^*$ with $\mathcal{g}^*$). Since in addition $O_H(p_0) \simeq G/G(p_0)$, then $\mathrm{Im}\, i(p_0) = \mathrm{Ker} L(p_0)^*$ which means that the sequence

$$0 \to \mathcal{g}(p_0) \xrightarrow{i(p_0)} \mathcal{g} \xrightarrow{L(p_0)^*} A^*$$

is exact. Therefore the dual sequence

$$0 \leftarrow \mathcal{g}(p_0)^* \xleftarrow{i(p_0)^*} \mathcal{g}^* \xleftarrow{L(p_0)} A$$

is also exact, which proves the proposition. □

This proposition, together with the well-known equivariance property of the canonical surjection $i(p_0)^*$,

$$i(p_0)^* \circ \mathrm{Ad}^*_G(k) = \mathrm{Ad}^*_{G(p_0)}(k) \circ i(p_0)^* \quad \forall \quad k \in G(p_0), \tag{16}$$

leads to the following fundamental proposition.

Proposition 3:

$$O_{H(p_0)}(M_0) \simeq O_{G(p_0)}(i(p_0)^*M_0) \times \mathrm{Ker}\, i(p_0)^*.$$

Proof: From the action (11) of $H(p_0)$ in $\mathcal{g}^*$, we see that a point M belongs to the orbit $O_{H(p_0)}(M_0)$ together with all the points $M + L(p_0,a)$, $a \in A$. Let $\mathcal{K}^*$ be any complementary subspace of $\mathrm{Ker}\, i(p_0)^*$ in $\mathcal{g}^*$; then it follows from the equality (15) that the orbit $O_{H(p_0)}(M_0)$ is a cylinder having $\mathrm{Ker}\, i(p_0)^*$ as generatrix and $O_{H(p_0)}(M_0) \cap \mathcal{K}^*$ as directrix. By taking into account the equivariance property (16), it follows immediately that this generatrix is diffeomorphic to the $G(p_0)$-orbit in $\mathcal{g}(p_0)^*$ of the image point $i(p_0)^*M_0$. □

By this proposition we remark that the orbit $O_H(p_0,M_0)$ is completely characterized by the orbit $O_H(p_0)$ in A^* and by the orbit $O_{G(p_0)}(i(p_0)^*M_0)$ in $\mathcal{g}(p_0)^*$. This result is the geometric analog of the one obtained by Wigner-Mackey's induction method applied to the search for unitary irreducible representations of the semidirect product $H = A \square G$.[9]

4. INVARIANT POLYNOMIAL FUNCTIONS ON $\mathcal{h}^*$

In this section we discuss the invariant polynomial functions in the light of the preceding orbit analysis.

We remark immediately from the co-adjoint action (10) that there exist invariant functions of the form $f = h \circ \Pi$ where $\Pi(p,M) = p$ and h is a H-invariant function on A^*. Of course, such nontrivial functions can exist only if there is no orbit $O_H(p_0)$ dense in A^*.

In order to study the structure of the other possible invariant functions, let us first consider an open covering $\{\mathcal{V}_\alpha\}$ of the orbit $O_H(p_0)$ and introduce the following local group actions on $\mathcal{V}_\alpha \times \mathcal{g}^*$ [see (13)]

$$S_{\beta\alpha}(a,g) = \varphi_\beta \circ \mathrm{Ad}^*(a,g) \circ \varphi_\alpha^{-1},$$

where the elements $g \in G$ are such that $T^*(g): \mathcal{V}_\alpha \to \mathcal{V}_\beta$. According to (9), (10), and (13) we obtain easily

$$S_{\beta\alpha}(a,g)(p,M) = (T^*(g)p, \mathrm{Ad}^*_G(k_{\beta\alpha}(p,g))M + L(p_0, T(\Lambda_\beta(T^*(g)p))a)), \tag{17}$$

where

$$k_{\beta\alpha}(p,g) = \Lambda_\beta(T^*(g)p)^{-1} g \Lambda_\alpha(p) \in G(p_0).$$

Note that H acts on $\mathcal{g}^*$ through the $H(p_0)$-action (11).

Now, let f be an invariant function on $\mathcal{h}^*$; since $(f \circ \varphi_\alpha^{-1}) \circ S_{\alpha\alpha}(a,g) = f \circ \mathrm{Ad}^*(a,g) \circ \varphi_\alpha^{-1} = f \circ \varphi_\alpha^{-1}$, the function $f \circ \varphi_\alpha^{-1}$ is locally invariant on $\mathcal{V}_\alpha \times \mathcal{g}^*$. Since G is acting transitively on $\mathcal{V}_\alpha$, $f \circ \varphi_\alpha^{-1}$ does not depend on the first variable; in other words it is of the form $f \circ \varphi_\alpha^{-1} = h$ where, according to (17), h is a $H(p_0)$-invariant function on $\mathcal{g}^*$. Moreover, by Proposition 3, any $H(p_0)$-invariant function on $\mathcal{g}^*$ factorizes as $h = \psi \circ i(p_0)^*$, where ψ is some $G(p_0)$-invariant function on $\mathcal{g}(p_0)^*$. Hence $f \circ \varphi_\alpha^{-1} = \psi \circ i(p_0)^*$ and therefore, according to (13), we have

$$f(p,M) = \psi(i(p_0)^* \mathrm{Ad}_G(\Lambda_\alpha(p)^{-1})M), \quad \forall (p,M) \in \mathcal{V}_\alpha \times \mathcal{g}^*. \tag{18}$$

Thus, to any invariant function f on $\mathcal{h}^*$ corresponds a unique invariant function ψ on $\mathcal{g}(p_0)^*$ for each chosen orbit $O_H(p_0)$ in A^* and we remark that ψ is constant if and only if $f = h \circ \Pi$.

However, we have to solve the problem in the opposite direction, that is, to reconstruct f starting from the functions ψ. In general this is a difficult problem for which very few specific results are available. Nevertheless it is manageable if we know in advance that f is a polynomial function because such a function is completely determined by one of its restrictions to an open neighborhood of a point. For our considered groups such a neighborhood can always be found. In the cases where an orbit $O_H(p_0)$ open in A^* exists, $\mathcal{V}_\alpha$ is an open set in A^*; in the cases where an open stratum S of generic orbits $O_H(p_0)$ exists, it turns out that it is always possible to find an open set $\mathcal{V}_\alpha$ in S and a mapping $\Lambda_\alpha: \mathcal{V}_\alpha \to G$ such that $T^*(\Lambda_\alpha(p))^{-1} p = p_0 \in \mathcal{V}_\alpha$ is smooth and constant on $O_H(p) \cap \mathcal{V}_\alpha$. Hence in all cases the functions [see (18)]

$$f_\alpha(p,M) = \psi(i(p_0)^* \mathrm{Ad}^*_G(\Lambda_\alpha(p)^{-1})M) \tag{19}$$

are defined on the open set $\mathcal{V}_\alpha \times \mathcal{g}^*$ in $\mathcal{h}^*$.

Let us remark that the function f, whose local restriction is f_α, is well defined, that is, does not depend on the chosen set $\mathcal{V}_\alpha$. Indeed, if $p \in \mathcal{V}_\alpha \cap \mathcal{V}_\beta$, then it follows from the fact that $\Lambda_\alpha(p)^{-1}\Lambda_\beta(p) \in G(p_0)$, from the equivariance property of $i(p_0)^*$, and from the $G(p_0)$-invariance of ψ that $f_\alpha(p,M) = f_\beta(p,M)$.

Another difficulty still remains: it may happen that neither $\Lambda_\alpha(p)$ nor $i(p_0)^* \mathrm{Ad}^*_G(\Lambda_\alpha(p)^{-1})M$ are polynomial functions of p. The problem is to find a judicious invariant function ψ in order to obtain a polynomial function $f_\alpha(p,M)$.

Example: We consider the well-known case of the Euclidean group $\mathcal{E}(3) = \mathbb{R}^3 \square O(3)$. According to (8), its adjoint representation is given by

$$\mathrm{Ad}(\mathbf{a},S)(\xi, X) = (S\xi - SXS^{-1}\mathbf{a}, SXS^{-1}),$$

where $\xi \in \mathbb{R}^3$ and X is a 3×3 antisymmetric matrix. By using the Lie algebra isomorphism $j: \mathbb{R}^3 \to o(3)$ with $j(\theta): \mathbf{v} \mapsto \theta \times \mathbf{v}$ and $[j(\theta), j(\theta')] = j(\theta \times \theta')$, the Lie algebra $e(3)$ can be realized as $\mathbb{R}^3 \times \mathbb{R}^3$ and, since $Sj(\theta)S^{-1} = j(\det(S)S\theta)$, we have

$$\mathrm{Ad}(\mathbf{a},S)(\xi,\theta) = (S\xi - \det(S)(S\theta) \times \mathbf{a}, \det(S)S\theta).$$

By setting $e(3)^* = \mathbb{R}^3 \times \mathbb{R}^3$ with the scalar product

$$\langle (\mathbf{p},\mathbf{l}),(\xi,\theta) \rangle = \mathbf{p} \cdot \xi + \mathbf{l} \cdot \theta$$

we obtain easily

$$\mathrm{Ad}^*(\mathbf{a},S)(\mathbf{p},\mathbf{l}) = (S\mathbf{p}, \det(S)(S\mathbf{l} + \mathbf{a} \times S\mathbf{p})).$$

The generic orbits in the first factor of $e(3)^*$ are spheres of radius $p = (\mathbf{p}^2)^{1/2}$ and they belong to the open stratum $\mathbb{R}^3_*$. Let $\mathbf{p}_0 = p\mathbf{e}_3$, $p > 0$, be a point in $\mathbb{R}^3_*$; its little group is $O(2)$. Since both adjoint and co-adjoint representations of $O(2)$ act as $\varphi \to \det(S)\varphi$, all even functions on $o(2)^* = \mathbb{R}$ are therefore invariant. Moreover we have $i(p_0)^*\varphi = \varphi \mathbf{e}_3$ and thus, $i(p_0)^* \mathbf{l} = \mathbf{e}_3 \cdot \mathbf{l}$. Let $\Lambda(\mathbf{p})$ denote a rotation such that $\Lambda(\mathbf{p})p\mathbf{e}_3 = \mathbf{p}$; since obviously $\Lambda(\mu\mathbf{p})\mathbf{e}_3 = \Lambda(\mathbf{p})\mathbf{e}_3$, it is not a polynomial function (it is not even a rational one for it depends explicitly on $p = (\mathbf{p}^2)^{1/2}$). Now, let us consider the invariant functions constructed according to the recipe (19):

$$f(\mathbf{p},\mathbf{l}) = \psi(i(p_0)^* \Lambda(\mathbf{p})^{-1}\mathbf{l}) = \psi(\mathbf{e}_3 \cdot \Lambda(\mathbf{p})^{-1}\mathbf{l}) = \psi(\Lambda(\mathbf{p})\mathbf{e}_3 \cdot \mathbf{l}) = \psi(\mathbf{p} \cdot \mathbf{l}/(\mathbf{p}^2)^{1/2}).$$

The argument of ψ is not a polynomial expression in $\mathbf{p}$; however, since $\mathbf{p} \to (\mathbf{p}^2)^{1/2}$ is an invariant function (of the form $f = h \circ \Pi$), we obtain all invariant polynomial functions (19) by choosing ψ as $\psi(\varphi) = \zeta((\mathbf{p}^2)^{1/2}\varphi)$ where ζ is any polynomial function of even degree on $\mathbb{R}$. Thus we recover the familiar result: the invariant polynomial functions on $e(3)^*$ are generated by the two generators

$$f_0(\mathbf{p},\mathbf{l}) = \mathbf{p}^2, \quad f_1(\mathbf{p},\mathbf{l}) = (\mathbf{p} \cdot \mathbf{l})^2.$$

It is not very useful for our needs to pursue this discussion without considering specific groups. Up to now, we have obtained some recipes to construct the generators of $P_H(\mathcal{h}^*)$ starting from those of $P_G(A^*)$ and of $P_{G(p_0)}(\mathcal{g}(p_0)^*)$ and we know that there are at most as many generators of

$P_H(\mathcal{h}^*)$ as there are generators of $P_G(A^*)$ and of $P_{G(p_0)}(\mathcal{g}(p_0)^*)$. In order to find such a set of generators, we adopt the following procedure:

(1) Write down the co-adjoint representation of H.

(2) Analyze the generic orbits $O_H(p_0)$ in A^* and determine the little group $G(p_0)$ of one of them.

(3) Find the generators of $P_G(A^*)$.

(4) Find the generators of $P_{G(p_0)}(\mathcal{g}(p_0)^*)$.

(5) Study the functions $\psi_j(i(p_0)^*\mathrm{Ad}_G^*(\Lambda(p)^{-1})M)$, where ψ_j is a generator of $P_{G(p_0)}(\mathcal{g}(p_0)^*)$, and, by inspection, determine the corresponding invariant polynomial functions on $\mathcal{h}^*$.

In the sequel, the (nonconstant) generators of $P_H(\mathcal{h}^*)$ will be called *fundamental invariants of H* and we do not write down the corresponding Casimir operators of $Z(\mathcal{h})$; it is sufficient to choose a basis of $\mathcal{h}$ and to use formula (4).

5. THE GENERAL AFFINE GROUP $\mathcal{A}(n,\mathbb{R}) = \mathbb{R}^n \square \mathrm{Gl}(n,\mathbb{R})$

The adjoint representation is given by (8), where $\xi \in \mathbb{R}^n$ and $X \in \mathcal{M}^n(\mathbb{R})$, the space of all $n \times n$ real matrices. By setting $a(n,R)^* = \mathbb{R}^n \times \mathcal{M}^N(\mathbb{R})$ with the scalar product

$$\langle (p,M),(\xi,X)\rangle = p^t\xi + \mathrm{tr}(M^tX),$$

then

$$\langle \mathrm{Ad}^*(a,g)(p,M),(\xi,X)\rangle = (T^*(g)p)^t a + (T^*(g)p)^t Xa + \mathrm{tr}(T^*(g)MT^*(g)^{-1}X), \quad (20)$$

where $T^*(g) = (g^t)^{-1}$. Since

$$(T^*(g)p)^tXa = \mathrm{tr}(a(T^*(g)p)^tX) = \mathrm{tr}(((T^*(g)p)a^t)^tX),$$

we obtain

$$\mathrm{Ad}^*(a,g)(p,M) = (T^*(g)p, T^*(g)MT^*(g)^{-1} + (T^*(g)p)a^t). \quad (21)$$

It is rather obvious that the group $\mathrm{Gl}(n,\mathbb{R})$ acts transitively on $\mathbb{R}^n_*$ through the representation T^* and it is easily verified that the little group of the point

$$p_0 = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \quad 0 \in \mathbb{R}^{n-1}$$

is equal to

$$G(p_0) = \left\{ \begin{bmatrix} A & b \\ 0 & 1 \end{bmatrix} \middle| A \in \mathrm{Gl}(n-1,\mathbb{R}), b \in \mathbb{R}^{n-1} \right\} \simeq \mathcal{A}(n-1,\mathbb{R}).$$

Since $\mathbb{R}^n_*$ is an orbit dense in $\mathbb{R}^n$, there is no nontrivial invariant function on $\mathbb{R}^n$; therefore there are at most as many fundamental invariants for $\mathcal{A}(n,\mathbb{R})$ as for $\mathcal{A}(n-1,\mathbb{R})$ and thus, as for $\mathcal{A}(1,\mathbb{R}) = \mathbb{R}\square\mathbb{R}_*$. For this latter group, the co-adjoint action (21) reduces to

$$\mathrm{Ad}^*(a,g)(p,m) = (g^{-1}p, m + g^{-1}pa), \quad g \in \mathbb{R}_*,\ a,p,m \in \mathbb{R}$$

and we see immediately that $\mathcal{A}(1,\mathbb{R})$ acts transitively on $\mathbb{R}_* \times \mathbb{R}$, which is an orbit dense in $a(1,\mathbb{R})^* = \mathbb{R} \times \mathbb{R}$. Hence there is no nontrivial invariant function on $a(1,\mathbb{R})^*$ and thus, we have proved

Proposition 4: The group $\mathcal{A}(n,\mathbb{R})$ has no fundamental invariant.[6]

6. THE SPECIAL AFFINE GROUP $\mathcal{SA}(n,\mathbb{R}) = \mathbb{R}^n \square \mathrm{Sl}(n,\mathbb{R})$

The adjoint representation is given by (8), where $\xi \in \mathbb{R}^n$ and $X \in \mathcal{M}_0^n(\mathbb{R})$, the space of all $n \times n$ real matrices of zero trace. By setting $sa(n,\mathbb{R})^* = \mathbb{R}^n \times \mathcal{M}_0^n(\mathbb{R})$ with the same scalar product as in the preceding case, we still obtain the formula (20) but now with X and M in $\mathcal{M}_0^n(\mathbb{R})$. However, the element $L(T^*(g)p,a)$ is not the same because $(T^*(g)p)a^t$ is in general not of zero trace. But if X is of zero trace and if N is not, we have

$$\mathrm{tr}(N^tX) = \mathrm{tr}(N_0^tX),$$

where

$$N_0 = N - (1/n)(\mathrm{tr}(N))I_n \in \mathcal{M}_0^n(\mathbb{R})$$

and thus

$$\mathrm{Ad}^*(a,g)(p,M) = (T^*(g)p, T^*(g)MT^*(g)^{-1} + (T^*(g)p)a^t - (1/n)((T^*(g)p)^ta)I_n). \quad (22)$$

In the same way as in the preceding case, we verify easily that $\mathrm{Sl}(n\mathbb{R})$ acts transitively on $\mathbb{R}^n_*$ and that the little group of the same point p_0 is equal to

$$G(p_0) = \left\{ \begin{bmatrix} A & b \\ 0 & 1 \end{bmatrix} | A \in \mathrm{Sl}(n-1,\mathbb{R}), b \in \mathbb{R}^{n-1} \right\} \simeq \mathcal{SA}(n-1,\mathbb{R}).$$

Again, we conclude that there are at most as many fundamental invariants for $\mathcal{SA}(n,\mathbb{R})$ as for $\mathcal{SA}(1,\mathbb{R}) = \mathbb{R}\square\{1\} = \mathbb{R}$. But since this latter group is now abelian, it follows, from the triviality of its co-adjoint representation, that all functions on $sa(1,R)^* = \mathbb{R}$ are invariant. The polynomial ones being generated by the identity function, it follows that there is at most one fundamental invariant for $\mathcal{SA}(n,\mathbb{R})$.

We claim that the following nonconstant polynomial function

$$\varphi_{(n)}(p,M) = \det(p,Mp,\ldots,M^{n-1}p)$$

is invariant. It is easily verified that, by the co-adjoint action (22), the terms M^kp are transformed into a linear combination of terms of the form $T^*(g)M^lp$ with $l = 0,\ldots,k$. Taking into account the total antisymmetry property of the determinant function, we obtain

$$\varphi_{(n)}(\mathrm{Ad}^*(a,g)(p,M)) = \det(T^*(g)p, T^*(g)Mp,\ldots,T^*(g)M^{n-1}p).$$

On the other hand, since $g \in \mathrm{Sl}(n,\mathbb{R})$, we have $\det(T^*(g)) = 1$ and, therefore,

$$\varphi_{(n)}(\mathrm{Ad}^*(a,g)(p,M)) = (\det(T^*(g)))\varphi_{(n)}(p,M) = \varphi_{(n)}(p,M).$$

The function $\varphi_{(n)}$ is of the form (19) and it is easily seen [by expressing p as $\Lambda(p)p_0$] that ψ is exactly $\varphi_{(n-1)}$. Then it follows from an induction argument that the function $\varphi_{(n)}$ cannot be expressed in terms of a unique invariant polynomial function of lower degree in M. Thus we have proved

Proposition 5: The unique fundamental invariant of $\mathcal{SA}(n,\mathbb{R})$ is

$$\varphi(p,M) = \det(p,Mp,\ldots,M^{n-1}p).$$

7. THE (PSEUDO)-EUCLIDEAN GROUP
$\mathscr{E}(p,q) = \mathbb{R}^n \square O(p,q)$

Of course $n = p + q$ and since $O(p,q) \simeq O(q,p)$, we can restrict ourselves to the cases $p \leqslant q$ (with $q \geqslant 2$). The adjoint representation is given by (8), where $\xi \in \mathbb{R}^n$ and $X \in o(p,q)$, which is the Lie algebra of all $n \times n$ real matrices satisfying

$$X^t I_{p,q} + I_{p,q} X = O_n \quad \text{with} \quad I_{p,q} = \begin{bmatrix} -I_p & 0 \\ 0 & I_q \end{bmatrix}.$$

By setting $e(p,q)^* = \mathbb{R}^n \times o(p,q)$ with the scalar product

$$\langle (p,M),(\xi,X)\rangle = p^t I_{p,q} \xi + \tfrac{1}{2} \operatorname{tr}(MX),$$

we obtain

$$\mathrm{Ad}^*(a,g)(p,M) = (gp, gMg^{-1} + L(gp,a)) \tag{23}$$

and it is not necessary to compute explicitly the element $L(gp,a)$.

The action of $O(p,q)$ on $\mathbb{R}^n$ is well known: the orbits are the quadrics given by the equation

$$Q(p) = p^t I_{p,q} p = \text{const},$$

and, except for the trivial one, they are of codimension one; in addition, for all cases of p and q, there is at least an open stratum, namely the stratum S containing the points

$$p_0 = \begin{bmatrix} 0 \\ m \end{bmatrix}, \quad 0 \in \mathbb{R}^{n-1}, \quad m > 0,$$

the little group of which being

$$G(p_0) = \left\{ \begin{bmatrix} A & 0 \\ 0 & 1 \end{bmatrix} \middle| A \in O(p,q-1) \right\} \simeq O(p,q-1).$$

We can conclude from the above that the invariant polynomial functions on $\mathbb{R}^n$ are generated by the quadratic form $Q(p)$. Furthermore we know (see Appendix) that the fundamental invariants of $O(p,q-1)$ are given by the characteristic coefficients $C_{2k}(N), k = 1,\ldots,[(n-1)/2]$, of the polynomial $\det(N - \lambda I_{n-1})$.

There remains to examine the invariant function (19). Let $\Lambda(p)$ be an element of $O(p,q)$ such that $\Lambda(p) p_0 = p$ for some p_0; since $\Lambda(p)\mu p_0 = \mu p$, we can put $\Lambda(\mu p) = \Lambda(p)$ and thus, the mapping $p \to \Lambda(p)$ is defined on the whole stratum S, but it is not a polynomial function. By virtue of (19) and (23) we have to examine the functions $C_{2k}(i(p_0)^*(\Lambda(p)^{-1} M \Lambda(p)))$ but instead of studying them separately we shall consider directly the function

$$\det(i(p_0)^*(\Lambda(p)^{-1} M \Lambda(p)) - \lambda I_{n-1}).$$

Contrarily to the matrix $i(p_0)^*(\Lambda(p)^{-1} M \Lambda(p))$, this function is not a complicated expression in p. In order to exhibit its simple dependence in p, let us introduce the projection operator

$$E(p_0) = \begin{bmatrix} I_{n-1} & 0 \\ 0 & 0 \end{bmatrix}$$

associated to the orthogonal complementary subspace of $\mathbb{R}p_0$. On one hand we have

$$E(p_0) N = \begin{bmatrix} i(p_0)^* N & * \\ 0 & 0 \end{bmatrix}$$

for any $N \in o(p,q)^*$ and, on the other hand, by setting

$$E(p) = \Lambda(p) E(p_0) \Lambda(p)^{-1},$$

we obtain the projection operator associated to the orthogonal complementary subspace of $\mathbb{R}p$; therefore it is given by

$$E(p) = I_n - (Q(p))^{-1} p p^t I_{p,q}$$

and we have the equality

$$\det(E(p)M - \lambda I_n) = -\lambda \det(i(p_0)^*(\Lambda(p)^{-1} M \Lambda(p)) - \lambda I_{n-1}).$$

Indeed, we obtain successively

$$\begin{aligned} &\det(E(p)M - \lambda I_n) \\ &\quad = \det(\Lambda(p) E(p_0) \Lambda(p)^{-1} M - \lambda I_n) \\ &\quad = \det(\Lambda(p)(E(p_0)\Lambda(p)^{-1} M \Lambda(p) - \lambda I_n)\Lambda(p)^{-1}) \\ &\quad = \det(E(p_0)\Lambda(p)^{-1} M \Lambda(p) - \lambda I_n) \\ &\quad = \det \begin{bmatrix} i(p_0)^*(\Lambda(p)^{-1} M \Lambda(p)) - \lambda I_{n-1} & * \\ 0 & -\lambda \end{bmatrix} \\ &\quad = -\lambda \det(i(p_0)^*(\Lambda(p)^{-1} M \Lambda(p)) - \lambda I_{n-1}). \end{aligned}$$

The expression $Q(p)E(p)M$ is a polynomial in p and since $Q(p)$ is invariant, $\det(Q(p)E(p)M - \lambda I_n)$ is an invariant polynomial function and we have

Proposition 6: There are $[(n+1)/2]$ fundamental invariants of $\mathscr{E}(p,q)$:

$$\begin{aligned} &\varphi_0(p,M) = Q(p) = p^t I_{p,q} p, \\ &\varphi_k(p,M) = C_{2k}((Q(p) I_n - p p^t I_{p,q}) M), \quad k = 1,\ldots,[(n-1)/2], \end{aligned}$$ [14]

where

$$\begin{aligned} &\det((Q(p) I_n - p p^t I_{p,q}) M - \lambda I_n) \\ &\quad = \sum_{l=0}^{n} C_l((Q(p) I_n - p p^t I_{p,q}) M) \lambda^{n-l}. \end{aligned}$$

8. THE (PSEUDO)-HERMITIAN GROUP
$\mathscr{H}(p,q) = \mathbb{C}^n \square U(p,q)$

Again $n = p + q$ and p,q can be restricted to the cases where $p \leqslant q$ and $n \geqslant 2$. The adjoint representation is still given by (8) where $\xi \in \mathbb{C}^n$ and $X \in u(p,q)$, which is the Lie algebra of all $n \times n$ complex matrices satisfying

$$X^\dagger I_{p,q} + I_{p,q} X = O_n.$$

The Lie algebra $h(p,q) = \mathbb{C}^n \times u(p,q)$ has to be considered as real. By introducing the real dual space $h(p,q)^* = \mathbb{C}^n \times u(p,q)$ equipped with the scalar product

$$\langle (p,M),(\xi,X)\rangle = \operatorname{Re}\{p^\dagger I_{p,q} \xi + \tfrac{1}{2} \operatorname{tr}(MX)\}$$

we obtain at once

$$\mathrm{Ad}^*(a,g)(p,M) = (gp, gMg^{-1} + L(gp,a)). \tag{24}$$

The search for the fundamental invariant of $\mathscr{H}(p,q)$ is exactly the same as the one of $\mathscr{E}(p,q)$. The generic orbits in $\mathbb{C}^n$ are the level sets (of real codimension one) of the quadratic form $Q(p) = p^\dagger I_{p,q} p$. There is always at least an open stratum, namely the one containing the points

$$p_0 = \begin{bmatrix} 0 \\ m \end{bmatrix}, \quad 0 \in \mathbb{C}^{n-1}, \quad m > 0$$

with little group

$$G(p_0)=\left\{\begin{bmatrix}A & 0\\ 0 & 1\end{bmatrix}\middle| A\in U(p,q-1)\right\}\simeq U(p,q-1).$$

The invariant functions on $\mathbb{C}^n$ are generated by the quadratic form $Q(p)$ and the fundamental invariants of $U(p,q-1)$ are given by the characteristic coefficients $C_l(\sqrt{-1}N), l=1,\ldots,n-1$ of the polynomial $\det(\sqrt{-1}N-\lambda I_{n-1})$.

Exactly in the same way as in the preceding section, we can prove

Proposition 7: There are n fundamental invariants of $\mathscr{H}(p,q)$

$$\varphi_0(p,M)=Q(p)=p^{\dagger}I_{p,q}p,$$
$$\varphi_k(p,M)=C_k(\sqrt{-1}(Q(p)I_n - pp^{\dagger}I_{p,q})M),\quad k=1,\ldots,n-1,^{14}$$

where

$$\det(\sqrt{-1}(Q(p)I_n - pp^{\dagger}I_{p,q})M-\lambda I_n)$$
$$=\sum_{l=0}^{n}C_l(\sqrt{-1}(Q(p)I_n - pp^{\dagger}I_{p,q})M)\lambda^{n-l}.$$

Remark: For the subgroup $\mathbb{C}^n\square SU(p,q)$, nothing is changed except for the invariant $\varphi_1(p,M)$ which is identically zero.

9. THE GROUP $St(n,\mathbb{R})=\mathbb{R}\square(R^{2n}\square Sp(n,\mathbb{R}))$

This group is not a classical one, but it appears in the next section as the stabilizer in $Sp(n+1,\mathbb{R})$ of a point in $\mathbb{R}^{2(n+1)}$ and we shall need its fundamental invariants. It can be defined as the following central extension of $\mathbb{R}^{2n}\square Sp(n,\mathbb{R})$ by the abelian group $\mathbb{R}$:

$$(\varphi,b,h)(\varphi',b',h')=(\varphi+\varphi'+b^tJ_nhb',b+hb',hh'),$$

where

$$\varphi,\varphi'\in\mathbb{R};\quad b,b'\in\mathbb{R}^{2n};\quad h,h'\in Sp(n,\mathbb{R});\quad J_n=\begin{bmatrix}J_1 & & \\ & \ddots & \\ & & J_1\end{bmatrix};\quad J_1=\begin{bmatrix}0 & -1\\ 1 & 0\end{bmatrix}.$$

Its Lie algebra is realized by the space $st(n,\mathbb{R})=\mathbb{R}\times\mathbb{R}^{2n}\times sp(n,\mathbb{R})$ and it is easily found that the adjoint representation is given by

$$\mathrm{Ad}(\varphi,b,h)(\xi,\eta,Y)=(\xi+2b^tJ_nh\eta-b^tJ_nhYh^{-1}b,\ h\eta-hYh^{-1}b,hYh^{-1}).$$

By setting $st(n,\mathbb{R})^*=\mathbb{R}\times\mathbb{R}^{2n}\times sp(n,\mathbb{R})$ with the scalar product

$$\langle(\mu,q,N),(\xi,\eta,Y)\rangle=\tfrac{1}{2}\mu\xi-q^tJ_n\eta+\tfrac{1}{2}\operatorname{tr}(NY),\qquad(25)$$

we obtain without difficulty

$$\mathrm{Ad}^*(\varphi,b,h)(\mu,q,N)=(\mu,hq+\mu b,hNh^{-1}-(b(hq)^t+hqb^t+\mu bb^t)J_n).$$

In order to determine the fundamental invariants of this group, it is sufficient to restrict ourselves to the open dense subset of $St(n,\mathbb{R})^*$ where $\mu\neq0$. On this subset we perform the change of variables

$$F(\mu,q,N)=(\mu,q,L=N+(1/\mu)qq^tJ_n),$$
$$F^{-1}(\mu,q,L)=(\mu,q,N=L-(1/\mu)qq^tJ_n);$$

then the modified action

$$\tilde{\mathrm{Ad}}^*(\varphi,b,h)=F\circ\mathrm{Ad}^*(\varphi,b,h)\circ F^{-1}$$

is simply equal to

$$\tilde{\mathrm{Ad}}^*(\varphi,b,h)(\mu,q,L)=(\mu,hq+\mu b,hLh^{-1}),\quad\mu\neq0.$$

We see immediately that the orbit of a point $(\mu_0\neq0,q_0,L_0)$ is $\{\mu_0\}\times\mathbb{R}^{2n}\times O_{Sp(n,R)}(L_0)$: it is a simultaneous level set of the $n+1$ functions

$$\tilde{\psi}_0(\mu,q,L)=\mu,$$
$$\tilde{\psi}_k(\mu,q,L)=C_{2k}(L),\quad k=1,\ldots,n,$$

where (see Appendix) $C_{2k}(L)$ is a characteristic coefficient of the polynomial $\det(L-\lambda I_{2n})$. Returning to the old variables we obtain *the* $n+1$ *fundamental invariants of* $St(n,\mathbb{R})$:

$$\psi_0(\mu,q,N)=\mu,$$
$$\psi_k(\mu,q,N)=C_{2k}(\mu N+qq^tJ_n),\quad k=1,\ldots,n.^{14}\qquad(26)$$

10. THE AFFINE CANONICAL GROUP $Can(n)=\mathbb{R}^{2n}\square Sp(n,\mathbb{R})$

Since $Sp(1,\mathbb{R})=Sl(2,\mathbb{R})$ we suppose that $n\geqslant2$. The adjoint representation is given by (8) where $\xi\in\mathbb{R}^{2n}$ and $X\in sp(n,\mathbb{R})$, which is the Lie algebra of all $2n\times2n$ real matrices satisfying

$$X^tJ_n+J_nX=O_{2n}.$$

By setting $can(n)^*=\mathbb{R}^{2n}\times sp(n,\mathbb{R})$ with the scalar product

$$\langle(p,M),(\xi,X)\rangle=p^tJ_n\xi+\tfrac{1}{2}\operatorname{tr}(MX),$$

we obtain

$$\mathrm{Ad}^*(a,g)(p,M)=(gp,gMg^{-1}+L(gp,a)).\qquad(27)$$

The group $Sp(n,\mathbb{R})$ acts transitively on $\mathbb{R}^{2n}_*$, but for our needs it is sufficient to show that there is an orbit of dimension $2n$ in $\mathbb{R}^{2n}$. This result follows at once from a dimension argument since, as is easily verified, the little group of the point

$$p_0=\begin{bmatrix}0\\ 1\end{bmatrix},\quad 0\in\mathbb{R}^{2n-1}$$

is equal to

$$G(p_0)=\left\{\begin{bmatrix}h & b & 0\\ 0 & 1 & 0\\ b^tJ_{n-1}h & \varphi & 1\end{bmatrix}\right\}\simeq St(n-1,\mathbb{R})\qquad(28)$$

where

$$h\in \mathrm{Sp}(n-1,\mathbb{R}),\quad b\in\mathbb{R}^{2(n-1)},\quad \varphi\in\mathbb{R}.$$

Hence, there is no invariant polynomial function on $\mathbb{R}^{2n}$ and the fundamental invariants of $G(p_0)$ are given by (26).

There remains to examine the functions (19), that is the functions $\psi_l(i(p_0)^*(\Lambda(p)^{-1}M\Lambda(p)))$, $l=0,\ldots,n-1$. For the same reasons as in the case of the (pseudo)-Euclidean group it is advantageous to study directly the function $\det(i(p_0)^*(\Lambda(p)^{-1}M\Lambda(p))-\lambda I_{2(n-1)})$.

First of all we identify the group $\mathrm{St}(n-1,\mathbb{R})$ with its matrix representation (28); then its Lie algebra $\mathrm{st}(n-1,\mathbb{R})$ is given by the subalgebra of $sp(n,\mathbb{R})$ consisting of all matrices satisfying the condition

$$Zp_0=0,$$

that is

$$Z=\begin{bmatrix} Y & \eta & 0\\ 0 & 0 & 0\\ \eta^t J_{n-1} & \zeta & 0\end{bmatrix},$$

$$Y\in sp(n-1,\mathbb{R}),\quad \eta\in\mathbb{R}^{2(n-1)},\quad \zeta\in\mathbb{R}. \tag{29}$$

Furthermore, we can define the dual space $st(n-1,\mathbb{R})^*$ as the subspace of $sp(n,\mathbb{R})$ consisting of all matrices of the form

$$K=\begin{bmatrix} N & 0 & q\\ -q^t J_{n-1} & 0 & \mu\\ 0 & 0 & 0\end{bmatrix},$$

$$N\in sp(n-1,\mathbb{R}),\quad q\in\mathbb{R}^{2(n-1)},\quad \mu\in\mathbb{R}, \tag{30}$$

since, in accordance with (25), we have

$$\langle K,Z\rangle=\tfrac{1}{2}\,\mathrm{tr}(KZ)=\tfrac{1}{2}\mu\zeta-q^t J_{n-1}\eta+\tfrac{1}{2}\,\mathrm{tr}(NY).$$

By setting

$$\Lambda(p)^{-1}M\Lambda(p)=\begin{bmatrix} N & p & q\\ -q^t J_{n-1} & \alpha & \mu\\ p^t J_{n-1} & \gamma & -\alpha\end{bmatrix}$$

[where the right-hand side is a block decomposition of a matrix belonging to $sp(n,\mathbb{R})=sp(n,\mathbb{R})^*$], we have

$$i(p_0)^*(\Lambda(p)^{-1}M\Lambda(p))=\begin{bmatrix} N & 0 & q\\ -q^t J_{n-1} & 0 & \mu\\ 0 & 0 & 0\end{bmatrix}$$

and, according to (26), the fundamental invariants of $\mathrm{Can}(n)$ are given by $\varphi_0(p,M)=\mu$ and by the nonzero characteristic coefficients of the polynomial $\det(\mu N+qq^t J_{n-1}-\lambda I_{2(n-1)})$. In order to exhibit the explicit form of these functions, some matrix computations are needed.

First we remark that for any element $L\in sp(n,\mathbb{R})$, the matrix

$$\rho(L)=\nu L+(Lp_0)(Lp_0)^t J_n$$

with

$$\nu=-(Lp_0)^t J_n p_0$$

satisfies obviously the equation $\rho(L)p_0=0$, that is, belongs to $st(n-1,\mathbb{R})$. In particular, according to (30), we have

$$\rho(K)=\begin{bmatrix} \mu N+qq^t J_{n-1} & 0 & 0\\ 0 & 0 & 0\\ 0 & 0 & 0\end{bmatrix},\quad \mu=\nu.$$

In addition we verify easily the property

$$\rho(i(p_0)^*L)=i(p_0)^*\rho(L).$$

Now, since according to (29)

$$\det(Z-\lambda I_{2n})=\det(i(p_0)^*Z-\lambda I_{2n})$$
$$=\lambda^2\det(Y-\lambda I_{2(n-1)}),$$

the following equality holds

$$\lambda^2\det(\mu N+qq^t J_{n-1}-\lambda I_{2(n-1)})$$
$$=\det(\rho(\Lambda(p)^{-1}M\Lambda(p))-\lambda I_{2n}).$$

Finally, since

$$\rho(\Lambda(p)^{-1}M\Lambda(p))=\nu\Lambda(p)^{-1}M\Lambda(p)$$
$$+(\Lambda(p)^{-1}M\Lambda(p)p_0)$$
$$\times(\Lambda(p)^{-1}M\Lambda(p)p_0)^t J_n$$
$$=\Lambda(p)^{-1}(\nu M+(Mp)(Mp)^t J_n)\Lambda(p)$$

and

$$\nu=-(\Lambda(p)^{-1}M\Lambda(p)p_0)^t J_n p_0=-p^t M^t J_n p$$
$$=p^t J_n Mp,$$

then

$$\det(\rho(\Lambda(p)^{-1}M\Lambda(p))-\lambda I_{2n})$$
$$=\det(\Lambda(p)^{-1}(\nu M+(Mp)(Mp)^t J_n$$
$$-\lambda I_{2n})\Lambda(p))$$
$$=\det(\nu M+(Mp)(Mp)^t J_n-\lambda I_{2n})$$

and we have proved

Proposition 8: There are n fundamental invariants of $\mathrm{Can}(n)$:

$$\varphi_0(p,M)=\nu=p^t J_n Mp,$$
$$\varphi_k(p,M)=C_{2k}(\nu M+(Mp)(Mp)^t J_n),$$
$$k=1,\ldots,n-1,$$[14]

where

$$\det(\nu M+(Mp)(Mp)^t J_n-\lambda I_{2n})$$
$$=\sum_{l=0}^{2n}C_l(\nu M+(Mp)(Mp)^t J_n)\lambda^{2n-l}.$$

11. THE COMPLEX INHOMOGENEOUS GROUPS

The following four complex inhomogeneous classical groups $\mathscr{A}(n,\mathbb{C})=\mathbb{C}^n\square\mathrm{Gl}(n,\mathbb{C})$, $\mathscr{SA}(n,\mathbb{C})=\mathbb{C}^n\square\mathrm{Sl}(n,\mathbb{C})$, $\mathscr{E}(n,C)=\mathbb{C}^n\square\mathrm{O}(n,\mathbb{C})$ and $\mathrm{Can}(n,\mathbb{C})=\mathbb{C}^{2n}\square\mathrm{Sp}(n,\mathbb{C})$ are nothing else but the complexifications of the four real groups $\mathscr{A}(n,\mathbb{R})$, $\mathscr{SA}(n,\mathbb{R})$, $\mathscr{E}(O,n)$ and $\mathrm{Can}(n)$. For this reason, their fundamental invariants are immediately known from those of their real forms by complexification. We will only give the results.

1. The general affine group $\mathscr{A}(n,\mathbb{C}) = \mathbb{C}^n \square \mathrm{Gl}(n,\mathbb{C})$

No fundamental invariant.

2. The special affine group $\mathscr{SA}(n,\mathbb{C}) = \mathbb{C}^n \square \mathrm{Sl}(n,\mathbb{C})$

Two fundamental invariants: the real and imaginary part of

$$\varphi(p,M) = \det(p, Mp, \ldots, M^{n-1}p),$$

where $sa(n,\mathbb{C})^* = sa(n,\mathbb{C}) = \mathbb{C}^n \times sl(n,\mathbb{C})$ with the scalar product

$$\langle (p,M),(\xi,X)\rangle = \mathrm{Re}\{p^t\xi + \mathrm{tr}(M^tX)\}.$$

3. The Euclidean group $\mathscr{E}(n,\mathbb{C}) = \mathbb{C}^n \square O(n,\mathbb{C})$

$2[(n+1)/2]$ fundamental invariants: the real and imaginary part of

$$\varphi_0(p,M) = Q(p) = p^tp,$$
$$\varphi_k(p,M) = C_{2k}((p^tpI_n - pp^t)M),$$
$$k = 1,\ldots,[(n-1)/2],$$

where

$$\det((p^tpI_n - pp^t)M - \lambda I_n) = \sum_{l=0}^{n} C_l((p^tpI_n - pp^t)M)\lambda^{n-l}$$

and $e(n,\mathbb{C})^* = e(n,\mathbb{C}) = \mathbb{C}^n \times o(n,\mathbb{C})$ with the scalar product

$$\langle (p,M),(\xi,X)\rangle = \mathrm{Re}\{p^t\xi + \tfrac{1}{2}\mathrm{tr}(MX)\}.$$

4. The affine canonical group $\mathrm{Can}(n,\mathbb{C}) = \mathbb{C}^{2n} \square \mathrm{Sp}(n,\mathbb{C})$

$2n$ fundamental invariants: the real and imaginary part of

$$\varphi_0(p,M) = \nu = p^tJ_nMp,$$
$$\varphi_k(p,M) = C_{2k}(\nu M + (Mp)(Mp)^tJ_n),$$
$$k = 1,\ldots,n-1,$$

where

$$\det(\nu M + (Mp)(Mp)^tJ_n - \lambda I_{2n}) = \sum_{l=0}^{2n} C_l(\nu M + (Mp)(Mp)^tJ_n)\lambda^{2n-l}$$

and $can(n,\mathbb{C})^* = can(n,\mathbb{C}) = \mathbb{C}^{2n} \times sp(n,\mathbb{C})$ with the scalar product

$$\langle (p,M),(\xi,X)\rangle = \mathrm{Re}\{p^tJ_n\xi + \tfrac{1}{2}\mathrm{tr}(MX)\}.$$

12. THE QUATERNIONIC INHOMOGENEOUS GROUPS

In this section we consider the natural inhomogeneous groups associated to the remaining four real forms of the classical complex groups (see Appendix): $U^*(2n)$, $SU^*(2n)$, $O^*(2n)$ and $Sp(p,q)$, $p+q=n$. All these groups consist of regular $2n \times 2n$ complex matrices satisfying the condition

$$J_n\bar{g}J_n^{-1} = g, \quad J_n = \begin{bmatrix} J_1 & & \\ & \ddots & \\ & & J_1 \end{bmatrix}, \quad J_1 = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}.$$

It follows that they can be completely decomposed into n^2 2×2 blocks $g_{\alpha\beta}$ subject to the condition

$$J_1\bar{g}_{\alpha\beta}J_1^{-1} = g_{\alpha\beta};$$

in other words, they are of the form

$$g_{\alpha\beta} = \begin{bmatrix} a & b \\ -\bar{b} & \bar{a} \end{bmatrix}, \quad a,b \in \mathbb{C}$$

and we recognize the well-known complex representation of a quaternion.[15] Therefore, these four groups can be realized by $n\times n$ matrices with quaternionic entries and thus have a natural action on the n-dimensional arithmetic vector space $\mathbb{H}^n$ over the quaternion field $\mathbb{H}$. Hence we are lead to consider the inhomogeneous groups $\mathbb{H}^n \square U^*(2n)$, $\mathbb{H}^n \square SU^*(2n)$, $\mathbb{H}^n \square O^*(2n)$, $\mathbb{H}^n \square \mathrm{Sp}(p,q)$ and it turns out that they are the quaternionic equivalents of the real groups $\mathscr{A}(n,\mathbb{R})$, $\mathscr{SA}(n,\mathbb{R})$, $\mathscr{E}(0,n)$, $\mathscr{H}(0,n)$, respectively.

Up to this change of the number field, the computations are very similar to the ones for the real cases and it will be enough to give a summary of the treatment. In order to facilitate the computations with the invariant polynomial functions and with the corresponding Casimir operators, we identify $\mathbb{H}^n$ with $\mathbb{C}^{2n}$ in the following way:

$$P \leftrightarrow p, \quad P = \begin{bmatrix} P_1 \\ \vdots \\ P_n \end{bmatrix} = [\,-J_n\bar{p}p\,],$$

$$P_i = \begin{bmatrix} \bar{q}_i & p_i \\ -\bar{p}_i & q_i \end{bmatrix}, \quad p = \begin{bmatrix} p_1 \\ q_1 \\ \vdots \\ p_n \\ q_n \end{bmatrix}. \tag{31}$$

The adjoint representations are all given by the formula (8) and the co-adjoint representations are of the form (20), (22), (23), and (24), respectively.

1. The group $\mathscr{A}(n,\mathbb{H}) = \mathbb{H}^n \square U^*(2n)$

We set $a(n,\mathbb{H})^* = a(n,\mathbb{H}) = \mathbb{C}^{2n} \times u^*(2n)$ with the (real) scalar product

$$\langle (p,M),(\xi,X)\rangle = \mathrm{Re}\{p^t\xi\} + \mathrm{tr}(M^tX).$$

The group $U^*(2n)$ acts transitively on $\mathbb{C}^{2n}_*$ and the little group of the point

$$p_0 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad 0 \in \mathbb{C}^{2n-1}$$

is equal to $\mathscr{A}(n-1,\mathbb{H})$. Since the group $\mathscr{A}(1,\mathbb{H}) = \mathbb{C}^2 \square U(2)$ acts freely and transitively on $\mathbb{C}^2_*$, no nontrivial invariant function exists on $a(1,\mathbb{H})^*$ and we are in the same situation as for the real and complex affine groups: *there is no invariant function of* $\mathscr{A}(n,\mathbb{H})$.

2. The group $\mathscr{SA}(n,\mathbb{H}) = \mathbb{H}^n \square SU^*(2n)$

We set $sa(n,\mathbb{H})^* = sa(n,H) = \mathbb{C}^{2n} \times su^*(2n)$ with the same scalar product as in the foregoing case.

The group $SU^*(2n)$ also acts transitively on $\mathbb{C}^{2n}$ and the little group of the same point p_0 is now equal to $\mathscr{SA}(n-1,\mathbb{H})$. The group $\mathscr{SA}(1,\mathbb{H}) = \mathbb{C}^2 \square SU(2)$ acts freely but not transitively on $\mathbb{C}^2_*$: the orbits are of real codimension one and separated by the invariant polynomial $p^\dagger p$.

Hence as for the real group $\mathscr{S}\mathscr{A}(n,\mathbb{R})$, *there is one fundamental* (real) *invariant of* $\mathscr{S}\mathscr{A}(n,H)$:

$$\varphi(p,M) = \det(p, J_n\bar{p}, Mp, MJ_n\bar{p}, \ldots, M^{n-1}p, M^{n-1}J_n\bar{p}).$$

3. The group $\mathscr{E}(n,\mathbb{H}) = \mathbb{H}^n \square O^*(2n)$

We set $e(n,\mathbb{H})^* = e(n,\mathbb{H}) = \mathbb{C}^{2n} \times o^*(2n)$ with the scalar product

$$\langle (p,M),(\xi,X)\rangle = \mathrm{Re}\{p^\dagger J_n \xi\} + \tfrac{1}{2}\,\mathrm{tr}(MX).$$

By using the quaternionic formalism, it is easy to show that the generic orbits in $\mathbb{C}^{2n}$ are of real codimension three and are associated to the little group $O^*(2(n-1))$; they are separated by the quaternionic invariant function P^tP or, according to (31), by the three real polynomial functions $\mathrm{Re}\{p^tp\}$, $\mathrm{Im}\{p^tp\}$ and $i\bar{p}^tJ_np$. As to the $n-1$ invariant functions (19), we proceed exactly as for the real Euclidean group, by introducing the projection operator $E(P)$ associated to the orthogonal complementary subspace of $\mathbb{H}P$:

$$E(P) = I_{2n} - P(P^tP)^{-1}P^t.$$

By setting

$$|P|^2 = \det_{\mathbb{C}}(P^tP) \in \mathbb{R}^+$$

we have

$$(P^tP)^{-1} = |P|^{-2}\bar{P}^t\bar{P};$$

hence

$$|P|^2 E(P) = |P|^2 I_{2n} - P(\bar{P}^t\bar{P})P^t$$

is a polynomial function in p [see (31)]. Thus *there are* $n+2$ (real) *fundamental invariants* of $\mathscr{E}(n,\mathbb{H})$

$$\varphi_0(p,M) = \mathrm{Re}\{p^tp\}, \quad \varphi_{00}(p,M) = \mathrm{Im}\{p^tp\},$$
$$\varphi_{000}(p,M) = i\bar{p}^tJ_np,$$
$$\varphi_k(p,M) = C_{2k}((|P|^2 I_{2n} - P(\bar{P}^t\bar{P})P^t)M),$$
$$k = 1,\ldots,n-1,^{14}$$

where

$$\det_{\mathbb{C}}(|P|^2 E(P)M - \lambda I_{2n}) = \sum_{l=0}^{2n} C_l(|P|^2 E(P)M)\lambda^{2n-l}.$$

4. The group $\mathscr{H}(p,q,\mathbb{H}) = \mathbb{H}^n \square \mathrm{Sp}(p,q)$

We set $h(p,q,\mathbb{H})^* = h(p,q,\mathbb{H})$ with the scalar product

$$\langle (p,M),(\xi,X)\rangle = \mathrm{Re}\{p^\dagger I_{2p,2q}\xi\} + \tfrac{1}{2}\,\mathrm{tr}(MX).$$

There is an open stratum in $\mathbb{C}^{2n}$ consisting of orbits of real codimension one associated to the little group $\mathrm{Sp}(p,q-1)$ and separated by the invariant polynomial function $Q(p) = p^\dagger I_{2p,2q}p$. The $n-1$ functions (19) may again be expressed with the help of a projection operator

$$E(p) = I_{2n} - \frac{1}{Q(p)}pp^\dagger I_{2p,2q}.$$

Thus, *there are n fundamental invariants of* $\mathscr{H}(p,q,\mathbb{H})$:

$$\varphi_0(p,M) = Q(p), = p^\dagger I_{2p,2q}p,$$
$$\varphi_k(p,M) = C_{2k}((Q(p)I_{2n} - pp^\dagger I_{2p,2q})M), \quad k = 1,\ldots,n-1,$$

where

$$\det((Q(p)I_{2n} - pp^\dagger I_{2p,2q})M - \lambda I_{2n}) = \sum_{l=0}^{2n} C_l((Q(p)I_{2n} - pp^\dagger I_{2p,2q})M)\lambda^{2n-l}.$$

CONCLUSION

In this article we have developed a general procedure for obtaining the invariant functions with respect to the coadjoint representation of a semidirect product of a Lie group by an abelian one. This procedure is of an inductive nature in the same sense as Mackey's one for constructing the unitary representations of such groups. Thereafter it has been applied to the inhomogeneous classical groups with the result that their invariant functions are generated by polynomial ones. As we have been able to give an explicit form of these polynomial functions, we have solved the problem of finding the Casimir operators of these groups.

This procedure can be applied to an arbitrary group of the considered structure in order to construct its Casimir operators (if they exist). An interesting question is: what happens if the abelian requirement in the first factor is dropped?

ACKNOWLEDGMENTS

I am indebted to Dr. J. Patera for having proposed this problem to me.

APPENDIX: THE FUNDAMENTAL INVARIANTS OF THE CLASSICAL LIE GROUPS

We list all classical Lie groups considered in the main text together with their Lie algebras and their fundamental invariants in Table I. The standard notations used by Helgason[16] have been adopted. All the listed groups are algebraic subgroups of $\mathrm{Gl}(n,\mathbb{C})$ for some n and their Lie algebras are subalgebras of $gl(n,\mathbb{C})$ considered as a real Lie algebra.

In the main text, the groups $\mathrm{Gl}(n,\mathbb{R})$, $\mathrm{Sl}(n,\mathbb{R})$, $\mathrm{U}(p,q)$, $\mathrm{O}(p,q)$, $\mathrm{Sp}(n,\mathbb{R})$ are referred as real groups, $\mathrm{Gl}(n,\mathbb{C})$, $\mathrm{Sl}(n,\mathbb{C})$, $\mathrm{O}(n,\mathbb{C})$, $\mathrm{Sp}(n,\mathbb{C})$ as complex groups and $\mathrm{U}^*(2n)$, $\mathrm{SU}^*(2n)$, $\mathrm{O}^*(2n)$, $\mathrm{Sp}(p,q)$ as quaternionic groups. This classification arises from Cartan's classification of the real forms of complex classical Lie algebras.[11]

The transpose and the conjugate of a matrix A are denoted by A^t and $\bar{A}$, respectively; in addition $A^\dagger = \bar{A}^t$. We use the matrices $I_{p,q} = -I_p \oplus I_q$ and $J_n = J_1 \oplus \cdots \oplus J_1$, where I_n denotes the unit $n \times n$ matrix and J_1 is the 2×2 antisymmetric matrix

$$J_1 = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}.$$

The fundamental invariants can be expressed in the form of characteristic coefficients $C_l(X)$ defined by

$$\det(X - \lambda I_n) = \sum_{l=0}^{n} C_l(X)\lambda^{n-l}.$$

TABLE I. All classical Lie groups considered in the main text together with their Lie algebras and their fundamental invariants.

$G \subseteq Gl(n, C)$	$g \subseteq gl(n, C)$	Fundamental invariants
$Gl(n, C)$	$gl(n, C)$	$\mathrm{Re}\{C_l(X)\}$, $\mathrm{Im}\{C_l(X)\}$, $l = 1,\ldots,n$
$Sl(n, C)$, $\det g = 1$	$sl(n, C)$, $\mathrm{tr}\, X = 0$	$\mathrm{Re}\{C_l(X)\}$, $\mathrm{Im}\{C_l(X)\}$, $l = 1,\ldots,n-1$
$O(n, C)$, $g^t g = I_n$	$O(n, C)$, $X^t + X = 0_n$	$\mathrm{Re}\{C_{2k}(X)\}$, $\mathrm{Im}\{C_{2k}(X)\}$, $k = 1,\ldots,(n/2)$
$Sp(n, C)$, $g^t J_n g = J_n$	$sp(n, C)$, $X^t J_n + J_n X = 0_{2n}$	$\mathrm{Re}\{C_{2k}(X)\}$, $\mathrm{Im}\{C_{2k}(X)\}$, $k = 1,\ldots,n$
$Gl(n, R)$, $\bar{g} = g$	$sl(n, R)$, $\bar{X} = X$	$C_l(X)$, $l = 1,\ldots,n$
$Sl(n, R)$, $\bar{g} = g$, $\det g = 1$	$sl(n, R)$, $\bar{X} = X$, $\mathrm{tr}\, X = 0$	$C_l(X)$, $l = 2,\ldots,n$
$U(p, q)$, $g^\dagger I_{p,q}\, g = I_{p,q}$	$u(p, q)$, $X^\dagger I_{p,q} + I_{p,q} X = 0_n$	$C_l(\sqrt{-1}X)$, $l = 1,\ldots,n = p+q$
$SU(p, q)$, $g^\dagger I_{p,q}\, g = I_{p,q}$, $\det g = 1$	$su(p, q)$, $X^\dagger I_{p,q} + I_{p,q} X = 0_n$, $\mathrm{tr}\, X = 0$	$C_l(\sqrt{-1}X)$, $l = 2,\ldots,n$
$O(p, q)$, $\bar{g} = g$, $g^t I_{p,q}\, g = I_{p,q}$	$o(p, q)$, $\bar{X} = X$, $X^t I_{p,q} + I_{p,q} X = 0_n$	$C_{2k}(X)$, $k = 1,\ldots,(n/2)$, $n = p+q$
$Sp(n, R)$, $\bar{g} = g$, $g^t J_n\, g = J_n$	$sp(n, R)$, $\bar{X} = X$, $X^t J_n + J_n X = 0_{2n}$	$C_{2k}(X)$, $k = 1,\ldots,n$
$U^*(2n)$, $J_n \bar{g} J_n^{-1} = g$	$u^*(2n)$, $J_n \bar{X} J_n^{-1} = X$	$C_l(X)$, $l = 1,\ldots,2n$
$Su^*(2n)$, $J_n \bar{g} J_n^{-1} = g$, $\det g = 1$	$su^*(2n)$, $J_n \bar{X} J_n^{-1} = X$, $\mathrm{tr}\, X = 0$	$C_l(X)$, $l = 2,\ldots,2n$
$O^*(2n)$, $J_n \bar{g} J_n^{-1} = g$, $g^t g = I_{2n}$	$o^*(2n)$, $J_n \bar{X} J_n^{-1} = X$, $X^t + X = 0_{2n}$	$C_{2k}(X)$, $k = 1,\ldots,n$
$Sp(p, q)$, $J_n \bar{g} J_n^{-1} = g$ $(n = p+q)$ $g^\dagger I_{2p,2q}\, g = I_{2p,2q}$	$sp(p, q)$, $J_n \bar{X} J_n^{-1} = X$ $X^\dagger I_{2p,2q} + I_{2p,2q} X = 0_{2n}$	$C_{2k}(X)$, $k = 1,\ldots,n = p+q$
also:	also:	
$g^t (J_n I_{2p,2q}) g = J_n I_{2p,2q}$	$X^t (J_n I_{2p,2q}) + (J_n I_{2p,2q}) X = 0_{2n}$	

Let us recall that

$$C_n(X) = \det X, \quad C_l(X) = (-1)^{n-1}\, \mathrm{tr}\, X,$$
$$C_l(\mu X) = \mu^l C_l(X).$$

Since the bilinear form $(X, Y) = \mathrm{tr}(XY)$ is nondegenerate for all considered Lie algebras $\mathcal{g}$, we can identify $\mathcal{g}^*$ with $\mathcal{g}$ and therefore $\mathrm{Ad}^* = \mathrm{Ad}$. According to Sec. 1, if $\{L_{ij}\}$ is a basis of $\mathcal{g}$, then the invariant functions $C_l(X)$ are polynomials in the variables $x_{ij} = (X, L_{ij})$:

$$C_l(X) = p_l(x_{ij})$$

and the corresponding Casimir operator is given by

$$I_l = \mathrm{Sym}(p_l(L_{ij})).$$

By developing this expression we recover its usual form.

The form $C_l(X)$ of the fundamental invariants is not completely standard. For compact groups it appears in the book of Kobayashi and Nomizu,[17] for the other groups we have no specific references. However, the result for the complex groups can be proved in the same way as for the compact real forms; for the remaining real forms, it follows directly from a Weyl's unitary trick argument. The main step for the case of complex groups consists of showing that the restriction mapping $P_G(\mathcal{g}) \to P_K(\mathcal{h})$ in injective (and actually bijective) where $\mathcal{h}$ is a Cartan subalgebra of $\mathcal{g}$ and K is the normalizer of $\mathcal{h}$ in G. This is true because the functions are polynomials and because an open neighborhood $\mathcal{V} \subset \mathcal{g}$ of a well chosen $Y \in \mathcal{h}$ always exists such that $\mathcal{V} \subseteq \mathrm{Ad}(G)\mathcal{h}$.

[1] G. Racah, *Group Theory and Spectroscopy*, Springer Tracts Mod. Phys. **37** (1965); L. C. Biedenharn, J. Math. Phys. **4**, 436 (1963); M. Moshinsky, J. Math. Phys. **4**, 1128 (1963); M. Umezawa, Nucl. Phys. **57**, 65 (1964).

[2] V. G. Kadyshevskii and I. T. Todorov, Sov. J. Nucl. Phys. **3**, 96 (1966).

[3] R. Mirman, J. Math. Phys. **9**, 39 (1968).

[4] F. W. Hehl, P. von der Heyde, G. D. Karlick, and J. M. Nester, Rev. Mod. Phys. **48**, 393 (1976); K. S. Stelle and P. C. West, Phys. Rev. D **21**, 1466 (1980).

[5] O. L. Weaver, R. Y. Cusson, and L. C. Biedenharn, Ann. Phys. (N.Y.) **102**, 493 (1976); G. Rosensteel and D. J. Rowe, Ann. Phys. (N.Y.) **126**, 363 (1980); V. Guillemin and S. Sternberg, Ann. Phys. (N.Y.) **127**, 220 (1981).

[6] A. P. Deemicev and N. F. Nelipa, Vestnik Mosk. Univ. Fiz. Astronomiya. **21**, No. 2, 3, 7; *ibid.* No. 4, 23 (1980); A. P. Deemicev, N. F. Nelipa, and M. Caician, *ibid.*, No. 4, 27, (1980).

[7] A. Kirillov, Russian Math. Surveys **17**, 53 (1962); *ibid, Eléments de la théorie des représentations* (Mir, Moscow, 1974).

[8] I. M. Gel'fand, Mat. Sbornik **26**, 103 (1950).

[9] J. Patera, R. T. Sharp, P. Winternitz, and H. Zassenhaus, J. Math. Phys. **17**, 986 (1976).

[10] G. W. Mackey, *Induced Representation of Groups and Quantum Mechanics* (Benjamin-Cummings, Reading, MA, 1968).

[11] N. Jacobson, *Lie Algebras* (Interscience, New York, 1962).

[12] M. Duflo, Funct. Analys. Applied **4**, 38 (1970).

[13] L. Abellanas and L. Martinez Alonso, J. Math. Phys. **16**, 1580 (1975).

[14] Some of these polynomial functions are defined modulo the function φ_0.

[15] W. Greub, *Linear Algebra, Grundl. d. Math. Wissen.* Vol. 97 (Springer, New York, 1967).

[16] S. Helgason, *Differential Geometry and Symmetric Spaces* (Academic, New York, 1962).

[17] S. Kobayashi and K. Nomizu, *Foundations of Differential Geometry* (Interscience, New York, 1969), Vol. II.

Representation matrices for U(4)

G. J. Balster, O. S. van Roosmalen,[a] and A. E. L. Dieperink
Kernfysisch Versneller Instituut, Rijksuniversiteit Groningen, The Netherlands

(Received 23 November 1982; accepted for publication 22 December 1982)

We propose an algorithm for the numerical calculation of matrix elements of general U(4) group elements, applicable to large totally symmetric representations of U(4). A possible generalization to the U(6) case is pointed out.

PACS numbers: 02.20.Qs, 02.70. + d

I. INTRODUCTION

The application of an algebraic model in which a U(4) group structure is proposed to describe rotation-vibrational degrees of freedom of diatomic molecules[1] stimulated our interest in the properties of this group. In particular, to calculate transition probabilities,[2] we found it necessary to calculate representation matrices $\langle|e^{i\epsilon\hat{G}}|\rangle$ of this group for totally symmetric representations $[N]$ with large N. In this paper we present an algorithm to calculate these matrix elements numerically using the concept of coherent states. In Sec. II we give an outline of the problem, and in Sec. III we describe the algorithm. In the Sec. IV we summarize the merits of the proposed formulation, and indicate how it can be generalized to the case of U(6), the symmetry group of the interacting boson model,[3] which has been applied to describe collective properties of nuclei.

II. OUTLINE OF THE PROBLEM

We will use a Fock representation of the U(4) structure in terms of identical bosons that can occupy a scalar s state and a vector p state. All unitary transformations on these four single-boson states define the U(4) group of interest here. All states formed by distributing N bosons among these four single boson states constitute the completely symmetric representation $[N]$ of U(4). Such states can be explicitly constructed by creating N bosons in the vacuum $|0\rangle$. The creation and annihilation operators are

$$s^\dagger, p^\dagger_\mu \quad \text{and} \quad s, \tilde{p}_\mu \equiv (-)^\mu p_{-\mu} \quad (\mu = -1,0,1), \tag{1}$$

respectively. They satisfy the usual commutation rules. The 16 generators $\hat{G}_i$, of the U(4) group can be expressed in terms of these operators:

$$\{\hat{G}_j, j = 1\cdots 16\} = \{(s^\dagger s)^{(0)}_0, (p^\dagger \tilde{p})^{(0)}_0, (p^\dagger \tilde{p})^{(1)}_\mu, (p^\dagger \tilde{p})^{(2)}_\mu, (p^\dagger s + s^\dagger \tilde{p})^{(1)}_\mu, i(p^\dagger s - s^\dagger \tilde{p})^{(1)}_\mu\}, \tag{2}$$

where the brackets denote angular momentum coupling. Since we have assigned, to the single-boson states, definite transformation properties under the rotation group O(3), there are only two group chains according to which we can label the N-boson states:

$$\mathrm{U}(4) \supset \mathrm{U}(3) \supset \mathrm{O}(3) \supset \mathrm{O}(2), \tag{3a}$$

$$\mathrm{U}(4) \supset \mathrm{O}(4) \supset \mathrm{O}(3) \supset \mathrm{O}(2). \tag{3b}$$

While the chain (3b) is appropriate for the description of molecular spectra[1] in practice it is more convenient to use the basis of the chain (3a) since the U(3) group consists of the transformations among the p-bosons and has n (the number of p-bosons) as an invariant. The quantum labels related to the chain (3a) are: $[N]$, n, l, m, and the corresponding states [the so-called U(3) basis]

$$|[N]nlm\rangle = A_{nl}(s^\dagger)^{N-n}(p^\dagger \cdot p^\dagger)^{(n-l)/2}\mathscr{Y}_{lm}(p^\dagger). \tag{4}$$

(The dot product is defined as $a\cdot b = \Sigma_\mu (-)^\mu a_\mu b_{-\mu}$). The coefficients A_{nl} are derived in Ref. 4:

$$A_{nl} = (-)^{(n-l)/2}[4\pi/(n+l+1)!!(n-l)!!(N-n)!]^{1/2} \tag{5}$$

and the $\mathscr{Y}_{lm}(p^\dagger)$ are solid harmonics in the p-boson creation operators.

A general element of the U(4) group can be expressed in terms of the generators

$$\hat{U}(\epsilon_j) = \exp\left(i\sum_j \epsilon_j \hat{G}_j\right). \tag{6}$$

We are interested in the evaluation of the following matrix elements:

$$\langle[N]nlm|\hat{U}(\epsilon_j)|[N]n'l'm'\rangle. \tag{7}$$

A 4×4 matrix representation for the generators can be obtained by considering their action on the four states $s^\dagger|0\rangle$ and $p^\dagger_\mu|0\rangle$. The unitary operator $\hat{U}$ can then be expressed simply as a 4×4 matrix $U^{(}$ by diagonalizing $\Sigma_j \epsilon_j \hat{G}_j$, exponentiating and transforming back to the original basis.

III. THE ALGORITHM

The procedure is based on three important properties of so-called coherent states,[5] which are product wavefunctions of the form

$$|[N]\alpha_\mu\rangle = \frac{1}{\sqrt{N!}}(\sqrt{1-\alpha\cdot\tilde{\alpha}^*}s^\dagger + \tilde{\alpha}\cdot p^\dagger)^N|0\rangle. \tag{8}$$

First, they have a simple behavior under the action of a group element (6),

$$\hat{U}(\epsilon_j)|[N]\alpha_\mu\rangle = \exp(i\beta N)|[N]\alpha'_\mu\rangle, \tag{9}$$

where β (real) and α'_μ can be determined from

[a] Present address: W. K. Kellogg Radiation Lab., California Institute of Technology, Pasadena, CA 91125.

 0022-2488/83/061392-03$02.50 © 1983 American Institute of Physics

$$U^f\begin{pmatrix}\alpha_1\\ \alpha_0\\ \alpha_{-1}\\ (1-\alpha\cdot\tilde{\alpha}^*)^{1/2}\end{pmatrix}=e^{i\beta}\begin{pmatrix}\alpha'_1\\ \alpha'_0\\ \alpha'_{-1}\\ (1-\alpha'\cdot\tilde{\alpha}'^*)^{1/2}\end{pmatrix} \tag{10}$$

Secondly, the overlap of such a state with a U(3) basis state is given by [compare with Eq. (4)]

$$\langle[N]nlm|[N]\alpha_\mu\rangle = (N!)^{1/2}A_{nl}(1-\alpha\cdot\tilde{\alpha}^*)^{(N-n)/2}(\alpha\cdot\alpha)^{(n-l)/2}\mathscr{Y}_{lm}(\alpha). \tag{11}$$

Thirdly, one can decompose any U(3) basis state in coherent states using

$$|[N]nlm\rangle = (2\pi A_{nl})^{-1}(2^N/N!)^{1/2}\times\int_0^{2\pi}d\chi\int_{-1}^{1}d\cos\theta\int_0^{2\pi}d\phi\times\exp(-in\chi)Y^l_m(\theta,\phi)|[N]\alpha_\mu(\chi,\theta,\phi)\rangle, \tag{12}$$

with

$$\alpha_\mu(\chi,\theta,\phi)=\frac{1}{\sqrt{2}}e^{i\chi}\begin{pmatrix}-(1/\sqrt{2})\sin\theta\,e^{-i\phi}\\ \cos\theta\\ (1/\sqrt{2})\sin\theta\,e^{i\phi}\end{pmatrix}. \tag{13}$$

An important point to note is that one can replace the three-dimensional integral in Eq. (12) by a summation in an *exact* fashion due to the fact that the states in the representation $[N]$ of U(4) are polynomials in the boson operators of (finite) order N. In particular the decompositon (12) of a general U(3) basis state remains exact if we replace the integration over ϕ by a $(2N+1)$-point, the one over χ by an $(N+1)$-point equidistant summation, and the integral over $\cos\theta$ by an $(N+1)$-point Gaussian summation. We replace symbolically: $\int\to\sum\!\!\!\!\!\!-$. Using Eqs. (9)–(13), one obtains the result

$$\langle[N]nlm|\hat{U}|[N]n'l'm\rangle = 2^{N/2}A_{nl}A^{-1}_{n'l'}\exp(i\beta N)\times\sum_0^{2\pi}\!\!\!\!\!\!-\,d\chi\sum_{-1}^{1}\!\!\!\!\!\!-\,d\cos\theta\sum_0^{2\pi}\!\!\!\!\!\!-\,d\phi\exp(-in'\chi)\times Y^{l'}_{m'}(\theta,\phi)(1-\alpha'\cdot\tilde{\alpha}'^*)^{(N-n)/2}(\alpha'\cdot\alpha')^{(n-l)/2}\mathscr{Y}_{lm}(\alpha'). \tag{14}$$

If one considers a general boson-number-conserving, rotationally invariant Hamiltonian,[6] its eigenstates can be expressed in the U(3) basis as

$$|[N]vlm\rangle=\sum_n a^v_n(l)|[N]nlm\rangle, \tag{15}$$

where the $a^v_n(l)$ are the expansion coefficients.

To obtain matrix elements of $\hat{U}$ between the eigenstates without the necessity of determining all matrix elements (7) (which is particularly convenient if the number of states in the representation $[N]$ is large and if one is interested in only a few matrix elements in the basis of eigenstates), one replaces in Eq. (14)

$$A_{nl}(1-\alpha'\cdot\tilde{\alpha}'^*)^{(N-n)/2}(\alpha'\cdot\alpha')^{(n-l)/2}$$

by

$$\sum_n a^v_n(l)A_{nl}(1-\alpha'\cdot\tilde{\alpha}'^*)^{(N-n)/2}(\alpha'\cdot\alpha')^{(n-l)/2}, \tag{16}$$

and

$$A^{-1}_{n'l'}\exp(-in'\chi)$$

by

$$\sum_{n'}a^{v'}_{n'}(l')A^{-1}_{n'l'}\exp(-in'\chi). \tag{17}$$

In particular, this last replacement will be efficient regarding computational time.

IV. DISCUSSION

There are two main reasons why the above described procedure is attractive. First, this method is numerically efficient. Because of the unique property (9) of coherent states one only needs to decompose the ket of a matrix element (7). In any other basis one would need to calculate *all* the matrix elements of $\hat{U}$ and use transformation brackets to return to the U(3) basis [for both the bra and the ket in Eq. (7)], which is inconvenient for large N {this advantage of the coherent state basis is partly compensated for by the fact that the decomposition (12) in order to be exact involves a number of coherent states that is larger than the dimension of the representation $[N]$}. In addition, one is usually interested in only few matrix elements of $\hat{U}(\epsilon_j)$ (particularly for large N) and these are easily provided by Eq. (14). Instead of transformation brackets only simple numerical procedures and functions are involved. Secondly, if one considers a typical Hamiltonian possessing a dynamical symmetry [corresponding to the group chains (3a) or (3b)] the ground state has special properties and can be decomposed rather easily. For example, in the U(3) limit one obtains

$$|[N],\mathrm{gs}\rangle=(1/\sqrt{N!})(s^\dagger)^N|0\rangle, \tag{18}$$

and in the O(4) limit

$$|[N],\mathrm{gs}\rangle=\mathscr{N}\int_{-1}^{1}d\cos\theta\int_0^{2\pi}d\phi\times[(1/\sqrt{2})s^\dagger+(1/\sqrt{2})\tilde{\alpha}(\chi=0,\theta,\phi)\cdot p^\dagger]^N|0\rangle. \tag{19}$$

Therefore, the evolution of the ground state of such a system under an operator (6) can be calculated rather simply. Also certain symmetries in the problem can easily be implemented, for instance, in the case of cylindrical symmetry one can omit the ϕ integration in (14).

We have written a computer code based upon this algorithm. It is available on request.

Extensions of these techniques to U(6) are straightforward, although the resulting expressions are more involved. The U(6) formula corresponding to Eq. (4) is given in Ref. 7, where also the generalizations of the spherical harmonics appropriate for the U(6) problem are treated. An equation analogous to (14) would involve a five-dimensional integral and contain more complicated functions. Simplifications in the spirit of the points mentioned above could be applied.

Let us finally emphasize that the special realization of the U(4) group, and in particular the O(3) structure assumed here, is by no means essential. An appropriate coherent state basis can also be found for other chains of subgroups.

ACKNOWLEDGMENTS

We wish to thank Professor F. Iachello for his stimulating interest. This work has been performed as part of the research program of the Stichting voor Fundamenteel Onderzoek der Materie (FOM) with financial support of the Nederlandse Organisatie voor Zuiver Wetenschappelijk Onderzoek (ZWO).

[1]F. Iachello, Chem. Phys. Lett. **78**, 581 (1981); F. Iachello and R. D. Levine, J. Chem. Phys. **77**, 3046 (1982).

[2]I. Benjamin, O. S. van Roosmalen, F. Iachello, and R. D. Levine (to be published).

[3]A. Arima and F. Iachello, Ann. Phys. (N.Y.) **99**, 253 (1976); **111**, 201 (1978); **123**, 468 (1979).

[4]M. Moshinsky, *The Harmonic Oscillator in Modern Physics: From Atoms and Quarks* (Gordon and Breach, New York, 1969), pp. 21, 22.

[5]R. Gilmore, C. M. Bowden, and L. M. Narducci, Phys. Rev. A **12**, 1019 (1975).

[6]O. S. van Roosmalen and A. E. L. Dieperink, Ann. Phys. (N.Y.) **139**, 198 (1982).

[7]E. Chacón and M. Moshinsky, J. Math. Phys. **18**, 870 (1977).

Sufficient conditions for weak minimum of a functional depending on n functions and their derivatives

Alak Ray
Department of Physics, University of California, Irvine, California 92717 and Tata Institute of Fundamental Research, Homi Bhabha Road, Bombay 400 005, India[a)]

(Received 18 August 1981; accepted for publication 4 June 1982)

The sufficient condition for positive definiteness of a quadratic form involving unsymmetric matrix operators is given in terms of a global criterion. This determines the sufficient condition for the weak minimum of a functional depending on n-functions and their derivatives. The criterion is formulated in terms of the absence of conjugate points in the interval of the variational integral and is a generalization of a similar criterion due to Newcomb in problems of hydromagnetic stability in plasma physics.

PACS numbers: 02.30. + g

I. INTRODUCTION

In a variety of physical problems the necessary and sufficient conditions for a functional to have a weak minimum prove to be interesting. An important example comes from plasma physics, where equilibrium magnetohydrodynamic configurations are subjected to tests of stability. In such cases, whether a configuration is stable or not depends on the positive definiteness (or lack thereof) of a bilinear functional embodying the second variation of the (potential) energy in the perturbed configuration. A number of necessary and sufficient conditions for the positive definiteness (and semidefiniteness) of the second variation exist; examples of these are Legendre's and Jacobi's conditions.[1] The second set of conditions have been used by Newcomb[2] to find the global stability criterion for a diffuse plasma pinch. This approach is more convenient than the previous variational formulation for stability due to Bernstein *et al.*[3] since Newcomb's criterion involves solution of a second-order ordinary differential equation as opposed to the arbitrary search in the space of perturbation functions for the positive (semi) definiteness of a quadratic form.[3] Aside from its application in physical situations, the problems of necessary and sufficient conditions for the existence of a weak minimum of a functional and the positive (semi) definiteness of quadratic forms associated with nonsymmetric matrix operators are of independent interest in mathematical physics. In this article we provide a generalization of the Jacobi-type conditions for the variational problem of functions involving n-dimensional vector functions. The nature of the generalization will be made apparent in the next section. To the best of the author's knowledge such a generalization has not previously appeared in the literature.

II. VARIATIONAL PROBLEM FOR A FUNCTIONAL OF n FUNCTIONS

Let $F(x,y,z)$ be a function with continuous partial derivatives up to order three with respect to all its arguments in a given range of its first argument $x \in (a,b) \subset \mathbb{R}, y, z \in \mathbb{R}^n$. Given this function, we construct the functional

$$W[y] = \int_a^b F(x,y,y')dx \tag{1}$$

with

$$y(a) = M_0 \quad \text{and} \quad y(b) = M_1,$$

where M_0 and M_1 are two fixed points in an n-dimensional vector space. Here, y denotes an n-dimensional row vector

$$y = \{y_1(x), y_2(x),..., y_n(x)\};$$

each component $y_i(x)$ may itself be a vector in the physical space. In addition, $y' = \{y_1'(x), y_2'(x),..., y_n'(x)\}$, where the primes denote differentiation with respect to x.

The functional (1) is made stationary by the row vector function $\hat{y} = \{\hat{y}_i(x)\}$, the coordinates of which satisfy the set of Euler–Lagrange equations

$$\frac{d}{dx} F_{y_i'} - F_{y_i} = 0 \quad (i = 1,...,n). \tag{2}$$

The subscripts y_i and y_i' on F denote partial differentiation of F with respect to those particular variables. As already stated, the necessary and sufficient conditions for a weak minimum of the functional (1) are expressed in terms of the second variation of (1). By Taylor expanding $y_i(x)$ about $\hat{y}_i(x)$,

$$y_i(x) = \hat{y}_i(x) + h_i(x), \tag{3}$$

the second variation is found to be

$$\delta^2 W[h] = \frac{1}{2}\int_a^b dx \left[\sum_{i,k=1}^{n} \left\{ F_{y_i'y_k'} h_i' h_k' + 2F_{y_i'y_k} h_i' h_k + F_{y_i y_k} h_i h_k \right\} \right]. \tag{4}$$

In this variation the end points are kept fixed at $y(a,b) = \hat{y}(a,b)$, i.e., $h_i(a) = h_i(b) = 0$.

In expression (2) the second derivatives $F_{y_i'y_k'}$, etc., are evaluated on the extremal curve $y_i = \hat{y}_i(x)$, $y_i' = \hat{y}_i'(x)$. The first variation of $W[y]$ about $\hat{y}$ is zero, since $\hat{y}$ is an extremal and gives the equilibrium configuration satisfied by the Euler–Lagrange set (2).

One can prove (e.g., Sec. 24 of Gel'fand and Fomin[1]) that the necessary and sufficient condition for the functional $W[y]$ to have a minimum for $y = \hat{y}$ is that

[a)] Current address.

 0022-2488/83/061395-04$02.50 © 1983 American Institute of Physics

$$\delta^2 W[h] > 0$$

for all coefficients in (4) evaluated for $y = \hat{y}$ and for all admissible h. (For the sufficient condition, one actually requires $\delta^2 W[h]$ to be strongly positive, i.e.,

$$\delta^2 W[h] \geqslant k \|h\|^2,$$

where k is a positive constant for $y = \hat{y}$ and $\|h\|$ is the norm of any function h in the linear space on which the quadratic functional $\delta^2 W[h]$ is defined. However, for a finite-dimensional space, strong positivity of a quadratic form is equivalent to positive definiteness of the quadratic form.) Therefore, the problem of finding a weak minimum in a finite-dimensional space can be reduced to finding the conditions under which the second variation (4) is positive definite.

It is obvious that the coefficients $F_{y_i y_k}$ and $F_{y_i' y_k'}$ are symmetric in the indices i and k; however, in general, the term $F_{y_i y_k'} \neq F_{y_i' y_k}$. The physical circumstances pertaining to a helical magnetic field in a plasma pinch as well as pre-flare loops in the coronal atmosphere of the sun in which these asymmetries arise are discussed elsewhere.[4] One therefore needs to determine the general conditions under which second variation (4) remains positive definite. A set of theorems in this connection that are useful in linking the differential equation approach of Newcomb[2] to the variational one of Bernstein *et al.*[3] to analyze the stability of a magnetized plasma configuration can be found in several texts on the calculus of variations (e.g., Ref. 1).

However, these theorems implicitly treat a special case in which $F_{y_i y_k'} = F_{y_i' y_k}$ and therefore are not directly applicable to a variety of physically realizable situations. In our example here, we give the global criterion for the positive definiteness for a quadratic form associated with a general asymmetric matrix operator and therefore the condition of weak minimum of the functional (1). We note that any term in the functional (4) involving symmetric operators $F_{y_i' y_k}$ can, through integration by parts, be absorbed in terms proportional to $h_i\, h_k$, while this is not possible with general, asymmetric operators $F_{y_i' y_k}$.

III. WEAK MINIMUM, JACOBI'S EQUATIONS, AND CONJUGATE POINTS

The functional $W[y]$ has a weak minimum for $y = \hat{y}$ if there exists an $\epsilon > 0$ such that $W[y] - W[\hat{y}]$ has a positive sign for all $y \neq \hat{y}$ in the domain of definition of the functional which satisfy $\|y - \hat{y}\|_1 < \epsilon$, where $\|\ \|$ denotes the norm in the space $\mathscr{D}_1$. The space $\mathscr{D}_1(a,b)$ is defined as consisting of all functions $y(x)$ defined on an interval $[a,b]$ which are continuous and have continuous first derivatives and the norm defined by

$$\|y\|_1 = \max_{a<x<b} |y(x)| + \max_{a<x<b} |y'(x)|.$$

The condition of weak minimum will be stated by introducing the concept of conjugate point, which is in turn associated with a set of differential equations known as Jacobi equations. Define the matrices P, Q, and R, where

$$P_{ik} = F_{y_i y_k},$$
$$Q_{ik} = F_{y_i y_k'},$$
$$R_{ik} = F_{y_i' y_k'}.$$

The Jacobi system of equations is a system of n equations, each involving n quantities $(h^i_1(x), h^i_2(x), \ldots, h^i_n(x))$ which form an n-dimensional row vector. The system can be symbolically written as

$$\frac{d}{dx}(RY') + (Q^T - Q)Y' + \left(\frac{dQ^T}{dx} - P\right)Y = 0. \quad (5)$$

Here, Q^T is the transposed matrix of Q and $Y(x)$ is the solution to be determined in the form of a $n \times n$ matrix function of x in the range $a \leqslant x \leqslant b$ which is formed from the n independent solutions

$$h^{(1)} = (h^{(1)}_1(x), h^{(1)}_2(x), \ldots, h^{(1)}_n(x))$$
$$h^{(2)}$$
$$\vdots$$
etc.

Also,

$$Y' \equiv \frac{dY}{dx}.$$

We now introduce the conjugate point by the

Definition: Let $Y(x)$ be a set of solutions of the system (5) satisfying the initial conditions

$$Y(a) = \varnothing \quad (6a)$$

and

$$Y'(a) = I, \quad (6b)$$

where $\varnothing$ and I are null and identity matrices, respectively. Then the point $\tilde{a}$ (different from a) is said to be conjugate to the point a if the determinant

$$\begin{vmatrix} Y_{11}(x) & Y_{12}(x) \ldots & Y_{1n}(x) \\ \vdots & & \\ Y_{n1}(x) & \ldots\ldots\ldots & Y_{nn}(x) \end{vmatrix}$$

vanishes for $x = \tilde{a}$ for a $Y_{ik}(x) = h^i_k(x)$ that is not identically equal to zero.

IV. SUFFICIENT CONDITION FOR A WEAK MINIMUM

A matrix $R(x)$ is said to be positive definite if the quadratic form

$$\sum_{i,k=1}^{n} R_{ik}(x) h_i(x) h_k(x)$$

is positive for all x in $[a,b]$ and arbitrary $h_1(x), \ldots, h_n(x)$.

We shall now state a few theorems concerning the necessary and sufficient conditions for the weak minimum of the functional (1). The theorems concerning necessary conditions (namely, Theorems 3 and 4) have been proved for general asymmetric matrices Q and can be found in Sagan.[4] We shall only prove Theorem 2, that gives the sufficient condition for the positive definiteness of a quadratic form associated with a general asymmetric matrix operator. Theorem 2 is a crucial step for Theorem 1 since, as already mentioned in

Sec. II (see Gel'fand and Fomin Sec. 24) the sufficient condition for the existence of a weak minimum is closely linked with positive definiteness of a quadratic form.

Theorem 1: Let γ be a curve in $\mathbb{R}^n$ given by $y_i(x)$.
1. The curve γ is an extremal; i.e., satisfies the system of Euler equations,

$$F_{y_i} - \frac{d}{dx} F_{y_i'} = 0 \quad (i = 1,...,n);$$

2. Along γ the matrix

$$R_{ik}(x) = \tfrac{1}{2} F_{y_i' y_k'}[x,y(x),y'(x)] \quad (i,k = 1,...n),$$

is positive definite;
3. The interval $[a,b]$ contains no points conjugate to the point a.
Then, the functional (1) has a weak minimum for the curve γ.

Theorem 2: If $R(x)$ is a positive definite matrix, and if the interval $[a,b]$ does not contain a point $\tilde{a}$ that is conjugate to a, then the quadratic form (4) is positive definite for all $h(x)$ such that $h(a) = h(b) = 0$.

Theorem 3 (Jacobi's necessary condition)**:** If the extremal

$$y_1 = y_1(x),...,y_n = y_n(x).$$

corresponds to a minimum of the functional (1) and if the matrix

$$R_{ik} = \tfrac{1}{2} F_{y_i' y_k'}(x,y(x),y'(x)), \quad (i,k = 1,...,n)$$

is positive definite along this extremal, then the open interval (a,b) contains no point conjugate to a.

Theorem 4: If the quadratic form (4), where $R(x)$ is a positive definite matrix, is positive definite for all $h(x)$ such that $h(a) = h(b) = 0$, then the interval $[a,b]$ contains no point conjugate to a.

To prove Theorem 2 the following strategy will be taken: First, we shall write expression (4) as an absolute square plus an extra term that can be equated to zero. The condition under which this can be done leads to a set of nonlinear first-order differential equations called Ricatti's equations. In general, the solutions of these nonlinear equations do not exist. With a further transformation we shall convert this nonlinear set of differential equations to a set of second-order linear differential equations called the Jacobi system of equations. Given a set of initial conditions, the solutions of such a system of equations exists. However, the transformation under which Ricatti's equations can be converted to the Jacobi system is nonsingular only if a certain determinant is nonvanishing in the interval $[a,b]$. This condition is related to the nonexistence of a point $\tilde{a}$ conjugate to a in $[a,b]$. Thus, if a conjugate point does not occur in $[a,b]$, the solution of the Jacobi equations, which exists, can be used to find the solution of the Ricatti system which, in turn, reduces the functional (4) to a positive definite form.

The functional (4) can be written symbolically as

$$\delta^2 W[h] = \int_a^b dx [h'(x) R h'^T(x) + hQh'^T + h'Q^T h^T + hPh^T], \tag{7}$$

where h^T is defined as the transpose of the row vector h.

We note that, because of the boundary conditions

$$h(a) = h(b) = 0,$$

the following integral is identically equal to zero:

$$\int_a^b dx \frac{d}{dx} [h(x) u(x) h^T(x)] = 0,$$

where $u(x)$ is an arbitrary $n \times n$ matrix. Carrying out the term by term differentiations and adding to (7) one has

$$\delta^2 W[h] = \int_a^b dx \left[h(x) \left\{ P + \frac{du}{dx} \right\} h^T(x) + h(x)\{Q + u\} h'^T(x) + h'(x)\{Q^T + u\} h^T(x) + h'(x) R h'^T(x).\right.$$

By using the invariance property of a scalar under transposition the above expression for $\delta^2 W[h]$ can be rewritten as

$$\delta^2 W[h] = \int_a^b dx \left[h \left\{ P + \frac{du}{dx} \right\} h^T + h\{2Q + u + u^T\} h'^T + h' R h'^T \right].$$

One can write the quantity in the square brackets in the following way:

$$[\ ] = [h' + h(Q + u)R^{-1}] R \times [h'^T + (R^{-1})^T (Q^T + u^T) h^T] \tag{8}$$

if

$$u = u^T$$

and

$$(Q + u)R^{-1}(Q^T + u) - P - \frac{du}{dx} = 0. \tag{9}$$

The system of equations (9) is the Ricatti system, which is of first order but nonlinear. We convert this nonlinear set of first-order differential equations to a set of linear second-order equations in the entire range $[a,b]$ so that we may obtain a solution, thereby demonstrating that the expression in square brackets in $\delta^2 W[h]$ can indeed be written as in (8).

We substitute

$$Q^T + u = RY'Y^{-1} \tag{10}$$

in Eq. (9), where Y is a $n \times n$ matrix and $Y' = dY/dx$.

On post multiplying the resulting equation by Y and rearranging terms one arrives at the Jacobi system of equations

$$\frac{d}{dx}(RY') + (Q^T - Q)Y' + \left(\frac{dQ^T}{dx} - P \right) Y = 0. \tag{5}$$

In this form (5) the term $Q^T \neq Q$ and one deals with the general unsymmetric form of the second variation (7). This is a system of n second-order *linear* differential equations. Given the set of boundary conditions (6a) and (6b), the solution of the Jacobi set of equations is guaranteed and tracing back the line of argument it can be shown that the second variation $\delta^2 W[h]$ can be written as an integral over x of the square bracket of the form (8). Since R is a positive definite matrix

and h is an arbitrary row vector, the integral $\delta^2 W[h]$ is necessarily positive definite.

However, in this chain of arguments the identity (10) represents a crucial step. In order to convert Ricatti equations into the Jacobi equations, Y^{-1} must exist. This means that the det $|Y|$ must not vanish anywhere in the closed interval $[a,b]$ except at the point $x = a$. This is precisely the condition of non-existence of a conjugate point at $x = \tilde{a}$ in the interval $[a,b]$. To reiterate, the conjugate point $x = \tilde{a}$ (conjugate to $x = a$) was defined to be the point at which

$$\det Y(\tilde{a}) = 0$$

subject to the boundary conditions (6a) and (6b) for the Jacobi system (5).

Therefore we have found the condition according to Theorem 2 for the positive definiteness of a second variation of the unsymmetric energy functional of an n-dimensional trial function. We note that because of the lack of symmetry $Q \neq Q^T$, the Jacobi system of equation (8) is not identical to its transposed system. Therefore, the solution $Y(x)$ will not be identical to $Y^T(x)$. In any case, the condition of vanishing of the det $Y(x)$ is independent of whether we use Eqs. (5) or its transpose.

It is to be noted (see Ref. 1, p. 108) that if the interval $[a,b]$ contains no points conjugate to a, then $[a,b]$ contains no points conjugate to $a - \epsilon$, for some sufficiently small ϵ, since the solution of the differential equation system (5) depends continuously on the initial conditions. Therefore the solution which satisfies the initial conditions $Y(a - \epsilon) = \emptyset, Y'(a - \epsilon) = I$, does not vanish anywhere in the interval $[a,b]$. In this argument it is implicitly assumed that $R(x) \neq 0$ everywhere in $[a,b]$; if this condition holds, Y need not vanish on the left boundary. Also, if $Y(x)$ is a solution of (5) that is not identically zero and satisfies the conditions $Y(a) = Y(c) = 0$ where $a < c \leqslant b$, then $CY(x)$ is also such a solution, where $C = \text{const} \neq 0$. One can impose some kind of normalization on $Y(x)$, and C can be chosen to make $Y'(a) = I$.

The plasma physical applications of these theorems and their astrophysical implications are discussed elsewhere.[5]

ACKNOWLEDGMENTS

The author wishes to thank Professor Meinhard Mayer for discussions and comments on the manuscript. He also wishes to thank members of the Irvine Solar Physics group for bringing this problem to the author's attention and for their interest in its completion.

This research was supported by grants from the National Science Foundation and the National Aeronautics and Space Administration.

[1] I. M. Gel'fand and S. V. Fomin, *Calculus of Variations* (Prentice Hall, Englewood Cliffs, N. J., 1963).
[2] W. A. Newcomb, Ann. Phys. **10**, 232 (1960).
[3] I. B. Bernstein, E. A. Frieman, M. D. Kruskal, and R. M. Kulsrud, Proc. Roy. Soc. London Ser. A **244**, 17 (1958).
[4] H. Sagan, *Introduction to the Calculus of Variations* (McGraw-Hill, New York, 1969).
[5] A. Ray and G. Van Hoven, Phys. Fluids **25**, 1355 (1982).

The fundamental identity for iterated spherical means and the inversion formula for diffraction tomography and inverse scattering

G. Beylkin
Courant Institute of Mathematical Sciences,[a] New York University, 251 Mercer Street, New York, New York 10012

(Received 21 September 1982; accepted for publication 23 December 1982)

A uniform derivation of the inversion formula for diffraction tomography and inverse scattering from the fundamental identity for iterated spherical means is presented.

PACS numbers: 02.30. + g, 03.80. + r, 03.40.Kf

I. INTRODUCTION

In the book *Plane Waves and Spherical Means,*[1] John derives a relation between the iterated spherical mean and the spherical mean of a function. Using this relation, we have obtained a generalization of an important inversion formula that was recently derived by Devaney for use in diffraction tomography[2,3] and inverse scattering.[4]

Scattering experiments are usually designed to yield the scattered field as a function of two unit vectors which represent the direction of the incident wave and the direction at which the scattered field is recorded. The inversion formulae[2–4] reconstruct the scatterer from the measured scattered field which is a function of two unit vectors.

The relation we derive here can also be considered as the representation of a function in terms of the iterated spherical mean of its Fourier transform, provided the Fourier transform has compact support.

Our derivation is independent of the dimension of the space and thus generalizes the formulae derived in Refs. 2–4 to spaces of n dimensions.

II. THE FUNDAMENTAL IDENTITY FOR ITERATED SPHERICAL MEANS

Let f be a continuous function in a domain $D \subset R^n$. Set

$$I(x,r) = \frac{1}{\omega_n} \int_{|\xi|=1} f(x + r\xi)\, d\omega_\xi, \tag{1}$$

where $\omega_n = 2\pi^{n/2}/\Gamma(n/2)$ is the surface area of the unit sphere in R^n, $x \in R^n$, $\xi \in R^n$ is a unit vector, and $d\omega_\xi$ is the solid angle differential form.

The function $I(x,r)$ is the average of the function f on a sphere of radius $|r|$ about the point x. The spherical mean $I(x,r)$ is even with respect to r.

Let us define the iterated spherical mean $M(x,\lambda,\mu)$ by the following formula:

$$M(x,\lambda,\mu) = \frac{1}{\omega_n} \int_{|\eta|=1} I(x + \lambda\eta,\mu)\, d\omega_\eta,$$

where λ,μ are real numbers and $\eta \in R^n$ is a unit vector. Alternatively, we can write

$$M(x,\lambda,\mu) = \frac{1}{\omega_n^2} \int_{|\xi|=1} \int_{|\eta|=1} f(x + \lambda\eta + \mu\xi)\, d\omega_\eta d\omega_\xi. \tag{2}$$

John obtained the fundamental identity for the iterated spherical mean $M(x,\lambda,\mu)$ (formula 4.9c[1]):

$$M(x,\lambda,\mu) = [2\omega_{n-1}/(2\lambda\mu)^{n-2}\omega_n] \times \int_{\mu-\lambda}^{\mu+\lambda} [(r+\mu-\lambda)(r+\mu+\lambda)(\lambda+r-\mu) \times (\lambda-r+\mu)]^{(n-3)/2} r I(x,r)\, dr. \tag{3}$$

In the next section we use this identity to derive a formula relating a function to the iterated spherical mean of its Fourier transform.

III. THE INVERSION FORMULA FOR DIFFRACTION TOMOGRAPHY AND INVERSE SCATTERING

We set $\lambda = \mu = k$ in (3) and obtain

$$M(x,k,k) = \frac{\omega_{n-1}}{2^{n-3} k^{2n-4} \omega_n} \times \int_0^{2k} (4k^2 - r^2)^{(n-3)/2} r^{n-2} I(x,r)\, dr. \tag{4}$$

Let $\hat{F}$ be a function with support inside the n-dimensional ball B_{2k},

$$B_{2k} = \{ p : |p| < 2k \}.$$

In scattering theory the surface of this ball is called the Ewald limiting sphere.

Consider the function

$$f_y(p) = [|p|/(4k^2 - |p|^2)^{(n-3)/2}] \hat{F}(p) e^{ip\cdot y},$$

where $p \in R^n$ and $y \in R^n$ is a parameter.

We compute the spherical mean $I_y(0,r)$ for the function $f_y(p)$. From the definition (1) it follows that

$$I_y(0,r) = \frac{1}{\omega_n} \frac{r}{(4k^2 - r^2)^{(n-3)/2}} \int_{|\xi|=1} \hat{F}(r\xi) e^{ir\xi\cdot y}\, d\omega_\xi.$$

Making use of relation (4), we find

$$M_y(0,k,k) = \frac{\omega_{n-1}}{2^{n-3} k^{2n-4} \omega_n^2} \int_0^{2k} r^{n-1}\, dr \times \int_{|\xi|=1} \hat{F}(r\xi) e^{ir\xi\cdot y}\, d\omega_\xi. \tag{5}$$

Since the support of the function $\hat{F}(p)$ is contained in the ball B_{2k} the function $F(y)$ defined by

$$F(y) = \frac{1}{(2\pi)^n} \int_{|p|<2k} \hat{F}(p) e^{ip\cdot y}\, dp \tag{6}$$

coincides with the inverse Fourier transform of $\hat{F}$. Thus we have from (5)

[a] Written while a summer employee at Schlumberger-Doll Research, P.O. Box 307, Ridgefield, CT.

 0022-2488/83/061399-02$02.50 © 1983 American Institute of Physics

$$M_y(0,k,k) = (8\pi^n \omega_{n-1}/k^{2n-4}\omega_n^2) F(y). \tag{7}$$

Also, from the definition (2) of the iterated spherical mean for the function $f_y(p)$ we obtain

$$M_y(0,k,k) = \frac{1}{k^{n-4}\omega_n^2} \int_{|\xi|=1} \int_{|\eta|=1} \frac{|\xi+\eta|}{(4-|\xi+\eta|^2)^{(n-3)/2}} \times \hat{F}(k\xi + k\eta) e^{ik(\xi+\eta)\cdot y}\, d\omega_\eta d\omega_\xi. \tag{8}$$

We can replace the unit vector η by $-\eta$ in the formula (8). Comparing (7) and (8), we finally obtain the representation

$$F(y) = \frac{k^n}{8\pi^n \omega_{n-1}} \int_{|\xi|=1} \int_{|\eta|=1} \frac{|\xi-\eta|}{(4-|\xi-\eta|^2)^{(n-3)/2}} \times \hat{F}(k\xi - k\eta) e^{ik(\xi-\eta)\cdot y}\, d\omega_\eta d\omega_\xi. \tag{9}$$

For $n = 2$ and for $n = 3$ the identity (9) reduces to the formulae obtained by Devaney.[2,4]

If the support of $\hat{F}$ in (6) is not restricted to the ball B_{2k}, then (6) defines the so-called low-pass-filtered version of the function whose Fourier transform is $\hat{F}$, and (9) provides a representation for the low pass filtered version of that function.

In diffraction tomography and in inverse scattering the function $\hat{F}(k\xi - k\eta)$ represents the measured data.[2–4]

In the case of diffraction tomography the function $F(y)$ defined by (6) is the low-pass-filtered version of the two-dimensional object profile.[2]

In the case of inverse scattering the function $F(y)$ defined by (6) is the low-pass-filtered version of the so called interatomic distance function.[4]

In both cases formula (9) provides the basis for reconstruction algorithms.[2–4]

We note that formula (9) can also be written in the form

$$F(y) = \frac{k^n}{2^{(n+2)/2}\pi^n \omega_{n-1}} \int_{|\xi|=1} \int_{|\eta|=1} \frac{(1-\xi\cdot\eta)^{1/2}}{(1+\xi\cdot\eta)^{(n-3)/2}} \times \hat{F}(k\xi - k\eta) e^{ik(\xi-\eta)\cdot y}\, d\omega_\eta d\omega_\xi. \tag{10}$$

ACKNOWLEDGMENTS

The author would like to thank A. J. Devaney for pointing out the problem and for valuable discussions.

[1] F. John, *Plane Waves and Spherical Means* (Interscience, New York, 1955).

[2] A. J. Devaney, "A filtered backpropagation algorithm for diffraction tomography," Ultrasonic Imaging **4**, 336–50 (1982).

[3] A. J. Devaney, "A computer simulation study of diffraction tomography," IEEE Trans. Biomedical Eng. (to be published).

[4] A. J. Devaney, "Inversion formula for inverse scattering within the Born approximation," Opt. Lett. **1**, 111–2 (1982).

Exact solution of a nonlinear Langevin equation with applications to photoelectron counting and noise-induced instability

K. Wódkiewicz[a] and M. S. Zubairy[b]
Max-Planck-Institut für Quantenoptik, D-8046 Garching bei München, Federal Republic of Germany

(Received 20 July 1982; accepted for publication 30 December 1982)

A stochastic differential equation of the Langevin type involving a linear and a quadratic noise is considered and an exact solution for the stochastic average of this equation is obtained. This solution is employed to calculate the effect of amplitude fluctuations of an electromagnetic field on the photoelectron counting, and an explicit expression is obtained for the photon counting probability. Finally, the steady state limit is studied and the possibility of an instability induced by the noise is discussed.

PACS numbers: 02.30.Hq

I. INTRODUCTION

A very wide class of physical problems ranging from radiative decay, optical pumping, chemical reactions, and laser threshold properties to population dynamics and multiphoton physics can be described by the following simple rate equation[1]:

$$\frac{d\rho}{dt} = R\rho. \tag{1}$$

It has been shown that in many physical applications the rate R depends on time and that the standard notion of a rate can be introduced only if a steady state limit of R can be obtained. When ρ describes the degrees of freedom of an open system exposed to an external driving noise $x(t)$, the rate R depends on the noise itself.[2]

If an atomic system is driven by an external source of light with a fluctuating electric field with amplitude $E(t)$, it is known that the one-photon effects are due to the intensity of the driving field $I(t) = E^2(t)$. This means that the simplest rate equation with a stark broadening[3,4] or a one-photon transition[5,6] will lead to a rate R, which depends on the square of the driving noise. Noise effects in electrical circuits[7] and photoelectron counting[8] also lead to rates which depend on quadratic Gaussian noise. A time-dependent rate with a linear and quadratic $\chi(t)$ occurs in the effect of external noise influence on the electrodynamics instability as well.[9] This leads to the formation of William domains in liquid crystals. In most of these applications the external noise has a finite coherence time τ_c and accordingly a well-defined power spectrum with bandwidth $1/\tau_c$.

In this paper we shall study the rate equation (1) with a general rate involving a linear and quadratic noise. We shall write this rate in the following general form:

$$R(t) = F(t) + ix(t)G(t) + ix^2(t)H(t), \tag{2}$$

where F, G, and H are arbitrary functions of time and the external Gaussian driving noise $x(t)$ is modeled on the Ornstein–Uhlenbeck process [see Eq. (8)] with the following correlation function:

$$\langle x(t)x(t')\rangle = (\Gamma/\tau_c)e^{-|t-t'|/\tau_c}. \tag{3}$$

In this case the rate equation (1) becomes a Langevin equation driven by a multiplicative linear and quadratic Ornstein–Uhlenbeck stochastic process with strength Γ/τ_c and coherence time τ_c.

In this paper we obtain an exact solution for the statistical average of the Langevin equation (1). The exact stochastic average satisfies again an exact rate equation:

$$\frac{d\langle\rho\rangle}{dt} = \mathscr{R}\langle\rho\rangle, \tag{4}$$

where the new rate $\mathscr{R}$ is completely determined from a function $F(t)$ and two nonlinear differential equations involving $G(t)$ and $H(t)$.

In Sec. II we solve $\langle\rho(t)\rangle$ in terms of the given functions F, G, and H, and we derive the two required differential equations. As an important example of our exact solution, we calculate, in Sec. III, the effect of amplitude fluctuations of an electric field on the photoelectron counting. Explicit expressions are obtained for the photon-counting probability in this case.

Finally in Sec. IV we obtain an exact form of the rate $\mathscr{R}$ for arbitrary functions F, G, and H in the steady state limit. The possibility of an instability induced by the noise is mentioned briefly.

II. EXACT STOCHASTIC AVERAGE OF ρ

In order to derive an exact equation for the mean value of ρ, we define the following auxiliary quantity:

$$g(t,\lambda) = e^{\lambda^2\Gamma/2\tau_c}e^{i\lambda x}\rho(t). \tag{5}$$

This quantity, after a stochastic average over $x(t)$, satisfies a λ-independent initial value:

$$\langle\rho(0)\rangle = \langle g(0,\lambda)\rangle = \rho(0), \tag{6}$$

where we have assumed that $\rho(0)$ is statistically independent of the noise $x(t)$.

From the definition (5) we conclude that the stochastic average of ρ is given by $\langle g(t,\lambda)\rangle$ evaluated at the point $\lambda = 0$:

$$\langle\rho(t)\rangle = \langle g(t,\lambda)\rangle|_{\lambda=0}. \tag{7}$$

From the definition of the Ornstein–Uhlenbeck stochastic process, $x(t)$ satisfies the following Langevin equation[10]:

[a] Permanent address: Institute of Theoretical Physics, Warsaw University, Warsaw 00-681, Poland.
[b] Permanent address: Institute for Modern Optics, Department of Physics and Astronomy, University of New Mexico, Albuquerque, New Mexico 87131.

 0022-2488/83/061401-04$02.50 © 1983 American Institute of Physics

$$\frac{dx}{dt} = -\frac{1}{\tau_c}x + x_w(t), \tag{8}$$

where $x_w(t)$ is a Gaussian white noise, i.e.,

$$\langle x_w(t)\rangle = 0,$$

$$\langle x_w(t)x_w(t')\rangle = \frac{2\Gamma}{\tau_c^2}\delta(t-t'). \tag{9}$$

We then obtain the following differential equation for $g(t,\lambda)$:

$$\frac{\partial g}{\partial t} = [M_0 + ix_w(t)M]g, \tag{10}$$

where M_0 and M are two linear operators given by the following expressions:

$$\begin{aligned} M_0 = &\left(-\frac{1}{\tau_c} + 2i\frac{\Gamma}{\tau_c}H(t)\right)\lambda\frac{\partial}{\partial\lambda} \\ &+ F(t) + i\frac{\Gamma}{\tau_c}H(t) \\ &- \frac{\Gamma}{\tau_c}G(t)\lambda - iH(t)\frac{\partial^2}{\partial\lambda^2} \\ &+ G(t)\frac{\partial}{\partial\lambda} - i\frac{\Gamma^2}{\tau_c^2}H(t)\lambda^2 + \frac{\Gamma}{\tau_c^2}\lambda^2, \end{aligned} \tag{11a}$$

$$M = \lambda. \tag{11b}$$

The stochastic differential equation for g with a white-noise $x_w(t)$ is known to lead to the following exact equation satisfied by the mean vlaue of $\langle g\rangle$[11,12]:

$$\frac{d\langle g\rangle}{dt} = \left(M_0 - \frac{\Gamma}{\tau_c^2}M^2\right)\langle g\rangle. \tag{12}$$

In order to perform the limit $\lambda = 0$ [see Eq. (7)], we introduce the annihilation and creation operators $a = \partial/\partial\lambda$ and $a^+ = \lambda$. With the help of this notation Eq. (12) can be written in the following form:

$$i\frac{\partial}{\partial t}\langle g\rangle = H(a^+,a,t)\langle g\rangle, \tag{13}$$

where

$$\begin{aligned} H(a^+,a,t) = i[&M_0(a^+,a,t) \\ &- (\Gamma/\tau_c^2)M^2(a^+,a,t)]. \end{aligned} \tag{14}$$

This quadratic "Hamiltonian" (H is not a self-adjoint operator) can be written in the following form:

$$\begin{aligned} H = &\omega(t)a^+a + f_1(t)a^2 + f_2(t)a^{+2} \\ &+ g_1(t)a + g_2(t)a^+ + c(t), \end{aligned} \tag{15}$$

where

$$\begin{aligned} &\omega(t) = i[-1/\tau_c + (2i\Gamma/\tau_c)H(t)], \\ &f_1(t) = H(t), \quad f_2(t) = (\Gamma^2/\tau_c^2)H(t), \\ &g_1(t) = iG(t), \quad g_2(t) = -i(\Gamma/\tau_c)G(t), \\ &c(t) = iF(t) - (\Gamma/\tau_c)H(t). \end{aligned} \tag{16}$$

Because we need only the value of $\langle g(t,\lambda)\rangle$ at the point $\lambda = 0$ [see Eq. (7)], the mean value of ρ is given by the vacuum expectation value of the time evolution operator associated with Eq. (13). This is easily realized by noting that, in view of Eq. (6), $a = \partial/\partial\lambda$ acting on the right gives zero, and $a^+ = \lambda$ acting on the left is zero by definition. We then obtain

$$\langle\rho(t)\rangle = \left\langle 0\middle| T\exp\left(-i\int_0^t d\tau\, H(a^+,a,\tau)\right)\middle|0\right\rangle\rho(0), \tag{17}$$

where T is the time ordering operator and the vacuum state $|0\rangle$ is defined so that $a|0\rangle = 0$. The simplest way of calculating this vacuum expectation value is to transform the time evolution operator in Eq. (17) into a normal product using the Wick theorem. This can also be achieved by a direct calculation of the propagator

$$K(\alpha,t;\alpha',t') = \langle\alpha|U(t,0)^{-1}U(t',0)|\alpha'\rangle \tag{18}$$

in a coherent state representation[13] with $\alpha = \alpha' = t' = 0$, where $U(t,0)$ is the time evolution operator of Eq. (13). The explicit form of this propagator can be determined in terms of a path integral. Since our "Hamiltonian" (15) is only quadratic in the annihilation and creation operators, an exact path integration can be carried out.[14] As a result we obtain

$$\begin{aligned} \frac{d}{dt}\langle\rho\rangle &= \mathscr{R}(t)\langle\rho\rangle \Rightarrow \langle\rho(t)\rangle \\ &= \exp\left[\int_0^t \mathscr{R}(\tau)\,d\tau\right]\rho(0), \end{aligned} \tag{19}$$

where the new rate $\mathscr{R}$ is given by the following formula[14]:

$$\begin{aligned} \mathscr{R}(t) = &-2if_1(t)X(t) - if_1(t)Z^2(t) \\ &- ig_1(t)Z(t) - ic(t), \end{aligned} \tag{20}$$

and the functions $X(t)$ and $Z(t)$ satisfy the following differential equations with the initial values $X(0) = Z(0) = 0$:

$$\frac{dX}{dt} = -2i\omega(t)X - 4if_1(t)X^2 - if_2(t), \tag{21a}$$

$$\begin{aligned} \frac{dZ}{dt} = &-i[\omega(t) + 4f_1(t)X]Z \\ &- i[g_2(t) + 2g_1(t)X]. \end{aligned} \tag{21b}$$

This set of nonlinear differential equation with the formula (20) completes our exact solution of the stochastic problem given by the Langevin equation (1) with $R(t)$ given by (2). For given functions F, H, and G we can calculate an exact form of the rate. Note that only the case $F = G = 0$ and $H = \text{const}$ has been investigated so far in the literature.[5,7] For arbitrary functions F, G, and H only approximate solutions in steady state have been obtained recently.[15]

III. AMPLITUDE FLUCTUATIONS IN THE PHOTOELECTRON COUNTING

The statistics of the counts obtained by a photodetector, which is illuminated by a statistically stationary light beam, is described by the probability $P_n(t)$ of obtaining counts in the time interval t. The distribution $P_n(t)$ is given by Mandel's formula[16]

$$P_n(t) = \langle e^{-U}U^n/n!\rangle, \tag{22}$$

where

$$U = \alpha\int_0^t I(\tau)\,d\tau, \tag{23}$$

α is a constant which represents the photosensitivity of the photocathode, $I(t) = E^2(t)$ is the intensity of the incident

field. We want to calculate the expectation value $\langle P_n(t)\rangle$ in the presence of the amplitude fluctuations of the incident field.

It is evident from Eq. (22) that $\langle P_n(t)\rangle$ can be determined from $\langle P_0(t)\rangle$ by the formula

$$\langle P_n(t)\rangle = \frac{(-t)^n}{n!}\frac{\partial^n\langle P_0(t)\rangle}{\partial t^n}. \tag{24}$$

Thus an evaluation of $\langle P_0(t)\rangle$ is sufficient to determine $\langle P_n(t)\rangle$.

If we write $E(t) = E_0 + \delta E(t)$, where $\delta E(t)$ is the amplitude fluctuation of the field, then it follows from Eq. (22) that $P_0(t)$ satisfies the following stochastic equation:

$$\frac{\partial P_0}{\partial t} = -\alpha(E_0^2 + 2E_0\delta E(t) + \delta E^2(t))P_0. \tag{25}$$

This equation has the same form as the Langevin equation [Eqs. (1) and (2)] with

$$F(t) \equiv -\alpha E_0^2, \tag{26a}$$

$$G(t) \equiv 2i\alpha E_0, \tag{26b}$$

$$H(t) \equiv i\alpha. \tag{26c}$$

If we assume $\delta E(t)$ to be an Ornstein–Uhlenbeck process with a correlation function of the form (3), then the mean value $\langle P_0(t)\rangle$ can be determined exactly using the metod developed in the last section. The X and Z equations (21a) and (21b) subject to the initial values $X(0) = Z(0) = 0$ are solved exactly and the resulting equations are

$$X(t) = -\frac{A}{4\alpha}\left(\frac{\sinh(At) + B\cosh(At)}{\cosh(At) + B\sinh(At)}\right) + \frac{AB}{4\alpha}, \tag{27a}$$

$$Z(t) = -2i\Gamma\alpha E_0 \times\left[\frac{1}{2\Gamma\alpha} - \frac{1}{(1+4\Gamma\alpha)[\cosh(At) + B\sinh(At)]} - \frac{1}{2\Gamma\alpha\sqrt{1+4\Gamma\alpha}}\left(\frac{\sinh(At) + B\cosh(At)}{\cosh(At) + B\sinh(At)}\right)\right], \tag{27b}$$

where

$$A = (1/\tau_c)\sqrt{1+4\Gamma\alpha}, \tag{28a}$$

$$B = (1+2\Gamma\alpha)/\sqrt{1+4\Gamma\alpha}. \tag{28b}$$

The mean value $\langle P_0(t)\rangle$ is then given, according to Eq. (20), by

$$\langle P_0(t)\rangle = e^{\int_0^t \mathscr{R}(\tau)\,d\tau}, \tag{29}$$

where

$$\mathscr{R}(\tau) = 2\alpha X(\tau) + \alpha Z^2(\tau) + 2iE_0\alpha Z(\tau) - \alpha E_0^2 - \Gamma\alpha/\tau_c. \tag{30}$$

The integrals in Eq. (29) can be evaluated, and they are given in the Appendix.

IV. STEADY STATE SOLUTION OF THE STOCHASTIC RATE EQUATION

As has been argued in the Introduction, the notion of a rate has a clear and physically appealing interpretation only when the evolution of ρ is given by an exponential function. This can be achieved in the steady state limit of $\mathscr{R}(t)$, which is obtained by a limiting procedure with the time t going to infinity (or, if this limit does not exist due to a periodic time dependence, we then substitute the values of the functions averaged over a period).

In this case we obtain from Eqs. (13) and (14)

$$\langle\rho\rangle = \exp(\bar{\mathscr{R}}t)\,\rho(0) \tag{31}$$

with

$$\bar{\mathscr{R}} = -2i\bar{f}_1\bar{X} - i\bar{f}_1\bar{Z}^2 - i\bar{g}_1\bar{Z} - i\bar{c}, \tag{32}$$

where the bar over the functions f_1, g_1, etc. indicates the appropriate limiting or averaged values. Setting $\dot{X} = 0$ and $\dot{Y} = 0$, we calculate from Eqs. (21a) and (21b) the steady state $\bar{X}$ and $\bar{Y}$, and accordingly we obtain

$$\bar{\mathscr{R}} = \frac{1}{2\tau_c} + \bar{F} - \frac{1}{2\tau_c}\sqrt{1-4i\Gamma\bar{H}} + \frac{i\bar{G}^2}{4\bar{H}(1-4i\Gamma\bar{H})} - \frac{i\bar{G}^2}{4\bar{H}}. \tag{33}$$

In the absence of quadratic noise, i.e., when $H = 0$,

$$\bar{\mathscr{R}} = -\Gamma\bar{G}^2 + \bar{F}, \tag{34}$$

and, if $\bar{F} > \Gamma\bar{G}^2$ the rate remains negative, i.e., $\langle\rho\rangle$ is approaching zero with increasing time. With a quadratic noise we can have a noise-induced instability, i.e., for some values of $\bar{H}$, $\bar{\mathscr{R}}$ can be positive and $\langle\rho\rangle$ diverges. The condition $\bar{\mathscr{R}} = 0$ gives a threshold condition for such a noise-induced change of the statistical properties of $\langle\rho\rangle$. Such a change of $\langle\rho\rangle$ due to quadratic noise has been observed experimentally in the effect of external noise in an electrohydrodynamic transition in liquid crystals characterized by the appearance of Williams domains.[3]

ACKNOWLEDGMENT

The authors would like to thank Professor H. Walther for his invitation to the Max-Planck Institute for Quantum Optics in Garching, where this work was done.

APPENDIX

Here we write the various integrals that occur in Eq. (29):

$$\int_0^t X(\tau)\,d\tau = \frac{ABt}{4\alpha} - \frac{1}{4\alpha}\times\ln(\cosh(At) + B\sinh(At)), \tag{A1}$$

$$\int_0^t Z(\tau)\,d\tau = -2i\Gamma\alpha E_0 \times\left\{\frac{t}{2\Gamma\alpha} - \frac{1}{(1+2\Gamma\alpha)(1+4\Gamma\alpha)} \times\ln\left(\frac{B(\cosh(At) + B\sinh(At))}{\sinh(At) + B\cosh(At)}\right) - \frac{1}{2\Gamma\alpha\tau_c}\ln(\cosh(At) + B\sinh(At))\right\}, \tag{A2}$$

$$\int_0^t Z^2(\tau)\,d\tau = -4\Gamma^2\alpha^2 E_0^2$$
$$\times\left[\frac{1}{4\Gamma^2\alpha^2}\left(1+\frac{1}{\tau_c^2 A^3}\right)t+\frac{1+4\Gamma^2\alpha^2}{A(1+4\Gamma\alpha)^2}\right.$$
$$\times\frac{\sinh(At)}{\cosh(At)+B\sinh(At)}-\frac{\tau_c}{\alpha\Gamma(1+2\Gamma\alpha)(1+4\Gamma\alpha)}$$
$$\times\ln\left(\frac{B(\cosh(At)+B\sinh(At))}{\sinh(At)+B\cosh(At)}\right)-\frac{\tau_c}{2\Gamma\alpha^2(1+4\Gamma\alpha)}$$
$$\times\ln(\cosh(At)+B\sinh(At))$$
$$\left.-\frac{\tau_c}{\Gamma\alpha(1+4\Gamma\alpha)^2}\left(\frac{1}{\cosh(At)+B\sinh(At)}-1\right)\right], \quad \text{(A3)}$$

where, as before,

$$A=(1/\tau_c)\sqrt{1+4\Gamma\alpha}, \quad \text{(A4)}$$

$$B=1+2\Gamma\alpha/\sqrt{1+4\Gamma\alpha}. \quad \text{(A5)}$$

[1]See, for example, R. Loudon, *Quantum Theory of Light* (Oxford U.P., Oxford, 1973).
[2]H. Haken, Rev. Mod. Phys. **47**, 65 (1975).
[3]A. T. Georges and P. Lambropoulos, Phys. Rev. A **20**, 991 (1979).
[4]P. Zoller, Phys. Rev. A **20**, 1019 (1979).
[5]P. Zoller and P. Lambropoulos, J. Phys. B **13**, 69 (1980).
[6]J. J. Yeh and J. H. Eberly, Phys. Rev. A **24**, 888 (1981).
[7]D. Slepian, Bell Syst. Tech. J. **37**, 165 (1958).
[8]See, for example, R. Barakat and J. Blake, Phys. Rep. **60**, 225 (1980).
[9]T. Kawakubo, A. Yanagita, and S. Kabashima, J. Phys. Soc. Jpn. **50**, 1451 (1981).
[10]See, for example, *Selected Papers on Noise and Stochastic Processes*, edited by N. Wax (Dover, New York, 1954).
[11]R. F. Fox, Phys. Rep. **48**, 179 (1978).
[12]K. Wódkiewicz, J. Math. Phys. **20**, 45 (1979); **23**, 2179 (1983).
[13]J. R. Klauder, Ann. Phys. (N.Y.) **11**, 123 (1960).
[14]M. Hillery and M. S. Zubairy, Phys. Rev. A **26**, 451 (1982).
[15]M. San Miguel and J. M. Sancho, Z. Phys. B: Condensed Matter **43**, 361 (1981).
[16]See, for example, L. Mandel and E. Wolf, Rev. Mod. Phys. **37**, 231 (1965).

The Painlevé property for partial differential equations. II: Bäcklund transformation, Lax pairs, and the Schwarzian derivative[a)]

John Weiss
La Jolla Institute, P. O. Box 1434, La Jolla, California 92038
Institute of Pure and Applied Science, University of California, San Diego, La Jolla, California 92093

(Received 20 July 1982; accepted for publication 7 January 1983)

In this paper we investigate the Painlevé property for partial differential equations. By application to several well-known partial differential equations (Burgers, KdV, MKdV, Bousinesq, higher-order KdV and KP equations) it is shown that consideration of the "singular manifold" leads to a formulation of these equations in terms of the "Schwarzian derivative." This formulation is invariant under the Moebius group (acting on dependent variables) and is shown to obtain the appropriate Lax pair (linearization) for the underlying nonlinear pde.

PACS numbers: 02.30.Jr

1. INTRODUCTION

We say[1] that a partial differential equation has the Painlevé property when the solutions of the pde are "single-valued" about the movable, singularity manifold. To be precise, if the singularity manifold is determined by

$$\phi(z_1,\ldots,z_n)=0 \tag{1.1}$$

and $u=u(z_1,\ldots,z_n)$ is a solution of the partial differential equation, then we assume that

$$u=\phi^{\alpha}\sum_{j=0}^{\infty}u_j\phi^j, \tag{1.2}$$

where

$$\phi=\phi(z_1,\ldots,z_n),\quad u_j=u_j(z_1,\ldots,z_n),\quad u_0\neq 0,$$

are analytic functions of (z_j) in a neighborhood of the manifold (1.1) and α is an integer. Substitution of (1.2) into the partial differential equation determines the value(s) of α and defines the recursion relations for u_j, $j\neq 0, 1, 2, \ldots$. When the anzatz (1.2) is correct, the pde is said to possess the Painlevé property and is conjectured to be integrable.

In this paper we demonstrate a connection between the Painlevé property and the occurrence of the Lax pairs (and Bäcklund transforms) for several well-known nonlinear pde's. These include the Burgers', KdV, MKdV, Bousinesq, higher-order KdV, and Kadomtsev–Petviashvili equations. This connection is most clearly formulated in terms of the "Schwarzian derivative"[2,3] of the function defining the singular manifold, (1.1). Consideration of the Painlevé property associated with the Schwarzian derivative leads to the identification of a wide class on nonlinear pde's that possess, in part, the Painlevé property. With some care, it appears possible to extend this identification to equations in any (finite) number of independent variables. Study of this class of equations is currently in progress.

In Sec. 2 we develop, at some length, the above procedures for the Kortweg–de Vries equation. Then in the succeeding sections the results for a number of equations are stated.

In Appendix A it is shown how an infinite sequence of "conserved densities," can be defined by the expansion of the solution (of the KdV equation) about the "singularity manifold."

In Appendix B the Harry Dym equation is obtained, by a charge of variables, from the KdV equation.

2. THE KORTEWEG–DE VRIES EQUATION

The KdV equation

$$u_t+uu_x+\delta u_{xxx}=0 \tag{2.1}$$

posses the Painlevé property.[1] The expansion about the singular manifold has the form

$$u=\phi^{-2}\sum_{j=0}^{\infty}u_j\phi^j. \tag{2.2}$$

The recursion relations for the u_j are presented in Appendix A. There it is found that the "resonances" occur at

$$j=-1,4,6. \tag{2.3}$$

Resonances are those values of j at which it is possible to introduce arbitrary functions into the expansion (2.2) (see Refs. 1, 4). For each nontrivial resonance there occurs a compatibility condition that must be satisfied if the solution is to have a single-valued expansion. The resonance at $j=-1$ corresponds to the "arbitrary" function ϕ defining the singular manifold (2.1). The compatability conditions at $j=4$ and 6 are satisfied identically.[1]

From the recusion relations, we find:

$$j=0:\quad u_0=-12\delta\phi_x^2; \tag{2.4}$$

$$j=1:\quad u_1=12\delta\phi_{xx}; \tag{2.5}$$

$$j=2:\quad \phi_x\phi_t+\phi_x^2u_2+4\delta\phi_x\phi_{xxx}-3\delta\phi_{xx}^2=0; \tag{2.6}$$

$$j=3:\quad \phi_{xt}+\phi_{xx}u_2-\phi_x^2u_3+\delta\phi_{xxxx}=0; \tag{2.7}$$

$$j=4:\quad \text{compatibility condition:}$$

$$\frac{\partial}{\partial x}(\phi_{xt}+\delta\phi_{xxxx}+\phi_{xx}u_2-\phi_x^2u_3)=0. \tag{2.8}$$

By (2.7) the compatibility condition (2.8) is satisfied identically. The compatability condition at $j=6$ is also satisfied identically,[1] and the KdV equation is thus shown to be

[a)] This work supported by Department of Energy Contract DOE 10923.

 0022-2488/83/061405-09$02.50 © 1983 American Institute of Physics

Painlevé.

We now specialize (2.2) by setting the "resonance" functions

$$u_4 = u_6 = 0\,. \tag{2.9}$$

Furthermore, by requiring

$$u_3 = 0\,, \tag{2.10}$$

it is easily demonstrated that

$$u_j = 0, \quad j \geqslant 3\,, \tag{2.11}$$

if

$$u_{2t} + u_2 u_{2x} + \delta u_{2xxx} = 0\,. \tag{2.12}$$

We thus obtain the following Bäcklund transformation[5]:

$$u = 12\delta \frac{\partial^2}{\partial x^2} \ln \phi + u_2\,, \tag{2.13}$$

where u and u_2 satisfy (2.1) and

$$\phi_x \phi_t + \phi_x^2 u_2 + 4\delta \phi_x \phi_{xxx} - 3\delta \phi_{xx}^2 = 0\,, \tag{2.14}$$

$$\phi_{xt} + \phi_{xx} u_2 + \delta \phi_{xxxx} = 0\,. \tag{2.15}$$

In Ref. 1 it was shown that the substitution

$$\phi_x = v^2 \tag{2.16}$$

reduces (2.14) and (2.15) to the Lax pair (in v) for the KdV equation.

However, here we shall proceed in a somewhat different manner. By eliminating u_2 from (2.14) and (2.15) we find that

$$\frac{\partial}{\partial x}\left[\frac{\phi_t}{\phi_x} + \delta\left\{+\frac{\partial}{\partial x}\left(\frac{\phi_{xx}}{\phi_x}\right) - \frac{1}{2}\left(\frac{\phi_{xx}}{\phi_x}\right)^2\right\}\right] = 0 \tag{2.17}$$

or

$$\frac{\phi_t}{\phi_x} + \delta\,\{\phi;x\} = \lambda\,, \tag{2.18}$$

where

$$\{\phi;x\} = \frac{\partial}{\partial x}\left(\frac{\phi_{xx}}{\phi_x}\right) - \frac{1}{2}\left(\frac{\phi_{xx}}{\phi_x}\right)^2$$

is the "Schwarzian derivative" of ϕ.[2] Equation (2.18) is invariant under the Moebius group.[3] That is, if

$$\psi = (a\phi + b)/(c\phi + d), \quad ad - bc \neq 0\,, \tag{2.19}$$

and ϕ satisfies (2.18), then ψ satisfies (2.18). This invariance and other well-known properties of the Schwarzian derivative suggests that the substitution

$$\phi = v_1/v_2\,, \tag{2.20}$$

where (v_1, v_2) satisfy the same system of linear equations, might be of use. With the notation

$$\theta_x = v_2 v_{1x} - v_1 v_{2x}\,, \tag{2.21}$$

$$\psi_t = v_2 v_{1t} - v_1 v_{2t}\,, \tag{2.22}$$

it is found that

$$\frac{\psi_t}{\theta_x} + \delta\left(\{\theta;x\} - 2\frac{v_{2xx}}{v_2} + 2\frac{\theta_{xx}}{\theta_x}\frac{v_{2x}}{v_2}\right) = \lambda\,. \tag{2.23}$$

If (v_1, v_2) satisfy

$$v_{xx} = av\,, \tag{2.24}$$

$$v_t = bv_x + cv\,, \tag{2.25}$$

then

$$\theta_{xx} = 0 \tag{2.26}$$

and

$$\psi_t = b\theta_x\,. \tag{2.27}$$

We find, from (2.23), that

$$b = 2\sigma a + \lambda\,. \tag{2.28}$$

By requiring

$$v_{xxt} = v_{txx} \tag{2.29}$$

and using (2.28), it is found that (2.24) and (2.25) are just the Lax pair for the KdV equation.

We note that the Bäcklund transform defined by (2.13) is essentially the formula employed in Ref. 6 for adding or subtracting a bound state to the potential function of the scattering problem (2.24). Furthermore, the constant term λ that appears in Eq. (2.18) is equivalent to the spectral parameter in the second-order scattering problem (2.24). Since the KdV equation is invariant under the Galilean transformation

$$u \to u + \lambda, \quad x \to x + \lambda t, \quad t \to t\,, \tag{2.30}$$

we find that if

$$\phi_t/\phi_x + \delta\,\{\phi;x\} = 0\,, \tag{2.31}$$

then

$$\psi(x,t) = \phi(x + \lambda t, t) \tag{2.32}$$

will satisfy (2.18). This observation will be employed in the next section.

3. THE MODIFIED KdV EQUATION

The modified KdV equation

$$w_t + \delta \frac{\partial}{\partial x}\left(w_{xx} - \frac{w^3}{2}\right) = 0 \tag{3.1}$$

possesses the Painlevé property.[1]

In general,

$$w = \phi^{-1} \sum_{j=0}^{\infty} w_j \phi^j \tag{3.2}$$

with resonances at

$$j = -1, 3, \text{ and } 4. \tag{3.3}$$

Since

$$w_2 = w_3 = w_4 = 0 \tag{3.4}$$

and w_1 satisfying (3.1) implies

$$w_j = 0\,, \quad j \geqslant 2\,, \tag{3.5}$$

the associated Bäcklund transform takes the following form:

$$w = 2\phi_x/\phi + w_1\,, \tag{3.6}$$

where

$$w_1 = \phi_{xx}/\phi_x \tag{3.7}$$

and (w, w_1) satisfy (3.1).

By substitution of (3.7) into (3.1) it is found that

$$\phi_t/\phi_x + \delta\{\phi;x\} = 0, \tag{3.8}$$

where $\{\phi;x\}$ is the Schwarzian derivative. Now this [Eq. (3.8)] is just Eq. (2.18) with $\lambda = 0$. Formally, the Bäcklund transform associated with the modified KdV equation {(3.6) and (3.7)} is degenerate since it does not depend on an arbitrary parameter (the spectral parameter vanishes). Iteration of (3.6) once recreates the initial solution. However, as was noted in the previous section, solutions of (3.8) are simply related to solutions of (3.8) by a partial Galilean transform:

$$x \to x + \lambda t. \tag{3.9}$$

Under the transformation (3.9), the MKdV becomes

$$w_t + \delta\frac{\partial}{\partial x}\left(w_{xx} - \frac{w^3}{2}\right) = \lambda w_x, \tag{3.10}$$

and the Bäcklund transform reads

$$w = 2\phi_x/\phi + w_1, \tag{3.11}$$

where

$$w_1 = \phi_{xx}/\phi_x \tag{3.12}$$

and

$$\phi_t/\phi_x + \delta\{\phi;x\} = \lambda. \tag{3.13}$$

Now identifying Eqs. (3.13) and (2.18), we find that

$$\text{MKdV:}\quad w - w_1 = \frac{2\phi_x}{\phi}, \tag{3.14}$$

$$\text{KdV:}\quad u - u_2 = 12\delta\frac{\partial}{\partial x}\left(\frac{\phi_x}{\phi}\right), \tag{3.15}$$

and

$$u - u_2 = 6\delta\frac{\partial}{\partial x}(w - w_1). \tag{3.16}$$

In addition, using (2.6), we find

$$u_2 + \lambda = 3\delta(w_{1,x} - \tfrac{1}{2}w_1^2), \tag{3.17}$$

$$u + \lambda = 3\delta(w_x - \tfrac{1}{2}w^2). \tag{3.18}$$

(3.17) and (3.18) are Miura transformations, mapping solutions of the MKdV into solutions of the KdV equation. Equations (3.14)–(3.18) indicate that the Miura transformation "commutes" with the Bäcklund transforms defined above. We note that Eq. (3.16) may be used to invert the Miura tranformation.

To complete the Bäcklund transform for the MKdV, we apply the reverse, partial Galilean transform [after application of (3.9) and (3.11)]

$$x \to x - \lambda t \tag{3.19}$$

to arrive at the original equation (3.1). This completes the Bäcklund transform for the MKdV.

4. THE BOUSINESQ EQUATION

The Bousinesq equation

$$u_{tt} + \frac{\partial^2}{\partial x^2}\left(\frac{u_{xx}}{3} + u^2\right) = 0 \tag{4.1}$$

possesses the Painlevé property.[1]

In general, it is found that

$$u = \phi^{-2}\sum_{j=0}^{\infty} u_j\phi^j, \tag{4.2}$$

where the resonances occur at

$$j = -1,4,5,6. \tag{4.3}$$

The compatability conditions at $j = 4,5$, and 6 are satisfied identically. We will require that

$$u_3 = 0 \tag{4.4}$$

and set

$$u_4 = u_5 = u_6 = 0 \tag{4.5}$$

(implying $u_j = 0 \geqslant j \geqslant 3$). We find the associated Bäcklund transform

$$u = 2\frac{\partial^2}{\partial x^2}\ln\phi + u_2, \tag{4.6}$$

where (u,u_2) satisfy (4.1), and

$$\phi_t^2 - \phi_{xx}^2 + \tfrac{4}{3}\phi_x\phi_{xxx} + 2u_2\phi_x^2 = 0, \tag{4.7}$$

$$\phi_{tt} + \tfrac{1}{3}\phi_{xxxx} + 2\phi_{xx}u_2 = 0. \tag{4.8}$$

Now, by eliminating u_2 in (4.7) and (4.8). it is found that

$$\frac{\partial}{\partial t}\left(\frac{\phi_t}{\phi_x}\right) + \frac{1}{2}\frac{\partial}{\partial x}\left(\frac{\phi_t}{\phi_x}\right)^2 + \frac{1}{3}\frac{\partial}{\partial x}\{\phi;x\} = 0, \tag{4.9}$$

where $\{\phi;x\}$ is the Schwarzian derivative.

As before, this equation is invariant under the Moebius group. We let

$$\phi = v_1/v_2 \tag{4.10}$$

and find

$$\frac{\partial}{\partial t}\left(\frac{\psi_t}{\theta_x}\right) + \frac{1}{2}\frac{\partial}{\partial x}\left(\frac{\psi_t}{\theta_x}\right)^2 + \frac{1}{3}\frac{\partial}{\partial x}\left(\{\theta;x\} - 2\frac{v_{2xx}}{v_2} + 2\frac{\theta_{xx}}{\theta_x}\frac{v_{2x}}{v_2}\right) = 0, \tag{4.11}$$

where

$$\theta_x = v_2v_{1x} - v_1v_{2x}, \tag{4.12}$$

$$\psi_t = v_2v_{1t} - v_1v_{2t}, \tag{4.13}$$

and

$$\{\theta;x\} = \frac{\partial}{\partial x}\left(\frac{\theta_{xx}}{\theta_x}\right) - \frac{1}{2}\left(\frac{\theta_{xx}}{\theta_x}\right)^2. \tag{4.14}$$

To solve (4.11), first we let (v_1,v_2) satisfy the following:

Second-order scattering problem:

$$v_{xx} = av, \tag{4.15}$$

$$v_t = bv_x + cv. \tag{4.16}$$

Then,

$$\theta_{xx} = 0,$$

$$\psi_t = b\theta_x,$$

and from (4.11)

$$b_t + bb_x = \tfrac{2}{3}a_x, \tag{4.17}$$

while by compatability of (4.15) and (4.16)

$$2c_x + b_{xx} = 0,$$

$$a_t = c_{xx} + 2ab_x + ba_x, \tag{4.18}$$

or, collecting equations,

$$a_t = -\tfrac{1}{2} b_{xxx} + 2ab_x + ba_x ,$$
$$b_t = -bb_x + \tfrac{2}{3} a_x . \tag{4.19}$$

Unfortunately, it is not possible to introduce an arbitrary, spectral parameter

$$a = u + \lambda \tag{4.20}$$

into Eq. (4.15) without obtaining from (4.19) equations that will depend explicitly on that parameter. Therefore, we shall consider the following:

Third-order scattering problem:

$$v_{xxx} = av_x + bv , \tag{4.21}$$
$$v_t = cv_{xx} + dv_x + ev . \tag{4.22}$$

Then,

$$\psi_t = c\theta_{xx} + d\theta_x , \tag{4.23}$$

and substitution into (4.11) determines

$$c^2 = 1 , \quad d = 0 \tag{4.24}$$
$$ce_x + \tfrac{2}{3} a_x = 0 . \tag{4.25}$$

By the consistency condition for (4.21) and (4.22) we find

$$3e_x + 2ca_x = 0 ,$$
$$a_t = 3e_{xx} + 2cb_x + ca_{xx} , \tag{4.26}$$
$$b_t + ae_x = e_{xxx} + cb_{xx} .$$

Letting

$$a = -\tfrac{3}{2} ce , \tag{4.27}$$

we find

$$ce_{tt} + \frac{\partial^2}{\partial x^2}\left(e^2 + c\frac{e_{xx}}{3}\right) = 0 , \tag{4.28}$$

where

$$c^2 = 1 .$$

Let

$$c = 1;$$

then

e is a solution of the Bousinesq equation

$$a = -\tfrac{3}{2} e ,$$
$$b = \lambda - \tfrac{3}{4} e_x - \tfrac{3}{4}\int^x e_t ,$$
$$c = 1 , \quad d = 0 . \tag{4.29}$$

Equations (4.29) determine the Lax pair for the Bousinesq equation.[7] We note the appearance of the spectral parameter λ.

5. A HIGHER-ORDER KORTEWEG–DE VRIES EQUATION

The higher-order KdV equation[8]

$$u_t = \frac{1}{4}\frac{\partial}{\partial_x}(u_{xxxx} + 5u_x^2 + 10u_{xx}u + 10u^3) \tag{5.1}$$

has an expansion, about the singular manifold, of the form

$$u = \phi^{-2}\sum_{j=0}^{\infty} u_j \phi^j . \tag{5.2}$$

The resonances occur at

(i) $r = -1,2,5,6,8$, when $u_0 = -2\phi_x^2$, (5.3)

(ii) $r = -3,-1,6,8,10$, when $u_0 = -6\phi_x^2$. (5.4)

We note the appearance of two solution branches, with branch (i) depending on five, and branch (ii) depending on four, arbitrary functions.[1,3] In this section we only consider branch (i).

Rather than verify by tedious calculation that the equation possesses the Painlevé property, we shall assume that the equation has an associated Bäcklund transform

$$u = -2\phi_x^2/\phi^2 + u_1/\phi + u_2 . \tag{5.5}$$

By substitution in (5.1) we find that

$$u_0 = -2\phi_x^2 , \quad u_1 = 2\phi_{xx} , \tag{5.6}$$

$$u_2 = -\frac{1}{2}\left[\frac{\partial}{\partial x}\left(\frac{\phi_{xx}}{\phi_x}\right) + \frac{1}{2}\left(\frac{\phi_{xx}}{\phi_x}\right)^2\right] + \lambda , \tag{5.7}$$

$$8\frac{\phi_t}{\phi_x} = 12\frac{\phi_{xxxxx}}{\phi_x} - 10\frac{\phi_{xx}\phi_{xxxx}}{\phi_x^2} + 60\frac{\phi_{xx}}{\phi_x}u_{2x} + 60u_2^2 + 20u_{2xx} + 20\left(\frac{4\phi_{xxx}}{\phi_x} - \frac{3\phi_{xx}^2}{\phi_x^2}\right)u_2 , \tag{5.8}$$

and u_2 must satisfy (5.1).

To verify the consistency of the above Bäcklund transform, we substitute (5.7) and (5.8) and obtain

$$4\frac{\phi_t}{\phi_x} = \frac{\partial^2}{\partial x^2}\{\phi;x\} + \frac{3}{2}\{\phi;x\}^2 + 10\lambda\{\phi;x\} + 30\lambda^2 , \tag{5.9}$$

where

$$\{\phi;x\} = \frac{\partial}{\partial x}\left(\frac{\phi_{xx}}{\phi_x}\right) - \frac{1}{2}\left(\frac{\phi_{xx}}{\phi_x}\right)^2$$

is the Schwarzian derivative.

From Eq. (5.9) we readily obtain the Lax pair. Letting

$$\phi = v_1/v_2 , \tag{5.10}$$

and

$$v_{xx} = av , \quad v_t = bv_x + cv , \tag{5.11}$$

we find from (5.9) that

$$4b = -2a_{xx} + 6a^2 - 20\lambda a + 30\lambda^2 . \tag{5.12}$$

With the consistency conditions

$$2c_x + b_{xx} = 0 ,$$
$$a_t = c_{xx} + 2ab_x + ba_x , \tag{5.13}$$
$$a = \lambda - u ,$$

we find

$$b = \tfrac{1}{2} u_{xx} + \tfrac{3}{2} u^2 + 2\lambda u + 4\lambda^2 ,$$
$$c = -\tfrac{1}{2} b_x + \alpha . \tag{5.14}$$

This defines the Lax pair for (5.1).[8]

6. THE KADOMTSEV–PETVIASHVILI, OR TWO-DIMENSIONAL KdV, EQUATION

The KP equation[9]

$$u_{tx} + u_x^2 + uu_{xx} + \sigma u_{xxxx} + u_{yy} = 0 \tag{6.1}$$

possesses the Painlevé property.[1]

It is found[1] that

$$u = \phi^{-2} \sum_{j=0}^{\infty} u_j \phi^j , \tag{6.2}$$

with resonances at

$$j = -1,4,5,6 . \tag{6.3}$$

The compatability conditions at $j = 4,5$, and 6 are satisfied identically. We set (truncate at constant level)

$$u_3 = u_4 = u_5 = u_6 = 0 .$$

The associated Bäcklund transform is

$$u = 12\sigma \frac{\partial^2}{\partial x^2} \ln \phi + u_2 , \tag{6.4}$$

where

$$\phi_t \phi_x + 4\sigma \phi_x \phi_{xxx} - 3\sigma \phi_{xx}^2 + \phi_y^2 + u_2 \phi_x^2 = 0 \tag{6.5}$$

and

$$\phi_{xt} + \sigma \phi_{xxxx} + \phi_{yy} + \phi_{xx} u_2 = 0 . \tag{6.6}$$

Solving for ϕ, we find that

$$\frac{\partial}{\partial x}\left(\frac{\phi_t}{\phi_x} + \sigma\{\phi;x\}\right) + \frac{\partial}{\partial y}\left(\frac{\phi_y}{\phi_x}\right) + \frac{1}{2}\frac{\partial}{\partial x}\left(\frac{\phi_y^2}{\phi_x^2}\right) = 0 . \tag{6.7}$$

We note that Eq. (6.7) is a combination of Eq. (2.17) for the KdV equation and Eq. (4.9) for the Bousinesq equation. The Lax pair[9] can be readily obtained from (6.7).

7. BURGERS' EQUATION

Burgers' equation

$$u_t + uu_x = \sigma u_{xx} \tag{7.1}$$

possesses the Painlevé property.[1] In general,

$$u = \phi^{-1} \sum_{j=0}^{\infty} u_j \phi^j \tag{7.2}$$

with resonances at

$$j = -1,2 . \tag{7.3}$$

The associated Bäcklund transform is

$$u = -2\sigma \phi_x/\phi + u_1 , \tag{7.4}$$

where (u,u_1) satisfy (7.1) and

$$\phi_t + u_1 \phi_x = \sigma \phi_{xx} . \tag{7.5}$$

From (7.1) and (7.5) we find

$$\frac{\partial}{\partial t}\left(\frac{\phi_t}{\phi_x}\right) - \frac{1}{2}\frac{\partial}{\partial x}\left(\frac{\phi_t}{\phi_x}\right)^2 - 2\sigma \frac{\partial^2}{\partial x^2}\left(\frac{\phi_t}{\phi_x}\right) + \sigma^2 \frac{\partial}{\partial x}\{\phi;x\} = 0 . \tag{7.6}$$

where $\{\phi;x\}$ is the Schwarzian derivative.

As before, we let

$$\phi = v_1/v_2 \tag{7.7}$$

and find

$$\frac{\partial}{\partial t}\left(\frac{\psi_t}{\theta_x}\right) - \frac{1}{2}\frac{\partial}{\partial x}\left(\frac{\psi_t}{\theta_x}\right)^2 - 2\sigma \frac{\partial^2}{\partial x^2}\frac{\psi_t}{\theta_x} + \sigma^2 \frac{\partial}{\partial x}\left(\{\phi;x\} - 2\frac{v_{2xx}}{v_2} + 2\frac{\theta_{xx}}{\theta_x}\frac{v_{2x}}{v_2}\right) = 0 . \tag{7.8}$$

If (v_1,v_2) satisfy

$$v_{xx} = av , \tag{7.9}$$

$$v_t = bv_x + cv , \tag{7.10}$$

then (7.8) obtains

$$b_t - bb_x - 2\sigma b_{xx} = 2\sigma^2 a_x , \tag{7.11}$$

while, by cross-differentiation,

$$2c_x + b_{xx} = 0 , \qquad a_t = c_{xx} + 2ab_x + ba_x , \tag{7.12}$$

or

$$b_t - bb_x - 2\sigma b_{xx} = 2\sigma^2 a_x , \qquad a_t = -\tfrac{1}{2}b_{xxx} + 2ab_x + ba_x . \tag{7.13}$$

If b is functionally independent of a, it is, in general, not possible to introduce as spectral parameter into the above without obtaining equations that depend on this parameter. However, if we set

$$a = -(1/\sqrt{2})b_x + \tfrac{1}{2}b^2 \tag{7.14}$$

and

$$\sigma = 1/\sqrt{2} ,$$

then the system (7.13) becomes

$$b_t = 2bb_x + (1/\sqrt{2})b_{xx} . \tag{7.15}$$

The Lax pair for this equation is

$$v_{xx} = [-(1/\sqrt{2})b_x + \tfrac{1}{2}b^2]\, v , \qquad v_t = bv_x - \tfrac{1}{2}b_x v . \tag{7.16}$$

By the consistency condition of Eqs. (7.16),

$$[-(1/\sqrt{2})\partial_x + b\,]\,[b_t - 2bb_x - (1/\sqrt{2})b_{xx}\,] = 0 . \tag{7.17}$$

Letting

$$b = u + \lambda, \quad x \to x - 2\lambda t,$$

we find

$$u_t = 2uu_x + (1/\sqrt{2})u_{xx} , \qquad v_{xx} = [-(1/\sqrt{2})u_x + \tfrac{1}{2}u^2 + u\lambda + \tfrac{1}{2}\lambda^2]\, v , \qquad v_t = (u - \lambda)v_x - \tfrac{1}{2}u_x v , \tag{7.18}$$

where λ is the spectral parameter.

8. THE SCHWARZIAN DERIVATIVE

The Schwarzian derivative

$$\{\phi;x\} = \frac{\partial}{\partial x}\left(\frac{\phi_{xx}}{\phi_x}\right) - \frac{1}{2}\left(\frac{\phi_{xx}}{\phi_x}\right)^2 \tag{8.1}$$

is a third-order, differential expression which is invariant in form under the Moebius group acting on the dependent variable. That is,

$$\left\{\frac{a\phi + b}{c\phi + d};x\right\} \equiv \{\phi;x\} . \tag{8.2}$$

Furthermore,

$$\{\phi;x\} = \{\psi;x\} + \{f;\psi\}\,\psi_x^2 , \tag{8.3}$$

$$\{\phi;x\} = h'^2\{\phi;z\} + \{h;x\} , \tag{8.4}$$

where

(i) $\phi = f(\psi)$ is an arbitrary change of dependent variable,

(ii) $z = h(x)$ is an arbitrary change of independent variable.

The Schwarzian derivative appears in a variety of contexts, including conformal mapping of curvelinear polygons,[10,11] algebraic solutions in differential equations,[10,12] and the theory of automorphic functions.[11]

In the preceding sections we have considered the Painlevé property and Bäcklund transformations for a number of integrable partial differential equations. Specifically, it is found that the equation for the "singular surface" (in the presence of a Bäcklund transform) is most naturally expressed in terms of the Schwarzian derivative. These equations are invariant under the Moebius group. By subjecting them to the well-known[2] "linearizing" transformation

$$\phi = v_1/v_2 , \tag{8.5}$$

the Lax pairs for these equations are readily found.

In this section we investigate the Painlevé property associated with equations formulated in terms of the Schwarzian derivative. By this means we seek to identify classes of integrable partial differential equations.

To begin consider the equation associated with the KdV–MKdV equations, (Secs. 2 and 3).

$$\phi_t/\phi_x + \sigma\{\phi;x\} = \lambda , \tag{8.6}$$

where

$$\{\phi;x\} = \frac{\partial}{\partial x}\left(\frac{\phi_{xx}}{\phi_x}\right) - \frac{1}{2}\frac{\phi_{xx}^2}{\phi_x^2} .$$

This equation is homogeneous in ϕ, and consists of two terms; one, ϕ_t/ϕ_x, invariant under arbitrary changes of dependent variable, $\phi = f(\psi)$; the other, a function of the Schwarzian derivative, invariant under the Moebius group,

$$\phi = (a\psi + b)/(c\psi + d) . \tag{8.7}$$

In consequence, (8.6) is invariant under (8.7). It is to be noted that all the equations [(2.18), (3.8), (4.9), (5.9), (6.6), and (7.6)] that were found have this structure.

Following the procedure defined before, we find for (8.6) that

$$\phi = \psi^{-1}\sum_{j=0}^{\infty}\phi_j\psi^j \tag{8.8}$$

with resonances at

$$j = -1,0,1 . \tag{8.9}$$

The compatability conditions at $j = 0$ and $j = 1$ are satisfied and Eq. (8.6) has the Painlevé property about the "movable" singularity (8.8).

The Bäcklund transform for (8.6) is

$$\phi = \phi_0/\psi + \phi_1 . \tag{8.10}$$

Direct calculation reveals that the most general form allowed for (8.10) is the transformation (8.7), the Moebius group.

Furthermore, every equation for the singular surface considered in this paper has an expansion of the form (8.8) and possesses the Painlevé property about singularities of this type. We note that the vanishing of ϕ_x in (8.1) introduces the possibility of movable essential singularities in these equations [e.g., (8.6)] and "manifolds of indeterminancy"[3] in the corresponding Bäcklund transforms [i.e., (2.13)]. Therefore, equations of this type cannot be said to possess the Painlevé property identically.

Nevertheless, it is possible to identify a wide class of equations whose only singularities of finite degree are of the form (8.8) and which possess the Painlevé property about these singularities.

It is fairly easy to show that an equation of the form

$$A\left(\frac{\phi_t}{\phi_x}\right) + B(\{\phi;x\}) = 0 , \tag{8.11}$$

where $A(\phi_t/\phi_x)$ and $B(\{\phi;x\})$ are constant coefficient multinomials in $(\partial^k/\partial t^j\partial x^i)(\phi_t/\phi_x)$ and $(\partial^l/\partial t^m\partial x^n)\{\phi;x\}$, respectively, with $i + j = k$, $m + n = l$, $\max(k) < \max(l) + 2$, will have these properties.

Indeed, any equation of the form

$$A(\phi_{x_1},\phi_{x_2},\ldots,\phi_{x_n}) + B(\{\phi;x_1\},\{\phi;x_2\},\ldots,\{\phi;x_n\}) = 0 , \tag{8.12}$$

where (i) A is a constant coefficient multinomial in

$$\frac{\partial^N}{\partial_{x_1}^{j_1}\cdots\partial_{x_n}^{j_n}}\left(\frac{\phi_{x_l}}{\phi_{x_m}}\right), \quad j_1 + \cdots + j_n = N ,$$

and (ii) B is a constant coefficient multinomial in

$$\frac{\partial^m}{\partial x_1^{j_1}\cdots\partial_{x_n}^{j_n}}\{\phi,x_k\} ,$$

$$j_1 + \cdots + j_m = m, \quad \max(N) < \max(m) + 2 , \tag{8.13}$$

will have the Painlevé property about the movable poles (or order -1). (Note that we require the highest-order derivatives occur in expressions involving the Schwarzian derivative.)

We conclude this section with an example of an equation of the form (8.10) that is related to nonintegrable equations:

$$\frac{\phi_t}{\phi_x} + \frac{\partial}{\partial x}\{\phi;x\} = 0 . \tag{8.14}$$

About the "movable poles" this equation has a solution of

the form:

$$\phi = \psi^{-1} \sum_{j=0}^{\infty} \phi_j \psi^j , \tag{8.15}$$

with resonances at

$$j = -1,0,1,2 . \tag{8.16}$$

It can be readily verified that the compatability conditions are satisfied. (The solution has the Painlevé property about the movable poles.) On the other hand, if we attempt to "integrate" (8.14) by

$$\phi = v_1/v_2 , \quad v_{xx} = av , \tag{8.17}$$

and

$$v_t = bv_x + cv , \tag{8.18}$$

we find

$$v_{xx} = \tfrac{1}{2}wv; \quad v_t = w_x v_x - \tfrac{1}{2} w_{xx} v , \tag{8.19}$$

and

$$w_t + w_{xxxx} = 2ww_{xx} + w_x^2 . \tag{8.20}$$

Unfortunately, it does not appear to be possible to introduce a spectral parameter into (8.19) and (8.20) in such a manner that Eq. (8.19) will not depend on this parameter. This restricts the type of solution that can be represented by (8.17) to be of a special type.

About the singularities of Eq. (8.20) we find that

$$w \simeq w_0/\psi^2 + \cdots \tag{8.21}$$

with resonances at

$$r = -1, \ (7 \pm i\sqrt{11})/2, \ 8 . \tag{8.22}$$

There exists a "Painlevé type" of expansion,

$$w = \sum_{j=0}^{\infty} w_j \psi^{j-2} , \tag{8.23}$$

where the compatability condition at $j = 8$ is satisfied identically; (8.23) represents a "special" form of the solution since the arbitrary functions introduced at the resonances $r = (7 \pm \sqrt{11})/2$ are not included in (8.23).

Furthermore, by the substitution

$$V = \phi_{xx}/\phi_x , \tag{8.24}$$

we find from (8.14) that

$$V_t + \frac{\partial}{\partial x}(V_{xxx} - V_x^2 - V^2 V_x) = 0 . \tag{8.25}$$

For this equation we find, by a leading order analysis,

$$V \sim V_0 \phi^{-1} , \tag{8.26}$$

where

$$V_0 = 3\phi_x , \quad -2\phi_x .$$

Letting

$$V_0 = 3\phi_x , \tag{8.27}$$

we find the resonances

$$r = -1, \ (7 \pm i\sqrt{11})/2 , \ 4 . \tag{8.28}$$

Again, the expansion is Painlevé at $r = 4$ and the equation allows the "special" Painlevé form of solution

$$V = \sum_{j=0}^{\infty} V_j \phi^{j-1} . \tag{8.29}$$

We note that the complex resonances in (8.22) and (8.28) are identical.

We now let

$$V_0 = -2\phi_x \tag{8.30}$$

and find the resonances

$$r = -1,2,4,5 . \tag{8.31}$$

In this case, we find that the equation has the Painlevé expansion

$$V = \sum_{j=0}^{\infty} V_j \phi^{j-1} , \tag{8.32}$$

where (ϕ, V_2, V_4, V_5) are arbitrary.

By forming the Bäcklund transform

$$V = V_0/\phi + V_1 \tag{8.33}$$

from (8.32), we find that

$$V_0 = -2\phi_x , \qquad V_1 = \phi_{xx}/\phi_x , \tag{8.34}$$

(V, V_1) satisfy (8.25) and

$$\frac{\phi_t}{\phi_x} + \frac{\partial}{\partial x}\{\phi; x\} = 0 . \tag{8.35}$$

Equation (8.35) is just Eq. (8.14).

If one attempts to form a Bäcklund transform (8.33) for the expansion (8.29), there is found

$$V_0 = 3\phi_x , \quad V_1 = -\tfrac{3}{2}(\phi_{xx}/\phi_x) . \tag{8.36}$$

(V, V_1) satisfy (8.25) and

$$\phi_t = 0, \quad \{\phi; x\} = 0 . \tag{8.37}$$

Although Eq. (8.14) does not lead to "completely integrable" equations, it is somewhat curious that it does determine equations with somewhat "special" properties.

We feel that further study of equations of the form (8.10) is warranted.

APPENDIX A: RECUSION RELATIONS FOR THE KdV EQUATION AND CONSERVED DENSITIES

For the KdV equation

$$u_t + \frac{\partial}{\partial x}\left(\frac{u^2}{2} + u_{xx}\right) = 0 \tag{A1}$$

we get

$$u = \phi^{-2} \sum_{j=0}^{\infty} u_j \phi^j . \tag{A2}$$

and find

$$u_{j-3,t} + (j-4)\phi_t u_{j-2} + \psi_{j-1,x} + (j-4)\phi_x \psi_j = 0 , \tag{A3}$$

where

$$\psi_j = \sum_{m=0}^{j} \frac{u_{j-m} u_m}{2} + \theta_{j-1,x} + (j-3)\phi_x \theta_j \tag{A4}$$

and

$$\theta_j = u_{j-1,x} + (j-2)\phi_x u_j \,. \tag{A5}$$

In (A2), (ϕ, u_4, u_6) are arbitrary.

In this appendix we show how the recursion relations (A3) may be used to formally define an infinite set of conserved densities for the KdV equation. By repeated application of (A3) for $j, j+1, \cdots$ it is easily demonstrated that

$$\frac{\partial}{\partial t} A_j + \frac{\partial}{\partial x} B_j = 0 \,, \tag{A6}$$

where

$$A_j = u_{j-3} + (j-4)\phi u_{j-2} + (j-4)(j-3)(\phi^2/2!)u_{j-1} + (j-4)(j-3)(j-2)(\phi^3/3!)u_j + \cdots \,, \tag{A7}$$

or, for $j \geqslant 5$,

$$A_j = \sum_{K=0}^{\infty} \frac{(j+K-5)!}{(j-5)!} u_{j+K-3} \frac{\phi^K}{K!} \,, \qquad A_{j+5} = \sum_{K=0}^{\infty} u_{j+2+K} \phi^K \,, \tag{A8}$$

and

$$B_j = \psi_{j-1} + (j-4)\phi\psi_j + (j-4)(j-3)(\phi^2/2!)\psi_{j+1} + (j-4)(j-3)(j-2)(\phi^3/3!)\psi_{j+2} + \cdots \,, \tag{A9}$$

or, for $j \geqslant 5$,

$$B_j = \sum_{K=0}^{\infty} \frac{(j+K-5)!}{(j-5)!} \psi_{j+k-1} \frac{\phi^K}{K!} \,, \qquad B_{j+5} = \sum_{K=0}^{\infty} \psi_{j+4+K} \phi^K \,. \tag{A10}$$

For the conserved densities A_j, we find that

$$A_0 = 12(\phi^4\phi_x)_x, \quad A_1 = -12(\phi^3\phi_x)_x \,, \quad A_2 = 12(\phi^2\phi_x)_x, \quad A_3 = -12(\phi\phi_x)_x \,, \quad A_4 = 12(\phi_x)_x \,. \tag{A11}$$

Thus, the densities A_0–A_4 are trivial, i.e., gradients. Furthermore,

$$A_5 = \sum_{K=2}^{\infty} u_K \phi^{K-2} = \sum_{K=0}^{\infty} u_{K+2} \phi^K \tag{A12}$$

and

$$A_{5+j} = \frac{1}{j!} \frac{\delta^j}{\delta\phi^j} A_5 \,, \tag{A13}$$

where, for constant ϵ,

$$\frac{\delta}{\delta\phi} f(\phi) = \frac{d}{d\epsilon} f(\phi + \epsilon)|_{\epsilon=0} \,.$$

We note that

$$u = 12 \frac{\partial^2}{\partial x^2} \ln\phi + A_5 \,. \tag{A14}$$

APPENDIX B: THE HARRY DYM EQUATION

From the KdV–MKdV equations we have obtained the equation for the singular surface:

$$\phi_t/\phi_x + \{\phi;x\} = \lambda \,, \tag{B1}$$

where

$$\{\phi;x\} = \frac{\partial}{\partial x}\left(\frac{\phi_{xx}}{\phi_x}\right) - \frac{1}{2}\frac{\phi_{xx}^2}{\phi_x^2}$$

is the Schwarzian derivative.

By the transform properties of the Schwarzian derivative,

$$\{\phi;x\} = -\phi_x^2 \{x;\phi\} \,. \tag{B2}$$

Now, under the change of variables;

$$x \to \phi, \quad t \to t, \quad \phi \to x \,, \tag{B3}$$

we note that

$$\phi_x = 1/x_\phi, \quad x_t = -\phi_t/\phi_x \tag{B4}$$

Therefore, we find

$$x_\phi^2 x_t = \lambda x_\phi^2 + \{x;\phi\} \,, \tag{B5}$$

or, by expanding $\{x;\phi\}$,

$$x_t = \lambda - \frac{1}{2}\frac{\partial^2}{\partial\phi^2}\left(\frac{1}{x_\phi^2}\right) + \frac{3}{2}\left[\frac{\partial}{\partial\phi}\left(\frac{1}{x_\phi}\right)\right]^2 . \tag{B6}$$

Letting

$$v = 1/x_\phi \,, \tag{B7}$$

we find that

$$v_t = v^3 v_{\phi\phi\phi} \,. \tag{B8}$$

Finally, letting

$$v = cu^{-1/2}, \quad -2c^3 = 1 \,, \tag{B9}$$

we find

$$u_t = \frac{\partial^3}{\partial\phi^3} u^{-1/2} \,, \tag{B10}$$

where

$$u = c^2/v^2 = c^2 x_\phi^2 \,. \tag{B11}$$

This equation is called the Harry Dym equation and is known to be integrable.[14]

Equation (B8) raises some interesting questions concerning the nature of the Painlevé property. In general, about the singular manifold, $\psi = 0$,

$$V = \psi^{2/3} \sum_{j=0}^{\infty} V_j \psi^{j/3} \tag{B12}$$

with resonances at

$$j = -1, 2, 4 \,.$$

We note that $\ln\psi$ terms do not arise in the expansion for Eq. (B8).

(B12) demonstrates that Eq. (B8) does not posses the Painlevé property, as it has been defined herein, although the system (B8) is (presumably) integrable. The Painlevé property seems to be a sufficient, but not necessary condition, for integrability and would appear to require reformulation for systems with nontransformable branch point behavior.

Finally, we note that Eqs. (B5) and (B6) are invariant under the Moebius group acting on the independent variable ϕ. Hence, by restricing ϕ to a fundamental region in the complex ϕ plane it should be possible to define an $x = x(\phi)$ as an

automorphic function[11]

$$x(\phi) = X\left(\frac{a\phi + b}{c\phi + d}\right) \tag{B13}$$

of the Moebius group.

[1]J. Weiss, M. Tabor, and G. Carnevale, "The Painlevé property for partial differential equations," J. Math. Phys. **24**, 522 (1983).

[2]Einar Hille, *Analytic Function Theory* (Ginn, Boston, 1962), Vol. 2.

[3]H. Schwerdtfeger, *Geometry of Complex Numbers* (Dover, New York, 1979).

[4]Y. F. Chang, M. Tabor, and J. Weiss, "Analytic structure of the Henon–Heiles Hamiltonian in integrable and nonintegrable regimes," J. Math. Phys. **23**, 531 (1982).

[5]R. M. Miura, Ed., *Bäcklund Transformations, the Inverse Scattering Method, Solitons, and Their Applications*, Lecture Notes in Mathematics, **515** (Springer-Verlag, New York, 1976).

[6]D. Deift and E. Trubowitz, "Inverse Scattering on the Line," Comm. Pure. Appl. Math **32**, 121 (1979).

[7]Hedley Morris, "Prolongation structures and a generalized inverse scattering problem," J. Math. Phys. **17**, 1867 (1976).

[8]Hedley Morris, Soliton solutions and the higher-order Korteweg–de Vries equations," J. Math. Phys. **18**, 530 (1977).

[9]V. S. Dryuma, "Analytic solutions of the two-dimensional Korteweg–de Vries equation," Pis'ma Zh. Eksp. Teor. Fiz. **19**, 753 (1974) [JETP Lett. **19**, 387 (1974)].

[10]Einar Hille, *Ordinary Differential Equation in the Complex Domain* (Wiley, New York, 1976).

[11]A. R. Forsyth, *Theory of Functions of a Complex Variable* (Dover, New York, 1965), Vol. 2.

[12]A. R. Forsyth, *Theory of Differential Equations* (Dover, New York, 1959), Vol. 4.

[13]W. F. Osgood, *Topics in the Theory of Functions of Several Complex Variables* (Dover, New York, 1966).

[14]K. M. Case, "Dual Hamiltonian Formulations and Completely Integrable Systems," in AIP Conference Proceedings, *Mathematical Methods in Hydrodynamics and Related Dynamical Systems Dec. 1981, La Jolla*, edited by M. Tabor and Y. M. Treve (Am. Inst. of Physics, New York, 1982).

The Hamiltonian structure of the nonabelian Toda hierarchy

M. Bruschi and O. Ragnisco
Istituto di Fisica dell'Università, Rome, Italy
Istituto Nazionale di Fisica Nucleare, Sezione di Roma, Rome, Italy

(Received 16 June 1982; accepted for publication 23 July 1982)

We show that a subset of the whole class of nonlinear differential-difference equations, associated with the discrete analog of the matrix Schrödinger operator, is endowed with a Hamiltonian structure and exhibits an infinite number of integrals of motion in involution. We also establish the relation between these integrals of motion and the transmission coefficient of the underlying linear problem, and show that such a relation implies that, for the whole class previously introduced, there exists an infinite number of conserved quantities.

PACS numbers: 02.30.Ks, 02.30.Tb

1. INTRODUCTION

Recently[1] the direct and inverse spectral problem for the discrete analog of the matrix Schrödinger operator has been solved; moreover, through two different approaches[2,3] the class of nonlinear differential-difference equations (NDDE's) corresponding to isospectral deformations of this operator and thus solvable by the inverse spectral transform has been derived; the simplest equation in this class is a nonabelian version of the celebrated Toda lattice.[4]

The purpose of the present paper is twofold. First we show that a subset of the class of NDDE's above mentioned is endowed with a Hamiltonian structure; to establish this result we follow the framework, essentially geometrical in nature, introduced in this subject by Magri.[5] Then, we show the connection between this geometrical approach and the usual analytic one, relying on the properties of the spectral problem.

According to this scheme, in Sec. 2 we recall the spectral problem under scrutiny and the associated class of NDDE's, which includes as a subset the nonabelian Toda hierarchy; in Sec. 3 and Sec. 4 we prove that this hierarchy exhibits a Hamiltonian structure; in Sec. 5 we relate the results obtained in the two previous sections with the properties of the spectral data and moreover we show that for the whole class of NDDE's introduced in Sec. 2 there exists an infinite number of conserved quantities. Most of the proofs are confined in Appendices A,B,C to simplify the exposition.

2. THE NONABELIAN TODA HIERARCHY

In Ref. 1, one has investigated the following spectral problem:

$$L\psi(n) = \lambda\psi(n), \tag{2.1a}$$

the linear operator L being defined as

$$L\psi(n) = V^{(1)}(n)\psi(n+1) + V^{(2)}(n)\psi(n) + \psi(n-1), \tag{2.1b}$$

where $V^{(1)}$ and $V^{(2)}$ are $N \times N$ matrices, depending on the integer variable n (and possibly on a continuous variable t, say "time"), obeying the following boundary conditions:

$$\lim_{n \to \pm\infty} V^{(1)}(n) - I = \lim_{n \to \pm\infty} V^{(2)}(n) = 0, \tag{2.2}$$

(by I we have denoted the $N \times N$ identity matrix), and $\psi = \psi(n,t;\lambda)$ is a fundamental matrix solution of the spectral problem (2.1). As remarked in Ref. 1, under conditions (2.2), the "eigenvalue" λ can be parametrized as

$$\lambda = z + z^{-1}, \tag{2.3}$$

where, for the continuous spectrum z spans the unit circle in the complex z-plane, while the discrete part of the spectrum consists of a finite number of points $z_j (j = 1,\ldots,\mathcal{N})$ lying inside the unit circle.

Under the assumption that the fields $V^{(1)}$ and $V^{(2)}$ approach their limits exponentially or faster, it was shown that one can associate to the linear operator L a set of spectral data which in turn, once given, allows one to reconstruct uniquely the fields $V^{(1)}$ and $V^{(2)}$.

In Refs. 2 and 3 one has derived the class of NDDE's corresponding to isospectral deformations of L, and the associated time-evolution of the spectral data. This class, solvable by the inverse spectral transform reads

$$\begin{pmatrix} V_t^{(1)}(n) \\ V_t^{(2)}(n) \end{pmatrix} = f_\nu(\mathscr{L}^\nu)\begin{pmatrix} [V^{(1)}(n),\sigma_\nu] \\ [V^{(2)}(n),\sigma_\nu] \end{pmatrix} + g_\nu(\mathscr{L})\begin{pmatrix} V^{(2)}(n)\Pi(n)\sigma_\nu\Pi^{-1}(n+1) - \Pi(n)\sigma_\nu\Pi^{-1}(n+1)V^{(2)}(n+1) \\ \Pi(n-1)\sigma_\nu\Pi^{-1}(n) - \Pi(n)\sigma_\nu\Pi^{-1}(n+1) \end{pmatrix}$$
$$(\Pi(n) = V^{(1)}(n)V^{(1)}(n+1)\ldots V^{(1)}(n+k)\ldots;\Pi^{-1}(n) = [\Pi(n)]^{-1}), \tag{2.4}$$

where $\{\sigma_\nu\}(\nu = 0,1,\ldots, N^2-1;\ \sigma_0 = I)$ is a basis in the linear space of $N \times N$ matrices, f_ν and g_ν are entire functions of their argument, and the recursion operator $\mathscr{L}$ is defined as[6]

 0022-2488/83/061414-08$02.50 © 1983 American Institute of Physics

$$\mathscr{L}_V\begin{pmatrix}\Phi^{(1)}(n)\\ \Phi^{(2)}(n)\end{pmatrix} = \begin{pmatrix}\Phi^{(1)}(n)V^{(2)}(n+1) + V^{(1)}(n)(\Phi^{(2)}(n) + \Phi^{(2)}(n+1)) + V^{(2)}(n)S(n) - S(n)V^{(2)}(n+1) + \sum_{j=n}^{\infty}[\Phi^{(2)}(j), V^{(1)}(n)]\\ V^{(2)}(n)\Phi^{(2)}(n) + \Phi^{(1)}(n) - S(n) + S(n-1) + \sum_{j=n}^{\infty}[\Phi^{(2)}(j), V^{(2)}(n)]\end{pmatrix} \tag{2.5}$$

$$\left(S(n) = \Pi(n)\left[\sum_{j=n}^{\infty}\Pi^{-1}(j)\Phi^{(1)}(j)\Pi(j+1)\right]\Pi^{-1}(n+1)\right).$$

In formula (2.5), $\Phi^{(1)}$ and $\Phi^{(2)}$ are matrices of rank N satisfying the asymptotic conditions

$$\lim_{n\to\pm\infty}\Phi^{(i)}(n) = 0 \quad (i = 1,2). \tag{2.6}$$

In the special case $f_\nu = 0, g_\nu = \delta_{\nu 0}\, g_0$, we get the subclass

$$\begin{pmatrix}V_t^{(1)}(n)\\ V_t^{(2)}(n)\end{pmatrix} = g_0(\mathscr{L}_V)\begin{pmatrix}V^{(2)}(n)V^{(1)}(n) - V^{(1)}(n)V^{(2)}(n+1)\\ V^{(1)}(n-1) - V^{(1)}(n)\end{pmatrix}, \tag{2.7}$$

whose simplest member, which obtains by setting $g_0(\mathscr{L}_V) = \mathbf{1}$, i.e., the identity operator, is just the nonabelian Toda equation; thus we will refer to the subclass (2.7) as to the "nonabelian Toda hierarchy."

3. BASIC CONCEPTS AND DEFINITIONS

Here and in the following section, we shall restrict considerations to the nonabelian Toda hierarchy; moreover, with no loss of generality, we will assume that the function g_0 introduced in Eq. (2.7) is just a power. According to Ref. 5, we will regard $V \equiv \begin{pmatrix}V^{(1)}\\ V^{(2)}\end{pmatrix}$ as a point in an infinite-dimensional manifold, hereafter denoted as "configuration space." To each point V in the configuration space, it is naturally associated the tangent space T_V, whose elements, which we shall denote by upper-case Greek characters, obey homogeneous boundary conditions and are related to the variations of V in the standard way:

$$\delta V = \epsilon\Phi. \tag{3.1}$$

A glance to Eqs. (2.6),(2.7) allows us to recognize that the operator $\mathscr{L}_V$, which yields a new evolution equations from a given one, maps the tangent space T_V into itself. Moreover, with any point V of the configuration space, one can associate the so-called cotangent space T_V^*, whose elements will be denoted by lower case Greek characters, by introducing the (nondegenerate) bilinear form:

$$\langle\alpha,\Phi\rangle = \mathrm{Tr}\sum_{n=-\infty}^{+\infty}(\alpha^{(1)}(n)\Phi^{(1)}(n) + \alpha^{(2)}(n)\Phi^{(2)}(n)), \tag{3.2}$$

where

$$\Phi \equiv \begin{pmatrix}\Phi^{(1)}\\ \Phi^{(2)}\end{pmatrix}\in T_V, \quad \alpha \equiv \begin{pmatrix}\alpha^{(1)}\\ \alpha^{(2)}\end{pmatrix}\in T_V^*. \tag{3.3}$$

In the following, a given evolution equation of the nonabelian Toda hierarchy will be written in the abstract form

$$V_t = H_j(V), \tag{3.4}$$

where the evolution operator $H_j(V)$, associating a tangent vector with any point of its domain, is given by

$$H_j(V) = \mathscr{L}_V^j\begin{pmatrix}V^{(2)}(n)V^{(1)}(n) - V^{(1)}(n)V^{(2)}(n+1)\\ V^{(1)}(n-1) - V^{(1)}(n)\end{pmatrix}. \tag{3.5}$$

By defining the operator $\mathscr{S}_V$ as[7]

$$\mathscr{S}_V\begin{pmatrix}\alpha^{(1)}(n)\\ \alpha^{(2)}(n)\end{pmatrix} = \begin{pmatrix}\alpha^{(2)}(n)V^{(1)}(n) - V^{(1)}(n)\alpha^{(2)}(n+1)\\ \alpha^{(1)}(n-1)V^{(1)}(n-1) - V^{(1)}(n)\alpha^{(1)}(n) + [\alpha^{(2)}(n), V^{(2)}(n)]\end{pmatrix}, \tag{3.6}$$

it is straightforward to recognize that

$$H_j(V) = \mathscr{L}_V^j\mathscr{S}_V Q_0(V), \tag{3.7a}$$

where

$$Q_0(V) = \begin{pmatrix}I\\ V^{(2)}(n)\end{pmatrix}. \tag{3.7b}$$

Let us now introduce the adjoint $\tilde{\mathscr{L}}_V$ of the operator $\mathscr{L}_V$ in the standard way

$$\langle\alpha,\mathscr{L}_V\Phi\rangle = \langle\tilde{\mathscr{L}}_V\alpha,\Phi\rangle. \tag{3.8a}$$

Through Eq. (3.8a) we get for $\tilde{\mathscr{L}}_V$ the explicit expression

$$\tilde{\mathscr{L}}_V\begin{pmatrix}\alpha^{(1)}(n)\\ \alpha^{(2)}(n)\end{pmatrix} = \begin{pmatrix}\alpha^{(2)}(n) + \alpha^{(2)}(n+1) + V^{(2)}(n+1)\alpha^{(1)}(n) + \Pi(n+1)\left[\sum_{j=n+1}^{\infty}\Pi^{-1}(j+1)\alpha^{(2)}(j)\Pi(j) - \Pi^{-1}(j)\alpha^{(2)}(j)\Pi(j-1)\right.\\ \left. + \Pi^{-1}(j+1)(V^{(2)}(j+1)\alpha^{(1)}(j) - \alpha^{(1)}(j)V^{(2)}(j))\Pi(j)\right]\Pi^{-1}(n)\\ \alpha^{(1)}(n-1)V^{(1)}(n-1) + V^{(1)}(n)\alpha^{(1)}(n) + \sum_{j=n}^{\infty}([\alpha^{(1)}(j), V^{(1)}(j)] + [\alpha^{(2)}(j), V^{(2)}(j)]) + V^{(2)}(n)\alpha^{(2)}(n)\end{pmatrix}. \tag{3.8b}$$

By direct computation, using (2.5),(3.6),(3.8b) one can easily establish the following "coupling" relation

$$\mathscr{L}_V \mathscr{S}_V \alpha = \mathscr{S}_V \tilde{\mathscr{L}}_V \alpha. \tag{3.9}$$

Thus, defining a class of operators $Q_j(V)$ by

$$Q_j(V) = \tilde{\mathscr{L}}^j{}_V Q_0(V), \tag{3.10}$$

the coupling condition (3.9) allows one to recast formula (3.7) in the form

$$H_j(V) = \mathscr{S}_V Q_j(V). \tag{3.11}$$

4. THE HAMILTONIAN STRUCTURE

To define the Hamiltonian structure associated with the nonabelian Toda hierarchy (2.7), we take advantage of the following theorems, concerning the properties of the operators introduced in the previous sections. The proofs of such theorems follow always by direct, although tedious, calculations.

Theorem I: The operator $Q_0(V)$, defined by Eq. (3.7b), is a potential operator, i.e., it satisfies the condition

$$\langle Q_0'(V)\cdot\Phi,\Psi\rangle = \langle Q_0'(V)\cdot\Psi,\Phi\rangle, \tag{4.1}$$

where the symbol $Q_0'(V)\cdot\Phi$ denotes the Gateaux derivative of the operator $Q_0(V)$ (see Appendix A).

Theorem II: The operator $\mathscr{L}_V$ [Eq. (2.5)] is a Nijenhuis operator, i.e., it satisfies the following condition:

$$\mathscr{L}_V'(\Phi;\mathscr{L}_V\Psi) - \mathscr{L}_V'(\psi;\mathscr{L}_V\Phi) = \mathscr{L}_V[\mathscr{L}_V'(\Phi;\psi) - \mathscr{L}_V'(\Psi;\Phi)], \tag{4.2}$$

where the symbol $\mathscr{L}_V'(\Phi;\psi)$ denotes the Gateaux derivative of the operator $\mathscr{L}_V$ (see Appendix A).

Theorem III: The operator $\mathscr{S}_V$ [Eq. (3.6)] is a symplectic operator,[8] i.e., it maps the cotangent space T_V^* into the tangent space T_V and satisfies the following two conditions:

(a) it is a skew-symmetric operator, namely

$$\langle\alpha,\mathscr{S}_V\beta\rangle = -\langle\beta,\mathscr{S}_V\alpha\rangle; \tag{4.3}$$

(b) it fulfills the Jacobi identity, namely,

$$\langle\alpha,\mathscr{S}_V'(\beta;\mathscr{S}_V\gamma)\rangle + \text{cyclic permutations} = 0. \tag{4.4}$$

Theorem IV: The symplectic operator $\mathscr{S}_V$ and the Nijenhuis operator $\mathscr{L}_V$ fulfill the following two "coupling" conditions:

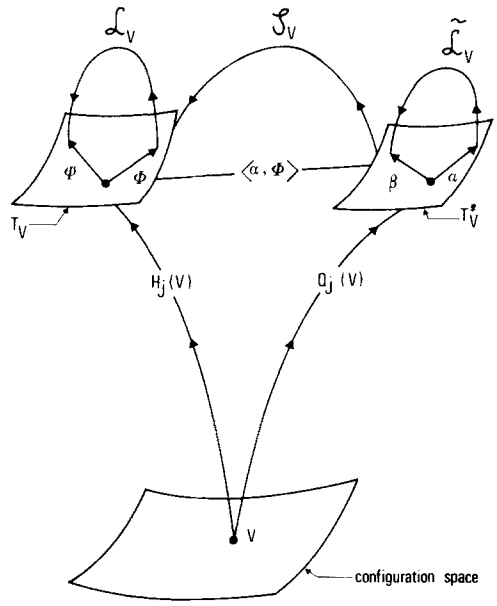


FIG. 1. A geometrical scheme of the symplectic Kahler manifold.

(a) $\mathscr{L}_V\mathscr{S}_V = \mathscr{S}_V\tilde{\mathscr{L}}_V$ [see Eq. (3.9)], (4.5)

(b) $\langle\alpha,\mathscr{L}_V'(\mathscr{S}_V\beta;\Phi) - \mathscr{L}_V'(\Phi;\mathscr{S}_V\beta)\rangle = \langle\beta,\mathscr{S}_V'(\alpha;\mathscr{L}_V\Phi) - \mathscr{L}_V'(\Phi;\mathscr{S}_V\alpha) - \mathscr{L}_V\mathscr{S}_V'(\alpha;\Phi)\rangle.$ (4.6)

Thus they define a symplectic Kahler manifold (see Ref. 5).

From the previous theorems one can derive the following corollary, whose proof is given in Appendix B.

Corollary I: All the operators $Q_j(V)$, defined by Eq. (3.10), are potential operators; thus they are the gradients of certain functionals $I_j[V]$

$$Q_j(V) = \nabla I_j[V], \quad j = 0,1,\cdots, \tag{4.7}$$

where

$$I_j(V) = \int_0^1 d\lambda\,\langle Q_j(\lambda V),V\rangle, \quad j = 0,1,\cdots. \tag{4.8}$$

Remark I: Being constructed through the action of a symplectic operator on potential operators (see 3.11) all the

TABLE I. The scheme of iteration for the nonabelian Toda hierarchy.

$$Q_0(V) = \begin{pmatrix} I \\ V^{(2)}_{(n)} \end{pmatrix} \xrightarrow{\mathscr{S}_V} H^{(1)}(V) = \begin{pmatrix} V^{(2)}_{(n)}V^{(1)}_{(n)} - V^{(1)}_{(n)}V^{(2)}_{(n+1)} \\ V^{(1)}_{(n-1)} - V^{(1)}_{(n)} \end{pmatrix}$$

$$\downarrow \tilde{\mathscr{L}}_V \qquad\qquad \downarrow \mathscr{L}_V$$

$$Q_1(V) = \begin{pmatrix} V^{(2)}_{(n)} + V^{(2)}_{(n+1)} \\ V^{(1)}_{(n-1)} + V^{(1)}_{(n)} + [V^{(2)}_{(n)}]^2 \end{pmatrix} \xrightarrow{\mathscr{S}_V} H_2(V) = \begin{pmatrix} V^{(1)}_{(n-1)}V^{(1)}_{(n)} - V^{(1)}_{(n)}V^{(1)}_{(n+1)} + [V^{(2)}_{(n)}]^2V^{(1)}_{(n)} - V^{(1)}_{(n)}[V^{(2)}_{(n+1)}]^2 \\ V^{(2)}_{(n-1)}V^{(1)}_{(n-1)} + V^{(1)}_{(n-1)}V^{(2)}_{(n)} - V^{(2)}_{(n)}V^{(1)}_{(n)} - V^{(1)}_{(n)}V^{(2)}_{(n+1)} \end{pmatrix}$$

$$\downarrow \tilde{\mathscr{L}}_V$$

$$Q_2(V) = \begin{pmatrix} V^{(1)}_{(n-1)} + V^{(1)}_{(n)} + V^{(1)}_{(n+1)} + [V^{(2)}_{(n)}]^2 + [V^{(2)}_{(n+1)}]^2 + V^{(2)}_{(n+1)}V^{(2)}_{(n)} \\ V^{(2)}_{(n-1)}V^{(1)}_{(n-1)} + (V^{(1)}_{(n-1)} + V^{(1)}_{(n)})V^{(2)}_{(n)} + V^{(2)}_{(n)}(V^{(1)}_{(n-1)} + V^{(1)}_{(n)}) + V^{(1)}_{(n)}V^{(2)}_{(n+1)} + [V^{(2)}_{(n)}]^3 \end{pmatrix} \xrightarrow{\mathscr{S}_V} H_3(V) =$$

operators $H_j(V)$ are, by definition, Hamiltonian operators. A geometrical scheme of the Kahler symplectic manifold is given in Fig. 1, while in Table I the scheme of iteration for the nonabelian Toda hierarchy is displayed.

Conjecture: From Table I, one can see that the first potentials and Hamiltonian operators of this hierarchy are purely local, in spite of the nonlocal structure of $\mathscr{L}_V$ and $\tilde{\mathscr{L}}_V$. We conjecture that this property holds true for the whole hierarchy.

It is now straightforward to prove that the NDDE's belonging to the class (2.7) are commuting flows leaving the symplectic Kahler manifold invariant. Indeed, the invariance of the Nijenhuis operator $\mathscr{L}_V$ along the flows $H_l(V)$, which is expressed by the following formula,

$$\mathscr{L}'_V(\Phi;H_l(V)) + \mathscr{L}_V H'_l\cdot\Phi - H'_l\cdot\mathscr{L}_V\Phi = 0, \tag{4.9}$$

is merely a consequence of the coupling condition (4.6) (see for the proof Appendix B).

Let us now consider the commutator between two flows $H_j(V),H_l(V)$, defined according to Ref. 5

$$\begin{aligned}[H_j,H_l] &\equiv H'_j\cdot H_l - H'_l\cdot H_j \\ &= \mathscr{L}_V H'_{j-1}\cdot H_l + \mathscr{L}'_V(H_{j-1};H_l) \\ &\quad - H'_l\cdot\mathscr{L}_V H_{j-1}.\end{aligned} \tag{4.10}$$

Identifying the tangent vector Φ with $H_{j-1}(V)$, formulas (4.9) and (4.10) imply

$$[H_j,H_l] = \mathscr{L}_V[H_{j-1},H_l], \tag{4.11}$$

and thus it follows by iteration that

$$[H_j,H_l] = \mathscr{L}_V^{j-l}[H_l,H_l] = 0. \tag{4.12}$$

Due to Theorems I–IV and to Corollary I, it is also possible to prove that the functionals $I_j[V]$, defined by Eq. (4.8), form an infinite set of integrals of motion in involution for any equation of the nonabelian Toda hierarchy. We are in fact enabled to introduce the following Poisson bracket:

$$\{I_j[V],I_l[V]\} = \langle Q_j(V),\mathscr{S}_V Q_l(V)\rangle, \tag{4.13}$$

which, as we show in Appendix B, is well-defined, since it is skew-symmetric and fulfills the Jacobi identity. Then, according to the classical scheme first introduced by Magri[9] and assuming with no restriction $j>l$, we have

$$\begin{aligned}\{I_j,I_l\} &= \langle Q_j,\mathscr{S}_V Q_l\rangle \\ &= -\langle\mathscr{S}_V Q_j,Q_l\rangle = -\langle\mathscr{S}_V\tilde{\mathscr{L}}_V Q_{j-1},Q_l\rangle \\ &= -\langle\mathscr{L}_V\mathscr{S}_V Q_{j-1},Q_l\rangle = \langle Q_{j-1},\mathscr{S}_V Q_{l+1}\rangle \\ &= \{I_{j-1},I_{l+1}\}.\end{aligned} \tag{4.14}$$

Now, if $j=l+2k$, by iteration we obtain

$$\begin{aligned}\{I_j,I_l\} &= \{I_{j-k},I_{l+k}\} \\ &= \{I_m,I_m\} = 0 \quad (m=j-k=l+k).\end{aligned} \tag{4.15a}$$

On the other hand, if $j=l+2k+1$, we get

$$\begin{aligned}\{I_j,I_l\} &= \{I_{m+1},I_m\} = \langle\tilde{\mathscr{L}}_V Q_m,\mathscr{S}_V Q_m\rangle \\ &= \langle Q_m,\mathscr{L}_V\mathscr{S}_V Q_m\rangle = \{I_m,I_{m+1}\} \\ &= -\{I_{m+1},I_m\}.\end{aligned} \tag{4.15b}$$

This clearly proves that all the I_j's are integral of motion in involution for any equation of the nonabelian Toda hierarchy; indeed, if the fields evolve according to (3.4), it follows that

$$I_{j,t} = \langle\nabla I_j,V_t\rangle = \langle Q_j,\mathscr{S}_V Q_l\rangle = \{I_j,I_l\} = 0. \tag{4.15c}$$

Remark II: It is worthwhile to notice that there exist two "extra-potentials," say Q_{-2} and Q_{-1}, defined as

$$Q_{-2} = \begin{pmatrix}[V^{(1)}(n)]^{-1} \\ 0\end{pmatrix}, \quad Q_{-1} = \begin{pmatrix}0 \\ I\end{pmatrix}, \tag{4.16}$$

related to each other and to Q_0 through the adjoint $\tilde{\mathscr{L}}_V$ of the Nijenhuis operator by the formulas

$$Q_{-1} = \tfrac{1}{2}\tilde{\mathscr{L}}_V Q_{-2}; \quad Q_0 = \tilde{\mathscr{L}}_V Q_{-1} - Q_{-2}. \tag{4.17}$$

These two potentials belong to the Ker of $\mathscr{S}_V$ and thus of course give rise to null Hamiltonian operators [see Eq. (3.11)]; however, they are easily recognized, by direct calculation, to be the gradients of two nontrivial functionals, namely,

$$Q_{-2}(V) = \nabla I_{-2}[V], \quad I_{-2}[V] = \ln\left[\det\prod_{j=-\infty}^{+\infty} V^{(1)}(j)\right], \tag{4.18a}$$

$$Q_{-1}(V) = \nabla I_{-1}[V], \quad I_{-1}[V] = \mathrm{Tr}\sum_{j=-\infty}^{+\infty} V^{(2)}(j), \tag{4.18b}$$

which are in involution between each other and with all the functionals $I_j[V]$ previously introduced.

5. SPECTRAL DATA AND CONSERVED QUANTITIES FOR THE WHOLE CLASS OF NDDE'S (2.4)

In this section we will establish the relation between the conserved quantities introduced in Sec. 4 and the spectral data associated to the linear problem (2.1).[10] We will also show that the functionals I_j's are actually integrals of motion for the whole class of NDDE's (2.4), although this larger class does not seem to be endowed with an Hamiltonian structure.

For our purposes it is convenient to introduce, together with the linear problem (2.1), its adjoint problem, reading[2]:

$$\bar{\psi}(n-1)V^{(1)}(n-1) + \bar{\psi}(n)V^{(2)}(n) + \bar{\psi}(n+1) = \lambda\bar{\psi}(n). \tag{5.1}$$

We consider the Jost solutions of problems (2.1), (5.1), $F^{(\pm)}$, $\bar{F}^{(\pm)}$, characterized by the asymptotic behaviors

$$\begin{aligned}&F^{(+)}(n,z)\sim z^n I, \quad (n\to+\infty), \\ &F^{(+)}(n,z)\sim z^n\hat{T}^{-1}(z) + z^{-n}\hat{R}(z)\hat{T}^{-1}(z), \quad (n\to-\infty), \\ &F^{-}(n,z)\sim z^{-n}T^{-1}(z) + z^n R(z)T^{-1}(z), \quad (n\to+\infty), \\ &F^{-}(n,z)\sim z^{-n}I, \quad (n\to-\infty), \\ &\bar{F}^{(+)}(n,z)\sim z^n I, \quad (n\to+\infty), \\ &\bar{F}^{(+)}(n,z)\sim z^n T^{-1}(z) + z^{-n}\hat{R}T^{-1}(z), \quad (n\to-\infty), \\ &\bar{F}^{(-)}(n,z)\sim z^{-n}\hat{T}^{-1}(z) + z^n R(z)\hat{T}^{-1}(z), \quad (n\to+\infty), \\ &\bar{F}^{(-)}(n,z)\sim z^{-n}I, \quad (n\to-\infty).\end{aligned} \tag{5.2}$$

Let us now define the "squared-eigenfunction" G as

$$G = \begin{pmatrix}G^{(1)}(n,z) \\ G^{(2)}(n,z)\end{pmatrix} = \begin{pmatrix}F^{(+)}(n;z)\dfrac{\hat{T}(z)}{z^{-1}-z}\bar{F}^{(-)}(n+1,z) \\ F^{(+)}(n,z)\dfrac{\hat{T}(z)}{z^{-1}-z}\bar{F}^{(-)}(n,z)\end{pmatrix}. \tag{5.3}$$

This function enjoys some remarkable properties, summarized in the following lemmas, whose proofs we give in Appendix C.

Lemma I: The "squared-eigenfunction" G satisfies the following "eigenvalue" equation

$$\tilde{\mathscr{L}}_V G = \lambda G - Q_{-1} + zQ_{-2}, \tag{5.4}$$

where the operator $\tilde{\mathscr{L}}_V$ is again the adjoint of the Nijenhuis operator $\mathscr{L}_V$ and the potentials Q_{-2}, Q_{-1} are given in Formula (4.16).

Lemma II: The "squared-eigenfunction" G admits the following McLaurin expansion in powers of the spectral parameter z

$$G = \sum_{l=1}^{\infty} G_l z^l, \tag{5.5a}$$

where, in particular,

$$G_1 = Q_{-1}, \quad G_2 = Q_0. \tag{5.5b}$$

This expansion holds true in the open circle of the complex z-plane:

$$C = \left\{ z : |z| < \min_{j=1,\dots,N} |z_j| \right\}, \tag{5.6}$$

where the z_j's are the zeroes of $\det \hat{T}^{-1}(z)$, i.e., the bound state eigenvalues of the linear problem (2.1) (see Ref. 2).

Corollary II.1: From Lemma I it follows that the coefficients G_l of the McLaurin expansion (5.5) obey the following recursion relation:

$$G_{l+1} = \tilde{\mathscr{L}}_V G_l - G_{l-1} \quad (l = 1,2,\cdots), \tag{5.7a}$$

having introduced for convenience

$$G_0 = Q_{-2}, \tag{5.7b}$$

so that, due to (4.17) and (5.5b),

$$G_1 = \tfrac{1}{2} \tilde{\mathscr{L}}_V G_0. \tag{5.7c}$$

Lemma III: In the circle C, defined through (5.6), the following formula holds

$$G = -\nabla \ln(\det \hat{T}). \tag{5.8}$$

Lemma IV: In the circle C, defined through (5.6), the following "trace formula" holds:

$$\frac{\partial}{\partial z} \ln(\det T) = (z^{-2} - 1)\mathrm{Tr} \sum_{n=-\infty}^{+\infty} \left[G^{(2)}(n,z) - \frac{I}{z^{-1} - z} \right]. \tag{5.9}$$

It is now straightforward to establish the relation between the integrals of motion I_j defined in the previous Section and the spectral data, namely the coefficients of the McLaurin expansion of $\ln(\det \hat{T})$.

Indeed, inserting into the rhs of formula (5.9) the power expansion (5.5), integrating from 0 to $z \in C$, and taking into account that $\hat{T}(0) = I$ [see Appendix C, Eq. (C7)], we get

$$\ln(\det \hat{T}) = \sum_{l=1}^{\infty} J_l z^l, \tag{5.10a}$$

where

$$J_l = l^{-1} \mathrm{Tr} \sum_{n=-\infty}^{+\infty} [G^{(2)}_{l-1}(n,z) - G^{(2)}_{l+1}(n,z)]. \tag{5.10b}$$

Moreover, by taking the gradient of (5.10a) and using Lemma III one immediately gets

$$G_l = \nabla J_l. \tag{5.11}$$

On the other hand, formulas (5.7) clearly imply that the G_l's are linear combinations of the potential operators Q_j's ($j = -2, -1, 0, \dots, l-2$) introduced in Sec. 4, and thus, recalling (4.7), it follows from formula (5.11) that the same linear relation connects the quantities J_l and the integrals of motion I_{-2}, I_{-1}, I_j ($j = 0, \dots, l-2$) defined in (4.18), (4.8) In the following we will show that the quantities J_l (or, equivalently, I_j) are integrals of motion not only for the nonabelian Toda hierarchy, but also for the whole class of NDDE's (2.4).

Indeed, let us now assume that V evolves according to any of the equation of class (2.4), which, with no loss of generality, we can write as

$$V_t = \mathscr{L}^k_V \bar{V}, \tag{5.12a}$$

where

$$\bar{V}(n) = \begin{pmatrix} [V^{(1)}(n), \sigma] \\ [V^{(2)}(n), \sigma] \end{pmatrix} + \begin{pmatrix} V^{(2)}(n) \Pi(n) \tau \Pi^{-1}(n+1) - \Pi(n) \tau \Pi^{-1}(n+1) V^{(2)}(n+1) \\ \Pi(n-1) \tau \Pi^{-1}(n) - \Pi(n) \tau \Pi^{-1}(n+1) \end{pmatrix}, \tag{5.12b}$$

σ and τ being two arbitrary n-independent (but possibly t-dependent) matrices.

Then, Lemma III, Eq. (5.12a) and formula (3.8a) clearly imply

$$\partial_t \ln(\det \hat{T}) = \langle \tilde{\mathscr{L}}^k_V G, \bar{V} \rangle. \tag{5.13}$$

By inserting into (5.13) the power expansion of G and $\ln(\det \hat{T})$ given respectively by formulas (5.5) and (5.10), one obtains

$$\sum_{l=1}^{\infty} z^l \left[\sum_{p=0}^{l+k} C_p^{(l,k)} \langle G_l, \bar{V} \rangle - J_{l,t} \right] = 0, \tag{5.14}$$

where the $C_p^{(l,k)}$'s are numerical coefficients whose explicit expression, though irrelevant for our purposes, can nevertheless be derived through the recursion relation (5.7).

But, since

$$\langle G_l, \bar{V} \rangle = 0, \tag{5.15}$$

as it will be proved in Appendix C, from formula (5.14) we can conclude that the quantities J_l (or, equivalently, I_j) form an infinite set of integrals of motion for the whole class of NDDE's (2.4) associated with the discrete analog of the matrix Schroedinger spectral problem.

APPENDIX A

In this appendix we recall the definition of the Gateaux derivative of an operator.

For operators like $Q_j(V)$ or $H_j(V)$, whose domain belongs to the configuration space, the Gateaux derivative is defined as follows[5]:

$$F'(V)\cdot\Phi = \partial_\epsilon F(V+\epsilon\Phi)\big|_{\epsilon=0}. \tag{A1}$$

For (tensor) operators like $\mathscr{L}_V, \tilde{\mathscr{L}}_V, \mathscr{S}_V$, whose domain belong either to the tangent or to the cotangent space, the Gateaux derivative is given by:

$$\mathscr{M}'_V(\Phi;\Psi) = \partial_\epsilon \mathscr{M}_{V+\epsilon\psi}\Phi\big|_{\epsilon=0}. \tag{A2}$$

APPENDIX B

Proof of formula (3.8b): The expression of $\tilde{\mathscr{L}}_V$ which one gets by applying the definition (3.8a) is actually the following one:

$$\tilde{\mathscr{L}}_V\begin{pmatrix}\alpha^{(1)}(n)\\ \alpha^{(2)}(n)\end{pmatrix} = \begin{pmatrix} \alpha^{(2)}(n)+\alpha^{(2)}(n+1)+V^{(2)}(n+1)\alpha^{(1)}(n)-\Pi(n+1)\Big[\sum_{j=-\infty}^{n}\Pi^{-1}(j+1)\alpha^{(2)}(j)\Pi(j)-\Pi^{-1}(j)\alpha^{(2)}(j)\Pi(j-1)\\ \qquad +\Pi^{-1}(j+1)(V^{(2)}(j+1)\alpha^{(1)}(j)-\alpha^{(1)}(j)V^{(2)}(j))\Pi(j)\Big]\Pi^{-1}(n)\\ \alpha^{(1)}(n-1)V^{(1)}(n-1)+V^{(1)}(n)\alpha^{(1)}(n)+V^{(2)}(n)\alpha^{(2)}(n)-\sum_{j=-\infty}^{n-1}([\alpha^{(1)}(j),V^{(1)}(j)]+[\alpha^{(2)}(j),V^{(2)}(j)]) \end{pmatrix}. \tag{B1}$$

It coincides with (3.8b) only in the submanifold of the cotangent space whose elements fulfill the conditions:

$$\sum_{j=-\infty}^{+\infty}([\alpha^{(1)}(j),V^{(1)}(j)]+[\alpha^{(2)}(j),V^{(2)}(j)])=0, \tag{B2a}$$

$$\sum_{j=-\infty}^{+\infty}(\Pi^{-1}(j+1)\alpha^{(2)}(j)\Pi(j)-\Pi^{-1}(j)\alpha^{(2)}(j)\Pi(j-1) +\Pi^{-1}(j+1)(V^{(2)}(j+1)\alpha^{(1)}(j)-\alpha^{(1)}(j)V^{(2)}(j))\Pi(j))=0. \tag{B2b}$$

It can be easily seen, by direct calculation, starting from formula (5.4) and taking into account the asymptotic behaviors (5.2) that conditions (B2) are indeed fulfilled by G, for any value of the spectral parameter $z\in C$: thus, they are fulfilled by all the coefficients G_l of its power expansion and consequently by all the potential operators Q_{-2}, Q_{-1}, Q_j $(j=1,\ldots)$. But this is just what we need, since in all our derivations $\tilde{\mathscr{L}}_V$ acts only on linear combinations of the potential operators.

Proof of Corollary I: By direct calculation, it follows that $Q_0(V)$ obeys the following condition

$$\langle\tilde{\mathscr{L}}_V Q'_0(V)\cdot\Phi,\psi\rangle - \langle\tilde{\mathscr{L}}_V Q'_0(V)\cdot\Psi,\Phi\rangle = \langle Q_0(V),\mathscr{L}'_V(\Phi;\psi)-\mathscr{L}'_V(\Psi;\Phi)\rangle, \tag{B3}$$

which ensures that $Q_j(V)=\tilde{\mathscr{L}}^j_V Q_0(V)$ is again a potential operator if $\tilde{\mathscr{L}}_V$ is the adjoint of a Nijenhuis operator (see Ref. 5). For the sake of completeness, we show now that formula (4.8b) implies formula (4.8c) and vice-versa. Let us first recall the standard definition of the gradient of a functional $F[V]$:

$$\langle\nabla F,\Phi\rangle = \partial_\epsilon F[V+\epsilon\Phi]\big|_{\epsilon=0}. \tag{B4}$$

Evaluating now the rhs of (B4) for the functionals $I_j(V)$ defined by (4.8), we get:

$$\partial_\epsilon I[V+\epsilon\Phi]\big|_{\epsilon=0} = \int_0^1 d\lambda\,\{\langle Q(\lambda V),\Phi\rangle + \langle Q'(\lambda V)\cdot\lambda\Phi,V\rangle\}. \tag{B5}$$

Taking into account the potential nature of $Q(V)$ [formula (4.1)], Eq. (B5) yields:

$$\partial_\epsilon I[V+\epsilon\Phi]\big|_{\epsilon=0} = \int_0^1 d\lambda\,\{\langle Q(\lambda V),\Phi\rangle + \langle Q'(\lambda V)\cdot V,\lambda\Phi\rangle\} = \int_0^1 d\lambda\partial_\lambda\langle Q(\lambda V),\lambda\Phi\rangle = \langle Q(V),\Phi\rangle \quad \text{Q.E.D.} \tag{B6}$$

Viceversa, if (4.7) is true, by definition we have:

$$\partial_\epsilon I[V+\epsilon\Phi]\big|_{\epsilon=0} = \langle Q(V),\Phi\rangle = \int_0^1 d\lambda\partial_\lambda\langle Q(\lambda V),\lambda\Phi\rangle = \int_0^1 d\lambda\,[\langle Q(\lambda V),\Phi\rangle + \langle Q'(\lambda V)\cdot V,\lambda\Phi\rangle]. \tag{B7}$$

Using again property (4.1), formula (B7) can be recast in the form:

$$\partial_\epsilon I[V+\epsilon\Phi]\big|_{\epsilon=0} = \int_0^1 d\lambda\,\{\langle Q(\lambda V),\Phi\rangle + \langle Q'(\lambda V)\cdot\lambda\Phi,V\rangle\} = \partial_\epsilon\int_0^1 d\lambda\,\langle Q(\lambda(V+\epsilon\Phi)),V+\epsilon\Phi\rangle\big|_{\epsilon=0}, \tag{B8}$$

which implies that all the functionals $I_j[V]$ take just the form given in Eq. (4.8). Note that to derive formula (B7) we have implicitly assumed $\partial_\lambda\langle Q(\lambda V),\lambda\Phi\rangle$ to be nonzero and integrable in the interval $0\leqslant\lambda\leqslant 1$. This property is fulfilled, as it can be easily recognized by iteration, by all the potentials introduced in this paper, with the only exception of Q_{-2} (4.16). Indeed the functional I_{-2}, given by (4.18a), cannot be cast in the form (4.8).

Proof of formula (4.9): To prove formula (4.9), we start from formula (4.6), identifying there α with Q_l and β with Q_k. With these positions, and taking into account the property (4.2) of the Nijenhuis operator $\mathscr{L}_V$, the l.h.s. of (4.6) becomes:

$$\langle\Phi,\tilde{\mathscr{L}}_V Q'_l\cdot\mathscr{S}_V Q_k\rangle - \langle\mathscr{S}_V Q_k,\tilde{\mathscr{L}}_V Q'_l\cdot\Phi\rangle. \tag{B9}$$

On the other hand, taking care of the property (4.1) of the

potential operators, and of the skew-symmetricity of $\mathscr{S}_V$, formula (B9) can be written as:

$$\langle Q_k, \mathscr{S}_V \tilde{\mathscr{L}}_V Q_l' \cdot \Phi - \mathscr{S}_V Q_l' \cdot \mathscr{L}_V \Phi \rangle, \tag{B10}$$

so that Eq. (4.6) implies

$$\begin{aligned} &\mathscr{S}_V \tilde{\mathscr{L}}_V Q_l' \cdot \Phi - \mathscr{S}_V Q_l' \cdot \mathscr{L}_V \Phi \\ &+ \mathscr{L}_V'(\Phi; \mathscr{S}_V Q_l) + \mathscr{L}_V \mathscr{S}_V'(Q_l; \Phi) \\ &- \mathscr{S}_V'(Q_l; \mathscr{L}_V \Phi) = 0. \end{aligned} \tag{B11}$$

But formula (B11) is easily identified with (4.9), recalling that, due to formula (3.11), we have

$$H_l' \cdot \Phi = \mathscr{S}_V'(Q_l; \Phi) + \mathscr{S}_V Q_l' \cdot \Phi. \tag{B12}$$

Proof of formula (4.13): Although the proof that the Poisson-bracket (4.13) is well defined is standard, we report it here for the sake of completeness. We have to prove that

$$\{I_j, I_k\} = -\{I_k, I_j\} \text{ (skew-symmetricity)} \tag{B13}$$

and that

$$\{I_j, \{I_k, I_l\}\} + \text{cyclic permutations} = 0 \text{ (Jacobi identity)}. \tag{B14}$$

Equation (B13) follows immediately from the definition (4.13), taking into account the skew-symmetricity of the symplectic operator $\mathscr{S}_V$ [formula (4.3)]. As for (B14), we notice that, from (4.13) and (4.3), we have

$$-\{I_j, \{I_k, I_l\}\} = \langle \nabla \langle Q_k, \mathscr{S}_V Q_l \rangle, \mathscr{S}_V Q_j \rangle. \tag{B15}$$

On the other hand, taking care of (B4), the rhs of (B15) can be rewritten as

$$\begin{aligned} &\langle \nabla \langle Q_k, \mathscr{S}_V Q_l \rangle, \mathscr{S}_V Q_j \rangle \\ &= \partial_\epsilon \langle Q_k(V + \epsilon \mathscr{S}_V Q_j), \mathscr{S}_{V+\epsilon \mathscr{S}_V Q_j} \cdot Q_l(V + \epsilon \mathscr{S}_V Q_j) \rangle|_{\epsilon=0} \\ &= \langle Q_k' \cdot \mathscr{S}_V Q_j, \mathscr{S}_V Q_l \rangle + \langle Q_k, \mathscr{S}_V'(Q_l; \mathscr{S}_V Q_j) \rangle + \langle Q_k, \mathscr{S}_V Q_l' \cdot \mathscr{S}_V Q_j \rangle. \end{aligned} \tag{B16}$$

By using (4.3) and (4.1) and summing over all cyclic permutations of the indices j,k,l, terms like the first and the third in the rhs of (B16) cancel each other; on the other hand the cancellation of terms like the second one in the rhs of (B16) is ensured by the property (4.4) of the symplectic operator $\mathscr{S}_V$.

APPENDIX C

Proof of Lemma I: It follows by direct calculation, taking care of the "eigenvalue" equations (2.1), (5.1), and of the asymptotic behaviors (5.2).

Proof of Lemma II: To prove this lemma, we introduce the two matrix functions $\chi(n,z)\bar{\chi}(n,z)$, through the formulas:

$$\chi(n,z) = z^{-n} F^{(+)}(n,z), \tag{C1a}$$

$$\bar{\chi}(n,z) = z^n \bar{F}^{(-)}(n,z). \tag{C1b}$$

They obviously fulfill the difference equations

$$\begin{aligned} &\chi(n-1,z) - \chi(n,z) + z V_{(n)}^{(2)} \chi(n,z) \\ &+ z^2 (V_{(n)}^{(1)} \chi(n+1,z) - \chi(n,z)) = 0, \end{aligned} \tag{C2a}$$

$$\begin{aligned} &\bar{\chi}(n+1,z) - \bar{\chi}(n,z) + z \bar{\chi}(n,z) V_{(n)}^{(2)} \\ &+ z^2 (\bar{\chi}(n-1,z) V_{(n-1)}^{(1)} - \bar{\chi}(n,z)) = 0. \end{aligned} \tag{C2b}$$

With the boundary conditions

$$\lim_{n \to +\infty} \chi(n,z) - I = \lim_{n \to -\infty} \bar{\chi}(n,z) - I = 0. \tag{C2c}$$

In Ref. 2 it has been shown that both $\chi, \bar{\chi}$ can be analytically continued inside the unit circle of the complex z-plane, so that we can consider their McLaurin expansion

$$\chi(n,z) = \sum_{l=0}^{\infty} \chi_l(n) z^l, \tag{C3a}$$

$$\bar{\chi}(n,z) = \sum_{l=0}^{\infty} \bar{\chi}_l(n) z^l, \tag{C3b}$$

which, once inserted in (C2a,b), taking care of (C2c), yield the recursion relations

$$\begin{aligned} \chi_l(n) = \sum_{j=n+1}^{\infty} &\{\chi_{l-2}(j) - V^{(1)}(j) \chi_{l-2}(j+1) \\ &- V^{(2)}(j) \chi_{l-1}(j)\}, \end{aligned} \tag{C4a}$$

$$\begin{aligned} \bar{\chi}_l(n) = \sum_{j=-\infty}^{n-1} &\{\bar{\chi}_{l-2}(j) - \bar{\chi}_{l-2}(j-1) V^{(1)}(j-1) \\ &- \bar{\chi}_{l-1}(j) V^{(2)}(j)\}, \end{aligned} \tag{C4b}$$

with

$$\chi_0 = \bar{\chi}_0 = I, \tag{C4c}$$

$$\chi_1(n) = -\sum_{j=n+1}^{\infty} V^{(2)}(j), \quad \bar{\chi}_1(n) = -\sum_{j=-\infty}^{n-1} V^{(2)}(j). \tag{C4d}$$

On the other hand, due to the asymptotic behaviors (5.2), we have

$$\begin{aligned} \frac{\hat{T}(z)}{z - z^{-1}} &= W^{-1}[\bar{F}^{(-)}, F^{(+)}] \\ &= (z \bar{\chi}(n,z) V^{(1)}(n) \chi(n+1,z) - z^{-1} \bar{\chi}(n+1,z) \chi(n,z))^{-1}, \end{aligned} \tag{C5}$$

where the Wronskian $W[\bar{\psi}, \psi]$ of two solutions of the spectral problems (2.1), (5.1) is defined as[2]

$$W[\bar{\psi}, \psi] = \bar{\psi}(n,z) V^{(1)}(n) \psi(n+1,z) - \bar{\psi}(n+1,z) \psi(n,z). \tag{C6}$$

Formula (C5) allows to evaluate recursively the coefficients of the McLaurin expansion of $\hat{T}(z)$, which holds true in the circle C defined in formula (5.6). We report here the first coefficients of this expansion

$$\hat{T}(z) = I + z \sum_{j=-\infty}^{+\infty} V^{(2)}(j) + O(z^2). \tag{C7}$$

Formulas (C4) and (C5) allow to evaluate the coefficients of the McLaurin series (5.5) of the "squared-eigenfunction" G.

Proof of Lemma III: We state from the spectral problem (2.1) written in terms of the Jost solution $F^{(+)}(n,z)$. A variation $\delta F^{(+)}(n,z)$ of this solution fulfills the nonhomogeneous difference equation

$$\delta F^{(+)}(n-1,z) + (V^{(2)}(n) - \lambda)\delta F^{(+)}(n,z) + V^{(1)}(n)\delta F^{(+)}(n+1,z) = -\delta V^{(2)}(n)F^{(+)}(n,z) - \delta V^{(1)}(n)F^{(+)}(n+1,z). \tag{C8}$$

This equation, with the boundary conditions $\lim_{n\to+\infty} \delta F^{(+)}(n,z) = 0$ can be immediately integrated by the standard method of variation of parameters, yielding

$$\begin{aligned}\delta F^{(+)}(n,z) = {} & \psi_1(n,z)W^{-1}[\bar\psi_2,\psi_1]\sum_{j=n+1}^{\infty}\bar\psi_2(j,z) \\ & \times[\delta V^{(2)}(j)F^{(+)}(j,z) + \delta V^{(1)}(j)F^{(+)}(j+1,z)] \\ & + \psi_2(n,z)W^{-1}[\bar\psi_1,\psi_2]\sum_{j=n+1}^{\infty}\bar\psi_1(j,z) \\ & \times[\delta V^{(2)}(j)F^{(+)}(j,z) + \delta V^{(1)}(j)F^{(+)}(j+1,z)],\end{aligned} \tag{C9}$$

where $\psi_1(n,z)[\bar\psi_1,(n,z)],\psi_2(n,z)[\bar\psi_2(n,z)]$ are two independent solutions of (2.1), (5.1), and $W[\bar\psi,\psi]$ is the Wronskian defined in formula (C6). By choosing $\psi_{1,2} = F^{(\pm)},\bar\psi_{1,2} = \bar F^{(\pm)}$, letting in formula (C9) $n\to-\infty$ and taking into account the asymptotic behaviors (5.2) one gets

$$\begin{aligned}&\operatorname{Tr}\hat T^{-1}\delta\hat T \\ &= \operatorname{Tr}\sum_{j=-\infty}^{+\infty}\left(F^{(+)}(j,z)\frac{\hat T}{z^{-1}-z}\bar F^{(-)}(j,z)\delta V^{(2)}(j)\right. \\ &\quad \left. + F^{(+)}(j+1,z)\frac{\hat T}{z^{-1}-z}\bar F^{(-)}(j,z)\delta V^{(1)}(j)\right) = \langle G,\delta V\rangle.\end{aligned} \tag{C10}$$

We notice that Eq. (C10), which has been derived for $|z| = 1$, can be analytically continued inside the unit circle [as the Jost solutions $F^{(+)},\bar F^{(-)}$ as well as the transmission coefficient $\hat T$ enjoy this property[1,2]] with the exception of the set $\{z_j:\det\hat T^{-1}(z_j) = 0\}$. But, as noticed in Refs. 1,2, this set is a finite one, so that we can always find a simply connected region belonging to the unit circle and including C [see formula (5.6)], where $\ln(\det\hat T)$ is a monodromic analytic function of z. This enables us to rewrite the lhs of (C10) as $\delta\ln(\det\hat T)$, and thus Lemma III is proved.

Proof of Lemma IV: From formulas (2.1), (5.1), (C6) we can derive the identity:

$$W[\bar F^{(-)},F_z^{(+)}](n) - W[\bar F^{(-)},F_z^{(+)}](n-1) = (1-z^{-2})\bar F^{(-)}(n,z)F^{(+)}(n,z) \tag{C11}$$

By summing up (C11) from $-\infty$ to $+\infty$ and taking care of the asymptotic behaviors (5.2), one gets

$$\frac{z}{z^{-1}-z}\operatorname{Tr}\hat T^{-1}\hat T_z = \operatorname{Tr}\sum_{n=-\infty}^{+\infty}\left[G^{(2)}(n,z) - \frac{I}{z^{-1}-z}\right]. \tag{C12}$$

Through the arguments already given in the proof of Lemma III, we can conclude that Eq. (C12) holds true for $z\in C$ [formula (5.6)], where the lhs can be written as $z/(z^{-1}-z)\partial_z\ln(\det\hat T)$, so that Lemma IV is proved.

Proof of formula (5.15): According to formula (3.3) of Ref. 2, identifying there $\bar\psi(n)$ with $(\hat T(z)/(z^{-1}-z)\bar F^{(-)}_{(n,z)}$ and $\chi(n)$ with $\sigma F^{(+)}(n,z)\Pi(n)\tau\Pi^{-1}(n+1)F^{(+)}(n+1,z)$, we get

$$\begin{aligned}&\sum_{j=-\infty}^{+\infty}\frac{\hat T(z)}{z^{-1}-z}\bar F^{(-)}(n,z)[V^{(1)}(n)F^{(+)}(n+1,z) + V^{(2)}(n)F^{(+)}(n,z)] \\ &\quad = \hat T(z)\sigma\hat T^{-1}(z) - \sigma + z[\hat T(z)\Pi(-\infty)\tau(\hat T(z)\Pi(-\infty))^{-1} - \tau]\end{aligned} \tag{C13}$$

whence, taking the trace, we immediately recover formula (5.15). Note that this proof relies just on the properties of the spectral problems (2.1), (5.1), and thus it holds whatever be the time-dependence of the matrices σ and τ.

ACKNOWLEDGMENTS

The authors wish first of all to warmly thank F. Magri, for his illuminating discussions and constant help throughout the accomplishment of this work. They also wish to acknowledge some useful discussions with D. Levi and A. Degasperis.

[1]M. Bruschi, S. V. Manakov, O. Ragnisco, and D. Levi, J. Math. Phys. **21**, 2749 (1980).
[2]M. Bruschi, D. Levi, and O. Ragnisco, J. Math. Phys. **22**, 2463 (1981). Unfortunately, this paper is affected by various misprints.
[3]M. Bruschi and O. Ragnisco, J. Phys. A: Math., Nucl., Gen. **14**, 1075 (1981).
[4]See for instance: M. Toda, Progr. Theor. Phys. Suppl. **59**, 1 (1976); H. Flaschka, Phys. Rev. **B9**, 1924 (1974); Progr. Theor. Phys. **51**, 705 (1974); S. V. Manakov, Sov. Phys. JETP **40**, 269 (1974).
[5]F. Magri: "A Geometrical Approach to the Nonlinear Solvable Equations," in *Nonlinear Evolution Equations and Dynamical Systems*, edited by M. Boiti, F. Pempinelli, and G. Soliani, Lecture Notes in Physics **120** (Springer–Verlag, Berlin, 1980). For a closely related approach see also B. Fuchssteiner and A. S. Fokas, Physica D **4**, 47 (1981).
[6]We have to notice that in Refs. 2,3 the definition of the recursion operator contains a misprint.
[7]We denote by α an element of the domain of $\mathscr{S}_V$, since it belongs to the cotangent space T^*_V, as it will be clear in the following.
[8]As it will be shown later (see Remark II), the operator has a nontrivial Ker; hence it should be more exactly denoted as a pre-symplectic operator.
[9]F. Magri, J. Math. Phys. **19**, 1156 (1978).
[10]For an analogous approach in the continuous case, see for instance: E. Olmedilla, L. Martinez Alonso, and F. Guil, Nuovo Cimento B, **61**, 49 (1981); O. Ragnisco, Lett. Nuovo Cimento **31**, 651 (1981); E. Olmedilla, L. Martinez Alonso, and F. Guil, J. Math. Phys. (to be published).

On the expansion of integrals containing Fermi distributions

Sven Åberg[a)]
Centre d'Etudes Nucléaires, Grenoble, France and Nordita, Blegdamsvej 17, DK-2100, Copenhagen, Denmark

(Received 8 September 1982; accepted for publication 14 January 1983)

We derive an expansion of integrals containing a general function multiplied by a Fermi function raised to an arbitrary power ν. When ν is an integer, direct expressions for the expansion coefficients are given. The expansion is found to converge quite rapidly if the diffuseness of the Fermi distribution is small, and when $\nu \gtrsim 1$.

PACS numbers: 02.30.Mv

In a recent paper,[1] Krivine and Treiner considered the expansion of integrals of the type

$$I_1 = \int_0^\infty r^n \rho^\nu(r) dr, \tag{1}$$

where ρ is a Fermi distribution

$$\rho(r) = \frac{1}{1 + \exp[(r - R)/a]} \tag{2}$$

and ν is positive and n is a positive integer. The expansion was performed in terms of powers of a/R which, for the nuclear physics applications discussed in Ref. 1, is a small number (~ 0.1). It is the purpose of this paper to discuss the expansion of integrals of the more general type

$$I = \int_0^\infty f(r) \rho^\nu(r) dr, \tag{3}$$

where f is a general function and ν is a real positive number. We will achieve an expansion which has a more rapid convergence compared to that of Ref. 1. When ν is a positive integer, direct expressions for the expansion coefficients will be given.

We start by integrating Eq. (3) by parts:

$$I = [F(r)\rho^\nu(r)]_0^\infty + \int_0^\infty F(r) P(r) dr, \tag{4}$$

where

$$-P(r) = \nu \rho^{\nu-1}(d\rho/dr) \tag{5}$$

and

$$F(r) = \int^r f(r') dr'. \tag{6}$$

F is expanded around $r = R_0$:

$$F(r) = \sum_{m=0}^\infty \frac{1}{m!} \left[\frac{d^m F}{dr^m}\right]_{R_0} (r - R_0)^m, \tag{7}$$

where

$$P(R_0) = \max P(r). \tag{8}$$

Solving Eq. (8) gives

$$R_0 = R - a \ln \nu, \tag{9}$$

which gives $R_0 = R$ only when $\nu = 1$. We require that f is continuous and repeatedly differentiable at $r = R_0$. Furthermore, f should vary rather slowly in the region around R_0. Substituting $s = (r - R_0)/a$ and using Eqs. (7) and (9), Eq. (4) may be written as

$$I = -F(0)\rho^\nu(0) + \sum_{m=0}^\infty \frac{1}{m!} \left[\frac{d^m F}{dr^m}\right]_{R_0} a^m J_{m,\nu} + \sigma(f,\nu), \tag{10}$$

where

$$J_{m,\nu} = \int_{-\infty}^\infty \frac{s^m e^s \, ds}{(1 + (1/\nu)e^s)^{\nu+1}}. \tag{11}$$

The rest function σ is due to the change of the lower integration limit in Eq. (11) from $-R_0/a$ to $-\infty$ and goes to zero faster than any power of a/R_0.

Generally the expansion coefficients [Eq. (11)] have to be calculated numerically. When ν is an integer the integrals in Eq. (11) may, however, be solved analytically. If the function

$$g_\nu(\gamma) = \int_{-\infty}^\infty \frac{e^{(\gamma+1)s} \, ds}{(1 + (1/\nu)e^s)^{\nu+1}} \tag{12}$$

is introduced, we obtain the relation

$$J_{m,\nu} = \left[\frac{d^m g_\nu}{d\gamma^m}\right]_{\gamma=0}. \tag{13}$$

By the change of variables $t = (1/\nu)e^s$, Eq. (12) becomes

$$g_\nu(\gamma) = \nu^{\gamma+1} \int_0^\infty \frac{t^\gamma \, dt}{(1+t)^{\nu+1}}, \tag{14}$$

where the integral is recognized as the beta function

$$g_\nu(\gamma) = \nu^{\gamma+1} B(1 + \gamma, \nu - \gamma). \tag{15}$$

The beta function can be written in terms of Γ functions:

$$B(1 + \gamma, \nu - \gamma) = \frac{\Gamma(1+\gamma)\Gamma(\nu - \gamma)}{\Gamma(1 + \nu)}. \tag{16}$$

Using well-known relations for the Γ functions, Eq. (15) becomes

$$g_\nu(\gamma) = \frac{\nu^{\gamma+1}}{\nu!} \frac{\pi\gamma}{\sin \pi\gamma} (\nu - 1 - \gamma)(\nu - 2 - \gamma)\cdots(1 - \gamma), \tag{17}$$

providing ν is an integer.

By expanding the three factors of Eq. (17),

$$\nu^\gamma = \sum_{k=0}^\infty e_k^{(1)} \gamma^k, \tag{18a}$$

$$\frac{\pi\gamma}{\sin \pi\gamma} = \sum_{k=0}^\infty e_k^{(2)} \gamma^k, \tag{18b}$$

[a)] On leave of absence from Department of Mathematical Physics, Lund, Sweden.

 0022-2488/83/061422-03$02.50 © 1983 American Institute of Physics

$$\frac{1}{(\nu-1)!}(\nu-1-\gamma)(\nu-2-\gamma)\cdots(1-\gamma)=\sum_{k=0}^{\infty} e_k^{(3)}\gamma^k, \tag{18c}$$

where

$$e_k^{(1)}=\frac{1}{k!}\ln^k\nu, \tag{19a}$$

$$e_k^{(2)}=\begin{cases}1, & k=0,\\ 0, & k \text{ odd},\\ \dfrac{2(2^{k-1}-1)|B_k|\pi^k}{k!}, & k \text{ even},\end{cases} \tag{19b}$$

$$e_k^{(3)}=\begin{cases}1, & k=0,\\ (-1)^k\displaystyle\sum_{l_1=1}^{\nu-1}\sum_{l_2=l_1}^{\nu-1}\cdots\sum_{l_k=l_{k-1}-1}^{\nu-1}\prod_{i=1}^{k}\frac{1}{l_i}, & \text{other } k,\end{cases} \tag{19c}$$

Eq. (17) may be written as

$$g_\nu(\gamma)=\sum_{m=0}^{\infty} d_m(\nu)\gamma^m, \tag{20a}$$

$$d_m(\nu)=\sum_{i=0}^{m}\sum_{j=0}^{i} e_j^{(1)}e_{i-j}^{(2)}e_{m-i}^{(3)}. \tag{20b}$$

The coefficients B_k in Eq. (19b) are Bernoulli numbers. Thus, providing ν is an integer, we have obtained a direct expression for the integrals of Eq. (11). The expansion coefficients in Eq. (10) are then

$$J_{m,\nu}=m!d_m(\nu), \tag{21}$$

where d_m are given by Eqs. (19) and (20).

For example, we get

$$J_{0,\nu}=1, \tag{22a}$$

$$J_{1,\nu}=\ln\nu-\sum_{k=1}^{\nu-1}\frac{1}{k}, \tag{22b}$$

$$J_{2,\nu}=\frac{\pi^2}{3}+\ln^2\nu-2\ln\nu\sum_{k=1}^{\nu-1}\frac{1}{k}+2\sum_{k=1}^{\nu-1}\sum_{n=k+1}^{\nu-1}\frac{1}{k\cdot n}. \tag{22c}$$

Equation (22a) also holds for noninteger values of ν. When $\nu=1$, the usual Sommerfeld expansion coefficients[2] are (of course) given by Eq. (21).

We now want to compare the above given expansion with that of Ref. 1. Putting $f(r)=r^n$, i.e., considering Eq. (1), Eq. (10) becomes

$$I^{(1)}=\frac{R_0^{n+1}}{n+1}\left[1+\frac{a}{R_0}(n+1)\times\sum_{m=0}^{n}\frac{1}{m+1}\binom{n}{m}\left(\frac{a}{R_0}\right)^m J_{m+1,\nu}\right]+\sigma, \tag{23}$$

which is an expression similar to that considered in Ref. 1:

$$I^{(2)}=\frac{R^{n+1}}{n+1}\left[1+\frac{a}{R}(n+1)\sum_{m=0}^{n}\binom{n}{m}\left(\frac{a}{R}\right)^m\eta_\nu^{(m)}\right]+\omega, \tag{24}$$

where

$$\eta_\nu^{(m)}=(-1)^m\int_0^\infty u^m\left[\frac{1+(-1)^m e^{-u\nu}}{(1+e^{-u})^\nu}-1\right]du. \tag{25}$$

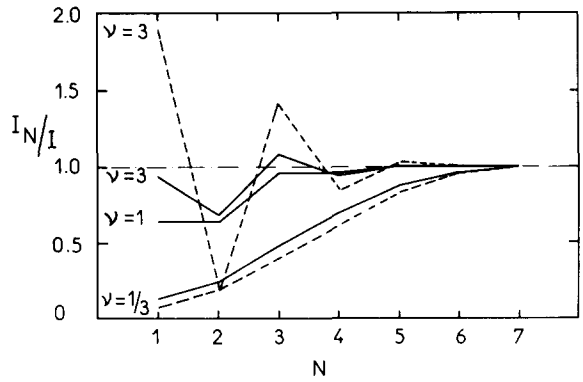


FIG. 1. The expansion of the integral (1) with $n=5$ and $a/R=0.1$ for some different values of ν. A truncation of the expansions Eq. (23) (expansion around $r=R_0$, solid curve) and Eq. (24) (expansion around $r=R$, dashed curve), compared to the exact value of the integral, is plotted vs the number of considered terms N. When $\nu=1$, the two expansions become identical.

If one integrates Eq. (25) by parts and makes the change of variable $u\to -u$ in a part of the integral, one obtains

$$\eta_\nu^{(m)}=\frac{\nu}{m+1}\int_{-\infty}^{\infty}\frac{u^{m+1}e^u\,du}{(1+e^u)^{\nu+1}}, \tag{26}$$

which, in a similar way as $J_{m,\nu}$, can be calculated as

$$\eta_\nu^{(m)}=\frac{\nu}{m+1}\left[\frac{d^{m+1}B(1+\gamma,\nu-\gamma)}{d\gamma^{m+1}}\right]_{\gamma=0}, \tag{27}$$

providing ν is an integer. If we, in the derivation of Eq. (23), perform the expansion of $F(r)$ [Eq. (7)] around $r=R$ instead of around the maximum of P, $r=R_0$, Eqs. (23) and (24) are identical and

$$J'_{m+1,\nu}=(m+1)\eta_\nu^{(m)}. \tag{28}$$

The prime on the J's imply that the expansion in Eq. (7) is performed around $r=R$. When ν is an integer, an analytical expression for $J'_{m,\nu}$ is given by Eq. (21) if all $e_j^{(1)}$ are put equal to 1. Note that the two expansions (23) and (24) are identical when $\nu=1$ and the usual Sommerfeld expansion[2] is then obtained in both cases. In Fig. 1, the two expansions are compared for some different values of ν, when $n=5$ and $a/R=0.1$. The two expansions, truncated after N terms and divided by a numerical solution of Eq. (1), are plotted vs N. For both expansions, the convergence is very poor for small values of ν. As $n=5$, Eqs. (23) and (24) only contain seven terms. The difference between I_7 and I_{exact} in Fig. 1 is thus a measure of the neglect of σ in Eq. (10). Figure 1 clearly shows how the convergence is improved when the expansion is performed around the maximum of P [Eq. (5)] instead of around $r=R$, which is the maximum of P only when $\nu=1$. The outcome of Fig. 1 is general and is also valid for other values of n, a/R, and ν. For $\nu<1$ we get a damped convergence, while the expansion shows an oscillating behavior when $\nu>1$. For higher values of n and/or lower values of a/R, the expansion converges more rapidly. For $a/R\lesssim 0.1$ and $\nu\gtrsim 3$, already the first term in the expansion (10) gives a very good approximation. The second term may then spoil this agree-

ment while the third term restores it again, as can be seen in Fig. 1 when $\nu = 3$.

ACKNOWLEDGMENTS

I am grateful to Gautam Ghosh and Peter Schuck for many helpful discussions. I would also like to express my gratitude for the kind hospitality at Centre d'Etude Nuclèaires and Institut Laue-Langevin, Grenoble.

[1]H. Krivine and J. Treiner, J. Math. Phys. **22**, 2484 (1981).
[2]A. Sommerfeld, Z. Phys. **47**, 1 (1928).

Conformal geometry of flows in n dimensions

Ivor Robinson
Program in Mathematical Sciences, The University of Texas at Dallas, Richardson, Texas 75080

Andrzej Trautman[a)]
Département de Mathématiques et de Statistique, Université de Montréal, Montréal, Québec, H3C 3J7, Canada

(Received 24 June 1982; accepted for publication 20 August 1982)

Flows generated by smooth vector fields are considered from the point of view of conformal geometry. A flow is defined to be conformally geodesic if it preserves the distribution of vector spaces orthogonal to the lines of the flow. It is shear-free if, moreover, it preserves the conformal structure on these vector spaces. Differential equations characterizing such flows are derived for the general case of an n-dimensional conformal space of arbitrary signature. In the special case of null flows in spacetime, one obtains a refined version of the theorem connecting null solutions of Maxwell's equations with null flows that are geodesic and shear-free.

PACS numbers: 02.40. + m, 04.20.Cv, 04.20.Jb

I. INTRODUCTION

It has been known since the work of Hermann Weyl[1] that there are two essential geometric ingredients in relativistic theories of spacetime: a family of null cones, defined by the rays of light, and the projective structure, related to freely falling particles. The latter structure is usually described by means of an infinitesimal connection, while the former leads to conformal geometry. When suitable compatibility relations between these structures are assumed, they can be shown to arise from Riemannian geometry.[2] There is a part of physics which depends only on the conformal structure: not only Maxwell's equations,[3] but essentially all equations of massless particles,[4,5] including the Yang–Mills equations,[6] are conformally invariant. Moreover, in the domain of phenomena where particles are produced and scattered with energies much larger than their masses, one can use an approximate description of their behavior, based on neglecting their masses and assuming conformal invariance of the fundamental laws. Conformal ideas also underlie the twistor program of Penrose.

Since the propagation of light is closely related to conformal geometry, it is not surprising that the geodesic, or nongeodesic, character of a null curve is invariant under conformal changes of the metric. This is not true of timelike and spacelike curves. Moreover, congruences of null geodesics are intimately linked to null solutions of Maxwell's equations. It is easy to see that a null electromagnetic field—a field which at any point has the same algebraic form as a plane wave—has a stress–energy tensor of the form $k^a k^b$. Since that tensor is always traceless, the vector field k is always null; and since it is also conserved, the field k is tangent to a congruence of geodesics. There is another, subtler property of k which, together with the null geodesic property, characterizes completely the vector fields associated with null electromagnetic fields.[7] The additional property is related to the absence of shear or distortion: One can visualize it by thinking of the congruence generated by k as representing a beam of light. The study of shear-free congruences of null geodesics (rays) played an important role in the investigation of exact solutions to Einstein's equations. Goldberg and Sachs[8] showed that an empty spacetime admits such a congruence of rays if, and only if, its curvature tensor is algebraically degenerate. This theorem was generalized in a conformally invariant manner by Robinson and Schild.[9] Moreover, a systematic search for solutions of Einstein's equations admitting congruences of rays led to the discovery of simple spherical gravitational waves[10] and of the Kerr black hole[11] as well as of other metrics.[12]

This paper contains a generalization of the notion of a congruence of rays to n-dimensional spaces of arbitrary signature. It turns out that, although in conformal geometry there is no way of assigning the geodesic property to a single nonnull curve, there is a well-defined notion of a conformally geodesic flow. All relevant notions—geodesity, expansion, vorticity, and shear—are defined in terms which are manifestly invariant under conformal transformations.

II. NOTATION

To a large extent, this paper follows the terminology and notation prevalent in differential geometry and mathematical physics.[13–15] All manifolds and maps are of class C^∞. If $f:M\to N$ is a map of a manifold M into another manifold N, then $Tf:TM\to TN$ is the *tangent*, or derived, map of the corresponding tangent bundles. If g is a covariant tensor field on N, then f^*g denotes its *pullback* to M. If k is a vector field on M and α is a p-form on M, then $k\rfloor\alpha$ is the $(p-1)$-form on M obtained by *contracting* α with k: that is, if $x\in M$ and $u_2,\ldots,u_p\in T_xM$, then

$$(k\rfloor\alpha)(u_2,\ldots,u_p)=\alpha(k(x),u_2,\ldots,u_p). \tag{2.1}$$

A vector field k on M *generates a flow* (φ_t) on M, i.e., a local one-parameter group of local transformations of M. For any (sufficiently small) $t\in R$, φ_t is a diffeomorphism of an open submanifold of M onto another such submanifold. If both t and s are sufficiently small,

$$\varphi_t\circ\varphi_s=\varphi_{t+s}\,, \tag{2.2}$$

[a)] Permanent address: Instytut Fizyki Teoretycznej Uniwersytetu Warszawskiego, Hoża 69, Warszawa, Poland.

 0022-2488/83/061425-05$02.50 © 1983 American Institute of Physics

provided that the domains of the maps occurring in this equation are restricted to make both its sides meaningful. Since φ_t is a (local) diffeomorphism, it can be used to pull back any tensor field A on M, and the last equation implies that

$$\varphi_t^*(\varphi_s^* A) = \varphi_{t+s}^* A . \tag{2.3}$$

Differentiating both sides with respect to s at $s = 0$, one obtains

$$\frac{d}{dt}\varphi_t^* A = \varphi_t^* \mathcal{L}_k A , \tag{2.4}$$

where

$$\mathcal{L}_k A = \frac{d}{dt}\varphi_t^* A \Big|_{t=0} \tag{2.5}$$

is the *Lie derivative* of A with respect to k. It is known that the Lie derivative of a p-form α can be evaluated from the formula

$$\mathcal{L}_k \alpha = k\rfloor d\alpha + d(k\rfloor\alpha) , \tag{2.6}$$

where d denotes the *exterior derivative*. The *exterior product* of forms is denoted by the wedge symbol. Both d and the interior product by k, $k\rfloor$, are antiderivations of the Cartan algebra of differential forms on M; e.g.,

$$k\rfloor(\alpha\wedge\beta) = (k\rfloor\alpha)\wedge\beta + (-1)^p\alpha\wedge(k\rfloor\beta) , \tag{2.7}$$

for any p-form α and form β.

By a *Riemannian space* we understand a manifold M with a metric tensor g which is nonsingular, but not necessarily positive-definite. In terms of local coordinates (x^a), $a = 1,\ldots,n$, the metric tensor may be written as $g_{ab}dx^a\otimes dx^b$; and a similar notation is used whenever it is convenient to represent tensors by their components. For example, if k and l are vector fields on M, then $\kappa = g(k)$ is a 1-form, $l\rfloor g(k) = g(k,l)$, and

$$k = k^a\partial/\partial x^a, \quad \kappa = k_a dx^a, \quad \text{where } k_a = g_{ab}k^b . \tag{2.8}$$

The *Levi-Civita connection* ∇ associated with g is metric and has no torsion, so that one can write

$$d\kappa = k_{a;b}dx^b\wedge dx^a = g_{ac}k^c{}_{;b}dx^b\wedge dx^a , \tag{2.9}$$

if it is accepted that $(k^a{}_{;b})$ denotes the components of k with respect to the local coordinates (x^a). One writes ∇_ℓ for covariant differentiation along a vector field ℓ.

A conformal structure $[g]$ on M is an equivalence class of Riemannian metrics on M. Two metrics, g and g', belong to the same class if and only if there is a positive function h on M such that $g' = hg$. One also says that the metrics g and g' are *conformally related* to each other. In *conformal geometry*, one is interested in the relations which derive from the conformal structure: i.e., those which do not change under a replacement of g by $g' = hg$. Such relations, properties, and structures are said to be *conformally invariant*. For example, assume M to be n-dimensional, and put $\gamma = \det(g_{ab})$. Then

$$\hat{g}_{ab} = |\gamma|^{-1/n}g_{ab} \tag{2.10}$$

defines on M a conformally-invariant tensor density g of weight $2/n$. The *signature* of g is also a conformal invariant. A *conformal spacetime* is a four-dimensional manifold with a conformal structure and the same signature as that of Minkowski space.

On an oriented Riemannian space, there is a standard definition of the *Hodge duals* of differential forms. In this paper, however, we use $\hat{g}$ to define duals relative to the conformal structure of M. Thus, if ω is a p-form of weight q, then $*\omega$ is the $(n-p)$-form, of weight $q - 1 + 2(n-p)/n$, defined by

$$\omega\wedge\hat{g}(u_{p+1})\wedge\cdots\wedge\hat{g}(u_n) = *\omega(u_{p+1},\ldots,u_n)\text{vol} \tag{2.11}$$

when $u_{p+1},\ldots,u_n$ are arbitrary vectors, and vol denotes the volume element on M. If M is even-dimensional, then the $(n/2)$-forms α and $*\alpha$ have the same weight. If k is a vector field and $\hat{\kappa} = \hat{g}(k)$ the 1-form corresponding to it in a conformally invariant way, then

$$k\rfloor *\omega = *(\omega\wedge\hat{\kappa}) \tag{2.12}$$

for any p-form ω.

III. CONFORMALLY GEODESIC FLOWS

Let k be a nowhere-vanishing vector field on M. It defines a distribution K of one-dimensional subspaces of the tangent spaces to M: For any $x\in M$, the subspace $K(x)$ consists of all vectors parallel to $k(x)$. The distribution K is invariant under the action of the flow (φ_t) generated by k. A conformal structure $[g]$ on M, together with K, defines a distribution $K^\perp$ of $(n-1)$-dimensional subspaces: $K^\perp(x)$ is the set of all vectors tangent to M at x and orthogonal to $k(x)$ relative to $[g]$. At any point $x\in M$, the sign $\epsilon(x)$ of $g(k(x), k(x))$ is a conformal invariant. We assume k to be such that ϵ is constant on M. The distributions K and $K^\perp$ define vector bundles over M. If $\epsilon\neq 0$, then $K(x)\cap K^\perp(x) = \{0\}$, and

$$TM = K\oplus K^\perp , \tag{3.1}$$

whereas in the null case, $\epsilon = 0$, $K\subset K^\perp$, and one can form the exact sequence of vector bundles over M,

$$0\to K\to K^\perp\to K^\perp/K\to 0 . \tag{3.2}$$

The fibers of the quotient bundle $K^\perp/K$ are $(n-2)$-dimensional, and correspond to the "screen spaces" considered in early analyses of the geometry of null geodesic congruences.[16,17]

Definition 1: The flow (φ_t) generated by k is said to be *conformally geodesic* if it preserves the distribution $K^\perp$:

$$T\varphi_t\circ K^\perp = K^\perp\circ\varphi_t . \tag{3.3}$$

Sometimes, when there is no danger of confusion, we describe a flow as *geodesic*, although it is actually conformally geodesic. For any $g\in[g]$, one defines the 1-form $\kappa = g(k) = g_{ab}k^b dx^a = k_a dx^a$, and obtains $K^\perp(x) = \ker\kappa(x)$. If $g' = hg$, then $\kappa' = g'(k) = h\kappa$ has the same kernels as κ.

Theorem 1: The following conditions are equivalent:

(i) $K^\perp$ is invariant under (φ_t);

(ii) $\kappa\wedge\mathcal{L}_k\kappa = 0$; (3.4)

(iii) there exists a metric $g\in[g]$ such that $g(k,k)$ is constant on M and the congruence defined by (φ_t) consists of affinely parameterized Riemannian geodesics of g.

Remark: Condition (ii) is invariant under a conformal change: that is, under a replacement of κ by $\kappa' = h\kappa$. In gen-

eral, it is not invariant under changes in the parameterization of the congruence: that is, under the replacement of k by fk, where f is a function on M. It is invariant under the latter change, however, if the congruence is null.

Proof of Theorem 1: Since $K^{\perp}(x) = \ker \kappa(x)$, invariance of K under φ_t is equivalent to $\kappa \wedge \varphi_t^* \kappa = 0$, and, consequently, from (2.4), to (ii). Next choose $g \in [g]$ so that $k \rfloor \kappa = g(k, k) = \text{const}$ on M. From (2.6),

$$\begin{aligned} \mathcal{L}_k \kappa &= k \rfloor d\kappa \\ &= k \rfloor (k_{a;b} dx^b \wedge dx^a) = k_{a;b} k^b dx^a , \end{aligned} \tag{3.5}$$

so that (ii) is equivalent to

$$k^{[a} k^{b]}{}_{;c} k^c = 0 . \tag{3.6}$$

If the vector field k is nonnull, then (3.6) implies

$$k^b{}_{;c} k^c = 0 , \tag{3.7}$$

and the parameter t of the flow (φ_t) is affine. If k is null, then (3.7) can be imposed by a conformal change of the metric tensor. This completes the proof of Theorem 1.

IV. THE PROPERTY OF BEING SHEAR-FREE

The vector bundle $K^{\perp} \to M$ has a conformal structure induced by $[g]$. In other words, for any $x \in M$, the vector space $K^{\perp}(x)$ is endowed with an equivalence class of scalar products, two scalar products being equivalent if and only if one is a positive multiple of the other. The scalar products on $K^{\perp}(x)$ need not be nonsingular, even though g has been assumed to be such.

If the flow (φ_t) preserves $K^{\perp}$, then the vector spaces $T\varphi_t(K^{\perp}(x))$ and $K^{\perp}(\varphi_t(x))$ coincide; and one can compare the conformal structure on $K^{\perp}(\varphi_t(x))$ with that on $K^{\perp}(x)$ by "dragging" (Lie transporting) tangent vectors by means of $T\varphi_t$. These considerations suggest the following:

Definition 2: A conformally geodesic flow (φ_t) is said to be *shear-free* if it preserves the conformal structure of $K^{\perp}$. In other words, (φ_t) is shear-free if, for any $t \in \mathbb{R}$ and $x \in M$, there exists $h \in \mathbb{R}$, $h > 0$, such that

$$u, v \in K^{\perp}(x) \quad \text{implies} \quad (\varphi_t^* g)(u, v) = h g(u, v) . \tag{4.1}$$

This is equivalent to the existence of a function h_t and a 1-form ξ_t such that

$$\varphi_t^* g = h_t g + \xi_t \otimes \kappa + \kappa \otimes \xi_t . \tag{4.2}$$

By differentiation, this implies

$$\mathcal{L}_k g = 2ag + \xi \otimes \kappa + \kappa \otimes \xi , \tag{4.3}$$

where a is a function and ξ is a 1-form on M. Introducing the tensor G with components

$$G_{abcd} = k_{[a} g_{b][c} k_{d]} , \tag{4.4}$$

one can replace (4.3) by the equivalent condition

$$\mathcal{L}_k G = 2AG , \tag{4.5}$$

where A is another function on M. It is straightforward to prove:

Theorem 2: The flow on $(M, [g])$ generated by k is conformally geodesic and shear-free if, and only if, $\kappa \wedge \mathcal{L}_k \kappa = 0$ and condition (4.3) or (4.5) holds.

Let g be now chosen in the way described under condition (iii) of Theorem 1. This implies

$$\mathcal{L}_k \kappa = 0 \quad \text{and} \quad A = a . \tag{4.6}$$

By contracting both sides of Eq. (4.3) with k one obtains

$$(2a + k \rfloor \xi) \kappa + (k \rfloor \kappa) \xi = 0 , \tag{4.7}$$

whereas taking the trace leads to

$$\operatorname{div} k = na + k \rfloor \xi , \tag{4.8}$$

where the divergence of k is computed with respect to the volume element defined by the Riemannian metric, $\mathcal{L}_k \text{ vol} = (\operatorname{div} k) \text{ vol}$. It follows from (4.7) and (4.8) that

$$a = m^{-1} \operatorname{div} k , \tag{4.9}$$

where m is $n - 1$ or $n - 2$ according as the flow is nonnull or null. In the first case, ξ is proportional to κ, and (4.3) reduces to $\mathcal{L}_k g = 2a(g - \kappa \otimes \kappa / k \rfloor \kappa)$; in the second case, they need not be proportional, and $k \rfloor \xi = -2m^{-1} \operatorname{div} k$. In these formulae, m is the number of dimensions of the relevant vector spaces whose conformal structure is preserved by the flow: for $\epsilon \neq 0$, they are the orthogonal spaces $K^{\perp}(x)$; for $\epsilon = 0$, the screen spaces $K^{\perp}(x)/K(x)$.

Remark: A flow of conformal automorphisms, characterized by $\mathcal{L}_k g = 2ag$, is geodesic and shear-free, but the converse is not true.

V. EXPANSION AND VORTICITY

Together with a, the scalar of *expansion*, one can consider

$$b^2 = m^{-1} g^{ab} g^{cd} k_{[a;c]} k_{[b;d]} \tag{5.1}$$

as a measure of the *vorticity* of the congruence associated with the flow. Taking account of (4.6), one obtains

$$k_{[a} k_{b];[c} k_{d]} + k_{[a} k_{c];[b} k_{d]} = 2a k_{[a} g_{b][c} k_{d]} \tag{5.2}$$

from (4.5), and hence

$$g^{ab} g^{cd} k_{(a;c)} k_{(b;d)} = ma^2 , \tag{5.3}$$

$$g^{ab} g^{cd} g^{ef} k_{(a;c)} k_{[b;e]} k_{[d;f]} = mab^2 . \tag{5.4}$$

The Ricci identity for the commutator of second covariant derivatives gives

$$k_{a;bc} k^c + k_{a;c} k^c{}_{;b} = R_{cabd} k^c k^d . \tag{5.5}$$

Using (5.3) and (5.4) to simplify the contractions of (5.5) with g^{ab} and $k^{[a;b]}$, respectively, we obtain the real and imaginary parts of

$$\mathcal{L}_k (a + ib) + (a + ib)^2 = m^{-1} R_{cd} k^c k^d . \tag{5.6}$$

This propagation equation for the complex scalar of expansion and vorticity was first obtained by Sachs[16] for the case $n = 4$, $\epsilon = 0$. Raychaudhuri[18] had previously investigated the case $n = 4$, $\epsilon \neq 0$.

VI. CONFORMAL INVARIANCE

It is possible to replace g by $\hat{g}$, the conformally invariant density of (2.10), in all results which do not refer explicitly to a preferred metric. Writing

$$\hat{\kappa} = \hat{k}_a dx^a = \hat{g}(k) , \tag{6.1}$$

for example, we may replace (3.4) by

$$\hat{\kappa} \wedge \mathcal{L}_k \hat{\varkappa} = 0 . \tag{6.2}$$

This shows that the flow generated by k is conformally geodesic if, and only if, there is a scalar a such that

$$\mathcal{L}_k \hat{\kappa} = -2mn^{-1} a \hat{\kappa} . \tag{6.3}$$

The coefficient has been chosen to make this equation compatible with (4.9). From (4.5) and (4.6), writing

$$\hat{G}_{abcd} = \hat{k}_{[a} \hat{g}_{b][c} \hat{k}_{d]} , \tag{6.4}$$

we derive

$$\mathcal{L}_k \hat{G} = 2(n-m) n^{-1} a \hat{G} \tag{6.5}$$

as a necessary and sufficient condition for the flow generated by k to be geodesic and shear-free. The vorticity of the flow is described by a 3-form density of weight $4/n$,

$$\Omega = (1/3!) \Omega_{abc} dx^a \wedge dx^b \wedge dx^c = \kappa \wedge d\kappa . \tag{6.6}$$

For a geodesic flow,

$$k \lrcorner \hat{\kappa} b^2 = \tfrac{1}{12} m^{-1} \hat{g}^{ps} \hat{g}^{qt} \hat{g}^{ru} \Omega_{pqr} \Omega_{stu} , \tag{6.7}$$

and, if $\epsilon = 0$,

$$\hat{\kappa}_a \hat{\kappa}_b b^2 = \tfrac{1}{4} m^{-1} g^{at} g^{ru} \Omega_{aqr} \Omega_{btu} . \tag{6.8}$$

In the null case, (6.2) is equivalent to

$$k \lrcorner \Omega = 0 . \tag{6.9}$$

Thus, a null flow is geodesic if and only if it is orthogonal to its vorticity 3-form. In three dimensions, this is equivalent to the vanishing of the vorticity. For $n > 3$, the condition (6.9) for a null flow to be geodesic is equivalent to the existence of an (n − 4)-form density ω of weight $1 - 4/n$, such that

$$\Omega = k \lrcorner {}^*\omega . \tag{6.10}$$

In four dimensions, $\omega = \pm 2b$. For $n > 4$, ω is determined modulo the exterior product of $\hat{\kappa}$ with an arbitrary $(n-5)$-form density.

VII. A SPECIAL CASE: NULL FLOWS IN CONFORMAL SPACETIME

The conformal structure of an oriented four-dimensional space of Lorentz signature induces a natural complex structure on the screen spaces associated with a null vector field k. Indeed, in this case $S(x) = K^{\perp}(x)/K(x)$ is an oriented plane with a conformal structure. Let J be a rotation in that plane through 90°. Clearly $J^2 = -I$, and J defines a complex structure on $S(x)$. The complexified space $\mathbb{C} \otimes S(x)$ may be represented as a direct sum $S^+(x) \oplus S^-(x)$, where

$$S^{\pm}(x) = \{ u \in \mathbb{C} \otimes S(x) : Ju = \pm iu \} . \tag{7.1}$$

Let $K^{\pm}(x)$ be the subspace of $\mathbb{C} \otimes K^{\perp}(x)$ projecting onto $S^{\pm}(x)$ by the canonical map $\mathbb{C} \otimes K^{\perp}(x) \to \mathbb{C} \otimes S(x)$. Clearly,

$$K^+(x) \cap K^-(x) = \mathbb{C} \otimes K(x) , \tag{7.2}$$

$$K^+(x) + K^-(x) = \mathbb{C} \otimes K^{\perp}(x) . \tag{7.3}$$

Each of the spaces $K^+(x)$ and $K^-(x)$ is a maximal, totally null subspace of $\mathbb{C} \otimes T_x M$. Indeed, if $v \in K^+(x)$, say, then, since k is null, the scalar square of v is equal to that of its image $u \in S^+(x)$ under the projection $K^+(x) \to S^+(x)$. Since J is an isometry, the square of u is equal to that of $Ju = iu$. The latter square is opposite to that of u; thus the square of u is zero. Therefore, the scalar product of any two vectors belonging to $K^+(x)$ is zero. $K^+(x)$ is maximal because it is two-dimensional. We remark that one has the exact sequence of complex vector bundles over M

$$0 \to K^+ \to \mathbb{C} \otimes K^{\perp} \to S^- \to 0 . \tag{7.4}$$

The preceding definitions easily lead to:

Theorem 3: The flow generated by a null vector field k on a spacetime is geodesic and shear-free if and only if it preserves the complex vector bundle $K^+ \to M$.

Now consider a nonvanishing null vector field k on M. Its flow need not be geodesic. The subspaces $K^{\pm}(x) \subset \mathbb{C} \otimes K^{\perp}(x)$ may be characterized by means of suitably chosen complex 2-forms. Let $E = \frac{1}{2} E_{ab} dx^a \wedge dx^b \neq 0$ be a form subject to

$$*E = iE, \quad k \lrcorner E = 0 , \tag{7.5}$$

where the dual, in agreement with (2.11), is given by

$$*E_{ab} = \tfrac{1}{2} \hat{g}_{ac} \hat{g}_{bd} \epsilon^{cdef} E_{ef} . \tag{7.6}$$

The complex conjugate form $\overline{E}$ satisfies

$$*\overline{E} = -i\overline{E}, \quad k \lrcorner \overline{E} = 0 . \tag{7.7}$$

At each point $x \in M$, the form E is defined by (7.5) up to a complex factor. The kernel of the map $\mathbb{C} \otimes T_x M \to \mathbb{C} \otimes T_x^* M$ defined by

$$v \mapsto v \lrcorner E \tag{7.8}$$

contains $\mathbb{C} \otimes K(x)$. It is two-dimensional; and since

$$g(u, v) E = g(u) \wedge (v \lrcorner E) + v * \lrcorner (u \lrcorner * E) \tag{7.9}$$

identically, for any 2-form E and vectors u and v, it is totally null. The kernel of (7.8) must therefore coincide with one of the spaces $K^+(x)$ and $K^-(x)$. If E corresponds to K^+, then $\overline{E}$ corresponds to K^-. The total nullity of K^+ finds expression in the equation

$$E_{ap} g^{pq} E_{qb} = 0 . \tag{7.10}$$

For any $g \in [g]$, one may restrict E further by requiring that

$$E_{ap} \overline{E}^{p}{}_{b} = k_a k_b ; \tag{7.11}$$

E is then determined up to a phase factor, and (as a special case of the unscrambling identity[19])

$$4G = E \otimes \overline{E} + \overline{E} \otimes E . \tag{7.12}$$

Writing $D = \mathcal{L}_k E - AE$, one sees that Eq. (4.5) is equivalent to

$$D \otimes \overline{E} + \overline{E} \otimes D = -\overline{D} \otimes E - E \otimes \overline{D} , \tag{7.13}$$

and, therefore, to $D = iBE$, for some real scalar B. Thus the flow generated by k is geodesic and shear-free if and only if

$$\mathcal{L}_k E = eE \tag{7.14}$$

for some complex function e $(= A + iB)$ on M. This provides an alternative proof of Theorem 3. Since (4.3) is the general (real) solution of

$$E_a{}^p (\mathcal{L}_k g)_{pq} E^q{}_b = 0 , \tag{7.15}$$

this too is necessary and sufficient. One can also derive it from (7.14), by taking the Lie derivative of (7.10).

For any vector field ℓ, the 2-form $E_{ap} \nabla_{\ell} E^{p}{}_{b} dx^a \wedge dx^b$ is self-dual, orthogonal to k, and therefore proportional to E.

Similarly, the self-dual and anti-self-dual parts of $\kappa \wedge \nabla_{/} \kappa$ are proportional to E and $\overline{E}$, respectively. Thus there exist 1-forms θ and θ' such that[19]

$$E_{ap} E^{p}{}_{b;c} = E_{ab} \theta'_{c} , \tag{7.16}$$

$$k_b k_{a;c} - k_a k_{b;c} = \overline{E}_{ab} \theta_c - E_{ab} \bar{\theta}_c . \tag{7.17}$$

Contracting the last equation with k, and using (3.5), we have

$$\kappa \wedge \mathscr{L}_k \kappa + (k \lrcorner \theta) \overline{E} + (k \lrcorner \bar{\theta}) E = 0 . \tag{7.18}$$

From (7.5) and (2.7), $k \lrcorner (\theta \wedge E) = (k \lrcorner \theta) E$; from (2.12), writing

$$*\sigma = \theta \wedge E , \tag{7.19}$$

we obtain $*(\sigma \wedge \kappa) = k \lrcorner (\theta \wedge E)$. It thus follows from Theorem 1 that the flow generated by k is conformally geodesic if and only if there exists a scalar c such that

$$\sigma = c\kappa . \tag{7.20}$$

Using (7.5), (7.10), and (7.11) to simplify the contractions of $\overline{E}^{ra}$ with (7.16) and E^{ra} with (7.17), we obtain

$$- E_{ap;q} k^{p} = k_a \theta'_q , \quad E_{ap} k^{p}{}_{;q} = k_a \theta_q . \tag{7.21}$$

From the contraction of E^{qb} with the last equation,

$$E_a{}^{p} (\mathscr{L}_k g)_{pq} E^{q}{}_{b} = 2 k_{(a} \sigma_{b)} . \tag{7.22}$$

Thus $\sigma = 0$ is equivalent to (7.15), a necessary and sufficient condition for the flow to be geodesic and shear-free. From (7.21) and the covariant derivative of (7.5), however, $\theta = \theta'$: from the contraction of (7.16), therefore,

$$i\sigma = *(*E \wedge *dE) . \tag{7.23}$$

One deduces:

Theorem 4: Let k be a null vector field on a conformal spacetime M. The following conditions are equivalent:

(i) the flow generated by k is conformally geodesic and shear-free;

(ii) $\sigma = 0$;

(iii) $K^{+}(x)$ is surface forming, i.e., it admits two-dimensional integral manifolds;

(iv) $\mathscr{L}_k E$ is proportional to E;

(v) there exists a nonvanishing complex function f on M such that $d(fE) = 0$;

(vi) there exists a real nonvanishing 2-form F on M which satisfies Maxwell's equations for empty space, $dF = 0 = d*F$, and the conditions $k \lrcorner F = 0 = k \lrcorner *F$.

Proof: The equivalence of (i), (ii), and (iv) has been established. From the expression (7.23) for σ, the equivalence of (ii), (iii), and (v) follows by the theorem of Frobenius. One demonstrates the equivalence of (v) and (vi) by writing $F = fE + \overline{fE}$.

[1] H. Weyl, Gottingen Nachr., 99 (1921).
[2] J. Ehlers, F. A. E. Pirani, and A. Schild, in *General Relativity*, volume in honour of J. L. Synge, edited by L. O'Raifeartaigh (Clarendon, Oxford, 1972), pp. 63–84.
[3] E. Bessel-Hagen, Math. Ann. **84**, 258 (1921).
[4] R. Penrose, Proc. R. Soc. (London) A **284**, 159 (1965).
[5] M. Flato, J. Simon, and D. Sternheimer, Ann. Phys. (N.Y.) **61**, 78 (1970).
[6] C. N. Yang, Phys Rev. D **16**, 330 (1977).
[7] I. Robinson, J. Math. Phys. **2**, 290 (1961).
[8] J. N. Goldberg and R. K. Sachs, Acta Phys. Polon. **22** (Suppl.), 13 (1962).
[9] I. Robinson and A. Schild, J. Math. Phys. **4**, 484 (1963).
[10] I. Robinson and A. Trautman, Phys. Rev. Lett. **4**, 431 (1960); Proc. R. Soc. (London) A **265**, 463 (1962).
[11] R. P. Kerr, Phys. Rev. Lett. **11**, 237 (1963).
[12] For a recent review, see D. Kramer, H. Stephani, M. MacCallum, and E. Herlt, *Exact Solutions of Einstein's Field Equations* (Cambridge U. P. and Deutsche Verlag der Wissenschaften, Cambridge and Berlin, 1980).
[13] S. Kobayashi and K. Nomizu, *Foundations of Differential Geometry* (Interscience-Wiley, New York, 1963), Vol. I.
[14] R. Abraham and J. E. Marsden, *Foundations of Mechanics*, 2nd ed. (Addison-Wesley, Reading, Mass., 1978).
[15] Y. Choquet-Bruhat, C. DeWitt-Morette, and M. Dillard-Bleick, *Analysis, Manifolds and Physics* (North-Holland, Amsterdam, 1977).
[16] R. K. Sachs, "Distance and the Asymptotic Behaviour of Waves in General Relativity," in *Recent Developments in General Relativity*, dedicated to Leopold Infeld (Pergamon Press and PWN, New York and Warsaw, 1962).
[17] J. Ehlers and W. Kundt, "Exact Solutions of the Gravitational Field Equations," in *Gravitation: An Introduction to Current Research*, edited by L. Witten (Wiley, New York, 1962).
[18] A. Raychaudhuri, Phys. Rev. **98**, 1123 (1955).
[19] I. Robinson "Null Congruences and Plebański-Schild Spaces," in *Spacetime and Geometry: The Alfred Schild Lectures* (University of Texas Press, Austin, 1982).

Homogeneous quadratic invariants for one-dimensional classical homogeneous quadratic Hamiltonians: Some properties of the auxiliary equation

L. S. Wollenberg
Department of Mathematics, Queen Elizabeth College, University of London, London W8 7AH, England

(Received 20 November 1981; accepted for publication 21 May 1982)

Using a previously derived characterization of the homogeneous quadratic invariants for an arbitrary one-dimensional classical homogeneous quadratic Hamiltonian we obtain various "nonlinear superposition laws." We establish a close connection between one of these (the quadratic superposition law) and the original characterization.

PACS numbers: 03.20. + i, 02.30.Hq

I. INTRODUCTION

We recently[1] characterized the homogeneous quadratic (in $q, \dot{q}$) invariants $J = J(q,\dot{q},t)$ for an arbitrary second-order ordinary differential equation

$$\ddot{q} + N\dot{q} + Kq = 0, \tag{1}$$

where N, K are (possibly complex) functions of t, as being of form

$$2J = c(\sigma^{-1}q)^2 + M^2(\sigma\dot{q} - \dot{\sigma}q)^2, = 2J_{\sigma,c} \quad \text{say,} \tag{2a}$$

for some[2] σ,c satisfying

$$\ddot{\sigma} + N\dot{\sigma} + K\sigma = c/M^2\sigma^3, \tag{2b}$$

where $\dot{M} = NM$ (the arbitrary multiplicative constant in M being chosen as is convenient).

We then used this to characterize the homogeneous quadratic (in q, p) invariants $I = I(q, p,t)$ for an arbitrary[3] classical homogeneous quadratic Hamiltonian

$$H = \tfrac{1}{2}(a^2p^2 + 2bqp + e^2q^2) \qquad (a \not\equiv 0), \tag{3}$$

where a,b,e are (possibly complex) functions of t, as being of form

$$2I = c(\sigma^{-1}q)^2 + [\sigma p - ((\dot{\sigma} - b\sigma)/a^2)q]^2, = 2I_{\sigma,c} \quad \text{say,} \tag{4a}$$

for some σ,c satisfying

$$\ddot{\sigma} - (2\dot{a}/a)\dot{\sigma} + K\sigma = ca^4/\sigma^3, \tag{4b}$$

where $K = (a^2e^2 - b^2) + 2\dot{a}b/a - \dot{b}$. In fact, since (see Ref. 1)

$$\det(2I_{\sigma,c}) = c, \tag{4c}$$

we have

$$c = \det(2I_{\sigma,c}) = \det(2I). \tag{4d}$$

[For other derivations of (4a) and (4b) see Ref. 4, and Theorem 7(ii), (iv) below.] We note that (3) leads to the q-equation

$$\ddot{q} + (2\dot{a}/a)\dot{q} + Kq = 0, \tag{5}$$

where K is as in (4b).

In particular, by taking (for *arbitrary* N, K)

$$a = \exp\left[-\frac{1}{2}\int_{t_0}^{t} N(t')\,dt'\right] \quad (\Rightarrow -2\dot{a}/a = N), \tag{6a}$$

$$b = 0, \tag{6b}$$

$$e^2 = K\exp\left[\int_{t_0}^{t} N(t')\,dt'\right] \quad (\Rightarrow a^2e^2 = K), \tag{6c}$$

and (since then $\dot{q}, = \partial H/\partial p, = a^2p$)

$$p = \dot{q}/a^2 = \dot{q}\exp\left[\int_{t_0}^{t} N(t')\,dt'\right] \tag{6d}$$

we recover (2a) and (2b) [with the particular choice $M = \exp[\int_{t_0}^{t} N(t')\,dt']$]. We will call Eq. (4b) the *auxiliary equation* (*for* H). (Note that in the literature this term is usually reserved for the $c = 1$ case.)

In classical mechanics "new" homogeneous quadratic (in q,p) invariants can be derived from "old" ones by addition, by scalar multiplication, and (see Ref. 5) by Poisson's theorem. (We note that Poisson's theorem holds for time-dependent invariants as well as for time-independent ones.)

In Sec. II of this paper we use the characterization (4a) and (4b) to examine the effect (Theorem 1) of these manipulations on the auxiliary equation, and pay particular attention to certain special cases (Theorem 2). In Sec. III we combine two of these special cases, to get what we call the *quadratic superposition law* (Theorem 3_s). We then prove a "converse" result (Theorem 3_n) and, in the course of further discussion, give $J_{\sigma,c}$ a physical interpretation. Finally, in Sec. IV, after some preliminary results (Theorem 5 and 6), we establish a close connection (Theorem 7) between Theorems 3_s and 3_n, and the characterization (4a), (4b).

Remark 1: It is worth pointing out how the invariant $J_{\sigma,c}$ can be "n-dimensionalized." Suppose σ,c satisfy (2b) and q_k ($k = 1,2,\ldots,n$) satisfy (1). Then, by linearity, for all constants[6] $k_1,k_2,\ldots,k_n$,

$$q = \sum_{r=1}^{n} k_r q_r$$

satisfies (1).

Hence, by (2a), for all constants $k_1,k_2,\ldots,k_n$,

$$\begin{aligned} 2J_{\sigma,c} &= c\left(\sigma^{-1}\sum_{r=1}^{n} k_r q_r\right)^2 + M^2\left[\sigma\left(\sum_{r=1}^{n} k_r\dot{q}_r\right) - \dot{\sigma}\left(\sum_{r=1}^{n} k_r q_r\right)\right]^2 \\ &= 2\left(\sum_{r=1}^{n} k_r^2 J_{r,r} + 2\sum_{\substack{r,s=1 \\ r\neq s}}^{n} k_r k_s J_{r,s}\right), \end{aligned}$$

 0022-2488/83/061430-09$02.50 © 1983 American Institute of Physics

where

$$2J_{r,r} = c(\sigma^{-1}q_r)^2 + M^2(\sigma\dot{q}_r - \dot{\sigma}q_r)^2$$

and

$$2J_{r,s} = c(\sigma^{-1}q_r)(\sigma^{-1}q_s) + M^2(\sigma\dot{q}_r - \dot{\sigma}q_r)(\sigma\dot{q}_s - \dot{\sigma}q_s),$$

is invariant.

We know already, from (2a), that each $J_{r,r}$ is invariant, but by choosing, say,

$$k_r = k_s = 1, \quad k_i = 0 \quad (i \neq r,s),$$

we find $J_{r,r} + J_{s,s} + 2J_{r,s}$ is invariant, $\Rightarrow J_{r,s}$ is invariant.

Thus, if σ,c satisfy (2b), then each $J_{r,s}$ $(r,s = 1,2,\ldots,n)$ is invariant. When applied to the time-dependent harmonic oscillator this reduces, when $M = 1$ and $c = 1$, to the Günther–Leach tensor.[7] So we see that the linearity of (1) can be used to motivate, and indeed to construct, the $J_{r,s}$ (and in particular the Günther–Leach tensor). For another derivation of the Günther–Leach tensor, see Ref. 8.

Remark 2: We reported in Ref. 1 a characterization of the homogeneous *linear* (in $q,\dot{q}$) invariants for (1). We hope to discuss the implications of this result in detail elsewhere (see Ref. 4).

II. "NEW" INVARIANTS FROM "OLD" ONES

We will say that σ satisfies (or σ, c satisfy) (σ,c) if (4b) holds. Thus q satisfies $(q,0)$.

It can be verified that if σ satisfies (σ,c), then any twice-differentiable solution $\bar{\sigma}$ of $\bar{\sigma}^2 = \sigma^2$ satisfies $(\bar{\sigma},c)$ and all such $\bar{\sigma},c$ lead to the same invariant, i.e.,

$$\bar{\sigma}^2 = \sigma^2 \Rightarrow I_{\bar{\sigma},c} = I_{\sigma,c}. \tag{7a}$$

Conversely, if $I_{\bar{\sigma},\bar{c}} = I_{\sigma,c}$, then comparing coefficients of p^2 leads to

$$\bar{\sigma}^2 = \sigma^2 \tag{7b}$$

and taking determinants leads to

$$\det(2I_{\bar{\sigma},\bar{c}}) = \det(2I_{\sigma,c}),$$

i.e., [see (4c)],

$$\bar{c} = c. \tag{7c}$$

Let[9] σ_k, c_k $(k = 1,2)$ satisfy (σ_k,c_k) $(k = 1,2)$, and let $\alpha_k = (\dot{\sigma}_k - b\sigma_k)/a^2$ $(k = 1,2)$. Then (4a) tell us that

$$\begin{aligned} 2I_{\sigma_k,c_k} &= c_k(\sigma_k^{-1}q)^2 + (\sigma_k p - \alpha_k q)^2 \\ &= (c_k\sigma_k^{-2} + \alpha_k^2)q^2 - 2\sigma_k\alpha_k qp + \sigma_k^2 p^2 \quad (k = 1,2) \end{aligned}$$

are invariants for H.

A. Use of addition

Let k_1 and k_2 be constants. Then clearly $k_1 I_{\sigma_1,c_1} + k_2 I_{\sigma_2,c_2}$ is homogeneous quadratic (in q,p), and invariant for H, hence [by (4a) and (4b)] of form I_{σ_3,c_3} for some σ_3, c_3 satisfying (σ_3,c_3). Thus

$$2(k_1 I_{\sigma_1,c_1} + k_2 I_{\sigma_2,c_2}) = 2I_{\sigma_3,c_3},$$

i.e.,

$$\begin{aligned} &[k_1(c_1\sigma_1^{-2} + \alpha_1^2) + k_2(c_2\sigma_2^{-2} + \alpha_2^2)]q^2 \\ &\quad - 2(k_1\sigma_1\alpha_1 + k_2\sigma_2\alpha_2)qp \\ &\quad + (k_1\sigma_1^2 + k_2\sigma_2^2)p^2 = (c_3\sigma_3^{-2} + \alpha_3^2)q^2 \\ &\quad - 2\sigma_3\alpha_3 qp + \sigma_3^2 p^2, \end{aligned}$$

where

$$\alpha_3 = (\dot{\sigma}_3 - b\sigma_3)/a^2.$$

In particular, comparing coefficients of p^2, q^2, respectively, gives

$$\sigma_3^2 = k_1\sigma_1^2 + k_2\sigma_2^2, \tag{8}$$

$$c_3\sigma_3^{-2} + \alpha_3^2 = (k_1c_1\sigma_1^{-2} + k_2c_2\sigma_2^{-2}) + (k_1\alpha_1^2 + k_2\alpha_2^2). \tag{9}$$

σ_3 is obtained by taking [cf. (7a)] any twice-differentiable solution of (8), and c_3 is obtained as follows. First note that (substituting for α_1, α_2, α_3) (9) becomes

$$\begin{aligned} &c_3\sigma_3^{-2} + (1/a^4)(\dot{\sigma}_3^2 - 2b\sigma_3\dot{\sigma}_3 + b^2\sigma_3^2) \\ &\quad = (k_1c_1\sigma_1^{-2} + k_2c_2\sigma_2^{-2}) + (1/a^4)[(k_1\dot{\sigma}_1^2 + k_2\dot{\sigma}_2^2) \\ &\quad - 2b(k_1\sigma_1\dot{\sigma}_1 + k_2\sigma_2\dot{\sigma}_2) + b^2(k_1\sigma_1^2 + k_2\sigma_2^2)]. \end{aligned} \tag{9'}$$

But (8)$\Rightarrow$

$$\sigma_3\dot{\sigma}_3 = k_1\sigma_1\dot{\sigma}_1 + k_2\sigma_2\dot{\sigma}_2. \tag{8'}$$

Thus, substituting (8) and (8') into (9'), we obtain

$$\begin{aligned} c_3\sigma_3^{-2} + \dot{\sigma}_3^2/a^4 &= (k_1c_1\sigma_1^{-2} + k_2c_2\sigma_2^{-2}) \\ &\quad + (1/a^4)(k_1\dot{\sigma}_1^2 + k_2\dot{\sigma}_2^2), \end{aligned} \tag{9''}$$

i.e.,

$$\begin{aligned} c_3 &= \sigma_3^2(k_1c_1\sigma_1^{-2} + k_2c_2\sigma_2^{-2}) \\ &\quad + (1/a^4)[\sigma_3^2(k_1\dot{\sigma}_1^2 + k_2\dot{\sigma}_2^2) - \sigma_3^2\dot{\sigma}_3^2], \end{aligned}$$

which, when we again use (8) and (8'), simplifies to

$$c_3 = k_1^2 c_1 + k_2^2 c_2 + k_1k_2 I(\sigma_1,\sigma_2), \tag{9'''}$$

where

$$I(\sigma_1,\sigma_2) \quad (\text{or} \quad I_{12}) = c_2\left(\frac{\sigma_1}{\sigma_2}\right)^2 + c_1\left(\frac{\sigma_2}{\sigma_1}\right)^2 + \left(\frac{W(\sigma_1,\sigma_2)}{a^2}\right)^2$$

and

$$W(\sigma_1,\sigma_2) \quad (\text{or} \quad W_{12}) = \sigma_1\dot{\sigma}_2 = \dot{\sigma}_1\sigma_2.$$

Of course, because of the way it has arisen, $I(\sigma_1,\sigma_2)$ is constant. This can also be verified directly. In fact, it is, modulo notation, a special case of (17a), (17b), (17c) in Ref. 1. By inspection of $I(\sigma_1,\sigma_2)$ we can see that the quantity $W(\sigma_1,\sigma_2)/a^2$ itself is not necessarily constant, although certainly it is if $c_1 = c_2 = 0$. (Note that in this case either $W_{12} = 0$ for all t or $W_{12} \neq 0$ for all t.)

B. Use of scalar multiplication

Let σ,c satisfy (σ,c). If we apply the results of Sec. IIA with $\sigma_1 = \sigma$, $c_1 = c$, $k_1 = \lambda$, $k_2 = 0$ (corresponding to considering $\lambda I_{\sigma,c}$), we find that $\bar{\sigma}$ given by any twice-differentiable solution of $\bar{\sigma}^2 = \lambda\sigma^2$ satisfies $(\bar{\sigma},\bar{c})$ with $\bar{c} = \lambda^2 c$.

C. Use of Poisson's theorem

Consider the Poisson bracket $\{I_{\sigma_1,c_1},I_{\sigma_2,c_2}\}$. This is homogeneous quadratic (in q, p), and (by Poisson's theorem) invariant for H, hence [by (4a) and (4b)] of form I_{σ_4,c_4} for some σ_4,c_4 satisfying (σ_4,c_4). Thus

$$2\{I_{\sigma_1,c_1},I_{\sigma_2,c_2}\} = 2I_{\sigma_4,c_4},$$

i.e.,

$$\begin{aligned}&2[\sigma_1\alpha_1(c_2\sigma_2^{-2}+\alpha_2^2)-(c_1\sigma_1^{-2}+\alpha_1^2)\sigma_2\alpha_2]q^2\\&\quad-2[\sigma_1^2(c_2\sigma_2^{-2}+\alpha_2^2)-(c_1\sigma_1^{-2}+\alpha_1^2)\sigma_2^2]qp\\&\quad+2(\sigma_1^2\sigma_2\alpha_2-\sigma_1\alpha_1\sigma_2^2)p^2\\&\quad=(c_4\sigma_4^{-2}+\alpha_4^2)q^2-2\sigma_4\alpha_4qp+\sigma_4^2p^2,\end{aligned}$$

where

$$\alpha_4 = (\dot\sigma_4 - b\sigma_4)/a^2.$$

In particular, comparing coefficients of p^2, qp, q^2, respectively, gives

$$\sigma_4^2 = 2\sigma_1\sigma_2(\sigma_1\alpha_2-\sigma_2\alpha_1), \tag{10}$$

$$\sigma_4\alpha_4 = (c_2\sigma_1^2\sigma_2^{-2}-c_1\sigma_2^2\sigma_1^{-2})+(\sigma_1^2\alpha_2^2-\alpha_1^2\sigma_2^2), \tag{11}$$

$$\begin{aligned}c_4\sigma_4^{-2}+\alpha_4^2 &= 2(c_2\sigma_1\alpha_1\sigma_2^{-2}-c_1\sigma_2\alpha_2\sigma_1^{-2})\\&\quad+2(\sigma_1\alpha_2-\alpha_1\sigma_2)\alpha_1\alpha_2.\end{aligned} \tag{12}$$

σ_4 is obtained by taking [cf. (7a)] any twice-differentiable solution of (10), and c_4 is obtained as follows. First note that (substituting for α_1,α_2) (10) becomes

$$\sigma_4^2 = (2\sigma_1\sigma_2/a^2)(\sigma_1\dot\sigma_2-\dot\sigma_1\sigma_2) \tag{10'}$$

$$= 2(W_{12}/a^2)\sigma_1\sigma_2. \tag{10''}$$

Now, from (12),

$$\begin{aligned}c_4 &= 2\sigma_4^2(c_2\sigma_1\alpha_1\sigma_2^{-2}-c_1\sigma_2\alpha_2\sigma_1^{-2})\\&\quad+2\sigma_4^2(\sigma_1\alpha_2-\alpha_1\sigma_2)\alpha_1\alpha_2-\sigma_4^2\alpha_4^2,\end{aligned} \tag{12'}$$

and, using (10) and (11) (and after considerable manipulation), we obtain

$$\begin{aligned}c_4 = &-2(\sigma_1\alpha_2-\alpha_1\sigma_2)^2(c_2\sigma_1^2\sigma_2^{-2}+c_1\sigma_2^2\sigma_1^{-2})\\&-(\sigma_1\alpha_2-\alpha_1\sigma_2)^4-(c_2\sigma_1^2\sigma_2^{-2}-c_1\sigma_2^2\sigma_1^{-2}).\end{aligned} \tag{12''}$$

Substituting for α_1,α_2, this becomes

$$\begin{aligned}c_4 = &-2(W_{12}/a^2)^2(c_2\sigma_1^2\sigma_2^{-2}+c_1\sigma_2^2\sigma_1^{-2})-(W_{12}/a^2)^2\\&-(c_2\sigma_1^2\sigma_2^{-2}-c_1\sigma_2^2\sigma_1^{-2})^2.\end{aligned}$$

A straightforward calculation shows that
$c_4 + I_{12}^2 = 4c_1c_2 \Rightarrow$

$$c_4 = 4c_1c_2 - I_{12}^2. \tag{12'''}$$

Collecting together our results so far, we see that we have proved:

Theorem 1: Let σ_k,c_k $(k=1,2)$ satisfy (σ_k,c_k) $(k=1,2)$. Let $W_{12}=\sigma_1\dot\sigma_2-\dot\sigma_1\sigma_2$. Then $I_{12}=c_2(\sigma_1/\sigma_2)^2$ $+c_1(\sigma_2/\sigma_1)^2+(W_{12}/a^2)^2$ is constant. Also:

(i) For all constants k_1,k_2 the function σ_3 given by any twice-differentiable solution of $\sigma_3^2=k_1\sigma_1^2+k_2\sigma_2^2$ satisfies (σ_3,c_3) with $c_3=k_1^2c_1+k_2^2c_2+k_1k_2I_{12}$.

(ii) For all constants λ the function $\bar\sigma_k$ given by any twice-differentiable solution of $\bar\sigma_k^2=\lambda\sigma_k^2$ satisfies $(\bar\sigma_k,\bar c_k)$ with $\bar c_k=\lambda^2c_k$.

(iii) The function σ_4 given by any twice-differentiable solution of $\sigma_4^2=2(W_{12}/a^2)\sigma_1\sigma_2$ satisfies (σ_4,c_4) with $c_4=4c_1c_2-I_{12}^2$.

D. Special case $c_1 = c_2 = 0$

If $c_1=c_2=0$, then the formulae for c_3,c_4 simplify, for in this case

$$I_{12} = (W_{12}/a^2)^2,$$

leading to

$$c_3 = k_1k_2(W_{12}/a^2), \quad c_4 = -(W_{12}/a^2)^4.$$

In particular, if $W_{12}\neq0$, we can "rescale" σ_4^2, as follows: Set $\lambda=k_3(W_{12}/a^2)^{-1}$ in Theorem 1(ii). Then we find that $\bar\sigma_4$ given by any twice-differentiable solution of

$$\bar\sigma_4^2 = \lambda\sigma_4^2 = 2k_3\sigma_1\sigma_2$$

satisfies $(\bar\sigma_4,\bar c_4)$ with $\bar c_4=\lambda^2c_4=-k_3^2(W_{12}/a^2)^2$.

This result holds *also* if $W_{12}=0$, for then (see Refs. 10 and 11) either $\sigma_2\propto\sigma_1$ or $\sigma_1\propto\sigma_2$; hence $\bar\sigma_4\propto\sigma_1$ or $\bar\sigma_4\propto\sigma_2$, so, of course, $\bar c_4=0$.

It is convenient to collect together these special cases. We have proved:

Theorem 2: Let σ_k $(k=1,2)$ satisfy $(\sigma_k,0)$ $(k=1,2)$. Let $W_{12}=\sigma_1\dot\sigma_2-\dot\sigma_1\sigma_2$. Then (W_{12}/a^2) is constant. Also:

(i) For all constants k_1,k_2 the function σ_3 given by any twice-differentiable solution of $\sigma_3^2=k_1\sigma_1^2+k\sigma_2^2$ satisfies (σ_3,c_3) with $c_3=k_1k_2(W_{12}/a^2)^2$.

(ii) The function σ_4 given by any twice-differentiable solution of $\sigma_4^2=2(W_{12}/a^2)\sigma_1\sigma_2$ satisfies (σ_4,c_4) with $c_4=-(W_{12}/a^2)^4$.

(iii) For all constants k_3 the function $\bar\sigma_4$ given by any twice-differentiable solution of $\bar\sigma_4^2=2k_3\sigma_1\sigma_2$ satisfies $(\bar\sigma_4,\bar c_4)$ with $\bar c_4=-k_3^2(W_{12}/a^2)^2$.

We give two further proofs of Theorem 2(iii). The first uses just Theorem 2(i); the second uses Theorem 2(i) and Theorem 1(i).

Proof No. 1: Write

$$\begin{aligned}\bar\sigma_4^2 &= 2k_3\left[\left(\frac{\sigma_1+\sigma_2}{2}\right)^2-\left(\frac{\sigma_1-\sigma_2}{2}\right)^2\right]\\&= k_1'\sigma_1'^2+k_2'\sigma_2'^2,\end{aligned}$$

where

$$\sigma_1' = \tfrac12(\sigma_1+\sigma_2), \quad \sigma_2' = \tfrac12(\sigma_1-\sigma_2),$$

and

$$k_1' = 2k_3, \quad k_2' = -2k_3.$$

σ_k' $(k=1,2)$ satisfy $(\sigma_k',0)$ $(k=1,2)$; hence, by Theorem 2(i), $\bar\sigma_4$ satisfies $(\bar\sigma_4,\bar c_4)$ with

$$\bar c_4 = k_1'k_2'(W(\sigma_1',\sigma_2')/a^2)^2 = -4k_3^2(W(\sigma_1',\sigma_2')/a^2)^2.$$

A straightforward calculation shows that
$W(\sigma_1',\sigma_2') = -\frac12W_{12}$; thus

$$\bar c_4 = -4k_3^2(-W_{12}/2a^2)^2 = -k_3^2(W_{12}/a^2)^2,$$

giving Theorem 2(iii).

Proof No. 2: Write

$$\begin{aligned}\bar\sigma_4^2 &= k_3[(\sigma_1+\sigma_2)^2-(\sigma_1^2+\sigma_2^2)]\\&= k_1'\sigma_1'^2+k_2'\sigma_2'^2,\end{aligned}$$

where

$$\sigma_1' = \sigma_1 + \sigma_2, \quad \sigma_2'^2 = \sigma_1^2 + \sigma_2^2$$

and

$$k_1' = k_3, \quad k_2' = -k_3.$$

σ_1' satisfies $(\sigma_1',0)$ and, by Theorem 2(i), σ_2' satisfies (σ_2',c_2') with $c_2' = 1\cdot 1\cdot(W_{12}/a^2)^2$. Hence, by Theorem 1(i), $\bar{\sigma}_4$ satisfies $(\bar{\sigma}_4,\bar{c}_4)$ with

$$\begin{aligned}\bar{c}_4 &= k_1'^2 c_1' + k_2'^2 c_2' + k_1' k_2' I(\sigma_1',\sigma_2') \\ &= k_3^2 \cdot 0 + k_3^2 (W_{12}/a^2)^2 - k_3^2 I(\sigma_1',\sigma_2').\end{aligned}$$

A straightforward, but laborious, calculation shows that $I(\sigma_1',\sigma_2') = 2(W_{12}/a^2)^2$; thus

$$\bar{c}_4 = k_3^2 (W_{12}/a^2)^2 - 2k_3^2 (W_{12}/a^2)^2 = -k_3^2 (W_{12}/a^2)^2,$$

giving Theorem 2(iii).

E. Direct verification

How difficult is it to verify Theorem 1(i)–(iii) directly? The author has managed 1(i) (the details being straightforward but laborious), 1(ii) (easy), but not 1(iii). Thus, to verify 1(iii), we need a new method.

A clue to what to do comes from noting that (as is easily verified)

$$\left(\frac{\sigma_1^2}{a^2}\right)\left(\frac{\sigma_2^2}{a^2}\right)^{\cdot} - \left(\frac{\sigma_1^2}{a^2}\right)^{\cdot}\left(\frac{\sigma_2^2}{a^2}\right) = \frac{2\sigma_1\sigma_2}{a^4}(\sigma_1\dot{\sigma}_2 - \dot{\sigma}_1\sigma_2) \tag{13}$$

and hence that [cf. (8) and (10′)]

$$\frac{\sigma_3^2}{a^2} = k_1\left(\frac{\sigma_1^2}{a^2}\right) + k_2\left(\frac{\sigma_2^2}{a^2}\right), \tag{14a}$$

$$\frac{\sigma_4^2}{a^2} = \left(\frac{\sigma_1^2}{a^2}\right)\left(\frac{\sigma_2^2}{a^2}\right)^{\cdot} - \left(\frac{\sigma_1^2}{a^2}\right)^{\cdot}\left(\frac{\sigma_2^2}{a^2}\right). \tag{14b}$$

Equations (14a) and (14b) suggest that we work with the variable $\theta = \sigma^2/a^2$ instead of σ. In θ terms the auxiliary equation becomes

$$\theta\ddot{\theta} - \tfrac{1}{2}\dot{\theta}^2 + \mu\theta^2 = 2c, \tag{15}$$

where $\mu = 2(K + \ddot{a}/a - 2\dot{a}^2/a^2)$. We will say that θ satisfies (or θ, c satisfy) (θ,c) if (15) holds. Notice that, although (15) is not linear, (15˙) *is*, for we find $(d/dt)(15) \Rightarrow$

$$\theta\dddot{\theta} + \dot{\mu}\theta^2 + 2\mu\theta\dot{\theta} = 0,$$

i.e., (assuming $\theta \not\equiv 0$),

$$\ddot{\theta} + 2\mu\dot{\theta} + \dot{\mu}\theta = 0. \tag{15˙}$$

Conversely, (15˙) $\Rightarrow \theta$ satisifies (θ,c) for *some* c. (Despite this, putting $I_{\sigma,c}$ in θ terms does not seem to be of any use).

In θ terms the problem of verifying Theorem 1(i)–(iii) directly becomes: Let θ_k, c_k $(k = 1,2)$ satisfy (θ_k,c_k) $(k = 1,2)$. Prove that:

(i)′ For all constants k_1,k_2 the function θ_3 given by [cf. (14a)] $\theta_3 = k_1\theta_1 + k_2\theta_2$ satisfies (θ_3,c_3) with $c_3 = k_1^2 c_1 + k_2^2 c_2 + k_1 k_2 I_{12}$.

(ii)′ For all constants λ the function $\bar{\theta}_k$ given by $\bar{\theta}_k = \lambda\theta_k$ satisfies $(\bar{\theta}_k,\bar{c}_k)$ with $\bar{c}_k = \lambda^2 c_k$.

(iii)′ The function θ_4 given by [cf. (14b)] $\theta_4 = \theta_1\dot{\theta}_2 - \dot{\theta}_1\theta_2$ satisfies (θ_4,c_4) with $c_4 = 4c_1c_2 - I_{12}^2$. The author has managed (i)′ [about as straightforward and laborious as Theorem 1(i)], (ii)′ (easy), and (iii)′ (difficult).

That θ_3 satisfies (θ_3,c_3) for *some* c_3 can be obtained immediately from the linearity of (15˙), for since θ_1,θ_2 (assumed $\not\equiv 0$, for otherwise $\theta_4 \equiv 0$) satisfy (θ_k,c_k) $(k = 1,2)$, they satisfy (15˙). Hence, by linearity, θ_3 satisfies (15˙). Thus θ_3 satisfies (θ_3,c_3) for some c_3.

III. QUADRATIC SUPERPOSITION LAW

The following result essentially combines parts (i) and (iii) of Theorem 2.

Theorem 3_s: Let σ_1,σ_2, W_{12} be as in Theorem 2. Then for all constants k_1, k_2, k_3 the function σ given by any twice-differentiable solution of

$$\sigma^2 = k_1\sigma_1^2 + k_2\sigma_2^2 + 2k_3\sigma_1\sigma_2 \tag{16}$$

satisfies (σ,c) with

$$c = (k_1k_2 - k_3^2)(W_{12}/a^2)^2. \tag{17}$$

We give two proofs [cf. the two further proofs of Theorem 2(iii)].

Proof No. 1: If $k_1 = k_2 = 0$, then the result follows immediately from Theorem 2(iii). So suppose k_1,k_2 are *not* both zero, say $k_1 \neq 0$, and write

$$\begin{aligned}\sigma^2 &= k_1[\sigma_1 + (k_3/k_1)\sigma_2]^2 + [(k_1k_2 - k_3^2)/k_1]\sigma_2^2 \\ &= k_1'\sigma_2'^2 + k_2'\sigma_2'^2,\end{aligned}$$

where

$$\sigma_1' = \sigma_1 + (k_3/k_1)\sigma_2, \quad \sigma_2' = \sigma_2$$

and

$$k_1' = k_1, \quad k_2' = (k_1k_2 - k_3^2)/k_1.$$

σ_k' $(k = 1,2)$ satisfy $(\sigma_k',0)$ $(k = 1,2)$; hence, by Theorem 2(i), σ satisfies (σ,c) with

$$c = k_1'k_2'[W(\sigma_1',\sigma_2'/a^2]^2 = (k_1k_2 - k_3^2[W(\sigma_1',\sigma_2')/a^2]^2.$$

But $W(\sigma_1',\sigma_2') = W[\sigma_1 + (k_3/k_1)\sigma_2,\sigma_2) = W(\sigma_1,\sigma_2)$; thus Theorem 3_s is proved.

Proof No. 2: If $k_3 = 0$, then the result follows immediately from Theorem 2(i). So suppose $k_3 \neq 0$, and write

$$\begin{aligned}\sigma^2 &= (k_1\sigma_1^2 + k_2\sigma_2^2) + (2k_3\sigma_1\sigma_2) \\ &= k_1'\sigma_1'^2 + k_2'\sigma_2'^2,\end{aligned}$$

where

$$\sigma_1'^2 = k_1\sigma_1^2 + k_2\sigma_2^2, \quad \sigma_2'^2 = 2k_3\sigma_1\sigma_2$$

and

$$k_1' = 1, \quad k_2' = 1.$$

By Theorem 2(i), σ_1' satisfies (σ_1',c_1') with $c_1' = k_1k_2(W_{12}/a^2)^2$. By Theorem 2(iii), σ_2' satisfies (σ_2',c_2') with $c_2' = -k_3^2(W_{12}/a^2)^2$. Hence, by Theorem 1(i), σ satisfies (σ,c) with

$$\begin{aligned}c &= k_1'^2 c_1' + k_2'^2 c_2' + k_1' k_2' I(\sigma_1',\sigma_2') \\ &= (k_1k_2 - k_3^2)(W_{12}/a^2)^2 + I(\sigma_1',\sigma_2').\end{aligned}$$

A straightforward, but laborious, calculation shows that $I(\sigma_1',\sigma_2') = 0$; thus Theorem 3_s is proved.

The author has managed to verify Theorem 3_s directly in σ terms (the details being straightforward but laborious), but not in θ terms.

An extension of Theorem 3_s will be reported elsewhere (see Ref. 4). Theorems 1, 2, and 3_s are examples of what have been called (see Refs. 12 and 13 and references therein) "nonlinear superposition laws." In what follows we shall be interested mainly in Theorem 3_s, which we will call the *quadratic superposition law*.

Does this rule give *every* solution of (σ,c)? I.e., must *every* solution of (σ,c) be expressible in the form (16) for some k_1,k_2,k_3 satisfying (17)? Is the expression unique? We prove:

Theorem 3_n: Let σ_1, σ_2, W_{12} be as in Theorem 2, and suppose $W_{12}\neq 0$. Then for any c the general solution of (σ,c) can be put in the form (16), where k_1,k_2,k_3 are constants satisfying

$$k_1k_2 - k_3^2 = c(W_{12}/a^2)^{-2}, \tag{17'}$$

in one and only one way.

Proof: Let c be given, and let k_1,k_2,k_3 be constants satisfying (17'). By Theorem 3_s, any twice-differentiable solution σ of

$$\sigma^2 = k_1\sigma_1^2 + k_2\sigma_2^2 + 2k_3\sigma_1\sigma_2 \tag{16'}$$

is a solution of (σ,c). Furthermore, since σ contains two independent arbitrary constants, it is[14] a possible form for the general solution of (σ,c).

We prove that the expression (16') is unique. Suppose that

$$\sigma^2 = k_1'\sigma_1^2 + k_2'\sigma_2^2 + 2k_3'\sigma_1\sigma_2, \tag{16''}$$

where k_1',k_2',k_3' are constants satisfying

$$k_1'k_2' - k_3'^2 = c(W_{12}/a^2)^{-2}, \tag{17''}$$

is also a possible form for the general solution of (σ,c). From (16') and (16'')

$$\lambda_1\sigma_1^2 + \lambda_2\sigma_2^2 + 2\lambda_3\sigma_1\sigma_2 = 0, \tag{18}$$

where $\lambda_i = k_i - k_i'$ $(i = 1,2,3)$:

$$\frac{d}{dt}(18) \Rightarrow \lambda_1\sigma_1\dot\sigma_1 + \lambda_2\sigma_2\dot\sigma_2 + \lambda_3(\dot\sigma_1\sigma_2 + \sigma_1\dot\sigma_2) = 0. \tag{18'}$$

Write (18) and (18') as

$$\begin{pmatrix}\sigma_1 & \sigma_2\\ \dot\sigma_1 & \dot\sigma_2\end{pmatrix}\begin{pmatrix}\lambda_1\sigma_1 + \lambda_3\sigma_2\\ \lambda_2\sigma_2 + \lambda_3\sigma_1\end{pmatrix} = \begin{pmatrix}0\\0\end{pmatrix}.$$

Since the coefficient matrix has determinant

$$\sigma_1\dot\sigma_2 - \sigma_2\dot\sigma_1 = W_{12}\neq 0,$$

we deduce that

$$\lambda_1\sigma_1 + \lambda_3\sigma_2 = \lambda_2\sigma_2 + \lambda_3\sigma_1 = 0.$$

In particular

$$\lambda_1\neq 0 \Rightarrow \sigma_1 \propto \sigma_2, \quad \Rightarrow W_{12} = 0,$$
$$\lambda_2\neq 0 \Rightarrow \sigma_2 \propto \sigma_1, \quad \Rightarrow W_{12} = 0,$$
$$\lambda_3\neq 0 \Rightarrow \sigma_1 \propto \sigma_2, \quad \Rightarrow W_{12} = 0.$$

But $W_{12}\neq 0$. Hence $\lambda_i = 0\,(i = 1,2,3), \Rightarrow k_i' = k_i\ (i = 1,2,3)$; thus Theorem 3_n is proved.

By taking (for given *arbitrary* N,K) a,b,e as in (6a), (6b), (6c), we obtain the following from Theorems 3_s and 3_n:

Theorem 4: Let σ_k $(k = 1,2)$ satisfy

$$\ddot\sigma_k + N\dot\sigma_k + K\sigma_k = 0 \quad (k = 1,2). \tag{19}$$

Let $W_{12} = \sigma_1\dot\sigma_2 - \dot\sigma_1\sigma_2$. Then $W_{12}\exp[\int_{t_0}^t N(t')\,dt']$ is constant, and for all constants k_1,k_2,k_3 the function σ given by any twice-differentiable solution of

$$\sigma^2 = k_1\sigma_1^2 + k_2\sigma_2^2 + 2k_3\sigma_1\sigma_2 \quad [\text{cf. (16)}] \tag{20}$$

satisfies

$$\ddot\sigma + N\dot\sigma + K\sigma = (c/\sigma^3)\exp\left[-2\int_{t_0}^t N(t')\,dt'\right] \tag{21}$$

with

$$c = (k_1k_2 - k_3^2)\left\{W_{12}\exp\left[\int_{t_0}^t N(t')\,dt'\right]\right\}^2. \tag{22}$$

In particular, if $W_{12}\neq 0$, then for any c the general solution of (21) can be put in the form (20), where k_1,k_2,k_3 are constants satisfying

$$k_1k_2 - k_3^2 = c\left\{W_{12}\exp\left[\int_{t_0}^t N(t')\,dt'\right]\right\}^{-2} \tag{22'}$$

in one and only one way.

Parts of Theorem 4, and special cases, have been given previously, in various forms, by various authors (see, for example, Refs. 15–21). Extensions of parts of Theorem 4 have been given by Reid,[22,23] and by Berkovich and Rozov.[24]

A direct proof of Theorem 4: Theorem 4 can be proved directly, by generalizing a device introduced by Eliezer and Gray[18] for the purpose of discussing the physical interpretation of the homogeneous quadratic invariants for the one-dimensional classical time-dependent harmonic oscillator.

Thus, following Eliezer and Gray,[18] let us associate with Eq. (19) the "auxiliary two-dimensional motion"

$$\ddot{\mathbf{r}} + N\dot{\mathbf{r}} + K\mathbf{r} = \mathbf{0}, \tag{23}$$

where $\mathbf{r} = (x,y)$ and x,y can be complex. (Note that $|\mathbf{r}|^2, = x^2 + y^2$, can also be complex.)

Now let us describe this two-dimensional motion using plane polar coordinates σ, ψ (i.e., where $x = \sigma\cos\psi$ and $y = \sigma\sin\psi$). Resolving (23) radially and transversely, we obtain

$$\ddot\sigma - \sigma\dot\psi^2 + N\dot\sigma + K\sigma = 0, \tag{24}$$

$$(1/\sigma)(\sigma^2\dot\psi)^{\cdot} + N\sigma\dot\psi = 0. \tag{25}$$

Now (25)$\Rightarrow$

$$(\sigma^2\dot\psi)^{\cdot} + N(\sigma^2\dot\psi) = 0$$

i.e.,

$$(M\sigma^2\dot\psi)^{\cdot} = 0,$$

where $\dot M = NM$ (the arbitrary multiplicative constant in M being chosen as is convenient), i.e.,

$$M\sigma^2\dot\psi = \text{const}, \quad = h \text{ say}. \tag{26}$$

(Note that h can be complex.)

Now by a standard formula

$$|\mathbf{r}\wedge\dot{\mathbf{r}}|^2 = (\sigma^2\dot\psi)^2;$$

thus (26)$\Rightarrow$

$$|\mathbf{r}\wedge\dot{\mathbf{r}}|^2 = h^2/M^2. \tag{27}$$

Also substituting (26) into (24) leads to

$$\ddot{\sigma} + N\dot{\sigma} + K\sigma = h^2/M^2\sigma^3. \tag{28}$$

Thus, setting $c = h^2$, we see that if $\mathbf{r}$ satisfies (23) and

$$|\mathbf{r}\wedge\dot{\mathbf{r}}|^2 = c/M^2, \tag{29}$$

then any twice-differentiable solution σ of $\sigma^2 = |\mathbf{r}|^2$ satisfies

$$\ddot{\sigma} + N\dot{\sigma} + K\sigma = c/M^2\sigma^3 \quad \text{[cf. (2b)]}. \tag{30}$$

Now suppose σ_k $(k = 1,2)$ satisfy (19), and let $W_{12} = \sigma_1\dot{\sigma}_2 - \dot{\sigma}_1\sigma_2$. Then for all (possibly complex) vectors $\mathbf{A}_1,\mathbf{A}_2$

$$\mathbf{r} = \sigma_1\mathbf{A}_1 + \sigma_2\mathbf{A}_2 \tag{31}$$

is a solution of (23). In particular,

$$\begin{aligned}\mathbf{r}\wedge\dot{\mathbf{r}} &= (\sigma_1\mathbf{A}_1 + \sigma_2\mathbf{A}_2)\wedge(\dot{\sigma}_1\mathbf{A}_1 + \dot{\sigma}_2\mathbf{A}_2)\\ &= \sigma_1\dot{\sigma}_2\mathbf{A}_1\wedge\mathbf{A}_2 + \sigma_2\dot{\sigma}_1\mathbf{A}_2\wedge\mathbf{A}_1 = W_{12}(\mathbf{A}_1\wedge\mathbf{A}_2).\end{aligned}$$

Hence (29) is satisfied if and only if

$$c = M^2W_{12}^2\,|\mathbf{A}_1\wedge\mathbf{A}_2|^2.$$

(That MW_{12} *is* constant follows from the usual Wronskian argument.) Hence, by previous considerations, any twice-differentiable solution σ of

$$\begin{aligned}\sigma^2 = |\mathbf{r}|^2 &= |\sigma_1\mathbf{A}_1 + \sigma_2\mathbf{A}_2|^2\\ &= |\mathbf{A}_1|^2\sigma_1^2 + |\mathbf{A}_2|^2\sigma_2^2 + 2(\mathbf{A}_1\cdot\mathbf{A}_2)\sigma_1\sigma_2\end{aligned}$$

satisfies (30) for this c.

Now set $k_1 = |\mathbf{A}_1|^2$, $k_2 = |\mathbf{A}_2|^2$, $k_3 = (\mathbf{A}_1\cdot\mathbf{A}_2)$. (It can be shown that k_1,k_2,k_3 can be arbitrary.) Then, by another standard formula,

$$|\mathbf{A}_1\wedge\mathbf{A}_2|^2 = |\mathbf{A}_1|^2\cdot|\mathbf{A}_2|^2 - (\mathbf{A}_1\cdot\mathbf{A}_2)^2 = k_1k_2 - k_3^2.$$

We have shown that for all constants k_1,k_2,k_3 any twice-differentiable solution σ of

$$\sigma^2 = k_1\sigma_1^2 + k_2\sigma_2^2 + 2k_3\sigma_1\sigma_2 \quad \text{[cf. (20)]} \tag{32}$$

satisfies (30) with

$$c = (MW_{12})^2(k_1k_2 - k_3^2). \tag{33}$$

Conversely, for any c, *every* solution σ of (30) arises as some twice-differentiable solution σ of $\sigma^2 = |\mathbf{r}|^2$ where $\mathbf{r}$ satisfies (23) and (29). For suppose σ satisfies (30), and define $\mathbf{r} = (\sigma\cos\psi, \sigma\sin\psi)$ with $M\sigma^2\dot{\psi} = h$ and $h^2 = c$. Then (in a standard notation)

$$\begin{aligned}\ddot{\mathbf{r}} + N\dot{\mathbf{r}} + K\mathbf{r} &= (\ddot{\sigma} - \sigma\dot{\psi}^2)\mathbf{e}_\sigma + (1/\sigma)(\sigma^2\dot{\psi})\cdot\mathbf{e}_\psi\\ &\quad + N(\dot{\sigma}\mathbf{e}_\sigma + \sigma\dot{\psi}\mathbf{e}_\psi) + K\sigma\mathbf{e}_\sigma\\ &= \frac{(c-h^2)}{M^2\sigma^3}\,\mathbf{e}_\sigma + \frac{1}{M\sigma}\,(M\sigma^2\dot{\psi})\cdot\mathbf{e}_\psi = \mathbf{0}\end{aligned}$$

and

$$|\mathbf{r}\wedge\dot{\mathbf{r}}|^2 = (\sigma^2\dot{\psi})^2 = (h/M)^2 = c/M^2.$$

I.e., (23) and (29) are satisfied.

In particular, for any c the general solution of (23) and (29) yields, via $\sigma^2 = |\mathbf{r}|^2$, the general solution of (30).

So now let c be given, and suppose $W_{12}\neq 0$. Then (see Refs. 10 and 11) σ_1, σ_2 are linearly independent, and, in fact, (31) gives the general solution of (23). Now (31) satisfies (29) if and only if

$$|\mathbf{A}_1\wedge\mathbf{A}_2|^2 = c(MW_{12})^{-2}; \tag{34}$$

hence the general solution of (23) and (29) is given by (31) with $\mathbf{A}_1$, $\mathbf{A}_2$ satisfying (34).

Hence, by previous considerations, putting

$$\sigma^2 = |\mathbf{r}|^2 = |\mathbf{A}_1|^2\sigma_1^2 + |\mathbf{A}_2|^2\sigma_2^2 + 2(\mathbf{A}_1\cdot\mathbf{A}_2)\sigma_1\sigma_2,$$

where $\mathbf{A}_1$, $\mathbf{A}_2$ satisfy (34), gives the general solution of (30).

Setting $k_1 = |\mathbf{A}_1|^2$, etc. as previously, we find for any c the general solution of (30) can be put in the form (32), where k_1,k_2,k_3 are constants satisfying

$$k_1k_2 - k_3^2 = c(MW_{12})^{-2}. \tag{35}$$

The uniqueness of the expression (32) is proved exactly as for (16′) in Theorem 3_n.

We recover Theorem 4 by taking

$$M = \exp\left[\int_{t_0}^{t} N(t')\,dt'\right].$$

We recover[25] Theorems 3_s and 3_n by taking $N = -2\dot{a}/a$, K as in (4b), and $M = 1/a^2$.

Remark 3: We have seen that (30) [≡(2b)] is the radial equation for a certain auxiliary two-dimensional motion. What value does $J_{\sigma,c}$ [see (2a), but here written in terms of x instead of q, and considered as arising from the projection of that particular auxiliary motion onto the x axis] take during it?

We have

$$2J_{\sigma,c} = c(\sigma^{-1}x)^2 + M^2(\sigma\dot{x} - \dot{\sigma}x)^2,$$

where $x = \sigma\cos\psi$, $M\sigma^2\dot{\psi} = h$, and $h^2 = c$. Thus

$$\begin{aligned}\dot{x} = \dot{\sigma}\cos\psi - \sigma\dot{\psi}\sin\psi, \quad &\Rightarrow \sigma\dot{x} = \sigma\dot{\sigma}\cos\psi - \sigma^2\dot{\psi}\sin\psi,\\ &\Rightarrow \sigma\dot{x} - \dot{\sigma}x = -\sigma^2\dot{\psi}\sin\psi.\end{aligned}$$

Hence during that motion

$$\begin{aligned}2J_{\sigma,c} &= c(\sigma^{-1}\cdot\sigma\cos\psi)^2 + M^2(-\sigma^2\dot{\psi}\cdot\sin\psi)^2\\ &= c\cdot\cos^2\psi + h^2\cdot\sin^2\psi = c(\cos^2\psi + \sin^2\psi) = c.\end{aligned}$$

Thus the constancy of $J_{\sigma,c}$ is a consequence of the constancy of $M\sigma^2\dot{\psi}$. This generalizes Eliezer and Gray's interpretation[18] of the homogeneous quadratic invariants for the one-dimensional classical time-dependent harmonic oscillator.

IV. CONNECTION BETWEEN QUADRATIC SUPERPOSITION LAW AND ORIGINAL CHARACTERIZATION

Do Theorems 3_s and 3_n have a physical interpretation (in terms of homogeneous quadratic invariants)?

In this section we establish a close connection between these theorems and the characterization (4a), (4b). We begin with some preliminary results.

Let σ_k $(k = 1,2)$ satisfy $(\sigma_k,0)$ $(k = 1,2)$, and let $W_{12} = \sigma_1\dot{\sigma}_2 - \dot{\sigma}_1\sigma_2$. Note that

$$2I_{\sigma_k,c_k} = (\sigma_k p - \alpha_k q)^2 \quad (k = 1,2),$$

thus each $(\sigma_k p - \alpha_k q)$ is invariant. Let k_1,k_2,k_3 be constants, and let

$$2I_{k_1,k_2,k_3} = k_1(\sigma_1 p - \alpha_1 q)^2 + k_2(\sigma_2 p - \alpha_2 q)^2 + 2k_3(\sigma_1 p - \alpha_1 q)(\sigma_2 p - \alpha_2 q),$$

which is clearly a homogeneous quadratic invariant.

Theorem 5:

$$I_{k_1,k_2,k_3} = I_{\sigma,c},$$

where σ is any differentiable solution of

$$\sigma^2 = k_1\sigma_1^2 + k_2\sigma_2^2 + 2k_3\sigma_1\sigma_2 \tag{36a}$$

and

$$c = (k_1k_2 - k_3^2)(W_{12}/a^2)^2. \tag{36b}$$

Proof: In detail

$$2I_{k_1,k_2,k_3} = (k_1\sigma_1^2 + k_2\sigma_2^2 + 2k_3\sigma_1\sigma_2)p^2 - 2[k_1\sigma_1\alpha_1 + k_2\sigma_2\alpha_2 + k_3(\sigma_1\alpha_2 + \alpha_1\sigma_2)]qp + (k_1\alpha_1^2 + k_2\alpha_2^2 + 2k_3\alpha_1\alpha_2)q^2,$$

and

$$2I_{\sigma,c} = (c\sigma^{-2} + \alpha^2)q^2 - 2\sigma\alpha qp + \sigma^2p^2,$$

where $\alpha = (\dot\sigma - b\sigma)/a^2$.

Thus we want to prove that

$$k_1\sigma_1^2 + k_2\sigma_2^2 + 2k_3\sigma_1\sigma_2 = \sigma^2, \tag{37a}$$

$$k_1\sigma_1\alpha_1 + k_2\sigma_2\alpha_2 + k_3(\sigma_1\alpha_2 + \alpha_1\sigma_2) = \sigma\alpha, \tag{37b}$$

$$k_1\alpha_1^2 + k_2\alpha_2^2 + 2k_3\alpha_1\alpha_2 = c\sigma^{-2} + \alpha^2, \tag{37c}$$

where σ is any differentiable solution of (36a) and c is as in (36b).

We start by defining σ to be any differentiable solution of (36a). Then certainly (37a) is satisfied. Now, substituting for α_1,α_2,α, Eqs. (37b) and (37c) become

$$k_1\sigma_1\left(\frac{\dot\sigma_1 - b\sigma_1}{a^2}\right) + k_2\sigma_2\left(\frac{\dot\sigma_2 - b\sigma_2}{a^2}\right) + k_3\left[\sigma_1\left(\frac{\dot\sigma_2 - b\sigma_2}{a^2}\right) + \left(\frac{\dot\sigma_1 - b\sigma_1}{a^2}\right)\sigma_2\right] = \sigma\left(\frac{\dot\sigma - b\sigma}{a^2}\right),$$

i.e.,

$$k_1\sigma_1\dot\sigma_1 + k_2\sigma_2\dot\sigma_2 + k_3(\sigma_1\sigma_2)^{\cdot} - b(k_1\sigma_1^2 + k_2\sigma_2^2 + 2k_3\sigma_1\sigma_2) = \sigma\dot\sigma - b\sigma^2, \tag{37b'}$$

and

$$k_1\left(\frac{\dot\sigma_1 - b\sigma_1}{a^2}\right)^2 + k_2\left(\frac{\dot\sigma_2 - b\sigma_2}{a^2}\right)^2 + 2k_3\left(\frac{\dot\sigma_1 - b\sigma_1}{a^2}\right)\left(\frac{\dot\sigma_2 - b\sigma_2}{a^2}\right) = c\sigma^{-2} + \left(\frac{\dot\sigma - b\sigma}{a^2}\right)^2,$$

i.e.,

$$(k_1\dot\sigma_1^2 + k_2\dot\sigma_2^2 + 2k_3\dot\sigma_1\dot\sigma_2) - 2b[k_1\sigma_1\dot\sigma_1 + k_2\sigma_2\dot\sigma_2 + k_3(\sigma_1\sigma_2)^{\cdot}] + b^2(k_1\sigma_1^2 + k_2\sigma_2^2 + 2k_3\sigma_1\sigma_2) = ca^4\sigma^{-2} + \dot\sigma^2 - 2b\sigma\dot\sigma + b^2\sigma^2. \tag{37c'}$$

But (36a)$\Rightarrow$

$$\sigma\dot\sigma = k_1\sigma_1\dot\sigma_1 + k_2\sigma_2\dot\sigma_2 + k_3(\sigma_1\sigma_2)^{\cdot}, \tag{36a'}$$

and clearly (36a), (36a')$\Rightarrow$(37b'). Thus (37b) is satisfied.

Now, substituting (36a) and (36a') into (37c'), we find that (37c') is satisfied if and only if

$$k_1\dot\sigma_1^2 + k_2\dot\sigma_2^2 + 2k_3\dot\sigma_1\dot\sigma_2 = ca^4\sigma^{-2} + \dot\sigma^2,$$

i.e., if and only if

$$\sigma^2(k_1\dot\sigma_1^2 + k_2\dot\sigma_2^2 + 2k_3\dot\sigma_1\dot\sigma_2) - \sigma^2\dot\sigma^2 = ca^4,$$

i.e., [using (36a) and (36a')] if and only if

$$(k_1\sigma_1^2 + k_2\sigma_2^2 + 2k_3\sigma_1\sigma_2)(k_1\dot\sigma_1^2 + k_2\dot\sigma_2^2 + 2k_3\dot\sigma_1\dot\sigma_2) - [k_1\sigma_1\dot\sigma_1 + k_2\sigma_2\dot\sigma_2 + k_3(\sigma_1\sigma_2)^{\cdot}]^2 = ca^4. \tag{37c''}$$

This simplifies, after some manipulation, to

$$(k_1k_2 - k_3^2)W_{12}^2 = ca^4.$$

Thus if we define c by (36b), then (37c''), and hence (37c), is satisfied. Hence Theorem 5 is proved.

Remark 4: In the above proof the σ_k need *not* satisfy $(\sigma_k,0)$, and, in fact, could be *arbitrary* functions of t. (Of course, then I_{k_1,k_2,k_3} might not be an invariant, and (W_{12}/a^2), and hence c, might not be constant.)

This shows that Theorem 5 is essentially algebraic in nature and, in particular, independent of Theorems 3_s and 3_n.

A shorter proof of Theorem 5 can be given. This second proof (we omit the details) *does* rely on the σ_k satisfying $(\sigma_k,0)$, and makes use of the characterization (4a), (4b) and of Theorem 3_s.

We have given the first, longer, proof because for our purposes it is useful to have a proof which does not rely on either of Theorems 3_s and 3_n or on the characterization 4(a), (4b).

Theorem 6: Let σ_k $(k = 1,2)$ satisfy $(\sigma_k,0)$ $(k = 1,2)$, and suppose $W_{12} \neq 0$. Then every homogeneous quadratic invariant I can be put in the form I_{k_1,k_2,k_3}, where k_1,k_2,k_3 are constants satisfying

$$k_1k_2 - k_3^2 = (W_{12}/a^2)^{-2}\cdot\det(2I), \tag{38}$$

in one and only one way.

Proof: We have noted previously that each $(\sigma_k p - \alpha_k q)$ is invariant. In particular,

$$\{\sigma_1 p - \alpha_1 q, \sigma_2 p - \alpha_2 q\} = (\sigma_1\alpha_2 - \alpha_1\sigma_2) = (1/a^2)(\sigma_1\dot\sigma_2 - \dot\sigma_1\sigma_2) = W_{12}/a^2.$$

(Thus Poisson's theorem tells us that W_{12}/a^2 is constant, a result seen previously.)

Furthermore,

$$Q = \mu(\sigma_1 p - \alpha_1 q) \quad [\mu = (W_{12}/a^2)^{-1}],$$

$$P = (\sigma_2 p - \alpha_2 q)$$

satisfy $\{Q,P\} = 1$, and hence (see Ref. 5) Γ: $(q,p)\to(Q,P)$ is a homogeneous linear canonical transformation. (Note that Q,P are also invariants).

Now let $I = I(q,p,t)$ be an arbitrary homogeneous quadratic invariant. Suppose $I(q,p,t) = J(Q,P,t)$ when expressed in (Q,P,t) terms. Since Γ is homogeneous linear, so is Γ^{-1} (its exact form does not matter) and hence J is homogeneous quadratic in Q,P. Thus

$$2J(Q,P,t) = k_1'Q^2 + k_2'P^2 + 2k_3'QP \tag{39}$$

for some k_1',k_2',k_3'.

Since Q,P,J are all invariants, by Poisson's theorem so are $\{J,P\}$, $\{Q,J\}$, $\{\{J,P\},P\} = k'_1$, $\{Q,\{Q,J\}\} = k'_2$, $\{Q,\{J,P\}\} = k'_3$. Thus k'_1,k'_2,k'_3 are constants (possibly obvious).

Also, since Γ (and hence Γ^{-1}) has unit determinant (this being the canonicity condition for one degree of freedom), it can be shown that $\det(2J) = \det(2I)$, i.e.,

$$k'_1k'_2 - k'^2_3 = \det(2I).$$

Now, expressing J in (q,p,t) terms, we obtain

$$2I = k'_1\mu^2(\sigma_1 p - \alpha_1 q)^2 + k'_2(\sigma_2 p - \alpha_2 q)^2 + 2k'_3\mu(\sigma_1 p - \alpha_1 q)(\sigma_2 p - \alpha_2 q),$$

and, setting $k_1 = \mu^2 k'_1$, $k_2 = k'_2$, $k_3 = \mu k'_3$ (all constants), we find

$$\begin{aligned} 2I &= k_1(\sigma_1 p - \alpha_1 q)^2 + k_2(\sigma_2 p - \alpha_2 q)^2 + 2k_3(\sigma_1 p - \alpha_1 q)(\sigma_2 p - \alpha_2 q), \\ &= 2I_{k_1,k_2,k_3}, \end{aligned}$$

where k_1,k_2,k_3 are constants satisfying

$$k_1k_2 - k_3^2 = (\mu^2 k'_1)\cdot k'_2 - (\mu k'_3)^2 = \mu^2(k'_1k'_2 - k'^2_3) = (W_{12}/a^2)^{-2}\cdot\det(2I).$$

We prove that the expression $I = I_{k_1,k_2,k_3}$ is unique. Suppose that $I = I_{l_1,l_2,l_3}$, where l_1,l_2,l_3 are constants satisfying

$$l_1l_2 - l_3^2 = (W_{12}/a^2)^{-2}\cdot\det(2I).$$

Then, when expressed in (Q,P,t) terms,

$$\begin{aligned} 2I &= l_1(Q/\mu)^2 + l_2P^2 + 2l_3(Q/\mu)P \\ &= (l_1/\mu^2)\cdot Q^2 + l_2P^2 + 2\cdot(l_3/\mu)QP. \end{aligned} \tag{40}$$

From (39) and (40)

$$l_1/\mu^2 = k'_1, \quad l_2 = k'_2, \quad l_3/\mu = k'_3,$$

i.e.,

$$l_1 = \mu^2 k'_1 = k_1, \quad l_2 = k'_2 = k_2, \quad l_3 = \mu k'_3 = k_3.$$

Thus Theorem 6 is proved.

Remark 5 (*cf. Remark 4*): The above proof *does* rely on the σ_k satisfying $(\sigma_k,0)$. A shorter proof of Theorem 6 can be given. This second proof (we omit the details) also relies on the σ_k satisfying $(\sigma_k,0)$, and makes use of the characterization (4a), (4b) and of Theorems 3_n and 5.

We have given the first, longer, proof because (as in the case of Theorem 5) for our purposes it is useful to have a proof which does not rely on either of Theorems 3_s and 3_n or on the characterization (4a), (4b).

Now, having proved Theorems 5 and 6 independently of either of Theorems 3_s and 3_n and of the characterization (4a), (4b), we are in a position to establish the main result of this section.

Let us use $(4a,b)_n$ [respectively, $(4a,b)_s$] to denote the result that conditions (4a) and (4b) are necessary (respectively, sufficient) for I to be a homogeneous quadratic (in q,p) invariant for the Hamiltonian (3).

Theorem 7: Let σ_k $(k = 1,2)$ satisfy $(\sigma_k,0)$ $(k = 1,2)$, and suppose $W_{12} \neq 0$. Then:

(i) $(4a,b)_n \Rightarrow$ Theorem 3_s;

(ii) Theorem $3_s \Rightarrow (4a,b)_n$;

(iii) $(4a,b)_s \Rightarrow$ Theorem 3_n;

(iv) Theorem $3_n \Rightarrow (4a,b)_s$.

Proof of (*i*): Let k_1,k_2,k_3 be constants, and let σ be given by any twice-differentiable solution of

$$\sigma^2 = k_1\sigma_1^2 + k_2\sigma_2^2 + 2k_3\sigma_1\sigma_2.$$

Let $c = (k_1k_2 - k_3^2)(W_{12}/a^2)^2$. We prove that σ satisfies (σ,c). Form the corresponding I_{k_1,k_2,k_3}. From Theorem 5

$$I_{k_1,k_2,k_3} = I_{\sigma',c'},$$

where σ' is any differentiable solution of $\sigma'^2 = k_1\sigma_1^2 + k_2\sigma_2^2 + 2k_3\sigma_1\sigma_2$ (so in particular $\sigma' = \sigma$ will do) and

$$c' = (k_1k_2 - k_3^2)(W_{12}/a^2)^2 = c.$$

Thus $I_{k_1,k_2,k_3} = I_{\sigma,c}$. But I_{k_1,k_2,k_3} is a homogeneous quadratic invariant and hence, by $(4a,b)_n$, of form $I_{\bar\sigma,\bar c}$ for some $\bar\sigma$, $\bar c$ satisfying $(\bar\sigma,\bar c)$. In particular $I_{\sigma,c} = I_{\bar\sigma,\bar c}$. Hence [cf. (7b), (7c)]

$$\sigma^2 = \bar\sigma^2 \tag{41}$$

and

$$c = \bar c. \tag{42}$$

From (41) we deduce [see paragraph preceding (7a)] that σ satisfies $(\sigma,\bar c)$. From (42) we deduce that σ satisfies (σ,c). {Previously Theorem 3_s was built up out of certain consequences [Theorem 2(i), (iii)] of the characterization (4a), (4b). Here we have given a more direct derivation of Theorem 3_s.}

Proof of (*ii*): Let I be an arbitrary homogeneous quadratic invariant. By Theorem 6 it can be put in the form I_{k_1,k_2,k_3}, where k_1,k_2,k_3 are constants satisfying (38), in one (and only one) way. By Theorem 5

$$I_{k_1,k_2,k_3} = I_{\sigma,c},$$

where σ is any differentiable solution of $\sigma^2 = k_1\sigma_1^2 + k_2\sigma_2^2 + 2k_3\sigma_1\sigma_2$ and

$$c = (k_1k_2 - k_3^2)(W_{12}/a^2)^2, \quad = \det(2I). \quad \text{[from (38)]}$$

But, for any such σ,

$$\dot\sigma = \sigma^{-1}[k_1\sigma_1\dot\sigma_1 + k_2\sigma_2\dot\sigma_2 + k_3(\sigma_1\sigma_2)^{\cdot}],$$

which is itself differentiable. Hence, in fact, any such σ is *twice*- differentiable and, in particular, by Theorem 3_s satisfies (σ,c') with

$$c' = (k_1k_2 - k_3^2)(W_{12}/a^2)^2 = c = \det(2I).$$

Thus $I\,(= I_{k_1,k_2,k_3}) = I_{\sigma,c}$, where σ,c satisfy (σ,c). [And in fact $c = \det(2I)$.]

Proof[25] *of* (*iii*): Let σ,c satisfy (σ,c). By $(4a,b)_s$ we know that $I_{\sigma,c}$ is a homogeneous quadratic invariant. By Theorem 6 we know that $I_{\sigma,c}$ can be put in the form I_{k_1,k_2,k_3}, where k_1,k_2,k_3 are constants satisfying

$$\begin{aligned} k_1k_2 - k_3^2 &= (W_{12}/a^2)^{-2}\cdot\det(2I_{\sigma,c}) \\ &= c(W_{12}/a^2)^{-2} \quad \text{[using (4c)]} \end{aligned} \tag{43}$$

in one and only one way. In particular, comparing coefficients of p^2 leads immediately to

$$\sigma^2 = k_1\sigma_1^2 + k_2\sigma_2^2 + 2k_3\sigma_1\sigma_2. \tag{44}$$

We have shown that σ can be put in the form (44), where k_1,k_2,k_3 are constants satisfying (43). The uniqueness of the expression (44) can be proved either exactly as for (16') in the

original proof (Sec. III) of Theorem 3_n, or as follows: Suppose that $\sigma^2 = l_1\sigma_1^2 + l_2\sigma_2^2 + 2l_3\sigma_1\sigma_2$, where l_1,l_2,l_3 are constants satisfying $l_1l_2 - l_3^2 = c(W_{12}/a^2)^{-2}$. Then [cf. details in Proof of (i)] $I_{l_1,l_2,l_3} = I_{\sigma,c}$. Thus $I_{\sigma,c} = I_{k_1,k_2,k_3} = I_{l_1,l_2,l_3}$. Hence, by uniqueness in Theorem 6, $l_1 = k_1$, etc.

Proof of (iv): Let $I = I_{\sigma,c}$, where σ,c satisfy (σ,c). By Theorem 3_n we know that σ can be put in the form (44), where k_1,k_2,k_3 are constants satisfying (43), in one (and only one) way. Form the corresponding I_{k_1,k_2,k_3}. Then [cf. details in Proof of (i)] $I_{k_1,k_2,k_3} = I_{\sigma,c}$. Thus $I = I_{\sigma,c} = I_{k_1,k_2,k_3}$, which is a homogeneous quadratic invariant. [Of course, $(4a,b)_s$ can also be verified directly.] Hence Theorem 7 is proved.

ACKNOWLEDGMENTS

The author would like to thank Professor Courtney Coleman (of Harvey Mudd College) for much encouragement, and for drawing his attention to Ref. 24. The author would also like to thank S. Fairthorne, M. E. Major, R. L. Perry, and A. E. Wollenberg for useful conversations.

[1]L. S. Wollenberg, Phys. Lett. A **79**, 269 (1980).
[2]Here and throughout the remainder of this paper c denotes a constant and σ,c can be complex.
[3]The $a\equiv 0$ case can always be brought to the form (3) by means of a suitable canonical transformation.
[4]L. S. Wollenberg, in preparation.
[5]H. Goldstein, *Classical Mechanics* (Addison-Wesley, Reading, MA, 1950).
[6]Here and throughout the remainder of this paper $k_1,k_2,\ldots,k_n$ can be complex.
[7]N. J. Günther and P. G. L. Leach, J. Math. Phys. **18**, 572 (1977).
[8]J. R. Ray and J. L. Reid, Phys. Lett. A **74**, 23 (1979).
[9]Here and throughout the remainder of this paper $c_1,c_2,\cdots$ denote constants and $\sigma_1,\sigma_2,\cdots,c_1,c_2,\cdots$ can be complex.
[10]W. Hurewicz, *Lectures on Ordinary Differential Equations* (M.I.T. Press, Cambridge, MA, 1958).
[11]J. C. Burkill, *The Theory of Ordinary Differential Equations* (Oliver and Boyd, Edinburgh, 1962).
[12]J. L. Reid and J. R. Ray, J. Math. Phys. **21**, 1583 (1980).
[13]J. R. Ray, Phys. Lett. A **78**, 4 (1980).
[14]So this proof relies on an assumption about the nature of the general solution of the auxiliary equation. Two other proofs, given below, avoid this assumption.
[15]E. Pinney, Proc. Amer. Math. Soc. **1**, 681 (1950).
[16]H. R. Lewis, Jr., Phys. Rev. Lett. **18**, 510, 636 (1967).
[17]H. R. Lewis, Jr., J. Math. Phys. **9**, 1976 (1968).
[18]C. J. Eliezer and A. Gray, SIAM J. Appl. Math. **30**, 463 (1976).
[19]J. W. Macki, J. Math. Phys. **18**, 1256 (1977).
[20]M. Lutzky, Phys. Lett. A **68**, 3 (1978).
[21]J. R. Ray and J. L. Reid, J. Math. Phys. **20**, 2054 (1979).
[22]J. L. Reid, Proc. Am. Math. Soc. **27**, 61 (1971).
[23]J. L. Reid, Proc. Am. Math. Soc. **38**, 532 (1973).
[24]L. M. Berkovich and N. Kh. Rozov, Differential Eq. **8**, 1609 (1972).
[25]This proof of Theorem 3_n is one of the other proofs referred to in Footnote 14.

Wobbling kinks in ϕ^4 and sine-Gordon theory

Harvey Segur
Aeronautical Research Associates of Princeton, Inc., P. O. Box 2229, 50 Washington Road, Princeton, New Jersey 08540

(Received 7 December 1982; accepted for publication 25 February 1983)

When the ϕ^4 model admits a kink solution, it also admits a wobbling kink, which satisfies the boundary conditions of a kink, but possesses an internal degree of freedom. In this paper we develop a formal perturbation series for the wobbling kink in ϕ^4 theory, and give the first two terms in the series explicitly. Then we prove that the formal series actually is asymptotic for a rather long time [$O(K \ln(1/\epsilon))$ for a certain K]. Finally, we construct an exact 3-soliton solution of the sine-Gordon equation that also has the properties of a wobbling kink. For the sine-Gordon equation, the wobbling kink seems to be mildly unstable.

PACS numbers: 03.40.Kf, 03.80. + r

1. INTRODUCTION

The ϕ^4 model of one-dimensional field theory, defined by the Hamiltonian

$$H = \frac{1}{2}\int_{-\infty}^{\infty} dX\left[\phi_T^2 + \phi_X^2 + \frac{1}{2}(\phi^2 - 1)^2\right], \tag{1.1}$$

is well known in many branches of physics. In Ref. 1, it was introduced as a simple continuum model of lightly doped polyacetylene [$(CH)_x$]. In that context, a solution of fundamental interest is the kink,

$$\phi(X, T) = X\tanh\{\tfrac{1}{2}(k + (2k)^{-1})X + \tfrac{1}{2}(k - (2k)^{-1})T + X_0\},$$

which separates two adjacent regions of different phase of bond alteration in trans-$(CH)_x$. Further discussion of the physical problem may be found in Ref. 2, in its references, and in Ref. 3.

On the basis of an approximate ansatz, Rice and Mele[4] found that (1.1) may admit more complicated kinklike solutions, with internal degrees of freedom, which we will call "wobbling kinks." A similar ansatz, applied to the sine-Gordon equation

$$u_{tt} - u_{xx} = \sin u, \tag{1.2}$$

indicates that it, too, may admit wobbling kinks.[5] The purpose of this paper is to establish rigorously that these two models admit wobbling kink solutions, at least for limited (but large) times.

Two definitions are required to make this discussion meaningful. (i) In ϕ^4 theory, an ordinary kink is an exact solution of the equation that satisfies the boundary conditions

$$\phi \to \pm 1 \text{ as } X \to \pm\infty$$

and is stationary in an appropriate coordinate system. (ii) A wobbling kink is a solution of (1.4) whose time average satisfies the boundary conditions of an ordinary kink, and which is periodic in time in an appropriate coordinate system. (A more general definition might allow almost periodic functions, but this generalization is not needed here.)

In order to state concisely my results for ϕ^4 theory, it is convenient to shift to the rest frame of a given kink. Let

$$\begin{aligned} x &= \tfrac{1}{2}(k + (2k)^{-1})X + \tfrac{1}{2}(k - (2k)^{-1})T + x_0, \\ t &= \tfrac{1}{2}(k - (2k)^{-1})X + \tfrac{1}{2}(k + (2k)^{-1})T + t_0. \end{aligned} \tag{1.3}$$

The governing equation becomes

$$\tfrac{1}{2}\phi_{tt} - \tfrac{1}{2}\phi_{xx} = \phi - \phi^3, \tag{1.4}$$

with a stationary kink solution

$$\phi(x, t) = \phi_0(x) = \tanh(x). \tag{1.5}$$

Let ϵ denote an ordering parameter ($0 < \epsilon \ll 1$) in a formal series solution of (1.4),

$$\phi(x, t) = \phi_0(x) + \sum_1^\infty \epsilon^n \phi_n(x, t), \tag{1.6a}$$

and let

$$\Phi_N(x, t) = \phi_0(x) + \sum_1^N \epsilon^n \phi_n(x, t) \tag{1.6b}$$

denote the N th truncation of the series. In Secs. 2 and 3, we demonstrate the following results.

(i) For sufficiently small ϵ, (1.4) admits a formal series solution corresponding to a wobbling kink. We may choose

$$\epsilon\phi_1(x, t) = \epsilon\,\mathrm{sech}\,x\tanh x\sin\sqrt{3}t, \tag{1.7}$$

so the frequency of oscillation is $\sqrt{3}$ to this order of approximation.

(ii) The Green's function of the perturbation problem is given explicitly, so any higher term ($n \geqslant 2$) in (1.6a) may be found by quadratures over known functions and/or lower terms in the series. To demonstrate the effectiveness of this procedure, we will find $\phi_2(x, t)$ in closed form.

(iii) Both $\phi_1(x, t)$ and $\phi_2(x, t)$ are uniformly bounded in (x, t).

(iv) Assume $\Phi_N(x, t)$ is uniformly bounded. Then there exists a constant K_N and an exact solution of (1.4), $\phi(x, t)$, such that

 0022-2488/83/061439-05$02.50 © 1983 American Institute of Physics

$$\phi(x,t) - \Phi_N(x,t) = o(\epsilon^N), \tag{1.8}$$

at least for times up to

$$|t| \leqslant O((1/2K_N)\ln(1/\epsilon)); \tag{1.9}$$

i.e., the formal series is asymptotic on a relatively long time scale. The constant K_N may be chosen to depend only on N, provided $\epsilon < 1$.

(v) The energy of the stationary kink is $H = 2/3$. The wobbling kinks are more energetic,

$$H = \tfrac{2}{3} + \tfrac{1}{2}\epsilon^2 + O(\epsilon^3). \tag{1.10}$$

The sine-Gordon equation also admits wobbling kinks, but they are rather different from the wobbling kinks of ϕ^4 theory. It is known that (1.2) admits N-soliton solutions.[6–8] In Sec. 4, we construct a particular 3-soliton solution that has the properties of a wobbling kink. One may think of it as a nonlinear superposition of a kink and a breather, traveling at the same speed.

2. THE FORMAL SERIES FOR ϕ^4

Substituting (1.6a) into (1.4) and equating to zero the coefficients of ϵ^n yields a sequence of linear differential equations to be satisfied. Let

$$Lv \equiv \left[\tfrac{1}{2}\partial_t^2 - \tfrac{1}{2}\partial_x^2 + 2 - 3\,\mathrm{sech}^2 x\right]v. \tag{2.1}$$

One finds that

$$L\phi_1 = 0, \tag{2.2a}$$

$$L\phi_2 = -3\phi_0(\phi_1)^2, \tag{2.2b}$$

and for $n \geqslant 3$,

$$L\phi_n = -3\phi_0\sum_{j=1}^{n-1}\phi_j\phi_{n-j} - \sum_{j=1}^{n-2}\sum_{k=1}^{n-2}\phi_j\phi_k\phi_{n-j-k}. \tag{2.2c}$$

Our objective is to find solutions of these equations that are uniformly bounded in (x, t)

Separating variables in (2.2a) [i.e., $\phi_1(x,t) = p(x)q(t)$] reduces that equation to a classical eigenvalue problem. The solutions that correspond to the discrete eigenvalues and that are bounded in x are

$$\bar{\phi}_1(x,t) = (A + Bt)\,\mathrm{sech}^2 x, \tag{2.3}$$

$$\phi_1(x,t) = C\,\mathrm{sech}\,x\tanh x\sin(\sqrt{3}(t+t_0)). \tag{1.7'}$$

The "translation mode" $\bar{\phi}_1$ represents the fact that (1.4) also admits exact kink solutions with a slightly different position (A) or velocity (B). In the present context, only the "shape mode" ϕ_1 is of interest. Without loss of generality, we may set $C = 1$, $t_0 = 0$ in (1.7').

Once ϕ_1 is chosen, ϕ_2 is found by solving (2.2b). The method of solution of (2.2b) indicates the general procedure for all of the higher-order equations. First, we express the forcing terms as a finite sum of simple harmonics in t:

$$L\phi_2 = -\tfrac{3}{2}\,\mathrm{sech}^2 x\tanh^3 x(1 - \cos(2\sqrt{3}t)).$$

Next, we require the solution to exhibit the same harmonics as those in the forcing terms:

$$\phi_2(x,t) = 1 \times f(x) + \cos(2\sqrt{3}t)g(x). \tag{2.4}$$

This assumption precludes unbounded growth in time, but secular terms may still appear if $f(x)$ or $g(x)$ grows without bound as $|x| \to \infty$. In any case, f and g satisfy ordinary differential equations:

$$\left[-\frac{1}{2}\left(\frac{d}{dx}\right)^2 + 2 - 3\,\mathrm{sech}^2 x\right]f = -\tfrac{3}{2}\,\mathrm{sech}^2 x\tanh^3 x. \tag{2.5}$$

Homogeneous solutions of (2.5) are

$$f_1 = \mathrm{sech}^2 x,$$
$$f_2 = 3x\,\mathrm{sech}^2 x + \tanh x + \sinh(2x), \tag{2.6}$$

and (2.5) now may be solved by variation of parameters. A particular solution is

$$f(x) = \tfrac{1}{2}\tanh x\,\mathrm{sech}^2 x - \tfrac{3}{4}x\,\mathrm{sech}^2 x. \tag{2.7}$$

Of the homogeneous solutions, $f_2(x)$ is unbounded and therefore must be suppressed; $f_1(x)$ is simply the translation mode reappearing at $O(\epsilon^2)$.

Similarly, $g(x)$ satisfies

$$-\left[\frac{1}{2}\left(\frac{d}{dx}\right)^2 + 4 + 3\,\mathrm{sech}^2 x\right]g = \tfrac{3}{2}\,\mathrm{sech}^2 x\tanh^3 x, \tag{2.8}$$

which can be solved by variation of parameters if we know two solutions of the homogeneous problem. More generally, as we go to higher order, we will need the solutions of homogeneous equations of the form

$$\left[\left(\frac{d}{dx}\right)^2 + p^2 + 6\,\mathrm{sech}^2 x\right]G = 0. \tag{2.9}$$

But these can all be found, using the "triangular representation" developed by Gel'fand and Levitan.[9] Set

$$G(x;p) = \exp(ipx) + \int_{-\infty}^{x} K(x,y)\exp(ipy)dy.$$

Then $K(x, y)$ satisfies the Gel'fand–Levitan–Marchenko equation,[7] and can be found in closed form for this particular potential:

$$K(x,y) = 3\,\mathrm{sech}\,x[\exp(y)\tanh x - \exp(2y)\,\mathrm{sech}\,x].$$

It follows that two linearly independent solutions of (2.9) are the real and imaginary parts of

$$G(x;p) = e^{ipx}\left[1 + \frac{3(1-ip)}{1+p^2}(\tanh x)(1+\tanh x) - \frac{3(2-ip)}{4+p^2}(1+\tanh x)^2\right] \tag{2.10}$$

as may be verified by direct substitution. Evaluating (2.10) with $p^2 = 8$ provides the functions required to solve (2.8). The final result at $O(\epsilon^2)$ is that the solution of (2.2b) is

$$\begin{aligned}\phi_2(x,t) = {} & \tfrac{1}{2}\tanh x\,\mathrm{sech}^2 x - \tfrac{3}{4}x\,\mathrm{sech}^2 x + c_1\,\mathrm{sech}^2 x \\ & - \tfrac{1}{8}\cos(2\sqrt{3}t) \\ & \times [(J_c(x)\cos px + J_s(x)\sin px)(3 - \tanh^2 x) \\ & + p(J_s(x)\cos px - J_c(x)\sin px)\tanh x \\ & + 3\,\mathrm{sech}^2 x\tanh x + c_2 G_c(x,p) + c_3 G_s(x,p)],\end{aligned} \tag{2.11}$$

where $p^2 = 8$,

$$J_c(x) + iJ_s(x) = \int_{-\infty}^{x} \exp(ip\xi)\operatorname{sech}^2 \xi \, d\xi, \tag{2.12}$$

$$G_c + iG_s = G(x;p),$$

G is defined by (2.10), and (c_1, c_2, c_3) are arbitrary real constants. It is easy to verify that $\phi_2(x, t)$ is uniformly bounded in (x, t).

Thus $\phi_0(x)$, $\phi_1(x, t)$, and $\phi_2(x, t)$ are each uniformly bounded. Higher terms in the series (1.6a) may exhibit secular growth. In fact, it seems likely that ϕ_3 may contain secular terms because (1.7) forces the frequency of the oscillation to be constant ($\omega = \sqrt{3}$), whereas the frequency of a nonlinear oscillator typically depends on amplitude. (*N. B.* I have not calculated ϕ_3.) If $\phi_3(x, t)$ turns out to be secular, it suggests that (1.7) should be replaced with

$$\phi_1(x, t;\epsilon) = \operatorname{sech} x \tanh x \sin(\omega t), \tag{2.13a}$$

$$\omega \sim \sqrt{3}(1 + \epsilon^2\omega_1 + \cdots), \tag{2.13b}$$

with a corresponding change in (2.2c). If the constant ω_1 may be chosen to eliminate the secular terms in ϕ_3, then (2.13) indicates the amplitude dependence of the frequency. This idea goes back to Stokes.[10]

It remains to calculate the energy of the stationary and wobbling kinks. Given ϕ_0, ϕ_1, and ϕ_2, it is straightforward to verify (1.10). Note that the $O(\epsilon)$ term in H vanishes identically because of (2.2a).

3. PROOF OF ASYMPTOTICITY FOR ϕ^4

In this section we assume that a formal series representation of a wobbling kink solution of (1.4) has been generated, up to and including $\phi_N(x, t;\epsilon)$ and that $(\phi_0, \phi_1, \ldots, \phi_N)$ all are smooth and uniformly bounded in (x, t). We saw in Sec. 2 that this is possible at least for $N = 1$ and 2. Denote $\Phi_N(x, t;\epsilon)$ as in (1.6b) and let $\hat{\phi}(x, t;\epsilon)$ denote the solution of (1.4) that satisfies initial conditions

$$\hat{\phi} = \Phi_N, \quad \partial_t\hat{\phi} = \partial_t\Phi_N \quad \text{at } t = 0, \tag{3.1a}$$

and boundary conditions

$$\hat{\phi} - \Phi_N \to 0 \quad \text{as } |x| \to \infty, \quad t > 0. \tag{3.1b}$$

By the usual theory of hyperbolic differential equations, this solution exists for some finite time and it is unique. Our objective here is to show that $\hat{\phi}$ exists for a rather long time, and that it remains close to Φ_N during this time. Specifically, we will show that there is a positive constant K_N such that if $t \leqslant O(1/2K_N \ln(1/\epsilon))$, then

$$\hat{\phi}(x, t;\epsilon) = \Phi_N(x, t;\epsilon) + o(\epsilon^N); \tag{3.2}$$

i.e., that the formal series developed in Sec. 2 actually is asymptotic for a fairly long time. The method of proof is a simple generalization of a method used by Kruskal[11] for ordinary differential equations.

Because Φ_N is known explicitly, it clearly must satisfy some definite differential equation. One shows that

$$\tfrac{1}{2}(\partial_t^2 - \partial_x^2)\Phi_N = \Phi_N - \Phi_N^3 + \epsilon^{N+1}F_N(x, t;\epsilon), \tag{3.3}$$

where $F_N = O(1)$ as $\epsilon \to 0$. For $N = 1$,

$$\epsilon^2 F_1 = \epsilon^2[3\phi_0\phi_1^2 + \epsilon\phi_1^3].$$

For $N = 2$,

$$\epsilon^3 F_2 = \epsilon^3[6\phi_0\phi_1\phi_2 + \phi_1^3] + 3\epsilon^4[\phi_0\phi_2^2 + \phi_1^2\phi_2] + 3\epsilon^5\phi_1\phi_2^2 + \epsilon^6\phi_2^3.$$

For $N \geqslant 3$,

$$\epsilon^{N+1}F_N = \epsilon^{N+1} \times 3\phi_0 \sum_{j=1}^{N} \phi_j\phi_{N+1-j} + 3\phi_0 \sum_{n=N+2}^{2N} \epsilon^n \sum_{j=n-N}^{N} \phi_j\phi_{n-j} + \sum_{n=N+3}^{2N} \epsilon^n \sum_{j=1}^{N} \sum_{k=A}^{B} \phi_j\phi_k\phi_{n-j-k} + \sum_{n=2N+1}^{3N} \epsilon^n \sum_{j=n-2N}^{N} \sum_{k=n-N-j}^{N} \phi_j\phi_k\phi_{n-j-k},$$

where $A = \max\{n - N - j, 1\}$, $B = \min\{n - 1 - j, N\}$. The only important feature of $F_N(x, t;\epsilon)$ for our purposes is that it be uniformly bounded. Thus, we may pick an $M_N < \infty$ such that if $|\epsilon| < 1$, then

$$|F_N(x, t;\epsilon)| \leqslant M_N. \tag{3.4}$$

Because F_N is a finite sum of uniformly bounded functions, it is clear that M_N exists. For $N = 1$, we may choose $M_1 = 4$.

Define

$$\delta_N(x, t;\epsilon) = \hat{\phi}(x, t;\epsilon) - \Phi_N(x, t;\epsilon). \tag{3.5}$$

Because Φ_N satisfies (3.3), while $\hat{\phi}$ satisfies (1.4) and (3.1), one shows that for $t > 0$, δ_N satisfies

$$\tfrac{1}{2}(\partial_t^2 - \partial_x^2)\delta_N = (1 - 3\Phi_N^2)\delta_N - 3\Phi_N\delta_N^2 - \delta_N^3 - \epsilon^{N+1}F_N, \qquad \delta_N \to 0 \quad \text{as } x \to \infty, \tag{3.6}$$

and at $t = 0$,

$$\delta_N = 0, \quad \partial_t\delta_N = 0.$$

We must show that (3.6) has a solution that is $o(\epsilon^N)$ over the time scale in question.

Let $R_N(\delta_N; x,t,\epsilon)$ denote the right-hand side of (3.6a). $R_N(\delta)$ is differentiable, so it certainly satisfies a Lipschitz condition. Specifically, if

$$|\delta_N| \leqslant \epsilon^N K, \quad |\bar{\delta}_N| \leqslant \epsilon^N K,$$

for some $K < \infty$, then there is a $K_N < \infty$ such that

$$|R_N(\delta_N;x,t,\epsilon) - R_N(\bar{\delta}_N;x,t,\epsilon)| \leqslant K_N^2 |\delta_N - \bar{\delta}_N|. \tag{3.7}$$

The best possible Lipschitz constant is not of interest here. If $\max(\epsilon)$ is small enough, $K_N^2 = 4$ for all $N \geqslant 1$.

The problem (3.6) is formally equivalent to

$$\delta_N(x,t,\epsilon) = \int_0^t d\tau \int_{x-t+\tau}^{x+t-\tau} d\xi \, R_N(\delta_N;\xi,\tau,\epsilon), \tag{3.8}$$

as the reader may verify directly. To solve (3.8), we consider an iterated mapping,

$$v_{p+1}(x,t,\epsilon) = \int_0^t d\tau \int_{x-t+\tau}^{x+t-\tau} d\xi\, R_N(v_p;\xi,\tau,\epsilon), \tag{3.9}$$

with $v_0 = 0$. First, we will show that the iterated mapping (3.9) has a fixed point, which is therefore a solution of (3.8). Second, we will bound the magnitude of this fixed point by controlling the rate of convergence of the mapping (3.9). This leads to the desired result, that the solution of (3.8) remains $o(\epsilon^N)$ over the time scale in question.

Because of (3.4),

$$|v_1 - v_0| \leqslant 2M_N \epsilon^{N+1} t^2/2!.$$

Because of (3.7),

$$\begin{aligned}|v_2 - v_1| &\leqslant \int_0^t d\tau \int_{x-t+\tau}^{x+t-\tau} d\xi\, K_N^2 |v_1 - v_0| \\ &\leqslant 4K_N^2 M_N \epsilon^{N+1} t^4/4!.\end{aligned}$$

In general,

$$|v_p - v_{p-1}| \leqslant (2K_N^2)^p \left(\frac{M_N}{K_N^2}\epsilon^{N+1}\right) \frac{t^{2p}}{(2p)!}, \tag{3.10}$$

which can be established by induction. As $p \to \infty$, holding $(K_N t)$ fixed, $v_p - v_{p-1} \to 0$ by (3.10). Therefore the sequence v_p converges to a fixed point, and this fixed point is a solution of (3.8). With more effort one may also show that this solution is unique and twice differentiable, so it solves (3.6) as well. It remains to estimate the magnitude of δ_N. Because

$$\delta_N = v_0 + \sum_{p=1}^{\infty} (v_p - v_{p-1}),$$

it follows that

$$\begin{aligned}|\delta_N| &\leqslant \frac{M_N \epsilon^{N+1}}{K_N^2} \sum_{p=1}^{\infty} \frac{(2K_N^2 t^2)^p}{(2p)!} \\ &= \frac{M_N \epsilon^{N+1}}{K_N^2} \left[\cosh(\sqrt{2}K_N t) - 1\right] \\ &\leqslant \frac{M_N \epsilon^{N+1}}{2K_N^2} \exp(\sqrt{2}K_N t).\end{aligned} \tag{3.11}$$

Thus, $\delta_N = o(\epsilon^N)$ provided

$$|t| \leqslant (1/2K_N) \ln(1/\epsilon); \tag{3.12}$$

this is the desired result.

In summary, we have now established that the formal series representation of the wobbling kink, as found in Sec. 2, is a proper asymptotic series, at least for times up to that given by (3.12). In other words, a solution of (1.4) exists that behaves approximately like a wobbling kink for a relatively long time.

4. THE SINE-GORDON EQUATION

Next we construct wobbling kink solutions of the sine-Gordon equation (1.2). It is known that (1.2) is completely integrable[6–8] and that it possesses N-soliton solutions. A wobbling kink for (1.2) is a special case of an N-soliton solution; one may think of it as a nonlinear superposition of an ordinary kink and a breather. Like all N-soliton solutions, it is intrinsically nonlinear and cannot be found by the naive sort of perturbation theory described in Sec. 2.

For the sine-Gordon equation, a stationary kink is given by

$$u(x,t) = -\pi + 4\arctan(me^x), \tag{4.1}$$

where m is an arbitrary real constant. The corresponding scattering data consist of one discrete eigenvalue at $\zeta = i/2$, and a "norming constant" m. The Hamiltonian for (1.2) is

$$H = \int dx \left[\frac{1}{2}(u_t)^2 + \frac{1}{2}(u_x)^2 + \cos u + 1\right], \tag{4.2}$$

and $H = 8$ for a stationary kink.

A stationary breather is given by

$$\begin{aligned}u(x,t) = &-\pi + 4\arctan\{(\beta/\alpha)\,\mathrm{sech}(2\beta x + x_0) \\ &\times \sin(2\alpha t + t_0)\},\end{aligned} \tag{4.3a}$$

where

$$\alpha^2 + \beta^2 = \tfrac{1}{4} \tag{4.3b}$$

and (x_0, t_0) are arbitrary real constants. The corresponding scattering data contain two discrete eigenvalues, $\zeta_1 = \alpha + i\beta, \beta > 0$ and $\zeta_2 = -\zeta_1^*$. There are also two complex norming constants m_1 and $m_2 = m_1^*$, whose real and imaginary parts determine x_0 and t_0, respectively. For a stationary breather, $H = 32\beta$, a fact most easily obtained from a trace formula for an N-soliton solution of (1.2)[12]:

$$H = 2i \sum_{j=1}^{N} \left[(\zeta_j)^{-1} + (\zeta_j^*)^{-1}\right]. \tag{4.4}$$

The scattering data for a wobbling kink solution are composed of scattering data for a breather plus those for a kink. We need three discrete eigenvalues,

$$\zeta_1 = i/2, \quad \zeta_2 = \alpha + i\beta, \quad \zeta_3 = -\zeta_2^*, \tag{4.5a}$$

with $\beta > 0$ and $\alpha^2 + \beta^2 = \frac{1}{4}$. We also need three norming constants:

$$m_1, \text{ real}, \quad m_2, \quad m_3 = m_2^*. \tag{4.5b}$$

Use these scattering data to define

$$\begin{aligned}M_{jk} = &(im_k/(\zeta_j + \zeta_k)) \\ &\times \exp\{-i(\zeta_j + \zeta_k)(x+t)/2 + i(x-t)/4\zeta_k\}.\end{aligned} \tag{4.6}$$

Then

$$u(x,t) = -\pi + 4\,\mathrm{Im}[\ln\{\det(I + iM)\}] \tag{4.7}$$

is an exact solution of (1.2).[8] If $m_2 = 0$, $m_1 \neq 0$, (4.7) reduces to (4.1); if $m_1 = 0$, $m_2 \neq 0$, it reduces to (4.3). With $m_1 \neq 0$, $m_2 \neq 0$, one shows that

$$\begin{aligned}\det(I + iM) = |m_2| e^{2\beta x} &\Bigg[\frac{2|m_2|\alpha}{\beta}\cos(2\beta x + x_0) \\ &- \frac{2m_1(1-2\beta)e^x}{4\alpha^2 + (1+2\beta)^2} \\ &\times \cos\{2\alpha t + \arg(m_2^*(\alpha + i(\beta + \tfrac{1}{2})))\}\Bigg] \\ &+ i|m_2| e^{2\beta x} \\ &\times \Bigg[2\sin\{2\alpha t + \arg(m_2^*(\alpha + i\beta))\} \\ &+ \frac{m_1}{|m_2|} e^{(1-2\beta)x} \\ &+ \frac{4m_1|m_2|\alpha^2(1-2\beta)^2}{\beta^2(4\alpha^2 + (1+2\beta)^2)^2} e^{(1+2\beta)x}\Bigg],\end{aligned}$$

where $\exp(x_0) = |m_2|\alpha/\beta$.

It is easy to verify from (4.7) and (4.8) that this solution actually is a kink, in the sense that $u \to \pm\pi$ as $x \to \pm\infty$, respectively. It is apparent from (4.8) that it oscillates in time, with frequency (2α), and that it is stationary in the sense that it has no mean motion. It therefore qualifies as a stationary wobbling kink, although it is not analogous to the ϕ^4 solution discussed in Sec. 2.

The energy of this solution, according to (4.4), is simply the sum of the energies of the stationary kink and the stationary breather: $H = 8 + 32\beta$, $0 < \beta < \frac{1}{2}$.

5. STABILITY

We have now established the existence of wobbling kinks, both for ϕ^4 and the sine-Gordon equations. To be of physical interest they should also be stable to perturbations, and this is a separate issue.

The stability of the wobbling kind for ϕ^4 theory seems to be an open question. An extremely limited form of stability can be established by the following generalization of the proof in Sec. 3.[13] Replace (1.3a) with the initial conditions

$$\hat{\phi} = \Phi_N + O(\epsilon^{N+1}), \quad \partial_t\hat{\phi} = \partial_t\Phi_N + O(\epsilon^{N+1}). \tag{5.1}$$

Then essentially the same proof given in Sec. 3 shows that the formal series for the wobbling kink is asymptotic to an exact solution of (1.4) over a long time scale, perhaps with an adjusted K_N in (3.13). In this form, the small terms in (5.1) constitute an arbitrary small perturbation of the initial conditions [i.e., "small" $= O(\epsilon^{N+1})$], and the argument in Sec. 3 establishes both existence and stability for the time scale in question. Unfortunately, the usual definition of stability has no such limitation on time scales, and this argument provides no information about longer times.

For the sine-Gordon equation, the following argument, due to David K. Campbell, indicates that the wobbling kinks are probably unstable. Each wobbling kink is composed of a kink and a breather. In order that these two modes not separate from each other in time, it is necessary that the three discrete eigenvalues in (4.5a) satisfy

$$|\zeta_1| = |\zeta_2| = |\zeta_3|, \tag{5.2}$$

where ζ_1 is purely imaginary and $\zeta_3 = -\zeta_2^*$. Consider initial data for the sine-Gordon equation consisting of a wobbling kink (4.7) plus a small perturbation. The wobbling kink is stable if and only if the discrete eigenvalues of the perturbed scattering data also satisfy (5.2). If the perturbation is real valued then $\zeta_3 = -\zeta_2^*$, both for the original and the perturbed scattering data. However, there seems to be no reason why $|\zeta_1| = |\zeta_2|$ should be preserved by an arbitrary small perturbation. If $|\zeta_1| \neq |\zeta_2|$ in the perturbed scattering data, then the breather and the kink will separate eventually, and the kink will cease to wobble. In this sense, the wobbling nature of the wobbling kink is probably unstable for the sine-Gordon equation. However, this instability is a rather mild one, inasmuch as it grows only algebraically rather than exponentially.

ACKNOWLEDGMENTS

The author is grateful to D. K. Campbell for repeatedly explaining the problem to him, and to both Campbell and Martin Kruskal for several valuable conversations.

This work was supported by the U. S. Army Reserach Office, under Contract #DAAG29-81-C-0009.

[1] M. J. Rice, Phys. Lett. A **71**, 152 (1979).
[2] D. K. Campbell and A. R. Bishop, Nucl. Phys. B **200**, 297 (1982).
[3] W. P. Su, J. R. Schrieffer, and A. J. Heeger, Phys. Rev. Lett. **42**, 1698 (1979).
[4] M. J. Rice and E. J. Mele, Solid State Comm. **35**, 487 (1980).
[5] D. K. Campbell, J. F. Schonfeld, and C. A. Wingate, "Resonance Structure in Kink-Antikink Interactions in ϕ^4 Theory," preprint.
[6] M. J. Ablowitz, D. J. Kaup, A. C. Newell, and H. Segur, Phys. Rev. Lett. **30**, 1262 (1973).
[7] M. J. Ablowitz and H. Segur, *Solitons and the Inverse Scattering Transform* (SIAM, Philadelphia, 1981).
[8] G. L. Lamb, *Elements of Soliton Theory* (Wiley-Interscience, New York, 1980).
[9] I. M. Gel'fand and B. M. Levitan, Amer. Math. Soc. Transl. Ser. 2, No. 1, p. 259 (1955).
[10] G. G. Stokes, Trans. Cambridge Philos. Soc. **8**, 841 (1847) [*Mathematics and Physics Papers* (Cambridge U. P., Cambridge, 1880), Vol. 1, p. 197. Now available from Johnson Reprint Services, New York].
[11] M. D. Kruskal, J. Math. Phys. **3**, 806 (1962).
[12] D. W. McLaughlin, J. Math. Phys. **16**, 96 (1975).
[13] This observation is due to Martin Kruskal.

Generalized linear inversion and the first Born theory for acoustic media

R. G. Keys and A. B. Weglein[a)]
Cities Service Company, Energy Resources Group, P.O. Box 3908, Tulsa, Oklahoma 74102

(Received 3 June 1982; accepted for publication 13 January 1983)

A procedure is derived which incorporates a generalized linear inverse viewpoint within a multidimensional Born inversion method. The method we present is a more general Born theory which can accommodate insufficient and inaccurate data. This general method reduces to the ordinary Born procedure when the data requirements of the latter technique are satisfied.

PACS numbers: 03.40.Kf, 43.40.Ph

I. INTRODUCTION

The inverse problem in wave propagation is to infer the characteristics of a medium from measurements of reflected or scattered waves. Two separate approaches to the inverse problem have evolved; both were derived from linearizations of the relationship between the medium parameters and the wave field. These two approaches are known as (1) the generalized linear inversion method, and (2) the first Born approximation. In this paper, the generalized linear inverse approach will be followed to construct a multidimensional Born inversion algorithm for the acoustic medium parameter.

The method of generalized linear inversion refers to that class of techniques which are designed to obtain estimates of model parameters from inexact or insufficient data. This approach was used, for example, by Backus and Gilbert[1] and Wiggins, Larner, and Wisecup,[2] and reviewed in Parker,[3] Sabatier,[4] and Aki and Richards.[5] The development of a generalized linear inversion algorithm is a two-step process. First, a linear relationship between the model parameters and the measured data is postulated. Then, using a selection criterion, a model is determined which yields the best fit to the measured data.

The method of generalized linear inversion has recently been applied to the one-dimensional seismic reflection problem by Cooke and Schneider.[6] The technique they present uses a linearized relationship between the acoustic wave field and the acoustic impedance of a one-dimensional medium. An inverse operator, which in their case will typically be a generalized inverse matrix, is required to solve for the variation of the parameters about some reference model. A brief summary of their procedure is given in Sec. II.

The first Born approximation for the acoustic wave field is a linear expression relating the wave field to an acoustic medium. The first Born approximation is obtained by replacing the unknown wave field in a Lippmann–Schwinger integral equation with the field which would exist under the same circumstances in the reference medium. In Sec. III, we present an overview of the Born theory approach to the acoustic inverse problem.

In the remaining sections of this paper, the generalized linear inversion approach will be applied to the acoustics problem to develop an inversion procedure. In Sec. IV, a linearization of the acoustic wave field with respect to the index of refraction of the medium will be derived. An optimal solution of the linearized problem will be determined in the following section. One of the results of this paper establishes a theoretical relationship between the generalized linear inversion method and the first Born approximation. This connection provides an avenue for transferring to the first Born theories some of the results which have been established for inaccurate and insufficient data in the generalized linear inversion problem. Consequently, an optimal Born model is derived which accommodates real world considerations, such as insufficient data and measurement errors. Furthermore, this more general Born result reduces to the ordinary Born inversion when the data requirements of the earlier theory are satisfied.

II. GENERALIZED LINEAR INVERSION

The problem of one-dimensional reflection seismology[7,8] is to determine the impedance profile from reflection seismic data. In a recent paper by Cooke and Schneider,[6] the generalized linear inverse approach to the one-dimensional seismic inverse problem is presented. They begin by assuming a Taylor series expansion of the acoustic wave field:

$$\psi(I) = \psi(I_0) + \frac{\partial \psi}{\partial I}(I_0)\cdot(I - I_0) + \frac{1}{2}\frac{\partial^2 \psi}{\partial I^2}(I_0)\cdot(I - I_0)^2 + \cdots,$$

where

I = the unknown impedance profile,
I_0 = the reference (or guessed) impedance profile,
$I - I_0$ = the error in the above guess,
$\psi(I)$ = the observed acoustic wave field,
$\psi(I_0)$ = the acoustic wave field generated by the impedance profile I_0,
$\frac{\partial \psi}{\partial I}(I_0)$ = the derivative operator.

The three steps involved in this generalized linear inverse approach are (1) the linearization of the Taylor series expansion, realized by truncating the series after the second term, (2) the discretization of the problem, which results in a finite order matrix for $\partial\psi(I_0)/\partial I$ and finite dimensional vectors for $\psi(I)$, $\psi(I_0)$, and $I - I_0$, and (3) a modified least-square error, generalized inverse technique, used to invert $\partial\psi/\partial I$ and solve for $I - I_0$.

[a)] Present address: Sohio Petroleum Company, One Lincoln Centre, Dallas, TX 75240.

 0022-2488/83/061444-06$02.50 © 1983 American Institute of Physics

Knowing I_0, and calculating ΔI provides a linear approximation for the impedance profile:

$$I = (I - I_0) + I_0 \approx \Delta I + I_0.$$

The technique presented by Cooke and Schneider can be applied iteratively by replacing I_0 with the new estimate for I and repeating steps 1–3.

The generalized linear inversion algorithm that will be derived later in this paper differs from the procedure developed by Cooke and Schneider in the following three ways. First, the proposed algorithm is a multidimensional technique. Second, the model parameter which will be determined is the index of refraction rather than the acoustic impedance. Third, the analysis is performed for the continuous case. There are no discretizations introduced into the problem; as a consequence, the optimum index of refraction is the solution of an integral equation.

III. BORN INVERSION THEORY

Within a given model, Born inversion consists of, first, linearizing an integral equation between the model parameters and the wave field; and second, solving the resulting Fredholm I integral equation for the model parameters. Linearization in this problem is accomplished by replacing, in the integral equation, the actual wave field with the corresponding wave field in a reference model. This linearization is equivalent to truncating the Born series[9] after the second term. In this paper, the expression "ordinary Born theory" refers to first Born approximations derived from general reference models, which are not necessarily constant background models.

We now present a brief review of the multidimensional Born theory as it pertains to the acoustic wave equation. The pressure field $p(\mathbf{x},t)$ in a constant density acoustic medium satisfies

$$\left(\nabla^2 - \frac{1}{c(\mathbf{x})^2}\frac{\partial^2}{\partial t^2}\right) p(\mathbf{x},t) = f(\mathbf{x},t),$$

where $f(\mathbf{x},t)$ is the source function, and $c(\mathbf{x})$ is the local acoustic velocity at the spatial coordinate $\mathbf{x}$. Let $P(\mathbf{x},\omega)$ and $F(\mathbf{x},\omega)$ denote the temporal Fourier transform of $p(\mathbf{x},t)$ and $F(\mathbf{x},t)$, respectively. Then

$$P(\mathbf{x},\omega) = \int_{\infty}^{\infty} p(\mathbf{x},t)e^{i\omega t}\,dt,$$

and

$$F(x,\omega) = \int_{-\infty}^{\infty} f(\mathbf{x},t)e^{i\omega t}\,dt.$$

$P(\mathbf{x},\omega)$ satisfies the differential equation

$$\left(\nabla^2 + \frac{\omega^2}{c(\mathbf{x})^2}\right) P(\mathbf{x},\omega) = F(\mathbf{x},\omega).$$

A Lippmann–Schwinger integral equation for P[9] which incorporates boundary conditions can be written as

$$P(\mathbf{x},\omega) = P_0(\mathbf{x},\omega) - \int G_0(\mathbf{x},\xi,\omega)\frac{\omega^2}{c_0(\xi)^2}\alpha(\xi)P(\xi,\omega)d\,\xi,$$

where P_0 is the solution of

$$\left(\nabla^2 + \frac{\omega^2}{c_0(\mathbf{x})^2}\right) P_0(\mathbf{x},\omega) = F(\mathbf{x},\omega),$$

and $c_0(\mathbf{x})$ and $\alpha(\mathbf{x})$ are the reference velocity and variation in the index of refraction, respectively; c_0 and α are related to the acoustic velocity c by the equation

$$\frac{1}{c(\mathbf{x})^2} = \frac{1}{c_0(\mathbf{x})^2}(1 + \alpha(\mathbf{x})).$$

$G_0(\mathbf{x},\xi,\omega)$ is the Green's function for the reference velocity problem

$$\left(\nabla^2 + \frac{\omega^2}{c_0(\mathbf{x})^2}\right) G_0(\mathbf{x},\xi,\omega) = \delta(\mathbf{x} - \xi).$$

The Born approximation consists of replacing $P(\mathbf{x},\omega)$ by $P_0(\mathbf{x},\omega)$ in the right-hand side of the Lippmann–Schwinger integral equation. The resulting equation is

$$P(\mathbf{x},\omega) = P_0(\mathbf{x},\omega) - \int G_0(\mathbf{x},\xi,\omega)\frac{\omega^2}{c_0(\xi)^2}\alpha(\xi)P_0(\xi,\omega)d\,\xi.$$

Born inversion involves solving this Fredholm I integral equation for $\alpha(\mathbf{x})$. For the multidimensional case, this problem has been addressed by a number of authors.[10–13]

IV. LINEARIZATION OF THE WAVE FIELD

The first step in developing linear inversion procedure is to find an approximation for the wave field which is linear with respect to the model parameter. It is convenient to choose the square of the index of refraction as the model parameter. This parameter, which we will denote by $m(\mathbf{x})$, is related to the acoustic velocity by the equation

$$m(\mathbf{x}) = V_0^2/c(\mathbf{x})^2,$$

where V_0 is a chosen constant reference velocity. The wave field associated with the model $m(\mathbf{x})$ satisfies the equation

$$[\nabla^2 + k^2 m(\mathbf{x})]\psi(\mathbf{x},k,m) = F(\mathbf{x},k) \quad \text{for} \quad \mathbf{x}\in\Omega, \tag{1}$$

where k is the wave number ($k = \omega/V_0$) and F is the source term. We will assume that Ω is a bounded domain.

Linearization of the wave field is accomplished by expanding the wave field in a generalized Taylor series about some reference model $m_0(\mathbf{x})$. The objective of the section is to derive a linear expression of the form

$$\psi_L(m) = \psi(m_0) + \psi'(m_0)\cdot(m - m_0). \tag{2}$$

In the above equation, ψ is the wave field; and $\psi_L(m)$ is the linearized approximation to $\psi(m)$ which becomes exact as m approaches m_0. $\psi'(m_0)$ is the directional derivative of ψ with respect to the model parameter.

In this section, a closed form expression for the directional derivative will be found. In addition, the connection between generalized linear inversion and the first Born theory will be made.

We begin by defining the set M of model parameters to be the space of complex-valued functions on Ω which are bounded almost everywhere, i.e., $M = L^{\infty}(\Omega)$. The norm for this space is the essential supremum, given by

$$\|m\|_{\infty} = \inf\{B : |m(\mathbf{x})| \leqslant B \text{ a.e. on } \Omega\} \quad \text{for } m\in M.$$

Suppose that for some $m_0(\mathbf{x})\in M$, the Green's function $G(\mathbf{x},\xi k,m_0)$, which satisfies

$$(\nabla^2 + k^2 m_0)G = \delta(\mathbf{x} - \xi), \quad \text{for } \mathbf{x},\xi\in\Omega,$$

with homogeneous boundary conditions, is unique. In the

following discussion, the wave number k is assumed to be real.

To ensure the existence of the directional derivative of the wave field, some constraint on the Green's function $G(\mathbf{x},\xi,k,m_0)$ is necessary. If the three-dimensional Green's function $G(\mathbf{x},\xi,k,m_0)$ has the form

$$G = \frac{\gamma(\mathbf{x},\xi,k,m_0)}{|\mathbf{x}-\xi|},$$

where γ is continuous for all $\mathbf{x}$ and ξ in the closure of Ω, then G is square integrable on Ω. The Green's function has this form, for example, when $m_0(\mathbf{x})$ is a constant. The constraint that will be imposed on the Green's function is that for a given k, there is some number C such that

$$\int_\Omega \int_\Omega |G(\mathbf{x},\xi,k,m_0)|^2 \, d\xi \, d\mathbf{x} < C^2. \tag{3}$$

In the following lemma, it will be shown that a neighborhood of m_0 exists for which the Lippmann–Schwinger equation corresponding to (1),

$$\psi(\mathbf{x},k,m) = \psi(\mathbf{x},k,m_0) - k^2 \int G(\mathbf{x},\xi,k,m_0)(m-m_0)(\xi)\psi(\xi,k,m)d\xi,$$

has a unique solution $\psi(\mathbf{x},k,m)$, provided that the Green's function $G(\mathbf{x},\xi k,m_0)$ satisfies the constraint (3), and also that the wave field $\psi(\mathbf{x},k,m_0)$ is square integrable on Ω.

Lemma 1: Let $m_0(\mathbf{x})\in M$. Suppose that for a fixed, real value of k, the problem

$$[\nabla^2 + k^2 m_0]G = \delta(\mathbf{x}-\xi),$$

with homogeneous boundary conditions, has a unique Green's function $G(\mathbf{x},\xi,k,m_0)$ for which there is some constant $C>0$ such that

$$\int_\Omega \int_\Omega |G(\mathbf{x},\xi,k,m_0)|^2 \, d\xi \, d\mathbf{x} < C^2.$$

In addition, suppose that the solution $\psi(\mathbf{x},k,m_0)$ of

$$[\nabla^2 + k^2 m_0]\psi = F,$$

with the specified boundary conditions, is square integrable on Ω.

Then the Lippmann–Schwinger equation,

$$\psi(\mathbf{x},k,m) = \psi(\mathbf{x},k,m_0) - k^2 \int_\Omega G(\mathbf{x},\xi,k,m_0)(m-m_0)\psi(\xi,k,m)\, d\xi,$$

has a unique solution for each m in some neighborhood of m_0.

Proof: Let $N(m_0,r) = \{m\in M: \|m-m_0\|_\infty < r\}$, where $r = \theta/[(k^2+1)C]$ and $0<\theta<1$. Let $m(\mathbf{x})\in N(m_0,r)$, and define a mapping T_m: on $L^2(\Omega)$ such that

$$T_m(\phi) = \psi(\mathbf{x},k,m_0) - k^2 \int_\Omega G(\mathbf{x},\xi,k,m_0)(m-m_0)(\xi)\phi(\xi,m)d\xi$$

for $\phi\in L^2(\Omega)$. We will show that T_m is contractive on $L^2(\Omega)$.

Let ϕ and ϕ' be contained in $L^2(\Omega)$. Then, by the Cauchy–Schwarz inequality,

$$\begin{aligned}\int_\Omega |T_m(\phi) - T_m(\phi')|^2 \, dx &= k^4 \int_\Omega \left| \int_\Omega G(\mathbf{x},\xi,k,m_0)(m-m_0)(\xi)(\phi-\phi')(\xi)d\xi \right|^2 d \\ &\leqslant k^4 \int_\Omega \int_\Omega |G(\mathbf{x},\xi,k,m_0)|^2 |m-m_0|^2 \, d\xi \, dx \\ &\quad \times \int_\Omega |\phi-\phi'|^2 \, d\xi.\end{aligned}$$

From condition (3), and the fact that $m\in N(m_0,r)$,

$$\begin{aligned}\int_\Omega |T_m(\phi) - T_m(\phi')|^2 \, d\mathbf{x} &\leqslant k^4 C^2 \|m-m_0\|_\infty^2 \, \|\phi-\phi'\|_2^2 \\ &\leqslant k^4 r^2 C^2 \|\phi-\phi'\|_2^2.\end{aligned}$$

Then, since $r = \theta/[(k^2+1)C]$, it follows that

$$\|T_m(\phi) - T_m(\phi')\|_2 \leqslant \theta \, \|\phi-\phi'\|_2, \quad \text{where } 0<\theta<1.$$

Therefore, T_m is a contractive map of $L^2(\Omega)$ into itself. This means that T_m has a unique fixed point $\psi(\mathbf{x},k,m)$ in $L^2(\Omega)$.[14] Thus, for each $m\in N(m_0,r)$, there is one and only one $\psi(\mathbf{x},k,m)$ such that

$$\psi(\mathbf{x},k,m) = \psi(\mathbf{x},k,m_0) - k^2 \int_\Omega G(\mathbf{x},\xi,k,m_0)(m-m_0)(\xi)\psi(\xi,k,m)d\xi.$$

As a corollary to the proof of Lemma 1, the wave field $\psi(\mathbf{x},k,m)$ is bounded in the L_2-norm for all $m\in N(m_0,r)$.

Corollary 1: Given the hypothesis of Lemma 1, there is some constant $K>0$ such that $\|\psi(\mathbf{x},k,m)\|_2 \leqslant K$ for all $m\in N(m_0,r)$.

Proof: Let $m\in N(m_0,r)$. Then $\|T_m(\psi(\mathbf{x},k,m)) - T_m(0)\|_2 \leqslant \theta \, \|\psi(\mathbf{x},k,m)\|_2$. Since $T_m(\psi(\mathbf{x},k,m)) = \psi(\mathbf{x},k,m)$ and $T_m(0) = \psi(\mathbf{x},k,m_0)$, then $\theta \, \|\psi(\mathbf{x},k,m)\|_2 \geqslant \|\psi(\mathbf{x},k,m) - \psi(\mathbf{x},k,m_0)\|_2 \geqslant \|\psi(\mathbf{x},k,m)\|_2 - \|\psi(\mathbf{x},k,m_0)\|_2$. Since $0<\theta<1$, then

$$\|\psi(\mathbf{x},k,m)\|_2 \leqslant \frac{1}{1-\theta} \, \|\psi(\mathbf{x},k,m_0)\|_2.$$

Corollary 2: Assuming the hypothesis of Lemma 1, the wave field $\psi(\mathbf{x},k,m)$ is continuous at m_0.

Proof: Let $m\in N(m_0,\epsilon)\subset N(m_0,r)$. From the Lippmann–Schwinger equation,

$$\begin{aligned}\int_\Omega |\psi(\mathbf{x},k,m) - \psi(\mathbf{x},k,m_0)|^2 \, d\mathbf{x} &= k^4 \int_\Omega \left| \int_\Omega G(\mathbf{x},\xi,k,m_0)(m-m_0)\,(\xi)\psi(\xi,k,m)d\xi \right|^2 d\mathbf{x} \\ &\leqslant k^4 \int_\Omega \int_\Omega |G(\mathbf{x},\xi,k,m_0)|^2 |m-m_0|^2 \, d\mathbf{x} \, d\xi \\ &\quad \times \int_\Omega |\psi(\xi,k,m)|^2 \, d\xi \\ &\leqslant k^4 C^2 \|m-m_0\|_\infty^2 \, \|\psi(\mathbf{x},k,m)\|_2^2 \leqslant k^4 C^2 \epsilon^2 K^2.\end{aligned}$$

Then $\|\psi(\mathbf{x},k,m) - \psi(\mathbf{x},k,m_0)\|_2 \leqslant k^2 CK\epsilon$. As ϵ goes to zero, $\psi(\mathbf{x},k,m)$ approaches $\psi(\mathbf{x},k,m_0)$, which shows that $\psi(\mathbf{x},k,m)$ is continuous at m_0.

In this last result, we will obtain the directional derivative of the wave field ψ at m_0.

Corollary 3: If the hypothesis of Lemma 1 holds, the derivative of the wave field ψ at m_0 in the direction Δm is

$$\psi'(\mathbf{x},k,m_0)\cdot\Delta m = \lim_{s\to 0}\frac{\psi(\mathbf{x},k,m_0+s\Delta m)-\psi(\mathbf{x},k,m_0)}{s}$$

$$= -k^2\int_\Omega G(\mathbf{x},\boldsymbol{\xi},k,m_0)\Delta m(\boldsymbol{\xi})\psi(\boldsymbol{\xi},k,m_0)d\boldsymbol{\xi}.$$

Proof: Let Δm be any nonzero vector in M. Let s be a scalar such that $0<|s|<r/\|\Delta m\|_\infty$. Then $m_0(\mathbf{x})+s\Delta m(\mathbf{x})$ is in $N(m_0,r)$. From Lemma 1,

$$\int_\Omega\left|\frac{\psi(\mathbf{x},k,m_0+s\Delta m)-\psi(\mathbf{x},k,m)}{s}+k^2\int_\Omega G(\mathbf{x},\boldsymbol{\xi},k,m_0)\Delta m(\boldsymbol{\xi})\psi(\boldsymbol{\xi},k,m_0)d\boldsymbol{\xi}\right|^2 d\mathbf{x}$$

$$=\int_\Omega\left|k^2\int_\Omega G(\mathbf{x},\boldsymbol{\xi},k,m_0)\Delta m(\boldsymbol{\xi})[\psi(\boldsymbol{\xi},k,m_0+s\Delta m)-\psi(\boldsymbol{\xi},k,m_0)]d\boldsymbol{\xi}\right|^2 d\mathbf{x}$$

$$\leqslant k^4\|\Delta m\|_\infty^2\int_\Omega\int_\Omega|G(\mathbf{x},\boldsymbol{\xi},k,m_0)|^2\,d\boldsymbol{\xi}\,d\mathbf{x}\int_\Omega|\psi(\boldsymbol{\xi},k,m_0+s\Delta m)-\psi(\boldsymbol{\xi},k,m_0)|^2\,d\boldsymbol{\xi}$$

$$\leqslant k^4\|\Delta m\|_\infty^2\,C^2\|\psi(\mathbf{x},k,m_0+s\Delta m)-\psi(\mathbf{x},k,m_0)\|_2^2.$$

From Corollary 2, ψ is continuous at m_0. Therefore, as s goes to zero, the right-hand side of the above inequality must vanish. Thus the directional derivative exists and is given by $\psi'(m_0)\cdot\Delta m = -k^2\int_\Omega G(\mathbf{x},\boldsymbol{\xi},k,m_0)\Delta m(\boldsymbol{\xi})\psi(\boldsymbol{\xi},k,m_0)d\boldsymbol{\xi}$.

Substituting the expression for the derivative into Eq. (2) yields the linearized approximation to the acoustic wave field:

$$\psi_L(\mathbf{x},k,m)=\psi(\mathbf{x},k,m_0)$$

$$-k^2\int_\Omega G(\mathbf{x},\boldsymbol{\xi},k,m_0)(m-m_0)(\boldsymbol{\xi})\psi(\boldsymbol{\xi},k,m_0)d\boldsymbol{\xi}. \tag{4}$$

The field $\psi(\mathbf{x},k,m_0)$ is the wave field in the reference medium; $G(\mathbf{x},\boldsymbol{\xi},k,m_0)$ is the associated Green's function. The right-hand side of Eq. (4) is identical to the Lippmann–Schwinger equation for $\psi(\mathbf{x},k,m)$, except that the field $\psi(\boldsymbol{\xi},k,m)$ inside the integral has been replaced by the reference wave field $\psi(\boldsymbol{\xi},k,m_0)$. Thus, the right-hand side of Eq. (4) is equivalent to the first Born approximation.

In the event that inversion of Eq. (4) is possible, generalized linear inversion and the first Born theory produce the same result. When the measured data are imperfect due to, for example, an insufficient measurement set or measurement errors, generalized inversion can be employed to obtain an optimal solution, i.e., a solution which in some sense provides the best match to the data. The calculation of such optimal solutions is described in the next section.

V. A LEAST-SQUARES SOLUTION FOR THE LINEARIZED PROBLEM

In this section, we consider the problem of estimating the acoustic model parameter $m(\mathbf{x})$ from observations of the wave field on a surface S contained in Ω. This problem is often complicated by presence of noise in the measured wave field; further difficulties arise when the extent of the measurement surface S is insufficient to guarantee a unique solution. Our objective is to find an index of refraction which minimizes the "difference" between the observed data and the linearized wave field.

Let $\psi_D(\mathbf{x},k)$ denote the measured data. $\psi_D(\mathbf{x},k)$ is defined only on the surface S, whereas the wave field is defined for all $\mathbf{x}$ in Ω. Let $\langle *,*\rangle_k$ be the inner product on the wave number space such that $\langle\phi,\psi\rangle_k=\int_{-\infty}^{\infty}\phi(k)\psi(k)^*W(k)dk$. The positive weight function $W(k)$ is necessary to guarantee convergence of the integral; W can be chosen based on the bandwidth of the data.

Let M_0 be the space of models which are square integrable on Ω, and let m_0 be a reference model. Finally, let $E(m)$ be the quadratic functional defined on M_0 by

$$E(m)=\int_S\|\psi_D(\mathbf{x},k)-\psi_L(\mathbf{x},k,m)\|_k^2\,d\mathbf{x}$$

$$+\lambda\int_\Omega|m(\mathbf{x})-m_0(\mathbf{x})|^2\,d\mathbf{x}, \tag{5}$$

where

$$\psi_L(\mathbf{x},k,m)=\psi(\mathbf{x},k,m_0)+\psi'(\mathbf{x},k,m_0)\cdot(m-m_0)$$

and λ is a nonnegative scalar. The norm inside the integral is the norm induced by the wave number inner product.

The quadratic functional E is a criterion for selecting the optimum model corresponding to a given set of measurements. The first term on the right-hand side of Eq. (5) represents the error between the linearized wave field and the measured data. The second term is a restriction on the size of the variation from the reference model. This restriction is consistent with the Born theory idea of searching for small variations in the model. The parameter λ controls the weight given to each of these two terms.

The model which minimizes E is found by substituting for ψ_L in Eq. (5). Then

$$E(m)=\lambda\int_\Omega|(m-m_0)(\boldsymbol{\xi})|^2\,d\boldsymbol{\xi}$$

$$+\int_S\|\psi_D(\mathbf{x},k)-\psi(\mathbf{x},k,m_0)$$

$$+k^2\int_\Omega G(\mathbf{x},\boldsymbol{\xi},k,m_0)\psi(\boldsymbol{\xi},k,m_0)(m-m_0)(\boldsymbol{\xi})d\boldsymbol{\xi}\|_k^2\,d\mathbf{x}.$$

Let $g(\boldsymbol{\xi})=-\int_S\langle\psi_D(\mathbf{x},k)-\psi(\mathbf{x},k,m_0),$
$k^2G(\mathbf{x},\boldsymbol{\xi},k,m_0)\psi(\boldsymbol{\xi},k,m_0)\rangle_k\,d\mathbf{x}$ and let $r(\boldsymbol{\xi},\boldsymbol{\eta})$ be the kernel:

$$r(\xi,\eta) = \int_S \langle k^2 G(\mathbf{x},\xi,k,m_0)\psi(\xi,k,m_0), k^2 G(\mathbf{x},\eta,k,m_0)\psi(\eta,k,m_0)\rangle_k \, d\mathbf{x}.$$

The kernel $r(\xi,\eta)$ is symmetric and positive semidefinite since for any function $m(\mathbf{x})$,

$$\int_\Omega \int_\Omega m(\eta)^* r(\xi,\eta) m(\xi) d\eta \, d\xi = \int_S \|k^2 \int_\Omega G(\mathbf{x},\xi,k,m_0)\psi(\xi,k,m_0) m(\xi) d\xi\|_k^2 \, d\mathbf{x} \geqslant 0, \quad (6)$$

provided the integrals are defined. The quadratic functional $E(m)$ can be written as

$$E(m) = \int_S \|\psi_D(\mathbf{x},k) - \psi(\mathbf{x},k,m_0)\|_k^2 \, d\mathbf{x} - \int_\Omega (m - m_0)^*(\xi) g(\xi) d\xi - \int_\Omega (m - m_0)(\xi) g(\xi)^* \, d\xi + \int_\Omega \int_\Omega (m - m_0)^*(\xi) r(\xi,\eta)(m - m_0)(\eta) d\xi \, d\eta + \lambda \int_\Omega |(m - m_0)(\xi)|^2 d\xi.$$

Suppose that $\tilde{m}(\mathbf{x})$ is the solution of the integral equation

$$\lambda(\tilde{m} - m_0)(\mathbf{x}) + \int_\Omega r(\mathbf{x},\xi)(\tilde{m} - m_0)(\xi) d\xi = g(\mathbf{x}) \quad \text{for } \mathbf{x}\in\Omega.$$

Let $\delta m = m - \tilde{m}$. Then, after replacing m by $\tilde{m} + \delta m$ in the equation for $E(m)$ and using the fact that $\tilde{m}$ is the solution of the above integral equation, it follows that

$$E(m) = E(\tilde{m}) + \int_\Omega \int_\Omega \delta m^*(\xi) r(\xi,\eta) \delta m(\eta) d\xi \, d\eta + \lambda \int_\Omega |\delta m(\mathbf{x})|^2 d\mathbf{x}.$$

Since $r(\xi,\eta)$ is positive semidefinite, it follows that $E(m) \geqslant E(\tilde{m})$ for all $m\in M_0$. If λ is positive, then strict inequality holds unless $\delta m(\mathbf{x}) = 0$, almost everywhere, or $m(\mathbf{x}) = \tilde{m}(\mathbf{x})$ almost everywhere. Thus, the solution of the Fredholm integral equation

$$\lambda(m - m_0)(\mathbf{x}) + \int_\Omega r(\mathbf{x},\xi)(m - m_0)(\xi) d\xi = g(\mathbf{x}) \quad (7)$$

minimizes E over the set of models M_0.

We now address the questions of (1) uniqueness and (2) the relationship of this solution to the ordinary Born theory. With regard to the question of uniqueness, let m_1 and m_2 both solve Eq. (7). The difference of these solutions, $m_1 - m_2$, then satisfies

$$\lambda(m_1 - m_2)(\xi) + \int_\Omega r(\xi,\eta)(m_1 - m_2)(\eta) d\eta = 0. \quad (8)$$

Multiplying Eq. (8) by $(m_1 - m_2)(\xi)$ and integrating over ξ, we find that $m_1 - m_2$ satisfies

$$\lambda \int_\Omega |(m_1 - m_2)(\xi)|^2 d\xi + \int_S \|k^2 \int_\Omega G(\mathbf{x},\xi,k,m_0)\psi(\xi,k,m_0) \times (m_1 - m_2)(\xi) d\xi\|_k^2 \, d\mathbf{x} = 0. \quad (9)$$

When $\lambda = 0$, Eq. (9) reduces to

$$\int_S \|k^2 \int_\Omega G(\mathbf{x},\xi,k,m_0)\psi(\xi,k,m_0)(m_1 - m_2)(\xi) d\xi\|_k^2 \, d\mathbf{x} = 0,$$

which implies that $\int_\Omega G(\mathbf{x},\xi,k,m_0)\psi(\xi,k,m_0)(m_1 - m_2)(\xi) \times d\xi = 0$ for all $\mathbf{x}\in S$ and $-\infty < k < \infty$.

Suppose S is a set such that if

$$\int_\Omega G(\mathbf{x},\xi,k,m_0)\psi(\xi,k,m_0)(m_1 - m_2)(\xi) d\xi = 0$$

for every $x\in S$ and for all k, then $(m_1 - m_2)(x) = 0$ for all $\mathbf{x}\in\Omega$. When the measurement set S satisfies this condition,[11,13] then the solution of Eq. (7), with $\lambda = 0$, is unique.

Now the quadratic functional $E(m)$, with $\lambda = 0$, will be minimized when m is the solution of

$$\psi_D(\mathbf{x},k) - \psi(\mathbf{x},k,m_0) + k^2 \int_\Omega G(\mathbf{x},\xi,k,m_0)\psi(\xi,k,m_0)(m - m_0)(\xi) d\xi = 0.$$

The model $m(\mathbf{x})$ which satisfies this equation is the ordinary Born theory approximation. Thus, when the solution of Eq. (7), with $\lambda = 0$, is unique, then our procedure must produce the ordinary Born solution. This shows how the generalized linear inversion method derived by minimizing $E(m)$ over the space of models, reduces to the ordinary Born theory when the measurement set S is sufficient to guarantee a unique result.

Equation (9), for $\lambda > 0$, has a unique solution for the case in which the measurement set S is not sufficient to guarantee a unique ordinary Born solution.

When the measurement S is not sufficient to guarantee a unique solution, then we may take λ to be positive. If $\lambda > 0$, the left-hand side of Eq. (9) will vanish only for $m_1 - m_2$.

VI. CONCLUSIONS

We have derived a procedure which incorporates a generalized linear inverse method in a multidimensional Born theory. The linearization for our procedure was accomplished by truncating a Taylor series expansion about a reference model parameter; linearization in the Born theory approach is achieved by truncating a Born series expansion about a reference wave field. Both the truncated Taylor series and the truncated Born series produce the same linear equation for the acoustic wave field in terms of the model parameter.

By introducing a minimization criterion, a more general Born approximation was derived which can accommodate insufficient and inaccurate data. This general method reduces to the ordinary Born procedure when the data requirements of the earlier procedure are satisfied.

[1]G. E. Backus and F. J. Gilbert, Philos. Trans. R. Soc. London **266**, 123 (1970).
[2]R. A. Wiggins, K. L. Larner, and R. D. Wisecup, Geophysics **41**(5), 922 (1976).
[3]R. L. Parker, Ann. Rev. Earth Planet Sci. **5**, 35 (1977).
[4]P. C. Sabatier, J. Geophys. **43**, 115 (1977).
[5]K. Aki and P. G. Richards, *Quantitative Seismology Theory and Methods*, Vol. 2 (Freeman, San Francisco, 1980).
[6]D. A. Cooke and W. A. Schneider, paper presented at the annual SEG meeting, Los Angeles, California, 1981.
[7]J. A. Ware and K. Aki, J. Acoust. Soc. Am. **5**, 911 (1969).
[8]R. G. Newton, Geophys. J. R. Astron. Soc. **65**, 191 (1981).
[9]C. J. Joachain, *Quantum Collision Theory* (North-Holland, Amsterdam, 1975).
[10]J. K. Cohen and N. Bleistein, Geophysics **44**, 1077 (1979).
[11]R. W. Clayton and R. H. Stolt, Geophysics **46**(11), 1559 (1981).
[12]S. Raz, Geophysics **46**(6), 837 (1981).
[13]A. B. Weglein, J. Acoust. Soc. Am. **71**, 1179 (1982).
[14]E. K. Blum, *Numerical Analysis and Computation: Theory and Practice* (Addison-Wesley, Reading, MA, 1972).

On the Radon–Nikodym property for σ-classes

Mirko Navara and Pavel Pták
Technical University of Prague, 166 27 Prague 6, Suchbátarova 2, Czechoslovakia

(Received 21 June 1982; accepted for publication 30 July 1982)

We show that a σ-class with the Radon–Nikodym property need not be a Boolean σ-algebra. This disproves the conjecture of Gudder and Zerbe.

PACS numbers: 03.65. – w

In Ref. 1, the authors introduce the Radon–Nikodym property for quantum logics. They show among other things that a "separable" σ-class possesses the Radon–Nikodym property if and only if the σ-class is a Boolean σ-algebra. The latter result does not hold true in general as we shall show in this paper.

Let us recall only the notions we shall need in the sequel (see Ref. 1 for a more detailed exposition). A σ-class is a pair (Ω,C), where C is a nonvoid collection of subsets of Ω closed under the formations of complements and countable disjoint unions. A measure on (Ω,C) is a nonnegative σ-additive real-valued function on C. If μ,ν are two measures on (Ω,C) and if $\mu(P)=0$ implies that $\nu(P)=0$, then ν is called absolutely continuous relative to μ (in symbols: $\nu\ll\mu$). A central observable on (Ω,C) is a function $f:\Omega\to R$ such that $f^{-1}(u,v)\cap P\in C$ for any interval $(u,v)\subseteq R$ and any set $P\in C$.

Let us now state the basic definition (see Ref. 1). A σ-class (Ω,C) is said to have the Radon–Nikodym property if any such couple μ,ν of measures on (Ω,C), $\nu\ll\mu$, admits the expression $\nu(P)=\int_P f d\mu$ for a central observable $f:\Omega\to R$. The question is if there is a non-Boolean σ-class with the Radon–Nikodym property. We are going to construct such an example.

Put $\Omega=I\times M$, where I is the unit interval $[0,1]$ and $M=\{a,b,c,d\}$. The collection C for the desired σ-class (Ω,C) is now defined by the following requirement. A set P, $P\subseteq\Omega$, belongs to C if and only if the set

$$S(P)=\{r\in I:(\{r\}\times M)\cap P \text{ has either exactly one or exactly three elements}\}$$

is at most countable. Our result is established once we prove the following two propositions.

Proposition 1: The couple (Ω,C) is a σ-class which is not a σ-algebra.

Proof: Obviously, $\emptyset\in C$. If $P\in C$ then $S(P)=S(\Omega-P)$. Finally, if $P_i\in C, P_i\cap P_j=\emptyset$, then $S(\cup_{i=1}^{\infty}P_i)\subseteq\cup_{i=1}^{\infty}S(P_i)$. It follows that (Ω,C) is a σ-class. On the other hand, (Ω,C) is not a σ-algebra since

$$(I\times\{a,b\})\cap(I\times\{b,c\})\notin C.$$

Proposition 2: If μ,ν are measures on (Ω,C), $\nu\ll\mu$, μ,σ-finite, then $\nu(P)=\int_P f\,d\mu$ for a central observable f on (Ω,C).

Proof: Denote by D the subset of C such that P belongs to D if and only if the set

$$V(P)=\{r\in I:(\{r\}\times M)\cap P \text{ has at least one and at most three elements}\}$$

is at most countable. We claim that (Ω,D) is a sub-σ-algebra of (Ω,C) and, moreover, D is the center of C, i.e., D consists of all sets of the form $g^{-1}(u,v)$, where g is a central observable on (Ω,C). The first statement is obvious. As for the second, if $g^{-1}(u,v)=P\in C-D$ then $V(P)$ is uncountable and $S(P)$ is countable. Then the set $J=\{r\in I:(\{r\}\times M)\cap P$ has exactly two elements$\}$ is uncountable. We can find a set $Q\subseteq J\times M$ such that for all $r\in J$ the set $(\{r\}\times M)\cap Q$ has exactly two elements and the set $(\{r\}\times M)\cap Q\cap P$ has exactly one element. Then $Q\in C$ and $P\cap Q\notin C$.

Let us now restrict μ, ν to the σ-algebra (Ω,D) and denote the restriction by μ_D,ν_D. By the classical Radon–Nikodym theorem, there exists a D-measurable function f such that $\nu_D(P)=\int_P f d\mu_D$ for any $P\in D$. We want to show that $\nu(P)=\int_P f d\mu$ for any $P\in C$. Observe first that μ_D (and therefore also ν_D) is carried by a countable subset of Ω. Indeed, since D contains the σ-algebra of all sets of the type $J\times M, J\subseteq I$, the measure μ is in fact defined on all subsets of I and therefore (see, e.g., Ref. 2) must be carried by a set $K\times M, K$ countable. Further, since any singleton belongs to D, the measure μ_D (and therefore also ν_D) can be written in the form $\mu_D=\Sigma_{i=1}^{\infty}p_i d_{\omega_i}$, where p_i are nonnegative numbers and d_{ω_i} are measures concentrated in points $\omega_i\in\Omega$. We need to show now that the only extension of $\nu_D=\Sigma_{i=1}^{\infty}q_i d_{\omega_i}$ over the entire σ-class C is the measure $\gamma=\Sigma_{i=1}^{\infty}q_i d_{\omega_i}$. Since $\nu(P)=\int_P f d\mu$ is such an extension, we obtain that $\nu=\gamma$.

Put $T=\cup_{i=1}^{\infty}\{\omega_i\}$ and consider $\nu(P)$ for a set $P\in C$. We have

$$\nu(P-T)+\nu((\Omega-P)-T)=\nu(\Omega-T)=0,$$

and therefore

$$\nu(P)=\nu(P\cap T)+\nu(P-T)=\nu(P\cap T),$$

where $P\cap T\in D$. This shows that any extension of ν_D over C must be of the desired type and the proof is complete.

[1] S. Gudder and J. Zerbe, "Generalized monotone convergence and Radon–Nikodym theorems," J. Math. Phys. **22**, 2553–61 (1981).
[2] S. Ulam, "Zur Mässtheorie in der algemeinen Mengenlehre," Fund. Math. **16**, 140–150 (1930).

 0022-2488/83/061450-01$02.50 © 1983 American Institute of Physics

Compatibility of the uncorrelated joint distribution with the classical limit

Roderick I. Sutherland
Department of Theoretical Physics, University of New South Wales, Sydney, Australia

(Received 8 April 1982; accepted for publication 20 August 1982)

It is shown that the uncorrelated joint distribution, when adopted for quantum mechanics, is not (despite initial appearances to the contrary) in conflict with the known correlations between position and momentum in the classical limit. It is therefore not excluded on these grounds as a possible joint distribution expression for quantum mechanics.

PACS numbers: 03.65.Bz

1. INTRODUCTION

In a previous paper[1] we have considered the question of a joint probability distribution $P(x, p;t)$ for position x and momentum p (at time t) in quantum mechanics.[2] For an ensemble of particles described by the wavefunction $\Psi(x;t)$, the joint distribution is required to satisfy the equations

$$\int_{-\infty}^{+\infty} P(x, p;t)\, dp = |\Psi(x;t)|^2,$$
$$\int_{-\infty}^{+\infty} P(x, p;t)\, dx = |\phi(p;t)|^2, \tag{1}$$

where $\phi(p;t)$ is the momentum wavefunction obtained by taking the Fourier transform of $\Psi(x;t)$. A number of people have looked at the question of a joint distribution for quantum mechanics[3–13] and various possible expressions satisfying Eqs. (1) have been constructed.[3,5–7] The simplest of these have been listed in Ref. 1. Most of the possibilities found so far have the failing that they are not positive definite and hence cannot be true probability distributions. The only joint distribution known to date which is positive definite is the "uncorrelated" distribution

$$P(x, p;t) = |\Psi(x;t)\phi(p;t)|^2. \tag{2}$$

This expression seems, however, to be open to the immediate criticism that it is not compatible with the correlations known to exist in the classical limit. It is the aim of the following considerations to demonstrate that, in fact, the uncorrelated joint distribution is not ruled out on these grounds.

2. ARGUMENT AGAINST THE UNCORRELATED DISTRIBUTION

In many situations at the classical level, the position x and the momentum p of a particle are correlated. For example, the motion of a particle in one dimension under the influence of a potential $V(x)$ is governed by the equation

$$p^2/2m + V(x) = E,$$

where m and E are the particle's mass and energy, respectively. This equation can be rewritten as

$$p(x) = \sqrt{2m(E - V(x))}, \tag{3}$$

which determines the momentum value corresponding to each position x. The probability distributions for the position and momentum of a particle in a potential $V(x)$ are thus strongly correlated—for a particular position x, the probability for the momentum value $p(x)$ given by (3) is unity and the probability for any other momentum value is zero.

Now, one may raise an objection against the joint distribution in (2) as follows. The way in which the momentum probability distribution depends on the position x is given by the conditional probability $P(p|x;t)$. Using the general relationship

$$P(x, p;t) = P(p|x;t)P(x;t),$$

we obtain from (2)

$$P(p|x;t) = P(p;t) = |\phi(p;t)|^2. \tag{4}$$

Since the right-hand side of (4) is not a function of x, the probability of a particular momentum value being found is independent of the particle's position. Since this result is quite general and holds for the classical limit, the introduction of the joint distribution given in (2) appears to conflict with the known classical correlations between x and p discussed above.

3. ALTERNATIVE VIEWPOINT

By adopting another approach, however, we obtain a different conclusion. It is well known that the mathematics of conventional quantum mechanics without any joint distribution is compatible with the classical limit. In particular, the correct correlation between x and p for each potential $V(x)$ arises in the relationship between the mean position $\langle x \rangle$ and the mean momentum $\langle p \rangle$. Now, the joint distribution is simply introduced as a supplement to the mathematical structure without any change in the existing equations. Hence the introduction of a joint distribution will not affect the consistency of the theory with classical mechanics.

We have above two apparently reasonable arguments which seem to imply opposite conclusions. We will now present the second argument in more detail and thence show why the uncorrelated joint distribution is not ruled out by the first argument.

Heisenberg's uncertainty relation

$$\Delta x \Delta p \gtrsim \hbar \tag{5}$$

implies that it is not possible to make the position distribution $|\Psi(x;t)|^2$ very narrow without broadening the corresponding moment distribution $|\phi(p;t)|^2$. This, however, does not conflict with the sharp trajectories of macroscopic particles since the quantity observed directly in the macroscopic world is velocity v rather than momentum and, for a particle

 0022-2488/83/061451-02$02.50 © 1983 American Institute of Physics

whose mass m is large, the distributions for position and velocity can be made simultaneously sharp, as is seen by putting $\Delta p = m\Delta v$ so as to obtain

$$\Delta x \Delta v \gtrsim \hbar/m. \tag{6}$$

It is, therefore, possible to construct a wavefunction which (as is necessary to describe the motion of a macroscopic particle) is highly localized and does not spread appreciably with time.

The classical and quantum mechanical modes of description are connected by the fact (Ehrenfest's theorem) that any relationship between x and v contained in the classical description also holds between the mean values $\langle x\rangle_t$ and $\langle v\rangle_t$ (at time t) calculated from the wavefunction in the quantum mechanical description. Some correlation between position and velocity is thus already contained in the wavefunction. If, in describing a macroscopic particle, we choose a wavefunction which is very narrow and has a negligible rate of spreading, then the only x and v values which, at any instant of time, have significant probability are those very close to $\langle x\rangle_t$ and $\langle v\rangle_t$, respectively. Hence, in this case, any relationship between $\langle x\rangle_t$ and $\langle v\rangle_t$ for an ensemble becomes, to a very good approximation, a relationship between the values x and v measured for a single particle, and so the correct classical limit is obtained. The required classical correlations have come from the form of the wavefunction itself rather than from any joint distribution.

4. RECONCILIATION OF CLASSICAL CORRELATIONS WITH THE UNCORRELATED DISTRIBUTION

We will now show how the above reasoning is reconciled with the argument against the uncorrelated joint distribution. If m is sufficiently large [see Eq. (6)], it is possible for both the x and v distributions corresponding to a particular wavefunction to be so narrow that they may be approximated by Dirac delta functions

$$P(x;t) \simeq \delta[x - \langle x\rangle_t],$$

$$P(v;t) \simeq \delta[v - \langle v\rangle_t].$$

Using the equations

$$P(p;t)\,dp = P(v;t)\,dv,$$

and

$$P(x,p;t)\,dp = P(x,v,t)\,dv$$

together with (2), we obtain

$$P(x,v;t) = P(x;t)P(v;t),$$

so that we have

$$P(x,v;t) \simeq \delta[x - \langle x\rangle_t]\delta[v - \langle v\rangle_t]. \tag{7}$$

The corresponding conditional probabilities

$$P(v|x;t) \simeq \delta[v - \langle v\rangle_t],$$

and

$$P(x|v;t) \simeq \delta[x - \langle x\rangle_t]$$

are independent of x and v, respectively. This means the joint distribution given by (7) is uncorrelated in the sense that the probability of a particle having a particular velocity value is not explicitly dependent on the particle's position and that the probability of the particle having a particular position is not explicitly dependent on its velocity. The joint distribution is, nevertheless, consistent with macroscopic observations because *the single x value and the single v value to which it assigns a nonzero probability at each instant of time satisfy the appropriate classical relationship between position and velocity*.

Most wavefunctions do not reduce to precise position and velocity values as we approach the classical domain. For example, solutions to the time-independent Schrödinger equation for simple potentials do not become sharply localized in the classical limit, and hence cannot provide a correlation between position and velocity for macroscopic particles. In examining a macroscopic particle to observe a correlation, however, we must perform position and velocity measurements involving minimal disruption of the particle's state. These measurement will change the initial wavefunction into a sharply localized, time-dependent wavefunction which will then contain the desired correlation.

The above discussion has been particularly concerned with the uncorrelated joint distribution. The same argument can, however, be used to demonstrate the compatibility with the known classical correlations of any joint distribution satisfying Eq. (1).

[1] R. I. Sutherland, J. Math. Phys. **23**, 2389 (1982).
[2] Joint distributions for quantum mechanics are usually discussed in one dimension. They can be readily generalized to three dimensions.
[3] E. P. Wigner, Phys. Rev. **40**, 749 (1932).
[4] J. E. Moyal, Proc. Cambridge Phil. Soc. **45**, 99 (1949).
[5] H. Margenau and R. N. Hill, Prog. Theor. Phys. **26**, 722 (1961).
[6] C. L. Mehta, J. Math. Phys. **5**, 677 (1964).
[7] L. Cohen, J. Math. Phys. **7**, 781 (1966).
[8] H. Margenau and L. Cohen in *Quantum Theory and Reality*, edited by M. Bunge (Springer Verlag, New York, 1967).
[9] L. E. Ballentine, Rev. Mod. Phys. **42**, 372 (1970).
[10] E. P. Wigner, in *Perspectives in Quantum Theory*, edited by W. Yourgrau and A. van der Merwe (M.I.T. Press, Cambridge, MA, 1971).
[11] L. Cohen, in *Contemporary Research in the Foundations and Philosophy of Quantum Theory*, edited by C. A. Hooker (Reidel, Dordrecht, Holland, 1973).
[12] N. Cartwright, Found. Phys. **4**, 127 (1974).
[13] L. Cohen, Found. Phys. **6**, 739 (1976).

Continuity statements and counterintuitive examples in connection with Weyl quantization

Ingrid Daubechies[a)]
Physics Department, Princeton University, Princeton, New Jersey 08544

(Received 11 March 1982; accepted for publication 27 August 1982)

We use the properties of an integral transform relating a classical function f with the matrix elements between coherent states of its quantal counterpart Qf, to derive continuity properties of the Weyl transform from classes of distributions to classes of quadratic forms. We also give examples of pathological behavior of the Weyl transform with respect to other topologies (e.g., bounded functions leading to unbounded operators).

PACS numbers: 03.65.Ca, 02.30. + g

I. INTRODUCTION

The Weyl correspondence or Weyl transform defines a map from the functions on the classical phase space to the operators on the quantum mechanical Hilbert space.[1]

The classical phase space is here a 2ν-dimensional real vector space E, equipped with a nondegenerate symplectic form σ. It is customary to consider E as the direct sum of position and momentum spaces:

$$E \ni v = (q, p), \quad q, p \in \mathbb{R}^{\nu},$$

with σ of the form

$$\sigma((q, p), (q', p')) = \tfrac{1}{2}(p \cdot q' - q \cdot p').$$

The Hilbert space $\mathscr{H}$ is a complex Hilbert space carrying an irreducible representation of the canonical commutation relations. Explicitly, $\exists\ W(v)$, unitary operators on $\mathscr{H}$, labeled by the points v of E, such that

$$\text{s-}\lim_{v \to 0} W(v) = \mathbb{1}_{\mathscr{H}},$$

$$W(v_1)W(v_2) = e^{i\sigma(v_1, v_2)}W(v_1 + v_2).$$

One usually writes the $W(v)$ as

$$W(p, q) = \exp[i(p \cdot Q - q \cdot P)] = \exp(-\tfrac{1}{2} ipq)\exp(ip \cdot Q)\exp(-iq \cdot P),$$

where Q_j (P_j) are the generators of the ν-parameter group $W(0, p)$ ($W(-q, 0)$); Q_j and P_j are called the position and momentum operators, respectively, and satisfy the usual commutation relations

$$[Q_j, P_k] = i\delta_{jk}\mathbb{1}$$

on a common core.

The Weyl transform of a function f on E is then an operator Qf defined by (formally)

$$Qf = 2^{-\nu}\int_E dv\ \tilde{f}(v)W(-v/2), \tag{1}$$

where $\tilde{f}$ is the symplectic Fourier transform of f:

$$\tilde{f}(v) = 2^{-\nu}\int dv'\ e^{i\sigma(v, v')}f(v').$$

One can check[1] that Eq. (1) defines an unambiguous extension ("symmetric ordering") of the usual correspondence

$$f_j(q, p) = q_j \Rightarrow Qf_j = Q_j,$$

$$g_j(q, p) = p_j \Rightarrow Qg_j = P_j.$$

To give a precise sense to Eq. (1), one has to specify in what sense the integral converges. It can easily be shown that this integral is well defined in the usual weak sense for $\tilde{f} \in L^1 + L^2$; however, one can give a meaning to Eq. (1) for much larger classes of f (see below).

It is possible to write the map $f \mapsto Qf$ without involving Fourier transforms[2]

$$Qf = 2^{\nu}\int dv\, f(v)W(2v)\Pi. \tag{2}$$

Here Π is the parity operator, i.e., an involutive, unitary operator satisfying

$$\Pi W(v) = W(-v)\Pi;$$

the operators $\Pi(v) = W(2v)\Pi$ used in Eq. (2) are called Wigner operators.[3] Like the Weyl operators, they satisfy specific multiplication rules; moreover, they are self-adjoint.

The inverse map, from the operators on $\mathscr{H}$ to the functions on E, is the Wigner transform.[4] Formally this transform can be written as

$$f(v) = 2^{\nu}\,\mathrm{Tr}[Qf\Pi(v)]. \tag{3}$$

One sees immediately that Eq. (3) is well defined for Qf trace-class; again, we shall see below how this correspondence can be extended to more general classes of operators.

The Weyl correspondence and its inverse have been used in several different contexts. One obvious field of applications has been the study of classical limits.[5,6] In Ref. 7 the Weyl correspondence and its properties are used to find the equations of motion for a particle with spin $\frac{1}{2}$ in an electromagnetic field in a semirelativistic approximation, in particular to derive the correct magnetodynamic effect. An approach of quantum mechanics related to the Weyl transform can be found in Ref. 8, where quantum effects are studied using only functions on phase space (no Hilbert space picture), with a noncommutative product, which is usually called the twisted product, and which is the transposition, through the Weyl correspondence, of the noncommutative operator product. See also Ref. 3 for several beautiful applications, and discussions of quantum phenomena by means of the Wigner transform.

We shall be concerned here with the "topological" properties of the Weyl correspondence. We have many topolo-

[a)] On leave from Dienst voor Theoretische Natuurkunde, Vrije Universiteit Brussel, Belgium, and from Interuniversitair Instituut voor Kernwetenschappen, Belgium.

 0022-2488/83/061453-09$02.50 © 1983 American Institute of Physics

gies at our disposal, both on the functions on the classical phase space, and on the operators on the Hilbert space; it is a natural question to ask for the continuity properties of the Weyl correspondence with respect to these topologies. Some answers to this question were given in Ref. 9, showing that the Weyl correspondence maps L^2 unitarily onto τ_2, the space of Hilbert–Schmidt operators, and in Ref. 10, where it was proved that all Schwartz functions yield trace-class operators, and all L^1 functions compact operators. There also exists an extensive literature on the properties of the Weyl transform and its inverse when attention is restricted to the pseudodifferential operators (see Refs. 5, 11, 12, and the references quoted therein).

It is our purpose here to prove some new continuity statements ("positive results") and show the existence of counterexamples illustrating the breakdown of continuity if other topologies are chosen ("negative results"). To derive these results, we use extensively the properties of the harmonic oscillator coherent states[13,14] and of an integral transform[15] relating the function f with the matrix elements of Qf between these coherent states. Basically this integral transform maps functions to *analytic* functions; it is well known that such integral transforms have very special properties, which we shall exploit in our proofs, using ideas going back to Refs. 16, 17, and 18. A first application of our integral transform can be found in Ref. 19, where we exhibit a larger class of functions than $\mathcal{S}$, yielding trace-class operators, and, by duality, put some restrictions on the distributions corresponding to bounded operators. The mathematical properties of this integral transform were studied in some more detail in Ref. 20; using the results obtained there, we shall see that we can sharpen the results of Ref. 19, and derive some new ones. These results constitute the first part of this article.

While the first part contains mostly continuity results, which can be considered to be "positive" results, the second part contains essentially "negative results," i.e., counterexamples and no-go theorems showing which kind of continuity cannot be expected. For instance,

$\exists\ f \in L^\infty$ such that Qf is unbounded,

$\exists\ A$ trace-class for which $Q^{-1}A \notin L^1(E)$.

II. POSITIVE RESULTS: CONTINUITY STATEMENTS

We shall derive here some continuity properties of the Weyl transform and its inverse, using an integral transform introduced in Ref. 15, connecting the function f with the matrix element of Qf between coherent states. We therefore start by giving a short review of the definition and properties of the coherent states and of the integral transform in question; for more details the reader is referred to Refs. 15 and 20.

A. The coherent states Ω^a; the integral transform $f \mapsto (\Omega^a, Qf\Omega^b)$

Let Ω be the ground state of the harmonic oscillator $P^2 + Q^2$ [alternatively, one can define Ω as the vector for which $(P_j - iQ_j)\Omega = 0\ \forall\ j$]. We define then

$$\forall\ a \in E: \quad \Omega^a = W(a)\Omega\ .$$

It is well known[13] that the following resolution of the identity holds:

$$\int_E da\, |\Omega^a)(\Omega^a| = \mathbb{1}_{\mathcal{H}}\ , \tag{4}$$

with $da = [1/(2\pi)^\nu]\, d^\nu x_a d^\nu p_a$, and where the integral converges in the weak sense. Inserting Eq. (4) twice, one sees that for every (bounded) operator A

$$A = \int_E da \int_E db\, |\Omega^a)(\Omega^a, A\Omega^b)(\Omega^b|\ , \tag{5}$$

which means that every (bounded) operator can be reconstructed from its matrix elements between coherent states [alternatively, one can say that in the Bargmann representation of the canonical commutation relations every (bounded) operator is given by an integral kernel]. Actually, the reconstruction of A from its coherent state matrix elements $(\Omega^a, A\Omega^b)$ works for much larger classes than only the bounded operators (it works, e.g., for all closed operators for which the span of the coherent states is a core).

Applying Eq. (5) to Eq. (2), we see that (formally)

$$Qf = \int_E da \int_E db\, |\Omega^a) \int_E dv\, f(v) 2^\nu (\Omega^a, W(2v)I\Omega^b)(\Omega^b|\ . \tag{6}$$

In Ref. 15 we used the notations

$$\begin{aligned}\{a,b\,|v\} &= 2^\nu(\Omega^a, W(2v)I\Omega^b)\\ &= 2^\nu \exp\left[\tfrac{1}{2} i(p_a x_b - p_b x_a + 2p_v x_a\right.\\ &\quad - 2p_a x_v + 2p_b x_b - 2p_v x_b)\\ &\quad \left. - \tfrac{1}{4}(2x_v - x_b - x_a)^2 - \tfrac{1}{4}(2p_v - p_b - p_a)^2\right]\end{aligned}$$

and introduced the integral transform

$$(If)(a,b) = \int dv\, f(v)\{a,b\,|v\}\ . \tag{7}$$

What Eq. (6) is telling us then is that the integral transform I can be used as a tool to study the Weyl correspondence $f \leftrightarrow Qf$. This integral transform was studied in some detail in Ref. 20. We review here some of its properties. Since, for fixed a,b, the function $\{a,b\,|\cdot\}$ is C^∞, with Gaussian decrease, the integral transform I can be defined for all tempered distributions, and also for some classes of nontempered distributions. The images If have quite remarkable analyticity properties:

$$\begin{aligned}If(a,b) &= \exp\left[-\tfrac{1}{4}(x_a^2 + p_a^2 + x_b^2 + p_b^2)\right]\\ &\quad \times F(p_a + ix_a, p_b - ix_b)\ ,\end{aligned} \tag{8}$$

where F is an analytic function (depending on f, of course) on $\mathbb{C}^\nu \times \mathbb{C}^\nu$. The set of all functions which can be written in such a form we denote by $Z(E_2)$.

Two special sets of elements of $Z(E_2)$ are given by

$$\begin{aligned}u_{[m,n]}(a,b) &= \exp\left[-\tfrac{1}{4}(x_a^2 + p_a^2 + x_b^2 + p_b^2)\right]\\ &\quad \times \frac{1}{([m!][n!])^{1/2}}\left(\frac{p_a + ix_a}{\sqrt{2}}\right)^{[m]}\left(\frac{p_b - ix_b}{\sqrt{2}}\right)^{[n]},\\ \omega^{[c,d]}(a,b) &= \exp\{\tfrac{1}{2} i(p_c x_a - p_a x_c + p_b x_d - p_d x_b)\\ &\quad + \tfrac{1}{4}[(x_a - x_c)^2 + (p_a - p_c)^2 + (x_b - x_d)^2\\ &\quad + (p_b - p_d)^2]\}\ .\end{aligned} \tag{9}$$

For every element ϕ of $Z(E_2)$, one can write a Taylor series for its analytic part, which can be considered as an expansion of ϕ with respect to the $u_{[m,n]}$:

$$\phi \in Z(E_2) \Rightarrow \phi(a,b) = \sum_{[m],[n]} \phi_{[m,n]} u_{[m,n]}(a,b) \tag{10}$$

with uniform and absolute convergence on compact sets. One can, moreover, show[20] that

$$\begin{aligned}\phi_{[m,n]} &= \int da \int db\, \phi(a,b)\, \overline{u_{[m,n]}(a,b)} \\ &= \int da \int db\, \phi(a,b) u_{[n,m]}(a,b)\end{aligned} \tag{11}$$

for all ϕ in $Z(E_2)$ such that the integral on the right-hand side converges absolutely (this is the case, e.g., for the elements of the $\mathcal{F}^{\rho}$-spaces defined below).

Analogously one shows that the following reproducing property holds[20]:

$$\begin{aligned}\phi(c,d) &= \iint dadb\, \phi(a,b)\, \overline{\omega^{(c,d)}(a,b)} \\ &= \iint dadb\, \phi(a,b) \omega^{(a,b)}(c,d)\end{aligned} \tag{12}$$

for all ϕ in $Z(E_2)$ for which the integrals converge absolutely. The $u_{[m,n]}$ and $\omega^{(c,d)}$ are related to the coherent states Ω^a in the following way:

$$\begin{aligned}&(\Omega^a, \Omega^c)(\Omega^d, \Omega^b) = \omega^{(c,d)}(a,b), \\ &\langle \Omega^a | [m] \rangle \langle [n], \Omega^b \rangle = u_{[m,n]}(a,b).\end{aligned} \tag{13}$$

In Ref. 20 we defined, $\forall\ \rho \in \mathbb{R}$, the spaces $\mathcal{F}^{\rho}$ and W^{ρ} as

$$\begin{aligned}\mathcal{F}^{\rho} = \Big\{ &\phi \in Z(E_2);\ \|\phi\|_{\rho}^2 \\ &= \iint dadb\, (1 + |a|^2 + |b|^2)^{\rho} |\phi(a,b)|^2 < \infty \Big\}\end{aligned}$$

[where $|a|^2 = \frac{1}{2}(x_a^2 + p_a^2)$] (see also Ref. 14 for the definition of $\mathcal{F}^{\rho}$),

W^{ρ} = closure under $\| \ \|_{\rho}^{s}$ of

$$\begin{aligned}&\{ f \in \mathcal{S}(E);\ (\|f\|_{\rho}^{s})^2 \\ &\quad = (f, (q^2 - \tfrac{1}{4}\Delta_q + p^2 - \tfrac{1}{4}\Delta_p + \nu)^{\rho} f) < \infty \}\end{aligned}$$

(the W^{ρ} are Sobolev-type space with respect to the operator $x^2 + p^2 - \frac{1}{4}\Delta_x - \frac{1}{4}\Delta_p$; they are the same spaces as used in the N representation of $\mathcal{S}$ and $\mathcal{S}'$[21]). One can then show that, $\forall\ \rho \in \mathbb{R}$, I defines an isomorphism from W^{ρ} to $\mathcal{F}^{\rho}$.

We shall now proceed to derive some properties of the Weyl correspondence from these properties of the integral transform I.

B. The Weyl transform as a map from the (tempered) distributions to quadratic forms on the span of the coherent states

For notational convenience, we define

$$D_{\text{coh}} = \text{linear span of the coherent states } \Omega^a.$$

D_{coh} is obviously dense in $\mathcal{H}$. For any quadratic form ϕ on D_{coh}, we shall use the notation:

$$K_{\phi}(a,b) = \phi(\Omega^a, \Omega^b).$$

In what follows we shall only consider quadratic forms ϕ for which the associated function K_{ϕ} is an element of $Z(E_2)$. In doing so, we do not put too severe a restriction on ϕ: For instance, all the quadratic forms associated with Schrödinger operators $p^2 + V(x)$, with V a tempered distribution, fall into this class (this includes, e.g., the Coulomb potential as soon as $\nu \geqslant 2$). The $\mathcal{F}^{\rho}$-topologies can then be used to define topologies on the quadratic forms on D_{coh}:

$$G^{\rho} = \{\phi \text{ quadratic form on } D_{\text{coh}};\ K_{\phi} \in \mathcal{F}^{\rho}\}$$

Equipped with the corresponding norm, $\|\phi\|_{\rho} = \|K_{\phi}\|_{\rho}$, the G^{ρ} constitute then a nested Hilbert space[22] of quadratic forms. One can now define the Weyl transform for all of $\mathcal{S}'$ by means of the integral transform I:

$$\begin{aligned}&\forall\ T \in \mathcal{S}'(E), \quad QT \text{ is a quadratic form on } D_{\text{coh}} \text{ defined by} \\ &\qquad K_{QT} = IT.\end{aligned} \tag{14}$$

It is easy to check that this definition of QT coincides with the direct definition by Eqs. (1) or (2) for $f \in L^1$ or $f \in L^2$, i.e., that

$$\int dv\, f(v) \{a,b|v\} = (\Omega^a, Qf\Omega^b)$$

for f in these classes, which justifies our definition (14) as an extension. This can be verified also for the pseudodifferential operators. The fact that I is an isomorphism from W_{ρ} to $\mathcal{F}_{\rho}$ now easily translates to Q:

$$\forall\ \rho \in \mathbb{R}, \quad Q:\ W^{\rho} \to G^{\rho}$$

is an isomorphism; one has[20]

$$\begin{aligned}&\|QT\|_{\rho} \leqslant K_{\rho} \|T\|_{\rho}^{s} \\ &\text{and} \\ &\|Q^{-1}\phi\|_{\rho}^{s} \leqslant K'_{\rho} \|\phi\|_{\rho}\end{aligned} \tag{15}$$

with

$$K_{\rho} = e^{-\rho/2} \cdot \begin{cases} 1, & \rho \leqslant 0, \\ \sqrt{e}(1 + \rho/2\nu)^{\nu + \rho/2}, & \rho \geqslant 0, \end{cases}$$

$$K'_{\rho} = \begin{cases} (1 + \rho/2\nu)^{1/2}, & \rho \geqslant 0, \\ \sqrt{e} e^{-|\rho|/2} (1 + |\rho|/2\nu)^{\nu + \rho/2}, & \rho \leqslant 0. \end{cases}$$

Remarks

1. The quadratic forms thus obtained need not be closable! A striking example of a nonclosable form is given by $T(q,p) = \delta(q)$. While both $T_1(q,p) = 1$ and $T_2(q,p) = \delta(p)\delta(q)$ lead to nice, in this case even bounded operators, the quadratic form QT [as defined by Eq. (14)] is not closable, and hence not associated with an operator.

2. If however a quadratic form ϕ in G^{ρ} *is* closable, then all the eigenstates $|[m]\rangle$ of $\frac{1}{2}(P^2 + Q^2)$ are in the form domain of the closure $\bar{\phi}$ of ϕ, and

$$\bar{\phi}(|[m]\rangle, |[n]\rangle) = (K_{\phi})_{[n,m]}, \tag{16}$$

where the $(K_{\phi})_{[m,n]}$ are the coefficients in a Taylor expansion for K_{ϕ} [see Eq. 10)]:

$$K_{\phi}(a,b) = \sum_{[m],[n]} (K_{\phi})_{[m,n]} u_{[m,n]}(a,b) \tag{17}$$

[Eq. (16) can easily be proved from the fact that suitable differentiations of the Ω^a yield the eigenvectors $|[m]\rangle$, in the same way as the $(K_\phi)_{[m,n]}$ can be obtained from K_ϕ by differentiating].

3. Actually one can give a sense to $\phi(|[m]\rangle, |[n]\rangle)$ for all ϕ in G^ρ, even if ϕ is not closable, in the following way. For $\mu \geqslant 0$, we define

$$D_\mu = \left\{\psi \in \mathscr{H};\ |\psi|_\mu^2 = \int da\ |(\psi, \Omega^a)|^2 (1+|a|^2)^\mu < \infty\right\}$$
$$= \left\{\psi \in \mathscr{H};\ |\psi|_\mu^2 = \sum_{[m]} |\langle\psi|[m]\rangle|^2 (|m|+\nu)^\mu < \infty\right\}. \quad (18)$$

For $\mu < 0$ we define the norms $|\ |_\mu$, $|\ |_\mu$ in exactly the same way; in this case D_μ is defined to be the closure of $\mathscr{H}$ with respect to either $|\ |_\mu$ or $|\ |_\mu$ (these two norms are equivalent for all μ). From the definitions of the D_μ, one immediately sees that, for any μ, D_μ is the dual space of $D_{-\mu}$ with respect to a suitable extension of the inner product on $\mathscr{H}$:

$$(\psi_1, \psi_2) = \int da\ (\psi_1, \Omega^a)(\Omega^a, \psi_2).$$

On the other hand, one can show that for any $\psi \in G^\rho$

$$\forall\ \psi_1, \psi_2 \in D_{\text{coh}},$$
$$\phi(\psi_1, \psi_2) = \iint dadb\ K_\phi(a,b)(\psi_1, \Omega^a)(\Omega^b, \psi_2) \quad (19)$$

[one has only to show this, for $\psi_1 = \Omega^c$, $\psi_2 = \Omega^d$; the statement then follows from Eq. (12)].

Since $(1+|a|^2)^{1/2}(1+|b|^2)^{1/2} \leqslant 1 + |a|^2 + |b|^2 \leqslant (1+|a|^2)(1+|b|^2)$, it is then obvious from our definitions and from the duality of D_μ and $D_{-\mu}$ that every ϕ in G^ρ can be extended, using Eq. (19), to a continuous map from $D_{-\rho}$ to D_ρ if $\rho \leqslant 0$, from $D_{-\rho/2}$ to $D_{\rho/2}$ if $\rho \geqslant 0$. Alternatively, one can also say that in this way we have extended $\phi \in G^\rho$ to a quadratic form on $D_{-\rho}$ if $\rho \leqslant 0$ and on $D_{-\rho/2}$ if $\rho \geqslant 0$. In particular, since the $|[m]\rangle$ are elements of all the D_μ, this means we have given a sense to $\phi(|[m]\rangle, |[n]\rangle)$, which is again given by $(K_\phi)_{[n,m]}$:

$$\phi(|[m]\rangle, |[n]\rangle) = \iint dadb\ K_\phi(a,b)\langle[m]|\Omega^a\rangle\langle\Omega^b|[n]\rangle$$
$$= \iint dadb\ K_\phi(a,b) u_{[m,n]}(a,b) \quad \text{[see Eq. 13)]}$$
$$= (K_\phi)_{[n,m]} \quad \text{[see Eq. (11)]}.$$

4. From Eq. (18) one easily sees that D_μ is exactly the domain of $(P^2+Q^2)^{\mu/2}$, or equivalently, the form domain of $(P^2+Q^2)^\mu$.

Take $\rho \geqslant 0$. Then, by the extension defined above, $\phi \in G^{-\rho}$ defines a continuous, more precisely a Hilbert–Schmidt map from D_ρ to $D_{-\rho}$. This means that ϕ is a quadratic form, relatively form-compact with respect to $(P^2+Q^2)^\rho$. (See Ref. 23 for the definition of relatively form-compactness).

5. Translating all this to the Weyl transform, we see now that $\forall\ \rho \geqslant 0$, $\forall\ T \in W^{-\rho}$: QT is a quadratic form, relatively form-compact with respect to $(P^2+Q^2)^\rho$.

6. In Ref. 20 the action of I on some classes of distributions "of type S," which contain also nontempered distributions, was studied. One can perform the same constructions and extensions as above for these larger classes; as a result one gets that Q, when applied to a $H(\overline{\alpha}, \overline{A})$[20,24] yields quadratic forms relatively compact with respect to

$$\exp[\tau(P^2+Q^2)^{1/2\alpha}], \quad \text{for } \alpha > \tfrac{1}{2},\ \forall\ \tau > 2A^{-1/\alpha}.$$

C. The distributions corresponding to bounded operators

Let A be a bounded operator. Equation (14) provides us with a simple rule to find the function or distribution $Q^{-1}A$, via the coherent states. We define $K_A(a,b) = (\Omega^a, A\Omega^b)$. One can easily check, from the properties of the coherent states, that K_A is an element of $Z(E_2)$. Moreover [use Eq. (4)],

$$\int_E da\ |K_A(a,b)|^2 = \|A\Omega^b\|^2 \leqslant \|A\|^2;$$

hence, $\forall\ \epsilon > 0$,

$$\int da \int db\,(1+|a|^2+|b|^2)^{-\nu-\epsilon}|K_A(a,b)|^2$$
$$\leqslant \int db\,(1+|b|^2)^{-\nu-\epsilon}\|A\|^2$$
$$= \|A\|^2 \frac{2}{\Gamma(\nu)}\int_0^\infty dx\,\frac{x^{2\nu-1}}{(1+x^2)^{\nu+\epsilon}}$$
$$= \|A\|^2 \Gamma(\epsilon)\Gamma(\nu+\epsilon)\Gamma(\nu)^{-1}.$$

This implies that A, considered as a quadratic form on D_{coh}, is an element of $\cap_{\epsilon>0} G^{-(\nu+\epsilon)}$, with

$$\forall\ \epsilon > 0, \quad \|A\|_{-(\nu+\epsilon)} \leqslant [\Gamma(\epsilon)/\Gamma(\nu+\epsilon)]^{1/2}\|A\|.$$

As a consequence of this, we can formulate the following restriction on the distributions corresponding to bounded operators: The Weyl transform QT of a distribution $T \in \mathscr{S}'(E)$ can be a bounded operator only if

$$T \in \bigcap_{\epsilon>0} W^{-(\nu+\epsilon)}$$

and $\exists\ K \geqslant 0$ such that

$$\forall\ \epsilon > 0, \quad \|T\|_{-(\nu+\epsilon)}^s$$
$$\leqslant e^{-(\nu+\epsilon-1)/2}[(3+\epsilon/\nu)/2]^{(3\nu+\epsilon)/2}|\Gamma(\epsilon)/\Gamma(\nu+\epsilon)|^{1/2}K. \quad (20)$$

The "only if" in this statement cannot be replaced by an "if and only if." This is again illustrated by the distribution $T(q,p) = \delta(q)$: it turns out that T is an element of $\cap_{\epsilon>0} W^{-(\nu+\epsilon)}$ and satisfies Eq. (20), though QT is not even closable, and certainly not bounded.

The topology induced on $B(\mathscr{H})$ by the $\|\ \|_{-(\nu+\epsilon)}$-norms is much weaker than the norm topology. Actually, it is even weaker than the strong topology: if $A_n \to_{n\to\infty} A$ strongly, then the $\|A_n\|$ are uniformly bounded (by the principle of uniform boundedness, e.g., Ref. 21).

Hence

$$\int da\,(1+|a|^2+|b|^2)^{-\nu-\epsilon}|K_A(a,b) - K_{A_n}(a,b)|^2$$
$$\leqslant (1+|b|^2)^{-\nu-\epsilon}\|A\Omega^b - A_n\Omega^b\|^2$$
$$\leqslant (1+|b|^2)^{-\nu-\epsilon}\sup(4\|A_n\|^2),$$

which is integrable in b. Together with

$\|A\Omega^b - A_n\Omega^b\| \to_{n\to\infty} 0$ this implies
$\|A - A_n\|_{-(\nu+\epsilon)} \to_{n\to\infty} 0 \ \forall\ \epsilon > 0$; hence
$Q^{-1}A_n \to_{n\to\infty} Q^{-1}A$ in each $W^{-(\nu+\epsilon)}$, $\epsilon > 0$.

Remarks

1. Note that this argument can also be used for unbounded operators: Whenever A_n and A are unbounded operators such that

$$(\Omega^a, A_n\Omega^b) \underset{n\to\infty}{\to} (\Omega^a, A\Omega^b) \quad \forall\ a,b$$

and $\exists\ K, l$ for which $|(\Omega^a, A_n\Omega^b)|^2 \leqslant K(1+|a|^2+|b|^2)^l$ uniformly in n then $Q^{-1}A_n \to_{n\to\infty} Q^{-1}A$ in $\mathscr{S}'(E)$ (i.e., "in the sense of the distributions").

2. Actually, the bound Eq. (20) can still be sharpened a little bit. Defining, for $\rho \in \mathbb{R}$, the norm $|\ |^s_\rho$ (equivalent to $\|\ \|^s_\rho$) by

$$|T|^s_\rho = (T, (q^2 - \tfrac{1}{4}\Delta_q + p^2 - \tfrac{1}{4}\Delta_p)^\rho T)^{1/2} \tag{21}$$

(we omit the extra ν in the definition of $\|\ \|^s_\rho$, which makes this norm larger than $\|\ \|^s_\rho$, for $\rho \leqslant 0$), one can show[25] that (20) can be replaced by

QT bounded only if $\exists\ K \geqslant 0$ such that

$$|T|^s_{-(\nu+\epsilon)} \leqslant \Gamma(\nu)^{-1/2}\zeta(1+\epsilon)^{1/2}K\,,$$

where ζ is Riemann's zeta function: $\zeta(x) = \Sigma_{k=1}^\infty k^{-x}$.

D. A class of functions yielding trace-class operators

The restriction (20) on the class of distributions corresponding to bounded operators is sharper than the one derived in Ref. 19. We can consequently also find a larger class of functions leading to trace-class operators.

To do this, we shall follow essentially the same method as in Ref. 19: we need as a preliminary lemma that

$$\forall\ B \text{ trace-class}, \quad \mathrm{Tr}\,B = \int da\,(\Omega^a, B\Omega^a)\,. \tag{22}$$

(This is fairly easy to prove; see, e.g., Ref. 19. Note that the absolute convergence of the integral on the right-hand side for a given operator B does not imply that B is trace-class).

Take now $f \in \mathscr{S}(E)$. We know[10] that Qf is trace-class. Moreover, the trace-norm of Qf is given by

$$\|Qf\|_{\mathrm{tr}} = \mathrm{Tr}|Qf| = \sup_{A\in B(\mathscr{H}),\, A\neq 0} \|A\|^{-1}\cdot|\mathrm{Tr}(A\cdot Qf)|\,.$$

But, by Eqs. (22) and (4),

$$\begin{aligned}
|\mathrm{Tr}(A\cdot Qf)| &= \left|\int da\,(\Omega^a, AQf\Omega^a)\right| \\
&= \left|\int da \int db\,(\Omega^a, A\Omega^b)(\Omega^b, Qf\Omega^a)\right| \\
&\leqslant \left(\iint dadb\,(1+|a|^2+|b|^2)^{-\nu-\epsilon}|(\Omega^a, A\Omega^b)|^2\right)^{1/2} \\
&\times \left(\iint dadb\,(1+|a|^2+|b|^2)^{\nu+\epsilon}|(\Omega^b, Qf\Omega^a)|^2\right)^{1/2} \\
&\leqslant |\Gamma(\epsilon)/\Gamma(\nu+\epsilon)|^{1/2}\|A\|\,\|If\|_{\nu+\epsilon} \\
&\leqslant |\Gamma(\epsilon)/\Gamma(\nu+\epsilon)|^{1/2}e^{-(\nu+\epsilon-1)/2}[(3+\epsilon/\nu)/2]^{(3\nu+\epsilon)/2}\|A\|\,\|f\|^s_{\nu+\epsilon}\,,
\end{aligned}$$

which implies

$$\begin{aligned}
\|Qf\|_{\mathrm{tr}} &\leqslant e^{-(\nu+\epsilon-1)/2}[(3+\epsilon/\nu)/2]^{(3\nu+\epsilon)/2} \\
&\times [\Gamma(\epsilon)/\Gamma(\nu+\epsilon)]^{1/2}\|f\|^s_{\nu+\epsilon}\,.
\end{aligned} \tag{23}$$

Since $\mathscr{S}(E)$ is dense in $W^{\nu+\epsilon}$, this implies then that all functions in $\cup_{\rho>\nu} W^\rho$ yield trace-class operators, with a bound on their trace-norm given by Eq. (23).

Remarks

1. Again, the bound Eq. (23) can be sharpened to

$$\|Qf\|_{\mathrm{tr}} \leqslant \Gamma(\nu)^{-1/2}\zeta(1+\epsilon)^{1/2}|f|^s_{\nu+\epsilon} \tag{24}$$

with $|f|^s_{\nu+\epsilon}$ defined as in Eq. (21).

2. Note that all the elements of W^ρ $(\rho > \nu)$ are L^1-functions (they are, of course, also square-integrable, which was to be expected since Q maps L^2 unitarily onto the Hilbert–Schmidt operators,[9,19] and since every trace-class operator is also a Hilbert–Schmidt operator).

3. In the same way as Eq. (23) one can also derive the following inequality[25]:

$$f_1(q,p) = f(q) \quad \text{(independent of } p\text{)},$$
$$f_2(q,p) = g(p) \quad \text{(independent of } q\text{)}.$$

Define

$$|||f_1|||^s_\rho = [(f, (q^2 - \tfrac{1}{4}\Delta_q + \nu/2)^\rho f)]^{1/2}\,,$$

with an analogous definition for $|||f_2|||^s_\rho$. Then

$$\begin{aligned}
&\|Qf_1Qf_2\|_{\mathrm{tr}} \\
&\leqslant 2^\nu[\Gamma(\epsilon)/\Gamma(\nu+\epsilon)]e^{-(\nu+\epsilon-1)}[(3+\epsilon/\nu)/2]^{2\nu+\epsilon} \\
&\times |||f_1|||^s_{\nu+\epsilon}|||f_2|||^s_{\nu+\epsilon}\,.
\end{aligned}$$

This condition on f_1 and f_2 is reminiscent of, but stronger (and our result therefore weaker) than the condition in Theorem XI.21 in Ref. 26. The reason why our condition is stronger is that our treatment, and hence this condition, are invariant under Fourier transforms (see below), while the condition in Ref. 26 is not.

E. The functions corresponding to the dyadics $|[m]\rangle\langle[n]|$

Again we use Eq. (14) to find the functions corresponding to these dyadics (remember that the $|[n]\rangle$ are the eigenstates of the harmonic oscillator $P^2+Q^2:\frac{1}{2}(P_j^2+Q_j^2)|[n]\rangle = (n_j+\frac{1}{2})|[n]\rangle$):

$$K_{[m],[n]}(a,b) = \langle\Omega^a|[m]\rangle\langle[n]|\Omega^b\rangle = u_{[m,n]}(a,b) \quad \text{[see Eq. (13)]}.$$

Hence

$$Q^{-1}(|[m]\rangle\langle[n]|) = I^{-1}u_{[m,n]_{\text{def}}} = h_{[n,m]} \quad \text{(see Refs. 13, 15)}.$$

These functions $h_{[n,m]}$ were studied in Refs. 15, 20, and 25 (among other things). They can be calculated explicitly:

$$h_{[k,l]}(q,p) = 2^{\nu}e^{-(q^2+p^2)}\prod_{j=1}^{\nu}\left[\sum_{s_j=0}^{\min(k_j,l_j)}(-2)^{-s_j}2^{(k_j+l_j)/2}\frac{\sqrt{k_j!l_j!}}{s_j!(k_j-s_j)!(l_j-s_j)!}(p_j+iq_j)^{l_j-s_j}(p_j-iq_j)^{k_j-s_j}\right],$$

and they have many interesting properties, as, e.g.,[25]

$$(-\tfrac{1}{4}\Delta_q-\tfrac{1}{4}\Delta_p+q^2+p^2)h_{[k,l]} = (|k|+|l|+\nu)h_{[k,l]}$$

which means they are related to the Hermite functions. Like the Hermite functions, they form an orthonormal base in L^2.[20] There exists also a connection with the Laguerre polynomials. One has, for instance, in the case $\nu=1$,

$$h_{kk}(q,p) = 2(-1)^k e^{-(x^2+p^2)}L_k(2x^2+2p^2), \tag{25}$$

while the nondiagonal h_{kl} can be related to generalized Laguerre polynomials.[3,27]

It is amusing to note that the functions we denote by $h_{[k,l]}$ have been discovered and rediscovered several times in the literature. The diagonal h_{kk} can already be found in Ref. 28 (25 years ago!); in Ref. 29 they are rediscovered, and used in a very elegant way to derive properties of the Laguerre polynomials from the properties of the Weyl transform. They were again found in Ref. 8. It is quite likely that these are not the only places in the literature where they were discovered... . The nondiagonal h_{kl} seem to be less popular; they can, nevertheless, also be found in Refs. 27 and 3. One can probably extend the methods used in Ref. 29 and use the h_{kl} and their relation with the Weyl transform to derive properties of the generalized Laguerre polynomials.

Their main interest to us, here, is that they can be used to compute matrix elements between harmonic oscillator eigenstates (such matrix elements are often used, e.g., in nuclear physics). Indeed, for any $T\in\mathcal{S}'(E)$ one has (see Sec. IIB)

$$\begin{aligned}\langle[m]|QT[n]\rangle &= (K_{QT})_{[n,m]}\\ &= (IT)_{[n,m]} \quad \text{[use Eq. (14)]}\\ &= \iint dadb\, IT(a,b)u_{[m,n]}(a,b) \quad \text{[see Eq. (11)]}\\ &= \int dv\, T(v)h_{[n,m]}(v) \quad \text{(see Ref. 15)}, \end{aligned}\tag{26}$$

where the last integral has to be understood as $T(h_{[n,m]})$, if the distribution T is not given by a genuine function. This means one can compute the matrix elements of QT between harmonic oscillator eigenstates by a direct integration on the classical phase space.

F. The relation between Q and the Fourier transform

As was pointed out in Refs. 15 and 20, the integral kernel $\{a,b|\cdot\}$ satisfies the following invariance property under Fourier transforms:

$$F_4(\{a,b|\cdot\})(v) = \{a,-b|v\},$$

where

$$(F_\alpha f)(v) = 2^{-\nu}|\alpha|^{\nu}\int dv'\, e^{i\alpha\sigma(v,v')}f(v')$$

(F_1 is exactly the symplectic Fourier transform $\tilde{}$ already defined in the introduction). This then implies, $\forall\ T\in\mathcal{S}'(E)$ [also for all the nontempered distributions in the $H(\overline{\alpha},\overline{A})$ spaces; see Ref. 20]

$$\forall\ a,b,\quad I(F_{-4}T)(a,b) = IT(a,-b).$$

Translated to the Weyl transform language, this means that

$$Q(F_{-4}T) = QT\cdot\Pi \tag{27}$$

($QT\cdot\Pi$ can be defined as a quadratic form on D_{coh}, and, if $T\in W^{\rho}$, on $D_{-\rho}$, without any problems, since these domains are invariant under Π). This also means that one can enlarge every class of distributions (providing it is not already invariant under F_{-4}), yielding operators with a specific property (provided this property is invariant under multiplication by the parity operator), just by applying F_{-4}, and so produce a new set of distributions with the desired property.

The classes of distributions we introduced in the preceding sections are invariant under F_{-4}, but this is not the case with other characterizations found in the literature. For instance, we know[10] that

$$\forall\ f\in L^1(E),\ Qf \text{ compact and } \|Qf\|\leqslant 2^{\nu}\|f\|_1 \tag{28}$$

[this can also be derived from Eq. (23): the bound Eq. (28) obviously holds for all f in L^1—see Eq. (2)—; since $W^{\nu+\epsilon}$ is dense in L^1, Eq. (23) implies then that, $\forall\ f\in L^1$, Qf is the norm-limit of trace-class operators, and hence compact). Us-

ing Eq. (27), we see then that

$$\forall\, f\in L^1,\ Q(F_{-4}f) \text{ compact and } \|Q(F_{-4}f)\|\leqslant 2^{\nu}\|f\|_1 . \tag{29}$$

Remark

Equation (29) can already be inferred from the formal expression (1); the invariance property of the kernel $\{a,b\,|\cdot\}$ under F_{-4} is another way of saying that Eqs. (1) and (2) are equivalent: formally $[F_4 W(2\cdot)](v) = W(2v)\Pi$.

III. NEGATIVE RESULTS: PATHOLOGIC BEHAVIOR

In this second half of this paper we want to present some results showing one should be careful about Wigner functions and the Weyl transform, and not always trust one's first intuition. These results complement the continuity statements in part A in showing which continuity properties can definitely *not* be expected.

A. A positive function leading to a nonpositive operator

This first section is really only a remark.

It is well known that the Wigner distribution corresponding to a density matrix (i.e., a positive, trace-class operator with trace 1) need not be positive everywhere.[4] {Actually, one can prove[30] that the *only* pure states for which the Wigner function is positive everywhere are the Gaussian states, i.e., the states of form $\psi(x) = N\exp[i\alpha x + \beta(x-x_0)^2]$ in the Schrödinger representation.} Note that the same phenomenon occurs for the so-called diagonal representation with respect to the coherent states, where one represents operators by an integral over dyadics $|\Omega^a)(\Omega^a|$: $A = \int da\, \phi_A(a)|\Omega^a)(\Omega^a|$. Here too it may happen that A is positive, even though ϕ_A is not positive everywhere. However, one always has $\phi_A \geqslant 0 \Rightarrow A \geqslant 0$. This is not true for the Weyl transform, where one *can* have a nonpositive operator stemming from a positive function.

Using the parity operator Π, we build here an example of a positive function f for which the Weyl transform Qf is not positive. Take any ψ such that $\Pi\psi = -\psi$, $\psi\neq 0$, (e.g., $\psi = |[n]\rangle$ with $|n|$ odd). Since the Wigner operators $\Pi(v)$ are strongly continuous, there exists an $r\geqslant 0$ such that

$$|v|\leqslant r \Rightarrow (\psi, \Pi(v)\psi) \leqslant -\tfrac{1}{2}\|\psi\|^2 .$$

Define

$$f(v) = \begin{cases} 1, & |v|\leqslant r , \\ 0, & |v| > r . \end{cases}$$

Then

$$\begin{aligned}(\psi, Qf\psi) &= 2^{\nu}\int dv\, f(v)(\psi, \Pi(v)\psi) \\ &= 2^{\nu}\int_{|v|\leqslant r} dv\, (\psi, \Pi(v)\psi) \\ &\leqslant -2^{\nu-1}\|\psi\|^2 \int_{|v|\leqslant r} dv \\ &= -2^{\nu} r^{2\nu}[1/\Gamma(\nu+1)]\|\psi\|^2 ,\end{aligned}$$

which shows that Qf is not positive.

Remarks

1. The function f we have constructed here is clearly in L^2, which means Qf is a Hilbert–Schmidt operator, and therefore bounded. We shall show later on that there exist positive, bounded functions f for which $\sigma(Qf)$ is not even bounded below.

2. Since the function f as we constructed it is discontinuous, it cannot be the Wigner distribution of a trace-class operator. It is obvious, however, that we could also have chosen f positive, C^∞, with support in $\{v, |v|\leqslant r\}$, without invalidating the conclusion that $\langle\psi, Qf\psi\rangle < 0$. This shows that there exist positive functions in $\mathscr{S}(E)$ with nonpositive, trace-class Weyl transforms.

B. Bounded functions leading to unbounded operators

As long as we concern ourselves with functions depending only on q or on p, we know that the operator corresponding to such a bounded function will always be bounded, with

$$\|Qf\|\leqslant\|f\| . \tag{30}$$

We shall show here that this is no longer the case once one considers functions depending on both q and p: Not only does the bound Eq. (29) not hold any longer, even with an extra constant K on the right-hand side, but there actually exist bounded functions whose Weyl transform is an unbounded operator.

We shall prove this in two steps. In a first step we construct a sequence of functions f_n such that $\|f_n\|_\infty = 1$, $\|Qf_n\| \to_{n\to\infty} \infty$, showing thereby that no K exists for which $\|Qf\|\leqslant K\|f\|_\infty$. We then prove (ad absurdum) that this implies $Q(L^\infty) \not\subset B(\mathscr{H})$.

Take, for simplicity reasons, $\nu = 1$. Define

$$f_n(v) = \begin{cases} 1 & \text{if } h_{nn}(v)\geqslant 0 , \\ -1 & \text{if } h_{nn}(v) < 0 . \end{cases}$$

Then

$$\begin{aligned}\langle n|Qf_n|n\rangle &= \int dv\, f_n(v)h_{nn}(v) \quad [\text{see Eq. (26)}] \\ &= \int dv\, |h_{nn}(v)| \\ &= (1/2\pi)\int dq \int dp\, 2e^{-(q^2+p^2)}|L_n(2q^2+2p^2)| \\ &\qquad [\text{see Eq. (25)}] \\ &= \tfrac{1}{2}\int du\, e^{-u/2}|L_n(u)| .\end{aligned}$$

To put a lower bound on this integral, we shall use asymptotic formulas for the Laguerre polynomials derived by Tricomi.[31] Let η be a small positive number (more specifically $\eta \in\,]0, \tfrac{4}{5}[$). Then, for $x\in[\eta(n+\tfrac{1}{2}), (1-\eta)(4n+2)]$,

$$L_n(x) = e^{x/2}(1/\sqrt{\pi k}\,\sqrt{\sin 2\theta})[\sin(\Theta + n\pi) + O(k^{-1})]$$

where

$$k = n + \tfrac{1}{2}, \quad \theta = \cos^{-1}\sqrt{x/4k} ,$$
$$\Theta = k(2\theta - \sin 2\theta) + \pi/4$$

and where the term $O(k^{-1})$ can be bounded by Mk^{-1}, *uniformly* in x for $x\in[\eta k, 4(1-\eta)k]$. (This last fact is not

stated explicitly in Ref. 31, but it can be derived from the details of Tricomi's proof). Consequently,

$$\begin{aligned}
&\int_0^\infty du\, e^{-u/2}|L_n(u)| \\
&\quad \geqslant \int_{\eta(n+1/2)}^{(1-\eta)(4n+2)} dx\,(1/\sqrt{\pi k \sin 2\theta}) \\
&\qquad \times |\sin(\Theta + n\pi) + O(k^{-1})| \\
&\quad \geqslant (8k/\sqrt{2\pi k}) \int_{\cos^{-1}\sqrt{1-\eta}}^{\cos^{-1}\sqrt{\eta/4}} d\theta \sqrt{\sin 2\theta} \\
&\qquad \times \{|\sin[k(2\theta - \sin 2\theta) + \pi/4]| - Mk^{-1}\} \\
&\quad \underset{n\to\infty}{\to} (8\sqrt{k}/\sqrt{2\pi})\Big[(2/\pi) \int_{\cos^{-1}\sqrt{1-\eta}}^{\cos^{-1}\sqrt{\eta/4}} \\
&\qquad \times d\theta \sqrt{\sin 2\theta}\,(1 - \pi M/2k)\Big]
\end{aligned}$$

(by an extension of the Riemann–Lebesgue lemma). This implies there exists an $\alpha > 0$ such that

$$\forall\ n, \quad \int_0^\infty du\, e^{-u/2}|L_n(u)| \geqslant \alpha\sqrt{n}\ ;$$

hence (31)

$$\|Qf_n\| \geqslant \langle n|Qf_n|n\rangle \geqslant \alpha\sqrt{n}\ .$$

Suppose now that $Q(L^\infty) \subset B(\mathscr{H})$. We show next that this implies that Q is closed from L^∞ to $B(\mathscr{H})$. Indeed, take $g_n \in L^\infty$ such that

$$\|g_n - g\|_\infty \underset{n\to\infty}{\to} 0, \quad \|Qg_n - B\| \underset{n\to\infty}{\to} 0\,.$$

Then,

$$\begin{aligned}
&\forall\ a,b, \quad (\Omega^a, B\Omega^b) \\
&\quad = \lim_{n\to\infty} (\Omega^a, Qg_n\Omega^b) \\
&\quad = \lim_{n\to\infty} \int dv\, g_n(v)\{a,b|v\} \\
&\quad = \int dv\, g(v)\{a,b|v\} \quad [\{a,b|\cdot\} \in L^1(E)] \\
&\quad = (\Omega^a, Qg\Omega^b)\,;
\end{aligned}$$

hence $B = Qg$. This implies Q is closed from L^∞ to $B(\mathscr{H})$; hence, by the closed graph theorem, $\exists\ K > 0$ such that $\|Qf\| \leqslant K\|f\|_\infty$. This is clearly in contradiction with Eq. (31), which allows us to conclude that $Q(L^\infty) \not\subset B(\mathscr{H})$; i.e., there exist bounded functions with unbounded Weyl transforms.

Remark

If f is a bounded function with unbounded Weyl transform Qf, then $g = f + \|f\|_\infty$ is a positive bounded function with unbounded Weyl transform. From g one can then construct a positive bounded function h such that Qh is not even bounded below. Let us suppose that Qg is bounded below (otherwise, we simply take $h = g$). Then Qg is not bounded above, since Qg is unbounded. Take now $h = \|g\|_\infty - g$. This is clearly a positive, bounded function; moreover, $Qg + Qh = \|g\|_\infty \mathbb{1}$, which implies Qh is not bounded below.

C. Non-absolutely-integrable Wigner distributions corresponding to positive, trace-class operators

All the functions leading to trace-class operators we have encountered here till now were absolutely integrable (e.g., the $h_{[m,n]}$, the class $\cup_{\rho>\nu} W^\rho$ in Sec. IID). Of course, this does not mean that every Wigner distribution for a trace-class operator must necessarily be in L^1. On the other hand, if Qf is trace-class, then

$$\lim_{R\to\infty} \int_{|v|<R} dv\, f(v) = \mathrm{Tr}\, Qf \leqslant \|Qf\|_{\mathrm{tr}}\ .$$

One knows, of course, that f need not be positive everywhere, even if Qf is positive, and hence that the integral of $|f|$ will be larger, in most cases, than the integral of f itself. The connected components of the domain where f is negative (for Qf positive, trace-class) have to be rather "small," however (otherwise, negative expectation values of Qf would be possible), and are physically thought of as being caused by "quantum fluctuations."[7] It thus does not seem unreasonable, at first sight, to hope that, even though

$$\|f\|_1 \leqslant \|Qf\|_{\mathrm{tr}}$$

cannot possibly hold for all trace-class operators Qf, one still would retain the property that for every trace-class operator A the associated Wigner distribution $Q^{-1}A$ would be in $L^1(E)$. It is a direct consequence of the result in the preceding section that this argument turns out to be deceiving: There do exist positive trace-class operators A for which $Q^{-1}A \notin L^1(E)$. Indeed, suppose there were none. Then

$$Q^{-1}(\tau_1(\mathscr{H})) \subset L^1(E)\,.$$

Again, this implies that $Q^{-1}:\tau_1(\mathscr{H}) \to L^1(E)$ is closed: Take A_n such that

$$\|A_n - A\|_{\mathrm{tr}} \underset{n\to\infty}{\to} 0\,, \quad \|Q^{-1}A_n - f\|_1 \underset{n\to\infty}{\to} 0\,;$$

then

$$\begin{aligned}
(\Omega^a, A\Omega^b) &= \lim_{n\to\infty} (\Omega^a, A_n\Omega^b) \\
&= \lim_{n\to\infty} \int dv\, (Q^{-1}A_n)(v)\{a,b|v\} \\
&= \int dv\, f(v)\{a,b|v\} \quad [\{a,b|\cdot\} \in L^\infty(E)]\ ;
\end{aligned}$$

hence $Q^{-1}A = f$ [$I:\mathscr{S}' \to Z(E_2)$ is injective]. By the closed graph theorem this then implies

$$\exists\ K > 0 \quad \text{such that} \quad \|Q^{-1}A\|_1 \leqslant K\|A\|_{\mathrm{tr}}\ .$$

But we already calculated that for $A_n = |n\rangle\langle n|$ (take $\nu = 1$)

$$\|Q^{-1}A_n\|_1 = \int dv\, |h_{nn}(v)| \geqslant \alpha\sqrt{n}\ ,$$

while $\|A_n\|_{\mathrm{tr}} = 1$. This is clearly a contradiction, implying that $Q^{-1}(\tau_1(\mathscr{H})) \not\subset L^1(E)$.

Remark

The inclusion in the other direction does not hold ei-

ther:

$$Q(L^1(E)) \not\subset \tau_1(\mathcal{H}).$$

This is an immediate consequence of the fact that all Wigner distributions of trace-class operators are continuous. Moreover, $Q^{-1}(\tau_1(\mathcal{H})) \subset L^2(E)$ (since every trace-class operator is Hilbert–Schmidt), which shows that not even all continuous L^1 functions lead to trace-class operators.

ACKNOWLEDGMENTS

It is a pleasure for me to thank A. Grossmann for his constant interest and his encouraging support. I would also like to thank O. Heilman for a helpful remark concerning Laguerre polynomials and J.-P. Antoine for indicating Ref. 26 to me and for his interest.

[1]H. Weyl, *The Theory of Groups and Quantum Mechanics* (Dover, New York, 1950).
[2]A Grossmann, Comm. Math. Phys. **48**, 191 (1976).
[3]J. P. Amiet and P. Huguenin, *Mécaniques classique et quantique dans l'espace de phase* (Université de Neuchâtel, Neuchâtel, Switzerland, 1981).
[4]E. Wigner, Phys. Rev. **40**, 749 (1932).
[5]A. Voros, "Développements semi-classiques," Thèse de Doctorat, Université de Paris-Sud, 1977.
[6]A. Grossmann, Comm. Math. Phys. **48**, 194 (1976).
[7]L. G. Suttorp and S. R. de Groot, Nuovo Cimento A **65**, 245 (1970); S. R. de Groot and L. G. Suttorp, *Foundations of Electrodynamics* (North-Holland, Amsterdam, 1972).
[8]F. Bayen, M. Flato, C. Fronsdal, A. Lichnerowics, and D. Sternheimer, Ann. Phys. (N.Y.) **111**, 111 (1978).
[9]I. Segal, Math. Scand. **13**, 31 (1963); J. C. Pool, J. Math. Phys. 7, 66 (1966).
[10]G. Loupias, "Sup la convolution gauche," Thèse de Doctorat, Université d'Aix—Marseille, 1966; G. Loupias and S. Miracle-Sole, Ann. Inst. H. Poincaré **6**, 39 (1967).
[11]A. Grossmann, G. Loupias, and E. M. Stein, Ann. Inst. Fourier **18**, 1 (1969).
[12]L. Hörmander, Comm. Pure Appl. Math. **32**, 359 (1979).
[13]J. R. Klauder, Ann. Phys. (N.Y.) **11**, 123 (1960); J. R. Klauder and E. C. G. Sudarshan, *Fundamentals of Quantum Optics* (Benjamin, New York, 1968).
[14]V. Bargmann, Comm. Pure Appl. Math. **14**, 187 (1961); **20**, 1 (1967).
[15]I. Daubechies and A. Grossmann, J. Math. Phys. **21**, 2080 (1980).
[16]N. Aronszajn, Trans. Am. Math. Soc. **68**, 337 (1950).
[17]J. McKenna and J. R. Klauder, J. Math. Phys. **5**, 878 (1964); J. R. Klauder, J. Math. Phys. **4**, 1055 (1965).
[18]C. L. Mehta and E. C. G. Sudarshan, Phys. Rev. B **138**, 274 (1965).
[19]I. Daubechies, Comm. Math. Phys. **75**, 229 (1980).
[20]I. Daubechies, A. Grossmann, and J. Reignier, J. Math. Phys., (to be published).
[21]M. Reed and B. Simon, *Modern Methods in Mathematical Physics, Vol. I: Functional Analysis* (Academic, New York, 1972).
[22]A. Grossmann, Comm. Math. Phys. **2**, 1 (1966).
[23]M. Reed and B. Simon, *Modern Methods in Mathematical Physics, Vol. IV: Analysis of Operators* (Academic, New York, 1978).
[24]A. Grossmann, J. Math. Phys. **6**, 54 (1965).
[25]I. Daubechies, "Representation of quantum mechanical operators by kernels on Hilbert spaces of analytic functions," doctoral thesis, Vrije Universiteit Brussel, 1980.
[26]M. Reed and B. Simon, *Modern Methods in Mathematical Physics, Vol. III: Scattering Theory* (Academic, New York, 1979).
[27]A. M. Nachin, Thèse de 3^e cycle, Université d'Aix—Marseille, 1972.
[28]U. Uhlhorn, Arkiv Fysik **11**, 87 (1957).
[29]J. Peetre, Le Matematiche (Catania) **27**, 301 (1972).
[30]R. L. Hudson, Rep. Math. Phys. **6**, 1249 (1974); C. Piquet, C. R. Acad. Sci. Paris A **279**, 107 (1974).
[31]F. G. Tricomi, Comm. Math. Helv. **22**, 150 (1949).

A rigged Hilbert space of Hardy-class functions: Applications to resonances

M. Gadella[a)]
Department of Physics, Center for Particle Theory, The University of Texas at Austin, Austin, Texas 78712

(Received 22 July 1981; accepted for publication 6 August 1982)

The explicit construction of a dense subspace Φ of square integrable functions on the positive half of the real line is given. This space Φ has the properties that: (1) it is endowed with a metrizable nuclear topology, (2) it is stable under multiplication by x, and (3) the functions in Φ have suitable analytical continuation to a half plane. The space Φ^* of functions which are conjugate to elements of Φ is also considered. Then the triplets $\Phi \subset \mathscr{L}^2(0,\infty) \subset \Phi'$ and $\Phi^* \subset \mathscr{L}^2(0,\infty) \subset \Phi^{*\prime}$ are used to give a description of resonances.

PACS numbers: 03.65.Db

I. INTRODUCTION

The aim of the present paper is to implement a model of decaying states, which has been recently proposed.[1] This model uses a Rigged Hilbert Space (RHS) of functions with some analytical properties in the second sheet of the Riemann surface associated with the transformation $E = p^2/2m$. However, it does not tell us what these functions are. Here we propose a mathematical apparatus which fits reasonably with the claims of Ref. 1.

From the physical point of view we are looking at a rather simple case. The total Hamiltonian has an absolutely continuous spectrum which coincides with the positive half of the real line (R^+). We also consider the problem of two-particle resonant scattering and assume that the analytically continued S-matrix has no singularities but pairs of simple poles in the p-plane near the real axis. In the energy representation we have pairs of simple poles in the second sheet near the positive real axis. In this second sheet we assume that the S matrix does not grow faster than a polynomial in E. The existence of potentials satisfying these criteria is briefly discussed elsewhere.[2,3] The definition of resonance is to some extent a matter of personal taste. We here endorse the philosophy that there is a bijective correspondence between resonances and simple poles in the second sheet, near the positive real axis in the energy representation.[4,5]

To describe resonances on the context of the RHS several attempts have already been made. A constructive description has been suggested by Baumgärtel.[6] Other authors implement these ideas on the context of the Friedrichs model by considering the resonances to be adequately described as poles in the second sheet.[7] We here develop a different model.

We split the resonant scattering in two parts.[1] If we define Φ_- as the space of functions in $\mathscr{L}^2(0,\infty)$ with properties to be specified later, then the set of decaying states will be some subspace of the dual (with respect to a certain topology) of Φ_-, which we call Φ_+. The set of growing states represented by Φ_- will be a space of functionals over Φ_+. Φ_- and Φ_+ are both spaces of square integrable functions of R^+ which can be uniquely extended to R. These extensions will call Ψ_- and Ψ_+, respectively, and they are easier to work with, though they represent the same space of states as Φ_- and Φ_+. Φ_+ and Φ_- are chosen for mathematical convenience only as we follow other authors.[4,8] in our reservations about their interpretation as preparable states.

We start by constructing an RHS[8,11] Φ of Hardy class functions in the upper half-plane. (Φ_+ and Φ_- correspond in a similar way to an RHS of Hardy class functions on lower and upper half-planes of the second sheet, respectively.) Φ is dense in $\mathscr{L}^2(0,\infty)$ with the usual topology, and it further has a nuclear topology that is finer than that inhereted from $\mathscr{L}^2(0,\infty)$.[12]

In the second section we closely follow Ref. 1, and we prove the statements that remain unproved there. We use the spaces Ψ_- and Ψ_+ for this purpose.

For the sake of completeness we give the definition of a class of functions we shall need but which are not familiar to all physicists. They are the Hardy-class functions[13-15] and can be defined on the unit circle, a half-plane, or even a more general domain, though we shall not consider this later case.

An analytic function $f(z)$ in the upper half-plane ($\mathbb{C}^+$) is called a Hardy class function (H^2_+) if and only if the integrals

$$\int_{-\infty}^{\infty} dx\, |f(x+iy)|^2 < \infty \tag{1.1}$$

are uniformly bounded in y. The definition in the lower half plane ($\mathbb{C}^-$) is similar (and are denoted H^2_-). Similarly an analytic function $\phi(z)$ in the open circle (U) is called a Hardy class function [$H(U)$] if and only if the integrals

$$\int_{-\pi}^{\pi} |f(re^{i\theta})|^2\, d\theta \tag{1.2}$$

are uniformly bounded in r.

The spaces H^p_+, H^p_-, and $H^p(U)$ are defined as above but with the substitution of p ($1 \leqslant p \leqslant < \infty$) for 2.

II. A RIGGED HILBERT SPACE OF HARDY-CLASS FUNCTIONS

In this section we shall set down the space of functions we shall use to implement mathematically the problem of resonances under consideration. To start with, let us construct a space Φ of $\mathscr{L}^2(0,\infty)$ functions $\phi(E)$ such that:

(1) $\phi(E)$ is the restriction to R^+ of a function in H^2_+;

(2) if $\phi(E) \in \Phi$, then $E^n \phi(E) \in \Phi$ for all $n \in N$;

(3) the space Φ can be endowed with a nuclear metriza-

[a)] Permanent address: Facultad de Ciencias, Universidad de Santander, Av. los Castros s/n, Santander, Spain.

 0022-2488/83/061462-08$02.50 © 1983 American Institute of Physics

ble topology;

(4) the space Φ is dense in $\mathscr{L}^2(0,\infty)$.

First of all, we note that an H^2_+ function is completely defined by its value in R^+. Let $\phi(E) \in \mathscr{L}^2(0,\infty)$. $\phi(E)$ will satisfy condition (1) if and only if its Mellin transform $f(s)$ fulfills[16]

$$\int_{-\infty}^{\infty} (1 + e^{2\pi s}) |f(s)|^2 \, ds < \infty. \tag{2.1}$$

The function $\phi(z) = \phi(re^{i\theta})$ defined as

$$\phi(re^{i\theta}) = (2n)^{-1/2} \int_{-\infty}^{\infty} f(s)(re^{i\theta})^{-is-1/2} \, ds, \tag{2.2}$$

$0 \leqslant \theta \leqslant n$,

is the function in H^2_+ whose boundary values in R^+ are given by $\phi(E)$ (a.e.).

In particular for negative values of E we have

$$\phi(-E) = (2n)^{-1/2} \int_{-\infty}^{\infty} f(s)(Ee^{i\pi})^{-is-1/2} \, ds. \tag{2.3}$$

To continue with the construction of the space Φ we need the following theorem[13-15] (Paley–Wiener).

Theorem: The Fourier transform is an isometry between $\mathscr{L}^2(-\infty,0)$ (resp. $\mathscr{L}^2(0,\infty)$) and H^2_+ (resp. H^2_-).

The inverse Fourier transform is an isometry between $\mathscr{L}^2(0,\infty)$ (resp. $\mathscr{L}^2(-\infty,0)$) and H^2_+ (resp. H^2_-).

H^2_+ and H^2_- are closed subspaces of $\mathscr{L}^2(R)$. Therefore, they are Hilbert spaces with the norm inhereted from $\mathscr{L}^2(R)$.

Now let us consider the functions $\mathscr{S}(R^+)$ of the Schwartz space with the support in R^+. Their inverse Fourier transforms are functions in H^2_+. Furthermore, if $\phi(\xi) \in \mathscr{S}(R^+)$, then $(-i)^n (d^n/d\xi^n)\phi(\xi) \in \mathscr{S}(R^+)$ and the inverse Fourier transform of this function is precisely E^n times the inverse Fourier transform of $\phi(\xi)$, which belongs to $\mathscr{S}(R)$.[17]

Consequently we define $\phi \in \Phi$ as the restriction to R^+ of any function $\phi_0 \in \Psi$, where $\Psi = F^{-1}(\mathscr{S}(R^+))$. ($F^{-1}$ denotes the inverse Fourier transform.) It is now clear from the above considerations that condition (2) is also fulfilled. We now need:

Proposition 1: Ψ is dense in H^2_+.

Proof: By the Paley–Wiener theorem, F^{-1} is a unitary mapping between $\mathscr{L}^2(0,\infty)$ and H^2_+. But a unitary mapping preserves dense sets. As $\mathscr{S}(R^+)$ is dense in $\mathscr{L}^2(0,\infty)$, Ψ must be dense in H^2_+.

Proposition 2: Φ is dense in $\mathscr{L}^2(0,\infty)$.

Proof: Consider H^2_{++} the space of the restrictions to R^+ of functions in H^2_+. Van Winter[16] has proven that H^2_{++} is dense in $\mathscr{L}^2(0,\infty)$. Ψ is dense in H^2_+; hence Φ is dense in H^2_{++}. Consequently, Φ is dense in $\mathscr{L}^2(0,\infty)$.

Proposition 3: Φ is a complete nuclear space.

Proof: $\mathscr{S}(R^+)$ is a complete nuclear space. Since the inverse Fourier transform is a bicontinuous bijection on $\mathscr{S}(R)$, Ψ is a complete nuclear space [as a subspace of $\mathscr{S}(R)$].

Now consider $\theta: \Psi \to \Phi$ such that:

$\theta(\phi_0(E)) = \phi_0(E)\theta(E) = \phi(E)$ [$\theta(E)$ being the Heaviside function].

Due to the fact that a function in H^2_+ is completely determined by its values in R^+, θ is one-to-one and onto. We may endow Φ with the initial topology with respect to θ^{-1} and the topology on Ψ. Then Φ is nuclear and complete [and also Frechet because $\mathscr{S}(R)$ is a Frechet space].[11]

The previous propositions tell us that Φ fulfills the required conditions. Let us denote it Φ_+. Henceforth Φ_- will be the space of complex conjugates of elements in Φ_+, so Φ_- will be the values in R^+ of functions which belong to the Schwartz space and are Hardy class in the lower half-plane (H^2_-). We shall call Ψ_+ and Ψ_-, respectively, the functions which have Φ_+ and Φ_- as their restrictions to R^+. Note that $\Psi_+ = \mathscr{S}(R) \cap H^2_+$ and $\Psi_- = \mathscr{S}(R) \cap H^2_-$. There exist bijective correspondences $\theta_\pm$, between Ψ_+ and Φ_+ and also between Ψ_- and Φ_-. The bijections $\theta_\pm$ are continuous, as are the inverses by the closed graph theorem.[18]
($\theta_\pm \Psi_\pm = \Phi_\pm$ by making use of the Heaviside function.) Consequently, we may identify Ψ_+ and Φ_+ as well as Ψ_- and Φ_-.

We have the following four Gel'fand triplets:

$$\Phi_+ \subset \mathscr{L}^2(0,\infty) \subset \Phi'_+, \quad \Phi_- \subset \mathscr{L}^2(0,\infty) \subset \Phi'_-\ ^{19} \tag{2.4}$$

and

$$\Psi_+ \subset H^2_+ \subset \Psi'_+; \quad \Psi_- \subset H^2_- \subset \Phi'_-. \tag{2.5}$$

We also define the sum spaces:

$$\Phi = \Phi_+ + \Phi_- \tag{2.6}$$

and

$$\Xi = \Psi_+ \oplus \Psi_-. \tag{2.7}$$

Ξ is the direct sum of Ψ_+ and Ψ_- because $\mathscr{L}^2(-\infty,\infty)$ is the direct sum of H^2_+ and H^2_-. Hence that Ξ itself is a Frechet complete nuclear space follows from Proposition 3. Φ is not the direct sum of Φ_+ and Φ_- because they do not have trivial intersection (see Appendix). That is an element $\phi \in \Phi_+ \cap \Phi_-$, which, considered as an element of Φ_+, has an extension to R^- which differs from that obtained if ϕ is originally considered as an element as Φ_-. This fact is important and it prevents us from identifying Φ and Ξ.

Remark: Ξ is not all of $\mathscr{S}(R)$. To see this, pick $f \in \mathscr{S}(R)$ then $f = g + h$, where $g \in H^2_+$ and $h \in H^2_-$ (g and h are unique). Take Fourier transforms: $\hat{f} = \hat{g} + \hat{h}$. As $g \in \Psi_+$ and $h \in \Psi_-$, g and h must be Schwartz functions. But, if $g, h \in \mathscr{S}(R)$, then $\hat{g}$, $\hat{h} \in \mathscr{S}(R)$. By the Paley–Wiener theorem, $g \in \mathscr{L}^2(0,\infty)$ and $h \in \mathscr{L}^2(-\infty,0)$, so for $\hat{g}$ and $\hat{h}$ to be in $\mathscr{S}(R)$ it is necessary that $\hat{g}(0) = \hat{h}(0) = 0$, which need not be true. (Choose f a Gaussian function.) Hence Ξ is a closed subspace of $\mathscr{S}(R)$.

$\phi^\mp \in \Phi_\pm$ can also be represented as follows:

$$\phi^\pm(E) = \langle {}^\pm E | \phi^\pm \rangle = \langle E | \phi \rangle, \tag{2.8}$$

where ϕ is some function in a subdomain of $\mathscr{D}(H_0)$, the domain of the free Hamiltonian H_0 in the Schrödinger representation. $\phi^\pm = \Omega^\pm \phi$, where $\Omega^\pm$ are the Møller wave operators.[20] $|E\rangle$ when E runs over R^+ is a complete set of eigenfunctionals for H_0[8-11]:

$$|E^\pm\rangle = \Omega^\pm |E\rangle \tag{2.9}$$

is a complete set of eigenfunctionals for H. (2.9) is defined through the equation

$$\langle \Omega^{\pm} E | \Omega^{\pm} \phi \rangle = \langle E | \phi \rangle. \tag{2.10}$$

The precise meaning of (2.10) will be clarified in a forthcoming article. The extension of (2.9) and (2.10) to negative (nonphysical) values of E is given analogously.

III. DECAYING AND GROWING GAMOW "STATE" VECTORS AND THEIR ASSOCIATION WITH THE RESONANCE POLES OF THE S MATRIX

In the present section we shall apply all the mathematical machinery we have developed so far to the model of resonances described in Ref. 1. Henceforth we shall assume that we are dealing with a two-body resonant scattering process in which the total Hamiltonian $H = H_0 + V$ has an absolutely continuous spectrum only, coinciding with R^+. Furthermore, the S matrix $S(\omega)$ is assumed to have no singularities save for a finite number of pairs of poles in the second sheet in the energy representation, and its behavior in the second sheet, as $|\omega|$ grows to infinity is taken to be regular (i.e., it either vanishes, is bounded, or grows slower than a polynimial in ω). Under such conditions the following equalities hold:

$$\begin{aligned}(\psi^{\text{out}}(t), S\phi^{\text{in}}(t)) &= (\Omega^- \psi^{\text{out}}(t), \Omega^+ \phi^{\text{in}}(t)) \\ &= (\psi^-(t), \phi^+(t)) = (\psi^-(0), \phi^+(0)) \\ &= \int_{R^+} \psi^{-*}(E) S(E + i0) \phi^+(E)\, dE, \end{aligned} \tag{3.1}$$

where $\psi^-(E) \in \Phi_+$ and $\phi^+(E) \in \Phi_-$. $\psi^-(E)$ and $\phi^+(E)$ are vectors in the energy representation. They respectively correspond to $\psi^-(0) = \Omega^- \psi^{\text{out}}$ and $\phi^+(0) = \Omega^+ \phi^{\text{in}}(0)$, which are the same states in the Schrödinger representation. In view of Eq. (2.8), the last integral can also be written:

$$\int_{R^+} \langle {}^-\psi | E^- \rangle S(E + i0) \langle {}^+E | \phi^+ \rangle\, dE, \tag{3.2}$$

where $|E^{\pm}\rangle$ are the generalized eigenvectors associated with the total Hamiltonian H, and they are related to the generalized eigenvectors $|E\rangle$ of the unperturbed Hamiltonian H_0 through Eqs. (2.9) and (2.10). The total Hamiltonian H plays the role of the multiplication operator on Φ.

The integral (3.2) can also be understood as the integral along the positive real line of $\langle {}^+\psi | E^- \rangle \in \Psi_-$ multiplied by $\langle {}^+E | \phi^+ \rangle \in \Psi_-$ times the boundary values in R^+ of the S matrix in the upper half-plane in the second sheet. $\langle {}^-\psi | E^- \rangle = \psi^{-*}(E) \in \Psi_-$ because its complex conjugate is in Ψ_+. We shall consider Ψ_+ and Ψ_-, instead of Φ_+ and Φ_-, as our space of states in the remainder of the present paper. We have chosen Φ_+ and Φ_- as the spaces of states because it is natural to define the scattering wavefunctions as square-integrable functions of the values of the Hamiltonian spectrum. We can choose Ψ_+ and Ψ_- instead of Φ_+ and Φ_- for three reasons. First, these spaces are (respectively) isomorphic. Secondly, Ψ_+ and Ψ_- have a trivial intersection so we can split the total space of states into two disjoint (apart from the zero vector) pieces which correspond to "decaying" and "growing" states. Finally we do so because we are not restricting ourselves to their values in R^+, and so are working with all the values of these functions (including the ones in the half planes).

Consider now the integral

$$\int_0^{\infty} \langle {}^-\psi | E - i\epsilon^- \rangle S_{II}(E - i\epsilon) \langle {}^+E - i\epsilon | \phi^+ \rangle\, dE, \tag{3.3}$$

which converges to (3.2) as $\epsilon \to 0$ by the Lebesgue theorem.[21] As we have a finite number of singularities in the second sheet, we can choose ϵ such that it is smaller than the imaginary part of any of the singularities.

In order to simplify the problem, we shall assume first that we have a lonely singularity in the lower half-plane in the second sheet. This will be located at the point $z_R = E_R - (i\Gamma/2)$. As in Ref. 1, the path of integration can be deformed in this half-plane; hence, (3.3) can be written

$$\begin{aligned} &-\delta(\epsilon) + (\psi^-(t), \phi^+(t)) \\ &\quad = \int_{C_-} d\omega\, \langle {}^-\psi | \omega^- \rangle S_{II}(\omega) \langle {}^+\omega | \phi^+ \rangle \\ &\qquad + \int_{\Delta} d\omega\, \langle {}^-\psi | \omega^- \rangle \frac{S_{-1}}{\omega - z_R} \langle {}^+\omega | \phi^+ \rangle, \end{aligned} \tag{3.4}$$

S_{-1} being the residium of $S(\omega)$ in the second sheet at z_R and $S_{II}(\omega)$ denoting the S matrix in the second sheet. The path C_- circles below the pole at z_R and is a circular path around z_R, as shown in Fig. 1. Also $\delta(\epsilon)$ is a complex function of ϵ which vanishes as $\epsilon \to 0$. On the other hand, if $\langle {}^-E | \psi^- \rangle \in \Psi_+$, then $\langle {}^-\psi | E^- \rangle = \langle {}^-E | \psi^- \rangle^* \in \Psi_-$ and $\langle {}^-\psi | \omega^- \rangle = \langle {}^-\omega^* | \psi^- \rangle^*$. The contour C_- can be deformed into a contour along the lower infinite semicircle. The path Δ in the second integral in (3.4) can also be deformed into an infinite semicircle, and the line $E - i\epsilon = y$ $(-\infty < E < \infty$, in the second sheet; see Fig. 2).

We now need the following result:

Lemma: Both integrals vanish along the semicircles as the radii go to infinity.

Before giving the proof we make a statement of a very useful theorem, known as the Titchmarsh theorem.

Theorem[13-15] **(Titchmarsh):** If $f \in H^p_-$ $(1 \leqslant p \leqslant \infty)$, then

$$f(\omega) = -\frac{1}{2\pi i} \int_{-\infty}^{\infty} \frac{f(t)}{t - \omega} dt, \quad \text{Im}\{\omega\} = y < 0, \tag{3.5}$$

and the integral vanishes for all $y > 0$.

Conversely, if $f \in \mathcal{L}^p(-\infty, \infty)$ $(1 \leqslant p \leqslant \infty)$ and (3.5) is fulfilled, then $f \in H^p_-$. Similar results can be proven when $f \in H^p_+$.

By applying the theorem it is clear that if $E^n \langle {}^+E | \phi^+ \rangle$

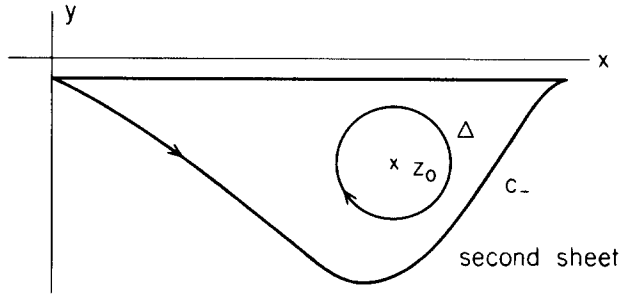


FIG. 1.

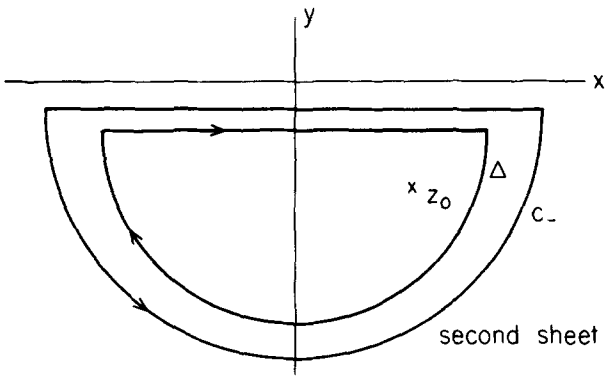


FIG. 2.

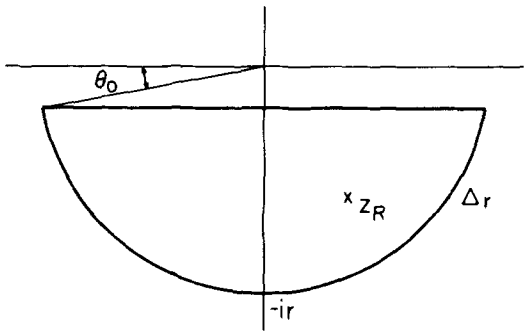


FIG. 2'.

is a function in H^2_-, then in the lower half-plane it is defined as $\omega^n \langle {}^+\omega|\phi^+\rangle$. This follows from the fact that

$$\frac{1}{2\pi i}\int_{-\infty}^{\infty}\frac{E\langle {}^+E|\phi^+\rangle\,dE}{E-\omega} = \frac{1}{2\pi i}\int_{-\infty}^{\infty}\langle {}^+E|\phi^+\rangle\,dE + \frac{1}{2\pi i}\int_{-\infty}^{\infty}\frac{\omega\langle {}^+E|\phi^+\rangle}{E-\omega}\,dE \tag{3.6}$$

The first integral in the right-hand side of (3.6) is the value of the Fourier transform of $\langle {}^+\phi|E^+\rangle$ at the origin, which is zero because of the Paley–Wiener theorem and the fact that it is a Schwartz function. The second one is

$$\frac{\omega}{2\pi i}\int_{-\infty}^{\infty}\frac{\langle {}^+E|\phi^+\rangle}{E-\omega}\,dE = \omega\langle {}^+\omega|\phi^+\rangle. \tag{3.7}$$

Extending by induction to all values of n, we obtain the desired result.

Proof of the Lemma: Consider the path Δ_r in Fig. 2'. By the Cauchy theorem it follows that:

$$2\pi i\langle {}^-\psi|z_R^-\rangle s_{-1}\langle {}^+z_R|\phi^+\rangle = \int_{-r\cos\theta_0}^{r\cos\theta_0} dE\,\frac{\langle {}^-\psi|E-i\epsilon^-\rangle s_{-1}\langle {}^+E-i\epsilon|\phi^+\rangle}{E-i\epsilon-z_R} + \int_{\theta_0}^{\pi-\theta_0}\frac{F(re^{i\theta})}{re^{i\theta}-i\epsilon-z_R}\,ire^{i\theta}\,d\theta, \tag{3.8}$$

where

$$F(re^{i\theta}) = s_{-1}\langle {}^-\psi|\omega^-\rangle\langle {}^+\omega|\phi^+\rangle, \tag{3.9}$$

and $\omega\in\Delta_r$. The last integral for very large r is bounded by

$$\mathcal{J} = \int_{\theta_0}^{\pi-\theta_0}|F(re^{i\theta})|\left[1+o\left(\frac{1}{r}\right)\right]d\theta, \tag{3.10}$$

$F(re^{i\theta})\in H^1_-$. Then

$$|F(re^{i\theta})|\leqslant K(r\sin\theta)^{-1}, \tag{3.11}$$

where K is a constant.[15] We have now: $\sin\theta\geqslant$ const θ on $(\theta_0,\pi/2]$ and $\sin\theta\geqslant$ const $(\pi-\theta)$ on $[\pi/2,\pi-\theta_0)$ whence

$$\mathcal{J}\leqslant\frac{K'}{r}\int_{\theta_0}^{\pi/2}\frac{d\theta}{\theta}\leqslant\frac{K'}{r}\log\frac{\pi r}{2\epsilon}, \tag{3.12}$$

which vanishes when $r\to\infty$. We have hence proven the first part of the lemma. If we not let $\epsilon\to 0$, we obtain the result:

$$2\pi i\langle {}^-\psi|z_R^-\rangle s_{-1}\langle {}^+z_R|\phi^+\rangle = \int_{-\infty}^{\infty}dE\frac{\langle {}^-\psi|E^-\rangle s_{-1}\langle {}^+E|\phi^+\rangle}{E-z_R}. \tag{3.13}$$

Note that the same result is induced from the Titchmarsh theorem, and hence we have an alternative proof to the first part of the lemma, Nevertheless, we need the above machinery to prove that the first integral in (3.4) vanishes. Let us proceed with the proof.

First, note that we have essentially proved that

$$\int_{C_-}\left|d\omega\,\langle {}^-\psi|\omega^-\rangle\frac{1}{\omega-z_R}\langle {}^+\omega|\phi^+\rangle\right|\to 0,\quad r\to\infty, \tag{3.14}$$

for all $\langle {}^+\omega|\phi^+\rangle\in\Psi_-$ and $\langle {}^-\omega|\psi^-\rangle\in\Psi_+$. What we do now is to make an assumption on the behavior of the S matrix in the second sheet. Suppose that for large ω there exists a polynomial $P(\omega)$ such that in the second sheet

$$|S_{II}(\omega)|\leqslant|P(\omega)|. \tag{3.15}$$

Then

$$\int_{C_-}|d\omega\,\langle {}^-\psi|\omega^-\rangle S_{II}(\omega)\langle {}^+\omega|\phi^+\rangle| \leqslant\int_{C_-}|d\omega\,\langle {}^-\psi|\omega^-\rangle P(\omega)\langle {}^+\omega|\phi^+\rangle| = \int_{C_-}|d\omega\,\langle {}^-\psi|\omega^-\rangle\frac{\omega-z_R}{\omega-z_R}P(\omega)\langle {}^+\omega|\phi^+\rangle|. \tag{3.16}$$

If we write $\langle {}^+\omega|\xi^+\rangle = (\omega-z_R)P(\omega)\langle {}^+\omega|\phi^+\rangle\in\Psi_-$, then

$$(3.16) = \int_{C_-}|d\omega\langle {}^-\psi|\omega^-\rangle\frac{1}{\omega-z_R}\langle {}^+\omega|\xi^+\rangle|\to 0, \quad r\to\infty \tag{3.17}$$

by (3.14). Hence, we have demonstrated that the first integral vanishes and the lemma is thus proved.

From (3.14) and (3.13) we have following:

$$(\psi^-(t),\phi^+(t)) = \text{background} - 2\pi i s_{-1}\langle {}^-\psi|z_R^-\rangle\langle {}^+z_R|\phi^+\rangle, \tag{3.18}$$

where background denotes

$$\int_{0\,II}^{-\infty}dE\,\langle {}^-\psi|E^-\rangle S_{II}(E)\langle {}^+E|\phi^+\rangle. \tag{3.19}$$

As we shall see later, the background term is the responsible for the decay being nonexponential.

Next we consider a sort of conjugate problem. Let us write

$$(\phi^+(t),\psi^-(t)) = \int_{R^+} dE\, \langle {}^+\phi\,|E^+\rangle S^*(E+i0)\langle {}^-E\,|\psi^-\rangle$$
$$= \int_{R^+} dE\, \langle {}^+\phi\,|E^+\rangle S(E-i0)\langle {}^-E\,|\psi^-\rangle, \tag{3.20}$$

where the second equality is motivated by the properties of the S matrix. Here $\langle {}^+\phi\,|E^+\rangle$ and $\langle {}^-E\,|\phi^-\rangle$ belong to H^2_+, and $S(E-i0)$ has a continuation in the upper half-plane in the second sheet, which is analytical except for a pole. This pole is located at $z_R = E_R + i\Gamma/2$ (second sheet). We can move and deform the path of integration into the upper half-plane in the second sheet in a similar way as we did before in the lower half-plane and obtain the result

$$-\delta'(\epsilon) + (\phi^+(t),\psi^-(t))$$
$$= \int_{C_+} \langle {}^+\phi\,|\omega^+\rangle S_{II}(\omega)\langle {}^-\omega|\psi^-\rangle\, d\omega$$
$$+ \int_{\Delta'} \langle {}^+\phi\,|\omega^+\rangle \frac{s_1}{\omega - z_R^*} \langle {}^-\omega|\psi^-\rangle\, d\omega, \tag{3.21}$$

where s_1 is the residuum of $S_{II}(\omega)$ at $E_R + i\Gamma/2$, $\delta'(\epsilon)$ is a function which vanishes as $\epsilon\to 0$, and C_+ and Δ' are paths as shown in Fig. 3. Again we can prove that the integral over both semicircles vanishes, and we finally obtain

$$(\phi^+(t),\psi^-(t))$$
$$= \text{background}^* + 2\pi i s_{-1}\langle {}^+\phi\,|z_R^{*+}\rangle\langle -z_R^*|\psi^-\rangle, \tag{3.22}$$

where again

$$\text{background}^* = \int_{0\,II}^{-\infty} dE\, \langle {}^+\phi\,|E^+\rangle S_{II}(E)\langle {}^-E\,|\psi^-\rangle, \tag{3.23}$$

which is the complex conjugate of (3.18).

If we omit in (3.18) the arbitrary vector $\psi^-\in\Psi_+$ and in (3.22) $\phi^+\in\Psi_-$, we have

$$\phi^+ = \int_{0\,II}^{-\infty} dE\, |E^-\rangle S_{II}(E)\langle {}^+E\,|\phi^+\rangle$$
$$- 2\pi i s_{-1}|z_R^-\rangle\langle {}^+z_R\,|\phi^+\rangle, \tag{3.24}$$
$$\psi^- = \int_{0\,II}^{-\infty} dE\, |E^+\rangle S_{II}(E)\langle {}^-E\,|\psi^-\rangle$$
$$+ 2\pi i s_1 |z_R^{*+}\rangle\langle {}^-z_R^*|\psi^-\rangle. \tag{3.25}$$

The integrals in (3.24) and (3.25) exist only in the weak sense. They have sense in the energy representation as well as in the Schrödinger representation. Let us assume for a while we are in the latter. ϕ^+ is a decaying state. It has been prepared in the remote past as ϕ^{in} and commences decay at time $t=0$. ψ^- is a growing state. It will be detected in the far future as ψ^{out} and will cease forming at time $t=0$.[22]

According to (3.24) and (3.25), ϕ^+ (ψ^-) as a functional on Ψ_+ (Ψ_-) is the sum of two objects. That they indeed belong to Ψ'_+ (Ψ'_-) follows from the next proposition.

Proposition 4: $|z_R^-\rangle$ and $|z_R^{*+}\rangle$ are continuous antilinear functionals over Ψ_+ and Ψ_-, repectively.

Proof: It is straightforward to prove that they are antilinear functionals, respectively, over these two spaces. We now need to prove their continuity. As in the linear case, an antilinear functional f on a locally convex space X is continuous if and only if there exists a finite number n of continuous seminorms $p_1, p_2, \ldots, p_n$ on X, and n constants $c_1, c_2, \ldots, c_n$ such that for every $x\in X$

$$|f(x)| \leqslant c_1 p_1(x) + c_2 p_2(x) + \cdots + c_n p_n(x). \tag{3.26}$$

We shal prove our statement with $|z_R^-\rangle$. The topology on Ψ_+ is the one inhereted from that on the Schwartz space. By the Tichmarsh theorem and (3.5) follows:

$$\langle {}^-\psi|z_R^-\rangle = \frac{1}{2\pi i}\int_{-\infty}^{\infty} \frac{dE\,\langle {}^-\psi|E^-\rangle}{E - z_R}$$
$$= \frac{1}{2\pi i}\int_{-\infty}^{\infty} dE\, \frac{(E-z_R)\langle {}^-\psi|E^-\rangle}{(E-z_R)^2}. \tag{3.27}$$

Hence

$$|\langle {}^-\psi|z_R^-\rangle| \leqslant \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{|(E-z_R)\langle {}^-\psi|E^-\rangle|}{|E-z_R|^2}\, dE$$
$$\leqslant \frac{1}{2\pi}\left[\int_{-\infty}^{\infty} \frac{dE}{|E-z_R|^2}\right] \sup_{E\in R} |(E-z_R)\langle {}^-\psi|E^-\rangle|\, dE$$
$$\leqslant \frac{1}{2\pi}\left[\int_{-\infty}^{\infty} \frac{dE}{|E-z_R|^2}\right]\Big[\sup_{E\in R}|E\,\langle {}^-\psi|E^-\rangle|$$
$$+ |z_R|\sup_{E\in R}|\langle {}^-\psi|E^-\rangle|\Big] dE. \tag{3.28}$$

The integral inside the parenthesis is clearly finite and both suprema in the bracket are seminorms (indeed norms) on $\mathcal{S}(R)$, hence on Ψ_+; therefore, $|z_R^-\rangle\in\Psi'_+$. There is an analogous result for $|z_R^{*+}\rangle\in\Psi'_-$.

Corollary 1: The integral in (3.24) ((3.25)) is a continuous antilinear functional on Ψ_+ (Ψ_-)

The proof is obvious.

Corollary 2: $|z_R^-\rangle$ ($|z_R^{*+}\rangle$) define an element in Φ'_+ (Φ'_-), its range in $\mathbb{C}$ coinciding with that it has as a functional on Ψ_+ (Ψ_-).

Proof: This element is $\theta^{-1}\circ|z_R^-\rangle$ ($\theta^{-1}\circ|z_R^{*+}\rangle$), which obviously is continuous.

The functionals $|z_R^-\rangle$ and $|z_R^{*+}\rangle$ are generalized eigenfunctionals of the total Hamiltonian H. The total Hamiltonian corresponds with the multiplication operator on Ξ. First note that H is continuous on Ψ_+ and Ψ_-, which follows directly from the fact that

$$\langle {}^\pm H\phi\,|E^\pm\rangle = \langle {}^\pm\phi\,|HE^\pm\rangle = E\,\langle {}^\pm\phi\,|E^\pm\rangle\in\Psi_\mp \tag{3.29}$$

with all $\langle {}^\pm\phi\,|E^\pm\rangle\in\Psi_\mp$.

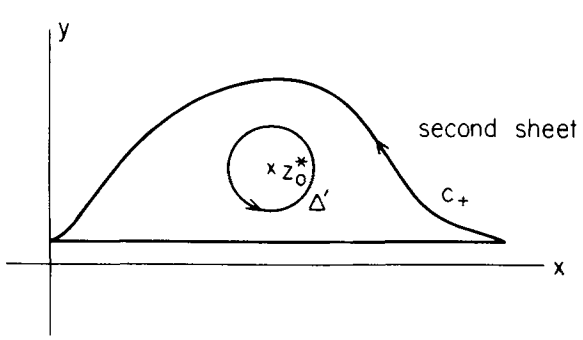


FIG. 3.

If we use the Titchmarsh theorem

$$\langle {}^-\psi|Hz_R^-\rangle = \langle {}^-H\psi|z_R^-\rangle = \int_{-\infty}^{\infty} dE\,\frac{\langle {}^-H\psi|E^-\rangle}{E-z_R}$$

$$= \int_{-\infty}^{\infty} dE\,\frac{E\langle {}^-\psi|E^-\rangle}{E-z_R} = z_R\langle {}^-\psi|z_R^-\rangle. \quad (3.30)$$

Hence

$$H|z_R^-\rangle = z_R|z_R^-\rangle. \quad (3.31)$$

In an analogous way we can show

$$H|z_R^{*+}\rangle = z_R^*|z^{*+}\rangle. \quad (3.32)$$

First of all, we should note that (3.31) and (3.32) are not valid on Ψ_+ and Ψ_-, but also on Φ_+ and Φ_-. The elements in $\Phi_\pm$ are functions on the energy E, but they can also be considered as the square-integrable functions of the position whose image in the energy representation is in $\Phi_\pm$. In the Schrödinger rerpesentation the evolution operator is $U(t) = e^{-itH}$, which is e^{-itE} in the energy representation. $U(t) = e^{-itE}$ is an operator on $\Phi_\pm$ and can be defined on $\Psi_\pm$ by adding the nonphysical negative values of the energy. The question is: How do $\Psi_\pm$ behave under the action of $U(t)$?

To begin with, observe that if $t<0, e^{-itE}\Psi_+ \subset \Psi_+$. In fact, if $\phi^-(E)\in\Psi_+$, $e^{-itE}\phi^-(E)$ belongs to $\mathscr{S}(R)$. Furthermore, its extension to the upper half of the plane ($\mathbb{C}^+$) is given by

$$e^{-itz}\phi^-(z) = e^{-itx}e^{ty}\phi^-(z) \quad (3.33)$$

when $z = x+iy\in\mathbb{C}^+$. If $t<0$, it is clear after (3.33) that this function belongs to H^2_+. Therefore, $e^{-itE}\phi^-(E)\in\Psi_+$. Analogously, if $t>0$, $e^{-itE}\psi^+(E)\in\Psi_-$.

On the other hand, if we have a Gel'fand triplet $\Phi\subset\mathscr{H}\subset\Phi'$ and a non-self-adjoint operator A on $\mathscr{H}$, we can define A on Φ' in the following way: (some authors call A^X to this extension) if $F\in\Phi$,

$$\langle A^*\phi|F\rangle = \langle\phi|AF\rangle \quad \text{for all } \phi\in\Phi \quad (3.34)$$

provided that Φ is stable under the action of A^* the Hilbert space adjoint of A.[8,10]

In our case $\Phi = \Psi_\pm$ and $A = U(t) = e^{-itH} = e^{-itE}$. $\Psi_\pm$ are subspaces of $\mathscr{L}^2(R)$. The Hilbert space adjoint of e^{-itH} is e^{itH}.

If $t>0$, $\psi^-\in\Psi_+$, and $e^{-itH}\Psi_+\subset\Psi_+$, then

$$\langle {}^-(e^{-itH})^*\psi|z_R^-\rangle = \langle {}^-e^{itH}\psi|z_R^-\rangle$$
$$= \langle {}^-\psi|e^{-itH}z_R^-\rangle \quad [\text{or} = \langle {}^-\psi|(e^{-itH})^X z_R^-\rangle] \quad (3.35)$$

by definition (3.33) and above considerations.

On the other hand,

$$\langle {}^-e^{itH}\psi|z_R^-\rangle = \frac{1}{2\pi i}\int_{-\infty}^{\infty} dE\,\frac{\langle {}^-e^{itH}\psi|E^-\rangle}{E-z_R}$$
$$= \frac{1}{2\pi i}\int_{-\infty}^{\infty} dE\,\frac{\langle {}^-\psi|E^-\rangle e^{-itE}}{E-z_R} = e^{-itz_R}\langle {}^-\psi|z_R^-\rangle, \quad (3.36)$$

which yields together with (3.35), omitting the arbitrary $\psi^-\in\Psi_+$,

$$U(t)|z_R^-\rangle = e^{-itz_R}|z_R^-\rangle = e^{-itE_R}e^{-t\Gamma/2}|z_R^-\rangle \quad (t>0). \quad (3.37)$$

Analogously,

$$U(t)|z_R^{*+}\rangle = e^{-itz_R^*}|z_R^{*+}\rangle = e^{-itE_R}e^{t\Gamma/2}|z_R^{*+}\rangle \quad (t<0). \quad (3.38)$$

(3.37) shows that the decaying state ϕ^+ is split into two parts. One, which corresponds with the functional $|z_R^-\rangle$, decays exponentially, and the other one, the background, is responsible for the deviations from the exponential law of the decaying state ϕ^+. Its counterpart (3.38) shows that the growing state is split into two parts. One of them, which corresponds with the functional $|z_R^{*+}\rangle$, grows exponentially, and the other one, the background, is responsible for deviations of the exponential law of the formation state.

Equation (3.37) ((3.38)) does not hold for $t<0$ ($t>0$) because Ψ_+ (Ψ_-) is not stable under e^{itH} with $t<0$ ($t>0$). The proof comes after the following result:

Proposition 5: For any $t>0$ ($t<0$), there exists a function $\phi\in\Psi_+$ ($\psi\in\Psi_-$) such that $U(t)\notin\Psi_+$ ($U(t)\notin\Psi_-$).

Proof: Pick any positive number t. The Fourier transform of $e^{-itE}\phi(E)$ is $\hat\phi(t+\xi)$, where $\hat\phi(\xi)$ is the Fourier transform of $\phi(E)\in\Psi_+$. $\hat\phi(\xi)$ is supported in $[0,\infty)$ because of the Paley–Wiener theorem; hence $\hat\phi(t+\xi)$ is supported in $[-t,\infty)$. Pick α such that $0<\alpha<t$, and let $f(\xi)$ be a $\mathscr{S}(R)$ function supported in $[\alpha,-\alpha]$. Its inverse Fourier transform $\check f(E)\notin H^2_+$. If we have $f(\xi-t)$, then this function is supported in $[0,\infty)$ and its inverse Fourier transform does belong to H^2_+. Let us call it $\psi(E)$. Then clearly $e^{-itE}\psi(E) = \check f(E)$. The instability of Ψ_- is proven analogously.

We learn from the last propositions the following important fact: Time evolution is not longer produced by a one-parameter group of unitary operators but by two semigroups. The one corresponding with negative values of time operators on Ψ_+, and that corresponding with positive values of time operates on Ψ_-.

We not prove our final result.

Proposition 6: $|z_R^-\rangle$ and $|z_R^{*+}\rangle$ are normalizable vectors in $\mathscr{H}$. In fact, $|z_R^-\rangle\in H^2_+$ and $|z_R^{*+}\rangle\in H^2_-$.

Proof: Let U be the unit circle in the complex plane. If $g\in H^2(U)$, we have for any $\omega_0\in U$[23]

$$|g(\omega_0)|^2 \leqslant \frac{k}{2\pi}\int_{-\pi}^{\pi}|g(e^{i\theta})|^2\,d\theta. \quad (3.39)$$

Since any function $f(z)\in H^2_+$ can be written as

$$f(z) = (1+z)^{-1}g\left(\frac{z-i}{z+i}\right), \quad (3.40)$$

where $g(\omega)\in H(U)$, we can write by picking $z_0\in\mathbb{C}^+$

$$|f(z_0^*)| = |1+z_0^*|^{-2}\left|g\left(\frac{z_0^*-i}{z_0^*+i}\right)\right|$$
$$\leqslant \frac{k_0}{2\pi}\int_{-\pi}^{\pi}|g(e^{i\theta})|^2\,d\theta = \frac{k_0}{\pi}\int_{-\infty}^{\infty}|f(x)|^2\,dx \quad (3.41)$$

with

$$k_0 = k/|1+z_0^*|^2. \quad (3.42)$$

Thus $f\to f^*(z_0^*)$ is a continuous antilinear functional on H^2_+. Hence there exists $\xi_{z_0}\in H^2_+$ such that

$$\langle f|\xi_{z_0}\rangle = f^*(z_0). \quad (3.43)$$

By considering $f \equiv |\psi^-\rangle$ and $z = z_0$, we have $|z_R^-\rangle = \xi_{z_0} \in H^2_+$. This proves the first part of our assertion. In the same manner we find that $|z_R^{*+}\rangle \in H^2_-$.

Remark 1: Note that Proposition 6 could have been directly proven by using the Titchmarsh theorem and the fact that $1/E - z_R^*$ and $1/E - z_R$ are in $\mathscr{L}^2(R)$. The norms of $|z_R^-\rangle$ and $|z_R^{*+}\rangle$, considered as functionals, respectively, on H^2_+ and H^2_-, are equal to

$$\frac{1}{2\pi}\left|\left|\frac{1}{E - z_R^*}\right|\right| = \frac{1}{2\pi}\left|\left|\frac{1}{E - z_R}\right|\right| = \frac{1}{(2\pi\Gamma)^{1/2}} \tag{3.44}$$

since

$$\xi_{z_0} = \frac{1}{2\pi i}\frac{1}{E - z_R^*} \quad \text{and} \quad \xi_{z_0^*} = \frac{1}{2\pi i}\frac{1}{E - z_R}. \tag{3.45}$$

Remark 2: $|z_R^-\rangle$ and $|z_R^{*+}\rangle$ belong to the Hilbert space $\mathscr{L}^2(-\infty, \infty)$. If, however, we consider them as functionals on Ψ_+ and Ψ_-, we note that their time evolutions do not coincide with the values of the curves $e^{-itH}\xi_{z_0}$ and $e^{-itH}\xi_{z_0^*}$ in $\mathscr{L}^2(-\infty, \infty)$, where e^{-itH} is defined over all $\mathscr{L}^2(-\infty, \infty)$ via Stone's theorem. Furthermore, $e^{-itH}\xi_{z_0}$ and $e^{-itH}\xi_{z_0^*}$ are defined for all values t (in the Stone therorem). The behavior of the time evolution of $|z_R^-\rangle$ and $|z_R^{*+}\rangle$ is consequently different from the above values in Hilbert space. On the other hand, it is easy to see that ξ_{z_0} and $\xi_{z_0^*}$ do not belong to the domain of H.[24] We recall that the action of H on $|z_R^-\rangle$ and $|z_R^{*+}\rangle$ is defined through the extension of H (sometimes called H^X) to Ψ'_+ and Ψ'_-. When considering the evolution operator $U(t)$, it has an extension to $\mathscr{H}$ given by the Stone theorem. Let us call it e^{-itH}. It is a unitary mapping preserving norms. When $t > 0$, $U(t)$ admits an extension to Ψ'_+ and, therefore to $\mathscr{H}$ as subspace of Ψ'_+. If we call it $(e^{-itH})^X$, we have

$$(e^{-itH})^X |z_R^-\rangle \neq e^{-itH}\xi_{z_0} \quad (t > 0) \tag{3.46}$$

because the first term of this inequality does not preserve norms [see (3.37)]. Moreover, both terms in (3.46) are indeed elements of different spaces with different topologies. The right hand side belongs to Ψ'_+ with its natural topology as the dual space of Ψ_+. The left hand side belongs to $\mathscr{H}$ with its Hilbert space topology. The same situation appears when $t < 0$.

Remark 3: If we have more than one pair of poles (3.18) can be written in the lower half-plane as

$$(\psi^-, \phi^+) = \text{background} - 2\pi i \sum_{k=1}^{N} \langle {}^-\psi | z_{R_k}^-\rangle \langle {}^+z_{R_k} | \phi^+\rangle s_{-k}. \tag{3.47}$$

An analogous formula for (3.22) can also be written. Here the poles in the second sheet are placed at the points z_{R_k}, and s_{-k} is the residium corresponding to z_{R_k}. The vectors $|z_{R_k}^-\rangle$ decay exponentially and belong to H^2_-, and ψ^- is a linear combination of them plus the background.

ACKNOWLEDGMENTS

The author wishes to express his gratitude to Professor A. Bohm for having suggested him this problem, for stimulating discussions and for his great encouragement, also to Professor J. E. Gilbert for helpful discussions and for giving him important suggestions on the material in the appendix, and finally to Professor K. Kraus for his continuing interest in the present research.

APPENDIX

Consider the space $\Phi_+ \cap \Phi_-$. We prove in this appendix that there exist nonzero functions belonging to it. The proof is as follows:

If $\phi \in \Phi_+$, consider $\phi_+(E)$, its continuation to R^- such that $\phi_+(E) \in H^2_+$. We already know that this continuation is unique. The same is true with $\phi_-(E) \in H^2_-$.

We first look for functions which can be extended in two different ways into R_-, where one extension is in H^2_+ and the other is in H^2_-.

Let $\mathfrak{H}$ be the Hilbert transform, i.e., if $f \in \mathscr{L}^2(-\infty, \infty)$

$$(\mathfrak{H}f)(t) = \frac{P}{\pi}\int_{-\infty}^{\infty} \frac{f(x)}{x - t}\,dx, \tag{A1}$$

where P denotes the principal value.

Now consider

$$\tfrac{1}{2}(\mathfrak{H} - iI)f = g, \tag{A2}$$

where I is the identity in $\mathscr{L}^2(-\infty, \infty)$. g is an H^2_+ function. This is a straightforward cosequence of a Titchmarsh theorem.[25] For the same reason

$$\tfrac{1}{2}(\mathfrak{H} + if) = h \tag{A3}$$

is a H^2_- function.

By choosing a square integrable function f with support in $\mathscr{L}^2(-\infty, \infty)$, we have[26]

$$\tfrac{1}{2}(\mathfrak{H} + iI)f|_{(0,\infty)} = \tfrac{1}{2}(\mathfrak{H} - iI)f|_{(0,\infty)} = \tfrac{1}{2}\mathfrak{H}f|_{(0,\infty)}. \tag{A4}$$

This implies that the functions we are looking for are the restrictions to $(0, \infty)$ of Hilbert transforms in $\mathscr{L}^2(-\infty, 0)$.

We now look for additional conditions we have to impose if we want $\mathfrak{H}f|_{(0,\infty)}$ to be the restriction of a Schwartz function.

First, note that

$$(\mathfrak{H}f)\hat{} = i\,\mathrm{sgn}(\xi)\hat{f}(\xi), \tag{A5}$$

where

$$\mathrm{sgn}(\xi) = \xi/|\xi|. \tag{A6}$$

From (A5) we obtain that $(\mathfrak{H}f)\hat{}$ and hence $\mathfrak{H}f$ would be a Schwartz function if and only if

$$\frac{d^n}{d\xi^n}\hat{f}(\xi)\bigg|_{\xi=0} = 0 \quad \text{for } n = 0,1,2,..., \tag{A7}$$

and $\hat{f}$ (and consequently f) are Schwartz functions.

Finally we look for the existence of a Schwartz function supported in $(-\infty, 0)$, such that its Fourier transform satisfies (A7).

If $\hat{f}$ is the Fourier transform of a function f supported in $(-\infty, 0)$, we have from the Paley–Wiener theorem that

$\hat{f} \in H^2_+$. The question this reduces to the existence of a function $g \in \Psi_+$ such that (A7) holds.

They have been several attempts to solve the equivalent problem in the unit circle. We use the method of Reference 27.

Let E a closed subset of ∂U = boundary of the unit circle. E is called a Carleson set if

$$\int_{-\pi}^{\pi} \log \rho(e^{i\theta}, E)\, d\theta > -\infty \tag{A8}$$

where $\rho(z,E)$ = distance between z and E.

Theorem (Taylor and Williams): Let E be a Carleson set. Then there exists functions in $A^\infty = C^\infty(\partial U) \cap H(U)$ such that they and their derivatives vanish in E.[27]

A single point is obviously a Carleson set. Hence we can find functions in A^∞ such that they and all their derivatives are zero in, say $-i$, Let $\phi(z)$ be such a function. As $\phi(z) \in A^\infty$,

$$g(\omega) = \phi\left(\frac{\omega - i}{\omega + i}\right) \in C^\infty(R) \cap H^2_+ . \tag{A9}$$

Let $f(\omega)$ be any function in Ψ_+. Consider the product $h(\omega) = f(\omega) g(\omega)$.

Indeed $h(\omega) \in \mathcal{S}(R) \cap H^2_+$. This implies that

$$h(x)|_{(0,\infty)} \in \Phi_+ . \tag{A10}$$

We have shown that there exists functions in Ψ_+ such that $f^{(n)}(0) = 0$. This implies that $\check{f}$ is supported in $(-\infty, 0)$ and

$$\mathcal{F}\check{f}|_{(0,\infty)} \in \Phi_+ \cap \Phi_- . \tag{A11}$$

[1] A. Böhm, "Resonance poles and quasistationary state vectors in the rigged Hilbert space formulation of quantum mechanics," J. Math. Phys. **22**, 2813 (1981).

[2] M. Reed and B. Simon, *Analysis of Operators* (Academic, New York, 1978).

[3] H. Nussenzveig, *Causality and Dispersion Relations* (Academic, New York, 1972).

[4] See arguments in favor of this statement in A. Bohm, *Quantum Mechanics* (Springer-Verlag, New York and Berlin, 1979); B. Simon, Int. J. Quant. Chem. **XIV**, 524 (1978), and also Ref. 1.

[5] For other arguments, see L. Fonda, G. C. Ghirardi, and A. Rimini, Rep. Prog. Phys., **41**, 578 (1978) and references quoted therein, mainly G. Calucci and C. G. Ghirardi, Phys. Rev. **169**, 1339 (1968).

[6] H. Baumgärtel, Math. Nachr. **75**, 133 (1976).

[7] E. C. G. Sudarshan, C. B. Chiu, and V. Gorini, Phys. Rev. D **18**, 2914 (1978); G. Parravicini, V. Gorini, and E. C. G. Sudarshan, J. Math. Phys. **21**, 2208 (1980).

[8] O. Melsheimer, J. Math. Phys. **15**, 902 (1974).

[9] A. Bohm, *The Rigged Hilbert Space and Quantum Mechanics*, Lecture Notes in Physics **78** (Springer, New York, 1978).

[10] J. P. Antoine, J. Math. Phys. **10**, 53 (1969).

[11] J. E. Roberts, Comm. Math. Phys. **3**, 98 (1966).

[12] I. M. Gel'fand and N. Y. Vilenkin, *Generalized Functions: Applications of Harmonic Analysis* (Academic, New York, 1964), Vol. 4; Albrecht Pietsch, *Nuclear Locally Convex Spaces* (Springer-Verlag, Berlin, 1972), and also Refs. 8, 9, 10 and 11.

[13] P. Duren, *Theory of H_p Spaces*, Pure and Applied Mathematics, Vol. 28 (Academic, New York, 1970).

[14] K. Hoffman, *Banach Spaces of Analytic Functions* (Prentice-Hall, Englewood Cliffs, N.J., 1962).

[15] P. Koosis, *Introduction to H_p Spaces*, London Math. Soc., Lecture Note Series No. 40 (Cambridge U.P., Cambridge, 1980).

[16] C. van Winter, Trans. Am. Math. Soc. **162**, 103 (1971); J. Math. Ann. **47**, 633 (1974).

[17] M. Reed and B. Simon, *Fourier Analysis. Self Adjointness* (Academic, New York, 1975).

[18] See, for instance, J. Horvath, *Topological Vectors Spaces and Distributions* (Addison-Wesley, London, 1966), p. 301.

[19] We always consider the duals as spaces of continuous antilinear functionals.

[20] W. Amrein, J. Jauch, and B. Sinha, *Scattering Theory in Quantum Mechanics* (Benjamin, New York, 1977); M. Reed and B. Simon, *Scattering Theory* (Academic, New York, 1979).

[21] See, for instance, A. Friedman, *Foundations of Modern Analysis* (Holt, Rinehart, and Winston, New York, 1970), p. 54.

[22] A. Bohm, Lett. Math. Phys. **3**, 455 (1979). and also Ref. 1.

[23] W. Rudin, *Real and Complex Analysis* (McGraw-Hill, New York, 1970), p. 348.

[24] A. Bohm, J. Math. Phys. **21**, 1040 (1980).

[25] E. C. Titchmarsh, *The Theory of Functions* (Oxford U.P., London and New York, 1939), p. 120.

[26] J. E. Gilbert, private communication.

[27] B. A. Taylor and D. L. Williams, Can J. Math. **XXII**(6), 1226 (1970).

Operator theory on the WKB method and Bremmer series

David Dudley[a] and Alan P. Wang
Department of Mathematics, Arizona State University, Tempe, Arizona 85287

(Received 16 September 1980; accepted for publication 30 July 1982)

The well-known WKB method and Bremmer series in mathematical physics have various modifications and applications. Basically, they are all limited to the finite-dimensional case. Such methods have been extended to the infinite-dimensional case in this paper, i.e., operators on a general normed linear space. This enables us to apply the methods to solve the problem of transport of radiation through a layer of cloud. The derivation is based on input-output states and linear operators. Some elementary properties are exploited, along with physical interpretations. The convergence of the series is established. An example is given to demonstrate the methods and the rate of convergence.

PACS numbers: 03.65.Db, 02.30.Lt, 03.65.Nk, 42.68.Mj

1. INTRODUCTION

Applications of the WKB approximation usually refer to the propagation of waves through an inhomogeneous medium,[1,2] for example, its application to the scalar Schrödinger equation

$$\frac{d^2y}{dx^2} + \frac{2m}{\hbar^2}[E - V(x)]\,y = 0, \tag{1.1}$$

where m is the mass, V is the potential, and E is the energy. Various modifications and applications are known.[3,4] It is well known that the WKB method offers a single term approximation to the solution of (1.1). Bremmer developed an infinite series solution with the WKB approximation as the starting term.[5,6] The approximation is valid under these conditions: the mass is large, the energy high and the potential smooth. The WKB solution cannot be valid near a classical turning point, where the momentum is zero. The power of Bremmer's series is that it starts with a well established first term approximation and each term in the series has a definite physical meaning. This feature is generally lost in most mathematical approximation schemes.

Our interest in this subject is strongly influenced by the work of Wing[7] and Bellman[8,9] in the area of invariant embedding. However, the results presented in this paper are from the "input-output system" approach involving states and operators, and the system is generalized to an infinite dimensional space. For example, (1.1) can be expressed as

$$\begin{pmatrix} y(x) \\ y'(x) \end{pmatrix}' = \begin{pmatrix} 0 & 1 \\ \frac{2m}{\hbar^2}[V(x) - E] & 0 \end{pmatrix} \begin{pmatrix} y(x) \\ y'(x) \end{pmatrix},$$

a special case of a state equation from the input-output point of view on an abstract space. The reason for considering operators on an abstract normed linear space is that our results can be applied to a broad class of physical systems with the same mathematical structure. Equations considered in this paper have applications in a wide range of physical problems such as the theory of transmission lines, radiative transfer, stochastic diffusion problems, and others.[10–12] One particular example we will consider is the problem of predicting transport of radiation through a layer of cloud involving operators that are of the integral form, such an operator maps a class of continuously differentiable functions to itself.

In Sec. 2, a short discussion of input-output systems is given, where the operators used are defined on a normed linear space.[13] Our solution form is closely related to the product integral as in Refs. 14, 15, and 16. The elementary properties are exploited in Secs. 3 and 4, where the Bremmer series for operators on a normed linear space is established, along with its physical interpretation. In the fifth section, the convergence of the series is discussed. In the last section a computational example is given for prediction of transport of radiation through an atmosphere.

2. THE INPUT-OUTPUT SYSTEM

We shall consider the system

$$\begin{pmatrix} u(y) \\ v(x) \end{pmatrix} = \begin{pmatrix} t(x,y) & \rho(x,y) \\ r(x,y) & \tau(x,y) \end{pmatrix} \begin{pmatrix} u(x) \\ v(y) \end{pmatrix}, \tag{2.1}$$

where the left side of the equation is the output and, on the right side, the transfer operator operates on the input. Input and output are assumed to be elements of a normed linear space X. Based on physical grounds, it is desirable to have the transfer operator be *continuous at the identity*, i.e.,

$$\lim_{y \to x} \left\| \begin{pmatrix} t(x,y) & \rho(x,y) \\ r(x,y) & \tau(x,y) \end{pmatrix} - \begin{pmatrix} E & 0 \\ 0 & E \end{pmatrix} \right\| = 0, \tag{2.2}$$

where E is the identity operator on X and $\| \ \|$ is the strong operator norm.[13] Associated with the transfer operator, we assume there exists a *generator*

$$M(z) = \begin{pmatrix} b(z) & a(z) \\ -c(z) & -d(z) \end{pmatrix} \tag{2.3}$$

that satisfies the equation

$$\frac{d}{dz}\begin{pmatrix} u(z) \\ v(z) \end{pmatrix} = M(z) \begin{pmatrix} u(z) \\ v(z) \end{pmatrix}, \quad x \leqslant z \leqslant y, \tag{2.4}$$

with boundary conditions $u(x)$ and $v(y)$ given. It is well known that the generator[10] is associated with the physical parameters of a particular system under consideration. Therefore, we may assume $M(z)$ is a bounded operator. The generator is related to the transfer operator by the following nonlinear differential system of operator equations:

[a] Current address: Phoenix College, Phoenix, Arizona 85013.

 0022-2488/83/061470-07$02.50 © 1983 American Institute of Physics

$$\frac{\partial}{\partial y} S(x,y) = \begin{pmatrix} (b+\rho c)\, t & a + b\rho + \rho d + \rho c\rho \\ \tau c t & \tau(d + c\rho) \end{pmatrix} \tag{2.5}$$

and

$$-\frac{\partial}{\partial x} S(x,y) = \begin{pmatrix} t(b + ar) & ta\tau \\ c + dr + rb + rar & (d + ra)\tau \end{pmatrix}, \tag{2.6}$$

where

$$S(x,y) = \begin{pmatrix} t(x,y) & \rho(x,y) \\ r(x,y) & \tau(x,y) \end{pmatrix} \tag{2.7}$$

and $S(x,y)$ is called the *scattering operator* or *scattering matrix*.

The purpose of this paper is to construct a solution to (2.4), (2.5), or (2.6) in the form of Bremmer's series. Details are presented in the next section. Since the following properties are used often, we develop them here. Let us consider the differential equation of operators

$$\frac{d}{dz} W(x,z) = h(z)\, W(x,z), \quad W(x,x) = E, \tag{2.8}$$

where $h(z)$, $W(x,z) \in L(X,X)$, i.e., linear operators from X into X; then

$$W(x, x+\eta) = E + h(x)\,\eta + o(\eta). \tag{2.9}$$

It can be shown that $W(x,z)$ has the property

$$W(x,z) = W(x_{n-1}, x_n) \cdots W(x_1, x_2)\, W(x_0, x_1),$$

which we define to be $\prod_{i=1}^{n} W(x_{i-1}, x_i)$ and where $\{x_i\}_{i=0}^{n}$ is a partition of $[x,z]$. Therefore,

$$\lim_{\delta \to 0} \left|\left| W(x,z) - \prod_{i=1}^{n} (E + h(x_{i-1})\,\Delta x_i) \right|\right| = 0, \tag{2.10}$$

where $\Delta x_i = x_i - x_{i-1}$ and $\delta = \max_i \{\Delta x_i\}$. Since

$$\exp[h(x_{i-1})\,\Delta x_i] = E + h(x_{i-1})\,\Delta x_i + o(\Delta x_i)$$

we obtain

$$W(x,z) = \lim_{\substack{n \to \infty \\ \delta \to 0}} \prod_{i=1}^{n} \exp[h(x_{i-1})\,\Delta x_i] \tag{2.11}$$

$$= \prod_{x}^{z} \exp[h(s)\,ds].$$

For details, see Ref. 15. The derivation of (2.11) can be made with other approaches. One interesting recent work by Dollard and Friedman[16] uses the product integral. The approach used here, exploiting the relation between operators and generators, is closely related to semigroup theory.

3. THE WKB METHOD

As pointed out in Sec. 1, many physical systems can be viewed as input-output systems. These systems are governed by the state equation (2.1) and the operator equations (2.5) and (2.6). The classic WKB method is now interpreted in this framework. In many physical systems the forward scattering is stronger than the backward scattering. In terms of a thin layer this means more is transmitted than is reflected.[17,18] Therefore, for a process involving multiple scattering, as interpreted here, a single scattering approximation is a good first step. For this reason we assume first there is no right

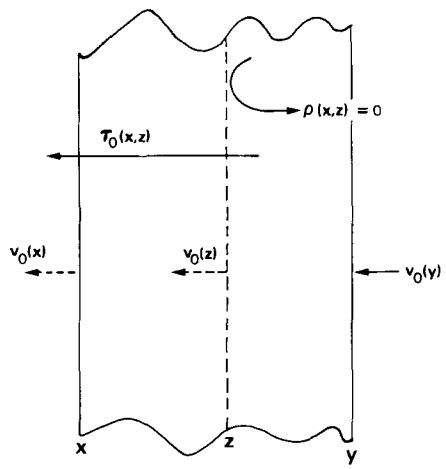


FIG. 1. Transmission in a nonreflective medium.

reflection, i.e., $\rho(x,z) = 0$ for all z, $x \leqslant z \leqslant y$, see Fig. 1. Let $u(x) = 0$ to simplify the mathematics.

By (2.5), with y replaced by z and $\rho(x,y) = 0$ we obtain

$$\frac{\partial}{\partial z} \tau_0(x,z) = \tau_0(x,z)\, d(z), \quad \tau_0(x,x) = E. \tag{3.1}$$

If τ_0^{-1} exists, then it satisfies

$$\frac{\partial}{\partial z} \tau_0^{-1}(x,z) = -d(z)\, \tau_0^{-1}(x,z), \quad \tau_0^{-1}(x,x) = E,$$

and from (2.8) and (2.11) we have

$$\tau_0^{-1}(x,z) = \prod_{x}^{z} \exp[-d(s)\,ds];$$

hence,

$$\tau_0(x,z) = \prod_{z}^{x} \exp[-d(s)\,ds]. \tag{3.2}$$

From (2.5), if $\tau^{-1}(x,y)$ exists, then

$$\frac{\partial}{\partial z}\rho(x,z) = a(z) + b(z)\,\rho(x,z) + \rho(x,z)\,\tau^{-1}(x,z)\,\frac{\partial}{\partial z}\tau(x,z). \tag{3.3}$$

From the assumption that $\tau(x,z) - E \to 0$ as $z \to x$, then $\tau^{-1}(x,z)$ exists for $z - x$ small. Hence, we obtain the local existence of solutions. Due to the nonlinear nature of the system (2.5), the global existence of solutions is more involved. This will be discussed in a separate paper involving critical points and maximal solutions. Replacing $\tau(x,z)$ in (3.3) by $\tau_0(x,z)$ in (3.2) and calling the resulting solution of (3.3) $\rho_1(x,z)$, we have

$$\frac{\partial}{\partial z}\rho_1(x,z) = a(z) + b(z)\,\rho_1(x,z) + \rho_1(x,z)\,d(z). \tag{3.4}$$

Equation (3.4) is linear in ρ_1, while the equation for ρ in (2.5) is nonlinear. Equation (3.4) can be solved using integrating factors and

$$\rho_1(x,z) = \int_x^z \prod_{\sigma}^{z} \exp[b(s)\,ds]\, a(\sigma)\, \tau_0(\sigma,z)\, d\sigma \tag{3.5}$$

or

$$u_1(z) = \rho_1(x, z)\, v_0(z) = \int_x^z \prod_\sigma^z \exp[b(s)\, ds]\; a(\sigma)\, v_0(\sigma)\, d\sigma\,, \tag{3.6}$$

where $v_0(\sigma) = \tau_0(\sigma, z)\, v_0(z)$ for $x \leqslant \sigma \leqslant z$, see Fig. 2.

In the physical model, the intensities at x and z are related by $v_0(x) = \tau_0(x,z)\, v_0(z)$ for particles experiencing no reflection being propagated from right to left (Fig. 1). In a similar manner the intensity is propagated from left to right, with no reflection, by

$$t_0(z_1, z_2) = \prod_{z_1}^{z_2} \exp[b(s)\, ds]\,,$$

and equation (3.6) can be written

$$u_1(z) = \int_x^z t_0(\sigma, z)\, a(\sigma)\, \tau_0(\sigma, z)\, v_0(z)\, d\sigma\,. \tag{3.7}$$

Now $u_1(z)$ denotes the intensity resulting from exactly one reflection, i.e., $\rho_1(x, z) \neq 0$. Equation (3.7) has a sound physical interpretation. The intensity at z, $v_0(z)$, is propagated to σ by $\tau_0(\sigma, z)$, reflected at σ by $a(\sigma)$, then propagated back to z by $t_0(\sigma, z)$. Finally, take the sum of all reflections for $x \leqslant \sigma \leqslant z$, see Fig. 3, i.e.,

$$u_1(z) = \lim_{n\to\infty} \sum_{i=1}^{n} t_0(\sigma_i, z)\, a(\sigma_i)\, \tau_0(\sigma_i, z)\, v_0(z)\, \Delta\sigma_i\,.$$

The scattering operator associated with a thin layer of thickness δ is related to the generator up to $O(\delta)$ by

$$S(x, x+\delta) = \begin{pmatrix} E + b(x)\,\delta & a(x)\delta \\ c(x)\delta & E + d(x)\delta \end{pmatrix}. \tag{3.8}$$

Since $c(x) \equiv 0$ in (3.4), particles experience only right hand reflection with no left-hand reflection, i.e., only one reflection is involved and no multiple scattering takes place. For additional discussion of the one-dimensional case see Ref. 7.

From (3.2) and (2.6) with $z = y$ input $u(x) = 0$, and $v(y) = v_0(y)$, we have

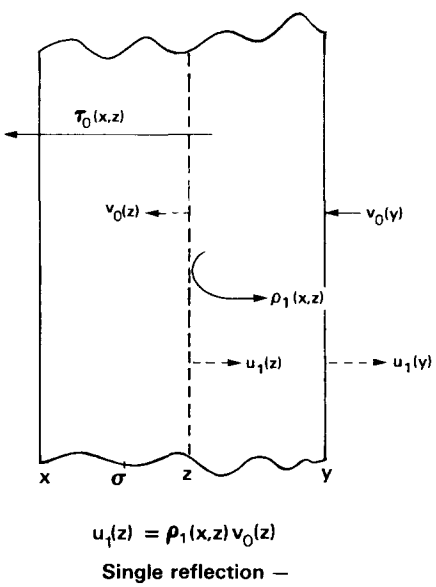


FIG. 2. The WKB approximation.

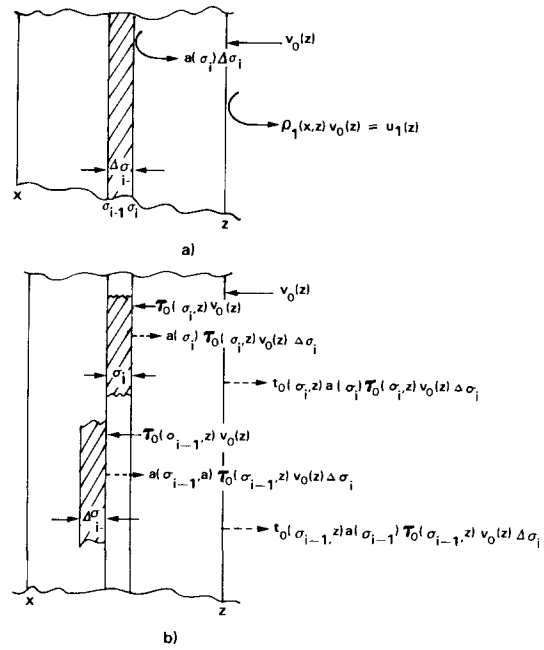


FIG. 3. Total reflection in a single reflective medium. (a) Reflection due to a thin medium. (b) Single reflection due to a large number of thin media. $u_1(z) \cong \Sigma_{i=1}^{n}\, t_0(\sigma_i, z)\, a(\sigma_i)\, \tau_0(\sigma_i, z)\, v(z) \Delta\sigma_i$, n large.

$$\begin{aligned} v_0(x) &= \tau_0(x, y)\, v_0(y), \\ u_1(y) &= \rho_1(x, y)\, v_0(y)\,. \end{aligned} \tag{3.9}$$

This is the WKB approximation for the solution of the input-output problem (2.1), where the output is found by considering only a single reflection.

4. THE GENERALIZED BREMMER SERIES

The purpose of this section is to construct a generalized Bremmer series as a solution of (2.4), of the form

$$u(z) = \sum_{n=0}^{\infty} u_{2n+1}(z) \quad \text{and} \quad v(z) = \sum_{n=0}^{\infty} v_{2n}(z)\,. \tag{4.1}$$

The convergence of these series will be discussed in the next section. To achieve our goal we shall consider the left-hand operators t and r. Under assumptions similar to those for τ_0 and ρ_1, let us denote by $t_1(z, y)$ the right transmission operator for particles experiencing no more reflection when travelling from z to y, and let $r_2(z, y)$ denote the operator which accounts for exactly one more reflection. Without repeating the steps, the results are

$$u_1(y) = t_1(z, y)\, u_1(z) \tag{4.2}$$

and

$$v_2(z) = \int_z^y \prod_\sigma^z \exp[-d(s)\, ds]\; c(\sigma)\, u_1(\sigma)\, d\sigma\,. \tag{4.3}$$

We have now generated $v_0(z)$, $u_1(z)$, and $v_2(z)$. In general,

$$u_{2n+1}(z) = \int_x^z \prod_\sigma^z \exp[b(s)\, ds]\; a(\sigma)\, v_{2n}(\sigma)\, d\sigma \tag{4.4}$$

and

$$v_{2n+2}(z) = \int_z^y \prod_\sigma^z \exp[-d(s)\, ds]\; c(\sigma)\, u_{2n+1}(\sigma)\, d\sigma\,, \tag{4.5}$$

for $n = 0,1,2,\ldots$. The term $u_{2n+1}(z)$ denotes the contributions of particles undergoing $2n+1$ reflections and $v_{2n+2}(z)$ denotes that for particles undergoing $2n+2$ reflections. Of course, the multiple scattering is the sum of all such terms; see Fig. 4 and Fig. 5.

This interpretation of multiple scattering is carried out in the fundamental work of Redheffer, who created the $*$-product, i.e.,

$$S(x, y) = S(x, z) * S(z, y) \tag{4.6}$$

For all $z \in [x, y]$, whenever $*$ is defined.[19] The above discussion all holds when an input $u(x) \neq 0$ is included in the system. Since all operators under consideration are linear, it is easy to see that the generalized Bremmer series constructed by (4.1), (4.4), and (4.5) is the desired solution of (2.4). Simply differentiate both sides of (4.4) and apply Leibnitz's Rule.

$$\frac{\partial}{\partial z} u_{2n+1}(z) = \int_x^z b(z) \prod_\sigma^z \exp[b(s)\,ds]\, a(\sigma)\, v_{2n}(\sigma)\, d\sigma + \prod_x^z \exp[b(s)\,ds]\, a(z)\, v_{2n}(z),$$

$$\frac{\partial}{\partial z} u_{2n+1}(z) = b(z)\, u_{2n+1}(z) + a(z)\, v_{2n}(z). \tag{4.7}$$

Likewise, from (4.5), we have

$$-\frac{\partial}{\partial z} v_{2n}(z) = c(z)\, u_{2n-1}(z) + d(z)\, v_{2n}(z). \tag{4.8}$$

Sum (4.7) and (4.8) over all n to see that $u(z)$ and $v(z)$ in (4.1) are the solution of (2.4), provided the series in (4.1) converge in the strong operator norm.

5. CONVERGENCE OF THE GENERALIZED BREMMER SERIES

In this section, the differential system (2.4) is replaced by the equivalent integral equations. The convergence of the Neumann series solution is discussed, and a criterion giving sufficient conditions for convergence of the Bremmer series is given.

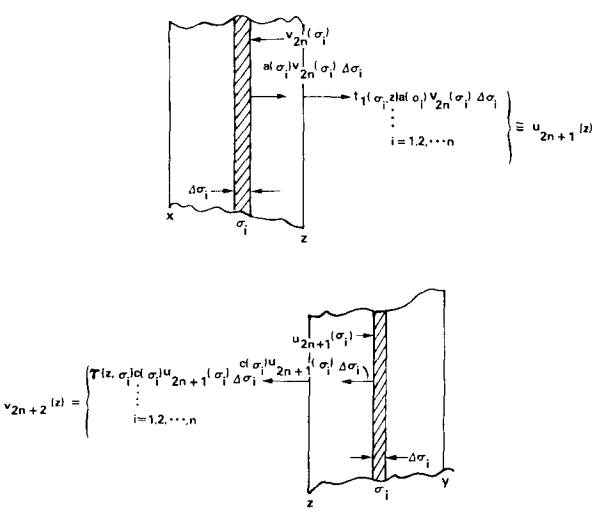


FIG. 4. The physical interpretations of the nth internal intensities.

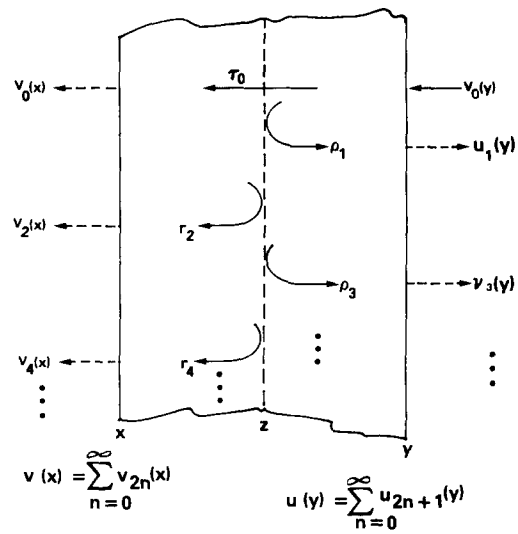


FIG. 5. The physical interpretation of the Bremmer's series due to multiple scattering.

Define $U(z) = \Pi_z^x \exp[b(s)\,ds]\, u(z)$ and $V(z) = \Pi_z^y \exp[-d(s)\,ds]\, v(z)$. Combining these with (2.4) we have that (U,V) satisfies

$$\frac{d}{dz}\begin{pmatrix} U \\ -V \end{pmatrix} = \begin{pmatrix} 0 & E(z) \\ F(z) & 0 \end{pmatrix}\begin{pmatrix} U \\ V \end{pmatrix}, \tag{5.1}$$

where $(U(x), V(y))$ are given boundary conditions and

$$E(z) = \prod_z^x \exp[b(s)\,ds]\, a(z) \prod_y^z \exp[-d(s)\,ds], \qquad F(z) = \prod_z^y \exp[-d(s)\,ds]\, c(z) \prod_x^z \exp[b(s)\,ds]. \tag{5.2}$$

Converting (5.1) into a pair of integral equations yields

$$U(z) = U(x) + \int_x^z E(\sigma)\, V(\sigma)\, d\sigma, \tag{5.3a}$$

$$V(z) = V(y) + \int_z^y F(\sigma)\, U(\sigma)\, d\sigma. \tag{5.3b}$$

Substituting (5.3b) into (5.3a),

$$U(z) = U(x) + \int_x^z E(\sigma)\, d\sigma\, V(y) + \int_x^z \int_\sigma^y E(\sigma)\, F(s)\, U(s)\, ds\, d\sigma.$$

Changing the order of integration in the iterated integral yields the inhomogeneous Fredholm equation

$$U(z) = U(x) + \int_x^z E(\sigma)\, d\sigma\, V(y) + \int_x^y K_1(s,z)\, U(s)\, ds, \tag{5.4a}$$

where

$$K_1(s, z) = \begin{cases} \displaystyle\int_x^s E(\sigma)\, d\sigma\, F(s), & s \leqslant z, \\ \displaystyle\int_x^z E(\sigma)\, d\sigma\, F(s), & z \leqslant s. \end{cases} \tag{5.4b}$$

Similarly for V we have

$$V(z) = V(y) + \int_z^y F(\sigma)\, d\sigma\, U(x) + \int_x^y K_2(s, z)\, V(s)\, ds \tag{5.5a}$$

where

$$K_2(s, z) = \begin{cases} \int_z^y F(\sigma)\, d\sigma\, E(s), & s \leqslant z, \\ \int_s^y F(\sigma)\, d\sigma\, E(s), & z \leqslant s. \end{cases} \tag{5.5b}$$

The Neumann series for (5.4a) and (5.5a) will converge to the solution provided

$$\int_x^y \int_x^y \| K_i(s,t) \|^2\, ds\, dt < 1 \quad (i = 1,2). \tag{5.6}$$

Let $L(s) \in L(X,X)$ for $x \leqslant s \leqslant y$. $\| \ \|$ will denote the usual operator norm in $L(X,X)$. Define $\| \ \|_0$ by

$$\| L \|_0 = \sup_{x \leqslant s \leqslant y} \| L(s) \| . \tag{5.7}$$

If $\| E \|_0 \| F \|_0 (y - x)^2 < 1$, then the Neumann series for (5.4a) and (5.5a) converge.

Proof:

$$\begin{aligned} &\int_x^y \int_x^y \| K_2(s,t) \|^2\, ds\, dt \\ &\quad = \int_x^y \int_x^t \left\| \int_t^y F(\sigma)\, d\sigma\, E(s) \right\|^2 ds\, dt \\ &\qquad + \int_x^y \int_t^y \left\| \int_s^y F(\sigma)\, d\sigma\, E(s) \right\|^2 ds\, dt \\ &\quad \leqslant \int_x^y \int_x^t \left[\int_t^y \| F(\sigma) \|\, d\sigma \right]^2 \| E(s) \|^2\, ds\, dt \\ &\qquad + \int_x^y \int_t^y \left[\int_s^y \| F(\sigma) \|\, d\sigma \right]^2 \| E(s) \|^2\, ds\, dt \\ &\quad \leqslant \left[\int_x^y \| F(\sigma) \|\, d\sigma \right]^2 \int_x^y \left[\int_x^y \| E(s) \|^2\, ds \right] dt \\ &\quad \leqslant \| F \|_0^2 \, \| E \|_0^2 \, (y - x)^4 < 1. \end{aligned}$$

Similarly for K_1.

If the analysis of Sec. 4 is applied to (5.1), we have $U(z) = \Sigma_{n=0}^{\infty} U_{2n+1}(z)$ and $V(z) = \Sigma_{n=0}^{\infty} V_{2n}(z)$ as the Bremmer series solution. Computing the Neumann series for (5.4a) and (5.5a) yields the same infinite series as the Bremmer series term-by-term. Hence, we have the following:

The generalized Bremmer series (4.1) is the solution to the differential system (2.4) provided $\|E\|_0 \, \|F\|_0 (y - x)^2 < 1$, where E and F are defined by (5.2).

6. AN APPLICATION

As an application of the above method, consider the problem for predicting transport of radiation through the atmospheres. In particular the problem is to estimate the amount of infrared radiation at wavelength 4.3μ through a layer of scattering media such as clouds.

The mathematical model for diffusion radiation has been constructed by one of the authors.[20] For the convenience of the reader, the equations for diffuse radiation are presented in the following operator form as a special case of equation (2.1),

$$\frac{d}{dz} \begin{pmatrix} D^+(z;\mathbf{S},\mathbf{S}_0) \\ D^-(z;\mathbf{S},\mathbf{S}_0) \end{pmatrix} = \begin{pmatrix} b(z) & a(z) \\ -c(z) & -d(z) \end{pmatrix} \begin{pmatrix} D^+(z;\mathbf{S},\mathbf{S}_0) \\ D^-(z;\mathbf{S},\mathbf{S}_0) \end{pmatrix}, \tag{6.1}$$

where

$D^+(z;\mathbf{S},\mathbf{S}_0)$ = the transmitted diffuse radiation intensity in the direction $\mathbf{S}$ due to incident radiation I_0 in the direction $\mathbf{S}_0$ on a layer of optical thickness z,

$D^-(z;\mathbf{S},\mathbf{S}_0)$ = the reflected diffuse radiation intensity in direction $\mathbf{S}$ due to incident radiation I_0 in the direction $\mathbf{S}_0$ on a layer of optical thickness z,

and where the operators

$$\begin{aligned} a(a)\cdot &= \int_t^z \int_{\Omega} \gamma(\mathbf{S},\mathbf{S}_0) \exp\left(\frac{z' - z}{u}\right) \cdot d\mathbf{S}_0 \frac{dz'}{u}, \\ b(z)\cdot &= \int_0^z \exp\left(\frac{z' - z}{u}\right) \delta(z' - 0) \cdot \frac{dz'}{u} \\ &\quad + \int_0^z \int_{\Omega_+} \gamma(\mathbf{S},\mathbf{S}_0) \\ &\quad \times \exp\left(\frac{z' - z}{u}\right) \cdot d\mathbf{S} \frac{dz'}{u}, \\ -c(z)\cdot &= \int_0^z \int_{\Omega_+} \gamma(\mathbf{S},\mathbf{S}_0) \exp\left(\frac{z' - z}{u}\right) \cdot d\mathbf{S}_0 \frac{dz'}{u}, \\ -d(z)\cdot &= \int_t^z \exp\left(\frac{z' - z}{u}\right) \delta(z' - t) \frac{dz'}{u} \\ &\quad + \int_t^z \int_{\Omega} \gamma(\mathbf{S},\mathbf{S}_0) \exp\left(\frac{z' - z}{u}\right) \cdot d\mathbf{S}_0 \frac{dz'}{u}. \end{aligned} \tag{6.2}$$

Here γ is called the scattering or phase function, δ is the delta function, Ω_+ is the half sphere where the transmitted radiation lies and Ω_- is the other half sphere where the reflected diffuse radiation lies. For more details, see Ref. 20.

The boundary conditions for an atmosphere of thickness t are

$$D^+(0;\mathbf{S},\mathbf{S}_0) = D^-(t;\mathbf{S},\mathbf{S}_0) = 0. \tag{6.3}$$

The scattering function for a homogeneous medium has the form

$$\gamma(\mathbf{S},\mathbf{S}_0) = \gamma(\theta), \tag{6.4}$$

with $\cos\theta = \mathbf{S}\cdot\mathbf{S}_0$, the dot product of the vectors $\mathbf{S}$ and $\mathbf{S}_0$. Of course in view of Eq. (2.1)

$$D^+(z; \mathbf{S}\cdot\mathbf{S}_0) = t(0, z)\, I_0$$

and

$$D^-(z; \mathbf{S}\cdot\mathbf{S}_0) = \rho(0, z)\, I_0$$

A remark on the scattering function is in order. For many physical applications of infrared the wavelength between 3.0–5.0μ is of great interest. In particular in military infrared applications,[21] the daytime spectra of objects at ambient temperatures show minima around 3.0–5.0μ. Therefore it is suggested that the 4.3μ band properties may be used for our study. According to the result of Deiermendjian,[22] clouds have maximum droplet size-distribution in the neighborhood of 5.0μ. Therefore, it is proper to use the scattering function[23] which is computed based on Deiermendjian's distribution function of water drops and on Mie scattering theory; see Fig. 6.

The physical condition under consideration is the reflection of clouds of various optical thickness, $t = 1.0$ and $t = 0.1$, due to an incident radiation of unit intensity normal to the cloud. By using Eq. (6.1) and the approximation method established in this report, one can try to compute the reflectivities of the cloud as a function of scattering angle θ.

To obtain the first approximation τ_0 of τ, it is convenient to use $\Pi_{i=1}^{n}\,[E - d(x_{i-1})\,\Delta x_i]$ as in Eq. (2.10), with $n = 20$ and $\Delta x_i = t/20$, $d(x_{i-1})$ operating on the unit normal input $\delta(\mathbf{S} - \mathbf{S}_0)$ by the Gauss quadrature formula.[24]

For example, the term

$$\begin{aligned} -d(z)\,\delta(\mathbf{S} - \mathbf{S}_0) &= \frac{1}{u}\exp\left(\frac{t-z}{u}\right) \\ &\quad + \int_z^t \gamma(\mathbf{S}_0)\left(\exp\frac{z'-z}{u}\right)\frac{dz'}{u} \\ &= \frac{1}{u}\exp\left(\frac{t-z}{u}\right) - \int_0^1 \gamma(\mathbf{S}_0) \\ &\quad \times\left(\exp\frac{(t-z)\,\tilde{z}}{u}\right)\frac{d\tilde{z}}{u} \end{aligned} \tag{6.5}$$

and the second term is approximated by means of the quadrature formula

$$\int_0^1 g(z)\,dz \doteq \sum_{k=1}^{N} c_k\, g(z_k), \tag{6.6}$$

where the c_k are called Christoffel numbers and the z_k are roots of the Legendre polynomials of order N. The integral is thus reduced to a summation. From this the first order transmitted term, $\tau_0(x, z)\,\delta(\mathbf{S} - \mathbf{S}_0)$, see Eq. (3.2), due to a normal input is obtained.

By repeating the process for operators a and b, values for $\rho_1(x, z)$ and $u_1(z)$ are obtained by using Eq. (3.5) and (3.6). Results are stored for the next computation. Equation (4.4) is repeated for the computation of u_3 and u_5. Computation is stopped at u_5, because these values are very close to 0. The following shows convergence of the Bremmer series for this particular case.

The scattering function γ used in this problem has the property that, the total area under it is less than one, due to the existence of some absorption. That is, the *albedo of single scattering*

$$\int_\Omega \gamma(\mathbf{S},\mathbf{S}_0)\,dS_0 \leqslant 1. \tag{6.7}$$

This important fact leads to the following matrix in equality for any Δz small

$$\begin{pmatrix} b(z) & a(z) \\ c(z) & d(z) \end{pmatrix} \Delta z \leqslant 0 \tag{6.8}$$

When we say a matrix $A \leqslant 0$, we mean $I^T A^T A I \leqslant 0$ for any vector I. This in turn gives us the desired inequalities

$$\| E \|_0 \leqslant 1 \quad \text{and} \quad \| F \|_0 \leqslant 1,$$

as defined in (5.7) which applies to this special case. This establishes the convergences of the Bremmer's series under consideration, i.e., the optical thickness is less than or equal to one.

However, a stronger result can be established which is certainly of great physical interest by observing that (6.7) and (6.8) also imply that

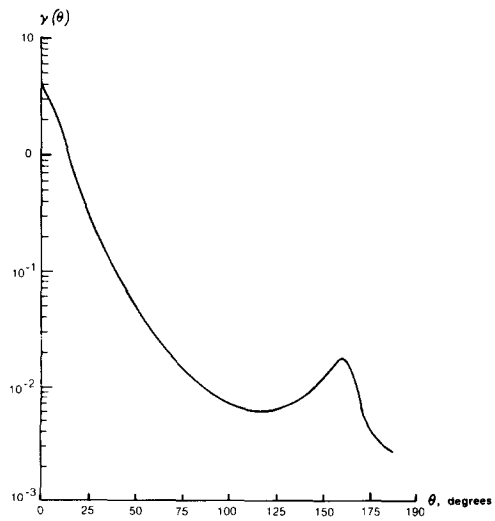


FIG. 6. Scattering function for incident radiation of 4.3-μ wavelength.

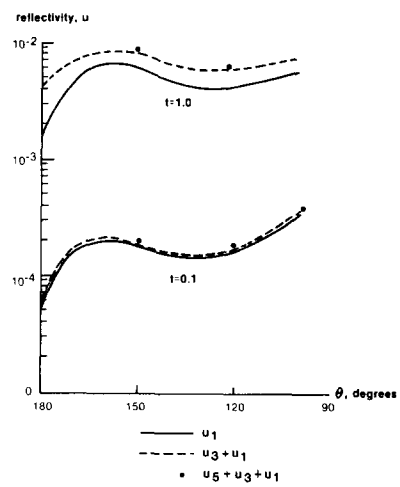


FIG. 7. Reflectivity of a cloud as a function of scattering angle θ with normal unit input.

$$\left|\left| \int_x^y E(\sigma)\, d\sigma \right|\right| \leqslant 1 \quad \text{and} \quad \left|\left| \int_x^y F(\sigma)\, d\sigma \right|\right| \leqslant 1,$$

for all $y \geqslant x$. It follows that the condition (5.6) is satisfied. Thus the convergence of the Bremmer series is established for any optical thickness for the system with albedo of single scattering less than or equal to one. Also note that for a given optical thickness the convergence is uniform.

All results are plotted in Fig. 7 showing the reflectivity of a cloud of two optical thicknesses as a function of scattering angle θ. Since the value of u_5 is very close to 0, quick convergence of this method is indicated. In the neighborhood of 180° the convergence seems to be slower. Likewise if the optical thickness is much larger than 1, many more iterations of the computation may be needed. However, the above theoretical result guarantees the convergence of this approximation method. This example demonstrates an application to a physical problem involving an integral operator. The problem cannot be solved by using the one-dimensional model as in Ref. 7. On the other hand, from a physical point of view, in this example we take anisotropy as well as multiple scattering into consideration.

[1]S. Ueno, R. Kalaba, and R. Bellman, "Some Mathematical Aspects of Multiple Scattering in a Finite Inhomogeneous Slab with Anistropic Scattering," Publ. Astron. Soc. Jpn. **22**, 75–83 (1970).

[2]E. Bahar, "Generalized WKB Method with Applications to Problems of Propagation in Nonhomogeneous Media," J. Math. Phys. **8**, 1735–1745 (1967).

[3]S. C. Miller, Jr. and R. H. Gord, Jr., "A WKB-Type Approximation to the Schrödinger Equation," Phys. Rev. **91**, 174–179 (1953).

[4]P. Lu and S. S. Wald, "Modified WKB Approximation Applied to the Solution of Repulsive Singular Potential," J. Math. Phys. **13**, 646–648 (1972).

[5]H. Bremmer, *The WKB Method as the First Term of a Geometric-Optical Series Theory of Electromagnetic Waves* (Interscience, New York, 1951).

[6]F. V. Atkinson, "Wave Propagation and Bremmer Series," J. Math. Anal. Appl. **1**, 176–255 (1960).

[7]G. Milton Wing, "Invariant Embedding and Generalizations of the WKB Method and the Bremmer Series," J. Math. Anal. Appl. **48**, 400–422 (1976).

[8]R. Bellman and R. Vasudevan, "Wave Equation with Sources, Invariant Embedding and Bremmer Series Solutions," J. Math. Anal. Appl. **48**, 17–30 (1974).

[9]R. Bellman, *Perturbation Techniques in Mathematics, Physics and Engineering* (Holt, Rinehart and Winston, New York, 1964).

[10]E. F. Beckenbach, *Modern Mathematics for Engineers*, Second Series (McGraw-Hill, New York, 1961), Chap. 12.

[11]R. Bellman and G. Milton Wing, *An Introduction to Invariant Embedding* (Wiley, New York, 1975).

[12]S. Chandrasekhar, *Radiative Transfer* (Oxford U.P., Oxford, 1950).

[13]A. V. Balakrishnan, *Applied Functional Analysis* (Springer-Verlag, New York, 1976).

[14]R. H. Martin, *Nonlinear Operators Differential Equations in Banach Space* (Wiley-Interscience, New York, 1976).

[15]R. Hecht-Nielson, "Transport Operators," Ph.D. thesis, Dept. of Mathematics, Arizona State University, Tempe, Arizona, 1974.

[16]J. D. Dolland and C. Freidman, "Product Integrals and Schrödinger Equations," J. Math. Phys. **18**, 1598–1607 (1977).

[17]A. P. Wang, "Multi-Scattering of Anisotropic Radiative Transfer Through a Layer of Cloud," J. Appl. Sci. Res. **23**, 221–236 (1970).

[18]G. Milton Wing, *An Introduction to Transport Theory* (Wiley, New York, 1962).

[19]R. Redheffer, "On the Relation of Transmission-Linear Theory to Scattering and Transfer," J. Math. & Phys. **41**, 1–41 (1962) (Cambridge, Mass.)

[20]A. P. Wang, "A Mathematical Model of Cloud," Astrophys. Space Sci. **70**, 447–459 (1980).

[21]William L. Wolfe, Ed., *Handbook of Military Infrared Technology* (Office of Naval Research, Washington, D.C., 1965), p. 145.

[22]D. Deiermendjian, "Scattering and Polarization Properties of Poly-dispersed Suspension with Partial Absorption," Appl. Opt. **3**, 171–189 (1964).

[23]D. Deiermendjian, "Theory of Solar Aureole—Parts I and II," Rand Report RM 2133, 1958.

[24]R. E. Bellman, R. E. Kalaba, and M. C. Prestrud, *Invariant Embedding and Radiative Transfer in a Slab of Finite Thickness* (America Elsevier, New York, 1963).

Large coupling behavior of the energy eigenvalues of two anharmonic oscillators[a)]

David Isaacson
Math Department, Rensselaer Polytechnic Institute, Troy, New York 12180

(Received 27 May 1982; accepted for publication 27 August 1982)

We give two methods for calculating the behavior of the energy levels of

$$H \equiv (\tfrac{1}{2})\{-\partial_x^2 - \partial_y^2 + g[(x^2-a^2)^2 + (y^2-a^2)^2] + b(x-y)^2\}$$

as the coupling constant g goes to infinity.

PACS numbers: 03.65.Ge

I. INTRODUCTION

Let $E_j = E_j(g)$ where $E_0 < E_1 \leqslant E_2 \leqslant E_3 \leqslant \cdots$ denote the eigenvalues of

$$\begin{aligned} H &= (\tfrac{1}{2})\{-\partial_x^2 - \partial_y^2 \\ &\quad + g[(x^2-a^2)^2 + (y^2-a^2)^2] + b(x-y)^2\} \\ &= (\tfrac{1}{2})\{-\Delta + v(x,y)\}. \end{aligned} \tag{1}$$

For g going to infinity we prove that

$$E_0(g) = (4a^2g)^{1/2} - 2/(4a^2) + o(1),$$

$$E_1(g) = E_0(g) + o(1),$$

$$E_2(g) = E_0(g) + 2a^2b + o(1),$$

$$E_3(g) = E_2(g) + o(1).$$

Thus $E_1(g) - E_0(g)$ goes to zero and $E_2(g) - E_0(g)$ goes to $2a^2b$ as g goes to infinity. We also prove that every eigenvalue of H has one of the expressions

$$(4a^2g)^{1/2}(n+m+1) - (1/4a^2) \times [3(n^2+m^2) + 3(n+m) + 2] + o(1) \tag{2}$$

or

$$(4a^2g)^{1/2}(n+m+1) - (1/4a^2) \times [3(n^2+m^2) + 3(n+m) + 2] + 2a^2b + o(1)$$

for some $n, m = 0, 1, 2, \cdots$ as g goes to infinity.

We describe two methods for calculating the behavior of the energy levels $E_j(g)$ as g goes to infinity.

In the first method we expand the function $V(x,y) = g[(x^2-a^2)^2 + (y^2-a^2)^2] + b(x-y)^2$ about each of its four minima, and rescale x and y so that as g goes to infinity the principal curvatures of each minima tend to constants. When this is done the four minima become widely separated by high barriers, which suggests that the leading ($O(g^{1/2})$) behavior of the eigenvalues of H is given by the union of the eigenvalues of $-\Delta + Q_i(x,y)$ where for $i = 1,2,3,4$ Q_i is the quadratic approximation to V in the neighborhood of its ith minima. To obtain approximations to the eigenvalues valid to $o(1)$ we take into account our knowledge of the higher order behavior of the eigenvalues of a single anharmonic oscillator.

In the second method we use degenerate perturbation theory directly to treat the term $b(x-y)^2$ as a perturbation of two uncoupled anharmonic oscillators.

We conclude with a proof of the validity of the formulas obtained by these methods.

II. THE CRITICAL POINTS OF $V(x, y)$

$V(x,y)$ has nine critical points. Since $V(x,y) \geqslant 0$ for all (x,y) and since $V(x,y) = 0$ only if $x = y = \pm a$ the two points $\pm(a,a)$ are the global minima of $V(x,y)$. The other seven critical points are found by solving

$$0 = \partial_x V(x,y) = 4g(x^2-a^2)x + 2b(x-y),$$

$$0 = \partial_y V(x,y) = 4g(y^2-a^2)y - 2b(x-y).$$

All solutions of these two cubic equations can be found by making the change of variables $u = (x+y)/\sqrt{2}$, $v = (x-y)/\sqrt{2}$ and adding and subtracting the resulting equations in u and v. The solutions are (i) $(0,0)$, (ii) $\pm(a,a)$, (iii) $\pm(a^2 - b/g)^{1/2}(1,-1)$, (iv) $\pm\tfrac{1}{2}(\pm(a^2+b/2g)^{1/2} + (a^2 - 3b/2g)^{1/2}, \pm(a^2+b/2g)^{1/2} - (a^2-3b/2g)^{1/2})$.

We note that for g large the two points (iii) are approximately $(a,-a)$ and $(-a,a)$ while the four points (iv) are approximately $(a,0)$, $(-a,0)$, $(0,a)$, $(0,-a)$.

It follows from evaluating the Hessian matrix

$$D^2V = \begin{bmatrix} 4g(3x^2-a^2) + 2b & -2b \\ -2b & 4g(3y^2-a^2) + 2b \end{bmatrix}$$

at the critical points (i) through (iv) that for large g $(0,0)$ is a local maximum, the four points (ii) and (iii) are local minima, and the four points (iv) are saddle points. The values of $V(x,y)$

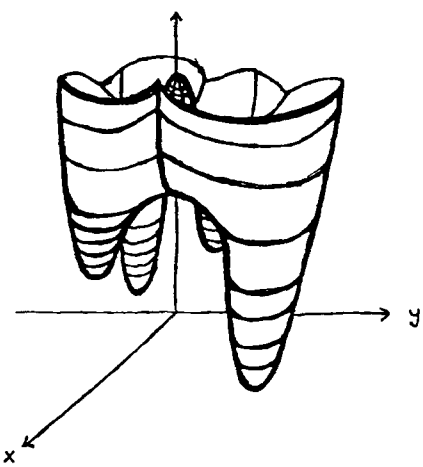


FIG. 1. Approximate sketch of $V(x,y)/2$ for g large.

[a)] This research is partially supported by the National Science Foundation under Grants No. INT81-15614 and MCS81-02426.

 0022-2488/83/061477-05$02.50 © 1983 American Institute of Physics

at these points are given by $V(\mathrm{i}) = 2ga^4$, $V(\mathrm{ii}) = 0$, $V(\mathrm{iii}) = 4a^2b - 2b^2/g$, and for g large $V(\mathrm{iv}) \approx ga^4 + ba^2$. Thus $V(x, y)$ has four minima which become separated from each other by barriers whose height increases as g approaches infinity. An approximate sketch of $V(x, y)/2$ for g large is given in Fig. 1.

III. EXPANSION AND RESCALING OF H ABOUT THE MINIMA OF V

In order to find the contribution to the spectra of H due to each of the four minima we expand $V(x, y)$ about each of them. We expand about the global minima $\pm(a, a)$ by letting $u = x \mp a$, $v = y \mp a$. We next scale by the unitary transformation U where $U^* u U = \epsilon u$, $U^* v U = \epsilon v$ and $\epsilon = (4ga^2)^{-1/4}$. H is then unitarily equivalent to the operators $A_\pm$ where

$$A_\pm = (\tfrac{1}{2})\epsilon^{-2}[-\partial_u^2 - \partial_v^2 + u^2 + v^2 \pm \epsilon/a(u^3 + v^3) + (\epsilon/2a)^2(u^4 + v^4) + b\epsilon^4(u - v)^2]. \quad (3)$$

When we expand about the other two minima $\pm(a^2 - b/g)^{1/2}(1, -1)$ by setting $u = x \mp (a^2 - b/g)^{1/2}$ and $v = y \pm (a^2 - b/g)^{1/2}$ and scale as before H becomes unitarily equivalent to the operators

$$B_\pm = (\tfrac{1}{2})\epsilon^{-2}[-\partial_u^2 - \partial_v^2 + (u^2 + v^2)(1 - 6b\epsilon^4) \pm (\epsilon/a)(1 - 4b\epsilon^4)^{1/2}(u^3 - v^3) + (\epsilon/2a)^2(u^4 + v^4) + b\epsilon^4(u - v)^2] + 2a^2b - (\epsilon^2 2ab)^2. \quad (4)$$

The formulas (3) and (4) for $A_\pm$ and $B_\pm$ show us that formally $A_\pm = \epsilon^{-2}\tfrac{1}{2}[-\partial_u^2 - \partial_v^2 + u^2 + v^2] + O(\epsilon^{-1}) = B_\pm$. Since the harmonic oscillator $\tfrac{1}{2}[-\partial_x^2 + x^2]$ has eigenvalues $j + \tfrac{1}{2}$ for $j = 0,1,2,\cdots$ we conjecture that the eigenvalues of H are given for $n, m = 0,1,2,\cdots$ by

$$\epsilon^{-2}(n + m + 1) + O(\epsilon^{-1}).$$

This conjecture is proven in the last section.

In order to obtain the eigenvalues of H to $o(1)$ we observe that

$$A_\pm = \epsilon^{-2}[\mathscr{A}_\pm(u) + \mathscr{A}_\pm(v)] + O(\epsilon^2),$$
$$B_\pm = \epsilon^{-2}[\mathscr{A}_\pm(u) + \mathscr{A}_\pm(-v)] + 2a^2b + O(\epsilon^2), \quad (5)$$

where

$$\mathscr{A}_\pm(x) = \tfrac{1}{2}[-\partial_x^2 + x^2 \pm (\epsilon/a)x^3 + (\epsilon/2a)^2 x^4].$$

We denote the eigenvalues and eigenfunctions of $\mathscr{A}_\pm$ by $\alpha_j^\pm = \alpha_j^\pm(\epsilon)$ and $\psi_j^\pm(x)$, respectively. Thus for $j = 0,1,2,\cdots$

$$\mathscr{A}_\pm(x)\psi_j^\pm(x) = \alpha_j^\pm \psi_j^\pm(x).$$

Since $\mathscr{A}_\pm(x)$ is unitarily equivalent to $\mathscr{A}_\pm(-x) = \mathscr{A}_\mp(x)$ [$R^* \mathscr{A}_\pm(x) R = \mathscr{A}_\pm(-x)$ where $R^* x R = -x$] and since the eigenvalues $\alpha_j^\pm$ are nondegenerate it follows that $\alpha_j^+ = \alpha_j^-$. In what follows we omit the $\pm$ superscripts on the α_j.

The expressions (5) for $A_\pm$ and $B_\pm$ suggest that H has eigenvalues given by

$$\epsilon^{-2}[\alpha_n(\epsilon) + \alpha_m(\epsilon)] + O(\epsilon^2)$$

and

$$\epsilon^{-2}[\alpha_n(\epsilon) + \alpha_m(\epsilon)] + 2a^2b + O(\epsilon^2).$$

The eigenvalues of H will be given formally to $o(1)$ if we replace the eigenvalues $\alpha_n(\epsilon)$ of $\mathscr{A}_\pm(x)$ in the preceding formulas by expressions accurate to $o(\epsilon^2)$. It is proven in Refs. 1 and 2 that the eigenvalues of $\mathscr{A}_\pm(x)$ become asymptotically degenerate in pairs and that they satisfy

$$\alpha_{2j}(\epsilon) = j + \tfrac{1}{2} - (3j^2 + 3j + 1)(\epsilon/2a)^2 + o(\epsilon^2)$$
$$\alpha_{2j+1}(\epsilon) = \alpha_{2j}(\epsilon) + o(\epsilon^2)$$

for $j = 0,1,2,\cdots$. It follows formally that H has eigenvalues given by

$$\epsilon^{-2}(n + m + 1) - (2a)^{-2}[3(n^2 + m^2) + 3(n + m) + 2] + o(1)$$

and

$$\epsilon^{-2}(n + m + 1) - (2a)^{-2}[3(n^2 + m^2) + 3(n + m) + 2] + 2a^2b + o(1).$$

Let $\lambda_{n,m} \equiv \epsilon^{-2}(n + m + 1) - (2a)^{-2}[3(n^2 + m^2) + 3(n + m) + 2]$. This formal calculation and symmetry considerations suggest that $E_0 = \lambda_{0,0} + o(1)$, $E_1 = \lambda_{0,0} + o(1)$ and that $E_2 = \lambda_{0,0} + 2a^2b + o(1)$. These formulas are derived by a different calculation in the next section. They are proven in the final section.

IV. DEGENERATE PERTURBATION THEORY

We describe another method for calculating the eigenvalues of H as g becomes large. Complete the square in the term $b(x - y)^2$, let $r^2 \equiv (a^2 - b/2g)$, and scale H by $x = \epsilon u$, $y = \epsilon v$, $\epsilon = (4r^2 g)^{-1/4}$ and $s \equiv (2r)^{-3} g^{-1/2}$. The original operator H then becomes unitarily equivalent to

$$(4r^2 s)^{-1}[H_0 - 2(4r^2 s)^2 buv] + a^2 b - b^2/4g,$$

where

$$H_0 = \tfrac{1}{2}[-\partial_u^2 - \partial_v^2 + s((u^2 - (4s)^{-1})^2 + (v^2 - (4s)^{-1})^2)].$$

In this scaling s goes to zero as g goes to infinity, and the term uv has the smallest coefficient ($O(s^2)$). We therefore treat this term as a perturbation of H_0. We use degenerate perturbation theory because the eigenvalues of H_0 are either degenerate or asymptotically degenerate as s approaches zero. To see this we denote the eigenvalues and eigenfunctions of

$$A(x) \equiv \tfrac{1}{2}[-\partial_x^2 + s(x^2 - (4s)^{-1})^2]$$

by a_j and $\psi_j(x)$ where $a_0 < a_1 < a_2 < \cdots$, $\psi_{2j}(x) = \psi_{2j}(-x)$, and $\psi_{2j+1}(x) = -\psi_{2j+1}(x)$. It then follows from Refs. 1 and 2

$$a_0 = \tfrac{1}{2} - s + o(s),$$
$$a_1 = a_0 + \Delta_1(s) \text{ and } \Delta_1(s) = o(s^n) \quad \text{for all } n,$$
$$a_2 = 1 + \tfrac{1}{2} - 7s + o(s),$$
$$a_3 = a_2 + \Delta_3(s) \text{ and } \Delta_3(s) = o(s^n) \quad \text{for all } n.$$

The lowest few eigenvalues e_j and corresponding eigenfunctions $\Psi_j(u, v)$ of H_0 are then given by $e_0 = 2a_0 = 1 - 2s + o(s)$; $\Psi_0(u, v) = \psi_0(u)\psi_0(v)$, $e_1 = e_2 = a_0 + a_1 = 2a_0 + \Delta_1(s) = 1 - 2s + o(s)$; $\Psi_1(u, v) = \psi_0(u)\psi_1(v)$ and $\Psi_2(u, v) = \psi_0(v)\psi_1(u)$, $e_3 = a_1 + a_1 = 2a_0 + 2\Delta_1(s) = 1 - 2s + o(s)$; $\Psi_3(u, v) = \psi_1(u)\psi_1(v)$, $e_4 = e_5 = a_2 + a_0 = 2 - 8s + o(s)$; $\Psi_4(u, v) = \psi_2(u)\psi_0(v)$, $\Psi_5(u, v) = \psi_0(u)\psi_2(v)$. Thus the lowest four eigenvalues of H_0

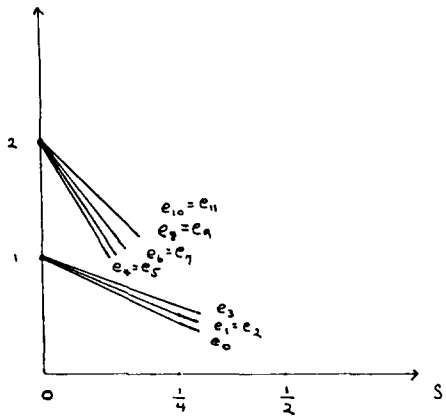


FIG. 2. Approximate sketch of lowest eleven eigenvalues of H_0.

are within $\Delta_1(s)$ (which is exponentially small[3]) of each other while the rest of the spectrum is separated from the lowest four by $O(1)$ (see Fig. 2). Since the perturbation is $O(s^2)$ we treat the four lowest eigenvalues of H by degenerate perturbation theory. We remark that the next lowest eight eigenvalues of H may be treated by degenerate perturbation theory and in principle all the higher eigenvalues may be treated similarly although we do not do so here.

The calculation is simplified if we introduce the notation

$$L = A + \mu B,$$

where $A \equiv H_0 - 2a_0$, $\mu \equiv \frac{1}{2}(4r^2s)^2 b$, $\Delta \equiv \Delta_1$, and $B \equiv -2uv$. Then $A\Psi_0 = 0$, $A\Psi_1 = \Delta\Psi_1$, $A\Psi_2 = \Delta\Psi_2$, $A\Psi_3 = 2\Delta\Psi_3$. Let $n(i)$ and $m(i)$ be integers such that for $i = 0,1,2,3$, $\Psi_i(u, v) = \psi_{n(i)}(u)\psi_{m(i)}(v)$. If we define

$$\langle \Psi_i, L\Psi_j \rangle \equiv \int\int_{-\infty}^{+\infty} \Psi_i^*(u, v) L\Psi_j(u, v)\, du\, dv$$

then

$$\begin{aligned}\langle \Psi_i, L\Psi_j \rangle &= e_j\delta_{ij} + \mu\langle \Psi_i, B\Psi_j \rangle \\ &= e_j\delta_{ij} - 2\mu\left[\int_{-\infty}^{+\infty} \psi^*_{n(i)}(u) u \psi_{n(j)}(u) du\right] \\ &\quad \times \left[\int_{-\infty}^{+\infty} \psi^*_{m(i)}(v) v \psi_{m(j)}(v) dv\right].\end{aligned}$$

From symmetry $\int_{-\infty}^{+\infty} \psi_i^*(x) x \psi_i(x) dx = 0$ for $i = 0, 1$. Let $\beta \equiv \int_{-\infty}^{+\infty} \psi_0^*(x) x \psi_1(x) dx$.

The first step in degenerate perturbation theory is to diagonalize the four by four matrix $(L_{ij}) \equiv (\langle \Psi_i, L\Psi_j \rangle)$. This task is simplified by observing that L and A have the same symmetries. More specifically $[U, L] = [U, A] = [R, L] = [R, A] = 0$ where $[C, D] = CD - DC$ is the commutator, $U\Psi(u,v) = \Psi(-u, -v)$, and $R\Psi(u,v) = \Psi(v,u)$. We diagonalize (L_{ij}) by taking as a basis for the span of $\{\Psi_j\}_0^3$ the eigenfunctions of U and R given by $\Phi_0 = \Psi_0$, $\Phi_1 = \Psi_3$, $\Phi_2 = (\Psi_1 + \Psi_2)/\sqrt{2}$, and $\Phi_3 = (\Psi_1 - \Psi_3)/\sqrt{2}$. The eigenvalues of (L_{ij}) are then the same as those of

$$(\langle \Phi_i, L\Phi_j \rangle)$$

$$= \begin{bmatrix} 0 & -2\mu\beta^2 & 0 & 0 \\ -2\mu\beta^2 & 2\Delta & 0 & 0 \\ 0 & 0 & \Delta - 2\mu\beta^2 & 0 \\ 0 & 0 & 0 & \Delta + 2\mu\beta^2 \end{bmatrix}.$$

It is now easily seen that the eigenvalues of (L_{ij}) are

$$\Delta - (\Delta^2 + (2\mu\beta^2)^2)^{1/2} = -2\mu\beta^2 + o(s),$$

$$\Delta - 2\mu\beta^2 = -2\mu\beta^2 + o(s),$$

$$\Delta + 2\mu\beta^2 = 2\mu\beta^2 + o(s),$$

$$\Delta + (\Delta^2 + (2\mu\beta^2)^2)^{1/2} = 2\mu\beta^2 + o(s).$$

It is shown in Ref. 4 that $\beta^2 = (4s)^{-1} + 2^{-1} + o(s)$. Since H is unitarily equivalent to

$$(4r^2s)^{-1}[L + 2a_0] + a^2b - b^2/4g$$

we have that the lowest four eigenvalues of H are given formally by degenerate perturbation theory as

$$\begin{aligned} E_0 &= (4r^2s)^{-1}[2a_0 - 2\mu\beta^2] + a^2b - b^2/4g + o(1) \\ &= 2ag^{1/2} - a^{-2}\tfrac{1}{2} + o(1), \\ E_1 &= E_0 + o(1), \\ E_2 &= (4r^2s)^{-1}[2a_0 + 2\mu\beta^2] + a^2b - b^2/4g + o(1) \\ &= 2ag^{1/2} - a^{-2}\tfrac{1}{2} + 2a^2b + o(1), \\ E_3 &= E_2 + o(1). \end{aligned}$$

The lowest four eigenvalues obtained by degenerate perturbation theory thus agree with the values $\lambda_{0,0}$ and $\lambda_{0,0} + 2a^2b$, found by the first method, to $o(1)$. We outline a rigorous proof that the lowest two eigenvalues of H are asymptotic to $\lambda_{0,0}$ and that the next highest two eigenvalues are asymptotic to $\lambda_{0,0} + 2a^2b$ in the next section.

V. ESTIMATES AND PROOFS

We sketch proofs of the facts that every eigenvalue of H has the form $\lambda_{n,m} + o(1)$ or $\lambda_{n,m} + 2a^2b + o(1)$ as g becomes large, and that $E_0 = \lambda_{0,0} + o(1)$, $E_1 = E_0 + o(1)$, $E_2 = E_0 + 2a^2b + o(1)$, and $E_3 = E_2 + o(1)$.

We first prove the weaker result that every eigenvalue of H is of the form $(4a^2g)^{1/2}(n + m + 1) + o(g^{1/2})$. Let Γ_i, $i = 1,2,3,4$ denote discs of radius $r_i = r_i(g)$ centered at the ith minima of $V(x, y)$. Define H^D and H^N to be the self-adjoint operators formed by replacing the Laplacian, Δ, in H with Δ_D and Δ_N, where Δ_D and Δ_N denote the Laplacian operators with Dirichlet and Neumann boundary conditions on $\partial\Gamma = \cup_{i=1}^4 \partial\Gamma_i$, respectively. We denote the eigenvalues and eigenfunctions of H^D and H^N by E_j^D, Ψ_j^D and E_j^N, Ψ_j^N, respectively. Here $E_0^D \leqslant E_1^D \leqslant \cdots$, $E_0^N \leqslant E_1^N \leqslant \cdots$ and the indexing is such that a degenerate eigenvalue will appear in the list of eigenvalues as many times as its multiplicity. From the minimax principle[5]

$$E_j^N \leqslant E_j \leqslant E_j^D.$$

To prove that E_j is of the form $(4a^2g)^{1/2}(n + m + 1) + o(g^{1/2})$ we show that E_j^D and E_j^N are both of the form $(4a^2g)^{1/2}(n + m + 1) + o(g^{1/2})$.

The spectrum of H^D is the union of the spectra of H_i^D

$\equiv H^D \restriction L_2(\Gamma_i)$ for $i = 1,2,3,4$ with the spectra of $H_5^D \equiv H^D \restriction L_2(\sim\Gamma)$. We first show that every eigenvalue of H_1^D is of the form $(4a^2g)^{1/2}(n+m+1)+o(g^{1/2})$. Make the unitary transformation that places the origin of our coordinates at the $x=a, y=a$ minima of $V(x,y)$. Then scale by $\epsilon = (4ga^2)^{-1/4}$. H_1^D becomes unitarily equivalent to $\epsilon^{-2}A_1^D$ where

$$A_1^D = \tfrac{1}{2}[-\Delta_1^D + P(u,v)],$$
$$P(u,v) \equiv Q(u,v) + \epsilon/a(u^3+v^3) + (\epsilon/2a)^2(u^4+v^4) + b\epsilon^4(u-v)^2,$$
$$Q(u,v) = u^2+v^2$$

and $-\Delta_1^D$ is the Laplacian $-\partial_u^2 - \partial_v^2$ with Dirichlet boundary conditions on the circle $u^2+v^2=\rho^2=\rho(\epsilon)^2$ with $\rho(\epsilon) = r_1(g)/\epsilon$. If we choose $\rho = \epsilon^{-1/4+\delta}$ for some $0<\delta<1/4$ then ρ goes to infinity as ϵ goes to zero, and there is a constant $M>0$ (independent of ϵ) for which

$$\max_{u^2+v^2\leqslant\rho^2} |P-Q| \leqslant M\epsilon^{4\delta}.$$

It follows that the eigenvalues e_j^D of A_1^D satisfy

$$a_j^D - M\epsilon^{4\delta} \leqslant e_j^D \leqslant a_j^D + M\epsilon^{4\delta},$$

where a_j^D denotes the eigenvalues of $L^D = \tfrac{1}{2}[-\Delta_1^D + Q(u,v)]$. We next show that $a_j^D = n+m+1+O(e^{-c\rho})$. Here $\alpha_{n,m} \equiv n+m+1$ are the eigenvalues of $L = \tfrac{1}{2}[-\Delta + Q(u,v)]$ on $L_2(R^2)$. The corresponding eigenfunctions of L are $\Psi_{n,m} = H_n(u)H_m(v)$ where $H_n(x)$ denotes the nth Hermite function. It follows from a more general result in Ref. 6 that the eigenvalues of L^D monotonically decrease to those of L as ρ goes to infinity. We give a simple proof that the rate of convergence is exponential.

Let $J = J(\lambda;\rho)$ be a C^∞ function of $\lambda = (u^2+v^2)^{1/2}$ which is identically one for $\lambda \leqslant \rho-1$, and is identically zero for $\lambda \geqslant \rho$, and $0 \leqslant J \leqslant 1$ for all other values of λ. The functions $J\Psi_{n,m}$ are then in the domains of L and L^D. Denote by a^D the eigenvalue of L^D closest to the eigenvalue $\alpha_{n,m}$ of L. Temple's inequality[7] implies that

$$|a^D - \alpha_{m,n}| \leqslant \|L^D(J\Psi_{n,m}) - \alpha_{n,m}(J\Psi_{n,m})\| \, \|J\Psi_{n,m}\|^{-1} = \|(L-\alpha_{n,m}I)J\Psi_{n,m}\| \, \|J\Psi_{n,m}\|^{-1}.$$

We show that this expression is exponentially small in ρ. First $\|J\Psi_{n,m}\|$ converges to one exponentially fast because

$$\|J\Psi_{n,m}\|^2 \leqslant \|\Psi_{n,m}\|^2 + \|(I-J)\Psi_{n,m}\|^2 = 1 + \int_{-\infty}^{+\infty}\int (1-J)^2\Psi_{n,m}^2 \, du\, dv;$$

however,

$$\int_{-\infty}^{+\infty}\int (1-J)^2\Psi_{n,m}^2 \, du\, dv \leqslant \iint_{\lambda>\rho-1} \Psi_{n,m}^2 \, du\, dv,$$

which for fixed n and m goes to zero exponentially fast in ρ. Second

$$\|(L-\alpha_{n,m}I)J\Psi_{n,m}\| = \|J(L-\alpha_{n,m}I)\Psi_{n,m} + [(L-\alpha_{n,m}I),J]\Psi_{n,m}\| = \|[L,J]\Psi_{n,m}\| = \tfrac{1}{2}\|[\Delta,J]\Psi_{n,m}\| \leqslant \|\nabla J\cdot\nabla\Psi_{n,m}\| + \tfrac{1}{2}\|(\Delta J)\Psi_{n,m}\| \leqslant \pi(2\rho-1) \times \max_{\rho-1\leqslant u^2+v^2\leqslant\rho} \{|\nabla J\cdot\nabla\Psi_{n,m}| + \tfrac{1}{2}|(\Delta J)\Psi_{n,m}|\}.$$

This expression goes to zero exponentially fast because of the exponential decay of the Hermite functions and their derivatives.

The preceding argument has shown that the eigenvalues E_j^D of H_1^D satisfy the equation $E_j^D = (4a^2g)^{1/2}(n+m+1)+O(g^{1/2-\delta})$ for any $0<\delta<1/4$. A similar argument shows that for $j=2,3,4$ the eigenvalues of H_j^D and H_j^N are given by the same formula. This is mainly due to the fact that the quadratic approximations to V in the neighborhood of each of its minima are the same.

The eigenvalues of H_5^N are bounded from below by $\min_{\sim\Gamma} V(x,y) \geqslant$ (constant) $g^{5/8-\delta/2}$. Thus the jth largest eigenvalue of H will, for large enough g, be of the form $(4a^2g)^{1/2}(n+m+1)+o(g^{1/2})$ as stated. We point out that since the lowest four eigenvalues of H^N and H^D (one corresponding to each minima of V) are all of the form $(4a^2g)^{1/2}+O(g^{1/2-\delta})$ the lowest four eigenvalues of H are also of the form $(4a^2g)^{1/2}+O(g^{1/2-\delta})$.

We prove next that for every n, $m = 0,1,2,\cdots$ there are eigenvalues E_j and E_k of H that satisfy $E_j = \lambda_{m,n} + o(1)$ and $E_k = \lambda_{m,n} + 2a^2b + o(1)$. From Sec. III H is unitarily equivalent to A_+ where $A_+ = \epsilon^{-2}[\mathscr{A}_+(u) + \mathscr{A}_+(v)] + (b\epsilon^2/2)(u-v)^2$ and $\mathscr{A}_+(x)\psi_n^+(x) = \alpha_n^+ \psi_n^+(x)$. We drop the $+$ in what follows.

Let E_j denote the eigenvalue of A closest to $\epsilon^{-2}(\alpha_{2n}+\alpha_{2m})$. Let $\Omega_{n,m} = \Omega_{n,m}(u,v) = [\psi_{2n}(u)\psi_{2m+1}(v) + \psi_{2n+1}(u)\psi_{2m}(v)]2^{-1/2}$. From Temple's inequality

$$|E_j - \epsilon^{-2}(\alpha_{2n}+\alpha_{2m})| \leqslant \|[A-\epsilon^{-2}(\alpha_{2n}+\alpha_{2m})]\Omega_{n,m}\|, = \epsilon^{-2}\|(\alpha_{2m+1}-\alpha_{2m})\psi_{2n}(u)\psi_{2m+1}(v) + (\alpha_{2n+1}-\alpha_{2n})\psi_{2n+1}(u)\psi_{2m}(v) + 2^{-1}b\epsilon^4(u-v)^2\Omega_{n,m}\| \leqslant \epsilon^{-2}[(\alpha_{2m+1}-\alpha_{2m}) + (\alpha_{2n+1}-\alpha_{2n})] + 2^{-1}b\epsilon^2\|(u-v)^2\Omega_{n,m}\|.$$

Since the gaps $\alpha_{2j+1}-\alpha_{2j}$ decrease faster than any power of ϵ, the first term is certainly $o(1)$. The second term is also $o(1)$. This follows from the formula

$$\|(u-v)^2\Omega_{n,m}\|^2 = \langle \Omega_{n,m}, (u-v)^4\Omega_{n,m}\rangle$$

and the one oscillator expectations

$$\langle\psi_{2j}(u), u^k\psi_{2l}(u)\rangle = 2^{-1}(-2a/\epsilon)^k\delta_{j,l} + O(\epsilon^{-k+1}),$$
$$\langle\psi_{2j+1}(u), u^k\psi_{2l+1}(u)\rangle = 2^{-1}(-2a/\epsilon)^k\delta_{j,l} + O(\epsilon^{-k+1}),$$
$$\langle\psi_{2j}(u), u^k\psi_{2l+1}(u)\rangle = -2^{-1}(-2a/\epsilon)^k\delta_{j,l} + O(\epsilon^{-k+1}).$$

Explicitly $\langle\Omega_{n,m}, (u-v)^4\Omega_{n,m}\rangle = \langle\Omega_{nm}, [u^4+v^4-4u^3v + 6u^2v^2 - 4uv^3]\Omega_{nm}\rangle = O(\epsilon^{-3})$ because the coefficients of the leading ϵ^{-4} term add up to zero. Thus $|E_j - \epsilon^{-2}(\alpha_{2n}$

$+\alpha_{2m})| = O(\epsilon^{1/2}) = o(1)$. It follows from the formulas for the one oscillator eigenvalues that $E_j = \lambda_{m,n} + o(1)$.

A similar estimate on the expression

$$\| [B_+ - (\epsilon^{-2}(\alpha_{2n} + \alpha_{2m}) + 2a^2 b)I]\Omega_{n,m}(u, -v)\|$$

yields an eigenvalue E_k of H of the form $\lambda_{n,m} + 2a^2 b + o(1)$.

We now combine the preceding argument with the knowledge that the four lowest eigenvalues of H are of the form $(4a^2 g)^{1/2} + O(g^{1/2-\delta})$ to prove the $E_0 = \lambda_{0,0} + o(1)$, $E_1 = E_0 + o(1)$, $E_2 = E_0 + 2a^2 b + o(1)$ and $E_3 = E_2 + o(1)$.

We show that H has four eigenvalues, two of the form $\lambda_{0,0} + o(1)$, and two of the form $\lambda_{0,0} + 2a^2 b + o(1)$. It then follows from the previously established facts that the four lowest eigenvalues of H are given by $(4a^2 g)^{1/2} + O(g^{1/2-\delta})$ and the next highest eigenvalue of H is of the form $2\cdot(4a^2 g)^{1/2} + O(g^{1/2-\delta})$ that the two pairs of eigenvalues of H given by $\lambda_{0,0} + o(1)$ and $\lambda_{0,0} + 2a^2 b + o(1)$ must be the lowest four eigenvalues of H.

That H has at least two eigenvalues of the form $\lambda_{0,0} + o(1)$ follows from the observation that the restrictions of A_+ to its invariant subspaces $\mathscr{H}_+ \equiv \{\Psi \mid \Psi(u, v) = \Psi(-u - 2a/\epsilon, -v - 2a/\epsilon)\}$ and $\mathscr{H}_- \equiv \{\Psi \mid \Psi(u, v) = -\Psi(-u - 2a/\epsilon, -v - 2a/\epsilon)\}$ have eigenvalues E_j and E_k that are of the form $\lambda_{0,0} + o(1)$. In other words the closest eigenvalue E_j of $A_+ \upharpoonright \mathscr{H}_+$ to $\lambda_{0,0}$ and the closest eigenvalue E_k of $A_+ \upharpoonright \mathscr{H}_-$ to $\lambda_{0,0}$ satisfy $|E_j - \lambda_{0,0}| = o(1) = |E_k - \lambda_{0,0}|$. This is proven by taking $\Omega_{0,0} \in \mathscr{H}_-$ and $\Phi(u, v) \equiv [\psi_0(u)\psi_0(v) + \psi_1(u)\psi_1(v)]2^{-1/2} \in \mathscr{H}_+$ as trial functions in Temple's inequality. In particular $|E_j - \lambda_{0,0}| \leqslant \|[A_+ - \lambda_{0,0}]\Omega_{0,0}\| = o(1)$ and $|E_k - \lambda_{0,0}| \leqslant \|[A_+ - \lambda_{0,0}]\Phi\| = o(1)$ by the preceding estimates. This proves that H possesses at least two eigenvalues (or one eigenvalue with multiplicity at least two) which differ from $\lambda_{0,0}$ by $o(1)$.

A similar calculation with the orthonormal trial functions $\psi_0 \equiv \Omega_{0,0}(u, -v)$ and $\psi_1 \equiv \Phi(u, -v)$ shows that $\|r_j\| = o(1)$, where $r_j \equiv [B_+ - \lambda]\psi_j, \lambda \equiv \lambda_{0,0} + 2a^2 b$, and $j = 0, 1$. We sketch an argument that shows that since ψ_0 and ψ_1 are perpendicular, B_+ has two eigenvalues (or one of multiplicity at least two) within $o(1)$ of λ.

Let P be the orthogonal projection onto the span of $\{\psi_0, \psi_1\}$ given by

$$P\psi = \psi_0\langle\psi_0, \psi\rangle + \psi_1\langle\psi_1, \psi\rangle.$$

Let $Q \equiv 1 - P$. Then

$$B_+ P\psi = \lambda P\psi + r_0\langle\psi_0, \psi\rangle + r_1\langle\psi_1, \psi\rangle$$

or $B_+ P = \lambda P + R$ where $R\psi = r_0\langle\psi_0, \psi\rangle + r_1\langle\psi_1, \psi\rangle$. We thus write

$$B_+ = (P + Q)B_+(P + Q) = E + D,$$

where $E = \lambda P + QB_+ Q$ and $D = 2^{-1}(PR + R^* P) + R^* Q + QR$. Thus as quadratic forms on $\mathscr{S} \times \mathscr{S}$ we have that $E - \|D\| \leqslant B_+ \leqslant E + \|D\|$. Since $\|P\| = \|Q\| = 1$ and $\|R\| = \|R^*\| \leqslant \|r_0\| + \|r_1\| = o(1)$ it follows that $\|D\| = o(1)$. From the mini-max principle and the fact that λ is an eigenvalue of E of multiplicity two it follows that B_+ has two eigenvalues of the form $\lambda + o(1)$. Hence the four lowest eigenvalues of H are $E_0 = \lambda_{0,0} + o(1)$, $E_1 = E_0 + o(1)$, $E_2 = \lambda_{0,0} + 2a^2 b + o(1)$, and $E_3 = E_2 + o(1)$.

We remark that the fact that H has four and only four eigenvalues within $O(1)$ of $\lambda_{0,0}$ may also be proven by showing that (in the notation of Sec. IV) $\|(z - L)^{-1} - (z - A)^{-1}\|$ is uniformly small for z any point on an appropriately chosen curve containing $\lambda_{0,0}$ in its interior.

ACKNOWLEDGMENTS

I would like to thank Eli Isaacson, Dan Marchesin, and Paulo Jorge Paes-Leme for many enjoyable conversations about the spectra of coupled anharmonic oscillators. In particular the formulas and proofs in this paper were given mainly for the purpose of checking the numerical calculations of the spectra of coupled systems of anharmonic oscillators in Ref. 8.

[1] D. Isaacson, "Singular perturbations and asymptotic eigenvalue degeneracy," Comm. Pure Appl. Math. **29**, 531–551 (1976).
[2] M. Reed and B. Simon, *Methods of Modern Mathematical Physics IV: Analysis of Operators* (Academic, New York, 1978), pp. 34–38.
[3] E. M. Harrel, "On the rate of asymptotic eigenvalue degeneracy," Comm. Math. Phys. **60**, 73–95 (1978).
[4] D. Isaacson, "The critical behavior of ϕ^4_1," Comm. Math. Phys. **53**, 257–275 (1977).
[5] R. Courant and D. Hilbert, *Methods of Mathematical Physics* (Interscience, New York, 1953), Vol. I, pp. 397–464.
[6] E. C. Titchmarsh, *Eigenfunction Expansions, Part Two* (Oxford, London, 1958), pp. 148–156.
[7] J. H. Wilkinson, *The Algebraic Eigenvalue Problem* (Clarendon, Oxford, 1965), pp. 170–188.
[8] D. Isaacson, E. L. Isaacson, D. Marchesin, and P. J. Paes-Leme, "Numerical analysis of the spectral properties of coupled oscillator Schrödinger operators. II. Two coupled anharmonic oscillators," SIAM J. Numer. Anal. **19**(1), 126–141 (1981).

On the dynamical symmetry of the Morse potential problem

Chia-Chung Sun and Zong-Hao Zeng
Institute of Theoretical Chemistry, Kirin University, Changchun, People's Republic of China

(Received 18 June 1982; accepted for publication 27 August 1982)

As is well known, the vibration–rotational Schrödinger equation with Morse potential for diatomic molecules has been solved approximately. In this article, a dynamical symmetry analysis of the same Schrödinger equation is performed to give the symmetries described by two kinds of groups, SO(2,1) and $H_4 \times SO(3,2)$.

PACS numbers: 03.65.Ge, 02.20. + b, 33.90. + h

I. INTRODUCTION

In the development of dynamical symmetry analysis of physical problems by the group-theoretical method, much valuable work was contributed by Barut,[1,2] Dothan,[3,4] Mukuda,[5] and Anderson[6,7] *et al.*[8,9] As is well known, the dynamical symmetries of some basic problems such as the hydrogen atom, the harmonic oscillator, and the symmetric top, etc., have been successfully performed. Also, in case the angular momentum quantum number l equals zero, the energy spectrum with respect to the Morse potential[10] has been discussed by means of the group SO(2,1).[11] In molecular physics, the Schrödinger equation with the Morse potential for arbitrary l has been solved approximately[12] to give explicit results of eigenvalues. In this article, a dynamical symmetry analysis of the problem with the Morse potential under the same approximation is performed. As a result, the direct product group $H_4 \times SO(3,2)$ can be used to adapt to the dynamical symmetry and the group H_4 as well as SO(2,1) can be employed to produce the energy spectrum.

2. GROUP SO(2,1) AND ENERGY SPECTRUM

For illustration of vibration–rotational spectra of diatomic molecules, Morse has proposed the following Schrödinger equation[10]:

$$-(1/2\mu)\nabla^2\psi + D(1 - e^{-a(r-r_e)})^2\psi = 0 \quad (\hbar = 1), \tag{1}$$

where D and a are empirical parameters, and r_e is the equilibrium distance between two nuclei of the molecule. By means of the conventional variable separation method, Eq. (1) can be separated into two parts. One is the angular part,

$$\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial Y}{\partial\theta}\right) + \frac{1}{\sin^2\theta}\frac{\partial^2 Y}{\partial\phi^2} = l(l+1)Y(\theta,\phi), \tag{2}$$

and the other is the radial part,

$$\left\{-\frac{1}{r^2}\frac{d}{dr}\left(r^2\frac{d}{dr}\right) + \frac{l(l+1)}{r^2} + 2\mu D\left[1 - e^{-a(r-r_e)}\right]^2\right\} \times R(r) = 2\mu E R(r). \tag{3}$$

It is known that Eq. (2), whose solution is spherical harmonics Y_{lm}, is the differential equation of angular momentum with dynamical symmetry of group SO(3,2).[13,7] In case the rotational quantum number equals zero, $l = 0$, Eq. (3) has been solved exactly by Morse,[10] but when $l \neq 0$, an approximate and reasonable solution has been obtained by keeping the terms up to $(y-1)^2$ in the expansion of r_e^2/r^2 in terms of $(y-1)^n$,[12]

$$\frac{r_e^2}{r^2} = C_0 + C_1 y + C_2 y^2,$$

where $y = e^{-a(r-r_e)}$ and

$$A = \frac{l(l+1)}{2\mu r_e^2}, \tag{4}$$

$$C_0 = A\left(1 - \frac{3}{ar_e} + \frac{3}{a^2 r_e^2}\right), \tag{5}$$

$$C_1 = A\left(\frac{4}{ar_e} - \frac{6}{a^2 r_e^2}\right), \tag{6}$$

$$C_2 = A\left(-\frac{1}{ar_e} + \frac{3}{a^2 r_e^2}\right). \tag{7}$$

Under the above approximation, Eq. (3) may be reduced to the following form:

$$-\frac{1}{r^2}\frac{d}{dr}\left(r^2\frac{dR}{dr}\right) + 2\mu \times [D + C_0 - E + (C_1 - 2D)e^{-a(r-r_e)} + (C_2 + D)e^{-2a(r-r_e)}]R = 0. \tag{8}$$

Now let us concentrate our attention upon the discussion of Eq. (8) by means of group theory. To begin with, let us introduce a new set of parameters

$$d^2 = (2\mu/a^2)(D + C_2),$$
$$b^2 = (8\mu/a^2)(D + C_0 - E), \tag{9}$$
$$\lambda = (\mu/a^2 d)(2D - C_1),$$

with $D + C_2 > 0$, $D + C_0 - E > 0$, and define a transformation by writing

$$x = 2dy. \tag{10}$$

Then Eq. (8) becomes

$$(\Gamma_3 + \lambda)\varphi = 0, \tag{11}$$

where $\varphi = rR$, and

$$\Gamma_3 = \frac{d}{dx}\left(x\frac{d}{dx}\right) - \frac{b^2}{4x} \pm \frac{x}{4}, \tag{12}$$

$$\Gamma_2 = -i\left(x\frac{d}{dx} + \tfrac{1}{2}\right). \tag{13}$$

It should be noted that the three infinitesimal operators constitute the algebra of group SO (2,1), i.e.,

$$[\Gamma_1,\Gamma_2] = -i\Gamma_3, \quad [\Gamma_2,\Gamma_3] = i\Gamma_1, \quad [\Gamma_3,\Gamma_1] = i\Gamma_2. \tag{14}$$

 0022-2488/83/061482-05$02.50 © 1983 American Institute of Physics

The corresponding Casimir operator can be written as[9]

$$C = \Gamma_1^2 + \Gamma_2^2 - \Gamma_3^2 = -\Phi(\Phi+1), \tag{15}$$

where $\Phi = \frac{1}{2}(-1 \pm b)$. It is known that, for the discrete representation $D^{-}_{(\Phi)}$[9] with $\Phi = -\frac{1}{2}(b+1)$, Γ_3 in Eq. (11) takes $-\lambda$ as an eigenvalue with

$$\lambda = v + (b+1)/2, \quad v = 0,1,2,\cdots, \tag{16}$$

to give

$$\Gamma_3 w_v^b(x) = -\left(v + \frac{b+1}{2}\right) w_v^b(x), \tag{17}$$

where $\{w_v^b\}$ are the eigenbases. By substituting the parameters of Eq. (9) into Eq. (16), we get

$$E_{vl} = D + C_0 - (a^2/2\mu)[\lambda(l) - v - \tfrac{1}{2}]^2. \tag{18}$$

This expression is in accordance with the quantum mechanical result obtained by Pekeris.[12] For the Morse potential problem with respect to $l = 0$, the spectrum generating group SO(2,1) has been proposed by Cordero *et al.*[11].

Now let us turn our attention to the eigenbases. By utilizing the commutation relations of ladder operators of SO(2,1)

$$[\Gamma_3, \Gamma_\pm] = \mp \Gamma_\pm, \quad [\Gamma_+, \Gamma_-] = 2\Gamma_3, \tag{19}$$

with

$$\Gamma_\pm = -\Gamma_1 \pm i\Gamma_2, \tag{20}$$

it is not difficult to obtain the following results:

$$\Gamma_+ w_v^b = [(v+1)(v+b+1)]^{1/2} w_{v+1}^b, \tag{21}$$

$$\Gamma_- w_v^b = [v(v+b)]^{1/2} w_{v-1}^b, \tag{22}$$

Substitution of $v = 0$ into Eqs. (17) and (22) gives

$$\Gamma_3 w_0^b = -[(b+1)/2] w_0^b, \tag{23}$$

$$\Gamma_- w_0^b = 0. \tag{24}$$

Thus we get immediately

$$w_0^b = [\Gamma(b+1)]^{-1/2} x^{b/2} e^{-x/2}. \tag{25}$$

Application of Eq. (21) to Eq. (25) v times yields

$$w_v^b = [v!\Gamma(v+b+1)]^{-1/2} \Gamma_+^v x^{b/2} e^{-x/2}. \tag{26}$$

For brevity, we only state without proof that the part involving Γ_+ on the right side of Eq. (26) can be expressed in terms of the Laguerre function

$$\Gamma_+^v x^{b/2} e^{-x/2} = x^{-b/2} e^{x/2} \frac{d^n}{dx^n}(x^{n+b} e^{-x}). \tag{27}$$

Clearly, $w_v^b(x)$ is a normalized Laguerre function. Consequently, the normalized radial function appearing in Eq. (8) can be written as

$$R(r) = (ab)^{1/2} r^{-1} w_v^b(x). \tag{28}$$

It should be noted that by means of Eq. (16), b can be expressed in terms of v and l, i.e.,

$$b(v,l) = 2\lambda(l) - 2v - 1. \tag{29}$$

When $r \to \infty$ ($x \to 0$), $R(r)$ must approach zero, so that $b > 0$. Thus from Eq. (14), we get

$$\lambda > \tfrac{1}{2} \quad \text{and} \quad \lambda - \tfrac{1}{2} > v. \tag{30}$$

Clearly, the total wave function can be written as

$$\psi_{vlm}(r,\theta,\phi) = (ab)^{1/2} r^{-1} w_v^b(x) Y_{lm}(\theta,\phi), \tag{31}$$

where $Y_{lm}(\theta,\phi)$ and $w_v^b(x)$ are, respectively, the normalized spherical harmonics and Laguerre function.

It should be pointed out that the ladder operators of SO(2,1) together with Eqs. (21) and (22) are employed only as a mathematical trick to derive the radial functions, as both $w_{v+1}^{b(v,l)}$ and $w_{v-1}^{b(v,l)}$, which appear, respectively, on the right sides of Eqs. (21) and (22), are not the radial functions which satisfy Eq. (8). The reason is that $w_{v+1}^{b(v,l)}$ ($w_{v-1}^{b(v,l)}$) corresponds to two different quantum numbers v and $v+1$ (v and $v-1$). So $\Gamma_\pm$ of SO(2,1) are not the ladder operators to meet the needs of our problem. Hence, SO(2,1) only acts as the spectrum generating group, but not the dynamical group. In the third part of this article, we shall show that group H_4 can be used to produce the energy spectrum as well.

3. GROUP SO(3,2) AND THE QUANTUM NUMBERS l AND m

Now let us discuss the dynamical symmetry of the total wave function given in Eq. (31). Let M_1, M_2, M_3 be the three components of angular momentum which satisfy the commutation relations

$$[M_3, M_\pm] = \pm M_\pm, \quad [M_+, M_-] = 2M_3, \tag{32}$$

where

$$M_\pm = M_1 \pm iM_2, \tag{33}$$

$$M_3 = -i\frac{\partial}{\partial\phi}, \quad M_\pm = \pm e^{\pm i\phi}\left(\frac{\partial}{\partial\theta} \pm i\,\mathrm{ctg}\theta \frac{\partial}{\partial\phi}\right). \tag{34}$$

The corresponding Casimir operator can be written as

$$L^2 = M_1^2 + M_2^2 + M_3^2, \qquad L^2 Y_{lm} = l(l+1) Y_{lm}. \tag{35}$$

Let us define

$$\hat{L} = (L^2 + \tfrac{1}{4})^{1/2} \tag{36}$$

and

$$\overline{Y}_{lm}(\theta,\phi,t_1) = e^{-i\hat{L}t_1} Y_{lm}(\theta,\phi) = Y_{lm}(\theta,\phi) e^{-i(l+1/2)t_1}. \tag{37}$$

Furthermore, we can define another three operators

$$\hat{L}_0 = i\frac{\partial}{\partial t_1}, \tag{38}$$

$$\hat{L}_\pm = \hat{L}_0^{1/2} e^{\pm it_1}\left[\cos\theta\left(i\frac{\partial}{\partial t_1} \pm \frac{1}{2}\right) \pm \sin\theta \frac{\partial}{\partial\theta}\right] \hat{L}_0^{-1/2} \tag{39}$$

which act on the space $\{\overline{Y}_{lm}\}$. By means of straightforward calculation, we get

$$\begin{aligned} &\hat{L}_0 \overline{Y}_{lm} = (l+\tfrac{1}{2})\overline{Y}_{lm}, \quad M_3 \overline{Y}_{lm} = m\overline{Y}_{lm}, \\ &M_\pm \overline{Y}_{lm} = [(l \pm m)(l \pm m + 1)]^{1/2} \overline{Y}_{l,m\pm1}, \\ &\hat{L}_+ \overline{Y}_{lm} = [(l+m+1)(l-m+1)]^{1/2} \overline{Y}_{l+1,m}, \\ &\hat{L}_- \overline{Y}_{lm} = [(l+m)(l-m)]^{1/2} \overline{Y}_{l-1,m}. \end{aligned} \tag{40}$$

From the six operators given in Eqs. (33), (34), (38), and (39), the ten operators

$$J_{12}=M_3,\quad J_{23}=\tfrac{1}{2}(M_+ + M_-),$$
$$J_{13}=\frac{1}{2i}(M_- - M_+),\quad J_{45}=\hat{L}_0,$$
$$J_{34}=-\tfrac{1}{2}(\hat{L}_+ + \hat{L}_-),\quad J_{35}=-\frac{1}{2i}(\hat{L}_- - \hat{L}_+),$$
$$J_{14}=i[J_{13},J_{34}],\quad J_{24}=i[J_{23},J_{34}],$$
$$J_{15}=i[J_{13},J_{35}],\quad J_{25}=i[J_{23},J_{35}] \tag{41}$$

can be introduced to satisfy the commutation relations of SO(3,2),

$$[J_{\alpha\beta},J_{\gamma\delta}]=i(g_{\alpha\delta}J_{\beta\gamma}+g_{\beta\gamma}J_{\alpha\delta}-g_{\alpha\gamma}J_{\beta\delta}-g_{\beta\delta}J_{\alpha\gamma}) \tag{42}$$

where the metric is defined as $g_{\alpha\beta}=(---++)$, and where we have made use of Eq. (32) and

$$[\hat{L}_0,\hat{L}_\pm]=\pm\hat{L}_\pm,\quad [\hat{L}_+,\hat{L}_-]=-2\hat{L}_0,$$
$$[M_3,\hat{L}_\pm]=0,[\hat{L}_0,M_\pm]=0. \tag{43}$$

These results have been obtained by Anderson *et al.*[7] with the aid of the space-time dilation technique.

It should be noted that the ladder operators $\hat{L}_+$ and $\hat{L}_-$ of SO(3,2) in Eq. (40) are used only to shift the quantum number l to $l+1$ or $l-1$ with respect to Y_{lm}, but the same quantum number l also appears in $b\,(v,l)$ of $w_v^{b(v,l)}$, so the operators of SO(3,2) should be further generalized to another form which can be employed to shift the same l in both Y_{lm} and $w_v^{b(v,l)}$.

The operator with respect to the quantity A in Eq. (4) can easily be obtained by use of Eq. (35) to give

$$\hat{A}=L^2/2\mu r_e^2. \tag{44}$$

Then the operators corresponding to the quantities C_0, C_1, C_2 in Eqs. (5)–(7) and the quantities d and λ in Eq. (9) can be expressed as

$$\hat{C}_0=\hat{A}\left(1-\frac{3}{ar_e}+\frac{3}{a^2r_e^2}\right),$$
$$\hat{C}_1=\hat{A}\left(\frac{4}{ar_e}-\frac{6}{a^2r_e^2}\right),$$
$$\hat{C}_2=\hat{A}\left(-\frac{1}{ar_e}+\frac{3}{a^2r_e^2}\right), \tag{45}$$
$$\hat{d}=\left[\frac{2\mu}{a^2}(D+\hat{C}_2)\right]^{1/2},$$
$$\hat{\lambda}=\frac{\mu}{a^2\hat{d}}(2D-\hat{C}_1).$$

From the operators $\hat{d}$ and $\hat{\lambda}$, it is not difficult to find that

$$\hat{d}Y_{lm}=dY_{lm},\quad \hat{\lambda}Y_{lm}=\lambda Y_{lm}, \tag{46}$$

where we have made use of Eq. (35). Also with respect to $x=2dy$ in Eq. (10), we can define the corresponding operator as

$$\hat{x}=2\hat{d}y \tag{47}$$

and

$$\frac{d}{d\hat{x}}=\frac{1}{2\hat{d}}\frac{d}{dy}=\frac{1}{a\hat{x}}\frac{d}{dr}. \tag{48}$$

Taking into account Eqs. (44)–(48), we can recast Eq. (11) into the form

$$\hat{V}w_v^b Y_{lm}=bw_v^b Y_{lm}, \tag{49}$$

with

$$\hat{V}=2(\hat{x})^{1/2}\left[\frac{d}{d\hat{x}}\left(\hat{x}^2\frac{d}{d\hat{x}}\right)-\frac{\hat{x}^2}{4}+\hat{\lambda}\hat{x}+\tfrac{1}{4}\right]^{1/2}\frac{1}{(\hat{x})^{1/2}}, \tag{50}$$

and $b=2\lambda-2v-1$. Taking the advantage of operators $\hat{V}$ and $\hat{L}$, we can define two transformations

$$D_1=e^{-i\hat{V}t_2}e^{-i\hat{L}t_1}(a\hat{V})^{-1/2}r,$$
$$D_1^{-1}=r^{-1}(a\hat{V})^{1/2}e^{i\hat{L}t_1}e^{i\hat{V}t_2}, \tag{51}$$
$$D_1D_1^{-1}=D_1^{-1}D_1=1$$

to yield

$$D_1\psi_{vlm}=f_{vlm}(r,\theta,\phi,t_1,t_2)$$
$$=w_v^b(x)Y_{lm}(\theta,\phi)e^{-i(l+1/2)t_1}e^{-ibt_2}, \tag{52}$$
$$D_1^{-1}f_{vlm}=\psi_{vlm}.$$

The operator of $-b/2$ which acts on space $\{f_{vlm}\}$ may take the form

$$\hat{v}_0=-\frac{i}{2}\frac{\partial}{\partial t_2}. \tag{53}$$

By means of Eqs. (47), (48), and (53), the three operators of SO(2,1) given in Eqs. (12) and (13) can be reformulated by writing

$$\hat{\Gamma}_1=\frac{\partial}{\partial\hat{x}}\left(\hat{x}\frac{\partial}{\partial\hat{x}}\right)+\frac{1}{4\hat{x}}\frac{\partial^2}{\partial t_2^2}\pm\frac{\hat{x}}{4},$$
$$\hat{\Gamma}_2=-i\left(\hat{x}\frac{\partial}{\partial\hat{x}}+\tfrac{1}{2}\right). \tag{54}$$

With the aid of Eqs. (53) and (54), another four operators can be introduced

$$\hat{\Gamma}_\pm=-\hat{\Gamma}_1\pm i\hat{\Gamma}_2,$$
$$\tilde{\Gamma}_+=(\hat{\Gamma}_+\hat{\Gamma}_-)^{-1/2}\hat{\Gamma}_+, \tag{55}$$
$$\tilde{\Gamma}_-=(\hat{\Gamma}_-\hat{\Gamma}_+)^{-1/2}\hat{\Gamma}_-.$$

Now let us define two transformations in terms of $\hat{v}_0$, $\hat{v}$, $\hat{\Gamma}_\pm$, $\hat{x}$, and $\hat{\lambda}$ such that

$$D_2=e^{+i2\hat{\lambda}t_2}\hat{x}^{\hat{v}_0}e^{\hat{x}/2}\tilde{\Gamma}_-^{\hat{v}}\,[\Gamma(-2\hat{v}_0+1)]^{1/2}, \tag{56}$$
$$D_2^{-1}=[\Gamma(-2\hat{v}_0+1)]^{-1/2}\tilde{\Gamma}_+^{\hat{v}}e^{-\hat{x}/2}\hat{x}^{-\hat{v}_0}e^{-i2\hat{\lambda}t_2}, \tag{57}$$

with

$$\hat{v}=\hat{v}_0+\hat{\lambda}-\tfrac{1}{2},\quad D_2D_2^{-1}=D_2^{-1}D_2=1. \tag{58}$$

Application of the operator of the gamma function $\Gamma(-2\hat{v}_0+1)$ in Eqs. (56) and (57) to f_{vlm} gives

$$\Gamma(-2\hat{v}_0+1)f_{vlm}=\Gamma(b+1)f_{vlm}. \tag{59}$$

Consequently we have, by use of Eqs. (21), (22), and (25),

$$D_2f_{vlm}=\eta_{vlm}(\theta,\phi,t_1,t_2)$$
$$=\overline{Y}_{lm}(\theta,\phi,t_1)e^{i(2v+1)t_2}, \tag{60}$$
$$D_2^{-1}\eta_{vlm}=f_{vlm}. \tag{61}$$

From the expression of η_{vlm} in Eq. (60), we know immediately that the transformation behavior of η_{vlm} under the action of the operators $\{J_{\alpha\beta}\}$ of SO(3,2) given in Eq. (41) is just the same as that of $\overline{Y}_{lm}$, so the relations given in Eq. (40) hold true for both $\overline{Y}_{lm}$ and η_{vlm}. Furthermore, application of D_2^{-1} to $\overline{J}_{\alpha\beta}\eta_{vlm}$ gives

$$D_2^{-1}J_{\alpha\beta}\eta_{vlm} = \overline{J}_{\alpha\beta}f_{vlm}, \tag{62}$$

where

$$\overline{J}_{\alpha\beta} = D_2^{-1}J_{\alpha\beta}D_2. \tag{63}$$

Equations (62) and (63) show that the relations in Eq. (40) also hold good for the newly defined operators $\{J_{\alpha\beta}\}$ of SO(3,2) with respect to $\{f_{vlm}\}$. By consecutive transformation of $\overline{J}_{\alpha\beta}f_{vlm}$ under D_1^{-1}, we get

$$D_1^{-1}\overline{J}_{\alpha\beta}f_{vlm} = \bar{\bar{J}}_{\alpha\beta}\psi_{vlm}, \tag{64}$$

where

$$\bar{\bar{J}}_{\alpha\beta} = D_1^{-1}\overline{J}_{\alpha\beta}D_1. \tag{65}$$

Clearly, the operators $\{\bar{\bar{J}}_{\alpha\beta}\}$ of SO(3,2) and $\{\psi_{vlm}\}$ satisfy the following relations:

$$\begin{aligned}
&\bar{\bar{J}}_{12}\psi_{vlm} = M_3\psi_{vlm} = m\psi_{vlm},\\
&(\bar{\bar{J}}_{23} \pm i\bar{\bar{J}}_{13})\psi_{vlm} = M_{\pm}\psi_{vlm}\\
&\qquad = [(l \mp m)(l \pm m + 1)]^{1/2}\psi_{vl,m\pm1},\\
&\bar{\bar{J}}_{45}\psi_{vlm} = \hat{L}_0\psi_{vlm} = (l + \tfrac{1}{2})\psi_{vlm},\\
&(-\bar{\bar{J}}_{34} + i\bar{\bar{J}}_{35})\psi_{vlm} = [(l + m + 1)(l - m + 1)]^{1/2}\psi_{v,l+1,m},\\
&(-\bar{\bar{J}}_{34} - i\bar{\bar{J}}_{35})\psi_{vlm} = [(l + m)(l - m)]^{1/2}\psi_{v,l-1,m}.
\end{aligned} \tag{66}$$

4. GROUP H_4 AND THE QUANTUM NUMBER v

Now let us deal with the ladder operators with respect to the quantum number v. For brevity, we only state the results without proof, but the details are given in the Appendix.

Let us define three operators as follows:

$$\tilde{v}_0 = \hat{v}_0 + \hat{\lambda}, \tag{67}$$

$$\tilde{v}_+ = (B_-B_+)^{-1/2}\hat{v}_+, \tag{68}$$

$$\tilde{v}_- = (B_+B_-)^{-1/2}\hat{v}_-, \tag{69}$$

where

$$B_{\pm} = e^{\mp it_2}\left[\mp(\hat{x})^{1/2}\frac{\partial}{\partial\hat{x}} + \frac{(\hat{x})^{1/2}}{2} + \frac{i}{2(\hat{x})^{1/2}}\frac{\partial}{\partial t_2}\right], \tag{70}$$

$$\begin{aligned}
\hat{v}_{\pm} &= e^{\pm 2it_2}\\
&\times\left[\hat{x}\frac{\partial^2}{\partial\hat{x}^2} \pm i\frac{\partial^2}{\partial\hat{x}\partial t_2} \mp \frac{1}{4\hat{x}}\left(2i\frac{\partial}{\partial t_2} \pm \frac{\partial^2}{\partial t_2^2}\right) - \frac{\hat{x}}{4}\right].
\end{aligned} \tag{71}$$

Under the transformation of D_1, the three operators $\tilde{v}_0$, $\tilde{v}_{\pm}$ can be turned into the forms

$$\underline{v}_0 = D_1^{-1}\tilde{v}_0 D_1, \tag{72}$$

$$\underline{v}_{\pm} = D_1^{-1}\tilde{v}_{\pm}D_1. \tag{73}$$

The identity operator together with the three operators are subjected to the commutation relations of the Lie algebra of group H_4,[14] which has been used to describe the one-dimensional harmonic oscillator,

$$[\underline{v}_0,\underline{v}_{\pm}] = \pm\underline{v}_{\pm}, \quad [\underline{v}_+,\underline{v}_-] = -1. \tag{74}$$

The three operators $\underline{v}_0$, $\underline{v}_+$, and $\underline{v}_-$ are just the ladder operators with respect to the quantum number v, i.e.,

$$\begin{aligned}
&\underline{v}_0\psi_{vlm} = (v + \tfrac{1}{2})\psi_{vlm},\\
&\underline{v}_+\psi_{vlm} = (v + 1)^{1/2}\psi_{v+1,lm}, \quad \underline{v}_-\psi_{vlm} = (v)^{1/2}\psi_{v-1,lm}.
\end{aligned} \tag{75}$$

From Eqs. (66) and (75), it is not difficult to find that the algebras of SO(3,2) and H_4 are commutative. Thus we conclude that the dynamical symmetry of our problem is described by the direct product of H_4 and SO(3,2), i.e.,

$$H_4 \times \mathrm{SO}(3,2).$$

The radial wave equation given in Eq. (8) is left invariant only under the action of the operators $\bar{\bar{J}}_{23} = M_1$, $-\bar{\bar{J}}_{13} = M_2$, $\bar{\bar{J}}_{12} = M_3$ of $\mathrm{SO}(3)\subset\mathrm{SO}(3,2)$, the operator $\bar{\bar{J}}_{45}$ of SO(3,2), and the operator $\underline{v}_0$ of H_4.

By means of the operator $\hat{V}$ given in Eq. (50), Eq. (8) can be transformed into the form

$$r^{-1}\hat{V}^2 rR(r) = b^2R(r). \tag{76}$$

Substituting Eqs. (51), (53), and (67) into Eq. (72), the operator $\hat{V}^2$ can be expressed in terms of $\underline{v}_0$ of H_4.

$$r^{-1}\hat{V}^2 r = 4(\underline{v}_0 - \hat{\lambda})^2. \tag{77}$$

From Eqs. (75)–(77), we immediately get

$$b^2 = 4(v + \tfrac{1}{2} - \lambda)^2. \tag{78}$$

Substitution of Eq. (9) into Eq. (78) again gives the result of eigenvalues given in Eq. (18). So the group H_4 can be employed as a spectrum generating group of our problem as well.

APPENDIX

In this appendix, we shall deal with the ladder operators with respect to the quantum number v. Let us define eight operators as follows:

$$\begin{aligned}
&B = i\frac{\partial}{\partial t},\\
&B_{\pm} = e^{\mp it}\left(\mp(x)^{1/2}\frac{\partial}{\partial x} + \frac{(x)^{1/2}}{2} + \frac{i}{2(x)^{1/2}}\frac{\partial}{\partial t}\right),\\
&B_{\pm\mp} = e^{\mp it}\left(\mp(x)^{1/2}\frac{\partial}{\partial x} - \frac{(x)^{1/2}}{2} + \frac{i}{2(x)^{1/2}}\frac{\partial}{\partial t}\right),\\
&\hat{\Gamma}_{\frac{1}{3}} = \frac{\partial}{\partial x}\left(x\frac{\partial}{\partial x}\right) + \frac{1}{4x}\frac{\partial^2}{\partial t^2} \pm \tfrac{1}{4},\\
&\hat{\Gamma}_2 = -i\left(x\frac{\partial}{\partial x} + \tfrac{1}{2}\right)
\end{aligned} \tag{A1}$$

which act on $\{\xi_{b,v}\}$ with

$$\xi_{b,v} = w_v^b(x)e^{-ibt} = w_v^{2\lambda-2v-1}(x)e^{i(2v+1-2\lambda)t}. \tag{A2}$$

The corresponding commutation relations are obtained by writing

$$\begin{aligned}
&[B,B_{\pm}] = \pm B_{\pm}, \quad [B,B_{\pm\mp}] = \pm B_{\pm\mp},\\
&[B_+,B_-] = -1, \quad [B_{+-},B_{-+}] = 1,\\
&[\hat{\Gamma}_3,B_{\pm}] = \mp\tfrac{1}{2}B_{\pm}, \quad [\hat{\Gamma}_3,B_{\pm\mp}] = \pm\tfrac{1}{2}B_{\pm\mp},
\end{aligned}$$

$$[\hat{\Gamma}_2,B_\pm]=\frac{i}{2}B_{\pm\mp},\quad [\hat{\Gamma}_2,B_{\pm\mp}]=\frac{i}{2}B_\pm,$$
$$[\hat{\Gamma}_1,B_\pm]=\mp\tfrac{1}{2}B_{\pm\mp},\quad [\hat{\Gamma}_1,B_{\pm\mp}]=\pm\tfrac{1}{2}B_\pm,$$
$$[\hat{\Gamma}_1,\hat{\Gamma}_2]=-i\hat{\Gamma}_3,\quad [\hat{\Gamma}_2,\hat{\Gamma}_3]=i\hat{\Gamma}_1,\quad [\hat{\Gamma}_3,\hat{\Gamma}_1]=i\hat{\Gamma}_2, \tag{A3}$$

and all the others are zero. With the aids of

$$\begin{aligned}
B_+B_- &= -\Gamma_3-\tfrac{1}{2}+\tfrac{1}{2}B,\\
B_-B_+ &= -\Gamma_3+\tfrac{1}{2}+\tfrac{1}{2}B,\\
B_{+-}B_{-+} &= -\Gamma_3+\tfrac{1}{2}-\tfrac{1}{2}B,\\
B_{-+}B_{+-} &= -\Gamma_3-\tfrac{1}{2}-\tfrac{1}{2}B,
\end{aligned} \tag{A4}$$

and

$$B\xi_{b,v}=b\xi_{b,v},\quad \Gamma_3\xi_{b,v}=-\left(v+\frac{b+1}{2}\right)\xi_{b,v}, \tag{A5}$$

it is not difficult to show that

$$\begin{aligned}
B_+\xi_{b,v} &= (v+b+1)^{1/2}\xi_{b+1,v},\\
B_-\xi_{b,v} &= (v+b)^{1/2}\xi_{b-1,v},\\
B_{+-}\xi_{b,v} &= (v)^{1/2}\xi_{b+1,v-1},\\
B_{-+}\xi_{b,v} &= (v+1)^{1/2}\xi_{b-1,v+1},\\
\hat{\Gamma}_+\xi_{b,v} &= [(v+1)(v+b+1)]^{1/2}\xi_{b,v+1},\\
\hat{\Gamma}_-\xi_{b,v} &= [v(v+b)]^{1/2}\xi_{b,v-1},
\end{aligned} \tag{A6}$$

where $\hat{\Gamma}_\pm=-\hat{\Gamma}_1\pm i\hat{\Gamma}_2$.

Let us define another three operators

$$\begin{aligned}
\tilde{v}_0 &= \hat{v}_0+\lambda,\\
\tilde{v}_+ &= (B_-B_+)^{-1/2}\hat{v}_+,\\
\tilde{v}_- &= (B_+B_-)^{-1/2}\hat{v}_-,
\end{aligned} \tag{A7}$$

where

$$\begin{aligned}
\hat{v}_0 &= -\tfrac{1}{2}B,\\
\hat{v}_\pm &= B_\mp B_{\mp\pm}\\
&= e^{\pm 2it}\left[x\frac{\partial^2}{\partial x^2}\pm i\frac{\partial^2}{\partial x\partial t}\right.\\
&\quad \left.\mp\frac{1}{4x}\left(2i\frac{\partial}{\partial t}\pm\frac{\partial^2}{\partial t^2}\right)-\frac{x}{4}\right].
\end{aligned} \tag{A8}$$

Taking into account Eq. (A3), we obtain the commutation relations of the group H_4,

$$[\tilde{v}_0,\tilde{v}_\pm]=\pm\tilde{v}_\pm,\quad [\tilde{v}_+,\tilde{v}_-]=-1. \tag{A9}$$

By means of Eq. (A6) we get

$$\begin{aligned}
\tilde{v}_0\xi_{2\lambda-2v-1,v} &= (v+\tfrac{1}{2})\xi_{2\lambda-2v-1},\\
\tilde{v}_+\xi_{2\lambda-2v-1,v} &= (v+1)^{1/2}\xi_{2\lambda-2(v+1)-1,v+1},\\
\tilde{v}_-\xi_{2\lambda-2v-1,v} &= (v)^{1/2}\xi_{2\lambda-2(v-1)-1,v-1}.
\end{aligned} \tag{A10}$$

[1] A. O. Barut and A. Bohm, Phys. Rev. B **139**, 1107 (1965).
[2] A. O. Barut and H. Kleinert, Phys. Rev. **156**, 1541 (1967); **157**, 1180 (1967).
[3] Y. Dothan, M. Gell-Mann, and Y. Ne'eman, Phys. Rev. Lett. **17**, 145 (1965).
[4] Y. Dothan, Phys. Rev. D **2**, 2944 (1970).
[5] N. Mukuda, L. Okaifeartaigh, and E. C. G. Sudanghan, Phys. Rev. Lett. **15**, 1045 (1965).
[6] R. L. Anderson, S. Kumei, and C. E. Wulfman, Phys. Rev. Lett. **28**, 988 (1972).
[7] R. L. Anderson, S. Kumei, and C. E. Wulfman, J. Math. Phys. **14**, 1527 (1973).
[8] J. Lanik, Nucl. Phys. B **2**, 263 (1967).
[9] B. G. Wybourne, *Classical Groups for Physicists* (Wiley, New York, 1974).
[10] P. A. Morse, Phys. Rev. **34**, 57 (1929).
[11] P. Cordero and G. C. Ghirardi, Nuovo Cimento A **2**, 217 (1971); P. Cordero and S. Hojman, Nuovo Cimento A **3**, 807 (1971).
[12] C. L. Pekeris, Phys. Rev. **45**, 98 (1929).
[13] E. Kyriakpoulos, Phys. Rev. **174**, 1846 (1968).
[14] R. Gilmore, *Lie Groups, Lie Algebras, and Some of Their Applications* (Wiley, New York, 1974).

Schrödinger spectral kernels: High order asymptotic expansions

T. A. Osborn
Department of Physics, University of Manitoba, Winnipeg, Manitoba R3T 2N2, Canada

R. Wong[a)]
Department of Mathematics, University of British Columbia, Vancouver, British Columbia V6T 1Y4, Canada

(Received 27 September 1982; accepted for publication 30 December 1982)

The large energy behavior of the spectral kernel for the N-body Schrödinger Hamiltonian is obtained. In a setting of a d-dimensional Euclidean space without boundaries, the Schrödinger Hamiltonian H is the sum of the negative Laplacian plus a real-valued local potential $v(x)$. The class of potentials studied is the family of bounded and continuous functions that are formed from the Fourier transforms of complex bounded measures. These potentials are suitable for the N-body problem, since they do not necessarily decrease as $|x|\to\infty$. Let $\{e(x,y;\lambda): \lambda\in\mathbb{R}\}$ be the family of spectral kernels generated by H. In the $\lambda\to\infty$ limit, explicit higher order asymptotic expansions are obtained for $e(x,y;\lambda)$ and its associated Riesz means. The asymptotic expansion is uniform in x and y and is accompanied by estimates of the error term.

PACS numbers: 03.65.Nk, 02.30. + g

I. INTRODUCTION

Let H be a Schrödinger Hamiltonian suitable for describing the N-body problem in nonrelativistic quantum mechanics. In the setting of a d-dimensional Euclidean space without boundaries, H is the self-adjoint operator composed of the sum of the negative Laplacian and a continuous bounded real-valued local potential $v(x)$, $x\in\mathbb{R}^d$. Let $\{E_\lambda: \lambda\in\mathbb{R}\}$ be the spectral family of H and $e(x,y;\lambda): \mathbb{R}^d\times\mathbb{R}^d\to\mathbb{C}$ be the associated spectral kernels. The objective of this paper is to derive higher-order asymptotic expansions of $e(x,y;\lambda)$ for large values of λ which are uniformly valid for x,y in any compact region within $\mathbb{R}^d$.

Suppose, for the moment, that H has a pure point spectrum. Take $\{\lambda_i\}$ be the nondecreasing sequence of eigenvalues of H that tend to $+\infty$ as $i\to\infty$. That is

$$H\psi_i = \lambda_i\psi_i, \tag{1.1}$$

where ψ_i is an L^2-eigenfunction solution of (1.1). The problem we investigate is a generalization of the study of the asymptotic density of λ_i as $i\to\infty$. The Schrödinger problem in an unbounded region typically has both point spectra and continuous spectra. In the presence of continuous spectra it is necessary to replace the density of λ_i by the values of $e(x,y;\lambda)$ as $\lambda\to\infty$. The asymptotic expansion of the spectral kernels $e(x,y;\lambda)$ for large classes of elliptic and parabolic differential equations has been the subject of extensive study.[1-10] However, this literature gives only the first term of the asymptotic expansion of $e(x,y;\lambda)$ for large values of λ. For the Schrödinger Hamiltonian we develop a general method for obtaining the nth-order asymptotic expansion of $e(x,y;\lambda)$. In essence our method for the nth-order expansion requires that the Hamiltonian H have two basic features. The first requirement is that $v(x)$ have a finite number of continuous bounded partial derivatives. Secondly, we require that the time evolution kernel for the continuous part of the spectrum decay as the time parameter $t\to\pm\infty$.

The mathematical problem we study is the following. Let q be the quantum scale factor

$$q = \hbar^2/2m, \tag{1.2}$$

where $\hbar$ is the rationalized value of Planck's constant and m is the mass of a spinless particle. If Δ_x is the Laplacian in $\mathbb{R}^d$, then take H_0 to be the self-adjoint extension of $-q\Delta_x$ which acts on the Hilbert space $\mathscr{H} = L^2(\mathbb{R}^d)$. The perturbation V is the bounded operator on $L^2(\mathbb{R}^d)$ defined by

$$(Vf)(x) = v(x)f(x), \quad f\in L^2(\mathbb{R}^d), \tag{1.3}$$

where $v(x)$ is initially taken as a continuous bounded real-valued function for all $x\in\mathbb{R}^d$. The Hamiltonian H is defined as

$$H = H_0 + V. \tag{1.4}$$

Because V is bounded ($\|V\| < \infty$), H and H_0 share a common domain $\mathscr{D}$. The spectrum Λ of H is bounded from below by $-\|V\|$.

The basic element in our analysis of the spectral kernel $e(x,y;\lambda)$ is the detailed kernel description of the analytic semigroup family generated by H. The semigroup of bounded operators $\{e^{-zH}: z\in\mathbb{C}, \operatorname{Re} z > 0\}$ is characterized uniquely by its associated family of kernels $U_z(x,y)$. Since the kernel $U_z(x,y)$ has an essential singularity at $z = 0$, it is useful to introduce the factorization

$$U_z(x,y) = U_z^{(0)}(x-y)F(x,y;z), \tag{1.5}$$

where $U_z^{(0)}(x-y)$ is the convolution kernel corresponding to the operator e^{-zH_0}. The function $F(x,y;z)$ is analytic in $D = \{z\in\mathbb{C}, \operatorname{Re} z > 0\}$, continuous in $\overline{D}$ and $F(x,y;0) = 1$ for all $x,y\in\mathbb{R}^d$. Previous studies of $F(x,y;z)$ show[11-15] that it has an asymptotic series representation in the form

$$F(x,y;z) = \sum_{n=0}^{M-1}\frac{(-z)^n}{n!}P_n(x,y) + E_M(x,y;z). \tag{1.6}$$

The coefficient functions $P_n(x,y)$ are known explicitly and the error term, $E_M(x,y;z)$, has a uniform bound in x,y of order $O(|z|^M)$ for all $z\in\overline{D}$.[11] The number of terms M in expansion (1.6) is proportional to the number of bounded derivatives the potential $v(x)$ possesses.

[a)] I. W. Killam Research Fellow. On leave from the University of Manitoba.

 0022-2488/83/061487-15$02.50 © 1983 American Institute of Physics

The relationship of the asymptotic formula (1.6) to the large-λ behavior of $e(x,y;\lambda)$ is a consequence of the inverse Laplace transform

$$e(x,y;\lambda) = \frac{1}{2\pi i}\int_{\beta-i\infty}^{\beta+i\infty} \frac{e^{z\lambda}}{z} U_z(x,y)\,dz. \tag{1.7}$$

We shall prove that (1.7) is valid for all $\beta > 0$. Introducing the small-z expansion of $U_z(x,y)$ eventually leads us to the higher-order asymptotic expansion of the spectral kernel, namely

$$e(x,y;\lambda) = \sum_{n=0}^{M-1} \frac{(-1)^n}{n!} P_n(x,y)\left(\frac{d}{d\lambda}\right)^n e_0(x,y;\lambda) + R_M(x,y;z). \tag{1.8}$$

In this formula $e_0(x,y;\lambda)$ denotes the spectral kernel for the Hamiltonian H_0, i.e.,

$$e_0(x,y;\lambda) = \left(\frac{\lambda^{1/2}}{2\pi q^{1/2}|x-y|}\right)^{d/2} J_{d/2}(\lambda^{1/2}q^{-1/2}|x-y|), \tag{1.9}$$

where $J_{d/2}$ is the Bessel function of the first kind. We derive estimates of $R_M(x,y;z)$ that depend on the smoothness of $v(x)$ as well as the nature of the $|t|\to\infty$ decay of the time evolution kernel.

The study of H and its spectral kernel $e(x,y;\lambda)$ is carried out for the class of potentials $v(x)$ that are represented by Fourier transforms of bounded complex measures supported on $\mathbb{R}^d$. Denote by $\mathscr{M}(\mathbb{R}^d)$ the set of all bounded complex measures defined on the Borel field $\mathscr{B}$ on $\mathbb{R}^d$. Each measure $\mu\in\mathscr{M}(\mathbb{R}^d)$ defines a potential function by a Fourier transform

$$v(x) = \int_{\mathbb{R}^d} e^{i\alpha x}\,d\mu(\alpha), \tag{1.10}$$

where αx indicates the scalar product of two vectors in $\mathbb{R}^d$. The potentials $v(x)$ are bounded and continuous. If $|\mu|(e)$ is the total variation of $\mu(e)$, we define a norm of μ in $\mathscr{M}(\mathbb{R}^d)$ by

$$\|\mu\| = \int_{\mathbb{R}^d} d\,|\mu(e)| = |\mu|(\mathbb{R}^d). \tag{1.11}$$

Clearly the operator norm of V in Eq. (1.4) satisfies the inequality

$$\|V\| \leqslant \|\mu\|. \tag{1.12}$$

The reality condition on $v(x)$ is satisfied if the measure μ obeys the reflection property $\mu(-e) = \overline{\mu(e)}$ for all $e\in\mathscr{B}$, where $-e = \{x\in\mathbb{R}^d: -x\in e\}$ and the bar denotes complex conjugation. We indicate by $\mathscr{M}^{\mathrm{r}}(\mathbb{R}^d)$ the space of measures $\mu\in\mathscr{M}(\mathbb{R}^d)$ that satisfy the reflection property. $\mathscr{F}^{\mathrm{r}}$ will represent the Fourier image of $\mathscr{M}^{\mathrm{r}}(\mathbb{R}^d)$. The elements of $\mathscr{F}^{\mathrm{r}}$ and $\mathscr{M}^{\mathrm{r}}(\mathbb{R}^d)$ are in a one-to-one correspondence because of the uniqueness of the transform (1.10) which states $v(x) = 0$ if and only if $\mu = 0$.

In order to control the partial derivatives of $v(x)$, we define for a positive integer M, the set of functions

$$\mathscr{F}^{\mathrm{r}}_M = \left\{v(x)\in\mathscr{F}^{\mathrm{r}}: \int_{\mathbb{R}^d} |\alpha|^n\,d\,|\mu|(\alpha) < \infty,\ \ n = 0,1,\ldots,M\right\}, \tag{1.13}$$

where μ is the measure associated with $v(x)$. In fact, if $v(x)\in\mathscr{F}^{\mathrm{r}}_M$, then there exists a smallest finite positive constant K (depending on μ and M) such that

$$\int_{\mathbb{R}^d} |\alpha|^n\,d\,|\mu|(\alpha) \leqslant K^n\|\mu\|, \quad \text{for } n = 0,1,\ldots,M. \tag{1.14}$$

We call K the *bound constant of measure μ in the family* $\mathscr{F}^{\mathrm{r}}_M$. If $v(x)\in\mathscr{F}^{\mathrm{r}}_M$ and D_x^L represents an arbitrary partial derivative in $\mathbb{R}^d$ multi-indexed by $L = (l_1,l_2,\ldots,l_d)$ with length $|L| \equiv l_1 + \cdots + l_d$ $(l_i \geqslant 0)$, then

$$|(D_x^L v)(x)| \leqslant K^{|L|}\|\mu\| \quad \text{for all } |L| \leqslant M. \tag{1.15}$$

The requirement of smoothness for the potential, $v(x)$, is the only restriction on the Hamiltonian H needed to prove expansion (1.6). If $v(x)\in\mathscr{F}^{\mathrm{r}}_{2M}$, then (1.6) is valid. Our analysis of the asymptotic expansion of the spectral kernel $e(x,y;\lambda)$ will impose one additional limitation on H. In order to ensure the existence of certain integrals, we need a growth estimate as $|t|\to\infty$ of the time-evolution kernel. In terms of the eigenfunction solutions of Eq. (1.1), the continuous spectrum part of the time evolution may be defined by the kernel

$$U^{\mathrm{c}}_{it}(x,y) = U_{it}(x,y) - \sum_{\lambda_i\in\sigma_{\mathrm{p}}(H)} e^{-i\lambda_i t}\psi_i(x)\,\overline{\psi_i(y)}, \tag{1.16}$$

where $\sigma_{\mathrm{p}}(H)$ is the point spectrum of Hamiltonian H. Let

$$E_\lambda = E^{\mathrm{p}}_\lambda + E^{\mathrm{c}}_\lambda \tag{1.17}$$

be the orthogonal decomposition of the spectral projectors E_λ into the projectors onto the point spectrum and the continuous spectrum subspaces of $\mathscr{H}$. If $H_{\mathrm{c}} \equiv E^{\mathrm{c}}_\lambda H$, then $U^{\mathrm{c}}_{it}(x,y)$ is the kernel of the operator $e^{-itH_{\mathrm{c}}}$. The type of growth estimate in t our analysis requires is

$$|U^{\mathrm{c}}_{it}(x,y)| < C\,|t|^\gamma \quad \text{as } |t|\to\infty. \tag{1.18}$$

To obtain an expansion of the form (1.8) requires $\gamma < 0$.

There is not an extensive mathematical literature on the large-t behavior of the kernel $U^{\mathrm{c}}_{it}(x,y)$. For the Schrödinger problem with decreasing potentials $v(x)$ as $|x|\to\infty$ there are a number of related results[16–23] that describe the large t behavior of the operator $e^{-itH_{\mathrm{c}}}$ in various different operator norms. However, these norm estimates do not immediately generalize to provide pointwise estimates like (1.18). Furthermore, the literature[16–23] that describes the t decay of $e^{-itH_{\mathrm{c}}}$ systematically assumes that $v(x)$ is decreasing for large $|x|$, and thus the large t estimates are not applicable to the N-body Schrödinger problem with N greater than 2.

We proceed by identifying a wide class of large t behavior. For each type of growth estimate in this class we will find the corresponding asymptotic expansion of $e(x,y;\lambda)$.

Let n be a nonnegative integer, $\gamma\in\mathbb{R}$, and C_u a finite constant independent of x and y. For $H = H_0 + V$, where $v(x)\in\mathscr{F}^{\mathrm{r}}$, we say H *belongs to class* $T_{\mathrm{u}}(\gamma;n)$ if as $|t|\to\infty$

$$\left|\left(\frac{d}{dt}\right)^j U^{\mathrm{c}}_{it}(x,y)\right| < C_{\mathrm{u}}\,|t|^\gamma \tag{1.19}$$

for $j = 0,1,\ldots,n$ and all $(x,y)\in\mathbb{R}^d\times\mathbb{R}^d$. The subscript u on $T_{\mathrm{u}}(\gamma;n)$ indicates the x,y uniform nature of the growth estimate (1.19). A related class of Hamiltonians is denoted by $T_{\mathrm{p}}(\gamma;n)$. We say H *belongs to class* $T_{\mathrm{p}}(\gamma;n)$ if

$$\int_{|t|>1} dt\,\frac{1}{|t|^{1+\gamma}} \sum_{j=0}^{n} \left|\left(\frac{d}{dt}\right)^j U^{\mathrm{c}}_{it}(x,y)\right| < \infty \tag{1.20}$$

for each $(x,y)\in\mathbb{R}^d\times\mathbb{R}^d$ and for $j=0,1,\ldots,n$.

The anticipated generic behavior of $U^c_{it}(x,y)$ for large t is the growth associated with the appropriate unperturbed time evolution kernels. For example, if $d=3$ and potential $v(x)$ is decreasing for large x, then one expects that $\gamma=-\frac{3}{2}$ in inequality (1.18). This generic decay is modified, as Jensen and Kato have shown,[21] if H has zero energy eigenfunctions or resonances, and as a consequence $\gamma=-\frac{1}{2}$ in (1.18). Our derivation of the asymptotic spectral expansion requires only that $\gamma<0$. For the N-body problem ($N>2$) that is composed of particles moving in three dimensions, then the large-t behavior should be controlled by the zero velocity separation of all possible two-cluster fragments. Since these clusters separate from each other in three dimensions, the associated decay of the N-body continuous spectrum time evolution kernel is again expected to be $\gamma=-\frac{3}{2}$. In the case where $U^c_{it}(x,y)$ has growth at $t=\pm\infty$ rather than decay, a higher-order asymptotic expansion related to the spectral kernel still exists. If γ is positive in inequality (1.18), then by replacing both $e(x,y;\lambda)$ and $e_0(x,y;\lambda)$ with their respective Riesz means a modified form of expansion (1.8) still holds.

The plan of this paper is as follows. In Sec. II the inverse Laplace transform relationship between the kernels $U_z(x,y)$ and $e(x,y;\lambda)$ is derived. In Sec. III we obtain a few of the basic properties of the kernel associated with the point spectrum projector E^p_λ. Section IV contains a large-parameter asymptotic expansion theorem for the type of oscillatory integral that appears in the transform connecting $U_z(x,y)$ to $e(x,y;\lambda)$. In Sec. V the results of the previous sections are combined with the decay classification schemes $T_u(\gamma,n)$ and $T_p(\gamma;n)$ to derive the spectral kernel expansion (1.8).

II. THE ANALYTIC SEMIGROUP CONSTRUCTION OF THE SPECTRAL KERNEL

The operator e^{-zH} is defined by the spectral integral

$$e^{-zH}=\int_\Lambda e^{-z\lambda}\,dE_\lambda,\quad z\in D, \tag{2.1}$$

where D is the right-half complex plane $\{z: z\in\mathbb{C}, \operatorname{Re} z>0\}$. The set of bounded operators on Hilbert space $\mathscr{H}=L^2(\mathbb{R}^d)$, $\{e^{-zH}: z\in D\}$, composes the analytic semigroup. This section explores the relationships between the kernels of the family $\{e^{-zH}\}$ and the kernels of the family of spectral projectors $\{E_\lambda:\lambda\in\mathbb{R}\}$. In essence we show how an appropriate contour integration of $U_z(x,y)$ in the domain D may be used to construct $e(x,y;\lambda)$.

We begin by summarizing the known properties of the kernels $U_z(x,y)$ and $e(x,y;\lambda)$. Our results always assume $v(x)\in\mathscr{F}^r$. It is well known[24] for the negative Laplacian H_0 that the kernel of the operator e^{-zH_0} is given by the function $U^0_z(x-y)$, where

$$U^0_z(x-y)\equiv(4\pi zq)^{-d/2}\exp[-|x-y|^2/4zq]. \tag{2.2}$$

Here $|x-y|$ denotes the metric length of the vector $x-y\in\mathbb{R}^d$. The following proposition describes the structure of the function $U_z(x,y)$, and the subsequent theorem indicates in what sense $U_z(x,y)$ is to be interpreted as the kernel of the operator e^{-zH} ($z\in\overline{D}\setminus\{0\}$).

Proposition 1: Let $v\in\mathscr{F}^r$. There exists a unique function $U_z(x,y)$: $\overline{D}\setminus\{0\}\times\mathbb{R}^d\times\mathbb{R}^d\to\mathbb{C}$ with the following properties:

(1) $U_z(x,y)=U^0_z(x-y)F(x,y;z)$. (2.3)

(2) The function $F(x,y;z)$: $\mathbb{R}^d\times\mathbb{R}^d\times\overline{D}\to\mathbb{C}$ is holomorphic in D and continuous in $\overline{D}$; furthermore, it is jointly continuous in $\mathbb{R}^d\times\mathbb{R}^d$. The function $U_z(\cdot,y)\in L^2(\mathbb{R}^d)$ for each $y\in\mathbb{R}^d$ and $z\in D$.

(3) $F(x,y;z)$ has the bound

$$|F(x,y,z)|\leqslant e^{|z|\,\|\mu\|}\quad\text{for } z\in\overline{D}. \tag{2.4}$$

Proof: This result is Proposition 2 of Ref. 11. It is obtained from the iterative series expression for the fundamental solution of the heat equation for operator H. □

Theorem 1: Let $v\in\mathscr{F}^r$.

(1) For all $z\in D$ and all $u(x)\in L^2(\mathbb{R}^d)$,

$$(e^{-zH}u)(x)=\int dy\; U_z(x,y)u(y). \tag{2.5}$$

(2) Suppose $t\neq0$. If $u(x)\in L^1(\mathbb{R}^d)\cap L^2(\mathbb{R}^d)$, then

$$(e^{-itH}u)(x)=\int dy\; U_{it}(x,y)u(y), \tag{2.6}$$

and, if $u(x)\in L^2(\mathbb{R}^d)$, then

$$(e^{-itH}u)(x)=\operatorname*{s-lim}_{B\to\infty}\int_{|y|<B} dy\; U_{it}(x,y)u(y). \tag{2.7}$$

Relations (2.5) and (2.6) are valid for a. a. $x\in\mathbb{R}^d$.

Proof: Reference 11 proves this result. A similar result is quoted by Simon[25] in the more restricted case where $v(x)$ is a C^∞ potential. See also Ref. 26–28. □

Theorem 2: Let $v\in\mathscr{F}^r$. Each member E_λ of the family of spectral projectors generated by H, is an integral operator on $L^2(\mathbb{R}^d)$. For each $\lambda\in\mathbb{R}$, E_λ has a kernel $e(x,y;\lambda)$: $\mathbb{R}^d\times\mathbb{R}^d\to\mathbb{C}$ with the following properties:

(1) For all $f\in L^2(\mathbb{R}^d)$ and a. a. $x\in\mathbb{R}^d$

$$(E_\lambda f)(x)=\int dy\; e(x,y;\lambda)f(y). \tag{2.8}$$

(2) For all $x\in\mathbb{R}^d$

$$e(x,x;\lambda)=\int dy\;|e(x,y;\lambda)|^2<\infty. \tag{2.9}$$

(3) Let s be the smallest integer larger than $d/4$. There exists a constant $K_d<\infty$ depending only on d such that for all $x\in\mathbb{R}^d$

$$0\leqslant e(x,x;\lambda)\leqslant K_d(1+\|\mu\|+\lambda)^{2s}. \tag{2.10}$$

(4) For each fixed $(x,y)\in\mathbb{R}^d\times\mathbb{R}^d$ and each finite interval $[a,b]$, the function $e(x,y;\lambda)$ is of bounded variation for $\lambda\in[a,b]$. In particular, $e(x,y;\lambda)$ defines a complex valued σ-finite measure on the measure space of Borel sets over the real line. For $\delta>0$ and $(x,y)\in\mathbb{R}^d\times\mathbb{R}^d$

$$\int_\Lambda(1+\|\mu\|+\lambda)^{-2s-\delta}d\,|e(x,y;\lambda)|<\infty. \tag{2.11}$$

Proof: We omit an explicit proof as these results are more or less well known. Using detailed estimates of the Green's functions implied by the estimates of $U_z(x,y)$ found in Ref. 11, it is straightforward, utilizing the methods of

Povzner[29] and Gårding,[3,30] to prove statements (2.8)–(2.11). □

A useful extension of Theorem 2 is a variant of a theorem proved by Gårding.[30]

Corollary 1: Let $v\in\mathscr{F}^r$. For each $z\in D$, the semigroup kernel $U_z(x,y)$ has the representation

$$U_z(x,y)=\int_\Lambda e^{-z\lambda}\,de(x,y;\lambda). \tag{2.12}$$

Proof: For a class of elliptic differential operators that include H, Gårding has proved the following. Let $\mathscr{B}_0$ be the set of complex Borel measurable functions on $\mathbb{R}$ having compact support. For $\alpha\in\mathscr{B}_0$ define the norm

$$\|\alpha\|_0=\sup_{\lambda\in\mathbb{R}}|\alpha(\lambda)|.$$

Every operator $T_\alpha: L^2(\mathbb{R}^d)\to L^2(\mathbb{R}^d)$ defined by

$$T_\alpha=\int_\Lambda \alpha(\lambda)\,dE_\lambda$$

is norm-bounded with $\|T_\alpha\|=\|\alpha\|_0$. Furthermore, each T_α is an integral operator with a kernel $T_\alpha(x,y)$: $\mathbb{R}^d\times\mathbb{R}^d\to\mathbb{C}$ given by

$$T_\alpha(x,y)=\int_\Lambda \alpha(\lambda)\,de(x,y;\lambda). \tag{2.13}$$

The kernels $T_\alpha(x,y)$ satisfy the Carleman property

$$\int dy\,|T_\alpha(x,y)|^2<\infty \quad \text{for a. a. } x\in\mathbb{R}^d.$$

We must show that Gårding's representation (2.13) extends to the operators e^{-zH} ($\mathrm{Re}\,z>0$). For $\lambda_0<\infty$, consider the operator

$$E_{\lambda_0}e^{-zH}=\int_\Lambda \chi_{[-\|\mu\|,\lambda_0)}(\lambda)e^{-z\lambda}\,dE_\lambda,$$

where $\chi_{[a,b)}$ is the characteristic function for the interval $[a,b)$. If $\mathrm{Re}\,z>0$, then $\chi_{[-\|\mu\|,\lambda_0)}(\lambda)\exp(-z\lambda)\in\mathscr{B}_0$, so representation (2.13) is applicable and yields

$$(E_{\lambda_0}e^{-zH})(x,y)=\int_\Lambda \chi_{[-\|\mu\|,\lambda_0)}(\lambda)e^{-z\lambda}\,de(x,y;\lambda). \tag{2.14}$$

Examine the limit $\lambda_0\to+\infty$. The semigroup property of e^{-zH} for $z=z_1+z_2$, and the commutivity of H and E_{λ_0} let us write

$$E_{\lambda_0}e^{-zH}=e^{-z_1H}E_{\lambda_0}e^{-z_2H},\quad z_1,z_2\in D.$$

The right-hand side is the product of three operators, each of which is an integral operator with a Carleman kernel. Thus

$$(E_{\lambda_0}e^{-zH})(x,y)=\int dy'\,U_{z_1}(x,y')\times\left[\int dy''\,e(y',y'';\lambda_0)U_{z_2}(y'',y)\right].$$

Since $U_{z_1}(x,\cdot)$ and $U_{z_2}(\cdot,y)$ are in $L^2(\mathbb{R}^d)$, the multiple integral may be interpreted as an inner product, so

$$(E_{\lambda_0}e^{-zH})(x,y)=(U_{z_1}(x,\cdot),E_{\lambda_0}U_{z_2}(\cdot,y)).$$

Because $E_{\lambda_0}\to I$ strongly in $L^2(\mathbb{R}^d)$ as $\lambda_0\to\infty$, we have

$$\lim_{\lambda_0\to\infty}(E_{\lambda_0}e^{-zH})(x,y)=(U_{z_1}(x,\cdot),U_{z_2}(\cdot,y))=U_z(x,y)\quad \mathrm{Re}\,z>0, \tag{2.15}$$

where the last equality has used the semigroup property for the kernels $U_{z_1}(x,y'')$ and $U_{z_2}(y'',y)$.

Finally take the $\lambda_0\to+\infty$ limit of the integral in Eq. (2.14). Observe that $(1+\|\mu\|+\lambda)^{2s+\delta}\exp(-\mathrm{Re}\,z\lambda)$ is bounded for $\lambda\in\Lambda$. This fact when combined with statement (2.11) means that

$$\int_\Lambda e^{-\mathrm{Re}\,z\lambda}\,d|e(x,y;\lambda)|<\infty,\quad \mathrm{Re}\,z>0.$$

Thus the dominated convergence theorem applies to the integral in Eq. (2.14) to give

$$\lim_{\lambda_0\to\infty}\int_\Lambda \chi_{[-\|\mu\|,\lambda_0)}(\lambda)e^{-z\lambda}\,de(x,y;\lambda)=\int_\Lambda e^{-z\lambda}\,de(x,y;\lambda).$$

Taken together, this last equation and Eq. (2.15) yield representation (2.12). □

Equation (2.12) of Corollary 1 constructs the kernel $U_z(x,y)$ from the family of spectral kernels. We obtain next the inverse transform. The following definitions and lemma assist in the construction of this inverse.

Let $C_{\tau\epsilon}$ be the contour along the imaginary axis of the right half-plane, $\overline{D}$, that begins at $-i\tau$ and ends at $+i\tau$ and has a semicircular indentation of radius ϵ ($\tau>\epsilon>0$) about the origin. This curve lies within the domain $\overline{D}\setminus\{0\}$ for all $\tau>0$ and all $\epsilon>0$. A second contour $C'_{\tau\beta}$ denotes the straight line joining $\beta-i\tau$ and $\beta+i\tau$, $\beta>0$. For $a\in\mathbb{R}$ define

$$I(a;\tau,\epsilon)=\frac{1}{2\pi i}\int_{C_{\tau\epsilon}}(e^{za}/z)\,dz,$$

$$J(a;\tau,\beta)=\frac{1}{2\pi i}\int_{C'_{\tau\beta}}(e^{za}/z)\,dz.$$

These functions are closely related to the unit step function

$$u(t)=\begin{cases}0 & (t<0)\\ \tfrac{1}{2} & (t=0)\\ 1 & (t>0).\end{cases}$$

Subsequently, the double limit in τ and ϵ occurs frequently so we adopt the abbreviation

$$\lim_{\tau,\epsilon}=\lim_{\tau\to\infty}\lim_{\epsilon\to+0}.$$

The relevant properties of functions $I(a;\tau,\epsilon\rho)$ and $J(a;\tau,\beta)$ are:

Lemma 1: For $a<\infty$

(1) $$\lim_{\tau,\epsilon}I(a;\tau,\epsilon)=\lim_{\tau\to\infty}J(a;\tau,\beta)=u(a). \tag{2.16}$$

(2) Let $A<\infty$. If $a<A$, then for $0<\epsilon\leqslant 1$ and $\tau\geqslant 1$

$$|I(a;\tau,\epsilon)|\leqslant(2/\pi)(1.852\cdots)+\tfrac{1}{2}e^{A}. \tag{2.17}$$

(3) If $a<A$, there exists a finite constant K dependent on β and A but independent of τ such that

$$|J(a;\tau,\beta)| \leqslant e^{\beta a} K(\beta, A), \quad \beta > 0. \tag{2.18}$$

Proof: We treat integral $I(a;\tau,\epsilon)$ first. Let $T \in \mathbb{R}$ and

$$\mathrm{Si}(T) = \int_0^T \frac{\sin t}{t}\, dt.$$

Denote by C_ϵ the semicircular contour in $\overline{D}$ that is centered at the origin $z = 0$ and has radius ϵ. From the reflection symmetry of the integrand of $I(a;\tau,\epsilon)$ with respect to $z \to \bar{z}$, one finds

$$I(a;\tau,\epsilon) = \frac{1}{\pi}\left[\mathrm{Si}(\tau a) - \mathrm{Si}(\tau\epsilon)\right] + \frac{1}{2\pi i}\int_{C_\epsilon} \frac{e^{za}-1}{z}\, dz + \frac{1}{2}. \tag{2.19}$$

$\mathrm{Si}(T)$ is continuous for all T, bounded by 1.852···, and has the limiting values $\mathrm{Si}(-\infty) = -\pi/2$, $\mathrm{Si}(0) = 0$, $\mathrm{Si}(+\infty) = \pi/2$. Further note that

$$\left|\frac{1}{2\pi i}\int_{C_\epsilon} \frac{e^{za}-1}{z}\, dz\right| \leqslant \frac{1}{2}|a|\epsilon.$$

Substituting the limiting values of the Si function shows that $u(a)$ is the $\tau \to \infty$, $\epsilon \to +0$ limit of $I(a;\tau,\epsilon)$.

To establish the bound (2.17) for $I(a;\tau,\epsilon)$, we need only examine the last two terms on the right of Eq. (2.19). But these terms are estimated by

$$\left|\frac{1}{2\pi i}\int_{C_\epsilon} \frac{e^{za}-1}{z}\, dz + \frac{1}{2}\right| \leqslant \frac{1}{2} e^{\epsilon A}$$

if $a < A$. Furthermore, the Si function contributions to Eq. (2.19) are all bounded. Thus estimate (2.17) follows.

The properties of $J(a;\tau,\beta)$ follow similarly. Take $\beta > 0$. Let C_1 denote the straight line segment $(-i\tau, -i\tau + \beta)$ and C_1^* the complex conjugate segment $(i\tau, i\tau + \beta)$. The contour $-C_{\tau\epsilon} - C_1 + C'_{\tau\beta} + C_1^*$ is closed. Applying Cauchy's theorem to the function $z^{-1}e^{za}$ gives

$$-I(a;\tau,\epsilon) + J(a;\tau,\beta) + \frac{1}{2\pi i}\int_{-C_1 + C_1^*} \frac{e^{za}}{z}\, dz = 0.$$

Obviously,

$$\left|\frac{1}{2\pi i}\int_{-C_1 + C_1^*} \frac{e^{za}}{z}\, dz\right| \leqslant e^{|a|\beta}\frac{\beta}{\pi\tau}.$$

So I and J have the same limit as $\tau \to \infty$.

In order to establish (3), first factor the integral $J(a;\tau,\beta)$

$$\frac{1}{2\pi i}\int_{C'_{\tau\beta}} \frac{e^{za}}{z}\, dz = \frac{e^{\beta a}}{2\pi i}\int_{-i\tau}^{i\tau} \frac{e^{za}}{z+\beta}\, dz.$$

To the rightmost integral add and subtract the integral $I(a;\tau,\epsilon)$. Taking the absolute value gives

$$\left|\int_{-i\tau}^{i\tau} \frac{e^{za}}{z+\beta}\, dz\right| \leqslant \left|\int_{C_{\epsilon\tau}} \frac{e^{za}}{z}\, dz\right| + \left|\int_{i\epsilon}^{i\tau} + \int_{-i\tau}^{-i\epsilon} + \left(\frac{1}{z+\beta} - \frac{1}{z}\right) e^{za}\, dz\right| + \left|\int_{-i\epsilon}^{i\epsilon} \frac{e^{za}}{z+\beta}\, dz\right| + \left|\int_{C_\epsilon} \frac{e^{za}}{z}\, dz\right|, \tag{2.20}$$

where C_ϵ denotes the semicircular contour of radius ϵ about the origin. All terms on the right of Eq. (2.20) have obvious bounds; estimate (2.17) bounds the first term; the second integral exists absolutely for $\tau = \infty$, and thus has a τ-independent bound; the third factor has bound $2\epsilon\beta^{-1}$, and fourth term is bounded by $\pi e^{A\epsilon}$ if $a < A$. □

The next proposition gives us one form for the inverse transform of Eq. (2.12).

Proposition 2: Let $v \in \mathcal{F}^r$. Choose (x,y) to be fixed but arbitrary point in $\mathbb{R}^d \times \mathbb{R}^d$. Let $\{U_z(x,y): z \in D\}$ be the family of kernels associated with the analytic semigroup $\{e^{-zH}: z \in D\}$. If λ is a point of continuity of $e(x,y;\lambda)$ $[\lambda \notin \sigma_p(H)]$, then

$$e(x,y;\lambda) = \lim_{\tau\to\infty} \frac{1}{2\pi i}\int_{\beta - i\tau}^{\beta + i\tau} \frac{e^{z\lambda}}{z} U_z(x,y)\, dz, \quad \beta > 0. \tag{2.21}$$

Proof: Since $\mathrm{Re}\, z > 0$, Corollary 1 may be used to represent the kernel $U_z(x,y)$. So we must examine the iterated integral

$$L(\tau) \equiv \frac{1}{2\pi i}\int_{\beta - i\tau}^{\beta + i\tau} \frac{e^{z\lambda}}{z}\left[\int_{-\|\mu\|}^{\infty} e^{-z\lambda'}\, de(x,y;\lambda')\right] dz.$$

Estimate (2.11) plus the fact that $e^{z\lambda}/z$ is bounded on the finite segment $(\beta + i\tau, \beta - i\tau)$ establishes that integral $L(\tau)$ is absolutely convergent. Employing Fubini's theorem to justify changing the order of integration leads to

$$L(\tau) = \int_{-\|\mu\|}^{\infty} J(\lambda - \lambda';\tau,\beta)\, de(x,y;\lambda'). \tag{2.22}$$

Set $A = \lambda + \|\mu\|$ and use estimate Lemma 1, (3), to provide a τ-independent bound

$$|J(\lambda - \lambda';\tau,\beta)| \leqslant e^{\beta(\lambda - \lambda')} K(\beta, \lambda + \|\mu\|).$$

Because estimate (2.11) implies that

$$\int_{-\|\mu\|}^{\infty} e^{-\beta\lambda'}\, d\,|e(x,y;\lambda')| < \infty, \quad \beta > 0,$$

the dominated convergence theorem applies to integral (2.22). Utilizing limit (2.16) for $J(\lambda - \lambda';\tau,\beta)$ gives

$$\lim_{\tau\to\infty} \frac{1}{2\pi i}\int_{\beta - i\tau}^{\beta + i\tau} \frac{e^{z\lambda}}{z} U_z(x,y)\, dz = \int_{-\|\mu\|}^{\infty} u(\lambda - \lambda')\, de(x,y;\lambda'), \quad \beta > 0.$$

Assuming λ is a point of continuity of $e(x,y;\lambda')$ yields statement (2.21). □

Although the line integral on the right of (2.21) depends on β, the resultant value of the integral is β independent. This is a manifestation of the holomorphic character of the functions $U_z(x,y)$. Nevertheless, one cannot choose $\beta = 0$. Integral (2.12) is not defined as an absolutely convergent integral in this case since $e^{-z\lambda}$ has no decrease as $\lambda \to \infty$.

It is advantageous to introduce the Reisz mean of E_λ. In the situation where $U_{it}^c(x,y)$ has a growth estimate at $t \to \pm\infty$ which does not increase more rapidly than a polynomial in $|t|$, then the Riesz-mean average of E_λ will still permit the derivation of an altered form of the high-order expansion (1.8) for the spectral kernel. For $\alpha \geqslant 0$ the αth-order Riesz-mean operator E_λ^α is defined by

$$E_\lambda^\alpha = \int_{-\|\mu\|}^{\lambda - 0} (1 - \lambda'_+/\lambda_+)^\alpha\, dE_{\lambda'}, \tag{2.23}$$

where $\lambda_+ = \lambda + \|\mu\|$, $\lambda'_+ = \lambda' + \|\mu\|$. The variables λ_+,

λ'_+ are nonnegative for $\lambda\in\Lambda$. Clearly, if $\alpha=0$, then $E^0_\lambda = E_{\lambda-0}$. Furthermore, the integrand of (2.23) is continuous, bounded, and of compact support; thus Gårding's representation (2.13) is applicable. So E^α_λ is an integral operator on $L^2(\mathbb{R}^d)$ with an x,y-jointly continuous Carleman kernel $e^\alpha(x,y;\lambda)$, constructed by the formula

$$e^\alpha(x,y;\lambda) = \int_{-\|\mu\|}^{\lambda-0} (1-\lambda'_+/\lambda_+)^\alpha \, de(x,y;\lambda').$$

The following proposition provides the inverse formula that allows the representation of $e^\alpha(x,y;\lambda)$ in terms of the semigroup kernels $U_z(x,y)$.

Proposition 3: Assume the same conditions as given in Proposition 2. For $\alpha\geqslant 0$ and λ a point of continuity of $e^\alpha(x,y;\lambda)$ $[\lambda\notin\sigma_p(H)]$ then

$$e^\alpha(x,y;\lambda) = \lim_{\tau\to\infty} \frac{\Gamma(1+\alpha)}{\lambda_+^\alpha 2\pi i} \int_{\beta-i\tau}^{\beta+i\tau} \frac{e^{z\lambda}}{z^{1+\alpha}} U_z(x,y)\,dz. \tag{2.24}$$

Proof: Formula (2.24) is a consequence of the identity

$$\frac{\Gamma(1+\alpha)}{\lambda_+^\alpha 2\pi i} \int_{\beta-i\infty}^{\beta+i\infty} \frac{e^{z(\lambda-\lambda')}}{z^{1+\alpha}}\,dz = \left(1-\frac{\lambda'_+}{\lambda_+}\right)^\alpha u(\lambda-\lambda'),$$

and an argument parallel to that used in Proposition 2. □

Propositions 2 and 3 construct the spectral kernels $e(x,y;\lambda)$ and $e^\alpha(x,y;\lambda)$ from contour integrals of the kernels $U_z(x,y)$, where $z=\beta+it$, $t\in\mathbb{R}$, $\beta>0$. It is natural to inquire what relationship exists between $e(x,y;\lambda)$ and the time evolution kernels. Because of the absence of decay of $U_{it}(x,y)$ for large x and y, a somewhat weaker result than Propositions 2 and 3 is to be expected. For $0<\epsilon<\tau<\infty$ and $\lambda\in\mathbb{R}$, we define the function $a^\alpha_\lambda(\tau,\epsilon;x,y)$ by the integral

$$a^\alpha_\lambda(\tau,\epsilon;x,y) = \frac{\Gamma(1+\alpha)}{\lambda_+^\alpha 2\pi i} \int_{C_{\tau\epsilon}} \frac{e^{z\lambda}}{z^{1+\alpha}} U_z(x,y)\,dz. \tag{2.25}$$

Except for the small indentation about $z=0$, $U_z(x,y) = U_{it}(x,y)$, in the integrand of Eq. (2.25). For finite τ and λ, function $a^\alpha_\lambda(\tau,\epsilon;x,y)$ is bounded and jointly continuous for $(x,y)\in\mathbb{R}^d\times\mathbb{R}^d$,

Lemma 2: Let $v\in\mathscr{F}^r$. For each given value of the parameters $\alpha\geqslant 0$, $0<\epsilon<\tau<\infty$, and $\lambda\in\mathbb{R}$, define a linear operator $A^\alpha_\lambda(\tau,\epsilon)$: $L^1(\mathbb{R}^d)\cap L^2(\mathbb{R}^d)\to L^2(\mathbb{R}^d)$ by

$$(A^\alpha_\lambda(\tau,\epsilon)u)(x) = \int dy\, a^\alpha_\lambda(\tau,\epsilon;x,y)u(y) \tag{2.26}$$

for $u\in L^1(\mathbb{R}^d)\cap L^2(\mathbb{R}^d)$. The function $A^\alpha_\lambda(\tau,\epsilon)u$ is also represented by

$$(A^\alpha_\lambda(\tau,\epsilon)u)(x) = \frac{\Gamma(1+\alpha)}{\lambda_+^\alpha 2\pi i} \int_{C_{\tau\epsilon}} \frac{e^{z\lambda}}{z^{1+\alpha}} (e^{-zH}u)(x)\,dz. \tag{2.27}$$

Proof: The contour $C_{\tau\epsilon}$ is of finite length; the integrand appearing in Eq. (2.25) is uniformly bounded with respect to $(x,y)\in\mathbb{R}^d\times\mathbb{R}^d$ for $z\in C_{\tau\epsilon}$. Furthermore, since $u(y)\in L^1(\mathbb{R}^d)$, Fubini's theorem lets us write

$$\int dy \left\{\frac{1}{2\pi i}\int_{C_{\tau\epsilon}} \frac{e^{z\lambda}}{z^{1+\alpha}} U_z(x,y)\,dz\right\} u(y) = \frac{1}{2\pi i}\int_{C_{\tau\epsilon}} \frac{e^{z\lambda}}{z^{1+\alpha}} \left[\int dy\, U_z(x,y)u(y)\right] dz.$$

This identity proves (2.27). To show $A^\alpha_\lambda(\tau,\epsilon)u\in L^2(\mathbb{R}^d)$, we note that the spectral integral (2.1) implies $\|e^{-zH}u\|\leqslant e^{\epsilon\|\mu\|}\|u\|$ if $z=\epsilon e^{i\phi}$, $(\pi/2\geqslant\phi\geqslant-\pi/2)$, and the unitarity of the time evolution operator means that $\|e^{-itH}u\| = \|u\|$. Combined, these estimates give the bound

$$\left\| \frac{1}{2\pi i}\int_{C_{\tau\epsilon}} \frac{e^{z\lambda}}{z^{1+\alpha}} (e^{-zH}u)\,dz \right\| \leqslant (2\tau+\pi\epsilon)\frac{1}{\epsilon^{1+\alpha}} e^{\epsilon\|\mu\|}\|u\|.$$

This shows $A^\alpha_\lambda(\tau,\epsilon)u\in L^2(\mathbb{R}^d)$. □

Proposition 4: Let $v\in\mathscr{F}^r$. If $\lambda\in\mathbb{R}$ is a point of continuity of E_λ, then, for all $u\in L^1(\mathbb{R}^d)\cap L^2(\mathbb{R}^d)$,

$$\lim_{\tau,\epsilon} \|[A^\alpha_\lambda(\epsilon,\tau) - E^\alpha_\lambda]u\| = 0, \quad \alpha\geqslant 0.$$

Proof: The derivation follows the same pattern as that for Propositions 2 and 3. However, the role of Corollary 1 is replaced by the spectral representation

$$(f,e^{-zH}u) = \int_\Lambda e^{-z\lambda}\, d(f,E_\lambda u), \quad f\in L^2(\mathbb{R}^d).$$

This representation is valid for all $z\in\overline{D}$ because the measure $d(f,E_\lambda u)$ has bounded variation on the interval $[-\|\mu\|,\infty]$,

$$\int_{-\|\mu\|}^{\infty} d|(f,E_\lambda u)| \leqslant \|f\|\,\|u\|. \qquad \square$$

A result closely related to Proposition 4 may be found in Sinha,[31] Eq. (1).

III. POINT SPECTRAL PROJECTOR

The point spectrum of H has the spectral projector E^p_λ. For Hamiltonians H with $v(x)\in\mathscr{F}^r$, we describe the basic properties of the kernel, $e_p(x,y;\lambda)$, that is associated with the projector E^p_λ. As noted in the Introduction, the contribution of the point spectrum to $U_{\beta+it}(x,y)$ is

$$\sum_{\lambda_j\in\sigma_p(H)} e^{-(\beta+it)\lambda_j}\psi_j(x)\,\overline{\psi_j(y)}.$$

A basic feature of this term is that it has no decay (or growth) as $|t|\to\infty$. For this reason, the integral (2.21) constructing $e(x,y;\lambda)$ cannot be absolutely convergent if $\tau=\infty$. However, the estimates given in this section allow one to show that the point spectrum contribution to integral (2.21) is uniformly convergent for $|x-y|$ contained in some finite interval, and sufficient large values of λ.

Lemma 3: Let $v\in\mathscr{F}^r$ and n be an integer. If $n>d/2$ and $\lambda_0<-\|\mu\|$, then the operator $(H-\lambda_0)^{-n}$ has a jointly continuous, uniformly bounded kernel in $\mathbb{R}^d\times\mathbb{R}^d$, $r^n(x,y;\lambda_0)$, with the properties

(1) $\|r^n(\cdot,y;\lambda_0)\| < K < \infty$,

where K is independent of y.

(2) $\lim_{x\to y} \|r^n(x,\cdot,\lambda_0) - r^n(y,\cdot;\lambda_0)\| = 0$, all $y\in\mathbb{R}^d$.

Proof: Joint continuity and boundedness of $r^n(x,y;\lambda_0)$ are well known (see Simon,[25] Theorem B.7.1). Briefly, these two properties follow from the Laplace transform formula

$$r^n(x,y;\lambda_0) = \frac{1}{(n-1)!}\int_0^\infty \beta^{n-1}e^{\beta\lambda_0}U_\beta(x,y)\,d\beta$$

and the known[11] continuity and boundedness estimates for the kernel $U_\beta(x,y)$.

The symmetry property $r^n(x,y;\lambda_0) = \overline{r^n(x,y;\lambda_0)}$ together with the formula

$$r^{2n}(x,y;\lambda_0) = \int dx'\, r^n(x,x';\lambda_0) r^n(x',y;\lambda_0)$$

gives the desired result (1) if we set $x = y$. Finally (2) is implied by the identity

$$\|r^n(x,\cdot;\lambda_0) - r^n(y,\cdot;\lambda_0)\|^2 = r^{2n}(x,x;\lambda_0) + r^{2n}(y,y;\lambda_0) - 2r^{2n}(x,y;\lambda_0)$$

and the continuity properties of $r^{2n}(x,y;\lambda_0)$ stated above. □

Lemma 4: Let $v \in \mathscr{F}^r$ and $\{\psi_i\}$ be the $L^2(\mathbb{R}^d)$ eigenfunctions of H. Then

(1) $\psi_i(x)$ is continuous and uniformly bounded in $\mathbb{R}^d$.

(2) There exists a constant $C(\lambda)$ depending only on λ such that

$$\sum_{\lambda_i < \lambda} |\psi_i(x)|^2 < C(\lambda) < \infty, \quad \text{all } x \in \mathbb{R}^d. \tag{3.1}$$

(3) The spectral projector for the point spectrum, E^{p}_λ, is an integral operator on $L^2(\mathbb{R}^d)$ with a kernel $e_{\mathrm{p}}(x,y;\lambda)$ that is jointly continuous and uniformly bounded in $\mathbb{R}^d \times \mathbb{R}^d$, and is given by

$$e_{\mathrm{p}}(x,y;\lambda) = \sum_{\lambda_i < \lambda} \psi_i(x)\, \overline{\psi_i(y)}. \tag{3.2}$$

Proof: (1) Since $(H - \lambda_0)^{-n}$ is an integral operator if $n > d/2$ and $\lambda_0 < -\|\mu\|$, the identity

$$\psi_i = (H - \lambda_0)^{-n}(H - \lambda_0)^n \psi_i$$

has the pointwise form

$$\psi_i(x) = (\lambda_i - \lambda_0)^n \int dy\, r^n(x,y;\lambda_0)\psi_i(y).$$

Because $r^n(x,\cdot;\lambda_0) \in L^2(\mathbb{R}^d)$ and $\psi_i \in L^2(\mathbb{R}^d)$, Hölder's inequality implies

$$|\psi_i(x)| \leqslant (\lambda_i - \lambda_0)^n \|r^n(x,\cdot;\lambda_0)\|\ \|\psi_i\|.$$

Furthermore, it is apparent that property (2) of Lemma 3 leads to the continuity of $\psi_i(x)$ in x.

(2) Since E^{p}_λ is a projection operator on $L^2(\mathbb{R}^d)$, one has the obvious inequality

$$\|f\|^2 \geqslant \|E^{\mathrm{p}}_\lambda f\|^2, \quad f \in L^2(\mathbb{R}^d). \tag{3.3}$$

Note that

$$\|E^{\mathrm{p}}_\lambda f\|^2 = (f, E^{\mathrm{p}}_\lambda f) = \sum_{\lambda_i < \lambda} |(f,\psi_i)|^2, \quad f \in L^2(\mathbb{R}^d).$$

Now choose $f = r^n(x,\cdot;\lambda_0)$. This is possible because of Lemma 3, (1). The inner product (f,ψ_i) is

$$(r^n(x,\cdot;\lambda_0),\psi_i) = \int dy\, r^n(x,y;\lambda_0)\psi_i(y) = \frac{1}{(\lambda_i - \lambda_0)^n}\psi_i(x). \tag{3.4}$$

Combining (3.3) and (3.4) gives us

$$\|r^n(x,\cdot;\lambda_0)\|^2 \geqslant \sum_{\lambda_i < \lambda} \frac{1}{|\lambda_i - \lambda_0|^{2n}} |\psi_i(x)|^2 \geqslant \frac{1}{|\lambda - \lambda_0|^{2n}} \sum_{\lambda_i < \lambda} |\psi_i(x)|^2. \tag{3.5}$$

Multiplying (3.5) by $|\lambda - \lambda_0|^{2n}$ determines $C(\lambda)$ to be

$$C(\lambda) = |\lambda - \lambda_0|^{2n} \|r^n(x,\cdot;\lambda_0)\|^2.$$

The boundedness of the sum in (3.2) is a result of the Schwartz inequality and estimate (3.1). The continuity in x,y of the sum in (3.2) is implied by the norm continuity of the kernel $r^n(x,y;\lambda_0)$. □

The following proposition requires that the point spectrum of H lie in a finite interval of the real axis. For the N-body problem a number of authors[32] have proved this result for various assumptions on the pair potentials entering the definition of $v(x)$. Suppose the vectors r_i for $i = 1,2,\ldots,N$ denote the positions of the N particles in $\mathbb{R}^3$, i.e., $x = (r_1, r_2, \ldots, r_N)$. If the function $\Phi_{ij}: \mathbb{R}^3 \to \mathbb{R}$ represents the pair potential between particles i and j, then $v(x)$ is defined by

$$v(x) = \sum_{i>j} \Phi_{ij}(r_i - r_j).$$

A typical result for the point spectrum of H in this case is (see Simon[32]) the following. If Φ_{ij} is of the form

$$\Phi_{ij}(|r_i - r_j|) = |r_i - r_j|^{-1} \int_0^\infty e^{-\alpha|r_i - r_j|}\, d\nu_{ij}(\alpha),$$

where each ν_{ij} is a finite measure, then H has no positive eigenvalues.

The integration paths $C_{\tau\epsilon}$ and $C'_{\tau\beta}$ entering the following proposition are those appearing in Lemma 1.

Proposition 5: Let $v \in \mathscr{F}^r$. Suppose $\Sigma = \sup \sigma_{\mathrm{p}}(H) < \infty$. Then for all $\lambda \notin \sigma_{\mathrm{p}}(H)$ and all $(x,y) \in \mathbb{R}^d \times \mathbb{R}^d$

$$e_{\mathrm{p}}(x,y;\lambda) = \lim_{\tau,\epsilon} \frac{1}{2\pi i} \int_{C_{\tau\epsilon}} \frac{e^{z\lambda}}{z} U^{\mathrm{p}}_z(x,y)\, dz, \tag{3.6}$$

where

$$U^{\mathrm{p}}_z(x,y) = \sum_j e^{-z\lambda_j} \psi_j(x)\, \overline{\psi_j(y)}. \tag{3.7}$$

Representation (3.6) remains valid if contour $C_{\tau\epsilon}$ is replaced by $C'_{\tau\beta}$ and $\lim_{\tau,\epsilon}$ by $\lim_\tau$.

Proof: The sum Σ_j may have an infinite number of terms and λ_j may have more than one cluster point. Nevertheless, the sum (3.7) is always absolutely and uniformly convergent for all x, y and all $\mathrm{Re}\, z \geqslant 0$. To see this, set $c_j = \psi_j(x)\, \overline{\psi_j(y)}$. We have

$$\sum_{\lambda_j < \lambda} |c_j| \leqslant \left(\sum_{\lambda_i < \lambda} |\psi_i(x)|^2\right)^{1/2} \left(\sum_{\lambda_j < \lambda} |\psi_j(y)|^2\right)^{1/2} \leqslant C(\Sigma) < \infty, \tag{3.8}$$

where $C(\Sigma)$ is the constant function in estimate (3.1) with $\lambda = \Sigma$. For $\mathrm{Re}\, z \geqslant 0$, $|e^{-z\lambda_i}| \leqslant e^{\|\mu\| \mathrm{Re}\, z}$, thus sum (3.7) is uniformly and absolutely convergent. Consider integral (3.6). For fixed finite τ and ϵ we can interchange the sum and integration in (3.6)

$$\frac{1}{2\pi i} \int_{C_{\tau\epsilon}} \frac{e^{z\lambda}}{z} U^{\mathrm{p}}_z(x,y)\, dz = \sum_j \frac{c_j}{2\pi i} \int_{C_{\tau\epsilon}} \frac{e^{z(\lambda - \lambda_j)}}{z} dz = \sum_j c_j I(\lambda - \lambda_j; \tau, \epsilon).$$

Since $|I(\lambda-\lambda_j;\tau,\epsilon)|\leqslant$const independent of ϵ and τ, this last sum is uniformly and absolutely convergent. Thus

$$\lim_{\tau\epsilon}\sum_j c_j I(\lambda-\lambda_j;\tau,\epsilon)=\sum_j c_j u(\lambda_j-\lambda)=\sum_{\lambda_j<\lambda} c_j, \quad (3.9)$$

where the last equality requires $\lambda\notin\sigma_p(H)$. Equation (3.9) is equivalent to statement (3.6). An identical proof follows for the case where $C_{\tau\epsilon}$ is replaced by $C'_{\tau\beta}$. □

IV. AN ASYMPTOTIC EXPANSION THEOREM

This section examines the asymptotic behavior of the oscillatory integrals (2.21) and (2.24) that occur in the inverse Laplace transform for the spectral kernels $e(x,y;\lambda)$ and $e^{\alpha}(x,y;\lambda)$. Our analysis is motivated by a recent paper of Olver.[33] We not only find the analytical form of the large λ expansion of these oscillatory integrals but also an explicit expression for the error term. Using Eq. (2.2) and Eq. (2.3) to express the integrand of (2.24) leads to the following formula for the Reisz-mean kernel:

$$e^{\alpha}(x,y;\lambda)=\frac{\Gamma(1+\alpha)}{\lambda_+^{\alpha}\,2\pi i}\int_{\beta-i\infty}^{\beta+i\infty}\frac{e^{z\lambda}}{z^{1+\alpha}}\left[\left(\frac{1}{4\pi zq}\right)^{d/2}\times\exp\left(\frac{-|x-y|^2}{4zq}\right)\right]F(x,y;z)\,dz. \quad (4.1)$$

The large λ expansion we seek results when Eq. (1.6) is inserted into integral (4.1).

To begin, we record several elementary facts.

Lemma 5: Let $q(t)$ be continuous in $(0,\infty)$. Then

$$\int_0^{\infty} q(t)e^{i\lambda t}\,dt=o(1) \quad \text{as } \lambda\to\infty,$$

provided that the integral converges uniformly at 0 and ∞ for all sufficiently large λ.

Proof: (Olver,[34] p. 73). □

Lemma 6: If the integral

$$\int_0^{\infty} f(t)\,dt$$

is convergent, then

$$\lim_{\epsilon\to+0}\int_0^{\infty} e^{-\epsilon t}f(t)\,dt=\int_0^{\infty} f(t)\,dt.$$

Proof: (Titchmarch,[35] p. 26). □

Lemma 7:

$$\frac{d^k}{dt^k}e^{a/t}=\frac{(-1)^k}{t^k}e^{a/t}\sum_{j=0}^{k-1}\frac{\Gamma(k)}{\Gamma(k-j)}\binom{k}{j}\left(\frac{a}{t}\right)^{k-j}.$$

Proof: (Gradstein and Ryzhik,[36] p. 19). □

Lemma 8: For $\nu>0$,

$$\lim_{\eta\to+0}\left[(1/2\pi)\int_{-\infty}^{\infty}(it)^{\nu-1}\exp(-\eta|t|+i\lambda t+i\mu/t)\,dt\right]=(\mu/\lambda)^{\nu/2}J_{-\nu}(2\sqrt{\lambda\mu}). \quad (4.2)$$

Proof: It is known that for $\mathrm{Re}\,a>0$ and $\mathrm{Re}\,p>0$,

$$\int_0^{\infty} t^{\nu-1}\exp(-pt-a/4t)\,dt=2(a/4p)^{\nu/2}K_{\nu}(\sqrt{ap});$$

see (Erdélyi, *et al.*[37] [4.5.29]). Put $a=4(\eta\pm i\mu)$ and let $\eta\to+0$. By Lebesgue's dominated convergence theorem, we have for $\mathrm{Re}\,\nu>0$ and $\mathrm{Re}\,p>0$

$$\int_0^{\infty} t^{\nu-1}\exp(-pt\pm i\mu/t)\,dt=2e^{\mp i\pi\nu/4}(\mu/p)^{\nu/2}K_{-\nu}(2e^{\mp i\pi/4}\sqrt{\mu p}),$$

since $K_{\nu}(z)=K_{-\nu}(z)$ for unrestricted z.

Therefore,

$$\lim_{\eta\to+0}\int_0^{\infty} t^{\nu-1}\exp[-(\eta-i\lambda)t+i\mu/t]\,dt=\pi ie^{-i\pi\nu/2}(\mu/\lambda)^{\nu/2}H^{(1)}_{-\nu}(2\sqrt{\lambda\mu})$$

and

$$\lim_{\eta\to+0}\int_0^{\infty} t^{\nu-1}\exp[-(\eta+i\lambda)t-i\mu/t]\,dt=-\pi ie^{i\pi\nu/2}(\mu/\lambda)^{\nu/2}H^{(2)}_{-\nu}(2\sqrt{\lambda\mu}).$$

Here we have used the connecting properties

$$H^{(1)}_{\nu}(iw)=(2/\pi i)e^{-i\pi\nu/2}K_{\nu}(w), \quad -\pi<\arg w<\pi/2,$$

$$H^{(2)}_{\nu}(-iw)=-(2/\pi i)e^{i\pi\nu/2}K_{\nu}(w), \quad -\pi/2<\arg w<\pi,$$

and the fact that $K_{\nu}(w)$ is an analytic function in $-\pi<\arg w<\pi$ and hence continuous there.

A combination of the above two equations, together with the connection formula

$$J_{\nu}(z)=\tfrac{1}{2}[H^{(1)}_{\nu}(z)+H^{(2)}_{\nu}(z)],$$

gives the desired result. □

Proposition 6: Let m be a positive integer, f be an m times continuously differentiable function on $(0,\infty)$, and

$$f(t)=\sum_{s=0}^{n-1}a_s t^{s+\alpha-1}+f_n(t), \quad (4.3)$$

where α is a positive number $\leqslant m$ and n satisfies

$$2m+1-\alpha\leqslant n<2m+2-\alpha. \quad (4.4)$$

If $f^{(j)}(t)$ is bounded at $t=\infty$ for $j=0,\ldots,m-1$, if each of the integrals

$$\int_1^{\infty}\exp(i\lambda t+i\mu/t)f^{(j)}(t)\,dt, \quad j=0,1,\ldots,m, \quad (4.5)$$

converges uniformly in λ and μ, and if, for $j=0,1,\ldots,m$, we have

$$f_n^{(j)}(t)=O(t^{n+\alpha-j-1}) \quad \text{as } t\to+0, \quad (4.6)$$

then there exists a function $B_{mn}(\lambda,\mu)$, which is continuous in $\lambda\in(0,\infty)$ and $\mu\in(-\infty,\infty)$, such that

$$\lim_{\eta\to+0}\int_0^{\infty} f_n(t)e^{-\eta t+i\lambda t+i\mu/t}\,dt=(i/\lambda)^m B_{mn}(\lambda,\mu). \quad (4.7)$$

Furthermore, as $\lambda\to\infty$, we have $B_{mn}(\lambda,\mu)=o(1)$ uniformly for μ restricted to any bounded interval.

Remark: If $f^{(j)}(t)\in L^1[1,\infty)$, $j=0,1,\ldots,m$, then the integrals in (4.5), of course, converge uniformly in λ and μ. Furthermore, since

$$f^{(j)}(T)=-\int_T^{\infty} f^{(j+1)}(t)\,dt, \quad j=0,1,\ldots,m-1,$$

we also have $f^{(j)}(T)\to 0$ as $T\to\infty$.

Proof: By integration by parts, we have

$$\int_0^\infty f_n(t)e^{-\eta t+i\lambda t+i\mu/t}\,dt=[1/(\eta-i\lambda)^m]$$
$$\times\int_0^\infty [f_n(t)e^{i\mu/t}]^{(m)}e^{-(\eta-i\lambda)t}\,dt, \tag{4.8}$$

all the integrated terms vanish since $n\geqslant 2m+1-\alpha$. Here we have used the fact that $f^{(j)}(t)$ is bounded at infinity for $j=0,\ldots,m-1$. Note that $[f_n(t)\,e^{i\mu/t}]^{(m)}$ is equal to

$$f_n^{(m)}(t)e^{i\mu/t}+\sum_{k=1}^{m}\binom{m}{k}f_n^{(m-k)}(t)\sum_{l=1}^{k}c_{k,l}t^{-k-l}e^{i\mu/t} \tag{4.9}$$

by Lemma 7, where

$$c_{k,l}=\frac{(-1)^k\Gamma(k)}{\Gamma(l)}\binom{k}{l}(i\mu)^l. \tag{4.10}$$

Thus, to analyze the integral on the right-hand side of Eq. (4.8), we must first consider integrals of the form

$$\int_0^\infty f_n^{(m-k)}(t)t^{-k-l}e^{-\eta t+i\lambda t+i\mu/t}\,dt, \tag{4.11}$$

where $l=0$ when $k=0$ and $1\leqslant l\leqslant k$, where $k\geqslant 1$. For easier understanding, let us first consider the two extremal cases: $k=0$; $l=k=m$.

Case (i): $l=k=m$: Since $n>2m-\alpha$, the integral

$$\int_0^1 f_n(t)t^{-2m}e^{i\lambda t+i\mu/t}\,dt$$

is absolutely convergent, and hence uniformly convergent in λ and μ. Similarly, since $f(t)$ is bounded at $t=\infty$ and $m\geqslant 1$, the integral

$$\int_1^\infty f(t)t^{-2m}e^{i\lambda t+i\mu/t}\,dt$$

converges absolutely and hence uniformly in λ and μ. Now consider the integrals

$$\int_1^\infty t^{s+\alpha-1-2m}e^{i\lambda t+i\mu/t}\,dt,\quad s=0,1,\ldots,n-1.$$

For $s=0,1,\ldots,n-2$, each of these integrals is absolutely convergent. Since

$$e^{i\mu/t}=1+(i\mu/t)e^{i\mu/c_t} \tag{4.12}$$

for some $c_t>t>0$, we have for $s=n-1$

$$\int_1^\infty t^{n+\alpha-2m-2}e^{i\lambda t+i\mu/t}\,dt=\int_1^\infty t^{n+\alpha-2m-2}e^{i\lambda t}\,dt$$
$$+i\mu\int_1^\infty t^{n+\alpha-2m-3}e^{i\lambda t+i\mu/c_t}\,dt. \tag{4.13}$$

The first integral is uniformly convergent for all large λ and the second integral is absolutely convergent. In view of the factor $i\mu$ in the front of the second integral, we can only conclude that the integral on the left-hand side of Eq. (4.13) is uniformly convergent for all large values of λ and all bounded values of μ. The same conclusion holds for the integral

$$\int_0^\infty f_n(t)t^{-2m}e^{i\lambda t+i\mu/t}\,dt,$$

which is equal to

$$\int_0^1 f_n(t)t^{-2m}e^{i\lambda t+i\mu/t}\,dt+\int_1^\infty f(t)t^{-2m}e^{i\lambda t+i\mu/t}\,dt$$
$$-\sum_{s=0}^{n-1}a_s\int_1^\infty t^{s+\alpha-1-2m}e^{i\lambda t+i\mu/t}\,dt.$$

Thus, by Lemma 6,

$$\lim_{\eta\to+0}\int_0^\infty f_n(t)t^{-2m}e^{-\eta t+i\lambda t+i\mu/t}\,dt$$
$$=\int_0^\infty f_n(t)t^{-2m}e^{i\lambda t+i\mu/t}\,dt. \tag{4.14}$$

Case (ii): $k=0$: Here we have

$$\int_0^\infty f_n^{(m)}(t)e^{-\eta t+i\lambda t+i\mu/t}\,dt. \tag{4.15}$$

Note that this integral will not converge under the conditions given in (4.4), if we put $\eta=0$ in Eq. (4.15). However, we shall show the limit of this integral exists as $\eta\to+0$. Choose $r<n$ such that

$$m-\alpha<r<m-\alpha+2 \tag{4.16}$$

and write

$$f_n(t)=-\sum_{s=r}^{n-1}a_s t^{s+\alpha-1}+f_r(t). \tag{4.17}$$

The integral in (4.15) is thus equal to

$$-2\sum_{s=r}^{n-1}a_s\frac{\Gamma(s+\alpha)}{\Gamma(s+\alpha-m)}e^{-i\pi(s+\alpha-m)/4}\left[\frac{\mu}{\eta-i\lambda}\right]^{(s+\alpha-m)/2}$$
$$\times K_{-(s+\alpha-m)}(2e^{-i\pi/4}\sqrt{\mu(\eta-i\lambda)})$$
$$+\int_0^\infty f_r^{(m)}(t)e^{-\eta t+i\lambda t+i\mu/t}\,dt. \tag{4.18}$$

It can be shown, as in case (i), that under the restriction given in (4.16), the last integral with $\eta=0$ converges uniformly for all large values of λ and all bounded values of μ. By Lemma 6 again,

$$\lim_{\eta\to+0}\int_0^\infty f_n^{(m)}(t)e^{-\eta t+i\lambda t+i\mu/t}\,dt$$
$$=-\pi i\sum_{s=r}^{n-1}a_s\frac{\Gamma(s+\alpha)}{\Gamma(s+\alpha-m)}e^{-i\pi(s+\alpha-m)/2}\left(\frac{\mu}{\lambda}\right)^{(s+\alpha-m)/2}$$
$$\times H^{(1)}_{-(s+\alpha-m)}(2\sqrt{\lambda\mu})+\int_0^\infty f_r^{(m)}(t)e^{i\lambda t+i\mu/t}\,dt. \tag{4.19}$$

The general case: Choose r such that

$$m+l-\alpha<r<m+l-\alpha+2 \tag{4.20}$$

(keep in mind that r depends on l). By the same argument we have

$$\lim_{\eta\to+0}\int_0^\infty f_n^{(m-k)}(t)\,t^{-k-l}e^{-\eta t+i\lambda t+i\mu/t}\,dt$$
$$=-\pi i\sum_{s=r}^{n-1}a_s\frac{\Gamma(s+\alpha)}{\Gamma(s+\alpha-m+k)}e^{-i\pi(s+\alpha-m-l)/2}\left(\frac{\mu}{\lambda}\right)^{(s+\alpha-m-l)/2}H^{(1)}_{-(s+\alpha-m-l)}(2\sqrt{\lambda\mu})$$
$$+\int_0^\infty f_r^{(m-k)}(t)\,t^{-k-l}e^{i\lambda t+i\mu/t}\,dt. \tag{4.21}$$

A combination of these results gives

$$\lim_{\eta\to+0}\int_0^\infty f_n(t)e^{-\eta t+i\lambda t+i\mu/t}\,dt$$
$$=\left[\frac{i}{\lambda}\right]^m\left\{-\pi i\sum_{k=0}^m\binom{m}{k}\sum_{l=0}^k c_{k,l}\sum_{s=r}^{n-1}a_s\frac{\Gamma(s+\alpha)}{\Gamma(s+\alpha-m+k)}e^{-i\pi(s+\alpha-m-l)/2}\right.$$
$$\left.\times\left[\frac{\mu}{\lambda}\right]^{(s+\alpha-m-l)/2}H^{(1)}_{-(s+\alpha-m-l)}(2\sqrt{\lambda\mu})+\sum_{k=0}^m\binom{m}{k}\sum_{l=0}^k c_{k,l}\int_0^\infty f_r^{(m-k)}(t)\,t^{-k-l}e^{i\lambda t+i\mu/t}\,dt\right\}, \tag{4.22}$$

where $c_{0,0}=1, c_{k,l}=0$ if $l=0$ and $1\leqslant k\leqslant m$, and $c_{k,l}$ is given by (4.10) if $1\leqslant l\leqslant k\leqslant m$. Recall that r depends on l through relation (4.20). To simplify the above expression, we put

$$C(k,l,s)=-\pi i\binom{m}{k}c_{k,l}a_s\frac{\Gamma(s+\alpha)}{\Gamma(s+\alpha-m+k)}e^{-i\pi(s+\alpha-m-l)/2}. \tag{4.23}$$

Then Eq. (4.22) reads

$$\lim_{\eta\to+0}\int_0^\infty f_n(t)e^{-\eta t+i\lambda t+i\mu/t}\,dt=\left[\frac{i}{\lambda}\right]^m B_{mn}(\lambda,\mu), \tag{4.24}$$

where

$$B_{mn}(\lambda,\mu)=\sum_{k=0}^m\sum_{l=0}^k\sum_{s=r}^{n-1}C(k,l,s)$$
$$\times\left(\frac{\mu}{\lambda}\right)^{(s+\alpha-m-l)/2}H^{(1)}_{-(s+\alpha-m-l)}(2\sqrt{\lambda\mu})$$
$$+\sum_{k=0}^m\binom{m}{k}\sum_{l=0}^k c_{k,l}\int_0^\infty f_r^{(m-k)}(t)\,t^{-k-l}e^{i\lambda t+i\mu/t}\,dt, \tag{4.25}$$

r depending on l and satisfying (4.20).

The condition $r>m+l-\alpha$ in (4.20) ensures that the exponents of (μ/λ) in (4.25) are positive. If $\lambda\mu$ is bounded away from 0 as $\lambda\to+\infty$, then clearly each term in the triple sum approaches zero as $\lambda\to\infty$, uniformly for μ in bounded intervals. In view of the approximation

$$H^{(1)}_\nu(x)=O(x^{-|\nu|})\quad\text{as } x\to+0,$$

the same result holds if $\lambda\mu\to0$ as $\lambda\to\infty$. The fact that each integral in the double sum in (4.25) also has this property follows from the following lemma. This proves the result that

$$\lim_{\lambda\to\infty}B_{mn}(\lambda,\mu)=0$$

uniformly in bounded μ. □

Lemma 9: Let $q(t,\mu)$ be continuous in any closed rectangle $[a,b]\times[c,d]$, where $0<a<b<\infty$ and $-\infty<c<d<\infty$. If the integral

$$F(\lambda,\mu)=\int_0^\infty q(t,\mu)e^{i\lambda t}\,dt$$

converges uniformly at 0 and ∞ for μ in $[c,d]$ and for all sufficiently large λ, then

$$F(\lambda,\mu)=o(1),\quad\text{as }\lambda\to\infty,$$

uniformly for μ in $[c,d]$.

Proof: For fixed μ, this is just the result stated in Lemma 5. For varying μ, it can be proved by modifying the proof given in (Olver,[34] p. 73). □

Proposition 6 obtains an error estimate for the remainder term associated with the asymptotic expansion of the oscillatory type integral (4.1) in the circumstance where $\beta=0$. The following theorem provides the form of the large-λ expansion and control of the remainder term when $\beta>0$.

Let λ be positive and μ be real, and consider the function $I(\lambda;\mu)$ defined by the contour integral

$$I(\lambda;\mu)=\frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty}\exp(\lambda z-\mu/z)G(z)\,dz, \tag{4.26}$$

where c is a positive number and $G(z)$ is an analytic function in $\mathrm{Re}\,z>0$ and continuous for $\mathrm{Re}\,z\geqslant0$ except possibly at the origin. Put $(z=\beta+it)$

$$G(z)=\sum_{s=0}^{n-1}a_s z^{s+\alpha-1}+G_n(z), \tag{4.27}$$

where α is a fixed real number.

Theorem 3: Assume:

(i) For each $j = 0,1,...,m$, $G_n^{(j)}(z) = O(|z|^{n+\alpha-j})$ as $|z| \to 0$, where m and n are related via the inequalities

$$2m - \alpha + 1 \leqslant n < 2m - \alpha + 2. \tag{4.28}$$

(ii) In the strip $0 \leqslant \beta \leqslant c$, $G(\beta + it) \to 0$ as $|t| \to \infty$ uniformly in β and $G^{(j)}(it) = O(1)$ as $|t| \to \infty$ for $j = 1,...,m-1$.

(iii) Each of the integrals

$$\int_1^\infty \exp(i\lambda t \pm i\mu/t) G^{(j)}(\pm it)\, dt, \quad j = 0,1,...,m, \tag{4.29}$$

converges uniformly in λ and μ. Then there exists a function $I_m(\lambda;\mu)$ such that

$$I(\lambda,\mu) = \sum_{s=0}^{n-1} a_s [\mu/\lambda]^{(s+\alpha)/2} J_{-s-\alpha}(2\sqrt{\lambda\mu}) + \frac{1}{\lambda^m} I_m(\lambda;\mu) \tag{4.30}$$

and

$$\lim_{\lambda \to \infty} I_m(\lambda;\mu) = 0, \tag{4.31}$$

the limit being uniformly valid for μ in any bounded subset of $(-\infty, \infty)$.

Proof: We first split the sum in (4.27) into two and write

$$G(z) = \sum_{s=0}^{<1-\alpha} a_s z^{s+\alpha-1} + G_\alpha(z) \tag{4.32}$$

with

$$G_\alpha(z) = \sum_{s>1-\alpha}^{n-1} a_s z^{s+\alpha-1} + G_n(z). \tag{4.33}$$

Recall the well-known identity (Watson,[38] p. 177)

$$J_\nu(z) = [(\tfrac{1}{2}z)^\nu/2\pi i] \int_{c-i\infty}^{c+i\infty} t^{-\nu-1} \exp(t - z^2/4t)\, dt, \quad \operatorname{Re}\nu > -1, \tag{4.34}$$

from which we have, upon inserting (4.32) into (4.26),

$$I(\lambda;\mu) = \sum_{s=0}^{<1-\alpha} a_s \left[\frac{\mu}{\lambda}\right]^{(s+\alpha)/2} J_{-s-\alpha}(2\sqrt{\lambda\mu}) + I_\alpha(\lambda;\mu), \tag{4.35}$$

where

$$I_\alpha(\lambda;\mu) = (1/2\pi i) \int_{c-i\infty}^{c+i\infty} \exp(\lambda z - \mu/z) G_\alpha(z)\, dz. \tag{4.36}$$

By assumptions (ii), (iii) with $j = 0$, and the strong form of Cauchy's theorem,[39] we can move the contour all the way to the imaginary axis. Thus, upon a change of variable,

$$I_\alpha(\lambda;\mu) = (1/2\pi) \int_{-\infty}^{\infty} \exp(i\lambda t + i\mu/t) G_\alpha(it)\, dt. \tag{4.37}$$

Furthermore, since the last integral exists (uniformly in λ and μ) as an improper Riemann integral by assumption (iii), Lemma 6 gives

$$I_\alpha(\lambda;\mu) = (1/2\pi) \lim_{\eta \to +0} \int_{-\infty}^{\infty} \exp(-\eta|t| + i\lambda t + i\mu/t) \times G_\alpha(it)\, dt. \tag{4.38}$$

Inserting (4.33) into (4.38) and applying Lemma 8, we obtain

$$I_\alpha(\lambda;\mu) = \sum_{s>1-\alpha}^{n-1} a_s [\mu/\lambda]^{(s+\alpha)/2} J_{-s-\alpha}(2\sqrt{\lambda\mu}) + (1/2\pi) \lim_{\eta \to +0} \int_{-\infty}^{\infty} \exp(-\eta|t| + i\lambda t + i\mu/t) G_n(it)\, dt. \tag{4.39}$$

Coupling (4.39) and (4.35) yields

$$I(\lambda;\mu) = \sum_{s=0}^{n-1} a_s [\mu/\lambda]^{(s+\alpha)/2} J_{-s-\alpha}(2\sqrt{\lambda\mu}) + (1/2\pi) \lim_{\eta \to +0} \int_{-\infty}^{\infty} \exp(-\eta|t| + i\lambda t + i\mu/t) G_n(it)\, dt. \tag{4.40}$$

To the last integral we apply Proposition 6 with $f_n(t) = G_n(\pm it)$. Hence we conclude that there exists a function $I_m(\lambda;\mu)$ such that Eqs. (4.30) and (4.31) hold. If the explicit form of $I_m(\lambda;\mu)$ is desired, one can use expression (4.25). □

V. SPECTRAL KERNELS AND THEIR ASYMPTOTIC FORMS

This section gives our specific results for the large-λ behavior of the spectral kernel $e(x,y;\lambda)$. To begin, we recall the known results[11] for the small z expansion of the semigroup kernel $U_z(x,y)$. The coefficient functions, $P_n(x,y)$, entering this expansion can be defined explicitly in terms of the measure, $\mu(\alpha)$, associated with potential $v(x)$. For $v \in \mathscr{F}^r_{2M}$, the subsidiary coefficient functions $D_{m,n+m}(x,y)$ with $m \leqslant M$ and $n \geqslant 0$ are defined by the following formulas. Set

$$\Theta(\xi_l,\xi_m) = \min\{\xi_l(1-\xi_m), \xi_m(1-\xi_l)\}, \tag{5.1}$$

where $0 \leqslant \xi_l \leqslant 1$, for $l = 1,2,\cdots$. To simplify the expressions, we use the multiple integral abbreviation,

$$\int d^n\xi \int d^n\mu = \int_0^1 d\xi_1 \cdots \int_0^1 d\xi_n \int d\mu(\alpha_1) \cdots \int d\mu(\alpha_n).$$

The functions $D_{m,n+m}(x,y)$ are real-valued, jointly and uniformly continuous in $\mathbb{R}^d \times \mathbb{R}^d$, and have the representation

$$D_{m,n+m}(x,y) = \binom{n+m}{m} \int d^n\xi \int d^n\mu \left[\sum_{l=1}^n \sum_{r=1}^n \Theta(\xi_l,\xi_r)\alpha_l\alpha_r\right]^m \times \exp\left(i \sum_{l=1}^m [\xi_l x + (1-\xi_l)y]\alpha_l\right) \tag{5.2}$$

for $n \geqslant 1$, $m \geqslant 0$. In Eq. (5.2), $\alpha_l\alpha_r$ denotes the scalar product of two vectors in $\mathbb{R}^d$. Furthermore, $D_{0,0}(x,y) = 1$ and $D_{m,m}(x,y) = 0$ if $m \geqslant 1$. The small z expansion for $U_z(x,y)$ is then given by

Theorem 4: Let $v(x)$ be a real-valued, bounded, and continuous potential represented by a complex bounded measure μ. Let $U_z(x,y)$ and $U_z^0(x-y)$ be the kernels of the integral operators e^{-zH} and e^{-zH_0} for $z \in D$, respectively. If $v(x) \in \mathscr{F}^r_{2M}$, then for all $z \in \overline{D} \setminus \{0\}$,

$$U_z(x,y) = U_z^0(x-y)\left[\sum_{n=0}^{M-1} \frac{(-z)^n}{n!} P_n(x,y) + E_M(x,y;z)\right], \tag{5.3}$$

where coefficient functions $P_n(x,y)$ are, $P_0(x,y) = 1$,

$$P_n(x,y) = \sum_{m=0}^{n-1} q^m D_{m,n}(x,y), \quad n = 1,2,...,M-1, \tag{5.4}$$

and have the x,y uniform bound

$$|P_n(x,y)| \leqslant (\|\mu\| + qn^2K^2/4)^n. \tag{5.5}$$

Here K is the bound constant of $v(x)$ in $\mathscr{F}^{\mathrm{r}}_{2M}$. The remainder term is of order $O(|z|^M)$ and has the bound

$$|E_M(x,y;z)| \leqslant [(|z|\ \|\mu\|)^M/M!] \times [(1 + q(M-1)^2K^2/4\|\mu\|)^M + \exp(|z|\ \|\mu\|)]. \tag{5.6}$$

Furthermore, if $v(x) \in \mathscr{F}^{\mathrm{r}}_{2(M+i)}$, then one may take i derivatives of expansion (5.3) with respect to $z \in \overline{D} \setminus \{0\}$. The ith derivative of the remainder term $E_M(x,y;z)$ is of order $O(|z|^{M-i})$.

Proof: See Ref. 11. □

Our basic approach to studying the large-λ expansion is to represent $e(x,y;\lambda)$ by the inverse Laplace transform in Proposition 2 and then utilize the asymptotic expansion Theorem 3. In order to use Theorem 3, it is necessary to establish that the integrand in (2.21) has uniform decay in a strip parallel to the time axis. For this purpose we need the following technical result:

Lemma 10: Let $v \in \mathscr{F}^{\mathrm{r}}$. If $t \neq 0$ and $\beta > 0$ then for a. a. $(x,y) \in \mathbb{R}^d \times \mathbb{R}^d$

$$U_{\beta+it}(x,y) = \int dy''\ U_\beta(x,y'')U_{it}(y'',y). \tag{5.7}$$

Proof: This plausible identity is the kernel analog of the operator equation

$$e^{-(\beta+it)H} = e^{-\beta H}e^{-itH}. \tag{5.8}$$

However, explicit proof is called for since $U_{it}(y'',y)$ is a kernel only in the limited sense of Theorem 1. Take $f(x) \in L^1(\mathbb{R}^d) \cap L^2(\mathbb{R}^d)$. Then, Theorem 1, (2), gives us

$$(e^{-itH}f)(y') = \int dy''\ U_{it}(y',y'')f(y'') \in L^2(\mathbb{R}^d). \tag{5.9}$$

Now use Theorem 1, (1) to determine $(e^{-\beta H}[e^{-itH}f])(x)$; thus utilizing both (5.8) and (5.9) yields

$$(e^{-(\beta+it)H}f)(x) = \int dy'\ U_\beta(x,y') \times \left[\int dy''\ U_{it}(y',y'')f(y'')\right]. \tag{5.10}$$

Since $f(y'') \in L^1(\mathbb{R}^d)$, $U_\beta(x,\cdot) \in L^1(\mathbb{R}^d)$, and $U_{it}(y',y'')$ is uniformly bounded [see inequality (2.4)], we have that the double integral in (5.10) is absolutely convergent. Interchange the order of integration to obtain

$$(e^{-(\beta+it)H}f)(x) = \int dy'' \times \left[\int dy'\ U_\beta(x,y')U_{it}(y',y'')\right]f(y''). \tag{5.11}$$

However, Theorem 1, (1), also gives us the representation

$$(e^{-(\beta+it)H}f)(x) = \int dy''\, U_{\beta+it}(x,y'')f(y''). \tag{5.12}$$

Subtract (5.12) from (5.11) to find

$$0 = \int dy'' \left[U_{\beta+it}(x,y'') - \int dy'\ U_\beta(x,y')U_{it}(y',y'')\right]f(y''). \tag{5.13}$$

If we can show the function of y'' in [ ] is $L^2(\mathbb{R}^d)$, then the integral (5.13) can be interpreted as the inner product in $L^2(\mathbb{R}^d)$. Since $L^1(\mathbb{R}^d) \cap L^2(\mathbb{R}^d)$ is dense in $L^2(\mathbb{R}^d)$, it would then follow that [ ] = 0 for a. a. $y'' \in \mathbb{R}^d$. Note that $U_{\beta+it}(x,\cdot) \in L^2(\mathbb{R}^d)$ by Proposition 1, parts (1) and (3). Thus it remains to consider the integral term in [ ]. Let the symbol * denote complex conjugation. Since H is self-adjoint, the kernels $U_z(x,y)$ satisfy

$$U_z(x,y) = [U_{z^*}(y,x)]^*.$$

So we can write

$$\int dy'\ U_\beta(x,y')U_{it}(y',y'') = \int dy'\ U_{-it}(y'',y')^*U_\beta(y',x)^* = [(e^{-itH}U_\beta(\cdot,x))(y'')]^*.$$

Because e^{-itH} is a unitary operator in $L^2(\mathbb{R}^d)$, the last function on the right is $L^2(\mathbb{R}^d)$. □

Consider, for the moment, the Bessel function form assumed by the spectral kernel for H_0. Let $\alpha \geqslant 0$ be the Riesz-mean index; Proposition 3, Eq. (2.24), with $H \to H_0$ states that

$$e_0^\alpha(x,y;\lambda) = \frac{\Gamma(1+\alpha)}{\lambda_+^\alpha}\frac{1}{2\pi i}\int_{\beta-i\infty}^{\beta+i\infty}\frac{e^{z\lambda}}{z^{1+\alpha}} \times \frac{1}{(4\pi zq)^{d/2}}\exp\left(\frac{-|x-y|^2}{4zq}\right)dz.$$

By using a standard inverse Laplace transform formula (Erdélyi,[36] [5.5.40]), one finds

$$e_0^\alpha(x,y;\lambda) = \frac{\Gamma(1+\alpha)}{\lambda_+^\alpha}\frac{1}{(4\pi q)^{d/2}} \times \left(\frac{2\lambda^{1/2}q^{1/2}}{|x-y|}\right)^{\alpha+d/2} J_{\alpha+d/2}(\lambda^{1/2}q^{-1/2}|x-y|). \tag{5.14}$$

Clearly, if $\alpha = 0$, then formula (1.9) is recovered.

Theorem 5: Let N be a positive integer and M be the integer satisfying

$$2N + d/2 + 1 \leqslant M < 2N + d/2 + 2.$$

Assume that the Hamiltonian H satisfies: (i) $v \in \mathscr{F}^{\mathrm{r}}_{2(M+N)}$; (ii) $\sup \sigma_{\mathrm{p}}(H) = \Sigma < \infty$; and (iii) for some $\gamma < 0$, $H \in T_n(\gamma;N)$. Then there exists a function $\Phi_N(x,y;\lambda)$, jointly continuous in $\mathbb{R}^d \times \mathbb{R}^d$, such that

$$e(x,y;\lambda) = \sum_{n=0}^{M-1}\frac{(-1)^n}{n!}P_n(x,y)\left(\frac{d}{d\lambda}\right)^n e_0(x,y;\lambda) + \frac{1}{\lambda^N}\Phi_N(x,y;\lambda) \tag{5.15}$$

and

$$\Phi_N(x,y;\lambda) \to 0 \quad \text{as } \lambda \to \infty$$

uniformly in any compact subset of $\mathbb{R}^d \times \mathbb{R}^d$.

Proof: We shall apply Theorem 3. The function $G(z)$ of Theorem 3 is determined by Eqs. (2.3) and (2.21) to be

$$G(z) = \frac{1}{(4\pi q)^{d/2}} \frac{F(x,y;z)}{z^{1+d/2}}$$
$$= \frac{1}{(4\pi q)^{d/2}} \frac{1}{z^{1+d/2}} \frac{U_z(x,y)}{U_z^0(x-y)}$$
$$= \frac{1}{z} \exp\left[\frac{|x-y|^2}{4zq}\right] U_z(x,y). \tag{5.16}$$

With this formula for $G(z)$ we have the identification

$$e(x,y;\lambda) = I(\lambda, -|x-y|^2/4q), \tag{5.17}$$

where $I(\lambda,\mu)$ is the contour integral (4.26). We demonstrate that $G(z)$ satisfies conditions (i), (ii), and (iii) of Theorem 3.

Examine first property (i). Since $v \in \mathcal{F}^r_{2(M+N)}$, Theorem 4 provides a small-z expansion of the form

$$G(z) = \frac{1}{(4\pi q)^{d/2}} \left[\sum_{n=0}^{M-1} \frac{(-1)^n}{n!} P_n(x,y) z^{n-1-d/2} + \frac{E_M(x,y;z)}{z^{1+d/2}}\right], \tag{5.18}$$

where for $z \in \overline{D} \setminus \{0\}$ and $l = 0,1,\ldots,N$

$$\left(\frac{d}{dz}\right)^l \left\{\frac{E_M(x,y;z)}{z^{1+d/2}}\right\} = O(|z|^{M-1-d/2-l}).$$

Thus property (i) is satisfied.

Consider requirement (ii) of Theorem 3. Set $z = \beta + it$ in Eq. (5.16); take $\beta \in [0,1]$ and $t^2 > 1$. We have

$$|G(\beta+it)| \leqslant |t|^{-1} \exp(|x-y|^2/4q) |U_{\beta+it}(x,y)|.$$

To estimate $U_{\beta+it}(x,y)$ recall that Proposition 1 gives us

$$|U_\beta(x,y'')| \leqslant e^{\beta\|\mu\|} U^0_\beta(x-y''). \tag{5.19}$$

If $H \in T_u(\gamma;N)$, then

$$|U^c_{it}(y'',y)| < C_u |t|^\gamma \quad \text{as} \quad |t| \to \infty. \tag{5.20}$$

On the other hand, the point spectrum component $U^p_{it}(x,y)$ of the time evolution kernel is bounded by a constant. This follows, since $\sup \sigma_p(H) = \Sigma < \infty$ and estimate (3.8) gives

$$\left|\sum_j e^{-i\lambda_j t} \psi_j(y'') \overline{\psi_j(y)}\right|$$
$$\leqslant \sum_j |\psi_j(y'') \overline{\psi_j(y)}| < C(\Sigma) < \infty. \tag{5.21}$$

Combining (5.20) and (5.21) leads to

$$|U_{it}(y'',y)| < C(\Sigma) + C_u |t|^\gamma \quad \text{as} \quad |t| \to \infty, \tag{5.22}$$

where both $C(\Sigma)$ and C_u are independent of x and y. Inserting inequalities (5.19) and (5.22) into identity (5.7) implies

$$|U_{\beta+it}(x,y)| < (C(\Sigma) + C_u |t|^\gamma) e^{\beta\|\mu\|} \int dy''\, U^0_\beta(x-y''). \tag{5.23}$$

Expression (2.2) for $U^0_\beta(x-y'')$ means that the dy'' integral in Eq. (5.23) is unity. Thus for large $|t|$ we have the β-uniform estimate

$$|G(\beta+it)| < (C(\Sigma) + C_u |t|^\gamma)[1/(4\pi q)^{d/2} t] \times \exp(|x-y|^2/4q + \|\mu\|), \quad \beta \in [0,1]. \tag{5.24}$$

Thus the first part of condition (ii) of Theorem 3 is satisfied since $\gamma < 0$.

Finally, we show that $G^{(j)}(it) = O(1)$ as $|t| \to \infty$ for $j = 1,\ldots,m-1$ and that requirement (iii) is valid. Again it is necessary to break $U_{it}(x,y)$ into continuous and discrete parts. Consider the point spectrum part first. We have

$$G^{(j)}_p(\pm it) \equiv [(1/\pm it) e^{\pm |x-y|^2/4itq} U^p_{it}(x,y)]^{(j)}. \tag{5.25}$$

First consider the case $j \neq 0$. Derivatives with respect to t of the left two factors in the square bracket expression in Eq. (5.25) always decay at least as fast as t^{-2} for large t. Thus, if we can show that the derivatives of the third factor in the square bracket are bounded, then the integrals in (4.29), for $j \neq 0$, are always uniformly convergent for all λ and for $|x-y|$ restricted to a finite interval. We now consider the time derivatives of $[U^p_{it}(x,y)]$. By definition

$$[U^p_{it}(x,y)]^{(j)} = \sum_l (-i\lambda_l)^j e^{-i\lambda_l t} \psi_l(x) \overline{\psi_l(y)}. \tag{5.26}$$

Again estimate (3.8) shows

$$|[U^p_{it}(x,y)]^{(j)}| \leqslant (\Sigma)^j \sum_l |\psi_l(x) \overline{\psi_l(y)}| < (\Sigma)^j C(\Sigma) < \infty. \tag{5.27}$$

Thus $[U^p_{it}(x,y)]^{(j)}$ is bounded in t and series (5.26) is uniformly convergent for all $t \in \mathbb{R}$ and all $(x,y) \in \mathbb{R}^d \times \mathbb{R}^d$. The fact that the integrals

$$\int_1^\infty \frac{1}{it} e^{i\lambda t \pm i\mu/t} \sum_l (-i\lambda_l)^j e^{-i\lambda_l t} \psi_l(x) \overline{\psi_l(y)}\, dt$$

for $j = 0,1,\cdots$ are uniformly convergent for $\lambda > \Sigma + 1$ and all $(x,y) \in \mathbb{R}^d \times \mathbb{R}^d$ follows from the uniform convergence of the series (5.26) and the t-integral

$$\int_1^\infty (1/t) e^{it(\lambda - \lambda_l) \pm i\mu/t}\, dt$$

for $\lambda > \Sigma + 1$. This argument, of course, applies also to the case $j = 0$ in (5.25). We therefore conclude that $G_p(\pm it)$ satisfies condition (iii) in Theorem 3 and $G^{(j)}_p(it) = O(1)$ as $|t| \to \infty$ for $j = 1,\ldots,m-1$.

Finally we must show the uniform convergence of the continuous spectrum portion of integral (4.29). Since $H \in T_u(\gamma;N)$, the continuous spectrum term has decay estimate

$$|G^{(j)}_c(\pm it)| = |[(1/\pm it) e^{\pm |x-y|^2/4itq} U^c_{it}(x,y)]^{(j)}| \leqslant C_u |t|^{\gamma-1},$$

for $j = 0,1,\ldots,N$. Since $\gamma < 0$, the integrals in (4.29) with G replaced by G_c are absolutely and uniformly convergent for $|x-y|$ in a finite interval and all $\lambda \in \mathbb{R}$. Thus property (iii) of Theorem 3 is established.

Theorem 5 now follows from (1.9), (5.17) and the identity (Watson,[38] p. 106)

$$\frac{d^n}{dz^n} z^{\nu/2} J_\nu(a\sqrt{z}) = (\tfrac{1}{2}a)^n z^{(\nu-n)/2} J_{\nu-n}(a\sqrt{z}). \quad \square \tag{5.28}$$

The most limiting restriction placed on H in Theorem 5 is the requirement $H\in T_u(\gamma;N)$ for $\gamma<0$. In the situation where $U_{it}^c(x,y)$ has growth as $|t|\to\infty$, or the inequalities estimating the large-t behavior have growth, then we can obtain an extension of Theorem 5, provided that the growth is not faster than a polynomial in $|t|$. We have

Theorem 6: Let N,M be positive integers. Assume that the Hamiltonian H satisifes: (i) $v\in\mathcal{F}^r_{2(M+N)}$; (ii) $\sup\sigma_p(H)=\Sigma<\infty$; and (iii) for some $\gamma<\infty$, $H\in T_u(\gamma;N)$. If N and M satisfy

$$2N+d/2+\alpha+1\leqslant M<2N+d/2+\alpha+2, \tag{5.29}$$

then for all $\alpha>\gamma$ there exists a function $\Phi_N^\alpha(x,y;\lambda)$, jointly continuous in $\mathbb{R}^d\times\mathbb{R}^d$, such that

$$\lambda_+^\alpha e^\alpha(x,y;\lambda)=\sum_{n=0}^{M-1}\frac{(-1)^n}{n!}P_n(x,y)\left(\frac{d}{d\lambda}\right)^n[\lambda_+^\alpha e_0^\alpha(x,y;\lambda)]+\frac{1}{\lambda^N}\Phi_N^\alpha(x,y;\lambda) \tag{5.30}$$

and

$$\Phi_N^\alpha(x,y;\lambda)\to 0 \quad \text{as } \lambda\to\infty \tag{5.31}$$

uniformly in any compact subset of $\mathbb{R}^d\times\mathbb{R}^d$.

Proof: The argument is basically the same as in Theorem 5 provided that we replace Proposition 2 by Proposition 3 and make use of Eq. (5.14). In order to convert the expansion of the form (4.30) into (5.30), we use Eq. (5.28) to establish that

$$\left(\frac{d}{d\lambda}\right)^n\left[\frac{\lambda_+^\alpha}{\Gamma(1+\alpha)}e_0^\alpha(x,y;\lambda)\right]=\frac{1}{(4\pi q)^{d/2}}\left[\frac{2q^{1/2}\lambda^{1/2}}{|x-y|}\right]^{\alpha+d/2-n}J_{\alpha+d/2-n}(\lambda^{1/2}q^{-1/2}|x-y|). \quad \square \tag{5.32}$$

To conclude, we remark that the results in Theorem 5 and 6 remain valid if the hypothesis $H\in T_u(\gamma;N)$ is replaced by $H\in T_p(\gamma;N)$. To see this, it clearly suffices to prove that for all continuous functions ψ and ϕ with compact supports

$$\begin{aligned}\lambda_+^\alpha&\int\int dxdy\ \overline{\psi(x)}\,\phi(y)e^\alpha(x,y;\lambda)\\&=\sum_{n=0}^{M-1}\frac{(-1)^n}{n!}\int\int dxdy\ \overline{\psi(x)}\phi(y)P_n(x,y)\\&\quad\times\left(\frac{d}{d\lambda}\right)^n[\lambda_+^\alpha e_0^\alpha(x,y;\lambda)]\\&\quad+\frac{1}{\lambda^N}\int\int dxdy\ \overline{\psi(x)}\phi(y)\Phi_N^\alpha(x,y;\lambda).\end{aligned} \tag{5.33}$$

We now sketch a proof of this result. The strong convergence of Proposition 4 also implies weak convergence; thus we have

$$(\psi,E_\lambda^\alpha\phi)=\lim_{\tau,\epsilon}(\psi,A_\lambda^\alpha(\tau,\epsilon)\phi).$$

In view of (2.27), this is equivalent to

$$\begin{aligned}&\int\int dxdy\ \overline{\psi(x)}\phi(y)e^\alpha(x,y;\lambda)\\&=\frac{\Gamma(1+\alpha)}{\lambda_+^\alpha}\lim_{\tau,\epsilon}\int dx\ \overline{\psi(x)}\\&\quad\times\left[\frac{1}{2\pi i}\int_{C_{\tau,\epsilon}}\frac{e^{z\lambda}}{z^{1+\alpha}}(e^{-zH}\phi)(x)\,dz\right].\end{aligned}$$

Note that the path $C_{\tau,\epsilon}$ has finite length. Thus, by Fubini's theorem,

$$\begin{aligned}&\int\int dxdy\ \overline{\psi(x)}\phi(y)e^\alpha(x,y;\lambda)\\&=\frac{\Gamma(1+\alpha)}{\lambda_+^\alpha}\lim_{\tau,\epsilon}\frac{1}{2\pi i}\int_{C_{\tau,\epsilon}}\frac{e^{\lambda z}}{z^{1+\alpha}}h(z)\,dz,\end{aligned} \tag{5.34}$$

where

$$h(z)=\int\int dxdy\ \overline{\psi(x)}\phi(y)U_z(x,y). \tag{5.35}$$

Inserting (5.2) into (5.35) yields the small-z expansion

$$\begin{aligned}h(z)=&\left(\frac{1}{4q\pi}\right)^{d/2}\sum_{n=0}^{M-1}\frac{(-1)^n}{n!}z^{n-d/2}\int\int dxdy\\&\times\overline{\psi(x)}\phi(y)P_n(x,y)\exp\left(\frac{-|x-y|^2}{4qz}\right)+\epsilon_M(z),\end{aligned} \tag{5.36}$$

where

$$\begin{aligned}\epsilon_M(z)=&\frac{1}{(4q\pi z)^{d/2}}\int\int dxdy\ \overline{\psi(x)}\phi(y)\\&\times\exp\left(\frac{-|x-y|^2}{4qz}\right)E_M(x,y;z).\end{aligned} \tag{5.37}$$

Now we proceed as in the first part of Theorem 3. By using (5.32), we arrive at

$$\begin{aligned}&\frac{\lambda_+^\alpha}{\Gamma(1+\alpha}\int\int dxdy\ \overline{\psi(x)}\phi(y)e^\alpha(x,y;\lambda)\\&=\frac{1}{\Gamma(1+\alpha)}\sum_{n=0}^{M-1}\frac{(-1)^n}{n!}\int\int dxdy\\&\quad\times\overline{\psi(x)}\phi(y)P_n(x,y)\left(\frac{d}{d\lambda}\right)^n[\lambda_+^\alpha e_0^\alpha(x,y;\lambda)]\\&\quad+\lim_{\eta\to0}\frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{1}{(it)^{1+\alpha}}e^{-\eta|t|+i\lambda t}\epsilon_M(it)\,dt.\end{aligned} \tag{5.38}$$

Substituting (5.37) into the last integral in (5.38) and interchanging the order of integration, the remainder term in (5.38) now takes the form

$$\begin{aligned}&[1/(4q\pi)^{d/2}]\int\int dxdy\ \overline{\psi(x)}\phi(y)\\&\quad\times\left[\lim_{\eta\to0}(1/2\pi)\int_{-\infty}^{\infty}(it)^{-d/2-\alpha-1}E_M(x,y;it)\right.\\&\quad\left.\times\exp(-\eta|t|+i\lambda t+i|x-y|^2/4qt)\,dt\right].\end{aligned}$$

To the inner integral, we apply Proposition 6. The resulting expression together with (5.38) immediately gives the desired expansion (5.33).

Finally, we point out that in the above argument we have bypassed the use of Theorem 3 and have gone directly to Proposition 6. In doing so, we have avoided the uniform decay requirement on $U_{\beta+it}(x,y)$ in condition (ii) of Theorem 3. The condition $H \in T_p(\gamma;N)$ will infer that the condition on the uniform convergence of the integrals in (4.5) is satisfied.

ACKNOWLEDGMENTS

The authors are indebted to many of their colleagues for constructive advice about various aspects of this work. In particular, we have benefited from discussions with Desiré Bollé, Yoshikazu Fujiwara, Steve Fulling, Richard Froese, Barry Simon, Ralph Phillips, and Stan Wilk. Both authors acknowledge support by the Natural Sciences and Engineering Research Council, Canada.

[1]H. Weyl, Math. Ann. **68**, 220 (1910).
[2]T. Carleman, Ark. Mat. Astron. Fys. **24**, 1 (1934).
[3]L. Gårding, K. Fysiogr. Sällsk. Lund Förh. **24**(21), 1 (1954); Math. Scand. **1**, 237 (1953).
[4]E. C. Titchmarsh, *Eigenfunction Expansions* (Oxford U.P., London, 1958), Part II.
[5]B. M. Levitan, Tr. Mosk. Mat. Obsc. **4**, 237 (1955) [Am. Math. Soc. Transl. **20**, 1 (1962)]; Izv. Akad. Nauk SSSR Ser. Mat. **16**, 325 (1952) [Am. Math. Soc. Transl. **101**, 192 (1973)].
[6]J. Odhnoff, Medd. Lunds Univ. Mat. Sem. **14**, 1 (1959).
[7]G. Bergendal, Medd. Lunds. Univ. Mat. Sem. **15**, 1 (1959).
[8]L. Hörmander, Acta. Math. **121**, 193–218 (1968); in *Belfer Graduate School of Science Annual Science Conference Proceedings: Some Recent Advances in the Basic Sciences, Vol. 2 (1965–1966)*, edited by A. Gelbart (Yeshiva U.P., New York, 1969), p. 155.
[9]S. Agmon and Y. Kannai, Israel J. Math. **5**, 1 (1967).
[10]C. Clark, SIAM Rev. **9**, 627 (1967).
[11]T. A. Osborn and Y. Fujiwara, J. Math. Phys. **24**, 1093 (1983).
[12]T. A. Osborn, R. G. Froese, and S. F. Howes, Phys. Rev. Lett **45**, 1987 (1980).
[13]S. F. J. Wilk, Y. Fujiwara, and T. A. Osborn, Phys. Rev. A **24**, 2187 (1981).
[14]Y. Fujiwara, T. A. Osborn, and S. F. J. Wilk, Phys. Rev. A **25**, 14 (1982).
[15]A. M. Perelomov, Ann. Inst. H. Poincaré **24**, 161 (1976).
[16]C. Wilcox, J. Math. Phys. **6**, 611 (1964).
[17]S. Steinberg, Arch. Ratl. Mech. Anal. **54**, 1348 (1977).
[18]J. Rauch, Commun. Math. Phys. **61**, 149 (1978).
[19]A. Jensen, Manuscripta Math. **25**, 61 (1978).
[20]M. Murata, Proc. Jpn. Acad. **54**, 42 (1978).
[21]A. Jensen and T. Kato, Duke Math. J. **46**, 583 (1979).
[22]T. Schonbek, Duke Math. J. **46**, 203 (1979).
[23]A. Jensen, Duke Math. J. **47**, 57 (1980).
[24]W. O. Amrein, J. M. Jauch, and K. B. Sinha, *Scattering Theory in Quantum Mechanics* (Benjamin, Reading, MA, 1977), Chap. 3.
[25]B. Simon, Bull. Am. Math. Soc. **7**, 447 (1982).
[26]D. Fujiwara, J. Anal. Math. **35**, 41 (1979).
[27]H. Kitada, J. Fac. Sci. Univ. Tokyo A **2**, 193 (1980); H. Kitada and H. Kumano-go, Osaka J. Math. **18**, 291 (1981).
[28]S. Zelditch, "Reconstruction of Singularities for Solutions of Schrödinger's Equation," Ph.D. thesis, Univ. of California, Berkeley, 1981.
[29]A. Povzner, Mat. Sb. **32**, 109 (1953) [Am. Math. Soc. Transl. Ser 2 **60**, 1 (1967)].
[30]L. Gårding, Comptes Rendus du Douzième Congrès des Mathématiciens Scandinaves, Lund, 1953, pp. 44–55.
[31]K. B. Sinha, Ann. Inst. H. Poincaré, Sec. A **26**, 263 (1977).
[32]J. Weidman, Bull. Am. Math. Soc. **73**, 452 (1967); S. Albeverio, Ann. Phys (N.Y.) **71**, 167 (1972); B. Simon, Math. Ann. **207**, 133 (1974).
[33]F. W. Olver, SIAM J. Math. Anal. **5**, 19 (1974).
[34]F. W. Olver, *Asymptotics and Special Functions* (Academic, New York, 1974).
[35]E. C. Titchmarsh, *The Theory of Functions* (Oxford U.P., London, 1939), 2nd ed.
[36]I. S. Gradshteyn and I. M. Ryzhik, *Tables of Integrals, Series and Products* (Academic, New York, 1965).
[37]A. Erdélyi, *et al.*, *Tables of Integrals Transforms* (McGraw-Hill, New York, 1954), Vol. 1.
[38]G. N. Watson, *A Treatise on the Theory of Bessel Functions* (Cambridge U.P., Cambridge, 1958), 2nd ed.
[39]M. Heins, *Selected Topics in the Classical Theory of Functions of a Complex Variable* (Holt, Rinehart, and Winston, New York, 1962), p. 135.

Inverse scattering theory for systems of the Zakharov–Shabat type

Helena M. A. de Castro
Instituto de Matemática, Universidade de São Paulo, São Paulo, Brazil

Walter F. Wreszinski
Instituto de Física, Universidade de São Paulo, São Paulo, Brazil

(Received 28 May 1982; accepted for publication 22 October 1982)

We present a mathematically rigorous inverse scattering theory for systems of the Zakharov–Shabat type. A qualitative spectral analysis of a class of non-self-adjoint operators, which is necessary for the construction, is also provided.

PACS numbers: 03.65.Nk, 02.30.Hq

INTRODUCTION AND SYNOPSIS

The application of inverse scattering theory to nonlinear equations of evolution is a subject of great interest, both physical and mathematical (see, for instance, Ref. 1 and references given there). In recent times, a mathematically rigorous inverse scattering theory for the Korteweg–de Vries (KdV) equation has been constructed.[2–4] This equation has the simplifying feature that the associated linear operator is self-adjoint. That is not the case for most of the other nonlinear evolution equations of physical interest. A global approach to treat the latter which includes the KdB equation as well, but as a somewhat singular limit, was developed by Ablowitz, Kaup, Newell, and Segur in Ref. 5 (hereafter referred to as AKNS) based in part on previous work by Zakharov and Shabat[6] on the nonlinear Schrödinger equation.

The AKNS theory has a number of mathematical disadvantages, however. The most serious one in our opinion, is the fact that the authors do not *prove* the isospectral property of the flow, but instead assume it to be true, and derive from this assumption a set of consistency relations, which are finally seen to be equivalent to the validity of a class of nonlinear equations. The main difficulty is that the operators forming a "Lax pair" (see, e.g., Ref. 1) depend in most cases on the eigenvalue itself. There exist, however, in several important cases,[7,6] Lax pairs without the latter defect, and for which the isospectral property may be proved.[8] We shall therefore assume henceforth that we are dealing with operators such as those considered in Refs. 1, 6, and 7, which we call for short Zakharov–Shabat-type operators.

The remaining analysis of Ref. 5 seems to us very elegant and natural. Their assumptions remain, however, unproved. In Sec. I we provide a qualitative spectral analysis of Zakharov–Shabat operators L, with exponentially decreasing "potentials" $q_{\pm}$ [see (1) of the next section]. We prove that under this assumption on $q_{\pm}$ the number of discrete eigenvalues of L is finite, which is an important point in the analysis of Ref. 5. We expect that a strong condition of this type is also necessary, because in Ref. 9 (see also the Appendix to Ref. 10) it was proven that the condition

$$\sup_{0\leqslant x<\infty} |p(x)|\exp(\epsilon\sqrt{x}) < \infty, \quad \epsilon > 0,$$

is necessary for the finiteness of the number of discrete eigenvalues of the non-self-adjoint Schröedinger operator $Ly \equiv -d^2y/dx^2 + p(x)y$ on $L^2(0,\infty)$ with boundary condition $y(0) = 0$. We are unable, however, to prove the absence of multiple zeroes of the Wronskians such as $a(\cdot)$ [see (10) of Sec. I for the definition), in contrast to the self-adjoint case (Ref. 3, p. 148). Some other properties assumed in Ref. 5 concerning the spectrum of L are also presented in Sec. I. We refrain from a deeper study of the spectral properties of L (which is probably feasible along the lines of Lyantse's appendix to Ref. 10; see also Ref. 11) because it is not required by inverse scattering theory, at least from the point of view of Ref. 5 and of Marchenko equations.[12]

In Sec. II we consider inverse scattering theory in detail. Many of the methods of Refs. 2 and 3 are applicable, and therefore some results are only stated or sketched. A difference lies in the problem of existence and unicity of the Marchenko equations. The solution of this problem in Ref. 2 makes essential use of unitarity of the scattering function, which does not hold in the non-self-adjoint case. The proof in Proposition 3 uses analyticity properties which follow from the assumption of exponential falloff. See also the remark following the proof of Proposition 3.

Finally, we remark that in some references 4×4 matrix operators are involved,[7] but the methods developed in Sec. I and II are applicable, because the structure of the operators is, *mutadis mutandis*, the same as (1) of Sec. I.

A more complete and general spectral analysis of the same class of potentials considered in this paper has been undertaken in Ref. 13. Although there is some overlap, particularly in the part concerning Jost functions, the emphasis is entirely different, and results which are more relevant for inverse scattering theory may be obtained with our more restricted class.

I. SPECTRAL ANALYSIS OF ZAKHAROV–SHABAT-TYPE OPERATORS

We consider the following operators on $\mathscr{H} = L^2(\mathbb{R},dx) \oplus L^2(\mathbb{R},dx)$

$$L(t) = L\begin{pmatrix} \partial/\partial x & -q_+(x,t) \\ q_-(x,t) & -\partial/\partial x \end{pmatrix} \tag{1}$$

with the condition

$$|q_{\pm}(x,0)| \leqslant C\exp(-\epsilon|x|), \quad \epsilon > 0, \quad \forall x\in\mathbb{R}, \tag{2a}$$

$$q_{\pm}(\cdot,0)\in C^{\infty}(\mathbb{R}). \tag{2b}$$

We wish to analyse the spectrum $\sigma(L(0))$ of $L(0)\equiv L$.

In Ref. 8 we proved that $\sigma_d(L(0)) = \sigma_d(L(t))$, where σ_d

 0022-2488/83/061502-07$02.50 © 1983 American Institute of Physics

denotes discrete spectrum, for a class of operators of type (1), and the results of the present section will be seen to imply that $\sigma(L(t)) = \sigma(L(0))$ [$\sigma(A)$ denotes the spectrum of A for any operator A on $\mathscr{H}$].

We write $L = L_0 + A$, where

$$L_0 \equiv i\begin{pmatrix} \partial_x & 0 \\ 0 & -\partial_x \end{pmatrix}, \quad \partial_x = \frac{\partial}{\partial_x}, \quad \text{and} \quad A \equiv i\begin{pmatrix} 0 & -q_+ \\ q_- & 0 \end{pmatrix}.$$

The following lemma is standard:

Lemma 1: L_0 is self-adjoint on $D(L_0) = D_p \oplus D_p$, where

$$D_p \equiv \{u \text{ a.c.} | u' \in L^2(\mathbb{R}, dx)\}$$

is the domain of $p = -i\partial_x$ on $L^2(\mathbb{R}, dx)$, and $\sigma(L_0) = \sigma_{\text{a.c.}}(L_0) = \sigma_e(L_0) = \mathbb{R}$, where σ_e denotes essential spectrum and a.c. stands for absolutely continuous.

Lemma 2: $\sigma_e(L) = \sigma_e(L_0) = \mathbb{R}$. Hence the complement $\sigma(L)/\mathbb{R}$ of $\mathbb{R}$ in $\sigma(L)$ consists of isolated point eigenvalues of finite multiplicities which may at most accumulate at $\mathbb{R}$.

Proof: $\|A(L_0 - \zeta)^{-1}\|_{\text{HS}} < \infty \ \forall \zeta \notin \mathbb{R}$, where $\|\cdot\|_{\text{HS}}$ is the Hilbert–Schmidt norm, since, by (2), $\|q_\pm\|_2 < \infty$ and $\|A(L_0 - \zeta)^{-1}\|_{\text{HS}}^2 \leqslant \int_{-\infty}^{\infty} dp \int_{-\infty}^{\infty} dp' |\tilde{q}_\pm (p - p')|^2 |(p' - \zeta)|^{-2} = \|q_\pm\|_2^2 \int_{-\infty}^{\infty} dp \, |(p - \zeta)|^{-2} < \infty$. Hence A is L_0-compact, and the result follows from Ref. 4, Theorem 5.35, Chap. IV, p. 244. ∎

We now define the open region

$$O_\epsilon = \{\zeta \mid |\text{Im}\zeta| < \epsilon/4\}. \tag{3}$$

The eigenvalue equation

$$Lv = \zeta v \tag{4}$$

shows that $v \in C^\infty$. The transformation

$$v = \begin{pmatrix} e^{-i\zeta x} & 0 \\ 0 & e^{i\zeta x} \end{pmatrix} u$$

transforms (4) into the equivalent equation

$$\partial_x \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} + \begin{pmatrix} 0 & -e^{2i\zeta x} & q_+ \\ e^{-2i\zeta x} & q_- & 0 \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} = 0.$$

We now use the following theorem (Ref. 15, Theorem XI.65, p. 159):

Theorem: Let $A(x)$ be a continuous function from $[R_0, \infty)$ to $L(X)$, the bounded operators on a finite-dimensional normed linear pace (R_0 is some finite real number). Assume that $\|A(x)\| \in L^1(R_0, \infty)$. Suppose that $u_0 \in X$. Then there exists a unique continuously differentiable function $u(x, u_0)$ so that

$$\frac{du}{dx} = A(x)u(x)$$

and $\lim_{x \to \infty} u(x) = u_0$. Every nonzero solution has a nonzero limit as $x \to \infty$. ∎

The above theorem proves the existence and uniticity of the following "Jost functions" in O_ϵ:

$$\begin{matrix} & x \to +\infty & & x \to -\infty \end{matrix}$$

$$\Phi: \quad \begin{pmatrix} a(\zeta) e^{-i\zeta x} \\ b(\zeta) e^{i\zeta x} \end{pmatrix}, \quad \begin{pmatrix} 1 \\ 0 \end{pmatrix} e^{-i\zeta x}, \tag{5}$$

$$\Psi: \quad \begin{pmatrix} 0 \\ 1 \end{pmatrix} e^{i\zeta x}, \quad \begin{pmatrix} -\bar{b}(\zeta) e^{-i\zeta x} \\ a(\zeta) e^{i\zeta x} \end{pmatrix}, \tag{6}$$

$$\bar{\Phi}: \quad \begin{pmatrix} \bar{b}(\zeta) e^{-i\zeta x} \\ -\bar{a}(\zeta) e^{i\zeta x} \end{pmatrix}, \quad -\begin{pmatrix} 0 \\ 1 \end{pmatrix} e^{i\zeta x}, \tag{7}$$

$$\bar{\Psi}: \quad \begin{pmatrix} 0 \\ 1 \end{pmatrix} e^{-i\zeta x}, \quad \begin{pmatrix} -\bar{a}(\zeta) e^{-i\zeta x} \\ -b(\zeta) e^{i\zeta x} \end{pmatrix}. \tag{8}$$

By Theorem 1 these functions are continuously differentiable and by the eigenvalue equation they are in fact C^∞. a, b, $\bar{a}$, $\bar{b}$ are defined (for $\zeta \in \mathbb{R}$ and then extended to O_ϵ) by the formulae

$$\Phi(x, \zeta) = a(\zeta)\bar{\Psi}(x, \zeta) + b(\zeta)\Psi(x, \zeta), \tag{9a}$$

$$\bar{\Phi}(x, \zeta) = \bar{b}(\zeta)\bar{\Psi}(x, \zeta) - \bar{a}(\zeta)\Psi(x, \zeta), \tag{9b}$$

which follow from the theorem upon picking a basis for X and, using local uniqueness, a, b, $\bar{a}$, $\bar{b}$ may be expressed as Wronskians; e.g.,

$$a(\zeta) = \Phi_1(x, \zeta)\Psi_2(x, \zeta) - \Phi_2(x, \zeta)\Psi_1(x, \zeta). \tag{10}$$

By the eigenvalue equation, a is independent of x.

Lemma 3: If $Lv = \zeta_0 v$ with $\text{Im}\zeta_0 > 0$, then $a(\zeta_0) = 0$. If $Lv = \zeta_0 v$ with $\text{Im}\zeta_0 < 0$, then $\bar{a}(\zeta_0) = 0$.

Proof: Assume $a(\zeta_0) \neq 0$ and that $Lv = \zeta_0 v$, $v \in D(L)$. It follows that $v \in C^\infty$ (therefore locally bounded) and hence

$$v = \alpha\Phi + \beta\Psi$$

because Φ and Ψ are linearly independent (as in Theorem 1 with α, β independent of x). Assume $\alpha \neq 0$. Then

$$e^{i\zeta_0 x} v(x, \zeta) \underset{x \to +\infty}{\to} \begin{pmatrix} a(\zeta) \\ 1 \end{pmatrix},$$

which contradicts

$$\lim_{x \to +\infty} \sup |v_i(x, \zeta)| < \infty, \quad i = 1, 2,$$

which is necessary for $v \in \mathscr{H}$. Similarly $\beta \neq 0$ leads to a contradiction. Hence $v \equiv 0$, and ζ_0 is not an eigenvalue. The second assertion is analogous. ∎

Lemma 4: Eigenvalues ζ_0 for which $\text{Im}\zeta_0 > 0$ are simple, as well as those ζ_0 such that $\text{Im}\zeta_0 < 0$.

Proof: Let $\text{Im}\zeta_0 > 0$, and let v, w be eigenvectors of L corresponding to ζ_0. It follows that $v_1 w_2 - v_2 w_1 = \text{const}$ (independent of x). The constant must be zero in order that

$$\lim_{|x| \to \infty} \sup \begin{Bmatrix} v_i(x) \\ w_i(x) \end{Bmatrix} = 0,$$

which is necessary for $v, w \in \mathscr{H}$. Hence $v = \lambda w$ for some $\lambda \in \mathbb{C}$. Similarly for $\text{Im}\zeta_0 < 0$. ∎

Proposition 1: $\sigma_d(L)$ is under assumption (3) a finite set.

Proof: The eigenvalue equation for Φ and Ψ is equivalent to the following Volterra integral equations:

$$\Phi(\zeta, x) = \begin{pmatrix} 1 \\ 0 \end{pmatrix} e^{-i\zeta x} + \int_{-\infty}^{x} dy \begin{pmatrix} 0 & e^{-i\zeta(x-y)} & q_+(y) \\ e^{i\zeta(x-y)} & q_-(y) & 0 \end{pmatrix} \Phi(\zeta, y), \tag{11a}$$

$$\Psi(\zeta, x) = \begin{pmatrix} 0 \\ 1 \end{pmatrix} e^{i\zeta x} - \int_{x}^{\infty} dy \begin{pmatrix} 0 & e^{-i\zeta(x-y)} & q_+(y) \\ e^{i\zeta(x-y)} & q_-(y) & 0 \end{pmatrix} \Psi(\zeta, y). \tag{12a}$$

We define

$$h(\zeta,x)\equiv e^{i\zeta x}\Phi(\zeta,x),\quad g(\zeta,x)\equiv e^{-i\zeta x}\Psi(\zeta,x).$$

By (10),

$$a(\zeta)=h_1(\zeta,x)g_2(\zeta,x)-h_2(\zeta,x)g_1(\zeta,x). \tag{13}$$

From (11) and (12) we obtain equations for h and g:

$$h(\zeta,x)=\begin{pmatrix}1\\0\end{pmatrix}+\int_{-\infty}^{x}dy\begin{pmatrix}0 & q_+(y)\\ e^{2i\zeta(x-y)} & (q-y)\quad 0\end{pmatrix}h(\zeta,y), \tag{11b}$$

$$g(\zeta,x)=\begin{pmatrix}0\\1\end{pmatrix}-\int_{x}^{\infty}dy\begin{pmatrix}0 & e^{-2i\zeta(x-y)} & q_+(y)\\ q_-(y) & & 0\end{pmatrix}g(\zeta,y). \tag{12b}$$

Let $h^{(n)}$ be the nth term in the Volterra iteration for h. Then by (2)

$$|h_i^{(n)}(\zeta,x)|\leqslant(C^n/n!)\left(\int_{-\infty}^{x}e^{-(\epsilon-|2\mathrm{Im}\zeta|)|y|}\right)^n,\quad i=1,2,$$

and from (3)

$$\sup_{i=1,2}\sup_{x\in\mathbf{R}}|h_i^{(n)}(\zeta,x)-h_i^{(m)}(\zeta,x)|$$
$$\leqslant(C^m/m!)\exp(4C/\epsilon),\quad n\geqslant m.$$

Hence the Volterra series for each component of h converges uniformly in x and ζ (for $\zeta\in\bar{O}_\epsilon$), and h_i is therefore analytic in O_ϵ. The same argument works for g_i, and hence $a(\cdot)$ is holomorphic in O_ϵ by (13).

The equation for h_1 is

$$h_1(\zeta,x)=1+\int_{-\infty}^{x}dy_1\int_{-\infty}^{y_1}dy_2\,q_+(y_1)q_-(y_2)e^{2i\zeta(y_1-y_2)}h_1(\zeta,y_2).$$

By the Riemann–Lebesgue lemma, dominated convergence and a similar argument as above, we obtain

$$\lim_{|\mathrm{Re}\zeta|\to\infty}h_1(\zeta,x)=1.$$

Similarly,

$$\lim_{|\mathrm{Re}\zeta|\to\infty}g_2(\zeta,x)=1,$$

$$\lim_{|\mathrm{Re}\zeta|\to\infty}h_2(\zeta,x)=0,$$

$$\lim_{|\mathrm{Re}\zeta|\to\infty}g_1(\zeta,x)=0.$$

Hence, by (10)

$$\lim_{|\mathrm{Re}\zeta|\to\infty}a(\zeta)=1. \tag{14}$$

Finally, Lemma 2, Lemma 3, analyticity of a in O_ϵ, and (14) imply Proposition 1. ■

Lemma 5: $h_i(\zeta,x)$ and $g_i(\zeta,x)$, $i=1,2$, are analytic in the region $\mathrm{Im}\zeta>0$. Moreover,

$$\sup_{i=1,2}\sup_{\zeta:\mathrm{Im}\zeta>0}\sup_{x\in\mathbf{R}}|h_i(\zeta,x)|<\infty,$$

$$\sup_{i=1,2}\sup_{\zeta:\mathrm{Im}\zeta>0}\sup_{x\in\mathbf{R}}|g_i(\zeta,x)|<\infty.$$

Proof: The proof follows from (11b) and (12b) along the lines of Proposition 1. ■

Lemma 6: If $\mathrm{Im}\zeta_0>0$ (resp. $\mathrm{Im}\zeta_0<0$) and $a(\zeta_0)=0$ (resp. $\bar{a}(\zeta_0)=0$), then $\zeta_0\in\sigma_d(L)$.

Proof: Let $a(\zeta_0)=0$ with $\mathrm{Im}\zeta_0>0$. Then $\Phi(\zeta_0,\cdot)=\alpha\Psi(\zeta_0,\cdot)$ for some $\alpha\neq0$. ($\alpha=0$ would contradict the boundary conditions.) Now

$$\Psi(\zeta_0,x)\underset{x\to\infty}{\to}\begin{pmatrix}0\\1\end{pmatrix}e^{i\zeta_0x},\quad \Psi(\zeta_0,x)\underset{x\to-\infty}{\to}\frac{1}{\alpha}\begin{pmatrix}1\\0\end{pmatrix}e^{-i\zeta_0x},$$

$$\Phi(\zeta_0,x)\underset{x\to-\infty}{\to}\begin{pmatrix}1\\0\end{pmatrix}e^{-i\zeta_0x},\quad \Phi(\zeta_0,x)\underset{x\to\infty}{\to}\alpha\begin{pmatrix}0\\1\end{pmatrix}e^{-i\zeta_0x},$$

whence $\Phi_1(\zeta_0,\cdot)\in L^2((-\infty,0),dx)$ and $\Phi_2(\zeta_0,\cdot)\in L^2((0,\infty),dx)$. Now, by Lemma 5,

$$\Psi_1(\zeta_0,x)=e^{i\zeta_0x}g_1(\zeta_0,x)\in L^2((0,\infty),dx)$$

and hence

$$\Phi_1(\zeta_0,\cdot)=\alpha\Psi_1(\zeta_0,\cdot)\in L^2((0,\infty),dx).$$

Similarly, $\Phi_2(\zeta_0,\cdot)\in L^2((-\infty,0),dx)$. Hence

$$\Phi=(1/\alpha)\Psi\in\sigma_d(L).$$ ■

Lemma 6 shows that the set of zeroes of $a(\cdot)$ in $\mathrm{Im}\zeta>0$ is identical with the set of discrete eigenvalues of L (analogous assertions for $\mathrm{Im}\zeta>0$). We have as a corollary of Lemma 5 and Proposition 1:

Proposition 2: $a(\cdot)$ is analytic in ζ for $\mathrm{Im}\zeta>0$. The set of zeroes of a in $\{\zeta:\mathrm{Im}\zeta\geqslant0\}$ is finite. Similarly for $\bar{a}(\cdot)$. ■

For inverse scattering theory we need an analysis of the real zeroes of $a(\cdot)$.

Lemma 7: $\sigma_d(L)\cap\mathbb{R}=\emptyset$.

Proof: Let $\xi\in\mathbb{R}$. Locally bounded solutions of $Lv=\xi v$ are of the form $v=\alpha\Psi+\beta\Psi$ because $\Psi(\xi,\cdot)$ and $\Psi(\xi,\cdot)$ are linearly independent. If $v\not\equiv0$, then

$$v\underset{x\to\infty}{\to}\begin{pmatrix}\alpha e^{i\xi x}\\ \beta e^{i\xi x}\end{pmatrix}$$

and hence $v\notin\mathscr{H}$. ■

Lemma 8: The set $\{\xi\in\mathbb{R}|a(\xi)=0\}$ is in general not empty, but it is a finite set.

Proof: Let $q_-=\alpha q_+$. The Volterra series for h, g may be summed at $\zeta=0$ yielding[5]

$$a(0)=\bar{a}(0)=\cosh(\sqrt{\alpha}p)$$

where $p\equiv\int_{-\infty}^{\infty}q_+(x)\,dx$. One obtains thus a real zero at $\zeta=0$ if

$$\sqrt{\alpha}\int_{-\infty}^{\infty}q_+(x)dx=i(2n+1)\pi/2.$$

These real zeroes of $a(\cdot)$ may be interpreted as "resonances" of L and play an important role in inverse scattering theory.

That the set of real zeroes of $a(\cdot)$ is finite follows from analyticity of $a(\cdot)$ in O_ϵ and (14). ■

Lemma 9: The set of zeroes of $a(\cdot)$ [resp. $\bar{a}(\cdot)$] is time-independent.

Proof: The proof follows from Theorem 1 of Ref. 8 for $\mathrm{Im}\zeta>0$ ($\mathrm{Im}\zeta<0$). Fpr $\zeta\in\mathbb{R}$ it may be constructed along the lines of the same theorem choosing the Banach space $B=L^\infty(\mathbb{R},dx)\oplus L^\infty(\mathbb{R}.dx)$ instead of $L^2(\mathbb{R},dx)\oplus L^2(\mathbb{R},dx)$. ■

In the self-adjoint case, it may be proved that the zeroes of $a(\cdot)$ are simple (Ref. 3, p. 148). No such general proof exists for the non-self-adjoint case. Furthermore, the existence of spectral singularities [real zeroes of $a(\cdot)$], which has no ana-

log in the non-self-adjoint case, complicates inverse scattering theory considerably. In principle, however, modifications due to multiple zeroes (real or nonreal) of $a(\cdot)$ are easy to take into account (see Ref. 5 and the following). The main features, without which the theory would have to be qualitatively modified, namely, finiteness in number and time independence of the real and nonreal zeroes of $a(\cdot)$ [$a(\cdot)$] are taken care of by the preceding results.

II. INVERSE SCATTERING THEORY FOR ZAKHAROV-SHABAT SYSTEMS

Define

$$m_\Phi(\zeta,x)\equiv\Phi(\zeta,x)e^{i\zeta x}-\begin{pmatrix}1\\0\end{pmatrix}, \tag{15a}$$

$$m_\Psi(\zeta,x)\equiv\Psi(\zeta,x)e^{-i\zeta x}-\begin{pmatrix}0\\1\end{pmatrix}, \tag{15b}$$

$$m_{\bar\Phi}(\zeta,x)=\bar\Phi(\zeta,x)e^{-i\zeta x}+\begin{pmatrix}0\\1\end{pmatrix}, \tag{15c}$$

$$m_{\bar\Psi}(\zeta,x)\equiv\bar\Psi(\zeta,x)e^{i\zeta x}-\begin{pmatrix}1\\0\end{pmatrix}. \tag{15d}$$

The following asymptotic representations may be obtained along the lines of Proposition 1 (see also Ref. 5) ($\mathrm{Im}\zeta\geqslant 0$):

$$h_1(\zeta,x)=1-(1/2i\zeta)\int_{-\infty}^{x}q_+(x')q_-(x')\,dx'+O(1/\zeta^2), \tag{16a}$$

$$h_2(\zeta,x)=-(1/2i\zeta)q_-(x)+O(1/\zeta^2). \tag{16b}$$

From these and analogous results for the other functions it follows that

Lemma 10: m_Φ and m_Ψ belong to the Hardy space $H^{2+}=\{h$ analytic in $\mathrm{Im}\zeta>0$ such that $\sup_{b>0}\int_{-\infty}^{\infty}|h(a+ib)|^2\,da<\infty\}$ and $m_{\bar\Phi}$ and $m_{\bar\Psi}$ belong to $H^{2-}=\{h$ analytic in $\mathrm{Im}\zeta<0$ such that $\sup_{b>0}\int_{-\infty}^{\infty}|h(a-ib)|^2\,da<\infty\}$. Hence the following integral representations [in the sense of $L^2(\mathbb{R},dy)$ for each $x\in\mathbb{R}$] exist and are unique:

$$m_\Psi(\zeta,x)=\int_0^\infty K^0(x,y)e^{2i\zeta y}\,dy,\quad \mathrm{Im}\zeta>0, \tag{17a}$$

$$m_{\bar\Psi}(\zeta,x)=\int_0^\infty \bar K^0(x,y)e^{-2i\zeta y}\,dy,\quad \mathrm{Im}\zeta<0, \tag{17b}$$

$$m_\Phi(\zeta,x)=-\int_{-\infty}^0 L^0(x,y)e^{-2i\zeta y}\,dy,\quad \mathrm{Im}\zeta>0, \tag{17c}$$

$$m_{\bar\Phi}(\zeta,x)=-\int_0^\infty \bar L^0(x,y)e^{2i\zeta y}\,dy,\quad \mathrm{Im}\zeta<0. \tag{17b}$$

Further, by the Volterra integral equations satisfied by the kernels $K^0,\bar K^0,L^0,\bar L^0$, it follows from (2) that each of them is C^∞ with respect to x and y and satisfy the estimates

$$\left|\overset{(-)}{K}{}^0(x,y)\right|\leqslant\text{const}\ \exp[-\epsilon'(x+y)],\quad x,y\geqslant 0, \tag{18a}$$

$$\left|\overset{(-)}{L}{}^0(x,y)\right|\leqslant\text{const}\ \exp[\epsilon'(x+y)],\quad x,y\leqslant 0, \tag{18b}$$

for some $\epsilon'>0$. In particular, the boundary values on the real axis in (17), which exists in the L^2 sense, may be taken pointwise. The kernels are also seen to obey the following boundary conditions:

$$K_1^0(x,0)=-q_+(x)=-L_1^0(x,0), \tag{19a}$$

$$K_2^0(x,0)=\int_x^\infty q_+(y)q_-(y)\,dy=\bar K_1^0(x,0), \tag{19b}$$

$$L_1^0(x,0)=-\bar L_2^0(x,0)=\int_{-\infty}^x q_+(y)q_-(y)\,dy, \tag{19c}$$

$$L_2^0(x,0)=\bar K_2^0(x,0)=q_-(x). \tag{19d}$$

Proof: The proof is identical to (and simpler than) the proof of analogous results in Ref. 3. ■

Remark: $\S K^{0(-)}(x,y)=\frac{1}{2}K^{(-)}(x,(y-x)/2)$ is the relation between our $L^{0(-)}(x,y)=\frac{1}{2}L^{(-)}(x,(y-x)/2)$ kernels and those of Ref. 5.

We now consider the inverse scattering theory, with the simplifying assumption that $a(\cdot)$ has N simple zeroes $\zeta_j\,|\,\mathrm{Im}\zeta_j>0, j=1,\ldots,N$, and one simple real zero $\zeta_0=\xi_0$. The case of a finite number of simple real zeroes is, of course, identical. Multiple zeroes imply obvious changes; see Ref. 5. We also assume that $\bar a(\cdot)$ has M simple zeroes $\bar\zeta_j\,|\,\mathrm{Im}\bar\zeta_j<0$ and a simple real zero $\bar\zeta_0$.

Let

$$f(\zeta,\zeta',x)\equiv\frac{e^{i\zeta' x}\Phi(\zeta',x)-a(\zeta')\binom{1}{0}}{a(\zeta')(\zeta'-\zeta)},$$

$$\mathrm{Im}\zeta>0,\quad \mathrm{Im}\zeta'\geqslant 0,\quad \text{but}\quad \zeta'\neq\xi_j, j=0,\ldots,N.$$

By (9), if $\zeta'\in O_\epsilon$, $\zeta'\neq\xi_j, j=0,\ldots,N$,

$$f(\zeta,\zeta',x)=\frac{m_{\bar\Psi}(\zeta',x)}{\zeta'-\zeta}+\frac{b(\zeta')}{a(\zeta')}\Psi(\zeta',x)e^{i\zeta' x}.$$

By (16) and analogous formulas for g (Ref. 5), $a(\zeta)\to_{|\zeta|\to\infty}1$ and $|[e^{i\zeta x}\Phi(\zeta,x)-a(\zeta)\binom{1}{0}]|=O(1/|\zeta|)$. Further, $|b(\zeta)|=O(1/|\zeta|^2)$ for $\zeta\in O_\epsilon$ and $|\Psi(\zeta,x)e^{i\zeta x}|=O(1)$. Let $R>\max_{j=0,\ldots,N}|\xi_j|$ and Γ_R be the contour

$$[-R,\xi_0-\epsilon]\cup[\xi_0+\epsilon,R\,]\cup\gamma_R\equiv\{Re^{i\theta};0<\theta<\pi\}$$

in the complex plane. Then, by Cauchy's theorem and (9)

$$\int_{\Gamma_R}f(\zeta,\zeta',x)\,d\zeta'=\left(\int_{-R}^{\xi_0-\epsilon}+\int_{\xi_0+\epsilon}^{R}\right)d\zeta'\left[\frac{m_{\bar\Psi}(\zeta',x)}{\zeta'-\zeta}+\frac{b(\zeta')}{a(\zeta')}\Psi(\zeta',x)e^{i\zeta' x}\right]+\int_{\gamma_R}f(\zeta,\zeta',x)\,d\zeta'$$

$$=2\pi i\sum_{j=0}^{N}\delta_j\frac{e^{i\zeta_j x}\Phi(\zeta_j,x)}{\dot a(\zeta_j)(\zeta_j-\zeta)},\quad \delta_j=\begin{cases}\frac{1}{2} & \text{if } j=0\\ 1 & \text{if } j\geqslant 1.\end{cases}$$

By the previous asymptotic relations, all the integrals are well defined, as improper complex Riemann integrals in the limit as $R\to\infty$ and

$$\lim_{R\to\infty}\int_{\gamma_R}f(\zeta,\zeta',x)\,d\zeta'=0.$$

Hence for any $\epsilon>0$

$$2\pi i\sum_{j=0}^{N}\delta_j\frac{e^{i\zeta_j x}\Phi(\zeta_j,x)}{\dot a(\zeta_j)(\zeta_j-\zeta)}=\left(\int_{-\infty}^{\xi_0-\epsilon}+\int_{\xi_0+\epsilon}^{\infty}\right)d\zeta'\left[\frac{m_{\bar\Psi}(\zeta',x)}{\zeta'-\zeta}+\frac{b(\zeta')}{a(\zeta')}\Psi(\zeta',x)e^{i\zeta' x}\right]. \tag{20}$$

Considering now the contour

$$\Gamma'_R = [-R,\zeta_0-\epsilon]\cup[\zeta_0+\epsilon,R]\cup\gamma'_R \equiv\{Re^{i\theta},\ 0<\theta<-\pi\},$$

we have

$$\lim_{R\to\infty}\int_{\gamma'_R} d\zeta'\,\frac{m_{\overline{\Psi}}(\zeta',x)}{\zeta'-\zeta} = \int_{-\infty}^{\infty} d\zeta'\,\frac{m_{\overline{\Psi}}(\zeta',x)}{\zeta'-\zeta} = -2\pi i m_{\overline{\Psi}}(\zeta,x)$$

by Cauchy's theorem. Taking now the limit of the rhs of (20), we obtain

$$\overline{\Psi}(\zeta,x)e^{i\zeta x} = \begin{pmatrix}1\\0\end{pmatrix} - \sum_{j=0}^{N}\delta_j\frac{b_j}{a(\zeta_j)}\frac{\Psi(\zeta_j,x)e^{i\zeta_j x}}{\zeta_j-\zeta} + \frac{1}{2\pi i}\,\text{p.v.}\int_{-\infty}^{\infty}\frac{b(\zeta')}{a(\zeta')}\frac{\Psi(\zeta',x)e^{i\zeta' x}}{\zeta'-\zeta}\,d\zeta',\quad \text{Im}\zeta<0, \tag{21a}$$

where $b_j>0$, $1=0,\ldots,N$ are defined by $\Phi(\zeta_j,x)=b_j\Psi(\zeta_j,x)$ {the "norming constants," independent of x [because $a(\cdot)$ is independent of x], and the principal value is well defined because $\dot{a}(\zeta_0)\neq 0$ by assumption and both b and Ψ are C^∞}. Analogous relations may be obtained:

$$\overline{\Phi}(\zeta,x)e^{-i\zeta x} = -\begin{pmatrix}0\\1\end{pmatrix} + \sum_{j=0}^{N}\delta_j\frac{\Phi(\zeta_j,x)e^{-i\zeta_j x}}{b_j\dot{a}(\zeta_j)(\zeta_j-\zeta)} - \frac{1}{2\pi i}\,\text{p.v.}\int_{-\infty}^{\infty}\frac{\bar{b}(\zeta')}{a(\zeta')}\frac{\Phi(\zeta',x)e^{-i\zeta' x}}{\zeta'-\zeta}\,d\zeta',\quad \text{Im}\zeta<0, \tag{21b}$$

$$\Psi(\zeta,x)e^{-i\zeta x} = \begin{pmatrix}0\\1\end{pmatrix} + \sum_{j=0}^{M}\delta_j\frac{\bar{b}_j}{\dot{\bar{a}}(\bar{\zeta}_j)}\frac{\overline{\Psi}(\bar{\zeta}_j,x)}{\bar{\zeta}_j-\zeta}e^{-i\bar{\zeta}_j x} + \frac{1}{2\pi i}\,\text{p.v.}\int_{-\infty}^{\infty}\frac{\bar{b}(\zeta')}{\dot{\bar{a}}(\zeta')}\frac{\overline{\Psi}(\zeta',x)e^{-i\zeta' x}}{\zeta'-\zeta}\,d\zeta',\quad \text{Im}\zeta<0, \tag{21c}$$

$$\Phi(\zeta,x)e^{i\zeta x} = \begin{pmatrix}1\\0\end{pmatrix} - \sum_{j=0}^{M}\delta_j\frac{1}{\bar{b}_j\dot{\bar{a}}(\zeta_j)}\times\frac{\overline{\Phi}(\bar{\zeta}_j,x)e^{i\bar{\zeta}_j x}}{\bar{\zeta}_j-\zeta} - \frac{1}{2\pi i}\,\text{p.v.}\int_{-\infty}^{\infty}\frac{b(\zeta')}{\bar{a}(\zeta')}\frac{\overline{\Phi}(\zeta',x)e^{i\zeta' x}}{\zeta'-\zeta}\,d\zeta',\quad \text{Im}\zeta>0. \tag{21d}$$

In (21), $\bar{b}_j$ are analogously defined by $\overline{\Phi}(\bar{\zeta}_i,x)=\bar{b}_j\overline{\Psi}(\bar{\zeta}_i,x)$. We now define for $x\in\mathbb{R}$

$$\overline{F}(x)\equiv\frac{i}{2}\sum_{j=0}^{M}\delta_j\frac{\bar{b}_j}{\dot{\bar{a}}(\bar{\zeta}_j)}e^{-2i\bar{\zeta}_j x} + \frac{1}{4\pi}\,\text{p.v.}\int_{-\infty}^{\infty}\frac{\bar{b}(\zeta')}{\bar{a}(\zeta')}e^{-2i\zeta' x}\,d\zeta',$$

$$\overline{G}(x)\equiv\frac{i}{2}\sum_{j=0}^{M}\delta_j\frac{e^{2i\bar{\zeta}_j x}}{\bar{b}_j\dot{\bar{a}}(\bar{\zeta}_j)} + \frac{1}{4\pi}\,\text{p.v.}\int_{-\infty}^{\infty}\frac{b(\zeta')}{\bar{a}(\zeta')}e^{2i\zeta' x}\,d\zeta',$$

$$F(x)\equiv-\frac{i}{2}\sum_{j=0}^{N}\delta_j\frac{b_j e^{2i\zeta_j x}}{\dot{a}(\zeta_j)} + \frac{1}{4\pi}\,\text{p.v.}\int_{-\infty}^{\infty}\frac{b(\zeta')}{a(\zeta')}e^{i2\zeta' x}\,d\zeta',$$

$$G(x)\equiv-\frac{i}{2}\sum_{j=0}^{N}\delta_j\frac{1}{b_j\dot{a}(\zeta_j)}e^{-2i\zeta_j x} + \frac{1}{4\pi}\,\text{p.v.}\int_{-\infty}^{\infty}\frac{\bar{b}(\zeta')}{a(\zeta')}e^{2i\bar{\zeta}' x}\,d\zeta'.$$

Lemma 11: F and $\overline{F}$ (resp. G and $\overline{G}$) are bounded continuous functions for $x\geqslant 0$ (resp. $x\leqslant 0$).

Proof: The proof is obvious except for the principal value part of the corresponding functions, which may be handled by introduction of a partition of unity, as in the following Theorem 1. ■

By Lemma 11, (18), and the convolution theorem we obtain from (21)

$$\mathbf{K}^0(x,y) = \mathbf{F}_1(x+y) + \int_0^\infty \mathbf{F}_2(x+y+z)\mathbf{K}^0(x,z)\,dz,\quad y\geqslant 0, \tag{22a}$$

$$\mathbf{L}^0(x,y) = \mathbf{G}_1(x+y) + \int_0^\infty \mathbf{G}_2(x+y+z)\mathbf{L}^0(x,z)\,dz,\quad y\leqslant 0, \tag{22b}$$

where

$$\mathbf{K}^0(x,y)\equiv\begin{pmatrix}K^0(x,y)\\ K^0(x,y)\end{pmatrix}\quad\text{and}\quad \mathbf{L}^0(x,y)\equiv\begin{pmatrix}\overline{L}^0(x,y)\\ L^0(x,y)\end{pmatrix},$$

$$\mathbf{F}_1(x)\equiv\begin{pmatrix}\overline{F}(x)\\ -F(x)\end{pmatrix}\quad\text{and}\quad \mathbf{F}_2(x)\equiv\begin{pmatrix}0 & \overline{F}(x)\\ -F(x) & 0\end{pmatrix},$$

$$\mathbf{G}_1(x)\equiv\begin{pmatrix}-G(x)\\ \overline{G}(x)\end{pmatrix}\quad\text{and}\quad \mathbf{G}_2(x)\equiv\begin{pmatrix}0 & G(x)\\ -\overline{G}(x) & 0\end{pmatrix}.$$

Equations (22) may be regarded as Volterra equations for F, $\overline{F}$, G, and $\overline{G}$. We obtain therefore as in in Ref. 2 (the preservation of these properties by time evolution depends on the particular model; for the KdV equation, see Ref. 4):

Lemma 12: F and $\overline{F}$ (resp. G and $\overline{G}$) are C^∞ and satisfy the estimates

$$\left|\overset{(-)}{F}(x)\right|\leqslant\text{const}\ \exp(-\epsilon'' x)\quad\text{if}\quad x\geqslant 0, \tag{23a}$$

$$\left|\overset{(-)}{G}(x)\right|\leqslant\text{const}\ \exp(-\epsilon''' x)\quad\text{if}\quad x\leqslant 0 \tag{23b}$$

for some $\epsilon''>0$, $\epsilon'''>0$. ■

Viewed as integral equations for $\mathbf{K}^0$, $\mathbf{L}^0$ with $F^{(1/N)}$, $G^{(1/N)}$ satisfying (23), Eqs. (22) are the *Marchenko equations for the Zakharov–Shabat systems*.

Define the Banach spaces

$$\mathbf{L}^p(\mathbb{R}_+,dx)\equiv\left\{f=\begin{pmatrix}f_1\\ f_2\end{pmatrix}\ \ \|f\|_p^p\equiv\int_0^\infty dx\,|f_1(x)|^p + \int_0^\infty dx\,|f_2(x)|^p<\infty\right\}$$

and $\mathbf{L}^p(\mathbb{R}_-,dx)$ analogously. Define also the integral operators $\mathbf{F}_{2x}$ on $\mathbf{L}^p(\mathbb{R}_+,dx)$ [resp. $\mathbf{G}_{2x}$ on $\mathbf{L}^p(\mathbb{R},dx)$] for each $x\in\mathbb{R}$

$$(\mathbf{F}_{2x}f)(y) = \int_0^\infty \mathbf{F}_{2x}(y+z)f(z)\,dz,\quad y\geqslant 0,$$

$$(\mathbf{G}_{2x}f)(y)\equiv\int_0^\infty \mathbf{G}_{2x}(y+z)f(z)\,dz,\quad y\leqslant 0,$$

where

$$\mathbf{F}_{2x}(y+z)\equiv\mathbf{F}_2(x+y+z)$$

and

$$\mathbf{G}_{2x}(y+z) = \mathbf{G}_2(x+y+z).$$

These operators are bounded by Lemma 11.

Lemma 13: $\mathbf{F}_{2x}$ (resp. $\mathbf{G}_{2x}$) are for each $x\in\mathbb{R}$ compact operators on $\mathbf{L}^p(\mathbb{R}_+,dx)$ (resp. $\mathbf{L}^p(\mathbb{R}_-,dx)$) for each $p<\infty$.

Proof: By Lemma 11 the conditions of M. Riesz's criterion (Ref. 16, Theorem XIII-66, reproduced for convenience as Theorem A-1 of the Appendix) are satisfied. ■

Theorem 1: Equations (22) have a unique solution in $\mathbf{L}^p(\mathbb{R}_\pm, dx)$. $\mathbf{K}^0$ and $\mathbf{L}^0$ are C^∞ and satisfy the estimates (18). The Jost functions $\Phi, \bar{\Phi}, \Psi, \bar{\Psi}$, defined by (15) and (17) satisfy the eigenvalue equation for (1) with the proper boundary conditions, with potentials $q_\pm$ defined by (19a) and(19d). It follows that $q_\pm$ satisfy (2).

Proof: By Lemma 12 existence and uniticity of solutions of (22) are reduced to the proof of nonexistence of solutions of the homogeneous equations

$$\mathbf{F}_{2x}\mathbf{K}^0 = \mathbf{K}^0, \tag{24a}$$

$$\mathbf{G}_{2k}\mathbf{L}^0 = \mathbf{L}^0 \tag{24b}$$

in $\mathbf{L}^p(\mathbb{R}_+, dx)$ and $\mathbf{L}^p(\mathbb{R}_-, dx)$, respectively. Consider (24a); the analysis for (24b) is similar. We may write it as a pair of equations

$$\bar{F}_x(-\lambda)\bar{K}_x^0(\lambda) = \tilde{K}_x^0(\lambda), \tag{25a}$$

$$-\tilde{F}_x(-\lambda)\tilde{K}_x^0(\lambda) = \bar{K}_x^0(\lambda), \tag{25b}$$

where

$$\overset{(-)}{F}_x(\lambda) \equiv \int_0^\infty dy\, e^{2i\lambda y} \overset{(-)}{F}(x+y), \tag{26a}$$

$$\overset{(-)}{K}_x(\lambda) \equiv \int_0^\infty dy\, e^{2i\lambda y} \overset{(-)}{K}{}^0(x,y). \tag{26b}$$

From (25a) and (25b) we obtain

$$[1 + \tilde{F}_x(-\lambda)\bar{F}_x(-\lambda)]\tilde{K}_x^0(\lambda) = 0. \tag{27}$$

Suppose first that a has no spectral singularities. Then no p.v. appears in the definition of F. Using

$$\delta_\pm(\lambda) = (1/\pi)\int_0^\infty e^{\pm 2i\lambda y}\, dy = \tfrac{1}{2}\delta(\lambda) \mp (1/2\pi i)\text{p.v.}(1/\lambda)$$

as a distribution on $\mathscr{E}$ (the space of C^∞ functions of arbitrary support), we obtain by suitable regularizations in the usual way

$$\tilde{F}_x(-\lambda) = -\frac{1}{4}\sum_{j=0}^{N} \delta_j \frac{b_j}{\dot{a}(\zeta_j)} e^{i\zeta_j x} \frac{1}{\lambda - \zeta_j} + \frac{1}{2i}\left[i\pi \frac{b(-\lambda)}{a(-\lambda)} + \text{p.v.}\int_{-\infty}^{\infty} d\zeta' \frac{1}{\lambda + \zeta'} \frac{b(\zeta')}{a(\zeta')} e^{2i\zeta' x}\right]. \tag{28}$$

We claim that $\tilde{F}_x$ is analytic in $O_{\epsilon/2}$ ($O_{\epsilon/2} = \{\zeta : |\text{Im}\zeta| < \epsilon/8\}$). This is obvious expect for the principal value term. Introducing a partition of unity $\chi_1 + \chi_2 = 1$, where $\chi_1, \chi_2 \in C^\infty$, $\chi_1 \in \mathscr{D}$, with supp χ_1 contained in a ball B_1 around $(-\lambda)$, $\chi_2 = 0$ in a ball $B_2 \subset B_1$, we have

$$\text{p.v.}\int_{-\infty}^{\infty} \frac{1}{\lambda + \zeta'} \frac{b(\zeta')}{a(\zeta')} e^{2i\zeta' x}\, d\zeta' = \int_{-\infty}^{\infty} d\zeta' \frac{\chi_1(\zeta')}{\lambda + \zeta'}\left[\frac{b(\zeta')e^{2i\zeta' x}}{a(\zeta')} - \frac{b(-\lambda)e^{-2i\lambda x}}{a(-\lambda)}\right] + \int_{-\infty}^{\infty} d\zeta' \frac{\chi_2(\zeta')}{\lambda + \zeta'} \frac{b(\zeta')e^{2i\zeta' x}}{a(\zeta')}.$$

Now B_1 and B_2 may be chosen *fixed* for sufficiently small variations of $(-\lambda)$ in $O_{\epsilon/2}$. In addition,

$$f(\zeta', \lambda) \equiv \left[\frac{b(\zeta')e^{2i\zeta' x}}{a(\zeta')} - \frac{b(-\lambda)e^{-2i\lambda x}}{a(-\lambda)}\right] \quad (\lambda + \zeta') = \sum_{n=1}^{\infty} \frac{c^{(n)}(\lambda)}{n!}(\lambda + \zeta')^{n-1},$$

where $c^{(n)}$ denotes the nth-order derivative of $b(\zeta)e^{2i\zeta x}/a(\zeta)$ at $\zeta = -\lambda$. By the analyticity properties of $a(\cdot)$ and $b(\cdot)$ $\sup_{\lambda \in O_{\epsilon/2}} |C^{(n)}(\lambda)/n!| \leqslant d^n$ for some constant d. Taking now the radius of B_1 smaller $1/(2d)$, it follows easily that

$$\frac{d}{d\lambda}\int_{-\infty}^{\infty} d\zeta'\, \chi_1(\zeta') f(\zeta', \lambda) = \int_{-\infty}^{\infty} d\zeta'\, \chi_1(\zeta')\left[\sum_{n=1}^{\infty} \frac{c^{(n+1)}(\lambda)}{n!}(\lambda + \zeta')^{n-1} + \sum_{n=1}^{\infty} \frac{c^{(n)}(\lambda)}{n!}(n-1)(\lambda + \zeta')^{n-2}\right],$$

where the two series in brackets converge absolutely and uniformly for $\zeta' \in B_1$. Hence the first integral is analytic in $\lambda \in O_{\epsilon/2}$, as is the second because $\chi_2 = 0$ in $B_2 \subset B_1$. The proof is the same for $\bar{F}_x$. Hence $\tilde{F}_x(-\lambda)\bar{F}_x(-\lambda) = -1$ cannot hold but for λ in a set with no cluster point; otherwise $\tilde{F}_x(-\lambda)\bar{F}_x(-\lambda) = -1$ *identically* in $O_{\epsilon/2}$, which is contradicted by by $\lim_{\lambda \to \pm\infty} \tilde{F}_x(-\lambda) = 0$, $\lim_{\lambda \to \pm\infty} \bar{F}_x(-\lambda) = 0$ [this follows from (26a) and Lemma 12]. Hence, by (27), $\tilde{K}_x^0(\lambda) = 0$ a.e. in λ, for each $x \in \mathbb{R}$. Similarly, $\bar{K}^0(\lambda) = 0$ for a.e. $\lambda \in \mathbb{R}$, for each $x \in \mathbb{R}$. In case of a spectra singularity, the convolution of p.v. $1/(\lambda - \zeta)$ with p.v. $[b(\zeta)/a(\zeta)]e^{2i\zeta x}$ is an analytic function for $\lambda \in O_{\epsilon/2} \setminus \{\zeta_0\}$, and the proof proceeds along the same lines.

The rest of the proof of Proposition 3 is similar to Faddeev's in Ref. 2. Estimates (18) follow as in Ref. 2, p. 159, and the C^∞ property from Lemma 12 and Eqs. (22). ■

Remark: A proof of existence and unicity of solutions of equations of type (22) was sketched in Ref. 17 (see also Ref. 10, Appendix by V. Lyantse, and consists in transforming the Fredholm equation into a Volterra equation by a suitable transformation of Volterra type. Following the author's recipe, we did not, however, end up with an equation of Volterra type, so that we believe that the proof is incorrect.

Some results of existence and unicity were given in Ref. 5. The conditions given there on the potentials do not seem, however, to be natural. We believe exponential falloff to be the main issue. It implies certain analyticity properties on a strip, which are used to exclude nontrivial solutions to the homogeneous equation. □

APPENDIX

We reproduce here for convenience M. Riesz's criterion (Theorem XIII.66, p. 248 of Ref. 6), which is used in the proof of Lemma 13.

Theorem A-1: Let $p < \infty$, and $S \subset L_1^p(\mathbb{R}^n)$, the unit ball in L^p. A necessary and sufficient condition that the norm closure of S be norm compact is that:

(1) $f \to 0$ in L^p sense at infinity uniformly in S, i.e., for any ϵ, there is a bounded set $K \subset \mathbb{R}^n$ so that

$$\int_{\mathbf{R}^n \setminus K} |f(x)|^p dx \leqslant \epsilon^p$$

for all $f \in S$;

(2) $f(\cdot - y) \to f$ uniformly in S as $y \to 0$, i.e., for any ϵ, there is a δ so that $f \in S$ and $|y| < \delta$ imply that

$$\int_{\mathbf{R}^n} |f(x-y) - f(x)|^p \, dx \leqslant \epsilon^p.$$

[1]A. C. Scott, F. Y. Chu, and D. W. McLaughlin, Proc. IEEE **61**, 1443 (1973).
[2]L. D. Faddeev, Am. Math. Soc. Transl. Ser. 2, **65**, 139–166 (1967).
[3]P. Deift and E. Trubowitz, Comm. Pure Appl. Math. **32**, 121 (1979).
[4]S. Tanaka, Osaka J. Math. **11**, 49 (1974).
[5]M. J. Ablowitz, D. J. Kaup, A. C. Newell, and H. Segur, Stud. Appl. Math. **53**, 249 (1974).
[6]V. E. Zakharov and A. B. Shabat, Zh. Eksp. Teor. Fiz. **61**, 118 (1971) [Sov. Phys. JETP **34**, 62 (1972)].
[7]L. D. Faddeev and L. A. Takhtadzhyan, Theor. Math. Phys. **21**, 160 (1974).
[8]H. M. A. de Castro and W. F. Wreszinski, "Isospectral Property for a Class of Non-Self-Adjoint Operators," J. Math. Phys. (in press).
[9]B. S. Pavlov, Dokl. Akad. Nauk SSSR **141**, 807 (1961); **146**, 1267 (1962).
[10]M. A. Naimark, *Linear Differential Operators* (Harrup, New York, 1968), Part II.
[11]R. K. Dodd and R. K. Bullough, Physica Scripta **20**, 514 (1979).
[12]Z. S. Agranovich and V. A. Marchenko, *The Inverse Problem of Scattering Theory* (Gordon and Breach, New York, 1963).
[13]N. Asano and Y. Kato, J. Math. Phys. **22**, 2780 (1981).
[14]T. Kato, *Perturbation Theory for Linear Operators*, 1st ed. (Springer-Verlag, Berlin, 1966).
[15]M. Reed and B. Simon, *Methods of Modern Mathematical Physics. Vol. III. Scattering Theory* (Academic, New York, 1979).
[16]M. Reed and B. Simon, *Methods of Modern Mathematical Physics. Vol. IV. Analysis of Operators* (Academic, New York, 1978).
[17]V. E. Lyantse, Dokl. Akad. Nauk SSSR **166**, 30 (1966) [Sov. J. Math. **166**, 27 (1966)].

Strong approximation of wave operators for short range and Coulomb interactions

Helmut Kröger
Département de Physique, Université Laval, Québec, P.Q., Canada G1K 7P4

(Received 29 July 1982; accepted for publication 3 December 1982)

For the two-body system with short range potentials the wave operators are shown to be strongly approximatable by exponentials of bounded or even finite-rank operators, which allow the calculation by expansion in a power series or even by some finite-dimensional matrix algebra. It is based on a bounded strong approximation of the kinetic energy Hamiltonian. For the Coulomb potential the same kind of result is obtained for the Dollard modified wave operators. A generalization for the N-body system is sketched.

PACS numbers: 03.65.Nk, 02.60. + y

I. INTRODUCTION

Quantum mechanical scattering processes are basically formulated in the time dependent language in terms of wave operators, which build up the S matrix.

In order to perform a calculation, usually this picture is transformed into the time-independent language. Thus either one solves the Schrödinger equation with certain boundary conditions or equivalent integral equations of Lippmannn–Schwinger type for the two-body system or Faddeev-type for the three-body system, e.g. Here I will propose an approach, which approximates the wave operators by an exponential of a bounded or even finite-rank operator. This in turn allows the calculation of the exponential operators by expansion in a convergent power series or even by diagonalizing a finite matrix.

In the case of short-range potentials, which for technical simplification are assumed to be bounded operators, this is achieved by strongly approximating the kinetic energy Hamiltonian by a sequence of bounded operators, yielding a sequence of bounded full Hamiltonians (the limits of which are unbounded, of course). In a second step the kinetic energy and the full Hamiltonian are strongly approximated by finite-rank operators. For the Coulomb potential modified wave operators have to be considered, because the ordinary ones do not exist. I have chosen those given by Dollard.[1-3] Similar strong approximations of the kinetic energy Hamiltonian and of the potential by a sequence of bounded operators and in a second step by finite rank operators are performed.

The approximated S matrix, constructed from the approximated wave operators, conserves unitarity. This approach avoids dealing with the singularities of integral equations. However, it yields S matrix elements only between Hilbert states, i.e., initial and final wavepackets, rather than on shell plane waves.

In this work spin dependence has been suppressed for sake of technical simplicity, but its inclusion is straightforward.

In Sec. II, the two-body case with short range forces will be considered, while Sec. III treats the two-body case with the long-range (Coulomb) interaction. In Sec. IV a generalization to the N-body case will be sketched.

II. TWO-BODY SYSTEM WITH SHORT RANGE PAIR INTERACTION

Assume that the center of mass motion can be split off and consider only the relative motion of the two particles. Consider the Hilbert space $\mathscr{H} = L_2(\mathbb{R}^3)$. Let H^0 denote the free Hamiltonian, which reads in momentum space, with $\mathbf{p}$ the relative momentum between the two particles and m the reduced mass,

$$h^0(p) = p^2/2m. \tag{2.1}$$

H^0 is a self-adjoint operator on the domain

$$D(H^0) = \left\{ f \in \mathscr{H} \mid \int_{\mathbb{R}^3} d\mathbf{p} \mid p^2 f(\mathbf{p})|^2 < \infty \right\}, \tag{2.2}$$

which is dense in $\mathscr{H}$. A proof can be found, e.g., in Ref. 4.

Let V denote the short range pair interaction, defined in coordinate space by the real valued function $v(\mathbf{r})$, $\mathbf{r} \in \mathbb{R}^3$, where $\mathbf{r}$ denotes the relative distance between the two particles. One requires $v \in L_2(\mathbb{R}^3)$ and

$$|v(\mathbf{r})| \leqslant cr^{-\beta} \quad \text{for } r > R \tag{2.3}$$

with some positive c, R, and $\beta \gtrneqq 1$ (this means that V is of short range). Moreover, I require that $v(\mathbf{r})$, $\mathbf{r} \in \mathbb{R}^3$ is a bounded function, which is not a necessary but technically simplifying assumption. Hence V is bounded and self adjoint on $\mathscr{H}$. Let

$$H = H^0 + V. \tag{2.4}$$

Then H is self-adjoint on the domain $D(H) = D(H^0)$ (see, e.g., Ref. 4).

The wave operators are defined as

$$\Omega^{(\pm)} = \underset{t \to \mp \infty}{\text{s-lim}}\ e^{iHt} e^{-iH^0 t}. \tag{2.5}$$

The operators H and H^0 are unbounded. Thus the exponential e^{iHt} is defined via the spectral representation

$$H = \int dE_\sigma\, \sigma, \tag{2.6}$$

$$e^{iHt} = \int dE_\sigma\, e^{i\sigma t}, \tag{2.7}$$

and $e^{iH^0 t}$ is defined analogously. s-lim denotes the strong limit. The existence of the wave operators for a potential as given above is proven in Ref. 4.

 0022-2488/83/061509-07$02.50 © 1983 American Institute of Physics

The S matrix is defined by the Møller operator Ω

$$S = \Omega^{(-)\dagger}\,\Omega^{(+)}. \tag{2.8}$$

Direct calculation of S-matrix elements, using the wave operators given by (2.5), would require the knowledge of the spectral family E_σ of H and H^0, but that would imply the knowledge of the scattering solution.

Now I quote some results on strong resolvent convergence given by Weidmann,[5] which will be used in the forthcoming.

Definition: Let T_n ($n\in\mathbb{N}$) and T be self-adjoint operators on a dense subset D of a Hilbert space $\mathscr{H}$. Then T_n is said to converge to T in the sense of strong resolvent convergence, if $(z - T_n)^{-1} \xrightarrow{s} (z - T)^{-1}$ for some $z\in\mathbb{C}\setminus\mathbb{R}$.

Proposition of strong resolvent convergence: Let T_n ($n\in\mathbb{N}$) and T be self-adjoint operators on a dense subset D of a Hilbert space $\mathscr{H}$. The sequence T_n converges to T in the sense of strong resolvent convergence if for every $f\in D$ holds $T_n f \to Tf$.

Theorem on strong resolvent convergence: Let T_n ($n\in\mathbb{N}$) and T be self-adjoint operators on a dense subset D of a Hilbert space $\mathscr{H}$, and let T_n converge to T in the sense of strong resolvent convergence. Then $u(T_n) \xrightarrow{s} u(T)$ for every continuous bounded function u defined on $\mathbb{R}$, in particular, $e^{itT_n} \xrightarrow{s} e^{itT}$ for $t\in\mathbb{R}$.

Now I introduce bounded strong approximations of H^0 and H.

Definition: For each $s\in\mathbb{R}$, $s\neq 0$ let

$$H_s^0 = (E/s)\arctan\,(sH^0/E), \tag{2.9}$$

where E is an arbitrary positive scaling factor of the dimension of energy and

$$H_s = H_s^0 + V, \tag{2.10}$$

which obey the following:

Proposition 1: The operators H_s^0 and H_s given by (2.9) and (2.10), respectively, are bounded and self-adjoint on $\mathscr{H}$. For every $\phi\in D(H^0)$, one has $H_s^0\phi \to H^0\phi$ and $H_s\phi \to H\phi$, whenever s tends to 0.

Proof: For real x one has

$$|\arctan(x)| \leqslant \pi/2; \tag{2.11}$$

hence

$$\| H_s^0\| \leqslant |E/s|\,\pi/2. \tag{2.12}$$

This and the assumption that V is bounded imply the boundedness of H_s. Obviously, H_s^0 is self-adjoint on $\mathscr{H}$ and so is H_s because of the assumption on V. Now I want to show that for each $\phi\in D(H^0)$ holds $H_s^0\phi \to H^0\phi$ with $s\to 0$. The momentum space representation is used and the convention $\hbar = 1$ is adopted in the following. Let $\epsilon > 0$ and $\phi\in D(H^0)$, which means $\int_{\mathbb{R}^3} d\mathbf{p}\,|(p^2/2m)\,\phi(\mathbf{p})|^2 < \infty$; thus there is a p_0, such that

$$\int_{p>p_0} d\mathbf{p}\,|(p^2/2m)\,\phi(\mathbf{p})|^2 < \epsilon. \tag{2.13}$$

One has for real x

$$|\operatorname{artan}(x)| \leqslant |x|, \tag{2.14}$$

and hence

$$\left|\frac{E}{s}\arctan\left(\frac{s}{E}\frac{p^2}{2m}\right)\right| \leqslant \frac{p^2}{2m} \tag{2.15}$$

for all p. Moreover, whenever s tends to 0, the function $(1/s)\arctan(sx) - x$ tends to 0 uniformly in any finite real interval $x\in[x_1, x_2]$ whenever s tends to 0; hence there is s_0, such that for $|s| < |s_0|$ and for $0\leqslant p\leqslant p_0$

$$\left|\frac{E}{s}\arctan\left(\frac{s}{E}\frac{p^2}{2m}\right) - \frac{p^2}{2m}\right| < \epsilon. \tag{2.16}$$

Now one can estimate

$$\begin{aligned}\| H_s^0\phi - H^0\phi\|^2 &= \int_{p<p_0} d\mathbf{p}\left|\left[\frac{E}{s}\arctan\left(\frac{s}{E}\frac{p^2}{2m}\right) - \frac{p^2}{2m}\right]\phi(\mathbf{p})\right|^2 \\ &\quad + \int_{p>p_0} d\mathbf{p}\left|\left[\frac{E}{S}\arctan\left(\frac{s}{E}\frac{p^2}{2m}\right) - \frac{p^2}{2m}\right]\phi(\mathbf{p})\right|^2.\end{aligned} \tag{2.17}$$

The first term on the rhs can be estimated by $\epsilon^2\|\phi\|^2$, using (2.16). The second term yields

$$\begin{aligned}&\int_{p>p_0} d\mathbf{p}\left|\left[\frac{E}{s}\arctan\left(\frac{s}{E}\frac{p^2}{2m}\right) - \frac{p^2}{2m}\right]\phi(\mathbf{p})\right|^2 \\ &\quad \leqslant \int_{p>p_0} d\mathbf{p}\left|2\frac{p^2}{2m}\phi(\mathbf{p})\right|^2 < 4\epsilon,\end{aligned} \tag{2.18}$$

using (2.15) and (2.13), which proves the claim for H_s^0. Thus, for each $\phi\in D(H^0)$, $H_s^0\phi \to H^0\phi$ implies $H_s\phi \to H\phi$. This leads to the following:

Theorem 1: Let V be as assumed above, and let H_s^0 and H_s be given by (2.9) and (2.10). Then for every $\phi\in\mathscr{H}$ one can find $t\in\mathbb{R}$, $t > 0$, $s\in\mathbb{R}$ such that

$$\| \Omega^{(+)}\phi - e^{-iH_s t}e^{iH_s^0 t}\phi\|$$

and

$$\|\Omega^{(-)}\phi - e^{iH_s t}e^{-iH_s^0 t}\phi\|$$

become arbitrarily small. The approximated S matrix

$$S_{t,s} = \left(e^{iH_s t}e^{-iH_s^0 t}\right)^\dagger e^{-iH_s t}e^{iH_s^0 t} \tag{2.19}$$

is unitary.

Proof:

$$\begin{aligned}&\| \Omega^{(+)}\phi - e^{-iH_s t}e^{iH_s^0 t}\phi\| \\ &\quad \leqslant \| \Omega^{(+)}\phi - e^{-iHt}e^{iH^0 t}\phi\| \\ &\qquad + \| e^{-iHt}e^{iH^0 t}\phi - e^{-iH_s t}e^{iH_s^0 t}\phi\|\end{aligned} \tag{2.20}$$

The first term on the rhs becomes small with sufficiently large t because the wave operators exist. Proposition 1 means that the proposition and the theorem on strong resolvent convergence can be applied which implies that, for a given $t\in\mathbb{R}$, $e^{iH_s^0 t} \xrightarrow{s} e^{iH^0 t}$ and $e^{-iH_s t} \to e^{-iHt}$ whenever s tends to 0.

All the exponentials $e^{iH^0 t}$, $e^{iH_s^0 t}$, e^{-iHt}, and $e^{-iH_s t}$ are unitary and hence uniformly bounded operators. For two uniformly bounded series of strongly converging operators the series of the products converges strongly to the product of the limits. Thus the second term on the rhs of (2.20) becomes arbitrarily

small. The proof for $\Omega^{(-)}$ is quite analogous. The approximated S matrix $S_{t,s}$ given by (2.19) is unitary, because each exponential is unitary.

The usefulness of this theorem lies in the fact that the exponentials are built from bounded operators, which allows one to calculate the exponentials e^B, $B\in\{\pm iH_s^0 t, \pm iH_s t\}$, e.g., by an expansion in a convergent power series

$$e^B = \sum_{\nu=0,1,2,\cdots} B^\nu/\nu! \tag{2.21}$$

or by the expression

$$e^B = \lim_{\nu\to\infty} (1 + B/\nu)^\nu. \tag{2.22}$$

The expressions (2.21) and (2.22), although they converge, may turn out to be numerically inconvenient because of large B, which arises from large times t or small s (which makes H_s^0 large). Thus I want to propose an additional approximation which allows to calculate exponential operators of large arguments.

Definition: Let H_s^0 and H_s be given by (2.9) and (2.10) and let $\{\phi_n\}_{n\in\mathbb{N}}$ be a complete set of $\mathscr{H}$. Denote by P_n the orthogonal projection on the subspace spanned by $\phi_1,\ldots,\phi_n$. Let

$$H_{s,n}^0 = P_n H_s^0 P_n, \tag{2.23}$$

$$H_{s,n} = P_n H_s P_n. \tag{2.24}$$

Proposition 2: $H_{s,n}^0$ and $H_{s,n}$ given by (2.23) and (2.24) are finite-rank self-adjoint operators on $\mathscr{H}$ obeying $H_{s,n}^0 \xrightarrow{s} H_s^0$ and $H_{s,n}\xrightarrow{s} H_s$, whenever n tends to infinity.

Proof: H_s^0 and H_s are bounded, self-adjoint operators on $\mathscr{H}$ by Proposition 1, P_n is bounded, $\| P_n \| = 1$, and P_n is self-adjoint on $\mathscr{H}$; hence $H_{s,n}^0$ and $H_{s,n}$ are bounded and self-adjoint on $\mathscr{H}$. Obviously, $H_{s,n}^0$ and $H_{s,n}$ are of finite rank. Because $\{\phi_n\}_{n\in\mathbb{N}}$ is a complete set, one has $P_n\xrightarrow{s}1$, whenever n tends to infinity. Consider the series $\{P_n\}_{n\in\mathbb{N}}$, $\{H_s^0\}_{n\in\mathbb{N}}$, $\{P_n\}_{n\in\mathbb{N}}$. The product of uniformly bounded strongly converging series converges strongly to the product of the limits; thus $H_{s,n}^0\xrightarrow{s}H_s^0$. The same kind of argument is valid for $H_{s,n}\xrightarrow{s}H_s$.

Theorem 2: Let $t\in\mathbb{R}$, $s\in\mathbb{R}$, $n\in\mathbb{N}$, and let $H_{s,n}^0$ and $H_{s,n}$ be given by (2.23) and (2.24), respectively. Then for every $\phi\in\mathscr{H}$ one can find $t>0$, s, n such that

$$\| \Omega^{(+)}\phi - e^{-iH_{s,n}t} e^{iH_{s,n}^0 t}\phi \|$$

and

$$\| \Omega^{(-)}\phi - e^{iH_{s,n}t} e^{-iH_{s,n}^0 t}\phi \|$$

become arbitrarily small. The approximated S matrix

$$S_{t,s,n} = (e^{iH_{s,n}t} e^{-iH_{s,n}^0 t})^\dagger e^{-iH_{s,n}t} e^{iH_{s,n}^0 t} \tag{2.25}$$

is unitary.

Proof:

$$\begin{aligned}\| \Omega^{(+)}\phi - e^{-iH_{s,n}t} e^{iH_{s,n}^0 t}\phi \| &\leqslant \| \Omega^{(+)}\phi - e^{-iH_s t} e^{iH_s^0 t}\phi \| + \| e^{-iH_s t} e^{iH_s^0 t}\phi \\ &\quad - e^{-iH_{s,n}t} e^{iH_{s,n}^0 t}\phi \|. \end{aligned}\tag{2.26}$$

Using Theorem 1, it remains to show that for a given $t\in\mathbb{R}$, $t>0$, $s\in\mathbb{R}$, $s\neq 0$, $e^{-iH_{s,n}t} e^{iH_{s,n}^0 t}\xrightarrow{s} e^{-iH_s t} e^{iH_s^0 t}$, whenever n to infinity. Proposition 2 means that the proposition and the theorem on strong resolvent convergence can be applied, which implies that for given $t\in\mathbb{R}$, $s\in\mathbb{R}$, $s\neq 0$

$$e^{iH_{s,n}^0 t}\xrightarrow{s} e^{iH_s^0 t},$$

$$e^{-iH_{s,n}t}\xrightarrow{s} e^{-iH_s t},$$

whenever n tends to infinity. All the exponentials are unitary and hence uniformly bounded operators. Thus the product of the series $\{e^{-iH_{s,n}t}\}_{n\in\mathbb{N}}$, $\{e^{iH_{s,n}^0 t}\}_{n\in\mathbb{N}}$ strongly converges to the product of the limits $e^{-iH_s t} e^{iH_s^0 t}$. The estimate of $\Omega^{(-)}\phi$ is quite analogous. The approximated S matrix given by (2.25) is unitary because all exponentials are unitary. The usefulness of this theorem lies in the fact that the wave operators and the S matrix are approximated by exponentials of finite-rank operators. Now one is able to calculate easily $e^{iB_n t}$, $B_n\in\{H_{s,n}, H_{s,n}^0\}$, if the argument of the exponential operator becomes large. One can diagonalize B_n (finite-dimensional matrix) and calculate $e^{iB_n t}$ in the eigen representation of B_n.

III. LONG-RANGE INTERACTION

In this section for the two-body case the strong approximation of the wave operators and the S matrix shall be carried over to the case of a long-range potential V, which is assumed to consist of a short range potential V^{sh}, as defined in Sec. II, and the long range Coulomb potential V^{c},

$$V = V^{\mathrm{sh}} + V^{\mathrm{c}}. \tag{3.1}$$

The Coulomb potential is defined in coordinate space by

$$v^{\mathrm{c}}(\mathbf{r}) = e_1 e_2/r, \tag{3.2}$$

giving rise to an unbounded operator because of the $1/r$ singularity at the origin. Let H^0 be as in Sec. II, and let

$$H = H^0 + V. \tag{3.3}$$

It is shown in Ref. 4 that, with $D(H) = D(H^0)$, H is a self-adjoint operator on $D(H)$.

It is well known that for long-range potentials the wave operators (2.5) do not exist.[1,4] So one has to look for a way out. I shall adopt here the work of Dollard,[1–3] who proved the existence of wave operators modified as follows. For $t\in\mathbb{R}$, $t\neq 0$, let

$$H^{\mathrm{oc}}(t) = H^0 t + \mathrm{sgn}(t)(m/2H^0)^{1/2} e_1 e_2 \log(4H^0|\,t\,|). \tag{3.4}$$

$H^{\mathrm{oc}}(t)$ is self-adjoint on the domain

$$\begin{aligned} D(H^{\mathrm{oc}}(t)) = \Big\{ f\in\mathscr{H} \Big| \int_{\mathbb{R}^3} d\mathbf{p} \Big| &\times [(p^2/2m)\, t + \mathrm{sgn}(t)(m/p)\, e_1 e_2 \log(2p^2|\,t\,|/m)] \\ &\times f(\mathbf{p})\Big|^2 < \infty \Big\}, \end{aligned}\tag{3.5}$$

with $D(H^{\mathrm{oc}}(t))$ being dense in $\mathscr{H}$. This can be verified quite analogously to H^0 being self-adjoint on $D(H^0)$ (see Ref. 4).

Dollard[1] proves the existence of Coulomb wave operators

$$\Omega^{(\pm)} = \underset{t \to \mp\infty}{\text{s-lim}}\, e^{iHt} e^{-iH^{0c}(t)}. \tag{3.6}$$

A more general form of modified wave operators has been suggested, e.g., in Ref. 6. As in Sec. II one aims at approximating the arguments of the exponential operators by some bounded and finite rank operators. Compared with the case of short range forces there are some differences. First, as mentioned above, V^c and hence V are not bounded. Secondly, in the Coulomb modified free Hamiltonian $H^{0c}(t)$ the second term on the rhs of (3.4) is unbounded because of the $1/p$ singularity at the origin in momentum space. Thus I introduce the following approximations.

Definition: For every $s, t \in \mathbb{R}$, $s, t \neq 0$, define V_s^c in coordinate space via

$$v_s^c(\mathbf{r}) = v^c(\mathbf{r})(2/\pi)\arctan(r/sR), \tag{3.7}$$

where R is an arbitrary positive scaling factor of dimension length. Furthermore, let

$$V_s = V^{sh} + V_s^c, \tag{3.8}$$

$$H_s = H_s^0 + V_s, \tag{3.9}$$

$$H_s^{0c}(t) = H_s^0 t + \operatorname{sgn}(t)(m/2H^0)^{1/2} \times e_1 e_2 \log(4H^0|t|)(2/\pi)\arctan(H^0/sE), \tag{3.10}$$

where E is an arbitrary positive scaling factor of the dimension of energy, and H_s^0 is given by (2.9). Then the following holds:

Proposition 3: The operators V_s^c, V_s, H_s, and $H_s^{0c}(t)$ given by (3.7)–(3.10), respectively, are bounded and self-adjoint on $\mathscr{H}$. For a given $\phi \in D(H^0)$ one has $V_s^c\phi \to V^c\phi$, $V_s\phi \to V\phi$, $H_s\phi \to H\phi$, and for a given $\phi \in D(H^{0c}(t))$ one has $H_s^{0c}(t)\phi \to H^{0c}(t)\phi$, whenever s tends to 0.

Proof: V_s^c is given in coordinate space by the function $v_s^c(\mathbf{r})$, $\mathbf{r} \in \mathbb{R}^3$, which is real-valued and bounded because of (2.14); hence V_s^c is a bounded and self-adjoint operator on $\mathscr{H}$. V^{sh} is bounded and self-adjoint on $\mathscr{H}$ by assumption in Sec. II; hence so is V_s. H_s^0 and $H_s^{0c}(t)$ are given for $\mathbf{p} \in \mathbb{R}^3$ by real-valued functions. These functions are also bounded, which was shown for H_s^0 in Proposition 1 and also holds for $H_s^{0c}(t)$ because for $x \in \mathbb{R}$, $x \geqslant 0$, the function $(1/x)\log(x^2)\arctan(x^2)$ has an absolute maximum. Hence H_s and $H_s^{0c}(t)$ are self-adjoint on $\mathscr{H}$.

Now let $\epsilon > 0$, $\phi \in D(H) = D(H^0)$; hence $V^c\phi \in \mathscr{H}$. In coordinate representation this reads $\int_{\mathbb{R}^3} d\mathbf{r}|(1/r)\phi(\mathbf{r})|^2 < \infty$; thus there are $0 < r_1 < r_2$ such that

$$\int_{r<r_1} d\mathbf{r}\,|(1/r)\phi(\mathbf{r})|^2 < \epsilon \tag{3.11}$$

and

$$\int_{r>r_2} d\mathbf{r}\,|(1/r)\phi(\mathbf{r})|^2 < \epsilon. \tag{3.12}$$

The function $(1/r)[1 - (2/\pi)\arctan(r/sR)]$ tends to 0 uniformly on the finite interval $[r_1, r_2]$, which excludes the origin, whenever s tends to 0, i.e., there is s_0, such that for

$$r_1 \leqslant r \leqslant r_2,\ |s| \leqslant s_0|$$
$$|(1/r)[1 - (2/\pi)\arctan(r/sR)]| < \epsilon. \tag{3.13}$$

Thus one can estimate

$$\begin{aligned}\|V_s^c\phi - V^c\phi\|^2 &= \int_{r<r_1} d\mathbf{r}|(e_1e_2/r)\,[1 - (2/\pi)\arctan(r/sR)]\,\phi(\mathbf{r})|^2 \\ &+ \int_{r_1<r<r_2} d\mathbf{r}\,|(e_1e_{21}/r)\,[1 - (2/\pi)\arctan(r/sR)]\,\phi(\mathbf{r})|^2 \\ &+ \int_{r>r_2} d\mathbf{r}\,|(e_1e_2/r)\,[1 - (2/\pi)\arctan(r/sR)]\,\phi(\mathbf{r})|^2.\end{aligned} \tag{3.14}$$

The first term on the rhs can be estimated by

$$\begin{aligned}&\int_{r<r_1} d\mathbf{r}\,|(e_1e_2/r)\,[1 - (2/\pi)\arctan(r/sR)]\,\phi(\mathbf{r})|^2 \\ &\leqslant \int_{r<r_1} d\mathbf{r}\,|(2e_1e_2/r)\,\phi(\mathbf{r})|^2 < 4e_1^2\,e_2^2\,\epsilon,\end{aligned} \tag{3.15}$$

using (2.11) and (3.11). Similarly the third term can be estimated by $4e_1^2\,e_2^2\,\epsilon$. The second term can be estimated by

$$\begin{aligned}&\int_{r_1<r<r_2} d\mathbf{r}\,|(e_1\,e_2/r)\,[1 - (2/\pi)\arctan(r/sR)]\,\phi(\mathbf{r})|^2 \\ &\leqslant \int_{r_1<r<r_2} d\mathbf{r}\,|\,e_1\,e_2\epsilon\phi(\mathbf{r})|^2 \\ &\leqslant e_1^2\,e_2^2\,\|\phi\|^2\,\epsilon^2\end{aligned} \tag{3.16}$$

for each $|s| \leqslant |s_0|$, using (3.13), which proves $V_s^c\phi \to V^c\phi$. For every $\phi \in D(H)$ that implies $V_s\phi = (V^{sh} + V_s^c)\phi \to (V^{sh} + V^c)\phi = V\phi$. For every $\phi \in D(H)$, that and Proposition 1 imply $H_s\phi = (H_s^0 + V_s)\phi \to (H^0 + V)\phi = H\phi$. Now putting

$$A(t) = \operatorname{sgn}(t)(m/2H^0)^{1/2}\,e_1\,e_2\log(4H^0|t|), \tag{3.17}$$

which Dollard called the anomalous term, and

$$A_s(t) = A(t)(2/\pi)\arctan(H^0/sE), \tag{3.18}$$

one can write (3.4) and (3.10), respectively, as

$$H^{0c}(t) = H^0 t + A(t), \tag{3.19}$$

$$H_s^{0c}(t) = H_s^0 t + A_s(t). \tag{3.20}$$

Because of Proposition 1, it remains to show for $\phi \in D(H^{0c}(t))$ $A_s(t)\phi \to A(t)\phi$. Let $\epsilon > 0$ and $\phi \in D(H^{0c}(t))$, i.e., one has in momentum representation $\int_{\mathbb{R}^3} d\mathbf{p}\,|(1/p)\log(2p^2|t|/m)\,\phi(\mathbf{p})|^2 < \infty$; thus there are $0 < p_1 < p_2$ such that

$$\int_{p<p_1} d\mathbf{p}\,|(1/p)\log(2p^2|t|/m)\,\phi(\mathbf{p})|^2 < \epsilon, \tag{3.21}$$

$$\int_{p>p_2} d\mathbf{p}\,|(1/p)\log[(2p^2/m|t|]\,\phi(\mathbf{p})|^2 < \epsilon. \tag{3.22}$$

The function

$$\frac{1}{p}\log\left(\frac{2p^2}{m}|t|\right)\left[1 - \frac{2}{\pi}\arctan\left(\frac{p^2}{2m}\cdot\frac{1}{sE}\right)\right]$$

tends to 0 uniformly on the finite interval $[p_1, p_2]$ which excludes the origin, whenever s tends to 0, i.e., there is s_0, such that for $p_1 \leqslant p \leqslant p_2$, $|s| \leqslant |s_0|$,

$$\left|\frac{1}{p}\log\left(\frac{2p^2}{m}|t|\right)\left[1-\frac{2}{\pi}\arctan\left(\frac{p^2}{2m}\frac{1}{sE}\right)\right]\right|<\epsilon. \tag{3.23}$$

Thus one can estimate

$$\begin{aligned}
&\|A_s(t)\phi - A(t)\phi\|^2 \\
&\quad=\int_{p<p_1} d\mathbf{p}\left|\frac{me_1e_2}{p}\log\left(\frac{2p^2|t|}{m}\right)\right. \\
&\quad\times\left.\left[1-\frac{2}{\pi}\arctan\left(\frac{p^2}{2m}\cdot\frac{1}{sE}\right)\right]\phi(\mathbf{p})\right|^2 \\
&\quad+\int_{p_1<p<p_2} d\mathbf{p}\left|\frac{me_1e_2}{p}\log\left(\frac{2p^2|t|}{m}\right)\right. \\
&\quad\times\left.\left[1-\frac{2}{\pi}\arctan\left(\frac{p^2}{2m}\frac{1}{sE}\right)\right]\phi(\mathbf{p})\right|^2 \\
&\quad+\int_{p>p_2} d\mathbf{p}\left|\frac{me_1e_2}{p}\log\left(\frac{2p^2|t|}{m}\right)\right. \\
&\quad\times\left.\left[1-\frac{2}{\pi}\arctan\left(\frac{p^2}{2m}\frac{1}{sE}\right)\right]\phi(\mathbf{p})\right|^2.
\end{aligned} \tag{3.24}$$

The first term on the rhs can be estimated by

$$\begin{aligned}
&\int_{p<p_1} d\mathbf{p}\left|\frac{me_1e_2}{p}\log\left(\frac{2p^2|t|}{m}\right)\right. \\
&\quad\times\left.\left[1-\frac{2}{\pi}\arctan\left(\frac{p^2}{2m}\frac{1}{sE}\right)\right]\phi(\mathbf{p})\right|^2 \\
&\quad\leqslant\int_{p<p_1} d\mathbf{p}\left|2\frac{me_1e_2}{p}\log\left(\frac{2p^2|t|}{m}\right)\phi(\mathbf{p})\right|^2 \\
&\quad\leqslant 4m^2e_1^2e_2^2\epsilon,
\end{aligned} \tag{3.25}$$

using (2.11) and (3.21). Similarly, the third term can be estimated by $4m^2e_1^2e_2^2\epsilon$. The second term can be estimated by

$$\begin{aligned}
&\int_{p_1<p<p_2} d\mathbf{p}\left|\frac{me_1e_2}{p}\log\left(\frac{2p^2|t|}{m}\right)\right. \\
&\quad\times\left.\left[1-\frac{2}{\pi}\arctan\left(\frac{p^2}{2m}\frac{1}{sE}\right)\right]\phi(\mathbf{p})\right|^2 \\
&\quad\leqslant\int_{p_1<p<p_2} d\mathbf{p}\,|me_1e_2\epsilon\phi(\mathbf{p})|^2\leqslant m^2e_1^2e_2^2\|\phi\|^2\epsilon^2
\end{aligned} \tag{3.26}$$

for each $|s|\leqslant|s_0|$, using (3.23), which finishes the proof of the proposition. This leads to the following:

Theorem 3: Let V be as assumed above, and let H_s and $H_s^{\mathrm{Oc}}(t)$ be given by (3.9) and (3.10), respectively. Then for every $\phi\in\mathscr{H}$ one can find $t\in\mathbb{R}$, $t>0$, $s\in\mathbb{R}$ such that

$$\|\Omega^{(+)}\phi - e^{-iH_st}e^{iH_s^{\mathrm{Oc}}(t)}\phi\|$$

and

$$\|\Omega^{(-)}\phi - e^{iH_st}e^{-iH_s^{\mathrm{Oc}}(t)}\phi\|$$

become arbitrarily small. The approximated S matrix

$$S_{t,s}=\left(e^{iH_st}e^{-iH_s^{\mathrm{Oc}}(t)}\right)^{\dagger}e^{-iH_st}e^{iH_s^{\mathrm{Oc}}(t)} \tag{3.27}$$

is unitary.

Proof: The proof runs quite analogous to that of Theorem 1, but now using Proposition 3, instead of Proposition 1 and thus shall not be repeated here. Theorem 3 establishes also for long range forces that the wave operators can be strongly approximated by exponentials of bounded operators and hence allow further treatment using (2.21) and (2.22).

Of course one can introduce the finite-rank approximation.

Definition: Let H_s and $H_s^{\mathrm{Oc}}(t)$ be given by (3.9) and (3.10) and let $\{\phi_n\}_{n\in\mathbb{N}}$ be a complete set of $\mathscr{H}$. Denote by P_n the orthogonal projection on the subspace spanned by $\{\phi_1,\ldots,\phi_n\}$. Let

$$H_{s,n}=P_nH_sP_n, \tag{3.28}$$

$$H_{s,n}^{\mathrm{Oc}}(t)=P_nH_s^{\mathrm{Oc}}(t)P_n. \tag{3.29}$$

Proposition 4: $H_{s,n}$ and $H_{s,n}^{\mathrm{Oc}}(t)$ given by (3.28) and (3.29) are finite-rank self-adjoint operators on $\mathscr{H}$ obeying $H_{s,n}\overset{s}{\to}H_s$ and $H_{s,n}^{\mathrm{Oc}}(t)\overset{s}{\to}H_s^{\mathrm{Oc}}(t)$, whenever n tends to infinity.

Proof: The proof is quite analogous to that of Proposition 2, but now with H_s^0 being replaced by $H_s^{\mathrm{Oc}}(t)$ and hence shall not be repeated here.

Theorem 4: Let $t\in\mathbb{R}$, $s\in\mathbb{R}$, $n\in\mathbb{N}$ and let $H_{s,n}$ and $H_{s,n}^{\mathrm{Oc}}(t)$ be given by (3.28) and (3.29), respectively. Then for every $\phi\in\mathscr{H}$ one can find $t>0$, s, n such that

$$\|\Omega^{(+)}\phi - e^{-iH_{s,n}t}e^{iH_{s,n}^{\mathrm{Oc}}(t)}\phi\|$$

and

$$\|\Omega^{(-)}\phi - e^{iH_{s,n}t}e^{-iH_{s,n}^{\mathrm{Oc}}(t)}\phi\|$$

become arbitrarily small. The approximated S matrix

$$S_{t,s,n}=\left(e^{iH_{s,n}t}e^{-iH_{s,n}^{\mathrm{Oc}}(t)}\right)^{\dagger}e^{-iH_{s,n}t}e^{iH_{s,n}^{\mathrm{Oc}}(t)} \tag{3.30}$$

is unitary.

Proof: The proof is quite analogous to that of Theorem 2, but now with H_s^0 being replaced by $H_s^{\mathrm{Oc}}(t)$ and hence shall not be repeated here. Thus also for long range forces one arrives at a strong approximation of the wave operators by exponentials of finite-rank operators.

IV. *N*-BODY SYSTEM

In this section I will sketch, without giving proofs, how the former results on two-body scattering for short range and long range Coulomb interaction can be generalized to the N-body system.

Assume there are N distinguishable particles, denoted by i,j, considered as "elementary" as opposite to "composite" particles, which are formed as bound states of the former ones. The elementary particle i has the mass m_i and the charge e_i. There are channels, denoted by α,β, describing the situation where some particles are elementary while some others are composite and all move freely relative to each other. The situation without a composite particle is called the free channel and is denoted by 0. The total interaction is assumed to consist of pair interactions

$$V=\sum_{1\leqslant i<j\leqslant N}V^{ij}, \tag{4.1}$$

where every V^{ij} consists of a short range potential $V^{\mathrm{sh}ij}$ and the Coulomb potential $V^{\mathrm{c}ij}$, as given in Sec. III. Let $V^{ij\alpha}$ denote the pair interaction V^{ij} if there is a composite particle in channel α and both elementary particles i, j are contained in that composite particle; otherwise, $V^{ij\alpha} = 0$ (in particular, $V^{ij0} = 0$). Corresponding to each channnel one defines an interaction

$$V^{\alpha} = \sum_{1 \le i < j \le N} V^{ij\alpha} \tag{4.2}$$

($V^0 = 0$). The kinetic energy of the single particle i is denoted by H_{0i} and the total kinetic energy is

$$H^0 = \sum_{i=1}^{N} H^{0i}. \tag{4.3}$$

Moreover, one introduces the channel Hamiltonian

$$H^{\alpha} = H^0 + V^{\alpha} \tag{4.4}$$

and the full Hamiltonian

$$H = H^0 + V. \tag{4.5}$$

Let me first consider the Coulomb forces to be switched off. Then one defines the wave operators corresponding to the channel α by

$$\Omega^{\alpha(\pm)} = \operatorname*{s-lim}_{t \to \mp \infty} e^{iHt} e^{-iH^{\alpha}t}. \tag{4.6}$$

More precisely, this expression is defined on D^{α} consisting of those elements of the N-body Hilbert space $\mathscr{H}$, which can be written as a tensor product of the bound state wavefunctions for all composite particles appearing in channel α and a wavepacket for the relative motion between the single particles and the c.m. of the composite particles. In the case $\alpha = 0$, $D^{\alpha} = \mathscr{H}$.

The partial S-matrix describing the scattering from channel α to β is given by

$$S^{\beta\alpha} = \Omega^{\beta(-)\dagger} \Omega^{\alpha(+)}. \tag{4.7}$$

The bounded and finite rank approximation of H^{α} and H can be defined in many ways. One choice is

$$H_s^0 = (E/s) \arctan(sH^0/E), \tag{4.8}$$

$$H_s^{\alpha} = H_s^0 + V^{\alpha}, \tag{4.9}$$

$$H_s = H_s^0 + V, \tag{4.10}$$

where $s \in \mathbb{R}$, $s \neq 0$, and E is an arbitrary positive energy. Let $\{\phi_n\}_{n \in \mathbb{N}}$ be a complete set of the N-body Hilbert space $\mathscr{H}$, let P_n denote the orthogonal projection on the subspace spanned by $\{\phi_1, \ldots, \phi_n\}$, and let

$$H_{s,n}^{\alpha} = P_n H_s^{\alpha} P_n, \tag{4.11}$$

$$H_{s,n} = P_n H_s P_n. \tag{4.12}$$

Thus one derives from Sec. II the following.

Conjecture: Let H_s^{α}, H_s, $H_{s,n}^{\alpha}$, and $H_{s,n}$ be given by (4.9)–(4.12), respectively. Then for every element $\phi \in D^{\alpha}$ one can find s, $t > 0$, n such that

$$\| \Omega^{(+)}\phi - e^{-iH_s t} e^{iH_s^{\alpha} t} \phi \|,$$

$$\| \Omega^{(-)}\phi - e^{iH_s t} e^{-iH_s^{\alpha} t} \phi \|,$$

$$\| \Omega^{(+)}\phi - e^{-iH_{s,n} t} e^{iH_{s,n}^{\alpha} t} \phi \|,$$

$$\| \Omega^{(-)}\phi - e^{iH_{s,n} t} e^{-iH_{s,n}^{\alpha} t} \phi \|$$

become arbitrarily small. The approximated partial S matrices

$$S_{s,t}^{\beta\alpha} = \left(e^{iH_s t} e^{-iH_s^{\beta} t}\right)^{\dagger} e^{-iH_s t} e^{iH_s^{\alpha} t}, \tag{4.13}$$

$$S_{s,t,n}^{\beta,\alpha} = \left(e^{iH_{s,n} t} e^{-iH_{s,n}^{\beta} t}\right)^{\dagger} e^{-iH_{s,n} t} e^{iH_{s,n}^{\alpha} t} \tag{4.14}$$

are unitary

Finally I will consider the modifications when in addition Coulomb forces are involved. According to Dollard,[3] the wave operators should be modified as follows. In the case of the free channel ($\alpha = 0$) let m_{ij}^{rel} denote the reduced mass of the particles i, j, let $H^{0\,\mathrm{rel}\,ij}$ denote the kinetic energy of the relative motion between particle i, j, and let

$$A^{0\mathrm{c}}(t) = \operatorname{sgn}(t) \sum_{1 \le i < j \le N} e_i e_j \left(\frac{m_{ij}^{\mathrm{rel}}}{2H^{0\,\mathrm{rel}\,ij}}\right)^{1/2} \times \log(4|t \mid H^{0\,\mathrm{rel}\,ij}), \tag{4.15}$$

$$H^{0\mathrm{c}}(t) = H^0 t + A^{0\mathrm{c}}(t). \tag{4.16}$$

Then the wave operators are defined on $D^0 = \mathscr{H}$ by

$$\Omega^{0(\pm)} = \operatorname*{s-lim}_{t \to \mp \infty} e^{iHt} e^{-iH^{0\mathrm{c}}(t)}. \tag{4.17}$$

In the case of channel $\alpha \neq 0$ let I, J denote an either single or composite particle. $I \subset \alpha$ means that particle I, if it is a single particle is contained as single particle in channel α, or if it is a composite particle it is contained as a composite particle in channel α. Let e_I denote the charge of the particle I, m_I the mass of particle I, m_{IJ}^{rel} the reduced mass of particles I, J, and $H^{0\,\mathrm{rel}\,IJ}$ the kinetic energy of the relative motion between particles I, J. Let

$$A^{\alpha\mathrm{c}}(t) = \operatorname{sgn}(t) \sum_{1 \le I < J \le N} e_I e_J \left(\frac{m_{IJ}^{\mathrm{rel}}}{2H^{0\,\mathrm{rel}\,IJ}}\right)^{1/2} \times \log(4|t\,|H^{0\,\mathrm{rel}\,IJ}), \tag{4.18}$$

$$H^{\alpha\mathrm{c}}(t) = H^{\alpha} t + A^{\alpha\mathrm{c}}(t). \tag{4.19}$$

Then the wave operators are defined on D^{α} by

$$\Omega^{\alpha(\pm)} = \operatorname*{s-lim}_{t \to \mp \infty} e^{iHt} e^{-iH^{\alpha\mathrm{c}}(t)}. \tag{4.20}$$

The S matrix is given by (4.7) but with the wave operators taken from (4.17) if $\alpha = 0$ or (4.20) if $\alpha \neq 0$. Bounded and finite rank approximations of H and $H^{\alpha\mathrm{c}}(t)$ can be defined as follows. For each pair potential V^{ij} and its Coulomb part $V^{\mathrm{c}ij}$ one introduces $V_s^{\mathrm{c}ij}$, given by (3.7) and

$$V_s^{ij} = V^{\mathrm{sh}ij} + V_s^{\mathrm{c}ij}. \tag{4.21}$$

Let $V_s^{ij\alpha}$ denote those pair interactions V_s^{ij}, such that the elementary particles i and j belong to the same composite particle in the channel α, let

$$V_s^{\alpha} = \sum_{1 \le i < j \le N} V_s^{ij\alpha} \tag{4.22}$$

and

$$V_s = \sum_{1 \le i < j \le N} V_s^{ij}. \tag{4.23}$$

Using (4.8), (4.22), and (4.23), oen defines

$$H_s^\alpha = H_s^0 + V_s^\alpha, \tag{4.24}$$

$$H_s = H_s^0 + V_s. \tag{4.25}$$

Using (4.15) in the case $\alpha = 0$, one defines in analogy to (3.17) and (3.18)

$$A_s^{0c}(t) = \operatorname{sgn}(t) \sum_{1<i<j<N} e_i e_j \left(\frac{m_{ij}^{\text{rel}}}{2H^{0\,\text{rel}\,ij}} \right)^{1/2} \times \log(4|t \mid H^{0\,\text{rel}\,ij}) \frac{2}{\pi} \arctan\left(\frac{H^{0\,\text{rel}\,ij}}{sE} \right), \tag{4.26}$$

and, using (4.8) and (4.26),

$$H_s^{0c}(t) = H_s^0 t + A_s^{0c}(t). \tag{4.27}$$

Using (4.18) in the case $\alpha \neq 0$, one defines

$$A_s^{\alpha c}(t) = \operatorname{sgn}(t) \sum_{1<I<J<N} e_I e_J \left(\frac{m_{IJ}^{\text{rel}}}{2H^{0\,\text{rel}\,IJ}} \right)^{1/2} \times \log(4|t \mid H^{0\,\text{rel}\,IJ}) \frac{2}{\pi} \arctan\left(\frac{H^{0\,\text{rel}\,IJ}}{sE} \right) \tag{4.28}$$

and, using (4.24) and (4.28),

$$H_s^{\alpha c}(t) = H_s^\alpha t + A_s^{\alpha c}(t). \tag{4.29}$$

The finite-rank approximation of $H_s^{\alpha c}(t)$ is obtained, using (4.27) if $\alpha = 0$ and (4.29) if $\alpha \neq 0$, by

$$H_{s,n}^{\alpha c}(t) = P_n H_s^{\alpha c}(t) P_n. \tag{4.30}$$

Using (4.25), the finite-rank approximation of H_s is given by

$$H_{s,n} = P_n H_s P_n. \tag{4.31}$$

From Sec. III one derives the following:

Conjecture: Let $H_s^{\alpha c}(t)$, H_s, $H_{s,n}^{\alpha c}(t)$, and $H_{s,n}$ be given by (4.27) or (4.29), (4.25), (4.30), and (4.31), respectively. Then for every element $\phi \in D^\alpha$ one can find s, $t > 0$, n such that

$$\| \Omega^{(+)}\phi - e^{-iH_s t} e^{iH_s^{\alpha c}(t)} \phi \|$$

$$\| \Omega^{(-)}\phi - e^{iH_s t} e^{-iH_s^{\alpha c}(t)} \phi \|$$

$$\| \Omega^{(+)}\phi - e^{-iH_{s,n} t} e^{iH_{s,n}^{\alpha c}(t)} \phi \|$$

$$\| \Omega^{(-)}\phi - e^{iH_{s,n} t} e^{-iH_{s,n}^{\alpha c}(t)} \phi \|$$

become arbitrarily small. The approximated S matrices

$$S_{s,t}^{\beta\alpha} = \left(e^{iH_s t} e^{-iH_s^{\beta c}(t)} \right)^\dagger e^{-iH_s t} e^{iH_s^{\alpha c}(t)}, \tag{4.32}$$

$$S_{s,t,n}^{\beta\alpha} = \left(e^{iH_{s,n} t} e^{-iH_{s,n}^{\beta c}(t)} \right)^\dagger e^{-iH_{s,n} t} e^{iH_{s,n}^{\alpha c}(t)} \tag{4.33}$$

are unitary.

V. CONCLUSION

A new method has been proposed in nonrelativistic few-body scattering theory which allows to approximate the S-matrix elements with arbitrary accuracy. However, this is restricted to S-matrix elements, where in the asymptotic ingoing and outgoing channel the relative motion between the fragments (elementary or composite particles) is described by a wavepacket. The basic idea is to approximate strongly a densely defined self-adjoint but unbounded Hamiltonian by a sequence of self-adjoint bounded or even finite-rank operators. This can be done in many ways. The particular choice used above was just to give an example. Then the Møller wave operators can be strongly approximated by exponentials of these bounded or finite-rank operators. These exponential operators, in turn can be calculated from a Taylor series or even by some finite-dimensional matrix algebra. The potential considered has the form of a sum of pair potentials, each of which has a short range part plus the long range Coulomb part. The short range potential is assumed to be a bounded operator, but that can be weakened to include, e.g., repulsive soft core potentials as long as the full Hamiltonian is a densely defined self-adjoint operator.

ACKNOWLEDGMENT

It is a pleasure to acknowledge helpful discussions with Professor J. Weidmann.

[1] J. D. Dollard, Rocky Mountain J. Math. **1**, 5 (1971).
[2] J. D. Dollard, Rocky Mountain J. Math. **2**, 317 (1972).
[3] J. D. Dollard, J. Math. Phys. **5**, 729 (1964).
[4] W. O. Amrein, J. M. Jauch, and K. B. Sinha, *Scattering Theory in Quantum Mechanics*, Lecture Notes and Supplements in Physics (Benjamin, Reading, Mass., 1977).
[5] J. Weidmann, *Linear Operators in Hilbert Spaces* (Springer-Verlag, New York, 1980).
[6] L. Hörmander, Math. Z. **146**, 69 (1976).

Continuity of wave and scattering operators with respect to interactions

E. Brüning and F. Gesztesy[a,b]
Fakultät für Physik, Universität Bielefeld, D-4800 Bielefeld 1, Federal Republic of Germany

(Received 11 February 1982; accepted for publication 28 May 1982)

Using time-dependent as well as time-independent methods we establish strong continuity of wave and scattering operators within the two Hilbert space formulation of scattering theory. Our applications to potential scattering include continuity of nonforward total cross sections, scattering amplitudes, and operators for Coulomb-type interactions in $\mathbb{R}^3$ and continuity of wave operators for strongly singular potentials in $\mathbb{R}^n$, $n \geqslant 3$.

PACS numbers: 03.65.Nk, 03.65.Db

1. INTRODUCTION

In quantum mechanics the technique of approximating a given (possibly singular) Hamiltonian H by a sequence of "regularized" Hamiltonians H_j, $j = 1,2,\ldots$ is used very often and powerful methods obtaining convergence of H_j to H in a certain sense as $j \to \infty$ have been developed. In this context it is natural to ask which additional conditions ensure that in the limit $j \to \infty$ quantities like the scattering amplitudes f_j, the wave operators $\Omega_\pm(H_j, H_0)$ and the scattering operators $S(H_j, H_0)$ associated with H_j and H_0 (a suitable "free" Hamiltonian) also converge to the corresponding quantities f, $\Omega_\pm(H,H_0)$, and $S(H,H_0)$ of H and H_0 in the limit $j \to \infty$. We shall present two different approaches to these problems, one based on time-dependent and the other on time-independent methods in scattering theory.

In Sec. 2 we discuss the time-dependent approach to continuity of wave and scattering operators with respect to the interactions H_j. Based on strong resolvent convergence of H_j to H (which turns out to be the appropriate way in which H_j should approximate H as $j \to \infty$) and a new criterion for the existence of wave operators due to Kato[1] we prove strong convergence of $\Omega_\pm(H_j,H_0)$ to $\Omega_\pm(H,H_0)$ in the general framework of two Hilbert space scattering theory.

Our applications to quantum scattering in $\mathscr{L}^2(\mathbb{R}^n)$, $n \geqslant 3$ with Hamiltonians of the type $H_j = -\Delta + V_j$, $H = -\Delta + V$ in Sec. 3, clearly show to what extent our approach generalizes previous results of Narnhofer[2] (see also Refs. 3, 4, and 5, p. 329) which were based on Cook's criterion[6] in its original form. Roughly speaking, we are able to treat potentials V_j, V, whose long-range parts are locally square integrable whereas their short-range parts are assumed to be only locally integrable. As a by-product, screening or cut-off procedures are shown to be applicable under extremely weak conditions on V_j, V.

From a physical point of view strong convergence of scattering operators immediately yields convergence of probabilities of scattering of particles into cones (cf. the end of Sec. 3) but this is not sufficient to ensure convergence of scattering amplitudes and cross sections. Therefore we study a stationary approach to the convergence of scattering operators in Sec. 4. Based on the fact that $S(H_j,H_0)$ and $S(H,H_0)$ are decomposable in the spectral representation of H_0 (assuming for a moment that H_0 has absolutely continuous spectrum) we discuss necessary and sufficient conditions on the on-shell scattering operators $S_j(k)$ and $S(k)$ [the corresponding decompositions of $S(H_j,H_0)$ and $S(H,H_0)$] such that a certain type of convergence of $S_j(k)$ to $S(k)$ implies strong convergence of $S(H_j,H_0)$ to $S(H,H_0)$ and vice versa as $j \to \infty$. (Actually by taking a slightly more general formulation we are not only able to treat continuity but also holomorphy of the scattering operators with respect to appropriate complex parameters).

In Sec. 5 we again give various applications to potential scattering. In Sec. 5A we consider Coulomb-type interactions in $\mathbb{R}^3$ and establish continuity of scattering amplitudes, on-shell scattering operators, nonforward total cross sections, and scattering operators with respect to Coulomb and short-range interactions. Holomorphy of the scattering operator for point interactions in the boundary condition parameter and continuity of scattering operators in the zero-range limit are treated in Secs. 5B and 5C. Finally holomorphy of scattering operators in the coupling constant and continuity of scattering phase shifts, etc., for spherically symmetric long-range potentials are investigated in Secs. 5D and 5E.

2. CONTINUITY OF WAVE OPERATORS—A TIME-DEPENDENT APPROACH

Throughout this paper $\overline{B_r(\alpha_0)} \subset \mathscr{B}$ denotes a closed sphere of radius $r > 0$ at α_0 in a finite-dimensional Banach space $\mathscr{B}$. For $\alpha \in \overline{B_r(\alpha_0)}$ H_α represents a self-adjoint operator in some (separable) Hilbert space $\mathscr{H}$. Let H_0 be a self-adjoint operator in another (separable) Hilbert space $\mathscr{H}_0$ and J a bounded linear operator from $\mathscr{H}_0$ to $\mathscr{H}$. Then, in the time-dependent approach to scattering with a pair of spaces,[7,8] the problem of continuity of wave operators naturally splits up into two parts:

(I) Existence of the wave operators $\Omega_\pm(H_\alpha,H_0;J)$, i.e., existence of the strong limits

$$\Omega_\pm(H_\alpha,H_0;J) = s\text{-}\lim_{t \to \pm\infty} e^{itH_\alpha} J e^{-itH_0} P_{ac}(H_0), \quad \alpha \in \overline{B_r(\alpha_0)}, \tag{2.1}$$

where $P_{ac}(H_0)$ denotes the projection onto the absolutely continuous subspace of H_0;

(II) strong continuity of the wave operators at $\alpha_0 \in \mathscr{B}$, i.e.,

[a] Alexander von Humboldt Research Fellow
[b] On leave of absence from Institut für Theoretische Physik, Universität Graz, Austria

 0022-2488/83/061516-13$02.50 © 1983 American Institute of Physics

$$s\text{-}\lim_{\alpha\to\alpha_0} \Omega_\pm(H_\alpha,H_0;J) = \Omega_\pm(H_{\alpha_0},H_0;J), \quad \alpha\in\overline{B_r(\alpha_0)}.$$

Since the scattering operator $S(H_\alpha,H_0;J)$ is given by

$$S(H_\alpha,H_0;J) = \Omega_+(H_\alpha,H_0;J)^*\Omega_-(H_\alpha,H_0;J), \quad \alpha\in\overline{B_r(\alpha_0)}, \tag{2.2}$$

(II) immediately yields weak continuity of $S(H_\alpha,H_0;J)$ at $\alpha=\alpha_0$. In those cases where $S(H_\alpha,H_0;J)$ is isometric for all $\alpha\in\overline{B_r(\alpha_0)}$ even strong continuity at $\alpha=\alpha_0$ is guaranteed. Therefore we first concentrate on the existence and strong continuity of $\Omega_\pm(H_\alpha,H_0;J)$. Our main results in this direction rely on ideas of Kato[1] who generalized previous attempts of Schechter[9] (see also Refs. 10–14) to extend Cook's criterion.[6] We have

Theorem 2.1: Let $\overline{B_r(\alpha_0)}$, J, H_α, H_0, $\mathscr{H}_0$, and $\mathscr{H}$ as introduced above. Assume furthermore $\mathscr{D}_0$ to be a total subset of $\mathscr{H}_{ac}(H_0)=P_{ac}(H_0)\mathscr{H}_0$ and suppose $F:\mathbb{R}\to\mathbb{R}$ to be continuous such that $xF(x)$ is bounded on $\mathbb{R}$ and $\mathscr{R}(F(H_0)P_{ac}(H_0))$ is dense in $\mathscr{H}_{ac}(H_0)$. Assume

(A^+): for any $\phi\in\mathscr{D}_0$ and $\alpha\in\overline{B_r(\alpha_0)}\setminus\{\alpha_0\}$ the function

$$\Psi_{\alpha,\phi}(t) = \| [H_\alpha F(H_\alpha)JF(H_0) - F(H_\alpha)JH_0F(H_0)]\, e^{-itH_0}\phi \|$$

belongs to $\mathscr{L}^1([0,\infty);dt)$;

(B^+): for any $\phi\in\mathscr{D}_0$ and any $\alpha\in\overline{B_r(\alpha_0)}\setminus\{\alpha_0\}$

$$s-\lim_{t\to\infty} [F(H_\alpha)J - JF(H_0)]\, e^{-itH_0}\phi = 0.$$

Then

(i) the wave operator $\Omega_+(H_\alpha,H_0;J)$ exists for all $\alpha\in\overline{B_r(\alpha_0)}\setminus\{\alpha_0\}$.

Suppose in addition

(C): for all $\alpha\in\overline{B_r(\alpha_0)}\setminus\{\alpha_0\}$

$$s-\lim_{\alpha\to\alpha_0}(H_\alpha - z)^{-1} = (H_{\alpha_0}-z)^{-1} \text{ for some } z\in\mathbb{C},$$

$\operatorname{Im} z\neq 0$;

(D^+): for any sequence $\{\alpha_n\}_{n\in\mathbb{N}}\subset\overline{B_r(\alpha_0)}\setminus\{\alpha_0\}$, $\alpha_n\xrightarrow[n\to\infty]{}\alpha_0$, for any $\epsilon>0$ and any $\phi\in\mathscr{D}_0$ there are $T_0=T_0(\epsilon,\phi,\{\alpha_n\})$ and $N_0=N_0(\epsilon,\phi)$ such that for all $n\geqslant N_0$

$$\int_{T_0}^{\infty} dt\, \Psi_{\alpha_n,\phi}(t) \leqslant \epsilon;$$

(E^+): condition (B^+) holds for all $\alpha\in\overline{B_r(\alpha_0)}$. Then

(ii) $\Omega_+(H_\alpha,H_0;J)$ exist for all $\alpha\in\overline{B_r(\alpha_0)}$ and are strongly continuous at α_0, i.e.,

$$s-\lim_{\substack{\alpha\to\alpha_0\\ \alpha\neq\alpha_0}} \Omega_+(H_\alpha,H_0;J) = \Omega_+(H_{\alpha 0},H_0;J).$$

A simple realization of conditions (A^+)–(E^+) is presented in Sec. 3 in the context of potential scattering. Here we discuss sufficient conditions for (B^+) and (E^+):

Lemma 2.1: Suppose that $\overline{B_r(\alpha_0)}$, J, H_α, H_0, $\mathscr{D}_0$, $\mathscr{H}_0$, $\mathscr{H}$, and F have the same meaning as in Theorem 2.1. Assume $\alpha\in\overline{B_r(\alpha_0)}\setminus\{\alpha_0\}$ and suppose that either

(B_1): $F(H_\alpha)J - JF(H_0)$ is compact or

(B_2^+): for any fixed $\phi\in\mathscr{D}_0$ there is a $p=p(\alpha,\phi)\in[1,\infty)$ such that the function

$$\Phi_{\alpha,\phi}(t) = \| [F(H_\alpha)J - JF(H_0)]\, e^{-itH_0}\phi \|$$

is an element of $\mathscr{L}^p([0,\infty);dt)$. Then condition ($B^+$) of Theorem 2.1 holds. If furthermore

(E_1): $F(H_\alpha)J - JF(H_0)$ is compact for all $\alpha\in\overline{B_r(\alpha_0)}$ or

(E_2): condition (B^+) holds and $F(H_\alpha)\xrightarrow[\alpha\to\alpha_0]{}F(H_{\alpha_0})$ in norm or

(E_3^+): condition (C) holds and for any sequence $\{\alpha_n\}_{n\in\mathbb{N}}\subset\overline{B_r(\alpha_0)}\setminus\{\alpha_0\}$, $\alpha_n\xrightarrow[n\to\infty]{}\alpha_0$, any $\epsilon>0$, and any $\phi\in\mathscr{D}_0$ there are $T_0=T_0(\epsilon,\phi,\{\alpha_n\})$ and $N_0=N_0(\epsilon,\phi)$ such that for all $n\geqslant N_0$ and some $p=p(\{\alpha_n\},\phi)\in[1,\infty)$

$$\int_{T_0}^{\infty} dt\, (\Phi_{\alpha_n,\phi}(t))^p < \epsilon$$

holds; then condition (E^+) of Theorem 2.1 is fulfilled.

Remark 2.1: (a) Clearly the statements of Theorem 2.1 and Lemma 2.1 hold for $\Omega_-(H_\alpha,H_0;J)$ as well if (A^+), (B^+), (B_2^+), (D^+), (E^+), and (E_3^+) are replaced by the corresponding conditions for $t\to-\infty$. We denote these conditions by (A^-), . . . ,(E_3^-).

(b) We note that conditions (A^+) and (B_2^+), and (D^+) and (E_3^+) are not necessarily independent of each other. In particular, if we choose $F(x)=(x-z)^{-1}$, $\operatorname{Im} z\neq 0$ then $\Psi_{\alpha,\phi}(t)=\Phi_{\alpha,\phi}(t)$ and, therefore, (A^+) and (B_2^+) coincide if we take $p=1$ in (B_2^+). Similarly, if (C) holds then (D^+) implies (E_3^+) if we again choose $p=1$.

Before giving the proofs of Theorem 2.1 and Lemma 2.1 we would like to add some further comments on possible extensions and applications of Theorem 2.1:

Remark 2.2: (a) If J' is another bounded linear operator from $\mathscr{H}_0$ to $\mathscr{H}$ and

$$s-\lim_{t\to\infty}(J'-J)\,e^{-itH_0}P_{ac}(H_0)=0$$

then (if they exist) $\Omega_+(H_\alpha,H_0;J)=\Omega_+(H_\alpha,H_0;J')$. In particular, if $\mathscr{H}_0=\mathscr{H}$, $J'=1$, and the condition $(1-J)F(H_0)$ is compact is added to the hypotheses (A^+)–(E^+) of Theorem 2.1, then existence and continuity of wave operators in potential scattering with strong local singularities of the interaction [cf. Refs. 2, 3, 5 (p. 347), 8 (pp. 21 and 68), 11, 15, 16, and Theorem 3.2] follows. (b) Theorem 2.1 obviously generalizes to situations where the free unitary group e^{-itH_0} is replaced by a unitary evolution of the type $e^{-ig(t,H_0)}$. In this case condition (A^+) should be replaced by (A'^+):

for any $\phi\in\mathscr{D}_0$ and any $\alpha\in\overline{B_r(\alpha_0)}\setminus\{\alpha_0\}$ the function

$$\Psi'_{\alpha,\phi}(t) = \| [H_\alpha F(H_\alpha)JF(H_0) - F(H_\alpha)J\dot{g}(t,H_0)F(H_0)]\, e^{-ig(t,H_0)}\phi \|$$

is an element of $\mathscr{L}^1([0,\infty);dt)$ and (since $e^{-ig(t,H_0)}$ in general has no group properties) the compactness conditions (B_1^+) and (E_1^+) instead of (B^+) and (E^+) should be used [$\dot{g}(t,H_0)$ denotes the derivative of $g(t,H_0)$ with respect to t]. For instance, existence and strong continuity of the modified wave operators for potential scattering in $\mathscr{L}^2(\mathbb{R}^n)$ with long-range interactions[17] [with tails $|\mathbf{x}|^{-\epsilon}$, $\epsilon>0$] follows in this way.

The proof of Theorem 2.1 uses the following argument (which therefore is stated as a lemma) several times.

Lemma 2.2: Let $U_\alpha(t)$, $t\in\mathbb{R}$, $\alpha\in\overline{B_r(\alpha_0)}$ be bounded linear operators on $\mathscr{H}_0$, $C:\mathscr{H}_0\to\mathscr{H}_0$ a bounded linear operator with dense range, and define $\tilde{U}_\alpha(t)=U_\alpha(t)\,C$. Suppose

(1) $\sup_{t\in\mathbb{R}}\sup_{\alpha\in\overline{B_r(\alpha_0)}}\|U_\alpha(t)\|=M<\infty$;

(2) for fixed $\alpha_1\in\overline{B_r(\alpha_0)}$, $s-\lim_{t\to\infty}\tilde{U}_{\alpha_1}(t)$ exists and is denoted by $\tilde{U}_{\alpha_1}(\infty)$;

(3) $s-\lim_{\substack{\alpha\to\alpha_0\\ \alpha\neq\alpha_0}}\tilde{U}_\alpha(\infty)$ and $s-\lim_{t\to\infty}\tilde{U}_{\alpha_0}(t)$ exist and are equal.

Then we have

(a) $s-\lim_{t\to\infty}U_{\alpha_1}(t)$ exists, we denote it by $U_{\alpha_1}(\infty)$;

(b) $s-\lim_{\substack{\alpha\to\alpha_0\\ \alpha\neq\alpha_0}}U_\alpha(\infty)$ and $s-\lim_{t\to\infty}U_{\alpha_0}(t)$ exist and are equal.

Now we indicate the

Proof of Theorem 2.1 (i): We introduce

$$\tilde{A}_\alpha(t)=F(H_\alpha)\,e^{itH_\alpha}Je^{-itH_0}F(H_0)P_{ac}(H_0),$$

$$t\in\mathbb{R},\quad \alpha\in\overline{B_r(\alpha_0)}$$

and note that by assumption (A^+)

$$\|\tilde{A}_\alpha(t)\phi-\tilde{A}_\alpha(s)\phi\|=\left\|\int_s^t d\tau\,\frac{d}{d\tau}\tilde{A}_\alpha(\tau)\phi\right\|\leqslant\int_s^t d\tau\,\Psi_{\alpha,\phi}(\tau)\to 0$$

as $t>s\to\infty$ for $\alpha\in\overline{B_r(\alpha_0)}\setminus\{\alpha_0\}$. Thus $\tilde{A}_\alpha(\infty)=s-\lim_{t\to\infty}\tilde{A}_\alpha(t)$ exists for all $\alpha\in\overline{B_r(\alpha_0)}\setminus\{\alpha_0\}$. Defining

$$A_\alpha(t)=F(H_\alpha)e^{itH_\alpha}Je^{-itH_0}P_{ac}(H_0),$$

$$t\in\mathbb{R},\quad \alpha\in\overline{B_r(\alpha_0)},$$

we infer from Lemma 2.2(a) that

$$A_\alpha(\infty)=s-\lim_{t\to\infty}A_\alpha(t)=\Omega_+(H_\alpha,H_0;F(H_\alpha)J)$$

exists for all $\alpha\in\overline{B_r(\alpha_0)}\setminus\{\alpha_0\}$. By condition (B^+) we obtain the existence of $s-\lim_{t\to\infty}e^{itH_\alpha}Je^{-itH_0}P_{ac}(H_0)F(H_0)$. Applying Lemma 2.2(a) again we get assertion (i) of Theorem 2.1. Next we perform the

Proof of Theorem 2.1 (ii): Let $\{\alpha_n\}_{n\in\mathbb{N}}\subset\overline{B_r(\alpha_0)}\setminus\{\alpha_0\}$, $\alpha_n\xrightarrow[n\to\infty]{}\alpha_0$ and choose $\epsilon>0$, $\phi\in\mathscr{D}_0$, T_0, and N_0 according to hypothesis (D^+). Then for all $n\geqslant N_0$ and all $t\geqslant T_0$

$$\|\tilde{A}_{\alpha_n}(t)\phi-\tilde{A}_{\alpha_n}(\infty)\phi\|\leqslant\int_t^\infty d\tau\left\|\frac{d}{d\tau}\tilde{A}_{\alpha_n}(\tau)\phi\right\|\leqslant\int_{T_0}^\infty d\tau\,\Psi_{\alpha_n,\phi}(\tau)\leqslant\epsilon.\tag{2.3}$$

In order to prove that $\{\tilde{A}_{\alpha_n}(\infty)\phi\}_{n\in\mathbb{N}}$ is a Cauchy sequence in $\mathscr{H}$ we first observe that due to assumption (C) $\{\tilde{A}_{\alpha_n}(t_1)\phi\}_{n\in\mathbb{N}}$ is a Cauchy sequence in $\mathscr{H}$ for any fixed $t_1\in\mathbb{R}$. Thus for fixed $t_1\geqslant T_0$ there is an $N_1\in\mathbb{N}$ such that $\|\tilde{A}_{\alpha_n}(t_1)\phi-\tilde{A}_{\alpha_m}(t_1)\phi\|<\epsilon$ for all $n,m\geqslant N_1$. Together with (2.3) this implies for all $n,m\geqslant\max(N_0,N_1)$

$$\|\tilde{A}_{\alpha_n}(\infty)\phi-\tilde{A}_{\alpha_m}(\infty)\phi\|\leqslant\|\tilde{A}_{\alpha_n}(\infty)\phi-\tilde{A}_{\alpha_n}(t_1)\phi\|+\|\tilde{A}_{\alpha_n}(t_1)\phi-\tilde{A}_{\alpha_m}(t_1)\phi\|+\|\tilde{A}_{\alpha_m}(t_1)\phi-\tilde{A}_{\alpha_m}(\infty)\phi\|\leqslant 3\epsilon.$$

Hence $\tilde{A}_{\alpha_0}(\infty)=s-\lim_{n\to\infty}\tilde{A}_{\alpha_n}(\infty)\phi$ exists. Again by hypothesis (C) we have for all $t\in\mathbb{R}$

$$\tilde{A}_{\alpha_n}(t)\phi-\tilde{A}_{\alpha_n}(\infty)\phi\xrightarrow[n\to\infty]{s}\tilde{A}_{\alpha_0}(t)\phi-\tilde{A}_{\alpha_0}(\infty)\phi$$

and by (2.3) we get for all $t>T_0$

$$\|\tilde{A}_{\alpha_0}(t)\phi-\tilde{A}_{\alpha_0}(\infty)\phi\|=\lim_{n\to\infty}\|\tilde{A}_{\alpha_n}(t)\phi-\tilde{A}_{\alpha_n}(\infty)\phi\|\leqslant\epsilon.$$

Therefore $s-\lim_{t\to\infty}\tilde{A}_{\alpha_0}(t)$ exists and

$$s-\lim_{\substack{\alpha\to\alpha_0\\ \alpha\neq\alpha_0}}\tilde{A}_\alpha(\infty)=s-\lim_{t\to\infty}\tilde{A}_{\alpha_0}(t)=\tilde{A}_{\alpha_0}(\infty).$$

Observe that $\tilde{A}_\alpha(t)=A_\alpha(t)\,F(H_0)$ and apply Lemma 2.2(b) to get that $A_\alpha(\infty)=s-\lim_{t\to\infty}A_\alpha(t)$ exists for all $\alpha\in\overline{B_r(\alpha_0)}$ and

$$s-\lim_{\substack{\alpha\to\alpha_0\\ \alpha\neq\alpha_0}}A_\alpha(\infty)=s-\lim_{t\to\infty}A_{\alpha_0}(t)=A_{\alpha_0}(\infty).$$

Because of assumption (E^+) we obtain for all $\alpha\in\overline{B_r(\alpha_0)}$

$$s-\lim_{t\to\infty}A_\alpha(t)=s-\lim_{t\to\infty}\Omega_\alpha(t)\,F(H_0),$$

where

$$\Omega_\alpha(t)=e^{itH_\alpha}Je^{-itH_0}P_{ac}(H_0).\tag{2.4}$$

Another application of Lemma 2.2(b) yields

$$s-\lim_{\substack{\alpha\to\alpha_0\\ \alpha\neq\alpha_0}}(s-\lim_{t\to\infty}\Omega_\alpha(t))=s-\lim_{t\to\infty}(s-\lim_{\substack{\alpha\to\alpha_0\\ \alpha\neq\alpha_0}}\Omega_\alpha(t))$$

and the proof is complete.

Proof of Lemma 2.1: (B_1) clearly implies (B^+). Suppose that (B_2^+) holds. Then, following Ref. 1 we note that the set of all vectors

$$\Psi(g,\phi)=\int_{\mathbf{R}}ds\,g(s)e^{-isH_0}\phi,$$

$$g\in C_0^\infty(\mathbb{R}),\quad \phi\in\mathscr{D}_0\tag{2.5}$$

is total in $\mathscr{H}_{ac}(H_0)$ and thus it suffices to control the limit of $[F(H_\alpha)J-JF(H_0)]e^{-itH_0}$ as $t\to\infty$ on vectors of the form (2.5). Observing that

$$e^{-itH_0}\Psi(g,\phi)=\int_{\mathbf{R}}ds\,g(s-t)e^{-isH_0}\phi$$

and assuming $\operatorname{supp}(g)\subseteq[a,b]$, we get

$$\|[F(H_\alpha)J-JF(H_0)]e^{-itH_0}\Psi(g,\phi)\|\leqslant\int_{a+t}^{b+t}ds\,|g(s-t)|\Phi_{\alpha,\phi}(s)\leqslant\|g\|_{p'}\left(\int_{a+t}^{b+t}ds(\Phi_{\alpha,\phi}(s)^p\right)^{1/p}\xrightarrow[t\to\infty]{}0,$$

$$\frac{1}{p} + \frac{1}{p'} = 1 \tag{2.6}$$

because of (B_2^+). Thus (B_2^+) implies (B^+). (E_1) obviously implies (E^+) and condition (E_2) clearly yields (E^+) since

$$\begin{aligned}\| [F(H_{\alpha_0})J - JF(H_0)]e^{-itH_0}\phi \| &\leqslant \| [F(H_\alpha)J - JF(H_0)]e^{-itH_0}\phi \| \\ &+ \| F(H_{\alpha_0}) - F(H_\alpha)\| \, \| J \| \, \|\phi\|.\end{aligned}$$

Assume condition (E_3^+). Take any sequence

$\{\alpha_n\}_{n\in\mathbb{N}} \subset \overline{B_r(\alpha_0)}\setminus\{\alpha_0\}$, $a_n \underset{n\to\infty}{\longrightarrow} \alpha_0$, any $\epsilon > 0$, any $\phi\in\mathscr{D}_0$, and choose T_0 and N_0 according to (E_3^+). Then for any vector $\Psi(g,\phi)$ of the type (2.5), all $n\geqslant N_0$, and all $t\geqslant T_0$ we get as in (2.6) (supp$(g)\subseteq[a,b]$)

$$\begin{aligned}\| \left[F(H_{\alpha_n})J - JF(H_0)\right]e^{-itH_0}\Psi(g,\phi)\| &\leqslant \| g\|_{p'} \left(\int_{T_0}^{\infty} ds\, (\Phi_{\alpha_n,\phi}(s))^p\right)^{1/p} \\ &\leqslant \| g\|_{p'}\epsilon^{1/p}\end{aligned}$$

by (E_3^+). Thus (C) implies

$$s - \lim_{t\to\infty} [F(H_{\alpha_0})J - JF(H_0)]e^{-itH_0}\Psi(g,\phi) = 0$$

and hence (E^+) holds.

Remark 2.3: By the explicit definition (2.1) of the wave operators it is evident that strong resolvent convergence of H_α to H_{α_0} is the appropriate way in which H_α should approach H_{α_0} as $\alpha\to\alpha_0$. In particular, this implies that

$$G(H_\alpha) \underset{\alpha\to\alpha_0}{\overset{s}{\longrightarrow}} G(H_{\alpha_0})$$

for any bounded and continuous function G on $\mathbb{R}$ and $s - \lim_{\alpha\to\alpha_0}\exp(itH_\alpha) = \exp(itH_{\alpha_0})$ uniformly in t on compact intervals.[18,19] Thus if condition (C) is assumed, problem (II) essentially reduces to the problem of interchanging the $t\to\pm\infty$ and the $\alpha\to\alpha_0$ limits. More precisely, problem (II) requires [cf. (2.4)]

$$s - \lim_{\substack{\alpha\to\alpha_0\\ \alpha\neq\alpha_0}} (s - \lim_{t\to\pm\infty} \Omega_\alpha(t)) = s - \lim_{t\to\pm\infty} (s - \lim_{\substack{\alpha\to\alpha_0\\ \alpha\neq\alpha_0}} \Omega_\alpha(t)),$$

where the existence of the inner limits is assumed. But necessary and sufficient conditions for the existence of the iterated limits and the equality of both are known to be the conditions of incomplete uniform convergence[20] of $\{\Omega_\alpha(t)\}$, i.e., for any sequence $\{\alpha_n\}_{n\in\mathbb{N}} \subset \overline{B_r(\alpha_0)}\setminus\{\alpha_0\}$, $\alpha_n\to\alpha_0$, any $\phi\in\mathscr{H}_{ac}(H_0)$, and any $\epsilon > 0$ there is a $T_0 = T_0(\epsilon,\phi) > \sigma$ such that for any $t\geqslant T_0$ there is a $N_0 = N_0(\epsilon,\phi,t)\in\mathbb{N}$ such that for all $n\geqslant N_0 \| \Omega_{\alpha_n}(t)\phi - \Omega_{\alpha_n}(\infty)\phi\| < \epsilon$.

In this context it seems to be impossible to realize these conditions of incomplete uniform convergence directly. We therefore propose to use a kind of "uniform smallness at infinity" in Theorem 2.1, condition (D^+). If we rely on the basic Cook estimate, condition (D^+) is the best possible one.

For previous results on continuity of wave and scattering operators based on the original form of Cook's criterion[6] we refer to Narnhofer[2] [see also Refs. 3, 4, and 5 (p. 329)]. For similar results based on trace norm convergence of the resolvents of H_α see Refs. 8 (pp. 27, 74, and 387), Ref. 18 (p. 551), Refs. 21, and Ref. 22 (p. 112).

Finally we state some consequences of Theorem 2.1 for the scattering operator $S(H_\alpha,H_0;J)$ in (2.2): If $\Omega_\pm(H_\alpha,H_0;J)$ are partially isometric with $\mathscr{R}(\Omega_-(H_\alpha,H_0;J))\subseteq\mathscr{R}(\Omega_+(H_\alpha,H_0;J))$ for all $\alpha\in\overline{B_r(\alpha_0)}$ (see Refs. 7 and 14 for sufficient conditions) then $S(H_\alpha,H_0;J)$ is isometric on $\mathscr{H}_{ac}(H_0)$ and thus Theorem 2.1 yields strong convergence of $S(H_\alpha,H_0;J)$ as $\alpha\to\alpha_0$. We summarize this in

Corollary 2.1: Assume the conditions of Theorem 2.1 to hold. Then $S(H_\alpha,H_0;J)$ is weakly continuous at $\alpha = \alpha_0$. If in addition, for all $\alpha\in\overline{B_r(\alpha_0)}$, $\Omega_\pm(H_\alpha,H_0;J)$ are partially isometric with $\mathscr{R}(\Omega_-(H_\alpha,H_0;J))\subseteq\mathscr{R}(\Omega_+(H_\alpha,H_0;J))$ then $S(H_\alpha,H_0;J)$ is strongly continuous at $\alpha = \alpha_0$.

3. APPLICATIONS TO SINGULAR POTENTIALS IN QUANTUM SCATTERING THEORY

In this section we show how to realize conditions $(A^\pm),\ldots,(E^\pm)$ of Theorem 2.1 in the framework of potential scattering. Furthermore, physical applications to the problem of scattering into cones are stressed.

We choose $\mathscr{H} = \mathscr{H}_0 = \mathscr{L}^2(\mathbb{R}^n)$, $n\geqslant 3$, $H_0 = -\Delta$ (the usual kinetic energy operator in $\mathscr{L}^2(\mathbb{R}^n)$). H_α are distinguished self-adjoint realizations of formal differential operators of the type $-\Delta + V_\alpha + W_\alpha$ in $\mathscr{L}^2(\mathbb{R}^n)$, where V_α, $W_\alpha\in\mathscr{L}^1_{loc}(\mathbb{R}^n\setminus\Sigma)$ are real and Σ is a compact set of Lebesgue measure zero. In addition we introduce the norms

$$\| V\|_{1,k} = \int_{\mathbb{R}^n} d^n x(1 + |\mathbf{x}|^2)^{-k/2}|V(\mathbf{x})|,$$

$$V\in\mathscr{L}^1(\mathbb{R}^n;(1 + |\mathbf{x}|^2)^{-k/2}d^n x), k > 0,$$

$$\| W\|_{2,l} = \left(\int_{\mathbb{R}^n} d^n x(1 + |\mathbf{x}|^2)^{-l/2}|W(\mathbf{x})|^2\right)^{1/2},$$

$$W\in\mathscr{L}^2(\mathbb{R}^n;(1 + |\mathbf{x}|^2)^{-l/2}\, d^n x),\ l > 0,$$

and the function $J_R(\mathbf{x})$

$$J_R(\mathbf{x})\in C^\infty(\mathbb{R}^n);\quad J_R(\mathbf{x}) = \begin{cases}0 & \text{if } |\mathbf{x}|\leqslant R,\\ 1 & \text{if } |\mathbf{x}|\geqslant 2R,\end{cases}$$

where $\Sigma\subseteq\{\mathbf{x}\in\mathbb{R}^n \mid |\mathbf{x}|\leqslant R\}$. Furthermore we use the notation $V_\pm(\mathbf{x}) = [|V(\mathbf{x})| \pm V(\mathbf{x})]/2$. Then we have

Theorem 3.1: Let $V_\alpha\in\mathscr{L}^1_{loc}(\mathbb{R}^n)$, $W_\alpha\in\mathscr{L}^2_{loc}(\mathbb{R}^n)$, $n\geqslant 3$, for all $\alpha\in\overline{B_r(\alpha_0)}$ and assume

(1) $V_\alpha \underset{\substack{\alpha\to\alpha_0\\ \alpha\neq\alpha_0}}{\longrightarrow} V_{\alpha_0}$ in $\mathscr{L}^1_{loc}(\mathbb{R}^n)$,

$W_\alpha \underset{\substack{\alpha\to\alpha_0\\ \alpha\neq\alpha_0}}{\longrightarrow} W_{\alpha_0}$ in $\mathscr{L}^1_{loc}(\mathbb{R}^n)$;

(2) for some $0 < k^\pm < n-2$ and $0 < l^\pm < n-2$ there are constants $M^\pm$ and $N^\pm$ such that

$$\| V_\alpha^\pm\|_{1,k^\pm} \leqslant M^\pm, \| W_\alpha^\pm\|_{2,l^\pm} \leqslant N^\pm,$$

for all $\alpha\in\overline{B_r(\alpha_0)}\setminus\{\alpha_0\}$,

(3) there are constants $0\leqslant\theta < 1$ and $0\leqslant K < \infty$ such that for all $\alpha\in\overline{B_r(\alpha_0)}\setminus\{\alpha_0\}$

(a) $\mathscr{D}(H_0^{1/2})\subseteq\mathscr{D}((V_\alpha^-)^{1/2})\cap\mathscr{D}((W_\alpha^-)^{1/2})$

and

(b) $\|(V_\alpha^-)^{1/2} g\|^2 + \|(W_\alpha^-)^{1/2} g\|^2$

$$\leqslant \theta \|H_0^{1/2} g\|^2 + K \|g\|^2, \; g \in \mathscr{D}(H_0^{1/2}).$$

Then, using quadratic forms, $H_\alpha = H_0 \dot{+} V_\alpha \dot{+} W_\alpha$, $\alpha \in \overline{B_r(\alpha_0)}$ are well defined self-adjoint operators in $\mathscr{L}^2(\mathbb{R}^n)$ and the wave operators $\Omega_\pm(H_\alpha, H_0) = \Omega_\pm(H_\alpha, H_0; 1)$ exist for all $\alpha \in \overline{B_r(\alpha_0)}$ and are strongly continuous at $\alpha = \alpha_0$.

Corollary 3.1: Assume that $V \in \mathscr{L}^1_{\rm loc}(\mathbb{R}^n)$, $W \in \mathscr{L}^2_{\rm loc}(\mathbb{R}^n)$, $n \geqslant 3$ such that $\|V^\pm\|_{1,k^\pm} < \infty$, $\|W^\pm\|_{2,l^\pm} < \infty$ for some $0 < k^\pm < n-2, 0 < l^\pm < n-2$, and suppose V^- and W^- fulfill 3(a) and 3(b) of Theorem 3.1. Let g_j be measurable functions on $\mathbb{R}^n$ such that

$$0 \leqslant g_j(\mathbf{x}) \leqslant 1, g_j(\mathbf{x}) \underset{j\to\infty}{\longrightarrow} 1 \text{ for a.e. } \mathbf{x} \in \mathbb{R}^n, \; j \in \mathbb{N},$$

and define

$$V_j(\mathbf{x}) = g_j(\mathbf{x}) V(\mathbf{x}), W_j(\mathbf{x}) = g_j(\mathbf{x}) W(\mathbf{x}), \; j \in \mathbb{N}.$$

Then V_j, W_j, satisfy the hypotheses of Theorem 3.1 and hence $H_j = H_0 \dot{+} V_j \dot{+} W_j, j \in \mathbb{N}$ and $H = H_0 \dot{+} V \dot{+} W$ defined through quadratic forms are self-adjoint in $\mathscr{L}^2(\mathbb{R}^n)$ and $s - \lim_{j\to\infty} \Omega_\pm(H_j, H_0) = \Omega_\pm(H, H_0)$.

Corollary 3.1 shows that cut-off (take g_j of compact support) and screening procedures [cf. Ref. 5 (p. 333) and Ref. 23] are applicable under extremely weak conditions on the interaction.

Our next result includes strongly repulsive singularities of the potentials.

Theorem 3.2: Let $0 \leqslant V_\alpha \in \mathscr{L}^1_{\rm loc}(\mathbb{R}^n \setminus \Sigma)$, $0 \leqslant W_\alpha \in \mathscr{L}^2_{\rm loc}(\mathbb{R}^n \setminus \Sigma)$, $n \geqslant 3, \alpha \in \overline{B_r(\alpha_0)}$ and assume

(1) $V_\alpha \underset{\substack{\alpha\to\alpha_0 \\ \alpha\neq\alpha_0}}{\longrightarrow} V_{\alpha_0}$ in $\mathscr{L}^1_{\rm loc}(\mathbb{R}^n \setminus \Sigma)$,

$$W_\alpha \underset{\substack{\alpha\to\alpha_0 \\ \alpha\neq\alpha_0}}{\longrightarrow} W_{\alpha_0} \text{ in } \mathscr{L}^1_{\rm loc}(\mathbb{R}^n \setminus \Sigma);$$

(2) for some $0 < k < n-2$ and $0 < l < n-2$ there are constants M and N such that

$$\|V_\alpha J_R\|_{1,k} \leqslant M, \; \|W_\alpha J_R\|_{2,l} \leqslant N,$$

for all $\alpha \in \overline{B_r(\alpha_0)} \setminus \{\alpha_0\}$. Then the closure of

$$\dot{Q}_{H_\alpha}(g,h) = (\nabla g, \nabla h) + (V_\alpha^{1/2} g, V_\alpha^{1/2} h) + (W_\alpha^{1/2} g, W_\alpha^{1/2} h),$$
$$g, h \in C_0^\infty(\mathbb{R}^n \setminus \Sigma), \; \alpha \in \overline{B_r(\alpha_0)},$$

uniquely defines self-adjoint operators H_α in $\mathscr{L}^2(\mathbb{R}^n)$, and $\Omega_\pm(H_\alpha, H_0)$ [which equal $\Omega_\pm(H_\alpha, H_0; J_R)$] exist for all $\alpha \in \overline{B_r(\alpha_0)}$ and are strongly continuous at $\alpha = \alpha_0$.

Remark 3.1: (a) For simplicity we took $V_\alpha \geqslant 0$, $W_\alpha \geqslant 0$, in Theorem 3.2. It is clear from Theorem 3.1 how to incorporate relatively form-bounded attractive interactions. (b) In Theorems 3.1 and 3.2. V_α are allowed to decay at infinity like $|\mathbf{x}|^{-2-\epsilon}$, $\epsilon > 0$ whereas W_α may decay like $|\mathbf{x}|^{-1-\delta}$, $\delta > 0$ as $|\mathbf{x}| \to \infty$.

Corollary 3.2: Under the conditions of Theorem 3.1 or 3.2 $S(H_\alpha, H_0) = S(H_\alpha, H_0; 1)$ is weakly continuous at $\alpha = \alpha_0$. If in addition to the assumptions of Theorem 3.1 or 3.2, $\mathscr{R}(\Omega_-(H_\alpha, H_0)) \subseteq \mathscr{R}(\Omega_+(H_\alpha, H_0))$, $\alpha \in \overline{B_r(\alpha_0)}$ also holds, then $S(H_\alpha, H_0)$ is strongly continuous at $\alpha = \alpha_0$.

Remark 3.2: (a) In the special case $\Sigma = \{0\}$, results similar to Corollary 3.1 under slightly stronger conditions on the potential have been derived by Robinson.[3] (b) For sufficient conditions implying completeness of $\Omega_\pm(H_\alpha, H_0)$ [i.e., $\mathscr{R}(\Omega_\pm(H_\alpha, H_0)) = \mathscr{H}_{ac}(H_0)$] compare Ref. 3, Ref. 8 (Chap. XI), Refs. 11–13, 16, 22 (p. 105), and Refs. 24; (in particular if V_α and W_α in Theorem 3.2 have compact support then asymptotic completeness is fulfilled[16]). (c) Extensions of Theorem 3.2 including hard cores are obviously possible.

We now turn to the proofs of these facts. In order to verify assumptions $(A^\pm) \cdots (E^\pm)$ of Theorem 2.1 we choose $F(x) = (x - z)^{-1}$, $\operatorname{Im} z \neq 0$. Thus using Lemma 2.1 and Remark 2.1(b), we only have to prove conditions (C) and $(D^\pm)$ since

$$(C) \wedge (D^\pm) \Rightarrow (E_3^\pm) \; (\text{if } p = 1) \Rightarrow (E^\pm)$$

and

$$(D^\pm) \Rightarrow (A^\pm) \Rightarrow (B_2) \; (\text{if } p = 1) \Rightarrow (B^\pm).$$

First we discuss condition (C). Actually, by following ideas of Simon[25] we prove a more general result than needed for Theorems 3.1 and 3.2, which is of its own interest.

Lemma 3.1: Let $\Omega \subset \mathbb{R}^n$, $n \geqslant 1$ be an open nonempty set. Let

$$Q_D(g,h) = \overline{(\nabla g, \nabla h)|_{\mathscr{C}_0^\infty(\Omega)}}$$

be the Dirichlet form in $\mathscr{L}^2(\Omega)$ and denote by $-\Delta_D$ the associated Dirichlet Laplacian. Assume in addition

(1) $0 \leqslant V_m$, $V \in \mathscr{L}^1_{\rm loc}(\Omega)$, $m \in \mathbb{N}$, $V_m \underset{m\to\infty}{\longrightarrow} V$ in $\mathscr{L}^1_{\rm loc}(\Omega)$,

$$0 \leqslant W_m, \; W \in \mathscr{L}^1_{\rm loc}(\Omega), \; m \in \mathbb{N}, \; W_m \underset{m\to\infty}{\longrightarrow} W \text{ in } \mathscr{L}^1_{\rm loc}(\Omega);$$

(2) there are constants $0 \leqslant \theta < 1$ and $0 \leqslant K < \infty$ such that for all $m \in \mathbb{N}$, $\mathscr{D}(W_m^{1/2}) \supseteq \mathscr{D}((-\Delta_D)^{1/2})$,

$$\|W_m^{1/2} g\|^2 \leqslant \theta \|(-\Delta_D)^{1/2} g\|^2 + K \|g\|^2, \quad g \in \mathscr{D}((-\Delta_D)^{1/2}).$$

Then the forms

$$Q_m(g,h) = Q_D(g,h) + (V_m^{1/2} g, V_m^{1/2} h) - (W_m^{1/2} g, W_m^{1/2} h),$$
$$\mathscr{D}(Q_m) = \mathscr{D}(Q_D) \cap \mathscr{D}(V_m^{1/2}),$$
$$Q(g,h) = Q_D(g,h) + (V^{1/2} g, V^{1/2} h) - (W^{1/2} g, W^{1/2} h),$$
$$\mathscr{D}(Q) = \mathscr{D}(Q_D) \cap \mathscr{D}(V^{1/2})$$

are closed. They uniquely determine self-adjoint Hamiltonians $H_m \geqslant -K$ and $H \geqslant -K$ and $\mathscr{C}_0^\infty(\Omega)$ is a form core for H_m and H. Moreover H_m converges to H in the strong resolvent sense

$$s - \lim_{m\to\infty} (H_m - z)^{-1} = (H - z)^{-1}, \quad z \in \mathbb{C} \setminus [K, \infty).$$

Proof: By assumptions (1) and (2), Q_m are closed. Take $g \in \mathscr{C}_0^\infty(\Omega)$, $\mathscr{M} = \operatorname{supp}(g)$; then

$$\| W_m^{1/2} g - W^{1/2} g \|^2$$
$$= \int_{\mathscr{M}} d^n x |\, g(\mathbf{x})|^2 \,|\, W_m^{1/2}(\mathbf{x}) - W^{1/2}(\mathbf{x})|^2$$
$$\leqslant (\sup_{\mathbf{x}\in\mathscr{M}} |\, g(\mathbf{x})|^2) \int_{\mathscr{M}} d^n x |\, W_m(\mathbf{x}) - W(\mathbf{x})| \underset{m\to\infty}{\longrightarrow} 0,$$

by condition (1). Consequently

$$\| W^{1/2} g\|^2 = \lim_{m\to\infty} \| W_m^{1/2} g\|^2 \leqslant \theta \| (-\Delta_D)^{1/2} g\|^2 + K \| g\|^2, \quad g\in\mathscr{C}_0^\infty(\Omega). \tag{3.1}$$

By Theorem 2 of Cycon,[26] $\mathscr{C}_0^\infty(\Omega)$ is a core for $Q_D(g,h) + (V^{1/2} g, V^{1/2} h)$ on $\mathscr{D}(Q_D)\cap\mathscr{D}(V^{1/2})$ and hence Q is closed. By the same Theorem $\mathscr{C}_0^\infty(\Omega)$ is a core for Q and Q_m.

Fix $g_0\in\mathscr{L}^2(\Omega)$. Following Simon[25] we prove

$s - \lim_{m\to\infty} (H_m + K + i)^{-1} g_0 = (H + K + i)^{-1} g_0$ by showing that any subsequence of $(H_m + K + i)^{-1} g_0$ has a subsequence which converges weakly to $(H + K + i)^{-1} g_0$.

Let $h_m = (H_m + K + i)^{-1} g_0$ be a subsequence. Then we know

$h_m\in\mathscr{D}(H_m)\subset\mathscr{D}(Q_m)$, $\|h_m\|\leqslant\| g_0\|$ for all $m\in\mathbb{N}$.

It follows that

$$(h_m, (H_m + K) h_m) \leqslant \| g_0\|^2$$

and

$$(h_m, (H_m + K) h_m) \geqslant (1-\theta) \| (-\Delta_D)^{1/2} h_m\|^2 + \| V_m^{1/2} h_m\|^2,$$

and thus

$$0\leqslant(1-\Theta)\| (-\Delta_D)^{1/2} h_m\|^2 + \| V_m^{1/2} h_m\|^2 \leqslant \| g_0\|^2.$$

Therefore the sequences $\{h_m\}$, $\{(-\Delta_D)^{1/2} h_m\}$, $\{V_m^{1/2} h_m\}$, and, using assumption (2), also $\{W_m^{1/2} h_m\}$ are strongly bounded in $\mathscr{L}^2(\Omega)$. Hence there exists a subsequence $\{h_{m_j}\}_{j\in\mathbb{N}}$ such that

$$h_{m_j} \underset{j\to\infty}{\overset{w}{\longrightarrow}} h, \quad (-\Delta_D)^{1/2} h_{m_j} \underset{j\to\infty}{\overset{w}{\longrightarrow}} \psi, \quad V_{m_j}^{1/2} h_{m_j} \underset{j\to\infty}{\overset{w}{\longrightarrow}} \xi,$$
$$W_{m_j}^{1/2} h_{m_j} \underset{j\to\infty}{\overset{w}{\longrightarrow}} \eta. \tag{3.2}$$

(3.2) together with (3.1) implies

$$(W^{1/2} g, h) = \lim_{j\to\infty} (W_{m_j}^{1/2} g, h_{m_j}) = \lim_{j\to\infty} (g, W_{m_j}^{1/2} h_{m_j}) = (g,\eta), \quad g\in\mathscr{C}_0^\infty(\Omega)$$

and thus

$$h\in\mathscr{D}(W^{1/2}) \text{ and } W^{1/2} h = w - \lim_{j\to\infty} W_{m_j}^{1/2} h_{m_j}. \tag{3.3}$$

In a similar way we get

$$h\in\mathscr{D}(V^{1/2})\cap\mathscr{D}((-\Delta_D)^{1/2})$$

and

$$V^{1/2} h = w - \lim_{j\to\infty} V_{m_j}^{1/2} h_{m_j},$$
$$(-\Delta_D)^{1/2} h = w - \lim_{j\to\infty} (-\Delta_D)^{1/2} h_{m_j}. \tag{3.4}$$

Using $g\in\mathscr{C}_0^\infty(\Omega)$, (3.3) and (3.4) we obtain after taking limits

$$(g, g_0) = (g, (H_{m_j} + K + i) h_{m_j})$$
$$= ((-\Delta_D)^{1/2} g, (-\Delta_D)^{1/2} h_{m_j}) + (V_{m_j}^{1/2} g, V_{m_j}^{1/2} h_{m_j})$$
$$= -(W_{m_j}^{1/2} g, W_{m_j}^{1/2} h_{m_j}) + (K + i)(g, h_{m_j})$$
$$= Q(g,h) + (K + i)(g,h).$$

Since $\mathscr{C}_0^\infty(\Omega)$ is a core for Q this implies

$$h\in\mathscr{D}(H), g_0 = (H + K + i) h$$

and

$$h = (H + K + i)^{-1} g_0 = w - \lim_{j\to\infty} (H_{m_j} + K + i)^{-1} g_0.$$

Therefore the resolvents $(H_{m_j} - z)^{-1}$, $z\in\mathbb{C}\setminus[-K,\infty)$ converge weakly, and hence strongly to $(H - z)^{-1}$.

For strong resolvent convergence using additional monotonicity arguments with respect to m see Refs. 3, 18 Chapt. VIII, and Refs. 27.

Proof of Theorem 3.1: As discussed after Remark 3.2 we only have to show that conditions (C) and ($D^\pm$) of Theorem 2.1 are satisfied; (C) follows by taking $\Omega = \mathbb{R}^n$ in Lemma 3.1. In order to prove the validity of ($D^\pm$) we choose $J = 1$ and[1]

$$\mathscr{D}_0 = (H_0 - z)\{\phi_a \,|\, \phi_a(\mathbf{x}) = e^{-(\mathbf{x}-a)^2/2}, \ a\in\mathbb{R}^n\} \tag{3.5}$$

and now have to control

$$\Psi_{\alpha,a}(t) = \| [H_\alpha - z)^{-1} - (H_0 - z)^{-1}]\, e^{-itH_0} (H_0 - z)\phi_a\|,$$
$$z\in\mathbb{C}\setminus[-K,\infty), \ \alpha\in\overline{B_r(\alpha_0)}.$$

We first show that by assumption (2) $\phi_a(t) = e^{-itH_0}\phi_a$ is in the form domain of $V_\alpha^\pm$ and in the domain of $W_\alpha^\pm$

$$\| (V_\alpha^\pm)^{1/2} \phi_a(t)\| \leqslant \| (1 + |\cdot|^2)^{-k^\pm/4} (V_\alpha^\pm)^{1/2}\| \times \| (1 + |\cdot|^2)^{k^\pm/4} \phi_a(t)\|_\infty$$
$$\leqslant \| V_\alpha^\pm\|_{1,k^\pm}^{1/2} c_{n,k^\pm} (1 + 4t^2)^{-(n-k^\pm)/4}$$
$$\leqslant (M^\pm)^{1/2} c_{n,k^\pm} (1 + 4t^2)^{-(n-k^\pm)/4}, \tag{3.6}$$

$$\| W_\alpha^\pm \phi_a(t)\| \leqslant \| W_\alpha^\pm\|_{2,l^\pm} c_{n,l^\pm} (1 + 4t^2)^{-(n-l^\pm)/4}$$
$$\leqslant N^\pm c_{n,l^\pm} (1 + 4t^2)^{-(n-l^\pm)/4}, \quad \alpha\in\overline{B_r(\alpha_0)}. \tag{3.7}$$

[For $\alpha = \alpha_0$ (3.6) and (3.7) follow by assumptions (2) and Fatou's lemma[28]]. If δ_z denotes a suitable multiple of the distance between z and $[-K,\infty)$ we have

$$\| (V_\alpha^+)^{1/2} (H_\alpha - z)^{-1}\| \leqslant \delta_z^{-1},$$
$$\| (V_\alpha^-)^{1/2} (H_\alpha - z)^{-1}\| \leqslant (\Theta^{-1} - 1)^{-1/2} \delta_z^{-1}, \alpha\in\overline{B_r(\alpha_0)}.$$

This allows us to establish

$$(H_\alpha - z)^{-1}(H_0 - z)\phi_a(t) - \phi_a(t)$$
$$= -[(V_\alpha^+)^{1/2} (H_\alpha - \bar{z})^{-1}]^* (V_\alpha^+)^{1/2} \phi_a(t)$$
$$+ [(V_\alpha^-)^{1/2} (H_\alpha - \bar{z})^{-1}]^* (V_\alpha^-)^{1/2} \phi_a(t)$$
$$+ (H_\alpha - z)^{-1} [W_\alpha^- - W_\alpha^+] \phi_a(t)$$

and then by (3.6) and (3.7)

$$\Psi_{\alpha,\mathbf{a}}(t)\leqslant\|(V_\alpha^+)^{1/2}(H_\alpha-\bar z)^{-1}\|\,\|(V_\alpha^+)^{1/2}\phi_{\mathbf{a}}(t)\|$$
$$+\|(V_\alpha^-)^{1/2}(H_\alpha-\bar z)^{-1}\|\,\|(V_\alpha^-)^{1/2}\phi_{\mathbf{a}}(t)\|$$
$$+\delta_z^{-1}\|W_\alpha^+\phi_{\mathbf{a}}(t)\|+\delta_z^{-1}\|W_\alpha^-\phi_{\mathbf{a}}(t)\|$$
$$\leqslant\delta_z^{-1}(M^+)^{1/2}c_{n,k^+}(1+4t^2)^{-(n-k^+)/4}$$
$$+\delta_z^{-1}(\Theta^{-1}-1)^{-1/2}(M^-)^{1/2}c_{n,k^-}(1+4t^2)^{-(n-k^-)/4}$$
$$+\delta_z^{-1}N^+c_{n,l^+}(1+4t^2)^{-(n-l^+)/4}$$
$$+\delta_z^{-1}N^-c_{n,l^-}(1+4t^2)^{-(n-l^-)/4},\quad\alpha\in\overline{B_r(\alpha_0)}.$$

Since by assumption $k^\pm<n-2$, $l^\pm<n-2$ conditions $(\mathrm{D}^\pm)$ are fulfilled.

Proof of Theorem 3.2: We first note that by Theorem 2 of Cycon[26] the closure Q_{H_α} of $\dot Q_{H_\alpha}$ equals

$$Q_{H_\alpha}(g,h)=Q_D(g,h)+(V_\alpha^{1/2}g,V_\alpha^{1/2}h)$$
$$+(W_\alpha^{1/2}g,W_\alpha^{1/2}h),$$
$$\mathscr{D}(Q_{H_\alpha})=\mathscr{D}(Q_D)\cap\mathscr{D}(V_\alpha^{1/2})\cap\mathscr{D}(W_\alpha^{1/2}),\quad\alpha\in\overline{B_r(\alpha_0)},$$

where now Q_D denotes the Dirichlet form on $\mathbb{R}^n\setminus\Sigma$,

$$Q_D(g,h)=(\overline{\nabla|_{\mathscr{C}_0^\infty(\mathbb{R}^n\setminus\Sigma)}}\,g,\overline{\nabla|_{\mathscr{C}_0^\infty(\mathbb{R}^n\setminus\Sigma)}}\,h)$$

and $-\Delta_D$ the associated Dirichlet Laplacian.

As before, to prove existence and continuity of the wave operators, we have to show that conditions (C) and $(\mathrm{D}^\pm)$ of Theorem 2.1 are satisfied. Taking $\Omega=\mathbb{R}^n\setminus\Sigma$ in Lemma 3.1 we infer that (C) holds. To prove $(\mathrm{D}^\pm)$ we take $\mathscr{D}_0$ as in (3.5) and define

$$(J_R\,g)(\mathbf{x})=J_R(\mathbf{x})\,g(\mathbf{x}),\quad g\in\mathscr{L}^2(\mathbb{R}^n).$$

By $J_R\mathscr{S}(\mathbb{R}^n)\subseteq\mathscr{D}(\Delta_D)$ we get for all $g\in\mathscr{L}^2(\mathbb{R}^n)$ and all $h\in\mathscr{S}(\mathbb{R}^n)$

$$(V_\alpha^{1/2}(H_\alpha-\bar z)^{-1}g,V_\alpha^{1/2}J_R\,h)$$
$$+(W_\alpha^{1/2}(H_\alpha-\bar z)^{-1}g,W_\alpha^{1/2}J_R\,h)$$
$$=Q_{H_\alpha}((H_\alpha-\bar z)^{-1}g,J_R\,h)-Q_D((H_\alpha-\bar z)^{-1}g,J_R\,h)$$
$$=(H_\alpha(H_\alpha-\bar z)^{-1}g,J_R\,h)$$
$$-((H_\alpha-\bar z)^{-1}g,(-\Delta_D)J_R\,h)$$
$$=(g,J_R\,h)+((H_\alpha-\bar z)^{-1}g,[z-(-\Delta_D)]J_R\,h),$$
$$\alpha\in\overline{B_r(\alpha_0)}.$$

[Clearly $J_R\mathscr{S}(\mathbb{R}^n)\subset\mathscr{D}(V_\alpha^{1/2})\cap\mathscr{D}(W_\alpha^{1/2})$, $\alpha\in\overline{B_r(\alpha_0)}\setminus\{\alpha_0\}$ by assumption (2). Using Fatou's lemma as in (3.6) and (3.7), $J_R\mathscr{S}(\mathbb{R}^n)\subset\mathscr{D}(V_{\alpha_0}^{1/2})\cap\mathscr{D}(W_{\alpha_0}^{1/2})$ follows.] This proves

$$(H_\alpha-z)^{-1}(-\Delta_D-z)J_Rh-J_Rh$$
$$=-[V_\alpha^{1/2}(H_\alpha-\bar z)^{-1}]^*V_\alpha^{1/2}J_Rh$$
$$-[W_\alpha^{1/2}(H_\alpha-\bar z)^{-1}]^*W_\alpha^{1/2}J_R\,h.$$

Since

$$(-\Delta_D)|_{J_R\mathscr{S}(\mathbb{R}^n)}=H_0|_{J_R\mathscr{S}(\mathbb{R}^n)}$$

and

$$J_R\mathscr{S}(\mathbb{R}^n)\subseteq\mathscr{D}(W_\alpha),\quad\alpha\in\overline{B_r(\alpha_0)}$$

by assumption (2), we get

$$(H_\alpha-z)^{-1}(H_0-z)J_R\phi_{\mathbf{a}}(t)-J_R\phi_{\mathbf{a}}(t)$$
$$=-[V_\alpha^{1/2}(H_\alpha-\bar z)^{-1}]^*V_\alpha^{1/2}J_R\phi_{\mathbf{a}}(t)$$
$$-(H_\alpha-z)^{-1}W_\alpha J_R\phi_{\mathbf{a}}(t),\quad\alpha\in\overline{B_r(\alpha_0)}.$$

Therefore

$$\Psi_{\alpha,\mathbf{a}}(t)$$
$$=\|[(H_\alpha-z)^{-1}J_R-J_R(H_0-z)^{-1}]\,e^{-iH_0t}(H_0-z)\phi_{\mathbf{a}}\|$$
$$\leqslant\|[(H_\alpha-z)^{-1}(H_0-z)-1]\,J_R\phi_{\mathbf{a}}(t)\|$$
$$+\|(H_\alpha-z)^{-1}[H_0,J_R]\phi_{\mathbf{a}}(t)\|$$
$$\leqslant\delta_z^{-1}M^{1/2}C_{n,k}(1+4t^2)^{(n-k)/4}$$
$$+\delta_z^{-1}NC_{n,l}(1+4t^2)^{-(n-l)/4}$$
$$+\delta_z^{-1}C_{n,R}(1+4t^2)^{-n/4},$$

which shows that $(\mathrm{D}^\pm)$ is valid. Thus for $\alpha\in\overline{B_r(\alpha_0)}$, $\Omega_\pm(H_\alpha,H_0;J_R)$ exist and are strongly continuous at $\alpha=\alpha_0$. Since $(1-J_R)(H_0-z)^{-1}$ is compact [cf. Remark 2.2(a)] $\Omega_\pm(H_\alpha,H_0)=\Omega_\pm(H_\alpha,H_0;J_R)$ and the proof is complete.

Finally, we state a simple physical application of the above results in the context of scattering into cones.

Let C denote the closed circular cone

$$C=\{\mathbf{x}\in\mathbb{R}^n|\,\mathbf{x}\cdot\mathbf{e}\geqslant b\,|\mathbf{x}|\},0<b\leqslant1,$$

$\mathbf{e}$ a unit vector in $\mathbb{R}^n$ and

$$C_R=\{\mathbf{x}\in\mathbb{R}^n|\;|\mathbf{x}|\geqslant R\}\cap C.$$

We introduce

$$(P_{C_R}\phi)(\mathbf{x})=\chi_{C_R}(\mathbf{x})\phi(\mathbf{x}),(Q_C\phi)(\mathbf{k})=\chi_C(\mathbf{k})\tilde\phi(\mathbf{k}),\phi\in\mathscr{L}^2(\mathbb{R}^n).$$

($\chi_{\mathscr{M}}$) denotes the characteristic function of the set $\mathscr{M}\subseteq\mathbb{R}^n$, $\tilde\phi$ the Fourier transform of ϕ) and note that the range of P_{C_R} consists of states localized in the cone C_R whereas that of Q_C are the states whose momenta lie in C.

Assume that in addition to the conditions of Theorem 3.1 or 3.2 $\mathscr{R}(\Omega_-(H_\alpha,H_0)\subseteq\mathscr{R}(\Omega_+(H_\alpha,H_0))$, $\alpha\in\overline{B_r(\alpha_0)}$. Then, by the scattering into cones formula [Ref. 5 (p. 280) and Refs. 29], the probability $P(\phi,C_R,H_\alpha)$ that a state ϕ will be found in the cone C_R as $t\to\infty$,

$$P(\phi,C_R,H_\alpha)=\lim_{t\to\infty}\|P_{C_R}\,e^{-itH_\alpha}\Omega_-(H_\alpha,H_0)\phi\|^2,\qquad(3.8)$$

is equal to the probability that its momentum lies in the cone C in momentum space

$$P(\phi,C_R,H_\alpha)=\lim_{t\to\infty}\|P_{C_R}e^{-itH_\alpha}\Omega_-(H_\alpha,H_0)\phi\|^2$$
$$=\|Q_CS(H_\alpha,H_0)\phi\|^2.$$

Thus applying Corollary 3.2 we get

Theorem 3.3: Assume the conditions of Theorem 3.1 [here we may take $R=0$ in (3.8)] or of Theorem 3.2. If in addition $\mathscr{R}(\Omega_-(H_\alpha,H_0))\subseteq\mathscr{R}(\Omega_+(H_\alpha,H_0))$, $\alpha\in\overline{B_r(\alpha_0)}$ then the probability of finding a particle in the state ϕ inside the cone C_R as $t\to\infty$ is continuous as $\alpha=\alpha_0$, i.e.,

$$\lim_{\alpha\to\alpha_0}P(\phi,C_R,H_\alpha)=P(\phi,C_R,H_{\alpha_0}).$$

Remark 3.3: This result generalizes to cases where $H_0=|\mathbf{P}|^2$ is replaced by $H_0=h(|\mathbf{P}|)$ (h a function with a postive derivative) or modified free evolutions $e^{-ig(t,|\mathbf{P}|)}$ instead of $e^{-it|\mathbf{P}|^2}$ [covering the long-range case discussed in Remark 2.2(b)] are used. Moreover it holds if we replace C by a countable union of disjoint convex cones.[29]

4. CONTINUITY OF SCATTERING OPERATORS—A STATIONARY APPROACH

Throughout this section we assume that the scattering operators $S_\alpha = S(H_\alpha, H_0; J)$ exist and are unitary for all $\alpha \in \overline{B_r(\alpha_0)}$. Let $\oplus_{m=0}^{\infty} \mathscr{L}^2(\Lambda_m; \mathscr{K}_m)$ be the spectral representation and U_0 be the spectral transformation of $H_0 P_{ac}(H_0)$ [cf. Ref. 5 (Chap. 5) and Ref. 30]

$U_0: \mathscr{H}_0 \to \oplus_{m=0}^{\infty} \mathscr{L}^2(\Lambda_m; \mathscr{K}_m)$,

$(U_0 H_0 P_{ac}(H_0) g)_m(k) = k^2 (U_0 g)_m(k)$, $k \in \Lambda_m$,

$g \in \mathscr{D}(H_0 P_{ac}(H_0))$, where Λ_m are disjoint open intervals such that $\mathbb{R} = \cup_{m=0}^{\infty} \overline{\Lambda_m}$ and each $\mathscr{K}_m$ is a separable Hilbert space of dimension m. Since S_α commutes with $H_0 P_{ac}(H_0)$ we get

$(U_0 S_\alpha U_0^{-1} g)_m(k) = S_\alpha^m(k) g_m(k)$, $k \in \Lambda_m$, $g_m(k) \in \mathscr{K}_m$,

where $S_\alpha^m(k)$ is unitary for a.e. $k \in \Lambda_m$. In order to keep notation as simple as possible we henceforth consider the special case where

$$U_0 \mathscr{H}_0 = \mathscr{L}^2(\Lambda; \mathscr{K}),$$

$$(U_0 S_\alpha U_0^{-1} g)(k) = S_\alpha(k) g(k), \quad k \in \Lambda, \ g(k) \in \mathscr{K},$$

$\mathscr{K}$ a separable Hilbert space; but note that obviously all results generalize to the framework discussed above.

The purpose of this section is to describe necessary and sufficient conditions on S_α and $S_\alpha(k)$ such that continuity (respectively, holomorphy) of S_α at $\alpha = \alpha_0$ implies continuity (respectively, holomorphy) of $S_\alpha(k)$ at $\alpha = \alpha_0$, and vice versa.

Assume μ is a σ-finite Borel measure on Λ and let $\{T_\alpha\}$, $\alpha \in \overline{B_r(\alpha_0)}$ be a family of bounded operators in $\mathscr{H}_0$ which are decomposable in $\mathscr{L}^2(\Lambda; d\mu; \mathscr{K})$; that is, if $U_0: \mathscr{H}_0 \to \mathscr{L}^2(\Lambda, d\mu; \mathscr{K})$ then $(U_0 T_\alpha U_0^{-1} g)(k) = T_\alpha(k) g(k)$, for a.e. $k \in \Lambda, g(k) \in \mathscr{K}$, where $T_\alpha(k)$ are bounded operators on $\mathscr{K}$. We call $\alpha \to T_\alpha(k)$ strongly continuous at $\alpha = \alpha_0$ in measure μ on a set $\Delta \subset \Lambda$, $\mu(\Delta) < \infty$, if and only if for all $g \in \mathscr{K}$ and all $\epsilon > 0$

$$\lim_{n \to \infty} \mu(\{k \in \Delta \mid \|T_{\alpha_n}(k) g - T_{\alpha_0}(k) g\| \geqslant \epsilon\}) = 0 \tag{4.1}$$

for all $\{\alpha_n\}_{n \in \mathbb{N}} \subset \overline{B_r(\alpha_0)} \setminus \{\alpha_0\}$ with $\alpha_n \underset{n \to \infty}{\longrightarrow} \alpha_0$. From now on assume that T_α are uniformly bound on $\overline{B_r(\alpha_0)}$ [note that $\sup_{\alpha \in \overline{B_r(\alpha_0)}} \|T_\alpha\| < \infty$, if $\alpha \to T_\alpha$ is strongly continuous on $\overline{B_r(\alpha_0)}$]. Then it suffices to study strong convergence of T_α on a total set in $\mathscr{H}_0$. Let $\mathscr{M}_0$ be a total set in $\mathscr{K}$; then $\{\chi_\Delta \otimes g / \Delta \subseteq \Lambda$ measurable with $\mu(\Delta) < \infty, g \in \mathscr{M}_0\}$ is total in $\mathscr{L}^2(\Lambda; d\mu; \mathscr{K})$. From $(U_0 T_\alpha U_0^{-1} \chi_\Delta \otimes g)(k) = \chi_\Delta(k) T_\alpha(k) g$, $g \in \mathscr{M}_0$ we infer that strong continuity of T_α at $\alpha = \alpha_0$ in $\mathscr{H}_0$ just means

$$\lim_{n \to \infty} \int_\Delta d\mu(k) \|T_{\alpha_n}(k) g - T_{\alpha_0}(k) g\|^2 = 0 \tag{4.2}$$

for any sequence $\{\alpha_n\}_{n \in \mathbb{N}} \subset \overline{B_r(\alpha_0)} \setminus \{\alpha_0\}$, $a_n \underset{n \to \infty}{\longrightarrow} \alpha_0$. Since T_α are uniformly bounded, (4.2) is equivalent to (4.1) by the dominated convergence theorem [Ref. 28 (p. 124)]. Thus we have

Lemma 4.1: Let $T_\alpha, \alpha \in \overline{B_r(\alpha_0)}$ be a family of uniformly bounded operators in $\mathscr{H}_0$ which are decomposable in $\mathscr{L}^2(\Lambda; d\mu; \mathscr{K})$. Then T_α is strongly continuous at $\alpha = \alpha_0$ if and only if for any $\epsilon > 0$, any $\Delta \subset \Lambda$ with $\mu(\Delta) < \infty$, any $g \in \mathscr{M}_0$, and any $\{\alpha_n\}_{n \in \mathbb{N}} \subset \overline{B_r(\alpha_0)} \setminus \{\alpha_0\}$ with $a_n \underset{n \to \infty}{\longrightarrow} \alpha_0$,

$$\lim_{n \to \infty} \mu(\{k \in \Delta \mid \|T_{\alpha_n}(k) g - T_{\alpha_0}(k) g\| \geqslant \epsilon\}) = 0$$

holds. In particular, if $s - \lim_{\alpha \to \alpha_0} T_\alpha(k) = T_{\alpha_0}(k)$ for μ a.e. $k \in \Lambda$ then, $s - \lim_{\alpha \to \alpha_0} T_\alpha = T_{\alpha_0}$ follows.

Turning to S_α and observing $\|S_\alpha\| = 1$ we get, by identifying μ with Lebesgue measure on Λ:

Theorem 4.1: S_α, $\alpha \in \overline{B_r(\alpha_0)}$, is strongly continuous at $\alpha = \alpha_0$ if and only if for any $\Delta \subset \Lambda$ with finite Lebesgue measure, $S_\alpha(k)$ is strongly continuous in measure on Δ at $\alpha = \alpha_0$. In particular, if $s - \lim_{\alpha \to \alpha_0} S_\alpha(k) = S_{\alpha_0}(k)$ for a.e. $k \in \Lambda$ then $s - \lim_{\alpha \to \alpha_0} S_\alpha = S_{\alpha_0}$ holds.

Next we extend this notion to holomorphic families of bounded operators T_α. In order to reduce this case to that considered above we recall that T_α, $\alpha \in \overline{B_r(\alpha_0)} \subset \mathbb{C}$ is holomorphic (in norm) in $B_r(\alpha_0)$ if and only if for all $\alpha, \alpha_1 \in B_r(\alpha_0)$

$$T_\alpha = T_{\alpha_1} + (\alpha - \alpha_1) \Gamma_{\alpha, \alpha_1}, \tag{4.3}$$

where $\Gamma_{\alpha, \alpha_1}$ is strongly continuous at $\alpha = \alpha_1$. Clearly $\Gamma_{\alpha_1, \alpha_1} = T'_{\alpha_1}$ is the complex derivative of T_α at $\alpha = \alpha_1$. For $\alpha \neq \alpha_1$, $\Gamma_{\alpha, \alpha_1} = (T_\alpha - T_{\alpha_1})/(\alpha - \alpha_1)$ is decomposable since T_α and T_{α_1} are. By taking the strong limit $\alpha \to \alpha_1$ and noting that an operator is decomposable if and only if it commutes with every diagonalizable operator we infer that $\Gamma_{\alpha, \alpha_1}$ is decomposable for all $\alpha \in B_r(\alpha_0)$. Assuming that the decomposition of $\Gamma_{\alpha, \alpha_1}$ is given by $\Gamma_{\alpha, \alpha_1}(k)$, i.e., $(U_0 \Gamma_{\alpha, \alpha_1} U_0^{-1} g)(k) = \Gamma_{\alpha, \alpha_1}(k) g(k)$, for a.e. $k \in \Lambda$, $g(k) \in \mathscr{K}$, then

$$\Gamma_{\alpha, \alpha_1}(k) = (T_\alpha(k) - T_{\alpha_1}(k))/(\alpha - \alpha_1)$$

for

$$\alpha \in B_r(\alpha_0) \setminus \{\alpha_1\}.$$

Thus we have

Lemma 4.2: T_α is holomorphic in $B_r(\alpha_0)$ if and only if (1) for any $\alpha_1 \in B_r(\alpha_0)$ there is a $0 < r_1 < r$ such that $(\overline{B_{r_1}(\alpha_1)} \subset B_r(\alpha_0))$,

$$\sup_{\alpha \in \overline{B_{r_1}(\alpha_1)}} \operatorname*{ess\,sup}_{k \in \Lambda} \|(T_\alpha(k) - T_{\alpha_1}(k))/(\alpha - \alpha_1)\| = M < \infty$$

and (2) for any $\epsilon > 0$, any $\Delta \subset \Lambda$ with $\mu(\Delta) < \infty$, any $g \in \mathscr{M}_0$, any $\alpha_1 \in B_r(\alpha_0)$, and any $\{\alpha_n\}_{\in \mathbb{N}} \subset \overline{B_r(\alpha_1)} \setminus \{\alpha_1\}$ with $\alpha_n \underset{n \to \infty}{\longrightarrow} \alpha_1$

$$\lim_{n \to \infty} \mu(\{k \in \Delta \mid \|(T_{\alpha_n}(k) g - T_{\alpha_1}(k) g)/(\alpha_n - \alpha_1) - T'_{\alpha_1}(k) g\| \geqslant \epsilon\}) = 0.$$

In particular, if (1) holds and for any $\alpha_1 \in B_r(\alpha_0)$

$$s - \lim_{\substack{\alpha \to \alpha_1 \\ \alpha \neq \alpha_1}} (T_\alpha(k) - T_{\alpha_1}(k)/(\alpha - \alpha_1)) = T'_{\alpha_1}(k)$$

for a.e. $k \in \Lambda$, then T_α is holomorphic in $B_r(\alpha_0)$.

Proof: (a) Assume T_α to be holomorphic in $B_r(\alpha_0)$. Choose $0 < r' < r$, $0 < \delta < 1$, and $\alpha \in \overline{B_{\delta r'}(\alpha_1)} \subset \overline{B_{r'}(\alpha_1)} \subset B_r(\alpha_0)$. Then

$$(T_\alpha g - T_{\alpha_1} g)/(\alpha - \alpha_1) = (2\pi i)^{-1} \oint_{|\alpha' - \alpha_1| = r'} d\alpha' (\alpha' - \alpha)^{-1} (\alpha' - \alpha_1)^{-1} T_{\alpha'} g,$$

and hence

$$\|(T_\alpha g - T_{\alpha_1} g)/(\alpha-\alpha_1)\| \leqslant (2\pi)^{-1} 2\pi r' \sup_{|\alpha'-\alpha_1|=r'} \times [\,|(\alpha'-\alpha)^{-1}(\alpha'-\alpha_1)^{-1}|\,\|T_{\alpha'} g\|\,] \leqslant [(1-\delta)r']^{-1} M \|g\|$$

since $T_{\alpha'}$ are continuous for all $\alpha' \in \overline{B_r(\alpha_1)}$ and thus uniformly bounded by M. Consequently, for any $\alpha_1 \in B_r(\alpha_0)$, Γ_{α,α_1} fulfills the hypothesis of Lemma 4.1 in $\overline{B_{r_1}(\alpha_1)}$ for some $0 < r_1 < r$ and (1) and (2) follow. (b) Conversely, assume conditions (1) and (2) to hold. Then for any $\alpha_1 \in B_r(\alpha_0)$, Γ_{α,α_1} is uniformly bounded for $\alpha \in \overline{B_{r_1}(\alpha_1)}$. Applying Lemma 4.1 again, Γ_{α,α_1} is strongly continuous in $\alpha = \alpha_1$ and using Eq. (4.3), T_α is holomorphic in $B_r(\alpha_0)$.

Again we state the result for S_α.

Theorem 4.2: S_α is holomorphic in $B_r(\alpha_0)$ if and only if (1) for any $\alpha_1 \in B_r(\alpha_0)$ there is a $0 < r_1 < r$ such that

$$\sup_{\alpha \in \overline{B_{r_1}(\alpha_1)}} \operatorname*{ess\,sup}_{k \in \Lambda} \|(S_\alpha(k) - S_{\alpha_1}(k))/(\alpha-\alpha_1)\| = M < \infty$$

and (2) for any $\Delta \in \Lambda$ with finite Lebesgue measure and any $\alpha_1 \in B_r(\alpha_0)$ $(S_\alpha(k) - S_{\alpha_1}(k))/(\alpha-\alpha_1)$, $\alpha \in \overline{B_{r_1}(\alpha_1)} \setminus \{\alpha_1\}$ is strongly continuous in measure on Δ at $\alpha = \alpha_1$. In particular, if (1) holds and for any $\alpha_1 \in B_r(\alpha_0)$,

$$\lim_{\substack{\alpha \to \alpha_1 \\ \alpha \neq \alpha_1}} (S_\alpha(k) - S_{\alpha_1}(k))/(\alpha-\alpha_1) = S'_{\alpha_1}(k)$$

for a.e. $k \in \Lambda$ then S_α is holomorphic in $B_r(\alpha_0)$.

Remark 4.1: In most applications $S_\alpha(k) - 1$ is a Hilbert–Schmidt operator in $\mathscr{K}$ for a.e. $k \in \Lambda$. Thus continuity (respectively, holomorphy) of S_α may be obtained from suitable properties of the integral kernel of $S_\alpha(k) - 1$ (see Sec. 5).

In the next section we illustrate applications of Theorems 4.1 and 4.2 in potential scattering. To fix notation we briefly consider the case where $\mathscr{H}_0 = \mathscr{L}^2(\mathbb{R}^n)$, $H_0 = -\Delta$. In this case $\Lambda = (0,\infty)$ and $\mathscr{K} = \mathscr{L}^2(S^{(n-1)})$, where $S^{(n-1)}$ denotes the unit sphere in $\mathbb{R}^n$. U_0 is given by

$$U_0: \begin{cases} \mathscr{L}^2(\mathbb{R}^n) \to \mathscr{L}^2((0,\infty); \mathscr{L}^2(S^{(n-1)})), \\ g(\mathbf{x}) \to (U_0 g)(k,\boldsymbol{\omega}) = k^{(n-1)/2} \tilde{g}(k\boldsymbol{\omega}), \quad \mathbf{x} \in \mathbb{R}^n, \boldsymbol{\omega} \in S^{(n-1)} \end{cases}$$

and

$$(U_0 H_0 g)(k,\boldsymbol{\omega}) = k^2 (U_0 g)(k,\boldsymbol{\omega}), \quad g \in \mathscr{D}(H_0).$$

For $\alpha \in \overline{B_r(\alpha_0)}$ let

$$(U_0 S_\alpha U_0^{-1} h)(k,\boldsymbol{\omega}) = S_\alpha(k) h(k,\boldsymbol{\omega}),$$

$$h(k,\boldsymbol{\omega}) \in \mathscr{L}^2(S^{(n-1)}), k > 0$$

and assume that

$$-2\pi i T_\alpha(k) = S_\alpha(k) - 1$$

is a Hilbert–Schmidt operator in $\mathscr{L}^2(S^{(n-1)})$. If

$$t_\alpha(k,\boldsymbol{\omega},\boldsymbol{\omega}') = -(4\pi^2)^{-1} k f_\alpha(k,\boldsymbol{\omega},\boldsymbol{\omega}'), k > 0$$

denotes the integral kernel of $T_\alpha(k)$, then $f_\alpha(k,\boldsymbol{\omega},\boldsymbol{\omega}')$ represents the scattering amplitude, and the on-shell scattering operator $S_\alpha(k)$ is given by

$$(S_\alpha(k)h)(\boldsymbol{\omega}) = h(\boldsymbol{\omega}) - (2\pi i)^{-1} k \int_{S^{(n-1)}} d\omega' f_\alpha(k,\boldsymbol{\omega},\boldsymbol{\omega}') h(\boldsymbol{\omega}'), \quad k > 0, h \in \mathscr{L}^2(S^{(n-1)}). \tag{4.4}$$

5. APPLICATIONS TO LONG-RANGE, ZERO-RANGE, AND SPHERICALLY SYMMETRIC INTERACTIONS

In this section we show how to use the ideas of Sec. 4 in order to get continuity of scattering operators and nonforward total cross sections in the presence of Coulomb-type interactions in $\mathbb{R}^3$, continuity of scattering operators in the zero-range limit, and continuity of phase shifts for interactions of the type $(\beta - (n-2)^2/4)/|\mathbf{x}|^2 + \gamma/|\mathbf{x}| + V_\alpha(|\mathbf{x}|)$, $\mathbf{x} \in \mathbb{R}^n$, $n \geqslant 2$. We also use Theorem 4.2 to obtain holomorphy of the scattering operator for point interactions with respect to the parameter in the boundary condition, and recover a well-known result for Rollnik potentials regarding holomorphy in the coupling constant.

A. Coulomb-type interactions in $\mathbb{R}^3$

In $\mathscr{L}^2(\mathbb{R}^3)$ we consider Hamiltonians (defined as quadratic forms) of the type

$$H_{\alpha,\gamma} = H_{c,\gamma} + V_\alpha, \quad \alpha \in \overline{B_r(\alpha_0)},$$

$$H_{c,\gamma} = -\Delta + \gamma/|\mathbf{x}|, \gamma \in [\gamma_0 - r', \gamma_0 + r'], \quad \gamma_0 \in \mathbb{R}, r' > 0,$$

where $V_\alpha \in \mathscr{L}^1(\mathbb{R}^3)$, V_α real and $V_\alpha \in R$, the Rollnik class [i.e.,

$$\|V\|_R^2 = \int_{\mathbb{R}^6} d^3x\, d^3y\, |V_\alpha(\mathbf{x})|\,|V_\alpha(\mathbf{y})|\,|\mathbf{x}-\mathbf{y}|^{-2} < \infty\Big].$$

In this case the unitary scattering operators in $\mathscr{L}^2(\mathbb{R}^3)$ and $\mathscr{L}^2(S^{(2)})$, respectively, are decomposed in a natural way into[31]

$$S_{\alpha,\gamma} = S^c_\gamma - 2\pi i T^{sc}_{\alpha,\gamma},$$

$$S_{\alpha,\gamma}(k) = S^c_\gamma(k) - 2\pi i T^{sc}_{\alpha,\gamma}(k), \quad k > 0,$$

where $S^c_\gamma, S^c_\gamma(k)$ denote the pure Coulomb scattering operators.[32,33] In particular, $T^{sc}_{\alpha,\gamma}(k)$ is a trace class operator in $\mathscr{L}^2(S^{(2)})$

$$(T^{sc}_{\alpha,\gamma}(k)h)(\boldsymbol{\omega}) = -(4\pi^2)^{-1} k \int_{S^{(2)}} d\omega' f^{sc}_{\alpha,\gamma}(k,\boldsymbol{\omega},\boldsymbol{\omega}') h(\boldsymbol{\omega}'), \quad h \in \mathscr{L}^2(S^{(2)}),$$

and $f^{sc}_{\alpha,\gamma}$, the "Coulomb–nuclear" scattering amplitude is given by[31]

$$f^{sc}_{\alpha,\gamma}(k,\boldsymbol{\omega},\boldsymbol{\omega}') = -(4\pi)^{-1}(v_\alpha \Psi^{(+)}_{c,\gamma}(k,\boldsymbol{\omega},.), ((u_\alpha G_{c,\gamma}(k) v_\alpha + 1)^{-1} u_\alpha \Psi^{(-)}_{c,\gamma})(k,\boldsymbol{\omega}',.)),$$

$$k^2 \notin \mathscr{E}_{\alpha,\gamma},$$

where

$$v_\alpha(\mathbf{x}) = |V_\alpha(\mathbf{x})|^{1/2}, u_\alpha(\mathbf{x}) = |V_\alpha(\mathbf{x})|^{1/2} \operatorname{sign} V_\alpha(\mathbf{x}),$$

and $u_\alpha G_{c,\gamma}(k) v_\alpha$ denotes the norm limit

$$u_\alpha G_{c,\gamma}(k) v_\alpha = n - \lim_{\epsilon \to 0_+} u_\alpha (H_{c,\gamma} - k^2 - i\epsilon)^{-1} v_\alpha, \quad k > 0,$$

and the exceptional set $\mathscr{E}_{\alpha,\gamma}$ (which is a compact subset of $[0,\infty)$ with Lebesgue measure zero) consists of those $k^2 \geqslant 0$ such that $u_\alpha G_{c,\gamma}(k) v_\alpha$ has an eigenvalue -1. If $\gamma \neq 0$ we always add $k^2 = 0$ to $\mathscr{E}_{\alpha,\gamma}$. Finally $\Psi^{(\pm)}_{c,\gamma}$ denote the Coulomb scattering wave functions

$$\Psi^{(-)}_{c,\gamma}(k,\omega,\mathbf{x}) = e^{-\pi\gamma/4k}\Gamma(1+i\gamma/2k)e^{ik\omega\cdot\mathbf{x}} \times {}_1F_1(-i\gamma/2k;1;ik(|\mathbf{x}|-\omega\cdot\mathbf{x})),$$

$$\Psi^{(+)}_{c,\gamma}(k,\omega,\mathbf{x}) = \overline{\Psi^{(-)}_{c,\gamma}(k,-\omega,\mathbf{x})}. \quad k>0 \tag{5.1}$$

[${}_1F_1(a;b;z)$ is the confluent hypergeometric function[34]]. In essentially the same way, the total scattering amplitude $f_{\alpha,\gamma}(k,\omega,\omega')$ splits up into

$$f_{\alpha,\gamma}(k,\omega,\omega') = f^c_\gamma(k,\omega,\omega') + f^{sc}_{\alpha,\gamma}(k,\omega,\omega'), k^2 \notin \mathscr{E}_{\alpha,\gamma},$$

where f^c_γ represents the pure Coulomb scattering amplitude[33]

$$f^c_\gamma(k,\omega,\omega') = -\lim_{\epsilon\to 0_+} \gamma \frac{\Gamma(1+i\gamma/2k)}{\Gamma(1-i\gamma/2k)} 2^{1+i\gamma/2k} \times (4k^2)^{\epsilon-1}[1-\omega\cdot\omega']^{\epsilon-1-i\gamma/2k}, \quad k>0.$$

With these preliminaries in mind we state

Theorem 5.1: Let $\alpha \in \overline{B_r(\alpha_0)}$, $\gamma \in [\gamma_0 - r', \gamma_0 + r']$. If $\lim_{\alpha\to\alpha_0}\|V_\alpha - V_{\alpha_0}\|_1 = 0$ and $\lim_{\alpha\to\alpha_0}\|V_\alpha - V_{\alpha_0}\|_R = 0$, then $f^{sc}_{\alpha,\gamma}(k,\omega,\omega')$ converges to $f^{sc}_{\alpha_0,\gamma_0}(k,\omega,\omega')$ as $\alpha\to\alpha_0$, $\gamma\to\gamma_0$ for a.e. $k>0$ and for all $\omega,\omega' \in S^{(2)}$. In addition, $S_{\alpha,\gamma}(k)$ converges in norm to $S_{\alpha_0,\gamma_0}(k)$ for a.e. $k>0$ and $S_{\alpha,\gamma}$ strongly converges to S_{α_0,γ_0} as $\alpha\to\alpha_0$, $\gamma\to\gamma_0$.

Proof: Let

$$\{\alpha_n\}_{n\in\mathbb{N}} \subset \overline{B_r(\alpha_0)}, \ \alpha_n \xrightarrow[n\to\infty]{} \alpha_0,$$

$$\{\gamma_m\}_{m\in\mathbb{N}} \subset [\gamma_0 - r', \gamma_0 + r'], \ \gamma_m \xrightarrow[m\to\infty]{} \gamma_0$$

and $\mathscr{E} = \mathscr{E}_{\alpha_0,\gamma_0} \cup \cup_{m=1,n=1}^{\infty} \mathscr{E}_{\alpha_n,\gamma_m}$. From (5.1), $\|V_{\alpha_n} - V_{\alpha_0}\|_1 \xrightarrow[n\to\infty]{} 0$, and[35] $|\Psi^{(\pm)}_{c,\gamma_m}(k,\omega,\mathbf{x})| \leqslant c(k)$, $k>0$, we infer that $v_{\alpha_n}(\mathbf{x})\Psi^{(+)}_{c,\gamma_m}(k,\omega,\mathbf{x})$ and $u_{\alpha_n}(\mathbf{x})\Psi^{(-)}_{c,\gamma_m}(k,\omega',\mathbf{x})$ are strongly continuous with respect to m and n. Next we prove

$$\lim_{\substack{n\to\infty \\ m\to\infty}} \|u_{\alpha_n} G_{c,\gamma_m}(k) v_{\alpha_n} - u_{\alpha_0} G_{c,\gamma_0}(k) v_{\alpha_0}\|_{H-S} = 0, \qquad k>0. \tag{5.2}$$

In order to do this we write

$$\begin{aligned}\|u_{\alpha_n} G_{c,\gamma_m}(k) v_{\alpha_n} - u_{\alpha_0} G_{c,\gamma_0}(k) v_{\alpha_0}\|_{H-S} &\leqslant \|u_{\alpha_n} G_{c,\gamma_m}(k) v_{\alpha_n} - u_{\alpha_n} G_{c,\gamma_0}(k) v_{\alpha_n}\|_{H-S} \\ &+ \|u_{\alpha_n} G_{c,\gamma_0}(k) v_{\alpha_n} - u_{\alpha_0} G_{c,\gamma_0}(k) v_{\alpha_0}\|_{H-S}.\end{aligned} \tag{5.3}$$

By assumption any subsequence of $\{V_{\alpha_n}(\mathbf{x})\}_{n\in\mathbb{N}}$ has a subsequence denoted by $\{V_{\alpha_j}(\mathbf{x})\}_{j\in\mathbb{N}}$ such that [Ref. 22 (p. 114)]

$$V_{\alpha_j}(\mathbf{x}) \xrightarrow[j\to\infty]{} V_{\alpha_0}(\mathbf{x}) \text{ pointwise a.e.} \tag{5.4}$$

and

$$|V_{\alpha_j}(\mathbf{x})| \leqslant W(\mathbf{x}) \text{ for some } W \in R. \tag{5.5}$$

Writing

$$G_{c,\gamma}(k,\mathbf{x},\mathbf{y}) = |\mathbf{x}-\mathbf{y}|^{-1} F(\gamma,k,\mathbf{x},\mathbf{y}), k>0 \tag{5.6}$$

where $G_{c,\gamma}(k,\mathbf{x},\mathbf{y})$ denotes the kernel of $(H_{c,\gamma} - k^2 - i\epsilon)^{-1}$ as $\epsilon\to 0_+$, then by the explicit representation of $F(\gamma,k,\mathbf{x},\mathbf{y})$ in terms of Whittaker functions, we get[35]

$$\lim_{m\to\infty} F(\gamma_m,k,\mathbf{x},\mathbf{y}) = F(\gamma_0,k,\mathbf{x},\mathbf{y}) \text{ pointwise,} \tag{5.7}$$

$$|F(\gamma_m,k,\mathbf{x},\mathbf{y})| \leqslant c_{\gamma_0}(k), k>0, \gamma_m \in [\gamma_0 - r', \gamma_0 + r'].$$

(5.4)–(5.7) and the Lebesgue dominated convergence theorem imply that both terms on the right-hand side of (5.3) converge to zero as $m,n\to\infty$. Thus

$$n - \lim_{\substack{n\to\infty \\ m\to\infty}} (u_{\alpha_n} G_{c,\gamma_m}(k) v_{\alpha_n} + 1)^{-1} = (u_{\alpha_0} G_{c,\gamma_0}(k) v_{\alpha_0} + 1)^{-1}, k^2 \notin \mathscr{E}$$

and therefore

$$\lim_{\substack{n\to\infty \\ m\to\infty}} f^{sc}_{\alpha_n,\gamma_m}(k,\omega,\omega') = f^{sc}_{\alpha_0,\gamma_0}(k,\omega,\omega'), \quad k^2 \notin \mathscr{E}.$$

Since $|f^{sc}_{\alpha_n,\gamma_m}(k,\omega,\omega')| \leqslant M_{\alpha_0,\gamma_0}(k), k^2 \notin \mathscr{E}$, $T^{sc}_{\alpha_n,\gamma_m}(k)$ converges in trace norm to $T^{sc}_{\alpha_0,\gamma_0}$ for $k^2 \notin \mathscr{E}$ in $\mathscr{L}^2(S^{(2)})$. Therefore, if $k^2 \notin \mathscr{E}$, then $S_{\alpha_n,\gamma_m}(k)$ converges in norm to $S_{\alpha_0,\gamma_0}(k)$, implying strong convergence of S_{α_n,γ_m} to S_{α_0,γ_0} as $n,m\to\infty$ (by Theorem 4.1).

Theorem 5.1 immediately yields

Corollary 5.1: Let $\alpha \in \overline{B_r(\alpha_0)}$, $\gamma \in [\gamma_0 - r', \gamma_0 + r']$, and assume $\lim_{\alpha\to\alpha_0}\|V_\alpha - V_{\alpha_0}\|_1 = 0$ and $\lim_{\alpha\to\alpha_0}\|V_\alpha - V_{\alpha_0}\|_R = 0$. (a) The probability $P(\phi,C,H_{\alpha,\gamma})$ of finding a particle in the state ϕ inside a cone C as $t\to\infty$ is uous in the interaction $H_{\alpha,\gamma}$,

$$\lim_{\substack{\alpha\to\alpha_0 \\ \gamma\to\gamma_0}} P(\phi,C,H_{\alpha,\gamma}) = P(\phi,C,H_{\alpha_0,\gamma_0}).$$

(b) Let C_i, $i=1,2$ be closed cones with vertices at the origin. Assume $C_1 \cap C_2 = \{0\}$ and define $\Omega_i = C_i \cap S^{(2)}$, $i=1,2$. Then the total scattering cross section $\sigma_{\alpha,\gamma}$ for scattering into C_2 from an initial state having momentum support in C_1 (see Refs. 36)

$$\sigma_{\alpha,\gamma}(k,C_1\to C_2) = (4\pi)^{-1}\int_{\Omega_2} d\omega \int_{\Omega_1} d\omega' |f_{\alpha,\gamma}(k,\omega,\omega')|^2, \quad k^2 \notin \mathscr{E}_{\alpha,\gamma},$$

is finite and continuous in α and γ

$$\lim_{\alpha\to\alpha_0} \sigma_{\alpha,\gamma}(k,C_1\to C_2) = \sigma_{\alpha_0,\gamma_0}(k,C_1\to C_2) \text{ for a.e. } k>0.$$

Remark 5.1: In the short-range case $\gamma = 0$, Theorem 5.1 is well known [cf. Ref. 22 (p. 112)]. From (5.6) and (5.7) we get that as $\alpha\to\alpha_0$

$$(H_{\alpha,\gamma} + E)^{-1} - (H_{\alpha_0,\gamma} + E)^{-1} \quad (E>0 \text{ sufficiently large})$$

tends to zero in trace norm if $\|V_\alpha - V_{\alpha_0}\|_1 \xrightarrow[\alpha\to\alpha_0]{} 0$ and $\|V_\alpha - V_{\alpha_0}\|_R \xrightarrow[\alpha\to\alpha_0]{} 0$ (cf. the proof in Ref. 22, p. 112). Thus $s - \lim_{\alpha\to\alpha_0} S_{\alpha,\gamma} = S_{\alpha_0,\gamma}$ by the trace class criterion mentioned after remark 2.3. (Note that by $\|V_\alpha - V_{\alpha_0}\|_R \leqslant C\|V_\alpha - V_{\alpha_0}\|_1^{1/3}\|V_\alpha - V_{\alpha_0}\|_2^{2/3}$, convergence of V_α to V_{α_0} in $\mathscr{L}^1(\mathbb{R}^3)$ and $\|V_\alpha\|_2 \leqslant C'$ are sufficient conditions for convergence of V_α to V_{α_0} in R.)

B. Point interactions in $\mathbb{R}^3$

Let H_α^0 denote the Hamiltonian with point interaction of parameter $\alpha\in\mathbb{C}$ centered at $\mathbf{x}=0$, i.e., H_α^0 is the closed extension of $-\Delta|_{\mathscr{C}_0^\infty(\mathbb{R}^3\setminus\{0\})}$ given by the boundary condition

$$[-4\pi\alpha|\mathbf{x}|\,g(|\mathbf{x}|)+(\hat{\mathbf{x}}\cdot\nabla)(|\mathbf{x}|\,g(|\mathbf{x}|))]|_{|\mathbf{x}|=0}=0,\quad \hat{\mathbf{x}}=\mathbf{x}/|\mathbf{x}| \tag{5.8}$$

in the partial wave subspace of angular momentum zero. (More precisely, after decomposition into partial waves and unitary transformation from $\mathscr{L}^2((0,\infty);r^2dr)$ into $\mathscr{L}^2((0,\infty))$ the functions g in (5.8) are contained in the set

$$\{g|\,g' \text{ locally absolutely continuous on } (0,\infty);\ -4\pi\alpha g(0_+)+g'(0_+)=0;\ g,g''\in\mathscr{L}^2((0,\infty))\}).$$

For a detailed discussion of H_α^0 see Refs. 37, 38, and the references cited therein. Here we state

Lemma 5.1: iH_α^0 generates a contraction semigroup $e^{-itH_\alpha^0}$, $t\geqslant 0$ if and only if $\operatorname{Im}\alpha\leqslant 0$.

Proof: We only have to show that iH_α^0 is accretive,[19] i.e., $(h,H_\alpha^0 h)\leqslant 0$ for all $h\in\mathscr{D}(H_\alpha^0)$, since then iH_α^0 is clearly maximal accretive. Let $H^0=-\overline{\Delta|_{\mathscr{C}_0^\infty(\mathbb{R}^3\setminus\{0\})}}$ and define

$$\phi_\pm(\mathbf{x})=(\exp i\sqrt{\pm i})|\mathbf{x}|)/4\pi|\mathbf{x}|,\quad \operatorname{Im}\sqrt{\pm i}>0$$

and

$$\mathscr{D}(H_\theta)=\{f+c\phi_+ + ce^{i\theta}\phi_-|\,f\in\mathscr{D}(H^0);\ c\in\mathbb{C}\},\theta\in\mathbb{C},$$
$$H_\theta(f+c\phi_+ + ce^{i\theta}\phi_-)=H^0 f+ic\phi_+ - ice^{i\theta}\phi_-.$$

Then

$$\operatorname{Im}([f+c\phi_+ + ce^{i\theta}\phi_-],H_\theta[f+c\phi_+ + ce^{i\theta}\phi_-]) = |c|^2\|\phi_+\|^2(1-e^{-2\operatorname{Im}\theta})$$

and hence

$$\operatorname{Im}(h,H_\theta h)\leqslant 0 \text{ for all } h\in\mathscr{D}(H_\theta)\Leftrightarrow \operatorname{Im}\theta\leqslant 0. \tag{5.9}$$

Since[39]

$$H_\theta=H_\alpha^0 \text{ if } \alpha=\frac{\cos(\pi/4)}{4\pi}(\tan(\theta/2)-1)$$

relation (5.9) finally implies

$$\operatorname{Im}(h,H_\alpha^0 h)\leqslant 0 \text{ for all } h\in\mathscr{D}(H_\alpha^0)\Leftrightarrow \operatorname{Im}\alpha\leqslant 0.$$

Consequently, if $\operatorname{Im}\alpha\leqslant 0$, the scattering operator

$$S(H_\alpha^0,H_0)=s-\lim_{t\to\infty}e^{itH_0}e^{-2itH_\alpha^0}e^{itH_0},$$

is a contraction and the corresponding on-shell scattering operator $S_\alpha^0(k)$ in $\mathscr{L}^2(S^{(2)})$ is given by[33]

$$S_\alpha^0(k)=1-\frac{2ik}{ik-4\pi\alpha}(Y_0,\cdot)Y_0,\quad \operatorname{Im}\alpha\leqslant 0,$$
$$Y_0(\omega)=(4\pi)^{-1/2}. \tag{5.10}$$

(See Refs. 11 and 40 for nonunitary scattering and dissipative interactions.)

Moreover we have

Theorem 5.2: Let $\operatorname{Im}\alpha\leqslant 0$. Then $S(H_\alpha^0,H_0)$ is norm-continuous in α except for $\alpha=0$ and $\alpha=\infty$. At $\alpha=0$ and $\alpha=\infty$ $S(H_\alpha^0,H_0)$ is only strongly continuous. In addition, $S(H_\alpha^0,H_0)$ is holomorphic in α for $\operatorname{Im}\alpha<0$.

Proof: We discuss separately $\alpha=0$, $\alpha=\infty$, $\alpha\neq 0$ and then apply Theorems 4.1 and 4.2.

(1) $\alpha_1=0$: Using (5.10) we obtain

$$\sup_{\alpha\in[-r,r]}\operatorname*{ess\,sup}_{k>0}\|(S_\alpha^0(k)-S_{\alpha=0}^0(k))/\alpha\| = \sup_{\alpha\in[-r,r]}2/|\alpha|=\infty.$$

(2) $\alpha_1=\infty$: Taking $\beta=1/\alpha$, $\tilde{S}_\beta^0(k)=S_{1/\alpha}^0(k)$ we get

$$\sup_{\beta\in[-r,r]}\operatorname*{ess\,sup}_{k>0}\|(\tilde{S}_\beta^0(k)-\tilde{S}_{\beta=0}^0(k))/\beta\| = \sup_{\beta\in[-r,r]}2/|\beta|=\infty.$$

(3) $\operatorname{Im}\alpha_1\leqslant 0$, $\alpha_1\neq 0$; There we have

$$\sup_{\alpha\in\overline{B_{r_1}(\alpha_1)}}\operatorname*{ess\,sup}_{k>0}\|((S_\alpha^0(k)-S_{\alpha_1}^0(k))/(\alpha-\alpha_1)\| = \sup_{\alpha\in\overline{B_{r_1}(\alpha_1)}}\left(\frac{2}{|\alpha_1|+|\alpha|}\right)=M(r_1,\alpha_1)<\infty.$$

Noting $S_\alpha^0(k)\xrightarrow[\alpha\to\alpha_1]{}S_{\alpha_1}^0(k)$ and $(S_\alpha^0(k)-S_{\alpha_1}^0(k))/(\alpha-\alpha_1)\xrightarrow[\alpha\to\alpha_1]{}(S_{\alpha_1}^0(k))'$ in norm for all $k>0$ and all α_1 with $\operatorname{Im}\alpha_1\leqslant 0$ (including $\alpha_1=\infty$) the proof is finished.

Note that for $\alpha=\infty$, H_∞^0 coincides with the Friedrichs or Dirichlet extension ($H_\infty^0=H_0=-\Delta$) whereas for $\alpha=0$, H_0^0 coincides with the Neumann extension of $-\Delta|_{\mathscr{C}_0^\infty(\mathbb{R}^3\setminus\{0\})}$.

C. Continuity in the zero-range limit

Let $e^{2a|\mathbf{x}|}V\in R$ for some $a>0$, V real, $\lambda:[0,1]\to\mathbb{R}$ real analytic with $\lambda(0_+)=1$, and denote

$$V_\epsilon(\mathbf{x})=\epsilon^{-2}V(\mathbf{x}/\epsilon),0<\epsilon\leqslant 1.$$

Using quadratic forms we define self-adjoint Hamiltonians

$$H=-\Delta+V(\mathbf{x})$$
$$H_\epsilon=-\Delta+\lambda(\epsilon)V_\epsilon(\mathbf{x}),0<\epsilon<1.$$

It is shown in Refs. 37 and 38 that as $\epsilon\to 0_+$, H_ϵ converges in the strong resolvent sense to the point interaction Hamiltonian H_α^0 where $\alpha\in\mathbb{R}$ depends on the spectral properties of $H=-\Delta+V$ at zero energy. Roughly speaking, we have $\alpha\neq\infty$ only if H has a zero energy resonance. If H has only zero energy bound states or if zero is neither a zero energy resonance nor a zero energy bound state of H then $\alpha=\infty$ and thus H converges to the free Hamiltonian $H_\infty^0=H_0$ in the strong resolvent sense (see Ref. 38 for a detailed discussion of these points). Regarding continuity, we have

Theorem 5.3: Let H_ϵ be as above and $S_\epsilon(k)$ be the on-shell scattering operator corresponding to $S(H_\epsilon,H_0)$. Then, as $\epsilon\to 0_+$, $S_\epsilon(k)$ tends to $S_\alpha^0(k)$ in norm for all $k>0$ and $S(H_\epsilon,H_0)$ converges strongly to $S(H_\alpha^0,H_0)$.

Proof: The scattering amplitude $f_\epsilon(k,\omega,\omega')$ corresponding to H_ϵ is given by[38]

$$f_\epsilon(k,\omega,\omega')=-(4\pi)^{-1}\epsilon\lambda(\epsilon)(ve^{i\epsilon k\omega\cdot},\times(\lambda(\epsilon)uG_{\epsilon k}v+1)^{-1}ue^{i\epsilon k\omega'\cdot}),\quad k^2\notin\mathscr{E}_\epsilon,$$

where

$$v(\mathbf{x}) = |V(\mathbf{x})|^{1/2}, u(\mathbf{x}) = |V(\mathbf{x})|^{1/2}\,\mathrm{sign}(V(\mathbf{x})),$$
$$G_{\epsilon k}(\mathbf{x},\mathbf{y}) = e^{i\epsilon k|\mathbf{x}-\mathbf{y}|}/4\pi|\mathbf{x}-\mathbf{y}|,$$

and $\mathscr{E}_\epsilon$ denotes the set of $k^2 \geqslant 0$ such that $\lambda(\epsilon)\, uG_{\epsilon k} v$ has an eigenvalue -1 ($\mathscr{E}_\epsilon$ is a finite discrete subset of $[0,\infty)$). For fixed $k>0$ and $\omega,\omega' \in S^{(2)}$ one can show by explicit expansion of $\epsilon(\lambda(\epsilon) uG_{\epsilon k} v + 1)^{-1}$ that $f_\epsilon(k,\omega,\omega')$ is real analytic[38] at $\epsilon = 0$. Moreover, for fixed $k>0$ and ϵ small enough $S_\epsilon(k)$ has a norm-convergent Taylor expansion around $\epsilon = 0$, completing the proof.

D. Holomorphy in the coupling constant

Here we consider $V \in R$, V real and define

$$H_\alpha = -\Delta + \alpha V, \quad \alpha \in \mathbb{C},$$

using quadratic forms.[18,22] Let $|\alpha| < 4\pi/\|V\|_R$ and

$$S(H_\alpha, H_0) = s - \lim_{t \to +\infty} e^{itH_0} e^{-2itH_\alpha} e^{itH_0}$$

then, applying Theorem 4.2 we get

Theorem 5.4: Let $V \in \mathscr{L}^1(\mathbb{R}^3) \cap R$. Then $S(H_\alpha, H_0)$ and for all $k \geqslant 0$, $S_\alpha(k)$, the corresponding on-shell scattering operators, are holomorphic in α for $|\alpha| < 4\pi/\|V\|_R$.

Proof: From

$$f_\alpha(k,\omega,\omega') = -(4\pi)^{-1}\alpha \sum_{n=0}^{\infty} (-1)^n \alpha^n (v e^{ik\omega\cdot}, (uG_k v)^n u e^{ik\omega'\cdot})$$

and

$$|(v e^{ik\omega\cdot}, (uG_k v)^n u e^{ik\omega'\cdot})| \leqslant \|V\|_1 (\|V\|_R/4\pi)^n$$

we infer that $S_\alpha(k)$ is holomorphic in α for $|\alpha| < 4\pi/\|V\|_R$. Similarly,

$$\|(S_\alpha(k) - 1)/\alpha\|_{H-S} \leqslant \frac{k}{8\pi^2} \sum_{n=0}^{\infty} |\alpha|^n \left(\int d\omega\, d\omega' |(v e^{ik\omega\cdot}, (uG_k v)^n u e^{ik\omega'\cdot})|^2 \right)^{1/2}$$
$$\leqslant 2 \sum_{n=0}^{\infty} |\alpha|^n (\|V\|_R/4\pi)^n < \infty \quad \text{if } |\alpha| < 4\pi/\|V\|_R.$$

Thus

$$\sup_{\alpha \in B_r(0)} \operatorname*{ess\,sup}_{k>0} \|S_\alpha(k) - 1)/\alpha\| \leqslant M(r) < \infty \text{ for all } r < 4\pi/\|V\|_R,$$

proving Theorem 5.4.

Remark 5.2: Using H_0-smoothness of $|V|^{1/2}$, Kato[41] derived Theorem 5.4 under the condition $V \in R$ only (actually he also got a slightly larger domain of holomorphy). The purpose of Theorem 5.4 was to show that within the restricted class $V \in \mathscr{L}_1(\mathbb{R}^3) \cap R$ holomorphy of $S(H_\alpha, H_0)$ and $S_\alpha(k)$ follows trivially from Theorem 4.2.

E. Spherically symmetric potentials

Now we discuss Hamiltonians $H_{\alpha,\beta,\gamma}$ in $\mathscr{L}^2(\mathbb{R}^n)$ which are self-adjoint realizations of differential expressions of the type

$$-\Delta + \frac{\beta - (n-2)^2/4}{|\mathbf{x}|^2} + \frac{\gamma}{|\mathbf{x}|} + V_\alpha(|\mathbf{x}|), \quad \mathbf{x} \in \mathbb{R}^n,$$
$$n \geqslant 2, \ \alpha \in \overline{B_r(\alpha_0)}, \ \beta > 0, \ \gamma \in \mathbb{R},$$

where $V_\alpha \in \mathscr{L}^1_{\mathrm{loc}}((0,\infty))$, V_α real and

$$\|V_\alpha\|_{k_0} = \int_0^\infty dr \frac{r}{1 + k_0 r} |V_\alpha(r)| < \infty \text{ for some } k_0 > 0.$$

Due to angular momentum conservation, $H_{\alpha,\beta,\gamma}$ and $S_{\alpha,\beta,\gamma}$, the corresponding unitary scattering operator, are reduced by the subspaces of fixed angular momentum l

$$H_{\alpha,\beta,\gamma} = \bigoplus_{l=0}^{\infty} U^{-1} h^{(l)}_{\alpha,\beta,\gamma} U \otimes 1,$$
$$\mathscr{L}^2(\mathbb{R}^n) \cong \mathscr{L}^2((0,\infty); r^{n-1}\, dr) \otimes \mathscr{L}^2(S^{(n-1)}),$$
$$S_{\alpha,\beta,\gamma} = \bigoplus_{l=0}^{\infty} U^{-1} S^{(l)}_{\alpha,\beta,\gamma} U \otimes 1,$$

where

$$U: \begin{cases} \mathscr{L}^2(0,\infty; r^{n-1} dr) \to \mathscr{L}^2((0,\infty)), \\ \rho(r) \to r^{(n-1)/2} \rho(r), \end{cases}$$

and $h^{(l)}_{\alpha,\beta,\gamma}$ is the Friedrichs extension of the minimal operator

$$\dot{h}^{(l)}_{\alpha,\beta,\gamma} = -\frac{d^2}{dr^2} + \frac{l(l+n-2) + \beta - 1/4}{r^2} + \frac{\gamma}{r} + V_\alpha(r), \quad l = 0,1,2,\ldots,$$
$$\mathscr{D}(\dot{h}^{(l)}_{\alpha,\beta,\gamma}) = \{\rho \in \mathscr{C}_0^\infty((0,\infty)) \,|\, \dot{h}^{(l)}_{\alpha,\beta,\gamma}\, \rho \in \mathscr{L}^2((0,\infty))\}. \tag{5.11}$$

(See Refs. 42 for a detailed discussion of $h^{(l)}_{\alpha,\beta,\gamma}$ and $S^{(l)}_{\alpha,\beta,\gamma}$). Since $S_{\alpha,\beta,\gamma}$ commutes with $H_0 = -\Delta$ and H_0 restricted to a subspace of fixed angular momentum has simple absolutely continuous spectrum, $S^{(l)}_{\alpha,\beta,\gamma}$ is diagonalizable and $S^{(l)}_{\alpha,\beta,\gamma}(k)$ can be written as

$$S^{(l)}_{\alpha,\beta,\gamma}(k) = e^{2i\delta^{(l)}_{\alpha,\beta,\gamma}(k)} \quad k > 0,$$

where $\delta^{(l)}_{\alpha,\beta,\gamma}(k)$ represent the phase shifts associated with $h^{(l)}_{\alpha,\beta,\gamma}$. As in Sec. 5A, $\delta^{(l)}_{\alpha,\beta,\gamma}(k)$ splits up into

$$\delta^{(l)}_{\alpha,\beta,\gamma}(k) = \delta^{(l)}_{\beta,\gamma}(k) + \tilde{\delta}^{(l)}_{\alpha,\beta,\gamma}(k),$$

where

$$\delta^{(l)}_{\beta,\gamma}(k) = \arg\{\Gamma[\tfrac{1}{2} + (l(l+n-2)+\beta)^{1/2} + i\gamma/2k]\} + \pi[n - 2 + 2l - 2(l(l+n-2)+\beta)^{1/2}]/4$$

denotes the phase shift for $h^{(l)}_{\beta,\gamma}$, the Friedrichs extension of (5.11) in the case $V_\alpha = 0$, and $\tilde{\delta}^{(l)}_{\alpha,\beta,\gamma}(k)$ can be chosen such that $\lim_{k\to\infty} \tilde{\delta}^{(l)}_{\alpha,\beta,\gamma}(k) = 0$.[42] The continuity result now reads

Theorem 5.5: Let $\alpha \in \overline{B_r(\alpha_0)}$, $\beta \in [\beta_0 - r', \beta_0 + r'] \subset (0,\infty)$, $\gamma \in [\gamma_0 - r'', \gamma_0 + r'']$ and assume

$$\lim_{\alpha \to \alpha_0} \|V_\alpha - V_{\alpha_0}\|_{k_0} = 0 \text{ for some } k_0 > 0.$$

Then for all $k > 0$

$$\lim_{\substack{\alpha\to\alpha_0 \\ \beta\to\beta_0 \\ \gamma\to\gamma_0}} \delta^{(l)}_{\alpha,\beta,\gamma}(k) = \delta^{(l)}_{\alpha_0,\beta_0,\gamma_0}(k)$$

and

$$n - \lim_{\substack{\alpha\to\alpha_0 \\ \beta\to\beta_0 \\ \gamma\to\gamma_0}} S^{(l)}_{\alpha,\beta,\gamma}(k) = S^{(l)}_{\alpha_0,\beta_0,\gamma_0}(k)$$

hold. In particular, $S^{(l)}_{\alpha,\beta,\gamma}$ and $S_{\alpha,\beta,\gamma}$ strongly converge to

$S^{(l)}_{\alpha_0,\beta_0,\gamma_0}$ and $S_{\alpha_0,\beta_0,\gamma_0}$ as $\alpha \to \alpha_0$, $\beta \to \beta_0$, $\gamma \to \gamma_0$.

The proof relies on Jost function techniques employed in Refs. 42, e.g., iterating the Volterra integral equation for the difference of the regular solutions corresponding to $h^{(l)}_{\alpha_0,\beta,\gamma}$ and $h^{(l)}_{\alpha,\beta,\gamma}$, continuity of the Jost functions and hence continuity of the phase shifts $\tilde{\delta}^{(l)}_{\alpha,\beta,\gamma}$ at $\alpha = \alpha_0$ is gained. We omit the details.

For related results in the short-range case $\gamma = 0$ and in the screening problem of Coulomb-interactions, see Refs. 43. We note that similar to Remark 5.1 $(h^{(l)}_{\alpha,\beta,\gamma} + E)^{-1} - (h^{(l)}_{\alpha_0,\beta,\gamma} + E)^{-1}$ ($E > 0$ large enough) tends to zero in trace norm as $\lim_{\alpha \to \alpha_0} \| V_\alpha - V_{\alpha_0} \|_{k_0} = 0$.

ACKNOWLEDGMENTS

We are indebted to the Universitätsschwerpunkt Mathematisierung (USP) of the University of Bielefeld for an invitation to a workshop in October 1980 that initiated our joint work. F. G. gratefully acknowledges constant encouragement by Prof. L. Streit and the warm hospitality of the Theoretical Physics Department of the University of Bielefeld as well as financial support by the Alexander von Humboldt Foundation.

[1]T. Kato, Commun. Math. Phys. **67**, 85 (1979).
[2]H. Narnhofer, Nuovo Cimento B **30**, 254 (1975).
[3]D. W. Robinson, Ann. Inst. H. Poincaré A **21**, 185 (1974).
[4]E. B. Davies, Math. Proc. Camb. Phil. Soc. **82**, 327 (1977).
[5]W. O. Amrein, J. M. Jauch, and K. B. Sinha, *Scattering Theory in Quantum Mechanics* (Benjamin, Reading, MA, 1977).
[6]J. M. Cook, J. Math. and Phys. **36**, 82 (1957).
[7]T. Kato, J. Funct. Anal. **1**, 342 (1967); J. Fac. Sci. Univ. Tokyo **24**, 503 (1977); A. L. Belopol'skii and M. S. Birman, Math. USSR Izv. **2**, 1117 (1968); M. S. Birman, Math. USSR Izv. **5**, 459 (1971); C. H. Wilcox, J. Funct. Anal. **12**, 257 (1973); M. Schechter, J. Math. Pures Appl. **53**, 373 (1974); Ann. Inst. H. Poincaré A **30**, 109 (1979).
[8]M. Reed andB. Simon, *Methods of Modern Mathematical Physics, III. Scattering Theory* (Academic, New York, 1979).
[9]M. Schechter, Duke Math. J. **44**, 863 (1977).
[10]B. Simon, Commun. Math. Phys. **53**, 151 (1977).
[11]B. Simon, Duke Math. J. **46**, 119 (1979).
[12]E. B. Davies, Duke Math. J. **47**, 171 (1980).
[13]V. Enss, "Geometric methods in spectral and scattering theory of Schrödinger operators" in *Rigorous Atomic and Molecular Physics*, edited by G. Velo and A. S. Wightman (Erice, 1980), p. 1.
[14]J. M. Combes and R. Weder, "New criterium for existence and completeness of wave operators and applications to scattering by unbounded obstacles," preprint, Univ. Mexico (1980).
[15]J. Kupsch and W. Sandhas, Commun. Math. Phys. **2**, 147 (1966).
[16]D. B. Pearson, Helv. Phys. Acta **48**, 639 (1975); P. Deift and B. Simon, J. Funct. Anal. **23**, 218 (1976); Yu. A. Semenov, Lett. Math. Phys. **1**, 457 (1977); M. Combescure and J. Ginibre, J. Funct. Anal. **29**, 54 (1978); W. O. Amrein, D. B. Pearson, and M. Wollenberg, Helv. Phys. Acta **53**, 335 (1980); M. A. Perelmuter and Yu. A. Semenov, J. Math Phys. **22**, 521 (1981).
[17]L. Hörmander, Math. Z. **146**, 69 (1976).
[18]T. Kato, *Perturbation Theory for Linear Operators in Hilbert Space* (Springer, New York, 1966).
[19]M. Reed and B. Simon, *Methods of Modern Mathematical Physics, II. Fourier Analysis, Self-Adjointness* (Academic, New York, 1975).
[20]G. Aumann, Reelle Funktionen (Springer, Berlin, 1969), p. 208.
[21]S. T. Kuroda, Nuovo Cimento **12**, 431 (1959); J. Math. Soc. Japan **11**, 247 (1959).
[22]B. Simon, *Quantum Mechanics for Hamiltonians defined as Quadratic Forms* (Princeton U. P., Princeton, NJ, 1971).
[23]J. D. Dollard, J. Math. Phys. **9**, 620 (1968).
[24]V. Enss, Commun. Math. Phys. **61**, 285 (1978); Ann. Phys. **119**, 117 (1979); P. A. Perry, Duke Math. J. **47**, 187 (1980).
[25]B. Simon, J. Operator Theory **1**, 37 (1979).
[26]H. L. Cycon, J. Operator Theory **6**, 75 (1981).
[27]W. G. Faris, *Self-Adjoint Operators, Lecture Notes in Mathematics 433* (Springer, Berlin, 1975); M. Schechter, Lett. Math. Phys. **1**, 265 (1976); B. Simon, Proc. Roy. Soc. Edinburgh **19**, 267 (1977); J. Funct. Anal. **28**, 377 (1978); Yu. A. Semenov, Studia Math. **64**, 77 (1979).
[28]N. Dunford and J. T. Schwartz, *Linear Operators I* (Interscience, New York, 1958).
[29]J. D. Dollard, Commun. Math. Phys. **12**, 193 (1969); J. M. Jauch, R. Lavine, and R. G. Newton, Helv. Phys. Acta **45**, 325 (1972); Ph. A. Martin, Helv. Phys. Acta **45**, 794 (1972).
[30]W. O. Amrein and D. B. Pearson, Lett. Math. Phys. **3**, 83 (1979).
[31]F. Gesztesy, J. Math. Phys. **23**, 74 (1982).
[32]J. D. Dollard, J. Math. Phys. **5**, 729 (1964); Rocky Mountain J. Math **1**, 5 (1971); H. Grosse, H. R. Grümm, H. Narnhofer, and W. Thirring, Acta Phys. Austriaca **40**, 97 (1974).
[33]I. W. Herbst, Commun. Math. Phys. **35**, 181 (1974).
[34]M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1972).
[35]J. C. Guillot, Ind. Univ. Math. J. **25**, 1105 (1976).
[36]S. Agmon, *Some new results in spectral and scattering theory of differential operators on* $\mathbf{R}^n$ (Ecole Polytechnique, Palaiseau, 1978); W. O. Amrein and D. B. Pearson, Ann. Inst. H. Poincaré A **30**, 89 (1979).
[37]S. Albeverio and R. Høegh-Krohn, J. Operator Theory **6**, 313 (1981).
[38]S. Albeverio, F. Gtesztesy, and R. Høegh-Krohn, Ann. Inst. H. Poincaré **A31**, 1 (1982).
[39]J. Zorbas, J. Math. Phys. **21**, 840 (1980).
[40]Ph. A. Martin, Nuovo Cimento B **30**, 217 (1975); E. B. Davies, Commun. Math. Phys. **71**, 277 (1980).
[41]T. Kato, Math. Ann. **162**, 258 (1966).
[42]F. Gesztesy, W. Plessas, and B. Thaller, J. Phys. A **13**, 2659 (1980); F. Gesztesy and B. Thaller, J. Phys. A **14**, 639 (1981).
[43]W. M. Frank, J. Math. Phys. **12**, 2529 (1971); J. Fröhlich and L. Streit, Acta Phys. Austriaca **47**, 125 (1977).

Time delay for long-range interactions[a)]

D. Bollé[b)]
Instituut voor Theoretische Fysica, Universiteit Leuven, B-3030 Leuven, Belgium

F. Gesztesy[c)]
Institut für Theoretische Physik, Universität Graz, A-8010 Graz, Austria

H. Grosse
Institut für Theoretische Physik, Universität Wien, A-1090 Wien, Austria

(Received 30 July 1982; accepted for publication 12 November 1982)

We discuss the concept of time delay and its relation to the S matrix for quantum mechanical scattering with potentials of the form $V(r) = \gamma/r + (\alpha - \frac{1}{4})/r^2 + V_{sc}(r)$. The spherically symmetric interactions $V_{sc}(r)$ are allowed to be $O(r^{-2-\epsilon})$ at infinity and $O(r^{-2+\epsilon})$ at zero. This improves considerably the existing results in the literature where the total $V(r)$ is required to be $O(r^{-5/2-\epsilon})$ at infinity. We study pure Coulomb time delay, relative time delay with respect to the long-range part $\gamma/r + (\alpha - \frac{1}{4})/r^2$ of the interaction, and total time delay. The corresponding results in classical scattering are also derived.

PACS numbers: 03.65.Nk, 03.80. + r

I. INTRODUCTION

This paper studies the concept of time delay and its connection with the S matrix for two particles interacting via Coulomb-type potentials plus spherically symmetric short-range potentials. We consider pure Coulomb time delay, relative time delay with respect to the long-range interaction and total time delay, in quantum mechanical as well as classical systems.

A recent overview of time delay theory and its applications has been given in Ref. 1. Since then, Lax and Philips[2] and Amrein and Wollenberg[3] have discussed time delay in the context of classical hyperbolic equations. Recently, Jensen[4,5] has derived results on time delay for general (not necessarily spherically symmetric) potentials, using properties of the generator of dilations. His potentials satisfy $V(x) = O(|x|^{-\beta}), \beta > 4$, at infinity (but see remark 6.7 of Ref. 4).

For spherically symmetric interactions, the weakest condition at infinity that one has been able to handle up to now is $V(r) = O(r^{-5/2-\epsilon}), \epsilon > 0$.[6] This is done in the time dependent approach of Martin.[7,8] Consequently, one could think that a generalization of this approach would allow one to discuss the problem in systems with long-range interactions. However, already for the pure Coulomb system, we show that Cook's estimate no longer guarantees the interchangeability of some time integrals and the infinite space limit that is crucial in this method (see Remark 3.2).

Therefore we will use a method that starts from a time-dependent definition of time delay and then employs the Abel-limit procedure together with distributional arguments (see also Ref. 9) and a rigorous extension of Smith's[10] stationary approach (see also Refs. 1 and 11) to the long range case. In addition, for a pure Coulomb system we will present a discussion of time delay within the algebraic approach[12-14] to scattering theory.

We obtain the following results. When the interaction is pure Coulomb, $V(r) = \gamma/r$, γ real, the time delay with respect to the Dollard[15,16]-modified free evolution is finite. However, the relation with the S matrix is not what one would expect by generalizing in a straightforward way the known short-range result.[1] To make this clear, we write the Dollard modification as follows:

$$U_{a,c}(t) = \exp\{-i[H_0 t + [\gamma \operatorname{sgn}(t)/2H_0^{1/2}] \times \log(4aH_0|t|)]\}, \quad t \neq 0, a > 0,$$

where we explicitly used the freedom to introduce an additional parameter a. Of course, one expects intuitively that the Coulomb time delay also depends on this parameter a. Indeed, we prove that the on-shell Coulomb time delay operator $\tau_{a,c}(k)$ satisfies

$$\tau_{a,c}(k) = -\frac{i}{2k} S_c^*(k) \frac{\partial}{\partial k} S_c(k) + \frac{\gamma}{2k^3}(1 - \log a)1, \quad k > 0,$$

where 1 denotes the identity operator on $L^2(S^2)$. This suggests that the choice $a = e$ is the most natural one since in that case this time delay formula exactly parallels the short-range result. However, the usual choice[15,16] is $a = 1$, leading to a supplementary factor $\gamma/2k^3$. An analog result is derived for classical scattering. All this forms the content of Sec. II.

In Sec. III we treat Coulomb-type interactions plus spherically symmetric short-range interactions. In particular, the interaction is of the form $\gamma/r + (\alpha - \frac{1}{4})/r^2 + V(r)$, $\alpha > 0$, γ real and $rV(r) \in L^1(0, \infty)$. With the help of methods used in Refs. 17 and 18, we then prove results for the total time delay $\tau_a(k)$ and the relative time delay $\tau_{sc}(k)$ with respect to the long-range interaction $V_\alpha(r) = \gamma/r + (\alpha - \frac{1}{4})/r^2$. Splitting up the total phase shift of the system into $\delta_l = \delta_l^{sc} + \delta_l^{\alpha}$, with obvious notation, we show that the on-shell

[a)] Supported in part by Fonds zur Förderung der Wissenschaftlichen Forschung in Österreich, Project nr. 4328.
[b)] Onderzoeksleider N. F. W. O., Belgium.
[c)] Supported in part by the Alexander von Humboldt Foundation.

 0022-2488/83/061529-13$02.50 © 1983 American Institute of Physics

relations are

$$\tau_{sc}(k) = \bigoplus_{l=0}^{\infty} \frac{1}{k} \frac{\partial}{\partial k} \delta_l^{sc}(k), \quad k>0,$$

where the connection with the S matrices is clear, once we choose how to split up the total S matrix (see, e.g., Ref. 19), and

$$\tau_a(k) = -\frac{i}{2k} S^*(k) \frac{\partial}{\partial k} S(k) + \frac{\gamma}{2k^3}(1 - \log a)1, \quad k>0.$$

Of course, the remark made before concerning the parameter a remains valid. Similar results are obtained for the corresponding classical system.

So, in addition to the long-range interaction $V_\alpha(r)$, we are able to treat spherically symmetric short-range potentials which behave like $r^{-2-\epsilon}$ at infinity and $r^{-2+\epsilon}$ at zero ($\epsilon>0$). This means that even in the pure short range case ($\gamma=0$, $\alpha=\frac{1}{4}$) our results are stronger than anything we know of in the literature.

Finally, in Sec. IV we give a second derivation of pure Coulomb time delay within the algebraic approach, using explicitly the SO(3,1) symmetry in the problem, expressed by an additional conserved quantity, i.e., the Runge–Lenz vector.

II. COULOMB TIME DELAY

In this section, we discuss the concept of Coulomb time delay and prove its relation with the S matrix in both quantum and classical mechanics.

A. Quantum mechanics

The method we will use starts from a time-dependent definition of time delay and employs then the Abel-limit procedure, together with distributional arguments (see also Ref. 9) and a rigorous extension of Smith's stationary approach[10] (cf. also Refs. 1 and 11) to the long-range case.

Let us begin by introducing some definitions and notations. In the Hilbert space $L^2(\mathbb{R}^3)$ we consider the self-adjoint kinetic energy operator H_0

$$H_0 = -\Delta \quad \text{on} \quad \mathcal{D}(H_0) = H^2(\mathbb{R}^3) \tag{2.1}$$

and the Coulomb Hamiltonian H_c

$$H_c = -\Delta + \gamma/|\mathbf{x}| \quad \text{on} \quad \mathcal{D}(H_c) = \mathcal{D}(H_0), \quad \gamma\in\mathbb{R}, \tag{2.2}$$

where we have used natural units $2m = \hbar = 1$. According to Dollard,[15,16] the corresponding unitary modified free evolution $U_{a,c}(t)$ reads

$$U_{a,c}(t) = \exp\{-i[H_0 t + [\gamma \operatorname{sgn}(t)/2H_0^{1/2}] \log(4aH_0|t|)]\}, \quad t\neq 0, \quad a>0, \tag{2.3}$$

where we explicitly used the freedom to introduce an additional parameter $a>0$.

Furthermore, let P_R be the projection onto states localized in the sphere of radius R, centered at the origin in configuration space. Denote by $\mathcal{S}_0(\mathbb{R})^3$ the following set:

$\mathcal{S}_0(\mathbb{R}^3) = \{\phi\in\mathcal{S}(\mathbb{R}^3) | \tilde{\phi}(\mathbf{k})$ vanishes in a (ϕ-dependent) neighborhood of the origin $\mathbf{k}=0\}$, (2.4)

where $\mathcal{S}(\mathbb{R}^3)$ is the usual Schwarz space and $\tilde{\phi}(\mathbf{k})$ is the Fourier transform of ϕ. For $\phi\in L^2(\mathbb{R}^3)$ the Fourier transform $\tilde{\phi}$ is defined by $\tilde{\phi}(\mathbf{k}) = \text{s-lim}_{L\to\infty} \int_{|\mathbf{x}|<L} d^3x (2\pi)^{-3/2} \times \exp(-i\mathbf{k}\cdot\mathbf{x})\phi(\mathbf{x})$.

To exploit the spherical symmetry of our problem, we introduce the functions $\phi_{l,m}\in\mathcal{S}_0(\mathbb{R}^3)$ which satisfy

$$\tilde{\phi}_{l,m}(\mathbf{k}) = \alpha_{l,m}(k) Y_{l,m}(\omega), \quad \alpha_{l,m}\in C_0^\infty(0,\infty), \quad l\in\mathbb{N}, \quad -l\leqslant m\leqslant l, \tag{2.5}$$

where $\mathbf{k} = k\omega$, $\omega\in S^2$, the unit sphere in $\mathbb{R}^3$ and $Y_{l,m}$ denote the spherical harmonics.

Next, we define the modified free and Coulomb transit time, associated with the state $\phi\in\mathcal{S}_0(\mathbb{R}^3)$, through a sphere of radius R by, respectively

$$T_{a,c}^R(\phi) = \int_{-\infty}^{+\infty} dt \,\| P_R U_{a,c}(t)\phi\|^2, \tag{2.6}$$

$$T_c^R(\phi) = \int_{-\infty}^{+\infty} dt \,\| P_R e^{-itH_c} \Omega_{a,c}^- \phi\|^2, \tag{2.7}$$

where $\Omega_{a,c}^{\pm}$ are the usual Coulomb wave operators with the a-dependence taken into account explicitly. Note that the Coulomb transit time does not depend on the parameter a, as is clear by writing out explicitly the norm in Eq. (2.7). The Coulomb time delay $\tau_{a,c}^R(\phi)$, associated with the state $\phi\in\mathcal{S}_0(\mathbb{R}^3)$, for a compact region $|\mathbf{x}|\leqslant R$ is then defined as

$$\tau_{a,c}^R(\phi) = \int_{-\infty}^{+\infty} dt \,[\| P_R e^{-itH_c}\Omega_{a,c}^-\phi\|^2 - \| P_R U_{a,c}(t)\phi\|^2]. \tag{2.8}$$

Now, one can ask the following three questions. Is the expression (2.8) meaningful for finite R? Does its limit for infinite space region, $R\to\infty$, exist? Is this limit related with the S matrix? We answer those questions in the following theorems, the proof of which we give afterwards.

Theorem 2.1: Let $\phi\in\mathcal{S}_0(\mathbb{R}^3)$. Then for all $N\in\mathbb{N}$ and all $R\in(0,\infty)$ there exist constants $C_{N,R}$ such that

$$\|P_R e^{-itH_c}\Omega_{a,c}^{\pm}\phi\| + \|P_R U_{a,c}(t)\phi\| \leqslant C_{N,R}(1+|t|)^{-N}. \tag{2.9}$$

In particular, $T_{a,c}^R(\phi)$, $T_c^R(\phi)$, and $\tau_{a,c}^R(\phi)$ defined by Eqs. (2.6)–(2.8) are finite.

Theorem 2.2: Let $\phi_{l,m}$ be as in (2.5). Then, for large R

$$T_{a,c}^R(\phi_{l,m}) \underset{R\to\infty}{=} \int_0^\infty k^2 dk \int_{S^2} d\omega |\tilde{\phi}_{l,m}(k\omega)|^2 \times\left[\frac{R}{k} + \frac{\gamma}{2k^3}\log(2kaR) - \frac{\gamma}{k^3}\right] + o(1). \tag{2.10}$$

Theorem 2.3: Let $\phi_{l,m}$ be as in (2.5). Then, for large R

$$T_c^R(\phi_{l,m}) \underset{R\to\infty}{=} \int_0^\infty k^2 dk \int_{S^2} d\omega\, |\tilde{\phi}_{l,m}(k\omega)|^2 \times\left[\frac{R}{k} + \frac{\gamma}{2k^3}\log(2kR) - \frac{\gamma}{2k^3} + \frac{1}{k}\frac{\partial}{\partial k}\delta_l^c(k)\right] + o(1), \tag{2.11}$$

where $\delta_l^c(k) = \arg[\Gamma(l+1+i\gamma/2k)]$ is the usual Coulomb phase shift.

We finally state the connection between the Coulomb time delay for infinite space region, defined as

$$\tau_{a,c}(\phi) = \lim_{R\to\infty} \tau_{a,c}^R(\phi) = \lim_{R\to\infty} [T_c^R(\phi) - T_{a,c}^R(\phi)], \tag{2.12}$$

whenever this limit exists, and the Coulomb S matrix. In order to do so, we need to introduce the spectral representation with respect to H_0. Let U_0 be the following isomorphism:

$$U_0: \begin{cases} L^2(\mathbb{R}^3) \to L^2((0,\infty);L^2(S^2)) \\ h(x) \to (U_0 h)(k,\omega) = k\tilde{h}(k\omega), \quad k>0, \ \omega\in S^2; \end{cases} \tag{2.13}$$

then the Coulomb scattering operator S_c satisfies

$$(U_0 S_c\, U_0^{-1} g)(k,\omega) = (S_c(k)\, g)(k,\omega), \tag{2.14}$$

where

$$S_c(k) = \exp[2i \arg\Gamma(\tfrac{1}{2} + (U_0 \mathbf{L}^2 U_0^{-1} + \tfrac{1}{4})^{1/2} + i\gamma/2k)] \tag{2.15}$$

and $\mathbf{L}^2$ denotes the square of the angular momentum operator. For a discussion of the properties of S_c we refer to the literature.[19–23] We then have

Theorem 2.4: (a) Let $\phi_{l,m}$ be as in (2.5). Then

$$\tau_{a,c}(\phi_{l,m}) = \int_0^\infty k^2\, dk \int_{S^2} d\omega\, |\tilde{\phi}_{l,m}(k\omega)|^2 \frac{1}{k}\frac{\partial}{\partial k} \times \left[\delta_l^c(k) - \frac{\gamma}{2k}(1-\log a)\right]. \tag{2.16}$$

(b) Furthermore, for any $\phi\in\mathscr{D}_0$, where $\mathscr{D}_0 = \{\phi\in\mathscr{S}_0(\mathbb{R}^3) | \operatorname{supp}\tilde{\phi} \text{ is compact}\}$, we have

$$\tau_{a,c}(\phi) = (\phi, \tau_{a,c}\,\phi), \tag{2.17}$$

where $\tau_{a,c}$ is essentially self-adjoint on $\mathscr{D}_0$ and commutes with H_0. In the spectral representation of H_0, the operator $\tau_{a,c}(k)$ is given by

$$\tau_{a,c}(k) = -\frac{i}{2k} S_c^*(k)\frac{\partial}{\partial k} S_c(k) + \frac{\gamma}{2k^3}(1-\log a)1, \quad k>0. \tag{2.18}$$

Remark 2.1: The Coulomb time delay (2.17)–(2.18) depends explicitly on a due to the a dependence of the modified free evolution $U_{a,c}(t)$. Theorem 2.4 suggests that the choice $a=e$ is the most natural one in the sense that in this way the time delay formula (2.18) exactly parallels the short range result. However, the common choice, following Refs. 15 and 16, is $a=1$.

Remark 2.2: It will become clear in the course of the proofs that we took P_R to be the projector onto states localized in a sphere of radius R only for computational convenience. Obviously, the Eqs. (2.6)–(2.8), and the results of Theorems 2.1 and 2.4(b) remain valid if P_R is replaced by P_Σ, where Σ is any compact subset of $\mathbb{R}^3$ and the limit $\Sigma\to\mathbb{R}^3$ is taken appropriately.

We now present the proofs of these theorems. We start with:

Proof of Theorem 2.1: Using the explicit expression for the wave operators

$$(\Omega_{a,c}^{\pm}\phi)(\mathbf{x}) = \int d^3k\, e^{\mp(i\gamma/2|\mathbf{k}|)\log a}\, \Psi_c^{\pm}(\mathbf{k},\mathbf{x})\tilde{\phi}(\mathbf{k}), \tag{2.19}$$

where $\phi\in\mathscr{S}_0(\mathbb{R}^3)$, we get after integrating by parts N times

$$|(e^{-itH_c}\Omega_{a,c}^{\pm}\phi)(\mathbf{x})| = \frac{1}{(2|t|)^N}\left|\int d^3k\, e^{-i|t|\mathbf{k}^2}[(\nabla_{\mathbf{k}}\cdot\mathbf{k}|\mathbf{k}|^{-2})^N \times e^{\mp(i\gamma/2|\mathbf{k}|)\log a}\Psi_c^{\pm}(\mathbf{k},\mathbf{x})\tilde{\phi}(\mathbf{k})]\right|. \tag{2.20}$$

Noting that $e^{\mp(i\gamma/2|\mathbf{k}|)\log a}\Psi_c^{\pm}(\mathbf{k},\mathbf{x})\tilde{\phi}(\mathbf{k})\in\mathscr{S}(\mathbb{R}^3)$ as a function of $\mathbf{k}$ and that its support is bounded away from $\mathbf{k}=0$, and inserting the explicit structure of the Coulomb scattering wave functions, viz.,

$$\Psi_c^-(\mathbf{k},\mathbf{x}) = (2\pi)^{-3/2} e^{-\pi\gamma/4k}\Gamma(1+i\gamma/2k) \times e^{i\mathbf{k}\cdot\mathbf{x}}\, {}_1F_1(-i\gamma/2k,1,i(kr-\mathbf{k}\cdot\mathbf{x})), \tag{2.21}$$

$$\Psi_c^+(\mathbf{k},\mathbf{x}) = \overline{\Psi_c^-(-\mathbf{k},\mathbf{x})}, \quad r=|\mathbf{x}|,$$

we arrive at

$$\|P_R e^{-itH_c}\Omega_{a,c}^{\pm}\phi\|^2 \leqslant \frac{1}{(2|t|)^{2N}}\int_{|\mathbf{x}|<R} d^3x\Big[\int d^3k\, |(\nabla_{\mathbf{k}}|\mathbf{k}|^{-2}\mathbf{k})^N \times e^{\mp(i\gamma/2|\mathbf{k}|)\log a}\Psi_c^{\pm}(\mathbf{k},\mathbf{x})\tilde{\phi}(\mathbf{k})|\Big]^2 \leqslant C'_{N,R}\,|t|^{-2N}, \quad |t|>0. \tag{2.22}$$

The proof of $\|P_R U_{a,c}(t)\phi\|^2 \leqslant C''_{N,R}\,|t|^{-2N}$, $|t|>0$, is similar. ■

Remark 2.3: An alternative proof of $\|P_R e^{-itH_c}\Omega_{a,c}^{\pm}\phi\|^2 \in L^1(\mathbb{R},dt)$ can be obtained from the smoothness of P_R with respect to H_c on any interval of the type $[\delta,\infty)$, $\delta>0$ (cf. Refs. 24–26 for the corresponding results in the short range case).

Next, we collect some simple results needed in the proof of Theorem 2.2:

Lemma 2.1: (a) As $R\to\infty$, both $\sin(k-p)R/\pi(k-p)$ and $\log(|k-p|R)\sin(k-p)R/(-C)\pi(k-p)$ converge to $\delta(k-p)$ in the sense of distributions on $(0,\infty)$. Here C denotes the constant of Euler.

(b) As $R\to\infty$,

$$\int_0^R dr\, kr[J_{l+1/2}(kr)]^2 \underset{R\to\infty}{=} \frac{R}{\pi} - \frac{\sin 2(kR - l\pi/2)}{2\pi k} + o(1), \tag{2.23}$$

and

$$\int_0^R dr\,(kr)^{1/2}(pr)^{1/2} J_{l+1/2}(kr)\, J_{l+1/2}(pr) \underset{R\to\infty}{=} \frac{\sin(k-p)R}{\pi(k-p)} - (-1)^l\frac{\sin(k+p)R}{\pi(k+p)} + o(1), \tag{2.24}$$

where $k,p>0$ and $l\in\mathbb{N}$.

Proof: Part (a) of this lemma follows from Ref. 27, p. 34. Part (b) can be proved, e.g., by induction after a simple exercise in recurrence relations and standard integrals of Bessel functions.[28] ■

Proof of Theorem 2.2: Using the finiteness of the modified free transit time (2.6), proved in Theorem 2.1, and the dominated convergence theorem, we introduce its Abel-limit form, viz.,

$$T^R_{a,c}(\phi_{l,m}) = \lim_{\epsilon \to 0^+} \int_{-\infty}^{\infty} dt\, e^{-\epsilon|t|}\, \|P_R\, U_{a,c}(t)\, \phi_{l,m}\|^2. \tag{2.25}$$

Employing

$$\begin{aligned}(U_{a,c}(t)\,\phi_{l,m})(\mathbf{x}) &= (1/r)\int_0^{\infty} dk\, k\alpha_{l,m}(k) e^{-i\{tk^2 + [\gamma\epsilon(t)/2k]\log(4a|t|k^2)\}} \\ &\quad \times (kr)^{1/2} J_{l+1/2}(kr)\, Y_{l,m}(\omega_{\mathbf{x}})\end{aligned} \tag{2.26}$$

together with Fubini's theorem and an explicit t integration, we obtain for the rhs of Eq. (2.25)

$$\begin{aligned}&\lim_{\epsilon \to 0^+} \int_0^R dr \int_0^{\infty} dk \int_0^{\infty} dp\, kp\alpha_{l,m}(k)\, \overline{\alpha_{l,m}(p)} \\ &\quad \times (kr)^{1/2}(pr)^{1/2} J_{l+1/2}(kr) J_{l+1/2}(pr) \\ &\quad \times \{(4k^2a)^{-i\gamma/2k}(4p^2a)^{i\gamma/2p}\Gamma(1 - i\gamma/2k + i\gamma/2p) \\ &\quad \times e^{(\pi\gamma/4)(1/p - 1/k)}(k^2 - p^2 - i\epsilon)^{-i\gamma/2k - i\gamma/2p} \\ &\quad \times 1/i(k^2 - p^2 - i\epsilon) + \text{c.c.}\}.\end{aligned} \tag{2.27}$$

After some algebraic manipulations [where we use $(k^2 - p^2)\,P(1/(k^2 - p^2)) = 1$ with P the principal value], we can write the term between { } of expression (2.27) in the limit $\epsilon \to 0^+$ as

$$\begin{aligned}&(\pi/k)\delta(k - p) + F(k,p)[\gamma/(k + p)kp] \\ &\quad \times [\log(4k^2a) - 2 - C - \log|k^2 - p^2| + G(k,p)]\end{aligned} \tag{2.28}$$

in the sense of distributions. Here

$$\begin{aligned}F(k,p) &= |\Gamma(1 - i\gamma/2k + i\gamma/2p)|\, e^{(\pi\gamma/4)(1/k - 1/p)} \\ &\quad \times e^{(\pi\gamma/2)(1/k - 1/p)\theta(p^2 - k^2)},\end{aligned} \tag{2.29}$$

where θ denotes the step function, and

$$\begin{aligned}G(k,p) &= \frac{(k + p)kp}{\gamma} \sin\left[\frac{\gamma}{2p}\log(4p^2a) - \frac{\gamma}{2k}\log(4k^2a)\right. \\ &\quad + \arg\Gamma\left(1 - \frac{i\gamma}{2k} + \frac{i\gamma}{2p}\right) \\ &\quad \left. + \frac{\gamma}{2}\left(\frac{1}{k} - \frac{1}{p}\right)\log|k^2 - p^2|\right] \\ &\quad - \frac{k^2 - p^2}{2}\,[\log(4k^2a) - 2 - C - \log|k^2 - p^2|]\end{aligned} \tag{2.30}$$

Clearly,

$$F(k,p) \underset{k \to p}{=} 1 + O(k^2 - p^2), \quad G(k,p) = O(k^2 - p^2). \tag{2.31}$$

Inserting the results (2.28)–(2.31) in expression (2.27), exchanging the r integral with the integrals over k and p by Fubini's theorem, and recalling Lemma 2.1b, we arrive at

$$\begin{aligned}T^R_{a,c}(\phi_{l,m}) \underset{R \to \infty}{=} &\int_0^{\infty} k^2 dk\, |\alpha_{l,m}(k)|^2 \{R/k - \sin[2(kR - l\pi/2)]/2k^2\} \\ &+ \int_0^{\infty} dk \int_0^{\infty} dp\, kp\alpha_{l,m}(k)\, \overline{\alpha_{l,m}(p)} F(k,p)[\gamma/(k + p)kp] \\ &\times \{\log 2ka - 2 - C - \log|k - p| + [G(k,p) - \log|k + p| + \log 2k]\} \\ &\times \{\sin(k - p)R/(k - p) - (-1)^l \sin(k + p)R/(k + p)\} + o(1).\end{aligned} \tag{2.32}$$

Finally, using Lemma 2.1a and the Riemann–Lebesgue lemma completes the proof of the theorem. ■

To prove Theorem 2.3, we first state

Lemma 2.2: Let $\phi_{l,m}$ be as in (2.5). Then

$$\begin{aligned}T^R_c(\phi_{l,m}) &= \int_0^{\infty} k^2\, dk \int_{S^2} d\omega\, |\tilde{\phi}_{l,m}(k\omega)|^2 \\ &\quad \times (2/k)\int_0^R dr[\psi_{l,c}(k,r)]^2,\end{aligned} \tag{2.33}$$

where

$$\begin{aligned}\psi_{l,c}(k,r) &= \Gamma(2l + 2)^{-1} 2^l e^{-\pi\gamma/4k}\, |\Gamma(l + 1 + i\gamma/2k)|\, e^{ikr} \\ &\quad \times (kr)^{l+1}\, {}_1F_1(l + 1 + i\gamma/2k;\, 2l + 1;\, -2ikr).\end{aligned} \tag{2.34}$$

Proof: Consider the operator $T^R_{c,K}$ defined by

$$T^R_{c,K} = E_0(K)\, \Omega^{-*}_{a,c}\, P_R\, \Omega^{-}_{a,c}\, E_0(K), \tag{2.35}$$

where K is a compact subset of $(0,\infty)$ and E_0 represents the spectral projectors associated with H_0. Remark that $T^R_{c,K}$ is a-independent. From Eq. (2.19) and the accompanying relation

$$(\Omega^{\pm *}_{a,c}\,\phi)(\mathbf{x}) = \int d^3k\, (2\pi)^{-3/2} e^{i\mathbf{k}\cdot\mathbf{x}}\, \hat{\phi}^{\pm}_{a,c}(\mathbf{k}), \tag{2.36}$$

where

$$\begin{aligned}\hat{\phi}^{\pm}_{a,c}(\mathbf{k}) &= \int d^3x\, e^{\pm(i\gamma/2|\mathbf{k}|)\log a} \\ &\quad \times \overline{\Psi^{\pm}_c(\mathbf{k},\mathbf{x})}\,\phi(\mathbf{x}), \quad \phi \in \mathscr{S}_0(\mathbb{R}^3),\end{aligned} \tag{2.37}$$

we infer that $T^R_{c,K}$ is a trace class operator with

$$\|T^R_{c,K}\|_1 = \int_{k^2 \in K} d^3k \int_{|y| < R} d^3y\, |\Psi^{-}_c(\mathbf{k},\mathbf{y})|^2 < \infty. \tag{2.38}$$

This implies, by a well-known Theorem of Jauch *et al.*[29] (cf. also Refs. 1,30,31) the existence of a essentially unique family of trace class operators $T^R_{c,K}(k)$, acting in $L^2(S^2)$, such that

$$T^R_c(\phi) = \int_{-\infty}^{+\infty} dt\, (e^{-itH_0}\phi, T^R_{c,K} e^{-itH_0}\phi)$$

$$= \int_0^\infty dk \int_{S^2} d\omega \ \overline{(U_0\phi)}(k,\omega)(T^R_{c,K}(k)U_0\phi)(k,\omega), \quad \phi \in \mathscr{D}_0, \tag{2.39}$$

where K is chosen such that $\{k^2 \in [0,\infty) | \mathbf{k} \in \text{supp}\phi\} \subseteq K$. In particular, for a.e. $k^2 \in (0,\infty)$, the kernel of $T^R_{c,K}(k)$ is given by

$$T^R_{c,K}(k,\omega,\omega') = \chi_K(k^2) \int_{|y|<R} d^3y \times \overline{\Psi_c^-(k\omega,\mathbf{y})}\, \Psi_c^-(k\omega',\mathbf{y}), \tag{2.40}$$

with χ_M the characteristic function of the set M. Insertion of Eq. (2.40) and of the partial wave expansion for Ψ_c^-, viz.,

$$\Psi_c^-(\mathbf{k},\mathbf{x}) = (2\pi)^{-3/2} 4r(kr)^{-1} \sum_{l=0}^{\infty} \sum_{m=-l}^{+l} i^l e^{i\delta_l^c(k)} \times \psi_{l,c}(k,r)\, \overline{Y_{l,m}(\omega_{\mathbf{k}})} Y_{l,m}(\omega_{\mathbf{x}}), \tag{2.41}$$

into Eq. (2.39) yields the result (2.33). ■

Lemma 2.3: As $r \to \infty$

$$\frac{2}{k}\left(\frac{\partial\psi_{l,c}}{\partial k^2}\frac{\partial\psi_{l,c}}{\partial r} - \psi_{l,c}\frac{\partial^2\psi_{l,c}}{\partial k^2 \partial r}\right)(k,r) \underset{r\to\infty}{=} \frac{r}{k} + \frac{\gamma}{2k^3}(\log 2kr - 1) + \frac{1}{k}\left(\frac{\partial}{\partial k}\delta_l^c\right)(k) - \frac{1}{2k^2}\sin 2\left[kr - \frac{\gamma}{2k}\log 2kr - \frac{l\pi}{2} + \delta_{l,c}(k)\right] + O(r^{-1}\log r) \tag{2.42}$$

uniformly in k if k varies in intervals of the type $[\delta,\infty)$, $\delta > 0$.

Proof: This follows directly from Eq. (2.34) and the asymptotic behavior of the confluent hypergeometric function ${}_1F_1$.[32] ■

With these results, we can give

Proof of Theorem 2.3: From the Schrödinger equation we find

$$[\psi_{l,c}(k,r)]^2 = \frac{d}{dr}\left(\frac{\partial\psi_{l,c}}{\partial k^2}\frac{\partial\psi_{l,c}}{\partial r} - \psi_{l,c}\frac{\partial^2\psi_{l,c}}{\partial k^2\partial r}\right)(k,r). \tag{2.43}$$

Application of Lemma 2.2, Lemma 2.3, and the Riemann–Lebesgue lemma immediately leads to Eq. (2.11). ■

Proof of Theorem 2.4: Part (a) of this theorem follows from combining the definition (2.12) with the results of Theorems 2.2 and 23. The results (2.17) and (2.18) of part (b) are obtained by noting that

$$\tau_{a,c}(k) = \bigoplus_{l=0}^{\infty} \tau_{a,c,l}(k) \tag{2.44}$$

where

$$\tau_{a,c,l}(k) = \frac{1}{k}\frac{\partial}{\partial k}\left[\delta_l^c(k) - \frac{\gamma}{2k}(1 - \log a)\right]. \tag{2.45}$$

The arguments to prove that $\tau_{a,c}$ is essentially self-adjoint on $\mathscr{D}_0$ parallel to those in Ref. 4. ■

We end this discussion of Coulomb time delay with some remarks concerning possible extensions of the results obtained so far.

Remark 2.4: (a) It is straightforward to see that $\phi_{l,m} \in \mathscr{S}_0(\mathbb{R}^3)$, respectively $\phi \in \mathscr{S}_0(\mathbb{R}^3)$, suffices for Theorems 2.2–2.4 to hold. The distributional arguments in the proof of Theorem 2.2 certainly extend to $\phi_{l,m} \in \mathscr{S}_0(\mathbb{R}^3)$. The proof of Theorem 2.3 that is presented here does not simply extend to such $\phi_{l,m}$'s. However, one can derive the results of Lemma 2.2 again on the basis of the distributional approach used in the proof of Theorem 2.2 without any trace class arguments. Thus Theorem 2.3 also hold for $\phi_{l,m} \in \mathscr{S}_0(\mathbb{R}^3)$. Consequently, Theorem 2.4 is valid again for $\phi_{l,m}$, $\phi \in \mathscr{S}_0(\mathbb{R}^3)$ and with $\mathscr{D}_0$ replaced by $\mathscr{S}_0(\mathbb{R}^3)$. (b) Replacing l by $-\frac{1}{2} + [l(l+1)+\alpha]^{1/2}$ and L^2 by $L^2 + (\alpha - \frac{1}{4})1$ [Eq. (2.15)], all the results of this section are valid for Hamiltonians of the form $-\Delta + \gamma/|\mathbf{x}| + (\alpha - \frac{1}{4})/|\mathbf{x}|^2$, $\alpha > 0$, $\mathbf{x} \in \mathbb{R}^3 \setminus \{0\}$, defined as the Friedrichs extension of the minimal operator on $C_0^\infty(\mathbb{R}^3 \setminus \{0\})$ (see also Sec. III). Thus they also extend to $n \geqslant 2$ dimensions with Hamiltonians of the type $-\Delta + \gamma/|\mathbf{x}| + [\alpha - (n-2)^2/4]/|\mathbf{x}|^2$, $\alpha > 0$, $\mathbf{x} \in \mathbb{R}^n \setminus \{0\}$.

B. Classical mechanics

The aim of this part of Sec. II is to derive the analog of Theorem 2.4 within the context of classical mechanics. We follow the treatment of Narnhofer and Thirring,[33] but, in addition, we discuss the a dependence of all quantities of interest.

We start by introducing the Coulomb-modified Møller transformations. Let $\Phi^t_{a,c}$ denote the modified free time evolution

$$\Phi^t_{a,c}: \begin{cases} \mathbb{R}^3\times\mathbb{R}^3\setminus\{0\} \to \mathbb{R}^3\times\mathbb{R}^3\setminus\{0\} \\ (\mathbf{y},\dot{\mathbf{y}}) \to \Phi^t_{a,c}(\mathbf{y},\dot{\mathbf{y}}) \end{cases} = (\mathbf{y} + \dot{\mathbf{y}}\{t - [2\gamma\,\text{sgn}(t)/|\dot{\mathbf{y}}|^3]\log[(a/e^2)|\dot{\mathbf{y}}|^2|t|)]\}, \dot{\mathbf{y}}), \quad a > 0, \ t \neq 0. \tag{2.46}$$

Furthermore, let $\Gamma = \{(\mathbf{z},\dot{\mathbf{z}}) \in \mathbb{R}^3\times\mathbb{R}^3 | z \wedge \dot{z} \neq 0\}$ and let Φ^t_c be the time evolution of a particle under the influence of the Coulomb potential $\gamma/|\mathbf{x}|$

$$\Phi^t_c: \begin{cases} \Gamma \to \Gamma \\ (\mathbf{x},\dot{\mathbf{x}}) \to \Phi^t_c(\mathbf{x},\dot{\mathbf{x}}) = (\mathbf{x}(t),\dot{\mathbf{x}}(t)) \end{cases}, \tag{2.47}$$

where $\mathbf{x}(t)$ solves Newton's equation of motion

$$\ddot{\mathbf{x}}(t) = [2\gamma/|\mathbf{x}(t)|^3]\mathbf{x}(t), \quad \mathbf{x}(0) = \mathbf{x}, \quad \dot{\mathbf{x}}(0) = \dot{\mathbf{x}}, \quad \gamma \in \mathbb{R}, \tag{2.48}$$

with $2m = 1$. Then the Møller transformations $\omega^\pm_{a,c}$ defined as

$$\omega^\pm_{a,c} = \lim_{t \to \pm\infty} \Phi_c^{-t} \circ \Phi^t_{a,c} \tag{2.49}$$

exist on Γ and are complete.[34]

The Coulomb scattering transformation S_c is defined by

$$S_c = \omega^{+-1}_{a,c} \circ \omega^-_{a,c}, \quad S_c(\mathbf{x}_-,\dot{\mathbf{x}}_-) = (\mathbf{x}_+,\dot{\mathbf{x}}_+). \tag{2.50}$$

It represents a canonical transformation from Γ onto itself. (If $\gamma < 0$, the existence of the parabolic orbit at energy $E = 0$ implies that S_c extends from Γ onto itself such that $(\mathbf{x}_-,0)$ is mapped to $(\mathbf{x}_+,0)$). By the intertwining property[34] of $\omega^\pm_{a,c}$, viz.,

$$\omega^\pm_{a,c} \circ \Phi^t_0 = \Phi^t_0 \circ \omega^\pm_{a,c}, \tag{2.51}$$

where Φ^t_0 denotes the free time evolution

$$\Phi^t_0: \begin{cases} \mathbb{R}^3\times\mathbb{R}^3 \to \mathbb{R}^3\times\mathbb{R}^3 \\ (\mathbf{y},\dot{\mathbf{y}}) \to (\mathbf{y} + \dot{\mathbf{y}}t, \dot{\mathbf{y}}), \end{cases} \tag{2.52}$$

S_c is independent of a. To describe the generator of S_c, we introduce the following canonical coordinates[33]:

$$k = |\dot{\mathbf{y}}|, \quad L = \tfrac{1}{2}|\mathbf{y} \wedge \dot{\mathbf{y}}|, \quad L_3 = \tfrac{1}{2}(\mathbf{y} \wedge \dot{\mathbf{y}})_3,$$

$$d = \mathbf{y}\cdot\dot{\mathbf{y}}/|\dot{\mathbf{y}}|, \quad \theta = \arccos \frac{L_1\dot{y}_2 - L_2\dot{y}_1}{k(L_1^2 + L_2^2)^{1/2}},$$

$$\zeta = \arccos \frac{L_2}{(L_1^2 + L_2^2)^{1/2}}, \tag{2.53}$$

such that

$$\Phi_0^t(d,\theta,\zeta,k,L,L_3) = (d + kt,\theta,\zeta,k,L,L_3). \tag{2.54}$$

One has then the following result:

Lemma 2.4: Let $k > 0$. Then S_c is given by

$$S_c(d,\theta,\zeta,k,L,L_3) = \left(d - 2\left(\frac{\partial}{\partial k}\delta_c\right)(k,L),\right.$$
$$\left.\theta - 2\left(\frac{\partial}{\partial L}\delta_c\right)(k,L),\zeta;k,L,L_3\right), \tag{2.55}$$

where

$$\delta_c(k,L) = L\,\mathrm{sgn}(\gamma)\mathrm{arccot}\,(2kL/|\gamma|) + (\gamma/4k)[-2 + \log(L^2 + \gamma^2/4k^2)]. \tag{2.56}$$

In particular,

$$2\left(\frac{\partial}{\partial L}\delta_c\right)(k,L) = 2\,\mathrm{sgn}(\gamma)\mathrm{arccot}(2kL/|\gamma|) \tag{2.57}$$

represents the Coulomb scattering angle and

$$2\left(\frac{\partial}{\partial k}\delta_c\right)(k,L) = -\frac{\gamma}{2k^2}\log\left(L^2 + \frac{\gamma^2}{4k^2}\right). \tag{2.58}$$

Proof: See Narnhofer and Thirring.[33] ■

Finally, the classical Coulomb time delay $\tau_{a,c}(k,L)$ is defined as

$$\tau_{a,c}(k,L) = \lim_{R\to\infty}\,[t_c^R(k,L) - t_{a,c}^R(k,L)], \quad k > 0, \tag{2.59}$$

where $t_{a,c}^R(k,L)$ is the modified free transit time through a sphere of radius R centered at the origin and $t_c^R(k,L)$ is the pure Coulomb transit time through that sphere. In the limit $R \to \infty$, we obtain from Eq. (2.46)

$$t_{a,c}^R(k,L)\underset{R\to\infty}{=} \frac{R}{k} + \frac{\gamma}{2k^3}\log(2kaR) - \frac{\gamma}{k^3} + o(1), \quad k > 0, \tag{2.60}$$

in complete analogy with the quantum result (2.10). Expanding $t_c^R(k,L)$, which is given by

$$t_c^R(k,L) = \int_{r_c}^R dr\left(k^2 - \frac{\gamma}{r} - \frac{L^2}{r^2}\right)^{-1/2},$$
$$r_c = \frac{\gamma}{2k^2} + \left(\frac{\gamma^2}{4k^4} + \frac{L^2}{k^2}\right)^{1/2}, \tag{2.61}$$

in the limit $R \to \infty$ yields

$$t_c^R(k,L)\underset{R\to\infty}{=} \frac{R}{k} + \frac{\gamma}{2k^3}\log(2kR) - \frac{\gamma}{2k^3} - \frac{\gamma}{4k^3}\log\left(L^2 + \frac{\gamma^2}{4k^2}\right) + o(1), \quad k > 0 \tag{2.62}$$

In that way we obtain the classical analog of Theorem 2.4.

Theorem 2.5: Let $k > 0$. Then the classical time delay $\tau_{a,c}(k,L)$ is given by

$$\tau_{a,c}(k,L) = -\frac{\gamma}{4k^3}\log\left(L^2 + \frac{\gamma^2}{4k^2}\right) + \frac{\gamma}{2k^3}(1 - \log a)$$
$$= \frac{1}{k}\frac{\partial}{\partial k}\left[\delta_c(k,L) - \frac{\gamma}{2k}(1 - \log a)\right]. \tag{2.63}$$

This result is in complete agreement with the corresponding quantum result (2.16).

It is clear that the analog of Remark 2.4(b) is valid in the classical case. For a treatment of classical time delay in an Hilbert space approach we refer to Refs. 35 and 36.

III. RELATIVE AND TOTAL TIME DELAY FOR SPHERICALLY SYMMETRIC COULOMB-TYPE INTERACTIONS

The concept of relative and total time delay for two particles interacting with a Coulomb plus a spherically symmetric short range potential is investigated both classically and quantum-mechanically. In particular, its connection with the S matrix of the system is derived.

A. Quantum mechanics

Let us assume that $V: (0,\infty) \to R$ is measurable, locally integrable and satisfies

$$rV(r) \in L^1(0,\infty). \tag{3.1}$$

With respect to the decomposition

$$L^2(\mathbb{R}^3) = L^2((0,\infty),r^2dr) \otimes L^2(S^2), \tag{3.2}$$

we define the self-adjoint Hamiltonian of our system, which is formally described by the differential expression $-\Delta + \gamma/|\mathbf{x}| + (\alpha - \frac{1}{4})/|\mathbf{x}|^2 + V(|\mathbf{x}|)$, as

$$H = \bigoplus_{l=0}^{\infty} H_l \otimes 1. \tag{3.3}$$

Here, H_l represents the Friedrichs extension[17] of the minimal operator $\dot{H}_l$, viz.,

$$(\dot{H}_l\, g)(r) = -g''(r) - \frac{2}{r}g'(r) + \frac{l(l+1) + \alpha - \frac{1}{4}}{r^2}g(r) + \frac{\gamma}{r}g(r) + V(r)\,g(r), \quad r > 0, \quad \gamma \in \mathbb{R}, \quad \alpha > 0, \tag{3.4}$$

with domain

$$\mathscr{D}(\dot{H}_l) = \{\, g \mid g' \text{ is locally absolutely continuous on } (0,\infty);$$
$$\mathrm{supp}(g) \subset (0,\infty) \text{ is compact;}$$
$$g, -g'' - 2/r\,g' + \frac{l(l+1) + \alpha - \frac{1}{4}}{r^2}g + \frac{\gamma}{r}g + Vg \in L^2((0,\infty);r^2\,dr)\Big\}.$$

Furthermore, we define the Hamiltonians H_α and $H_{\alpha,l}$ according to Eqs. (3.3) and (3.4), but assuming V to vanish identically. Then it is well known[23] that the wave operators

$$\Omega_{a,\alpha}^{\pm} = \underset{t\to\pm\infty}{\text{s-lim}}\, e^{itH_\alpha}\, U_{a,c}(t) \tag{3.5}$$

and

$$\Omega_a^{\pm} = \underset{t \to \pm\infty}{\text{s-lim}}\, e^{itH} U_{a,c}(t) \tag{3.6}$$

exist and are asymptotically complete (in particular, the singular continuous spectrum and the positive point spectrum of H and H_α are empty).

Next, we state our results. The proofs will be given at the end of the section.

Lemma 3.1: Let $\phi \in \mathscr{D}_0 = \{ \phi \in \mathscr{S}_0(\mathbb{R}^3) | \operatorname{supp}\tilde{\phi}$ is compact$\}$. Then

$$\int_{-\infty}^{+\infty} dt \left[\| P_R\, e^{itH} \Omega_a^{\pm} \phi \|^2 + \| P_R\, e^{itH_\alpha} \Omega_{a,\alpha}^{\pm} \phi \|^2 \right] < \infty. \tag{3.7}$$

Using this, we can define the transit times $T^R(\phi)$ and $T_\alpha^R(\phi)$, for a state $\phi \in \mathscr{D}_0$, through a sphere of radius R centered at the origin of the scattering system described by H and H_α, respectively, viz.,

$$T^R(\phi) = \int_{-\infty}^{+\infty} dt \| P_R\, e^{-itH} \Omega_a^- \phi \|^2, \tag{3.8}$$

$$T_\alpha^R(\phi) = \int_{-\infty}^{+\infty} dt \| P_R\, e^{-itH_\alpha} \Omega_{a,\alpha}^- \phi \|^2. \tag{3.9}$$

Again, $T^R(\phi)$ and $T_\alpha^R(\phi)$ are independent of a. The relative and total time delay, $\tau_{sc}(\phi)$ and $\tau_a(\phi)$, associated with the state ϕ, are then defined as

$$\tau_{sc}(\phi) = \lim_{R \to \infty} [T^R(\phi) - T_\alpha^R(\phi)] \tag{3.10}$$

and

$$\tau_a(\phi) = \lim_{R \to \infty} [T^R(\phi) - T_{a,c}^R(\phi)], \tag{3.11}$$

whenever these limits exist. The behavior of T^R for large R is obtained in the following:

Theorem 3.1: Let $\phi_{l,m}$ be as in (2.5). Let V satisfy condition (3.1). Then, as $R \to \infty$, we have

$$T^R(\phi_{l,m}) \underset{R \to \infty}{=} \int_0^\infty k^2\, dk \int_{S^2} d\omega |\tilde{\phi}_{l,m}(k\omega)|^2 \left[\frac{R}{k} + \frac{\gamma}{2k^3} \times \log(2kR) - \frac{\gamma}{2k^3} + \frac{1}{k} \frac{\partial}{\partial k} \delta_l(k) \right] + o(1), \tag{3.12}$$

where δ_l denotes the phase shift associated with H_l.

Splitting up $\delta_l(k)$ according to

$$\delta_l(k) = \delta_l^\alpha(k) + \delta_l^{sc}(k), \tag{3.13}$$

where

$$\delta_l^\alpha(k) = \arg\Gamma(\tfrac{1}{2} + [l(l+1) + \alpha]^{1/2} + i\gamma/2k) + \tfrac{1}{2}\pi \times \{ l + \tfrac{1}{2} - [l(l+1) + \alpha]^{1/2} \} \tag{3.14}$$

represents the phase shift associated with $H_{\alpha,l}$, we obtain the following result concerning relative time delay:

Theorem 3.2: (a) Let $\phi_{l,m}$ be as in (2.5). Let V satisfy condition (3.1). Then

$$\tau_{sc}(\phi_{l,m}) = \int_0^\infty k^2 dk \int_{S^2} d\omega \times |\tilde{\phi}_{l,m}(k\omega)|^2 \frac{1}{k} \frac{\partial}{\partial k} \delta_l^{sc}(k). \tag{3.15}$$

(b) Furthermore, for any $\phi \in \mathscr{D}_0$, we have

$$\tau_{sc}(\phi) = (\phi, \tau_{sc}\, \phi), \tag{3.16}$$

where τ_{sc} is essentially self-adjoint on $\mathscr{D}_0$ and commutes with H_0. Its energy shell components $\tau_{sc}(k)$ fulfill

$$\tau_{sc}(k) = \bigoplus_{l=0}^{\infty} \frac{1}{k} \frac{\partial}{\partial k} \delta_l^{sc}(k), \quad k > 0. \tag{3.17}$$

Note that from Eq. (3.17) the relation between $\tau_{sc}(k)$ and the S matrices involved in the problem is clear, once one chooses how to split the total S matrix (see, e.g., Ref. 19).

For the total time delay, we can write

Theorem 3.3: (a) Let $\phi_{l,m}$ be as in (2.5) and V satisfy condition (3.1). Then

$$\tau_a(\phi_{l,m}) = \int_0^\infty k^2\, dk \int_{S^2} d\omega |\tilde{\phi}_{l,m}(k\omega)|^2 \times \frac{1}{k} \frac{\partial}{\partial k} \left[\delta_l(k) - \frac{\gamma}{2k} (1 - \log a) \right]. \tag{3.18}$$

(b) Furthermore, for any $\phi \in \mathscr{D}_0$

$$\tau_a(\phi) = (\phi, \tau_a\, \phi), \tag{3.19}$$

where τ_a is essentially self-adjoint on $\mathscr{D}_0$ and commutes with H_0. The energy shell components $\tau_a(k)$ of τ_a are given by

$$\tau_a(k) = -\frac{i}{2k} S^*(k) \frac{\partial}{\partial k} S(k) + \frac{\gamma}{2k^3} (1 - \log a) 1, \quad k > 0, \tag{3.20}$$

where $S(k)$ in $L^2(S^2)$ denotes the total on-shell scattering operator for the system described by H.

Remark 3.1: (a) Note that our methods, being a combination of time-dependent arguments, distributional methods and an extension of the stationary approach due to Smith, give rather strong results under extremely weak assumptions on the interaction. Loosely speaking, our conditions allow short-range potentials $V(r)$ which behave like $r^{-2-\epsilon}$ at infinity and $r^{-2+\epsilon}$ at zero. In addition we are able to include the long-range interaction $\gamma/r + (\alpha - \frac{1}{4})/r^2$, $\gamma \in \mathbb{R}$, $\alpha > 0$. So, even in the pure short range case ($\gamma = 0$, $\alpha = \frac{1}{4}$) Theorem 3.3 is stronger than the results obtained in the literature. So far, the weakest condition on the behavior at infinity in the spherically symmetric case is $r^{-5/2-\epsilon}$, $\epsilon > 0$.[6]

(b) Extensions of our results to general nonspherically symmetric potentials are in preparation. For a recent geometric approach to time delay for (general) short range interactions ($\gamma = 0$, $\alpha = \frac{1}{4}$) we refer to Jensen.[4,5] In the context of the classical Lax–Philips scattering theory, time delay has been discussed recently in Refs. 2 and 3. The notion of relative time delay in the presence of Coulomb interactions has been considered very briefly by Said[37] in the spirit of Ref. 29 (i.e., within Krein's approach). Relative time delay in the presence of a uniform electric field has been treated in Ref. 38. Of course, Remark 2.4(a) and the second part of Remark 2.4(b) do also apply to the results of this section.

Let us now turn to the proofs

Proof of Lemma 3.1: Using the trace methods employed in the proof of Lemma 2.2, we arrive at

$$T^R(\phi)=\int_0^\infty dk\int_{S^2}d\omega\ \overline{(U_0\phi)(k,\omega)}\times(T_K^R(k)U_0\phi)(k,\omega)<\infty. \tag{3.21}$$

[A similar result holds for $T_\alpha^R(\phi)$]. Here T_K^R (k) is a trace class operator in $L^2(S^2)$ with kernel

$$T_K^R(k,\omega,\omega')=\chi_K(k^2)\int_{|y|<R}d^3y\times\overline{\Psi^-(k,\omega,y)}\Psi^-(k,\omega',y), \tag{3.22}$$

where Ψ^- is now given by

$$\Psi^-(k,x)=(2\pi)^{-3/2}4\pi(kr)^{-1}\times\sum_{l=0}^{\infty}\sum_{m=-l}^{+l}i^l e^{i\delta_l(k)}\psi_l(k,r)\,\overline{Y_{l,m}(\omega_k)}\,Y_{l,m}(\omega_x), \tag{3.23}$$

with ψ_l a solution of the radial Schrödinger equation such that

$$\psi_l(k,r)\underset{r\to\infty}{\overset{k>0}{\sim}}\sin[kr-(\gamma/2k)\log 2kr-l\pi/2+\delta_l(k)]. \ \blacksquare \tag{3.24}$$

Proof of Theorem 3.1: Following the proof of Lemma 2.2 step by step and noting that Eq. (2.43) stays valid with $\psi_{l,c}$ replaced by ψ_l, we obtain

$$T^R(\phi_{l,m})=\int_0^\infty k^2\,dk\int_{S^2}d\omega\,|\tilde\phi_{l,m}(k\omega)|^2\times\frac{2}{k}\left(\frac{\partial\psi_l}{\partial k^2}\frac{\partial\psi_l}{\partial r}-\psi_l\frac{\partial^2\psi_l}{\partial k^2\partial r}\right)(k,R). \tag{3.25}$$

In order to determine the large R behavior of the rhs of Eq. (3.25), we first introduce

$$F_l^\alpha(k,r)=2^{\lambda-1}e^{-\pi\gamma/4k}\Gamma(2\lambda-1)|\Gamma(\lambda+i\gamma/2k)|(kr)^\lambda\times e^{-ikr}{}_1F_1(\lambda-i\gamma/2k;2\lambda;2ikr) \tag{3.26}$$

and

$$G_l^\alpha(k,r)=-i2^\lambda e^{\pi\gamma/4k}e^{i\pi\lambda}e^{-i\delta_l^\alpha(k)}\times(kr)^\lambda e^{-ikr}U(\lambda-i\gamma/2k;2\lambda;2ikr), \tag{3.27}$$

where $\lambda=\frac{1}{2}+[l(l+1)+\alpha]^{1/2}$ and $U(a,b,z)$ denotes the irregular confluent hypergeometric function.[32] Secondly, we define

$$\hat g_l^\alpha(k,r,r')=(1/k)\begin{cases}F_l^\alpha(k,r')G_l^\alpha(k,r), & r'\leqslant r,\\ F_l^\alpha(k,r)G_l^\alpha(k,r'), & r'\geqslant r,\end{cases} \tag{3.28}$$

and

$$g_l^\alpha(k,r,r')=(1/k)[F_l^\alpha(k,r)G_l^\alpha(k,r')-F_l^\alpha(k,r')G_l^\alpha(k,r)]. \tag{3.29}$$

Since the interaction satisfies $[r/(1+r)]|V(r)|\in L^1(0,\infty)$ [see (3.1)], the function $F_l(k,r)$ given by

$$F_l(k,r)=e^{-i\delta_l^{sc}(k)}\psi_l(k,r), \tag{3.30}$$

obeys the following integral equations[17,18]:

$$F_l(k,r)=F_l^\alpha(k,r)-\int_0^\infty dr'\,\hat g_l^\alpha(k,r,r')V(r')F_l(k,r'), \tag{3.31}$$

$$F_l(k,r)=F_l^\alpha(k,r)+\int_0^r dr'\,g_l^\alpha(k,r,r')V(r')F_l(k,r'), \tag{3.32}$$

Furthermore,

$$e^{-2i\delta_l^{sc}(k)}=1+(2i/k)\int_0^\infty dr\,V(r)\,F_l^\alpha(k,r)F_l(k,r). \tag{3.33}$$

Using the bounds[17,18]

$$|F_l^\alpha(k,r)|\leqslant a_l(k_0)(r/(1+r))^\lambda, \tag{3.34}$$

$$|G_l^\alpha(k,r)|\leqslant b_l(k_0)(r/(1+r))^{1-\lambda}, \tag{3.35}$$

and

$$|g_l^\alpha(k,r,r')|\leqslant(1/k)c_l(k_0)(r/(1+r))^\lambda(r'/(1+r'))^{1-\lambda},\quad r\geqslant r',\ k\geqslant k_0>0, \tag{3.36}$$

we infer after iterating Eq. (3.32) that

$$|F_l(k,r)|\leqslant d_l(k_0)(r/(1+r))^\lambda,\quad k\geqslant k_0>0, \tag{3.37}$$

where $a_l(k_0)$, $b_l(k_0)$, $c_l(k_0)$, and $d_l(k_0)$ are appropriate constants. In addition, since $r|V(r)|\in L^1(0,\infty)$ [see (3.1)], differentiating Eq. (3.32) with respect to k and noting that

$$\left|\frac{\partial}{\partial k}F_l^\alpha(k,r)\right|\leqslant\tilde a_l(k_0)r\left(\frac{r}{1+r}\right)^{\lambda-1}, \tag{3.38}$$

$$\left|\frac{\partial}{\partial k}G_l^\alpha(k,r)\right|\leqslant\tilde b_l(k_0)r\left(\frac{r}{1+r}\right)^{1-\lambda}, \tag{3.39}$$

and

$$\left|\frac{\partial}{\partial k}g_l^\alpha(k,r,r')\right|\leqslant\tilde c_l(k_0)r\left(\frac{r}{1+r}\right)^{\lambda-1}\left(\frac{r'}{1+r'}\right)^{2-\lambda},\quad r\geqslant r',\ k\geqslant k_0>0, \tag{3.40}$$

we can prove after iteration that

$$\left|\frac{\partial}{\partial k}F_l(k,r)\right|\leqslant\tilde d_l(k_0)r\left(\frac{r}{1+r}\right)^{\lambda-1},\quad k\geqslant k_0>0, \tag{3.41}$$

where $\tilde a_l(k_0),\ldots,\tilde d_l(k_0)$ are again appropriate constants. So, Eqs. (3.33), (3.37), and (3.41) show that, for $k>0$, $\exp[-2i\delta_l^{sc}(k)]$ is continuously differentiable with respect to k. Next, combining Eqs. (3.28), (3.30), (3.31), and (3.33), we arrive at

$$\psi_l(k,r)=e^{i\delta_l^{sc}(k)}F_l^\alpha(k,r)+\sin[\delta_l^{sc}(k)]G_l^\alpha(k,r)-e^{i\delta_l^{sc}(k)}F_l^\alpha(k,r)\frac{1}{k}\int_r^\infty dr'\,V(r')G_l^\alpha(k,r')F_l(k,r')+e^{i\delta_l^{sc}(k)}G_l^\alpha(k,r)\frac{1}{k}\int_r^\infty dr'\,V(r')F_l^\alpha(k,r')F_l(k,r'). \tag{3.42}$$

Therefore, using Eq. (3.42), we get that the integrand of Eq. (3.25) can be written in the limit $R\to\infty$ as

$$|\tilde\phi_{l,m}(k,\omega)|^2\left(\frac{R}{k}+\frac{\gamma}{2k^3}\log 2kR-\frac{\gamma}{2k^3}+\frac{1}{k}\frac{\partial}{\partial k}\delta_l(k)-\frac{1}{2k^2}\sin\left\{2\left[kR-\frac{\gamma}{2k}\log 2kR-\frac{l\pi}{2}+\delta_l(k)\right]\right\}+o(1)\right), \tag{3.43}$$

uniformly in k for $k\geqslant k_0>0$. Applying then the Riemann–Lebesque lemma to the oscillating term in this result (3.43) completes the proof. ■

Proof of Theorem 3.2: Eq. (3.15) follows immediately from definition (3.10) and Eqs. (3.12) and (3.13). Part (b) is proved similarly as in Theorem 2.4. ■

Proof of Theorem 3.3: The result (3.18) follows from definition (3.11) and Eqs. (2.10) and (3.12). Again part (b) is proved similarly as in Theorem 2.4. ■

Remark 3.2: One could think of generalizing the approach of Martin (see Refs. 1, 6, and 7) to the quantum mechanical systems treated here. However, in order to guarantee the interchangeability of the $R \to \infty$ limit with certain time integrals in this approach, conditions like

$$\| [- e^{itH} \Omega^{-} - U(t)] \phi \| \in L^1((-\infty, t_0); dt), \quad \phi \in \mathscr{S}_0(\mathbb{R}^3),$$

play a crucial role. But already in the pure Coulomb case, using Cook's estimate (which seems to be unavoidable) and a splitting of $U_{a,c}(t)$ into three pieces [similar to Ref. 15, Eq. (44)], we obtain, after some manipulations, that

$$\| \left[e^{-itH_c} \Omega_{a,c}^{-} - U_{a,c}(t) \right] \phi \| \leqslant \int_{-\infty}^{t} ds \left\| \frac{d}{ds} \left[e^{isH_c} U_{a,c}(s) \right] \phi \right\| = O(|t|^{-1}), \quad \phi \in \mathscr{S}_0(\mathbb{R}^3), \quad \text{as } t \to -\infty.$$

B. Classical mechanics

In this part of Sec. III, we present the classical analog of Theorems 3.2 and 3.3. Throughout this discussion, we assume that $|V(\mathbf{x})| \leqslant \text{const}\,(1 + |\mathbf{x}|)^{-2-\epsilon}$, $|(\partial/\partial x_i) V(\mathbf{x})| \leqslant \text{const}\,(1 + |\mathbf{x}|)^{-3-\epsilon}$, $|(\partial^2/\partial x_i \partial x_j) V(\mathbf{x})| \leqslant \text{const}\,(1 + |\mathbf{x}|)^{-4-\epsilon}$, $\epsilon > 0$, $i, j = 1,2,3$.

Let us start by defining the time evolution

$$\Phi^t : \begin{cases} \Gamma \to \Gamma \\ (\mathbf{x}, \dot{\mathbf{x}}) \to (\mathbf{x}(t), \dot{\mathbf{x}}(t)), \end{cases} \tag{3.44}$$

where $\mathbf{x}(t)$ now solves the following equation:

$$\ddot{\mathbf{x}}(t) = \frac{2\gamma}{|\mathbf{x}(t)|^3} \mathbf{x}(t) + \frac{4\alpha - 1}{|\mathbf{x}(t)|^4} \mathbf{x}(t) - 2(\nabla V)(|\mathbf{x}(t)|), \tag{3.45}$$

$$\mathbf{x}(0) = \mathbf{x}, \quad \dot{\mathbf{x}}(0) = \dot{\mathbf{x}}, \quad \gamma \in \mathbb{R}, \quad \alpha > \tfrac{1}{4}.$$

Following Ref. 34, we know that the Møller transformations $\omega_a^{\pm}$ defined by

$$\omega_a^{\pm} = \lim_{t \to \pm\infty} \Phi^{-t} \circ \Phi_{a,c}^{t} \tag{3.46}$$

exist on all of Γ and are complete. In addition, the scattering transformation S given by

$$S = \omega_a^{+\,-1} \circ \omega_a^{-} \tag{3.47}$$

is a canonical transformation of Γ onto itself. Finally, let δ_α denote the generator of the scattering transformation S_α that corresponds to the system described by Eq. (3.45) with V taken identically zero. It can be written as

$$\begin{aligned} \delta_\alpha(k,L) = {} & (L^2 + \alpha - \tfrac{1}{4})^{1/2} \operatorname{sgn}(\gamma) \operatorname{arccot}[(2k/|\gamma|) \\ & \times (L^2 + \alpha - \tfrac{1}{4})^{1/2}] + (\gamma/4k)[-2 \\ & + \log(L^2 + \alpha - \tfrac{1}{4} + \gamma/4k^2)] + \tfrac{1}{2}\pi \\ & \times [L - (L^2 + \alpha - \tfrac{1}{4})^{1/2}], \\ & k > 0. \end{aligned} \tag{3.48}$$

As a first result, we obtain

Lemma 3.2: Let $k > 0$. Then S is given by

$$S(d,\theta,\xi;k,L,L_3) = \left(d - 2\left(\frac{\partial}{\partial k}\delta\right)(k,L), \theta - 2\left(\frac{\partial}{\partial L}\delta\right)(k,L), \xi; k, L, L_3\right), \tag{3.49}$$

where

$$\begin{aligned} \delta(k,L) = {} & \delta_\alpha(k,L) + \lim_{R \to \infty} \left\{ \int_{r'}^{R} dr \right. \\ & \times \left[k^2 - \frac{L^2 + \alpha - \frac{1}{4}}{r^2} - \frac{\gamma}{r} - V(r) \right]^{1/2} \\ & \left. - \int_{r'_\alpha}^{R} dr \left(k^2 - \frac{L^2 + \alpha - \frac{1}{4}}{r^2} - \frac{\gamma}{r} \right)^{1/2} \right\}, \end{aligned} \tag{3.50}$$

with r' and r'_α the classical turning points. In particular,

$$\begin{aligned} 2\left(\frac{\partial}{\partial L}\delta\right)(k,L) = {} & \pi - 2L \int_{r'}^{\infty} dr \frac{1}{r^2} \\ & \times \left[k^2 - \frac{L^2 + \alpha - \frac{1}{4}}{r^2} - \frac{\gamma}{r} - V(r) \right]^{-1/2} \end{aligned} \tag{3.51}$$

represents the scattering angle and

$$\begin{aligned} 2\left(\frac{\partial}{\partial k}\delta\right)(k,L) = {} & 2\left(\frac{\partial}{\partial k}\delta_\alpha\right)(k,L) + \lim_{R \to \infty} \left\{ 2k \int_{r'}^{R} dr \right. \\ & \times \left[k^2 - \frac{L^2 + \alpha - \frac{1}{4}}{r^2} - \frac{\gamma}{r} - V(r) \right]^{-1/2} \\ & \left. - 2k \int_{r'_\alpha}^{R} dr \left(k^2 - \frac{L^2 + \alpha - \frac{1}{4}}{r^2} - \frac{\gamma}{r} \right)^{-1/2} \right\}. \end{aligned} \tag{3.52}$$

Proof: We can follow the arguments of Ref. 33. ■

Next, we define the classical relative time delay $\tau_{sc}(k,L)$ and the classical total time delay $\tau_a(k,L)$ to be

$$\tau_{sc}(k,L) = \lim_{R \to \infty} [t^R(k,L) - t_\alpha^R(k,L)] \tag{3.53}$$

and

$$\tau_a(k,L) = \lim_{R \to \infty} [t^R(k,L) - t_{a,c}^R(k,L)]. \tag{3.54}$$

Here $t_{a,c}^R(k,L)$ is the modified free transit time through a sphere of radius R centered at the origin, and has already been discussed in Sec. IIB. [see Eq. (2.60)]. Similarly, $t^R(k,L)$ and $t_\alpha^R(k,L)$ are the transit times through that sphere of radius R for the systems described by the scattering transformation S, respectively S_α. They are given by

$$t^R(k,L) = \int_{r'}^{R} dr \left[k^2 - \frac{L^2 + \alpha - \frac{1}{4}}{r^2} - \frac{\gamma}{r} - V(r) \right]^{-1/2}, \quad k > 0, \tag{3.55}$$

and

$$t_\alpha^R(k,L) = \int_{r'_\alpha}^{R} dr \left(k^2 - \frac{L^2 + \alpha - \frac{1}{4}}{r^2} - \frac{\gamma}{r} \right)^{-1/2}, \quad k > 0. \tag{3.56}$$

Using this and taking the derivative of Eq. (3.50) with respect to k, we immediately arrive at the analog of the Theorems 3.2 and 3.3.

Theorem 3.4: The classical relative time delay $\tau_{sc}(k,L)$ satisfies

$$\tau_{sc}(k,L) = \frac{1}{k}\left(\frac{\partial}{\partial k}\delta_{sc}\right)(k,L). \tag{3.57}$$

The classical total time delay is given by

$$\tau_a(k,L) = \frac{1}{k}\frac{\partial}{\partial k}\left[\delta(k,L) - \frac{\gamma}{2k}(1-\log a)\right]. \tag{3.58}$$

In these formulas,

$$\delta(k,L) = \delta_\alpha(k,L) + \delta_{sc}(k,L) \quad \text{and} \quad k>0. \tag{3.59}$$

IV. AN ALGEBRAIC APPROACH TO COULOMB TIME DELAY

Exploiting the well-known SO(3,1) symmetry of the Coulomb scattering problem, we present an algebraic approach to Coulomb time delay, both in quantum and classical mechanics.

A. Quantum mechanics

According to the SO(3,1) symmetry, the Runge–Lenz vector F defined by

$$\begin{aligned}4\mathbf{F} &= (\dot{\mathbf{x}}(t)\wedge\mathbf{L}) - (\mathbf{L}\wedge\dot{\mathbf{x}}(t)) + 2\gamma\mathbf{x}(t)/|\mathbf{x}(t)| \\ &= i\dot{\mathbf{x}}(t)\mathbf{L}^2 - i\mathbf{L}^2\dot{\mathbf{x}}(t) + 2\gamma\mathbf{x}(t)/|\mathbf{x}(t)|\end{aligned} \tag{4.1}$$

is conserved together with H_c and $\mathbf{L}$. [In this section, we use the Heisenberg picture, i.e., $\mathbf{x}(t) = \exp(itH_c)\mathbf{x}\exp(-itH_c)$, $\mathbf{x} = \mathbf{x}(0)$, etc.] Following Refs. 13, 14, and 22, this property enables us to investigate Coulomb time delay from an algebraic point of view.

Let $E_c = \theta(H_c)$ be the projector onto the (absolutely) continuous spectral subspace of H_c and let

$$E_c\dot{\mathbf{x}}E_c = \Omega^{\pm}_{a,c}\dot{\mathbf{x}}_{\pm}\Omega^{\pm *}_{a,c}. \tag{4.2}$$

(For simplicity, we suppress the a dependence of $\dot{\mathbf{x}}$ and $\mathbf{x}$.) Then it is shown in Ref. 12 that $E_c\dot{\mathbf{x}}(t)E_c$ converges to $\dot{\mathbf{x}}_{\pm}E_c$ as $t\to\pm\infty$ in strong resolvent sense, i.e.,

$$\begin{aligned}&\text{s-}\lim_{t\to\pm\infty}[E_c\dot{x}_j(t)E_c - z]^{-1} = (\dot{x}_{\pm j}E_c - z)^{-1}, \\ &z\in C\setminus\mathbb{R}, \quad j=1,2,3.\end{aligned} \tag{4.3}$$

Furthermore, also

$$E_c[\mathbf{x}(t)/t]E_c \underset{t\to\pm\infty}{\to} \pm\dot{\mathbf{x}}_{\pm}E_c, \tag{4.4}$$

$$E_c[\mathbf{x}(t)/|\mathbf{x}(t)|]E_c \underset{t\to\pm\infty}{\to} \pm(\dot{\mathbf{x}}_{\pm}/|\dot{\mathbf{x}}_{\pm}|)E_c \tag{4.5}$$

in strong resolvent sense.[14]

Denoting by A the generator of dilations,

$$A = \tfrac{1}{4}(\mathbf{x}\cdot\dot{\mathbf{x}} + \dot{\mathbf{x}}\cdot\mathbf{x}), \tag{4.6}$$

we introduce

$$\begin{aligned}\tilde{A}(t) &= \tilde{U}_{a,c}(t)A\tilde{U}^*_{a,c}(t) \\ &= A - \tfrac{1}{2}\dot{\mathbf{x}}^2 t + \gamma|\dot{\mathbf{x}}|\operatorname{sgn}(t)/(|\dot{\mathbf{x}}| + c/|t|)^2 \\ &\quad\times\log[(a/e^2)(|\dot{\mathbf{x}}| + c/|t|)^2|t|], \quad c>0,\end{aligned} \tag{4.7}$$

where $\tilde{U}_{a,c}(t)$ is obtained from $U_{a,c}(t)$ by changing $|\dot{\mathbf{x}}|$ into $|\dot{\mathbf{x}}| + c/|t|$. Furthermore, we define

$$D(t) = e^{itH_c}\tilde{A}(t)e^{-itH_c}. \tag{4.8}$$

The existence of $\Omega^{\pm}_{a,c}$ then immediately implies convergence of $D(t)$ in strong resolvent sense

$$\text{s-}\lim_{t\to\pm\infty}[E_cD(t)E_c - z]^{-1} = (E_cD_{\pm}E_c - z)^{-1}, \quad z\in C\setminus\mathbb{R}, \tag{4.9}$$

where

$$D_{\pm} = \Omega^{\pm}_{a,c}A\Omega^{\pm *}_{a,c}. \tag{4.10}$$

In order to establish the relation between the generator of dilations and time delay (cf. Refs. 4, 5, 25, and 39), we calculate the commutator between $D(t)$ and $\mathbf{F}$.

Let $E_c(K)$ be the spectral projector of H_c associated with the compact interval $K\subset(0,\infty)$. Since we will turn to a fixed angular momentum later on, we restrict ourselves to Hilbert space vectors that are linear combinations of finitely many angular momentum eigenfunctions. We call this set $\mathscr{D}_f(L^2)$. Both $L^2E_c(K)$ and $\mathbf{F}E_c(K)$ (see Remark 4.1) are defined on $\mathscr{D}_f(L^2)$.

Let $\mathscr{D}_t = \mathscr{D}_f(L^2)\cap\mathscr{D}[E_c(K)D(t)E_c(K)]$. As a weak commutator on $\mathscr{D}_t\times\mathscr{D}_t$ we have that

$$\begin{aligned}&4iE_c(K)[D(t),\mathbf{F}]E_c(K) \\ &\quad= E_c(K)\{(\mathbf{L}\wedge\dot{\mathbf{x}}(t)) - (\dot{\mathbf{x}}(t)\wedge\mathbf{L})\}E_c(K) \\ &\quad\quad-\gamma E_c(K)\left\{\dot{\mathbf{x}}(t)\frac{t}{|\mathbf{x}(t)|}\right. \\ &\quad\quad\left.-(\dot{\mathbf{x}}(t)\cdot\mathbf{x}(t))\frac{\mathbf{x}(t)t}{|\mathbf{x}(t)|^3} + \text{h.c.}\right\}E_c(K) \\ &\quad\quad+2i\gamma E_c(K)\left[\frac{|\dot{\mathbf{x}}|\operatorname{sgn}(t)}{[|\dot{\mathbf{x}}(t)| + c/|t|]^2}\right. \\ &\quad\quad\left.\times\log\left\{\frac{a}{e^2}\left[|\dot{\mathbf{x}}(t)| + \frac{c}{|t|}\right]^2|t|\right\},\mathbf{F}\right]E_c(K).\end{aligned} \tag{4.11}$$

Next we prove that in the limit $t\to\pm\infty$ only the first term on the rhs of Eq. (4.11) gives a nonzero contribution. To do so we rewrite, e.g., the term $-\gamma E_c(K)\dot{\mathbf{x}}(t)[t/|\mathbf{x}(t)|]E_c(K)$ as

$$\begin{aligned}&-\gamma\{E_c(K)(|H_c|+1)^2\}\{(|H_c|+1)^{-2}[\dot{\mathbf{x}}(t)^2+1]\dot{\mathbf{x}}(t)\} \\ &\quad\times\{[\dot{\mathbf{x}}(t)^2+1]^{-1}[t/|\mathbf{x}(t)|]\}E_c(K).\end{aligned} \tag{4.12}$$

From this we infer, by recalling the results (4.3)–(4.5), that the second term on the rhs of Eq. (4.11) is a sequence of bounded operators which converges strongly to zero as $t\to\pm\infty$. To treat the last term of Eq. (4.11), we first note that

$$E_c(K)\frac{|\dot{\mathbf{x}}|}{[|\dot{\mathbf{x}}(t)| + c/|t|]^2}\log\left\{\frac{a}{e^2}\left[|\dot{\mathbf{x}}(t)| + \frac{c}{|t|}\right]^2|t|\right\}E_c(K)$$

is also a bounded operator (see Remark 4.2). Using then the fact that

$$\begin{aligned}&\text{s-}\lim_{t\to\pm\infty}\left[E_c(K)\frac{t}{|\dot{\mathbf{x}}(t)|}E_c(K), E_c(K)\frac{\mathbf{x}(t)}{|\mathbf{x}(t)|}E_c(K)\right] \\ &\quad= \frac{\dot{\mathbf{x}}_{\pm}}{2|\dot{\mathbf{x}}_{\pm}|^4}E_c(K),\end{aligned} \tag{4.13}$$

we deduce that the third term on the rhs of Eq. (4.11) converges to zero in strong resolvent sense as $t\to\pm\infty$.

Let $\mathscr{D}_{\pm} = \mathscr{D}_f(L^2)\cap\mathscr{D}(E_c(K)D_{\pm}E_c(K))$. Since for self-adjoint operators strong resolvent convergence is equivalent to strong graph convergence,[40–42] there exists for any $\psi_{\pm}\in\mathscr{D}_{\pm}$ a $\psi_{\pm}(t)\in\mathscr{D}_t$ such that

$$\operatorname*{s-lim}_{t\to\pm\infty}\psi_\pm(t)=\psi_\pm \tag{4.14}$$

and

$$\operatorname*{s-lim}_{t\to\pm\infty}D(t)\psi_\pm(t)=D_\pm\psi_\pm. \tag{4.15}$$

This proves

$$\begin{aligned}&4i(E_c(K)D_\pm E_c(K)\psi_\pm,E_c(K)\mathbf{F}E_c(K)\phi_\pm)\\&\quad-4i(E_c(K)\mathbf{F}E_c(K)\psi_\pm,E_c(K)D_\pm E_c(K)\phi_\pm)\\&\quad=(E_c(K)\psi_\pm,E_c(K)[(\mathbf{L}\wedge\dot{\mathbf{x}}_\pm)\\&\quad-(\dot{\mathbf{x}}_\pm\wedge\mathbf{L})]E_c(K)\phi_\pm),\ \phi_\pm,\ \psi_\pm\in\mathscr{D}_\pm.\end{aligned} \tag{4.16}$$

Noting then that[13,14]

$$\begin{aligned}4\mathbf{F}&=i\dot{\mathbf{x}}_\pm\mathbf{L}^2-i\mathbf{L}^2\dot{\mathbf{x}}_\pm\pm2\gamma\dot{\mathbf{x}}_\pm/|\dot{\mathbf{x}}_\pm|\\&=(\mathbf{L}\wedge\dot{\mathbf{x}}_\pm)-(\dot{\mathbf{x}}_\pm\wedge\mathbf{L})\pm2\gamma\dot{\mathbf{x}}_\pm/|\dot{\mathbf{x}}_\pm|,\end{aligned} \tag{4.17}$$

we finally obtain

$$\begin{aligned}&E_c(K)[D_\pm,(\dot{\mathbf{x}}_\pm\mathbf{L}^2-\mathbf{L}^2\dot{\mathbf{x}}_\pm\mp2i\gamma\dot{\mathbf{x}}_\pm/|\dot{\mathbf{x}}_\pm|)]E_c(K)\\&\quad=i(\dot{\mathbf{x}}_\pm\mathbf{L}^2-\mathbf{L}^2\dot{\mathbf{x}}_\pm)E_c(K),\end{aligned} \tag{4.18}$$

as a weak commutator on $\mathscr{D}_+$.

Next, recalling Eq. (4.10) together with the relation

$$e^{isA}e^{itH_0}e^{-isA}=e^{it\exp(-2s)H_0},\quad s,t\in\mathbb{R}, \tag{4.19}$$

we deduce that

$$e^{isD_\pm}e^{itH_c}e^{-isD_\pm}=e^{it\exp(-2s)H_c},\quad s,t\in\mathbb{R}. \tag{4.20}$$

From this we infer that $D_- - D_+$ commutes with H_c.

Since S_c and $D_\pm$ do not depend on a, we define the a-independent part $\hat{\tau}_c$ of the Coulomb time delay operator $\tau_{a,c}$ by

$$\hat{\tau}_c=(1/2H_c)(D_- - D_+)E_c. \tag{4.21}$$

In the spectral representation of H_c E_c, $\hat{\tau}_c$ is represented by $\hat{\tau}_c(k)$ and

$$\hat{\tau}_c(k)=\bigoplus_{l=0}^{\infty}\hat{\tau}_{c,l}(k). \tag{4.22}$$

More explicitly, let U_c be the following isomorphism:

$$U_c:\begin{cases}E_cL^2(\mathbb{R}^3)\to L^2((0,\infty);L^2(S^2))\\h(x)\to(U_ch)(k,\omega)=k\hat{h}_c(k\omega),\quad k>0,\ \omega\in S^2,\end{cases} \tag{4.23}$$

where

$$\hat{h}_c(k\omega)=\operatorname*{s-lim}_{R\to\infty}\int_{|\mathbf{x}|<R}d^3x\ \overline{\Psi_c^-(k\omega,\mathbf{x})}h(\mathbf{x}), \tag{4.24}$$

then the operator $\hat{\tau}_c$ satisfies

$$(U_c\hat{\tau}_cU_c^{-1}g)(k,\omega)=(\hat{\tau}_c(k)g)(k,\omega),\quad g\in\mathscr{D}(\hat{\tau}_c(k)) \tag{4.25}$$

Similarly, since $\dot{\mathbf{x}}_\pm$ commute with H_c, we have

$$\begin{aligned}&(U_c\dot{\mathbf{x}}_\pm U_c^{-1}g)(k,\omega)=(\dot{\mathbf{x}}_\pm(k)g)(k,\omega),\quad g\in\mathscr{D}(\dot{\mathbf{x}}_\pm(k)),\\&\tfrac{1}{4}[\dot{\mathbf{x}}_\pm(k)]^2=k^2\mathbf{1}.\end{aligned} \tag{4.26}$$

Finally, employing the relation[13,14]

$$\begin{aligned}&\int_0^\infty dk\int_{S^2}d\omega((U_c\,g_{l+1})(k,\omega),(\dot{x}_{+3}(k)U_c\,g_l)(k,\omega))\\&\quad=\int_0^\infty dk\int_{S^2}d\omega\,\frac{l+1-i\gamma/2k}{l+1+i\gamma/2k}\\&\qquad\times((U_c\,g_{l+1})(k,\omega),(\dot{x}_{-3}(k)U_c\,g_l)(k,\omega))\quad l\in\mathbb{N},\end{aligned} \tag{4.27}$$

where $g_l\in\mathscr{D}(\dot{x}_{\pm3})$ and $g_l(k,\omega)=g_{l,m}(k)Y_{l,m}(\omega)$, we obtain from Eq. (4.18) and (4.22)

$$\begin{aligned}\hat{\tau}_{c,l+1}(k)-\hat{\tau}_{c,l}(k)&=-\frac{\gamma(l+1)}{2k^3[(l+1)^2+\gamma^2/4k^2]}\\&=\frac{1}{k}\frac{\partial}{\partial k}[\delta_{l+1}^c(k)-\delta_l^c(k)].\end{aligned} \tag{4.28}$$

Remark 4.1: $F_jE_c(K)$ is defined on $\mathscr{D}_f(L^2)$. To prove this, we first note that $\dot{\mathbf{x}}(t)E_c(K)$ is bounded since

$$\begin{aligned}\|\,|\dot{\mathbf{x}}(t)|E_c(K)\|&=\|\,|\dot{\mathbf{x}}(t)|(|H_c|+1)^{-1}\\&\quad\times(|H_c|+1)E_c(K)\|<\infty.\end{aligned}$$

From definition (4.1) we deduce that there exist constants $c_i\in\mathbb{R}^+$ such that for $\psi\in\mathscr{D}_f(L^2)$ we have

$$\|F_jE_c(K)\psi\|\leqslant c_1\|L^2\dot{x}_jE_c(K)\psi\|+c_2\|\dot{x}_jE_c(K)\psi\|+c_3\|\psi\|.$$

Now, since $\dot{x}_j$ changes the angular momentum at most by 1, we have

$$\|L^2\dot{x}_jE_c(K)\psi\|\leqslant c_4\|\dot{x}_jE_c(K)\psi\|,$$

for some constant $c_4\in\mathbb{R}^+$. Therefore, we obtain

$$\|F_jE_c(K)\psi\|<\infty,\quad\psi\in\mathscr{D}_f(L^2).$$

Remark 4.2: The operator $E_c(K)|\dot{\mathbf{x}}(t)|^2E_c(K)$ is bounded. Indeed, if $\gamma>0$, we have

$$\begin{aligned}&\tfrac{1}{4}E_c(K)|\dot{x}(t)|^2E_c(K)\\&\quad=H_cE_c(K)-[\gamma E_c(K)/|\mathbf{x}(t)|]E_c(K)\\&\quad\leqslant\sup(K)\cdot E_c(K).\end{aligned}$$

Therefore, $E_c(K)|\dot{\mathbf{x}}(t)|^2E_c(K)$ is bounded.

If $\gamma<0$, we have

$$\begin{aligned}&\tfrac{1}{4}E_c(K)|\dot{\mathbf{x}}(t)|^2E_c(K)\\&\quad=H_cE_c(K)+[|\gamma|E_c(K)/|\mathbf{x}(t)|]E_c(K).\end{aligned}$$

Furthermore, in the sense of forms we can write

$$\begin{aligned}&[E_c(K)/|\mathbf{x}(t)|]E_c(K)\leqslant dH_cE_c(K)\\&\quad+[\gamma^2(1-d)^2/d]E_c(K),\quad0<d<1.\end{aligned}$$

This follows from the scaling behavior of the Coulomb ground state energy. Therefore, there exist values of d such that

$$\begin{aligned}&\tfrac{1}{4}E_c(K)|\dot{\mathbf{x}}(t)|^2E_c(K)\leqslant\{(1+d)H_c\\&\quad+\gamma^2(1-d)^2/d\}E_c(K)\leqslant\text{const.}\,E_c(K)\end{aligned}$$

is bounded.

B. Classical mechanics

With the help of asymptotic completeness, i.e., the existence of $(\mathbf{x},\dot{\mathbf{x}})\in\Gamma$ and $(\mathbf{x}_+,\dot{\mathbf{x}}_+)\in\Gamma$ such that for given $(\mathbf{x}_-,\dot{\mathbf{x}}_-)\in\Gamma$

$$\lim_{t\to\pm\infty}|\Phi_c^t(\mathbf{x},\dot{\mathbf{x}})-\Phi_{a,c}^t(\mathbf{x}_\pm,\dot{\mathbf{x}}_\pm)|=0 \tag{4.29}$$

(we again suppress the a-dependence of $\mathbf{x},\dot{\mathbf{x}}$), we infer the existence of the following limits:

$$\lim_{t\to\pm\infty}|\mathbf{x}(t)-\dot{\mathbf{x}}(t)t+[2\dot{\mathbf{x}}(t)\gamma\,\mathrm{sgn}(t)/|\dot{\mathbf{x}}(t)|^3]\times\log[(a/e^2)|\dot{\mathbf{x}}(t)|^2|t|]|=\mathbf{x}_\pm \tag{4.30}$$

and

$$\lim_{t\to\pm\infty}\dot{\mathbf{x}}(t)=\dot{\mathbf{x}}_\pm,\quad\lim_{t\to\pm\infty}[\mathbf{x}(t)/|\mathbf{x}(t)|]=\pm\dot{\mathbf{x}}_\pm/|\dot{\mathbf{x}}_\pm|. \tag{4.31}$$

In particular, Eq. (4.31) immediately implies the following representation of the Runge–Lenz vector F:

$$\begin{aligned}2\mathbf{F}&=(\dot{\mathbf{x}}(t)\wedge\mathbf{L})+\gamma\mathbf{x}(t)/|\mathbf{x}(t)|\\&=(\dot{\mathbf{x}}_\pm\wedge\mathbf{L})\mp\gamma\dot{\mathbf{x}}_\pm/|\dot{\mathbf{x}}_\pm|.\end{aligned} \tag{4.32}$$

Solving for $\dot{\mathbf{x}}_+$ as a function of $\dot{\mathbf{x}}_-$ gives[43]

$$\dot{\mathbf{x}}_+=\frac{L^2-\gamma^2/4k^2}{L^2+\gamma^2/4k^2}\dot{\mathbf{x}}_-+\frac{\gamma/k}{L^2+\gamma^2/4k^2}\times(\dot{\mathbf{x}}_-\wedge\mathbf{L}),\quad k=\tfrac{1}{2}|\dot{\mathbf{x}}_\pm|. \tag{4.33}$$

Consequently, using Eq. (2.55), we deduce that the Coulomb scattering angle $2[(\partial/\partial L)\delta_c](k,L)$ reads

$$\begin{aligned}2\left(\frac{\partial}{\partial L}\delta_c\right)(k,L)&=\mathrm{sgn}(\gamma)\arccos[(\dot{\mathbf{x}}_+,\dot{\mathbf{x}}_-)/|\dot{\mathbf{x}}_\pm|^2]\\&=\mathrm{sgn}(\gamma)\arccos[(L^2-\gamma^2/4k^2)/\\&\quad\times(L^2+\gamma^2/4k^2)]\\&=2\,\mathrm{sgn}(\gamma)\,\mathrm{arccot}(2kL/|\gamma|).\end{aligned} \tag{4.34}$$

(We note again that δ_c has been derived in Ref. 33.)

We now define

$$\hat{\tau}_c=(\mathbf{x}_-\cdot\dot{\mathbf{x}}_-)/2|\dot{\mathbf{x}}_-|^2-(\mathbf{x}_+\cdot\dot{\mathbf{x}}_+)/2|\dot{\mathbf{x}}_+|^2. \tag{4.35}$$

Because the Coulomb scattering transformation and its generator are independent of a, it is clear that $\hat{\tau}_c$ only gives the a-independent part of Coulomb time delay. From Eq. (2.55) we immediately obtain the following result:

$$\hat{\tau}_c(k,L)=\frac{1}{k}\left(\frac{\partial}{\partial k}\delta_c\right)(k,L). \tag{4.36}$$

Next, we introduce

$$D(t)=\tfrac{1}{2}\mathbf{x}(t)\cdot\dot{\mathbf{x}}(t)-\tfrac{1}{2}\dot{\mathbf{x}}(t)^2t+[\gamma\,\mathrm{sgn}(t)/2|\dot{\mathbf{x}}(t)|]\log[(a/e^2)|\dot{\mathbf{x}}(t)|^2|t|]. \tag{4.37}$$

We then exploit the conservation law for the Runge–Lenz vector. Observing that $\mathbf{x}\cdot\dot{\mathbf{x}}$ generates dilations, we first arrive at

$$[\mathbf{x}(t)\cdot\dot{\mathbf{x}}(t),\mathbf{F}]=(\dot{\mathbf{x}}(t)\wedge\mathbf{L}). \tag{4.38}$$

Furthermore, since

$$\lim_{t\to\pm\infty}[\dot{\mathbf{x}}(t)^2t,\mathbf{F}]=\lim_{t\to\pm\infty}2\gamma\{(\dot{\mathbf{x}}(t)\cdot\dot{\mathbf{x}}(t))t\mathbf{x}(t)/|\mathbf{x}(t)|^3-t\dot{\mathbf{x}}(t)/|\mathbf{x}(t)|\}=0 \tag{4.39}$$

and also

$$\lim_{t\to\pm\infty}[\ln|t|/|\dot{\mathbf{x}}(t)|,\mathbf{x}(t)/|\mathbf{x}(t)|]=\lim_{t\to\pm\infty}(\ln|t|/2t)[1/|\dot{\mathbf{x}}(t)|,\mathbf{x}(t)/|\dot{\mathbf{x}}(t)|]=\lim_{t\to\pm\infty}(\ln|t|/t)\dot{\mathbf{x}}_\pm/|\dot{\mathbf{x}}_\pm|^4=0, \tag{4.40}$$

we get on the one hand

$$[|\dot{\mathbf{x}}_\pm|^2\hat{\tau}_c,\mathbf{F}]=\lim_{t\to\pm\infty}[D(-t)-D(t),\mathbf{F}]=((\dot{\mathbf{x}}_--\dot{\mathbf{x}}_+)\wedge\mathbf{L}). \tag{4.41}$$

On the other hand, we know that $\partial F_3/\partial d=0$ and that δ_c and hence $\hat{\tau}_c$ can only depend on k and L. Consequently, we obtain

$$\begin{aligned}[|\dot{\mathbf{x}}_\pm|^2\hat{\tau}_c,F_3]&=-\frac{\partial F_3}{\partial\theta}\frac{\partial(|\dot{\mathbf{x}}_\pm|^2\hat{\tau}_c)}{\partial L}\\&=-[F_3,L]\frac{\partial(|\dot{\mathbf{x}}_\pm|^2\hat{\tau}_c)}{\partial L}\\&=-\frac{1}{L}(\mathbf{L}\wedge\mathbf{F})_3\frac{\partial(|\dot{\mathbf{x}}_\pm|^2\hat{\tau}_c)}{\partial L}.\end{aligned} \tag{4.42}$$

Putting all this together, we conclude

$$\left(\frac{\partial}{\partial L}\hat{\tau}_c\right)(k,L)=-\frac{\gamma L}{2k^3(L^2+\gamma^2/4k^2)}. \tag{4.43}$$

Equation (4.43) is again remarkably similar to the corresponding quantum mechanical result (4.28). We end with the obvious remark that if one is just interested in the result (4.43) and not in an algebraic derivation of it, one can, of course, immediately find it by differentiating Eq. (4.34) with respect to k.

ACKNOWLEDGMENTS

The present work was started while the authors were visiting the L. P. T. H. E. at Orsay. We would like to thank K. Chadan for his warm hospitality and for encouraging discussions. We are also indebted to W. Amrein, Ph. Martin, and H. Zingl for stimulating discussions on various aspects of this work. Two of us (D. B. and F. G.) gratefully acknowledge the kind hospitality extended to them by the University of Graz, respectively, the University of Bielefeld and the University of Leuven. They also wish to thank the C. N. R. S., France, respectively the Alexander von Humboldt Foundation, for financial support during the early stages of this work.

[1]Ph. A. Martin, Acta Phys. Aust. Suppl. **23**, 157 (1981).
[2]P. D. Lax and R. S. Philips, in *Topics in Functional Analysis*, Advances in Mathematics, Supplementary Studies Vol. 3 (Academic, New York, 1978), p. 197.
[3]W. O. Amrein and M. Wollenberg, Proc. Roy. Soc. Edinburgh A **87**, 219 (1981).
[4]A. Jensen, Commun. Math. Phys. **82**, 435 (1981).
[5]A. Jensen, "On Lavine's formula for time delay," Preprint CPT, CNRS Marseille, February 1982.
[6]K. Gustafson and K. Sinha, Lett. Math. Phys. **4**, 381 (1980).
[7]Ph. A. Martin, Commun. Math. Phys. **47**, 221 (1976).
[8]W. O. Amrein, J. M. Jauch, and K. B. Sinha, *Scattering Theory in Quantum Mechanics* (Benjamin, Reading, Mass., 1977).

[9]D. Bollé and T. A. Osborn, J. Math. Phys. **20**, 1121 (1979).
[10]F. T. Smith, Phys. Rev. **118**, 349 (1960).
[11]D. Bollé and T. A. Osborn, Phys. Rev. D **11**, 3417 (1975).
[12]R. B. Lavine, J. Funct. Anal. **5**, 368 (1970).
[13]H. Grosse, H. Narnhofer, and W. Thirring, in *Proceedings of the Adriatic Summer Meeting in Particle Physics.* (Rovinj), edited by M. Martinis, S. Pallua, and N. Zovko (North-Holland, Amsterdam, 1973), p. 22.
[14]H. Grosse, H. R. Grümm, H. Narnhofer, and W. Thirring, Acta Phys. Aust. **40**, 97 (1974).
[15]J. D. Dollard, J. Math. Phys. **5**, 729 (1964).
[16]J. D. Dollard, Rocky Mountain J. Math. **1**, 5 (1971).
[17]F. Gesztesy, W. Plessas, and B. Thaller, J. Phys. A **13**, 2659 (1980).
[18]F. Gesztesy and B. Thaller, J. Phys. A **14**, 639 (1981).
[19]F. Gesztesy, J. Math. Phys. **23**, 74 (1982).
[20]A. G. Gibson and C. Chandler, J. Math. Phys. **15**, 1366 (1974).
[21]I. W. Herbst, Commun. Math. Phys. **35**, 181 (1974).
[22]W. Thirring, *A Course in Mathematical Physics, Vol. 3: Quantum Mechanics of Atoms and Molecules* (Springer-Verlag, Vienna and New York, 1981).
[23]M. Reed and B. Simon, *Methods of Modern Mathematical Physics, Vol. III: Scattering Theory* (Academic, New York, 1979).
[24]R. Lavine, "Constructive Estimates in Scattering Theory", preprint, University of Rochester, 1973.
[25]R. Lavine, in *Scattering Theory in Mathematical Physics*, edited by J. A. LaVita and J. P. Marchand (D. Reidel, Dordrecht, Holland, 1974), p. 141.
[26]R. Lavine, in *Atomic Scattering Theory*, edited by J. Nuttall (Univ. of Western Ontario, London, Canada 1978), p. 45.
[27]I. M. Gel'fand and G. E. Shilov, *Generalized Functions* (Academic, New York, 1964), Vol. I.
[28]I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products* (Academic, New York, 1965).
[29]J. M. Jauch, K. B. Sinha, and B. N. Misra, Helv. Phys. Acta **45**, 398 (1972).
[30]Ph. Martin and B. Misra, in *Scattering Theory in Mathematical Physics*, edited by J. A. LaVita and J. P. Marchand (D. Reidel, Dordrecht, Holland, 1974), p. 173.
[31]L. W. McMillan and T. A. Osborn, Ann. Phys. (N. Y.) **126**, 1 (1980).
[32]M. Abramovitz and I. A. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1972).
[33]H. Narnhofer and W. Thirring, Phys. Rev. A **23**, 1688 (1981).
[34]I. W. Herbst, Commun. Math. Phys. **35**, 193 (1974).
[35]D. Bollé, Acta Phys. Aust. Suppl. **23**, 587 (1981).
[36]D. Bollé and J. D'Hondt, J. Phys. A **14**, 1663 (1981).
[37]T. Said, C. R. Acad. Sci. Paris B **281**, 521 (1975).
[38]K. Sinha, Rep. Math. Phys. **14**, 65 (1978).
[39]H. Narnhofer, Phys. Rev. D **22**, 2387 (1980).
[40]J. Glimm and A. Jaffe, Commun. Pure Appl. Math. **22**, 401 (1969).
[41]M. Reed and B. Simon, *Methods of Modern Mathematical Physics, Vol. 1: Functional Analysis* (Academic, New York, 1972).
[42]R. Wüst, Math. Z. **131**, 339 (1973).
[43]W. Thirring, *A Course in Mathematical Physics, Vol. 1: Classical Dynamical Systems* (Springer-Verlag, Vienna and New York, 1978).

On the optical approximation in two-channel systems

P. Exner[a)]
Laboratory of Theoretical Physics, JINR, 141980 Dubna, USSR

I. Úlehla
Nuclear Centre, Charles University, Areál Troja, Povltavská ul., 18000 Prague, Czechoslovakia

(Received 14 October 1981; accepted for publication 27 August 1982)

The optical approximation for two-channel quantum mechanical systems is treated within Kato theory of smooth perturbations. Concerning the difference between the optical and exact scattering matrices, the theorem deduced recently by Davies is generalized: essentially the same assertion is proved without unphysical neglection of the "diagonal" interaction terms.

PACS numbers: 03.65.Nk, 24.10.Ht

1. INTRODUCTION

Complexity of many scattering problems in nuclear and particle physics often hinders us from obtaining exact solutions to them. In such a situation, one usually selects a subspace of the whole state Hilbert space from which the needed quantities can be extracted, while influence of the rest is taken into account in some phenomenological way. In this paper, we shall treat a mathematical model for scattering in a two-channel system,[1] i.e., such that its state Hilbert space $\mathscr{H}$ can be expressed as an orthogonal sum $\mathscr{H}_0 \oplus \mathscr{H}_1$. Motivation of the problem comes from the optical model of neutron scattering on nuclei,[2–4] in the case of which the subspace $\mathscr{H}_0$ would refer to the direct scattering of the neutron; we borrow from this model the phenomenological method of expressing the second-channel influence.[3] This study represents a direct continuation of a recent work by Davies.[5]

The model under consideration is described by its Hamiltonian, which is assumed to be $H = H_0 + H_1 + \lambda V$ with $P_i H_i \subset H_i P_i$ and $P_i H_j = H_j P_i = 0$ for $i \neq j$, where P_i are projections on $\mathscr{H}_i$, $i = 1,2$, and λ is the coupling constant. For the sake of simplicity we shall use the symbol H_i for $H_i \restriction \mathscr{H}_i$ too. We denote $V_{ij} = P_i V P_j$, further it is useful to introduce $h_1 = H_1 + \lambda V_{11}$ and $v = V - V_{11}$, then the considered Hamiltonian can be written as

$$H = H_0 + h_1 + \lambda v \tag{1.1}$$

with $v_{11} = 0$. We adopt the following assumptions:

(S) the operators H_0, h_1 are self-adjoint, reduced by P_i and their parts in $\mathscr{H}_1$, $\mathscr{H}_0$, respectively, are zero.

(B) the operator v is Hermitian, i.e., bounded and self-adjoint.

Further we shall assume H_0 to be the free Hamiltonian of the first ($i = 0$) channel, then the first-channel-part of the S-matrix is

$$S_{00} = \text{s-lim}_{t\to\infty} S_{00}^t, \quad S_{00}^t = P_0 e^{iH_0 t} e^{-2iHt} e^{iH_0 t} P_0. \tag{1.2}$$

The *optical approximation* consists of choosing a suitable $V_{\text{opt}} \in \mathscr{B}(\mathscr{H}_0)$ such that

$$S_{\text{opt}} = \text{s-lim}_{t\to\infty} S_{\text{opt}}^t, \quad S_{\text{opt}}^t = e^{iH_0 t} e^{-2iH_{\text{opt}} t} e^{iH_0 t} \tag{1.3}$$

with $H_{\text{opt}} = H_0 + V_{\text{opt}}$ is in some sense near to S_{00}. In what follows we consider the *Feshbach-type optical potential*[3]:

$$V_{\text{opt}}(E) = \lambda v_{00} - \lambda^2 \lim_{\epsilon\to 0+} v_{01}(h_1 - E - i\epsilon)^{-1} v_{10}. \tag{1.4}$$

Recently Davies has found[5] a class of Hamiltonians (1.1) for which S_{opt} corresponding to the potential (1.4) approximates strongly S_{00}. His argument is based on Kato theory of smooth perturbations.[6] However, the latter can be applied only if the spectrum of h_1 is continuous; thus one has to exhibit conditions under which S is approximated weakly by S' referring to some h_1' with this property. Assume now the situation when a neutron striking the nucleus can be either elastically scattered or form a compound nucleus, which is relatively long-lived so that there are two natural time scales appropriate to the system. The results of Ref. 5 can be then interpreted physically as follows: the optical approximation is justified by the existence of a short-time scale (comparing to the other two) which characterizes formation of the compound nucleus; in addition, the eigenvalues of h_1 should be numerous and close together.

However, in order to simplify the deductions, Davies assumes $E = 0$ and $v_{00} = 0$. Especially the last assumption is seriously limiting when realistic physical situations are considered. As an example, consider the simplest optical-model situation[3a,7] when

$$\mathscr{H} = L^2(\mathbb{R}^3) \otimes L^2(\mathbb{R}^{3A}), \tag{1.5}$$

$$H = -\Delta_n + h_N + \lambda v(x_n, x_N). \tag{1.6}$$

Let h_N possess an eigenvector $u_0 \in L^2(\mathbb{R}^{3A})$ referring to the eigenvalue ϵ_0 (ground state of the target nucleus) and choose

$$\mathscr{H}_0 = L^2(\mathbb{R}^3) \otimes \{u_0\}_{\text{lin}}, \quad \text{i.e.,} \quad P_0 = I_n \otimes P_{u_0}, \tag{1.7}$$

then

$$H = (-\Delta_n + \epsilon_0)P_0 + (-\Delta_n + h_N)P_1 + \lambda \sum_{i,j=0}^{1} v_{ij} \tag{1.8}$$

with

$$(v_{00}\psi)(x_n, x_N) = u_0(x_N)(u_0, v(x_n, .)u_0)_N (u_0, \psi(x_n, .))_N, \tag{1.9}$$

$(.\,,.)_N$ being the inner product in $L^2(\mathbb{R}^{3A})$; thus v_{00} is nonzero for reasonable potentials v.

The aim of the present paper is to show that the results of Davies are essentially preserved if one gives up the mentioned assumptions.[8] This assertion is formulated and proved in Secs. 5–8: the method is based again on the smooth-perturbation theory but technically is more complicated. Before doing that, we discuss briefly two related problems: in Sec. 2 we show how the potential (1.4) can be derived

[a)] On leave of absence from Nuclear Centre, Charles University, Prague, Czechoslovakia.

 0022-2488/83/061542-06$02.50 © 1983 American Institute of Physics

in a time-dependent way, while Sec. 3 is devoted to the approximation of h_1 by h_1' with continuous spectrum. These considerations yield antagonistic requirements on the initial wavepackets which will be mentioned in concluding remarks.

2. THE OPTICAL POTENTIAL FOR THE TWO-CHANNEL SYSTEM

Starting from the simplest case referring to (1.5)–(1.9), Feshbach derived the expression (1.4) by a formal time-independent way.[3a] This procedure applies to more general situations too, when the selected subspace in the "target" state space is multidimensional etc.[3,7] There are also other time-independent derivations of the optical potential.[4]

In the present section, we shall show how the optical potential (1.4) can be formally obtained within the time-dependent approach. We shall start from the relations

$$e^{-i(A+B)t} = e^{-iAt} - i\int_0^t e^{-i(A+B)s}Be^{-iA(t-s)}ds$$
$$= e^{-iAt} - i\int_0^t e^{-iA(t-s)}Be^{-i(A+B)s}ds, \quad (2.1)$$

which are valid (rigorously) for any self-adjoint A and Hermitian B. We apply these relations repeatedly to expand S_{00}^t and $S_{\rm opt}^t$, further we pass to the limit $t\to\infty$. Comparing now the resulting expansions of S_{00} and $S_{\rm opt}$ up to the second order terms in v and $V_{\rm opt}$, respectively, we obtain the following two candidates to the role of the optical potential:

$$V_{\rm opt}^{(+)} = \lambda v_{00} - i\lambda^2\int_0^\infty e^{iH_0t}v_{01}e^{-ih_1t}v_{10}dt, \quad (2.2a)$$

$$V_{\rm opt}^{(-)} = \lambda v_{00} - i\lambda^2\int_0^\infty v_{01}e^{-ih_1t}v_{10}e^{iH_0t}dt. \quad (2.2b)$$

Up to now we have followed Davies[5]; he rightly points out that besides the ambiguity in choosing the right one, the potentials (2.2) have a more serious defect that neither of them is dissipative: counterexamples can be found even for rank-one operators v_{01}.

Nevertheless, (2.2) can suggest the correct choice in the following way: suppose that the initial state of the system (for t large negative) is a wavepacket ϕ_E with narrow energy distribution concentrated around the mean E. It lies entirely in $\mathscr{H}_0$, and therefore one can replace $e^{iH_0t}\phi_E$ approximately by $e^{iEt}\phi_E$ obtaining in this way

$$V_{\rm opt}\phi_E = \lambda v_{00}\phi_E - i\lambda^2\int_0^\infty dt\, v_{01}e^{-ih_1t}v_{10}e^{iEt}\phi_E$$
$$= \lambda v_{00}\phi_E - i\lambda^2\lim_{\epsilon\to0^+}\int_0^\infty dt\, v_{01}e^{-i(h_1-E-i\epsilon)t}v_{10}\phi_E$$
$$= \lambda v_{00}\phi_E - \lambda^2\lim_{\epsilon\to0^+} v_{01}(h_1-E-i\epsilon)^{-1}v_{10}\phi_E.$$

Thus we have arrived just to Feshbach energy-dependent optical potential (1.4), which can be easily seen to obey the dissipativity condition.

3. APPROXIMATION OF h_1

As pointed out in the introduction, application of Kato theory in the following sections is conditioned by the fact that h_1 has a purely continuous spectrum. Usually it is not so [cf. e.g. (1.8)], thus one has first to replace h_1 by h_1' with this property, which is near to h_1, say, in the operator norm. The question is, whether the related S-matrices are in some sense near on $\mathscr{H}_0$. It can be answered positively if we restrict ourselves to a certain subset of states: let $\mathscr{L}$ be a dense subspace in $\mathscr{H}_0$ endowed by a norm $|\,.\,|$ such that

$$\int_{\mathbb{R}} \|ve^{-iH_0t}\phi\|(1+|t|^\alpha)dt \leqslant |\phi|, \quad \phi\in\mathscr{L} \quad (3.1)$$

for some α, $0<\alpha\leqslant1$. The following assertion is valid:

Proposition 1: Assume (S), (B) and $\|h_1 - h_1'\| < \epsilon$, then for each $\phi, \psi\in\mathscr{L}$ and $t\geqslant0$ we have

$$|(\phi,S_t\psi) - (\phi,S_t'\psi)| \leqslant 2\lambda^2\epsilon^\alpha|\phi|\,|\psi|, \quad (3.2)$$

$$\|S_t\phi - S_t'\phi\| \leqslant 2\lambda\epsilon^\alpha(1+t^\alpha)|\phi|. \quad (3.3)$$

The proof is essentially the same as in Ref. 5. The only difference caused by nonzero v_{00} is the presence of terms linear in λ in the expansion of $(\phi,S_t\psi)$ and $(\phi,S_t'\psi)$. However, these terms do not depend on h_1, h_1' so they cancel in subtraction. Derivation of (3.3) does not employ the assumption $v_{00}=0$.

We see that S' approximates S on $\mathscr{L}$ in the weak operator topology. Strong approximation is also possible but in finite time intervals only; no essentially stronger inequalities can be expected to hold as discussed in Ref. 5.

4. SOME NOTATIONS

$v_{10} = C_1^*C_0$, $v_{01} = C_0^*C_1$, $v_{00} = B^*A$, where A, B, $C_i\in\mathscr{B}(\mathscr{H})$, $C_1 = P_0C_1P_1$ and the operators A, B, C_0 coincide with their parts in $\mathscr{H}_0$; moreover, $A = \sqrt{|v_{00}|}$, $B = AW^*$, where W: $\overline{\mathrm{Ran}v_{00}}\to\overline{\mathrm{Ran}v_{00}}$ is a partial isometry,
$\phi_C^t(s) = \chi_{[-t,t]}(s)C_0\exp(-iH_0s)\phi$, $\phi_B^t(s) = \chi_{[-t,t]}(s)B\times\exp(-iH_0t)\phi$, and similarly $\phi_A^t(s)$ is defined for each $\phi\in\mathscr{H}_0$, where χ_M is characteristic function of the set M,

$$x_{11}(t) = C_0e^{-iH_0t}C_0^*\Theta(t), \quad x_{10}(t) = C_0e^{-iH_0t}B^*\Theta(t),$$
$$x_{01}(t) = Ae^{-iH_0t}C_0^*\Theta(t), \quad x_{00}(t) = Ae^{-iH_0t}B^*\Theta(t)$$

and

$$z_{11}(t) = C_1e^{-ih_1t}C_1^*\Theta(t),$$

where Θ is Heaviside function,
$\|\cdot\|$, $(.,.)$ Hilbert-space norm and the corresponding inner product in $\mathscr{H}$, $\mathscr{H}_0$
$\|\cdot\|_2$, $\langle .,.\rangle$ Hilbert-space norm and the corresponding inner product in $L^2(\mathbb{R};\mathscr{H}_0)$,
$\|\cdot\|_u$ operator norm in $\mathscr{B}(\mathscr{H}_0)$,
$\|\cdot\|_{uf}$ operator norm in $\mathscr{B}(L^2(\mathbb{R};\mathscr{H}_0))$,
$\|f\|_1 = \int_{\mathbb{R}}\|f(x)\|_u dx$ and $\|f\|_\infty = \sup\,\mathrm{ess}\,\{\|f(x)\|_u : x\in\mathbb{R}\}$ for a measurable operator-valued function $f:\mathbb{R}\to\mathscr{B}(\mathscr{H}_0)$,
$\hat f(.)$ is Fourier transform of f: $\hat f(x) = \int_{\mathbb{R}} e^{ixy}f(y)dy$,
F is the operator of "left multiplication" by $f:\mathbb{R}\to\mathscr{B}(\mathscr{H}_0)$ on $L^2(\mathbb{R};\mathscr{H}_0)$, $(F\psi)(x) = f(x)\psi(x)$, and similarly for $\hat F$; clearly $\|F\|_{uf} = \|f\|_\infty$,
$\tilde F$ is convolution of F, i.e., the operator on $L^2(\mathbb{R};\mathscr{H}_0)$ defined by $(\tilde F\psi)(x) = \int_{\mathbb{R}} f(y)\psi(x-y)dy$.

5. THE MAIN RESULT

The central trick of Kato theory consists of factorizing the perturbation into product of a pair of operators which

are smooth w.r.t. the unperturbed Hamiltonian. In the considered case, we choose suitable factorization using notation from the previous section and assume the quantities $\|x_{ij}\|_1$, $i, j = 0,1$, and $\|z_{11}\|_1$ to be *finite*. Since the operators C_0, C_1 are given up to a multiplicative constant, $C_1^* C_0 = (\alpha C_1)^*(\alpha^{-1} C_0)$ with a positive α, and similarly for λv, the last assumption can be reformulated without loss of generality as follows[9]

$$\|x_{00}\|_1 = 1, \quad \max\{\|x_{10}\|_1, \|x_{01}\|_1, \|x_{11}\|_1\} = 1, \tag{N}$$

$$\|z_{11}\|_1 \equiv \eta < \infty.$$

With these prerequisites, we can formulate the above mentioned assertion which shall be proved in the next three sections:

Theorem: Assume (S), (B), and (N), then for $|\lambda| < (2\sqrt{\eta_0})^{-1}$, $\eta_0 = \max\{1,\eta\}$ and an arbitrary $\psi \in \mathscr{H}_0$ the estimate

$$\|S_{00}^t \psi - S_{\text{opt}}^t(E)\psi\| \leqslant \lambda^2 \{\alpha(E) + \sqrt{2}(\beta(E) + 2\gamma(E))\|\psi\|\}(\sqrt{2}(1 - 2|\lambda|\sqrt{\eta_0}))^{-1} \tag{5.1}$$

holds, where $S_{\text{opt}}^t(E)$ refer to the optical potential (1.4) and

$$\alpha(E) = \|(\tilde{Z}_{11} - \hat{Z}_{11}(E))\psi_C^t\|_2, \tag{5.2a}$$

$$\beta(E) = \|(\tilde{Z}_{11} - \hat{Z}_{11}(E))X_{11}\|_{uf}, \tag{5.2b}$$

$$\gamma(E) = \|(\tilde{Z}_{11} - \hat{Z}_{11}(E))X_{10}\|_{uf}. \tag{5.2c}$$

6. THE BASIC ESTIMATES

Proposition 2: If (N) is valid, then the vectors ϕ_C^t, ϕ_A^t, ϕ_B^t belong to $L^2(\mathbb{R};\mathscr{H}_0)$ and

$$\|\phi_K^t\|_2^2 \leqslant 2\|\phi\|^2, \quad K = C,A,B \tag{6.1}$$

for each $t > 0$ and all $\phi \in \mathscr{H}_0$.

Proof: Consider first ϕ_C^t, we have

$$\|\phi_C^t\|_2^2 \leqslant \int_{\mathbb{R}} \|C_0 e^{-iH_0 s}\phi\|^2 ds$$

and according to Ref. 6, the rhs is $\leqslant 2\pi\|C_0\|_{H_0}^2\|\phi\|^2$; thus one has to show that C_0 is H_0-smooth, $\|C_0\|_{H_0} < \infty$. Kato derived six equivalent expressions of the norm $\|.\|_{H_0}$, among them

$$\|C_0\|_{H_0}^2 = a_3 = \frac{1}{2\pi}\sup_{z,\psi} |((R(z) - R(\bar{z}))C_0^*\psi, C_0^*\psi)|$$

(Ref. 6, Theorem 5.1), where $R(z) = (H_0 - z)^{-1}$, $z \in \mathbb{C}\setminus\mathbb{R}$ and $\psi \in D(C_0^*)$, $\|\psi\| = 1$. Using

$$R(z) - R(\bar{z}) = i\int_{\mathbb{R}} \exp(i\xi s - \eta|s| - iH_0 s)ds,$$

where $z = \xi + i\eta$, $\eta > 0$, together with simple estimates, we get

$$2\pi\|C_0\|_{H_0}^2 \leqslant \int_{\mathbb{R}} \|C_0 e^{-iH_0 s} C_0^*\|_u ds; \tag{6.2}$$

further the integrand in (6.2) is an even function of s so $(*)$ and (N) give (6.1). As for $K = A,B$, the partial isometry W is Hermitian and commutes with A because v_{00} is Hermitian; thus $A^* = B^* W$ so that

$$2\pi\|A\|_{H_0}^2 \leqslant 2\int_0^\infty \|Ae^{-iH_0 s}A^*\|_u ds \leqslant 2\|x_{00}\|_1.$$

and similarly for ϕ_B^t. ∎

Remark: The inequality (6.1) differ from the one used in Ref. 5 by factor 2 on the rhs. It cannot be avoided as shown by the following example: let $U_s = e^{-iH_0 s}$ be translations in $L^2(\mathbb{R})$, $(U_s\phi)(x) = \phi(x+s)$ and let C be a one-dimensional projection containing unit vector ψ_0 in its range. Then (6.1) corresponds to the inequality

$$\int_{\mathbb{R}} |(\psi_0, U_s\phi)|^2 ds \leqslant K\|\phi\|^2 \int_{\mathbb{R}} |(\psi_0, U_s\psi_0)| ds$$

with $K = 1$. Choosing $\phi = \phi_\alpha = \alpha^{-1}\chi_{[0,\alpha^2]}$ for $\alpha \geqslant 1$ and $\psi_0 = \phi_1$, we get

$$K \geqslant 1 - (3\alpha^2)^{-1}$$

so that $K = \frac{1}{2}$ is not possible and $K = 1$ is saturated for $\alpha \to \infty$. Consequently one must add the factor $\sqrt{2}$ in appropriate places, especially in Theorem 3.1 of Ref. 5.

Proposition 3: If (N) is valid, then the operators $\tilde{X}_{ij}$, $\tilde{Z}_{11}$ and $\hat{Z}_{11}(E)$ are bounded:

$$\|\tilde{X}_{ij}\|_{uf} \leqslant 1, \quad i,j = 0,1, \tag{6.3a}$$

$$\|\tilde{Z}_{11}\|_{uf} \leqslant \eta \tag{6.3b}$$

and

$$\|\hat{Z}_{11}(E)\|_{uf} \leqslant \eta, \quad \forall E \in \mathbb{R}. \tag{6.3c}$$

Proof: Let us take $f: \mathbb{R} \to \mathscr{B}(\mathscr{H}_0)$ with $\|f\|_1 < \infty$ and $g \in L^2(\mathbb{R};\mathscr{H}_0)$. It holds

$$\|\tilde{F}g\|_2^2 = \int_{\mathbb{R}} \left| \left| \int_{\mathbb{R}} f(y)g(x-y)dy \right| \right|^2 dx$$

$$\leqslant \int_{\mathbb{R}} dy \|f(y)\|_u \int_{\mathbb{R}} dz \|f(z)\|_u \int_{\mathbb{R}} dx \|g(x-y) \times \|g(x-z)\|$$

so that Hölder inequality applied to the last integral gives $\|\tilde{F}g\|_2^2 \leqslant \|f\|_1^2 \|g\|_2^2$, i.e.,

$$\|\tilde{F}\|_{uf} \leqslant \|f\|_1, \tag{6.4}$$

which proves (6.3a and 6.3b). Further $\|\hat{Z}_{11}(E)\|_{uf} = \|\hat{z}_{11}(E)\|_u$ because $\hat{Z}_{11}(E)$ acts as a "constant operator" on $L^2(\mathbb{R};\mathscr{H}_0)$. Consequently, $\|\hat{Z}_{11}(E)\|_{uf} \leqslant \|z_{11}\|_1$, which proves (6.3c). ∎

7. DYSON EXPANSIONS OF S_{00}^t AND S_{opt}^t

Using the notation introduced above, the optical potential (1.4) can be written as follows:

$$V_{\text{opt}}(E) = \lambda B^* A - i\lambda^2 C_0 \hat{Z}_{11}(E) C_0^* \tag{7.1}$$

(one can perform the limit $\epsilon \to 0+$ because $\|z_{11}\|_1 < \infty$). Both v and $V_{\text{opt}}(E)$ are bounded so S_{00}^t and S_{opt}^t may be expressed via Dyson expansion.[10] Consider first S_{00}^t and ϕ, $\psi \in \mathscr{H}_0$: it holds

$$(\phi, S_{00}^t\psi) = (\phi,\psi) + \sum_{k=1}^{\infty} (-i\lambda)^k \int_{-t}^{t} ds_1 \int_{-t}^{s_1} ds_2 \cdots \int_{-t}^{s_{k-1}} ds_k (\phi, e^{iH_0 s_1} \times v e^{-i(H_0 + h_1)(s_1 - s_2)} v \cdots v e^{-iH_0 s_k}\psi).$$

We substitute $v = v_{00} + v_{10} + v_{01}$ and use the factorization

from Sec. 4, then the last expression can be rewritten as follows:

$$(\phi,S^t_{00}\psi)=(\phi,\psi)+\sum_{k=1}^{\infty}(-i\lambda)^k s^t_k, \tag{7.2}$$

where

$$\begin{aligned}
s^t_1 &= \langle\phi^t_B,\psi^t_A\rangle,\\
s^t_2 &= \langle\phi^t_B,\tilde{X}_{00}\psi^t_A\rangle+\langle\phi^t_C,\tilde{Z}_{11}\psi^t_C\rangle,\\
s^t_3 &= \langle\phi^t_B,\tilde{X}_{00}\tilde{X}_{00}\psi^t_A\rangle+\langle\phi^t_B,\tilde{X}_{01}\tilde{Z}_{11}\psi^t_C\rangle\\
&\quad+\langle\phi^t_C,\tilde{Z}_{11}\tilde{X}_{10}\psi^t_A\rangle,\\
s^t_4 &= \langle\phi^t_B,\tilde{X}_{00}\tilde{X}_{00}\tilde{X}_{00}\psi^t_A\rangle+\langle\phi^t_B,\tilde{X}_{00}\tilde{X}_{01}\tilde{Z}_{11}\psi^t_C\rangle\\
&\quad+\langle\phi^t_B,\tilde{X}_{01}\tilde{Z}_{11}\tilde{X}_{10}\psi^t_A\rangle\\
&\quad+\langle\phi^t_C,\tilde{Z}_{11}\tilde{X}_{10}\tilde{X}_{00}\psi^t_A\rangle+\langle\phi^t_C,\tilde{Z}_{11}\tilde{X}_{11}\tilde{Z}_{11}\psi^t_C\rangle
\end{aligned} \tag{7.3}$$

etc. In order to illustrate the structure of a general s^t_k, let us exhibit the allowed pairings: containing

$$\begin{aligned}
&\tilde{Z}_{11}\text{: } \tilde{X}_{11}\tilde{Z}_{11},\ \tilde{X}_{01}\tilde{Z}_{11};\ \tilde{Z}_{11}\tilde{X}_{11},\ \tilde{Z}_{11}\tilde{X}_{10},\\
&\tilde{X}_{11}\text{: } \tilde{Z}_{11}\tilde{X}_{11};\ \tilde{X}_{11}\tilde{Z}_{11},\\
&\tilde{X}_{10}\text{: } \tilde{Z}_{11}\tilde{X}_{10};\ \tilde{X}_{10}\tilde{X}_{01},\ \tilde{X}_{10}\tilde{X}_{00},\\
&\tilde{X}_{01}\text{: } \tilde{X}_{10}\tilde{X}_{01},\ \tilde{X}_{00}\tilde{X}_{01};\ \tilde{X}_{01}\tilde{Z}_{11},\\
&\tilde{X}_{00}\text{: } \tilde{X}_{00}\tilde{X}_{00},\ \tilde{X}_{10}\tilde{X}_{00};\ \tilde{X}_{00}\tilde{X}_{00},\ \tilde{X}_{00}\tilde{X}_{01},
\end{aligned} \tag{7.4}$$

We need to know the number n_k of terms contained in s^t_k. Notice that each operator product in s^t_k is due to (7.4) uniquely determined by positions of the operators $\tilde{Z}_{11}$; let c^m_{k-1} denote in how many ways m operators $\tilde{Z}_{11}$ can be distributed over $k-1$ places. According to (7.4), we have $c^m_{k-1}=c^{m-1}_{k-3}+c^{m-1}_{k-4}+\cdots$ so that $c^m_{k-1}=\binom{k-m}{m}$ for $0\leqslant m\leqslant[k/2]$, where [.] denotes entire part. Consequently,

$$n_k=\sum_{m=0}^{[k/2]}\binom{k-m}{m}. \tag{7.5}$$

Let us pass now to $S^t_{\rm opt}$. Dyson expansion gives

$$(\phi,S^t_{\rm opt}\psi)=(\phi,\psi)+\sum_{k=1}^{\infty}b^t_k, \tag{7.6}$$

where

$$\begin{aligned}
b^t_1 &= -i\lambda\langle\phi^t_B,\psi^t_A\rangle+(-i\lambda)^2\langle\phi^t_C,\hat{Z}_{11}(E)\psi^t_C\rangle,\\
b^t_2 &= (-i\lambda)^2\langle\phi^t_B,\tilde{X}_{00}\psi^t_A\rangle\\
&\quad+(-i\lambda)^3\langle\phi^t_B,\tilde{X}_{01}\hat{Z}_{11}(E)\psi^t_C\rangle\\
&\quad+(-i\lambda)^3\langle\phi^t_C,\hat{Z}_{11}(E)\tilde{X}_{10}\psi^t_A\rangle\\
&\quad+(-i\lambda)^4\langle\phi^t_C,\hat{Z}_{11}(E)\tilde{X}_{11}\hat{Z}_{11}(E)\psi^t_C\rangle,
\end{aligned} \tag{7.7}$$

etc. In order to make comparison of the expansions (7.2) and (7.6) possible, one has to rearrange the latter w.r.t. the powers of $-i\lambda$. It may be done if the series converges absolutely. Applying Propositions 2 and 3, we get

$$\begin{aligned}
|b^t_1| &\leqslant|\lambda|\|\phi^t_B\|_2\|\psi^t_A\|_2+|\lambda|^2\|\phi^t_C\|_2\|\psi^t_C\|_2\|\hat{Z}_{11}(E)\|_{uf}\\
&\leqslant 2\|\phi\|\|\psi\|(|\lambda|+2\eta|\lambda|^2),\\
|b^t_2| &\leqslant 2\|\phi\|\|\psi\|(|\lambda|^2+2|\lambda|^3\eta+|\lambda|^4\eta^2),\\
&\cdots\cdots\cdots\cdots\\
|b^t_k| &\leqslant 2\|\phi\|\|\psi\||\lambda|^k\sum_{s=0}^{k}\binom{k}{s}(\eta|\lambda|)^s,
\end{aligned}$$

so

$$\sum_{k=1}^{\infty}|b^t_k|\leqslant 2\|\phi\|\|\psi\|\sum_{k=1}^{\infty}|\lambda|^k(1+\eta|\lambda|)^k,$$

and therefore the series converges absolutely for $|\lambda|(1+\eta|\lambda|)<1$, i.e.,

$$|\lambda|<(1/2\eta)((1+4\eta)^{1/2}-1). \tag{7.8}$$

It is easy to see that after the rearrangement, (7.6) differs from (7.2) just by replacement of all $\tilde{Z}_{11}$ by $\hat{Z}_{11}(E)$.

8. COMPARISON OF THE EXPANSIONS

Now we can complete the proof comparing term-by-term the expansions (7.2) and (7.6). We denote

$$(\phi,(S^t_{00}-S^t_{\rm opt})\psi)=\sum_{k=2}^{\infty}(-i\lambda)^k d^t_k, \tag{8.1}$$

then Propositions 2 and 3 together with (5.2) give

$$\begin{aligned}
|d^t_2| &\leqslant\|\phi^t_C\|_2\alpha(E)\leqslant\sqrt{2}\|\phi\|\alpha(E),\\
|d^t_3| &\leqslant\|\phi^t_B\|_2\|\tilde{X}_{01}\|_{uf}\alpha(E)+\|\phi^t_C\|_2\|\psi^t_A\|_2\gamma(E)\\
&\leqslant\sqrt{2}\|\phi\|\{\alpha(E)+\sqrt{2}\|\psi\|\gamma(E)\},\\
|d^t_4| &\leqslant\sqrt{2}\|\phi\|\alpha(E)+2\sqrt{2}\|\phi\|\gamma(E)\sqrt{2}\|\psi\|\\
&\quad+|\langle\phi^t_C,(\tilde{Z}_{11}-\hat{Z}_{11}(E))\tilde{X}_{11}\tilde{Z}_{11}\psi^t_C\rangle|\\
&\quad+|\langle\phi^t_C,\hat{Z}_{11}(E)\tilde{X}_{11}(\tilde{Z}_{11}-\hat{Z}_{11}(E))\psi^t_C\rangle|\\
&\leqslant\sqrt{2}\|\phi\|\{(1+\eta)\alpha(E)+\sqrt{2}\|\psi\|(2\gamma(E)+\eta\beta(E))\}
\end{aligned}$$

etc. In order to estimate general d^t_k, let us rewrite it as

$$d^t_k=\sum_{m=1}^{[k/2]}d^t_k(m), \tag{8.2}$$

where $d^t_k(m)$ contains terms with just m operators $\tilde{Z}_{11}$ or $\hat{Z}_{11}(E)$. The number of terms in $d^t_k(m)$ is $mc^m_{k-1}=m\binom{k-m}{m}$ because similarly as above one must add and subtract $(m-1)c^m_{k-1}$ terms in order to single out the differences $\tilde{Z}_{11}-\hat{Z}_{11}(E)\equiv D_{11}$. They can be divided into three groups containing terms in which D_{11} is followed by $\psi^t_C,\tilde{X}_{11}$ and $\tilde{X}_{10}$; numbers of their elements are denoted $c^m_{k-1}(\alpha)$, $c^m_{k-1}(\beta)$ and $c^m_{k-1}(\gamma)$, respectively. It holds

$$\begin{aligned}
c^m_{k-1}(\alpha) &= \binom{k-m-1}{m-1},\\
c^m_{k-1}(\beta) &= c^{m+1}_k(\gamma)=m\binom{k-m-1}{m}
\end{aligned} \tag{8.3}$$

as shown in Appendix, and therefore (6.1) and (6.3) together with (5.2) imply

$$\begin{aligned}
|d^t_k(m)| &\leqslant\sqrt{2}\|\phi\|\eta^{m-1}\Big\{\binom{k-m-1}{m-1}\alpha(E)\\
&\quad+\sqrt{2}\|\psi\|\Big[(m-1)\binom{k-1-m}{m-1}\beta(E)\\
&\quad+m\binom{k-1-m}{m}\gamma(E)\Big]\Big\}.
\end{aligned} \tag{8.4}$$

In order to make use of this inequality, one has to estimate

$$n_k(i)=\sum_{n=1}^{[k/2]}\eta^{m-1}c^m_{k-1}(i),\quad i=\alpha,\beta,\gamma. \tag{8.5}$$

Assume first $\eta\leqslant 1$. One obtains easily the relations

$$n_k(\alpha) = n_{k-1}(\alpha) + n_{k-2}(\alpha),$$
$$n_k(\beta) = n_{k-2}(\alpha) + n_{k-2}(\beta) + n_{k-1}(\beta),$$
$$n_k(\gamma) = n_{k+1}(\beta),$$

which imply by induction $n_k(\alpha) \leqslant 2^{k-3}$ and $n_k(\beta) = n_{k-1}(\gamma) \leqslant 2^{k-3}$. On the other hand, if $\eta \geqslant 1$, then

$$n_k(\alpha) = \eta^{[k/2]-1} \sum_{m=1}^{[k/2]} \binom{k-m-1}{m-1} \leqslant \eta^{-1}(\sqrt{\eta})^k 2^{k-3},$$

and similarly for $n_k(\beta)$, $n_k(\gamma)$. Together we obtain

$$\begin{aligned} n_k(\alpha) &\leqslant \eta_0^{-1}(\sqrt{\eta_0})^k 2^{k-3}, \\ n_k(\beta) &\leqslant \eta_0^{-1}(\sqrt{\eta_0})^k 2^{k-3}, \\ n_k(\gamma) &\leqslant \eta_0^{-1}(\sqrt{\eta_0})^k 2^{k-2}, \end{aligned} \tag{8.6}$$

where $\eta_0 = \max\{1, \eta\}$. The relations (8.1)–(8.6) imply

$$\begin{aligned} |(\phi,(S_{00}^t - S_{\mathrm{opt}}^t)\psi)| &\leqslant \sum_{k=2}^{\infty} |\lambda|^k \sum_{m=1}^{[k/2]} |d_k^t(m)| \\ &\leqslant \sqrt{2}\|\phi\| \sum_{k=2}^{\infty} |\lambda|^k \{n_k(\alpha)\alpha(E) + \sqrt{2}\|\psi\| \sum_{i=\beta,\gamma} n_k(i) i(E)\} \\ &\leqslant (\sqrt{2})^{-1}\|\phi\| |\lambda|^2 \sum_{l=0}^{\infty} (2|\lambda|\sqrt{\eta_0})^l \{\alpha(E) \\ &\quad + \sqrt{2}\|\psi\|(\beta(E) + 2\gamma(E))\}, \end{aligned}$$

i.e., the following inequality

$$\begin{aligned} |(\phi(S_{00}^t - S_{\mathrm{opt}}^t)\psi)| &\leqslant \lambda^2\|\phi\|\{\alpha(E) + \sqrt{2}\|\psi\|(\beta(E) \\ &\quad + 2\gamma(E))\}(\sqrt{2}(1 - 2|\lambda|\sqrt{\eta_0}))^{-1} \end{aligned} \tag{8.7}$$

for each $\phi \in \mathscr{H}_0$, which is equivalent to (5.1). finally, $(2\sqrt{\eta_0})^{-1} < (2\eta)^{-1}(\sqrt{1+4\eta} - 1)$ so the condition (7.8) is fulfilled; thus the proof is finished.

9. CONCLUDING REMARKS

Comparing to the results of Ref. 5, the estimate (5.2) contains the term proportional to $\gamma(E)$ and the parameter η_0 which are present due to nonzero v_{00}. The right-hand-sides of the two estimates depend on the coupling constant in a slightly different way because of different processes of estimation. What is more surprising, the substantially more complicated structure of Dyson expansions in the case of nonzero v_{00} [we have $\Sigma_{i=\alpha,\beta,\gamma} n_k(i) \lesssim 2^{k-1}$ terms in the k-th order of (8.1), in contrast to mere $\frac{1}{2}k$ for even k only if $v_{00} = 0$] reduces to the combinatorial factor only which in effect multiplies the coupling constant by two in (5.1).

The physical interpretation mentioned in the introduction does not change with nonzero v_{00}. Assume for simplicity[8] $E = 0$, then the optical approximation can be intuitively expected from (5.1), (5.2) to be good if $z_{11}(.)$ is sharply peaked around zero [in a time scale appropriate for $\psi_C^t(.)$, $x_{11}(.)$ and $x_{10}(.)$]. This assertion can be formulated rigorously as a direct generalization to Theorem 3.2 of Ref. 5. Let us notice that the appearance of three natural time scales is proper not only to the optical model: remember the well-known problem of deviations from the exponential decay law.[11] In most models, there are two regions in which these deviations affect the decay law significantly: for times much larger[12] and much smaller[13] than the mean life of the considered system.

The most problematic part of the discussed approximation lies in replacement of h_1 by h_1' with continuous spectrum.[14] As shown in Sec. 3, it needs the initial wavepackets ϕ to be such that $t \mapsto \|ve^{-iH_0 t}\phi\|$ decreases sufficiently rapidly with $|t|$; such a behavior could be expected for ϕ with slowly varying energy density. On the other hand, considerations of Sec. 2 show that the use of the energy-dependent optical potential (1.4) can be justified if the energy density is sharply peaked. The arguments leading to this contradiction are, of course, only roughly qualitative; a more careful analysis and model examples are needed in order to decide whether and under which conditions the scheme works.

Let us finally mention, that though (1.4) is "the best" optical potential in view of Sec. 2, in practical calculations it is often replaced by some local one,[2,4,7] i.e., by operator of multiplication by a complex function. Such potentials can be handled more easily because one can adapt for them methods elaborated for real local potentials[15]; nevertheless, there is a lot of open problems related to them, especially in scattering theory.

ACKNOWLEDGMENT

The authors are grateful to Dr. J. Blank and Dr. M. Havlíček for useful discussions.

APPENDIX

Consider first $c_{k-1}^m(\alpha)$. In the corresponding terms, D_{11} must stand first from the right and its left neighbor is $\tilde{X}_{11}$ or $\tilde{X}_{01}$ due to (7.4). Thus we have $k-3$ places for arranging of $m-1$ operators Z_{11} [i.e., $\tilde{Z}_{11}$ or $\hat{Z}_{11}(E)$], symbolically

$$\overbrace{\quad\quad}^{\frac{m-1}{k-1}} \tilde{X} D_{11},$$

which gives $c_{k-1}^m(\alpha) = c_{k-3}^{m-1} = \binom{k-m-1}{m-1}$. As for $c_{k-1}^m(\beta)$, we have the following diagrams

$$\begin{aligned} &D_{11}\tilde{X}_{11}Z_{11}\tilde{X} \overbrace{\quad\quad}^{\frac{m-2}{k-5}}, \\ &\overbrace{\quad\quad}^{\frac{m-s-2}{k-l-6}} \tilde{X} D_{11}\tilde{X}_{11}Z_{11}\tilde{X} \overbrace{\quad}^{\frac{s}{l}} \quad, l = 0,1,\ldots,k-6, \\ &\overbrace{\quad\quad}^{\frac{m-2}{k-5}} \tilde{X} D_{11}\tilde{X}_{11}Z_{11}, \end{aligned}$$

which give

$$\begin{aligned} c_{k-1}^m(\beta) &= 2\binom{k-m-2}{m-2} \\ &\quad + \sum_{l=0}^{k-6} \sum_{s=0}^{[(l+1)/2]} \binom{k-l-m+s-3}{m-s-2}\binom{l+1-s}{s}, \end{aligned}$$

i.e.,

$$c_{k-1}^m(\beta) = \sum_{r=0}^{k-4} \sum_{s=0}^{[r/2]} \binom{k-r-m+s-2}{m-s-2}\binom{r-s}{s}.$$

Analogously, the diagrams

$$D_{11}\tilde{X}_{10}\tilde{X}\frac{m-1}{k-4}\Big|,$$

$$\Big|\frac{m-s-1}{k-l-5}\tilde{X}D_{11}\tilde{X}_{10}\tilde{X}\frac{s}{l}\Big|,\ l=0,1,\ldots,k-5,$$

$$\Big|\frac{m-1}{k-4}\tilde{X}D_{11}\tilde{X}_{10}$$

yield

$$c^m_{k-1}(\gamma) = \sum_{r=0}^{k-3}\sum_{s=0}^{[r/2]}\binom{k-r-m+s-2}{m-s-1}\binom{r-s}{s}$$
$$= c^{m+1}_k(\beta).$$

In view of the last relation, it is sufficient to evaluate for instance $c^m_{k-1}(\gamma)$. We change summation indices to $p=r-s$ and s; it holds $0\leqslant s\leqslant p\leqslant s+k-2m-1\leqslant k-m-2$ and $p+s\leqslant k+2s-2m-1\leqslant k-3$ so[16]

$$c^m_{k-1}(\gamma) = \sum_{s=0}^{m-1}\sum_{p=s}^{s+k-2m-1}\binom{k-m-2-p}{m-1-s}\binom{p}{s}$$
$$= \sum_{s=0}^{m-1}\binom{k-m-1}{m},$$

and therefore (8.3) is valid.

[1]With the referee's comment in mind, we stress that the term "channel" is used in this paper as well as in Ref. 5 below in a conditional sense only. The channel projections P_0, P_1 here are related, e.g., to the projections Λ, M of Ref. 7; thus both the channels and especially $\mathscr{H}_1$ can contain a lot of the "real" scattering channels characterized by internal quantum numbers.

[2]A. I. Akhiezer and I. Ya. Pomeranchuk, *Some Problems in the Nuclear Theory* (Russian) (Gostekhizdat, Moscow, 1948); H. Feshbach, C E. Porter, and V. F. Weisskopf, Phys. Rev. **96**, 448–464 (1954).

[3]H. Feshbach, Ann. Phys. **5**, 357–390 (1958); H. Feshbach, Ann. Phys. **19**, 287–313 (1962).

[4]I. Úlehla, L. Gomolčák, and Z. Pluhař, *Optical Model of the Atomic Nucleus* (Czech. Acad. Sci. Publ., Prague, 1964).

[5]E. B. Davies, Ann. Inst. H. Poincaré **A29**, 395–413 (1978).

[6]T. Kato, Math. Annalen **162**, 258–279 (1966).

[7]J. R. Taylor, *Scattering Theory-The Quantum Theory of Nonrelativistic Collisions* (Wiley, New York, 1972), § 19.4.

[8]As for the assumption $E=0$, it can be always fulfilled by redefinition $H\to H'=H-E$; however, we let mostly E to be nonzero.

[9]Instead of the normalization (N), one can choose some other one. For instance, after fixing λ by $\|x_{00}\|_1=1$, one can choose C_0 so that some of the remaining $\|x_{ij}\|_1$'s would equal one. In such a case, however, the following considerations would contain three parameters instead of one and the estimation would be much more complicated.

[10]M. Reed and B. Simon, *Methods of Modern Mathematical Physics* II. *Fourier Analysis, Self-Adjointness* (Academic, New York, 1975), Sec. X.12. Notice that in the following we deal with matrix elements of S'_{00}, $S'_{\rm opt}$, i.e., with "weak" solutions to the evolution equations only. Notice further that the use of Dyson expansion is not conditioned by Hermiticity of the interaction.

[11]L. Fonda, G. C. Ghirardi, and A. Rimini, Rep. Progr. Phys. **41**, 587–631 (1978); P. Exner, Rep. Math. Phys. **17**, 275–285 (1980), and references given in these papers.

[12]L. A. Khalfin, Zh. Eksp. Teor. Fiz. **33**, 1371–1382 (1957) [Sov. Phys. JETP **6**, 1053–1063 (1958)]; P. T. Matthews and A. Salam, Phys. Rev. **115**, 1079–1084 (1958);G. Höhler, Z. Phys. **152**, 542–565 (1958).

[13]A. Beskow and J. Nilsson, Arkiv. Fysik **34**, 561–569 (1967); M. Havlíček and P. Exner, Czech. J. Phys. B **23**, 594–600 (1973); A. Degasperis, L. Fonda, and G. C. Ghirardi, Nuovo Cimento **21A**, 471–484 (1974); B. Misra, and K. Sinha, Helv. Phys. Acta **50**, 99–104 (1977); B. Misra and E. C. G. Sudarshan, J. Math. Phys. **18**, 756–763 (1977).

[14]The analogous problem concerns the spectrum of H_0 which must be continuous if v_{00} is nonzero, but this is seldom a problem.

[15]J. Blank, P. Exner, and M. Havlíček, Czech. J. Phys. B **29**, 1325–1341 (1979); P. Exner and G. I. Kolerov, Phys. Lett. **83A**, 203–206 (1981).

[16]A. P. Prudnikov, Yu.A. Brychkov, and O. I. Marichev, *Integrals and Series* (Nauka, Moscow, 1981), cf. 4.2.5.39; one has to correct an evident typographical error in this formula.

Splines and the Galerkin method for solving the integral equations of scattering theory

M. Brannigan
Department of Statistics and Computer Sciences, University of Georgia, Athens, Georgia 30602

D. Eyre
National Research Institute for Mathematical Sciences of the CSIR, PO Box 395, Pretoria 0001, Republic of South Africa

(Received 20 July 1982; accepted for publication 2 September 1982)

This paper investigates the Galerkin method with cubic B-spline approximants to solve singular integral equations that arise in scattering theory. We stress the relationship between the Galerkin and collocation methods. The error bound for cubic spline approximates has a convergence rate of $O(h^4)$, where h is the mesh spacing. We test the utility of the Galerkin method by solving both two- and three-body problems. We demonstrate, by solving the Amado–Lovelace equation for a system of three identical bosons, that our numerical treatment of the scattering problem is both efficient and accurate for small linear systems.

PACS numbers: 03.80. + r, 02.30.Rz, 05.30.Jp

I. INTRODUCTION

The aim of this paper is to investigate the Galerkin technique with cubic B-spline approximants to solve integral equations that arise in scattering theory. In particular, we consider a numerical treatment of the integral equations that describe two- and three-body problems.

The integral equation approach of Faddeev[1] provides a framework for the investigation of the three-body problem. In general, the Faddeev equations are singular multidimensional integral equations of the second kind. Projection methods,[2,3] which expand the solution of the integral equation in some finite set of basis functions, have been used successfully to solve both the bound state[4] and scattering[5] problem. A difficulty with this approach, however, is that in practice the efficiency of the method rests on there being a suitable choice of basis functions. Often these functions must be tailored to suit a particular problem. If the basis functions are chosen to be spline functions, then no special effort is required to select an optimum basis set.

Splines have been used in the numerical treatment of scattering and bound state problems.[6] The use of splines and the Galerkin method has been proposed[7] as a practical means of solving the Faddeev equations. Splines have also been used in a degenerate kernel method to solve integral equations that describe three-body scattering.[8]

In a previous paper[9] (hereinafter referred to as I) we carried out an investigation of the projection collocation method, using spline approximants, and applied this technique to solve integral equations that arise in two-body scattering. The present paper is devoted to a closely related variational method using the Galerkin technique. (The use of a variational method with compactly supported trial functions is often called "the finite-element method.")

Section II gives a mathematical formulation of the method and an error analysis. In Sec. III we review the integral equation approach to two- and three-body problems. We include only those aspects which are relevant to the numerical treatment described in Sec. IV. Section V gives our numerical results, including a numerical comparison between the Galerkin and collocation methods, and in Sec. VI we state our main conclusions.

II. THEORY

The general problem, from which our physical problem is taken, is the solution to the operator equation

$$(\mathscr{I} - \mathscr{K})f = y \tag{2.1}$$

where $f, y \in C[a,b]$, the space of continuous real-valued functions defined on the closed interval $[a,b]$, $\mathscr{K}$ and $\mathscr{I}$ are linear operators mapping $C[a,b]$ into itself, with $\mathscr{I}$ the identity operator.

Since it is not always possible to find an analytic solution of this equation for the physical problems under investigation, we seek to find an approximation to f which satisfies some interpolation property or variational principle. To this end, we choose as our approximating functions the finite linear space of cubic splines with a given set of knots.
To be precise, let π_n be a partition of the interval $[a,b]$, defined by the knots $a = t_1 < t_2 < \cdots < t_n = b$, with a mesh spacing h_n, where $h_n = \max\{(t_{i+1} - t_i): 1 \leqslant i < n\}$. On this partition together with the extended knots $t_{-2} \leqslant t_{-1} \leqslant t_0 \leqslant t_1$ and $t_n \leqslant t_{n+1} \leqslant t_{n+2} \leqslant t_{n+3}$, we can construct the cubic B-splines. Numerically stable methods for this construction can be found in Ref. 10. We will denote by B_{ni} the cubic B-spline defined on π_n and which is nonzero over the interval (t_{i-2}, t_{i+2}). It is then known that $\{B_{ni}: i = 0,\ldots,n+1\}$ forms a basis for the linear space S_n of cubic splines with partition π_n and which has C_2 continuity over $[a,b]$. Using this linear space, we can now approximate the function f by the linear combination

$$\sum_{i=0}^{n+1} \alpha_{ni} B_{ni}, \tag{2.2}$$

and seek an appropriate set of coefficients $\{\alpha_{ni}\}$.

Consider now the function $r \in C[a,b]$ given by

$$r = \sum_{i=0}^{n+1} \alpha_{ni}(\mathscr{I} - \mathscr{K})B_{ni} - y. \tag{2.3}$$

 0022-2488/83/061548-07$02.50 © 1983 American Institute of Physics

One such method of solving for α_{ni} is to choose $n+2$ abscissa values $s_j, j = 0,...,n+1$, and then solve the linear system of equations given by

$$r(s_j) = 0, \quad j = 0,...,n+1. \tag{2.4}$$

This is the methoid of collocation and a discussion of this method is given in I. The Galerkin techinque, which we discuss in this paper, supposes that we have defined on $C[a,b]$ the usual inner product

$$(\psi,\phi) = \int_a^b \psi(x)\phi(x)\,dx. \tag{2.5}$$

We now choose the coefficients α_{ni}, $i = 0,...,n+1$, by solving the system of linear equations

$$(r,B_{ni}) = 0, \quad i = 0,...,n+1. \tag{2.6}$$

To analyze this method, let G_n be the mapping of $C[a,b]$ onto S_n such that, for $\psi \in C[a,b]$, $G_n\psi$ is the best approximation of ψ by S_n in the norm defined by our inner product. That is,

$$G_n\psi = \sum_{i=0}^{n+1} \lambda_i B_{ni}, \tag{2.7}$$

where the λ_i are the solutions (see Ref. 11) to the linear system

$$\sum_{i=0}^{n+1} \lambda_i (B_{ni},B_{nj}) = (\psi,B_{nj}), \quad j = 0,...,n+1. \tag{2.8}$$

That G_n is a projection operator follows easily.

If we now consider $G_n r$, then we have to solve

$$\sum_{i=0}^{n+1} \lambda_i (B_{ni},B_{nj}) = (r,B_{nj}), \quad j = 0,...,n+1. \tag{2.9}$$

Due to the linear independence of the splines, the solution to the above system is $\lambda_i = 0$ for all i, that is, $G_n r = 0$ (zero function), if and only if $(r,B_{nj}) = 0$ for $j = 0,...,n+1$. Hence our Galerkin procedure gives us the solution $g \in S_n$, such that

$$G_n(\mathscr{I} - \mathscr{K})g = G_n y. \tag{2.10}$$

This relationship, together with the fact that G_n is a projection operator, shows that the Galerkin procedure is a projection method.

In the scattering problems that we consider the kernel K is singular and of the principal value type. That is, the operator $\mathscr{K}$ has the form

$$\mathscr{K}f = \int_a^b K(\cdot,t)f(t)\,dt, \tag{2.11}$$

where

$$K(\cdot,t) = \tilde{K}(\cdot,t)/(t-u), \tag{2.12}$$

for some $u \in [a,b]$.

The usual error analysis of integral equations, such as that set out in Ref. 2, does not apply because K is not continuous. However, as can be seen from our numerical results, we do obtain "convergence." In order to justify these results, we are able to show that the Galerkin method does converge. We first show how the method set out above is the solution of a finite system.

To solve the system of equations $(r,B_{ni}) = 0$, we need to perform two integrations. We carry out these integrations using a standard numerical quadrature, namely

$$\int_a^b F(x)\,dx = \sum_{i=0}^{p} \omega_i F(x_i) \tag{2.13}$$

for some suitably chosen x_i, $i = 1,...,p$. Thus, for any $k = 0,...,n+1$,

$$\begin{aligned}(r,B_{nk}) &= \int_a^b r(t)B_{nk}(t)\,dt \\ &= \sum_{j=1}^{p} \omega_j r(t_j)B_{nk}(t_j),\end{aligned} \tag{2.14}$$

while

$$\begin{aligned}r(t_j) &= \sum_{i=0}^{n+1} \alpha_{ni}\left[B_{ni}(t_j) - \int_a^b \frac{\tilde{K}(t_j,t)}{t-u} B_{ni}(t)dt\right] - y(t_j) \\ &= \sum_{i=0}^{n+1} \alpha_{ni}\left[B_{ni}(t_j) - \sum_{q=1}^{p} \omega_q \frac{\tilde{K}(t_j,t_q)}{t_q - u} B_{ni}(t_q)\right] - y(t_j).\end{aligned} \tag{2.15}$$

In this representation we may assume that $t_q \neq u$ for any $q = 1,...,p$.

We thus see that our problem is totally defined by the grid points $(t_j,t_q), j,q = 1,...,p$. Given this grid, we are able to approximate $\tilde{K}(s,t)$ by a function $(t-u)k(s,t)$ such that

$$(t_q - u)k(t_j,t_q) = \tilde{K}(t_j,t_q), \quad j,q = 1,...,p. \tag{2.16}$$

We have, therefore, an equivalent formulation to Eq. (2.15), namely

$$r(t_j) = \sum_{i=0}^{n+1} \alpha_{ni}\left[B_{ni}(t_j) - \sum_{q=1}^{p} \omega_q k(t_j,t_q)B_{ni}(t_q)\right] - y(t_j). \tag{2.17}$$

The matrix of coefficients formed from Eqs. (2.15) and (2.17), corresponding to the system in Eq. (2.6), are equal; hence equal solutions result. Equation (2.17) would be formed if we were solving the integral equation

$$f(s) - \int_a^b k(s,t)f(t)\,dt = y(s), \tag{2.18}$$

and our method cannot distinguish between Eqs. (2.1) and (2.18).

If now f_1 is the solution to Eq. (2.18), then from Theorem 2 in Ref. 2, p. 51, using the uniform norm, we obtain a bound for the finite solution $g_n \in S_n$ to the system in Eq. (2.6) given by

$$\|f_1 - g_n\| \leqslant \|(I - G_n\mathscr{K}_1)^{-1}\|\,\|f_1 - G_n f_1\|, \tag{2.19}$$

where $\mathscr{K}_1$ is the integral in Eq. (2.18).

One property of least squares approximation is that there exists a strictly increasing set of abscissae τ_i, $i = 0,...,n+1$, such that $t_{i-2} < \tau_i \leqslant t_{i+2}$, where the error of approximation is zero. Hence the set $\{\tau_i\}$ is a set of collocation points depending on our function f_1. We are, therefore, able to use the error bounds derived for cubic spline interpolation. In fact it can be shown (see Ref. 11), using the Peano kernel theorem, that

$$\|f_1 - G_n f_1\| \leqslant \tfrac{5}{384}\,\|f_1^{(4)}\|h_n^4, \tag{2.20}$$

where $f_1 \in C^4[a,b]$.

We have thus shown that our method does converge as $O(h_n^4)$ to some function f_1, and for some $\epsilon > 0$ we have an N

such that every solution $g_m \in S_n$, for $m > N$, of Eq. (2.6) satisfies

$$\| f_1 - g_m \| < \epsilon. \tag{2.21}$$

Now consider the solution f_2 to Eq. (2.1). Let us impose the following condition on the kernel: We shall say that the kernel in Eq. (2.1) is well conditioned if for any given $\delta > 0$ we can find an approximation $k(s,t)$, satisfying Eq. (2.16), such that the solutions f_1, f_2 of Eqs. (2.18) and (2.1), respectively, satisfy

$$\| f_1 - f_2 \| < \delta. \tag{2.22}$$

Assuming now that the kernel in Eq. (2.1) is well conditioned, then for $m > M$ we find

$$\| f_2 - g_m \| \leqslant \| f_2 - f_1 \| + \| f_1 - g_m \| < \delta + \epsilon, \tag{2.23}$$

and hence the solutions to Eq. (2.16) converges to the solution of our original Eq. (2.1).

A useful device to improve a solution is the Neumann iteration scheme,[12] according to which a sequence of functions $\{\bar{g}_i\}$, in which our initial approximation is $g_m = \bar{g}_1$, is constructed from the relationship

$$\bar{g}_{i+1}(\cdot) = y(\cdot) + \int_a^b K(\cdot, t)\bar{g}_i(t)\, dt. \tag{2.24}$$

This iterative improvement technique will converge provided $\bar{g}_1$ is a sufficiently good initial guess. In our algorithm we use one iteration which we find sufficient.

III. INTEGRAL EQUATIONS

In this section we review the integral equations that describe two-body and three-body problems.

Let us consider the two-body problem. The partial-wave off-shell amplitude $M(q,\kappa;z)$ is given by a solution of the integral equation

$$M(q,\kappa;z) = v(q,\kappa) - (2/\pi)\int_0^\infty [v(q,q')q'^2 dq'/(q'^2 - z)] \times M(q',\kappa;z), \quad q \in [0,\infty). \tag{3.1}$$

Here $v(q,q')$ is a two-body potential and z is a complex energy variable. Let κ be the incident two particle momentum, and set $z = \kappa^2$. The half-shell K matrix $M(q,\kappa) = M(q,\kappa;\kappa^2)$ is given by a solution of the integral equation

$$M(q,\kappa) = v(q,\kappa) - (2/\pi)\int_0^\infty [v(q,q')q'^2 dq'/(q'^2 - \kappa^2)] M(q',\kappa), \tag{3.2}$$

where the integral is evaluated by a principal value prescription and M is a real-valued function. The s-wave two-body phase shift $\eta(\kappa)$ is given by

$$M(\kappa,\kappa) = -[\kappa \cot \eta(\kappa)]^{-1}. \tag{3.3}$$

Next, we consider the homogeneous form of Eq. (3.1). The homogeneous equation

$$\psi(q,z) = -(2/\pi)\int_0^\infty [v(q,q')q'^2 dq'/(q'^2 - z)]\psi(q',z) \tag{3.4}$$

has a nontrivial solution only at $z = -\epsilon_i$, where ϵ_i is real and corresponds to the energy of the ith bound state. For simplicity, we shall assume that the potential v can support only one bound state with energy ϵ.

We now consider the three-body problem. We shall consider only a simple system of three identical particles, each with spin zero and isospin zero, interacting via a separable two-particle potential of the form

$$v(q,q') = \lambda \phi(q)\phi(q'), \tag{3.5}$$

where $\phi(q)$ is some smooth vertex function. For the above potential it is known that the Faddeev equations reduce to an integral equation in one momentum variable.[13] In the elastic scattering sector (negative three-body energy) the partial-wave half-shell K matrix $X(p,k)$ is given by a solution of the integral equation

$$X(p,k) = Z(p,k;k) - (2/\pi)\int_0^\infty [Y(p,p';k)p'^2 dp'/(p'^2 - k^2)] \times X(p',k), \tag{3.6}$$

where k is the momentum of the incident particle in the three-body c.m. frame. The s-wave three-body phase-shift $\delta(k)$ is given by

$$X(k,k) = -\tfrac{3}{4}[k \cot \delta(k)]^{-1}. \tag{3.7}$$

The driving term Z of Eq. (3.6) is defined by the integral

$$Z(p,p';k) = -\int_{-1}^{1} \frac{\phi(q_1)\phi(q_2)}{p^2 + p'^2 + pp'x - \frac{3}{4}k^2 + \epsilon}\, dx, \tag{3.8}$$

where

$$q_1 = (\tfrac{1}{4}p^2 + p'^2 + pp'x)^{1/2}, \qquad q_2 = (p^2 + \tfrac{1}{4}p'^2 + pp'x)^{1/2}. \tag{3.9}$$

The function Y in the kernel of Eq. (3.6) is constructed from the solution $M(q,q';z)$ of Eq. (3.1). Because the interaction v is separable and supports a single bound state with energy ϵ, we can write

$$M(q,q';z) = \phi(q)[S(z)/(\epsilon + z)]\phi(q'). \tag{3.10}$$

The energy of the two-particle subsystem z is determined by energy conservation in the three-body system, viz.,

$$\tfrac{3}{4}p'^2 + z = \tfrac{3}{4}k^2 - \epsilon, \tag{3.11}$$

where p' is the off-shell two-cluster momentum. The function Y takes the form

$$Y(p,p';k) = \tfrac{4}{3}Z(p,p';k)S(\tfrac{3}{4}(k^2 - p'^2) - \epsilon). \tag{3.12}$$

In order to carry out model calculations, we have chosen for the two-particle interaction a one-term separable potential of the Yamaguchi[4] type. In momentum space, this potential can be written

$$v(q,q') = \lambda/(q^2 + \beta^2)(q'^2 + \beta^2). \tag{3.13}$$

Over the scattering region, $\kappa^2 > 0$, one obtains for the two-body K matrix the result

$$M(q,\kappa) = \left(\frac{\kappa^2 + \beta^2}{q^2 + \beta^2}\right)\left[-\beta + \frac{\kappa^2 + \beta^2}{2\beta} + \frac{(\kappa^2 + \beta^2)^2}{2\beta(\beta + \alpha)^2}\right], \tag{3.14}$$

where

$$\alpha \equiv \sqrt{\epsilon} = -\beta + \sqrt{-\lambda/2\beta}. \tag{3.15}$$

This simple analytic solution can be used to test the accuracy of our numerical methods for solving Eqs. (3.2) and (3.4).

For the Amado–Lovelace equation we take a Yamaguchi form for the vertex function, i.e.,

$$\phi(q) = C/(q^2 + \beta^2). \tag{3.16}$$

The constant

$$C = [2\alpha\beta(\alpha+\beta)^3]^{1/2}, \tag{3.17}$$

is chosen such that the bound state wavefunction is normalized to unity. For the function S, defined by Eq. (3.10), one obtains

$$S(-\chi^2) = (\beta+\chi)^2(\alpha+\chi)/\alpha(\alpha+\beta)(\alpha+\chi+2\beta), \tag{3.18}$$

where $\chi > 0$.

IV. APPROXIMATION METHODS

In this section we describe our numerical treatment of Eqs. (3.2), (3.4), and (3.6). Because of the similarity between Eqs. (3.2) and (3.6), we shall restrict our discussion to the former equation.

It is convenient to map the momentum variable q onto a finite interval $[-1, +1]$ using

$$q(x) = \xi[(1+x)/(1-x)], \tag{4.1}$$

where ξ is a constant scale parameter. Using the above transformation, Eq. (3.2) can be written

$$M(q(x),\kappa) = v(q(x),\kappa) - 2\xi^3 \frac{2}{\pi}\int_{-1}^{1} v(q(x),q'(x'))\left(\frac{1+x'}{1-x'}\right)^2 \times \frac{M(q'(x'),\kappa)\,dx'}{[\xi^2(1+x')^2 - \kappa^2(1-x')^2]}. \tag{4.2}$$

A description of the numerical treatment of Eq. (4.2) using the collocation method can be found in (*I*). We now describe the Galerkin method. Let M_n be the approximation $G_n M$ to M given by

$$M_n(q(x),\kappa) = \sum_{i=0}^{n+1} \alpha_{ni}(\kappa) B_{ni}(x). \tag{4.3}$$

The linear system of $n+2$ equations for α_{ni} is

$$\sum_{i=0}^{n+1} \alpha_{ni}(\kappa)[(B_{ni},B_{nj}) + (\mathscr{K}(\kappa^2)B_{ni},B_{nj})] = (v(\kappa),B_{nj}), \quad j = 0,\ldots,n+1, \tag{4.4}$$

where

$$v((\kappa),B_{nj}) = \int_{-1}^{1} v(q(x),\kappa)B_{nj}(x)\,dx, \tag{4.5}$$

$$(\mathscr{K}(\kappa^2)B_{ni},B_{nj}) = 2\xi^3\frac{2}{\pi}\int_{-1}^{1}\int_{-1}^{1} \frac{v(q(x),q'(x'))[(1+x')/(1-x')]^2}{[\xi^2(1+x')^2 - \kappa^2(1-x'^2)]} \times B_{ni}(x)B_{nj}(x')\,dx dx'. \tag{4.6}$$

The integral over the singularity in Eq. (4.6) can be evaluated by the method of subtraction.[15]

The first iterative improvement $M_n^{(2)}$ to the approximate solution $M_n^{(1)} = M_n$, defined by Eq. (4.3), is given by the formula

$$M_n^{(2)}(q(x),\kappa) = v(q(x),\kappa)) - \sum_{i=0}^{n+1} \alpha_{ni}(\kappa) I_{ni}(q(x),\kappa), \tag{4.7}$$

where

$$I_{ni}(q(x),\kappa) = 2\xi^3\frac{2}{\pi}\int_{-1}^{1} \frac{v(q(x),q'(x'))[(1+x')/(1-x')]^2}{[\xi^2(1+x')^2 - \kappa^2(1-x')^2]} \times B_{ni}(x')dx'. \tag{4.8}$$

We note that Eq. (4.8) is a one-dimensional integral and therefore simpler to evaluate than the two-dimensional integral in Eq. (4.6).

For the bound state problem we define the approximation

$$\psi_n(q(x),z) = \sum_{i=0}^{n+1} \beta_{ni}(z) B_{ni}(x). \tag{4.9}$$

After applying the Galerkin technique, Eq. (3.4) now reads

$$\beta_{ni}(z) = -\sum_{j,k=0}^{n+1} [(B_{ni},B_{nj})]^{-1}(\mathscr{K}(z)B_{nk},B_{nj})\beta_{nk}(z). \tag{4.10}$$

Equation (4.10) has a nontrivial solution only at $z = -\epsilon$.

V. NUMERICAL RESULTS

The cubic B-splines were evaluated using the iterative technique of Cox.[10] On the set of points $\{x_i\}$ we define

$$B_{ni}^{(1)}(x) = \begin{cases} (x_i - x_{i-1})^{-1}, & x_{i-1} \leqslant x \leqslant x_i, \\ 0, & \text{otherwise.} \end{cases} \tag{5.1}$$

The B-spline of degree m is constructed from the recursion relation

$$B_{ni}^{(m)}(x) = \frac{(x - x_{i-m})B_{ni-1}^{(m-1)}(x) + (x_i - x)B_{ni}^{(m-1)}(x)}{x_i - x_{i-m}}. \tag{5.2}$$

In this notation the cubic B-spline is given by $B_{ni}(x) = B_{ni+2}^{(4)}(x)$. Cox has shown that this procedure is both fast and accurate.

In the case of uniformly spaced knots, the matrix $B_{ij} = (B_{ni},B_{nj})$ is symmetric and has the following special form[16]:

$$\begin{bmatrix} B_{00} & B_{01} & B_{02} & B_{03} & 0 & & & \\ B_{01} & B_{11} & B_{12} & B_{13} & B_{03} & & & \\ B_{02} & B_{12} & B_{22} & B_{23} & B_{13} & & & \\ B_{03} & B_{13} & B_{23} & B_{33} & B_{23} & & & \\ 0 & B_{03} & B_{13} & B_{23} & & & & \\ & & & & B_{33} & B_{23} & B_{13} & B_{03} \\ & & & & B_{23} & B_{22} & B_{12} & B_{02} \\ & & & & B_{13} & B_{12} & B_{11} & B_{01} \\ & & & & B_{03} & B_{02} & B_{01} & B_{00} \end{bmatrix} \tag{5.3}$$

Here each element along the diagonal, indicated by the solid

TABLE I. Error in the largest eigenvalue showing solution of the Galerkin method with n knots. The exact eigenvalue is 1.0.

n	Error
3	1.4(− 4)
4	9.2(− 5)
5	5.5(− 7)
6	4.8(− 6)
7	5.8(− 8)
8	1.1(− 8)

line, is a constant along that diagonal. All the matrix elements are therefore determined by the 4×4 matrix B_{ij}, where $i,j = 0,1,2,3$. We obtain for the elements of this matrix

$$\frac{1}{80\,640h}\begin{bmatrix} 20 & 129 & 60 & 1 \\ & 1208 & 1062 & 120 \\ & & 2396 & 1191 \\ & & & 2416 \end{bmatrix}, \tag{5.4}$$

where h is the spacing between the knots.

As a numerical example for the integral equations, we consider a two-particle system with the following parameters: we choose the binding energy $\epsilon = 0.0536\ 95\ \text{fm}^{-2}$ and range parameter $\beta = 1.444\ 01\ \text{fm}^{-1}$. These parameters are chosen so that the Yamaguchi potential in Eq. (3.13) will approximately describe low energy neutron–proton (n–p) scattering in the s-wave spin-1 channel.

We now test the utility of the approximation methods described in the previous section. We use only uniformly spaced knots over the interval $[-1, +1]$.

We first solve the eigenvalue problem given by Eq. (4.10). The largest eigenvalue for the exact solution is 1.0.

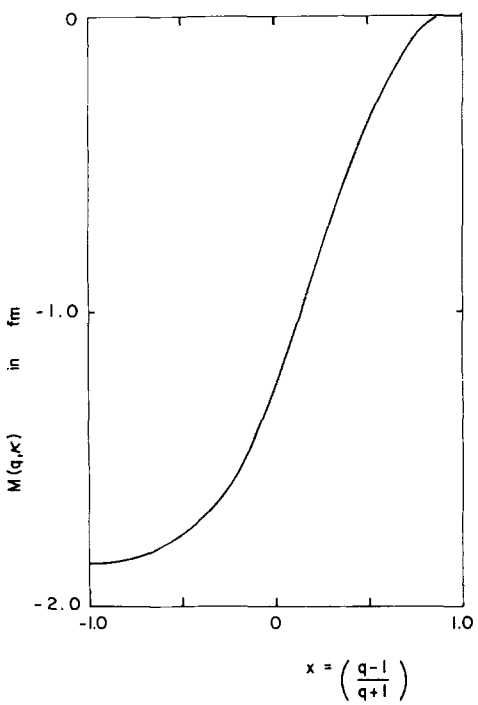


FIG. 1. K matrix given by Eq. (4.2) at a scattering energy of $\kappa^2 = 1\ \text{fm}^{-2}$.

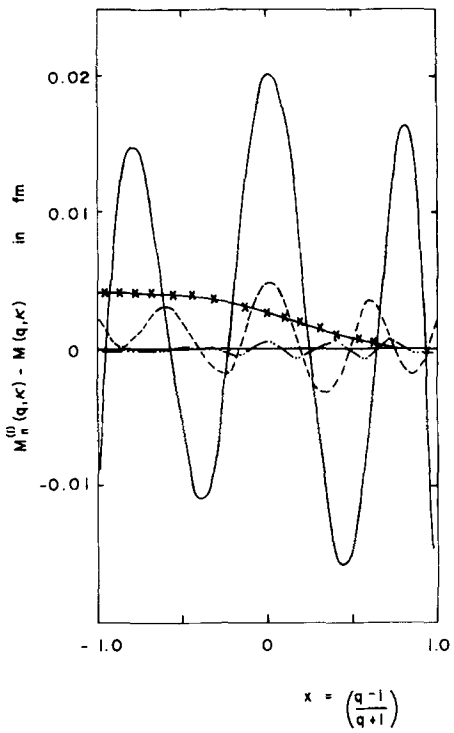


FIG. 2. Approximate K-matrix solutions using the Galerkin method. Shown is the error function $M_n^{(i)}(q,\kappa) - M(q,\kappa)$ for the cubic B-spline interpolate with n knots. The —— curve is for $n = 4$, the ----- curve for $n = 6$, and the ——·——·—— curve for $n = 8$. The xxxx curve is the result for $n = 4$ after a single iteration $i = 2$.

Table I shows the error in the largest computed eigenvalue using the Galerkin method with n knots.

Figure 1 illustrates the scattering solution given by Eq. (3.14) at a scattering energy of $\kappa^2 = 1\ \text{fm}^{-2}$. Figure 2 illustrates the difference between the K matrix, shown in Fig. 1, and the calculated value determined by the Galerkin procedure. The error function is shown for various numbers of knots. Also shown in Fig. 2 is the error function obtained for 4 knots after applying one iteration of the iterative improvement scheme. The iterative improvement, given by Eqs. (4.7) and (4.8), requires little additional computation.

To further illustrate the convergence properties of our approximation, we tabulate in Table II the $L_2[-1,1]$-norm of the error function. [The L_2-norm for a function $f(x)$, $x\in[a,b]$, is given by $(f,f)^{1/2}$, where the inner product is defined in Eq. (2.5).] The results verify a theoretical convergence rate of $O(h^4)$. The iterative improvement yields approximately 1 order of magnitude reduction in the L_2-norm of the error function for each value of n.

TABLE II. L 2-norm of the error function, $M_n^{(i)}(q,\kappa) - M(q,\kappa)$, for $i = 1$ and 2 with n knots at a scattering energy $\kappa^2 = 1\ \text{fm}^{-2}$.

n	$\|M_n^{(1)} - M\|_{L2}$	$\|M_n^{(2)} - M\|_{L2}$
4	2.8(− 4)	1.6(− 5)
6	1.0(− 5)	1.7(− 6)
8	2.2(− 7)	1.4(− 8)
10	2.2(− 8)	6.4(− 9)

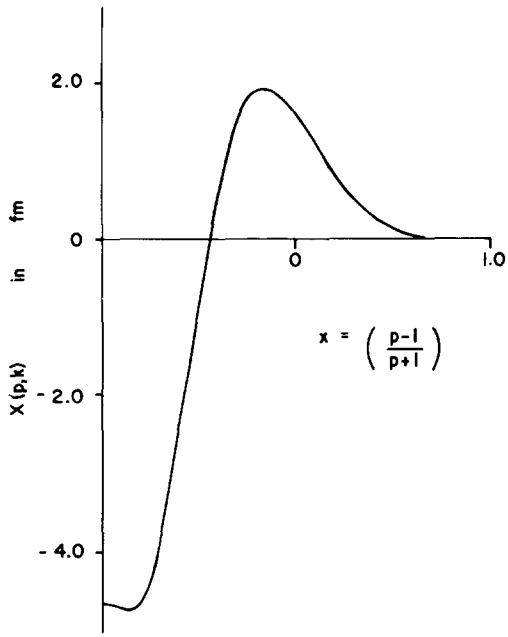


FIG. 3. K-matrix solution of the Amado–Lovelace equation at scattering energy $(k/k_B)^2 = 0.8$.

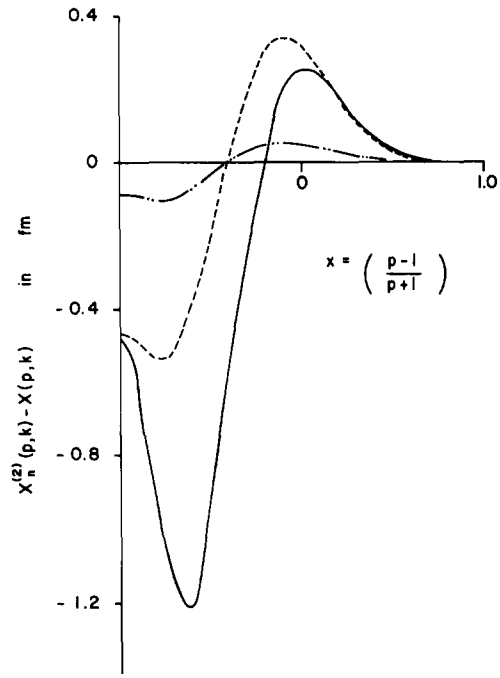


FIG. 5. Approximate solution of the Amado–Lovelace equation using the collocation method. Error function and curves as described in Fig. 4.

We now describe the behavior of Galerkin method solutions of the Amado–Lovelace equation in the elastic scattering sector. The two-particle parameters, ϵ and β, are chosen to fit the n–p system and taken on the values given above. As a specific example, we fix the scattering energy at $(k/k_B)^2 = 0.8$, where $k_B = 2\alpha/\sqrt{3} = 0.2676\,\mathrm{fm}^{-1}$ is the momentum at breakup threshold.

To obtain a reference solution of the Amado–Lovelace equation, we have used the well-known Kowalski–Noyes[17] procedure. A 24-point Gaussian quadrature was used to obtain the reference solution shown in Fig. 3. Notice that the half-shell K matrix $X(p,k)$ has a node and, therefore, exhibits a more complicated structure than the problem described in our previous example.

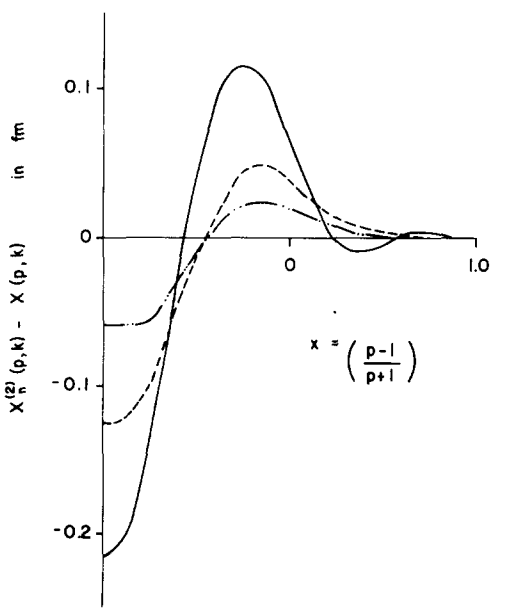


FIG. 4. Approximate solution of the Amado–Lovelace equation using the Galerkin method. The error function $X_n^{(2)}(p,k) - X(p,k)$ for the cubic B-spline with n knots after a single iteration $i = 2$. The —— curve is for $n = 4$, the ----- curve for $= 6$, and the ——·····—— curve for $n = 8$.

Figure 4 illustrates the difference between the approximate solution, after applying the iterative improvement scheme, and the reference solution. Only a relatively small number of mesh points has been used in the Galerkin procedure.

It is interesting to make a numerical comparison between the Galerkin method (described in this paper) and the method of collocation described in I. For this purpose, we show in Fig. 5 the results obtained from the collocation method after applying the iterative improvement scheme. We remark that the iterative improvement scheme, defined for all orders by Eq. (2.24), is the same for both methods. The difference between the solutions shown in Figs. 4 and 5 results from the way in which the coefficients defined by the expansion in Eq. (4.3) are calculated.

VI. CONCLUSIONS

We have obtained numerical solutions for both two- and three-body problems using the Galerkin method with cubic spline trial functions. The convergence rate of our solutions is consistent with a convergence rate of $O(h^4)$ for the theoretical error bound. For the method to be useful in practice, however, it is not sufficient to demonstrate a fast convergence. Efficiency of the method is determined by the size of the linear system that must be used for an accurate solution. We have demonstrated, by numerical solution of the Amado–Lovelace equation for three identical bosons, that accurate solutions can be obtained using this method with a small number of mesh points.

Comparison of these results with the method of collocation has shown, in general, that for a given number of knots the Galerkin method gives a better result. This is to be expected as the set $\{\tau_i\}$ of collocation points found in Sec. II are "best" in some sense. One can consider the Galerkin technique as a collocation procedure where the collocation points are optimal in the last-squares sense. This advantage of the Galerkin method over the collocation method is offset by the fact that the Galerkin method requires evaluation of a two-dimensional integral for the matrix elements $(\mathscr{K}(\kappa)B_{ni}, B_{nj})$ in Eq. (4.6). For this reason it is desirable to obtain accurate solutions with small linear systems.

To obtain our results, we have used uniformly spaced knots. Improved results for a given number of knots n should be possible if these knots are concentrated in regions where the solution function has more structure. A study of this optimization problem is deferred to future work.

[1]L. D. Faddeev, *Mathematical Aspects of the Three-Body Problem in Quantum Scattering Theory*, Steklov Math. Institute **69** (1963) (Davey, New York, 1965).
[2]K. E. Atkinson, *A Survey of Numerical Methods for the Solution of Fredholm Integral Equations of the Second Kind* (SIAM, Philadelphia, 1976).
[3]C. de Boor and B. Swartz, SIAM J. Numer. Anal. **10**, 582–606 (1973).
[4]Y. E. Kim, J. Math. Phys. **10**, 1491–4 (1969); W. Glockle and R. Offerman, Phys. Rev. C **16**, 2039–49 (1977).
[5]F. Sohre and H. Ziegelmann, Phys. Lett. B **34**, 579–80 (1971); E. Harms, Phys. Lett. B **41**, 26–8 (1973); Nancy M. Larson and J. H. Heatherington, Phys. Rev. C **9**, 699–708 (1974).
[6]H. R. Fiebig, Comp. Phys. Comm. **23**, 135–143 (1981); **20**, 181–9 (1980).
[7]J. Horácek and L. Malina, Czech. J. Phys. B **27**, 134–49 (1972).
[8]T. A. Osborn and D. Eyre, Nucl. Phys. A **327**, 125–38 (1979).
[9]M. Brannigan and D. Eyre, J. Math. Phys., to be published.
[10]M. G. Cox, J. Inst. Math. Appl. **10**, 134–49 (1972).
[11]M. J. D. Powell, *Approximation Theory and Methods* (Cambridge U.P., Cambridge, 1981).
[12]I. H. Sloan, Math. Comp. **30**, 758–64 (1976).
[13]C. Lovelace, Phys. Rev. **135**, B1225–49 (1964); R. D. Amado, Ann. Rev. Nucl. Sci. **19**, 61–98 (1969).
[14]Y. Yamaguchi, Phys. Rev. **95**, 1628–34 (1954).
[15]B. Noble and B. Beighton, J. Inst. Math. Appl. **26**, 431–46 (1980).
[16]G. Hämmerlin and L. L. Schumaker, Numer. Math. **34**, 125–41 (1980).
[17]K. L. Kowalski, Phys. Rev. Lett. **15**, 798–800 (1965); H. P. Noyes, Phys. Rev. Lett. **15**, 538–40 (1965).

Low-energy sum rules for two-particle scattering[a)]

D. Bollé[b)]
Instituut voor Theoretische Fysica, Universiteit Leuven, B-3030 Leuven, Belgium

S. F. J. Wilk
Department of Physics, University of Manitoba, Winnipeg, Manitoba, R3T 2N2 Canada

(Received 29 June 1982; accepted for publication 17 December 1982)

We consider the low-energy expansion of the operators describing a two-particle scattering system and derive recursion relations for the expansion coefficients. The particles of the system are allowed to interact via general (not necessarily spherically symmetric) potentials and can also form zero-energy bound states and/or resonances. Using these expansions, we prove low-energy sum rules for such a system. They relate the negative-energy moments of the trace of the resolvent difference to the bound-state eigenvalues. As an application, we first discuss the low-energy expansion of the two-body total cross section. In particular, we derive a closed expression for the expansion coefficients. We also find that there is always a finite characteristic length in the problem, even in the presence of zero-energy resonances. Secondly, we study the low-temperature behavior of the second virial coefficient.

PACS numbers: 03.80. + r, 11.20.Dj, 11.50.Li, 03.65.Nk

I. INTRODUCTION

Sum rules for the two-body scattering problem have been extensively studied. The most well-known example is Levinson's theorem,[1] which states that for a given partial wave l, the value of the phase shift at zero energy is proportional to the number of bound states having angular momentum l. Extensions of these results include the study of higher-order energy moments of the phase shift[2–6] and a discussion of Levinsons's theorem and higher-order sum rules for general (not necessarily spherically symmetric) potentials.[7–13] Here the partial-wave phase shift has to be replaced by a global quantity describing the scattering phenomena in the system, like the logarithmic derivative of the determinant of the S-matrix, the time delay, the imaginary part of the trace of the resolvent difference. Briefly, these papers provide us with a systematic study of the positive moments for general potentials allowing also zero-energy resonances and/or bound states.

Concerning the sum rules for negative moments, there is no such systematic treatment for these general potentials. The partial wave case for central potentials was analyzed in Refs. 14 and 15. In Ref. 16, the first negative-moment rule has been derived for a rather restricted class of noncentral potentials, remarking that a detailed and systematic treatment of the low-energy behavior of the trace of the resolvent difference for general potentials was not available. (Note that in these papers it is assumed that there are no zero-energy eigenstates in the problem.) Recently, however, there appeared some new discussions of the low-energy behavior of the Schrödinger operator,[17,18] This inspired us to look again at the (negative) moment problem for general two-particle systems.

Our aim in this paper is twofold. First, we want to discuss the low-energy expansion for the trace of the resolvent difference, allowing zero-energy eigenstates of the two-particle Hamiltonian. In particular, we would like to obtain a closed expression for the expansion coefficients. We then want to use these results to prove sum rules for the negative (integer and half-integer) energy moments. Secondly, we want to stress again the usefulness of these results in the discussion of some physical quantities. We consider two particular applications, namely, the low-energy behavior of the total cross section and the low-temperature behavior of the second virial coefficient. Thereby, we generalize existing results in the literature.

A systematic study of negative-moment sum rules requires the low-energy expansion of the trace of the resolvent difference (or any other global quantity describing the scattering phenomena in the system). The first few coefficients of such an expansion can be obtained from Refs. 17 and 18. (In the following, we will refer to the work of Albeverio, Gesztesy, and Høegh-Krohn as AGH). Of course, these coefficients depend on a case distinction according to possible zero-energy bound states or resonances. To be able, in principle, to compute explicitly all coefficients in the various cases, we derive a closed expression for them on the basis of some of the results of AGH and the use of the integral equation for the t-matrix. The structure of the recursion relations is similar in all cases. This forms the content of Sec. II.

In Sec. III we derive the N th negative (N is allowed to be zero) moment sum rule by employing contour integration over the spectrum of the Hamiltonian. These rules relate the energy-integral of the N th negative energy moment of the trace of the imaginary part of the resolvent difference, to the N th negative power of the bound-state energies. A number of low-energy subtractions, depending on N and determined on the basis of the results in Sec. II are needed in the integral. A surface term is also present. We then extend these rules to allow complex powers of the moments. In particular, we present the form of the half-integer negative-moment rules. Finally, we discuss in more detail the subtractions and the physical meaning of the $N = 0$, $N = -1/2$, and $N = -1$ result.

[a)] Work supported in part by a grant from the Natural Sciences and Engineering Research Council, Canada.
[b)] Onderzoeksleider N. F. W. O., Belgium.

 0022-2488/83/061555-09$02.50 © 1983 American Institute of Physics

The theoretical and practical importance of the study of the low-energy behavior of nonrelativistic quantum mechanics is discussed in the Introduction of AGH[17] (see also the references cited there). In Sec. IV we consider two applications of our results, which underline again this importance. They also generalize well-known results in scattering theory. The first one is a detailed discussion of the low-energy behavior of the two-body total cross section for scattering systems with our without zero-energy eigenstates. We present a closed expression for the coefficients in its low-energy expansion and study in detail the first few of them. As a result we find that there is always a finite characteristic length in the system, even if there are zero-energy resonances present. It reduces to the scattering length when there are no zero-energy eigenstates. Secondly, we discuss the low-temperature behavior of the second virial coefficient for a Boltzmann gas. The negative energy bound states give the most important contributions. Then we get a contribution proportional to the number of zero-energy bound states and resonances, counting degeneracy. Finally, the leading scattering contribution is always given by the characteristic length in the system.

II. LOW-ENERGY EXPANSION OF TWO-PARTICLE OPERATORS

In this section we discuss the low-energy expansion for the two-body transition operator and for the trace of the two-body resolvent difference. The results obtained here are inspired by Refs. 17, 18, and 10 but go further in the sense that recursion relations are obtained for the coefficients appearing in these expansions.

Let us begin by introducing some notations and definitions. In the Hilbert space $\mathscr{H} = L^2(\mathbb{R}^3)$, we introduce the two-particle Hamiltonian

$$H = H_0 + V, \tag{2.1}$$

where $H_0 = -\Delta$ denotes the usual kinetic-energy operator (we use natural units such that $\hbar^2 = 2\mu = 1$, where μ is the two-body reduced mass). The potential V is assumed to be a real measurable function on $\mathbb{R}^3$ in the Rollnik class R, i.e.,

$$\|V\|_R^2 = \int d^3\mathbf{x}\, d^3\mathbf{y}\, \frac{|V(\mathbf{x})||V(\mathbf{y})|}{|\mathbf{x}-\mathbf{y}|^2} < \infty. \tag{2.2}$$

Using the notations

$$R(z) = (H-z)^{-1}, \quad R_0(z) = (H_0 - z)^{-1}, \tag{2.3}$$

with z the two-body complex energy, and

$$v(\mathbf{x}) = |V(\mathbf{x})|^{1/2}, \quad u(\mathbf{x}) = |V(\mathbf{x})|^{1/2}\,\mathrm{sgn}\, V(\mathbf{x}), \tag{2.4}$$

we obtain from the resolvent (Lippmann–Schwinger) equation

$$R(z) - R_0(z) = -R_0(z)vT(z)uR_0(z), \tag{2.5}$$

where the operator $T(z)$ is given by

$$T(z) = (uR_0(z)v + 1)^{-1}. \tag{2.6}$$

This operator satisfies the integral equation

$$T(z) = 1 - uR_0(z)vT(z), \tag{2.7}$$

where the kernel $uR_0(z)v$ is Hilbert–Schmidt[19] in Π_η, with Π_η defined as the set of points in the z-plane a distance η or greater away from the positive real axis.

We are interested in the low-energy expansion of $T(z)$. Such an expansion has been discussed by AGH[17] (see also Ref. 18). It turns out that its explicit form strongly depends on whether H has a zero-energy resonance and/or bound state or not. AGH then derive the first few expansion coefficients in the different cases. We go one step further in the sense that we derive recursion relations which allow us, in princple, to compute explicitly all the coefficients in the various cases.

We start by defining the different cases (see Refs. 10, 17, 18, and 20). The zero-energy resonances and bound states are solutions of

$$uR_0(0)v\phi_j = -\phi_j, \quad j = 1,2,\ldots,N_0, \tag{2.8}$$

for some $\phi_j \in L^2(\mathbb{R}^3)$. Since $uR_0(0)v$ is Hilbert–Schmidt, N_0 is necessarily finite. The functions ψ_j defined by

$$\psi_j(\mathbf{x}) \equiv -(R_0(0)v\phi_j)(\mathbf{x}), \quad j = 1,2,\ldots,N_0$$
$$\phi_j(\mathbf{x}) = (u\psi_j)(\mathbf{x}), \tag{2.9}$$

are zero-energy resonance functions. They satisfy

$$H\psi_j = (-\Delta + V)\psi_j = 0, \quad j = 1,2,\ldots,N_0, \tag{2.10}$$

in the sense of distributions. In general, they need not be in $L^2(\mathbb{R}^3)$. We then have

Case I: There exists no resonance function ψ_j or, equivalent, -1 is not an eigenvalue of $uR_0(0)v$.

Case II: There exist only one resonance function ψ [i.e., -1 is a simple eigenvalue of $uR_0(0)v$] and ψ is not in $L^2(\mathbb{R}^3)$.

Case III: There exist $N_0 \geqslant 1$ resonance functions ψ_j, $j = 1,\ldots,N_0$ which are all in $L^2(\mathbb{R}^3)$

Case IV: There exist $N_0 \geqslant 2$ resonance functions ψ_j, $j = 1,\ldots,N_0$ and at least one of them is not in $L^2(\mathbb{R}^3)$.

Next, we recall the following results.

Proposition 2.1: Assume $V \in L^2(\mathbb{R}^3) \cap R$ and let $\psi(\mathbf{x})$ be a zero-energy resonance function satisfying Eqs. (2.8) and (2.9). Then $\psi \in L^2_{\mathrm{loc}}(\mathbb{R}^3)$. If in addition $|\mathbf{x}|\,V(\mathbf{x}) \in L^1(\mathbb{R}^3)$, then $\psi \in L^2(\mathbb{R}^3)$ is equivalent to

$$(v,\phi) = \int d^3x\, V(\mathbf{x})\psi(\mathbf{x}) = 0. \tag{2.11}$$

Proof: See AGH.[17] ■

For other proofs of this result with alternative conditions on V, we refer to Refs. 10 and 20.

Lemma 2.1: Let $V \in L^1(\mathbb{R}^3) \cap R$. Then $(uR_0(0)v + 1 + \epsilon)^{-1}$ has a norm convergent Laurent expansion around $\epsilon = 0$, viz.,

$$(uR_0(0)v + 1 + \epsilon)^{-1} = \frac{P}{\epsilon} + \sum_{m=0,\infty} (-\epsilon)^m T_0^{m+1}, \tag{2.12}$$

where P is the projection onto the N_0-dimensional eigenspace of $uR_0(0)v$ to the eigenvalue -1 and is given by

$$P = \sum_{j=1,N_0} \frac{(\tilde{\phi}_j,\cdot)\phi_j}{(\tilde{\phi}_j,\phi_j)}, \quad N_0 = \dim P, \tag{2.13}$$

where ϕ_j are the solutions of Eq. (2.8) and

$$\tilde{\phi}_j = (\mathrm{sgn}\, V)\phi_j, \tag{2.14}$$

$$(\tilde{\phi}_j,\phi_l) = 0, \quad j \neq l, \quad j = 1,\ldots,N_0, \tag{2.15}$$

and T_0 is a bounded operator given by

$$T_0 = \frac{1}{2\pi i}\oint_C dz(z+1)^{-1}(uR_0(0)v - z)^{-1}$$
$$= n - \lim_{\epsilon \to 0^+} (uR_0(0)v + 1 + \epsilon)^{-1}Q, \quad Q = 1 - P. \tag{2.16}$$

Here the contour C goes around the isolated eigenvalue -1 of $uR_0(0)v$ counterclockwise.[21,22]

Proof: See AGH.[17] ■

In the following, we will abbreviate $\Sigma_{j=1,N_0}$ by Σ_j.

Lemma 2.2: Let $e^{2a|x|}V(\mathbf{x}) \in R$. Then $uR_0(z)v$ is analytic in z for $z \in \Pi_\eta$ and

$$uR_0(z)v = \sum_{n=0,\infty} (i\sqrt{z})^n r_n, \tag{2.17}$$

where the r_n are Hilbert–Schmidt operators with the kernel representation

$$(r_n f)(\mathbf{x}) = \frac{1}{4\pi n!}\int d^3y\, u(\mathbf{x})|\mathbf{x}-\mathbf{y}|^{n-1}v(\mathbf{y})f(\mathbf{y}), \quad f \in L^2(\mathbb{R}^3). \tag{2.18}$$

Proof: See, e.g., Ref. 18. ■

The main results of this section can then be stated. We begin with

Theorem 2.1: Let $e^{2a|x|}V(\mathbf{x}) \in R$ for some $a > 0$, and assume Case I to IV.

Let $PT(z) = P(z)$ and $QT(z) = Q(z)$. Then

$$P(z) = \sum_{n=-2,\infty} (i\sqrt{z})^n p_n, \tag{2.19}$$

$$Q(z) = \sum_{n=0,\infty} (i\sqrt{z})^n q_n, \tag{2.20}$$

where the coefficients p_n and q_n satisfy the recursion relations

$$p_n = a_n - \sum_{\substack{k=0,n+1 \\ l=0,n+1-k}} a_{-2+k} r_{n+4-k-l} p_{-2+l}$$
$$- \sum_{\substack{k=0,n+1 \\ l=0,n+1-k}} a_{-2+k} r_{n+2-k-l} q_l, \tag{2.21}$$

$$q_n = T_0\delta_{n0} - \sum_{k=0,n+1} T_0 r_{n+2-k} p_{-2+k}$$
$$- \sum_{k=0,n-1} T_0 r_{n-k} q_k, \tag{2.22}$$

with T_0 given by Eq. (2.16) and

$$a_{-2} = \sum_{i,j} (\tilde{\phi}_j,\cdot)\phi_i (\tilde{\phi}, r_2\phi)^{-1}_{ij} + \frac{1}{b} a_{-1}, \tag{2.23}$$

$a_n = 0$ for $n \geqslant -1$ in Case III,

$$= \frac{b^{n+3}}{4\pi}\sum_{i,j} (\tilde{\phi}_j,\cdot)\phi_i c_i c_j^* \quad \text{for } n \geqslant -1 \text{ in the other cases.} \tag{2.24}$$

Finally,

$$b = -4\pi\left\{\sum_{i,j} (v,\phi_i)(\tilde{\phi}, r_2\phi)^{-1}_{ij}(\tilde{\phi}_j,u)\right\}^{-1}, \tag{2.25}$$

$$c_i = \sum_k (\tilde{\phi}, r_2\phi)^{-1}_{ik}(\tilde{\phi}_k,u), \tag{2.26}$$

$$c_j^* = \sum_k (v,\phi_k)(\tilde{\phi}, r_2\phi)^{-1}_{kj}. \tag{2.27}$$

Here $(\tilde{\phi}, r_2\phi)^{-1}_{ij}$ denotes the inverse matrix of $(\tilde{\phi}_i, r_2\phi_j)$.

Proof: From Ref. 17 (Theorems 3.1–3.4), we infer that an expansion of the type

$$T(z) = \sum_{n=-2,\infty} (i\sqrt{z})^n t_n \tag{2.28}$$

is possible in all cases. The coefficients t_n are, of course, case dependent. Introducing $P(z) = PT(z)$ and $Q(z) = QT(z)$, we know that all the possible zero-energy singularities are contained in $P(z)$ and that $Q(z)$ is regular as $z \to 0^+$. Therefore, the expansions (2.19) and (2.20) are valid. Furthermore, from Eq. (2.7) we get the following set of coupled equations:

$$P(z) = P - PuR_0(z)vP(z) - PuR_0(z)vQ(z), \tag{2.29}$$

$$Q(z) = Q - QuR_0(z)vP(z) - QuR_0(z)vQ(z). \tag{2.30}$$

Next, employing the explicit form of the projection operator P [Eq. (2.13)] together with Eq. (2.8), we write

$$PuR_0(z)vP = -P + i\sqrt{z}\sum_{j,k} \frac{(\tilde{\phi}_k,\cdot)\phi_j}{(\tilde{\phi}_j,\phi_j)(\tilde{\phi}_k,\phi_k)}$$
$$\times\left[\frac{(\tilde{\phi}_j,u)(v,\phi_k)}{4\pi} + i\sqrt{z}\,(\tilde{\phi}_j, r_2\phi_k)\right]$$
$$+ P\tilde{R}_0^{(3)}(z)P, \tag{2.31}$$

where

$$\tilde{R}_0^{(n)} = \sum_{j=n,\infty} (i\sqrt{z})^j r_j. \tag{2.32}$$

Using all this information in Eqs. (2.29) and (2.30) and knowing that $(1 + uR_0(0)v)Q$ is invertible (Lemma 2.1), we arrive at

$$i\sqrt{z}\left\{\sum_{j,k} \frac{(\tilde{\phi}_k,\cdot)\phi_j}{(\tilde{\phi}_j,\phi_j)(\tilde{\phi}_k,\phi_k)}\left[\frac{(\tilde{\phi}_j,u)(v,\phi_k)}{4\pi} + i\sqrt{z}\,(\tilde{\phi}_j, r_2\phi_k)\right]\right\} P(z)$$
$$= P - P\tilde{R}_0^{(3)}(z)P(z) - P\tilde{R}_0^{(1)}(z)Q(z), \tag{2.33}$$

$$Q(z) = T_0 - T_0\tilde{R}_0^{(1)}(z)P(z) - T_0\tilde{R}_0^{(1)}(z)Q(z). \tag{2.34}$$

We can then invert the rank N_0 operator appearing in Eq. (2.33) between the brackets { } on the N_0-dimensional eigenspace of $uR_0(0)v$, since the determinant of the matrix between the square brackets [ ] is different from zero. The result, $\tilde{P}_{-2}(z)$, is given by

$$\tilde{P}_{-2}(z) = \frac{1}{(i\sqrt{z})^2}\sum_{j,k} (\tilde{\phi}_k,\cdot)\phi_j\left[(\tilde{\phi}, r_2\phi)^{-1}_{jk} + \frac{b}{4\pi}\frac{c_j c_k^*}{1 - bi\sqrt{z}}\right], \tag{2.35}$$

where b, c_j, and c_k^* are defined by Eqs. (2.25)–(2.27). In Case III, where all $(v,\phi_j) = 0, j = 1,2,\ldots,N_0$, it is straightforward to see that the second term in Eq. (2.35) is absent. Thus in all the cases we can write down the following expansion for $\tilde{P}_{-2}(z)$:

$$\tilde{P}_{-2}(z) = \sum_{n=-2,\infty} (i\sqrt{z})^n a_n, \tag{2.36}$$

where the a_n are given by Eqs. (2.23) and (2.24). Furthermore, the coupled set of Eqs. (2.33) and (2.34) now reads

$$P(z) = \tilde{P}_{-2}(z) - \tilde{P}_{-2}(z)\tilde{R}_0^{(3)}(z)P(z) - \tilde{P}_{-2}(z)\tilde{R}_0^{(1)}(z)Q(z), \tag{2.37}$$

$$Q(z) = T_0 - T_0\tilde{R}_0^{(1)}(z)P(z) - T_0\tilde{R}_0^{(1)}(z)Q(z). \tag{2.38}$$

Inserting the expansions (2.19), (2.20), (2.32), and (2.36) in this set of equations, we obtain after some lengthy but straightforward manipulations the recursion relations (2.21) and (2.22). ■

It is of course clear that in case I the general result (2.21) and (2.22) can be much simplified since $P=0$ and $Q=1$. We then immediately get

Case I:

$$t_n = t_0\delta_{n0} - \sum_{k=0,n-1} t_0 r_{n-k} t_k \quad (T_0 = t_0). \tag{2.39}$$

In Case II, where there is only one function ϕ and $(v,\phi)\neq 0$, we see from Eqs. (2.23)–(2.27) that

$$b = \frac{-4\pi(\tilde{\phi},r_2\phi)}{|(v,\phi)|^2}, \tag{2.40}$$

and

$$a_{-2} = 0,$$

$$a_n = b^{n+1}\frac{4\pi(\tilde{\phi},\cdot)\phi}{|(v,\phi)|^2} \quad \text{for } n\geqslant -1. \tag{2.41}$$

Consequently, the recursion relations (2.21) and (2.22) can be written as

Case II:

$$p_n = b^{n+1}a_{-1} - \sum_{\substack{k=0,n-1\\ l=0,k}} b^{n-k-1}a_{-1}r_{3+k-l}p_{-1+l} - \sum_{\substack{k=0,n\\ l=0,k}} b^{n-k}a_{-1}r_{1+k-l}q_l, \tag{2.42}$$

$$q_n = T_0\delta_{n0} - \sum_{k=0,n} T_0 r_{n+1-k}p_{-1+k} - \sum_{k=0,n-1} T_0 r_{n-k}q_k. \tag{2.43}$$

Remark that in this case it suffices, strictly speaking, to separate out only the term $i\sqrt{z}\,[4\pi(\tilde{\phi},r_2\phi)]^{-1}\,|(v,\phi)|^2(\tilde{\phi},\cdot)\phi$ in Eq. (2.31), the reason being that this rank 1 operator is already invertible since $(v,\phi)\neq 0$. Doing this, we arrive at an equation for p_n which has a somewhat different structure than Eq. (2.42) but which gives the same coefficients p_n, as it should be.

We have already indicated in the course of the proof of Theorem 2.1 that only the coefficient a_{-2} survives in Case III [see Eqs. (2.23)–(2.25)]. Consequently, the recursion relations become

Case III:

$$p_n = a_{-2}\delta_{n,-2} - \sum_{k=0,n+1} a_{-2}r_{n+4-k}p_{-2+k} - \sum_{k=0,n} a_{-2}r_{n+2-k}q_k, \tag{2.44}$$

$$q_n = T_0\delta_{n0} - \sum_{k=0,n+1} T_0 r_{n+2-k}p_{-2+k} - \sum_{k=0,n-1} T_0 r_{n-k}q_k. \tag{2.45}$$

In this and all other cases, we have verified that the coefficients given by AGH[17] [for $\lambda(\epsilon)=1$] satisfy these recursion relations.

Finally, we have

Proposition 2.2: Let $e^{2a|\mathbf{x}|}V(\mathbf{x})\in R$ for some $a>0$ and assume Cases I to IV. Then the following expansion holds

$$\mathrm{Tr}[R(z)-R_0(z)] = \frac{1}{2}\sum_{n=-2,\infty}(-i\sqrt{z})^n\Delta_{n+1}, \tag{2.46}$$

where the coefficients Δ_n are real and given by

$$\Delta_n = (-1)^{n+1}\sum_{j=0,n-s}(j+1)\mathrm{Tr}(r_{j+1}t_{n-j}), \tag{2.47}$$

with $s=0$ in Case I, $s=-1$ in Case II, and $s=-2$ in the other cases (III and IV). Here $\mathrm{Tr}(r_n t_m)$ in Eq. (2.47) formally denotes the integral over the forward matrix elements of $r_n t_m$, viz.,

$$\mathrm{Tr}(r_n t_m) \equiv \int d^3x\,(r_n t_m)(\mathbf{x},\mathbf{x}) \quad \text{for any } n,m. \tag{2.48}$$

(We will keep this notation throughout the rest of the paper.) The coefficients r_n are given by Eq. (2.18). The coefficients t_n are case dependent and given by the recursion relations (2.39), respectively, (2.42) and (2.43); (2.44) and (2.45); (2.21) and (2.22).

Proof: For $V\in L^1\cap R$ we know that $uR_0(z)$ is Schmidt class and that the difference $R(z)-R_0(z)$ is trace class.[19] So from Eq. (2.5) we infer, by employing the cyclic invariance property of the trace,

$$\begin{aligned}\mathrm{Tr}[R(z)-R_0(z)] &= -\mathrm{Tr}[uR_0^2(z)vT(z)]\\ &= -\mathrm{Tr}\left[u\left(\frac{d}{dz}R_0(z)\right)vT(z)\right].\end{aligned} \tag{2.49}$$

For $e^{2a|\mathbf{x}|}V(\mathbf{x})\in R$ we can then use expression (2.17) and the expansions (2.19) and (2.20) with the appropriate coefficients to obtain, after straightforward manipulations,

$$\mathrm{Tr}[R(z)-R_0(z)] = \frac{1}{2}\int d^3x \sum_{n=s-1,\infty}(i\sqrt{z})^n \times \sum_{j=0,n+1-s}(j+1)(r_{j+1}t_{n+1-j})(\mathbf{x},\mathbf{x}), \tag{2.50}$$

where $s=0$ in Case I, $s=-1$ in Case II and $s=-2$ in Cases III and IV. The coefficients t_{n+1-j} are case dependent and given by the respective recursion relations. It is easy to check that $n=s-1,\infty$ may be replaced by $n=-2,\infty$ because $\Delta_{-1}=0$ in Case I and $\Delta_{-2}=0$ in Cases III and IV. Furthermore, we assume that the interchange of the summation over n and the integral over $\mathbf{x}$ is possible in some region around $z=0$. (Remark that for our conditions on the potential, the expansions in $\sqrt{z}$ we have used are Taylor or Laurent expansions around the zero-energy limit.) Finally, to prove that the Δ_n are real, we first infer from Eqs. (2.17) and (2.18) that the r_j are unitarily equivalent to their adjoint, viz.,

$$r_j = (\mathrm{sgn}\,V)r_j^\dagger(\mathrm{sgn}\,V). \tag{2.51}$$

Furthermore, using Eq. (2.7), we arrive at

$$T^\dagger(z) = (\mathrm{sgn}\,V)T^*(z)(\mathrm{sgn}\,V), \tag{2.52}$$

where, of course

$$T^\dagger(z) = \sum_{n=-2,\infty}(-i\sqrt{z^*})^n t_n^\dagger. \tag{2.53}$$

Consequently,

$$t_n^\dagger = (\operatorname{sgn} V) t_n (\operatorname{sgn} V). \tag{2.54}$$

Using this information in Eq. (2.47), we immediately see that Δ_n is real. ■

The sign conventions in Eqs. (2.46) and (2.47) are chosen such that the Δ_n correspond to physical quantities, as will become clear later on. The use of the strong condition $e^{2a|x|}V(\mathbf{x}) \in R$ throughout this section leads to strong results in the sense that the expansions we get for $T(z)$ and $\operatorname{Tr}[R(z) - R_0(z)]$ are Taylor or Laurent expansions. If we only assume $V \in R$ and $(1 + |\mathbf{x}|)^n V(\mathbf{x}) \in L^1(\mathbb{R}^3)$, we then obtain asymptotic expansions, the order of which depends on the value of n (see also Refs. 17 and 18).

III. NEGATIVE MOMENT SUM RULES

The derivation of the negative energy moment sum rules is based on the analyticity properties of the function $z^{-N} \operatorname{Tr}[R(z) - R_0(z)]$, $N \geqslant 0$, in the complex energy plane. We know already that this function is analytic for $z \in \Pi_\eta$[7,9,13]. Furthermore, from Refs. 7, 9, 13, and 23 we also know that at high energies

$$\operatorname{Tr}[R(z) - R_0(z)] = (1/8\pi i\sqrt{z})(v,u) + O(z^{-3/2}). \tag{3.1}$$

Finally, the low-energy behavior is specified in Eqs. (2.46), (2.47), and (2.48). Using this information, we can apply Cauchy's theorem to the function

$$F_N(z) = z^{-N}\left\{\operatorname{Tr}[R(z) - R_0(z)] - \frac{\delta_{N,0}}{8\pi i\sqrt{z}}(v,u) - \frac{1}{2}\sum_{n=-2,2N-2}(-i\sqrt{z})^n \Delta_{n+1}\right\}, \tag{3.2}$$

and state that

$$\oint_C dz\, F_N(z) = 0, \quad N \geqslant 0, \tag{3.3}$$

where C is a closed contour around the spectrum of H, as shown in Fig. 1. The number of subtraction terms, $2N-2$, is chosen in such a way that $F_N(z)$ is integrable for $z \to 0$. We can then prove the following result.

Theorem 3.1: Let $e^{2a|x|}V(\mathbf{x}) \in R$ for some $a > 0$ and assume Cases I to IV. Then, for $N \geqslant 0$,

$$\int_0^\infty dE \frac{1}{E^N}\left\{2\operatorname{Im}\operatorname{Tr}[R(E) - R_0(E)] + \frac{\delta_{N,0}}{4\pi\sqrt{E}}(v,u) - \sum_{n=0,N-1}(-1)^n E^{n-1/2}\Delta_{2n}\right\} = -2\pi \sum_{j=1,N_b}(-\chi_j^2)^{-N} + \pi(-1)^{N+1}\Delta_{2N-1}, \tag{3.4}$$

where N_b is the number of discrete eigenstates of the Hamiltonian H with eigenvalue $-\chi_j^2 < 0$, counting degeneracy. The coefficients Δ_n are given by Eqs. (2.47) and (2.48) and are case dependent.

Proof: As is standard by now,[7,9,10,13] we start from the contour integral (3.3) of the analytic function $F_N(z)$, and consider the contributions from the different contour segments C_j, C_Γ, and C_η (see Fig. 1). First, it is straightforward to evaluate the C_j integrals around the discrete eigenvalues of H, which appear as poles of the resolvent $R(z)$. The result is

$$\sum_{j=1,N_b}\oint_{C_j} dz\, F_N(z) = 2\pi i \sum_{j=1,N_b}(-\chi_j^2)^{-N}, \tag{3.5}$$

where N_b is the number of negative eigenstates, with eigenvalue $-\chi_j^2 < 0$, counting degeneracy. Secondly, we determine the contribution from C_Γ in the limit $\Gamma \to \infty$. On the basis of Eq. (3.1), or from more detalied results in the literature (e.g., Ref. 9, Lemma 6; Ref. 7), we infer that the C_Γ integration of the first two terms in $F_N(z)$ gives zero in the limit $\Gamma \to \infty$. So we only have to consider the C_Γ integral of the low-energy subtraction terms containing Δ_{n+1}. We obtain then

$$\lim_{\Gamma\to\infty} -\frac{1}{2}\oint_{C_\Gamma} dz\, z^{-N} \sum_{n=-2,2N-2}(-i\sqrt{z})^n \Delta_{n+1} = -\pi i(-1)^{N+1}\Delta_{2N-1}. \tag{3.6}$$

Finally, we discuss the C_η integral. We start with the circle around the origin. Using the results of Sec. II and the fact that the $\delta_{N,0}$ term does not contribute, we arrive at

$$\frac{i}{2}\int_{3\pi/2}^{\pi/2} d\theta\,(\eta e^{i\theta})^{1-N} \sum_{n=2N-1,\infty}(-i)^n(\eta e^{i\theta})^{n/2}\Delta_{n+1} \sim \eta^{1/2} + O(\eta), \tag{3.7}$$

which goes to zero in the limit $\eta \to 0$. Next, using the reflection property of the resolvent, the rest of the C_η integral can be written as

$$\lim_{\Gamma\to 0}\lim_{\eta\to 0} 2i\int_0^\Gamma dE\, \operatorname{Im} F_N(E + i\eta). \tag{3.8}$$

It is easy to check that only the following terms will survive

$$\lim_{\Gamma\to\infty}\lim_{\eta\to 0} 2i\int_0^\Gamma \mathrm{dE}\, \mathrm{E}^{-N} \operatorname{Im}\{\operatorname{Tr}[R(E + i\eta) - R_0(E + i\eta)] + \frac{i(v,u)}{8\pi(E+i\eta)^{1/2}}\delta_{N,0} - \frac{1}{2}\sum_{n=-2,2N-2}(-i)^n(E+i\eta)^{n/2}\Delta_{n+1}\Big\}. \tag{3.9}$$

Splitting up the sum over n in Eq. (3.9) into its even and odd parts, the third term of the integrand can be written as

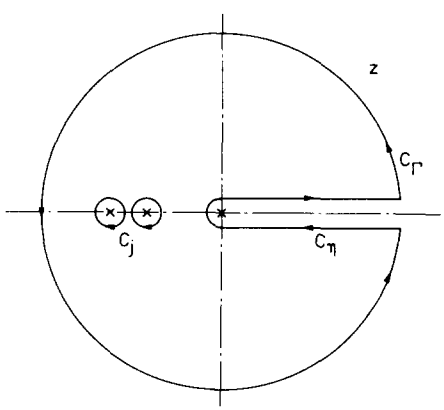


FIG. 1. Contour in the complex energy plane around the spectrum of H.

$$E^{-N}\mathrm{Im}\Big\{-\frac{1}{2}\sum_{m=-1,N-1}(-i)^{2m}\times\sum_{k=0,m}\binom{m}{k}(i\eta)^k E^{m-k}\Delta_{2m+1}-\frac{1}{2}E^{-1/2}\Big(1+\frac{i\eta}{E}\Big)^{-1/2}\times\sum_{m=0,n-1}(-i)^{(2m-1)}\sum_{k=0,m}\binom{m}{k}(i\eta)^k E^{m-k}\Delta_{2m}\Big\}. \tag{3.10}$$

So, passing the limit $\eta\to 0$ through the integral (3.9), we see from expression (3.10) that only the $k=0$ terms survive. Furthermore, recalling the result of Proposition 2.2 that the coefficients Δ_n are real, we have that only the second part of expression (3.10) contributes to the imaginary part. Consequently, the integral (3.9) becomes

$$2i\int_0^\Gamma dE\,E^{-N}\Big\{\mathrm{Im\,Tr}[R(E)-R_0(E)]+\frac{(v,u)}{8\pi E^{1/2}}\delta_{N,0}-\frac{1}{2}\sum_{n=0,N-1}(-1)^n E^{n-1/2}\Delta_{2n}\Big\}. \tag{3.11}$$

Adding the results (3.5)–(3.7), and (3.11), we obtain the negative-moment sum rules. ■

Let us discuss in more detail indentity (3.4), which is one of the principal results of this paper. We start with the remark that $\mathrm{Im\,Tr}[R(E)-R_0(E)]$ can be written down in terms of the logarithmic derivative of the S-matrix,[7] or in terms of time delay,[9,13,24]

$$2\,\mathrm{Im\,Tr}[R(E)-R_0(E)] = -i\,\mathrm{tr}\,s^\dagger(E)\frac{d}{dE}s(E)=\mathrm{tr}\,q(E). \tag{3.12}$$

Here, the matrix elements defining the on-shell S-matrix $s(E)$ are derived from the usual S-matrix elements by subtracting off the energy conserving δ-function, as follows,

$$S(\mathbf{p},\mathbf{p}')=\delta(E-E')(2/p)s(\hat{p},\hat{p}',E), \tag{3.13}$$

where $\hat{p}$ denotes the unit vector. The trace tr is taken over the two-dimensional angular space $L^2(\hat{p})$. The operator $q(E)$ is the on-shell time delay and is, in fact, a generalization of the phase shift. (For a review of time delay and its properties, we refer to Ref. 24.)

The rules (3.4) can also be extended to allow complex energy moments, $E^{-(\lambda+i\mu)}$, $\lambda\geqslant 0$, of the trace of the resolvent difference. They couple the discontinuity of the imaginary part of the trace of the resolvent difference with that of the real part. We will only write down explicitly one case of special interest, namely, the rules for λ half-integer ($\lambda=M+1/2$, $M\geqslant 0$) and $\mu=0$.

Theorem 3.2: Let $e^{2a|\mathbf{x}|}V(\mathbf{x})\in R$ for some $a>0$ and assume Cases I to IV. Then for $M\geqslant 0$,

$$\int_0^\infty dE\,\frac{1}{E^{M+1/2}}\Big\{2\,\mathrm{Re\,Tr}[R(E)-R_0(E)]-\sum_{n=-1,M-1}(-1)^n E^n\Delta_{2n+1}\Big\}$$
$$=-2\pi i\sum_{j=1,N_b}(-\chi_j^2)^{-M-1/2}-\tfrac{1}{4}(v,u)\delta_{M,0}+\pi(-1)^{M+1}\Delta_{2M}. \tag{3.14}$$

Proof: The proof of relation (3.14) is completely analogous to that of identity (3.4) in Theorem 3.1. The function $F_M(z)$ we integrate in this case is

$$F_M(z)=\frac{1}{z^{M+1/2}}\Big\{\mathrm{Tr}[R(z)-R_0(z)]-\frac{1}{2}\sum_{n=-2,2M-1}(-i\sqrt{z})^n\Delta_{n+1}\Big\}. \tag{3.15}$$

■

The set of identities (3.4) and (3.14) generalizes the results obtained in Refs. 14–16. First of all, they are valid for noncentral potentials. Secondly, we are able to treat zero-energy bound states and/or resonances whose presence has been shown to be important in some physical phenomena. As a well-known example, we mention the Efimov effect.[25] For other examples, we refer to Sec. IV and to Refs. 17, 26, and 27.

Next, it is instructive to write down the explicit form of the relations (3.4) and (3.14) for $N=0$, 1 and $M=0$ in the four cases. As we have already remarked at the end of Sec. II, we don't need the strong condition $\exp(2a|\mathbf{x}|)V(\mathbf{x})\in R$ for the validity of these specific sum rules. We only need $V\in R$ and $(1+|\mathbf{x}|)^n V(\mathbf{x})\in L^1(\mathbb{R}^3)$ for some n, the value of which depends on N, M and the specific Cases I–IV we are considering. For $N=0$, Eq. (3.4) simplifies to

$$\int_0^\infty dE\Big\{2\,\mathrm{Im\,Tr}[R(E)-R_0(E)]+\frac{(v,u)}{4\pi\sqrt{E}}\Big\}=-2\pi N_b-\pi\Delta_{-1}. \tag{3.16}$$

From the formula (2.47), (2.48) we obtain that $\Delta^{\mathrm{I}}_{-1}=0$, $\Delta^{\mathrm{II}}_{-1}=1$, $\Delta^{\mathrm{III}}_{-1}=2N_0$ and $\Delta^{\mathrm{IV}}_{-1}=2(N_0-1)+1$, where N_0 is the number of zero-energy eigenstates. Remark that we write $\Delta^{\mathrm{IV}}_{-1}$ in that particular way because of the following. Proposition 2.1 tells us that we can always choose in Case IV a particular set of linear combinations of the resonance functions such that $(v,\phi_{j_0})\neq 0$ for some j_0 and $(v,\phi_j)=0$ for $j=1,\ldots,N_0$, $j\neq j_0$. So in Case III, zero is an eigenvalue with multiplicity N_0, whereas in Case IV, its multiplicity is always N_0-1. Thus, in the four cases the rhs of the generalized Levinson's theorem (3.16) can be written as $-2\pi(N_b+N_0'+q)$, where N_0' is the number of zero-energy bound states counting degeneracy and $q=1/2$ when there is a zero-energy resonance (Cases II and IV) and zero otherwise. So we see that Eq. (3.16) generalizes the $N=0$ result in Refs. 9 and 13 by taking eventual zero-energy eigenstates into account. Furthermore, it completely agrees with the work of Newton.[10]

For $N=1$, the sum rule (3.4) reads

$$\int_0^\infty dE\,\frac{1}{E}\Big\{2\,\mathrm{Im\,Tr}[R(E)-R_0(E)]-\frac{1}{\sqrt{E}}\Delta_0\Big\}=-2\pi\sum_{j=1,N_b}(-\chi_j^2)^{-1}+\pi\Delta_1. \tag{3.17}$$

Using then again the formula (2.47), (2.48) we obtain the following expressions for the coefficients Δ_n, $n = 0,1$.

Case I:

$$\Delta_0 = -(1/4\pi)(v,t_0u), \Delta_1 = -\Delta_0^2 + 2\,\mathrm{Tr}\, r_2 t_0, \quad (3.18)$$

with $t_0 = T_0$ [see Eq. (2.16)] and r_2 defined by expression (2.18). As remarked already by AGH,[17] $(-\Delta_0)$ is the scattering length. For the other cases, the following expressions are obtained after lengthy but rather straightforward calculations.

Case II:

$$\Delta_0 = -4\pi \frac{(\tilde{\phi}, r_2\phi)}{|(v,\phi)|^2}, \quad (3.19)$$

$$\Delta_1 = -\Delta_0^2 + 2\frac{(\tilde{\phi}, r_2\phi)}{|(v,\phi)|^2}(v,T_0u) + 8\pi\frac{(\tilde{\phi}, r_3\phi)}{|(v,\phi)|^2} - \frac{2}{|(v,\phi)|^2}[(v,\phi)(\tilde{\phi}, r_2T_0u) + (\tilde{\phi},u)(v,T_0r_2\phi)] + 2\,\mathrm{Tr}\, r_2T_0. \quad (3.20)$$

As is well known, the scattering length is infinite in this case. Nevertheless, when we will study the low-energy behavior of the total cross section in Sec. IVA, we will see that even in this case Δ_0 is some kind of characteristic length of the problem.

Introducing the shorthand notation

$$(0)_{ij} \equiv (\tilde{\phi}, 0\phi)_{ij}, \quad (3.21)$$

for any operator 0, we write down the Δ's for Cases III and IV

Case III:

$$\Delta_0 = -\frac{1}{4\pi}(v,T_0u) - \sum_{i,j=1,N_0}(r_2)^{-1}_{ij}(r_3)_{ji}, \quad (3.22)$$

$$\Delta_1 = -\frac{1}{16\pi^2}(v,T_0u)^2 + 2\,\mathrm{Tr}\, r_2T_0 - 2\sum_{i,j}(r_2)^{-1}_{ij}(r_2T_0r_2)_{ji} - \sum_{i,j,k,l}(r_2)^{-1}_{ij}(r_3)_{jk}(r_2)^{-1}_{kl}(r_3)_{li} + 2\sum_{i,j}(r_2)^{-1}_{ij}(r_4)_{ji}. \quad (3.23)$$

Again, T_0 is given by Eq. (2.16) and the r_n are determined by Eq. (2.18). Remark that $(-\Delta_0)$ contains an additional contribution compared with Case I (see also Sec. IVA).

Case IV:

$$\Delta_0 = 4\pi\left[\sum_{i,j}(v,\phi_i)(r_2)^{-1}_{ij}(\tilde{\phi}_j,u)\right]^{-1}\left\{\frac{1}{4\pi}\sum_{i,j,k,l}(v,\phi_i)(r_2)^{-1}_{ij}(r_3)_{jk}(r_2)^{-1}_{kl}(\tilde{\phi}_l,u) - 1\right\} - \sum_{i,j}(r_2)^{-1}_{ij}(r_3)_{ji} \quad (3.24)$$

$$\Delta_1 = -16\pi^2\left[\sum_{i,j}(v,\phi_i)(r_2)^{-1}_{ij}(\tilde{\phi}_j,u)\right]^{-2}\left\{1 - \frac{1}{4\pi}\sum_{i,j,k,l}(v,\phi_i)(r_2)^{-1}_{ij}(r_3)_{jk}(r_2)^{-1}_{kl}(\tilde{\phi}_l,u)\right\}^2 - 8\pi\left[\sum_{i,j}(v,\phi_i)(r_2)^{-1}_{ij}(\tilde{\phi}_j,u)\right]^{-1}\left\{\frac{1}{4\pi}\sum_{i,j}[(v,T_0r_2\phi_i)(r_2)^{-1}_{ij}(\tilde{\phi}_j,u) + (v,\phi_i)(r_2)^{-1}_{ij}(\tilde{\phi}_j,r_2T_0u)\right\} + \frac{1}{4\pi}\sum_{i,j,k,l}(v,\phi_i)(r_2)^{-1}_{ij}[(r_4)_{jk} - (r_2T_0r_2)_{jk}](r_2)^{-1}_{kl}(\tilde{\phi}_l,u) - \frac{1}{4\pi}\sum_{i,j,k,l,m,n}(v,\phi_i)(r_2)^{-1}_{ij}(r_3)_{jk}(r_2)^{-1}_{kl}(r_3)_{lm}(r_2)^{-1}_{mn}(\tilde{\phi}_n,u) - \frac{1}{4\pi}(v,T_0u)\Big\} + 2\sum_{i,j}(r_2)^{-1}_{ij}[(r_4)_{ji} - (r_2T_0r_2)_{ji}] - \sum_{i,j,k,l}(r_2)^{-1}_{ij}(r_3)_{jk}(r_2)^{-1}_{kl}(r_3)_{li} + 2\,\mathrm{Tr}\, r_2T_0. \quad (3.25)$$

The remark made in Case II [after Eqs. (3.19), (3.20)] about Δ_0 is also true here.

Finally, we state the negative moment sum rule (3.14) for $M = 0$

$$\int_0^\infty dE\,\frac{1}{\sqrt{E}}\left\{2\,\mathrm{Re}\,\mathrm{Tr}[R(E) - R_0(E)] + \frac{1}{E}\Delta_{-1}\right\} = -2\pi i\sum_{j=1,N_b}(-\chi_j^2)^{-1/2} - \tfrac{1}{4}(v,u) - \pi\Delta_0, \quad (3.26)$$

where Δ_{-1} and Δ_0 are discussed in detail above. This rule is particular in the sense that there is an additional surface term present on the rhs namely, $-\frac{1}{4}(v,u)$. This term comes from the high-energy behavior of the resolvent difference [see Eq. (3.1)]. It does not appear in the partial wave version[14] of Eq. (3.26) because the partial wave resolvent difference goes more rapidly to zero at high energies.[26,27] Of course, the higher-order half-integer rules (3.14), $M \geqslant 1$, don't have this extra contribution.

In the following section, we discuss some applications of the results obtained up to now.

IV. APPLICATIONS

As we have remarked already in the Introduction, it has been recognized in the literature (see the Introduction of AGH,[17] and especially the references cited there) that the study of the low-energy behavior of nonrelativistic scattering systems has theoretical and practical importance. In this section, we choose to briefly study two applications of the results we have obtained before. Our aim hereby is to underline again the importance of this low-energy study, but also to generalize some well-known results in scattering theory.

A. Two-body total cross section

We derive a low-energy expansion for the two-body total cross section together with a closed expression for the coefficients. We discuss the first coefficients in detail in the

four cases I–IV. This leads us to the observation that there is always a finite characteristic length in the scattering problem, even when there are zero-energy resonances and bound states.

We start with

Lemma 4.1: Let $e^{2a|\mathbf{x}|}V(\mathbf{x})\in R$ for some $a>0$ and assume Case I to IV. Then the total cross section $\sigma_{\rm tot}(E)$ has the following expansion for low energies:

$$\sigma_{\rm tot}(E)=\sum_{n=[(s-1)/2],\infty} E^n\sigma_n, \tag{4.1}$$

with

$$\sigma_n=4\pi(-1)^{n+1}\sum_{j=0,n+[(1-s)/2]} \mathrm{Tr}\, r_{j+1}t_{2(n-j)+1}, \tag{4.2}$$

where $s=0$ in Case I, $s=-1$ in Case II, and $s=-2$ in the other cases, and where the symbol [ ] in the summation denotes "integer part of." Again, Tr denotes the integral over the forward matrix element.

Proof: Employing the optical theorem for two-body scattering,[26,27] the validity of which does not depend on the nature of the spectrum of H, viz.,

$$\sigma_{\rm tot}(E)=\int\frac{d\hat p\,d\hat q}{4\pi}\,|f(\hat p,\hat q,E)|^2=\frac{1}{\sqrt{E}}\int d\hat p\,\mathrm{Im}f(\hat p,\hat p,E), \tag{4.3}$$

where $\hat p$ denotes a unit vector in the direction of $\mathbf{p}$ and where

$$f(\hat p,\hat q,E)=-(1/4\pi)(vT(E+i0)u)(\hat p\sqrt{E},\hat q\sqrt{E}), \tag{4.4}$$

we arrive at

$$\sigma_{\rm tot}(E)=-\frac{1}{\sqrt{E}}\int d\mathbf{x}\,d\mathbf{y}\,\frac{\sin\sqrt{E}\,|\mathbf{x}-\mathbf{y}|}{\sqrt{E}\,|\mathbf{x}-\mathbf{y}|}\,\mathrm{Im}(vT(E+i0)u)(\mathbf{y},\mathbf{x}). \tag{4.5}$$

Inserting the low-energy expansion for $T(z)$, given by Eq. (2.28), and the small-argument expansion of the sine function, we obtain the results (4.1), (4.2) after some straightforward manipulations. ■

Next, we specify explicitly the first few coefficients σ_n in the four cases. Remark that we will present more coefficients in Case IV than are available from AGH.

Case I:

$$\sigma_0=4\pi\Delta_0^{\,2}, \tag{4.6}$$

$$\sigma_1=-4\pi\Delta_0^{\,4}-2\Delta_0(v,t_0r_2t_0u)-2(v,t_0r_3t_0u), \tag{4.7}$$

where Δ_0 is given by the first of Eq. (3.18).

Case II:

$$\sigma_{-1}=4\pi, \tag{4.8}$$

$$\sigma_0=-4\pi\Delta_0^{\,2}, \tag{4.9}$$

where Δ_0 is given by Eq. (3.19).

Case III:

$$\sigma_{-1}=0, \tag{4.10}$$

$$\sigma_0=\frac{1}{4\pi}(v,T_0u)^2+4\pi\sum_{i,j,k,l}(r_2)_{ij}^{-1}(r_3)_{jk}(r_2)_{kl}^{-1}(r_3)_{li}. \tag{4.11}$$

Case IV:

$$\sigma_{-1}=4\pi, \tag{4.12}$$

$$\begin{aligned}\sigma_0=&-(4\pi)^3\Big[\sum_{i,j}(v,\phi_i)(r_2)_{ij}^{-1}(\tilde\phi_j,u)\Big]^{-2}\\&\times\Big\{1-\Big[\frac{1}{4\pi}\sum_{i,j,k,l}(v,\phi_i)(r_2)_{ij}^{-1}(r_3)_{jk}(r_2)_{kl}^{-1}(\tilde\phi_l,u)\Big]^2\Big\}\\&-8\pi\Big[\sum_{i,j}(v,\phi_i)(r_2)_{ij}^{-1}(\tilde\phi_j,u)\Big]^{-1}\\&\times\sum_{i,j,k,l,m,n}(v,\phi_i)(r_2)_{ij}^{-1}(r_3)_{jk}(r_2)_{kl}^{-1}(r_3)_{lm}(r_2)_{mn}^{-1}(\tilde\phi_n,u)\\&+4\pi\sum_{i,j,k,l}(r_2)_{ij}^{-1}(r_3)_{jk}(r_2)_{kl}^{-1}(r_3)_{li}.\end{aligned} \tag{4.13}$$

In Case I, Eqs. (4.6) and (4.7) tell us that, in the absence of zero-energy bound states and resonances, the scattering in the low-energy limit is isotropic and determined by the scattering length, even if the potential is not spherically symmetric.[12,16,17,27] In Case III, where we have N_0 zero-energy bound states, this isotropy is broken by the presence of the second-term in Eq. (4.11) (see also Ref. 17). However, we can still say that the low-energy limit is determined in function of Δ_0, when we compare Eq. (4.11) with Eq. (3.22). In the presence of zero-energy resonances (Cases II and IV), the cross section behaves like $4\pi/E$ and the scattering length becomes infinite. This is well known. (See, e.g., Ref. 28 for a recent rigorous discussion.) Nevertheless, we see by comparing Eqs. (4.9) and (4.13) with, respectively, Eqs. (3.19) and (3.24) that the next term σ_0 can be written again in terms of Δ_0. From this we conclude that in the four cases, there is a finite *characteristic length* in the scattering system defined by

$$\Delta_0=\lim_{z\to+0^+}2\sqrt{z}\,\mathrm{Im}\,\mathrm{Tr}[R(z)-R_0(z)]. \tag{4.14}$$

In Case I, it reduces to minus the standard scattering length.

In this respect the sum rule (3.17) is quite interesting because it connects different "observables" in a general scattering situation, namely, the bound-state energies, the time delay [see Eq. (3.12)], and the characteristic length Δ_0, with the potential. For that reason it could probably be used, e.g., in nuclear scattering to eliminate ambiguities in the determination of these observables and to evaluate approximations to the S-matrix by rational functions. In its restricted partial wave form (in Case I), that sum rule has already been used in this direction.[14,15,29]

We end this part of Sec. IV with the following interesting observation. Consider Case I. For spherically symmetric potentials, we know that [see Eq. (3.12)]

$$2\,\mathrm{Im}\,\mathrm{Tr}[R(E)-R_0(E)]=2\sum_{l=0,\infty}(2l+1)\frac{d}{dE}\delta_l(E), \tag{4.15}$$

where δ_l is the phase shift for partial wave l. On one hand, from Eq. (2.47), we obtain that the lhs of Eq. (4.15) is given by

$$\Delta_0/\sqrt{E}-\Delta_2\sqrt{E}+O(E^{3/2}), \tag{4.16}$$

where Δ_0 is given by Eq. (3.18) and

$$\Delta_2=\Delta_0^{\,3}+(3/4\pi)(v,t_0r_2t_0u)-3\,\mathrm{Tr}\,r_3t_0. \tag{4.17}$$

On the other hand, using the effective range approximation[26] for δ_l, viz.,

$$E^{l+1/2}\cot\delta_l(E) = -1/a_l' + \tfrac{1}{2}r_l'E + O(E^2), \tag{4.18}$$

where a_l' is the scattering length and r_l' the effective range for partial wave l, we arrive at the following expression for the rhs of Eq. (4.15):

$$-a_0'/\sqrt{E} + (a_0'^3 - \tfrac{1}{2}a_0'^2 r_0' - 9\,a_1')\sqrt{E} + O(E^{3/2}). \tag{4.19}$$

Consequently, comparing Eqs. (4.16), (4.17) with Eq. (4.19) we get, as we have already mentioned, (see also Ref. 27)

$$-\Delta_0 = (1/4\pi)(v,t_0u) = a_0'. \tag{4.20}$$

Furthermore, the coefficient Δ_2 couples the $l=0$ and $l=1$ partial waves. We identify

$$-\tfrac{1}{3}\,\mathrm{Tr}\,(r_3 - \tilde{r}_3)t_0 = a_1',$$
$$\tilde{r}_3(\mathbf{x},\mathbf{y}) = (24\pi)^{-1}u(\mathbf{x})(x^2+y^2)v(\mathbf{y}),$$
$$8\pi\frac{(v,t_0r_2t_0u)}{|(v,t_0u)|^2} - 32\pi^2\frac{\mathrm{Tr}\tilde{r}_3t_0}{|(v,t_0u)|^2} = r_0'. \tag{4.21}$$

Similar results could be derived for the higher Δ_n's, $n\geqslant 3$ and in Cases II–IV.

B. Second virial coefficient

Another application that is particularly interesting to discuss briefly is the low-energy behavior of the quantum mechanical two-body partition function. The two-body interaction part of the partition function for Boltzmann particles can be written as[30]

$$a_2 = -\sqrt{2}\lambda^3\,\mathrm{Tr}(e^{-\beta H} - e^{-\beta H_0}), \tag{4.22}$$

where λ is the thermal wavelength, $\lambda = (4\pi\beta)^{1/2}$. We then introduce the Watson transform,[31] which connects the statistical operator with the resolvent, viz.,

$$e^{-\beta H} = -\frac{1}{2\pi i}\oint_C dz e^{-\beta z}R(z), \tag{4.23}$$

where C is again the contour shown in Fig. 1. Using this transform in Eq. (4.22), we can work out the contour integral similarly as in Sec. III. We obtain, for Cases I to IV,

$$a_2 = -\sqrt{2}\lambda^3\left\{\sum_{j=1,N_b} e^{\beta\chi_j^2} + \frac{\Delta_{-1}}{2} + \frac{1}{\pi}\int_0^\infty dE\, e^{-\beta E}\,\mathrm{Im}\,\mathrm{Tr}[R(E) - R_0(E)]\right\}. \tag{4.24}$$

The same formula (for $\beta = it$) has been written down by Jensen and Kato.[18]

Starting from Eq. (4.24), we can determine the low-temperature behavior of a_2^c, the continuum part of a_2. We can use, e.g., the method proposed in Ref. 16, or follow Ref. 18, Sec. 10. We get

$$a_2^c = \frac{-\sqrt{2}\lambda^3}{2\pi}\sum_{n=0,N}(-1)^n\frac{\Gamma(n+1/2)}{\beta^{n+1/2}}\Delta_{2n} + O(\beta^{-N-1/2}). \tag{4.25}$$

From Eqs. (4.24) and (4.25), we infer that the negative-energy bound states form the most important contribution at low temperatures. The next-order contribution is given by $\frac{1}{2}\Delta_{-1} = N_0' + q$, with N_0' the number of zero-energy bound states counting degeneracy and $q = \frac{1}{2}$ when there is a zero-energy resonance (Cases II and IV) and zero otherwise. Finally, the leading scattering contribution is always given by the characteristic length in the system [see Eq. (4.14)]. This result generalizes the one obtained in Ref. 16.

We end with the remark that it is straightforward to determine the possible influence of the zero-energy eigenstates in the high-temperature expansion of the second virial coefficient a_2 (see Ref. 13). Using the general $N=0$ sum rule, written down in Eq. (3.16), it is easy to show that they are exactly cancelled, just like the negative-energy bound states.

ACKNOWLEDGMENTS

One of the authors (D.B) would like to thank F. Gesztesy for very interesting discussions. The other (S.W.) gratefully acknowledges the warm hospitality extended to him by the Institute for Theoretical Physics of the University of Leuven.

[1] N. Levinson, K. Dan. Vidensk. Selsk. Mat.-Fys. Medd. **25**, 1 (1949).
[2] R. G. Newton, Phys. Rev. **101**, 1588 (1956).
[3] V. S. Buslaev and L. D. Faddeev, Dokl. Akad. Nauk. SSSR **132**, 13 (1960) [Sov. Math. Dokl. **1**, 451 (1960)].
[4] I. C. Percival, Proc. Phys. Soc. London **80**, 1290 (1962).
[5] I. C. Percival and M. J. Roberts, Proc. Phys. Soc. London **82**, 519 (1963).
[6] R. D. Puff, Phys. Rev. A **11**, 154 (1975).
[7] V. S. Buslaev, Dokl. Akad. Nauk. SSSR **143**, 1067 (1962) [Sov. Phys. Dokl. **7**, 295 (1962)]; in *Topics in Mathematical Physics*, edited by Mh. Sh. Birman (Consultants Bureau, New York, 1967), Vol. 1, p. 69.
[8] V. S. Buslaev, in *Topics in Mathematical Physics*, edited by Mh. Sh. Birman (Consultants Bureau, New York, 1971), Vol. 4, p. 43.
[9] T. A. Osborn and D. Bollé, J. Math. Phys. **18**, 432 (1977).
[10] R. G. Newton, J. Math. Phys. **18**, 1348 (1977).
[11] T. Dreyfus, J. Phys. A **9**, L187 (1976); Helv. Phys. Acta **51**, 321 (1978); J. Math. Anal. Appl. **64**, 114 (1978).
[12] R. G. Newton, J. Math. Phys. **18**, 1582 (1977) and references therein.
[13] D. Bollé and H. Smeesters, Phys. Lett. A **62**, 290 (1977); D. Bollé, Ann. Phys. (N.Y.) **121**, 131 (1979).
[14] M. J. Roberts, Proc. Phys. Soc. London **83**, 503 (1964).
[15] D. A. Kirhznits and N. Zh. Takibaev, *Proceedings of the European Symposium on Few-Particle Problems* (Potsdam, 1977), p. 98; Yad. Fiz. **16**, 435 (1972) [Sov. J. Nucl. Phys. **16**, 242 (1973)].
[16] D. Bollé, Phys. Lett. A **74**, 26 (1979).
[17] S. Albeverio, F. Gesztesy, and R. Høegh-Krohn, Ann. Inst. Henri Poincaré, Sec. A **37**, 1 (1982).
[18] A. Jensen and T. Kato, Duke Math. J. **46**, 583 (1979).
[19] B. Simon, *Quantum Mechanics for Hamiltonians Defined as Quadratic Forms* (Princeton University, Princeton, NJ, 1971).
[20] M. Klaus and B.Simon, Ann. Phys. (N.Y.) **130**, 251 (1980).
[21] T. Kato, *Perturbation theory for linear operators* (Springer, New York, 1966).
[22] M. Reed and B. Simon, *Methods of Modern Mathematical Physics IV, Analysis of Operators* (Academic, New York, 1978).
[23] S. F. J. Wilk, Y. Fujiwara, and T. A. Osborn, Phys. Rev. A **24**, 2187 (1981).
[24] Ph. Martin, Acta Phys. Austriaca Suppl. **23**, 157 (1981) and references therein.
[25] V. N. Efimov, Phys. Lett. B **33**, 563 (1970); Sov. J. Nucl. Phys. **12**, 589 (1971); Nucl. Phys. A **210**, 157 (1973).
[26] R. G. Newton, *Scattering theory of waves and particles* (McGraw-Hill, New York, 1966).
[27] W. O. Amrein, J. M. Jauch, and K. B. Sinha, *Scattering theory in quantum mechanics* (Benjamin, Reading, MA, 1977).
[28] V. Enss and B. Simon, Commun. Math. Phys. **76**, 177 (1980).
[29] N. Zh. Takibaev, Yad. Fiz. **26**, 81 (1977) [Sov. J. Nucl. Phys. **26**, 42 (1977)].
[30] K. Huang, *Statistical mechanics* (Wiley, New York, 1964).
[31] K. M. Watson, Phys. Rev. **103**, 489 (1956).

Charged spinning mass field involving rational functions[a)]

Y. Chen, D. S. Guo, and F. J. Ernst
Department of Physics, Illinois Institute of Technology, Chicago, Illinois 60616

(Received 20 October 1982; accepted for publication 30 December 1982)

Using a microcomputer, we have derived a new nine-parameter charged spinning mass field. This solution is the simplest member of an infinite sequence of rational function solutions of the coupled Einstein–Maxwell field equations.

PACS numbers: 04.20.Jb

Now that there exist methods for generating arbitrarily complicated asymptotically flat, stationary, axially symmetric exact solutions of the vacuum and electrovac Einstein field equations, attention is likely to shift to problems which have so far remained largely untouched. We regard the geometrical analysis and physical interpretation of exact solutions to be one of the tasks of paramount importance. Except for the charged Kerr solution, which has been studied extensively, the only other solution which has received any attention at all has been the $\delta = 2$ Tomimatsu–Sato vacuum solution. In addition to being the first solutions to be discovered after the Kerr metric, the T–S solutions have the advantage of being expressible in terms of rational functions. In the present paper we shall show how by a limiting process electrovac generalizations of the T–S solutions may be obtained from the multiple-Cosgrove solutions[1] discussed recently by Guo and Ernst. In particular, we explicitly work out the limit in the case of the double-Cosgrove solution. We hope that the now rational function solution, which involves a total of nine parameters, will become the subject of future analysis. In special cases our solution reduces to the five-parameter vacuum solution[2] of Kinnersley and Chitre and to the charged T–S solution[3] of Ernst.

THE DOUBLE COSGROVE SOLUTION

The double-Cosgrove solution worked out by Guo and Ernst corresponds to the complex potentials[4]

$$\mathscr{E} = 1 - 2iN/D \quad \text{and} \quad \Phi = -2N'/D, \tag{1}$$

where

$$\begin{aligned} N = {} & \Delta_{456}\Delta_{32}R_1(K_1 - K_1^*)(K_1 - K_2^*)(K_1^* - K_2) \\ & + \Delta_{423}\Delta_{65}R_1(K_1 - K_1^*)(K_2 - K_2^*)(K_2 - K_1^*) \\ & + \Delta_{523}\Delta_{61}R_4(K_1 - K_1^*)(K_2^* - K_1)(K_2 - K_2^*) \\ & + \Delta_{623}\Delta_{51}R_4(K_1 - K_1^*)(K_2 - K_2^*)(K_1 - K_2^*) \\ & + \Delta_{123}\Delta_{65}R_4(K_1 - K_2)(K_1 - K_2^*)(K_2 - K_2^*) \end{aligned} \tag{2a}$$

and

$$\begin{aligned} D = {} & \Delta_{423}\Delta_{156}(K_2 - K_1)(K_2^* - K_1^*) \\ & + \Delta_{523}\Delta_{164}(K_2^* - K_1)(K_2 - K_1^*) \\ & + \Delta_{623}\Delta_{145}(K_2^* - K_1)(K_2 - K_1^*). \end{aligned} \tag{2b}$$

The expression for N' is like the expression for N except that primes appear on the two-index deltas.

The three-index deltas are determinants

$$\Delta_{ijk} = \det(\psi_i \psi_j \psi_k) \tag{3}$$

of certain eigenvectors

$$\psi_i = \begin{pmatrix} Q_i \\ R_i \\ S_i \end{pmatrix}, \tag{4}$$

while the two-index deltas are subdeterminants

$$\Delta_{ij} = \begin{vmatrix} R_i & R_j \\ S_i & S_j \end{vmatrix} \tag{5a}$$

and

$$\Delta'_{ij} = \begin{vmatrix} Q_i & Q_j \\ R_i & R_j \end{vmatrix}. \tag{5b}$$

The Cosgrove transformation corresponds to the K–C group element

$$v(\tau) = \exp[J\eta(\tau)], \tag{6}$$

where

$$\eta(\tau) = -i \ln[(\tau - K)/(\tau - K^*)],$$

$$J = (\tfrac{1}{3})\, iI + (2/E)hh^{\dagger}\mathfrak{E},$$

$$E = 2ih^{\dagger}\mathfrak{E}h,$$

and

$$\mathfrak{E} = \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & -i/2 \end{pmatrix}.$$

The complex column matrix

$$h = \begin{pmatrix} h_1 \\ h_2 \\ h_3 \end{pmatrix} \tag{7a}$$

is completely arbitrary.

In addition to h we may select the other two (degenerate) eigenvectors of J to be

$$h' = \begin{pmatrix} h_1^* \\ h_2^* \\ 0 \end{pmatrix}, \tag{7b}$$

$$h'' = \begin{pmatrix} -h_3^* h_1 \\ -h_3^* h_2 \\ E - h_3^* h_3 \end{pmatrix}. \tag{7c}$$

Then, using the method described by Guo and Ernst, we may construct the vectors ψ_1, ψ_2, and ψ_3 from $P_0(K^*)h$, $P_0(K)h'$, and $P_0(K)h''$, respectively, where

[a)] Research supported in part by National Science Foundation Grants PHY-79-08627 and PHY-82-05608.

 0022-2488/83/061564-04$02.50 © 1983 American Institute of Physics

$$P_0(\tau)=\begin{pmatrix} -(r-\tau+z)/(2r) & i(r+\tau-z)/(2r) & 0 \\ -i/(2r) & 1/(2r) & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

and

$$r=r(\tau)=[(z-\tau)^2+\rho^2]^{1/2}.$$

We find that

$$\begin{aligned} Q_1 &= -i[R_2^* r(K^*)-R_1(K^*-z)], \\ R_1 &= h_1+ih_2, \quad S_1=2ir(K^*)h_3, \\ Q_2 &= -i[R_1^* r(K)-R_2(K-z)], \\ R_2 &= h_1^*+ih_2^*, \quad S_2=0, \\ Q_3 &= h_3[R_2^* r(K)-R_1(K-z)], \\ R_3 &= ih_3^* R_1, \quad S_3=2r(K)[R_1^* R_1-R_2^* R_2]. \end{aligned} \tag{8}$$

We have also verified that ψ_4, ψ_5, and ψ_6 have exactly the same form as ψ_1, ψ_2, and ψ_3, respectively, but with h_1, h_2, and h_3 replaced by new h_4, h_5, and h_6 (related indirectly to the parameters of the J matrix of the second Cosgrove transformation) and with the constant $K=K_1$ replaced by a new K_2.

The N-fold Cosgrove solution[5] with $N>2$ has been worked out by Wang, Guo, and Wu. The limiting procedure we shall now discuss could be applied to the N-fold Cosgrove solution too.

THE LIMITING PROCESS

The general N-fold Cosgrove solution involves the Weyl canonical fields z and ρ in a complicated manner through N fields $r(K_i)$ and N fields $r(K_i^*)$. In order to obtain a rational metric of the T-S type, one must through a limiting process reduce these $2N$ fields to precisely one $r(K)$ and one $r(K^*)$, in terms of which the so-called *oblate* spheroidal coordinates x and y may be defined. (We generate directly the $q^2-p^2=1$ case involving oblate coordinates. The $q^2+p^2=1$ case with prolate coordinates may be obtained by analytic continuation of our final result.) In the vacuum case a similar limiting procedure was introduced by Kramer and Neugebauer[6] and described in detail by Oohara and Sato.[7]

In our case, we take a limit in which K_2 tends towards K_1, and h_4, h_5, and h_6 tend toward h_1, h_2, and h_3, respectively. Specifically, we express the components of ψ_4, ψ_5, and ψ_6 in the form

$$\begin{aligned} Q_4 &= -i[R_5^* r(K^*+\epsilon)-R_4(K^*-z+\epsilon)], \\ R_4 &= R_1\exp(\lambda_1\epsilon), \quad S_4=2ir(K^*+\epsilon)h_6, \\ Q_5 &= -i[R_4^* r(K+\epsilon)-R_5(K-z+\epsilon)], \\ R_5 &= R_2\exp(\lambda_2\epsilon), \quad S_5=0, \\ Q_6 &= h_6^*[R_5^* r(K+\epsilon)-R_4(K-z+\epsilon)], \\ R_6 &= ih_6^* R_4, \quad S_6=2r(K+\epsilon)[R_4^* R_4-R_5^* R_5], \end{aligned} \tag{9}$$

where $h_6=h_3\exp(\lambda_3\epsilon)$.

We shall consider the limit as ϵ goes to zero, under which ψ_4, ψ_5, and ψ_6 tend toward ψ_1, ψ_2, and ψ_3, respectively. In this case N and D are found to have the limiting forms

$$\begin{aligned} N= & -4i[\Delta_{32}-i(\Delta_{32'}-\Delta_{23'})] \\ & \times[\Delta_{123}R_1-2i(\Delta_{1'23}R_1-\Delta_{123}R_{1'})] \\ & -4\Delta_{32}[\Delta_{123}R_{1'}-(\Delta_{1'23}+\Delta_{12'3}-\Delta_{13'2})R_1 \\ & +2i(\Delta_{1'2'3}-\Delta_{1'3'2})R_1-2i(\Delta_{12'3}-\Delta_{13'2})R_{1'}] \end{aligned} \tag{10}$$

and

$$D=(\Delta_{123})^2+4(\Delta_{1'13}\Delta_{2'23}+\Delta_{1'12}\Delta_{3'32}). \tag{11}$$

There is a similar expression for N'. The primes signify differentiation of the designated matrix column with respect to the parameter ϵ, followed by evaluation at $\epsilon=0$. Common factors in N and D and in N' have been cancelled out. After taking the limit, we assigned the value $-i$ to K in keeping with the prevalent convention of suppressing the mass parameter of the T–S solutions. The x–y coordinates are then introduced via the relations

$$r(-i)=x+iy, \quad r(+i)=x-iy,$$

corresponding to $z=xy$ and $\rho=(x^2+1)^{1/2}(1-y^2)^{1/2}$.

THE NEW NINE-PARAMETER ELECTROVAC SOLUTION

We found it convenient to perform our work using the notation $r=r(-i)$, $r^*=r(+i)$, and $-i-z=rs$, $+i-z=r^*s^*$, introducing x and y only at the very end of the computation, when we also introduced p and q, the NUT parameter γ and a new complex parameter h by writing

$$\begin{aligned} R_1^* &= iR_2(q+p)\exp(i\gamma), \\ R_2^* &= iR_1(q-p)\exp(i\gamma), \\ h_3 &= ihR_2^*, \end{aligned} \tag{12}$$

which involves no loss of generality. In order to facilitate comparison with the solution of Kinnersley and Chitre, we also wrote

$$\begin{aligned} \lambda_2^*-\lambda_1 &= \alpha+i\beta, \\ \lambda_3-\lambda_2^* &= \sigma+i\tau. \end{aligned} \tag{13}$$

Our nine parameters are the mass, the NUT parameter γ, the Tomimatsu–Sato rotation parameter q, the Kinnersley–Chitre α and β parameters, the electric and magnetic charge parameters (in the complex h), and σ and τ, which appear only in the nonvacuum case.

It should be noted that our nine-parameter solution could not have been obtained from the five-parameter vacuum solution of Kinnersley and Chitre merely by applying the charging transformation[8] of Ernst. The latter transformation produces only two additional real parameters. It is anticipated that analysis of the nine-parameter solution will reveal that the new parameters σ and τ are associated with the electric and magnetic dipole moments.

We find it convenient to use Kinnersley's representation[9]

$$\mathscr{E}=(U-W)/(U+W), \quad \Phi=V/(U+W), \tag{14}$$

of the complex potentials. The complete evaluation of U, V, and W was carried out on a Z80 home computer with 64 kbytes of memory, using Ernst's equation evaluator,[10] a symbolic manipulation language based upon the CP/M-80

version of muLISP. We are in the process of setting up an 8088 computer with 320 kbytes of memory, which should permit us to perform much more elaborate calculations using a higher level symbolic manipulation language based upon the CP/M-86 version of muLISP.

It turns out that V and W are linear homogeneous in the variables r, r^*, s, and s^*, while U is quadratic homogeneous in these variables. In fact, we may write

$$\begin{aligned} U &= \exp(i\gamma)(\tfrac{1}{2}U_{rr}r^2 + U_{rr^*}rr^* + \tfrac{1}{2}U_{r^*r^*}r^{*2} \\ &\quad + U_{ss^*}ss^* + U_{rs^*}rs^* + U_{r^*s}r^*s), \\ V &= V_r r + V_{r^*}r^* + V_s s + V_{s^*}s^*, \\ W &= W_r r + W_{r^*}r^* + W_s s + W_{s^*}s^*, \end{aligned} \tag{15}$$

where the constant coefficients are given by the following expressions:

$$\begin{aligned} U_{rr} &= \tfrac{1}{2}(q+p)^2, \\ U_{rr^*} &= -(1-h^*h)(\tfrac{1}{2}+\alpha^2+\beta^2) \\ &\quad + h^*h[2(\sigma\alpha+\tau\beta)+\sigma^2+\tau^2], \\ U_{r^*r^*} &= \tfrac{1}{2}(q-p)^2 \\ &\quad \times[(1-h^*h)^2 - 4h^*h(\sigma^2+\tau^2)], \\ U_{ss^*} &= 1-h^*h, \\ U_{rs^*} &= -(1-h^*h)(\alpha-i\beta)+h^*h(\sigma-i\tau), \\ U_{r^*s} &= (1-h^*h)(\alpha+i\beta)-h^*h(\sigma+i\tau); \end{aligned} \tag{16a}$$

$$\begin{aligned} V_r &= h^*(q+p)[1-i\alpha \\ &\quad -\beta - i(\sigma-i\tau)], \\ V_{r^*} &= -h^*(q-p)[(1-h^*h) \\ &\quad \times(1-i\alpha-\beta)-i(\sigma-i\tau) \\ &\quad -2(\alpha+i\beta)(\sigma-i\tau)+ih^*h(\sigma+i\tau)], \\ V_s &= ih^*(q+p), \\ V_{s^*} &= ih^*(q-p) \\ &\quad \times[(1-h^*h)-i(\sigma-i\tau)]; \end{aligned} \tag{16b}$$

$$\begin{aligned} W_r &= (q+p)(1-i\alpha-\beta), \\ W_{r^*} &= -(q-p)\{(1-h^*h)(1-i\alpha+\beta) \\ &\quad -2h^*h[\tau+(\alpha+i\beta) \\ &\quad \times(\sigma-i\tau)+\sigma^2+\tau^2]\}, \\ W_s &= i(q+p), \\ W_{s^*} &= i(q-p)[(1-h^*h) \\ &\quad -ih^*h(\sigma-i\tau)]. \end{aligned} \tag{16c}$$

In order to express the complex potentials in terms of the oblate spheroidal coordinates x and y, we simply substitute

$$r = x+iy \quad \text{and} \quad s = (-i-xy)/(x+iy), \tag{17}$$

and observe that

$$\begin{aligned} r^3r^* &= (x^4-1)-(y^4-1) \\ &\quad +2ixy(x^2+y^2), \\ r^2r^*s^* &= xy(-x^2+y^2-2) \\ &\quad -i(-x^2+y^2+2x^2y^2), \\ r^2r^{*2} &= (x^2+y^2)^2, \end{aligned}$$

and

$$\begin{aligned} rr^*ss^* &= \tfrac{1}{2}[(x^2+y^2)^2 \\ &\quad -(x^4-1)-(y^4-1)], \end{aligned}$$

while

$$r^2r^* = x(x^2+y^2)+iy(x^2+y^2)$$

and

$$\begin{aligned} rr^*s &= -y(1-y^2)-y(x^2+y^2) \\ &\quad -ix(x^2+1)+ix(x^2+y^2). \end{aligned}$$

Thus, including an extra factor rr^* in U, V, and W, we finally arrive at the following expressions in terms of the oblate spheroidal coordinates:

$$\begin{aligned} U &= \exp(i\gamma)[U_1(x^4-1)+U_2(y^4-1) \\ &\quad +U_3xy(x^2+y^2)+U_4(x^2+y^2)^2 \\ &\quad +U_5xy(-x^2+y^2-2)+U_6 \\ &\quad \times(-x^2+y^2+2x^2y^2)], \end{aligned} \tag{18a}$$

$$\begin{aligned} V &= V_1x(x^2+1)+V_2y(1-y^2) \\ &\quad +V_3x(x^2+y^2)+V_4y(x^2+y^2), \end{aligned} \tag{18b}$$

and

$$\begin{aligned} W &= W_1x(x^2+1)+W_2y(1-y^2) \\ &\quad +W_3x(x^2+y^2)+W_4y(x^2+y^2), \end{aligned} \tag{18c}$$

where the constant coefficients are given by:

$$\begin{aligned} U_1 &= p^2+\tfrac{1}{2}h^*h-\tfrac{1}{2}h^*h(q-p)^2 \\ &\quad \times[1-\tfrac{1}{2}h^*h+2(\sigma^2+\tau^2)], \\ U_2 &= -q^2+\tfrac{1}{2}h^*h+\tfrac{1}{2}h^*h(q-p)^2 \\ &\quad \times[1-\tfrac{1}{2}h^*h+2(\sigma^2+\tau^2)], \\ U_3 &= 2ipq+ih^*h(q-p)^2 \\ &\quad \times[1-\tfrac{1}{2}h^*h+2(\sigma^2+\tau^2)], \\ U_4 &= -(\alpha^2+\beta^2)+h^*h \\ &\quad \times[(\alpha+\sigma^2)+(\beta+\tau)^2], \\ U_5 &= 2i\beta(1-h^*h)-2ih^*h\tau, \\ U_6 &= 2i\alpha(1-h^*h)-2ih^*h\sigma; \end{aligned} \tag{19a}$$

$$\begin{aligned} V_1 &= h^*[2p+(q-p) \\ &\quad \times(h^*h+2i\sigma+2\tau)], \\ V_2 &= ih^*[-2q+(q-p) \\ &\quad \times(h^*h+2i\sigma+2\tau)], \\ V_3 &= -2ih^*[\alpha p - i(\beta+i\sigma+\tau)q] \\ &\quad +h^*(q-p)[2(\alpha+i\beta)(\sigma-i\tau) \\ &\quad +h^*h(-i\alpha+\beta-i\sigma+\tau)], \\ V_4 &= -2ih^*[i\alpha q+(\beta+i\sigma+\tau)p] \\ &\quad -ih^*(q-p)[2(\alpha+i\beta)(\sigma-i\tau) \\ &\quad +h^*h(-i\alpha+\beta-i\sigma+\tau)]; \end{aligned} \tag{19b}$$

$$\begin{aligned} W_1 &= 2p+h^*h(q-p) \\ &\quad \times[1+2i\sigma+2\sigma], \\ W_2 &= -2iq+ih^*h(q-p) \\ &\quad \times[1+2i\sigma+2\sigma], \end{aligned}$$

$$
\begin{aligned}
W_3 = & -2i(\alpha p - i\beta q) \\
& + h^* h\,(q-p)[-i\alpha + \beta \\
& - 2i\sigma + 2(\sigma^2 + \tau^2) \\
& + 2(\alpha + i\beta)(\sigma - i\tau)], \qquad (19c) \\
W_4 = & -2i(i\alpha q + \beta p) \\
& - ih^* h\,(q-p) \\
& \times[-i\alpha + \beta - 2i\sigma + 2(\sigma^2 + \tau^2) \\
& + 2(\alpha + i\beta)(\sigma - i\tau)].
\end{aligned}
$$

It will be noticed that when $h = 0$, this reduces to the five-parameter vacuum solution of Kinnersley and Chitre (the $q^2 - p^2 = 1$ case). Furthermore, when $\alpha = \beta = \sigma = \tau = 0$, our solution reduces to the charged T–S solution of Ernst. A similar procedure will, when applied to the N-fold Cosgrove solution worked out by Wang, Guo, and Wu, result in rational function electrovac generalizations of the higher Kinnersley–Chitre–Tomimatsu–Sato solutions.

In closing, we should like to emphasize once again that we feel that the time has arrived when one must begin to analyze geometrically and physically some of the many solutions the new solution-generating techniques have afforded us. Our nine-parameter rational function solution might be the place to begin.

[1]D. S. Guo and F. J. Ernst, J. Math. Phys. **23**, 1359 (1982).
[2]W. Kinnersley and D. M. Chitre, J. Math. Phys. **19**, 2037 (1978).
[3]F. J. Ernst, Phys. Rev. D **7**, 2520 (1973).
[4]F. J. Ernst, Phys. Rev. **168**, 1415 (1968).
[5]S. K. Wang, H. Y. Guo, and K. Wu, preprint.
[6]D. Kramer and G. Neugebauer, Phys. Lett. A **75**, 259 (1980).
[7]K. Oohara and H. Sato, Progr. Theor. Phys. **65**, 1891 (1981).
[8]F. J. Ernst, J. Math. Phys. **15**, 1409 (1974).
[9]W. Kinnersley, J. Math. Phys. **14**, 651 (1973).
[10]The source listing of the equation evaluation will be submitted for publication elsewhere.

Magnetization of all stationary cylindrically symmetric vacuum metrics

B. R. Iyer and C. V. Vishveshwara
Raman Research Institute, Bangalore, 560080 India

(Received 13 May 1982; accepted for publication 29 July 1982)

New exact solutions of the Einstein–Maxwell equations are obtained by a Harrison-type transformation applied to the most general stationary cylindrically symmetric vacuum spacetime. These solutions include generalizations of the well-known Melvin Magnetic Universe, and the Weyl–Levi-Cività magnetic solution in addition to a new class of magnetic stationary axisymmetric spacetimes.

PACS numbers: 04.20.Jb

I. INTRODUCTION

The importance of magnetic fields in astrophysical situations is well recognized and one convenient way to introduce a magnetic field in some given spacetimes is by a Harrison[1]-type transformation. Ernst[2] used this method to obtain the solution corresponding to a black hole in an external magnetic field. Though the method leads to nonasymptotically flat spacetimes and sourceless electromagnetic fields the Ernst solution has been useful and employed as an idealized model by various authors[3] to study the effect of magnetic fields in different physical situations. However, especially in the Kerr[4] case, the forms of the fields are very complicated and it is difficult to understand their structure. Consequently it would be instructive if one could study the electromagnetic fields that arise from a Harrison-type transformation in a case where the form of the functions involved is simpler.

To initiate such an investigation we apply a Harrison-type transformation to the stationary cylindrically symmetric vacuum metrics because the most general form of these solutions has been given in an especially simple form by Vishveshwara and Winicour.[5] Further, these solutions include as a special case the flat Minkowski metric so that we expect to obtain similar but more general spacetimes than the Melvin magnetic Universe (MMU).[6] Indeed as we shall see later the solutions include generalizations of the twisted MMU (TMMU)[1] and the twisted Weyl–Levi-Cività magnetic solution (TWLCM). In addition a new class of magnetic stationary metrics is generated which is axisymmetric and no longer retain their original cylindrical symmetry.

Recently Hiscock[7] has pointed out that the magnetized Kerr black hole is asymptotically non-Melvin unlike the magnetized Schwarzschild black hole although both the metrics in their original nonmagnetized versions are asymptotically flat. This follows by noting that in the magnetic Kerr case the electric fields persist even in the asymptotic region. It would be of obvious interest to check whether asymptotically the magnetic Kerr solution goes into any one of the above new classes of solutions. These above three reasons motivate us to obtain this new class of magnetized solutions starting from the most general stationary cylindrically symmetric vacuum spacetime.

In the next section, the Harrison transformation used by Ernst[2] to obtain magnetized solutions is given. Next following Vishveshwara and Winicour[5] the most general stationary cylindrically symmetric vacuum metric is written down and its relevant properties summarized for later use. The Harrison transformation is then applied and it is shown that the equation for ω' can be explicitly integrated to obtain the new magnetic solutions. In Sec. IV these solutions are classified. These solutions as mentioned earlier yield generalizations of known magnetic solutions as well as a new class of magnetic solutions. After obtaining these solutions the electromagnetic fields seen by a locally nonrotating observer are also calculated.

II. THE MAGNETIZATION PRESCRIPTION

In this section, we summarize the magnetization procedure of Ernst applicable to any axially symmetric stationary solution of Einstein's equations. We do this in some detail to make the calculation self contained. The prescription involves making a Harrison-type transformation that generates in general new solutions of Einstein Maxwell equations from old ones. We confine our summary to cases where the starting solution is a vacuum metric. Following Ernst, any axially symmetric stationary solution may be written in the form

$$ds^2 = f^{-1}[-2P^{-2}\,d\xi d\xi^* + \rho^2\,dT^2] - f(d\phi - \omega\,dT)^2, \tag{1}$$

where $f<0$. For any metric of the above type, the twist potential φ and the complex gravitational potential $\mathscr{E}$ are given by

$$\nabla\varphi = \frac{f^2}{\rho}(i\nabla\omega), \tag{2}$$

$$\mathscr{E} = f + i\varphi, \tag{3}$$

where[8]

$$\nabla \equiv \frac{\partial}{\partial\rho} + i\frac{\partial}{\partial z}. \tag{4}$$

To obtain the magnetized version of the metric given by Eq. (1) one replaces f and ω in Eq. (1) and f' and ω' defined by

$$f' = |\Lambda|^{-2} f, \tag{5}$$

$$\nabla\omega' = |\Lambda|^2\nabla\omega + \frac{\rho}{f}[\Lambda^*\nabla\Lambda - \Lambda\nabla\Lambda^*], \tag{6}$$

where[8]

$$\Lambda = 1 - \bar{B}_0^2\mathscr{E}. \tag{7}$$

The nontrivial part of the procedure is the integration of Eq. (6) to obtain ω'. For vacuum metrics—to which we confine

 0022-2488/83/061568-06$02.50 © 1983 American Institute of Physics

ourselves—Eq. (6) is equivalent to

$$d(\omega' - \omega) = \bar{B}_0^4 [\varphi d\chi - (f^2 + \varphi^2) d\omega], \tag{8}$$

where χ is a new "potential" defined by

$$\nabla\chi = -\frac{i\rho}{f^2}\nabla(f^2 + \varphi^2). \tag{9}$$

It is simpler to obtain ω' from Eqs. (8) and (9) and we shall employ them in our computation. Under this transformation the complex gravitational potential $\mathscr{E}$ goes into $\mathscr{E}'$ where

$$\mathscr{E}' = \Lambda^{-1}\mathscr{E}, \tag{10}$$

whilst the complex electromagnetic potential changes from $\Phi = 0$ to Φ' given by

$$\Phi' = -\bar{B}_0\Lambda^{-1}\mathscr{E}. \tag{11}$$

It should be noted that in the axially symmetric stationary case $A_1 = A_2 = 0$. The complex electromagnetic field is related to the vector potential by

$$\Phi' = A_3 + iA'_0 \tag{12}$$

with A'_0 satisfying

$$\nabla A_0 + \omega'\nabla A_3 = i\rho f'^{-1}\nabla A'_0. \tag{13}$$

The combination Φ' is of interest since the electromagnetic field components as measured by a locally nonrotating observer is related to it by

$$B^{\mathrm{NR}}_{\hat{\rho}} + iE^{\mathrm{NR}}_{\hat{\rho}} = -P\frac{\partial\Phi'}{\partial z}, \tag{14a}$$

$$B^{\mathrm{NR}}_{\hat{z}} + iE^{\mathrm{NR}}_{\hat{z}} = P\frac{\partial\Phi'}{\partial\rho}. \tag{14b}$$

We now proceed to apply this method to the spacetimes of our interest.

III. APPLICATION TO STATIONARY CYLINDRICALLY SYMMETRIC METRICS

Following the treatment of Ref. (5) we see that all stationary cylindrically symmetric vacuum metrics may be taken in the form

$$ds^2 = e^{2\psi}(d\tau^2 + d\sigma^2) + \lambda_{00}dt^2 + 2\lambda_{01}dtd\phi + \lambda_{11}d\phi^2, \tag{15}$$

with

$$\lambda_\alpha = A_\alpha\tau^{1+b} + B_\alpha\tau^{1-b}, \tag{16}$$

$$e^{2\psi} = c\tau^{(b^2-1)}, \tag{17}$$

where α runs over (00), (01), (11) and $b^2 > 0$. The τ and σ are related to the usual coordinates ρ and z as

$$\tau = \sqrt{2}\rho, \quad \sigma = \sqrt{2}z, \tag{18}$$

while the A's and B's in Eq. (16) satisfy the normalization conditions

$$A_{00}A_{11} - A_{01}^2 = B_{00}B_{11} - B_{01}^2 = 0, \tag{19a}$$

$$A_{00}B_{11} + A_{11}B_{00} - 2A_{01}B_{01} = -\tfrac{1}{2}. \tag{19b}$$

The mass and angular momentum per unit value of z are given by

$$m = \tfrac{1}{4} + \tfrac{1}{2}b(A_{11}B_{00} - A_{00}B_{11}), \tag{20}$$

$$j = \tfrac{1}{2}b(A_{01}B_{11} - A_{11}B_{01}). \tag{21}$$

By comparing the metric given by Eq. (15), with the general form of Eq. (1) we identify

$$d\xi = \frac{d\rho + idz}{\sqrt{2}}, \tag{22}$$

$$P^{-2} = 2\lambda_{11}e^{2\psi}, \tag{23}$$

$$f = -\lambda_{11} = -(A_{11}\tau^{1+b} + B_{11}\tau^{1-b}), \tag{24}$$

$$\omega = -\frac{\lambda_{01}}{\lambda_{11}} = -\frac{A_{01}\tau^b + B_{01}\tau^{-b}}{A_{11}\tau^b + B_{11}\tau^{-b}}. \tag{25}$$

To proceed with the magnetization we first compute the twist potential φ using Eq. (2) which in component form reads

$$\frac{\partial\varphi}{\partial\rho} = -\frac{f^2}{\rho}\frac{\partial\omega}{\partial z}, \tag{26a}$$

$$\frac{\partial\varphi}{\partial z} = \frac{f^2}{\rho}\frac{\partial\omega}{\partial\rho}. \tag{26b}$$

Using

$$\frac{\partial\omega}{\partial z} = 0, \tag{27a}$$

$$\frac{\partial\omega}{\partial\tau} = \frac{2bD\tau}{\lambda_{11}^2}, \tag{27b}$$

where

$$D \equiv A_{11}B_{01} - A_{01}B_{11}, \tag{28}$$

it follows that

$$\varphi = 4bDz + c_1, \tag{29}$$

where c_1 is a (real) constant of integration. Consequently, from Eqs. (3), (5) and (7) we have

$$\mathscr{E} = -\lambda_{11} + i(4bDz + c_1), \tag{30}$$

$$f' = -\frac{\lambda_{11}}{|\Lambda|^2}, \tag{31}$$

where

$$\Lambda = (1 + \bar{B}_0^2\lambda_{11}) - i\bar{B}_0^2(4bDz + c_1) \tag{32}$$

We now proceed to the nontrivial part of the prescription viz., the determination of ω'. As mentioned earlier we do it via the potential χ. Equations (9) defining χ are the two partial differential equations

$$\frac{\partial\chi}{\partial\rho} = \frac{\rho}{f^2}\frac{\partial}{\partial z}(f^2 + \varphi^2), \tag{33a}$$

$$\frac{\partial\chi}{\partial z} = -\frac{\rho}{f^2}\frac{\partial}{\partial\rho}(f^2 + \varphi^2). \tag{33b}$$

In this case f and φ are given by Eqs. (24) and (29) so that Eqs. (33) become

$$\frac{\partial\chi}{\partial\tau} = \frac{4bD(4bDz + c_1)\tau^{2b-1}}{(A_{11}\tau^{2b} + B_{11})^2}, \tag{34a}$$

$$\frac{\partial\chi}{\partial z} = -2\left[1 + b - \frac{2bA_{11}}{(A_{11}\tau^{2b} + B_{11})}\right]. \tag{34b}$$

We shall now write Eq. (8) as

$$\frac{\partial}{\partial \rho}(\omega' - \omega) = \overline{B}_0^4\left[\varphi \frac{\partial \chi}{\partial \rho} - (f^2 + \varphi^2)\frac{\partial \omega}{\partial \rho}\right], \tag{35a}$$

$$\frac{\partial}{\partial z}(\omega' - \omega) = \overline{B}_0^4\left[\varphi \frac{\partial \chi}{\partial z} - (f^2 + \varphi^2)\frac{\partial \omega}{\partial z}\right]. \tag{35b}$$

Employing Eqs. (24), (27), (29) and (34), Eqs. (35) become after some simplification

$$\frac{\partial}{\partial \tau}(\omega' - \omega) = 2bD\overline{B}_0^4\left[\frac{(4bDz + c_1)^2\tau^{2b-1}}{(A_{11}\tau^{2b} + B_{11})^2} - \tau\right], \tag{36a}$$

$$\frac{\partial}{\partial z}(\omega' - \omega) = -2\overline{B}_0^4(4bDz + c_1) \times\left(1 + b - \frac{2bB_{11}}{(A_{11}\tau^{2b} + B_{11})}\right). \tag{36b}$$

Integrating Eq. (36b) gives

$$\omega' - \omega = -2\overline{B}_0^4 z(2bDz + c_1) \times\left(1 + b - \frac{2bB_{11}}{(A_{11}\tau^{2b} + B_{11})}\right) + c_1'(\tau). \tag{37}$$

Differentiating Eq. (37) w.r.t τ and using Eq. (36a) after some algebra and use of the normalization Eq. (19), we find $c_1'(\tau)$ satisfies the equation

$$\frac{\partial c_1'}{\partial \tau} = 2bD\overline{B}_0^4\left[-\tau + \frac{c_1^2\tau^{2b-1}}{(A_{11}\tau^{2b} + B_{11})^2}\right], \tag{38}$$

which can be integrated to yield

$$c_1' = -bD\overline{B}_0^4\left[\tau^2 + 1\frac{c_1^2}{bA_{11}(A_{11}\tau^{2b} + B_{11})}\right] + c_2, \quad A_{11} \neq 0, \tag{39}$$

where c_2 is another constant of integration. Equation (37) thus becomes

$$\omega' = \omega - \overline{B}_0^4\left[bD\left\{\tau^2 + 4z^2\left(1 + b - \frac{2bB_{11}}{(A_{11}\tau^{2b} + B_{11})}\right)\right\} + 2zc_1\left(1 + b - \frac{2bB_{11}}{(A_{11}\tau^{2b} + B_{11})}\right) + \frac{Dc_1^2}{A_{11}(A_{11}\tau^{2b} + B_{11})}\right] + c_2, \quad A_{11} \neq 0. \tag{40}$$

For completeness we mention that Eqs. (34) can be integrated to obtain the potential χ. The result is

$$\chi = -2z\left(1 + b - \frac{2bA_{11}}{(A_{11}\tau^{2b} + B_{11})}\right) - \frac{2Dc_1}{A_{11}(A_{11}\tau^{2b} + B_{11})} + c_2', \quad A_{11} \neq 0. \tag{41}$$

In the case $A_{11} = 0$ the equations are simpler and we only quote the results

$$\omega' = \omega - 2c_1\overline{B}_0^4(1 - b)z + c_2, \tag{42}$$

with

$$\chi = -2(1 - b)z + c_2'. \tag{43}$$

Having obtained f' and ω' the magnetization procedure is complete. The new magnetic solution as mentioned before is given by Eq. (1) with f and ω replaced by f' and ω' respectively and is of the form

$$ds^2 = |\Lambda|^2\left[2e^{2\psi}(d\rho^2 + dz^2) - \frac{\rho^2}{\lambda_{11}}dt^2\right] + \frac{\lambda_{11}}{|\Lambda|^2}(d\phi - \omega' dt)^2, \tag{44}$$

where

$$|\Lambda|^2 = (1 + \overline{B}_0^2\lambda_{11})^2 + \overline{B}_0^4(4bDz + c_1)^2. \tag{45}$$

IV. THE NEW MAGNETIZED SOLUTIONS

Before examining the new solutions, we make the following remarks which would clarify our classification of the solutions. From Eqs. (21) and (28) the angular momentum per unit length of the original metric is seen to be

$$j = -\tfrac{1}{2}bD. \tag{46}$$

Hence,

$$j = 0 \Leftrightarrow D = 0. \tag{47}$$

Further from Eqs. (28) and (19),

$$D^2 = -\tfrac{1}{2}A_{11}B_{11}. \tag{48}$$

Since $\lambda_{11} \neq 0$, it follows that $D = 0$ iff $A_{11} = 0$ or $B_{11} = 0$ but not both. It is easy to show with a little algebra that $D = 0$ iff $\omega =$ constant and consequently we have

$$j = 0 \Leftrightarrow D = 0 \Leftrightarrow \omega = \text{constant} \Leftrightarrow A_{11} = 0 \text{ or } B_{11} = 0 \text{ (but not both)}. \tag{49}$$

Further it is easy to see that any overall constant c_2 in ω' can be transformed away by a coordinate transformation of the type

$$\overline{\phi} = \phi - c_2 t, \quad \overline{t} = t. \tag{50}$$

Such a transformation does not affect the general form of the metric and hence there is no loss of generality in setting $c_2 = 0$ in Eqs. (40) and (42) for ω'. Also as expected if $\overline{B}_0 = 0$ in Eq. (45) one gets back to the original unmagnetized metric. With these remarks we classify the solutions as follows:

A. D = 0 solutions

As pointed out earlier these correspond to solutions with constant ω which can be transformed away by a suitable transformation of the form Eq. (50). Hence they correspond to magnetized versions of *static cylindrically symmetric solutions*. Moreover since $D = 0$, $\varphi = c_1$ and therefore in general these solutions have a constant twist. These solutions fall into two classes according to whether $B_{11} = 0$ or $A_{11} = 0$, each representative of two well known solutions as seen below.

1. Generalized Twisted Melvin Magnetic Universe (GTMMU)

This class corresponds to the particular choice of constants

$$B_{11} = 0, \quad A_{11} = \tfrac{1}{2}, \quad A_{01} = 0, \quad c = \tfrac{1}{2}, \tag{51}$$

which implies

$$B_{01} = 0 \quad \text{and} \quad B_{00} = -1. \tag{52}$$

For the original unmagnetized metric

$$m = \tfrac{1}{4}(1 - b), \tag{53}$$

so that $m \geqslant 0$ iff $b \leqslant 1$. The metric is given by Eq. (45) where

$$\lambda_{11} = \frac{\rho^2}{\epsilon^2}, \quad e^{2\psi} = \tfrac{1}{2}\tau^{b^2 - 1}, \tag{54}$$

$$\omega' = -2\overline{B}_0^4 c_1(1 + b)z, \tag{55}$$

$$|\Lambda|^2 = \left(1 + \frac{\overline{B}_0^2 \rho^2}{\epsilon^2}\right)^2 + \overline{B}_0^4 c_1^2, \tag{56}$$

$$\epsilon^2 = \tau^{1-b}, \quad \epsilon > 0. \tag{57}$$

If $b = 1$ and $c_1 = 0$ we get the well known Melvin Magnetic Universe (MMU) whereas if $b = 1$ and $c_1 \neq 0$ we get the Twisted Melvin Magnetic Universe (TMMU) discussed by Harrison.[1] In the case of a "twisted" solution the angular velocity ω' varies montonically with z becoming 0 on the $z = 0$ plane. The case $b \neq 1, c_1 \neq 0$ corresponds to a Generalized Melvin Magnetic Universe (GMMU).

It should be noticed that other solutions, e.g., $B_{11} = B_{01} = 0, A_{11} = \frac{1}{2}, B_{00} = -1, c = \frac{1}{2}$, but $A_{01} \neq 0$, $A_{00} \neq 0$ can be transformed into a suitable one of the above solutions by a coordinate transformation of type Eq. (50). Similarly, solutions with $A_{11} \neq \frac{1}{2}, B_{11} \neq -1$, can be transformed by a scaling of coordinates t *and* ϕ.[9]

2. Twisted Weyl–Levi-Cività magnetic solution (TWLCM)

It is known that the special case

$$A_{11} = 0, \quad B_{11} = \tfrac{1}{2}, \quad B_{01} = 0, \tag{58}$$

corresponds to the Weyl–Levi-Cività solution whose source is an infinite rod of linear density

$$m = \tfrac{1}{4}(1 + b), \tag{59}$$

or the infinite mass cylindrical limit of the augmented Schwarzschild solution.[10] Our solution for the above choice of parameters has a constant twist c_1 leading to the above nomenclature. In this case

$$\lambda_{11} = \frac{\rho^2}{\bar{\epsilon}^2}, \quad e^{2\psi} = c\tau^{b^2 - 1}, \tag{60}$$

$$\omega' = -2\overline{B}_0^4 c_1(1 + b)z, \tag{61}$$

$$|\Lambda|^2 = \left(1 + \frac{\overline{B}_0^2 \rho^2}{\bar{\epsilon}^2}\right)^2 + \overline{B}_0^4 c_1^2, \tag{62}$$

where

$$\bar{\epsilon}^2 = \tau^{1+b}, \quad \bar{\epsilon} > 0. \tag{63}$$

If $c_1 = 0$ we get the Weyl–Levi-Cività magnetic solution (WLCM). The particular case $b = 1$ corresponds to the cylindrical limit of the Schwarzschild solution and in this case

$$\lambda_{11} = \tfrac{1}{2}, \quad e^{2\psi} = c \tag{60a}$$

$$\omega' = -4\overline{B}_0^4 c_1 z, \tag{61a}$$

$$|\Lambda|^2 = [1 + (\overline{B}_0^2/2)]^2 + \overline{B}_0^4 c_1^2, \tag{62a}$$

with

$$\bar{\epsilon} = \tau. \tag{63a}$$

In this case if $c_1 = 0$ then the new solution is reducible to the original solution by a scaling of coordinates so that the Harrison transformation is an identity transformation. The peculiarity of this case will be more apparent when we look at the electromagnetic fields.

As before if $B_{01} \neq 0$ the solution may be transformed to one of the above by a coordinate transformation. We thus find that under the magnetization procedure the static solutions retain their original cylindrical symmetry only if there is no twist, i.e., $c_1 = 0$.

B. $D \neq 0$, $c_1 = 0$ solution

In this case the original background is stationary. ω is not a constant and therefore cannot be globally transformed away. These solutions correspond to nonvanishing j and consequently the magnetized version of these solutions form a new class of solutions. The relevant function for this metric are given by

$$\lambda_{11} = A_{11}\tau^{1+b} + B_{11}\tau^{1-b}, \tag{64}$$

$$\omega' = \omega - bD\overline{B}_0^4\left[\tau^2 + 4z^2\left(1 + b - \frac{2bB_{11}}{(A_{11}\tau^{2b} + B_{11})}\right)\right], \tag{65}$$

$$|\Lambda|^2 = (1 + \overline{B}_0^2\lambda_{11})^2 + 16b^2D^2\overline{B}_0^4 z^2. \tag{66}$$

We thus see that for this class of solutions even though $c_1 = 0$ the metrics are axially symmetric and have lost their original cylindrical symmetry under the magnetization transformation. Moreover, these solutions have reflection symmetry with respect to $z = 0$ plane (the z^2 dependence) and are qualitatively very different from the previously discussed metrics with constant twist.

C. $D \neq 0$, $c_1 \neq 0$ solution

These correspond to the most general of our solutions and represent "twisted" versions of the previous solutions. The most general form is given by Eq. (45) with ω' given by Eq. (40).

V. THE ELECTROMAGNETIC FIELDS

We conclude this investigation by computing the electromagnetic fields generated by the Harrison type transformation or in other words the magnetization prescription. As mentioned in Sec. II the fields as seen by a locally nonrotating observer (LNRO) are directly calculable from the potential Φ'. For any metric of the form given by Eq. (1) a LNRO is specified by a tetrad whose nonvanishing components are given by

$$\begin{aligned} &e_{\hat{0}}{}^t = f^{1/2}\rho^{-1}, \\ &e_{\hat{1}}{}^\rho = e_{\hat{2}}{}^z = f^{1/2}P, \\ &e_{\hat{3}}{}^\phi = f^{-1/2}, \quad e_{\hat{0}}{}^\phi = \omega e_{\hat{0}}{}^t. \end{aligned} \tag{67}$$

Employing Eqs. (3) and (7) in Eq. (11), Eq. (12) yields in general

$$A_3 = -\frac{\overline{B}_0[f - \overline{B}_0^2(f^2 + \varphi^2)]}{|\Lambda|^2}, \tag{68a}$$

$$A_0' = -\frac{\overline{B}_0\varphi}{|\Lambda|^2}, \tag{68b}$$

which for our specific case gives

$$A_3 = \frac{\overline{B}_0[\lambda_{11} + \overline{B}_0^2(\lambda_{11}^2 + \varphi^2)]}{|\Lambda|^2}, \tag{69a}$$

$$A_0' = -\frac{\overline{B}_0\varphi}{|\Lambda|^2}, \tag{69b}$$

with φ and $|\Lambda|^2$ being given by Eqs. (29) and (45) respectively.

Using Eqs. (14), (12) and (69), after a straightforward computation the electromagnetic field components may be written down in all the cases as

$$B_{\hat{\rho}}^{NR} = -\frac{8bDP\varphi(1 + \overline{B}_0^2\lambda_{11})}{|\Lambda|^4}, \tag{70a}$$

$$B_{\hat{z}}^{NR} = \frac{\sqrt{2}\overline{B}_0 P\lambda_{11}\tau[(1 + \overline{B}_0^2\lambda_{11})^2 - \overline{B}_0^4\varphi^2]}{|\Lambda|^4}, \tag{70b}$$

$$E_{\hat{\rho}}^{NR} = \frac{4bD\overline{B}_0 P[(1 + \overline{B}_0^2\lambda_{11})^2 - \overline{B}_0^4\varphi^2]}{|\Lambda|^4}, \tag{70c}$$

$$E_{\hat{z}}^{NR} = 2\frac{\sqrt{2}\overline{B}_0^3 P\lambda_{11,\tau}\varphi[1 + \overline{B}_0^2\lambda_{11}]}{|\Lambda|^4}, \tag{70d}$$

with λ_{11}, P, D, φ and $|\Lambda|^2$ as defined by Eqs. (16), (23), (28), (29) and (45) respectively. To see the structure more clearly we write down the particular forms of the fields in each of the cases discussed in the previous section.

A1. GTMMU fields

In this case,

$$B_{\hat{\rho}}^{NR} = E_{\hat{\rho}}^{NR} = 0,$$

and

$$B_{\hat{z}}^{NR} = (1 + b)\overline{B}_0\epsilon^b \frac{[(1 + \overline{B}_0^2\rho^2/\epsilon^2)^2 - \overline{B}_0^4 c_1^2]}{[(1 + \overline{B}_0^2\rho^2/\epsilon^2)^2 + \overline{B}_0^4 c_1^2]^2}, \tag{71a}$$

$$E_{\hat{z}}^{NR} = \frac{2(1 + b)\overline{B}_0^3 c_1\epsilon^b(1 + \overline{B}_0^2\rho^2/\epsilon^2)}{[(1 + \overline{B}_0^2\rho^2/\epsilon^2)^2 + \overline{B}_0^4 c_1^2]^2}. \tag{71b}$$

Although the metric in general has a z dependence the electromagnetic fields in a LNRF are independent of z. Further as $\rho \to \infty$, $B_{\hat{z}}^{NR} \sim \tau^{-(b^2 + 3b + 4)}$, and $E_{\hat{z}}^{NR} \sim \tau^{-(b^2 + 5b + 6)}$ so that as $\rho \to \infty$, $B_{\hat{z}}^{NR}$ and $E_{\hat{z}}^{NR} \to 0$. It is interesting to note that as $\rho \to 0$, $B_{\hat{z}}^{NR} \sim \tau^{b(1-b)/2}$, $E_{\hat{z}}^{NR} \sim \tau^{b(1-b)/2}$ so that for solutions with linear density positive, $m > 0$, i.e., $b < 1$, both $B_{\hat{z}}$ and $E_{\hat{z}}$ vanish on the axis.[11]

A2. TWLCM fields

In this case too

$$B_{\hat{\rho}}^{NR} = E_{\hat{\rho}}^{NR} = 0, \tag{72a}$$

while

$$B_{\hat{z}}^{NR} = \frac{(1 - b)\overline{B}_0}{\overline{\epsilon}^b} \frac{[(1 + \overline{B}_0^2\rho^2/\epsilon^2)^2 - \overline{B}_0^4 c_1^2]}{[(1 + \overline{B}_0^2\rho^2/\overline{\epsilon}^2)^2 + \overline{B}_0^4 c_1^2]^2}, \tag{72b}$$

$$E_{\hat{z}}^{NR} = \frac{2(1 - b)\overline{B}_0^3 c_1}{\overline{\epsilon}^b} \frac{(1 + \overline{B}_0^2\rho^2/\overline{\epsilon}^2)}{[(1 + \overline{B}_0^2\rho^2/\overline{\epsilon}^2)^2 + \overline{B}_0^4 c_1^2]^2}. \tag{72c}$$

Here also the fields are z independent and as $\rho \to \infty$, $B_{\hat{z}}^{NR} \sim \tau^{-[(b-3/2)^2 + 7/4]} \to 0$. However $E_{\hat{z}}^{NR} \sim \tau^{-(b-3)(b-2)}$ so that $E_{\hat{z}}^{NR} \to 0$ unless $2 < b < 3$. In this case on the axis the fields are divergent. $b = 1$ is a degenerate case as seen from, Eq. (72), since there are no electromagnetic fields. Thus the cylindrical limit of the Schwarzschild solution under the Harrison transformation does not pick up any electromagnetic field but a constant vector potential even if $c_1 \neq 0$.

In both the cases it should be noted that if the constant twist $c_1 = 0$ then only a magnetic field in the z direction is produced but no electric field. The magnetic field for GMMU is given by

$$B_{\hat{z}}^{NR} = \frac{(1 + b)\overline{B}_0\epsilon^b}{(1 + \overline{B}_0^2\rho^2/\epsilon^2)^2}, \tag{73}$$

which for MMU becomes

$$B_{\hat{z}}^{NR} = \frac{2\overline{B}_0}{(1 + \overline{B}_0^2\rho^2)^2}, \tag{74}$$

whilst that of WLCM becomes

$$B_{\hat{z}}^{NR} = \frac{(1 - b)\overline{B}_0}{\overline{\epsilon}^b(1 + \overline{B}_0^2\rho^2/\overline{\epsilon}^2)^2}. \tag{75}$$

For the TMMU however

$$B_{\hat{z}}^{NR} = 2\overline{B}_0 \frac{[(1 + \overline{B}_0^2\rho^2)^2 - \overline{B}_0^4 c_1^2]}{[(1 + \overline{B}_0^2\rho^2)^2 + \overline{B}_0^4 c_1^2]^2}, \tag{76a}$$

$$E_{\hat{z}}^{NR} = \frac{4\overline{B}_0^3 c_1(1 + \overline{B}_0^2\rho^2)^2}{[(1 + \overline{B}_0^2\rho^2)^2 + \overline{B}_0^4 c_1^2]^2}. \tag{76b}$$

B. $D \neq 0$, $c_1 = 0$

The structure of the electromagnetic fields in this case is very different from the previous case because of the nontrivial rotation present in the metric. The fields are no longer z independent and are given as

$$B_{\hat{\rho}}^{NR} = -\frac{32b^2D^2\overline{B}_0^3 P(1 + \overline{B}_0^2\lambda_{11})z}{|\Lambda|^4}, \tag{77a}$$

$$B_{\hat{z}}^{NR} = \frac{\sqrt{2}\overline{B}_0 P\lambda_{11,\tau}[(1 + \overline{B}_0^2\lambda_{11})^2 - 16b^2D^2\overline{B}_0^4 z^2]}{|\Lambda|^4}, \tag{77b}$$

$$E_{\hat{\rho}}^{NR} = \frac{4bD\overline{B}_0 P[(1 + \overline{B}_0^2\lambda_{11})^2 - 16b^2D^2\overline{B}_0^4 z^2]}{|\Lambda|^4}, \tag{77c}$$

$$E_{\hat{z}}^{NR} = \frac{8\sqrt{2}bD\overline{B}_0^3 P\lambda_{11,\tau}(1 + \overline{B}_0^2\lambda_{11})z}{|\Lambda|^4}. \tag{77d}$$

It is easy to see that in this case also the fields vanish as $\rho \to \infty$ for a fixed value of z. Using the tetrad components listed in Eq. (67) and its inverse the coordinate components of the electromagnetic field tensor may be obtained. We have done this and verified explicitly that these fields satisfy the generally covariant Maxwell's equations without any sources.

However, no one of the above solutions correspond to the asymptotic form of the magnetized Kerr solution.

[1] B. K. Harrison, J. Math. Phys. **9**, 1744 (1968).

[2] F. J. Ernst, Phys. Rev. **167**, 1175 (1968), **168**, 1415 (1968); J. Math. Phys. **17**, 54 (1976).

[3] See e.g., D. V. Gal'tsov and V. I. Petukhov, Zh. Eksp. Teor. Fiz. **74**, 801 (1978) [Sov. Phys. JETP **47**, 419 (1978)]; N. Dadhich, C. Hoenselaers and C. V. Vishveshwara, J. Phys. A **12**, 215 (1979); N. Dadhich and P. J. Wiita,

J. Phys. A **15**, 2645 (1982).
[4]F. J. Ernst and W. J. Wild, J. Math. Phys. **17**, 182 (1976).
[5]C. V. Vishveshwara and J. Winicour, J. Math. Phys. **18**, 1280 (1977).
[6]M. A. Melvin, Phys. Lett. **8**, 65 (1964).
[7]W. A. Hiscock, J. Math. Phys. **22**, 1828 (1981).
[8]Our $\overline{B}_0$ corresponds to Ernst's $B_0/2$ and his index 4 to our 0, e.g. A_4, A'_4 to our A_0 and A'_0.
[9]The magnetic field $\overline{B}_0$ is scaled suitably. However, the new ϕ has periodicity different from 2π.
[10]J. Winicour, A. J. Janis, and E. T. Newman, Phys. Rev. **176**, 1507 (1968).
[11]The $\rho \to 0$ limit may not be realized since the vacuum solution would be matched onto some suitable interior solution to take care of the singularity of the metric on the axis.

Some exact solutions of Einstein–Dirac–Maxwell zero mass scalar equations

K. D. Krori and R. Bhattacharya
Mathematical Physics Forum, Cotton College, Gauhati-781 001, India
Institute of Advanced Study in Science and Technology, Assam Science Society, Gauhati-781 001, India

D. Nandy
St. Anthony's College, Shillong-793 003, India

(Received 18 June 1982; accepted for publication 2 September 1982)

In this paper, some exact solutions of Einstein–Dirac–Maxwell zero mass scalar field equations are given.

PACS numbers: 04.20.Jb

I. INTRODUCTION

Recently two exact solutions of the Einstein–Dirac field equations have been given in the plane-symmetric metric.[1,2] In one of the solutions,[1] the Dirac field is taken as time-independent while in the other[2] the Dirac field is time-dependent. Later Krori, Chaudhury, and Bhattacharya[3] have presented some exact solutions of the Einstein–Dirac–Maxwell equations in cylindrical and plane-symmetric metrics.

Weinberg[4] and Wilczek[5] have recently suggested that there should exist a pseudoscalar boson, the so-called axion, of negligible mass. Peccei and Quinn[6] have shown that such a particle would explain the absence of CP nonconservation in strong interactions in particle physics. It is well known that any light particle has a potential for playing a major role in stellar energy loss so that there may exist a cosmic background of these particles.[7]

Viewed in the above perspective, the Einstein zero mass scalar theory acquires some relevance. This adds to the importance of an exact solution, derived earlier by Singh,[8] of Einstein's field equations in the presence of zero mass scalar field. In this solution both the metric tensor and the scalar field have plane symmetry. It may, therefore, be of some interest to present some exact solutions of Einstein's field equations in the presence of Dirac, Maxwell, and zero mass scalar fields in the plane-symmetric metric. We derive some such solutions in this paper. Our solutions are quite different from those of Refs. 1, 2, and 8.

II. FIELD EQUATIONS

The field equations of the Einstein–Maxwell–Dirac zero mass scalar field are

$$R_{\mu\nu} - \tfrac{1}{2}g_{\mu\nu}R = -8\pi[E_{\mu\nu} + S_{\mu\nu} + T_{\mu\nu}], \tag{1}$$

$$F^{\alpha\beta}{}_{;\beta} = 0, \tag{2}$$

$$F_{\alpha\beta;\nu} + F_{\beta\nu;\alpha} + F_{\nu\alpha;\beta} = 0, \tag{3}$$

$$\gamma^{\mu}\nabla_{\mu}\psi = 0, \tag{4}$$

$$g^{\mu\nu}\phi_{;\mu\nu} = 0, \tag{5}$$

where the energy-momentum tensor for electromagnetic, Dirac, and scalar fields are, respectively, given by

$$E_{\mu\nu} = -F_{\mu\alpha}F_{\nu}{}^{\alpha} + \tfrac{1}{4}g_{\mu\nu}F_{\alpha\beta}F^{\alpha\beta}, \tag{6}$$

$$T_{\mu\nu} = \tfrac{1}{4}[\psi^{+}\gamma_{\mu}\nabla_{\nu}\psi + \psi^{+}\gamma_{\nu}\nabla_{\mu}\psi - (\nabla_{\mu}\psi^{+})\gamma_{\nu}\psi - (\nabla_{\nu}\psi^{+})\gamma_{\mu}\psi], \tag{7}$$

$$S_{\mu\nu} = \phi_{,\mu}\phi_{,\nu} - \tfrac{1}{2}g_{\mu\nu}(g^{lm}\phi_{,l}\phi_{,m}). \tag{8}$$

We use units in which $\hbar = c = 1$. We adopt the conventions of Jauch and Rohrlich[9] for Dirac γ matrices and the notations of Brill and Wheeler[10] with regard to ψ^{+}, ψ^{*}, and $\nabla_{\mu}\psi$.

We choose the plane-symmetric line element

$$ds^2 = e^{2u}(dt^2 - dx^2) - e^{2v}(dy^2 + dz^2), \tag{9}$$

where u and v are functions of x alone.

A. Time-independent Dirac field

When the Dirac field ψ is time-independent, Eqs. (4) and (9) give

$$\psi = e^{-(v + u/2)}\psi_0, \tag{10}$$

where ψ_0 is a constant spinor. The nonvanishing components of $T_{\mu\nu}$ are

$$T_{20} = \tfrac{1}{4}e^{-u}(v_{,1} - u_{,1})\psi^{+}\gamma^{1}\gamma^{2}\gamma^{0}\psi, \tag{11}$$

$$T_{30} = \tfrac{1}{4}e^{-u}(v_{,1} - u_{,1})\psi^{+}\gamma^{1}\gamma^{2}\gamma^{0}\psi, \tag{12}$$

where a comma denotes the differentiation with respect to x. Since $R_{20} = R_{30} = 0$ this implies

$$T_{20} = T_{30} = 0. \tag{13}$$

Equations (11)–(13) give

$$\psi_0 = \alpha_0 \begin{pmatrix} 1 \\ \pm 1 \\ i \\ \pm i \end{pmatrix}, \tag{14}$$

where α_0 is a constant. Thus ψ is obtained from Eq. (10) when u and v are obtained from the solutions.

Equations (2), (3), and (9) give the electromagnetic field,

$$F_{01} = C_1 e^{-2v}, \qquad F_{23} = C_2 e^{-2v}, \tag{15}$$

where C_1 and C_2 are constants.

With the help of Eq. (9), Eq. (1) reduces to

$$v_{,1}^2 + 2u_{,1}v_{,1} = ae^{2u-4v} + be^{-4v}, \tag{16}$$

$$2v_{,11} - 2u_{,1}v_{,1} + 3v_{,1}^2 = ae^{2u-4v} - be^{-4v}, \tag{17}$$

$$u_{,11} + v_{,11} + v_{,1}^2 = -ae^{2u-4v} - be^{-4v}. \tag{18}$$

Here

$$a = -4\pi(C_1^2 + C_2^2), \tag{19}$$

$$b = 4\pi d^2 \tag{20}$$

 0022-2488/83/061574-03$02.50 © 1983 American Institute of Physics

and

$$\phi_{,1} = de^{-2v}, \tag{21}$$

where d is a constant. Adding Eqs. (16) and (17) one gets

$$v_{,11} + 2v_{,1}^2 = ae^{2u-4v}. \tag{22}$$

B. Time-dependent Dirac field

We assume the Dirac field ψ is a function of x and t and without loss of generality we choose

$$\psi = \psi_0(x)e^{-i\omega t}, \tag{23}$$

where $\psi_0(x)$ is a spinor function and ω is a real constant. Equations (4), (9), and (23) give

$$\psi = e^{-(v+u/2)}(\cos\omega x + i\gamma^1\gamma^0 \sin\omega x)e^{-i\omega t}\psi_C, \tag{24}$$

where ψ_C is an arbitrary constant spinor. The nonvanishing components of $T_{\mu\nu}$ are

$$T_{00} = T_{11} = \tfrac{1}{4}e^{-u}\psi + (4i\omega\gamma^0)\psi, \tag{25}$$

$$T_{10} = T_{01} = \tfrac{1}{4}e^{-u}\psi^+(-4i\omega\gamma^1)\psi, \tag{26}$$

$$T_{20} = T_{02} = \tfrac{1}{4}e^{-u}\psi^+[-2i\omega\gamma^2 + \gamma^1\gamma^2\gamma^0(v_{,1} - u_{,1})]\psi, \tag{27}$$

$$T_{30} = T_{03} = \tfrac{1}{4}e^{-u}\psi^+[-2i\omega\gamma^3 + \gamma^1\gamma^3\gamma^0(v_{,1} - u_{,1})]\psi. \tag{28}$$

Since $R_{01} = R_{02} = R_{03} = 0$, this implies

$$T_{01} = T_{02} = T_{03} = 0. \tag{29}$$

Equations (26)–(29) give

$$\psi_C = \begin{pmatrix} S \\ \pm S \\ q \\ \pm q \end{pmatrix} e^{i\lambda}, \tag{30}$$

where S, q, λ are real constants. Thus ψ is obtained from (24), when u and v are obtained from the solutions.

In this case the field equations are

$$v_{,1}^2 + 2u_{,1}v_{,1} = ae^{2u-4v} + be^{-4v} - 8\pi e^{2u}T_{11}, \tag{31}$$

$$2v_{,11} - 2u_{,1}v_{,1} + 3v_{,1}^2 = ae^{2u-4v} - be^{-4v} + 8\pi e^{2u}T_{00}, \tag{32}$$

$$u_{,11} + v_{,11} + v_{,1}^2 = -ae^{2u-4v} - be^{-4v}. \tag{33}$$

Adding (31) and (32) one gets

$$v_{,11} + 2v_{,1}^2 = ae^{2u-4v}. \tag{34}$$

Equations (33) and (34) are identical to Eqs. (18) and (22). Thus the solutions of Eqs. (16)–(18) are the solutions of Eq. (31)–(33).

III. SOLUTIONS

Eliminating u from Eqs. (16)–(18), one gets

$$v_{,11} + 3v_{,1}^2 - 2k_1v_{,1}e^{-2v} + be^{-4v} = 0, \tag{35}$$

where k_1 is a nonzero positive integrational constant.

Putting $p = 1/\alpha(v)$, where $dv/dx = p$, we have

$$\frac{d\alpha}{dv} = \alpha[3 - 2k_1e^{-2v}\alpha + be^{-4v}\alpha^2]$$

$$\equiv \alpha[f_1 - f_2\alpha + f_3\alpha^2]. \tag{36}$$

Eqution (36) is a special case of the Abel equation of the first kind.[11]

Putting $\alpha(v) = \phi(z)\psi(v)$, where $z = \int\psi(v)f_2dv$ and $\psi(v) = e^{\int f_1dv}$ in (36), we get

$$\frac{d\phi}{dz} = \phi^2 + g(z)\phi^3, \tag{37}$$

where

$$g(z) = \psi(v)\frac{f_3}{f_2}.$$

Assuming $dz/d\tau + 1/\tau\phi(z) = 0$ and using Eq. (37), we get

$$\tau^2\left[\frac{d^2v}{d\tau^2} + \left(\frac{dv}{d\tau}\right)^2\right] + \frac{b}{4k_1^2} = 0. \tag{38}$$

Putting $dv/d\tau = P$ and $P\tau = x$, one gets

$$\tau\frac{dx}{d\tau} = -[(x - \tfrac{1}{2})^2 - \tfrac{1}{4}(1 - b/4k_1^2)]. \tag{39}$$

Case a: when $b < k_1^2$ and $\tau(dv/d\tau) > \frac{1}{2}(A+1)$, Eq. (39) gives

$$e^{2v} = \frac{(\tau^A - 1)^2}{\tau^{A-1}}$$

and the complete solutions of Eqs. (16)–(18) are

$$e^{2v} = \frac{(\tau^A - 1)^2}{\tau^{A-1}}, \tag{40}$$

$$e^{2u} = \frac{\tau^A}{(\tau^A - 1)^2}, \tag{41}$$

$$2k_1x + k_2 = \frac{\tau^{A+1}}{A+1} + \frac{\tau^{1-A}}{1-A} - 2\tau, \tag{42}$$

where

$$A^2 = 1 - b/k_1^2, \tag{43}$$

$$a = 4(b - k_1^2), \tag{44}$$

and k_2 is a constant of integration.

Case b: When $b < k_1^2$ and $\tau\, dv/d\tau < \frac{1}{2}(A+1)$ Eq. (39) gives

$$e^{2v} = \frac{(\tau^A + 1)^2}{\tau^{A-1}}$$

and the complete solutions of Eqs. (16)–(18) are

$$e^{2v} = \frac{(\tau^A + 1)^2}{\tau^{A-1}}, \tag{45}$$

$$e^{2u} = \frac{k_3\tau^A}{(\tau^A + 1)^2}, \tag{46}$$

$$2k_1x + k_2 = \frac{\tau^{A+1}}{A+1} + \frac{\tau^{1-A}}{1-A} + 2\tau, \tag{47}$$

where k_3 is a constant of integration, and

$$4(k_1^2 - b) = ak_3. \tag{48}$$

In this case we get a solution with negative a if k_3 is negative. But then e^{2u} becomes negative, as (46) shows. This means that x and t exchange their roles in the metric.

Case c: when $b = k_1^2$ Eq. (39) gives

$$e^{2v} = k_5^2\tau[\ln(k_4\tau)]^2$$

and the complete solutions of Eqs. (16)–(18) are

$$e^{2v} = k_5^2\tau\{\ln(k_4\tau)\}^2, \tag{49}$$

$$e^{2u} = \frac{1}{k_6^2\{\ln(k_4\tau)\}^2}, \tag{50}$$

$$\frac{2k_1}{k_5^2}x + k_7 = \tau[\{\ln(k_4\tau)\}^2 - 2\{\ln(k_4\tau)\} + 2], \tag{51}$$

where k_4, k_5, k_6, k_7 are constants of integration and

$$a + 4k_1^2k_6^2 = 0. \tag{52}$$

Case d: when $b > k_1^2$ Eq. (39) gives

$$e^{2v} = k_9\tau[\cos\{\ln(k_8\tau)\}^{B/2}]^2$$

and the complete solutions of Eqs. (16)–(18) are

$$e^{2v} = k_9\tau[\cos\{\ln(k_8\tau)\}^{B/2}]^2, \tag{53}$$

$$e^{2u} = \frac{k_{10}^2}{\cos^2\{\ln(k_8\tau)\}^{B/2}}, \tag{54}$$

$$\frac{2k_1}{k_9}x + k_{11} = \frac{\tau}{2(1+B^2)}[1 + B^2 + \cos\{B\ln(k_8\tau)\} + B\sin\{B\ln(k_8\tau)\}], \tag{55}$$

where k_8, k_9, k_{10}, k_{11} are constants of integration and

$$B^2 = b/k_1^2 - 1, \tag{56}$$

$$ak_{10}^2 + b - k_1^2 = 0. \tag{57}$$

ACKNOWLEDGMENTS

The authors express their profound gratitude to the government of Assam, Dispur, for all facilities provided at Cotton College, Gauhati-781 001, India to prepare this paper. One of them (D.N.) is particularly grateful to Mrs. S. Nandy and Mr. and Mrs. A. K. Aditya for unstinted encouragement.

[1] T. M. Davis and J. R. Ray, Phys. Rev. D **9**, 331 (1974).
[2] K. R. Pechenick and J. M. Cohen, Phys. Rev. D **19**, 1635 (1979).
[3] K. D. Krori, T. Chaudhury, and R. Bhattacharya, Phys. Rev. D **25**, 1429 (1982).
[4] S. Weinberg, Phys. Rev. Lett. **40**, 223 (1978).
[5] F. Wilczek, Phys. Rev. Lett. **40**, 279 (1978).
[6] R. D. Peccei and Quinn, Phys. Rev. D **16**, 1791 (1979).
[7] D. A. Dicus, E. W. Kolb, V. L. Teplitz, and R. V. Wagner, Phys. Rev. D **22**, 839 (1980).
[8] T. Singh, Gen. Relativ. Gravit. **5**, 657 (1974).
[9] J. M. Jauch and F. Rohrlich, *The Theory of Photons and Electrons* (Springer, New York, 1976), Appendix A_2.
[10] D. Brill and J. A. Wheeler, Rev. Mod. Phys. **29**, 465 (1957).
[11] G. M. Murphy, *Ordinary Differential Equations and Their Solutions* (Van Nostrand, New York, 1960).

On spacelike congruences in general relativity

M. Tsamparlis
Department of Physics, Division of Mechanics, University of Athens, Penepistimiopolis, Athens 621, Greece

D. P. Mason
Department of Applied Mathematics, University of the Witwatersrand, Jan Smuts Avenue, Johannesburg 2001, South Africa

(Received 5 February 1982; accepted for publication 18 June 1982)

The theory of spacelike congruences in general relativity is briefly reviewed and the physical interpretation of the rotation tensor $\mathscr{R}_{ab}$, the expansion $\mathscr{E}$, and the shear tensor $\mathscr{S}_{ab}$, of the curves is discussed. It is proved that if the unit tangent vector to any curve of the congruence is everywhere orthogonal to the 4-velocity field u^a of a self-gravitating fluid, then observers comoving with the fluid can be employed along a curve of the congruence if and only if the curves are material curves in the fluid. A congruence of vortex lines is studied in detail. Starting from the Ricci identity for u^a and using Einstein's equations, general expressions in terms of the kinematic quantities and fluid variables are derived for $\mathscr{R}_{ab}$, $\mathscr{E}$, and $\mathscr{S}_{ab}$ for a vortex congruence. It is found that $\mathscr{E}$ and $\mathscr{S}_{ab}$ depend explicitly on the gravitational field only through the magnetic part of the Weyl tensor, and $\mathscr{R}_{ab}$ only through a term proportional to the total energy flux q^a derived from Einstein's equations. With the aid of Maxwell's equations, properties of congruences of magnetic field lines, electric field lines, and a certain combination of vortex and magnetic field lines are determined. For a congruence of magnetic field lines in an electrically conducting fluid and assuming the magnetohydrodynamic approximation of vanishing electric field, it is proved that for a comoving observer, $\mathscr{R}_{ab} = 0$ if and only if the conduction current in the fluid is orthogonal to the magnetic field. The propagation equations for $\mathscr{R}_{ab}$, $\mathscr{E}$, and $\mathscr{S}_{ab}$ along a curve of a spacelike congruence are considered. These equations are developed in full for the special case of a congruence of material curves in a fluid. The divergence, or constraint, equation for the rotation vector is also derived. Where appropriate, corresponding results in Newtonian gravitation theory are given for comparison.

PACS numbers: 04.40. + c

1. INTRODUCTION

Spacelike vectors, such as the vorticity vector and the electric and magnetic field vectors, play an important role in relativistic fluid dynamics and relativistic electrodynamics of continuous media. In general relativity, the theory of spacelike congruences, i.e., families of non-intersecting spacelike curves, was first considered by Greenberg,[1] who established the basic theory and applied it briefly to a congruence of vortex lines. Although other applications have been discussed by several authors,[2–8] the study of spacelike congruences in general relativity has not received the same attention as the study of null or timelike congruences. In this paper we develop the theory of spacelike congruences and consider further applications.

Since several of the applications considered in this paper are concerned with charged electrically conducting fluids we gather together in Sec. 2 Maxwell's equations and the conservation equations for a charged perfect fluid; these results will be referred to throughout the paper. The general theory of spacelike congruences is briefly reviewed and alternative approaches suggested in Sec. 3. To observe the deformations of the spacelike curves and define the expansion $\mathscr{E}$, the shear tensor $\mathscr{S}_{ab}$, and the rotation tensor $\mathscr{R}_{ab}$ of the curves of the congruence, the 4-velocities w^a of the observers employed along a curve cannot be arbitrarily related but must satisfy a transport law for w^a derived by Greenberg.[1] It is proved in Sec. 4 that if the unit tangent vector n^a to any curve of the congruence is everywhere orthogonal to the 4-velocity field u^a of a self-gravitating fluid, then observers comoving with the fluid can be used all along a curve of the congruence if and only if the curves are material curves in the fluid. A material curve in a fluid is a curve which consists at all times of the same fluid particles and hence it moves with the fluid as the fluid evolves; it is sometimes said to be "frozen-in" to the fluid. This result establishes the central role in the general theory played by congruences of curves frozen-in to the fluid. A congruence of vortex lines is studied in detail in Sec. 5. With the aid of Einstein's field equations general expressions in terms of the kinematic quantities and fluid variables are derived for $\mathscr{R}_{ab}$, $\mathscr{E}$, and $\mathscr{S}_{ab}$ for a vortex congruence and the role played by the "free" gravitational field and that part of the curvature determined locally through Einstein's equations is examined. Relativistic generalizations of Helmholtz's three theorems on vortex tubes are obtained. Sections 6 and 7 are concerned, respectively, with congruences of magnetic and electric field lines in a charged electrically conducting fluid. Expressions for the rotation tensor are derived with the aid of Maxwell's equations. In the barotropic flow of a charged perfect fluid the magnetic field and vorticity do not in general separately satisfy conservation laws. For this reason Bekenstein and Oron[9] introduced a spacelike vector field defined as a certain sum of the

 0022-2488/83/061577-17$02.50 © 1983 American Institute of Physics

vorticity and magnetic field. This vector, which we will denote by Ω^a, has interesting properties and in Sec. 8 a congruence of Ω-lines in a charged fluid is briefly considered. Propagation equations for $\mathscr{R}_{ab}$, $\mathscr{E}$, and $\mathscr{S}_{ab}$ along a curve of a spacelike congruence are considered in Sec. 9, the special case of a congruence of material curves in a fluid being analyzed in detail. The divergence, or constraint, equation for the rotation vector is discussed in Sec. 10. Finally, concluding remarks are made in Sec. 11.

For comparison, we give, where appropriate, corresponding results in Newtonian gravitation theory. This allows us to isolate the purely general-relativistic effects.

2. MAXWELL'S EQUATIONS AND THE CONSERVATION EQUATIONS FOR A CHARGED PERFECT FLUID

We collect together for future reference Maxwell's equations and the conservation equations for a charged perfect fluid with 4-velocity field u^a. It is assumed throughout the paper that the dielectric permittivity and magnetic permeability of the fluid are unity. This section also serves to introduce notation and make definitions required in the rest of the paper.[10]

Maxwell's equations are

$$F_{[ab;c]} = 0\,, \quad F^{ab}{}_{;b} = J^a\,, \tag{2.1}$$

where

$$J^a{}_{;a} = 0\,. \tag{2.2}$$

The 4-current J^a, and the electromagnetic field tensor F^{ab} can be split by u^a into

$$J^a = qu^a + \mathscr{J}^a\,, \tag{2.3}$$

$$F^{ab} = u^aE^b - u^bE^a - \eta^{abcd}u_cH_d\,, \tag{2.4}$$

where η^{abcd} is the alternating tensor[11] and

$$q = -u_aJ^a\,, \quad \mathscr{J}^a = h^a{}_bJ^b\,, \tag{2.5}$$

$$E^a = F^{ab}u_b\,, \quad H^a = \tfrac{1}{2}\eta^{abcd}u_bF_{cd}\,. \tag{2.6}$$

q is the charge density, $\mathscr{J}^a$ the conduction current, E^a and H^a the electric and magnetic field 4-vectors, all measured by an observer moving with 4-velocity u^a, and $h^{ab} = g^{ab} + u^au^b$. Maxwell's equations (2.1) can be decomposed with respect to u^a into the following constraint and propagation equations[12]:

$$H^a{}_{;b}h^b{}_a = 2\omega^aE_a\,, \tag{2.7}$$

$$E^a{}_{;b}h^b{}_a = q - 2\omega^aH_a\,, \tag{2.8}$$

$$h^a{}_b\dot{H}^b = u^a{}_{;b}H^b - \theta H^a - I^a(E)\,, \tag{2.9}$$

$$h^a{}_b\dot{E}^b = u^a{}_{;b}E^b - \theta E^a + I^a(H) - \mathscr{J}^a, \tag{2.10}$$

where

$$I^a(E) = \eta^{abcd}u_b(\dot{u}_cE_d - E_{c;d})\,, \tag{2.11}$$

$$I^a(H) = \eta^{abcd}u_b(\dot{u}_cH_d - H_{c;d})\,, \tag{2.12}$$

$$\omega^a = \tfrac{1}{2}\eta^{abcd}u_bu_{c;d}\,, \quad \theta = u^a{}_{;a}\,. \tag{2.13}$$

ω^a is the vorticity vector and θ is the expansion of the fluid. An over dot denotes covariant differentiation along a fluid particle world line, so for example $\dot{H}^a = H^a{}_{;b}u^b$. The spacelike vectors $I^a(E)$ and $I^a(H)$ play an important part in the study of congruences of electric and magnetic field lines. If we operate on (2.11) and (2.12) with $\eta_{arst}u^t$ then a direct calculation yields the following two useful mathematical identities:

$$E_{[r;s]} = u_{[r}\dot{E}_{s]} + \dot{u}_{[r}E_{s]} + u^tE_{t;[r}u_{s]} + \tfrac{1}{2}\eta_{rstf}\,u^tI^f(E)\,, \tag{2.14}$$

$$H_{[r;s]} = u_{[r}\dot{H}_{s]} + \dot{u}_{[r}H_{s]} + u^tH_{t;[r}u_{s]} + \tfrac{1}{2}\eta_{rstf}\,u^tI^f(H)\,. \tag{2.15}$$

The total energy-momentum tensor of a charged perfect fluid is

$$T^{ab} = \mu u^au^b + ph^{ab} + T^{ab}_{EM}\,, \tag{2.16}$$

where p is the isotropic fluid pressure, μ is the total energy density of the fluid as measured by u^a (excluding electromagnetic energy), and T^{ab}_{EM} is the Minkowski electromagnetic energy-momentum tensor:

$$T^{ab}_{EM} = F^{ac}F^b{}_c - \tfrac{1}{4}g^{ab}F_{cd}F^{cd}\,. \tag{2.17}$$

But Maxwell's equations (2.1) imply that

$$T^{ab}_{EM;b} = -F^{ab}J_b\,; \tag{2.18}$$

hence, with the aid of (2.3) and (2.4), it is easily verified that the energy and momentum conservation equations for a charged perfect fluid assume the form

$$\dot{\mu} + (\mu + p)\theta = E_a\mathscr{J}^a\,, \tag{2.19}$$

$$(\mu + p)\dot{u}^a = -p_{,b}h^{ab} + qE^a + \eta^{abcd}u_b\mathscr{J}_cH_d\,. \tag{2.20}$$

If the flow is barotropic, so that $p = p(\mu)$, then (2.20) may be written as

$$u^a = -(\ln\, r)_{,b}h^{ab} + \frac{q}{\mu + p}E^a + \frac{1}{\mu + p}\eta^{abcd}u_b\mathscr{J}_cH_d\,, \tag{2.21}$$

where

$$r = \exp\left(\int_{p_0}^{p}\frac{dp}{p + \mu}\right). \tag{2.22}$$

Finally, using (2.3), the equation of charge conservation (2.2) becomes

$$\dot{q} + q\theta = -\mathscr{J}^a{}_{;a}\,. \tag{2.23}$$

3. GENERAL THEORY OF SPACELIKE CONGRUENCES

A. Rotation, expansion and shear

The theory of spacelike congruences in general relativity was established by Greenberg.[1]

Consider the spacelike congruence

$$x^a = x^a(\eta^\alpha, p)\,, \tag{3.1}$$

where the three parameters $\eta^\alpha (\alpha = 1,2,3)$ specify the particular spacelike curve of the congruence and where p is some parameter along the curves which we take to be the arc length measured along each curve from some initial cross section of the congruence. At any given point P on any one of the curves $\mathscr{C}$, the unit tangent vector to $\mathscr{C}$ at P is defined by

$$n^a = \left(\frac{dx^a}{dp}\right)_{\eta^\alpha}\,; \tag{3.2}$$

we have $n_an^a = +1$ and also, since we assume that n^a forms a vector field, $n_an^a{}_{;b} = 0$.

Consider now two neighboring curves $\mathscr{C}$ and $\mathscr{C}'$ of the congruence. The vector δx^a from the point $P(p)$ on $\mathscr{C}$ to the point $P'(p)$ on $\mathscr{C}'$, where P and P' have the same value of the parameter p, is called a connecting vector:

$$\delta x^a = \frac{\partial x^a}{\partial \eta^\alpha}\,\delta\eta^\alpha\,. \tag{3.3}$$

It can be proved that[1]

$$\underset{n}{\mathscr{L}}\,\delta x^a = 0\,, \tag{3.4}$$

where $\underset{n}{\mathscr{L}}$ stands for the Lie derivative with respect to the vector field n^a.

We now introduce an observer at P with unit 4-velocity w^a such that w^a is orthogonal to n^a at P; we have $w_a w^a = -1$, $w_a n^a = 0$. We note that, given a spacelike vector n^a, there is not a unique timelike unit vector w^a orthogonal to it. For example, the vector $v^a = w^a + \lambda^a$ is another timelike unit vector orthogonal to n^a at P provided

$$\lambda_a n^a = 0\,, \quad \lambda_a \lambda^a + 2 w_a \lambda^a = 0\,. \tag{3.5}$$

This freedom of choice of observer is essential to the theory.[1] To observe the deformations of the curves of the congruence, the observer w^a erects a screen orthogonal to the curve $\mathscr{C}$ at P, so that the congruence of curves passes perpendicularly through the screen at P. Because the connecting vector δx^a need not lie on the screen at P, we introduce the projection tensor

$$p_{ab} = g_{ab} + w_a w_b - n_a n_b\,. \tag{3.6}$$

Then, since $p_{ab} w^b = 0$ and $p_{ab} n^b = 0$, the vector

$$\perp\delta x^a = p^a_b \delta x^b \tag{3.7}$$

is orthogonal to both w^a and n^a at P and hence is a spacelike vector on the screen at P. With the aid of (3.4), a direct calculation gives

$$p^a_b(\perp\delta x^b)^* = A^a{}_b \perp\delta x^b + B^a{}_b \delta x^b\,, \tag{3.8}$$

where

$$A_{ab} = p^c_a\, p^d_b\, n_{c;d}\,, \tag{3.9}$$

$$B_{ab} = p_{ac}(\overset{*}{w}{}^c - \mathring{n}^c) w_b\,, \tag{3.10}$$

and we have introduced the notation

$$\overset{*}{A}{}^a = \frac{D}{dp} A^a = A^a{}_{;b} n^b\,, \tag{3.11}$$

$$\mathring{A}{}^a = A^a{}_{;b} w^b\,. \tag{3.12}$$

Except at the given point P, the motions of the observers employed along the curve $\mathscr{C}$ have still to be specified. We require that the 4-velocities w^a of the observers used along $\mathscr{C}$ are related by

$$p^a_c \overset{*}{w}{}^c = p^a_c \mathring{n}^c\,, \tag{3.13}$$

$$(w_a w^a)^* = 0\,, \quad (w_a n^a)^* = 0\,. \tag{3.14}$$

(3.13) is the necessary and sufficient condition for $B_{ab} = 0$, and this ensures that the rotation and shear tensors of the curves of the congruence lie on the screen. Condition (3.14) ensures that along $\mathscr{C}$, w^a is always a unit vector orthogonal to n^a. Equations (3.13) and (3.14) are equivalent to the single condition

$$\overset{*}{w}{}^a = \mathring{n}^a + (\mathring{n}_b w^b) w^a - (\overset{*}{n}_b w^b) n^a\,, \tag{3.15}$$

which is Greenberg's transport law for w^a. With $B_{ab} = 0$, (3.8) reduces to

$$p^a_b(\perp\delta x^b)^* = A^a{}_b \perp\delta x^b\,. \tag{3.16}$$

Decomposing A_{ab} into its irreducible parts we have

$$A_{ab} = \mathscr{R}_{ab} + \tfrac{1}{2}\mathscr{E} p_{ab} + \mathscr{S}_{ab}\,, \tag{3.17}$$

where

$$\mathscr{R}_{ab} = A_{[ab]} = p^c_a\, p^d_b\, n_{[c;d]}\,, \tag{3.18}$$

$$\mathscr{E} = A^a{}_a = p^{ab} n_{a;b} = n^a{}_{;a} + \mathring{n}_a w^a\,, \tag{3.19}$$

$$\mathscr{S}_{ab} = A_{(ab)} - \tfrac{1}{2} A^c{}_c\, p_{ab} = p^c_a\, p^d_b\, n_{(c;d)} - \tfrac{1}{2}\mathscr{E} p_{ab}\,. \tag{3.20}$$

Clearly,

$$\mathscr{R}_{ab} w^b = 0\,, \quad \mathscr{R}_{ab} n^b = 0\,;$$
$$\mathscr{S}_{ab} w^b = 0\,, \quad \mathscr{S}_{ab} n^b = 0\,, \quad \mathscr{S}^a{}_a = 0\,. \tag{3.21}$$

$\mathscr{R}_{ab}$, $\mathscr{E}$, and $\mathscr{S}_{ab}$ are called, respectively, the rotation tensor, the expansion, and the shear tensor of the curves of the congruence, as measured by observer w^a. These names are justified in the next subsection.

Finally, (3.9) and (3.17) imply the following useful identity:

$$n_{a;b} = \mathscr{R}_{ab} + \tfrac{1}{2}\mathscr{E} p_{ab} + \mathscr{S}_{ab} + \overset{*}{n}_a n_b - \mathring{n}_a w_b + w_a(n^t w_{t;b}) + w_a w_b(n^t \mathring{w}_t) - w_a n_b(n^t \overset{*}{w}_t)\,. \tag{3.22}$$

B. Physical interpretation of $\mathscr{R}_{ab}$, $\mathscr{S}_{ab}$, and $\mathscr{E}$

In order to determine the physical interpretation of $\mathscr{R}_{ab}$, $\mathscr{E}$, and $\mathscr{S}_{ab}$ we follow the same general procedure as used to investigate the vorticity, expansion, and shear of a timelike congruence of particle world lines in a fluid[12,13]; this provides an alternative approach to that given by Greenberg.[1] We split $\perp\delta x^a$, the projection onto the screen of the connecting vector δx^a, into a relative distance δL and a direction e^a on the screen

$$\perp\delta x^a = \delta L e^a\,, \tag{3.23}$$

where

$$e_a e^a = +1\,, \quad e_a w^a = 0\,, \quad e_a n^a = 0\,. \tag{3.24}$$

On substituting (3.23) into (3.16) and using (3.17), the following two equations can be derived:

$$\frac{(\delta L)^*}{\delta L} = \mathscr{S}_{ab} e^a e^b + \tfrac{1}{2}\mathscr{E}\,, \tag{3.25}$$

$$p^a_b \overset{*}{e}{}^b = [\mathscr{S}^{ab} + \mathscr{R}^{ab} - (\mathscr{S}_{cd} e^c e^d) p^{ab}]\, e_b\,. \tag{3.26}$$

The physical interpretation is aided if we introduce the concept of a flux tube: a flux tube is defined to be the 2-dimensional surface swept out by the curves of the congruence which pass through a given simple closed curve in 3-dimensional space. We will denote by δA the cross-sectional area of

the flux tube as measured by observer w^a at the point P on any one of the curves of the tube; δA can be taken to be the area of the screen erected by observer w^a at P.

(i) $\mathscr{R}_{ab}$.

If $\mathscr{S}_{ab}=0$, $\mathscr{E}=0$, but $\mathscr{R}_{ab}\neq 0$, then (3.25) and (3.26) reduce to

$$\frac{(\delta L)^*}{\delta L}=0\,, \tag{3.27}$$

$$p_b^a \overset{*}{e}{}^b = \mathscr{R}^{ab}e_b\,. \tag{3.28}$$

Since $(D/dp)\delta L=0$, there is no change in shape or area of the cross section of a flux tube along the length of the tube. However, the directions of the projections on the screen of all connecting vectors from a given curve $\mathscr{C}$ to all neighboring curves undergo a rotation along $\mathscr{C}$ with respect to a basis of vectors $\{e^a_{(i)}\}i=1,2$ transported along $\mathscr{C}$ according to the law

$$p_b^a \overset{*}{e}{}^b_{(i)}=0\,, \tag{3.29}$$

$$(e^a_{(i)}w_a)^*=0\,,\quad (e^a_{(i)}n_a)^*=0\,. \tag{3.30}$$

Such a basis will be called a nonrotating basis along $\mathscr{C}$.

Thus, if $\mathscr{R}_{ab}$ alone is nonzero, the curves of the congruence undergo a rotation along $\mathscr{C}$ but the cross section of a flux tube formed by curves of the congruence remains unchanged in shape and area along the tube.

Clearly, the instantaneous axis of rotation associated with $\mathscr{R}_{ab}$ is perpendicular to the screen and hence parallel to n^a. A vector parallel to n^a, constructed from $\mathscr{R}_{ab}$ and which describes this rotation, is

$$\mathscr{R}^a=\tfrac{1}{2}\eta^{abcd}w_b\mathscr{R}_{cd}\,. \tag{3.31}$$

Since $\mathscr{R}_{cd}$ lies on the screen, it is easily verified that $\mathscr{R}^a$ is parallel to n^a:

$$\mathscr{R}^a=(n_t\mathscr{R}^t)n^a\,. \tag{3.32}$$

Equation (3.31) can be inverted[11] to give

$$\mathscr{R}_{ab}=\eta_{abcd}\mathscr{R}^c w^d\,, \tag{3.33}$$

and we define

$$\mathscr{R}^2=\mathscr{R}_a\mathscr{R}^a=\tfrac{1}{2}\mathscr{R}_{ab}\mathscr{R}^{ab}\,. \tag{3.34}$$

A second vector can be associated with $\mathscr{R}_{ab}$ defined by

$$\mathscr{Q}^a=\tfrac{1}{2}\eta^{abcd}n_b\mathscr{R}_{cd}\,. \tag{3.35}$$

It can be verified that $\mathscr{Q}^a$ is timelike and parallel to w^a:

$$\mathscr{Q}^a=-(w_t\mathscr{Q}^t)w^a\,. \tag{3.36}$$

$\mathscr{Q}^a$ does not appear to play an important role in the theory.

(ii) $\mathscr{S}_{ab}$.

If $\mathscr{R}_{ab}=0$, $\mathscr{E}=0$, but $\mathscr{S}_{ab}\neq 0$, then (3.25) and (3.26) reduce to

$$\frac{(\delta L)^*}{\delta L}=\mathscr{S}_{ab}e^ae^b\,, \tag{3.37}$$

$$p_b^a \overset{*}{e}{}^b=[\mathscr{S}^{ab}-(\mathscr{S}_{cd}e^ce^d)p^{ab}]e_b\,. \tag{3.38}$$

In the usual way, e^a is defined to be a principal axis of shear if

$$\mathscr{S}_{ab}e^b=\lambda e_a\,, \tag{3.39}$$

where the scalar λ is the eigenvalue corresponding to e^a. Since $\mathscr{S}_{ab}w^b=0$ and $\mathscr{S}_{ab}n^b=0$, both w^a and n^a are principal axes of shear corresponding to the eigenvalue zero. Because $\mathscr{S}_{ab}$ is a symmetric tensor, the other two principal axes of shear must be orthogonal to both w^a and n^a and hence must lie on the screen. From (3.38) and (3.39), the directions of the two principal axes of shear on the screen remain unchanged with respect to a nonrotating basis along $\mathscr{C}$, although all other directions on the screen suffer a change in orientation along $\mathscr{C}$. Further, if δA is the cross-sectional area of a flux tube, then using (3.37) and (3.39) we have

$$\frac{(\delta A)^*}{\delta A}=\frac{(\delta L_1)^*}{\delta L_1}+\frac{(\delta L_2)^*}{\delta L_2}=\lambda_1+\lambda_2=\mathscr{S}^a_a=0\,, \tag{3.40}$$

where δL_1 and δL_2 are lengths measured along the principal axes of shear on the screen.

Thus, if $\mathscr{S}_{ab}$ alone is nonzero, there is a change of shape in the cross section of a flux tube along the length of the tube, but δA and the directions of the principal axes of shear on the screen remain invariant.

(iii) $\mathscr{E}$.

If $\mathscr{R}_{ab}=0$, $\mathscr{S}_{ab}=0$, but $\mathscr{E}\neq 0$, then (3.25) and (3.26) reduce to

$$\frac{(\delta L)^*}{\delta L}=\tfrac{1}{2}\mathscr{E}\,, \tag{3.41}$$

$$p_b^a \overset{*}{e}{}^b=0\,. \tag{3.42}$$

From (3.42), all directions on the screen remain unchanged with respect to a nonrotating basis along a curve $\mathscr{C}$ of the congruence. Also, $(D/dp)\delta L$ is independent of direction e^a and hence there is no change in the shape of the cross section of a flux tube along the length of a tube, but by (3.41),

$$\frac{(\delta A)^*}{\delta A}=2\,\frac{(\delta L)^*}{\delta L}=\mathscr{E}\,. \tag{3.43}$$

Thus if $\mathscr{E}$ alone is nonzero, the cross section of a flux tube suffers a change in area without a change in orientation or shape along the tube.

Since $\mathscr{R}_{ab}$ is absent from (3.25) and the effect of $\mathscr{S}_{ab}$ is to leave δA invariant, we have even if both $\mathscr{R}_{ab}\neq 0$ and $\mathscr{S}_{ab}\neq 0$,

$$\frac{(\delta A)^*}{\delta A}=\mathscr{E}\,. \tag{3.44}$$

If we introduce the representative length L along $\mathscr{C}$ defined by the equation

$$\frac{L}{L}^*=\tfrac{1}{2}\mathscr{E}\,, \tag{3.45}$$

then $\delta A\alpha L^2$.

C. Orthonormal tetrad

At any point P on a given curve $\mathscr{C}$ of the congruence, an orthonormal tetrad of vectors

$$w^a,n^a,r^a,s^a \tag{3.46}$$

can be introduced, where r^a and s^a are mutually orthogonal unit vectors on the screen erected by observer w^a at P. w^a

satisfies the transport law (3.15) along $\mathscr{C}$. Using a method developed by Pirani[14] in the theory of timelike congruences, Greenberg[1] determined the transport law for r^a (and s^a) along the spacelike congruence for which $\mathscr{R}_{ab}$ is the observed rotation of the curves:

$$\overset{*}{r}{}^a = (\overset{\circ}{n}_b r^b)w^a - (\overset{*}{n}_b r^b)n^a . \tag{3.47}$$

Equation (3.47) is the transport law for r^a (and s^a) which eliminates rotation of r^a (and s^a) along $\mathscr{C}$.

Equation (3.47) agrees with the results of the previous subsection. For, in order that r^a be a nonrotating vector along $\mathscr{C}$ (where rotation is defined in terms of $\mathscr{R}_{ab}$ as in the previous subsection) and always lie on the screen of the observers employed along the congruence, it must satisfy (3.29) and (3.30)

$$p_b^a \overset{*}{r}{}^b = 0 , \quad (r^a w_a)^* = 0 , \quad (r^a n_a)^* = 0 . \tag{3.48}$$

But (3.48) is equivalent to the single transport law

$$\overset{*}{r}{}^a = (\overset{*}{w}_b r^b)w_a - (\overset{*}{n}_b r^b)n^a , \tag{3.49}$$

and on using (3.15) for $\overset{*}{w}_b$, (3.47) is derived.

Under the transport law (3.15) for w^a and either (3.47) or (3.49) for r^a and s^a, the orthonormal tetrad (3.46) remains an orthonormal tetrad along $\mathscr{C}$.

D. Newtonian theory

Suppose

$$x^\alpha = x^\alpha(\eta^\beta, p) \tag{3.50}$$

is a congruence of curves in 3-dimensional Euclidean space,[10] where p denotes arc length along a curve from some initial cross section of the congruence. The unit tangent vector n^α at any point P on a given curve $\mathscr{C}$, and the connecting vector δx^α are defined as in general relativity; it is easily verified that

$$\underset{n}{\mathscr{L}}\, \delta x^\alpha = 0 . \tag{3.51}$$

Since δx^α need not lie on a screen orthogonal to the curves at P, we introduce the projection tensor

$$p^{\alpha\beta} = h^{\alpha\beta} - n^\alpha n^\beta , \tag{3.52}$$

where $h^{\alpha\beta}$ is the flat 3-space metric tensor, and define

$$\perp\delta x^\alpha = p^\alpha_\beta \delta x^\beta . \tag{3.53}$$

Since $n_\alpha\, p^\alpha_\beta = 0$, $\perp\delta x^\alpha$ lies on the screen at P. With the aid of (3.51), we find that

$$p^\alpha_\beta (\perp\delta x^\beta)^* = A^\alpha{}_\beta \perp\delta x^\beta , \tag{3.54}$$

where

$$A_{\alpha\beta} = p^\lambda_\alpha\, p^\mu_\beta n_{\lambda,\mu} \tag{3.55}$$

and we have introduced the notation $\overset{*}{A}{}^\alpha = (D/dp)A^\alpha = A^\alpha{}_{,\beta} n^\beta$. A significant difference between Newtonian theory and general relativity is that in Newtonian theory it is not necessary to employ special observers along the congruence to ensure that the rotation and shear tensors lie on the screen. Decomposing $A_{\alpha\beta}$ into its irreducible parts, we have

$$A_{\alpha\beta} = \mathscr{R}_{\alpha\beta} + \tfrac{1}{2}\mathscr{C} p_{\alpha\beta} + \mathscr{S}_{\alpha\beta} , \tag{3.56}$$

where the rotation tensor, the expansion, and the shear tensor are given respectively by

$$\mathscr{R}_{\alpha\beta} = A_{[\alpha\beta]} = p^\lambda_\alpha\, p^\mu_\beta n_{[\lambda,\mu]} , \tag{3.57}$$

$$\mathscr{C} = A^\alpha{}_\alpha = p^{\alpha\beta} n_{\alpha,\beta} = n^\alpha{}_{,\alpha} , \tag{3.58}$$

$$\mathscr{S}_{\alpha\beta} = A_{(\alpha\beta)} - \tfrac{1}{2}A^\tau{}_\tau\, p_{\alpha\beta} = p^\lambda_\alpha\, p^\mu_\beta n_{(\lambda,\mu)} - \tfrac{1}{2}n^\tau{}_{,\tau}\, p_{\alpha\beta} . \tag{3.59}$$

Clearly,

$$\mathscr{R}_{\alpha\beta} n^\beta = 0 , \quad \mathscr{S}_{\alpha\beta} n^\beta = 0 , \quad \mathscr{S}^\alpha_\alpha = 0 . \tag{3.60}$$

It follows directly from (3.55) and (3.56) that

$$n_{\alpha,\beta} = \mathscr{R}_{\alpha\beta} + \tfrac{1}{2}\mathscr{C} p_{\alpha\beta} + \mathscr{S}_{\alpha\beta} + \overset{*}{n}_\alpha n_\beta . \tag{3.61}$$

The rotation vector $\mathscr{R}^\alpha$ is defined by[15]

$$\mathscr{R}^\alpha = \tfrac{1}{2}\eta^{\alpha\beta\gamma}\mathscr{R}_{\beta\gamma} . \tag{3.62}$$

Since $\mathscr{R}_{\beta\gamma}$ lies on the screen, it follows that $\mathscr{R}^\alpha$ is parallel to n^α

$$\mathscr{R}^\alpha = (n_\tau \mathscr{R}^\tau) n^\alpha , \tag{3.63}$$

and hence $\mathscr{R}^\alpha$ is along the instantaneous axis of rotation of the curves of the congruence. Equation (3.62) can be inverted to give

$$\mathscr{R}_{\alpha\beta} = \eta_{\alpha\beta\gamma}\mathscr{R}^\gamma , \tag{3.64}$$

and we define

$$\mathscr{R}^2 = \mathscr{R}_\alpha \mathscr{R}^\alpha = \tfrac{1}{2}\mathscr{R}_{\alpha\beta}\mathscr{R}^{\alpha\beta} . \tag{3.65}$$

If we write $\perp\delta x^\alpha = \delta L e^\alpha$ where $e_\alpha e^\alpha = +1$, $e_\alpha n^\alpha = 0$, then (3.54) can be split into two equations which in form are identical to (3.25) and (3.26) but with Latin indices replaced by Greek indices. The tensors $\mathscr{R}_{\alpha\beta}$, $\mathscr{C}$, and $\mathscr{S}_{\alpha\beta}$ therefore have the same physical interpretation as in general relativity. In particular, (3.44) is also valid in Newtonian theory.

At any point P on a given curve $\mathscr{C}$ of the congruence, we can introduce the orthonormal triad of vectors

$$n^\alpha, r^\alpha, s^\alpha , \tag{3.66}$$

where r^α and s^α lie on the screen at P. The transport law for r^α (and s^α) along $\mathscr{C}$ is

$$p^\alpha_\beta \overset{*}{r}{}^\beta = 0 , \quad (r_\alpha n^\alpha)^* = 0 , \tag{3.67}$$

or equivalently, the single transport equation

$$\overset{*}{r}{}^\alpha = -(r_\beta \overset{*}{n}{}^\beta) n^\alpha . \tag{3.68}$$

Equation (3.68), together with a similar equation for s^α, eliminates rotation of r^α and s^α along $\mathscr{C}$ and ensures that the orthonormal triad (3.66) remains an orthonormal triad along $\mathscr{C}$.

4. COMOVING OBSERVERS

Let u^a denote the 4-velocity field of a self-gravitating fluid. Many of the spacelike vector fields associated with the

fluid are defined in terms of u^a and are orthogonal to u^a, for example, the vorticity vector

$$\omega^a = \tfrac{1}{2}\eta^{abcd}u_b u_{c;d} \tag{4.1}$$

and, in an electromagnetic fluid, the electric and magnetic field 4-vectors

$$E^a = F^{ab}u_b\ , \quad H^a = \tfrac{1}{2}\eta^{abcd}u_b F_{cd}\ . \tag{4.2}$$

A spacelike vector field gives rise to a spacelike congruence of curves and if the vector field is everywhere orthogonal to u^a, then the unit tangent vector n^a to a curve of the congruence will also be orthogonal to u^a; thus a comoving observer, $w^a = u^a$, can be employed to observe the deformations of the curves at any one specified point P on a curve $\mathscr{C}$ of the congruence. Because the 4-velocities of the observers employed along $\mathscr{C}$ are not independent but are related through Greenberg's transport law for w^a, it will not in general be possible to use comoving observers at points on $\mathscr{C}$ other than P. Clearly, it is of value to determine when comoving observers can be employed all along a given curve of a spacelike congruence.

Theorem: If the unit tangent vector n^a to any given curve of a spacelike congruence is everywhere orthogonal to the 4-velocity field u^a of a self-gravitating fluid, then comoving observers can be employed all along any curve of the congruence if and only if the curves are material curves in the fluid.

Proof: (i) Suppose the curves are material curves in the fluid.

Since $n^a u_a = 0$, we can choose $w^a = u^a$ at some given initial point P on a curve $\mathscr{C}$ of the congruence. At other points along $\mathscr{C}$, $w^a = u^a + \lambda^a$ where λ^a satisfies conditions (3.5) with w^a replaced by u^a. At any point on $\mathscr{C}$, Greenberg's transport law (3.15) therefore assumes the form

$$\frac{D}{dp}\lambda^a = h^a_b\dot{n}^b - \overset{*}{u}{}^a - (\overset{*}{n}_b u^b)n^a + A^a(\lambda;u,n)\,, \tag{4.3}$$

where dot denotes covariant differentiation along u^a and

$$\begin{aligned} A^a(\lambda;u,n) = {} & \lambda^a[(u^b + \lambda^b)\dot{n}_b + n_{b;c}(u^b + \lambda^b)\lambda^c] \\ & + \lambda^c[n^a{}_{;c} - n^a\overset{*}{n}_c + u^a(n_{b;c}u^b \\ & + n_{b;c}\lambda^b + \dot{n}_c)]\ . \end{aligned} \tag{4.4}$$

All terms in A^a contain at least one component of $\boldsymbol{\lambda}$ as a factor; if $\boldsymbol{\lambda} = \mathbf{0}$, for example at P, then $A^a = 0$.

Since $\mathscr{C}$ is a material curve in the fluid it moves with the fluid and hence consists at all times of the same fluid particles. Thus any two neighboring fluid particles initially on $\mathscr{C}$ always lie in the instantaneous direction of the unit tangent vector n^a to $\mathscr{C}$ at the particles. If δl is the distance as measured by u^a between the particles at any instant, then the vector $\delta l n^a$ links at all times these two particles; since also $n^a u_a = 0$, $\delta l n^a$ lies in the rest space of u^a and therefore is a relative position vector. But if $X^a_\perp$ is a relative position vector it must satisfy the equation[12,13]

$$h^a_b\dot{X}^b_\perp = u^a{}_{;b}X^b_\perp\ ; \tag{4.5}$$

hence

$$h^a_b(\delta l n^b)^{\cdot} = u^a{}_{;b}(\delta l n^b)\,. \tag{4.6}$$

On contracting (4.6) with n_a we find that[16]

$$\frac{(\delta l)^{\cdot}}{\delta l} = -\overset{*}{n}_b u^b\,, \tag{4.7}$$

and hence (4.6) becomes

$$h^a_b\dot{n}^b = \overset{*}{u}{}^a + (\overset{*}{n}_b u^b)n^a\,. \tag{4.8}$$

Substituting into (4.3) from (4.8) we obtain

$$\frac{D}{dp}\lambda^a = A^a(\lambda;u,n)\,. \tag{4.9}$$

Since this system of first-order differential equations for the components of $\boldsymbol{\lambda}$ is *homogeneous*, due to the form of (4.4) for $A^a(\lambda;u,n)$, its solution subject to the initial condition $\lambda^a = 0$ at P is $\lambda^a \equiv 0$ along $\mathscr{C}$; hence comoving observers can be employed all along $\mathscr{C}$.

(ii) Conversely, suppose comoving observers can be employed all along a curve $\mathscr{C}$ of the congruence. Then $w^a = u^a$ along $\mathscr{C}$ and hence $\lambda^a \equiv 0$ along $\mathscr{C}$. From (4.3) and (4.4) Greenberg's transport law therefore reduces to

$$h^a_b\dot{n}^b = \overset{*}{u}{}^a + (\overset{*}{n}_b u^b)n^a\,. \tag{4.10}$$

Define

$$f = \exp\left(-\int_{\tau_0}^{\tau}\overset{*}{n}_b u^b\,d\tau\right), \tag{4.11}$$

where integration is performed along the world line of u^a from some arbitrary initial proper time τ_0; then

$$\frac{\dot{f}}{f} = -\overset{*}{n}_b u^b\,, \tag{4.12}$$

and (4.10) may be rewritten as

$$h^a_b(fn^b)^{\cdot} = u^a{}_{;b}(fn^b)\,. \tag{4.13}$$

Since also $n^a u_a = 0$, it follows that $\epsilon f n^a$ is a relative position vector where ϵ is a constant. Hence if $\epsilon f n^a$ links two neighboring fluid particles at one instant, it does so at all later times. The same fluid particles therefore always lie on $\mathscr{C}$ as the fluid evolves and hence we conclude that $\mathscr{C}$ is a material curve in the fluid. This completes the proof.

Material curves in a fluid are sometimes said to be frozen-in to the fluid. If we define

$$N^a = h^a_b\dot{n}^b - \overset{*}{u}{}^a - (\overset{*}{n}_b u^b)n^a\,, \tag{4.14}$$

where $n^a u_a = 0$, then from the details of the proof of the above theorem, it is clear that comoving observers can be employed all along a given curve $\mathscr{C}$, and $\mathscr{C}$ will be frozen-in to the fluid, if and only if $N^a = 0$. When the nature of the congruence has been specified N^a can be calculated with the aid of the propagation equation for n^a and this will be done for four particular spacelike congruences in the following sections. We note from (4.3) that $N^a(P) = \overset{*}{\lambda}{}^a(P)$, where P is the initial point on $\mathscr{C}$ where we choose $\lambda^a = 0$. Also, if at a point $N^a \neq 0$, then N^a must lie on the screen erected by a comoving observer at that point; for (4.14) can be expressed as

$$N^a = p^a_b(\dot{n}^b - \overset{*}{u}{}^b)\,, \tag{4.15}$$

where

$$p^a_b = g^a_b + u^a u_b - n^a n_b \,. \tag{4.16}$$

5. CONGRUENCE OF VORTEX LINES

A. Comoving observers

If u^a is the 4-velocity field of a self-gravitating fluid, then the vorticity vector ω^a of the fluid at any point is defined by (4.1). Since $\omega^a u_a = 0$, ω^a is a spacelike vector and we define $\omega = (\omega_a \omega^a)^{1/2}$. The unit tangent vector to a vortex line[17] at any point is $n^a = \omega^a/\omega$; clearly $n^a u_a = 0$.

To evaluate N^a defined by (4.14), we note that the vorticity propagation equation is[12,13]

$$h^a_b \dot{\omega}^b = u^a{}_{;b}\omega^b - \theta\omega^a + \tfrac{1}{2}\eta^{abcd}u_b \dot{u}_{c;d}, \tag{5.1}$$

from which we deduce that

$$h^a_b \dot{n}^b = \overset{*}{u}{}^a + (\overset{*}{n}_b u^b)n^a + \frac{1}{2\omega} p^a_t \eta^{tbcd} u_b \dot{u}_{c;d} \,, \tag{5.2}$$

where p^a_b is the projection tensor for a comoving observer given by (4.16). Substituting into (4.14) from (5.2) gives

$$N^a = \frac{1}{2\omega} p^a_t \eta^{tbcd} u_b \dot{u}_{c;d} \,; \tag{5.3}$$

hence comoving observers can be employed all along a vortex line (and vortex lines are material lines in the fluid) if and only if

$$p^a_t \eta^{tbcd} u_b \dot{u}_{c;d} = 0 \,. \tag{5.4}$$

It can be verified that (5.4) agrees with the condition derived by Greenberg.[18] To examine (5.4), $\dot{u}^a$ must be either specified or derived, for example, with the aid of the momentum conservation equation. Clearly, for "dust", since $\dot{u}^a = 0$, Eq. (5.4) is satisfied and all observers employed along a vortex congruence can be comoving. This is in agreement with the well-known result that vortex lines in dust are material lines. We note further special cases of interest for which (5.4) is satisfied.

(i) Acceleration potential.

Suppose there exists locally a function r (an acceleration potential) such that

$$\dot{u}_a = -h^b_a (\ln r)_{,b} \,. \tag{5.5}$$

r can exist either due to a physical property of the fluid or due to a property of the flow. There are at least two examples of the former.[12] If the fluid is a perfect fluid with equation of state $p = p(\mu)$, then the momentum conservation equation can be written precisely in the form (5.5), where r is given by (2.22). Secondly, if the constitutive equation

$$q^a = -\varkappa h^{ab}(T_{,b} + T\dot{u}_b) \tag{5.6}$$

is valid, where q^a is the energy flux relative to u^a, T is the absolute temperature, and $\varkappa$ is the coefficient of heat conduction, but $q^a = 0$ and $\varkappa > 0$, then (5.5) holds with $r = T$. An isometric flow is an example of the latter possibility[19–21]; no matter what the physical nature of the fluid is, if there exists a timelike Killing vector field ξ_a parallel to u_a, then $\dot{u}^a$ has the form (5.5) with $r = (-\xi_a \xi^a)^{-1/2}$.

Using (5.5), a direct calculation gives

$$\eta^{tbcd} u_b \dot{u}_{c;d} = -2\omega(\ln r)\dot{} n^t \tag{5.7}$$

and hence, since $p^a_t n^t = 0$, Eq. (5.4) is satisfied. Comoving observers can therefore be employed all along a vortex line. Again, this is consistent with the well-known result that vortex lines are frozen-in to a fluid with an acceleration potential.[12]

(ii) Charged perfect fluid.

Consider the barotropic flow of a charged perfect fluid with vanishing magnetic field as measured by u^a. Then the momentum conservation equation (2.21) reduces to

$$\dot{u}_a = -(\ln r)_{,b} h^b_a + \frac{q}{\mu + p} E_a \,, \tag{5.8}$$

where r is defined by (2.22), and E_a satisfies [by Maxwell's equation (2.9), and (2.14)]

$$E_{[a;b]} = u_{[a}\dot{E}_{b]} + \dot{u}_{[a}E_{b]} + u^t E_{t;[a}u_{b]} \,. \tag{5.9}$$

Using (5.8) and (5.9), condition (5.4) becomes

$$p^a_t \eta^{tbcd} u_b E_c e_{,d} = 0 \,, \tag{5.10}$$

where $e = rq(\mu + p)^{-1}$. If the additional assumption is made that e is constant everywhere,[22] then (5.10) is satisfied; comoving observers can be employed all along a vortex line, and the vortex lines are material lines in the charged fluid.

B. Rotation, expansion, and shear of a congruence of vortex lines

For a congruence of vortex lines, (3.9) with $n^a = \omega^a/\omega$ becomes

$$A_{ab} = \frac{1}{\omega} p^c_a \, p^d_b \omega_{c;d} \,. \tag{5.11}$$

Thus, given an observer $w^a (w^a n_a = 0)$ at any point P on a vortex line, the rotation tensor, the expansion, and the shear tensor can be calculated if $\omega_{c;d}$ is known. Starting from the Ricci identity for u^a and making use of Einstein's field equations, we derive a general expression for $\omega_{c;d}$, in terms of the kinematic quantities and fluid variables, which is valid irrespective of the nature of the fluid, which may even contain an electromagnetic field. Because ω^a is defined in terms of the fluid 4-velocity u^a, the result we obtain for $\omega_{c;d}$ is expressed in terms of u^a no matter the 4-velocity w^a of the observer used at P.

The Ricci identity for u^a is

$$u_{a;bc} - u_{a;cb} = R_{tabc} u^t \,. \tag{5.12}$$

We operate[11] on (5.12) with $\eta^{bcrs}u_s$ and use the expansion

$$u_{a;b} = \sigma_{ab} + \tfrac{1}{3}\theta h_{ab} + \omega_{ab} - \dot{u}_a u_b \,, \tag{5.13}$$

where σ_{ab} is the shear tensor of the fluid and $\omega_{ab} = \eta_{abcd}\omega^c u^d$ is the vorticity tensor. The resulting expression for $\omega_{a;b}$ can be simplified slightly with the aid of the constraint equation for vorticity[12,13]:

$$\omega^a{}_{;b} h^b_a = \omega_a \dot{u}^a \,. \tag{5.14}$$

We find that

$$\begin{aligned}\omega_{a;b} &= g_{ab}(\omega_c \dot{u}^c) + \tfrac{1}{3}\eta_{ab}{}^{st}\theta_{,s}u_t + \eta_a{}^{rst}\sigma_{br;s}u_t - \tfrac{1}{2}\eta_a{}^{cds}R_{tbcd}u^t u_s \\ &\quad - \omega_a(\theta u_b + 2\dot{u}_b) + u_a(\omega^c u_{c;b}) \\ &\quad + u_b(u_a \omega^c{}_{;c} + u_{a;c}\omega^c - \dot{\omega}_a) \,.\end{aligned} \tag{5.15}$$

We now express the Riemann curvature tensor R_{abcd} in terms of the Ricci tensor R_{ab} and the Weyl tensor C_{abcd}. This allows us to introduce dynamics into the purely kinematical equation (5.15) through Einstein's field equations and also to distinguish between the curvature contribution to $\omega_{a;b}$ made by the local energy distribution and by the surrounding fluid. We have

$$R^{ab}{}_{cd} = C^{ab}{}_{cd} + 2g^{[a}_{[c}R^{b]}_{d]} - \tfrac{1}{3}Rg^{[a}_{[c}g^{b]}_{d]}\,, \tag{5.16}$$

and Einstein's field equations

$$R_{ab} = T_{ab} - \tfrac{1}{2}Tg_{ab} + \Lambda g_{ab}\,, \tag{5.17}$$

where $T = T^a_a$, Λ is the cosmological constant, and T_{ab} is the total energy–momentum tensor of the fluid, which can be decomposed with respect to u^a as

$$T_{ab} = \mu u_a u_b + ph_{ab} + 2q_{(a}u_{b)} + \pi_{ab}\,. \tag{5.18}$$

μ is the energy density and q^a is the energy flux as measured by u^a, p is the isotropic pressure and π_{ab} is the trace-free anisotropic pressure tensor. These quantities include contributions by all sources, for example by an electromagnetic field as well as by the fluid; in particular q^a represents processes such as heat conduction and diffusion as well as the electromagnetic flux. Finally, the Weyl tensor can be split by u^a into its trace-free electric and magnetic parts:

$$E_{ab} = C_{arbs}u^r u^s\,, \quad H_{ab} = \tfrac{1}{2}\eta_{arpq}C^{pq}{}_{bs}u^r u^s\,, \tag{5.19}$$

$$E_{ab} = E_{(ab)}\,, \quad H_{ab} = H_{(ab)}\,, \quad E^a_a = H^a_a = 0\,,$$

$$E_{ab}u^b = H_{ab}u^b = 0\,. \tag{5.20}$$

We find that

$$\tfrac{1}{2}\eta_a{}^{cds}R_{tbcd}u^t u_s = -H_{ab} + \tfrac{1}{2}\eta_{abcd}q^c u^d\,; \tag{5.21}$$

the symmetric part contributes to $\mathscr{E}$ and $\mathscr{S}_{ab}$ only, the skew-symmetric part to $\mathscr{R}_{ab}$ only. Equation (5.15) becomes

$$\begin{aligned}\omega_{a;b} = {} & g_{ab}(\omega_c \dot{u}^c) + H_{ab} + \tfrac{1}{2}\eta_{ab}{}^{st}(\tfrac{2}{3}\theta_{,s} - q_s)u_t + \eta_a{}^{rst}\sigma_{br;s}u_t \\ & - \omega_a(\theta u_b + 2\dot{u}_b) + u_a(\omega^c u_{c;b}) \\ & + u_b(u_a\omega^c{}_{;c} + u_{a;c}\omega^c - \dot{\omega}_a)\,.\end{aligned} \tag{5.22}$$

Equation (5.22) for $\omega_{a;b}$ consists of three kinds of terms. The first kind depend on the kinematic quantities θ, σ_{ab}, ω^a, $\dot{u}^a$ and are determined by local symmetries. The second kind is dynamic in the sense that it arises from Einstein's field equations and is represented in (4.22) only by the term containing the total energy flux q^a. The third kind describes the free gravitational field and is fixed by asymptotic assumptions; this is represented by H_{ab} only, $\omega_{a;b}$ being independent of E_{ab}.

The 4-velocity w^a of the observer has yet to be specified. Since $\omega_a u^a = 0$, we can always choose to use a comoving observer at the given point P to evaluate $\mathscr{R}_{ab}$, $\mathscr{E}$ and $\mathscr{S}_{ab}$ at P (although as a result the other observers used along the congruence may not be comoving observers) and we will suppose that this choice has been made in the rest of this subsection. The projection tensor p^a_b is then given by (4.16), and this leads immediately to the simplification that terms containing u_a and u_b in (5.22) do not contribute to A_{ab}. (Clearly the term containing ω_a does not contribute to A_{ab}.)

In Newtonian theory, (5.12) is replaced by

$$v_{\alpha,\beta\gamma} - v_{\alpha,\gamma\beta} = 0\,, \tag{5.23}$$

where v_α is the 3-velocity of the fluid. If we operate[15] on (5.23) with $\eta^{\beta\gamma\tau}$, use the expansion

$$v_{\alpha,\beta} = \sigma_{\alpha\beta} + \tfrac{1}{3}\theta h_{\alpha\beta} + \omega_{\alpha\beta}\,, \tag{5.24}$$

where $\omega_{\alpha\beta} = \eta_{\alpha\beta\gamma}\omega^\gamma$ and note that $\omega^\alpha{}_{,\alpha} = 0$, then we obtain in place of (5.22) the purely kinematical result

$$\omega_{\alpha,\beta} = \tfrac{1}{3}\eta_{\alpha\beta}{}^\tau\theta_{,\tau} + \eta_\alpha{}^{\lambda\tau}\sigma_{\beta\lambda,\tau}\,. \tag{5.25}$$

$\omega_{\alpha,\beta}$ does not contain gravitational field terms.

(i) Rotation tensor

From (3.18) with $n^a = \omega^a/\omega$, the rotation tensor of a congruence of vortex lines is

$$\mathscr{R}_{ab} = \frac{1}{\omega}p^c_a\,p^d_b\omega_{[c;d]}\,. \tag{5.26}$$

Hence, from (5.22), and employing a comoving observer (4.16) at P, we have

$$\mathscr{R}_{ab} = \frac{1}{\omega}p^c_a\,p^d_b\eta_{[c}{}^{rst}\sigma_{d]r;s}u_t + \frac{1}{2\omega}p^c_a\,p^d_b\eta_{cd}{}^{st}(\tfrac{2}{3}\theta_{,s} - q_s)u_t\,. \tag{5.27}$$

From (3.57) and (5.25), the corresponding result in Newtonian theory is

$$\mathscr{R}_{\alpha\beta} = \frac{1}{\omega}p^\nu_\alpha\,p^\tau_\beta\eta_{[\nu}{}^{\gamma\mu}\sigma_{\tau]\gamma,\mu} + \frac{1}{2\omega}p^\nu_\alpha\,p^\tau_\beta\eta_{\nu\tau}{}^\mu\tfrac{2}{3}\theta_{,\mu}\,. \tag{5.28}$$

In general relativity, $\mathscr{R}_{ab}$ is independent of the free gravitational field H_{ab} due to the surrounding fluid, but it does depend on the contribution to the curvature made by the local energy distribution at P as represented by the term containing q^a. In Newtonian theory, if the fluid is rigid, i.e., $\sigma_{\alpha\beta} = 0$ and $\theta = 0$, then $\mathscr{R}_{\alpha\beta} = 0$, which is consistent with uniform vorticity in a rigid rotation. In general relativity, if the fluid is kinematically rigid,[19] i.e., $\sigma_{ab} = 0$ and $\theta = 0$, then the vorticity is not necessarily uniform[23] and $\mathscr{R}_{ab}$ is not necessarily zero.

To determine the necessary and sufficient condition for a congruence of vortex lines to have zero rotation as measured by a comoving observer at P, we consider the rotation vector

$$\mathscr{R}^a = \tfrac{1}{2}\eta^{abcd}u_b\mathscr{R}_{cd}\,. \tag{5.29}$$

If we substitute from (5.27) we find that

$$\mathscr{R}^a = -\frac{1}{2\omega}(n_b I^b(\omega))n^a\,, \tag{5.30}$$

where

$$I^a(\omega) = q^a - \tfrac{2}{3}\theta_{,b}h^{ab} + h^{ab}\sigma_{bc;d}h^{cd}\,. \tag{5.31}$$

$I^a(\omega)$ is a spacelike vector and is a generalization of the energy flux. From (3.34) we have

$$\mathscr{R}^2 = \frac{1}{4\omega^2}(n_b I^b(\omega))^2 \tag{5.32}$$

and hence $\mathscr{R} = 0$ if and only if

$$\omega_b I^b(\omega) = 0\,; \tag{5.33}$$

in a kinematically rigid fluid, for example, u^a measures zero rotation of a vortex congruence at P if and only if q^a is orthogonal to ω^a at P.

From (3.62), (3.65), and (5.28), the corresponding results in Newtonian theory are

$$\mathscr{R}^{\alpha} = -\frac{1}{2\omega}(n_{\beta}I^{\beta}(\omega))n^{\alpha}, \tag{5.34}$$

$$\mathscr{R}^{2} = \frac{1}{4\omega^{2}}(n_{\beta}I^{\beta}(\omega))^{2}, \tag{5.35}$$

where

$$I_{\alpha}(\omega) = -\tfrac{2}{3}\theta_{,\alpha} + \sigma_{\alpha}{}^{\beta}{}_{,\beta}. \tag{5.36}$$

Equation (5.27) for $\mathscr{R}_{ab}$ can be derived alternatively from the (0ν) components of Einstein's field equations,[12,13]

$$h^{ab}(\tfrac{2}{3}\theta_{,b} - \sigma_{bc;d}h^{dc}) - \eta^{abcd}u_{b}(\omega_{c;d} + 2\omega_{c}\dot{u}_{d}) = q^{a}, \tag{5.37}$$

which follows directly from the Ricci identity for u_a and Einstein's equations. Equation (5.28) for $\mathscr{R}_{\alpha\beta}$ can be derived from the Newtonian analogue of (5.37)[13]:

$$\tfrac{2}{3}\theta^{,\alpha} - \sigma^{\alpha\beta}{}_{,\beta} - \eta^{\alpha\beta\gamma}\omega_{\beta,\gamma} = 0. \tag{5.38}$$

(ii) Expansion

From (3.19) with $n^a = \omega^a/\omega$, the expansion of a congruence of vortex lines is

$$\mathscr{E} = \frac{1}{\omega}p^{ab}\omega_{a;b}. \tag{5.39}$$

With the aid of (4.16) and (5.22), we find for a comoving observer at P,

$$\mathscr{E} = \frac{1}{\omega}\left(2\omega n_{c}\dot{u}^{c} - n^{b}\sigma_{br;s}\frac{\omega^{rs}}{\omega} - H_{ab}n^{a}n^{b}\right). \tag{5.40}$$

From (3.58) and (5.25), the corresponding expression in Newtonian theory is

$$\mathscr{E} = -\frac{1}{\omega}n^{\beta}\sigma_{\beta\lambda,\tau}\frac{\omega^{\lambda\tau}}{\omega}. \tag{5.41}$$

In both Newtonian theory and general relativity, $\mathscr{E}$ does not depend explicitly on the expansion scalar θ of the fluid. In general relativity, $\mathscr{E}$ is independent of q^a which represents the contribution to the curvature due to the local energy distribution at P; only the free gravitational field, represented by H_{ab}, contributes directly to $\mathscr{E}$. In Newtonian theory, if the fluid is shear-free, then $\mathscr{E} = 0$, but this is not true in general relativity.

Equation (4.45) can be derived alternatively from the constraint equation for H_{ab}:

$$H_{ab} = 2\dot{u}_{(a}\omega_{b)} - h^{s}_{a}h^{t}_{b}(\omega_{(s}{}^{d;c} + \sigma_{(s}{}^{d;c})\eta_{t)fdc}u^{f}, \tag{5.42}$$

which is derived directly from the Ricci identity for u^a without the aid of Einstein's field equations. Equation (5.41) follows from the Newtonian analog of (5.42).[13]

$$(\omega_{(\alpha}{}^{\nu,\tau} + \sigma_{(\alpha}{}^{\nu,\tau})\eta_{\beta)\nu\tau} = 0. \tag{5.43}$$

(iii) Shear tensor

From (3.20) with $n^a = \omega^a/\omega$, the shear tensor of a congruence of vortex lines is

$$\mathscr{S}_{ab} = \frac{1}{\omega}p^{c}_{a}p^{d}_{b}\omega_{(c;d)} - \tfrac{1}{2}\mathscr{E}p_{ab}. \tag{5.44}$$

Hence, using (5.22) and (5.40), we have for a comoving observer at P,

$$\mathscr{S}_{ab} = \frac{1}{\omega}p^{c}_{a}p^{d}_{b}(\eta_{(c}{}^{rst}\sigma_{d)r,s}u_{t} + H_{cd}) + \frac{1}{2\omega}\left(n^{t}\sigma_{tc;d}\frac{\omega^{cd}}{\omega} + H_{cd}n^{c}n^{d}\right)p_{ab}. \tag{5.45}$$

The corresponding result in Newtonian theory follows directly from (3.59), (5.25) and (5.41):

$$\mathscr{S}_{\alpha\beta} = \frac{1}{\omega}p^{\lambda}_{\alpha}p^{\mu}_{\beta}\eta_{(\lambda}{}^{\gamma\tau}\sigma_{\mu)\gamma,\tau} + \frac{1}{2\omega}\left(n^{\tau}\sigma_{\tau\nu,\lambda}\frac{\omega^{\nu\lambda}}{\omega}\right)p_{\alpha\beta}. \tag{5.46}$$

In both Newtonian theory and general relativity, the shear tensor does not depend explicitly on θ. In general relativity, $\mathscr{S}_{ab}$ is independent of $\dot{u}^a$ and, as was the case with $\mathscr{E}$, it depends explicitly only on the free gravitational field H_{ab}, being independent of q^a. For a shear-free flow in Newtonian theory $\mathscr{S}_{\alpha\beta} = 0$, but in general relativity $\mathscr{S}_{ab}$ can be nonzero even if $\sigma_{ab} = 0$.

Equations (5.45) and (5.46) can be derived alternatively from (5.42) and (5.43) respectively.

In this subsection we have considered a comoving observer and determined the values for $\mathscr{R}_{ab}$, $\mathscr{E}$ and $\mathscr{S}_{ab}$ measured by such an observer at any point P on a vortex line. Corresponding expressions for other observers w^a ($w^a n_a = 0$) can be calculated by operating on (5.22) with the appropriate projection tensor p^{ab}. The essential features of the results are preserved. In particular, the role played by the free gravitational field and that part determined locally through Einstein's field equations, is determined by the symmetry properties of the terms on the right-hand side of (5.21) and not by the choice of observer. The Newtonian results are independent of choice of observer.

C. Vortex tubes

We first note the three theorems of Helmholtz on vortex tubes in Newtonian theory[24] and then consider their extensions to general relativity.

In Newtonian theory, it follows from (3.44), and (3.58) with $n^{\alpha} = \omega^{\alpha}/\omega$, that

$$\frac{D}{dp}\ln(\omega\delta A) = \frac{1}{\omega}\omega^{\alpha}{}_{,\alpha}, \tag{5.47}$$

where δA is the cross-sectional area of a vortex tube. But $\omega^{\alpha}{}_{,\alpha} \equiv 0$, and hence $\omega\delta A$ is constant along a vortex tube; it therefore represents a characteristic of the tube and is called the strength of the vortex tube. This is the first of the Helmholtz theorems and is purely kinematical. Helmholtz's other two theorems depend on the properties of the fluid and/or of the flow. For the barotropic flow of a perfect fluid, the second and third Helmholtz theorems state, respectively, that vortex tubes move with the fluid (or equivalently, vortex lines are material lines) and that the strength of a vortex tube remains constant as the tube moves with the fluid. Both theorems are derived with the aid of the vorticity propagation equation. For a viscous fluid, for example, the strength of a vortex tube is not in general an integral of the equations of motion.

To obtain relativistic generalizations of these theorems, consider a comoving observer at any point P on one of the

vortex lines of a vortex tube. Then from (3.19) with $n^a = \omega^a/\omega$ and (3.44), we have at P

$$\frac{D}{dp}\ln(\omega\delta A) = \frac{1}{\omega}\omega^a{}_{;b}h^b{}_a\,, \tag{5.48}$$

where δA is the cross-sectional area of the vortex tube at P as measured by u^a. Equation (5.48) differs from the Newtonian result (5.47) in two important respects. Firstly, since

$$\omega^a_{;b}h^b{}_a = \omega^a\dot{u}_a \tag{5.49}$$

the right-hand side of (5.48) is not in general zero.[25] Secondly, (5.48) was evaluated for a comoving observer and hence in general is valid only at P; it cannot be integrated along a vortex tube to yield a conserved quantity unless the vortex lines are material lines. When comoving observers cannot be employed along a vortex line, an alternative expression to (5.48) would have to be derived in terms of observers satisfying Greenberg's transport law. We consider two special cases.

Firstly, suppose there exists locally an acceleration potential r defined by (5.5). Then with the aid of (5.49), (5.48) may be expressed as

$$\frac{D}{dp}\ln(r\omega\delta A) = 0\,. \tag{5.50}$$

But (5.50) is valid all along a vortex line because comoving observers can be employed along a vortex congruence in a fluid with an acceleration potential; hence $r\omega\delta A$ is conserved along a vortex tube[1] and can be defined as the strength of the tube. The second Helmholtz theorem that the vortex tube moves with the fluid is valid (since the vortex lines are material lines) and it can be proved with the aid of the vorticity propagation equation (5.1) that the strength of the tube $r\omega\delta A$ remains constant as the tube moves with the fluid.[26]

Secondly, consider again the barotropic flow of a charged perfect fluid with vanishing magnetic field as measured by u^a. Then $\dot{u}^a$ is given by (5.8), and since Maxwell's equation (2.7) with $H^a = 0$ implies that $\omega_a E^a = 0$, (5.50) is again derived from (5.48) and (5.49). However, (5.50) can now be integrated along the vortex line if and only if (5.10) is satisfied. We will therefore assume that e is constant everywhere[22]; (5.50) then implies that $r\omega\delta A$ is constant along a vortex tube. This extension of Helmholtz's first theorem is not a purely kinematical result since Maxwell's equations and the momentum conservation equation were used in its derivation. Clearly, Helmholtz's second theorem is valid under the above assumptions. To determine if the strength of the tube is preserved as the tube moves with the charged fluid, consider the propagation equation for vorticity (5.1). A direct calculation using (5.8) and (5.9) gives

$$\eta^{abcd}u_b\dot{u}_{c;d} = -2\frac{\dot{r}}{r}\omega^a + \frac{1}{r}\eta^{abcd}u_bE_ce_{,d} = -\frac{2\dot{r}}{r}\omega^a \tag{5.51}$$

since we assume e is constant everywhere, and hence (5.1) may be expressed as

$$h^a_b(l^3r\omega^b)^{\cdot} = u^a{}_{;b}(l^3r\omega^b)\,, \tag{5.52}$$

where the characteristic length l is defined by $\dot{l}/l = \theta/3$. Using the argument of Ellis,[26] it follows directly from (5.52) that the strength of the tube $r\omega\delta A$ is preserved as the tube moves with the charged fluid. A special case of this result, valid for charged dust, was established by Raychaudhuri and De.[27]

6. CONGRUENCE OF MAGNETIC FIELD LINES

A. Comoving observer

If the fluid is charged and electrically conducting there will be a mutual interaction between the fluid and the electromagnetic field. In this section we consider the properties of a congruence of magnetic field lines in such a fluid. Now, the electromagnetic field tensor F_{ab} can be split by any given 4-velocity field into local electric and magnetic fields, but in an electromagnetic fluid it is natural to use the 4-velocity field u^a of the fluid to perform this splitting. We do that here, and the electric and magnetic field 4-vectors as measured by u^a are defined by (4.2). Along a magnetic field line $n^a = H^a/H$ where $H = (H_aH^a)^{1/2}$; clearly $n_au^a = 0$.

To determine when comoving observers can be employed all along a given magnetic field line of the congruence we evaluate N^a defined by (4.14). It follows directly from the propagation equation for H^a, (2.9), that

$$h^a_b\dot{n}^b = \overset{*}{\dot{u}}{}^a + (\overset{*}{n}{}^bu_b)n^a - \frac{1}{H}p^a_bI^b(E)\,, \tag{6.1}$$

where the spacelike vector field $I^a(E)$ is defined by (2.11) and p^a_b is the projection tensor for a comoving observer, given by (4.16). Substituting from (6.1) into (4.14), we obtain

$$N^a = -\frac{1}{H}p^a_bI^b(E)\,. \tag{6.2}$$

Comoving observers can therefore be employed all along a magnetic field line (or equivalently, magnetic field lines are material lines) if and only if

$$p^a_bI^b(E) = 0\,. \tag{6.3}$$

An important special case for which (6.3) is satisfied is the magnetohydrodynamic (MHD) approximation of vanishing electric field E^a as measured by u^a. Physically this can occur in a fluid for the idealized limit of infinite electric conductivity K: Ohm's law in the comoving frame is $\mathcal{J}^a = KE^a$, and hence if $K\to\infty$ we require $E^a\to 0$ to keep the conduction current $\mathcal{J}^a$ finite.[28] If $E^a = 0$, then from definition (2.11), $I^a(E) = 0$, and (6.3) is trivially satisfied; in the MHD approximation comoving observers can be used along a congruence of magnetic field lines. Magnetic field lines are, of course, material lines in a fluid with $E^a = 0$.[12]

B. Rotation

An expression for the rotation of a congruence of magnetic field lines can be derived from Maxwell's equations without the necessity of actually solving these equations for H^a. As with a congruence of vortex lines, since $u_an^a = 0$ we can always choose to employ a comoving observer at any point P on a magnetic field line, although as a result the other observers employed along the field line will not be comoving unless (6.3) is satisfied.

From (3.18) with $n^a = H^a/H$, the rotation of a con-

gruence of magnetic field lines is

$$\mathscr{R}_{ab} = \frac{1}{H} p^c_a\, p^d_b H_{[c;d]}\,. \tag{6.4}$$

Substituting from the mathematical identity (2.15) into (6.4), we obtain for a comoving observer at P,

$$\mathscr{R}_{ab} = \frac{1}{2H} p^c_a\, p^d_b \eta_{cdrs} u^r I^s(H)\,, \tag{6.5}$$

where p^a_b is given by (4.16) and $I^a(H)$ is defined by (2.12). Inverting[11] Eq. (6.5) we find that

$$\mathscr{R}^a = -\frac{1}{2H}(n_b I^b(H))n^a\,, \tag{6.6}$$

where $\mathscr{R}^a$ is defined by (5.29).

Equations (6.5) and (6.6) are mathematical identities, being direct consequences of definition (2.12) for $I^a(H)$. Dynamics is introduced through Maxwell's equation (2.10) which relates $I^a(H)$ to E^a and the conduction current $\mathscr{J}^a$. An important special case is the MHD approximation; if $E^a = 0$ then from (2.10), $I^a(H) = \mathscr{J}^a$, and (6.5) and (6.6) become

$$\mathscr{R}_{ab} = \frac{1}{2H} p^c_a\, p^d_b \eta_{cdrs} u^r \mathscr{J}^s\,, \tag{6.7}$$

$$\mathscr{R}^a = -\frac{1}{2H}(n_b \mathscr{J}^b)n^a\,. \tag{6.8}$$

Hence, in a perfectly conducting fluid, the magnetic field lines have zero rotation as measured by u^a if and only if

$$H_a \mathscr{J}^a = 0\,. \tag{6.9}$$

This result has the additional significance that (6.3) is satisfied in the MHD approximation, and hence all observers employed along a magnetic field line can be comoving observers. We note the formal similarity between (5.30) for the rotation of vortex lines and the two expressions (6.6) and (6.8) for the rotation of magnetic field lines.

Because Maxwell's equation (2.10) yields only $H_{[c;d]}$ and not $H_{c;d}$, it cannot be used to calculate $\mathscr{S}_{ab}$ and $\mathscr{E}$. The remaining Maxwell equations do not provide an expression for $H_{[c;d]}$ as they stand. In the absence of such an expression, information on $\mathscr{E}$ can be obtained from the divergence equation (2.7): with the aid of (3.19), we find for a comoving observer

$$\mathscr{E} = \frac{2}{H}\omega_a E^a - (\ln H)^*\,. \tag{6.10}$$

Equation (6.10) has been discussed by Date[2] for the case $E^a = 0$. If $\omega_a E^a = 0$, then $\mathscr{E}$ decreases as H increases, which is consistent with the results of the next subsection. To complete the derivation of $\mathscr{E}$, $(\ln H)^*$ would have to be determined.

C. Magnetic flux tubes

Counterparts of the three Helmholtz theorems for vortex tubes can be derived for magnetic flux tubes.

It follows from (3.19) with $n^a = H^a/H$ and (3.44) that, for a comoving observer at any point P on one of the magnetic field lines of a magnetic flux tube,

$$\frac{D}{dp}\ln(H\delta A) = \frac{1}{H} H^a_{;b} h^b_a\,, \tag{6.11}$$

where δA is the cross-sectional area of the flux tube at P as measured by u^a. Hence, using Maxwell's equation (2.7), we have

$$\frac{D}{dp}\ln(H\delta A) = \frac{2}{H}\omega_a E^a\,. \tag{6.12}$$

We assume that (6.3) is satisfied so that (6.12) is valid along the magnetic field line and not only at P. We also make the further assumption that $\omega_a E^a = 0$. Then $H\delta A$ is constant along a magnetic flux tube,[3] and can be defined as the strength of the tube.

To obtain counterparts of the second and third Helmholtz theorems, we note that since (6.3) is assumed satisfied, the magnetic field lines are material lines and hence the flux tube moves with the fluid. Also, consideration of the propagation equation for H^a, (2.9), shows that for the strength of the tube $H\delta A$ to remain constant during the motion we require $I^a(E) = 0$.[29]

A special case for which both $\omega_a E^a = 0$ and $I^a(E) = 0$ is the MHD approximation $E^a = 0$.

7. CONGRUENCE OF ELECTRIC FIELD LINES

A. Comoving observers

In this section we consider a congruence of electric field lines in an electromagnetic fluid with 4-velocity field u^a. The roles played by the propagation equations for H^a and E^a in the previous section are interchanged; with a congruence of electric field lines, (2.10) determines N^a defined by (4.14), and (2.9) determines the rotation of the electric field lines. The unit tangent vector to an electric field line at any point is $n^a = E^a/E$ where $E = (E_a E^a)^{1/2}$ and E^a is defined by (4.2); clearly $n_a u^a = 0$.

To evaluate N^a we note that the propagation equation for E^a, (2.10), may be expressed in terms of n^a as

$$h^a_b \dot{n}^b = \overset{*}{\dot{u}}{}^a + (\overset{*}{n}{}^b u_b)n^a + \frac{1}{E} p^a_b (I^b(H) - \mathscr{J}^b)\,, \tag{7.1}$$

where p^a_b is given by (4.16). Substituting from (7.1) into (4.14) gives

$$N^a = \frac{1}{E} p^a_b (I^b(H) - \mathscr{J}^b)\,. \tag{7.2}$$

Hence, comoving observers can be employed all along an electric field line (equivalently: electric field lines are material lines) if and only if

$$p^a_b (I^b(H) - \mathscr{J}^b) = 0\,. \tag{7.3}$$

If Ohm's law $\mathscr{J}^a = KE^a$ is satisfied locally then, since $p^a_b E^b = 0$, (7.3) reduces to

$$p^a_b I^b(H) = 0\,, \tag{7.4}$$

which has the same form as condition (6.3). If further $H^a = 0$ everywhere, then by definition (2.12), $I^a(H) = 0$, and (7.4) is trivially satisfied; in a charged fluid with vanishing magnetic field as measured by u^a and in which Ohm's law is valid, comoving observers can be used all along a congruence of

electric field lines, and the electric field lines are material lines in the fluid.

B. Rotation

From (3.18) with $n^a = E^a/E$, the rotation of a congruence of electric field lines is

$$\mathscr{R}_{ab} = \frac{1}{E} p^c_a\, p^d_b E_{[c;d]}\,. \tag{7.5}$$

Since the identities (2.14) and (2.15) for $H_{[c;d]}$ and $E_{[c;d]}$ have the same form, the rotation of a congruence of electric field lines as measured by a comoving observer at a given point P can be obtained from (6.5) and (6.6) on replacing H by E:

$$\mathscr{R}_{ab} = \frac{1}{2E} p^c_a\, p^d_b \eta_{cdrs} u^r I^s(E)\,, \tag{7.6}$$

$$\mathscr{R}^a = -\frac{1}{2E}(n_b I^b(E))n^a\,. \tag{7.7}$$

Dynamics can be introduced into the mathematical identities (7.6) and (7.7) through Maxwell's equation (2.9), which relates $I^a(E)$ to H^a. An interesting special case is a charged fluid with vanishing magnetic field as measured by u^a: if $H^a = 0$, then from (2.9), $I^a(E) = 0$ and hence the electric field lines have zero rotation as measured by a comoving observer at P. This result does not depend on Ohm's law, but if Ohm's law is satisfied, then all observers along the congruence related to the comoving observer at P through Greenberg's transport law also measure zero rotation since all such observers are comoving. Charged dust distributions with vanishing magnetic field have been studied by Raychaudhuri and De.[27]

Maxwell's equation (2.9) cannot be used to calculate $\mathscr{S}_{ab}$ and $\mathscr{E}$ because it yields $E_{[c;d]}$ and not $E_{(c;d)}$. An expression for $\mathscr{E}$ can be obtained from the divergence equation (2.8): using (3.19) we find for a comoving observer,

$$\mathscr{E} = \frac{1}{E}(q - 2\omega^a H_a) - (\ln E)^*\,. \tag{7.8}$$

The charge enters naturally into (7.8) and its role is clearly exhibited.

C. Electric flux tubes

Counterparts of the three Helmholtz theorems for vortex tubes can also be established for electric flux tubes.

From (3.19) with $n^a = E^a/E$, Eq. (3.44), and Maxwell's equation (2.8), we have for a comoving observer at any point P on one of the electric field lines of an electric flux tube,

$$\frac{D}{dp}\ln(E\delta A) = \frac{1}{E} E^a{}_{;b} h^b_a = \frac{1}{E}(q - 2H_a\omega^a)\,. \tag{7.9}$$

We assume that (7.3) is satisfied so that (7.9) holds all along an electric field line and not merely at P, and we further consider a fluid for which

$$q - 2\omega_a H^a = 0\,; \tag{7.10}$$

then $E\delta A$ is constant along the flux tube, and can be taken as a measure of the strength of the tube.

Since we assume that (7.3) is valid, the electric flux tube moves with the fluid. Also, consideration of the propagation equation for E^a, (2.10), shows that the strength of the tube, $E\delta A$, is conserved as the tube moves with the fluid provided

$$I^a(H) - \mathscr{J}^a = 0\,. \tag{7.11}$$

8. CONGRUENCE OF COMBINED VORTEX AND MAGNETIC FIELD LINES

In the idealized approximation of infinite electric conductivity, the magnetic field satisfies certain conservation laws,[29] as does the vorticity in the barotropic flow of an uncharged perfect fluid.[26] However, in the barotropic flow of a charged perfect fluid with finite electric conductivity, the magnetic field and vorticity do not in general separately satisfy conservation laws and at most only a combination of magnetic field and vorticity is conserved. For this reason, Bekenstein and Oron[9] introduced the spacelike vector field (in the notation of this paper)

$$\Omega^a = r\omega^a - \tfrac{1}{2}eH^a\,, \tag{8.1}$$

where r is defined by (2.22) and $e = rq(\mu + p)^{-1}$. Ω^a has useful properties when the magnetic field is force-free, i.e., from (2.20)

$$\eta^{abcd} u_b \mathscr{J}_c H_d = 0\,, \tag{8.2}$$

and when e is constant everywhere. Force-free magnetic fields have been studied extensively in nonrelativistic theory and may have application in astrophysics.[30–33] A particular case for which (8.2) is satisfied is a charged fluid with null electric conductivity. If the fluid is also a perfect fluid with suitable equation of state then $\dot{e} = 0$,[22] and the more restrictive assumption of e constant everywhere will then be valid if the fluid is homogeneous.

A spacelike curve in the fluid everywhere tangent to the vector field Ω^a will be called an Ω-line. We will examine briefly the properties of a congruence of Ω-lines in a charged perfect fluid, assuming only that the flow is barotropic and that the magnetic field is force-free. We will impose restrictions on e later to obtain special cases of our general results.

A. Comoving observers

The unit tangent vector to an Ω-line at any point is $n^a = \Omega^a/\Omega$ where $\Omega = (\Omega_a \Omega^a)^{1/2}$; we have $n_a u^a = 0$. With the aid of the propagation equations for vorticity and magnetic field, it is proved in Appendix A that if the magnetic field is force-free then the propagation equation for n^a assumes the form

$$h^a_b \dot{n}^b = \overset{*}{u}{}^a + (\overset{*}{n}{}^b u_b)n^a + \frac{1}{2\Omega} p^a_s(\eta^{sbcd} u_b E_c e_{,d} - \dot{e}H^s)\,, \tag{8.3}$$

where p^a_s is given by (4.16). Substituting from (8.3) into (4.14) for N^a, we find that

$$N^a = \frac{1}{2\Omega} p^a_s(\eta^{sbcd} u_b E_c e_{,d} - \dot{e}H^s)\,. \tag{8.4}$$

Comoving observers can therefore be employed all along an Ω-line (and Ω-lines are material lines) if and only if

$$p^a_s(\eta^{sbcd} u_b E_c e_{,d} - \dot{e}H^s) = 0\,. \tag{8.5}$$

If $\dot{e}=0$, then (8.5) depends on H^a only through the projection tensor p^a_s while if e is constant everywhere then (8.5) is identically satisfied.

B. Flux tubes

Consider a flux tube composed of Ω-lines. From (3.19) with $n^a=\Omega^a/\Omega$ and (3.44), a comoving observer at any point P on one of the Ω-lines of the flux tube finds at P

$$\frac{D}{dp}\ln(\Omega\delta A)=\frac{1}{\Omega}\Omega^a{}_{;b}h^b{}_a\,. \tag{8.6}$$

But with the aid of the divergence equation for vorticity (5.49), the momentum conservation equation (2.21) subject to the force-free condition (8.2), and Maxwell's equation (2.7), it is proved in Appendix A that

$$\Omega^a{}_{;b}h^b{}_a=-\tfrac{1}{2}e_{,a}H^a\,; \tag{8.7}$$

thus

$$\frac{D}{dp}\ln(\Omega\delta A)=-\frac{1}{2\Omega}e_{,a}H^a\,. \tag{8.8}$$

We assume that (8.5) is satisfied so that (8.8) holds all along an Ω-line, and we also consider a fluid for which $e_{,a}H^a=0$. Then $\Omega\delta A$ is constant along the flux tube and is defined to be the strength of the tube. This provides a counterpart of the first Helmholtz theorem although the result is not purely kinematical even if e is constant everywhere.

Since we assume (8.5) is satisfied, the flux tube moves with the fluid. Further, consideration of the propagation equation for Ω^a [Appendix A, Eq. (A4)] shows that $\Omega\delta A$ will remain constant as the tube moves with the fluid, provided

$$\eta^{abcd}u_bE_ce_{,d}-\dot{e}H^a=0\,. \tag{8.9}$$

This result, for the case e constant everywhere, was obtained by Bakenstein and Oron.[9]

We note that all three conditions, Eq. (8.5), $e_{,a}H^a=0$, and Eq. (8.9), are satisfied if e is constant everywhere.

9. PROPAGATION EQUATIONS

A. Propagation equations for A_{ab}

The equations governing the propagation of $\mathscr{R}_{ab}$, $\mathscr{E}$, and $\mathscr{S}_{ab}$ along a curve of a spacelike congruence are obtained from the skew part, the trace, and the trace-free symmetric part, respectively, of the propagation equation for A_{ab}.

The propagation equation for A_{ab} is derived from the Ricci identity for n^a:

$$n_{a;bc}-n_{a;cb}=R_{tabc}n^t\,. \tag{9.1}$$

If we contract (9.1) with n^c and note that

$$n_{a;cb}n^c=\overset{*}{n}_{a;b}-n_{a;c}n^c{}_{;b}\,, \tag{9.2}$$

then we obtain

$$(n_{a;b})^*=\overset{*}{n}_{a;b}-n_{a;c}n^c{}_{;b}+R_{tabc}n^tn^c\,. \tag{9.3}$$

Substituting from (3.22) [with (3.17)] for $n_{a;b}$ into (9.3), projecting on the indices a and b, and using the transport law (3.13), we obtain the following propagation equation for A_{ab} (renaming the free indices):

$$p^c_a\,p^d_b\overset{*}{A}_{cd}=-A_{ac}A^c{}_b+p^c_a\,p^d_bR_{tcds}n^tn^s$$
$$+p^c_a\,p^d_b(\overset{*}{n}_{c;d}-\overset{*}{n}_c\overset{*}{n}_d+\mathring{n}_c\mathring{n}_d-2\mathring{n}_c(n^tw_{t;d}))\,. \tag{9.4}$$

In Newtonian theory, in place of the Ricci identity (9.1), we have

$$n_{\alpha,\beta\gamma}-n_{\alpha,\gamma\beta}=0\,, \tag{9.5}$$

and it can be verified that the propagation equation for $A_{\alpha\beta}$ is

$$p^\lambda_\alpha\,p^\tau_\beta\overset{*}{A}_{\lambda\tau}=-A_{\alpha\tau}A^\tau{}_\beta+p^\lambda_\alpha\,p^\tau_\beta(\overset{*}{n}_{\lambda,\tau}-\overset{*}{n}_\lambda\overset{*}{n}_\tau)\,. \tag{9.6}$$

This equation, unlike (9.4), does not contain a gravitational field term.

B. Propagation equation for $\mathscr{R}_{ab}$

The propagation equation for $\mathscr{R}_{ab}$ is the skew part of (9.4). Using (3.17) for A_{ab}, the skew part of (9.4) can be written as

$$p^c_a\,p^d_b(\overset{*}{\mathscr{R}}_{cd}-\overset{*}{n}_{[c;d]}-2n^tw_{t;[c}\mathring{n}_{d]})$$
$$-2\mathscr{S}_{c[a}\mathscr{R}_{b]}{}^c+\mathscr{E}\mathscr{R}_{ab}=0\,. \tag{9.7}$$

With the aid of (3.15) and (3.22), Eq. (9.7) may be expressed alternatively as

$$\underset{n}{\mathscr{L}}\,\mathscr{R}_{ab}=p^c_a\,p^d_b(\overset{*}{n}_{[c;d]}+2n^tw_{t;[c}\mathring{n}_{d]})\,. \tag{9.8}$$

Both (9.7) and (9.8) are purely kinematical results and, further, they are independent of the Riemann tensor R_{abcd}. These observations also apply to the propagation equation for the vorticity tensor ω_{ab} along a fluid particle world line.[12,13]

On taking the skew part of (9.6), a similar calculation yields the corresponding Newtonian results:

$$p^\lambda_\alpha\,p^\tau_\beta(\overset{*}{\mathscr{R}}_{\lambda\tau}-\overset{*}{n}_{[\lambda,\tau]})-2\mathscr{S}_{\tau[\alpha}\mathscr{R}_{\beta]}{}^\tau+\mathscr{E}\mathscr{R}_{\alpha\beta}=0\,, \tag{9.9}$$

$$\underset{n}{\mathscr{L}}\,\mathscr{R}_{ab}=p^\lambda_\alpha\,p^\tau_\beta\overset{*}{n}_{[\lambda,\tau]}\,. \tag{9.10}$$

Simplification in the propagation equation for $\mathscr{R}_{ab}$ can result either due to a symmetry property of the vector field n^a or due to the physical nature of the congruence. As an example of the former, it is easily verified that if there exists a spacelike Killing vector field parallel to n^a then the right hand side of (9.8) vanishes and hence

$$\underset{n}{\mathscr{L}}\,\mathscr{R}_{ab}=0\,. \tag{9.11}$$

This result was established by Prasad and Sinha[6] for a congruence of magnetic field lines, but it is true in general. As examples of the latter possibility, we establish the following two results.

Theorem: (i) If $\mathscr{R}_{ab}$ is the rotation tensor, as measured by a comoving observer, of a congruence of magnetic field lines in the MHD approximation of vanishing electric field,

then

$$\underset{n}{\mathscr{L}}\left(\frac{H^2}{(n_c\mathscr{J}^c)}\mathscr{R}_{ab}\right)=0\,, \tag{9.12}$$

where $n^a=H^a/H$, $\mathscr{J}^c$ is the conduction current, and it is assumed that $n_c\mathscr{J}^c\neq 0$.

(ii) If $\mathscr{R}_{ab}$ is the rotation tensor, as measured by a comoving observer, of a congruence of vortex lines in a fluid with an acceleration potential r, then

$$\underset{n}{\mathscr{L}}\left(\frac{r\omega^2}{n_cI^c(\omega)}\mathscr{R}_{ab}\right)=0\,, \tag{9.13}$$

where $n^a=\omega^a/\omega$, $I^c(\omega)$ is defined by (5.31), and it is assumed that $n_cI^c(\omega)\neq 0$.

Proof: (i) Equation (9.12) can be derived more readily if instead of using (9.8), we note that if $E^a=0$, then[34]

$$\underset{n}{\mathscr{L}}F_{ab}=0\,, \tag{9.14}$$

where F_{ab} is the electric field tensor (2.4) in the MHD approximation:

$$F_{ab}=\eta_{abcd}H^cu^d\,. \tag{9.15}$$

Equation (9.14) is a direct consequence of both the constraint equation (2.7) and the propagation equation (2.9) for H^a with $E^a=0$. But when $E^a=0$, comoving observers can be employed all along a magnetic field line and we suppose that this choice has been made. With $w^a=u^a$, it follows from (3.33), (6.8), and (9.15) that

$$F_{ab}=-\frac{2H^2}{(n_c\mathscr{J}^c)}\mathscr{R}_{ab}\,. \tag{9.16}$$

The result (9.12) follows immediately from (9.14) and (9.16).

(ii) With the aid of the propagation and constraint equations for the vorticity vector in a fluid with an acceleration potential,[12] it can be verified that

$$\underset{n}{\mathscr{L}}(r\omega_{ab})=0\,, \tag{9.17}$$

where $\omega_{ab}=\eta_{abcd}\omega^cu^d$. Also, comoving observers can be employed all along a vortex line in a fluid with an acceleration potential. Using (3.33) with $w^a=u^a$, and (5.30), we find that

$$\omega_{ab}=-\frac{2\omega^2}{(n_cI^c(\omega))}\mathscr{R}_{ab}\,. \tag{9.18}$$

Equation (9.13) follows immediately from (9.17) and (9.18). This completes the proof.

Unlike (9.8), Eqs. (9.12) and (9.13) are not purely kinematical results; Maxwell's equations were used in the derivation of (9.12) and Einstein's equations in (9.13). Equations (9.16) and (9.18) show the close relation between the rotation tensor and the electromagnetic field tensor on the one hand and the vorticity tensor on the other. A more general result, of which (9.12) and (9.13) are special cases, is derived in Appendix B.

Finally, we note that in the MHD approximation[34] and for a fluid with an acceleration potential, it can also be proved that

$$\underset{u}{\mathscr{L}}F_{ab}=0\,,\quad \underset{u}{\mathscr{L}}(r\omega_{ab})=0\,; \tag{9.19}$$

hence, using (9.16) and (9.18) we have

$$\underset{u}{\mathscr{L}}\left(\frac{H^2}{(n_c\mathscr{J}^c)}\mathscr{R}_{ab}\right)=0\,, \tag{9.20}$$

$$\underset{u}{\mathscr{L}}\left(\frac{r\omega^2}{n_cI^c(\omega)}\mathscr{R}_{ab}\right)=0\,, \tag{9.21}$$

where $\mathscr{R}_{ab}$ is defined for a comoving observer.

C. Propagation equation for $\mathscr{E}$

The propagation equation for $\mathscr{E}$ is the trace of (9.4). On contracting a with b in (9.4), using expansion (3.17) for A_{ab}, and expressing the Riemann tensor in terms of the Weyl and Ricci tensors through (5.16), we find that

$$\begin{aligned}\overset{*}{\mathscr{E}}+\tfrac{1}{2}\mathscr{E}^2=\;&2(\mathscr{R}^2-\mathscr{S}^2)\\&-C_{abcd}w^bw^dn^an^c-\tfrac{1}{2}R_{ab}w^aw^b\\&-\tfrac{1}{2}R_{ab}n^an^b-\tfrac{1}{6}R+p^{ab}(\overset{*}{n}_{a;b}-\overset{*}{n}_a\overset{*}{n}_b+\mathring{n}_a\mathring{n}_b\\&-2\mathring{n}_a(n^iw_{i;b}))\,,\end{aligned} \tag{9.22}$$

where the magnitude $\mathscr{S}$ of $\mathscr{S}_{ab}$ is defined by $\mathscr{S}^2=\frac{1}{2}\mathscr{S}_{ab}\mathscr{S}^{ab}$.

Equation (9.22) is purely kinematical. Consider now a spacelike congruence in a self-gravitating fluid, and assume that the unit tangent vector n^a to any curve of the congruence is everywhere orthogonal to the 4-velocity field u^a of the fluid. We will first analyze the special case in which the curves are material curves. This simplifies (9.22) for two reasons: comoving observers can be employed along a curve and the propagation equation for n^a along u^a is known and given by (4.8). We will then discuss briefly the modifications to the propagation equation for $\mathscr{E}$ in the general case when the curves are not frozen-in to the fluid.

Employing a comoving observer along the congruence, and making use of (4.8) and Einstein's field equations (5.17) where T_{ab} is decomposed with respect to u^a according to (5.18), we find that

$$\left.\begin{aligned}&\overset{*}{\mathscr{E}}+\tfrac{1}{2}\mathscr{E}^2\\&2\frac{\overset{**}{L}}{L}\end{aligned}\right\}=\begin{aligned}&2(\mathscr{R}^2-\mathscr{S}^2)\\&-E_{ab}n^an^b-\tfrac{2}{3}(\mu+\Lambda)-\tfrac{1}{2}\pi_{ab}n^an^b\\&+h^{ab}\overset{*}{n}_{a;b}+\overset{*}{u}^a\overset{*}{u}_a-2\overset{*}{u}^b(n^iu_{i;b})\end{aligned}\,, \tag{9.23}$$

where the representative length L is defined by (3.45).

From (9.6) the equation corresponding to (9.22) and (9.23) in Newtonian theory is

$$\left.\begin{aligned}&\overset{*}{\mathscr{E}}+\tfrac{1}{2}\mathscr{E}^2\\&2\frac{\overset{**}{L}}{L}\end{aligned}\right\}=2(\mathscr{R}^2-\mathscr{S}^2)+\overset{*}{n}{}^{\tau}{}_{,\tau}\,. \tag{9.24}$$

The terms on the right-hand side of (9.23) can be divided into three groups. Firstly, there is the rotation and shear of the congruence and these terms have the same form as in Newtonian theory. The rotation and shear tend to cause the curves to diverge and converge, respectively. The second group of terms describes the influence of the gravitational

field and represent a purely general-relativistic effect: $\mathscr{E}$ is affected directly by the fluid at each point through μ and π_{ab}, and also by the free gravitational field due to the surrounding fluid as represented by E_{ab}. The third group of terms depend on n^a and u^a, the latter being purely general relativistic in origin.

It is of interest to compare (9.23) with the Raychaudhuri equation which is the propagation equation for θ along a fluid particle world line[12,13]

$$\left.\begin{matrix}\dot{\theta}+\frac{1}{3}\theta^2\\ 3\dfrac{\ddot{l}}{l}\end{matrix}\right\} = 2(\omega^2-\sigma^2)+\dot{u}^a{}_{;a}-\tfrac{1}{2}(\mu+3p)+\Lambda\,, \tag{9.25}$$

where $\dot{l}/l=\theta/3$. The propagation of θ is not influenced by the free gravitational field, π_{ab} can occur only through the divergence of $\dot{u}^a$, there is explicit dependence on the pressure p, and if the cosmological constant is positive it tends to cause the curves to diverge in (9.25) but to converge in (9.23).

When the curves are not material curves we must return to Eq. (9.22). The contribution by the rotation and shear have the same form as when the curves are frozen-in to the fluid but of course $\mathscr{R}$ and $\mathscr{S}$ cannot now be defined with respect to comoving observers all along the congruence. To evaluate the gravitational field terms in (9.22) we can write $w^a=u^a+\lambda^a$ and solve Greenberg's transport law for λ^a subject to some initial condition on λ^a. It is found that with $\lambda^a\neq 0$, the propagation equation for $\mathscr{E}$ contains, besides the terms describing the gravitational field in (9.23), additional terms which depend on H_{ab}, p, and q^a as well as on E_{ab}, μ and π_{ab}. To develop the remaining terms in (9.22), knowledge of the propagation equation for n^a along u^a is required which, unlike the case of material curves, depends on the physical nature of the congruence.

D. Propagation equation for $\mathscr{S}_{ab}$

As with the propagation equation for $\mathscr{E}$, we will consider a spacelike congruence in a self-gravitating fluid, which may contain an electromagnetic field, and such that the unit tangent vector n^a to any curve of the congruence in everywhere orthogonal to the 4-velocity field u^a of the fluid. We will derive here for simplicity the propagation equation for $\mathscr{S}_{ab}$ for the special case in which the curves of the congruence are material curves. The more general case will contain extra terms which can be calculated by setting $w^a=u^a+\lambda^a$.

The propagation equation for $\mathscr{S}_{ab}$ is the symmetric trace-free part of (9.4). As previously, we express the Riemann tensor in terms of the Ricci and Weyl tensors using (5.16). The Ricci tensor can in turn be expressed in terms of the total energy–momentum tensor through Einstein's equations (5.17) where T_{ab} is decomposed with respect to u^a as in (5.18). The Weyl tensor can also be split with respect to u^a into its electric and magnetic components:

$$C_{ab}{}^{cd} = -8u_{[a}E_{b]}{}^{[c}u^{d]}+4g^{[c}_{[a}E^{d]}_{b]}+2\eta_{abrs}u^rH^{s[c}u^{d]} + 2\eta^{cdrs}u_rH_{s[a}u_{b]}\,, \tag{9.26}$$

where E_{ab} and H_{ab} are defined by (5.19). After a lengthy calculation, in which use is also made of (9.23) for $\overset{*}{\mathscr{E}}$, we find that

$$\begin{aligned}p^c_a\,p^d_b\,\overset{*}{\mathscr{S}}_{cd} = p^c_a\,p^d_b\Big[&\overset{*}{n}_{(c;d)}-\overset{*}{n}_c\overset{*}{n}_d+\dot{n}_c\dot{n}_d-2n^tu_{t;(c}\dot{n}_{d)}\Big]\\ &-\mathscr{R}_a\mathscr{R}_b-\mathscr{S}_{ac}\mathscr{S}^c_b-\mathscr{E}\mathscr{S}_{ab}\\ &-E_{ab}+2n_{(a}E_{b)c}n^c-\tfrac{1}{2}\pi_{ab}+n_{(a}\pi_{b)t}n^t\\ &+n_an_b(\mathscr{R}^2-E_{cd}n^cn^d-\tfrac{1}{2}\pi_{cd}n^cn^d)\\ &+p_{ab}\Big[\mathscr{S}^2-\tfrac{1}{2}E_{cd}n^cn^d-\tfrac{1}{4}\pi_{cd}n^cn^d-\tfrac{1}{2}h^{cd}\overset{*}{n}_{c;d}\\ &-\tfrac{1}{2}\overset{*}{u}^c\overset{*}{u}_c+\overset{*}{u}^c(n^tu_{t;c})\Big]\,.\end{aligned} \tag{9.27}$$

The corresponding result in Newtonian theory can be derived from (9.6)

$$\begin{aligned}p^\lambda_\alpha\,p^\tau_\beta\,\overset{*}{\mathscr{S}}_{\lambda\tau} = p^\lambda_\alpha\,p^\tau_\beta\overset{*}{n}_{\lambda,\tau}&-\overset{*}{n}_\alpha\overset{*}{n}_\beta\\ &-\mathscr{R}_\alpha\mathscr{R}_\beta-\mathscr{S}_{\alpha\tau}\mathscr{S}^\tau_\beta-\mathscr{E}\mathscr{S}_{\alpha\beta}+\mathscr{R}^2n_\alpha n_\beta\\ &+p_{\alpha\beta}(\mathscr{S}^2-\tfrac{1}{2}\overset{*}{n}{}^\tau_{,\tau})\,.\end{aligned} \tag{9.28}$$

In the first set of terms in (9.27), $\dot{n}^a$ can be replaced by $\overset{*}{u}{}^a$ using either (3.13) or, equivalently, the propagation equation for n^a along u^a, Eq. (4.8). It is then easily verified that the trace of the right-hand side of (9.27) vanishes as required, the term proportional to p_{ab} ensuring that this is the case. The terms involving the rotation, expansion, and shear of the congruence have the same form in general relativity as in Newtonian theory. When the curves are material curves, we see that the distortion induced in the congruence by the gravitational field depends only on π_{ab} and E_{ab}; this effect is purely general-relativistic. The tensors π_{ab} and E_{ab} also govern the propagation of σ_{ab} along a fluid particle world line.[12,13] Finally, we note that for a congruence of vortex lines, $\mathscr{S}_{ab}$ depends explicitly on H_{ab} in (5.45), while the propagation of $\mathscr{S}_{ab}$ along a frozen-in vortex line is controlled by E_{ab}.

10. DIVERGENCE EQUATION FOR ROTATION

In this section we consider the implication of the first Bianchi identity

$$R_{t[abc]}=0\,. \tag{10.1}$$

If we apply (10.1) to (9.1), then

$$n_{[a;bc]}=0\,, \tag{10.2}$$

and if we operate on (10.2) with $\eta^{dabc}w_d$ and use expansion (3.22) for $n_{a;b}$ then we find that

$$\begin{aligned}\mathscr{R}^a{}_{;b}\,p^b_a = \tfrac{1}{2}\eta^{abcd}n_aw_b&\Big[\overset{*}{n}_{c;d}-\overset{*}{\mathscr{R}}_{cd}-w_{c;d}(n^t\overset{*}{w}_t)\Big]\\ &-\tfrac{1}{2}\eta^{abcd}w_aw_{b;c}\left[\mathring{n}_d+n^tw_{t;d}\right]\,.\end{aligned} \tag{10.3}$$

Equation (10.3) is the divergence equation for the rotation vector and is a purely kinematical result. It can also be regarded as a constraint equation for rotation in an initial value problem for propagation along the congruence because it does not contain components of the derivative of the rotation perpendicular to the screen erected by w^a: the left hand side of (10.3) is the 2-dimensional divergence of $\mathscr{R}^a$ on the screen

and with $\overset{*}{\mathscr{R}}_{cd}$ on the right-hand side only its components on the screen contribute because of the operator $\eta^{abcd}n_a w_b$.

In Newtonian theory, the divergence equation for rotation is obtained by operating on the identity

$$n_{[\alpha,\beta\gamma]} = 0 \tag{10.4}$$

with $\eta^{\alpha\beta\gamma}$ and using (3.61) for $n_{\alpha,\beta}$:

$$\mathscr{R}^{\alpha}{}_{\beta}\, p^{\beta}_{\alpha} = \tfrac{1}{2}\eta^{\alpha\beta\gamma}n_{\alpha}(\overset{*}{n}_{\beta,\gamma} - \overset{*}{\mathscr{R}}_{\beta\gamma}) . \tag{10.5}$$

Equation (10.5) can be regarded as a constraint equation for rotation in Newtonian theory, the operator $\eta^{\alpha\beta\gamma}n_{\alpha}$ ensuring that with $\overset{*}{\mathscr{R}}_{\beta\gamma}$, only its components perpendicular to n^{α} contribute.

11. CONCLUDING REMARKS

Clearly, there are many other spacelike congruences of interest besides those considered in this paper, such as congruences defined in terms of the Poynting vector, the energy flux q^a, and the 4-acceleration $\dot{u}^a$. For spacelike congruences associated with a self-gravitating fluid we have seen that a central role is played in the theory by congruences of material curves; comoving observers can be employed all along a material curve and this greatly simplifies results such as the propagation equations for $\mathscr{R}_{ab}$, $\mathscr{E}$, and $\mathscr{S}_{ab}$.

The main difference between general relativity predictions for spacelike congruences and corresponding Newtonian results, is that the Newtonian results in general do not contain gravitational field terms. This was found to be the case with the propagation equations as well as with the expressions derived for the rotation, expansion, and shear of a vortex congruence. These Newtonian results were purely kinematical unlike their general-relativistic analogs.

Flux vectors play an important role in the determination of the rotation of the curves of a congruence. In the MHD approximation, the rotation of a congruence of magnetic field lines is determined directly at a given point by the conduction current $\mathscr{J}^a$ at that point and by a generalization of the conduction current $I^a(E)$ when $E^a \neq 0$. For a congruence of vortex lines, the general-relativistic result shows that $\mathscr{R}^a$ depends only on a generalization of the total energy flux q^a, the contribution due to q^a itself being absent in Newtonian theory.

ACKNOWLEDGMENT

One of the authors (M.T.) would like to thank the Applied Mathematics Department of the University of the Witwatersrand for their hospitality during the South African winter of 1980.

APPENDIX A

A. Propagation equation for Ω^a and $n^a = \Omega^a/\Omega$

We first use the momentum conservation equation (2.21), subject to the force-free magnetic field condition (8.2), and the identity (2.14) to rewrite the last term of the vorticity propagation Eq. (5.1) as

$$\tfrac{1}{2}\eta^{abcd}u_b\dot{u}_{c;d} = -\frac{\dot{r}}{r}\omega^a + \frac{1}{2r}\eta^{abcd}u_b E_c e_{,d} - \frac{e}{2r}I^a(E) . \tag{A1}$$

Equation (5.1) becomes

$$h^a_b(r\omega^b)^{\cdot} = u^a{}_{;b}r\omega^b - \theta r\omega^b + \tfrac{1}{2}\eta^{abcd}u_b E_c e_{,d} - \tfrac{1}{2}eI^a(E) , \tag{A2}$$

and on replacing $I^a(E)$ in (A2) using Maxwell's equation (2.9), we find that

$$h^a_b\dot{\Omega}^b = u^a{}_{;b}\Omega^b - \theta\Omega^a + \tfrac{1}{2}\eta^{abcd}u_b E_c e_{,d} - \tfrac{1}{2}\dot{e}H^a . \tag{A3}$$

Since $\theta = 3\dot{l}/l$, (A3) can be written concisely as

$$h^a_b(l^3\Omega^b)^{\cdot} = u^a{}_{;b}(l^3\Omega^b) + \tfrac{1}{2}l^3(\eta^{abcd}u_b E_c e_{,d} - \dot{e}H^a) , \tag{A4}$$

which is the required propagation equation for Ω^a.

To obtain the propagation equation for $n^a = \Omega^a/\Omega$, we contract (A3) with Ω_a; this gives

$$\dot{\Omega} = -(\theta + \overset{*}{n}_b u^b)\Omega^2 + \tfrac{1}{2}n_a(\eta^{abcd}u_b E_c e_{,d} - \dot{e}H^a) \tag{A5}$$

and with the aid of this result and (A3) we obtain

$$h^a_b\dot{n}^b = \overset{*}{u}{}^a + (\overset{*}{n}_b u^b)n^a + \frac{1}{2\Omega}p^a_s(\eta^{sbcd}u_b E_c e_{,d} - \dot{e}H^s) , \tag{A6}$$

where p^a_s is defined by (4.16).

B. Constraint equation for Ω^a

From definition (8.1) of Ω^a, we have

$$\Omega^a{}_{;b}h^b_a = (r\omega^a)_{;b}h^b_a - \tfrac{1}{2}eH^a{}_{;b}h^b_a - \tfrac{1}{2}e_{,a}H^a . \tag{A7}$$

But, using (2.21) for $\dot{u}^a$ subject to the force-free condition (8.2), the divergence equation for voricity (5.14) may be expressed as

$$(r\omega^a)_{;b}h^b_a = e\omega_a E^a . \tag{A8}$$

On substituting from (A8) and Maxwell's equation (2.7) in (A7), we obtain finally

$$\Omega^a{}_{;b}h^b_a = -\tfrac{1}{2}e_{,a}H^a . \tag{A9}$$

APPENDIX B

In this Appendix, we show that if the curves of the congruence are material curves with unit tangent vector n^a everywhere orthogonal to u^a and if n^a satisfies the constraint equation

$$(fn^a)_{;b}h^b_a = 0 \tag{B1}$$

for some fluid variable f, then

$$\mathscr{L}_n\left(\frac{f}{(n_c\mathscr{R}^c)}\mathscr{R}_{ab}\right) = 0 , \tag{B2}$$

where $\mathscr{R}_{ab}$ and $\mathscr{R}^a$ are defined for a comoving observer. We assume that $n_c\mathscr{R}^c \neq 0$; from (3.32) this is true if $\mathscr{R} \neq 0$.

To derive (B2), we first define

$$K_{ab} = \eta_{abcd}n^c u^d . \tag{B3}$$

Then, with the aid of the propagation equation (4.8) for n^a, which is valid because the curves are material curves, it can be shown that

$$\underset{n}{\mathscr{L}} K_{ab} = (n^c{}_{;d} h^d{}_c) K_{ab} \ . \tag{B4}$$

But if f is any fluid variable, it follows from (B4) that

$$\begin{aligned} \underset{n}{\mathscr{L}} (fK_{ab}) &= (fn^c)_{;d} h^d{}_c K_{ab} \\ &= 0 \end{aligned} \tag{B5}$$

if (B1) is satisfied. Further, comoving observers can be employed all along any given curve of the congruence. With $w^a = u^a$, a direct calculation using (3.32), (3.33), and (B3) gives

$$K_{ab} = \mathscr{R}_{ab} / n_c \mathscr{R}^c \ . \tag{B6}$$

The result (B2) follows immediately from (B5) and (B6).

Equations (9.12) and (9.13) are special cases of (B2) with $f = H$ and $f = r\omega$, respectively.

[1]P. J. Greenberg, J. Math. Anal. Applic. **30**, 128–43 (1970).
[2]T. H. Date, Ann. Inst. Henri Poincaré **24**, 417–27 (1976).
[3]G. Prasad, Ann. Inst. Henri Poincaré **30**, 17–26 (1979).
[4]G. Prasad, Gen. Rel. Grav. **11**, 177–88 (1979).
[5]G. Prasad, Inter. J. Theor. Phys. **18**, 329–43 (1979).
[6]G. Prasad and B. B. Sinha, I Nuovo Cimento **52B**, 105–12 (1979).
[7]G. Prasad and B. B. Sinha, I Nuovo Cimento **54B**, 401–12 (1979).
[8]G. Prasad and B. B. Sinha, Ann. Inst. Henri Poincaré **32**, 15–20 (1980).
[9]J. D. Bekenstein and E. Oron, Phys. Rev. D **18**, 1809–19 (1978).
[10]The notation and conventions of Refs. 12 and 13 will be followed throughout. Latin indices run over the four coordinates of space-time, Greek indices over the three spatial coordinates. A semicolon denotes covariant differentiation with respect to the metric tensor g_{ab} of space-time [signature $(-+++)$], and units are used for which the velocity of light and Einstein's gravitational constant are both unity. In Newtonian theory, $\{x^\alpha\}$ are fixed curvilinear coordinates, and a comma denotes covariant differentiation with respect to the flat 3-space metric tensor $h_{\alpha\beta}$ [signature $(+++)$]. Thus $h_{\alpha\beta,\gamma} = 0$.
[11]η^{abcd} is defined by $\eta^{abcd} = \eta^{[abcd]}$, $\eta^{0123} = (-g)^{-1/2}$, where $g = \det[g_{ab}]$. We make use of the following identities:
$\eta^{abcd}\eta_{qrst} = -4!\,\delta^{[a}_q \delta^b_r \delta^c_s \delta^{d]}_t$, $\eta^{abcd}\eta_{arst} = -3!\delta^{[b}_r \delta^c_s \delta^{d]}_t$,
$\eta^{abcd}\eta_{abst} = -4\delta^{[c}_s \delta^{d]}_t$.
[12]G. F. R. Ellis, in *Cargèse Lectures in Physics, Vol. 6*, edited by E. Schatzmam (Gordon and Breach, New York, 1973), pp. 1–60.
[13]G. F. R. Ellis, in *General Relativity and Gravitation*, edited by R. K. Sachs (Academic, New York, 1971), pp. 104–82.
[14]F. A. E. Pirani, Bull. Acad. Polon. Sci. **5**, 143–46 (1957).
[15]$\eta^{\alpha\beta\gamma}$ is defined by $\eta^{\alpha\beta\gamma} = \eta^{[\alpha\beta\gamma]}$, $\eta^{123} = h^{-1/2}$, where $h = \det[h_{\alpha\beta}]$. We make use of the following identities:
$\eta^{\alpha\beta\gamma}\eta_{\nu\lambda\mu} = 3!\,\delta^{[\alpha}_\nu \delta^\beta_\lambda \delta^{\gamma]}_\mu$,
$\eta^{\alpha\beta\gamma}\eta_{\alpha\lambda\mu} = 2\,\delta^{[\beta}_\lambda \delta^{\gamma]}_\mu$, $\eta^{\alpha\beta\gamma}\eta_{\alpha\beta\mu} = 2\,\delta^\gamma_\mu$.
[16]Equation (4.7) can be written in a more familiar form. Since $n_b u^b = 0$, we have $\overset{*}{n}_b u^b = -u_{b;c} n^b n^c$, and using

$$u_{b;c} = \theta_{bc} + \omega_{bc} - \dot{u}_b u_c \ ,$$

where $\theta_{bc} = \theta_{(bc)}$ is the expansion tensor and $\omega_{bc} = \omega_{[bc]}$ is the vorticity tensor, we find that

$$\frac{(\delta l)^{\cdot}}{\delta l} = \theta_{bc}\, n^b n^c$$

in agreement with the general theory.[12,13]
[17]The manner in which a vortex line may be constructed in the fluid, originating at some point P, is explained in detail by Greenberg, Ref. 1, p. 139.
[18]Reference 1, p. 140.
[19]A. Trautman, in *Lectures in General Relativity* (*Brandeis Summer Institute in Theoretical Physics*, Vol. 1), edited by S. Deser and K. Ford (Prentice Hall, Englewood Cliffs, NJ, 1964), pp. 215, 217.
[20]C. D. Ciubotariu, Phys. Lett. A **40**, 369–70 (1972).
[21]M. J. Dupré and S. I. Rosencrans, J. Math. Phys. **19**, 1532–5 (1978).
[22]An easily demonstrated case in which e is constant, at least along the world line of u^a, is a charged fluid with null electric conductivity and equation of state $p = (\gamma - 1)\mu$, where γ is a constant $(1 \leqslant \gamma \leqslant 2)$. The result follows from (2.22) and the conservation equations (2.19) and (2.23) with $\mathscr{J}^a = 0$; we find that $e \propto q\mu^{-1/\gamma}$ and constant along u^a. To examine the physical interpretation of e, we note that $\theta = (\delta V)^{\cdot}/\delta V$ where δV is a comoving volume element. With the aid of (2.19) with $\mathscr{J}^a = 0$, we find that $e \propto q\delta V$ along u^a; if the nonconducting fluid is uniformly charged then along the world line of u^a, e is proportional to the charge per particle.
[23]M. Tsamparlis and D. P. Mason, Gen. Rel. Grav. **14**, 717–726 (1982).
[24]J. Serrin, in *Handbuch der Physik, Vol. 8/1)*, edited by S. Flügge (Springer, Berlin, 1959), pp. 162–3.
[25]Equation (5.48) can always be rewritten in a form with the right-hand side zero. For, if f is any fluid variable, it follows from (5.48) that $(D/dp)\ln(f\omega\delta A) = (1/f\omega)(f\omega^a)_{;b}\, h^b_a$. But if we choose $f = \exp(-\int_{p_0}^{p} \dot{u}_a\, n^a\, dp)$, where integration is along the vortex line, then (5.49) becomes $(f\omega^a)_{;b}\, h^b_a = 0$, and the result follows. Essentially, $f = r$ in the two applications of this subsection.
[26]Reference 12, p. 27.
[27]A. K. Raychaudhuri and U. K. De, J. Phys. A: Gen. Phys. **3**, 263–8 (1970).
[28]A justification of the MHD approximation, even when the anisotropy of the electric conductivity due to the magnetic field is taken into account, is given in Ref. 9, p. 1810.
[29]Reference 12, pp. 35 and 36.
[30]R. Lüst and A. Schlüter, Z. Astrophys. **34**, 263–82 (1954).
[31]S. Chandrasekhar, Proc. Nat. Acad. Sci. **42**, 1–5 (1956).
[32]S. Chandrasekhar, Astrophys. J. **124**, 232–43 (1956).
[33]S. Chandrasekhar and P. C. Kendall, Astrophys. J. **126**, 457–60 (1957).
[34]A. Lichnerowicz, *Relativistic Magnetohydrodynamics* (Benjamin, Amsterdam, 1967), p. 115.

Summational invariants in the mechanics of mass points[a)]

J. M. Amigó and H. Reeh
Institut für Theoretische Physik der Universität, Bunsenstrasse 9, D-3400 Göttingen, West Germany

(Received 19 March 1982; accepted for publication 11 June 1982)

For systems of point particles, summational invariants (the analog to symmetries of the S-matrix, and in general depending on momenta and coordinates of the particle and explicitly on time) are defined. For systems with interaction, their most general form is derived.

PACS numbers: 05.20.Gg

1. INTRODUCTION AND STATEMENTS

Consider a system of N mass points. A state of the system is represented by a point (or, equivalently, by an orbit) in phase space Γ

$$\vec{x}_1(t), \vec{p}_1(t), \ldots, \vec{x}_N(t), \vec{p}_N(t),$$

and an observable by a (real) function

$$F(\vec{x}_1, \vec{p}_1, \ldots, t)$$

which, as indicated, may have an additional dependence on t. The F belong to a suitable class of functions. Requiring that an observable should vary continuously with the state, one is led to the class of continuous functions on Γ. If, instead, one represents a state by a distribution on Γ (which, e.g., could be a characteristic function of a region), one is led for the observables to the larger class of locally L^1-functions. For symmetry reasons, we allow the dependence on t to be locally L^1 too.

An observable is called conserved, if it is constant in time.

In case the interaction between the particles is vanishing with increasing distance sufficiently fast, there are orbits which for $t \to \pm\infty$ look like orbits of free particles, i.e., there exists an asymptotic condition:

$$|\vec{x}_i(t) - \vec{x}_i^{\text{in}}(t)| + |\vec{p}_i(t) - \vec{p}_i^{\text{in}}| \to 0 \quad \text{as } t \to -\infty,$$

$$|\vec{x}_i(t) - x_i^{\text{out}}(t)| + |\vec{p}_i(t) - \vec{p}_i^{\text{out}}| \to 0 \quad \text{as } t \to +\infty.$$

We assume asymptotic completeness: The set of all incoming states is the same as the set of all outgoing states and the same as the set of all orbits of free particles, up to sets of measure 0. [The orbits, e.g., are characterized by the values of the $\vec{x}_i(t), \vec{p}_i(t)$ at $t = 0$.]

We call a conserved quantity additive (or more precisely asymptotically additive) in case

$$F(\vec{x}_1, \vec{p}_1, \ldots, t) = \sum_{i=1}^{N} f_i(\vec{x}_i^{\text{in}}(t), \vec{p}_i^{\text{in}}, t) = \sum_{i=1}^{N} f_i(\vec{x}_i^{\text{out}}(t), \vec{p}_i^{\text{out}}, t)$$

for all orbits with asymptotic condition. The f_i may be different for different kinds of particles.

Lemma 1: Let F be conserved and additive. Then there is a decomposition

$$F = \sum_{i=1}^{N} f_i(\vec{x}_i^{\text{in}}, \vec{p}_i^{\text{in}}, t) = \sum_{i=1}^{N} f_i(\vec{x}_i^{\text{out}}, \vec{p}_i^{\text{out}}; t)$$

with

$$\frac{d}{dt} f_i(\vec{x}, \vec{p}; t) = 0$$

for every orbit of a free particle (d/dt in distributive sense).

Proof: Consider the case of two particles. (The general case follows along the same line.) From

$$\frac{d}{dt} f_1(\vec{x}_1, \vec{p}_1, t) + \frac{d}{dt} f_2(\vec{x}_2, \vec{p}_2, t) = 0,$$

and since asymptotic completeness allows $\vec{x}_1, \vec{p}_1$ and $\vec{x}_2, \vec{p}_2$ to vary independently, it follows that

$$\frac{d}{dt} f_1(\vec{x}_1, \vec{p}_1, t) = c_1(t),$$

$$\frac{d}{dt} f_2(\vec{x}_2, \vec{p}, t) = c_2(t)$$

with $c_1(t) + c_2(t) = 0$. Hence,

$$f_i(\vec{x}, \vec{p}, t) = g_i(\vec{x}, \vec{p}, t) + k_i(t),$$

where

$$\frac{d}{dt} g_i(\vec{x}, \vec{p}, t) = 0$$

and

$$\frac{d}{dt}(k_1(t) + k_2(t)) = c_1(t) + c_2(t) = 0.$$

With $k := k_1 + k_2$, $F - k$ has the desired decomposition. □

In case of $f_i(\vec{x}, \vec{p}, t)$, $i = 1, \ldots, N$ satisfying

$$\sum_{i=1}^{N} f_i(\vec{x}_i^{\text{in}}, \vec{p}_i^{\text{in}}, t) = \sum_{i=1}^{N} f_i(\vec{x}_i^{\text{out}}, \vec{p}_i^{\text{out}}, t) \tag{1}$$

and

$$\frac{d}{dt} f_i(\vec{x}_i, \vec{p}_i, t) = 0 \tag{2}$$

for all free orbits, where the f_i are not necessarily related to an F which is conserved along the actual path, one has the analog of a symmetry of the S-matrix which in classical mechanics is also called "collision invariant" or "summational invariant." At first sight, one may expect that there are more of them compared to the additive conservation laws. It will turn out, however, that this is not the case.

In addition to the asymptotic condition, we need a cluster property: We assume that the interaction between the particles is decreasing sufficiently fast such that we may choose in the limit initial conditions for any of the particles such that this particle never interacts with the rest and its contribution drops from Eq. (1).

[a)] A preliminary report of this work can be found in the proceedings of the conference "Fundamental Problems of Theoretical and Mathematical Physics," Dubna, August 23–27, 1979.

 0022-2488/83/061594-09$02.50 © 1983 American Institute of Physics

In general, it is very difficult to list all conservation laws for a given system of particles. It is easy, however, to list all its summational invariants and additive conservation laws.

A special case is well known from the kinetic theory of gases in connection with the Boltzmann equation. There one is led to the equation

$$f_1(\vec{p}_1)+f_2(\vec{p}_2)=f_1(\vec{p}_3)+f_2(\vec{p}_4) \tag{3}$$

whenever

$$\vec{p}_1+\vec{p}_2=\vec{p}_3+\vec{p}_4\,, \tag{4a}$$

$$\omega_1(\vec{p}_1)+\omega_2(\vec{p}_2)=\omega_1(\vec{p}_3)+\omega_2(\vec{p}_4)\,, \tag{4b}$$

with $\omega_i(\vec{p})=\vec{p}^{\,2}/2m_i$ or $(\vec{p}^{\,2}+m_i^2)^{1/2}$. This, apparently, is the case of a summational invariant of a system of two particles where the f_i depend only on the momenta.

The solution of Eqs. (3) and (4) is given by

Theorem 1: Let $f_i(\vec{p})$, $i=1,2$, be locally L^1. Assume Eq. (3) on the scattering manifold (4), at least for subsets which correspond after transformation to the lab systems of particle 1 and 2 to open subsets U_i of the states allowed by Eq. (4) in the lab systems. The U_i may be different for different intervals of the scattering energy in the lab systems. Then

$$f_i(\vec{p})=a^0\omega_i(\vec{p})+\vec{a}\cdot\vec{p}+c_i \tag{5}$$

(almost everywhere) with constant $\vec{a}$, a^0, c_i.

Remarks: 1. Only the constant c_i may depend on the kind of the particle (e.g., mass or color). This is the only classical remnant of internal symmetries of elementary particles.

2. In case there is no interaction between the particles, the U_i shrink to a point at scattering angle 0 and the f_i may be arbitrary functions.

3. The theorem is formulated for space dimension $d=3$. There is an analog for $d=2$. However, there is no analog in case $d=1$. Indeed, when $m_1=m_2$, the scattering manifold is reduced to two points and every function $f(\vec{p})=f_1(\vec{p})=f_2(\vec{p})$ is a solution of Eqs. (3) and (4).

A proof of Theorem 1 will be given in the next section.

Proofs of the assertion of Theorem 1 considering f_i from slightly different classes of functions and assuming Eq. (3) on the whole of the scattering manifold defined by Eq. (4) were given in Refs. 1 and 2 for the nonrelativistic case, in Refs. 3 and 4 for the relativistic case, and in Ref. 5. All these proofs are rather complicated and suffer from the fact that they really need Eq. (3) on the whole of Eq. (4). But there are simple examples of interactions not fulfilling this assumption (e.g., a smooth interaction potential of finite range and finite height for which, for sufficiently high interaction energy, there is no scattering in backward directions). It should be mentioned, however, that Ref. 4 is more general in the respect that it works for $f_i(\vec{p})$ which need only be measurable. Presumably, there are other proofs in unpublished lecture notes on kinetic theory of gases.

In case of a general summational invariant, the f_i depend on $\vec{p}$, $\vec{x}$ and explicitly on t:

Theorem 2: Consider a system of N particles with central force interactions such that scattering in a neighborhood of forward scattering between each two particles occurs and depends there continuously on the impact parameter ρ. If the interaction has infinite range, it is assumed to decrease such that for the scattering angle θ in the lab systems $|\theta|\leqslant\text{const}/\rho^2$, $|\partial\theta/\partial\rho|\leqslant\text{const}/\rho^3$ for large ρ. Assume the f_i to be locally L^1. Then

$$f_i(\vec{x},\vec{p},t)=a^0\frac{\vec{p}^{\,2}}{2m_i}+\vec{a}\vec{p}+\vec{b}\,(\vec{x}\times\vec{p})+\vec{c}\vec{s}_i+d\frac{\vec{p}\cdot\vec{s}_i}{m_i}+e\frac{\vec{s}_i^{\,2}}{m_i}+k_i$$

with $\vec{s}_i=m_i\vec{x}-\vec{p}t$ in the Galilei-invariant case, and

$$f_i(\vec{x},\vec{p},t)=a^0(\vec{p}^{\,2}+m_i^2)^{1/2}+\vec{a}\vec{p}+\vec{b}\,(\vec{x}\times\vec{p})+\vec{c}\,\vec{s}_i+k_i$$

with $\vec{s}_i=(\vec{p}^{\,2}+m_i^2)^{1/2}\vec{x}-\vec{p}t$ in the Lorentz-invariant case, with constant $a^0,\ldots,k_i$.

Here, as forward scattering we denote scattering in the limit of the impact parameter approaching the range of the interaction (finite or infinite)—observe that in case of scattering by a potential $V(r)=\alpha/r^2$, $\theta\approx\text{const}/\rho^2$ for large ρ.

Remark: In comparison with Theorem 1, the arbitrary open sets U_i are replaced by a neighborhood of forward scattering in the lab systems of each particle. In case that the f_i have no explicit dependence on t; this more restrictive assumption is not necessary as will be seen later on (Lemma 3). Up to now, however, we do not have a proof for the general case without this assumption. [There are interaction potentials for which the scattering angle is discontinuous in the impact parameter at forward scattering. For spherically symmetric potentials it is continuous if (i) $V(r)$ has finite range r_0, $V(r)$ is continuous in a neighborhood of r_0, $V'(r)$ exists in a neighborhood on the left of r_0, and $\lim_{r\uparrow r_0}V'(r)\leqslant 0$. (ii) $V(r)$ has infinite range, $V'(r)$ exists for sufficiently large r, and $|r^3V'(r)|\leqslant\alpha<\infty$ for $r\to\infty$.]

See also the remark following the proof of Theorem 2 for the Galilei-invariant case in the next section.

2. PROOFS

As a first step we consider:

Lemma 2: Let $f_i(\vec{p})$, $i=1,2$, be continuously differentiable. Then the assertion of Theorem 1 holds.

Proof: (This generalizes an idea which one of the authors learned from D. Maison): Consider for a scattering process

$$\begin{aligned}&f_1(\vec{p}_1)+f_2(\vec{p}_2)=f_1(\vec{p}_3)+f_2(\vec{p}_4)\,,\\&\vec{p}_4=\vec{p}_1+\vec{p}_2-\vec{p}_3\,,\\&h_1(\vec{p}_1^{\,2})+h_2(\vec{p}_2^{\,2})=h_1(\vec{p}_3^{\,2})+h_2(\vec{p}_4^{\,2})\,,\end{aligned} \tag{6}$$

where $h_i(\vec{p}^{\,2})=\omega_i(\vec{p})$, with $\vec{p}_4=0$ (time reversed process to scattering of a particle with momentum $\vec{p}_3$ in the lab system). By varying the impact parameter for fixed $\vec{p}_1,\vec{p}_2$, we get a small change of $\vec{p}_3$ by $\delta\vec{p}_3$ and

$$0=\delta\vec{p}_3(\vec{\nabla}f_1(\vec{p}_3)-\vec{\nabla}f_2(\vec{p}_4)) \tag{7}$$

whenever

$$0=\delta\vec{p}_3(\vec{p}_3h'(\vec{p}_3^{\,2})-\vec{p}_4h'_1(\vec{p}_4^{\,2}))\,, \tag{8}$$

where still $\vec{p}_4=0$.

We may replace $f_i(\vec{p})$ by

$$\hat{f}_i(\vec{p}) := f_i(\vec{p}) - \vec{p}\vec{\nabla} f_2(0),$$

$i = 1,2$. Then Eq. (6) stays true for $\hat{f}_i$ instead of f_i, and Eq. (7) reduces to

$$0 = \delta\vec{p}_3 \cdot \vec{\nabla} \hat{f}_1(\vec{p})$$

whenever $\delta\vec{p}_3 \perp \vec{p}_3$ according to Eq. (8). Hence,

$$\hat{f}_1(\vec{p}_3) =: g_2(\vec{p}_3^2)$$

depends only on $\vec{p}_3^2$ and

$$f_1(\vec{p}) = \vec{p}\vec{\nabla} f_2(0) + g_1(\vec{p}^{\,2})$$

(because of continuity of f_1, g_1 cannot have an additional dependence on the energy interval), and

$$\vec{\nabla} f_1(0) = \vec{\nabla} f_2(0). \tag{9}$$

In case the particles are different (otherwise simply drop the index of f, g), we repeat the procedure after permuting the role of particles 1 and 2 and get

$$f_2(\vec{p}) = \vec{p}\vec{\nabla} f_2(0) + g_2(\vec{p}^{\,2}),$$

where (9) was applied.

For showing that $g_i(\vec{p}^{\,2}) = a^0 h_i(\vec{p}^{\,2}) + c_i$, we introduce $g_i(\vec{p}^{\,2})$ instead of $f_i(\vec{p})$ into (6) and obtain in a neighborhood of $\vec{p}_4 = 0$, but now with $\vec{p}_4 = \vec{p}_1 + \vec{p}_2 - \vec{p}_3 \neq 0$,

$$0 = \delta\vec{p}_3(\vec{p}_3 g_1'(\vec{p}_3^2) - \vec{p}_4 g_2'(\vec{p}_4^2)) \tag{10}$$

when

$$0 = \delta\vec{p}_3(\vec{p}_3 h_1'(\vec{p}_3^2) - \vec{p}_4 h_2'(p_4^2)). \tag{11}$$

Multiplication of Eq. (10) by $h_1'(p_3^2)$ yields, because of (11),

$$0 = \delta\vec{p}_3 \vec{p}_4(g_1'(\vec{p}_3^2) h_2'(p_4^2) - g_2'(p_4^2) h_1'(\vec{p}_3^2)).$$

Now $\delta\vec{p}_3 \cdot \vec{p}_4 = \delta\vec{p}_3(\vec{p}_1 + \vec{p}_2 - \vec{p}_3) \neq 0$ [in the case of central forces one may argue as follows: According to Eq. (11), $\delta\vec{p}_3 \cdot \vec{p}_4 = 0$ would imply $\delta\vec{p}_3 \cdot \vec{p}_3 = 0$ and $\delta\vec{p}_3 \perp (\vec{p}_1 + \vec{p}_2)$, thus $\delta\vec{p}_3 = 0$ since $\vec{p}_1 + \vec{p}_2$ and $\vec{p}_3$ cannot be parallel for $\vec{p}_4 \neq 0$ and $\delta\vec{p}_3, \vec{p}_3, \vec{p}_4$ are coplanar—in the general case one can always find a change of scattering parameters resulting in a change of the energies of the particle after collision.] Hence,

$$g_1'(\vec{p}_3^2) = \frac{g_2'(\vec{p}_4^2)}{h_2'(\vec{p}_4^2)} h_1'(p_3^2),$$

which by continuity extends to $\vec{p}_4 = 0$,

$$g_1'(\vec{p}^{\,2}) = \frac{g_2'(0)}{h_2'(0)} h_1'(\vec{p}^{\,2}),$$

writing $\vec{p}$ instead of $\vec{p}_3$.

Exchanging the role of particle 1 and 2, we obtain

$$g_2'(\vec{p}^{\,2}) = \frac{g_1'(0)}{h_1'(0)} h_2'(\vec{p}^{\,2}).$$

Integration yields

$$g_1(\vec{p}^{\,2}) = a h_1(\vec{p}^{\,2}) + c,$$
$$g_2(\vec{p}^{\,2}) = b h_2(\vec{p}^{\,2}) + d.$$

Inserting into Eq. (6), it follows $a = b$. □

Proof of Theorem 1:

(i) *Nonrelativistic case*: For f_i locally L^1 we have Eq. (3) for $\vec{p}_1, \vec{p}_2, \vec{p}_3, \vec{p}_4$ from a certain subset V of the scattering manifold $\mathcal{M}$ defined by Eq. (4). Looking at the Galilei-boosted scattering process, we get

$$f_1(\vec{p}_1 - m_1\vec{v}) + f_2(\vec{p}_2 - m_2\vec{v}) = f_1(\vec{p}_3 - m_1\vec{v}) + f_2(\vec{p}_4 - m_2\vec{v}).$$

$\mathcal{M}$ and V are Galilei invariant.

Put now for a test function $\chi \in \mathcal{D}(\mathbb{R}^3)$,

$$F_\chi^i(\vec{p}) = \int f_i(\vec{p} - m_i\vec{v})\chi(\vec{v})\, d^3\vec{v},$$

which, essentially, is a convolution. Then, F_χ^i is C^∞ in $\vec{p}$ and satisfies Eq. (3) on $V \subset \mathcal{M}$. Lemma 2 yields

$$F_\chi^i = a^0(\chi)\vec{p}^{\,2}/2m_i + \vec{a}(\chi)\vec{p} + c_i(\chi) \tag{12}$$

for every $\chi \in \mathcal{D}(\mathbb{R}^3)$. In particular, we consider a sequence $\chi_n(\vec{x}) \to \delta(\vec{x})$ in distributive sense. Then, as a convolution,

$$F_{\chi_n}^i \to f_i \quad \text{in } \mathcal{D}'(\mathbb{R}^3),$$

i.e., almost everywhere on $\mathbb{R}^3$.

To show that the limits on the rhs of Eq. (12) exist separately we smear it with $\varphi(\vec{p}) \in \mathcal{D}(\mathbb{R}^3)$ and get

$$\int f_i(\vec{p})\varphi(\vec{p})\, d^3\vec{p} = \lim_{n\to\infty} \left\{ a^0(\chi_n) \int \frac{\vec{p}^{\,2}}{2m_i} \varphi(\vec{p})\, d^3\vec{p} + \vec{a}(\chi_n) \int \vec{p}\varphi(\vec{p})\, d^3\vec{p} + c_i(\chi_n) \int \varphi\, d^3\vec{p} \right\}.$$

Since this holds for all φ, the separate limits on the rhs exist. Hence (5) is proved.

(ii) *Relativistic case*: Consider a Lorentz transformation $\Lambda \in L^\uparrow_+$ and the left and right invariant Haar measure $\mu(\Lambda)$ ($L^\uparrow_+$ is unimodular!). The constraints (4) are Lorentz invariant. Hence, again we may replace $f_i(\vec{p})$ in Eq. (3) by

$$F_\chi^j(\vec{p}) = \int f_j(\overrightarrow{\Lambda^{-1}p_j})\chi(\Lambda)\, d\mu(\Lambda), \quad j = 1,2,$$

$p_j = ((\vec{p}^{\,2} + m_j^2)^{1/2}, \vec{p})$, $\chi \in \mathcal{D}(L^\uparrow_+)$. $F_\chi^j(\vec{p})$ is $C^\infty(\mathbb{R}^3)$ since it can be interpreted as convolution: Let $e_j = (m_j, 0)$ and consider a boost $\Lambda_{j,\vec{p}} \in L^\uparrow_+$ such that $\Lambda_{j,\vec{p}} e_j = p_j$. Then

$$f_j(\overrightarrow{\Lambda^{-1}p_j}) = f_j(\overrightarrow{\Lambda^{-1}\Lambda_{j,\vec{p}}e_j}) =: f_{e_j}(\Lambda^{-1}\Lambda_{j,\vec{p}})$$

and

$$F_\chi^j(\vec{p}) = \int_{L^\uparrow_+} f_{e_j}(\Lambda^{-1}\Lambda_{j,\vec{p}})\chi(\Lambda)\, d\mu(\Lambda) = (f_{e_j} * \chi)(\Lambda_{j,\vec{p}}).$$

The function $f_{e_j} : L^\uparrow_+ \to \mathbb{R}$ is locally μ integrable (see Appendix A). Then the convolution exists and is $C^\infty(\mathbb{R}^3)$.

By Lemma 2,

$$(f_{e_j} * \chi)(\Lambda_{j,\vec{p}}) = a_\mu(\chi) p_j^\mu + c_j(\chi)$$

and we can now complete the proof as in the nonrelativistic case. □

Remark: It is now apparent that for Theorem 1 one needs only assume Eq. (3) in a weak sense, i.e., after averaging with a test function over Galilei boosts or Lorentz transformations, respectively. This is satisfying from a physical point of view since it corresponds to averaging over observations from slightly different reference systems. In this sense, it is clear that Theorem 1 (and correspondingly Theorem 2) can be extended to $f_i \in \mathscr{D}'(\mathbb{R}^3)$.

As a first step for proving Theorem 2, let us consider the case where $f_i(\vec{x}, \vec{p}, t)$ have no explicit dependence on t.

Lemma 3: Consider a two-particle system with central force interaction and let $f_i(\vec{x}, \vec{p})$ be continuously differentiable with

$$f_1(\vec{x}_1, \vec{p}_1) + f_2(\vec{x}_2, \vec{p}_2) = f_1(\vec{x}_3, \vec{p}_3) + f_2(\vec{x}_4, \vec{p}_4) \tag{13}$$

for $\vec{x}_1, \vec{p}_1, \ldots \vec{x}_4, \vec{p}_4$ from a (Galilei- or Poincaré-) invariant subset V of the manifold defined by

$$\begin{aligned} &\omega_1(\vec{p}_1) + \omega_2(\vec{p}_2) = \omega_1(\vec{p}_3) + \omega_2(\vec{p}_4)\,, \\ &\vec{p}_1 + \vec{p}_2 = \vec{p}_3 + \vec{p}_4\,, \\ &\vec{x}_1 \times \vec{p}_1 + \vec{x}_2 \times \vec{p}_2 = \vec{x}_3 \times \vec{p}_3 + \vec{x}_4 \times \vec{p}_4\,, \end{aligned} \tag{14}$$

V specifying the realizable scattering processes. Assume scattering occurs for every energy E and some impact parameter ρ (depending on E) in the lab systems of particles 1 and 2 and is continuous in the sense that the data of the outgoing particles vary continuously in (E, ρ) from an open neighborhood of the point considered. Then

$$f_i = a^0 \omega_i(\vec{p}) + \vec{a}\vec{p} + \vec{b}\,(\vec{x} \times \vec{p}) + c_i\,.$$

Proof: Choose the origin within the scattering plane. f_i may be written as

$$f_i(\vec{x}, \vec{p}) = \hat{f}_i(\vec{p}, \vec{x} \times \vec{p})\,,$$

with $\hat{f}_i$ continuously differentiable for $\vec{p} \neq 0$. For seeing this, decompose

$$\vec{x} = \left(\vec{x} \cdot \frac{\vec{p}}{|\vec{p}|}\right) \frac{\vec{p}}{|\vec{p}|} + \frac{\vec{p}}{|\vec{p}|} \times \left(\vec{x} \times \frac{\vec{p}}{|\vec{p}|}\right),$$

showing that f_i may be written as

$$f_i(\vec{x}, \vec{p}) = g_i\left(\vec{p}, \vec{x} \times \frac{\vec{p}}{|\vec{p}|}, \vec{x} \cdot \frac{\vec{p}}{|\vec{p}|}\right), \quad \vec{p} \neq 0\,.$$

This cannot depend on $x \cdot (\vec{p}/|\vec{p}|)$ since $d f_i(\vec{x}, \vec{p})/dt = 0$ for free orbits. Put

$$f(\vec{x}, \vec{p}) = g(\vec{p}, \vec{x} \times (\vec{p}/|\vec{p}|)) = \hat{f}(\vec{p}, \vec{x} \times \vec{p})\,.$$

Insertion into Eq. (13), at first for $\vec{p}_i \neq 0$, yields

$$\begin{aligned} &g_1\left(\vec{p}_1, \vec{x}_1 \times \frac{\vec{p}_1}{|\vec{p}_1|}\right) + g_2\left(\vec{p}_2, \vec{x}_2 \times \frac{\vec{p}_2}{|\vec{p}_2|}\right) \\ &\quad = g_1\left(\vec{p}_3, \vec{x}_3 \times \frac{\vec{p}_3}{|\vec{p}_3|}\right) + g_2\left(\vec{p}_4, \vec{x}_4 \times \frac{\vec{p}_4}{|\vec{p}_4|}\right). \end{aligned} \tag{15}$$

Go now to the rest system $\vec{p}_2 \to 0$ with $\vec{e} = \vec{p}_2/|\vec{p}_2|$ fixed. By Eq. (15),

$$g_2(\vec{p}_2, \vec{x}_2 \times (\vec{p}_2/|\vec{p}_2|)) \to g_2(0, \vec{x}_2 \times \vec{e})\,.$$

Since the data of the other particles do not depend on how $\vec{p}_2 \to 0$, Eq. (15) implies that $g_2(0, \vec{x}_2 \times \vec{e})$ does not depend on $\vec{e}$, hence also not on $\vec{x}_2$. Hence, $\hat{f}_2(\vec{p}_2, \vec{x}_2 \times \vec{p}_2)$ for $\vec{p} \to 0$ is independent on $\vec{x}_2$ and we write it as $\hat{f}_2(0, 0)$.

For a nontrivial scattering process in the lab system of particle 2, Eq. (13) now reads

$$\begin{aligned} &\hat{f}_1(\vec{p}_1, \vec{x}_1 \times \vec{p}_1) + \hat{f}_2(0, 0) \\ &\quad = \hat{f}_1(\vec{p}_3, \vec{x}_3 \times \vec{p}_3) + \hat{f}_4(\vec{p}_4, \vec{x}_4 \times \vec{p}_4)\,. \end{aligned} \tag{16}$$

Put $\vec{L} = \vec{x} \times \vec{p} = (0, 0, L)$. Perform a small shift of the origin, i.e., $\vec{x}_i \to \vec{x}_i + \delta\vec{a}$ within the plane. Take $\delta\vec{a}$ parallel to $\vec{p}_4$. Then Eq. (16) implies

$$\delta L_1 \frac{\partial}{\partial L} f_1(\vec{p}_1, \vec{L}_1) = \delta L_3 \frac{\partial}{\partial L} f_1(\vec{p}_3, \vec{L}_3)$$

with, due to Eq. (14), $\delta L_1 = \delta L_3 \neq 0$. Hence,

$$\frac{\partial}{\partial L} f_1(\vec{p}_1, \vec{L}_1) = \frac{\partial}{\partial L} f_1(\vec{p}_3, \vec{L}_3) = \text{const} \tag{17}$$

for all $\vec{p}_3, \vec{L}_2$ which for fixed $\vec{p}_1, \vec{L}_1$ may be reached by varying the position $\vec{x}_2$. Since $\vec{x}_2$ may be shifted parallel or orthogonal to $\vec{p}_1, \vec{L}_3$ may be changed independently of $\vec{p}_3$, and Eq. (17) holds in a full neighborhood of outgoing scattering data of particle 3. According to the assumption the outgoing states vary continuously if we vary the initial data. Hence

$$\frac{\partial}{\partial L} f_1(\vec{p}_1, \vec{L}_1) = \text{const}$$

in a neighborhood of $\vec{p}_1, \vec{L}_1$. By continuity, this extends to all possible values of $\vec{p}_1, \vec{L}_1$, and we get

$$f_1 = \vec{b}_1 \vec{L} + h_1(\vec{p})\,.$$

Interchanging the role of particles 1 and 2, we obtain

$$f_2 = \vec{b}_2 \vec{L} + h_2(\vec{p})\,.$$

Insertion into Eq. (13) yields $\vec{b}_1 = \vec{b}_2$, and Lemma 2 completes the proof. □

We now turn to f_i with an explicit dependence on t.

Lemma 4: Let $f(\vec{x}, \vec{p}, t)$ be continuously differentiable and independent of t for all free orbits. Then

$$f(\vec{x}, \vec{p}, t) = g(\vec{s}, \vec{p})\,,$$

where $g(\vec{s}, \vec{p})$ is continuously differentiable on $\mathbb{R}^6$, $\vec{s} = m\vec{x} - \vec{p}t$ or $(\vec{p} + m^2)^{1/2}\vec{x} - \vec{p}t$ in the Galilei- or Lorentz-invariant case.

Proof: The transformation $(\vec{x}, \vec{p}, t) \to (\vec{s}, \vec{p}, t)$ is C^∞ and it has Jacobian $(p^2 + m^2)^{3/2}$ or m^3, respectively. Independence of f on t leads to the asssertion. □

Lemma 5: Consider the Galilei-invariant case. Assume central force interaction, in the case of infinite range such that the scattering angle θ in each lab system decreases at least as const$/\rho^2$ where ρ is the impact parameter and $|\partial\theta/\partial\rho| \leqslant \text{const}/\rho^3$ for $\rho \to \infty$. Assume that $f_i(\vec{s}, \vec{p})$, $i = 1,2$ is four times continuously differentiable, and

$$f_1(\vec{s}_1, \vec{p}_1) + f_2(\vec{s}_2, \vec{p}_2) = f_1(\vec{s}_3, \vec{p}_3) + f_2(\vec{s}_4, \vec{p}_4) \tag{18}$$

on an invariant subset V of the manifold defined by

$$\begin{aligned} &\omega_1(\vec{p}_1) + \omega_2(\vec{p}_2) = \omega_1(\vec{p}_3) + \omega_2(\vec{p}_4)\,, \\ &\vec{p}_1 + \vec{p}_2 = \vec{p}_3 + \vec{p}_4\,, \\ &\vec{x}_1 \times \vec{p}_1 + \vec{x}_2 \times \vec{p}_2 = \vec{x}_3 \times \vec{p}_3 + \vec{x}_4 \times \vec{p}_4\,, \\ &\vec{s}_1 + \vec{s}_2 = \vec{s}_3 + \vec{s}_4\,. \end{aligned} \tag{19}$$

V is assumed to contain a full neighborhood of forward scattering in the lab systems of particles 1 and 2. Here, for every

energy, the scattering angle is assumed to be continuous in the impact parameter. Then

$$f_i = \vec{a}\vec{p} + a^0\omega_i(\vec{p}) + \vec{b}(\vec{x}\times\vec{p}) + \vec{c}\vec{s}_i$$
$$+ d\vec{s}_i\vec{p}/m_i + e\vec{s}_i^{\,2}/m_i + k_i \quad (\vec{s}_i := m_i\vec{x} - \vec{p}t)\,.$$

Proof: We present here only the case of infinite range of the interaction which needs more care and a few additional steps compared to the case of finite range.

(i) Together with $f_i(\vec{s},\vec{p}), f_i(\vec{s}+\vec{p}\tau,\vec{p})$ also fulfills Eq. (18) (shift of origin in time), hence, also

$$\hat{f}_i(\vec{s},\vec{p}) := p_k \frac{\partial}{\partial s_k} f_i(\vec{s},\vec{p})\,,$$
$$\hat{\hat{f}}_i := p_k p_l \frac{\partial}{\partial s_k}\frac{\partial}{\partial s_l} f_i(\vec{s},\vec{p}), \quad \hat{\hat{f}}_i := \cdots$$

as do (no summation convention with respect to i!)

$$m_i \frac{\partial}{\partial s_k} f_i(\vec{s},\vec{p}), \quad m_i \frac{\partial}{\partial p_k} f_i(\vec{s},\vec{p})$$

(by shift in direction k of the origin in space and Galilei boost, respectively).

Consider scattering close to the forward direction in the lab system of particle 2. For a small scattering angle, $\vec{p}_4 = |\vec{p}_1|\cdot\theta$. Choose the origin in time such that

$$\frac{\vec{s}_2}{m_2} - \frac{\vec{s}_1}{m_1} \perp \vec{p}_1\,. \tag{20}$$

Keep $\vec{s}_1, \vec{p}_1, \vec{p}_2 = 0$ and Eq. (20) fixed, vary the impact parameter ρ by varying s_2.

Equation (18) for f_i implies in the limit of forward scattering

$$\lim_{\rho\to\infty} \hat{f}_2(\vec{s}_4,\vec{p}_4) = 0\,.$$

The same holds for $m_i\partial/\partial s_k \cdots m_i\partial\hat{f}_i(\vec{s},\vec{p})/\partial s_1$. Hence, for $\rho\to\infty$,

$$\hat{\hat{f}}_2(\vec{s}_4,\vec{p}_4) \to 0$$

faster than $|\vec{p}_4|^2 \approx (|\vec{p}_1|\cdot\theta)^2 = \text{const}/\rho^4$. The same holds for $\vec{\partial}\hat{\hat{f}}_i/\partial\vec{s}$ and $\vec{\partial}\hat{\hat{f}}_i/\partial\vec{p}$.

Consider now Eq. (18) for $\hat{\hat{f}}_i$. A variation $\vec{s}_2 \to \vec{s}_2 + \delta\vec{s}_2$ [with the restriction Eq. (20)!] gives

$$0 = \left(\delta\vec{p}_3\frac{\vec{\partial}}{\partial\vec{p}} + \delta\vec{s}_3\frac{\vec{\partial}}{\partial\vec{s}}\right)\hat{\hat{f}}_1(\vec{s}_3,\vec{p}_3)$$
$$+ \left(\delta\vec{p}_4\frac{\vec{\partial}}{\partial\vec{p}} + \delta\vec{s}_4\frac{\partial}{\partial\vec{s}}\right)\hat{\hat{f}}_2(\vec{s}_4,\vec{p}_4)$$

on the manifold Eq. (19). In the limit $\rho\to\infty$, the second summand vanishes of higher order in $1/\rho$ (here use is made of $|\theta| \leqslant \text{const}/\rho^2$ and $|\partial\theta/\partial\rho| \leqslant \text{const}/\rho^3$) and it remains that

$$\left(\delta\vec{p}_3\frac{\vec{\partial}}{\partial\vec{p}} + \delta\vec{s}_3\frac{\vec{\partial}}{\partial\vec{s}}\right)\hat{\hat{f}}_1(\vec{s}_1,\vec{p}_1) = 0\,. \tag{21}$$

Energy conservation implies $\delta\vec{p}_3 \perp \vec{p}_3 = \vec{p}_1$ for forward scattering. Conservation of angular momentum implies after short calculation [taking Eq. (20) into account]

$$\delta\vec{s}_3 \times \vec{p}_1 = 0$$

(up to terms of higher order in $1/\rho$ than $|\delta\vec{p}_3| = |\delta\vec{p}_4|$).

Next, consider a scattering process obtained from the first by bringing the position of particle 2 into mirror position with respect to the path $(\vec{s}_1,\vec{p}_1)$. For central force interaction, instead of Eq. (21) one obtains

$$\left(-\delta\vec{p}_3\frac{\vec{\partial}}{\partial\vec{p}} + \delta\vec{s}_3\frac{\partial}{\partial\vec{s}}\right)\hat{\hat{f}}_1(\vec{s}_1,\vec{p}_1) = 0\,. \tag{22}$$

By summing Eqs. (21) and (22),

$$\delta\vec{p}_3\frac{\partial}{\partial\vec{p}}\hat{\hat{f}}_1(\vec{s},\vec{p}) = 0 \tag{23}$$

for $\delta\vec{p}_3 \perp \vec{p}$, and therefore

$$\hat{\hat{f}}_1(\vec{s},\vec{p}) = g_1(\vec{s},\vec{p}^{\,2})\,.$$

Interchanging the role of particles 1 and 2, we get

$$\hat{\hat{f}}_2(\vec{s},\vec{p}) = g_2(\vec{s},\vec{p}^{\,2})\,.$$

In case $\delta\vec{s}_3 \neq 0$, by taking the difference of Eqs. (21) and (22),

$$\delta\vec{s}_3\frac{\vec{\partial}}{\partial\vec{s}}g_1(\vec{s},\vec{p}^{\,2}) = 0$$

for $\delta\vec{s}_3$ and $\vec{p}$ parallel. This implies that g_1 does not depend on $\vec{s}$. In case $\delta\vec{s}_3 = 0$, consider

$$\frac{d}{d\tau}g_i(\vec{s}+\vec{p}\tau,\vec{p}^{\,2})|_{\tau=0} = \vec{p}\frac{\vec{\partial}}{\partial\vec{s}}g_i(\vec{s},\vec{p}^{\,2})\,.$$

They satisfy the same equations as the $\hat{\hat{f}}_i$. Insertion into Eq. (23) leads to

$$\delta\vec{p}_3\frac{\vec{\partial}}{\partial\vec{s}}g_1(\vec{s},\vec{p}^{\,2}) = 0\,,$$

now implying that g_1 is independent of $\vec{s}$. The same follows for g_2.

By Lemma 2,

$$\hat{\hat{f}}_i(\vec{s},\vec{p}) = g_i(p^2) = \alpha p^2/2m_i + \beta_i.$$

Since for $\vec{p}\to 0$, $\hat{\hat{f}}_i$ vanishes at third order, $\alpha = \beta_i = 0$,

$$\hat{\hat{f}}_i(\vec{s},\vec{p}) = 0\,.$$

(ii) Any function of $\vec{s},\vec{p}$ may, for $\vec{p}\neq 0$, be written as a function of $\vec{p}$, $\vec{s}\times\vec{p}/m$, and $\vec{s}\cdot\vec{p}/m$. Putting

$$\hat{\hat{f}}_i(\vec{s},\vec{p}) =: \hat{\hat{h}}_i(\vec{p},\vec{s}\times\vec{p}/m_i,\vec{s}\cdot\vec{p}/m_i)$$

we get from

$$0 = \hat{\hat{f}}_i = \vec{p}\frac{\vec{\partial}}{\partial\vec{s}}\hat{f}_i = \frac{\vec{p}^{\,2}}{m_i}\partial_3\hat{h}_i\left(\vec{p},\frac{\vec{s}\times\vec{p}}{m_i},\frac{\vec{s}\cdot\vec{p}}{m_i}\right)$$

(where ∂_3 denotes the derivative with respect to the third variable) that

$$\hat{\hat{f}}_i = k_i\left(\vec{p},\frac{\vec{s}\times\vec{p}}{m_i}\right),$$

and by Lemma 3

$$\hat{\hat{f}}_i = a_0\vec{p}^{\,2}/2m_i$$

since other possible contributions do not vanish for $\vec{p}\to 0$ of second order as $\hat{f}_i$ does.

(iii) From

$$\alpha\frac{\vec{p}^{\,2}}{2m_i} = \vec{p}\frac{\vec{\partial}}{\partial\vec{s}}f_i(\vec{s},\vec{p}) = \frac{\vec{p}^{\,2}}{m_i}\partial_3\hat{h}_i\left(\vec{p},\frac{\vec{s}\times\vec{p}}{m_i},\frac{\vec{s}\cdot\vec{p}}{m_i}\right)$$

we get

$$\hat{h}_i = \hat{f}_i = \alpha \frac{\vec{s}\cdot\vec{p}}{2m_i} + \gamma_i\left(\vec{p}, \frac{\vec{s}\times\vec{p}}{m_i}\right).$$

As in the proof of Lemma 3 (using a suitable shift of the origin in space), it now follows that γ_i is linear in $\vec{s}\times\vec{p}/m_i$,

$$\hat{f}_i = \alpha \frac{\vec{s}\,\vec{p}}{2m_i} + \vec{b}\frac{\vec{s}\times\vec{p}}{m_i} + \gamma_i(\vec{p}),$$

where $\gamma_i(\vec{p})$ vanishes for $\vec{p}\to 0$.

Because of angular momentum conservation,

$$\check{f}_i := \frac{\alpha}{2}\frac{\vec{s}\cdot\vec{p}}{m_i} + \gamma_i(\vec{p})$$

is a solution of Eqs. (18) and (19) too. Averaging over rotations of $\vec{s}, \vec{p}$ leads to

$$\overline{\check{f}_i} := \frac{\alpha}{2}\frac{\vec{s}\cdot\vec{p}}{m_i} + \bar{\gamma}_i(\vec{p}^{\,2}),$$

which again is a solution. Applying an infinitesimal boost in direction 1, one sees that

$$(\alpha/2)s_l + 2m_i\, p_l \bar{\gamma}_i'(\vec{p}^{\,2})$$

(no summation over i!) solves Eqs. (18) and (19) as does $(\alpha/2)s_l$, and hence $2m_i\, p_l\bar{\gamma}_i'(p^2)$. Application of Lemma 2 to the last gives

$$2m_i\, p_l\bar{\gamma}_i'(p^2) = a_{l0}\frac{p^2}{2m_i} + a_{lk}p_k\,.$$

Choosing $p_l = 0, p_k \neq 0$ for $k \neq l$, one finds that $a_{lk} = \beta\cdot\delta_{lk}, a_{l0} = 0$. Hence $\bar{\gamma}_i' = \beta/2m_i$ and

$$\check{f}_i = \frac{\alpha}{2}\frac{\vec{s}\,\vec{p}}{m_i} + \beta\frac{\vec{p}^{\,2}}{2m_i}$$

(the integration constant vanishes since $\bar{\gamma}(0) = 0$!).

Because of energy conservation, we now see that either $\alpha = 0$, or $\vec{s}\cdot\vec{p}/m_i$ is a solution of Eqs. (18) and (19). Hence $\gamma_i(\vec{p})$ in $\check{f}_i$ is a solution too and Lemma 2 implies

$$\gamma_i(\vec{p}) = \vec{a}\vec{p} + a_0\frac{\vec{p}^{\,2}}{2m_i}$$

and

$$\hat{f}_i = \frac{\alpha}{2}\frac{\vec{s}\,\vec{p}}{m_i} + \vec{a}\vec{p} + a_0\frac{\vec{p}^{\,2}}{2m_i} + \vec{b}\frac{\vec{s}\times\vec{p}}{m_i}.$$

(iv) From

$$\vec{p}\,\frac{\vec{\partial}}{\partial\vec{s}} f_i(\vec{s},\vec{p}) = \frac{\alpha}{2}\frac{\vec{s}\cdot\vec{p}}{m_i} + \vec{a}\vec{p} + a_0\frac{\vec{p}^{\,2}}{2m_i} + \vec{b}\frac{\vec{s}\times\vec{p}}{m_i}$$
$$= \frac{\vec{p}^{\,2}}{m_i}\partial_3 h_i\left(\vec{p}, \frac{\vec{s}\times\vec{p}}{m_i}, \frac{\vec{s}\cdot\vec{p}}{m_i}\right)$$

one gets after division by $\vec{p}^{\,2}/m_i$ an expression for $\partial_3 h_i$ which can be integrated. After reordering the various terms, this may be written as

$$f_i(\vec{s},\vec{p}) = \frac{\alpha}{2}\frac{\vec{s}^{\,2}}{2m_i^2} + \vec{a}\,\vec{s} + \frac{a_0}{2}\frac{\vec{s}\,\vec{p}}{m_i}$$
$$+ \vec{b}\frac{(\vec{s}\times\vec{p})(\vec{s}\cdot\vec{p})}{m_i p^2} + \gamma_i\left(\vec{p}, \frac{\vec{s}\times\vec{p}}{m_i}\right).$$

According to the assumption, f_i is continuous. Hence b must vanish since the discontinuous coefficient cannot be compensated by γ_i which does not depend on $\vec{s}\cdot\vec{p}$.

Together with f_i,

$$\frac{\alpha}{4}\frac{\vec{s}^{\,2}}{m_i} + a_0\frac{\vec{s}\,\vec{p}}{m_i} + \gamma_i\left(\vec{p}, \frac{\vec{s}\times\vec{p}}{m_i}\right)$$

is also a solution. In case $\alpha = 0$, conclude as in (iii). In case $\alpha \neq 0$, we know from (iii) that $\vec{s}\cdot\vec{p}/m_i$ is a solution, hence we need only consider

$$\frac{\alpha}{4}\frac{\vec{s}^{\,2}}{m_i} + \gamma_i\left(\vec{p}, \frac{\vec{s}\times\vec{p}}{m_i}\right).$$

Now again as in the proof of Lemma 3 (using a suitable shift of the origin and taking into account that $\vec{s}$ is a solution!), it follows that

$$\gamma_i\left(\vec{p}, \frac{\vec{s}\times\vec{p}}{m_i}\right) = \gamma_i(\vec{p}) + \vec{b}\frac{\vec{s}\times\vec{p}}{m_i} + c_i\,.$$

Averaging over rotations of $\vec{s}, \vec{p}$ as above leads to

$$\gamma_i(\vec{p}) = \vec{d}\vec{p} + d_0\vec{p}^{\,2}/2m_i + c_i.$$

Collecting the terms, f_i gets the desired form. □

Proof of theorem 2 for a two-particle system in the Galilei invariant case: Lemmas 4 and 5 prove the assertion in the case of four times continuously differentiable $f_i(\vec{x},\vec{p},t)$. We extend it to general f_i as in case of Theorem 1: We define

$$F_\chi^2(\vec{x},\vec{p},t)$$
$$= \int f_i(\vec{x}-\vec{y}, \vec{p}-m_i\vec{v}, t-\tau)\chi(\vec{y},\vec{v},\tau)\,d^3\vec{y}\,d^3\vec{v}\,d\tau$$

$\chi \in \mathcal{D}(\mathbb{R}^7)$, and take advantage of the Galilei invariance of the problem. According to Lemma 5,

$$F_\chi^i = a^0(\chi)\frac{\vec{p}^{\,2}}{2m_i} + \vec{a}(\chi)\vec{p} + \vec{b}(\chi)(\vec{x}\times\vec{p}) + \vec{c}(\chi)\vec{s}_i$$
$$+ d(\chi)\frac{\vec{s}_i\,\vec{p}}{m_i} + e(\chi)\frac{\vec{s}_i^{\,2}}{m_i} + k_i(\chi)\,. \tag{24}$$

As in the proof of Theorem 1, consider a δ-sequence χ_n. Then $F_\chi^i \to f_i$ in $\mathcal{D}'(\mathbb{R}^7)$. To show that $a^0(\chi_n)$,... converge too, smear Eq. (24) with a suitable $\varphi \in \mathcal{D}(\mathbb{R}^7)$ to separate the various terms on the right-hand side. □

Remark: The solutions $\vec{s}_i\cdot\vec{p}/m_i$ and $\vec{s}_i^{\,2}/m_i$ of the equations in addition to the well-known solutions related to the generators of the Galilei group, occur only in cases of very special interactions. Applying an infinitesimal time translation, it is clearly seen that if $\vec{s}_i^{\,2}/m_i$ is a solution then so is $\vec{s}_i\,\vec{p}/m_i$. Concerning the latter, one has the following:

Define for a two-particle system

$$\vec{p}_{\rm rel} := \frac{m_2\vec{p}_1 - m_1\vec{p}_2}{m_1+m_2},$$

$$\vec{s}_{\rm rel} := \frac{m_2\vec{s}_1 - m_1\vec{s}_2}{m_1+m_2} = \mu\vec{x}_{\rm rel} - \vec{p}_{\rm rel}t,$$

$$\mu := \frac{m_1 m_2}{m_1+m_2}.$$

Then $\vec{s}_{\rm rel}\cdot\vec{p}_{\rm rel}/\mu$ is invariant for the scattering processes for a given interaction if and only if $\vec{s}_1\cdot\vec{p}_1/m_1 + \vec{s}_2\cdot\vec{p}_2/m_2$ is invariant for the scattering processes of the same interaction. [This follows easily by computation taking into account that $(\vec{s}_1+\vec{s}_2)(\vec{p}_1+\vec{p}_2)$ is conserved too.]

Consider next the time delay due to scattering in the center-of-mass system. In case of central symmetric potential scattering, this is defined as

$$T := 2 \lim_{R\to\infty} \left\{ \int_\rho^R \frac{dr}{(\vec{p}^{\,2} - \rho^2 \vec{p}^{\,2}/r^2)^{1/2}} - \int_{r_0}^R \frac{dr}{(\vec{p}^{\,2} - 2V(r) - \rho^2\vec{p}^{\,2}/r^2)^{1/2}} \right\}, \tag{25}$$

where $V(r)$ is the potential, r_0 the largest zero of the second radicand, and ρ the impact parameter.

After drawing the free and interacting trajectories of a particle with fixed scattering center, a simple geometric consideration yields that

$$T = \frac{\vec{s}\,\vec{p} - \vec{s}'\vec{p}'}{\vec{p}^{\,2}},$$

where $\vec{s}, \vec{p}, \vec{s}', \vec{p}'$ denote the asymptotic variables before and after scattering. Hence, $\vec{s}_i\,\vec{p}_i/m_i$ is summational invariant if and only if $T = 0$. This, however, under reasonable assumptions, occurs only in case $V(r) = \alpha/r^2$ (see Appendix B). It can be checked that for $V(r) = \alpha/r^2$, $\vec{s}_i^{\,2}/m_i$ is summational invariant too.

For an N-particle system, the situation remains the same. To see that, one observes that $\Sigma_i\,\vec{s}_i\,\vec{p}_i/m_i$ is infinitesimal generator for the dilation defined by

$$(\vec{x}_i, \vec{p}_i, t) \to (\vec{x}_i' = \alpha\vec{x}_i, \vec{p}_i' = (1/\alpha)\vec{p}_i, t' = \alpha^2 t),$$

leaving invariant the action integral

$$\int \sum_i \frac{p_i^2}{2m_i} - \sum_{i\neq j} \frac{\alpha_{ij}}{|\vec{x}_i - \vec{x}_j|^2}\, dt$$

in case of two-particle $1/r^2$-interactions. Similarly for $\Sigma\,\vec{s}_i^{\,2}/m_i$ (together with the 10 generators of the Galilei group, one is led to the 12-parameter Schrödinger group[6]).

Lemma 6: Consider the Lorentz-invariant case. Under the assumptions of Lemma 5, it follows that

$$f_i = \vec{a}\vec{p} + a^0\omega_i(\vec{p}) + \vec{b}\,(\vec{x}\times\vec{p}) + \vec{c}\vec{s}_i + k_i$$

$$(\vec{s}_i := (\vec{p}^{\,2} + m_i^2)^{1/2}\vec{x} - \vec{p}t).$$

Proof: Again, we consider only the case of interaction with infinite range.

(i) Together with $f_i(\vec{s}, \vec{p})$, now also,

$$\hat{f}_i(\vec{s}, \vec{p}) := p_k \frac{\partial}{\partial s_k} f_i(\vec{s}, \vec{p}), \quad \hat{\hat{f}}_i := \cdots, \tag{26}$$

$$(\vec{p}^{\,2} + m_i^2)^{1/2} \frac{\partial}{\partial s_k} f_i(\vec{s}, \vec{p}), \tag{27}$$

$$p^0 \frac{\partial}{\partial p_k} f_i(\vec{s}, \vec{p}) + \epsilon_{kjl} L_j \frac{\partial}{\partial s_l} f_i(\vec{s}, \vec{p}) \tag{28}$$

fulfill Eq. (18) (infinitesimal shift of space–time origin, infinitesimal velocity transformation in direction k, $\vec{L} := \vec{x}\times\vec{p}$).

Consider scattering close to forward scattering in the lab system of particle 2 and choose the time origin, as in the proof of Lemma 5, such that

$$\frac{\vec{s}_2}{m_2} - \frac{\vec{s}_1}{(\vec{p}_1^{\,2} + m_1^2)^{1/2}} \perp \vec{p}_1.$$

Keep $\vec{s}_1, \vec{p}_1, \vec{p}_2 = 0$ fixed, vary the impact parameter ρ by varying $\vec{s}_2$.

As in the proof of Lemma 5, it follows that $\hat{\hat{f}}_2(\vec{s}_4, \vec{p}_4)$ vanishes for $\rho \to \infty$ faster than $(p_4)^2 = \text{const}/\rho^4$. The same follows for $\vec{\partial}\hat{\hat{f}}_2/\partial\vec{s}$, $\vec{\partial}\hat{\hat{f}}_2/\partial\vec{p}$ as can be seen by applying the $\hat{}$ operation to Eqs. (27) and (28).

We can now proceed as in the proof of Lemma 5 [observe, however, that now $\vec{L} = \vec{s}\times\vec{p}/(p^2 + m^2)^{1/2}$] to conclude that

$$\hat{\hat{f}}_i(\vec{s}, \vec{p}) = g_i(\vec{p}^{\,2}) = 0.$$

(ii) Rewriting the f's for $\vec{p} \neq 0$ as functions

$$h\left(\vec{p}, \frac{\vec{s}\times\vec{p}}{(\vec{p}^{\,2} + m^2)^{1/2}}, \frac{\vec{s}\cdot\vec{p}}{(\vec{p}^{\,2} + m^2)^{1/2}}\right)$$

we may argue as in the proof of Lemma 5 to conclude now that

$$\hat{\hat{f}}_i(\vec{s}, \vec{p}) = 0$$

since nonzero contributions allowed by Lemma 3 do not vanish at second order for $\vec{p} \to 0$.

(iii) From

$$0 = \vec{p}\frac{\vec{\partial}}{\partial\vec{s}}\hat{f}_i(\vec{s}, \vec{p}) = \frac{\vec{p}^{\,2}}{(\vec{p}^{\,2} + m_i^2)^{1/2}} \times \partial_3\hat{h}_i\left(\vec{p}, \frac{\vec{s}\times\vec{p}}{(\vec{p}^{\,2} + m_i^2)^{1/2}}, \frac{\vec{s}\cdot\vec{p}}{(\vec{p}^{\,2} + m_i^2)^{1/2}}\right)$$

we get

$$\hat{f}_i = \hat{h}_i = \gamma_i\left(\vec{p}, \frac{\vec{s}\times\vec{p}}{(\vec{p}^{\,2} + m_i^2)^{1/2}}\right) = \vec{a}\vec{p} + \vec{b}\frac{\vec{s}\times\vec{p}}{(\vec{p}^{\,2} + m_i^2)^{1/2}}$$

since other contributions allowed by Lemma 3 do not vanish for $\vec{p} \to 0$.

(iv) From

$$\vec{p}\frac{\vec{\partial}}{\partial\vec{s}} f_i(\vec{s}, \vec{p}) = \vec{a}\vec{p} + \vec{b}\frac{\vec{s}\times\vec{p}}{(\vec{p}^{\,2} + m_i^2)^{1/2}} = \frac{\vec{p}^{\,2}}{(\vec{p}^{\,2} + m_i^2)^{1/2}} \times \partial_3 h_i\left(\vec{p}, \frac{\vec{s}\times\vec{p}}{(\vec{p}^{\,2} + m_i^2)^{1/2}}, \frac{\vec{s}\cdot\vec{p}}{(\vec{p}^{\,2} + m_i^2)^{1/2}}\right)$$

one gets after division by $\vec{p}^{\,2}/(\vec{p}^{\,2} + m_i^2)^{1/2}$ an expression for $\partial_3 h_i$ which can be integrated. After reordering the terms, this may be written as

$$f_i = \vec{a}\,\vec{s} + \vec{b}\frac{\vec{s}\times\vec{p}}{\vec{p}^{\,2}}\frac{\vec{s}\cdot\vec{p}}{(\vec{p}^{\,2} + m_i^2)^{1/2}} + \gamma_i\left(\vec{p}, \frac{\vec{s}\times\vec{p}}{(\vec{p}^{\,2} + m_i^2)^{1/2}}\right).$$

By continuity of f_i, again, $\vec{b} = 0$. Since $\vec{a}\cdot\vec{s}$ is a solution of Eq. (18), Lemma 3 applied to γ_i leads to the final form of f_i. □

Remarks: 1. Observe that the essential difference compared to the Galilei-invariant case is due to $\omega_i(0) = m_i \neq 0$. Clearly, the result is different in case of zero-mass particles.

2. We believe that the assertions of Lemmas 5 and 6 stay true in case of an arbitrary invariant open subset not necessarily containing a neighborhood of forward scattering, but we have not been able to prove it.

Proof of theorem 2 for a two-particle system in the relativistic case: To extend the assertion of Lemma 6 to noncontinuously differentiable functions, we proceed as in the proof of Theorem 1: Consider a Poincaré transformation

$\{a, \Lambda\} \in \mathscr{P}^{\uparrow}_{+}$ and the left and right invariant Haar measure $\mu(\{a, \Lambda\})$ on $\mathscr{P}^{\uparrow}_{+}$. The constraints are Poincaré invariant, hence the $f_i(x, \vec{p})[x := (t, \vec{x})]$, may be replaced by

$$F^j_\chi(x, \vec{p}) = \int f_j(\{a, \Lambda\}^{-1}(x, \vec{p}))\chi(\{a, \Lambda\})d\mu(\{a, \Lambda\}),$$

where $\chi \in \mathscr{D}(\mathscr{P}^{\uparrow}_{+})$. Let e_j and $\Lambda_{j\vec{p}}$ be defined as in the proof of Theorem 1(ii). Then

$$\begin{aligned} f_j(\{a, \Lambda\}^{-1}(x, \vec{p})) &= f_j(\{a, \Lambda\}^{-1}\{x, \Lambda_{j\vec{p}}\}(0, \vec{e}_j)) \\ &=: f_{(0,e_j)}(\{a, \Lambda\}^{-1}\{x, \Lambda_{j\vec{p}}\}) \end{aligned}$$

is locally μ integrable (Appendix A!), and

$$f_{(0,e_j)} * \chi \in C^\infty(\mathscr{P}^{\uparrow}_{+})$$

implying that $F^j_\chi \in C^\infty(\mathbb{R}^7)$. The assertion follows now from Lemma 6 as in the Galilei-invariant case. □

Proof of theorem 2 for N-particle systems: Due to the assumed cluster property, the proof for an N-particle system immediately reduces to the proof of the two-particle subsystems. □

3. CONCLUDING REMARKS

Theorem 2 shows that the situation concerning summational invariants in the mechanics of mass points is very much the same as for elementary particles in the framework of Galilei- or Lorentz-invariant quantum field theory (see Refs. 7 and 8 and the references quoted therein). The proof presented in the last section is rather close to physical details of actual scattering processes; in particular this applies to the first part of the proof of Lemma 5. One should believe that there is a simpler, more general and more abstract proof covering all the different cases simultaneously. On the other hand, the exceptional cases covered by Lemma 5 may be an indication that a more abstract approach is not so simple.

ACKNOWLEDGMENTS

We benefitted from discussions with K. Baumann, P. Breitenlohner, W. D. Garber, and D. Maison. The final version of this paper was completed during a stay of the second named author at the Max Planck Institut für Physik (Werner Heisenberg Institut), Munich. The kind hospitality of the theory group is gratefully acknowledged.

APPENDIX A

The Poincaré group $\mathscr{P}^{\uparrow}_{+}$ operates on $(x, \vec{p}) \in \mathbb{R}^{4+3}$ as follows:

$$(x, \vec{p}) \to \{a, \Lambda\}(x, \vec{p}) = (\Lambda x + a, \overrightarrow{\Lambda p}), \quad p^0 = (\vec{p}^{\,2} + m^2)^{1/2}. \tag{A1}$$

The transformation Eq. (A1) has a nonzero functional determinant for all $(x, \vec{p})$. Define

$$f_{(x,\vec{p})}(a, \Lambda) = f(\Lambda x + a, \Lambda \vec{p}).$$

Lemma A1: $f_{(x,\vec{p})}$ is locally μ integrable for all $x, \vec{p}$.

Proof: $f_{(x,\vec{p})}$ is locally μ integrable iff

$$\int |f(\Lambda x + a, \overrightarrow{\Lambda p})\varphi(\{a, \Lambda\})|\, d\mu(\{a, \Lambda\}) < \infty \tag{A2}$$

$\forall (x, \vec{p}), \forall \varphi \in \mathscr{D}(\mathscr{P}^{\uparrow}_{+})$. To prove Eq. (A2), we apply Fubini's theorem. The function f is (Lebesgue) measurable and we may suppose, as far as integration is concerned, that f is a Borel function. The composition $f_{x,\vec{p}}$ is therefore a Borel function as well, and so is

$$B(a, \Lambda; x, \vec{p}) := f(\Lambda x + a, \overrightarrow{\Lambda p})\psi(x, \vec{p})\varphi(\{a, \Lambda\})$$

for all $\varphi \in \mathscr{D}(\mathscr{P}^{\uparrow}_{+})$ and $\psi \in \mathscr{D}(\mathbb{R}^{4+3})$.

That the iterated integral

$$\begin{aligned} &\int \left\{ \int |B(a, \Lambda; x, \vec{p})|\, d^4x\, d^3\vec{p} \right\} d\mu(\{a, \Lambda\}) \\ &\quad = \int |\varphi(\{a, \Lambda\})| \\ &\qquad \times \left\{ \int f(\Lambda x + a, \overrightarrow{\Lambda p})\psi(x, \vec{p})\, d^4x\, d^3\vec{p} \right\} d\mu(\{a, \Lambda\}) \end{aligned}$$

is finite follows after transformation of variables for the integral in $\{\cdots\}$, observing that the functional determinant does not vanish and that the integral is C^∞ in $\{a, \Lambda\}$. Hence, the other iterated integral exists and, in particular, Eq. (A2) holds for fixed φ for almost all $(x, \vec{p})$.

Consider an arbitrary compact set $K \in \mathscr{P}^{\uparrow}_{+}$. Choose a φ with $\varphi|_{K_\epsilon} = 1$, $K_\epsilon \supset K$.

$$\int_K |f(\{a, \Lambda\}(x, \vec{p}))|\, d\mu(\{a, \Lambda\}) \tag{A3}$$

exists for almost all $(x, \vec{p})$, e.g., for $x_0, \vec{p}_0$. Let $\{\hat{a}, \hat{\Lambda}\} \in U_\epsilon$, a neighborhood of the identity, and adjust K_ϵ such that $K \cdot \{\hat{a}, \hat{\Lambda}\} \subset K_\epsilon$ for all $\{\hat{a}, \hat{\Lambda}\} \in U_\epsilon$. Then

$$\int_K |f(\{a, \Lambda\}\{\hat{a}, \hat{\Lambda}\}(x, \vec{p}))|\, d\mu(\{a, \Lambda\})$$

exists. Hence, Eq. (A3) holds for all $\hat{x}$,$\mathbf{p}$ from a full neighborhood of $(x_0, \vec{p}_0)$. This implies the statement. □

APPENDIX B

Lemma: Consider real $V(r)$, two times differentiable for $r > 0$ with

(i) $\forall\ r_0 > 0\ \exists\ c(r_0) < \infty : \left| \frac{1}{r}\frac{d}{dr}(r^2 V(r)) \right| \leqslant c\ \ \forall r > r_0$;

(ii) $\int_a^\infty \frac{1}{r}\left| \frac{d}{dr}(r^2 V(r)) \right| dr < \infty\ \ \forall a > 0$;

(iii) $\int_a^\infty \left| \frac{d}{dr}\left(\frac{1}{r}\frac{d}{dr}(r^2 V(r)) \right) \right| dr < \infty\ \ \forall a > 0$.

Let T be defined by Eq. (25). Then $T = 0$ for all $\rho > 0$ and sufficiently large p^2 if and only if $V = \text{const}/r^2$.

Remark: (ii) requires a decay faster than $1/r$, (iii) a decay up to $1/r$ at infinity.

Proof: After change of variables,

$$s^2 := r^2 - 2r^2 V(r)/p^2.$$

[The Jacobian does not vanish due to (i) if p^2 is sufficiently large] T can be written as

$$\begin{aligned} T &= \frac{1}{|p|}\int_{\rho^2}^\infty \frac{ds^2}{(s^2 - \rho^2)^{1/2}} \left(1 - 1/(1 - \frac{1}{p^2}\frac{1}{r}\frac{d}{dr}(r^2 V(r))) \right) \\ &= -\frac{2}{|p|^3}\int_{r_0}^\infty \frac{d/dr(r^2 V(r))}{(r^2 - (2r^2 V(r)/p^2) - \rho^2)^{1/2}}\, dr, \end{aligned}$$

proving the if part.

Conversely, $T=0$ for all ρ and sufficiently large p^2 implies

$$0=\int_{r_0(p)}^{\infty}\frac{d/dr(r^2V(r))}{(r^2-(2r^2V/p^2)-\rho^2)^{1/2}}\,dr\,.$$

Hence, in the limit $p^2\to\infty$

$$0=\int_{\rho}^{\infty}\frac{d/dr(r^2V(r))}{(r^2-\rho^2)^{1/2}}\,dr=\int_{\rho^2}^{\infty}\frac{d/dr^2(r^2V(r))}{(r^2-\rho^2)^{1/2}}\,d(r^2)$$
$$=:\beta(\rho^2)\,. \tag{B1}$$

Consider now with $s:=r^2$, $\alpha(s):=d(r^2V(r))/d(r^2)$,

$$\beta(t)=\int_t^{\infty}\frac{\alpha(s)}{(s-t)^{1/2}}\,ds \tag{B2}$$

with

$$\text{(ii')}\quad\int_a^{\infty}\frac{|\alpha(s)|}{(s)^{1/2}}\,ds<\infty\quad\forall a>0$$

and

$$\text{(iii')}\quad\int_a^{\infty}|\alpha'(s)|\,ds<\infty\quad\forall a>0$$

following from (ii) and (iii).

Equation (B2) is Abel's integral equation for $\alpha(s)$. It can be solved as follows:

$$\beta(t)=\int_0^{\infty}\alpha(u+t)\,d\sqrt{u}$$

$$\beta'(t)=\int_0^{\infty}\alpha'(u+t)\,d\sqrt{u}=\int_t^{\infty}\frac{\alpha'(s)}{(s-t)^{1/2}}\,ds$$
$$=\int_0^{\infty}\alpha'(s)\frac{\theta(s-t)}{(s-t)^{1/2}}\,ds\,.$$

$\beta'(t)<\infty$ due to (iii'). Because of (iii') one gets further

$$\int_{-\infty}^{\infty}\beta'(t)\frac{\theta(t-a)}{(t-a)^{1/2}}\,dt$$
$$=\int_0^{\infty}\alpha'(s)\left\{\int_{-\infty}^{\infty}\frac{\theta(s-t)}{(s-t)^{1/2}}\frac{\theta(t-a)}{(t-a)^{1/2}}\,dt\right\}ds\,,$$

with (see, e.g., Ref. 9, pp. 120 and 121)

$$\{\quad\}=(\Gamma(\tfrac{1}{2}))^2\theta(s-a)\,.$$

Hence,

$$\alpha(a)=-\int_a^{\infty}\alpha'(s)\,ds=\int_a^{\infty}\beta'(t)\frac{dt}{(t-a)^{1/2}}\frac{1}{(\Gamma(\frac{1}{2}))^2}\,.$$

Hence, $T=0$ implies $\alpha(a)=0\ \forall a$, i.e., $V=\text{const}/r^2$. □

Remark: There also is an inverse scattering formula giving the potential from the limit of the time delay for large energies for every impact parameter by solving an Abel's integral equation[10]; in the above notation

$$v(a)=-\frac{1}{(\Gamma(\frac{1}{2}))^2}\int_a^{\infty}\frac{\beta(t)}{(t-a)^{1/2}}\,dt,\quad v(r^2):=r^2V(r)$$

following from (B2) under the assumption that $\int_a^{\infty}|v'(s)|\,ds<\infty\ \forall a>0$. This, however, requires a decay of $V(r)$ faster than $1/r^2$ at infinity.

[1]E. H. Kennard, *Kinetic Theory of Gases* (McGraw-Hill, New York, 1939).
[2]H. Grad, Comm. Pure Appl. Math. **2**, 331 (1949).
[3]N. A. Chernikov, Acta Phys. Pol. **26**, 1069 (1969).
[4]J. J. Dijkstra, Proc. K. Ned. Akad. Wet. Ser. B **81**, 265, 277 (1979).
[5]Th. Klinker, Diplomarbeit, Göttingen, 1979.
[6]U. Niederer, Helv. Phys. Acta **45**, 802 (1972).
[7]W. D. Garber, J. Math. Phys. **24**, 1256 (1983).
[8]W. D. Garber and H. Reeh, Commun. Math. Phys. **70**, 169 (1979).
[9]I. M. Gel'fand and G. E. Schilow, *Verallgemeinerte Funktionen (Distributionen)* I (VEB Deutschev Verlag der Wissenschaften, Berlin, 1960), pp. 120 and 121.
[10]K. Chadan and P. C. Sabatier, *Inverse Problems in Quantum Scattering Theory* (Springer-Verlag, New York, 1977), p. 288.

Generalized parastatistics and internal symmetry

R. Y. Levine and Y. Tomozawa
Department of Physics, University of Michigan, Ann Arbor, Michigan 48109

(Received 29 June 1982; accepted for publication 2 January 1983)

We exhibit generalized parastatistical double commutation relations which allow a solution less restrictive than Green's ansatz. This leads to local interaction terms in the Lagrangian consistent with an internal symmetry among Green indices. The generalized commutation relations are covariant under this symmetry transformation rather than invariant. The formulation may be applicable to the generation problem as well as supersymmetric model building.

PACS numbers: 05.30. – d

I. INTRODUCTION

Green[1] generalized the usual quantum statistics by postulating double commutation relations among fields as alternative solutions of Heisenberg's equation of motion. Taking paraboson relations in this discussion, define

$$H = \tfrac{1}{2}\sum_{\mathbf{k}} \omega_{\mathbf{k}} \{a_{\mathbf{k}}, a_{\mathbf{k}}^{\dagger}\}, \tag{1}$$

and require

$$[H, a_{\mathbf{k}}] = -\omega_{\mathbf{k}} a_{\mathbf{k}}. \tag{2}$$

One has generalized conditions[1]

$$\begin{aligned}
&[\{a_{\mathbf{k}}^{\dagger}, a_{\mathbf{k}'}\}, a_{\mathbf{k}''}] = -2\delta_{\mathbf{k}'',\mathbf{k}} a_{\mathbf{k}'},\\
&[\{a_{\mathbf{k}}, a_{\mathbf{k}'}\}, a_{\mathbf{k}''}] = 0,\\
&[\{a_{\mathbf{k}}, a_{\mathbf{k}'}\}, a_{\mathbf{k}''}^{\dagger}] = 2\delta_{\mathbf{k}'',\mathbf{k}'} a_{\mathbf{k}} + 2\delta_{\mathbf{k}',\mathbf{k}''} a_{\mathbf{k}},\\
&[\{a_{\mathbf{k}}^{\dagger}, a_{\mathbf{k}'}^{\dagger}\}, a_{\mathbf{k}''}^{\dagger}] = 0.
\end{aligned} \tag{3}$$

Parafermi relations are similar with all commutators on the lhs of Eqs. (3) and (1). We exhibit these later in Eqs. (31)–(34).

A solution of Eqs. (3) is given by Green's ansatz

$$a_{\mathbf{k}} = \sum_{i=1}^{p} a_{\mathbf{k}}^{(i)}, \tag{4}$$

where the commutation relations among operators with the same Green index are normal boson relations, while those of different Green components are "abnormal,"

$$[a_{\mathbf{k}}^{(i)}, a_{\mathbf{k}'}^{(i)\dagger}] = \delta_{\mathbf{k},\mathbf{k}'}, \quad \text{etc.}, \tag{5a}$$

and

$$\{a_{\mathbf{k}}^{(i)}, a_{\mathbf{k}'}^{(j)}\} = \{a_{\mathbf{k}}^{(i)}, a_{\mathbf{k}'}^{(j)\dagger}\} = 0 \quad (i \neq j). \tag{5b}$$

Greenberg and Messiah[2] showed that the ansatz, Eq. (4), is a unique solution in the case of a unique vacuum state, $|0\rangle$, such that

$$a_{\mathbf{k}}|0\rangle = 0, \tag{6a}$$

and derived

$$a_{\mathbf{k}}^{(i)} a_{\mathbf{k}'}^{(j)\dagger}|0\rangle = \delta_{\mathbf{k},\mathbf{k}'}\delta_{ij}|0\rangle, \tag{6b}$$

which can be written

$$a_{\mathbf{k}} a_{\mathbf{k}'}^{\dagger}|0\rangle = \delta_{\mathbf{k},\mathbf{k}'} p|0\rangle \tag{6c}$$

for integer p.

One may ask the significance of the simple sum in Eq. (4) in light of the fact that Eq. (5) indicates that Green components $a_{\mathbf{k}}^{(i)}$ and $a_{\mathbf{k}}^{(i)\dagger}$ are independent operators. The form of Eq. (4) is due to the postulated double commutators in Eq. (3), and may be unsatisfactory because it results in an undue constraint on the theory. In Sec. II of this article we generalize the double commutator to accommodate mixing of the Green components as independent fields. In Sec. III we consider local interaction terms with the Green index representing an internal symmetry. A discussion follows in Sec. IV.

II. GENERALIZATION OF PARASTATISTICS

Define the Hamiltonian and equations of motion for boson fields $A_{\mathbf{k}}^{(i)}$,

$$H = \tfrac{1}{2}\sum_{\mathbf{k},i} \omega_{\mathbf{k}} \{A_{\mathbf{k}}^{(i)}, A_{\mathbf{k}}^{(i)\dagger}\} \tag{7}$$

and

$$[H, A_{\mathbf{k}}^{(i)}] = -\omega_{\mathbf{k}} A_{\mathbf{k}}^{(i)}, \tag{8}$$

where $i = 1, 2, \ldots, p$. The solution to Eqs. (7) and (8) is given by

$$[\{A_{\mathbf{k}}^{(i)}, A_{\mathbf{k}'}^{(j)\dagger}\}, A_{\mathbf{k}''}^{(l)}] = -2\delta_{\mathbf{k}',\mathbf{k}''} \sum_{n} \gamma_n^{ijl} A_{\mathbf{k}}^{(n)}, \tag{9a}$$

$$\begin{aligned}
[\{A_{\mathbf{k}}^{(i)}, A_{\mathbf{k}'}^{(j)}\}, A_{\mathbf{k}''}^{(l)\dagger}] &= 2\delta_{\mathbf{k}',\mathbf{k}''} \sum_{n} \rho_n^{ijl} A_{\mathbf{k}}^{(n)}\\
&\quad + 2\delta_{\mathbf{k},\mathbf{k}''} \sum_{n} \rho_n^{jil} A_{\mathbf{k}'}^{(n)},
\end{aligned} \tag{9b}$$

$$[\{A_{\mathbf{k}'}^{(i)}, A_{\mathbf{k}'}^{(j)}\}, A_{\mathbf{k}''}^{(l)}] = 0, \tag{9c}$$

with Hermitian conjugation of the above, where γ and ρ are complex numbers. From Eqs. (7) and (8),

$$\sum_{i} \gamma_n^{iil} = \delta_{ln}. \tag{10}$$

The Jacobi identity

$$[\{A,B\},C] + [\{B,C\},A] + [\{C,A\},B] = 0, \tag{11}$$

applied to Eq. (9), gives

$$\rho_n^{ijl} = \gamma_n^{ilj}. \tag{12}$$

We further impose a symmetry relation

$$\rho_n^{ijl} = \rho_n^{jil}, \tag{13}$$

which implies, from Eq. (12),

$$\gamma_n^{ijl} = \gamma_n^{lji}. \tag{14}$$

Similar to Greenberg and Messiah,[2] we derive further

 0022-2488/83/061603-05$02.50 © 1983 American Institute of Physics

constraints due to the existance of a unique vacuum. We have, from Eq. (9),

$$[\{A^{(i)}_{\mathbf{k}},A^{(j)\dagger}_{\mathbf{k}'}\},A^{(l)}_{\mathbf{k}''}A^{(m)\dagger}_{\mathbf{k}'''}]$$
$$= -2\delta_{\mathbf{k}',\mathbf{k}''}\left(\sum_n \gamma_n^{jil}A^{(n)}_{\mathbf{k}}\right)A^{(m)\dagger}_{\mathbf{k}'''}$$
$$+ 2\delta_{\mathbf{k},\mathbf{k}'''}A^{(l)}_{\mathbf{k}''}\left(\sum_n (\gamma_n^{jim})^* A^{(n)\dagger}_{\mathbf{k}'}\right). \tag{15}$$

Assume there exists a unique vacuum, $|0\rangle$. Then it is easy to show that

$$A^{(j)}_{\mathbf{k}}|0\rangle = 0 \tag{16}$$

from Heisenberg's equation of motion, provided $\omega_{\mathbf{k}} \neq 0$ (an exception occurs for zero modes, $\omega_{\mathbf{k}} = 0$, which may lead to symmetry breaking). From Eq. (9) this implies there exists numbers $C^{ij}_{\mathbf{k},\mathbf{k}'}$ such that

$$A^{(i)}_{\mathbf{k}}A^{(j)\dagger}_{\mathbf{k}'}|0\rangle = C^{ij}_{\mathbf{k},\mathbf{k}'}|0\rangle. \tag{17}$$

Taking the vacuum expectation value on both sides of Eq. (17) yields

$$C^{jl}_{\mathbf{k},\mathbf{k}'} = C^{lj*}_{\mathbf{k}',\mathbf{k}}. \tag{18}$$

Operating both sides of Eq. (15) on the vacuum yields

$$(C^{jl}_{\mathbf{k}',\mathbf{k}})^* C^{lm}_{\mathbf{k}'',\mathbf{k}'''} - (C^{ml}_{\mathbf{k}''',\mathbf{k}''})^* C^{ij}_{\mathbf{k},\mathbf{k}'}$$
$$= -2\delta_{\mathbf{k}',\mathbf{k}''}\sum_n \gamma_n^{jil}C^{nm}_{\mathbf{k},\mathbf{k}'''} + 2\delta_{\mathbf{k},\mathbf{k}'''}\sum_n (\gamma_n^{jim})^* C^{ln}_{\mathbf{k}'',\mathbf{k}'}. \tag{19}$$

The lhs of Eq. (19) vanishes due to Eq. (18), and inspection of the rhs of Eq. (19) gives the condition

$$C^{ij}_{\mathbf{k},\mathbf{k}'} = C^{ij}\delta_{\mathbf{k},\mathbf{k}'}, \tag{20}$$

with

$$C^{ij} = (C^{ji})^*, \tag{21}$$

and implies the equation

$$\sum_n C^{nm}\gamma_n^{jil} = \sum_n (\gamma_n^{jim})^* C^{ln}. \tag{22}$$

Up to this point all relations are implied by Eqs. (9), (12), and (16). In order to find an explicit solution to Eq. (22), we rewrite Eq. (17) with Eq. (20),

$$A^{(i)}_{\mathbf{k}}A^{(j)\dagger}_{\mathbf{k}'}|0\rangle = \delta_{\mathbf{k},\mathbf{k}'}C^{ij}|0\rangle. \tag{23}$$

From the assumed uniqueness of the physical vacuum, we observe that it is natural to postulate

$$C^{ij} = \delta^{ij}. \tag{24}$$

First, notice that the uniqueness of the physical vacuum would lead to a less restrictive condition $C_{ij} = C_i\delta_{ij}$ $(C_i > 0)$. But this condition can be reduced to Eq. (24) by normalization of the fields $A^i_k \to \sqrt{C_i}A^i_k$. Secondly, we should point out that $C^{ij} \neq \delta^{ij}$ implies that the vacuum is a superposition of states containing arbitrary numbers of particles. This can be seen by repeated application of the operator $A_iA^\dagger_j$ on the vacuum, $|0\rangle$. From Eqs. (22) and (24) we have

$$\gamma_m^{jil} = (\gamma_l^{jim})^*. \tag{25}$$

In order to find an explicit solution for γ_n^{jil} we take the special case

$$\gamma_n^{jil} = \sum_m \alpha_{im}\beta_{mj}\omega_{lm}\xi_{mn}. \tag{26}$$

Conditions Eqs. (10), (12), (13), (14), and (25) then imply

$$\beta = \alpha^\dagger, \quad \omega = \alpha, \quad \xi = \alpha^\dagger, \tag{27}$$

where α is a unitary matrix. We prove this in the Appendix. Hence, the matrix γ is given by

$$\gamma_n^{jil} = \sum_m \alpha^*_{im}\alpha^*_{jm}\alpha_{lm}\alpha^*_{nm}. \tag{28}$$

Comparison of Eqs. (5), (9a), and (28) implies $A_{\mathbf{k}}$ can be written

$$A^{(i)}_{\mathbf{k}} = \sum_j \alpha_{ij}a^{(j)}_{\mathbf{k}}, \tag{29}$$

where $a^{(j)}_{\mathbf{k}}$ is a Green component satisfying the commutation relations (5). We have shown that Eq. (29) is a unique solution of the double commutators (9), under the vacuum constraints (16) and (24), the symmetry relation (13), and the ansatz (26).

Note that mathematically a more general solution in the form of Eq. (26) exists without the symmetry relation, Eq. (13). Using an argument similar to the proof in the Appendix, Eqs. (9), (10), and (25) imply the nonsingular solution,

$$\gamma_n^{jil} = \sum_m \frac{\alpha_{im}\alpha^\dagger_{mj}\xi_{lm}\xi^\dagger_{mn}}{(\alpha^\dagger\alpha)_{mm}}, \tag{30}$$

where ξ is a unitary matrix and α is an arbitrary nonsingular matrix. Equation (30) results from the fact that Eq. (9a) and the equation

$$[\{A^i_{\mathbf{k}},A^j_{\mathbf{k}'}\},B^l_{\mathbf{k}''}] = -2\delta_{\mathbf{k}',\mathbf{k}''}\sum_n \gamma_n^{jil}B^n_{\mathbf{k}},$$

where

$$A^i_{\mathbf{k}} = \sum_j \frac{\alpha_{ij}}{\sqrt{(\alpha^\dagger a)_{jj}}}a^j_{\mathbf{k}}$$

and

$$B^i_{\mathbf{k}} = \sum_j \xi_{ij}a^j_{\mathbf{k}} \quad (\xi\xi^\dagger = 1)$$

are not differentiated in Heisenberg's equation of motion (8). The physical interpretation of this solution is clear, in that the symmetry of the Hamiltonian is reflected in the commutation relations defining the particles.

The entire argument above can be duplicated for parafermions. We exhibit the key relations:

$$[[B^{(i)}_{\mathbf{k}},B^{(j)\dagger}_{\mathbf{k}'}],B^{(l)}_{\mathbf{k}''}] = 2\delta_{\mathbf{k}',\mathbf{k}''}\sum_n \eta_n^{jil}B^{(n)}_{\mathbf{k}}, \quad \text{etc.,} \tag{31}$$

and

$$\eta_n^{jil} = \sum_m \beta_{im}\beta^*_{jm}\beta_{lm}\beta^*_{nm}, \tag{32}$$

with β unitary. The operator $B^{(i)}_{\mathbf{k}}$ can be represented by

$$B^{(i)}_{\mathbf{k}} = \sum_j \beta_{ij}b^{(j)}_{\mathbf{k}}, \tag{33}$$

where

$$\{b^{(i)}_{\mathbf{k}},b^{(i)\dagger}_{\mathbf{k}'}\} = \delta_{\mathbf{k},\mathbf{k}'}, \tag{34a}$$
$$[b^{(i)}_{\mathbf{k}},b^{(j)\dagger}_{\mathbf{k}'}] = [b^{(i)}_{\mathbf{k}},b^{(j)}_{\mathbf{k}'}] = 0 \quad (i \neq j). \tag{34b}$$

Equations (28) and (29) for parabosons, as well as Eqs.

(32) and (33) for parafermions, imply the covariance of the commutation relations (9) and (31) under unitary transformation of the Green components of the parafields. That is,

$$A'^{(i)}_{\mathbf{k}} = \sum_j \Lambda^{ij} A^{(j)}_{\mathbf{k}} \tag{35}$$

corresponds to Eq. (9) with

$$(\gamma')^{ijl}_n = \sum_{i',j',l',n'} \Lambda_{ii'} \Lambda^{*}_{jj'} \Lambda_{ll'} \Lambda^{*}_{nn'} \gamma^{i'j'l'}_{n'}. \tag{36}$$

A similar equation holds for parafermions. (For this reason, we call Eqs. (9) and (31) *covariant commutation relations*.) It should be noted that Green's parastatistical commutation relations, as well as normal commutation relations, are invariant under unitary transformations.

It should be noted that, without the assumption of Eq. (26), solutions for γ^{ijl}_n exist which are associated with relative single commutations relations more general than Green's ansatz. An example of this is shown in the next section.

III. LOCAL INTERACTIONS

From Eqs. (29) and (33) it is suggested that the Green index can be viewed as an internal symmetry index. One is then led to the question of local interaction terms which satisfy an internal symmetry. We exhibit an example of an interaction obeying an SU(2) symmetry.

Consider paraboson and parafermion fields, ϕ_i and ψ_i ($i = 1,2,3$) (ϕ could be scalar or gauge boson), which are vectors under SU(2) with "i" a Green index as defined in Eqs. (9) and (31). Also, define canonical Green components ϕ^g_i and ψ^g_i ($i = 1,2,3$), which satisfy normal Green commutation relations Eqs. (5) and (34). By an SU(2) symmetry on the Green index we mean that α and β [in Eq. (29) and (33), respectively] are orthogonal matrices. In other words,

$$\psi_i = \sum_j \beta_{ij} \psi^g_j, \quad \phi_i = \sum_j \alpha_{ij} \phi^g_j, \quad i,j = 1,2,3. \tag{37}$$

Assume relative commutation relations between ψ^g_i and ϕ^g_i, ($i = 1,2,3$):

$$[\psi^g_i, \phi^g_i] = 0 \tag{38a}$$

and

$$\{\psi^g_i, \phi^g_j\} = 0 \quad (i \neq j). \tag{38b}$$

Consider the interaction term ($i,j,k = 1,2,3$)

$$\epsilon_{ijk} \bar{\psi}_i \psi_j \phi_k = \epsilon_{ijk} \bar{\psi}^g_i \psi^g_j \phi^g_k = \bar{\psi}^g_1 \psi^g_2 \phi^g_3 - \cdots, \tag{39}$$

where the first equality follows from SU(2) invariance. Locality of the interaction term in Eq. (39) follows due to the fact that the rhs contains Green fields of different indices, and provided relative commutation relations between parafields are of the normal type, Eq. (38). (As a result, ϕ^g_j and ψ^g_j commute with this interaction at spacelike separation.) A Klein transformation,[3]

$$\phi^g_i \to \phi'^g_i = \phi^g_i (-1)^{N_j + N_k} \quad (i,j,k \text{ cyclic}, \psi_j \text{ unchanged}), \tag{40}$$

where N_j is the number operator of the jth quantum number, will change the relative commutation to all commutators:

$$[\psi^g_i, \phi'^g_j] = 0, \quad (\text{all } i,j). \tag{41}$$

This will violate locality of the interaction term in Eq. (39) with ϕ^g_j replaced by ϕ'^g_j; e.g., $\bar{\psi}^g_1 \psi^g_2 \phi'^g_3$ and $\bar{\psi}^g_1 \psi^g_3 \phi'^g_2$ anticommute at spacelike separation instead of commute. This indicates that this Klein transformation has a physical implication in the sense that ψ and ϕ' cannot form a local O(3) Yukawa interaction. Of course, if ψ^g_j in Eq. (40) is altered by a Klein transformation similar to ϕ^g_j in Eq. (40), all commutation relations remain the same and locality of the interactions is maintained. Equations (40) and (41) reveal the relationship between locality of interactions and commutation relations between parafields.

As is mentioned earlier, we note that the fields ϕ_i of the Yukawa interaction in Eq. (39) can be Lorentz scalar or vector fields. The latter case naturally arises in the case of gauge field theory to which our discussion can be applied. In other words, the symmetry between Green indices can be a local symmetry. Also note that a scalar ϕ^4 interaction, which is essential in spontaneously broken gauge theory, is easily made local; for example,

$$\left(\sum_i (\phi^2_i)\right)^2.$$

The above results can be generalized to a lemma: Any group invariant interaction between generalized Green components is local provided suitable relative double commutation relations are chosen among the components. We show this for SO(N) and SU(N) symmetries. In general, two types of identifications of the Green index with an internal symmetry representation are possible; e.g., either spinor or vector representations. If the vector (adjoint) representation ϕ^g_i is chosen for fields satisfying single commutation-anticommutation relations, Eq. (5) [and Eq. (34) for parafermions], the spinor representation ϕ^α_β, where

$$\phi^\alpha_\beta = \frac{1}{\sqrt{2}} \sum_i \sigma^\alpha_{i\beta} \phi^g_i \tag{42}$$

satisfies only double commutation relations. [In Eq. (42) $\{\frac{1}{2}\sigma^\alpha_{i\beta}\}$ are the generators of SU(N).] On the other hand, if the spinor representation ϕ^g_α is assumed to satisfy single commutation–anticommutation relations, one has single commutation–anticommutation relations for $\phi^{g\alpha}_\beta$ by the composite rule:

$$\{\phi^{g\alpha}_\beta, \phi^{g\alpha}_\gamma\} = \{\phi^{g\alpha}_\beta, \phi^{g\alpha\dagger}_\gamma\} = \{\phi^{g\beta}_\alpha, \phi^{g\gamma}_\alpha\} = \{\phi^{g\beta}_\alpha, \phi^{g\gamma\dagger}_\alpha\} = 0 \tag{43}$$

for $\beta \neq \gamma$ and $\beta,\gamma \neq \alpha$; and the rest commute at spacelike separation. In this case, the fields in an adjoint representation

$$\phi_i = \sum_{\alpha,\beta} \sigma^\alpha_{i\beta} \phi^{g\beta}_\alpha \tag{44}$$

satisfy only double commutation relations (9), in which γ^{ijl}_n is more general than the solution in Eq. (26).

The interaction term which is local (the group-invariant term associated with the lemma above), in the case of spinor fields satisfying Eq. (43), is given by

$$\sum_{\alpha\beta} \bar{\psi}^\alpha \psi^\beta \phi^\alpha_\beta. \tag{45}$$

Equation (45) is local because terms with repeated indices

commute with all other fields at spacelike separation. Similarly, in a theory in which only vector and tensor fields exist, the commutation relations in Eq. (38) yield group-invariant local interaction terms. In the case of only the vector (adjoint) representations allowed, one has only the SU(2) group, with local interaction equation (39), or its generalization, $SU(2)\times\cdots\times SU(2)$. This is the only case where the group invariant interaction does not have repeated indices on the fields. Incidently, the uniqueness of the SU(2) group, mentioned above, is a property of group invariance and is independent of the statistics.

Taking the viewpoint that locality of interactions is an essential ingredient in constructing a field theory, one notes that general double commutation relations (9) and (31) may break the internal symmetry between Green components by causing invariant terms to be nonlocal. A mechanism based on this idea may replace the Higg's mechanism for spontaneously breaking internal gauge symmetry with mass and mass-splitting generated dynamically. This concept of dynamical symmetry breaking by a locality condition may also be applied to a global internal symmetry or supersymmetry.

IV. DISCUSSION

The liberation of the Green index as described in Sec. II made possible the assignment of an internal symmetry to that index. Note that the $A^{(i)}$'s do not obey bilinear commutation relations in contrast to normal parastatistics in which physical fields satisfy bilinear commutation relations (Green components). We are departing from the usual point of view in parastatistics that says that only the full parastatistics field has independent status, not the individual components. We also noted that the double commutation relations proposed in the text, Eqs. (9) and (31), are covariant under group transformations, while the canonical commutation relations, as well as ordinary parastatistical double commutation rules, are invariant. In fact, this invariance is the reason for the restricted form of the Green ansatz. This implies the object γ_n^{ijl} in the double commutation relations [Eqs. (9) and (31)] forms a basis for the representations of the symmetry group [see Eq. (36)]. The three conditions on γ_n^{ijl}, Eqs. (10), (13), and (25), are invariant conditions which are analogous to the tracelessness condition on tensor representations of the symmetry group. This condition selects irreducible representations.

It was emphasized that locality of the group-invariant interaction terms is associated with a specific subset of possible double commutators [Eqs. (38) and (43) and group transformations on components in these equations]. Also mentioned was the fact that the spinor rules for commutators [Eq. (43)] and the vector rules [Eq. (38)] resulted in a different formulation of the theory. An interesting question is whether the two formulations lead to different physical results. Note that if canonical commutation relations for the fields are assumed, the vector and spinor forms of the field [related by Eq. (44)] both satisfy canonical commutation relations. This follows due to the invariance of the canonical commutation relations under the internal symmetry group.

It may be possible that this type of symmetry could exist between the three generations of quark and leptons. However, it should be mentioned that the representations of the Higgs scalars for symmetry breaking between Green components are constrained. This is due to the fact that anticommutivity between the scalars of different components (or, more generally, double commutators) does not allow a constant shift in the scalar fields, unless anticommuting c-numbers (Grassmann algebra) are introduced. Therefore, the Higgs in standard spontaneously broken gauge theories would have to be either a singlet, or the adjoint representation; or, more generally, a tensor of even rank, in the Green index symmetry. (For example, the spinor representation is not allowed.) For this reason, symmetry breaking of the horizontal gauge symmetry, which explains mass splittings between generations, is subject to the constraints described above, if the generation symmetry is associated with generalized Green indices. In Sec. III we suggested the basis of a possible alternative symmetry breaking mechanism.

Also, this concept may be applicable to a supersymmetric theory with parastatistics. This has been considered by the authors elsewhere.[4]

ACKNOWLEDGMENT

This work is supported in part by the U. S. Department of Energy.

APPENDIX: REPRESENTATION OF MATRIX γ_n^{ijl}

We want to find the representation for γ_n^{ijl} satisfying conditions (10), (12), (14), and (25):

$$\sum_i \gamma_n^{iil} = \delta_{ln}, \tag{A1}$$

$$\gamma_n^{ijl} = \gamma_n^{lji}, \tag{A2}$$

$$\gamma_n^{ijl} = (\gamma_l^{jin})^*. \tag{A3}$$

In this Appendix all summation of indices are explicitly shown and a double index does not imply summation. Assume the form

$$\gamma_n^{ijl} = \sum_m \alpha_{im}\beta_{mj}w_{lm}\xi_{mn}, \tag{A4}$$

where α,β,w,ξ are nonsingular. This corresponds to a decomposition of an ordinary matrix,

$$\gamma^{ij} = \sum \alpha_{im}\beta_{mj}$$

or, in matrix notation, $\gamma = \alpha\beta$. Equation (A4) is probably the most general form compatible with the symmetry requirements; Eqs. (A1)–(A3).

Multiplying Eq. (A2) by $\beta^{-1}_{jk}\xi^{-1}_{nm}$ and summing over j and n, we obtain

$$\alpha_{im}w_{lm} = \alpha_{lm}w_{im}. \tag{A5}$$

This implies, by nonsingularity of w,

$$w = \alpha E, \tag{A6}$$

where E is a diagonal matrix which can be absorbed into the definition of β or ξ. As a result, without loss of generality, we write

$$w = \alpha. \tag{A7}$$

Multiplying Eq. (A1) by $\alpha_{il}^{-1}\xi_{nj}^{-1}$ and summing over l and n yields

$$(\alpha^{-1}\xi^{-1})_{ij} = \delta_{ij}(\beta\alpha)_{ii}. \tag{A8}$$

Define the diagonal matrix, D,

$$\alpha^{-1}\xi^{-1} \equiv D \equiv (\delta_{ij}d_i), \tag{A9}$$

where

$$d_i \equiv (\beta\alpha)_{ii}. \tag{A10}$$

Note $d_i \neq 0$ because of the nonsingularity of $(\xi\alpha)$. Therefore,

$$\xi = D^{-1}\alpha^{-1}. \tag{A11}$$

Multiplying Eq. (A3) by $(\alpha^{-1})_{sl}(\xi^{-1})_{nq}(\beta^{\dagger})_{ti}^{-1}$ and summing over l, q, and i yields

$$\delta_{sq}(\beta^{\dagger-1})_{ts}\beta_{sj} = (\alpha^{\dagger})_{tj}(\alpha^{-1}\xi^{\dagger})_{st}(\alpha^{\dagger}\xi^{-1})_{tq}. \tag{A12}$$

From δ_{sq} on the lhs of Eq. (A12) one has that $(\alpha^{-1}\xi^{\dagger})$ and $(\alpha^{\dagger}\xi^{-1})$ on the rhs are diagonal matrices, which then implies $((\beta^{\dagger})^{-1}\alpha)$ on the lhs is diagonal. Define

$$\alpha^{-1}\xi^{\dagger} \equiv F_1, \quad \alpha^{\dagger}\xi^{-1} \equiv F_2,$$
$$\text{and} \tag{A13}$$
$$(\beta^{\dagger})^{-1}\alpha \equiv F_3.$$

Substituting Eq. (A13) into Eq. (A12), one obtains

$$\beta = F_1F_2F_3^{-1}\alpha^{\dagger}. \tag{A14}$$

Note that the F_i are not independent and are related by

$$F_2^{\dagger}F_1 = 1 \tag{A15}$$

and

$$F_3^{-1}F_3^{\dagger}F_1F_2 = 1. \tag{A16}$$

Equations (A10) and (A14) imply

$$\alpha^{\dagger}\alpha = F_2D^{-1} = G, \tag{A17}$$

where G is a positive definite diagonal matrix. From Eq. (A14), using Eq. (A17),

$$\beta\alpha = F_1F_2F_3^{-1}F_2D^{-1} \equiv F, \tag{A18}$$

where F is diagonal. Therefore, from Eq. (A10) and the fact that F is diagonal, we have

$$F = D. \tag{A19}$$

In other words,

$$\beta = D\alpha^{-1}. \tag{A20}$$

Recalling Eq. (A8), $\xi = D^{-1}\alpha^{-1}$, we arrive at the expression

$$\gamma_n^{jil} = \sum_m \alpha_{im}\alpha_{mj}^{-1}\alpha_{lm}\alpha_{mn}^{-1}, \tag{A21}$$

where

$$\alpha^{\dagger}\alpha = G. \tag{A22}$$

Define

$$\alpha' = \alpha X, \tag{A23}$$

which results in

$$\alpha'^{\dagger}\alpha' = X^{\dagger}\alpha^{\dagger}\alpha X = X^{\dagger}GX. \tag{A24}$$

Choose

$$X = \sqrt{G}^{-1}, \tag{A25}$$

using the fact that G is a positive definite diagonal matrix, and note that X is diagonal. We arrive at Eq. (27) in the text by substituting

$$\alpha = \alpha' X^{-1} \tag{A26}$$

into Eq. (A21). Notice that X cancels, yielding Eq. (A21) (removing primes) with

$$\alpha^{\dagger}\alpha = 1. \tag{A27}$$

[1] G. Gentile, Nuovo Cimento **17**, 493 (1940); T. Okayama, Progr. Theoret. Phys. **7**, 517 (1952); H. S. Green, Phys. Rev. **90**, 270 (1953); I. E. McCarthy, Proc. Cambridge Philos. Soc. **51**, 131 (1955); S. Kamefuchi and Y. Takahashi, Nucl. Phys. **36**, 177 (1962); L. O'Raifeartaigh and C. Ryan, Proc. Roy. Irish Acad. Sect. A **62**, 93 (1963); S. Kamefuchi, Matscience Rep. **24**, (1963).

[2] O. W. Greenberg and A. Messiah, Phys. Rev. B **138**, 1155 (1965).

[3] O. Klein, J. Phys. Radium **9**, 1 (1938); K. Nishijima, Progr. Theoret. Phys. **5**, 187 (1950); S. Oneda and H. Umezawa, Progr. Theoret. Phys. **9**, 685 (1953); T. Kinoshita, Phys. Rev. **96**, 199 (1954); **110**, 978 (1958); G. Luders, Z. Naturforsch. **139**, 251 (1958); H. Araki, J. Math. Phys. **2**, 267 (1961).

[4] R. Y. Levine and Y. Tomozawa, Preprint UM HE 82-14, University of Michigan, 1982.

Fluctuation properties of the imperfect Bose gas

E. Buffet
Department of Mathematical Physics, University College, Belfield, Dublin, 4, Ireland

J. V. Pulè
Department of Mathematical Physics, University College, Belfield, Dublin, 4, Ireland and Dublin Institute for Advanced Studies, 10, Burlington Road, Dublin, 4, Ireland

(Received 6 May 1982; accepted for publication 18 June 1982)

The fluctuations of the occupation numbers of the different levels of the imperfect Bose gas are computed. They are neither normal, Gaussian, nor independent of the way the infinite volume limit is taken. The fluctuations of the overall density are on the contrary normal, Gaussian, and shape independent.

PACS numbers: 05.30.Jp

1. INTRODUCTION

It has been known for some time that for the free Bose gas the canonical and grand canonical ensembles are not strictly equivalent, although they give rise to the same equation of state. This lack of equivalence manifests itself in the fact that in the two-phase region, the grand canonical distribution of the number density is not asymptotically degenerate (that is, it is not concentrated at one point) but it is on the contrary exponential. This asymptotic distribution, which we shall call the *Kac density*, has its Laplace transform given by (see Refs. 1 and 2)

$$\lim_{V\to\infty} \langle \exp[\lambda (N^V/V)] \rangle_{\bar{\rho}V}^{\text{g.c.}}$$

$$= \begin{cases} \dfrac{e^{\lambda\rho_C}}{1-\lambda(\bar{\rho}-\rho_C)}, & \bar{\rho}>\rho_C, \quad (1a) \\ e^{\lambda\bar{\rho}}, & \bar{\rho}<\rho_C. \quad (1b) \end{cases}$$

As a result, canonical averages need not coincide with grand canonical ones in the two-phase region. In particular the occurrence of Bose–Einstein condensation in the canonical-free Bose gas does not follow from the usual grand canonical treatment. Moreover, it should be clear that in case of a conflict between grand canonical and canonical predictions, only the latter should be relied on. Consequently, we think that a complete description of the behavior of the ground-state occupation number in the canonical ensemble is of interest.

Kac's conjecture was that the lack of equivalence of ensembles was a pathology of the free gas which would disappear in the presence of a repulsive interaction, however small. To test this idea, Davies[3] studied in great detail a mean-field interacting Bose gas previously introduced by Huang[4] and hereafter referred to as the *imperfect Bose gas*. He concluded that indeed the Kac density of the imperfect Bose gas is a delta function even in the two-phase region. The corresponding problem for a general superstable interaction has been recently studied by Lewis and Pulè.[5]

A natural extension of Davies' work consists in investigating in more detail the density fluctuations in the grand canonical imperfect Bose gas, and to see if they are as expected normal and Gaussian in the thermodynamic limit. This question has already been tackled by Wreszinski,[6] but we find that his analysis is not conclusive, even though his result is correct.

Once the question of the fluctuations of the overall density has been settled, one can ask for more detailed information such as the behavior of each energy level occupation number, and especially that of the ground state. We find that to solve this problem for the imperfect Bose gas it is most convenient to first treat the same problem for the canonical-free Bose gas. Our results show that in the two-phase region of the free Bose gas, the canonical fluctuations of the ground-state occupation number are larger than normal and that their distribution is neither Gaussian nor independent of the way the infinite volume limit has been taken (we refer to this last property as a *shape dependence*). Similar results have been obtained in less general situations by Ziff, Uhlenbeck, and Kac.[7]

Finally, we take advantage of the equivalence of ensembles for the imperfect Bose gas to extend the above analysis to this system.

The work is organized as follows. In Sec. 2 we have gathered all our results concerning the free Bose gas in the canonical ensemble. Not only is this study a necessary preliminary to that of the imperfect Bose gas, but we think it is of interest in itself for the reasons discussed at the beginning of this introduction. After a general section where we fix the framework, we give a direct proof of Bose–Einstein condensation in the canonical formalism, followed by a computation of the asymptotic distribution of the ground-state density fluctuations. We next turn to the imperfect Bose gas, whose overall density fluctuations are proved to be normal and Gaussian in Sec. 3. Finally, in Sec. 4 we show how the results of Sec. 2 can be used to demonstrate the existence of Bose–Einstein condensation in the imperfect Bose gas and to obtain the asymptotic distribution of the ground-state density fluctuations for this system.

Throughout this paper we work for definiteness with Dirichlet boundary conditions, but our results can be proved in a similar manner for Neumann or periodic boundary conditions.

 0022-2488/83/061608-09$02.50 © 1983 American Institute of Physics

2. THE FREE BOSE GAS IN THE CANONICAL ENSEMBLE

2.1. General setting and notation

(a) Let Λ^1 be an open bounded connected domain of $\mathbb{R}^3$ with unit volume. We suppose that Λ^1 contains the origin and that its boundary is piecewise continuously differentiable. For each $V>0$ we define Λ^V to be the region of volume V obtained from Λ^1 by isotropic dilation:

$$\Lambda^V = \{x\in\mathbb{R}^3 : V^{-1/3}x\in\Lambda^1\}. \tag{2}$$

The space of one-particle wavefunctions is $\mathscr{H}^V = L^2(\Lambda^V)$ and the one-particle Hamiltonian H^V is the self-adjoint operator on $\mathscr{H}^V$ determined by $-\frac{1}{2}\Delta$ with Dirichlet boundary conditions. Let E_k^V, $k=1,2,3,\cdots$ be the eigenvalues of H^V ordered such that

$$0<E_1^V<E_2^V\leqslant E_3^V\leqslant\cdots. \tag{3}$$

We denote by $\mathscr{H}_n^V = S(\overset{n}{\otimes}\mathscr{H}^V)$ the space of n-particle symmetric wave functions and by H_n^V the Hamiltonian determined by H^V on $\mathscr{H}_n^V$. We also introduce the Fock space on $\mathscr{H}^V$, $\mathscr{F}(\mathscr{H}^V)=\overset{\infty}{\underset{n=0}{\oplus}}\mathscr{H}_n^V$, and the corresponding Hamiltonian $\tilde{H}^V$.

The canonical partition function of the n-particle free Bose gas is

$$Z_n^V = \mathrm{Tr}_{\mathscr{H}_n^V}\exp(-\beta H_n^V). \tag{4}$$

The grand canonical partition function at chemical potential μ ($\mu<E_1^V$) is

$$Q_\mu^V = \mathrm{Tr}_{\mathscr{F}(\mathscr{H}^V)}\exp[-\beta(\tilde{H}^V-\mu N^V)], \tag{5}$$

with N^V the number of particles operator.

The operator for the number of particles in the k th level is denoted by N_k^V.

(b) For the purpose of studying fluctuations it is useful to think of N^V, N_k^V as random variables (notice their commutative nature) rather than operators. One can then define the probability that, in the canonical ensemble with $n=\rho V$ total number of particles, the occupation number of the k th energy level is xV:

$$\begin{aligned}\mathbb{P}_k^V(x|\rho) &= \mathrm{Prob}\{N_k^V/V=x\,|\,N^V/V=\rho\}\\ &= \langle\mathscr{P}_k^V(xV)\rangle_{\rho V}^{\mathrm{can.}}.\end{aligned} \tag{6}$$

In (6) we have introduced $\mathscr{P}_k^V(m)$, the orthogonal projection onto the subspace of $\mathscr{H}_{\rho V}^V$ such that exactly m particles are in k th energy state.

We now obtain a more explicit form for (6).

From the identity

$$\begin{aligned}&\mathrm{Tr}_{\mathscr{F}(\mathscr{H}^V)}\exp[-\beta(\tilde{H}^V-\mu N^V-\sigma N_k^V)]\\ &\quad=\sum_{n=0}^{\infty}z^n\sum_{l=0}^{n}\xi^l\mathrm{Tr}_{\mathscr{H}_n^V}[\mathscr{P}_k^V(l)\exp(-\beta H_n^V)],\end{aligned} \tag{7}$$

where $z=e^{\beta\mu}$ and $\xi=e^{\beta\sigma}$, we see that, up to the normalization factor $Z_{\rho V}^V$, the coefficient of $z^{\rho V}\xi^{xV}$ in (7) is precisely $\mathbb{P}_k^V(x|\rho)$. On the other hand, since in the grand canonical ensemble N_k^V is geometrically distributed, we have an alternative expression for the left-hand side of (7):

$$\begin{aligned}&\frac{1-z\exp(-\beta E_k^V)}{1-z\xi\exp(-\beta E_k^V)}Q_\mu^V\\ &\quad=[1-z\exp(-\beta E_k^V)]\sum_{m=0}^{\infty}z^m\xi^m\\ &\qquad\times\exp(-m\beta E_k^V)\sum_{p=0}^{\infty}z^pZ_p^V.\end{aligned} \tag{8}$$

Identifying the coefficients of $z^r\xi^s$ in the right-hand sides of (7) and (8) we obtain, finally,

$$\mathbb{P}_k^V(x|\rho)=\begin{cases}0, & x>\rho,\\ \dfrac{\exp(-\rho V\beta E_k^V)}{Z_{\rho V}^V}, & x=\rho,\\ \dfrac{\exp(-xV\beta E_k^V)}{Z_{\rho V}^V}[Z_{(\rho-x)V}^V-\exp(-\beta E_k^V)Z_{(\rho-x)V-1}^V], & 0\leqslant x\leqslant\rho-1/V.\end{cases} \tag{9}$$

Our objective in the next two sections is an asymptotic study of $\mathbb{P}_k^V(dx|\rho)$ in the limit $V\to\infty$. This can be done most conveniently through its Laplace transform; from (9) it follows immediately that

$$\langle\exp(\lambda N_k^V)\rangle_{\rho V}^{\mathrm{can.}}=\frac{e^{-\lambda}Z_{\rho V}^V+\exp[\rho V(\lambda-\beta E_k^V)](1-e^{-\lambda})\Sigma_{m=0}^{\rho V}\exp[-m(\lambda-\beta E_k^V)]Z_m^V}{Z_{\rho V}^V}. \tag{10}$$

Formula (10) is the main tool of our subsequent study of the canonical free Bose gas.

(c) It is known that convergence of Laplace transforms of probability distributions imply convergence of the distributions themselves, at least on intervals of continuity (i.e., intervals which do not have as one of their end points an atom of the limit probability measure) (see Ref. 8, Theorem 2, Chap. XIII. 1 and criterion 2 Chap. VIII.1). If, moreover, the sequence of absolute moments of a given order is bounded, then the sequence of moments of any lower order converges to the corresponding moment of the limit distribution (see Ref. 9, corrollary in Sec. 11.4).

We end this section with the definition of several objects which are basic for the description of the free Bose gas.

(d) Let

$$F_V(\eta) = (1/V)\max\{k:\eta_k^V \leqslant \eta\}, \tag{11}$$

with $\eta_k^V = E_k^V - E_1^V$, i.e., $VF_V(\eta - E_1^V)$ is the *distribution of the eigenvalues* of H^V.

A classical result of Weyl gives the asymptotic behavior of $F_1(\eta)$ (cf. Ref. 10 for example):

$$\lim_{\eta\to\infty} \eta^{-3/2}F_1(\eta) = \sqrt{2}/3\pi^2. \tag{12}$$

Notice that by virtue of the scaling relation

$$F_V(\eta) = (1/V)F_1(\eta V^{2/3}) \tag{13}$$

the property (12) is equivalent to

$$\lim_{V\to\infty} F_V(\eta) = (\sqrt{2}/3\pi^2)\eta^{3/2} \equiv F(\eta). \tag{14}$$

One can also deduce from (12) that the following inequality holds uniformly in V:

$$0 \leqslant F_V(\eta) - 1/V < C\eta^{3/2}. \tag{15}$$

(e) Let

$$\tilde{\rho}_C^V = \int_{0+}^{\infty} \frac{F_V(d\eta)}{e^{\beta\eta} - 1} \tag{16}$$

and let

$$\rho_C = (2\pi\beta)^{-3/2} g_{3/2}(1), \tag{17}$$

where the function $g_{3/2}$ is a member of the following family:

$$g_\alpha(\zeta) = \sum_{n=1}^{\infty} \frac{\zeta^n}{n^\alpha}. \tag{18}$$

Then a standard result of the usual analysis of the Bose-Einstein condensation gives[2]:

$$\lim_{V\to\infty} \tilde{\rho}_C^V = \rho_C. \tag{19}$$

(f) Define finally, for $\sigma < 0$,

$$p_V(\sigma) = -\frac{1}{\beta}\int_{0-}^{\infty} \ln(1 - e^{\beta(\sigma-\eta)})F_V(d\eta). \tag{20}$$

The function $p_V(\sigma)$ is nothing but the grand canonical pressure of the free Bose gas at chemical potential $\sigma + E_1^V$:

$$p_V(\sigma) = (1/\beta V)\ln Q_{\sigma+E_1^V}^V. \tag{21}$$

2.2. Bose–Einstein condensation in the canonical ensemble

In this section, we obtain the asymptotic distribution of the ground-state occupation number for the canonical-free Bose gas. That its first moment is the same as in the grand canonical ensemble, hence showing Bose–Einstein condensation, was proved by Fraser[11] and can also be deduced by combining the works of Cannon[12] and Lewis and Pulè.[2] Our result shows that higher moments are different in the two ensembles: we find that the distribution of the random variable N_1^V/V converges in the infinite volume limit to the degenerate distribution $\delta_{\max(0,\,\rho-\rho_C)}$. Equivalently, in the Laplace transformed version we prove the following:

Theorem 1: For the free Bose gas, the following limit holds when $\lambda \geqslant 0$:

$$\lim_{V\to\infty} \langle \exp[\lambda N_1^V/V]\rangle_{\rho V}^{\text{can.}} = \begin{cases} \exp[\lambda(\rho - \rho_C)], & \rho > \rho_C, \quad (22a) \\ 1, & \rho < \rho_C, \quad (22b) \end{cases}$$

where ρ_C is defined in (17). The convergence also holds for every moment.

Proof: Define

$$\tilde{p}_V(\sigma) = -\frac{1}{\beta V}\sum_{k=2}^{\infty} \ln\{1 - \exp(\beta\sigma) \times \exp[-\beta(E_k^V - E_1^V)]\} \tag{23}$$

$$= -\frac{1}{\beta}\int_{0+}^{\infty} \ln[1 - \exp(\beta\sigma)\exp(-\beta\eta)]F_V(d\eta). \tag{24}$$

Define also

$$K_V(x) = \begin{cases} Z_r^V \exp[-\beta(V\tilde{p}_V(0) - rE_1^V)], & r/V < x \leqslant (r+1)/V, \quad r\in\mathbb{N}, \\ 0, & x \leqslant 0. \end{cases} \tag{25}$$

Using (10) and (25), it is easy to check that

$$\langle \exp(\lambda N_1^V/V)\rangle_{\rho V}^{\text{can.}} = \frac{e^{\lambda\rho}\int_{0-}^{\rho+1/V} + e^{-\lambda x}K_V(dx)}{K_V(\rho + 1/V)}. \tag{26}$$

On the other hand, consider

$$\int_{0-}^{\infty} e^{-\lambda x}K_V(dx) = \lambda\int_0^{\infty} e^{-\lambda x}K_V(x)\,dx. \tag{27}$$

With (25) the right-hand side of (27) reads

$$(1 - e^{-\lambda/V})e^{-V\beta\tilde{p}_V(0)}\sum_{r=0}^{\infty} \exp[-r(\lambda/V - \beta E_1^V)]Z_r^V, \tag{28}$$

and can be written by (21) as

$$\exp\{\beta V[p_V(-\lambda/\beta V) - \tilde{p}_V(0)] + \ln(1 - e^{-\lambda/V})\}. \tag{29}$$

The exponent in (29) can be evaluated by making use of the explicit formulas (20) and (24) for p_V and $\tilde{p}_V$. The result is

$$-V\int_{0+}^{\infty}\ln\left(1-\frac{e^{-\lambda/V}-1}{e^{\beta\eta}-1}\right)F_V(d\eta) \tag{30}$$

$$=-\lambda(1+O(1/V))\tilde{\rho}_C^V. \tag{31}$$

Since $\tilde{\rho}_C^V$ tends to ρ_C [see (19)] we have thus shown

$$\lim_{V\to\infty}\int_{0-}^{\infty}e^{-\lambda x}K_V(dx)=e^{-\lambda\rho_C}=\int_{0-}^{\infty}e^{-\lambda x}\delta_{\rho_C}(dx). \tag{32}$$

This implies the convergence of the measures themselves,

$$\lim_{V\to\infty}K_V(dx)=\delta_{\rho_C}(dx), \tag{33}$$

in the sense that (33) holds for the measure of any open bounded interval (a,b) with $a,b\neq\rho_C$ [see Sec. 2.1.(c)] Equivalently, with $\chi_{(a,b)}$ the indicator function, we have for the distribution functions,

$$\lim_{V\to\infty}K_V(x)=\chi_{(\rho_C,\infty)}(x),\quad x\neq\rho_C. \tag{34}$$

The result (22a) follows directly from (33), (34), and (26). Since all the measures involved have their support on the positive half of the real line, boundedness of the moments follows immediately from the above, and hence their convergence.

The result (22b) is a special case of our Theorem 3. ■

We conclude this section with two comments.

(i) One can show along the lines of the above proof that on the intensive scale every excited-state occupation number is asymptotically degenerately distributed at 0:

$$\lim_{V\to\infty}\langle\exp(\lambda N_k^V/V)\rangle_{\rho V}^{\text{can.}}=1\quad\forall\rho,\ \forall k>1. \tag{35}$$

(ii) In order to illustrate the lack of equivalence of ensembles in the two-phase region of the free Bose gas, we give below the grand canonical equivalent of (22). As usual we mean by $\langle\cdot\rangle_{\bar{\rho}V}^{\text{g.c.}}$ the grand canonical average at a value $\mu_V(\bar{\rho})$ of the chemical potential such that the mean density is $\bar{\rho}$. The computation is much easier than the canonical one, due to the fact that in the grand canonical ensemble N_1^V is geometrically distributed. The result is that N_1^V/V is asymptotically exponentially distributed: when $\lambda<1/(\bar{\rho}-\rho_C)$,

$$\lim_{V\to\infty}\langle\exp(\lambda N_1^V/V)\rangle_{\bar{\rho}V}^{\text{g.c.}}$$

$$=\begin{cases}\dfrac{1}{1-\lambda(\bar{\rho}-\rho_C)}, & \bar{\rho}>\rho_C, \qquad (36a)\\ 1, & \bar{\rho}<\rho_C. \qquad (36b)\end{cases}$$

Notice that, although (22) and (36) are completely different in the two-phase region, they give rise, however, to identical first moments. All the other moments differ. Alternatively, (36) can be deduced from (22) with the use of the Kac density (1).

2.3. Canonical fluctuations of the occupation numbers

The first problem one comes across when investigating the fluctuations of some extensive observable A^V in the limit $V\to\infty$ is this of finding the relevant scale; namely the value of α (if any) such that $V^{\alpha}((A^V-\langle A^V\rangle_V)/V)$ attains asymptotically a nondegenerate finite distribution.

The second problem is to compute explicitly this distribution, and to study its possible shape dependence.

In noncritical situations, it is generally believed[13] that $\alpha=\frac{1}{2}$ and a shape-independent Gaussian distribution provide the answer to both problems.

In contrast to that supposedly universal behavior, very little is known on both problems (even at a heuristic level) in the critical case.

For the ground-state occupation number of the free Bose gas, the solution is contained in formula (36) for the *grand canonical* ensemble and in the next theorem for the *canonical* ensemble.

Theorem 2: Consider the free Bose gas. Put

$$\left\langle\frac{N_1^V}{V}\right\rangle_{\rho V}^{\text{can.}}=\rho_1^V. \tag{37}$$

Then the following limit holds for $\lambda\geqslant 0$:

$$\lim_{V\to\infty}\langle\exp[\lambda V^{1/3}(N_1^V/V-\rho_1^V)]\rangle_{\rho V}^{\text{can.}}$$

$$=\begin{cases}e^{g(\lambda/\beta)}, & \rho>\rho_C, \qquad (38a)\\ 1, & \rho<\rho_C, \qquad (38b)\end{cases}$$

where

$$g(u)=\int_{0+}^{\infty}\left[\frac{u}{x}-\ln\left(1+\frac{u}{x}\right)\right]F_1(dx), \tag{39}$$

with F_1 as defined in (11). The convergence also holds for every moment.

Proof: The proof begins very much as in Theorem 1. Define, with $\tilde{\rho}_C^V$ and $\tilde{p}_V$ as in (16) and (23):

$$L_V(x)=\begin{cases}Z_r^V\exp[-\beta(V\tilde{p}_V(0)-rE_1^V)], & V^{1/3}(r/V-\tilde{\rho}_C^V)<x\leqslant V^{1/3}((r+1)/V-\tilde{\rho}_C^V),\quad r\in\mathbb{N},\\ 0, & x\leqslant -V^{1/3}\tilde{\rho}_C^V.\end{cases} \tag{40}$$

This gives, using (10),

$$\langle\exp[\lambda V^{1/3}(N_1^V/V-\rho+\tilde{\rho}_C^V)]\rangle_{\rho V}^{\text{can.}}=\int_{-\infty}^{\alpha_V}e^{-\lambda x}L_V(dx)/K_V(\rho+1/V), \tag{41}$$

with K_V as in (25) and

$$\alpha_V = V^{1/3}(\rho - \tilde{\rho}_C^V) + V^{-2/3}. \tag{42}$$

On the other hand,

$$\ln \int_{-\infty}^{\infty} e^{-\lambda x} L_V(dx) = \beta V \left[p_V\left(-\frac{\lambda V^{-2/3}}{\beta} \right) - \tilde{p}_V(0) \right] + \lambda V^{1/3} \tilde{\rho}_C^V + \ln(1 - e^{-\lambda V^{-2/3}}) \tag{43}$$

$$= V \int_{0+}^{\infty} \left[-\ln \frac{e^{\beta\eta} - e^{-\lambda V^{-2/3}}}{e^{\beta\eta} - 1} + \frac{\lambda V^{-2/3}}{e^{\beta\eta} - 1} \right] F_V(d\eta) \tag{44}$$

$$= \int_{0+}^{\infty} \left[-\ln \frac{e^{\beta x V^{-2/3}} - e^{-\lambda V^{-2/3}}}{e^{\beta x V^{-2/3}} - 1} + \frac{\lambda V^{-2/3}}{e^{\beta x V^{-2/3}} - 1} \right] F_1(dx). \tag{45}$$

It can be shown that the integrand in (45) satisfies for some constant $d > 0$ the uniform bound

$$|[\cdots]| \leqslant d\lambda^2 / x^2. \tag{46}$$

This bound, together with (12) allows for the application of the Lebesgue dominated convergence theorem which yields

$$\lim_{V \to \infty} \int_{-\infty}^{\infty} e^{-\lambda x} L_V(dx) = e^{g(\lambda/\beta)}, \tag{47}$$

with g as defined in (39).

Now the estimate

$$0 \leqslant \int_{\alpha_V}^{\infty} e^{-\lambda x} L_V(dx)$$

$$\leqslant e^{-(\lambda/2)\alpha_V} \int_{-\infty}^{\infty} e^{-(\lambda/2)x} L_V(dx) \tag{48}$$

in conjunction with (41), (42), and (47) gives for $\rho > \rho_C$

$$\lim_{V \to \infty} \langle \exp[\lambda V^{1/3}(N_1^V/V - \rho + \tilde{\rho}_C^V)] \rangle_{\rho V}^{\text{can.}} = e^{g(\lambda/\beta)}. \tag{49}$$

To complete the proof it is enough to notice that a variant of the bound (46) gives the boundedness (and thus the convergence) of the moments:

$$\langle [V^{1/3}(N_1^V/V - \rho + \tilde{\rho}_C^V)]^m \rangle_{\rho V}^{\text{can.}}. \tag{50}$$

In particular, since $g'(0) = 0$ we have

$$\lim_{V \to \infty} V^{1/3}(\rho_1^V - \rho + \tilde{\rho}_C^V) = 0, \tag{51}$$

which allows us to center the random variable N_1^V so that (49) and (51) imply the announced result (38a) in the two-phase region.

For a proof of (38b) we refer to our Theorem 3 where a more general result is obtained. ■

Several comments are in order here:

(i) From (38a) and (39) we see that in the two-phase region the fluctuations of the ground-state occupation number are larger than normal (the relevant scaling factor is of the order of the square root of the surface area of Λ^V rather than the square root of the volume) and that their distribution is not Gaussian.

Moreover, we stress that they are *shape dependent* because they involve the distribution F_1 of the eigenvalues of $-\frac{1}{2}\Delta$ *on the unit volume* Λ^1. More specifically, the second and third moments are given by ($\rho > \rho_C$):

$$\lim_{V \to \infty} \langle [V^{1/3}(N_1^V/V - \rho_1^V)]^2 \rangle_{\rho V}^{\text{can.}}$$

$$= \frac{1}{\beta^2} \int_{0+}^{\infty} \frac{F_1(dx)}{x^2}$$

$$= \frac{1}{\beta^2} \sum_{k=2}^{\infty} \frac{1}{(E_k^1 - E_1^1)^2}, \tag{52}$$

$$\lim_{V \to \infty} \langle [V^{1/3}(N_1^V/V - \rho_1^V)]^3 \rangle_{\rho V}^{\text{can.}}$$

$$= \frac{-2}{\beta^3} \int_{0+}^{\infty} \frac{F_1(dx)}{x^3} = \frac{-2}{\beta^3} \sum_{k=2}^{\infty} \frac{1}{(E_k^1 - E_1^1)^3}. \tag{53}$$

Notice that the odd moments do not vanish, so that the distribution is not symmetric. It can be checked that the distribution is infinitely divisible.

(ii) The function $e^{g(\lambda/\beta)}$ is decreasing in β and goes to 1 as β goes to infinity. In that sense, the fluctuations are increasing with the temperature at a given density $\rho > \rho_C$. This is to be contrasted with the temperature-independent behavior shown by the grand canonical fluctuations [see (36a)].

(iii) Comparison between (38a) and (36a) shows that the canonical fluctuations of N_1^V are much smaller than the grand canonical ones.

(iv) A similar study can be carried out for the excited levels. One gets the following result: For any $k > 1$ and $\lambda < \beta(E_k^1 - E_1^1)$,

$$\lim_{V \to \infty} \langle \exp(\lambda V^{1/3} N_k^V / V) \rangle_{\rho V}^{\text{can.}}$$

$$= \begin{cases} \dfrac{1}{1 - \lambda/\beta(E_k^1 - E_1^1)}, & \rho > \rho_C, \quad (54a) \\ 1, & \rho < \rho_C. \quad (54b) \end{cases}$$

It is an easy calculation to check that this coincides with the grand canonical prediction. Notice that the shape dependence shown by (54a) is only through $E_k^1 - E_1^1$, whereas for the ground-state density fluctuations the whole spectrum of $-\frac{1}{2}\Delta$ on Λ^1 is involved [see (38a), (39), (52), and (53)].

We finally settle the question of fluctuations in the one-phase region. The following result shows that for all occupation numbers (including the ground state) these fluctuations are abnormally small.

Theorem 3: For any $\alpha < 1, \lambda \geqslant 0, k \in \mathbb{N}$ the following limit holds for the free Bose gas when $\rho < \rho_C$:

$$\lim_{V \to \infty} \langle \exp(\lambda V^{\alpha} N_k^V / V) \rangle_{\rho V}^{\text{can.}} = 1. \tag{55}$$

Proof: Define the canonical-free energy density at volume V and for n particles:

$$q_V(n/V) = -(1/\beta V)\ln Z_n^V. \tag{56}$$

Extend this definition to any density by linear interpolation. It can be shown (see Ref. 3) that the resulting free energy density of the free Bose gas is convex (we emphasize that this holds already for *finite volume*). Hence when $x > y$

$$q_V(x) - q_V(y) \leqslant (x-y) q'_{V-}(x), \tag{57}$$

with q'_{V-} being the left-hand derivative of q_V.

Using (56) and (57), after an easy resummation (10) gives

$$\left|\langle \exp(\lambda V^{\alpha} N_k^V/V)\rangle_{\rho V}^{\text{can.}} - \exp(-\lambda V^{\alpha-1})\right| \leqslant \left| [1-\exp(-\lambda V^{\alpha-1})] \frac{e^{\rho V \gamma_V} - e^{-\gamma_V}}{1-e^{-\gamma_V}} \right|. \tag{58}$$

In (58) we have set

$$\gamma_V = \lambda V^{\alpha-1} - \beta E_k^V + \beta q'_{V-}(\rho). \tag{59}$$

Now, because each function q_V is convex, we have when $\rho < \rho_C$ and $\alpha < 1$ [with $q(x) = \lim_{V\to\infty} q_V(x)$]:

$$\lim_{V\to\infty} \gamma_V = +\beta q'(\rho) < 0. \tag{60}$$

Hence,

$$\lim_{V\to\infty} \left| \frac{e^{\rho V \gamma_V} - e^{-\gamma_V}}{1-e^{-\gamma_V}} \right| = \left| \frac{1}{1-e^{+\beta q'(\rho)}} \right| < \infty, \tag{61}$$

and the right-hand side of (58) goes to 0. ■

That $\alpha = 1$ is actually the relevant scale, i.e., that $\langle \exp(\lambda N_k^V)\rangle$ attains a finite limit different from 1, can be seen most easily in the grand canonical ensemble; it must then also hold in the canonical ensemble, since in the regime $\bar{\rho} < \rho_C$ the Kac density is a delta function [see (1b)].

3. THE STATISTICS OF THE GRAND CANONICAL NUMBER DENSITY FOR THE IMPERFECT GAS

As the free Bose gas is a highly idealized model, it is of evident interest to know if the results (1), (22), and (38) enjoy some stability property with respect to the introduction of an interaction. Since the treatment of a realistic interacting Bose gas is a major unsolved problem, we shall content ourselves with a mean-field-type interaction, which is certainly a reasonable first step. We thus consider the *imperfect Bose gas* defined by its Hamiltonian

$$\tilde{H}^V + \frac{a}{2} V \left(\frac{N^V}{V}\right)^2, \tag{62}$$

where $\tilde{H}^V$ is the free Hamiltonian and the parameter a is a positive constant.

It is known that the mean-field interaction drastically modifies the density distribution (1) (see Ref. 3 for a study of this particular model, or Ref. 5 for a more general discussion). We shall see in Sec. 4 that on the contrary the individual energy level results (22) and (38) carry over to the imperfect gas.

In this section, after having recalled general results on the imperfect Bose gas, we concentrate on its density fluctuations. Our main result is that they are normal and Gaussian for all values of the chemical potential. In this sense the introduction of a mean-field interaction suppresses the pathological density fluctuations of the free Bose gas (1).

The grand canonical ensemble for this model in the region Λ^V defines a probability measure $\mathbb{P}_\mu^V$ for each value of the chemical potential μ in $\mathbb{R}$, which assigns to the random variable N^V/V describing the density of particles in the region Λ^V the probability distribution

$$\mathbb{P}_\mu^V\left\{\frac{N^V}{V} = \frac{n}{V}\right\} = \frac{Z_n^V e^{\beta(\mu n - (a/2)n^2/V)}}{\tilde{Q}_\mu^V},$$

where $\tilde{Q}_\mu^V$ is the grand canonical partition function for the imperfect gas,

$$\tilde{Q}_\mu^V = \sum_{n=0}^{\infty} Z_n^V e^{\beta(\mu n - (a/2)n^2/V)},$$

with Z_n^V as in (4).

The grand canonical pressure for this model

$$p(\mu) = \lim_{V\to\infty} (1/\beta V)\ln \tilde{Q}_\mu^V$$

is given by

$$p(\mu) = \begin{cases} \dfrac{\mu^2}{2a} + \dfrac{1}{\beta(2\pi\beta)^{3/2}} g_{5/2}(1), & \mu > \mu_C \equiv a\rho_C, \\ \dfrac{a}{2}(x(\mu))^2 + \dfrac{1}{\beta(2\pi\beta)^{3/2}} g_{5/2}(e^{\beta(\mu - ax(\mu))}), & \mu \leqslant \mu_C, \end{cases}$$

where $x(\mu)$ is the unique solution of

$$x = (2\pi\beta)^{-3/2} g_{3/2}(e^{\beta(\mu - ax)}).$$

Here $\rho_C, g_{3/2}, g_{5/2}$ are as defined in (17) and (18). The density $\bar{\rho}_V(\mu) = \langle N^V/V\rangle_\mu^{\text{imp.,g.c.}}$ therefore converges to $\bar{\rho}(\mu) = dp(\mu)/d\mu$ which is given by

$$\bar{\rho}(\mu) = \begin{cases} \mu/a, & \mu > \mu_C, \\ x(\mu), & \mu \leqslant \mu_C. \end{cases}$$

Since $\mu \to \bar{\rho}(\mu)$ is continuous $\mathbb{P}_\mu^V$ converges for each μ to the probability measure $\delta_{\bar{\rho}(\mu)}$.[5] We therefore look at the distribution of the scaled random variable $V^{1/2}(N^V/V - \bar{\rho}_V(\mu))$ with corresponding probability measure $\tilde{\mathbb{P}}_\mu^V$ related to $\mathbb{P}_\mu^V$ by $\tilde{\mathbb{P}}_\mu^V(A) = \mathbb{P}_\mu^V(\bar{\rho}_V(\mu) + V^{1/2}A)$. The behavior of $\tilde{\mathbb{P}}_\mu^V$ for large V is given by the next theorem, but first we prove the following lemma for the corresponding measure for the free gas, which we denote by $\tilde{\mathbb{P}}_{0,\mu}^V$.

Lemma: Suppose $\mu_V \to \mu < 0$; then in the free case, that is for $a = 0$, $\tilde{\mathbb{P}}_{0,\mu_V}^V$ converges to a Gaussian measure centered at 0 with variance $\sigma_0^2(\mu)$ given by

$$\sigma_0^2(\mu) = (2\pi\beta)^{-3/2} g_{1/2}(e^{\beta\mu}).$$

Proof: It is sufficient to consider the Laplace transform of $\tilde{\mathbb{P}}_{0,\mu_V}^V$,

$$\phi_V^0(\lambda) = \langle \exp[-\lambda V^{1/2}(N^V/V - \bar{\rho}_V^0(\mu_V))]\rangle_{\mu_V}^{\text{free,g.c.}}$$

with $\lambda > 0$. A simple computation gives

$$\ln \phi_V^0(\lambda) = \lambda V^{1/2}\bar{\rho}_V^0(\mu_V) + \beta V [p_V(\mu_V - E_1^V - \lambda/\beta V^{1/2}) - p_V(\mu_V - E_1^V)],$$

where $\bar{\rho}_V^0(\mu) = \langle N^V/V\rangle_\mu^{\text{free,g.c.}}$.

Therefore,

$$\ln \phi_V^0(\lambda)$$

$$= -V\int_{0-}^{\infty}\left\{\ln\frac{1-\exp[\beta(\mu_V - E_1^V - \eta) - \lambda V^{-1/2}]}{1-\exp[\beta(\mu_V - E_1^V - \eta)]} - \frac{\lambda V^{-1/2}}{\exp[-\beta(\mu_V - E_1^V - \eta)]-1}\right\}F_V(d\eta),$$

which tends to

$$\frac{\lambda^2}{2}\int_0^{\infty}\frac{e^{\beta(\mu-\eta)}}{[1-e^{\beta(\mu-\eta)}]^2}F(d\eta),$$

and this is equal to $(\lambda^2/2)(2\pi\beta)^{-3/2}g_{1/2}(e^{\beta\mu})$ (see Lemma 2.3 in Ref. 3). ■

Theorem 4: $\tilde{\mathbb{P}}_\mu^V$ converges to a Gaussian measure centered at 0 with variance $\sigma^2(\mu)$ given by

$$\sigma^2(\mu) = \begin{cases}(\beta a)^{-1}, & \mu > \mu_C, \quad (63a)\\ \left(\beta a + \dfrac{(2\pi\beta)^{3/2}}{g_{1/2}(e^{\beta[\mu - a\bar\rho(\mu)]})}\right)^{-1}, & \mu < \mu_C. \quad (63b)\end{cases}$$

In each case the moments also converge to moments of the Gaussian measure.

Proof: We consider first the case $\mu < \mu_C$. Let $x_V(\mu)$ be the unique solution of $x = \bar\rho_V^0(\mu - ax)$. Then $x_V(\mu) \to \bar\rho(\mu)$ as $V \to \infty$ and if $\mu_V = \mu - ax_V(\mu)$,

$$\mu_V \to \mu - a\bar\rho(\mu) < 0.$$

As before let
$\phi_V(\lambda) = \langle \exp[-\lambda V^{1/2}(N^V/V - \rho_V(\mu))]\rangle_\mu^{\text{imp.,g.c.}}$ with $\lambda > 0$; then

$$\phi_V(\lambda)\exp[-\lambda V^{1/2}(\bar\rho_V(\mu) - x_V(\mu))]$$

$$= \frac{\Sigma_{n=0}^{\infty}\exp[-\lambda V^{1/2}(n/V - x_V(\mu))]Z_n^V\exp[\beta(\mu n - (a/2)n^2/V]}{\Sigma_{n=0}^{\infty}Z_n^V\exp[\beta(\mu n - (a/2)n^2/V]}$$

$$= \frac{\Sigma_{n=0}^{\infty}\exp[-\lambda V^{1/2}(n/V - x_V(\mu))]Z_n^V\exp\{\beta[(\mu - ax_V(\mu))n - (a/2)V(n/V - x_V(\mu))^2]\}}{\Sigma_{n=0}^{\infty}Z_n^V\exp\{\beta[(\mu - ax_V(\mu))n - (a/2)V(n/V - x_V(\mu))^2]\}}$$

$$= \frac{\Sigma_{n=0}^{\infty}\exp[-\lambda V^{1/2}(n/V - \bar\rho_V^0(\mu_V))]\exp[-\beta(a/2)V(n/V - \bar\rho_V^0(\mu_V))^2]Z_n^V\exp(\beta\mu_V n)}{\Sigma_{n=0}^{\infty}\exp[-\beta(a/2)V(n/V - \bar\rho_V^0(\mu_V))^2]Z_n^V\exp(\beta\mu_V n)}$$

$$= \frac{\int_{-\infty}^{\infty}\exp(-\lambda x)\exp[-\beta(a/2)x^2]\tilde{\mathbb{P}}_{0,\mu_V}^V(dx)}{\int_{-\infty}^{\infty}\exp[-\beta(a/2)x^2]\tilde{\mathbb{P}}_{0,\mu_V}^V(dx)},$$

which converges to

$$\exp\left\{\frac{\lambda^2}{2}\left[\beta a + \frac{(2\pi\beta)^{3/2}}{g_{1/2}(e^{\beta[\mu - a\bar\rho(\mu)]})}\right]^{-1}\right\}.$$

We now turn to the case $\mu > \mu_C$. This time we consider

$$\phi_V(\lambda)\exp[-\lambda V^{1/2}(\bar\rho_V(\mu) - \bar\rho(\mu))]$$

$$= \frac{\Sigma_{n=0}^{\infty}\exp[(-\lambda V^{1/2}(n/V - \bar\rho(\mu))]Z_n^V\exp[-\beta(a/2)V(n/V - \bar\rho(\mu))^2]}{\Sigma_{n=0}^{\infty}Z_n^V\exp[-\beta(a/2)V(n/V - \bar\rho(\mu))^2]},$$

since in this case $\bar\rho(\mu) = \mu/a$. Define the distribution function M_V by

$$M_V(x) = \frac{1}{V^{1/2}}\sum_{n\in I_V(x)}Z_n^V\exp[-\beta(V\bar p_V(0) - nE_1^V)]$$

$$= \frac{1}{V^{1/2}}\sum_{n\in I_V(x)}K_V\left(\frac{n+1}{V}\right),$$

where

$$I_V(x) = \{n : n\in\mathbb{N}, \bar\rho(\mu)V \leqslant n \leqslant \bar\rho(\mu)V + xV^{1/2}\}.$$

Then $M_V(x)$ tends to x as $V\to\infty$ because, since $K_V((n+1)/V) \leqslant 1$,

$$\frac{1}{V^{1/2}}\sum_{n\in I_V(x)}\left|K_V\left(\frac{n+1}{V}\right) - 1\right|$$

$$= \frac{1}{V^{1/2}}\sum_{n\in I_V(x)}1 - K_V\left(\frac{n+1}{V}\right)$$

$$\leqslant x(1 - K_V(\bar\rho(\mu)))$$

because $K_V(n/V)$ is increasing with n [see (25)]. Now

$$\frac{1}{V^{1/2}}\sum_{n>\bar\rho(\mu)V}\exp[-\lambda V^{1/2}(n/V - \bar\rho(\mu))]Z_n^V \times\exp[-\beta(a/2)V(n/V - \bar\rho(\mu))^2] \times\exp[-\beta(\tilde p_V(0)V - \bar\rho(\mu)E_1^V)]$$

$$= \int_{0-}^{\infty}\exp[-x(\lambda + \beta E_1^V V^{1/2})] \times\exp[-\beta(a/2)x^2]M_V(dx).$$

Since $M_V(x) \leqslant x$ and $V^{1/2}E_1^V$ converges to 0, by the Lebesgue dominated convergence theorem this last integral converges to

$$\int_0^{\infty}e^{-\lambda x}e^{-\beta(a/2)x^2}\,dx.$$

The summation over n with $n < \bar\rho(\mu)V$ is treated similarly. In both cases all the moments are bounded and therefore the

convergence of the moments and centering of the measures proceed as in the other theorems. ■

It is easy to see from the proof of the foregoing theorem that the following corollary is true:

Corollary: If $t\in\mathbb{R}$ and $\alpha<\frac{1}{2}$ then the measure $e^{tV^{\alpha}(\rho-\bar{\rho}_V(\mu))}\mathbb{P}_\mu^V(d\rho)$ converges also to $\delta_{\bar{\rho}(\mu)}(d\rho)$.

Remarks:

(i) As a by-product of Theorem 4 we have a new proof of Davies' result[3] to the effect that in the grand canonical imperfect Bose gas the overall density is asymptotically degenerately distributed for all values of the chemical potential:

$$\lim_{V\to\infty}\langle e^{\lambda N^V/V}\rangle_\mu^{\text{imp.,g.c.}}=e^{\lambda\bar{\rho}(\mu)}. \tag{64}$$

(ii) Notice that if one lets the interaction parameter a tend to 0 in (62) the variance (63a) in the two-phase region blows up. However (64) is independent of a as long as $a>0$. Consequently, the free case (1) cannot be recovered from the imperfect one by letting a go to 0 after the thermodynamic limit.

4. THE STATISTICS OF THE OCCUPATION NUMBERS IN THE IMPERFECT BOSE GAS

In this section we continue our study of the imperfect Bose gas; having calculated in Sec. 3 the fluctuations of the overall density, we now turn to the occupation densities of each energy level.

The first remark is that *in the canonical ensemble* all the results of Sec. 2 can be immediately transferred from the free gas to the imperfect gas. Indeed in the canonical ensemble the random variable N^V is fixed at ρV and consequently the mean-field interaction term $(a/2)V(N^V/V)^2$ is only a constant which does not affect any expectation value:

$$\langle A_V\rangle_{\rho V}^{\text{imp.,can.}}=\langle A_V\rangle_{\rho V}^{\text{free,can.}}. \tag{65}$$

Thus Theorems 1 and 2 hold also for the canonical imperfect gas. Moreover, there are good reasons to believe that the same results will cover the more interesting case of the imperfect gas in the *grand canonical ensemble*. This is clear at a formal level from

$$\langle A_V\rangle_\mu^{\text{imp.,g.c.}}=\int_0^\infty\langle A_V\rangle_{\rho V}^{\text{free,can.}}\mathbb{P}_\mu^V(d\rho) \tag{66}$$

and from the convergence of $\mathbb{P}_\mu^V(d\rho)$ to $\delta_{\bar{\rho}(\mu)}(d\rho)$.

As a first application of this scheme, we show that the imperfect Bose gas in the grand canonical ensemble exhibits the Bose–Einstein condensation phenomenon, with macroscopic occupation of the ground state. This result is not new, but we think that our approach is simpler, both conceptually and technically, than the one available so far.[14,15]

Theorem 5: For the imperfect Bose gas (with interaction parameter $a>0$) the following limit holds when $\lambda\geqslant 0$:

$$\lim_{V\to\infty}\langle\exp(\lambda N_1^V/V)\rangle_\mu^{\text{imp.,g.c.}}$$

$$=\begin{cases}\exp[\lambda(\bar{\rho}(\mu)-\rho_C)], & \mu>\mu_C \quad (67a)\\ 1, & \mu<\mu_C. \quad (67b)\end{cases}$$

The convergence also holds for every moment.

We do not give a separate proof of Theorem 5, as it is a straightforward consequence of our next theorem.

We now broaden the scale to get the fluctuations of the ground-state density.

Theorem 6: For the imperfect Bose gas (with interaction parameter $a>0$) the following limit holds when $\lambda\geqslant 0$:

$$\lim_{V\to\infty}\langle\exp[\lambda V^{1/3}(N_1^V/V-m_1^V(\mu))]\rangle_\mu^{\text{imp.,g.c.}}$$

$$=\begin{cases}e^{g(\lambda/\beta)}, & \mu>\mu_C\\ 1, & \mu<\mu_C,\end{cases} \tag{68}$$

with

$$m_1^V(\mu)=\langle N_1^V/V\rangle_\mu^{\text{imp.,g.c.}} \tag{69}$$

and $g(\lambda/\beta)$ as in Theorem 2.

The convergence also holds for every moment.

Proof: Put

$$h_V(\rho)=\langle\exp(\lambda V^{1/3}N_1^V/V)\rangle_{\rho V}^{\text{free,can.}}. \tag{70}$$

Since we shall be working with λ fixed, we omit the explicit dependence of h_V on λ.

We have

$$\langle\exp[\lambda V^{1/3}(N_1^V/V-\bar{\rho}_V(\mu)+\tilde{\rho}_C^V)]\rangle_\mu^{\text{imp.,g.c.}}$$

$$=\int_0^\infty h_V(\rho)\exp[-\lambda V^{1/3}(\bar{\rho}_V(\mu)-\tilde{\rho}_C^V)]\mathbb{P}_\mu^V(d\rho) \tag{71}$$

with $\tilde{\rho}_C^V$ as in (16); our aim is to show that (71) converges to $e^{g(\lambda/\beta)}$ as $V\to\infty$ when $\bar{\rho}>\rho_C$.

We know by Theorem 2 that for $\rho>\rho_C$

$$\lim_{V\to\infty}h_V(\rho)\exp[-\lambda V^{1/3}(\rho-\tilde{\rho}_C^V)]=e^{g(\lambda/\beta)}. \tag{72}$$

Moreover this limit holds uniformly in ρ on compact intervals in (ρ_C,∞). This can be extracted from the proof of the Theorem 2 using the fact that in (41) both the numerator and the denominator are increasing functions of ρ.

Keeping this in mind, we split the integral (71) into three parts, ρ running, respectively, in $[0,\bar{\rho}(\mu)-\delta)$, $[\bar{\rho}(\mu)-\delta,\bar{\rho}(\mu)+\delta]$, $(\bar{\rho}(\mu)+\delta,\infty)$ for some $\delta>0$.

On the first interval we have, since $\bar{\rho}(\mu)>\rho_C$ and $\tilde{\rho}_C^V\to\rho_C$:

$$\tilde{\rho}_C^V<K(\bar{\rho}_V(\mu)-\rho) \tag{73}$$

for some constant $K>0$, and V large enough.

This together with

$$h_V(\rho)\leqslant e^{\lambda V^{1/3}\rho} \tag{74}$$

yields the estimate

$$\int_0^{\bar{\rho}(\mu)-\delta}h_V(\rho)\exp[-\lambda V^{1/3}(\bar{\rho}_V(\mu)-\tilde{\rho}_C^V)]\mathbb{P}_\mu^V(d\rho)$$

$$\leqslant\int_0^{\bar{\rho}(\mu)-\delta}\exp[\lambda V^{1/3}(1-K)(\rho-\bar{\rho}_V(\mu))]\mathbb{P}_\mu^V(d\rho). \tag{75}$$

The right-hand side of (75) converges to 0 as $V\to\infty$ by the corollary in Sec. 3.

We dispose in the same way of the third integral.

For the remaining part, we have

$$\int_{\bar{\rho}(\mu)-\delta}^{\bar{\rho}(\mu)+\delta}h_V(\rho)\exp[-\lambda V^{1/3}(\bar{\rho}_V(\mu)-\tilde{\rho}_C^V)]\mathbb{P}_\mu^V(d\rho)$$

$$=e^{g(\lambda/\beta)}\int_{\bar{\rho}(\mu)-\delta}^{\bar{\rho}(\mu)+\delta}\exp[-\lambda V^{1/3}(\bar{\rho}_V(\mu)-\rho)]\mathbb{P}_\mu^V(d\rho)$$

$$+\int_{\bar{\rho}(\mu)-\delta}^{\bar{\rho}(\mu)+\delta}\{h_V(\rho)\exp[-\lambda V^{1/3}(\rho-\tilde{\rho}_C^V)]$$
$$-\exp[g(\lambda/\beta)]\}$$
$$\times\exp[-\lambda V^{1/3}(\bar{\rho}_V(\mu)-\rho)]\mathbb{P}_\mu^V(d\rho).$$

Using again the corollary and the fact that the limit in (72) holds uniformly, the conclusion follows. The convergence of moments and the replacement of $\bar{\rho}_V(\mu)-\tilde{\rho}_C^V$ by $m_1^V(\mu)$ are dealt with as in Theorem 2. ■

Comments:

(i) The remarks of Sec. 2.3 on shape dependence apply equally here, as do the explicit formulas (52) and (53) for the second and third moment.

(ii) One can also show that the higher level occupation numbers of the imperfect gas are asymptotically exponentially distributed exactly as in formula (54). The proof is straightforward.

(iii) Notice that all the results of this paragraph are independent of the value of the interaction parameter a, *provided that* $a>0$. In particular the free case *cannot* be recovered from the imperfect one by letting a go to 0 after the thermodynamic limit. The singular character of the limit $a\to 0$ for the infinite gas is already apparent in the asymptotic form of the overall density distribution [see Remark (ii) in Sec. 3].

ACKNOWLEDGMENT

We are grateful to Professor J. T. Lewis for his interesting suggestions and remarks.

[1]M. Kac, unpublished notes (1971).
[2]J. T. Lewis and J. V. Pulè, Commun. Math. Phys. **36**, 1 (1974).
[3]E. B. Davies, Commun. Math. Phys. **28**, 69 (1972).
[4]K. Huang, *Statistical Mechanics* (Wiley, New York, 1963).
[5]J. T. Lewis and J. V. Pulè, "On the statistics of the number density operator in the grand canonical ensemble" (to appear).
[6]W. F. Wreszinski, Helv. Phys. Acta. **46**, 844 (1974).
[7]R. M. Ziff, G. E. Uhlenbeck, and M. Kac, Phys. Rep. **32**, 169 (1977).
[8]W. Feller, *An Introduction to Probability Theory and its Applications*, Vol. II (Wiley, New York, 1966).
[9]M. Loève, *Probability Theory*, Vol. I (Springer-Verlag, New York, 1977).
[10]M. Reed and B. Simon, *Methods of Modern Mathematical Physics*, Vol. IV (Academic, New York, 1978).
[11]A. R. Fraser, Philos. Mag. **42**, 165 (1951).
[12]J. T. Cannon, Commun. Math. Phys. **29**, 89 (1973).
[13]Whenever this scale applies, A^V is said to have *normal fluctuations*.
[14]M. Fannes and A. Verbeure, Phys. Lett. A **76**, 31 (1980).
[15]M. Fannes and A. Verbeure, J. Math. Phys. **21**, 1809 (1980).

Cycle statistics of steady states for master-equation systems

Teruaki Nakagomi
Department of Physics, Kyoto University, Kyoto, Japan

(Received 29 June 1982; accepted for publication 26 November 1982)

A systematic graph theoretical treatment of steady-state master-equation systems is presented, in order to give a general formulation to Hill's theory of cycle fluxes which characterizes the cyclic-action structure of open systems in nonequilibrium steady states accompanied by nonzero steady flows of energy or materials between the systems and their environments, such as laser actions or enzymatic reactions. In our formalism, a steady state is described by an ensemble of cycles C's with each of which a probability current $\omega(C)$ and an occurrence probability $u(C)$ are associated. A steady flow is expressed as an average over the cycles with $\omega(C)$, i.e., $\Sigma_C\, \omega(C)\times$(elementary flow generated in C). A similar cycle-average formula by $u(C)$ is also given to state quantities. The Kirchhoff formula for steady states is also referred to.

PACS numbers: 05.40. + j, 02.10. + w, 02.30.Hq

I. INTRODUCTION

In many cases, open systems have nonequilibrium steady states, in which steady flows of energy, materials, and so on are maintained between the systems and their environments. Microscopically or on the level of elementary processes, such flows can be described as being generated by cyclic actions of the systems which do not cause any change in the macroscopic or average steady states. Typical examples are found in lasing systems or enzymes catalyzing chemical reactions.

Hill[1–4] has characterized such cyclic-action structure of steady-state open systems on the basis of master-equation description. Consider a master equation,

$$\frac{d}{dt}p_t(a) = \sum_{b\in\mathscr{S}} p_t(b)\,\gamma(b\to a) - \sum_{b\in\mathscr{S}} p_t(a)\,\gamma(a\to b), \tag{1.1}$$

where $\mathscr{S}$ is a set of element-states, p_t is a probability distribution over $\mathscr{S}$ at time t and identified with a macroscopic state, and $\gamma(a\to b)$ is the transition probability rate (tpr) from $a\in\mathscr{S}$ to $b\in\mathscr{S}$. The probability current $j(a\to b)$ from a to b is defined by $j(a\to b) = p(a)\,\gamma(a\to b)$. Hill has shown illustratively that in the steady state $p = p_{\rm st}$ it is written in the form

$$j(a\to b) = \sum_C \omega(C), \tag{1.2}$$

where the summation is over all cycles C's (cyclic sequences of transitions without repetition) including the transition $a\to b$, and $\omega(C)$ is a positive quantity defined diagrammatically for each cycle C. Similarly, the net probability current from a to b, $J(a\to b) = j(a\to b) - j(b\to a)$, is expressed as

$$J(a\to b) = \sum_C [\omega(C) - \omega(\bar{C})], \tag{1.3}$$

where the summation is the same as in (1.2), and $\bar{C}$ is the reversed cycle of C.

In Hill's terminology, the quantity $j(a\to b)$ is called the transition flux along $a\to b$, and $\omega(C)$ the (one-way) cycle flux around C, with the corresponding net quantities $J(a\to b)$ and $\omega(C) - \omega(\bar{C})$. Hill and Chen[3] have further discussed the cycle flux $\omega(C)$ as a frequency of cycle completion in relation to Monte Carlo simulations of the master equation. Hill's cycle theory is reviewed in his book[4] with interesting applications.

In the present paper, the above-mentioned cycle theory is refined and enlarged with attention to the following three points.

The first point is to set up our argument on a mathematical basis, which is necessary for establishing the cycle theory in a general form. A special formulation of graph theory is used in this paper. Iwahori's formulation[5,6] serves as a good reference, which deals with the relation between graphs and stochastic (or transition) matrices. Kohler and Vollmerhaus[7] also use graph theory to discuss the meaning of $\omega(C)$ in a different context from ours.

The second is concerned with the multiplicity of transitions in the following sense. An open system exhibiting a nonequilibrium steady state usually has several transitions with different physical properties and origines between a pair of element states. For example, in a many-reservoir open system with element states $a,b,c,\cdots$, the reservoirs can cause independently the same transition $a\to b$. Such transitions must be distinguished (written, e.g., $a\xrightarrow{1}b$, $a\xrightarrow{2}b,\cdots$) and be separately given the tpr's in order to define real flows as mentioned below, although they need not be separated to specify the master equation. The multiplicity of transitions is reflected in the graphs by the multiplicity of arcs joining the same pair of vertices, which affects various graph theoretical definitions including the cycles.

The third point is concerned with the definition of real flows. The quantity $j(a\to b)$ or $J(a\to b)$ does not correspond to any of real flows of energy, molecules, and others between a system and its environment, but represents the flow of probability in the abstract space of element states. So we call it a probability current in order to distinguish it from the real flows. On a microscopic level, a flow is specified by the elementary flows $f(a\xrightarrow{i}b)$ generated in individual transitions $a\xrightarrow{i}b$. If the flow is observed on a macroscopic level in a state p, it will be given by the average of $f(a\xrightarrow{i}b)$ with the statistical weight equal to the probability current as follows:

$$\sum p(a)\,\gamma(a\xrightarrow{i}b)\,f(a\xrightarrow{i}b), \tag{1.4}$$

where the summation is over all transitions $a\xrightarrow{i}b$. If $a\xrightarrow{i}b$ indicates the transition caused by the ith reservoir for each i

 0022-2488/83/061617-08$02.50 © 1983 American Institute of Physics

and $f(\cdot)$ means the energy flow from a specified reservoir i_0 to the system, then $f(\cdot)$ is given by

$$f(a \xrightarrow{i} b) = \begin{cases} \epsilon(b) - \epsilon(a) & \text{if } i = i_0 \\ 0 & \text{otherwise} \end{cases} \tag{1.5}$$

with the energy function ϵ of the system. This definition (1.4) of flows allows nonzero steady flows.

In our theory, roughly speaking, the quantity (1.4) in the steady state p_{st} is reexpressed by an average over cycles as follows:

$$\sum_C \omega(C) f^\circ(C), \tag{1.6}$$

where $f^\circ(C)$ is the elementary flow generated in the cycle C and the quantity $\omega(C)$ corresponds to that in (1.1). Formula (1.6) shows that the steady state can be regarded as an ensemble of cycles with the statistical weight $\omega(C)$. Moreover, it is shown that a similar cycle-average formula holds for state quantities. So we propose to call our formalism "cycle statistic of steady states."

In order to make our discussion definite, we introduce a scheme of transition systems (Sec. II), which gives a mathematical representation to the above-mentioned master-equation systems with multiplicity of transitions. On the basis of this scheme, we present graph theory (Sec. III) and lemmas (Sec. IV) necessary for deriving the fundamental formulas (Secs. V and VI). One of the lemmas is concerned with the well-known Kirchhoff formula[8–10] for steady states. In a remark in Sec. VII, the significance of the cycle statistics is discussed in relation to the mesoscopic thermodynamics.[11]

II. THE SCHEME OF TRANSITION SYSTEMS

In the usual master-equation description, a physical system is specified by a set $\mathcal{S}$ of element states and a nonnegative function γ on $\mathcal{S} \times \mathcal{S}$ whose value $\gamma(a \to b)$ for $(a,b) \in \mathcal{S} \times \mathcal{S}$ represents the tpr from a to b. As mentioned in Sec. I, however, this scheme $(\mathcal{S}, \gamma)$ is insufficient for the definition of real flows when there is the multiplicity of transitions. We need the individually defined tpr $\gamma(a \xrightarrow{i} b)$ rather than the total tpr $\gamma(a \to b) = \Sigma_i \gamma(a \xrightarrow{i} b)$. Moreover, we must eliminate unrealizable transitions, which have zero tpr, so as to make γ be defined as a strictly positive function. This elimination has no physical effect, but is required in our graph theoretical argument. We introduce, then, the set $\mathcal{T}$ of the realizable transitions on which the positive tpr γ is defined. In this scheme, a physical system is specified by the triplet $(\mathcal{S}, \mathcal{T}, \gamma)$, which will be called a *transition system* in the following formal definitions.

D 2.1: A *transition system* $(\mathcal{S}, \mathcal{T}, \gamma)$ is defined by the specification of a nonempty finite set $\mathcal{S}$, a finite set $\mathcal{T}$ of *transitions* over $\mathcal{S}$, and a strictly positive function γ on $\mathcal{T}$. The elements in $\mathcal{S}$ are called *elements states* or *points*. A *transition* over $\mathcal{S}$ means an object τ with which two points $I(\tau)$ and $F(\tau)$ in $\mathcal{S}$ (called the *initial* and *final* points of τ, respectively) are associated. The function γ is called the *transition probability rate* (tpr).

For $a,b \in \mathcal{S}$ the following subsets of $\mathcal{T}$ are defined:

$$\mathcal{T}_{a\cdot} = \{\tau \in \mathcal{T};\ I(\tau) = a\}, \tag{2.1}$$

$$\mathcal{T}_{\cdot b} = \{\tau \in \mathcal{T};\ F(\tau) = b\}, \tag{2.2}$$

and

$$\mathcal{T}_{a,b} = \mathcal{T}_{a\cdot} \cap \mathcal{T}_{\cdot b}. \tag{2.3}$$

Remark: We may write, if necessary, the transitions in $\mathcal{T}_{a,b}$ as $a \to b$ (if $|\mathcal{T}_{a,b}| = 1$) or $a \xrightarrow{i} b$, $i = 1,\ldots,n$ (if $n \equiv |\mathcal{T}_{a,b}| > 1$), where $|A|$ for a set A means the number of the elements belonging to A. In this notation, $I(a \xrightarrow{i} b) = a$ and $F(a \xrightarrow{i} b) = b$. In illustrations, the element states are represented as dots, and the transitions as arrows drawn between the dots.

The following concepts, *states*, *steady states*, *transition quantities*, and *state quantities*, are associated with a transition system $(\mathcal{S}, \mathcal{T}, \gamma)$.

D 2.2: A *state* is a probability distribution over $\mathcal{S}$. A *steady state* is a state p which satisfies the condition

$$\sum_{\tau \in \mathcal{T}_{\cdot a}} \gamma(\tau) p(I(\tau)) - \sum_{\tau \in \mathcal{T}_{a\cdot}} \gamma(\tau) p(a) = 0, \quad \forall\ a \in \mathcal{S}. \tag{2.4}$$

D 2.3: A *transition quantity* is a real function f on $\mathcal{T}$. A *state quantity* is a real function g on $\mathcal{S}$. In a state p, they have the values, respectively,

$$J(f;p) = \sum_{\tau \in \mathcal{T}} p(I(\tau))\, \gamma(\tau) f(\tau), \tag{2.5}$$

$$\langle g, p \rangle = \sum_{a \in \mathcal{S}} g(a)\, p(a), \tag{2.6}$$

called, respectively, the (*average*) *flow* of f and the *average* of g in p.

Remark: Let f be a transition quantity. If there is a state quantity g (called a *potential* of f) such that $f(\tau) = g(F(\tau)) - g(I(\tau))$ for all $\tau \in \mathcal{T}$, then we have

$$J(f;p_t) = \frac{d}{dt} \langle g, p_t \rangle, \tag{2.7}$$

where p_t is a time-dependent state satisfying the master equation

$$\frac{d}{dt} p_t(a) = \sum_{\tau \in \mathcal{T}_{\cdot a}} \gamma(\tau)\, p_t(I(\tau)) - \sum_{\tau \in \mathcal{T}_{a\cdot}} \gamma(\tau)\, p_t(a). \tag{2.8}$$

In this case, therefore, the flow is zero in the steady state given by the time-independent solution of (2.8). Note that not all transition quantities have potentials, as can be seen from the simple example in Sec. I. Equation (2.7) does not hold if f has no potentials. Hence, one can see that the above definition (2.5) allows nonzero steady flows.

III. GRAPH THEORETICAL DEFINITIONS

We present graph theoretical definitions associated with a transition system $(\mathcal{S}, \mathcal{T}, \gamma)$ (cf. the remark at the end of this section).

D 3.1: A *path* is a finite sequence of transitions $(\tau_1,\ldots,\tau_n)$ $(n \geqslant 1)$ such that $F(\tau_i) = I(\tau_{i+1})$ for $i = 1,\ldots,n-1$. The

points $I(\tau_1)$ and $F(\tau_n)$ are called the *initial* and *final* points of the path, respectively, and the integer n the *length* of the path. A path $(\tau_1,...,\tau_n)$ is called *simple* if the points $I(\tau_1)$, $F(\tau_1),...,F(\tau_n)$ are different from each other. A path $(\tau_1,...,\tau_n)$ is called *cyclic* if the subpath $(\tau_1,...,\tau_{n-1})$ is simple and $I(\tau_1)=F(\tau_n)$. Let $G\subset\mathcal{T}$. A *G-path* is a path $(\tau_1,...,\tau_n)$ such that $\tau_i\in G$ for $i=1,...,n$.

Note that if there is a path from a point a to b, then we can make a simple path from a to b by eliminating cyclic parts in the former path.

D 3.2: A *cycle* is a subset C of $\mathcal{T}$ such that the elements of C can be arrayed so as to make a cyclic path. Note that if $F(\tau)=I(\tau)$, then the one-element set $\{\tau\}$ is itself a cycle.

The set of all cycles in $\mathcal{T}$ is denoted by CCL. We also define the notation for a cycle C:

$$V(C)=\{I(\tau);\ \tau\in C\}\quad(=\{F(\tau);\ \tau\in C\}). \tag{3.1}$$

D 3.3: Let $|\mathcal{S}|\geqslant 2$. An *oriented rooted tree* (ort) is a subset T of $\mathcal{T}$ with which a point a (called the *root* of T) is associated such that for any $b\in\mathcal{S}-\{a\}$ there exists a unique T-path from b to a. Note that this uniqueness implies that (i) the T-path from b to a is simple, (ii) the root a is unique, and (iii) there is no cycles in T (cf. Appendix B).

In the case of $|\mathcal{S}|=1$, for convenience, we define the ort by the empty set $\varnothing$ and its root by the only one element in $\mathcal{S}$.

The set of all ort's with the root a is denoted by ORT(a). We also put $\mathrm{ORT}=\cup_{a\in\mathcal{S}}\mathrm{ORT}(a)$.

D 3.4: An *oriented rooted forest* (orf) is a subset F of $\mathcal{T}$ which includes a cycle C (called the *root cycle* of F) such that $F-\{\tau\}\in\mathrm{ORT}(I(\tau))$ for *some* transition $\tau\in C$. Note that this definition implies the facts that (i) there are no cycles in F except the root cycle C and (ii) $F-\{\tau\}\in\mathrm{ORT}(I(\tau))$ for any $\tau\in C$.

The set of all orf's with the root cycle C is denoted by ORF(C). We also put $\mathrm{ORF}=\cup_{C\in\mathrm{CCL}}\mathrm{ORF}(C)$.

D 3.5: For every subset $G\subset\mathcal{T}$, a positive value $W(G)$ (called the *weight* of G) is defined by[12]

$$W(G)=\prod_{\tau\in G}\gamma(\tau). \tag{3.2}$$

D 3.6: For $a,b\in\mathcal{S}$, the notation $a\sim b$ means that either $a=b$ or there exist both paths from a to b and from b to a. An *attractor* is a nonempty subset A of $\mathcal{S}$ such that (i) $a\sim b$ if $a,b\in A$ and (ii) there are no paths from a to b if $a\in A$ and $b\notin A$. From this definition it follows that there is at least one attractor, and any two different attractors have no intersection (cf. Appendix A).

D 3.7: If $a\sim b\ \forall a,b\in\mathcal{S}$, then the transition system is called *strongly connected*. In this case, $\mathcal{S}$ itself is the only one attractor; moreover, the sets ORT(a) and ORF(C) are nonempty for every point a and every cycle C (for the proof of this fact, consult Ref. 5 or 6).

D 3.8: Let A be a nonempty subset of $\mathcal{S}$. Put $\mathcal{T}_A=\{\tau\in\mathcal{T};\ I(\tau)\in A \text{ and } F(\tau)\in A\}$. Then we have a new transition system $(A,\ \mathcal{T}_A,\ \gamma)$, which is called the *A-restriction*. It is evident that the A-restriction for an attractor A is strongly connected.

Example: Let $\mathcal{S}=\{a,b,c,d,e\}$ and $\mathcal{T}=\{a\xrightarrow{1}b,\ a\xrightarrow{2}b,\ b\to c,\ c\to c,\ c\to d,\ d\to b,\ d\to e,\ e\to d,\ e\to a\}$. Then this system is strongly connected, and we have the cycles, ort's and orf's as follows:

$$\mathrm{CCL}=\{C_1,...,C_5\},\ \text{where } C_i=\{a\xrightarrow{i}b\to c\to d\to e\to a\}$$

$(i=1,2,)$, $C_3=\{b\to c\to d\to b\}$, $C_4=\{e\to d\to e\}$, and $C_5=\{c\to c\}$. Here we have used the abbreviation such as $a\to b\to c\cdots$ for $a\to b$, $b\to c,\cdots$. This abbreviation is also used below.

$\mathrm{ORT}(a)=\{T_a\}$, $\mathrm{ORT}(b)=\{T_b^1,...,T_b^6\}$, $\mathrm{ORT}(c)=\{T_c^1,...,T_c^6\}$, $\mathrm{ORT}(d)=\{T_d^1,...,T_d^4\}$, and $\mathrm{ORT}(e)=\{T_e^1,T_e^2\}$, where $T_a=\{b\to c\to d\to e\to a\}$, $T_b^i=\{c\to d\to e\to a\xrightarrow{i}b\}$, $T_b^{i+2}=\{c\to d,\ e\to d\to b,\ a\xrightarrow{i}b\}$, $T_b^{i+4}=\{c\to d\to b,\ e\to a\xrightarrow{i}b\}$, $T_c^i=\{d\to e\to a\xrightarrow{i}b\to c\}$, $T_c^{i+2}=a\xrightarrow{i}b,\ e\to d\to b\to c\}$, $T_c^{i+4}=\{e\to a\xrightarrow{i}b,\ d\to b\to c\}$, $T_d^i=\{e\to a\xrightarrow{i}b\to c\to d\}$, $T_d^{i+2}=\{a\xrightarrow{i}b\to c\to d,e\to d\}$, and $T_e^i=\{a\xrightarrow{i}b\to c\to d\to e\}$ for $i=1,2$.

$\mathrm{ORF}(C_i)=\{F_i\}$ $(i=1,2)$, $\mathrm{ORF}(C_3)=\{F_3^1,...,F_3^4\}$, $\mathrm{ORF}(C_4)=\{F_4^1,F_4^2\}$, and $\mathrm{ORF}(C_5)=\{F_5^1,...,F_5^6\}$, where $F_i=C_i$ $(i=1,2)$, $F_3^i=\{e\to a\xrightarrow{i}b\}\cup C_3$ $(i=1,2)$, $F_3^{i+2}=\{e\to d,\ a\xrightarrow{i}b\}\cup C_3$ $(i=1,2)$, $F_4^i=a\xrightarrow{i}b\to c\to d\}\cup C_4$ $(i=1,2)$ and $F_5^i=T_c^i\cup C_5$ $(i=1,...,6)$.

Note that $F_i=T_a\cup\{a\xrightarrow{i}b\}$ $(i=1,2)$, $F_3^iT_d^i\cup\{d\to b\}$ $(i=1,...,4)$, and $F_4^i=T_e^i\cup\{e\to d\}$ $(i=1,2)$, which shows the accordance with the definition of orf's in D3.4.

If the transition $d\to e$ is erased from $\mathcal{T}$, then we have an example which is not strongly connected; there are no paths from any of b, c, and d to either a or e. In this case, there is one attractor $\{b,c,d\}$.

Remark: The pair $(\mathcal{S},\mathcal{T})$ corresponds to a graph in the usual graph theory.[13] The ort's and orf's are subgraphs of it. The concepts defined in this section will find their correspondents in Ref. 5 or 6 and partly in Ref. 7, if they confine themselves to the case without the multiplicity of transitions (or the case of 1-graphs in the terminology of Ref. 13). Especially, the correspondent of the orf is called variously: flux diagram,[1,4] oriented rooted pseudotree,[6] mapping tree,[5] and forest with basis.[7] Also the attractors are equivalent to what are called the ergodic classes in the theory of Markov chains.[5,6,14]

IV. LEMMAS

Lemma 1: Let $(\mathcal{S},\mathcal{T},\gamma)$ be a transition system. Let ϕ and ψ be real functions on ORF and on $\mathcal{T}$, respectively. Then we have the identity[12]

$$\sum_{\tau\in\mathcal{T}}\ \sum_{T\in\mathrm{ORT}(I(\tau))}\phi(T\cup\{\tau\})\psi(\tau)=\sum_{C\in\mathrm{CCL}}\ \sum_{F\in\mathrm{ORF}(C)}\phi(F)\sum_{\tau\in C}\psi(\tau). \tag{4.1}$$

The proof of this lemma is given in Appendix B. Here,

note that $T \cup \{\tau\} \in \mathrm{ORF}$ if $T \in \mathrm{ORT}(I(\tau))$ as shown in Appendix B. Lemma 1 is used in the proof of the second lemma and in Sec. V.

The second lemma corresponds to the so-called Kirchhoff formula[15] for the steady states of master equations. Our formula is more general than the usual one in that the multiplicity of transitions is taking into consideration in enumerating the ort's.

Lemma 2: (Hill,[1,4] generalized Kirchhoff formula): Let $(\mathscr{S}, \mathscr{T}, \gamma)$ be a strongly connected transition system. Then the system has a unique steady state p_{st}, and it is given by the formula

$$p_{\mathrm{st}}(a) = Z^{-1} \sum_{T \in \mathrm{ORT}(a)} W(T), \quad \forall\, a \in \mathscr{S}, \tag{4.2}$$

where

$$Z = \sum_{T \in \mathrm{ORT}} W(T). \tag{4.3}$$

Here, note that $\mathrm{ORT}(a) \neq \emptyset$ for $\forall\, a \in \mathscr{S}$ since $(\mathscr{S}, \mathscr{T}, \gamma)$ is strongly connected (cf. D3.7), and hence we have $Z > 0$ and $p_{\mathrm{st}}(a) > 0$. The proof of Lemma 2 is given in Appendix C.

The third lemma is concerned with the steady states of a transition system which is not necessarily strongly connected. For such a system, the uniqueness of a steady state is no longer guaranteed. As proved in Appendix A, any transition system has at least one attractor. Since the A-restriction for an attractor A is strongly connected (as mentioned in D3.8), Lemma 2 is applicable to the A-restriction, which has a unique steady state distributed over A.

Lemma 3: Let $(\mathscr{S}, \mathscr{T}, \gamma)$ be a transition system. Let $A_1, \ldots, A_n$ $(n \geqslant 1)$ be the attractors of $(\mathscr{S}, \mathscr{T}, \gamma)$ and q_i be the steady state of the A_i-restriction $(A_i, \mathscr{T}_{A_i}, \gamma)$ for $i = 1, \ldots, n$. The the state $p_{\alpha_1, \ldots, \alpha_n}$ defined by the following formula (4.4) is always a steady state of $(\mathscr{S}, \mathscr{T}, \gamma)$:

$$p_{\alpha_1, \ldots, \alpha_n} = \alpha_1 \bar{q}_1 + \cdots + \alpha_n \bar{q}_n \tag{4.4}$$

with

$$\alpha_1 + \cdots + \alpha_n = 1 \quad \text{and} \quad \alpha_1 \geqslant 0 \quad \text{for } i = 1, \ldots, n, \tag{4.5}$$

where $\bar{q}_i$ is defined by $\bar{q}_i(a) = q_i(a)$ for $a \in A_i$ and $\bar{q}_i(a) = 0$ for $a \notin A_i$. Conversely, any steady state of $(\mathscr{S}, \mathscr{T}, \gamma)$ can be written in the form (4.4)

The proof of Lemma 3 is given in Appendix D.

V. CYCLE STATISTICS FOR TRANSITION QUANTITIES

Lemma 3 suggests that it is enough to discuss only the strongly connected case as long as we are concerned with steady states. Any statement in the strongly connected case can be extended to the general case in a linear-combination form [cf. Sec. VII(1)]. So we deal with a strongly connected system $(\mathscr{S}, \mathscr{T}, \gamma)$, which is fixed throughout this section. By Lemma 2, the system has a unique steady state p_{st}, which can be written in the form (4.2).

With each transition quantity f, we associate a new function f° on CCL as follows:

$$f^\circ(C) = \sum_{\tau \in C} f(\tau) \quad \text{for } C \in \mathrm{CCL}. \tag{5.1}$$

It is evident that $f^\circ(C)$ means the elementary flow of f generated in the cycle C (or in the cyclic path determined by C).

Theorem (fundamental formula): Let f be a transition quantity. Then the flow of f in the steady state p_{st} is written in the form

$$J(f; p_{\mathrm{st}}) = \sum_{C \in \mathrm{CCL}} f^\circ(C)\, \omega(C) \tag{5.2}$$

with

$$\omega(C) = Z^{-1} \sum_{F \in \mathrm{ORF}(C)} W(F), \tag{5.3}$$

where Z is defined by (4.3) and $W(F)$ is the weight of F (cf. D3.5).

Proof: Substitution of (4.2) into (2.5) yields

$$J(f; p_{\mathrm{st}}) = Z^{-1} \sum_{\tau \in \mathscr{T}} \sum_{T \in \mathrm{ORT}(I(\tau))} \gamma(\tau)\, W(T) f(\tau). \tag{5.4}$$

From D3.5, it follows that $W(A \cup B) = W(A)\, W(B)$ if $A \cap B = \emptyset$; especially, $\gamma(\tau)\, W(T) = W(T \cup \{\tau\})$ if $\tau \notin T$. Moreover, if $T \in \mathrm{ORT}(I(\tau))$, then $\tau \notin T$ [cf. Appendix B]. Therefore, Eq. (5.4) becomes

$$J(f; p_{\mathrm{st}}) = Z^{-1} \sum_{\tau \in \mathscr{T}} \sum_{T \in \mathrm{ORT}(I(\tau))} W(T \cup \{\tau\}) f(\tau),$$

which, by Lemma 1, leads to

$$= Z^{-1} \sum_{C \in \mathrm{CCL}} \left\{ \sum_{F \in \mathrm{ORF}(C)} W(F) \right\} \left\{ \sum_{\tau \in C} f(\tau) \right\}$$

$$= \sum_{C \in \mathrm{CCL}} \omega(C) f^\circ(C). \quad \text{QED}$$

The quantity $\omega(C)$ defined by (5.3) is constructed solely from γ and is independent of the transition quantity f. Since the transition system considered here is strongly connected, we have $\mathrm{ORT}(C) \neq \emptyset$ (cf. D3.7) and hence $\omega(C) > 0$ for any cycle C.

The expression (5.3) agrees with Hill's diagrammatic definition of the cycle fluxes, and Hill's formula (1.2) can be derived from our formula (5.2). As mentioned in Sec. I, the probability current $j(\tau, p)$ along a transition τ in a state p is defined by

$$j(\tau, p) = p(I(\tau))\, \gamma(\tau). \tag{5.5}$$

For each $\tau \in \mathscr{T}$, define a function δ_τ on $\mathscr{T}$ as $\delta_\tau(\tau') = \delta_{\tau,\tau'}$ (Kronecker's symbol). Setting $f = \delta_\tau$ in (2.5) and in (5.2), we have, respectively,

$$J(\delta_\tau; p) = j(\tau, p) \tag{5.6}$$

and

$$J(\delta_\tau; p_{\mathrm{st}}) = \sum_{C \in \mathrm{CCL}(\tau)} \omega(C), \tag{5.7}$$

where

$$\mathrm{CCL}(\tau) = \{C \in \mathrm{CCL};\ C \ni \tau\}. \tag{5.8}$$

In the steady state p_{st}, therefore, we obtain

$$j(\tau, p_{\mathrm{st}}) = \sum_{C \in \mathrm{CCL}(\tau)} \omega(C). \tag{5.9}$$

This accords with the formula (1.2).

Comparing (5.2) and (2.5), we know that the summation

of the products $j \times f$ over $\mathscr{T}$ is rearranged into the summation of $\omega \times f^{\circ}$ over CCL, and we can interpret the quantity $\omega(C)$ as the steady probability current around the cycle C. This interpretation is also supported by (5.9).

Example: Consider the case of $\gamma(\tau) = \gamma(= \text{const})$. Then we have $W(G) = \gamma^{|G|}$. The fact that $|T| = |\mathscr{S}| - 1$ and $|F| = |\mathscr{S}|$ for $T \in$ ORT and $F \in$ ORF is easily checked by definition. Therefore, it follows that

$$Z = |\text{ORT}|\, \gamma^{|\mathscr{S}|-1} \tag{5.10}$$

and

$$\omega(C) = \gamma |\text{ ORF}(C)|/|\text{ORT}|. \tag{5.11}$$

For the special examples treated in Sec. III, we have $Z = 19\gamma^4$, $\omega(C_1) = \omega(C_2) = \gamma/19$, $\omega(C_3) = 4\gamma/19$, $\omega(C_4) = 2\gamma/19$, and $\omega(C_5) = 6\gamma/19$.

VI. CYCLE STATISTICS FOR STATE QUANTITIES

In this section, we also deal with the strongly connected transition system $(\mathscr{S}, \mathscr{T}, \gamma)$, but we exclude the case in which $|\mathscr{S}| = 1$ and $|\mathscr{T}| = 0$.

Define three functions θ on $\mathscr{S}$, u on CCL, and p_C on $\mathscr{S}$ for $C \in$ CCL as follows:

$$\theta(a)^{-1} = \sum_{\tau \in \mathscr{T}_{a\cdot}} \gamma(\tau), \tag{6.1}$$

$$u(C) = \omega(C)\langle \theta \rangle_C, \tag{6.2}$$

$$p_C(a) = \begin{cases} \theta(a)/\langle \theta \rangle_C & \text{if } a \in V(C), \\ 0 & \text{otherwise,} \end{cases} \tag{6.3}$$

where $\langle \cdot \rangle_C$ means the summation over $V(C)$. Here, note that $\mathscr{T}_{a\cdot} \neq \emptyset$ for any point a, which is evident by the definition of the strong connectedness (D3.7). Hence $\theta^{-1}(a) > 0$, and θ is well defined.

As corollaries to the theorem in Sec. V, we have the following facts:

Corollary 1: The following three identities hold:

$$\sum_{C \in \text{CCL}} u(C) = 1, \tag{6.4}$$

$$\sum_{a \in \mathscr{S}} p_C(a) = 1, \tag{6.5}$$

$$p_{\text{st}}(a) = \sum_{C \in \text{CCL}} u(C)\, p_C(a). \tag{6.6}$$

Proof: Define a function θ' on $\mathscr{T}$ by $\theta'(\tau) = \theta(I(\tau))$. From (2.5), we have $J(\theta'; p_{\text{st}}) = 1$. On the other hand, it follows from the fundamental formula (5.2) that $J(\theta'; p_{\text{st}}) = \Sigma_C\, \omega(C)\langle \theta \rangle_C = \Sigma_C\, u(C)$. Thus we obtain Eq. (6.4).

The identity (6.5) is evident by definition.

In order to prove (6.6), we define a function f_a on $\mathscr{T}$ for $a \in \mathscr{S}$ by

$$f_a(\tau) = \begin{cases} \theta(a) & \text{if } I(\tau) = a, \\ 0 & \text{otherwise.} \end{cases} \tag{6.7}$$

From (2.5), we have $J(f_a; p_{\text{st}}) = p_{\text{st}}(a)$. On the other hand, we obtain from (5.2)

$$J(f_a; p_{\text{st}}) = \sum_{C \in \text{CCL}(a)} \omega(C)\, \theta(a)$$

$$= \sum_{\text{CCL}} u(C)\, p_C(a), \tag{6.8}$$

where $\text{CCL}(a) = \{C \in \text{CCL};\ a \in V(C)\}$ and we have used the fact that

$$f^{\circ}_a(C) = \begin{cases} \theta(a) & \text{if } C \in \text{CCL}(a), \\ 0 & \text{otherwise.} \end{cases} \tag{6.9}$$

Thus the identity (6.6) has been proved. QED

Corollary 2: Let g be a state quantity. Then the average of g in the steady state p_{st} is written in the form

$$\langle g, p_{\text{st}} \rangle = \sum_{C \in \text{CCL}} u(C)\langle g, p_C \rangle. \tag{6.10}$$

Proof: This follows straightforwardly from the identity (6.6). QED

Since $\gamma > 0$ and $\theta > 0$, we have $u > 0$. The identity (6.4) shows that the positive function u can be regarded as a probability distribution over CCL. The function p_C, which is also positive, defines a state concentrated within $V(C)$. Formula (6.6) shows that the steady state p_{st} is expressed as an average of the states p_C, which are distributed over CCL with the probability weight $u(C)$. From these facts, it seems reasonable to interpret as follows: $u(C)$ is the probability of the cycle C occurring, and $p_C(a)$ is the conditional probability of the element state appearing in the point a given that the cycle C has occurred. However, this conditional probability is not the same concept as in the ordinary probability theory, as will be remarked in the next section.

Example: In the case of $\gamma(\tau) = \gamma(= \text{const})$ treated in the previous section, u and p_C are independent of γ, and we have

$$u(C) = \frac{|\text{ORF}(C)|}{|\text{ORT}|} \sum_{x \in V(C)} \frac{1}{|\mathscr{T}_{x\cdot}|}, \tag{6.11}$$

$$p_C(y) = \left(\sum_{x \in V(C)} \frac{|\mathscr{T}_{y\cdot}|}{|\mathscr{T}_{x\cdot}|} \right)^{-1}. \tag{6.12}$$

For the special example of Sec. III, these are as follows: $u(C_1) = u(C_2) = 7/38$, $u(C_3) = 20/38$, $u(C_4) = 2/19$, $u(C_5) = 6/19$, and p_{C_1} is given by $p_{C_1}(a) = p_{C_1}(d) = p_{C_1}(e) = 1/7$ and $p_{C_1}(b) = p_{C_1}(C) = 2/7$. p_{C_i} $(i > 1)$ are also easily obtained.

VII. CONCLUDING REMARKS

(1)Nonconnected case: In the general case in which the strong connectedness does not necessarily assumed, the fundamental formula (5.2) should be altered to

$$J(f; p_{\alpha_1}, \ldots, \alpha_n) = \sum_{i=1}^{n} \sum_{C \in \text{CCL}_i} \alpha_i\, \omega(C) f^{\circ}(C), \tag{7.1}$$

where the notation in Lemma 3 has been used, and CCL_i is the set of cycles defined for the A_i-restriction $(A_i, \mathscr{T}_{A_i}, \gamma)$ for each attractor A_i, and also $\omega(C)$ $(C \in \text{CCL}_i)$ is defined by the formula (5.3) for the A_i-restriction. The other formula are also altered in a similar manner.

(2) Conditional probabilities: If we adopt the interpretation of u and p_C mentioned in Sec. VI, then the quantity defined by

$$r_a(C) = u(C)\, p_C(a)/p_{\rm st}(a) \tag{7.2}$$

for $a \in \mathscr{S}$ and $C \in \mathrm{CCL}$ is interpreted as the conditional probability of the cycle C occurring given that the element state a has appeared. It should be noted that $u(C)$ is the probability of the occurrence of the *cycle* C, and not the probability of the *event* that the element state is observed in $V(C)$. The latter is given by $\mathrm{Prob}(V(C)) = \Sigma_{a\in V(C)}\, p_{\rm st}(a)$. Correspondingly, the state p_C is not simply the normalized restriction of $p_{\rm st}$ to $V(C)$ except in the special case in which $V(C)\cap V(C') = \emptyset$ for any two different cycles C and C'. These facts are actually reflected in the branching property of the uncertainty measure (or entropy) $H(p) = -\langle p, \ln p\rangle$. In the steady state $p_{\rm st}$, we have

$$H(p_{\rm st}) = -\sum_C u(C)\ln[u(C)] + \sum_C u(C)\, H(p_C) + \sum_C \sum_{a\in V(C)} u(C)\, p_C(a)\ln[r_a(C)]\,. \tag{7.3}$$

The first two summation parts represent the usual branching property.[16,17] The last part indicates the cross effect of the cycles, where the cross of cycles means that $V(C)\cap V(C') \neq \emptyset$. In fact, the last part vanishes in the exceptional case mentioned above.

(3) Stochastic heat engine: Hill[4] has discussed the second law of thermodynamics in his cycle theory. His discussion can be reformulated in the scheme of mesoscopic thermodynamics presented by the author,[11,18] and a new physical interpretation is given to the cycles in the steady state. In this scheme, an open system is described by a transition system $(\mathscr{S}, \mathscr{T}, \gamma)$ equipped with a state quantity s (called the *residual entropy*) and with a transition quantity σ (called the *elementary negentropy consumption*), where $\mathscr{T}$ is reversible, i.e., for any $\tau \in \mathscr{T}$ there is its reverse transition τ^{-1} such that $I(\tau) = F(\tau^{-1})$, $F(\tau^{-1}) = I(\tau)$, $(\tau^{-1})^{-1} = \tau$, and $\sigma(\tau^{-1}) = -\sigma(\tau)$. A main axion of the mesoscopic thermodynamics is the *individual detailed balance* (idb) condition

$$\gamma(\tau) = \gamma(\tau^{-1})\exp\{[\sigma(\tau) + s(F(\tau)) - s(I(\tau))]/k_B\}\,. \tag{7.4}$$

Let us consider the application of the cycle statistics to this scheme. By the reversibility of $\mathscr{T}$, every cycle C has its reverse C^{-1}, and from (7.4) we have

$$\omega(C) = \omega(C^{-1})\exp[\sigma^{\circ}(C)/k_B], \tag{7.5}$$

where $\sigma^{\circ}(C)$ is defined by (5.1) for σ, representing the negentropy consumed by the system while it follows the cycle C. The negentropy source is in the environment, and a quantity of negentropy consumption means increase of the environmental entropy by the same quantity.[19] The cycles, then, can be compared to thermodynamic cycles along which heat engines operate accompanied by consumption of negentropy (or production of entropy). It is reasonable to consider the open system in the steady state as a *stochastic heat engine*, which operates along the cycles C's with the stochastic weight $\omega(C)$. The stochastic heat engine is allowed to take such a cycle C that violates the second law "$\sigma^{\circ}(C)\geqslant 0$." However, macroscopically or on the level of average, the second law is always guaranteed in the sense that the cycle average of $\sigma^{\circ}(C)$, namely the flow of σ, is always nonnegative, i.e.,

$$\begin{aligned} J(\sigma; p_{\rm st}) &= \sum_C \sigma^{\circ}(C)\,\omega(C) \\ &= \tfrac{1}{2}\sum_C \sigma^{\circ}(C)\,[\omega(C) - \omega(C^{-1})] \\ &= \tfrac{1}{4}\sum_C \sigma^{\circ}(C)\tanh[\sigma^{\circ}(C)/2k_B]\,[\omega(C) + \omega(C^{-1})] \\ &\geqslant 0, \end{aligned}$$

where we have used the fact that $\sigma^{\circ}(C^{-1}) = -\sigma^{\circ}(C)$ and the condition (7.5), and the last inequality follows from $\omega(C) > 0$ and $x\tanh x \geqslant 0$.

The idea of the stochastic heat engine will provide a good framework for discussing the problem of Maxwell's demon.[18] If Maxwell's demon is expressed as a stochastic heat engine, then it is consistent with the second law, even though it has anti-second-law actions on the level of elementary processes. The idea serves also to attach physical meaning to each cycle, and hence it is useful when we approximate the formula (5.2) by collecting physically important cycles. We will discuss the stochastic heat engine in more detail elsewhere.

ACKNOWLEDGMENTS

The author wishes to thank Professor H. Hasegawa for his continual encouragement. He is also indebted to Professor N. Iwahori for valuable comments on graph theory and to Professor T. L. Hill for useful remarks on Ref. 4.

APPENDIX A: EXISTENCE OF ATTRACTOR

Let $(\mathscr{S}, \mathscr{T}, \gamma)$ be a transition system. We show that it has at least one attractor.

The binary relation $\sim$ is defined in D3.6. It is evidently an equivalence relation on $\mathscr{S}$. Hence, the set $\mathscr{S}$ is divided into equivalence classes by the relation $\sim$. Let $[\mathscr{S}]$ denote the family of the equivalence classes. The definition of attractors (D3.6) implies that an attractor is one of the equivalence classes. It is evident that an equivalence class α is an attractor if and only if there exist no transitions $\tau \in \mathscr{T}$ such that $I(\tau) \in \alpha$ and $F(\tau) \notin \alpha$.

In the case of $|[\mathscr{S}]| = 1$, the system is strongly connected and we have the only one attractor $\mathscr{S}$.

Next we consider the case $|[\mathscr{S}]| > 1$. *Suppose* there are no attractors. Then for any $\alpha \in [\mathscr{S}]$, there exists a transition τ such that $I(\tau) \in \alpha$ and $F(\tau) \notin \alpha$. Hence, we can obtain a infinite series of $\alpha_i \in [\mathscr{S}]$ $(i = 1,2,\ldots)$ such that $\alpha_i \neq \alpha_{i+1}$ and $\alpha_i \to \alpha_{i+1}$ for $i = 1,2,\cdots$, where the notation $\alpha \to \beta$ for $\alpha, \beta \in [\mathscr{S}]$ means that there exists $\tau \in \mathscr{T}$ such that $I(\tau) \in \alpha$ and $F(\tau) \in \beta$. Since the set $[\mathscr{S}]$ is finite ($|[\mathscr{S}]| \leqslant |\mathscr{S}| < \infty$), we can cut out finite subseries $\beta_1,\ldots,\beta_n$ from the series $\alpha_1, \alpha_2, \cdots$ such that (i) $n \geqslant 2$, (ii) $\beta_1 \to \beta_2 \to \cdots \to \beta_n \to \beta_1$, and (iii) $\beta_1,\ldots,\beta_n$ are different from each other. The property (ii) implies that $\beta_1 = \beta_2 = \cdots = \beta_n$ since $\beta_1,\ldots,\beta_n$ are equivalence classes. This *contradicts* either (i) or (iii). QED

APPENDIX B: PROOF OF LEMMA 1

If $|\mathscr{S}| = 1$, then ORF = CCL = $\{\{\tau\}; \tau \in \mathscr{T}\}$ and ORT = $\{\varnothing\}$; hence, the identity (4.1) is evident. In the case $|\mathscr{S}| > 1$, the proof is carried out in three steps.

(1) Let $\tau \in \mathscr{T}$ and $T \in \mathrm{ORT}(I(\tau))$. Then $T \cup \{\tau\}$ is an orf whose root cycle includes τ. This fact will be proved in this step.

In the case of $F(\tau) = I(\tau)$, the one-element set $\{\tau\}$ is a cycle, and it is evident that $T \cup \{\tau\} \in \mathrm{ORF}(\{\tau\})$.

Next we consider the case $F(\tau) \neq I(\tau)$. There is a unique T-path $(\tau_1,...,\tau_n)$ from $F(\tau)$ to $I(\tau)$. By the uniqueness, the T-path is necessarily simple. Therefore, the set $C \equiv \{\tau_1,...,\tau_n,\tau\}$ is a cycle. Moreover, $\tau \notin T$, because, if $\tau \in T$, the T-path from a point in $V(C)$ to $I(\tau)$ is not unique. Thus it follows that $T \cup \{\tau\} \supset C \ni \tau$ and $Y \cup \{\tau\} - \{\tau\} = T$, which means $T \cup \{\tau\} \in \mathrm{ORF}(C)$.

(2) Define the following sets $A(a)$ and $B(C)$ for $a \in \mathscr{S}$ and $C \in \mathrm{CCL}$:

$$A(a) = \{(T,\tau);\ T \in \mathrm{ORT}(a) \text{ and } \tau \in \mathscr{T}_{a\cdot}\}, \tag{B1}$$

$$B(C) = \{(F,\tau);\ F \in \mathrm{ORF}(C) \text{ and } \tau \in C\} \tag{B2}$$

and put $A = \cup_{a \in \mathscr{S}} A(a)$ and $B = \cup_{C \in \mathrm{CCL}} B(C)$. Define a mapping $m: A \to B$ by

$$m(T,\tau) = (T \cup \{\tau\},\tau) \quad \text{for } (T,\tau) \in A, \tag{B3}$$

where $m(T,\tau) \in B$ is evident from the step (1).

Here we show that m is *one-to-one*. Suppose $m(T,\tau) = m(T',\tau')$. Then

$$T \cup \{\tau\} = T' \cup \{\tau'\} \quad \text{and} \quad \tau = \tau'. \tag{B4}$$

From the definition of the set A, we have $T \in \mathrm{ORT}(I(\tau))$ and $T' \in \mathrm{ORT}(I(\tau'))$, which implies that

$$T \not\ni \tau \quad \text{and} \quad T' \not\ni \tau', \tag{B5}$$

as shown in the step (1). From (B4) and (B5), we obtain $(T,\tau) = (T',\tau')$.

Next we show that m is *onto*. Let $(F,\tau) \in B$, namely $F \in \mathrm{ORF}(C)$ and $\tau \in C$ for some cycle C. From D3.4 (ii) it follows that $F - \{\tau\} \in \mathrm{ORT}(I(\tau))$. Therefore, $(F - \{\tau\}, \tau) \in A(I(\tau))$ and $m(F - \{\tau\}, \tau) = (F, \tau)$.

(3) The left-hand side of Eq. (4.1) in Lemma 1 is transformed as follows:

$$\begin{aligned}\mathrm{lhs} &= \sum_{a \in \mathscr{S}} \sum_{\tau \in \mathscr{T}_{a\cdot}} \sum_{T \in \mathbf{ORT}(a)} \phi(T \cup \{\tau\})\, \psi(\tau) \\ &= \sum_{a \in \mathscr{S}} \sum_{(T,\tau) \in A(a)} \phi(T \cup \{\tau\})\, \psi(\tau) \\ &= \sum_{(T,\tau) \in A} \phi(T \cup \{\tau\})\, \psi(\tau),\end{aligned} \tag{B6}$$

where we have used the evident fact that $\mathscr{T}_{a\cdot} \cap \mathscr{T}_{b\cdot} = \varnothing$ and $A(a) \cap A(b) = \varnothing$ if $a \neq b$. Similarly, the right-hand side of Eq. (4.1) becomes

$$\mathrm{rhs} = \sum_{(F,\tau) \in B} \phi(F)\, \psi(\tau). \tag{B7}$$

Since the mapping $m: A \to B$ is one-to-one and onto, Eq. (B7) can be rewritten as

$$\mathrm{rhs} = \sum_{(T,\tau) \in A} \phi(m(T,\tau))\, \psi(\tau). \tag{B8}$$

From (B6), (B8), and (B3), we obtain lhs = rhs. QED

APPENDIX C: PROOF OF LEMMA 2

(1) As mentioned in Sec. IV, p_{st} of Eq. (4.2) is well defined as a strictly positive state. First of all, we prove that p_{st} satisfies the condition (2.4) for steady states.[20] For $a,b \in \mathscr{S}$, define a function $\psi_{a,b}$ on $\mathscr{T}$ by

$$\psi_{a,b}(\tau) = \begin{cases} 1 & \text{if } \tau \in \mathscr{T}_{a,b}, \\ 0 & \text{otherwise.} \end{cases} \tag{C1}$$

Put

$$I_{a,b} = \sum_{C \in \mathrm{CCL}} \left\{ \sum_{F \in \mathbf{ORF}(C)} W(F) \right\} \left\{ \sum_{\tau \in C} \psi_{a,b}(\tau) \right\}. \tag{C2}$$

From (C1), it follows that

$$\begin{aligned}\sum_{b \in \mathscr{S}} \sum_{\tau \in C} \psi_{b,a}(\tau) &= \sum_{a \in \mathscr{S}} \sum_{\tau \in C} \psi_{a,b}(\tau) \\ &= \begin{cases} 1 & \text{if } a \in V(C), \\ 0 & \text{otherwise.} \end{cases}\end{aligned} \tag{C3}$$

Hence

$$\sum_b I_{b,a} = \sum_b I_{a,b}. \tag{C4}$$

On the other hand, from Lemma 1 it follows that

$$I_{a,b} = \sum_{\tau \in \mathscr{T}} \sum_{T \in \mathbf{ORT}(I(\tau))} W(T)\, \gamma(\tau)\, \psi_{a,b}(\tau), \tag{C5}$$

where we have used the fact that $W(T \cup \{\tau\}) = W(T)\, \gamma(\tau)$ (cf. the proof in Sec. V). By summing Eq. (C5) over $b \in \mathscr{S}$, we have

$$\begin{aligned}\sum_{b \in \mathscr{S}} I_{a,b} &= \sum_{\tau \in \mathscr{T}_{a\cdot}} \sum_{T \in \mathbf{ORT}(I(\tau))} W(T)\, \gamma(\tau)\, \psi_{a,b}(\tau) \\ &= Z \sum_{\tau \in \mathscr{T}_{a\cdot}} p_{\mathrm{st}}(a)\, \gamma(\tau).\end{aligned} \tag{C6}$$

Similarly,

$$\sum_{b \in \mathscr{S}} I_{b,a} = Z \sum_{\tau \in \mathscr{T}_{\cdot a}} p_{\mathrm{st}}(I(\tau))\, \gamma(\tau). \tag{C7}$$

Equations (C4), (C6), and (C7) show that the state p_{st} satisfies the condition (2.4).

(2) Next, we prove the uniqueness of the steady state.[21] Let PS denote the set of strictly positive states. Let SS denote the set of steady states. Define a function Y on PS by

$$Y(p) = \sum_{\tau \in \mathscr{T}} p(I(\tau))\, \gamma(\tau) \ln \frac{p(F(\tau))\, p_{\mathrm{st}}(I(\tau))}{p_{\mathrm{st}}(F(\tau))\, p(I(\tau))}. \tag{C8}$$

The function has the following properties

$$p \in \mathrm{SS} \cap \mathrm{PS} \Rightarrow Y(p) = 0 \tag{C9}$$

and

$$Y(p) = 0 \Leftrightarrow p = p_{\mathrm{st}}, \tag{C10}$$

where the fact (C9) is derived by multiplying both sides of Eq. (2.4) by $\ln[p(a)/p_{\mathrm{st}}(a)]$ and taking the summation over $a \in \mathscr{S}$. The fact (C10) can be obtained by using the inequality $\log x \leqslant x - 1$ (with equality iff $x = 1$) and the assumption of

the strong connectedness. The facts (C9) and (C10) imply that $SS \cap PS = \{p_{st}\}$. Moreover, we have $SS = \{p_{st}\}$ because $SS \subset PS$ as shown below.

Suppose there exist a steady state p and a point a such that $p(a) = 0$. By the assumption of the strong connectedness, for any $b \in \mathscr{S} - \{a\}$ there exists a path $(\tau_1,...,\tau_n)$ from b to a. Put $a_{i-1} = I(T_i)$ $(n = 1,...,n)$ with $a_0 = b$ and $a_n = a$. The condition (2.4) implies the inequality

$$p(I(\tau))\,\gamma(\tau) \leqslant p(F(\tau))\,\mu(F(\tau)) \tag{C11}$$

for any $\tau \in \mathscr{T}$, where μ is defined by

$$\mu(x) = \sum_{\tau \in \mathscr{T}_x} \gamma(\tau). \tag{C12}$$

Setting $\tau = \tau_i$ in (C11), we obtain the inequality

$$P(a_{i-1})\,\gamma(\tau_i) \leqslant p(a_i)\,\mu(a_i), \quad i = 1,...,n, \tag{C13}$$

which implies

$$p(b) \leqslant \left(\prod_{i=1}^{n} \frac{\mu(a_i)}{\gamma(\tau_i)}\right) p(a) = 0. \tag{C14}$$

Thus, we have $p(b) = 0$ for all $b \in \mathscr{S}$, which is *contradictory*. QED

APPENDIX D: PROOF OF LEMMA 3

Let p be a steady state. If we show that $p(a) = 0$ for any point $a \notin X \equiv \cup_i A_i$, then Lemma 3 is evident by the linearity of the condition (2.4). So we show that fact below.

For $i = 1,...,n$, define

$$\mathscr{G}_i = \{\tau \in \mathscr{T};\ I(\tau) \notin A_i \quad \text{and} \quad F(\tau) \in A_i\}. \tag{D1}$$

By summing both sides of Eq. (2.4) over $a \in A_i$, we obtain

$$\sum_{\tau \in \mathscr{G}_i} p(I(\tau))\,\gamma(\tau) = 0. \tag{D2}$$

Since $p \geqslant 0$ and $\gamma > 0$, Eq. (D2) implies $p(I(\tau)) = 0$ for $\tau \in \mathscr{G}_i$. Therefore, $p(b) = 0$ for any $b \in Y \equiv \{I(\tau), \tau \in \cup_i \mathscr{G}_i\}$.

Let $a \notin X$. From the discussion in Appendix A, we know that there exists a path $(\tau_1,...,\tau_n)$ from a to a point in X. Put $a_{i-1} = I(\tau_i)$ $(n = 1,...,n)$ with $a_0 = a$. In the sequence $a_1,...,a_n$, let a_m be the first point that is included in X. Then $a_{m-1} \in Y$, and hence $p(a_{m-1}) = 0$. If $m = 1$, then $p(a) = p(a_0) = 0$. In the case $m > 1$, by the same technique as used in deriving (C14), we obtain

$$p(a) \leqslant \left(\prod_{i=1}^{m-1} \frac{\mu(a_i)}{\gamma(\tau_i)}\right) p(a_{m-1}) = 0. \quad \text{QED} \tag{D3}$$

[1] T. L. Hill, J. Theor. Biol. **10**, 442 (1966).
[2] T. L. Hill and O. Kedem, J. Theor. Biol. **10**, 399 (1966).
[3] T. L. Hill and Y. Chen, Proc. Nat. Acad. Sci. USA **72**, 1291 (1975).
[4] T. L. Hill, *Free Energy Transduction in Biology* (Academic, New York, 1977).
[5] Nobuko Iwahori, *Graphs and Stochastic Matrices* (Sangyo-Tosho, Tokyo, 1974) (in Japanese).
[6] Nobuko Iwahori, J. Fac. Sci., Univ. Tokyo **13**, 139 (1966); **14**, 285 (1967); **15**, 13 (1968).
[7] H.-H. Kohler and E. Vollmerhaus, J. Math. Biol. **9**, 275 (1980).
[8] J. Schnakenberg, Rev. Mod. Phys. **48**, 571 (1976).
[9] H. Haken, *Synergetics, An Introduction* (Springer-Verlag, Berlin, 1978), 2nd ed.
[10] W. Weidlich, Z. Phys. B **30**, 345 (1978).
[11] T. Nakagomi, J. Stat. Phys. **26**, 567 (1981).
[12] We will use the following convention. If the product sign Π is applied over the empty set, then it is equal to 1. Also, if the summation sign Σ is applied over the empty set, then it is equal to zero.
[13] C. Berge, *Graphs and Hypergraphs* (North-Holland, Amsterdam, 1973).
[14] J. L. Doob, *Stochastic Processes* (Wiley, New York, 1953).
[15] The Kirchhoff formula for the steady solution of the master equation is first presented by Hill in Ref. 1, see also Ref. 4. As a mathematical statement, the formula is included in the Bott-Mayberry theorem for trees and matrices, see R. Bott and J. P. Mayberry, in *Economic Activity Analysis*, edited by O. Morgenstein (Wiley, New York, 1954), pp. 391–400 with Math. Rev. **16**, 665 (1955) and see Refs. 5 and 6. In addition, there are many discussions on the formula in Refs. 8–10.
[16] A. J. Khinchin, *Mathematical Foundations of Information Theory* (Dover, New York, 1957).
[17] I. J. Taneja, J. Stat. Phys. **14**, 263 (1976).
[18] T. Nakagomi, "Negentropy Consumption for Information Transfer. Mesoscopic Thermodynamics of Chemically Driven Tape-Copying Systems," preprint, Kyoto Univ. (1982).
[19] For details about the relation between negentropy and entropy, see Refs. 11 and 18.
[20] The essential idea in this proof is the same as in Hill's (Ref. 1 or Appendix 2 in Ref. 4).
[21] Although the uniqueness can be verified in the general theory of Markov chains (see, e.g., Ref. 14), we present here a simple proof on the basis of the condition (2.4) alone and without introducing any analytical structure of time evolution. The idea of the function $Y(p)$ is owed to Ref. 8.

An equation of Hammerstein type arising in particle transport theory

V. C. Boffi and G. Spiga
Laboratorio di Ingegneria Nucleare dell'Università di Bologna, 16, via dei Colli, 40136 Bologna, Italy

(Received 15 October 1982; accepted for publication 6 January 1983)

An equation of Hammerstein type arising in particle transport theory is derived. A sufficient condition for the existence of a solution to it is obtained via a simple application of the contracting mapping principle.

PACS numbers: 05.60. + w, 02.30.Rz, 02.40. + m

I. INTRODUCTION

As well known (see, for instance, Ref. 1), an equation of Hammerstein type is a nonlinear integral equation of the form

$$\psi(x) + \int_a^b K(x,y)\varphi\,[y,\psi(y)]\,dy = 0 \tag{1}$$

with a, b real, $\psi(x)$ being the unknown function. Equations of Hammerstein type are today receiving a certain deal of attention. Most of the recent literature is in Russian, and unfortunately is not easily available. Among the papers published in other languages the reader can, for instance, refer to those quoted under Refs. 2 and 3, the latter containing further bibliographical informations. As most of the literature deals with the theory of abstract Hammerstein equations (problems of existence, positivity, and bifurcation of the solution to such abstract equations are mainly investigated), it may be of some interest to illustrate here an equation of Hammerstein type that we have encountered in particle transport theory. To our knowledge, in fact, such type of equations is not very common in the applications, even if also the original paper by Hammerstein, that goes back to 1930, has been inspired by an application concerning forced oscillation problems.[4]

The paper is organized as follows.

After the problem of particle transport theory to be dealt with is stated, a detailed derivation of the Hammerstein-type equation associated with it is developed by resorting to a standard spherical harmonics technique, and laying particular emphasis on the most characterizing physical and mathematical implications. A study of the existence of a solution to the equation so derived, in which the integration interval becomes infinite (we have indeed $a = 0$ and $b = +\infty$, respectively), is then attempted by exploiting the method of the successive approximations. This classical approach fails, however, because of the shape of the kernel $K(x, y)$, that is not as smooth as necessary, in the physical cases of interest, over its infinite domain $(0,\infty)\times(0,\infty)$ of definition. At last, anyway, we can succeed in proving the existence of a solution to the equation under examination by a straightforward application of the contracting mapping principle. This is made possible on the basis of certain assumptions concerning the ratios of the shape function of the external source and, respectively, of the physical kernel relevant to the particles considered, to the removal collision frequency of the host medium which those particles diffuse through.

A. Statement of the problem

The study of the diffusion of the particles of a mixture is of basic importance in many fields of modern physics and engineering. One of the most advanced tools to face this study is represented by the nonlinear *integrodifferential Boltzmann* (or transport) *equation*. In the case of a mixture of two species of particles, of mass m and M, respectively, and that will be distinguished as test particles (t.p.) and field particles (f.p.), respectively, the relevant transport equation reads as[5]

$$\frac{Df}{Dt} = -\left[\hat{\nu}_R(\mathbf{v}) + \nu_R(\mathbf{x},\mathbf{v},t) + \nu_S(\mathbf{x},\mathbf{v},t)\right] f(\mathbf{x},\mathbf{v},t) + G(\mathbf{x},\mathbf{v},t), \quad \mathbf{x},\mathbf{v}\in\mathbb{R}_3, \quad t\in(0,\infty), \tag{2}$$

and is to be equipped with the appropriate initial and boundary conditions.

In Eq. (2), where the unknown function f, depending on position $\mathbf{x}$, velocity $\mathbf{v}$, and time t, is the distribution function of the t.p.,

$$\frac{Df}{Dt} \equiv \left(\frac{\partial}{\partial t} + \mathbf{v}\cdot\nabla_{\mathbf{x}} + \frac{\mathbf{F}}{m}\cdot\nabla_{\mathbf{v}}\right) f \tag{3}$$

is the substantial derivative of f, ∇ being the gradient operator, and $\mathbf{F} = \mathbf{F}(\mathbf{x}, t)$ the external force acting on the t.p.,

$$\hat{\nu}_R(\mathbf{v}) = \hat{g}_R(|\mathbf{v}|)N, \tag{4}$$

where N is the fixed total density of the f.p., is the collision frequency for the removal of the t.p. by the f.p.,

$$\nu_\alpha(\mathbf{x},\mathbf{v},t) = \int_{\mathbf{R}_3} g_\alpha(|\mathbf{v}-\mathbf{v}'|) f(\mathbf{x},\mathbf{v}',t)\,d\mathbf{v}' \quad (\alpha = S, R) \tag{5}$$

is the collision frequency for the t.p.–t.p. α-event, α standing for either S (scattering) or R (removal), and

$$G(\mathbf{x},\mathbf{v},t) = Q(\mathbf{x},\mathbf{v},t) + I_S(\mathbf{x},\mathbf{v},t) = Q(\mathbf{x},\mathbf{v},t) + \int_{\mathbf{R}_3}\int_{\mathbf{R}_3} g_S(|\mathbf{v}'-\mathbf{v}''|)\Pi_S(\mathbf{v}',\mathbf{v}''\to\mathbf{v}) \times f(\mathbf{x},\mathbf{v}',t) f(\mathbf{x},\mathbf{v}'',t)\,d\mathbf{v}'d\mathbf{v}'', \tag{6}$$

where $\Pi_S(\mathbf{v}',\mathbf{v}''\to\mathbf{v})$ denotes the scattering probability distribution, includes all the gain terms, the external source $Q(\mathbf{x},\mathbf{v},t)$, and the scattering-in integral $I_S(\mathbf{x},\mathbf{v},t)$, respectively, only binary collisions being considered.

For more details on Eq. (2) the reader can also refer, for instance, to Refs. 6 and 7. [It must be noted that in these monographs the rhs of Eq. (2) has mostly an equivalent, but different formulation. We use here, as illustrated in Ref. 5, the so-called scattering-kernel formulation.] Further results

 0022-2488/83/061625-05$02.50 © 1983 American Institute of Physics

concerning existence and uniqueness as well as convergence of some iterative schemes for the solution to Eq. (2)—in connection with various specializations of the nonnegative functions $Q, \hat{g}_R, g_S, g_R$, and Π_S and various geometries—can also be found in the recent review articles of Refs. 8 and 9.

Here we shall consider the special case in which t.p., emitted at a constant, spatially uniform rate $Q_0 S(\mathbf{v})$ (with $\int_{\mathbf{R}_3} S(\mathbf{v})\, d\mathbf{v} = 1$) in an infinite host medium consisting of certain f.p., either are removed by these f.p. or remove each other, the removal effects being described by some $\hat{g}_R(|\mathbf{v}|)$ and $g_R(|\mathbf{v} - \mathbf{v}'|)$ to be specified according to the physics of the problem. All the system is subject to the action of an external constant force $\mathbf{F}$. Equation (2) reduces thus to

$$\frac{\mathbf{F}}{m} \cdot \nabla_{\mathbf{v}} f + \hat{g}_R(|\mathbf{v}|) N f(\mathbf{v}) + f(\mathbf{v}) \int_{\mathbf{R}_3} g_R(|\mathbf{v} - \mathbf{v}'|) f(\mathbf{v}')\, d\mathbf{v}' = Q_0 S(\mathbf{v}), \tag{7}$$

which is a nonlinear integrodifferential equation to be integrated upon suitable vanishing conditions at infinity, namely, for $v_\alpha \to \pm\infty$, v_α ($\alpha = 1,2,3$) being the general component of $\mathbf{v}$ along the x_α-axis.

Equation (7) is essentially applicable for studying physical situations in which removal of t.p. between themselves is a dominant event. Thus, for instance, in the study of the chemical and biological action of radiations, the annihilation of certain radiations between themselves is the basic important event.[10] Another example that has many applications in the study of materials and their properties is represented by the annihilation at low energy of a positron with a free electron at rest, when positron and electron can be regarded as indistinguishable particles.[11]

B. The P_L approximation to Eq. (7)

Using polar coordinates v, θ, φ with polar axis in the direction of $\mathbf{F}$, one gets

$$\frac{\mathbf{F}}{m} \cdot \nabla_{\mathbf{v}} = \frac{F}{m}\left(\mu \frac{\partial}{\partial v} + \frac{1-\mu^2}{v} \frac{\partial}{\partial \mu}\right) \tag{8a}$$

with $\mu = \cos\theta$. If now we choose the new dependent variable such that

$$v^2 f(\mathbf{v}) = f(v, \mu, \varphi), \tag{8b}$$

Eq. (7) becomes

$$\frac{F}{m}\left(\mu \frac{\partial f}{\partial v} + \frac{1-\mu^2}{v} \frac{\partial f}{\partial \mu} - \frac{2\mu}{v} f\right) + N\hat{g}_R(v) f(v, \mu, \varphi) + f(v, \mu, \varphi) \int_0^{2\pi} d\varphi' \int_{-1}^{1} d\mu' \int_0^{+\infty} g_R[(v^2 + v'^2 - 2vv'\mu_0)^{1/2}] f(v', \mu', \varphi')\, dv' = Q_0 S(v, \mu, \varphi) \tag{9}$$

with

$$\mu_0 = \mu\mu' + [(1-\mu^2)(1-\mu'^2)]^{1/2} \cos(\varphi - \varphi'). \tag{9a}$$

If S does not depend on φ, the same occurs to f. Let $P_l(z)$ denotes the l th Legendre polynomial and expand $g_R(|\mathbf{v} - \mathbf{v}'|)$ as

$$g_R(|\mathbf{v} - \mathbf{v}'|) = \sum_{l=0}^{\infty} \frac{2l+1}{2} g_R^l(v, v') P_l(\mu_0), \tag{10}$$

$$g_R^l(v, v') = \int_{-1}^{1} g_R[(v^2 + v'^2 - 2vv'\mu_0)^{1/2}] P_l(\mu_0)\, d\mu_0$$

so that Eq. (9) can be rewritten as

$$\frac{F}{m}\left(\mu \frac{\partial f}{\partial v} + \frac{1-\mu^2}{v} \frac{\partial f}{\partial \mu} - \frac{2\mu}{v} f\right) + N\hat{g}_R(v) f(v, \mu) + f(v, \mu) \sum_{l=0}^{\infty} \frac{2l+1}{2} P_l(\mu) \int_0^{\infty} g_R^l(v, v')\, dv'\, 2\pi \times \int_{-1}^{1} P_l(\mu') f(v', \mu')\, d\mu' = Q_0 S(v, \mu), \tag{11}$$

once $P_l(\mu_0)$ has been expressed according to the addition theorem for Legendre polynomials.[12]

Expanding analogously the unknown $f(v, \mu)$ as

$$f(v, \mu) = \sum_{l=0}^{\infty} \frac{2l+1}{4\pi} f_l(v) P_l(\mu),$$

$$f_l(v) = 2\pi \int_{-1}^{1} P_l(\mu) f(v, \mu)\, d\mu \tag{12}$$

and recalling the recursion and differentiation formulas for Legendre polynomials yields finally

$$\frac{F}{m} \sum_{l=0}^{\infty} \frac{2l+1}{4\pi} \left\{ \frac{df_l}{dv}\left[\frac{l+1}{2l+1} P_{l+1}(\mu) + \frac{l}{2l+1} P_{l-1}(\mu)\right] + \frac{1}{v} f_l\left[- \frac{(l+1)(l+2)}{2l+1} P_{l+1}(\mu) + \frac{l(l-1)}{2l+1} P_{l-1}(\mu)\right]\right\} + N\hat{g}_R(v) \sum_{l=0}^{\infty} \frac{2l+1}{4\pi} f_l(v) P_l(\mu) + f(v, \mu) \sum_{l=0}^{\infty} \frac{2l+1}{2} P_l(\mu) \int_0^{\infty} g_R^l(v, v') f_l(v')\, dv' = Q_0 S(v, \mu), \tag{13}$$

which is a new form of Eq. (7).

In order to generate the appropriate system of infinite nonlinear integrodifferential equations defining the infinite unknown Legendre coefficients, $f_l(v)$ of $f(v, \mu)$, let us multiply both sides of Eq. (13) by $2\pi P_k(\mu)$ and then integrate over $(-1, 1)$. There results

$$\frac{F}{m}\left[\frac{k}{2k+1}\frac{df_{k-1}}{dv}+\frac{k+1}{2k+1}\frac{df_{k+1}}{dv}\right.$$

$$\left.-\frac{k(k+1)}{2k+1}\frac{1}{v}f_{k-1}(v)+\frac{k(k+1)}{2k+1}\frac{1}{v}f_{k+1}(v)\right]+N\hat{g}_R(v)f_k(v)$$

$$+\sum_{l=0}^{\infty}\frac{2l+1}{2}\sum_{j=0}^{k+l}\alpha_j^{kl}f_j(v)\int_0^{\infty}g_R^l(v,v')f_l(v')\,dv'=Q_0\,S_k(v), \tag{14}$$

where the numbers α_j^{kl} are defined by the equality

$$P_k(\mu)P_l(\mu)=\sum_{j=0}^{k+l}\alpha_j^{kl}\,P_j(\mu). \tag{14a}$$

Determining α_j^{kl} is, in principle, straightforward, even though the actual task becomes rather awkward in practice, when k, or l, increases. The spherical harmonics (or P_L) approach would consist now in setting $f_l(v)=0$ for $l>L$ and in determining the first $L+1$ moments $f_0, f_1, \ldots, f_L$ by solving the set of $L+1$ nonlinear integrodifferential equations obtained by setting $k=0,1,\ldots,L$ in Eq. (14). The choice of a reliable value of L must be made according to the order of anisotropy of the external source and to the strength of the external force. We report here, as an example, the equations relevant to the P_1 approximation, which are, of course, in order only for almost isotropic source and weak external field. They read as

$$\frac{F}{m}\frac{df_1}{dv}+N\hat{g}_R(v)f_0(v)+\tfrac{1}{2}f_0(v)\int_0^{\infty}g_R^0(v,v')f_0(v')\,dv'+\tfrac{3}{2}f_1(v)\int_0^{\infty}g_R^1(v,v')f_1(v')\,dv'=Q_0\,S_0(v), \tag{15a}$$

$$\frac{F}{3m}\left[\frac{df_0}{dv}-\frac{2}{v}f_0(v)\right]+N\hat{g}_R(v)f_1(v)+\tfrac{1}{2}f_1(v)\int_0^{\infty}g_R^0(v,v')f_0(v')\,dv'+\tfrac{1}{2}f_0(v)\int_0^{\infty}g_R^1(v,v')f_1(v')\,dv'=Q_0\,S_1(v), \tag{15b}$$

and their solution is already not easy at all.

Of course, other approaches may be proposed for the solution of Eq. (7) different from the P_L approximation so classical in linear transport theory. We note, however, that beside the exact calculation of the relevant distribution function of the t.p., another main goal of the present investigation (we shall be mostly focusing on this in the sequel) is the understanding and the resolution of the nonlinearity associated with the pure removal effects taking place in the mixture considered. Indeed, the system, Eq. (14), and its derivation constitute an illuminating sample about the nature and the structure of such nonlinearity.

C. The isotropic forceless case: Eq. (15a) as an equation of Hammerstein type

It is easily realized that in the limit of $F\to 0$ and for isotropic source (i.e., $S_k=0$ when $k\geqslant 1$) the system, Eq. (14), is solved by $f_k(v)=0$ for any $k\geqslant 1$, and with $f_0(v)$ solution to the autonomous nonlinear integral equation

$$N\hat{g}_R(v)f_0(v)+\tfrac{1}{2}f_0(v)\int_0^{\infty}g_R^0(v,v')f_0(v')\,dv'=Q_0\,S_0(v). \tag{16}$$

The same result is achieved directly starting from Eq. (7), which, for $\mathbf{F}=0$ and $S(\mathbf{v})=(4\pi)^{-1}S_0(v)$, is easily seen to be solved by $f(\mathbf{v})=(4\pi)^{-1}f_0(v)$, provided $f_0(v)$ satisfies Eq. (16). If we further set

$$\psi(v)=2Q_0\,S_0(v)/f_0(v)-2N\hat{g}_R(v) \tag{17a}$$

and

$$\varphi(v,u)=2Q_0\,S_0(v)/[2N\hat{g}_R(v)+u], \tag{17b}$$

then Eq. (16) can be recast, in terms of ψ, as

$$\psi(v)=\int_0^{\infty}g_R^0(v,v')\varphi[v',\psi(v')]\,dv', \tag{18}$$

and $f_0(v)$ turns out to be

$$f_0(v)=\varphi[v,\psi(v)]. \tag{19}$$

Equation (18) is clearly of the type like Eq. (1) with $g_R^0=-K$ and $a=0$, $b=\infty$, respectively.

As long as the existence of a solution to Eq. (18) and the convergence of the relevant method of successive approximations are concerned, there exist classical conditions that, following Ref. 1 [in which, however, the interval of integration is finite ($a=0$, $b=1$)], can be formulated as follows:

(i) The kernel $g_R^0(v,v')$ belongs to the class $L_2(0,\infty)$, or, at least, the function

$$A^2(v)=\int_0^{\infty}[\,g_R^0(v,v')\,]^2\,dv' \tag{20}$$

exists almost everywhere in $(0,\infty)$ and is summable there. In Eq. (20), we have

$$g_R^0(v,v')=\int_{-1}^{1}g_R[(v^2+v'^2-2vv'\mu_0)^{1/2}]\,d\mu_0$$

$$=\frac{1}{vv'}\int_{|v-v'|}^{v+v'}xg_R(x)\,dx, \tag{21}$$

as there results from Eq. (10) with $l=0$.

(ii) The function $\varphi(v,u)$ satisfies uniformly a Lipschitz condition of the type

$$|\varphi(v,u_1) - \varphi(v,u_2)| < C(v)|u_1 - u_2|. \tag{22}$$

(iii) The function

$$\varphi(v,0) = Q_0 S_0(v)/N\hat{g}_R(v) \tag{23}$$

belongs to $L_2(0,\infty)$.

These two last conditions do depend on $\hat{g}_R(v)$. [The dependence on $S_0(v)$ is not as critical as the one on $\hat{g}_R(v)$.] We remark that the case $\hat{g}_R(v) = 0$, that is also of physical interest, must be excluded if we want to fulfill condition (iii).

Sufficient condition for the iterative scheme ($n = 0,1,2,...$)

$$\psi_{n+1}(v) = \int_0^\infty g_R^0(v,v')\varphi\,[v',\psi_n(v')]\,dv', \tag{24a}$$

$$\psi_0(v) = 0, \tag{24b}$$

to converge almost everywhere to a solution of Eq. (18) is then proved by standard technique to be

$$\int_0^\infty A^2(v)C^2(v)\,dv < 1. \tag{25}$$

(This convergence turns actually to be also almost uniform.)

Let us check now the validity of the condition (25) for the following physical behaviors of $g_R(|\mathbf{v} - \mathbf{v}'|) > 0$, $\hat{g}_R(v)$ being also a positive function.

(a) $g_R(|\mathbf{v} - \mathbf{v}'|) = \text{const} = C_R > 0$: In this case, $g_R^0(v,v') = 2C_R$, and therefore condition (i) is not satisfied. In spite of this, Eq. (7) can even be solved in closed form for any source function $S(\mathbf{v})$ and in the presence of an external force $\mathbf{F}$. Equation (7) is now

$$\frac{F}{m}\frac{\partial f}{\partial v_3} + N\hat{g}_R(v)f(\mathbf{v}) + C_R\, nf(\mathbf{v}) = Q_0 S(\mathbf{v}), \tag{26}$$

where

$$n = \int_{\mathbf{R}_3} f(\mathbf{v})\,d\mathbf{v} \tag{26a}$$

is the (unknown) total density of the t.p. considered, and we suppose, without loss of generality, that the force $\mathbf{F}$ is parallel to the x_3 axis. If also $\hat{g}_R(v) = \text{const} = \hat{C}_R > 0$, then Eq. (26) is solved straightforwardly. Indeed, by integrating both sides of Eq. (26) over the domain of the velocities and recalling both the normalization for $S(\mathbf{v})$ and the condition at infinity, we get for n the continuity equation

$$C_R n^2 + \hat{C}_R\, Nn = Q_0, \tag{27}$$

from which

$$n = [\,-N\hat{C}_R + (N^2\hat{C}_R^2 + 4Q_0 C_R)^{1/2}]/(2C_R) > 0. \tag{28}$$

The solution $f(\mathbf{v})$ to Eq. (26) is thus explicit and reads as

$$f(\mathbf{v}) = (m/F)Q_0\, e^{-(m/F)(\hat{C}_R N + C_R n)v_3} \times \int_{-\infty}^{v_3} e^{+(m/F)(\hat{C}_R N + C_R n)v_3'} S(v_1,v_2,v_3')\,dv_3'. \tag{29}$$

In the limit of $F \to 0$, Eq. (29) reduces to

$$f(\mathbf{v}) = [Q_0/(\hat{C}_R N + C_R n)]S(\mathbf{v}). \tag{30}$$

(b) $g_R(|\mathbf{v} - \mathbf{v}'|) = B_R|\mathbf{v} - \mathbf{v}'|^{-1}$ ($B_R > 0$): In this case

$$g_R^0(v,v') = 2B_R\,[(1/v)U(v - v') + (1/v')U(v' - v)], \tag{31a}$$

where $U(z)$ denotes the unit step function, and

$$A^2(v) = 8B_R^2/v, \tag{31b}$$

which clearly is not of class $L_1(0,\infty)$. Consequently, condition (i) is again not satisfied.

(c) $g_R(|\mathbf{v} - \mathbf{v}'|) = D_R|\mathbf{v} - \mathbf{v}'|$ ($D_R > 0$): Also in this case, condition (i) is not satisfied as follows from the shape of $g_R^0(v,v')$, which now reads as

$$g_R^0(v,v') = \tfrac{2}{3}D_R\,[(v'^2/v + 3v)U(v - v') + (v^2/v' + 3v')U(v' - v)]. \tag{32}$$

D. Existence of a solution to Eq. (16)

In this section, we shall give a sufficient condition for the existence of a solution to Eq. (16) via a simple application of the contracting mapping principle. Introducing the operator A defined as

$$(A\varphi)(v) = -[1/2N\hat{g}_R(v)]\varphi(v) \times \int_0^\infty g_R^0(v,v')\varphi(v')\,dv' + Q_0 S_0(v)/N\hat{g}_R(v), \tag{33}$$

Eq. (16) is rewritten for $\hat{g}_R(v) > 0$, in the operational form

$$f_0 = Af_0. \tag{34}$$

Since the integral of f_0 with respect to v represents a finite physical quantity, namely, the total density of the t.p. [compare Eq. (26a)], it seems appropriate to refer to the Banach space $L_1(0,\infty)$. There results then, for $\varphi,\psi \in L_1(0,\infty)$,

$$|A\varphi| \leqslant \frac{Q_0}{N}\frac{S_0(v)}{\hat{g}_R(v)} + \frac{1}{2N\hat{g}_R(v)}|\varphi(v)| \int_0^\infty g_R^0(v,v')|\varphi(v')|\,dv' \tag{35a}$$

and

$$|A\varphi - A\psi| \leqslant \frac{1}{2N\hat{g}_R(v)}\Big[|\psi(v) - \varphi(v)| \int_0^\infty g_R^0(v,v')|\psi(v')|\,dv' + |\varphi(v)| \int_0^\infty g_R^0(v,v')|\psi(v') - \varphi(v')|dv'\Big]. \tag{35b}$$

We shall make the following assumptions:

(i) $S_0(v)/\hat{g}_R(v) \in L_1(0,\infty)$, namely,

$$\int_0^\infty [S_0(v)/\hat{g}_R(v)]\,dv = \beta < \infty; \tag{36a}$$

(ii) there exists a positive constant γ such that

$$g_R^0(v,v')/\hat{g}_R(v) \leqslant \gamma \tag{36b}$$

for any $v \in (0,\infty)$ and $v' \in (0,\infty)$.

Consequently, we verify that

$$\|A\varphi\| \leqslant \beta Q_0/N + (\gamma/2N)\|\varphi\|^2 \tag{37a}$$

and

$$\|A\varphi - A\psi\| \leqslant (\gamma/2N)(\|\varphi\| + \|\psi\|)\|\varphi - \psi\|. \tag{37b}$$

Introducing the ball B_ρ of $L_1(0,\infty)$ with center at the origin and radius ρ, we easily realize that A maps B_ρ into itself if

$$\beta Q_0/N + \gamma\rho^2/2N \leqslant \rho \tag{38a}$$

and is a contraction there if

$$\gamma\rho/N < 1. \tag{38b}$$

Values of ρ for which both conditions are fulfilled do exist provided

$$Q_0 < N^2/2\beta\gamma \tag{39}$$

and are explicitly given by

$$(N/\gamma)[1 - (1 - 2\beta\gamma Q_0/N)^{1/2}] \leqslant \rho < N/\gamma. \tag{40}$$

The contracting mapping principle guarantees thus that Eq. (34), under the assumptions Eqs. (36) and (39), has a unique solution in any ball B_ρ of $L_1(0,\infty)$ satisfying Eq. (40).[13] Such a solution is the limit in the L_1- norm of the successive approximation scheme ($n = 1,2,\cdots$)

$$f_n = A f_{n-1}, \quad f_0 = 0. \tag{41}$$

As a final remark it is worth noticing that the condition Eq. (36b), which replaces the more classical one related to Eq. (20) [condition (i) of the preceding section], is actually fulfilled in some cases of practical interest, in which the latter is not. It is trivially true for the case of constant collision frequencies and turns out to be in order also for $g_R(x) = B_R\, x^{-1}$ [see case (b) of the preceding section], when also $\hat{g}_R(x) = \hat{B}_R\, x^{-1}$, since then $\gamma = 2B_R\, \hat{B}_R^{-1} < \infty$.

[1] F. G. Tricomi, *Integral Equations* (Interscience, New York, 1957).
[2] P. M. Fitzpatrik and W. V. Petryshyn, Trans. Am. Math. Soc. **238**, 321 (1978).
[3] H. R. Thieme, Manuscripta Math. **29**, 49 (1979).
[4] A. Hammerstein, Acta Math. **54**, 117 (1930).
[5] M. M. R. Williams, *Mathematical Methods in Particle Transport Theory* (Butterworths, London, 1971).
[6] C. Cercignani, *Theory and Application of the Boltzmann Equation* (Scottish Academic, Edinburgh, 1975).
[7] C. Truesdell and R. G. Muncaster, *Fundamentals of Maxwell's Kinetic Theory of a Simple Monoatomic Gas* (Academic, New York, 1980).
[8] P. F. Zweifel, "The Boltzmann Equation and Its Properties," in *Lecture Notes on "Kinetic Theory and Boltzmann Equation,"* CIME Summer Course, Montecatini, 10–18 June 1981 (to be published).
[9] A. Palczewski, "The Cauchy Problem for the Boltzman Equation: A Survey of Recent Results," in *Lecture Notes on "Kinetic Theory and Boltzmann Equation,"* CIME Summer Course, Montecatini, 10–18 June 1981 (to be published).
[10] M. Haissinsky, *The Chemical and Biological Action of Radiations* (Academic, London, 1961).
[11] E. Segre, *Nuclei and Particles: an Introduction to Nuclear and Subnuclear Physics* (Benjamin, New York, 1964).
[12] M. Abramowitz and I. A. Stegun, ed., *Handbook of Mathematical Functions* (Dover, New York, 1965).
[13] M. A. Krasnosel'skii *et al.*, *Approximate Solutions of Operator Equations* (Wolters–Noordhoff, Groeningen, 1972).

Two-dimensional, steady-state diffusion of a passive scalar in a flow field[a)]

Leaf Turner
Los Alamos National Laboratory, University of California, Los Alamos, New Mexico 87545
Carl M. Bender
Department of Physics, Washington University, St. Louis, Missouri 63130

(Received 26 October 1982; accepted for publication 6 January 1983)

The two-dimensional, steady-state equation governing the diffusion of a passive scalar field Ψ, $\nabla\cdot(\mathbf{u}\Psi)=\eta\nabla^2\Psi$, is shown to yield a set of analytic solutions for η-independent flows $\mathbf{u}$. For the case of vorticity flow, this equation can be reduced to an equation solved by Darboux. For the case of potential flow, a special class of solutions is found, of which that demonstrated by Matthaeus and Montgomery is a special case.

PACS numbers: 05.60. + w, 03.70. + k, 02.30.Jr, 47.30. + s

The dynamics of the two-dimensional diffusion of an incompressibly convected passive scalar field $\Psi(x,y)$ is governed by

$$\frac{D\Psi}{Dt}\equiv\frac{\partial\Psi}{\partial t}+\mathbf{u}\cdot\nabla\Psi=\frac{\partial\Psi}{\partial t}+\nabla\cdot(\mathbf{u}\Psi)=\eta\nabla^2\Psi, \quad (1)$$

where the two-dimensional flow $\mathbf{u}$ is given by

$$\mathbf{u}(x,y)=u_1(x,y)\hat{x}+u_2(x,y)\hat{y}, \quad \nabla\cdot\mathbf{u}=0,$$

and where η is the diffusivity. Equation (1) describes not only the diffusion of a scalar quantity such as a chemical concentration or a temperature field, but also implicitly the diffusion of the two-dimensional (vector) magnetic field defined as the curl of $\Psi(x,y)\hat{z}$. We are using $\hat{x}$, $\hat{y}$, and $\hat{z}$ to denote the unit vectors directed along the positive x, y, and z axes, respectively.

In this paper, we shall show how to generate infinite sets of analytic solutions to the steady-state counterpart of Eq. (1).

$$\mathbf{u}\cdot\nabla\Psi=\eta\nabla^2\Psi, \quad (2)$$

where $\mathbf{u}$ is assumed to be an η-independent flow.[1]

First we shall consider the more intricate analysis required for vorticity flow (for which $\nabla\times\mathbf{u}$ is nonzero), leaving the treatment of potential flow (where $\mathbf{u}=\nabla\Omega$) until later.

For the vorticity flow case, we shall assume a WKB form for the scalar field[2]:

$$\Psi(x,y)=T(x,y)\exp[B(x,y)/\eta], \quad (3)$$

where T and B are η-independent functions. If we insert Eq. (3) into Eq. (2) and segregate the coefficients of like powers of η, we find[3]:

$$\mathbf{u}\cdot\nabla B=(\nabla B)^2, \quad (4a)$$

$$\mathbf{u}\cdot\nabla T=T\nabla^2B+2\nabla T\cdot\Delta B, \quad (4b)$$

$$\nabla^2T=0. \quad (4c)$$

We shall focus on the nontrivial situation where both ∇T and ∇B are nonzero. We shall regard T as a prescribed field satisfying Eq. (4c). Because the flow is solenoidal (incompressible), we find that Eq. (4a) implies

$$\nabla\Phi(x,y)=\hat{z}\times\nabla B(x,y)-\chi(x,y)\nabla B(x,y), \quad (5)$$

where Φ is the stream function of the flow; i.e.,

$$\mathbf{u}=\nabla\times\hat{z}\Phi=\nabla\Phi\times\hat{z}, \quad (6)$$

and where χ is some function which, according to Eq. (5), must satisfy

$$\nabla^2B=\frac{\partial\chi\partial B}{\partial x\partial y}-\frac{\partial\chi\partial B}{\partial y\partial x}\equiv\frac{\partial(\chi,B)}{\partial(x,y)}. \quad (7)$$

We note next that Eqs. (4b), (4c), and (6) imply that

$$\nabla\cdot[\nabla(BT)]=-\nabla\cdot(\hat{z}\times T\nabla\Phi),$$

which can be integrated readily to yield

$$\nabla\Phi=\frac{1}{T}\hat{z}\times\nabla(BT)-\frac{\nabla\Xi}{T}, \quad (8)$$

where $\Xi(x,y)$ is a function yet to be determined. A comparison of Eq. (5) with Eq. (8) yields

$$\hat{z}\times\nabla T=\frac{\nabla\Xi}{B}-\chi T\frac{\nabla B}{B}. \quad (9)$$

Note that Eq. (4c) signifies that there exists a function $\tilde{T}(x,y)$ related to $T(x,y)$ by the Cauchy–Riemann equation:

$$\nabla\tilde{T}\equiv\hat{z}\times\nabla T. \quad (10)$$

Thus $T+i\tilde{T}$ is an analytic function of $x+iy$. Therefore $\tilde{T}$, of course, also satisfies

$$\nabla^2\tilde{T}=0.$$

Taking the curl of Eq. (9) leads us to

$$\nabla(\chi-BT\chi)\times\nabla B=0,$$

or, equivalently, to

$$\chi=BT\chi+\int^B xH'(x)dx$$

for an arbitrary function $H(x)$. Inserting this expression back into Eq. (9) and noting Eq. (10), we obtain

$$\nabla\tilde{T}=\nabla(\chi\,\mathrm{T})+\nabla\mathrm{H(B)}$$

or, equivalently,

$$\chi(x,y)=\frac{\tilde{T}(x,y)-H(B(x,y))}{T(x,y)}. \quad (11)$$

Application of Eq. (7) leads us to a nonlinear patial differential equation for B, where $T(x,y)$ [and hence $\tilde{T}(x,y)$] and $H(B)$ are regarded as prescribed functions. Because $\nabla T\neq0$, we can simplify the further analysis by shifting independent var-

a) Work was performed under the auspices of the U.S. Department of Energy.

 0022-2488/83/061630-03$02.50 © 1983 American Institute of Physics

iables from x and y to T and $\tilde{T}$[4]:

$$\nabla_T^2 B = -\frac{1}{T^2}\left\{T\frac{\partial B}{\partial T} + [\tilde{T} - H(B)]\frac{\partial B}{\partial \tilde{T}}\right\}, \tag{12}$$

where

$$\nabla_T^2 \equiv \frac{\partial^2}{\partial T^2} + \frac{\partial^2}{\partial \tilde{T}^2}.$$

We shall henceforth restrict our analysis to the linear case by prescribing that $H(B) \equiv 0$.

The structure of the resulting linear differential equation motivates a second change to a new set of independent variables, z and θ, where

$$z = \ln(r)$$

and

$$T = r\cos\theta,$$

$$\tilde{T} = r\sin\theta.$$

We thereby obtain the equation

$$\frac{\partial^2 B}{\partial z^2} + \frac{\partial^2 B}{\partial \theta^2} + \frac{1}{\cos^2\theta}\frac{\partial B}{\partial z} = 0.$$

This equation has separable solutions of the form

$$B = b(\theta)\exp(-\alpha z) = b(\theta)/r^{\alpha}, \tag{13}$$

where $b(\theta)$ satisfies the equation

$$\cos^2\theta\, b''(\theta) + (\alpha^2\cos^2\theta - \alpha)b(\theta) = 0. \tag{14}$$

For the case when $\alpha = n(n-1)$, this equation was found by Darboux to have the general solution[5]:

$$b_n(\theta) = \cos^n\theta\left(\frac{1}{\cos\theta}\frac{d}{d\theta}\right)^n\{c_1\sin[n(n-1)\theta] + c_2\cos[n(n-1)\theta]\}, \tag{15}$$

where c_1 and x_2 are arbitrary constants.

We note that Eq. (11) yields

$$\chi = \tilde{T}/T.$$

Equation (5) then can be rewritten in the form

$$\frac{\partial \Phi}{\partial z} = -\tan\theta\frac{\partial B}{\partial z} - \frac{\partial B}{\partial \theta}, \tag{16a}$$

$$\frac{\partial \Phi}{\partial \theta} = \frac{\partial B}{\partial z} - \tan\theta\frac{\partial B}{\partial \theta}. \tag{16b}$$

One can solve Eqs. (16) for $\partial B/\partial z$ and $\partial B/\partial\theta$. If one equates

$$\frac{\partial}{\partial\theta}\left(\frac{\partial B}{\partial z}\right) \text{ to } \frac{\partial}{\partial z}\left(\frac{\partial B}{\partial \theta}\right),$$

one obtains the partial differential equation satisfied by Φ:

$$\cos^2\theta\left(\frac{\partial^2\Phi}{\partial\theta^2} + \frac{\partial^2\Phi}{\partial z^2}\right) - \cos 2\theta\frac{\partial\Phi}{\partial z} - \sin 2\theta\frac{\partial\Phi}{\partial\theta} = 0.$$

For the separable case,

$$\Phi = \phi(\theta)\exp(-\alpha z), \tag{17}$$

ϕ satisfies the equation

$$\frac{d^2\phi}{dy^2} + \left[\frac{\alpha^2 + 2\alpha}{(1+y^2)^2} - \frac{\alpha}{1+y^2}\right]\phi = 0, \tag{18}$$

where $y = \tan\theta$. Using Eqs. (13), (15), (16a), and (17), we extract the solution of Eq. (18) when $\alpha = n(n-1)$:

$$\phi_n(\theta) = -\tan\theta\, b_n(\theta) + \frac{1}{n(n-1)}\frac{db_n(\theta)}{d\theta}, \tag{19}$$

which obviously satisfies Eq. (16b) by virtue of Eq. (14).

We have thus developed an infinite number of sets of analytic solutions to the steady-state diffusion equation with (nonzero) vorticity flow.[6] For any $t(x, y)$ satisfying Eq. (4c), one uses the Cauchy–Riemann equation, Eq. (10), to derive a $\tilde{T}(x, y)$. For any value of n greater than 1, one can use Eqs. (13) and (15) to find a B and Eqs. (17) and (19) to determine the associated incompressible fluid flow.

For example, let us choose $n = 2$ and $c_1 = 0$. We find that

$$B = \frac{T^2}{(T^2 + \tilde{T}^2)^2}$$

and

$$\Phi = \frac{-2T\tilde{T}}{(T^2 + \tilde{T}^2)^2}.$$

Note that this flow happens to have zero vorticity. To demonstrate a particular case, we shall set

$$T = \frac{x}{x^2 + y^2} \text{ and } \tilde{T} = \frac{-y}{x^2 + y^2}.$$

We then find that the steady-state field

$$\Psi = \frac{x}{x^2 + y^2}\exp\left(\frac{x^2}{\eta}\right) \tag{20}$$

can be maintained in the presence of a diffusivity η by the prescribed flow

$$\mathbf{u} = 2(x\hat{x} - y\hat{y}). \tag{21}$$

Note that the actual steady-state field that is realized in a prescribed flow is not uniquely determined until the boundary conditions are given. We shall present below another example of a Ψ that can be maintained in the presence of the same diffusivity by the same prescribed flow.[7]

Let us now derive a set of solutions of eq. (2) for the case of potential flow; namely, for the case when

$$\mathbf{u}(x, y) = \nabla\Omega(x, y) \neq 0, \quad \nabla^2\Omega(x, y) = 0. \tag{22}$$

Without loss of generality, we look for nontrivial solutions $\Psi(x, y)$ of the form

$$\Psi(x, y) = \Lambda(x, y)\exp\left(\frac{\Omega(x, y)}{2\eta}\right). \tag{23}$$

We find that Eqs. (2) and (22) yield

$$\nabla^2\Lambda - (\Lambda/4\eta^2)(\nabla\Omega)^2 = 0. \tag{24}$$

Equation (22) allows us to define $\tilde{\Omega}(x, y)$ by the Cauchy–Riemann equation

$$\tilde{\Omega}(x, y) = \hat{z}\times\nabla\Omega(x, y). \tag{25}$$

Observe that $\Omega(x, y) + i\tilde{\Omega}(x, y)$ is an arbitrary analytic function of $x + iy$. We then can reexpress Eq. (24) using Ω and $\tilde{\Omega}$ as independent variables:

$$\nabla_{\Omega}^2\Lambda = \Lambda/4\eta^2, \tag{26}$$

where

$$\nabla_{\Omega}^2 \equiv \frac{\partial^2}{\partial\Omega^2} + \frac{\partial^2}{\partial\tilde{\Omega}^2}.$$

Equation (26) is readily solvable by means of the standard technique of separation of variables.

As an example of a solution, we cite

$$\Lambda(x,y)=\sin\left[\frac{(4\gamma^2-1)^{1/2}}{2\eta}\Omega(x,y)\right]\times\cosh\left[\frac{\gamma}{\eta}\tilde{\Omega}(x,y)\right]. \tag{27}$$

By use of Eqs. (22)–(27), we note that not only the field of Eq. (20), but also the field

$$\Psi=\sin\left[\frac{(4\gamma^2-1)^{1/2}}{2\eta}(x^2-y^2)\right]\times\cosh\left[\frac{2\gamma}{\eta}xy\right]\exp\left[\frac{x^2-y^2}{2\eta}\right]$$

can be maintained in the presence of a diffusivity η by the prescribed flow of Eq. (21). This particular solution has already been noted by Matthaeus and Montgomery.[8]

ACKNOWLEDGMENT

One of us (C.M.B.) would like to thank the Los Alamos National Laboratory for its hospitality.

[1]We originally became interested in this equation because of a remark by D. Montgomery that the reasoning given in H. K. Moffatt, *Magnetic Field Generation in Electrically Conducting Fluids* (Cambridge U. P., Cambridge, 1978), pp. 62–63 [that in the limit as $\eta\to0^+$, the solution Ψ of Eq. (2) vanishes within a domain bounded by a streamline] is inadequate because it assumes that, for small η, Ψ has a regular (Taylor-like) perturbation series in powers of η. We examined this equation further and using a WKB argument found the rather uninteresting result that Moffatt's conclusion is indeed correct. The WKB argument proceeds as follows: Let $\Psi(x,y)\sim\exp\{\eta^{\alpha}\Sigma_{k=0}^{\infty}\eta^{k/2}S_k(x,y)\}$ $(\eta\to0^+)$. The relevant distinguished limits are $\alpha=0$ and $\alpha=-1$. (There are special solutions with $\alpha=-\frac{1}{2}$, but these are treated in the same manner as the solutions for $\alpha=-1$.) For the case $\alpha=-1$, S_0 satisfies $\mathbf{u}\cdot\nabla S_0=(\nabla S_0)^2$. Integrating this equation over the region bounded by a closed streamline and assuming that S_0 is real and smooth within the region imply that $\nabla S_0=0$ and, consequently, that S_0 is a constant. (If S_0 is complex, the conclusion that S_0 vanishes remains valid, following from the condition that the solution Ψ not have arbitrarily short wavelength in the limit $\eta\to0^+$.) Thus, this case reduces to the distinguished limit $\alpha=0$ for which Ψ does indeed have a regular perturbation series in powers of η. While this result is not especially exciting, we discovered in the course of this investigation that the WKB ansatz is particularly useful for generating solutions of Eq. (2).

[2]C. M. Bender and S. A. Orszag, *Advanced Mathematical Methods for Scientists and Engineers* (McGraw-Hill, New York, 1978).

[3]This procedure is fruitless for the one-dimensional Schrödinger equation $\eta^2\Psi''(x)=Q(x)\Psi(x)$. If we demand that the first-order WKB (physical optics) approximation, $\Psi(x)=T(x)\exp[B(x)/\eta]$, be an exact solution, as we do for Eq. (2), the three equations analogous to those in Eq. (4) combine to give a condition that $Q(x)$ must satisfy in order that any solution exist at all. The result is that $Q(x)=(\gamma x+\delta)^{-4}$, where γ and δ are constants. In contrast, it is remarkable that this procedure applied to Eq. (2) does yield a rich spectrum of solutions.

[4]When an ordinary differential equation is autonomous (i.e., when the independent variable does not appear explicitly in the differential equation), the usual procedure is to transform the differential equation into one of lower order in which the new independent variable is the old dependent variable. When this technique is applied to partial differential equations, as we are doing in this paper, it is referred to as a hodograph transformation. the hodograph transformation has been used traditionally in the study of one-dimensional gas flow. See P. R. Garabedian, *Partial Differential Equations* (Wiley, New York, 1964).

[5]G. Darboux, *Lecons sur la Théorie* Génerale des Surfaces (Chelsea, New York, 1972), Vol. 2.

[6]Other WKB solutions of Eq. (2) that do not arise from solutions of the Darboux equation are $B=\arctan(\tilde{T}/T)$, $\Phi=\ln[T/(T^2+\tilde{T}^2)]$ and $B=\ln(T)$, $\Phi=\ln(\tilde{T}/T)$.

[7]Note also that the physics of Eq. (2) may cease to remain valid in the vicinity of singularities, such as that of Eq. (20). As a result, a solution of Eq. (2) may have to be patched onto a solution of a more dynamically detailed equation valid in the neighborhood of such singularities. We emphasize that the purpose of this paper has been to generate solutions of Eq. (2) without serious consideration of boundary conditions or singularities.

[8]W. H. Matthaeus and D. Montgomery, J. Plasma Phys. **25**, 11 (1981).

On systems of local operators and the duality condition

Eyvind H. Wichmann[a)]
Department of Physics, University of California, Berkeley, California 94720

(Received 7 September 1982; accepted for publication 3 December 1982)

Systems of von Neumann algebras of local operators are considered, and a particular procedure for the construction of such a system from any given set of intrinsically local operators is described. The condition of duality is discussed with particular emphasis on a special form of this condition which appears in previous work by Bisognano and Wichmann. Subsystems generated by any set of local operators in an original irreducible system of local operators are considered. Some properties of the irreducible vacuum representation of such a subsystem are derived assuming the special form of the duality condition for the original system. It is shown that this special condition also holds for the subsystem. It is furthermore shown that weak additivity holds for the reduced system if it holds for the original system.

PACS numbers: 11.10.Cd, 03.70. + k

I. INTRODUCTION

In this paper we shall be concerned with systems of von Neumann algebras of (bounded) "local" operators,[1] and in particular we shall consider subsystems generated by an arbitrary set of local operators.

The so-called duality condition for a single Hermitian scalar field was discussed, within the framework of standard quantum field theory, in a paper by Bisognano and Wichmann,[2] hereafter referred to as BW-I. This discussion was generalized to the case of an arbitrary number of finite-component quantum fields in a second paper by the same authors,[3] hereafter referred to as BW-II. The general question of whether it is always possible to find a system of local von Neumann algebras such that a given set of local quantum fields, subject to the usual assumptions of standard quantum field theory,[4] are locally affiliated with the von Neumann algebras was not solved in the quoted papers, and still remains an open question. It was shown, however, that if certain additional conditions are imposed on the (generally unbounded) quantum field operators, then a set of von Neumann algebras properly related to the fields exists. The additional conditions discussed by Bisognano and Wichmann are not as such unreasonable physically, but it is clearly highly desirable to settle the question of whether they are in fact necessary. The present study contributes nothing to the settlement of this question. We shall, however, base our discussion on a notion which seems to follow almost inevitably from the results in the quoted papers, namely the following. *If* an irreducible set of finite-component local quantum fields are locally affiliated with a system of local von Neumann algebras, *then* these von Neumann algebras satisfy what we shall here call the special condition of duality. A precise statement of this condition is given in Definition 3 in the next section, but we shall shortly present a more informal description in this introductory section. Throughout this paper the notation and terminology will closely follow BW-II, and we will depend on results from both BW-I and BW-II. The present study is, however, only concerned with *bounded* local operators, and *no* assumption is made about the possible existence of local quantum fields.

The general premises will be spelled out in some detail in the next section. They are, by and large, the usual ones in this context. In particular von Neumann algebras will be regarded as associated with certain special subsets of Minkowski spacetime. The subsets in question are double cones and their causal complements, and wedge regions bounded by two characteristic planes. All (open) regions of the latter kind are Poincaré-equivalent to the special wedge-regions W_R and W_L defined by

$$W_R = \{x|x^3 > |x^4|\}, \quad W_L = \{x|x^3 < -|x^4|\} \tag{1}$$

with reference to the customary parametrization of Minkowski space $\mathscr{M}$ in terms of Cartesian coordinates $x = (x^1,x^2,x^3,x^4)$, where x^4 is the time coordinate.

We consider a von Neumann algebra $\mathscr{A}(W_R)$, which we regard as associated with W_R, and a von Neumann algebra $\mathscr{A}(W_L)$, which we regard as associated with W_L. The precise sense of this "association" is specified in Definition 1 in the next section. The notion involves some physically motivated conditions on the relationship between $\mathscr{A}(W_R)$ and $\mathscr{A}(W_L)$ and the unitary operators which represent the Poincaré group, as well as a *condition of locality*. This latter condition says that if $X \in \mathscr{A}(W_R)$ and $Y \in \mathscr{A}(W_L)$, then X commutes with Y if either X or Y is a boson operator, whereas X and Y anticommute if both are fermion operators. The *condition of duality* is a (substantial) strengthening of this condition: It says that $\mathscr{A}(W_L)$ is precisely equal to the set of *all* operators Y which satisfy the above condition of locality relative to the operators in $\mathscr{A}(W_R)$.

In the theory of local von Neumann algebras the condition of duality is particularly interesting in the situation in which the vacuum vector Ω is cyclic and separating for $\mathscr{A}(W_R)$ [as well as for $\mathscr{A}(W_L)$]. The Tomita–Takesaki theory[5] then applies, and by Tomita's theorem the pair $(\mathscr{A}(W_R),\Omega)$ determines a unique antiunitary involution J, the *modular conjugation*, and a unique one-parameter unitary group $\{V(t)|t \in R^1\}$, the *modular automorphism group*, which among other interesting properties satisfy the relations $J\mathscr{A}(W_R)J = \mathscr{A}(W_R)'$ and $V(t)\mathscr{A}(W_R)V(t)^{-1}$

[a)] Work supported in part by the National Science Foundation under Grant No. PHY-79-23251.

 0022-2488/83/061633-12$02.50 © 1983 American Institute of Physics

$= \mathcal{A}(W_R)$. Here and in the following, $\mathcal{A}'$ denotes the commutant of the von Neumann algebra $\mathcal{A}$.

Our special condition of duality is the condition of duality with the *additional* condition that the modular automorphism group $\{V(t)\}$ is the group of all velocity transformations in the 3-direction (within the Lorentz group), and that the modular conjugation J is related in a certain specific manner to a *TCP* operator Θ. In the absence of fermion operators, J is simply the product of Θ and a rotation by angle π about the 3-axis. The adjective "special" is employed because the author can imagine local systems of von Neumann algebras which satisfy the condition of duality but fail to satisfy the special condition. Just how special or general the special condition really is, is not very clear to the author at this time. In spite of this, it seems meaningful and interesting to consider consequences of this condition, because in view of the results in BW-I and BW-II it emerges as a *necessary* condition which an irreducible system of local von Neumann algebras must satisfy if there is to exist an irreducible set of local finite-component quantum fields locally affiliated with the algebras.

The nature of the systems of local algebras under consideration is specified precisely in Definition 1 in the next section, and further explained in Theorem 1. The idea is exactly the same as in BW-I and BW-II, namely, the entire system of algebras is determined by the particular algebra $\mathcal{A}(W_R)$. The term "*AB*-system" is used as a technical term to describe such a system, as it was used in BW-II. Definition 2 represents an attempt to formula an adequate definition of an arbitrary set $\mathcal{S}$ of local and relatively local operators. In Theorem 2 it is shown how such a set $\mathcal{S}$ defines an *AB*-system of local von Neumann algebras, in terms of a specific construction based on geometrical notions. If an operator $X \in \mathcal{S}$ is locally associated with a double cone K, then $U(\lambda)XU(\lambda)^{-1}$ is locally associated with W_R for all Poincaré transformations λ such that the image K_λ of K under λ is contained in W_R. In this manner each operator in $\mathcal{S}$ defines a set of operators locally associated with W_R, and the algebra $\mathcal{A}_s(W_R)$ is defined as the von Neumann algebra generated by the union of all such sets. By Theorem 1, this, in effect, defines a local *AB*-system. To the set $\mathcal{S}$ corresponds a sub-Hilbert space $\mathcal{H}_s$ of $\mathcal{H}$, defined as the smallest Poincaré-invariant subspace which contains the vacuum vector Ω and which is mapped into itself by the elements of $\mathcal{S}$. The restriction $\hat{\mathcal{A}}_s(W_R)$ of $\mathcal{A}_s(W_R)$ to the subspace $\mathcal{H}_s$ is a von Neumann algebra which in view of Theorem 1 gives rise to an irreducible *AB*-system of local algebras (on $\mathcal{H}_s$). This is also discussed in Theorem 2. The considerations are quite elementary, and do not involve the duality condition.

The special duality condition is defined in Definition 3. In Theorem 3 a criterion is presented, with reference to the special condition of duality, for an *AB*-system to be generated by its strictly local operators. This is a minor extension of Theorem 6 in BW-II.

Theorem 4 is concerned with the *AB*-systems generated by any subset $\mathcal{S}$ of the set of all strictly local operators of a given *AB*-system. It is assumed that the original *AB*-system is generated by its local operators and that it satisfies the special condition of duality, in which case quite detailed conclusions can be drawn about the sub-*AB*-system. As an informal summary one can say that the restriction of the subsystem to the subspace $\mathcal{H}_s$ has the *same* properties as the original system: In particular, it is *TCP*-covariant, and it satisfies the special condition of duality. Furthermore, it satisfies the so-called condition of *weak additivity* if the original system satisfies this condition.

Proofs are given in Sec. III. In conclusion some further general comments are made in Sec. IV, about local internal symmetries and related subjects.

II. SURVEY OF PREMISES AND RESULTS

We shall first review some simple geometrical notions which appeared in BW-I and BW-II, and which are relevant to our discussion. The Lorentz "metric" is so defined that $x \cdot x$ is positive if x is timelike. For any subset R of Minkowski space $\mathcal{M}$ we define the *causal complement* R^c of R by

$$R^c = \{x | x \in \mathcal{M},\ (x-y)\cdot(x-y) < 0 \quad \text{for all } y \in R\} \tag{2}$$

and (with the obvious symbolism) we then have $R \subset R^{cc}$, $R^{ccc} = R^c$.

The open forward light cone with x as apex is denoted $V_+(x)$, and the open backward light cone with x as apex is denoted $V_-(x)$. If $y \in V_+(x)$ we define the *open double cone* $C(x,y)$ by

$$C(x,y) = V_+(x) \cap V_-(y). \tag{3}$$

The set of all $C(x,y)$ was denoted $\mathcal{D}_c$ in BW-I and BW-II. We shall here define the set $\mathcal{D}$ *as the set of all closed double cones with a nonempty interior*, i.e.,

$$\mathcal{D} = \{\overline{C} | C \in \mathcal{D}_c\}. \tag{4}$$

It turns out to be convenient to exclude the degenerate closed double cones, which are either single points or else lightlike line segments, and this is the reason for the above definition of $\mathcal{D}$. We note here that if $K \in \mathcal{D}$, then K^c is open and $K = K^{cc}$.

Let π' and π'' be two nonparallel characteristic 3-planes: Their normals are thus lightlike. The wedge region $W = W(\pi', \pi'')$ consisting of all points "below" π' which are "above" π'' is then Poincaré-equivalent to either one of the special wedges W_R and W_L in (1). As in BW-I, we define $\mathcal{W}$ *as the set of all such open wedges* W. For any $W \in \mathcal{W}$ the causal complement W^c is the closure of an element in $\mathcal{W}$, and in particular we have $W_R{}^c = \overline{W}_L$.

Among the simple geometrical relationships between wedges and double cones we note the following two, which are of crucial importane for the theory. If $K \in \mathcal{D}$, then

$$K = \cap\{W | W \in \mathcal{W},\ W \supset K\} \tag{5a}$$

and

$$K^c = \cup\{W | W \in \mathcal{W},\ W \subset K^c\} = \cup\{W | W \in \mathcal{W},\ \overline{W}^c \supset K\}. \tag{5b}$$

The Hilbert space of physical states is separable, and it carries a strongly continuous unitary representation $\lambda \to U(\lambda)$ of the universal covering group $\bar{\mathcal{P}}$ of the Poincaré group $\overline{L}_0$. The elements of $\bar{\mathcal{P}}$ are ordered pairs (g,x), where x is any element of $\mathcal{M}$, and where g is any element of $\mathcal{P} = \mathrm{SL}(2,C)$, regarded as the universal covering group of the

homogeneous Lorentz group L_0. The usual "spectral condition" is assumed: The momentum spectrum associated with the spectral resolution of the translation subgroup $x \to T(x) = U(I,x)$ is confined to the closed forward light cone $\overline{V}_+$, and there exists a vacuum (unit) vector Ω uniquely determined (up to a multiple) by the condition $U(\lambda)\Omega = \Omega$. As in BW-II, we define

$$U_0 = U(-I,0), \quad Z = (I + iU_0)/(1 + i) \tag{6}$$

and U_0 thus represents a rotation by angle 2π about any axis. We trivially have $U_0^2 = I$, $Z^2 = U_0$, and $ZU(\lambda) = U(\lambda)Z$ for any $\lambda \in \bar{\mathcal{G}}$. The operator Z is introduced because it permits a compact statement of the usual conditions of locality (reflecting the *normal* connection between spin and statistics). For a full discussion if this device, see Sec. 3. in BW-II. For any operator X and any set of operators $\mathcal{S}$, we employ the notation

$$X^z = ZXZ^{-1}, \quad \mathcal{S}^z = \{X^z | X \in \mathcal{S}\} \tag{7}$$

The locality condition can then by stated as follows. If X and Y are two local operators associated with spacelike separated regions, then $[X, Y^z] = 0$.

The one-parameter subgroup of velocity transformations in the 3-direction will play a special role, and hence we introduce a special notation for the corresponding representatives:

$$U(\exp(t\sigma_3/2),0) = V(t). \tag{8}$$

Specifically, $V(t)$ thus represents the velocity transformation which in its action on the rest state of a particle leads to a state in which the particle has the velocity $tgh\,(t)$ in the 3-direction. The representation $t \to V(t)$ of the real line can be extended to a representation of the complex plane, on a dense domain of analytic vectors. In our discussion the operator $V(i\pi)$ will play a role: It is a strictly positive self-adjoint operator on a certain domain D_+.

An antiunitary operator Θ which satisfies the condition

$$\Theta U(g,x)\Theta^{-1} = U(g,-x), \quad \text{for all } (g,x) \in \bar{\mathcal{G}}, \tag{9}$$

is a possible candidate for a *TCP* transformation, and if this condition holds we shall say that Θ *is a TCP operator relative to* $U(\bar{\mathcal{G}})$. The above condition on Θ is not very restrictive, and it is thus merely a necessary condition which the *TCP* operator in a physical theory must satisfy. In a theory based on the notion of local observables it is also required that the *TCP* operator act on the local observables (by conjugation) in a certain specific manner, and this latter condition is, as is well known, very restrictive.

The geometrical interpretation of a theory of local operators is based on the well-known homomorphism of $\bar{\mathcal{G}}$ onto the Poincaré group $\overline{L}_0$. To every $\lambda \in \bar{\mathcal{G}}$ thus corresponds an element of $\overline{L}_0$ interpreted as a point transformation of physical Minkowski space. If R is any subset of $\mathcal{M}$, we denote by R_λ the image of R under the Poincaré transformation corresponding to λ. We follow the standard conventions, and the homogeneous subgroup $\{(g,0) | g \in \mathcal{G}\}$ of $\bar{\mathcal{G}}$ thus corresponds precisely to the Poincaré transformations which leave the origin fixed. For any subset R of $\mathcal{M}$ we also employ the notation

$$R_\Theta = \{-x | x \in R\} \tag{10}$$

and the mapping $R \to R_\Theta$ is thus a geometric *TCP* transformation (with respect to the origin).

We shall now specify the nature of the systems of von Neumann algebras with which we shall be concerned.

Definition 1: (1) Let there correspond, to each $W \in \mathcal{W}$, a von Neumann algebra $\mathcal{A}(W)$, and let there correspond, to each $K \in \mathcal{D}$, two von Neumann algebras, denoted $\mathcal{B}(K)$ and $\mathcal{A}(K^c)$. The collection $(AB) = \{\mathcal{A}(W), \mathcal{B}(K), \mathcal{A}(K^c)\}$ shall be called a *local Poincaré-covariant AB-system* (of von Neumann algebras) if and only if the following conditions hold:

$$\text{(a)} \quad \mathcal{B}(K) = \cap\{\mathcal{A}(W) | W \in \mathcal{W}, \quad W \supset K\}, \tag{11a}$$

$$\mathcal{A}(K^c) = \{\mathcal{A}(W) | W \in \mathcal{W}, \quad W \subset K^c\}'' \tag{11b}$$

for all $K \in \mathcal{D}$. Furthermore,

$$\mathcal{A}(W) \supset \mathcal{A}(W_i) \quad \text{whenever} \quad W \supset W_i, \tag{12a}$$

$$\mathcal{B}(K) \supset \mathcal{B}(K_i) \quad \text{whenever} \quad K \supset K_i, \tag{12b}$$

$$\text{(b)} \quad U(\lambda)\mathcal{A}(W)U(\lambda)^{-1} = \mathcal{A}(W_\lambda), \tag{13a}$$

$$U(\lambda)\mathcal{B}(K)U(\lambda)^{-1} = \mathcal{B}(K_\lambda),$$

$$U(\lambda)\mathcal{A}(K^c)U(\lambda)^{-1} = \mathcal{A}(K_\lambda{}^c) \tag{13b}$$

for all $\lambda \in \bar{\mathcal{G}}$, all $K \in \mathcal{D}$, and $W \in \mathcal{W}$.

$$\text{(c)} \quad \mathcal{A}(W)^z \subset \mathcal{A}(\overline{W}^c)', \quad \mathcal{B}(K)^z \subset \mathcal{A}(K^c)',$$

$$\mathcal{A}(K^c)^z \subset \mathcal{B}(K)'. \tag{14}$$

(2) For such an AB-system we define a von Neumann algebra $\mathcal{A}(\mathcal{M})$ and a $*$-algebra $\mathcal{U}$ by

$$\mathcal{A}(\mathcal{M}) = \{\mathcal{A}(W) | W \in \mathcal{W}\}'', \quad \mathcal{U} = \cup\{\mathcal{B}(K) | K \in \mathcal{D}\}, \tag{15}$$

and $\mathcal{U}$ shall be called the *set of all local operators of the systems.*

(3) An AB-system shall be said to be *TCP-covariant* if and only if there exists an antiunitary operator Θ which satisfies the condition (9) *and* the conditions

$$\Theta\mathcal{A}(W)\Theta^{-1} = \mathcal{A}(W_\Theta), \quad \Theta\mathcal{B}(K)\Theta^{-1} = \mathcal{B}(K_\Theta),$$

$$\Theta\mathcal{A}(K^c)\Theta^{-1} = \mathcal{A}(K_\Theta{}^c) \tag{16}$$

for all $K \in \mathcal{D}$, $W \in \mathcal{W}$.

(4) An AB-system shall be said to satisfy the *condition of duality* if and only if equality holds in (14), i.e.,

$$\mathcal{A}(W)^z = \mathcal{A}(\overline{W}^c)', \quad \mathcal{B}(K)^z = \mathcal{A}(K^c)', \quad \mathcal{A}(K^c)^z = \mathcal{B}(K)' \tag{17}$$

for all $K \in \mathcal{D}$, $W \in \mathcal{W}$.

This definition of an AB-system is equivalent to the definition given in Sec. 6 of BW-II, except that we have here included a number of statements which appeared as conclusions in Theorem 5 in BW-II. The author felt that it would be useful to list in this manner all the properties of an AB-system in which we shall be interested, although these properties are not logically independent. The interdependence is partly shown by the following:

Theorem 1: (1) Let $\mathcal{A}$ be a von Neumann algebra which satisfies the conditions

$$\text{(a)} \quad U(\lambda)\mathcal{A}U(\lambda)^{-1} \subset \mathcal{A}, \tag{18}$$

for all $\lambda \in \bar{\mathcal{G}}$ such that $W_{R,\lambda} \subset W_R$

with *equality* whenever $W_{R,\lambda} = W_R$.

(b) $$U_{\pi 1}\mathcal{A}^z U_{\pi 1}{}^{-1} \subset \mathcal{A}', \tag{19}$$

where $U_{\pi 1}$ represents a rotation by angle π about the 1-axis.

Then the algebra $U(\lambda)\mathcal{A}U(\lambda)^{-1}$ depends on λ only through the set $W = W_{R,\lambda} \in \mathcal{W}$; hence

$$\mathcal{A}(W_{R,\lambda}) \equiv U(\lambda)\mathcal{A}U(\lambda)^{-1}, \quad \text{all } \lambda \in \bar{\mathcal{G}}, \tag{20}$$

defines a unique von Neumann algebra $\mathcal{A}(W)$ for every $W \in \mathcal{W}$, and, in particular, we have $\mathcal{A}(W_R) = \mathcal{A}$.

For each $K \in \mathcal{D}$, let the von Neumann algebras $\mathcal{B}(K)$ and $\mathcal{A}(K^c)$ be *defined* in terms of the algebras $\mathcal{A}(W)$ by (11a) and (11b) in Definition 1. Then the set of von Neumann algebras $(AB) = \{\mathcal{A}(W), \mathcal{B}(K), \mathcal{A}(K^c)\}$ is a local Poincaré-covariant AB -system. We shall say that the wedge algebra $\mathcal{A}$ *generates* this system.

(2) If in addition to the conditions in (1) the algebra $\mathcal{A}$ also satisfies the condition that there exists an antiunitary *TCP* operator Θ relative to $U(\bar{\mathcal{G}})$, i.e., Θ satisfies the condition (9), such that

$$\Theta\mathcal{A}\ \Theta^{-1} = U_{\pi 1}\mathcal{A}U_{\pi 1}{}^{-1}, \tag{21}$$

then the AB-system defined in (1) is *TCP*-covariant with Θ as the *TCP* operator.

(3) If in addition to the condition (1a) the algebra $\mathcal{A}$ also satisfies the condition that

$$U_{\pi 1}\mathcal{A}^z U_{\pi 1}{}^{-1} = \mathcal{A}', \tag{22}$$

then the AB-system generated by $\mathcal{A}$ satisfies the condition of duality.

The proof of this theorem follows at once from the discussion in Sec. 6 of BW-II. (The proof is in effect contained in Theorem 5 of the quoted paper.) This theorem is of considerable interest because it shows, first of all, that the single wedge algebra $\mathcal{A}(W_R)$ of an AB-system effectively determines the entire system. Secondly, we see that the duality condition and the condition of *TCP* covariance for the AB-system hold if and only if the corresponding conditions hold for the single wedge algebra $\mathcal{A}(W_R)$. Finally, the theorem shows how every von Neumann algebra which satisfies the premises of the theorem *defines* an AB-system in accordance with the construction in the theorem. This is of importance for a procedure which we shall discuss next, in which an AB-system is constructed from some given set of local operators. The strategy is to first construct a wedge algebra $\mathcal{A} = \mathcal{A}(W_R)$ which satisfies the premises of Theorem 1, after which the construction of the AB-system is completed on the basis of the theorem. The above remarks are also offered as an explanation of the unorthodox (and not very satisfactory) term "AB-system" in place of other terms, such as "local nets," which have been employed in the literature to describe local systems of algebras. We are here dealing with a very specific structure, as described in Definition 1, and we wish to avoid conflicts with notions or assumptions in the works of other authors.

We now consider a set $\mathcal{S}$ of bounded operators with the following properties: (a) If $X \in \mathcal{S}$, then $X^* \in \mathcal{S}$; (b) to every $X \in \mathcal{S}$ corresponds a $K(X) \in \mathcal{D}$, and the correspondence is such that $K(X) = K(X^*)$ and such that for any two $X_1, X_2 \in \mathcal{S}$,

$$[X_1{}^z, U(\lambda)X_2U(\lambda)^{-1}] = 0 \tag{23}$$

for all $\lambda \in \bar{\mathcal{G}}$ such that $K(X_1)^c \supset K(X_2)_\lambda$. It is then almost trivial that the von Neumann algebra $\mathcal{A}_s(W_R)$ defined by

$$\mathcal{A}_s(W_R) = \{U(\lambda)XU(\lambda)^{-1} | X \in \mathcal{S},\ \lambda \in \bar{\mathcal{G}},\ K(X)_\lambda \subset W_R\}'' \tag{24}$$

satisfies the premises in part (1) of Theorem 1, and hence $\mathcal{S}$ defines, in effect, a local Poincaré-covariant AB-system such that $X \in \mathcal{B}_s(K(X))$ for all $X \in \mathcal{S}$.

We can say that $\mathcal{S}$ is a (self-adjoint) set of local and relatively local operators. The conditions (23) are intrinsic in the sense that they do not depend on the prior existence of some local AB-system *outside* $\mathcal{S}$. The specific assignments $X \to K(X)$ should be regarded as essential for the characterization of $\mathcal{S}$ as a specific set of *local* operators. If we are given an arbitrary set $\mathcal{S}$ of bounded operators [satisfying the condition (a)], then $\mathcal{S}$ can be interpreted as a set of *local* operators if and only if *some* mapping $X \to K(X)$ of $\mathcal{S}$ into $\mathcal{D}$ is possible such that the conditions (b) hold. If such a mapping is possible, it is certainly not unique, because if for each X we choose an arbitrary $K_1(X)$ such that $K_1(X) \supset K(X)$, then the mapping $X \to K_1(X)$ also satisfies the conditions in (b). It could even happen that two different mappings exist such that for a particular X the double cones $K_1(X)$ and $K_2(X)$ are spacelike separated. The point of this is that the algebra $\mathcal{A}_s(W_R)$, and thereby the corresponding AB-system may, in general, depend on the specific mapping $X \to K(X)$.

If $\mathcal{S}$ is given as a subset of the set $\mathcal{U}$ of all local operators in an *already* explicitly specified local AB-system, then the conditions in (b) are satisfied for any mapping $X \to K(X)$ such that $X \in \mathcal{B}(K(X))$. If we now wish to construct a local AB-system from $\mathcal{S}$, the procedure described above is not satisfactory: We should base the construction on the fact that any $X \in \mathcal{S}$ is locally associated with a *set* of double cones.

We shall now try to characterize the kind of sets of local and relatively local operators which can serve as the "initial objects" in the construction of a local AB-system.

Definition 2: (1) A set $\mathcal{S}$ of bounded operators shall be called a *primary set of local operators* if and only if the following conditions hold:

(a) To each $K \in \mathcal{D}$ corresponds a subset $\mathcal{S}(K)$ of $\mathcal{S}$, such that

$$\mathcal{S} = \cup\{\mathcal{S}(K) | K \in \mathcal{D}\}. \tag{25}$$

(b) If $X \in \mathcal{S}(K)$, then $X^* \in \mathcal{S}(K)$.

(c) $$U(\lambda)\mathcal{S}(K_1)^z U(\lambda)^{-1} \subset \mathcal{S}(K_2)' \tag{26}$$

for all $K_1, K_2 \in \mathcal{D}$ and all $\lambda \in \bar{\mathcal{G}}$, such that $K_{1,\lambda} \subset K_2^c$.

(2) If $\mathcal{S}$ is a primary set of local operators, then $\mathcal{P}_s$ is defined as the *smallest* $*$-algebra which contains $\mathcal{S}$ and which is invariant under conjugation by all $U(\lambda)$, $\lambda \in \bar{\mathcal{G}}$. Differently stated: $\mathcal{P}_s$ is the set of all *finite* linear combinations (over the complex field) of all products of a *finite* number of elements in the set $\{U(\lambda)XU(\lambda)^{-1} | X \in \mathcal{S}, \lambda \in \bar{\mathcal{G}}\}$.

Concerning this definition, we note that it is *not* assumed that $\mathcal{S}$ itself is invariant under conjugation by elements of $U(\bar{\mathcal{G}})$, and hence $U(\lambda)\mathcal{S}(K)U(\lambda)^{-1}$ need not be equal to $\mathcal{S}(K_\lambda)$. It is also not assumed that $\mathcal{S}(K_i) \subset \mathcal{S}(K)$ for $K_i \subset K$. The locality condition (26) is, however, consistent with Poincaré covariance, and it is clear that we could con-

struct a larger primary set from $\mathscr{S}$ which would not have the above-mentioned "defects." In a sense the algebra, $\mathscr{P}_s$ is such a set, if we associate the elements of $\mathscr{P}_s$ with double cones in a natural and obvious manner. This algebra can thus be regarded as the "most obvious set of local operators generated by $\mathscr{S}$." Note also that $\mathscr{P}_s$ is simply an algebra: No topological notions are involved in its definition. Finally, we note that our definition is perfectly consistent with the not uninteresting special case in which $\mathscr{S}$ contains only one single self-adjoint operator X associated with one single double cone K_0, i.e., $\mathscr{S}(K)$ is empty if $K \neq K_0$, and $\mathscr{S}(K_0) = \{X\}$.

We shall now describe a specific construction of an AB-system from a given primary set of local operators.

Theorem 2: (1) Let $\mathscr{S} = \cup\{\mathscr{S}(K) | K \in \mathscr{D}\}$ be a primary set of local operators, with an associated $*$-algebra $\mathscr{P}_s$, in the sense of Definition 2. Let the von Neumann algebra $\mathscr{A}_s(W_R)$ be defined by

$$\mathscr{A}_s(W_R) = \{U(\lambda)\mathscr{S}(K)U(\lambda)^{-1} | K \in \mathscr{D},\ \lambda \in \bar{\mathscr{V}},\ K_\lambda \subset W_R\}''. \tag{27}$$

Then $\mathscr{A}_s(W_R)$ satisfies the premises in part (1) of Theorem 1, and, hence, by the construction in the Theorem, determines a local Poincaré-covariant AB-system $(AB)_s = \{\mathscr{A}_s(W), \mathscr{B}_s(K), \mathscr{A}_s(K^c)\}$ such that

$$\mathscr{S}(K) \subset \mathscr{B}_s(K), \quad \text{for all } K \in \mathscr{D}. \tag{28}$$

We shall say that $(AB)_s$ is *generated by* $\mathscr{S}$.

The AB-system $(AB)_s$ has the special property that all its algebras are generated by its *local* operators, i.e., for all K and W,

$$\mathscr{A}_s(W) = \{\mathscr{B}_s(K) | K \in \mathscr{D},\ K \subset W\}'', \tag{29a}$$

$$\mathscr{A}_s(K^c) = \{\mathscr{B}_s(K_i) | K_i \in \mathscr{D},\ K_i \subset K^c\}''. \tag{29b}$$

(2) Let $\mathscr{U}_s$ and $\mathscr{A}_s(\mathscr{M})$ be defined in terms of the AB-system $(AB)_s$ as in part (2) of Definition 1. Let the sub-Hilbert space $\mathscr{H}_s = F_s\mathscr{H}$ with the corresponding projection F_s be defined by

$$\mathscr{H}_s = \overline{\mathscr{A}_s(\mathscr{M})\Omega} \tag{30a}$$

where the overbar denotes strong closure.

Then:

$$\mathscr{S} \subset \mathscr{P}_s \subset \mathscr{U}_s \subset \mathscr{A}_s(\mathscr{M}) = \mathscr{P}_s'' \subset \{F_s\}', \tag{30b}$$

$$\overline{\mathscr{P}_s\Omega} = \overline{\mathscr{A}_s(W)\Omega} = \mathscr{A}_s(\mathscr{M})\Omega = \mathscr{H}_s, \quad \text{all } W \in \mathscr{W}. \tag{30c}$$

Furthermore,

$$F_s\Omega = \Omega, \quad U(\lambda)F_s = F_sU(\lambda) \quad \text{for all } \lambda \in \bar{\mathscr{V}} \tag{30d}$$

(3) For any bounded operator Q on $\mathscr{H}$, let $\rho(Q)$ denote the restriction of F_sQF_s to the subspace $\mathscr{H}_s$. Then $\lambda \to \rho(U(\lambda))$ is a strongly continuous unitary representation of $\bar{\mathscr{V}}$ on $\mathscr{H}_s$ which satisfies the same spectral conditions as $U(\bar{\mathscr{V}})$ with Ω (interpreted as a vector in $\mathscr{H}_s$) as the vacuum vector.

Let $\rho(\mathscr{S}) = \{\rho(X) | X \in \mathscr{S}\}$ and $\rho(\mathscr{S}(K)) = \{\rho(X) | X \in \mathscr{S}(K)\}$. Then the set $\rho(\mathscr{S}) = \cup\{\rho(\mathscr{S}(K)) | K \in \mathscr{D}\}$ is a primary set of local operators in the sense of Definition 2, and hence by part (1) of the present theorem generates a local, Poincaré-covariant AB-system $(AB)_{\rho s} = \{\hat{\mathscr{A}}_s(W), \hat{\mathscr{B}}_s(K), \hat{\mathscr{A}}_s(K^c)\}$ of von Neumann algebras on $\mathscr{H}_s$. It satisfies the condition

$$\hat{\mathscr{A}}_s(W) = \rho(\mathscr{A}_s(W)) \quad \text{for all } W \in \mathscr{W} \tag{31}$$

and the vector Ω is *cyclic and separating* for $\hat{\mathrm{A}}_s(W)$ (within the Hilbert space $\mathscr{H}_s$).

Furthermore,

$$\hat{\mathscr{A}}_s(\mathscr{M}) = \rho(\mathscr{A}_s(\mathscr{M})) = B(\mathscr{H}_s), \tag{32}$$

where $B(\mathscr{H}_s)$ is the algebra of all bounded operators on $\mathscr{H}_s$.

The proof of this theorem is quite simple, but, so as not to disrupt the discussion, we postpone it to the next section. The assertions are probably not at all surprising.

After these generalities we shall now consider the special condition of duality and its implications.

Definition 3: A local, *TCP*- and Poincaré-covariant AB-system $\{\mathscr{A}(W), \mathscr{B}(K), \mathscr{A}(K^c)\}$ of von Neumann algebras shall be said to satisfy the *special condition of duality* if and only if $\mathscr{A}(W_R)$ satisfies the following conditions (a), (b), and (c):

(a) The vector Ω is cyclic and separating for $\mathscr{A}(W_R)$.

(b) The linear manifold $\mathscr{A}(W_R)\Omega$ is a core for the self-adjoint operator $V(i\pi)$, and

$$JV(i\pi)X\Omega = X^*\Omega, \quad \text{for all } X \in \mathscr{A}(W_R). \tag{33}$$

Here J is an antiunitary involution related to the *TCP* operator Θ by

$$J = ZU_{\pi 3}\Theta, \tag{34}$$

where $U_{\pi 3}$ represents a rotation by angle π about the 3-axis and Z is defined by (6).

(c) $$J\mathscr{A}(W_R)J = \mathscr{A}(W_R)' = \mathscr{A}(W_L)^z. \tag{35}$$

An equivalent statement of these conditions is that relative to the cyclic and separating vector Ω the group $\{V(t) | t \in R^1\}$ is the modular automorphism group, and J is the modular conjugation for the algebra $\mathscr{A}(W_R)$.

We explained in the Introduction why the adjective "special" is used. Concerning the duality condition in general, within the framework of Definition 1, it might perhaps at first seem that given any AB-system the duality property could be arrived at by the cheap device of redefining the algebras $\mathscr{A}(K^c)$ as $\mathscr{A}_1(K^c) = (\mathscr{B}(K)^z)'$. This can, of course, be done, but then the relation (11b) no longer holds. The same device *cannot* be applied to the wedge algebras to remedy the situation. The attempt to redefine $\mathscr{B}(K)$ as $\mathscr{B}_1(K) = (\mathscr{A}(K^c)^z)'$ is likely to be even less successful in the sense that the resulting algebras $\mathscr{B}_1(K)$ may fail to satisfy the condition of locality. In view of the relations (11a) and (11b) the duality condition $\mathscr{B}(K)' = \mathscr{A}(K^c)^z$ is trivial only in the sense that it is a trivial consequence of the nontrivial duality condition $\mathscr{A}(W_R)' = \mathscr{A}(W_L)^z$.

One might say that the duality condition is interesting only if it is the case that the algebras $\mathscr{A}(K^c)$ are generated (as von Neumann algebras) by the *local* operators they contain. If the special duality condition holds, there is a simple criterion for this situation to obtain.

Theorem 3: Let $\{\mathscr{A}(W), \mathscr{B}(K), \mathscr{A}(K^c)\}$ be a local, *TCP*- and Poincaré-covariant AB-system which satisfies the special condition of duality, and let the set $\mathscr{U}$ of all local operators in the systems be defined as in part (2) of Definition 1.

(1) Then the conditions

$$\mathscr{A}(W) = \{\mathscr{B}(K)|K\in\mathscr{D},\ K\subset W\}'' = \cap\{\mathscr{A}(K^c)|K\in\mathscr{D},\ K^c\supset W\}, \tag{36a}$$

$$\mathscr{A}(K^c) = \{\mathscr{B}(K_i)|K_i\in\mathscr{D},\ K_i\subset K^c\}'' \tag{36b}$$

hold if and only if $\mathscr{U}$ is irreducible, which is the case if and only if $\mathscr{U}\Omega$ is dense in the Hilbert space $\mathscr{H}$.

(2) In particular, the above conditions hold if there exists a $K_0\in\mathscr{D}$ such that $\mathscr{B}(K_0)\Omega$ is dense in $\mathscr{H}$, in which case

$$\mathscr{A}(W) = \{U(\lambda)\mathscr{B}(K_0)U(\lambda)^{-1}|\lambda\in\bar{\mathscr{G}},\ K_{0,\lambda}\subset W\}''. \tag{37}$$

The proof will be given in the next section. It depends on the circumstance that $\{V(t)\}$ is the modular automorphism group for the pair $(\mathscr{A}(W_R),\Omega)$. The statements are pleasing, because $\mathscr{U}$ ought to be irreducible in a bona fide theory of local operators.

We shall now pursue the ideas in Theorem 2 in the case when the primary set $\mathscr{S}$ is a subset of the set $\mathscr{U}$ of local operators in a given AB-system. As we shall see, quite specific conclusions can be drawn if the special condition of duality holds for the given AB-system.

Theorem 4: (1) Let $(AB)_0 = \{\mathscr{A}(W), \mathscr{B}(K), \mathscr{A}(K^c)\}$ [with the associated algebras $\mathscr{U}$ and $\mathscr{A}(\mathscr{M})$ defined as in (15)] be a local, *TCP*- and Poincaré-covariant AB-system with Θ as the *TCP* operator. It is furthermore assumed that $(AB)_0$ satisfies the special condition of duality in definition 3 *and* that $\mathscr{U}$ is irreducible [and hence $(AB)_0$ satisfies the conditions (36a) and (36b) in Theorem 3].

Let $\mathscr{S} = \cup\{\mathscr{S}(K)|K\in\mathscr{D}\}$ be a subset of $\mathscr{U}$ such that, for each $K\in\mathscr{D}$,

$$\mathscr{S}(K)\subset\mathscr{B}(K). \tag{38}$$

Then $\mathscr{S}$ satisfies the premises of Theorem 2 as a primary set of local operators, and hence all the conclusions of this theorem apply. Let the notation pertaining to the AB-system $(AB)_s = \{\mathscr{A}_s(W), \mathscr{B}_s(K), \mathscr{A}_s(K^c)\}$ generated by $\mathscr{S}$, and the AB-system (of operators on $\mathscr{H}_s$) $(AB)_{\rho s} = \{\hat{\mathscr{A}}_s(W), \hat{\mathscr{B}}_s(K), \hat{\mathscr{A}}_s(K^c)\}$ generated by $\rho(\mathscr{S})$, be exactly as in Theorem 2. Then

$$\Theta F_s = F_s\Theta,\quad JF_s = F_sJ, \tag{39}$$

where J is given in (34) in Definition 3. The AB-system $(AB)_s$ is *TCP*-covariant with Θ as the *TCP* operator. Furthermore,

$$\mathscr{A}_s(W) = \mathscr{A}(W)\cap\mathscr{A}_s(\mathscr{M}) = \mathscr{A}(W)\cap\{F_s\}', \tag{40a}$$

$$\mathscr{B}_s(K) = \mathscr{B}(K)\cap\mathscr{A}_s(\mathscr{M}) = \mathscr{B}(K)\cap\{F_s\}', \tag{40b}$$

$$\mathscr{A}_s(K^c)\subset\mathscr{A}(K^c)\cap\mathscr{A}_s(\mathscr{M})\subset\mathscr{A}(K^c)\cap\{F_s\}', \tag{40c}$$

with *equality* instead of inclusion whenever $\mathscr{B}(K)\Omega$ is dense in $\mathscr{H}$.

(2) In conformity with the notation in Theorem 2, let $\rho(Q)$ denote the restriction of F_sQF_s to $\mathscr{H}_s$ for any linear or antilinear operator Q on $\mathscr{H}$. Let

$$\Theta_s = \rho(\Theta),\quad J_s = \rho(J), \tag{41}$$

The AB-systems $(AB)_{\rho s}$ is a local, *TCP*- and Poincaré-covariant AB-system, with Θ_s as the *TCP* operator, and it satisfies the special condition of duality with J_s playing the role of J in Definition 3. The algebras in the system satisfy the relations

$$\hat{\mathscr{A}}_s(W) = \rho(\mathscr{A}_s(W)) = \rho(\mathscr{A}(W)), \tag{42a}$$

$$\hat{\mathscr{B}}_s(K) = \rho(\mathscr{B}_s(K)) = \rho(\mathscr{B}(K)), \tag{42b}$$

$$\hat{\mathscr{A}}_s(K^c) = \rho(\mathscr{A}_s(K^c)) = \rho(\mathscr{A}(K^c)). \tag{42c}$$

(3) If, for a particular $K\in\mathscr{D}$, the manifold $\mathscr{B}(K)\Omega$ is dense in $\mathscr{H}$, then $\hat{\mathscr{B}}_s(K)\Omega$ is dense in $\mathscr{H}_s$. For such a K, let J_K be the modular conjugation and let $\{V_K(t)|t\in R^1\}$ be the modular automorphism group defined by the pair $(\mathscr{B}(K),\Omega)$. Then

$$J_KF_s = F_sJ_K,\quad V_K(t)F_s = F_sV_K(t)\quad \text{for all } t \tag{43}$$

and $J_{sK} = \rho(J_K)$ is the modular conjugation and $\{V_{sK}(t) = \rho(V_K(t))|t\in R^1\}$ is the modular automorphism group defined by the pair $(\hat{\mathscr{B}}_s(K),\Omega)$.

(4) With the premises in (1) the AB-systems $(AB)_0$ and $(AB)_s$ are *identical* if and only if $\mathscr{H}_s = \mathscr{H}$.

Let $\mathscr{S}_1$ be another subset of $\mathscr{U}$ which satisfies the same premises as $\mathscr{S}$ in (1) above, and let $\mathscr{H}_{s1}$ be the subspace defined by $\mathscr{S}_1$ as in Theorem 2. Then the AB-system $(AB)_s$ is *identical* with the AB-system generated by $\mathscr{S}_1$ if and only if $\mathscr{H}_s = \mathscr{H}_{s1}$.

The proof of this somewhat overloaded theorem is given in the next section. As we said in the Introduction, one pleasing aspect of these results is that the "reduced" subsystem $(AB)_{\rho s}$ inherits so many of the characteristics of the original system $(AB)_0$.

Let us here discuss some particular features of the results. The assertions about uniqueness in part (4) say that the AB-system $(AB)_s$ depends on $\mathscr{S}$ *solely* through the subspace $\mathscr{H}_s$. Among other things this means, since $\mathscr{H}_s$ is determined by the algebra $\mathscr{P}_s$, that $(AB)_s$ only depends on what operators are included in $\mathscr{S}$, but *not* on how these operators are associated with specific double cones, beyond the general conditions (38). To illustrate this, we can consider the case of the local algebras associated with a single free Hermitian scalar field (of mass $m>0$). Let $r(x)$ be a real test function such that the average $\phi[r]$ of the field over $r(x)$ is not zero. Let us suppose that $r(x)$ is of compact support, and let K_0 be *any* double cone which contains the support of $r(x)$. Let $X = \sin(\phi[r])$, $\mathscr{S}(K) = 0$ if $K\neq K_0$, and $\mathscr{S}(K_0) = \{X\}$. It is well known that in this case $\mathscr{H}_s = \mathscr{H}$, and it follows that the AB-system $(AB)_s$ is equal to the AB-system $(AB)_0$ generated by all (local) Weyl operators of this field theory.

The statements in part (4) follow by inspection of the right member in (40a). In view of the relations (40a)–(40c) we have an alternative scheme for the definition of the AB-system $(AB)_s$ in the *special* case that $\mathscr{B}(K)\Omega$ is dense in $\mathscr{H}$ for every $K\in\mathscr{D}$, namely as follows. Given the subset $\mathscr{S}$ of operators in $\mathscr{U}$, we *define* the von Neumann algebra $\mathscr{A}_s(\mathscr{M})$ by $\mathscr{A}_s(\mathscr{M}) = \{U(\lambda)XU(\lambda)^{-1}|X\in\mathscr{S},\lambda\in\bar{\mathscr{G}}\}''$, after which we *define* the algebras $\mathscr{A}_s(W)$, $\mathscr{B}_s(K)$, and $\mathscr{A}_s(K^c)$ by (40a)–(40c). Note that in the special case under consideration the three members in (40c) are actually equal. This clear and straightforward approach thus leads to the same end result as the approach we have followed. Note, however, that with the above approach it still has to be shown that the resulting algebras form an AB-system in the sense of Definition 1, that the relations (39) hold, and that the system $(AB)_{\rho s}$ satisfies the special condition of duality.

With our construction of $\mathscr{A}_s(W_R)$ we necessarily have $V(t)\mathscr{A}_s(W_R)V(t)^{-1} = \mathscr{A}_s(W_R)$. With the special condition of duality this means that the subalgebra $\mathscr{A}_s(W_R)$ is invariant under the modular automorphism group of the pair $(\mathscr{A}(W_R),\Omega)$. The relations (39) and the *TCP* covariance of the system $(AB)_s$ follow from this, and it also follows that the system $(AB)_{\rho s}$ satisfies the special condition of duality. If the original AB-system satisfies a condition of duality different from the special condition, there is no guarantee that our "geometrical" construction of $\mathscr{A}_s(W_R)$ leads to an algebra invariant under the modular automorphism group for $\mathscr{A}(W_R)$ relative to Ω. If it does not, the situation could be "remedied" by including conjugations by elements of the group in the process for constructing the sub-wedge algebra, but the purely geometrical flavor of our construction is then lost. Such an approach will also not lead to an understanding of the nature of *all* local sub-AB-systems of a given AB-system.

We have included the condition that the original AB-system is generated by its *local* operators in the premises of the theorem because its seems like a physically reasonable requirement and because it leads to a better theorem. The somewhat subtle manner in which this assumption influences matters will be clear from the proof.

We shall now turn to the proofs of Theorems 2–4, after which we present some further general considerations.

III. PROOFS OF THEOREMS 2, 3, AND 4

In what follows we shall depend on a standard result concerning von Neumann algebras in general,[6] which we here recall in the form of a lemma.

Lemma 1: Let $\mathscr{A}$ be a von Neumann algebra of operators on a Hilbert space $\mathscr{H}$, an let F be a self-adjoint projection in the commutant $\mathscr{A}'$. Let $\mathscr{H}_F = F\mathscr{H}$, and, for any operator X on $\mathscr{H}$, let $\rho(X)$ denote the restriction of the operator FXF to $\mathscr{H}_F$. Similarly, if $\mathscr{S}$ is any set of operators on $\mathscr{H}$, we employ the notation $\rho(\mathscr{S}) = \{\rho(X)|X\in\mathscr{S}\}$. Then:

(a) The sets $\rho(\mathscr{A})$ and $\rho(\mathscr{A}')$ are von Neumann algebras, and

$$\rho(\mathscr{A})' = \rho(\mathscr{A}'). \tag{44}$$

(b) If the set $\mathscr{S}_a \subset \mathscr{A}$ generates $\mathscr{A}$ and is such that $X^*\in\mathscr{S}_a$ whenever $X\in\mathscr{S}_a$, then

$$\rho(\mathscr{A}) = \rho(\mathscr{S}_a)''. \tag{45}$$

(c) If the set $\mathscr{S}_c \subset \mathscr{A}'$ generates $\mathscr{A}'$ and is such that $X^*\in\mathscr{S}_c$ whenever $X\in\mathscr{S}_c$, and if $\mathscr{S}_c$ is, furthermore, closed under multiplication, then

$$\rho(\mathscr{A}') = \rho(\mathscr{S}_c)''. \tag{46}$$

Proof of Theorem 2: (1) Let the premises and notation be exactly as stated in Theorem 2. Part (1) of this theorem is trivial and needs no further discussion. We consider now the $*$-algebra $\mathscr{P}_s$. It follows at once from the definition of $\mathscr{P}_s$ that

$$U(\lambda)\mathscr{P}_s U(\lambda)^{-1} = \mathscr{P}_s, \quad \text{all } \lambda\in\bar{\mathscr{V}}. \tag{47}$$

From the definition (27) of $\mathscr{A}_s(W_R)$ it follows that this algebra is generated by elements in $\mathscr{P}_s$, and in view of (47) above the same is true of every $\mathscr{A}_s(W)$, for any $W\in\mathscr{W}$. Furthermore, every element $X\in\mathscr{P}_s$ is contained in some wedge algebra $\mathscr{A}_s(W)$, for some W. It follows that

$$\mathscr{P}_s'' = \{\mathscr{A}_s(W)|W\in\mathscr{W}\}'' = \mathscr{A}_s(\mathscr{M}). \tag{48}$$

Since $\mathscr{P}_s$ is a $*$-algebra it follows that

$$\overline{\mathscr{P}_s\Omega} = \overline{\mathscr{A}_s(\mathscr{M})\Omega} = \mathscr{H}_s = F_s\mathscr{H}. \tag{49}$$

The relation $U(\lambda)F_s = F_s U(\lambda)$ for all $\lambda\in\bar{\mathscr{V}}$ is trivial.

(2) We now recall a well-known consequence of the spectral condition which we shall refer to as the Reeh–Schlieder principle. Let $\phi\in\mathscr{H}$, and let R be any nonempty open subset of Minkowski space $\mathscr{M}$. Then

$$\overline{\text{span}}\{T(x)\phi\,|x\in\mathscr{M}\} = \overline{\text{span}}\{T(x)\phi\,|x\in R\}, \tag{50}$$

where the bar denotes closure.

We shall apply this principle to show that

$$\overline{\mathscr{A}_s(W_R)\Omega} = \overline{\mathscr{P}_s\Omega} = \mathscr{H}_s. \tag{51}$$

Let $X\in\mathscr{P}_s$. There then exists a K such that $X\in\mathscr{B}_s(K)$. In view of the nature of the sets K and W_R there obviously exists a nonempty open set R of translations such that $K_\lambda \subset W_R$, and hence $U(\lambda)XU(\lambda)^{-1}\in\mathscr{A}_s(W_R)$, for all $\lambda\in R$. It follows readily from the Reeh–Schlieder principle that

$$X\Omega\in\overline{\mathscr{A}_s(W_R)\Omega}, \tag{52}$$

which establishes (51) since it is trivial that the left member of (51) is a subspace of $\mathscr{H}_s$.

(3) The definition of the projection F_s trivially implies that $F_s\Omega = \Omega$ and that $\mathscr{A}_s(\mathscr{M}) \subset \{F_s\}'$. Since we have $ZF_s = F_sZ$, where Z is the operator defined in (6), it follows that if two bounded operators X and Y commute with F_s, then

$$[\rho(X)^z,\rho(Y)] = \rho([X^z,Y]), \tag{53}$$

where we define $\rho(X)^z = \rho(Z)\rho(X)\rho(Z)^{-1}$, and hence we have $\rho(X)^z = \rho(X^z)$ for any X. It follows that the set $\rho(\mathscr{S})$ $= \cup\{\rho(\mathscr{S}(K))|K\in\mathscr{D}\}$, with $\rho(\mathscr{S}(K)) = \{\rho(X)|X\in\mathscr{S}(K)\}$, is a primary set of local operators on $\mathscr{H}_s$ in the sense of Definition 2. All the conclusions in part (1) of Theorem 2 thus apply to $\rho(\mathscr{S})$ and the AB-system $(AB)_{\rho s} = \{\hat{\mathscr{A}}_s(W), \hat{\mathscr{B}}_s(K), \mathscr{A}_s(K^c)\}$ generated by $\rho(\mathscr{S})$.

(4) Since $F_s\in\mathscr{A}_s(\mathscr{M})' \subset \mathscr{A}_s(W_R)'$ we conclude from part (a) of Lemma 1 that $\rho(\mathscr{A}_s(\mathscr{M}))$ and $\rho(\mathscr{A}_s(W_R))$ are von Neumann algebras. Since $\mathscr{P}_s'' = \mathscr{A}_s(\mathscr{M})$ and since $\rho(\mathscr{P}_s)''$ $= \hat{\mathscr{A}}_s(\mathscr{M})$, it follows from part (b) of Lemma 1 that $\hat{\mathscr{A}}_s(\mathscr{M})$ $= \rho(\mathscr{A}_s(\mathscr{M}))$.

Let $\mathscr{S}_R$ be defined as the set of operators

$$\mathscr{S}_R = \cup\{U(\lambda)\mathscr{S}(K)U(\lambda)^{-1}|K\in\mathscr{D},\ \lambda\in\bar{\mathscr{V}},\ K_\lambda\subset W_R\}. \tag{54}$$

We then have $\mathscr{S}_R'' = \mathscr{A}_s(W_R)$ and $\rho(\mathscr{S}_R)''$ $= \hat{\mathscr{A}}_s(W_R)$, and it follows from part (b) of Lemma 1 that $\hat{\mathscr{A}}_s(W_R) = \rho(\mathscr{A}_s(W_R))$.

(5) In view of the above and in view of (51) it follows that Ω is a cyclic vector of $\hat{\mathscr{A}}_s(W_R)$, and hence of $\hat{\mathscr{A}}_s(W)$ for any $W\in\mathscr{W}$. It follows that Ω is a cyclic vector of $\hat{\mathscr{A}}_s(W_L)^z$, and since $\hat{\mathscr{A}}_s(W_R)' \supset \hat{\mathscr{A}}_s(W_L)^z$ we conclude that Ω is a separating vector for $\hat{\mathscr{A}}_s(W_R)$. In this discussion we, of course, regard Ω as a vector in $\mathscr{H}_s$, and all the considerations apply to this Hilbert space.

(6) From the fact that $\hat{\mathscr{A}}_s(W_R)\Omega$ is dense in $\mathscr{H}_s$ it fol-

lows by a well-known argument[7] based on the spectral condition that the set $\{T_s(x)\hat{\mathscr{A}}_s(W_R)T_s(x)^{-1}|x\in\mathscr{M}\}$ is irreducible on $\mathscr{H}_s$, and since this set is included in $\hat{\mathscr{A}}_s(\mathscr{M})$ it follows that $\hat{\mathscr{A}}_s(\mathscr{M})$ is equal to the algebra $B(\mathscr{H}_s)$ of all bounded operators on $\mathscr{H}_s$. We have here employed the notation $T_s(x) = \rho(T(x))$ for the representatives on $\mathscr{H}_s$ of the translation group.

From $\rho(\mathscr{A}_s(\mathscr{M})) = B(\mathscr{H}_s)$ we finally conclude that $\mathscr{A}_s(\mathscr{M})\Omega = \mathscr{H}_s$.

We have thus proved all the assertions in Theorem 2.

Proof of Theorem 3: As we said in the Introduction, this theorem is a minor extension of Theorem 6 in BW-II. The proof of the present theorem is based on the same reasoning.

We apply Theorem 2, with $\mathscr{S} = \mathscr{U}$ and $\mathscr{S}(K) = \mathscr{B}(K)$, and we adhere to the notation in that theorem. In the present case we have $\mathscr{P}_s = \mathscr{S} = \mathscr{U}$, and it immediately follows from (30b), (30c), and (32) that $\mathscr{A}_s(W_R)\Omega$ is dense in $\mathscr{H}$ if and only if $\mathscr{U}\Omega$ is dense in $\mathscr{H}$, which is the case if and only if $\mathscr{U}'' = B(\mathscr{H})$ (i.e., $\mathscr{U}$ is irreducible). If the condition (36a) holds, we have $\mathscr{A}_s(W_R) = \mathscr{A}(W_R)$, and, since $\mathscr{A}(W_R)\Omega$ was assumed to be dense, it follows that $\mathscr{U}\Omega$ is dense.

Suppose now that $\mathscr{U}\Omega$ is dense. The vector Ω is then a cyclic and separating vector for $\mathscr{A}_s(W_R)$, and we also have $V(t)\mathscr{A}_s(W_R)V(t)^{-1} = \mathscr{A}_s(W_R)$ for all t. Since, furthermore, $\mathscr{A}_s(W_R)\subset\mathscr{A}(W_R)$ and since by assumption on the latter algebra satisfies the special condition of duality, it follows from Theorem 2 in BW-I that $\mathscr{A}_s(W_R) = \mathscr{A}(W_R)$. This completes the proof.

We shall now turn our attention to Theorem 4. Since the proof is somewhat lengthy and involves careful attention to a number of details, we shall decompose the proof into a number of easily digestible propositions.

Proposition 1: Let the premises and notation be exactly as in the statement of Theorem 4, *except* that the condition that the AB-system $\{\mathscr{A}(W), \mathscr{B}(K), \mathscr{A}(K^c)\} = (AB)_0$ be generated by the local operators in $\mathscr{U}$ *is omitted*. Then $\Theta F_s = F_s\Theta$, i.e., the subspace $\mathscr{H}_s$ is a *TCP*-invariant.

Proof: Let $X\in\mathscr{A}_s(W_R)$. Then $X, X^*\in\mathscr{A}_s(W_R)\subset\mathscr{A}(W_R)$, and by the special condition of duality in Definition 3, and with the notation in that definition,

$$JX\Omega = V(i\pi)X^*\Omega. \tag{55}$$

Since $V(t)\mathscr{H}_s = \mathscr{H}_s$ for all t, and since $X^*\Omega\in\mathscr{H}_s$, it follows that $V(i\pi)X^*\Omega\in\mathscr{H}_s$. We have thus shown that $J\mathscr{A}_s(W_R)\Omega\subset\mathscr{H}_s$, and since by Theorem 2, $\mathscr{A}_s(W_R)\Omega$ is dense in $\mathscr{H}_s$, it follows that $J\mathscr{H}_s\subset\mathscr{H}_s$. Since J is an involution, we have $J\mathscr{H}_s = \mathscr{H}_s$, and hence $JF_s = F_sJ$ and $\Theta F_s = F_s\Theta$, where the latter relation follows from the relation $J = ZU_{\pi 3}\Theta$ and the Poincaré invariance of F_s.

Proposition 2: Let the premises and notation be exactly as in the Proposition 1 above. Let θ_s and J_s denote the restriction of the operator Θ, respectively J, to the subspace $\mathscr{H}_s$. Then the AB-system $(AB)_{\rho s} = \{\hat{\mathscr{A}}_s(W), \hat{\mathscr{B}}_s(K), \hat{\mathscr{A}}_s(K^c)\}$ is *TCP*-covariant, with Θ_s as the *TCP* operator. Furthermore, $(AB)_{\rho s}$ satisfies the special condition of duality, with J_s as the modular conjugation for the pair $(\hat{\mathscr{A}}_s(W_R),\Omega)$.

These conclusions in part (2) of Theorem 4 thus hold irrespective of whether $(AB)_0$ is generated by its local operators or not.

Proof: (1) We first note that in view of Proposition 1 the above definition of J_s makes sense. The operator $J_s = \rho(J)$ is thus an antiunitary involution which satisfies the relations

$$J_s\rho(X) = \rho(JX), \quad \rho(X)J_s = \rho(XJ) \tag{56}$$

for any $X\in B(\mathscr{H})$, where $\rho(Q)$ is defined as in Theorem 4.

We write $U_s(\lambda) = \rho(U(\lambda))$, and hence $V_s(t) = \rho(V(t))$. Since F_s commutes with $V(t)$ it also commutes (in the strong sense of von Neumann) with the positive self-adjoint operator $V(i\pi)$. The operator $V_s(i\pi) = \rho(V(i\pi))$ is thus a well-defined positive self-adjoint operator whose domain is precisely equal to F_sD_+ (interpreted as a subset of $\mathscr{H}_s$) if D_+ is the domain of $V(i\pi)$.

(2) Let $X_s\in\hat{\mathscr{A}}_s(W_R)$. By Theorem 2 we have $\hat{\mathscr{A}}_s(W_R) = \rho(\mathscr{A}_s(W_R))$, and hence there exists an $X\in\mathscr{A}_s(W_R)\subset\mathscr{A}(W_R)$ such that $X_s = \rho(X)$ and $X_s^* = \rho(X^*)$. By the special condition of duality we have

$$JV(i\pi)X\Omega = X^*\Omega, \tag{57}$$

which trivially implies that

$$(F_sJF_s)(F_sV(i\pi)F_s)(F_sXF_s)F_s\Omega = (F_sXF_s)^*F_s\Omega. \tag{58}$$

Interpreting this relation as a relation between vectors in $\mathscr{H}_s$ and operators on $\mathscr{H}_s$, we conclude that $\hat{\mathscr{A}}_s(W_R)\Omega$ is in the domain of $V_s(i\pi)$ and that

$$J_sV_s(i\pi)X_s\Omega = X_s^*\Omega, \quad \text{for all } X_s\in\hat{\mathscr{A}}_s(W_R) \tag{59}$$

(3) We have $V_s(t)\Omega = \Omega = J_s\Omega$, and by Theorem 2, Ω is a cyclic (and separating) vector for $\hat{\mathscr{A}}_s(W_R)$. By the construction of this von Neumann algebra we furthermore have $V_s(t)\hat{\mathscr{A}}_s(W_R)V_s(t)^{-1} = \hat{\mathscr{A}}_s(W_R)$ for all real t. In view of (59) above it then follows by Theorem 2 in BW-I that

$$J_s\hat{\mathscr{A}}_s(W_R)J_s = \hat{\mathscr{A}}_s(W_R)'. \tag{60}$$

(4) From this relation, and from that fact that the AB-system $(AB)_{\rho s}$ is local, it follows that

$$J_s\hat{\mathscr{A}}_s(W_R)J_s \supset \hat{\mathscr{A}}_s(W_L)^z. \tag{61}$$

Conjugating both members by $U_{\pi 1}$ representing a rotation by angle π about the 1-axis, we obtain

$$J_s\hat{\mathscr{A}}_s(W_L)J_s \supset \hat{\mathscr{A}}_s(W_R)^z,$$

$$\text{hence} \quad \hat{\mathscr{A}}_s(W_L)^z \supset J_s\hat{\mathscr{A}}_s(W_R)J_s. \tag{62}$$

In view of (61) it follows that

$$J_s\hat{\mathscr{A}}_s(W_R)J_s = \hat{\mathscr{A}}_s(W_R)' = \hat{\mathscr{A}}_s(W_L)^z \tag{63}$$

and hence trivially

$$\Theta_s\hat{\mathscr{A}}_s(W_R)\Theta_s = \hat{\mathscr{A}}_s(W_L). \tag{64}$$

We have thus shown that the AB-system $(AB)_{\rho s}$ is *TCP*-covariant, with Θ_s as the *TCP* operator, and that it satisfies the special condition of duality.

Proposition 3: Let the premises and notation by precisely as in Propositions 1 and 2. Then:

(a)

$$\hat{\mathscr{A}}_s(W) = \rho(\mathscr{A}_s(W)) = \rho(\mathscr{A}(W)), \tag{65}$$

$$\mathscr{A}_s(W) = \mathscr{A}(W) \cap \mathscr{A}_s(\mathscr{M}) = \mathscr{A}(W) \cap \{F_s\}' \tag{66}$$

for any $W \in \mathscr{W}$.

(b) The AB-system $(AB)_s = \{\mathscr{A}_s(W), \mathscr{B}_s(K), \mathscr{A}_s(K^c)\}$ is *TCP*-covariant with Θ as the *TCP* operator.

The above conclusions in Theorem 4 thus hold irrespective of whther $(AB)_0$ is generated by local operators or not.

Proof: (1) By Theorem 2 we have $\rho(\mathscr{A}_s(W_R)) = \hat{\mathscr{A}}_s(W_R)$ and also $F_s \in \mathscr{A}_s(W_R)'$. By part (a) of Lemma 1 it follows that

$$\hat{\mathscr{A}}_s(W_R)' = \rho(\mathscr{A}_s(W_R))' = \rho(\mathscr{A}_s(W_R)') \supset \rho(\mathscr{A}(W_L)^z). \tag{67}$$

By Proposition 2 the AB-system $(AB)_{\rho s}$ satisfies the condition of duality, and we have

$$\hat{\mathscr{A}}_s(W_R)' = \hat{\mathscr{A}}_s(W_L)^z = \rho(\mathscr{A}_s(W_L)^z) \subset \rho(\mathscr{A}(W_L)^z). \tag{68}$$

The relations (67) and 68) taken together imply that

$$\rho(\mathscr{A}(W_L)^z) = \rho(\mathscr{A}_s(W_L)^z). \tag{69}$$

Conjugating both members by $ZU(\lambda)$, we obtain (65).

(2) Let $X \in \mathscr{A}(W_R) \cap \{F_s\}'$, and hence $F_s X = X F_s = F_s X F_s$. It follows from (65) that there exists an $X_s \in \mathscr{A}_s(W_R)$ such that $F_s X_s F_s = F_s X F_s$ and hence $X F_s = X_s F_s$. The latter relation implies that $X\Omega = X_s \Omega$ and since $(X - X_s) \in \mathscr{A}(W_R)$, and since Ω is a separating vector for this algebra, we conclude that $X = X_s$. We have thus proved that $\mathscr{A}(W_R) \cap \{F_s\}' \subset \mathscr{A}_s(W_R)$, and, since we trivially have $\mathscr{A}_s(W_R) \subset \mathscr{A}(W_R) \cap \mathscr{A}_s(\mathscr{M}) \subset \mathscr{A}(W_R) \cap \{F_s\}'$ the relation (66) follows.

(3) The assertion in part (b) follows trivially if we conjugate the first and the third members in (66) by Θ and note that $\Theta F_s \Theta = F_s$ by Proposition 1.

Proposition 4: Let the premises and notation be exactly as in Propositions 1–3. Then the following relations hold for all $K \in \mathscr{D}$:

(a)

$$\mathscr{B}_s(K) = \mathscr{B}(K) \cap \mathscr{A}_s(\mathscr{M}) = \mathscr{B}(K) \cap \{F_s\}'; \tag{70}$$

(b)

$$\hat{\mathscr{B}}_s(K) \supset \rho(\mathscr{B}_s(K)), \tag{71}$$

$$\hat{\mathscr{A}}_s(K^c) = \rho(\mathscr{A}_s(K^c)), \tag{72}$$

$$\rho(\mathscr{B}_s(K))' \supset \rho(\mathscr{A}(K^c))^z, \tag{73}$$

$$\rho(\mathscr{A}_s(K^c))' \supset \rho(\mathscr{B}(K))^z; \tag{74}$$

(c)

$$\hat{\mathscr{B}}_s(K)^z = \hat{\mathscr{A}}_s(K^c)' = \rho(\mathscr{A}_s(K^c)') \supset \rho(\mathscr{B}(K))^z \supset \rho(\mathscr{B}_s(K))^z, \tag{75}$$

$$\rho(\mathscr{B}_s(K)^z)' \supset \rho(\mathscr{A}(K^c)) \supset \rho(\mathscr{A}_s(K^c)) = \hat{\mathscr{A}}_s(K^c) = (\hat{\mathscr{B}}_s(K)^z)'. \tag{76}$$

Proof: (1) The identifies in (70) follow at once from the relation (66) of Proposition 3, and from the basic definition of double-cone algebras as intersections of wedge algebras.

(2) Any element in $\rho(\mathscr{B}_s(K))$ is of the form $\rho(X)$ with $X \in \mathscr{B}_s(K)$. Then $X \in \mathscr{A}_s(W)$ for all $W \supset K$ It follows that $\rho(X) \in \rho(\mathscr{A}_s(W)) = \hat{\mathscr{A}}_s(W)$ for all $W \supset K$, in view of (65) in Proposition 3. In view of the definition of $\hat{\mathscr{B}}_s(K)$ this implies that $\rho(X) \in \hat{\mathscr{B}}_s(K)$. We have thus proved the relation (71).

(3) We have $F_s \in \mathscr{A}_s(K^c)'$ and $\mathscr{A}_s(K^c)$ is generated (as a von Neumann algebra) by the set $\cup\{\mathscr{A}_s(W) | W \in \mathscr{W}, W \subset K^c\}$. By part (b) of Lemma 1 it follows that $\rho(\mathscr{A}_s(K^c))$ is generated by the set $\cup\{\rho(\mathscr{A}_s(W)) | W \in \mathscr{W}, W \subset K^c\}$. Since by Proposition 3 we have $\rho(\mathscr{A}_s(W)) = \hat{\mathscr{A}}_s(W)$ the relation (72) follows, in view of the definition of $\hat{\mathscr{A}}_s(K^c)$.

(4) By part (a) of Lemma 1 we have $\rho(\mathscr{B}_s(K))' = \rho(\mathscr{B}_s(K)')$. Since $\mathscr{B}_s(K) \subset \mathscr{B}(K)$, we have, taking into account the condition of locality, $\mathscr{B}_s(K)' \supset \mathscr{B}(K)' \supset \mathscr{A}(K^c)^z$, and hence $\rho(\mathscr{B}_s(K)') \supset \rho(\mathscr{A}(K^c)^z) = \rho(\mathscr{A}(K^c))^z$, from which (73) follows. The relation (74) is proved by exactly the same kind of reasoning.

(5) We consider the assertions in part (c). By Proposition 2 the AB-system $(AB)_{\rho s}$ satisfies the condition of duality, which implies that the first two members in (75) are equal and that the last two members in (76) are equal. The remaning identities and inclusion relations in (75) and (76) then follow trivially from the relations (70)–(74) and from the fact that $\mathscr{B}_s(K) \subset \mathscr{B}(K)$ and $\mathscr{A}_s(K^c) \subset \mathscr{A}(K^c)$. This completes the proof.

If we consider (75), we note that all the sets in these relations are equal if $\hat{\mathscr{B}}_s(K) = \rho(\mathscr{B}_s(K))$. As far as the author can see, it is not possible to conclude that equality holds without additional assumptions. We shall next show that equality in fact holds if the original AB-system $\{\mathscr{A}(W), \mathscr{B}(K), \mathscr{A}(K^c)\}$ is generated by local operators, as we would expect to be the case in a physically meaningful theory. Our demonstration of this involves some simple geometrical considerations which we shall now discuss.

Proposition 5: Let $W_1 = W(\pi_1', \pi_1'')$ and $W_2 = W(\pi_2', \pi_2'')$ be two wedges in $\mathscr{W}$ bounded by the characteristic planes

$$\begin{aligned} \pi_1' &= \{x | x \cdot u_1' = c_1'\}, \quad \pi_1'' = \{x | x \cdot u_1'' = c_1''\}, \\ \pi_2' &= \{x | x \cdot u_2' = c_2'\}, \quad \pi_2'' = \{x | x \cdot u_2'' = c_2''\}, \end{aligned} \tag{77}$$

i.e.,

$$\begin{aligned} W_1 &= \{x | x \cdot u_1' < c_1', \quad x \cdot u_1'' > c_1''\}, \\ W_2 &= \{x | x \cdot u_2' < c_2', \quad x \cdot u_2'' > c_2''\}. \end{aligned} \tag{78}$$

Here u_1', u_1'', u_2', and u_2'' are forward lightlike 4-vectors such that u_1' and u_1'' are not proportional and such that u_2' and u_2'' are not proportional. We shall say that W_1 *is in a general position relative to* W_2 (and vice versa) if and only if the 4-vectors u_1', u_1'', u_2', and u_2'' are linearly independent.

(a) If the two wedges W_1 and W_2 are in a general relative position, then the subset $C_{12} = \overline{W}_1^c \cap \overline{W}_2^c$ of Minkowski space $\mathscr{M}$ is a nonempty open convex cone, and, if K is any double cone in $\mathscr{D}$, then there exists a nonempty open subset T_K of $\mathscr{M}$, regarded also as a subset of the translation group in $\bar{\mathscr{P}}$, such that $K_\lambda \subset C_{12}$ for all $\lambda \in T_K$.

(b) Let $K\in\mathscr{D}$ be a double cone, and let $W_0\in\mathscr{W}$ be a wedge such that $W_0\supset K$. Let $\mathscr{W}_K$ denote the set of all wedges $W\in\mathscr{W}$ such that $W\supset K$ *and* such that W_0 and W are in a general relative position. The set $\mathscr{W}_K$ is nonempty, and, if $W_1\in\mathscr{W}$ and $W_1\supset K$, then there exists a $W_2\in\mathscr{W}_K$ such that W_1 is in a general position relative to W_2.

Proof: (1) Let $W_1=W(\pi_1',\pi_1'')$ and $W_2=W(\pi_2',\pi_2'')$ be two wedges such that (77) and (78) apply. Since Minkowski space is four-dimensional, we note that the four characteristic planes π_1', π_1'', π_2', and π_2'' have precisely one common point of intersection x_0 if and only if the four normals of these planes are linearly independent, and hence if and only if W_1 and W_2 are in a general relative position. Let us now assume that this is the case. For any $x\in\mathscr{M}$ we can regard the scalar products of x with the four vectors u_1', u_1'', u_2', u_2'' as the coordinates of x relative to an oblique system of coordinate axes. We immediately conclude that the complement of the union of the four characteristic planes is the union of 16 disjoint sets, each one of which is a nonempty open convex cone, and one of these sets is precisely $\overline{W}_1{}^c\cap\overline{W}_2{}^c$. The remaining assertion in part (a) is trivial in view of the fact that double cones are bounded sets.

(2) We consider part (b), and we first note that, given any $K\in\mathscr{D}$ and any $W\in\mathscr{W}$, we can always so parallel-translate W that K is contained in the translated wedge. In view of this, the issue in part (b) is really the following. Given two pairs (u_0',u_0'') and (u_1',u_1'') of nonproportional forward lightlike vectors, does there exist a pair (u_2',u_2'') of nonproportional forward lightlike vectors such that the four vectors u_0', u_0'', u_2' and u_2'' are linearly independent, and such that the four-vectors u_1', u_1'', u_2', and u_2'' are also linearly independent? With a moments reflection we readily conclude that the answer is in the affirmative, and this establishes the assertions in part (b).

Proposition 6: Let the premises and notation be exactly as in Propositions 1–4, except that we now assume in addition that the AB-system $(AB)_0=\{\mathscr{A}(W),\mathscr{B}(K),\mathscr{A}(K^c)\}$ is generated by the set of local operators $\mathscr{U}$. Then, for all $K\in\mathscr{D}$,

$$\hat{\mathscr{B}}_s(K)=\rho(\mathscr{B}_s(K))=\rho(\mathscr{B}(K)), \tag{79a}$$

$$\hat{\mathscr{A}}_s(K^c)=\rho(\mathscr{A}_s(K^c))=\rho(\mathscr{A}(K^c)), \tag{79b}$$

and

$$\mathscr{A}_s(K^c)\subset\mathscr{A}(K^c)\cap\mathscr{A}_s(\mathscr{M})\subset\mathscr{A}(K^c)\cap\{F_s\}'. \tag{80}$$

These three sets are *equal* whenever $\mathscr{B}(K)\Omega$ is dense in $\mathscr{H}$.

Proof: (1) We want to show that $\hat{\mathscr{B}}_s(K)\subset\rho(\mathscr{B}_s(K))$ for any $K\in\mathscr{D}$, from which the relations (79a) and (79b) follow at once when we take into account the results (75) and (76) in Proposition 4.

Let $\hat{X}\in\hat{\mathscr{B}}_s(K)$, and let W_1, $W_2\in\mathscr{W}$ be such that $W_1\supset K$ and $W_2\supset K$, *and* such that W_1 and W_2 are in a general relative position in the sense of the definition in Proposition 5. In view of (65) in Proposition 3, we have $\hat{X}\in\hat{\mathscr{A}}_s(W_1)=\rho(\mathscr{A}_s(W_1))$ and $\hat{X}\in\hat{\mathscr{A}}_s(W_2)=\rho(\mathscr{A}_s(W_2))$. It follows that there exists an $X_1\in\mathscr{A}_s(W_1)$ and an $X_2\in\mathscr{A}_s(W_2)$ such that $\hat{X}=\rho(X_1)=\rho(X_2)$ and hence

$$X_1F_s=F_sX_1F_s=F_sX_2F_s=X_2F_s. \tag{81}$$

The equality of the first and fourth members implies that

$$X_1\Omega=X_2\Omega. \tag{82}$$

The operator (X_1-X_2) is an element of the von Neumann algebra $(\mathscr{A}_s(W_1)\cup\mathscr{A}_s(W_2))''$. We shall show that Ω is a separating vector for this algebra, which is equivalent to the statement that Ω is a cyclic vector for the commutant $\mathscr{A}_s(W_1)'\cap\mathscr{A}_s(W_2)'$, and we can then conclude from (82) that $X_1=X_2$.

(2) By Proposition 5 the set $C=\overline{W}_1{}^c\cap\overline{W}_2{}^c$ is a nonempty open cone which contains a translate of every double cone. We define a von Neumann algebra $\mathscr{A}_C$ by

$$\mathscr{A}_C=\{\mathscr{B}(K)|K\in\mathscr{D},\quad K\subset C\}'', \tag{83}$$

and we then have $\mathscr{A}_C\subset\mathscr{A}(\overline{W}_1{}^c)\cap\mathscr{A}(\overline{W}_2{}^c)$. Taking into account the facts in part (a) of Proposition 5, we conclude by a simple application of the Reeh–Schlieder principle in analogy with the reasoning in step (2) in the proof of Theorem 2 that

$$\overline{\mathscr{A}_C\Omega}=\overline{\mathscr{U}\Omega}=\mathscr{H}. \tag{84}$$

Hence Ω is a cyclic vector for $\mathscr{A}_C$, and trivially also for $\mathscr{A}_C{}^z$. Since $\mathscr{A}_C{}^z\subset\mathscr{A}(\overline{W}_1{}^c)^z\cap\mathscr{A}(\overline{W}_2{}^c)^z\subset\mathscr{A}_s(W_1)'\cap\mathscr{A}_s(W_2)'$, Ω is also a cyclic vector for the last algebra in this string. With reference to the discussion in step (1) above we conclude that $X_1=X_2$.

(3) Let $W_0\in\mathscr{W}$ and $W_0\supset K$, and let $X\in\mathscr{A}_s(W_0)$ be such that $\hat{X}=\rho(X)$. From what has been proved it follows that if $W\in\mathscr{W}$, $W\supset K$, and W is in a general position relative to W_0, then $X\in\mathscr{A}_s(W)$. Let $\mathscr{W}_K$ be the set of all W which satisfy these geometrical conditions. Let $W_1\in\mathscr{W}$, $W_1\supset K$. By part (b) of Proposition 5 there then exists a $W_2\in\mathscr{W}_K$ such that W_1 and W_2 are in a general position. In view of what was proved in (1) and (2) above, it follows that $X\in\mathscr{A}_s(W_1)$. Hence

$$X\in\cap\{\mathscr{A}_s(W)|W\in\mathscr{W},\ W\supset K\}=\mathscr{B}_s(K), \tag{85}$$

and we have thus shown that $\hat{\mathscr{B}}_s(K)\subset\rho(\mathscr{B}_s(K))$, and thereby proved the relations (79a)–(79b) in view of what was said in step (1).

(4) We now consider (80). Let $X\in\mathscr{A}(K^c)\cap\{F_s\}'$. By (79b), there then exists an $X_s\in\mathscr{A}_s(K^c)$ such that $F_sX_sF_s=X_sF_s=F_sXF_s=XF_s$, and we then have $X\Omega=X_s\Omega$. Now $(X-X_s)\in\mathscr{A}(K^c)$, and *if* $\mathscr{B}(K)\Omega$ is dense, *then* Ω is a separating vector for $\mathscr{A}(K^c)$, and hence $X=X_s$. In *this* case we thus have $\mathscr{A}(K^c)\cap\{F_s\}'\subset\mathscr{A}_s(K^c)$, which implies that the three sets in (80) are equal since the stated inclusions hold trivially, always. This completes the proof.

Proposition 7: Let the premises and notation be exactly as in Theorem 4, and hence the premises in Propositions 1–4 and 6 are assumed. In addition, we assume that for a *particular* $K\in\mathscr{D}$ the set $\mathscr{B}(K)\Omega$ is dense in $\mathscr{H}$. The vector Ω is then cyclic and separating for $\mathscr{B}(K)$. Let J_K be the modular conjugation, and let $\{V_K(t)|t\in R^1\}$ be the modular automorphism group for the pair $(\mathscr{B}(K),\Omega)$, with the parametrization of $V_K(t)$ so chosen that

$$J_KV_K(i\pi)X\Omega=X^*\Omega,\quad \text{for all } X\in\mathscr{B}(K). \tag{86}$$

Then:

(a) The set $\hat{\mathscr{B}}_S(K)\Omega$ is dense in $\mathscr{H}_s$, and Ω is a cyclic

and separating vector (in $\mathscr{H}_s$) for $\hat{\mathscr{B}}_s(K)$.

(b)

$$J_K F_s = F_s J_K, \quad V_K(t) F_s = F_s V_K(t) \quad \text{for all } t. \tag{87}$$

The operator $J_{sK} = \rho(J_K)$ is the modular conjugation, and $\{V_{sK}(t) = \rho(V_K(t)) | t \in R^1\}$ is the modular automorphism group, for the pair $(\hat{\mathscr{B}}_s(K), \Omega)$.

Proof: (1) In this discussion the double cone K, with the stated properties, is fixed. Since $\mathscr{B}(K)\Omega$ was assumed dense in $\mathscr{H}$, we have

$$\overline{F_s \mathscr{B}(K)\Omega} = \mathscr{H}_s. \tag{88}$$

By Proposition 6 we have $F_s \mathscr{B}(K) F_s = F_s \mathscr{B}_s(K) F_s$, and from this and from (88) it follows that $\hat{\mathscr{B}}_s(K)\Omega$ is dense in $\mathscr{H}_s$. The vector Ω is thus a cyclic vector for $\hat{\mathscr{B}}_s(K)$ (regarded as a vector in $\mathscr{H}_s$), and, since it is trivially a cyclic vector for the commutant $\hat{\mathscr{A}}_s(K^c)^z$ it follows that Ω is also separating for $\hat{\mathscr{B}}_s(K)$.

(2) Let $X \in \mathscr{B}(K)$, and hence (86) applies. By Proposition 6 there exists an $X_s \in \mathscr{B}_s(K) \subset \mathscr{B}(K)$ such that $F_s X F_s = F_s X_s F_s = X_s F_s = F_s X_s$, and, hence,

$$F_s X \Omega = X_s \Omega, \quad F_s X^* \Omega = X_s^* \Omega. \tag{89}$$

Now (86) also applies to X_s, and we thus have, for any $X \in \mathscr{B}(K)$,

$$J_K V_K(i\pi) F_s X \Omega = F_s X^* \Omega = F_s J_K V_K(i\pi) X \Omega. \tag{90}$$

By the Tomita–Takesaki theory the domain $\mathscr{B}(K)\Omega$ is a core for the positive self-adjoint operator $V_K(i\pi)$, and, since F_s is a bounded self-adjoint operator, the identity between the first and third members in (90) implies that F_s commutes with J_K and that F_s commutes in the strong sense of von Neumann with $V_K(i\pi)$, and hence we conclude that the relations (87) hold.

(3) In view of (87) the operator $J_{sK} = \rho(J_K)$ is an antiunitary involution on $\mathscr{H}_s$, and $\{V_{sK}(t) = \rho(V_K(t)) | t \in R^1\}$ is a one-parameter unitary group on $\mathscr{H}_s$. If D_{+K} is the domain of the self-adjoint operator $V_K(i\pi)$, then $F_s D_{+K}$ is the domain of the self-adjoint operator $V_{sK}(i\pi) = \rho(V_K(i\pi))$. Furthermore, if D_c is a core for $V_K(i\pi)$, then $F_s D_c$ is a core for $V_{sK}(i\pi)$. Here, of course, we regard $F_s D_{+K}$ and $F_s D_c$ as (dense) domains within the Hilbert space $\mathscr{H}_s$. From what has been said it thus follows that $\hat{\mathscr{B}}_s(K)\Omega$ is a core for $V_{sK}(i\pi)$.

The relation (90) holds for any $X_s \in \mathscr{B}_s(K)$, and for any such X_s it is equivalent to the relation

$$(F_s J_K F_s)(F_s V_K(i\pi) F_s)(F_s X_s F_s)\Omega = (F_s X_s F_s)^* \Omega. \tag{91}$$

If we interpret this relation as a relation between vectors in $\mathscr{H}_s$ and operators on $\mathscr{H}_s$, and note that every operator in $\hat{\mathscr{B}}_s(K)$ is of the form $\rho(X_s)$ with $X_s \in \mathscr{B}_s(K)$, we conclude that

$$J_{sK} V_{sK}(i\pi) X \Omega = X^* \Omega, \quad \text{for all } X \in \hat{\mathscr{B}}_s(K). \tag{92}$$

This completes the proof of the assertions in part (b), and it follows that

$$J_{sK} \hat{\mathscr{B}}_s(K) J_{sK} = \hat{\mathscr{B}}_s(K)' = \hat{\mathscr{A}}_s(K^c)^z, \tag{93a}$$

$$V_{sK}(t) \hat{\mathscr{B}}_s(K) V_{sK}(t)^{-1} = \hat{\mathscr{B}}_s(K), \quad \text{for all real } t. \tag{93b}$$

We note here that these two relations can also easily be proved directly, on the basis of (87) and (79a), without any further reliance on the Tomita–Takesaki theory.

Final comments on the proof of Theorem 4: At this point we note that Propositions 1–4 and 6–7 and taken together contain the proofs of all the assertions in parts (1)–(3) of Theorem 4. The assertions in part (4) follow trivially from the relations (40a).

As we promised to do, we have kept track of the role of the assumption that the original AB-system $(AB)_0$ is generated by *local* operators. As we have seen, this assumption is not relevant for Propositions 1–4. It did enter into the proof of Proposition 6. From the nature of the proof we see that the apparently weaker assumption that Ω is a cyclic vector for $\mathscr{A}(W_1) \cap \mathscr{A}(W_2)$ whenever the wedges W_1 and W_2 are in a general relative position is sufficient for the conclusion. Finally, we note that Proposition 7 applies only if $\mathscr{B}(K)\Omega$ is dense for some K, in which case $(AB)_0$ is necessarily generated by local operators.

IV. CONCLUDING REMARKS

It is interesting to consider Theorem 4 under the additional assumption of *weak additivity*,[8] which is a very plausible assumption physically if local fields play any role in the theory. In the present context this is the assumption that $\mathscr{B}(K)\Omega$ is dense in $\mathscr{H}$ for any $K \in \mathscr{D}$. By Theorem 4 weak additivity holds for any sub-AB-system $(AB)_{\rho s}$ (on the subspace $\mathscr{H}_s$) if it holds for the original AB-system $(AB)_0$ *and* if the special condition of duality holds for $(AB)_0$. With weak additivity there is a modular automorphism group $\{V_K(t)\}$ and a modular conjugation operator J_K corresponding to every double cone K. By part (3) of Theorem 4 these objects for the subsystem $(AB)_{\rho s}$ are related in a very simple manner to the corresponding objects for the original system $(AB)_0$: The operators for the subsystem are simply the restrictions of the operators for the original system to the subspace $\mathscr{H}_s$. These circumstances suggest that the group Γ generated by all modular automorphisms and modular conjugations might be an interesting object for study. With the special condition of duality this group always contains the Poincaré group as a subgroup. Whether it has any other geometrical significance is not known. It is interesting to note, however, that Γ has been determined by Hislop and Longo[9] for the special case of a theory of a free massless scalar field. In this case the group generated by all modular automorphisms associated with wedges and double cones turns out to be identical with the conformal group. This result seems to depend in a very essential way on the very special properties of massless fields: See also the discussion of such fields by Buchholz.[10]

In this paper we have associated von Neumann algebras with wedges, double cones, and their causal complements. A natural generalization is to define the algebras $\mathscr{B}(K)$ for compact convex subsets of Minkowski space which satisfy the condition $K = K^{cc}$ and which have a nonempty interior. Such sets K satisfy the crucial relations (5a) and (5b), and hence the considerations of this paper apply with very minor modifications. The group Γ is now a very much larger group, even in the case of massless fields.

A *local internal symmetry* is a unitary operator G which

maps every local algebra onto itself by conjugation, and which preserves the vacuum state, $G\Omega = \Omega$. Relative to our notion of an AB-system we thus have $G\mathcal{A}(W)G^{-1} = \mathcal{A}(W)$ for every wedge W, which trivially implies $G\mathcal{B}(K)G^{-1} = \mathcal{B}(K)$ for every double cone K. It was shown by Landau and Wichmann[11] that G then commutes with all translations $T(x)$. It was shown in BW-I and BW-II that if the special condition of duality holds, then G also commutes with all Poincaré transformations $U(\lambda)$, as well as with the TCP operator Θ. The proof is very simple, and depends on the fact that G must commute with all modular automorphisms and all modular conjugations. Hence G commutes with the group Γ introduced above. *Without* the special condition of duality it can well happen that G commutes only with the translations, but *not* with the homogeneous Lorentz transformations. An example of this has been given by Landau.[12] The example involves an infinite-component free field, and is thus unphysical, but it demonstrates convincingly that the special condition of duality is indeed special.

Let $\mathcal{G}$ be a group of local internal symmetries [for an AB-system $(AB)_0$]. We can then consider the sub-AB-system generated by the set $\mathcal{S}$ of all local operators which commute with all elements of $\mathcal{G}$, i.e., we have $\mathcal{S} = \cup\{\mathcal{S}(K)|K\in\mathcal{D}\}$ with $\mathcal{S}(K) = \mathcal{B}(K)\cap\mathcal{G}'$. This situation has been discussed by Doplicher, Haag, and Roberts,[13] who presented good arguments why it is reasonable to assume that $\mathcal{G}$ is compact. If $\mathcal{G}$ is compact it is easy to show, with reference to the notation in our Theorem 2, that the subspace $\mathcal{H}_s$ is precisely equal to the set of all vectors in $\mathcal{H}$ which are invariant under $\mathcal{G}$. In the quoted paper a much stronger result was obtained, namely that $\mathcal{A}_s(\mathcal{M}) = \mathcal{G}'$. With the premises of Theorem 4, including the assumption of weak additivity, we then have the simple relations $\mathcal{A}_s(W) = \mathcal{A}(W)\cap\mathcal{G}'$, $\mathcal{B}_s(K) = \mathcal{B}(K)\cap\mathcal{G}'$, and $\mathcal{A}_s(K^c) = \mathcal{A}(K^c)\cap\mathcal{G}'$. This is obviously a particularly important case of a sub-AB-system, but it is easy to show by examples that a sub-AB-system constructed as in Theorem 4 need not be determined by a group of local internal symmetries in this fashion.

ACKNOWLEDGMENTS

It is a pleasure to thank Joseph J. Bisognano, Detlev Buchholz, and Peter D. Hislop for many conversations on these topics and for helpful suggestions.

[1]R. Haag and D. Kastler, J. Math. Phys. **5**, 848 (1964); S. Doplicher, R. Haag, and J. E. Roberts, Commun. Math. Phys. **13**, 1 (1969). The literature on this subject is quite extensive. For additional references, see the above, as well as the references in Refs. 2 and 3.
[2]J. J. Bisognano and E. H. Wichmann, J. Math. Phys. **16**, 985 (1975).
[3]J. J. Bisognano and E. H. Wichmann, J. Math. Phys. **17**, 303 (1976).
[4]R. F. Streater and A. S. Wightman, *PCT, Spin and Statistics, and All That* (Benjamin, New York, 1964).
[5]M. Takesaki, *Tomita's Theory of Modular Hilbert Algebras and Its Applications* (Springer-Verlag, Berlin, 1970).
[6]J. Dixmier, *Les algèbres d'opérateurs dans l'espace hilbertien* (Gauthier-Villars, Paris, 1957), see theorem 1 on p. 18.
[7]R. F. Streater and A. S. Wightman, Ref. 4, p. 141.
[8]H. J. Borchers, Commun. Math. Phys. **10**, 269 (1968).
[9]P. D. Hislop and R. Longo, Commun. Math. Phys. **84**, 71 (1982).
[10]D. Buchholz, *Proceedings of the International Conference on Operator Algebras, Ideals and Their Application in Theoretical Physics*, edited by Baumgärtel, Lassner, Pietsch, and Uhlmann (Teubner, Leipzig, 1978), pp. 146–53.
[11]L. J. Landau and E. H. Wichmann, J. Math. Phys. **11**, 306 (1970).
[12]L. J. Landau, Commun. Math. Phys. **17**, 156 (1970).
[13]S. Doplicher, R. Haag, and J. E. Roberts, Ref. 1, Sec. III.

Potential densities for quantum field theories in the scattering picture[a)]

M. O. Farrukh
Department of Physics, Faculty of Sciences, Lebanese University, Hadeth, Beirut, Lebanon

(Received 20 September 1979; accepted for publication 23 December 1982)

We introduce causality into the general field theory in the scattering picture, introduced in a previous paper, by expanding the effective potential density, which is constructed in analogy with local nonrelativistic potentials and which is shown not to be microcausal. Thus the usual divergencies of conventional quantum field theory could be avoided. In the nonrelativistic limit, the expansion of the U matrix reduces to the conventional time-ordered expansion, which shows that the theory is, in a sense, macrocausal.

PACS numbers: 11.10.Cd, 11.20.Dj

I. INTRODUCTION

In a previous paper,[1] we have shown that there exists a picture of quantum mechanics, most adequate for quantum field theory, called the scattering picture. If $\mathscr{H}$ is the Hilbert space of states and H_0 is the free Hamiltonian, then for any $t \in \mathscr{R}$ there exists a closed subspace $\mathscr{H}(t) \subseteq \mathscr{H}$ and a unitary operator $U(t): \mathscr{H} \to \mathscr{H}(t)$ such that[2]:

(1) $U(t) = e^{itH_0} U(0) e^{-itH_0}$,
(2) s-$\lim_{t \to -\infty} U(t) = I$, the identity operator on $\mathscr{H}$,
(3) s-$\lim_{t \to +\infty} U(t) = S$, the scattering matrix, which is a unitary transformation of $\mathscr{H}$ into itself,

where s-lim denotes the strong limit.

In this picture, the states are time-independent. If $\phi_{\text{in}} \in \mathscr{H}$ is a state in the scattering picture, then the equation of motion of the same state in the interaction picture is given by

$$\phi_I(t) = U(t)\phi_{\text{in}}, \quad t \in R. \tag{1}$$

On the other hand, if $\{A_0(t); t \in \mathscr{R}\}$ is the equation of motion of an operator in the interaction picture, then the corresponding operator in the scattering picture is given by

$$\hat{A}(t) = U^{\dagger}(t) E(t) A_0(t) U(t), \tag{2}$$

where $E(t)$ is the projection operator on $\mathscr{H}(t)$.

In the absence of bound states, $\{U(t); t \in \mathscr{R}\}$ becomes a one-parametric family of unitary operators on $\mathscr{H}$. In this case, Eq. (2) becomes

$$\hat{A}(t) = U^{-1}(t) A_0(t) U(t). \tag{3}$$

Also, one introduces an effective potential[3] $F(t) = e^{itH_0} F e^{-itH_0}$ such that $D = -\int_{-\infty}^{\infty} F(t)\, dt$ exists, where

$$U(t) = \exp\left\{ -i \int_{-\infty}^{t} F(t')\, dt' \right\}. \tag{4}$$

The S matrix thus takes the form $S = e^{iD}$, where D is the phase matrix.

The potential $V(t)$ in the interaction picture has the form

$$V(t) = i \frac{dU(t)}{dt} U^{-1}(t). \tag{5}$$

Applying (5) to the derivative of (1), we get

$$i \frac{d}{dt} \phi_I(t) = V(t) \phi_I(t), \tag{6}$$

which is the Schrödinger equation in the interaction picture. If we solve this equation and compare the solution with the original equation (1), we get

$$U(t) = T \exp\left\{ -i \int_{-\infty}^{t} V(t')\, dt' \right\}, \tag{7}$$

where T denotes time ordering.

Equations (3) and (4) can be readily generalized to the relativistic case. Given a Poincaré-covariant effective potential density[3] $\{G(x); x \in \mathscr{R}^4\}$ on a Fock space $\mathscr{F}$ such that $D = -\int G(x)\, d^4x$ exists, we write

$$U(x) = \exp\left\{ -i \int_{y \in \Gamma(x)} G(y)\, d^4y \right\}, \tag{8}$$

where $\Gamma(x)$ is the past cone of the point $x \in \mathscr{R}^4$. Also[4]

$$\Psi(x) = U^{-1}(x) \Phi(x) U(x), \tag{9}$$

where $\Phi(x)$ is a free field and $\Psi(x)$ is the interacting field in the scattering picture. The S matrix is again given by $S = e^{iD}$, satisfying the conditions of unitarity and Poincaré invariance.

The object of the present paper is to introduce causality into the theory. In the conventional Hamiltonian approach to quantum field theory,[5] an interaction Hamiltonian density $\{\mathscr{H}(x); x \in \mathscr{R}^4\}$ is introduced, and the potential V is defined as follows:

$$V(t) = \int_{x_0 = t} \mathscr{H}(x)\, d^3x, \tag{10}$$

from which we get

$$S = I + \sum_{n=1}^{\infty} \frac{(-i)^n}{n!} \int dx_1 \int dx_2 \cdots \int dx_n \times T\{\mathscr{H}(x_1)\mathscr{H}(x_2)\cdots\mathscr{H}(x_n)\}. \tag{11}$$

But the expression $T\{\mathscr{H}(x_1)\mathscr{H}(x_2)\cdots\mathscr{H}(x_n)\}$ is not covariant unless $[\mathscr{H}(x), \mathscr{H}(y)] = 0$ for $(x-y)^2 < 0$. Hence, in order to get a unitary invariant S matrix, the interaction Hamiltonian density must satisfy the conditions of Hermiticity, covariance, and local commutativity. However, these conditions, when taken together, give rise to all the known divergencies, and the theory is not consistent unless $\mathscr{H}(x) = 0$.

We have shown before that given an S matrix, there exist many interpolating fields, giving rise to that S matrix.

[a)] Work supported by the Lebanese University Council.

 0022-2488/83/061645-07$02.50 © 1983 American Institute of Physics

This result has been known before,[6] and the following definition of causality has been adopted: An S matrix is causal if there exists one interpolating field satisfying local commutativity from which that S matrix can be derived. However, this definition cannot be applied in our case because, by Haag's theorem,[7] this cannot be done unless the S matrix is trivial.

Our solution to the problem of causality, to which the remaining part of this paper is devoted, may be outlined as follows:

We aim at finding a sequence of functions $\{K_n(x;x_1,x_2,...,x_n);n\geqslant 1\}$, called the associated causality functions, such that, given a covariant self-adjoint density $\mathscr{H}(x)$, called the potential density, the U matrix can be expressed as follows:

$$U(x)=I+\sum_{n=1}^{\infty}(-i)^n\int dx_1\int dx_2\cdots\int dx_n \times K_n(x;x_1,x_2,...,x_n)\mathscr{H}(x_1)\mathscr{H}(x_2)\cdots\mathscr{H}(x_n). \quad (12)$$

The U matrix, expressed in this way, should satisfy the following conditions required by general field theory:

(1) Unitarity: $U^{\dagger}(x)=U^{-1}(x)$,

(2) Covariance: $U(\Lambda,a)U(x)U^{-1}(\Lambda,a)=U(\Lambda x+a)$, $(\Lambda,a)\in E(1,3)$,

(3) Asymptotic limits: s-$\lim_{x\downarrow}U(x)=I$, s-$\lim_{x\uparrow}U(x)=S$,

where $\lim_{x\downarrow}$ and $\lim_{x\uparrow}$ denote respectively the limits as x tends to the infinite past and future along a timelike direction. Besides, the K functions should have the following property: In the nonrelativistic limits as $c\to\infty$,

$$K_n(x;x_1,x_2,...,x_n)\to\theta(t-t_1)\theta(t-t_2)\cdots\theta(t-t_n) \quad (13)$$

so that the form (12) reduces to (7), where V is formally given by (10). Thus, our theory is causal in the sense that the time-ordered expansion is retained in the nonrelativistic limit. This concept of causality restricts the form of the causality functions, leaving the dependence of the potential density on the free field variables arbitrary to a great extent. This arbitrariness in the dependence of the U matrix on the free field variables makes possible the construction of practical models, giving our approach an advantage over the rather similar approach,[6,8] which uses retarded functions subject to the condition of local commutativity.

II. CAUSALITY FUNCTIONS

In this section we construct all the functions required by causality, as defined in the introduction.

Let n be a positive integer and denote by Γ_n the set of all permutations on $(1,2,...,n)$. Let $p\in\Gamma_n$ and denote by $\nu(p)$ the number of pairs of consecutive numbers which appear in p in order. Formally,

$$\nu(p)=\mathrm{Card}\{(j,j+1);p_j^{-1}<p_{j+1}^{-1};j=1,2,...,n-1\}, \quad (14)$$

where p_j is the number in the jth position and p_j^{-1} is the position of the number j.

As an example, if $n=4$ and $p=(1,4,2,3)$, then the pairs $(1,2)$ and $(2,3)$ appear in p in order, while $(3,4)$ does not appear in order. Hence $\nu(p)=2$.

Now, for any permutation $p\in\Gamma_n$, we define the characteristic function

$$C_p=(-1)^{n-1-\nu(p)}[n-1-\nu(p)]![\nu(p)]! \quad (15)$$

Thus $C_p=-2$ in the above example.

We define now the nonrelativistic phase generating functions

$$\gamma_n(t_1,t_2,...,t_n)=\sum_{p\in\Gamma_n}C_p\theta(t_{p_1}-t_{p_2})\theta(t_{p_2}-t_{p_3})\cdots\theta(t_{p_{n-1}}-t_{p_n}) \quad (16)$$

and the associated nonrelativistic phase generating functions:

$$\tau_n(t;t_1,t_2,...,t_n)=\prod_{i=1}^{n}\theta(t-t_i)\times\gamma_n(t_1,t_2,...,t_n), \quad (17)$$

where $\theta(t)$ is the Heaviside step function, defined to be 1 if $t>0$ and 0 if $t<0$.

We want now to express the effective potential in terms of the potential in the nonrelativistic case. Starting with (7), putting $U(t)=e^{Y(t)}$, and utilizing the standard expansion of $\log(1+z)$, we show in the Appendix the following relation:

$$Y(t)=\sum_{n=1}^{n}\frac{(-i)^n}{n!}\int dt_1\int dt_2\cdots\int dt_n \times\tau_n(t;t_1,t_2,...,t_n)V(t_1)V(t_2)\cdots V(t_n). \quad (18)$$

The effective potential is $F(t)=i(d/dt)Y(t)$ and the phase matrix is $D=-i\lim_{t\to\infty}Y(t)$ or

$$D=\sum_{n=1}^{\infty}\frac{(-i)^{n+1}}{n!}\int dt_1\int dt_2\cdots\int dt_n \times\gamma_n(t_1,t_2,...,t_n)V(t_1)V(t_2)\cdots V(t_n). \quad (19)$$

We turn now to the relativistic case. Clearly, the expressions given in (16) and (17) are not covariant. Thus, we should replace the Heaviside step function θ by a covariant function η. This can be done as follows: If $\theta(t-t')=1$, then $t>t'$, or an event taking place at t is causally affected by an event taking place at t', while if $\theta(t-t')=0$, then $t<t'$, or an event taking place at t is not causally affected by an event taking place at t'.

In the relativistic case, if $x,y\in\mathscr{R}^4$, then we should have $\eta(x-y)=1$ if x depends causally on y and $\eta(x-y)=0$ if x does not depend causally on y. Thus, $\eta(x)=1$ if and only if x is a timelike vector pointing towards the future; otherwise, $\eta(x)=0$. Hence

$$\eta(x)=\theta(x^0)\theta(x^2), \quad (20)$$

where $x=(x^0,x)$ and $x^2=(x^0)^2-|x|^2$.

The function $\eta(x)$ is Lorentz invariant, while $\eta(x-y)$ is Poincaré-invariant. Besides, as $c\to\infty$, $x^2=c^2t^2-|x|^2\to\infty$, and $\eta(x)\to\theta(t)$. Hence, the substitution $\theta(t)\to\eta(x)$ retains relativistic invariance and causality, and preserves the nonrelativistic limits as $c\to\infty$. Therefore, we define the covariant phase generating functions as

$$C_n(x_1,x_2,...,x_n)=\sum_{p\in\Gamma_n}C_p\eta(x_{p_1}-x_{p_2})\eta(x_{p_2}-x_{p_3})\cdots\eta(x_{p_{n-1}}-x_{p_n}) \quad (21)$$

and the associated phase generating functions as

$$T_n(x;x_1,x_2,\ldots,x_n) = \prod_{i=1}^{n} \eta(x-x_i)\times C_n(x_1,x_2,\ldots,x_n). \tag{22}$$

It follows readily from (22) that

$$\lim_{x\downarrow} T_n(x;x_1,x_2,\ldots,x_n) = 0,$$

$$\lim_{x\uparrow} T_n(x;x_1,x_2,\ldots,x_n) = C_n(x_1,x_2,\ldots,x_n). \tag{23}$$

Also, we prove in the Appendix the following important relation:

$$C_n(x_n,x_{n-1},\ldots,x_1) = (-1)^{n-1} C_n(x_1,x_2,\ldots,x_n). \tag{24}$$

The same relation holds for the T functions.

Define now the density

$$Z(x) = \sum_{n=1}^{\infty} \frac{(-i)^n}{n!} \int dx_1 \int dx_2 \cdots \int dx_n \; T_n(x;x_1,x_2,\ldots,x_n)\mathscr{H}(x_1)\mathscr{H}(x_2)\cdots\mathscr{H}(x_n) \tag{25}$$

and the operator

$$Z = \sum_{n=1}^{\infty} \frac{(-i)^n}{n!} \int dx_1 \int dx_2 \cdots \int dx_n \times C_n(x_1,x_2,\ldots,x_n)\mathscr{H}(x_1)\mathscr{H}(x_2)\cdots\mathscr{H}(x_n), \tag{26}$$

where $\mathscr{H}(x)$ is a self-adjoint Poincaré-covariant density, called the potential density, whose form is to be discussed later. Strong convergence in (25) and (26) is assumed. From (23), (25), and (26), the following limits hold formally:

$$\lim_{x\downarrow} Z(x) = 0 \quad \text{and} \quad \lim_{x\uparrow} Z(x) = Z. \tag{27}$$

Also, from (24)–(26) it follows that

$$Z^{\dagger}(x) = -Z(x) \quad \text{and} \quad Z^{\dagger} = -Z. \tag{28}$$

The phase matrix is $D = -iZ$ and the effective potential density is proved in the Appendix to be

$$G(x) = (1/8\pi)\Box^2 iZ(x). \tag{29}$$

We proceed now to construct the U matrix itself. For $1 \leqslant j \leqslant n$, we define the partitioning function:

$$P_{n,j}(x_1,x_2,\ldots,x_n) = \sum \frac{1}{n_1!n_2!\cdots n_j!} C_{n_1}(x_1,\ldots,x_{n_1}) \times C_{n_2}(x_{n_1+1},\ldots,x_{n_1+n_2})\cdots C_{n_j}(x_{n-n_j+1},\ldots,x_n), \tag{30}$$

where the summation is carried over all values $n_1,n_2,\ldots,n_j \geqslant 1$ such that $n_1 + n_2 + \cdots + n_j = n$.

As an illustration, $P_{3,2}(x_1,x_2,x_3) = (1/1!2!)C_1(x_1)C_2(x_2,x_3) + (1/2!1!)C_2(x_1,x_2)C_1(x_3)$. Knowing that $C_1(x) = 1$, we conclude that $P_{3,2}(x_1,x_2,x_3) = \frac{1}{2}C_2(x_2,x_3) + \frac{1}{2}C_2(x_1,x_2)$.

The covariant causality functions are defined as follows:

$$J_n(x_1,x_2,\ldots,x_n) = \sum_{j=1}^{n} \frac{1}{j!} P_{n,j}(x_1,x_2,\ldots,x_n) \tag{31}$$

and the associated covariant causality functions are defined as follows:

$$K_n(x;x_1,x_2,\ldots,x_n) = \prod_{i=1}^{n} \eta(x-x_i)\times J_n(x_1,x_2,\ldots,x_n). \tag{32}$$

It follows readily from (32) that

$$\lim_{x\downarrow} K_n(x;x_1,x_2,\ldots,x_n) = 0,$$

$$\lim_{x\uparrow} K_n(x;x_1,x_2,\ldots,x_n) = J_n(x_1,x_2,\ldots,x_n). \tag{33}$$

Also, one can prove the important relation

$$J_n(x_n,x_{n-1},\ldots,x_1) = J_n(-x_1,-x_2,\ldots,-x_n). \tag{34}$$

Now, defining the U matrix as $U(x) = e^{Z(x)}$ and the S matrix as $S = e^{Z}$, it can be easily shown that the following expansions are valid:

$$U(x) = I + \sum_{n=1}^{\infty} (-i)^n \int dx_1 \int dx_2 \cdots \int dx_n \times K_n(x;x_1,x_2,\ldots,x_n)\mathscr{H}(x_1)\mathscr{H}(x_2)\mathscr{H}(x_n) \tag{35}$$

$$S = I + \sum_{n=1}^{\infty} (-i)^n \int dx_1 \int dx_2 \cdots \int dx_n \times J_n(x_1,x_2,\ldots,x_n)\mathscr{H}(x_1)\mathscr{H}(x_2)\cdots\mathscr{H}(x_n). \tag{36}$$

By construction, U is unitary and Poincaré-covariant, while S is unitary and Poincaré-invariant. From (33), (35), and (36), the following limits are formally satisfied:

$$\lim_{x\downarrow} U(x) = I \quad \text{and} \quad \lim_{x\uparrow} U(x) = S. \tag{37}$$

It is evident that in the nonrelativistic limit, (35) reduces to (7), from which it is derived, utilizing (10) for the definition of V.

III. LOCALITY

In the conventional approaches to quantum field theory,[9] causality and locality mean the same thing. The Poincaré invariance of the time-ordered expansion of the S matrix in the Hamiltonian approach to quantum field theory implies that the interaction Hamiltonian should be a normally ordered polynomial in the field varaibles. But the polynomial form is responsible for all the known divergencies. Thus, the conventional notion of causality is too restrictive.

On the other hand, our concept of causality does not imply any conditions on the potential density, apart from hermiticity and Poincaré covariance. So, extra restrictions should be implied from outside on the potential density in order to meet physical requirements.

First of all, if we want crossing relations to hold, the potential density should have the form

$$\mathscr{H}(x) = \sum_{n=4}^{\infty} \int dy_1 \int dy_2 \cdots \int dy_n \; R{:}\Phi(y_1)\Phi(y_2)\cdots\Phi(y_n) \times \Delta_n(x;y_1,y_2,\ldots,y_n), \tag{38}$$

where R denotes the R product, which is the original product with the terms containing less than two creation and two annihilation operators dropped, and Δ_n is a real distribution which is symmetric in the y variables. The presence of

the R product is required physically, due to the stability of vacuum and one particle states, and is necessary mathematically, because vacuum fluctuations and self-energy terms yield infinite volume divergencies, due to the noncompactness of the space–time continuum. Reality and Poincaré invariance of the Δ distributions imply the following form:

$$\Delta_n(x; y_1, y_2, \ldots, y_n) = \int d^4p_1 \int d^4p_2 \cdots \int d^4p_n \; e^{ip_1(y_1 - x)} \cdots \times e^{ip_n(y_n - x)} \tilde{\Delta}_n(p_1, p_2, \ldots, p_n), \tag{39}$$

where the vertex function $\tilde{\Delta}_n$, which is the Fourier transform of Δ_n, is Lorentz-invariant, symmetric, and satisfies the reality condition

$$\tilde{\Delta}_n^*(p_1, p_2, \ldots, p_n) = \tilde{\Delta}_n(-p_1, -p_2, \ldots, -p_n). \tag{40}$$

Apart from these conditions, no other restrictions are implied by crossing on the vertex function.

We proceed now to relate $\tilde{\Delta}_4$ to classical potential theory. Let $V = V(r)$ be a real potential such that

$$\int_0^\infty r|V(r)|\, dr < +\infty. \tag{41}$$

The Fourier transform $\tilde{V}$ of V thus exists and is given by

$$\tilde{V}(|k|) = \int d^3r\, V(|r|) e^{-ik\cdot r} = \frac{4\pi}{k} \int_0^\infty rV(r) \sin kr\, dr. \tag{42}$$

Defining

$$M(x) = \sum_{n=0}^{\infty} \frac{x^n}{(2n+1)!}, \tag{43}$$

it follows that $\sin x = xM(-x^2)$ and (42) reduces to

$$\tilde{V}(|k|) = 4\pi \int_0^\infty V(r) M(-k^2 r^2)\, r^2\, dr. \tag{44}$$

Putting $z = -4k^2$, we define

$$\bar{V}(z) = 4\pi \int_0^\infty V(r) M(\tfrac{1}{4} z r^2) r^2\, dr, \quad z < 0. \tag{45}$$

We assume now that the function $\bar{V}$ is analytic in a region of the complex plane containing the whole real line and that $\lim_{z \to +\infty} \bar{V}(z) = 0$. The other limit as $z \to -\infty$ follows readily from (42). The necessity of these extra conditions will shortly become clear.

Consider now a theory of two nonrelativistic indistinguishable particles of potential energy operator $\hat{V}$, satisfying

$$\langle r_1 r_2 | \hat{V} | \psi \rangle = V(|r_1 - r_2|) \langle r_1 r_2 | \psi \rangle. \tag{46}$$

It follows after some manipulations that $\hat{V}$ has the following kernel in momentum space:

$$\langle p_1' p_2' | \hat{V} | p_1 p_2 \rangle = \delta^{(3)}(p_1' + p_2' - p_1 - p_2) \{ \bar{V}(-|p_1 - p_2 - p_1' + p_2'|^2) + \bar{V}(-|p_1 - p_2 + p_1' - p_2'|^2) \}. \tag{47}$$

Consider now a theory of a relativistic scalar neutral particle of potential density

$$\mathscr{H}(x) = \int dy_1\, dy_2\, dy_3\, dy_4\, R{:}\Phi(y_1)\Phi(y_2)\Phi(y_3)\Phi(y_4){:} \times \Delta_4(x; y_1, y_2, y_3, y_4) \tag{48}$$

and define the noncovariant relativistic state

$$|p_1 p_2\rangle = (2E_{p_1} 2E_{p_2})^{-1/2} a^\dagger(p_1) a^\dagger(p_2)|0\rangle, \tag{49}$$

where $a^\dagger$ is the creation operator and $E_p = (m^2 + p^2)^{1/2}$. It follows from (48) and (49) that

$$\left\langle p_1' p_2' \left| \int \mathscr{H}(0,x)\, d^3x \right| p_1 p_2 \right\rangle = 6(E_{p_1} E_{p_2} E_{p_1'} E_{p_2'})^{-1/2} \times \delta^{(3)}(p_1' + p_2' - p_1 - p_2) \tilde{\Delta}_4(-p_1', p_1, -p_2', p_2). \tag{50}$$

Let us put

$$\tilde{\Delta}_4(p_1, p_2, p_3, p_4) = (m^2/6) \times \{ \bar{V}[(p_1 + p_2 - p_3 - p_4)^2] + \bar{V}[(p_1 - p_2 + p_3 - p_4)^2] + \bar{V}[(p_1 - p_2 - p_3 + p_4)^2] \}. \tag{51}$$

It is obvious that (51) requires $\bar{V}(z)$ to be defined for all real values of z. If we use the metric $(c^2, -1, -1, -1)$ and define the 4 momentum $p = (E_p, p)$, it follows that $E_p = (m^2 + p^2/c^2)^{1/2}$. Making use of the assumption $\lim_{z \to +\infty} \bar{V}(z) = 0$, and using (47), (50), and (51), it follows that

$$\lim_{c \to \infty} \left\langle p_1' p_2' \left| \int \mathscr{H}(0,x)\, d^3x \right| p_1 p_2 \right\rangle = \langle p_1' p_2' | \hat{V} | p_1 p_2 \rangle, \tag{52}$$

Thus, if $\tilde{\Delta}_4$ is related to classical potential theory, it should have the form (51).

The higher order distributions cannot be directly related to classical potentials. However, in view of (51), we may introduce the following definition of locality:

A density $\mathscr{D}(x)$ is simply local if and only if there exist a number $n \in \mathscr{N}$, $n \geqslant 4$, and a number $m \in \mathscr{N}$, $m \leqslant n/2$, together with a function $F(z)$ which is analytic in a region containing the whole real axis, such that

$$\mathscr{D}(x) = \int dy_1 \int dy_2 \cdots \int dy_n\; R{:}\Phi(y_1)\Phi(y_2)\cdots\Phi(y_n){:} \times \Delta_F^{n,m}(x; y_1, y_2, \ldots, y_n), \tag{53}$$

where

$$\Delta_F^{n,m}(x; y_1, y_2, \ldots, y_n) = \int dp_1 \int dp_2 \cdots \int dp_n\; e^{ip_1(y_1 - x)} \cdots \times e^{ip_n(y_n - x)} \tilde{\Delta}_F^{n,m}(p_1, p_2, \ldots, p_n) \tag{54}$$

and

$$\tilde{\Delta}_F^{n,m}(p_1, p_2, \ldots, p_n) = F\{(p_1 + p_2 + \cdots + p_{n-m} - p_{n-m+1} - \cdots - p_n)^2\} + \text{distinct permutations}. \tag{55}$$

A density $\mathscr{D}(x)$ is local if and only if it is a linear combination of simply local densities.

It is clear from the above definition that our notion of locality implies Poincaré covariance, annihilation of vacuum and one particle states, crossing relations and discrete

symmetries, as will be shown later.

We are now in a position to state the following assumption:

The potential density which is responsible for the interaction is a self-adjoint local density, such that the integrals (25) *and* (26) *and the limits in* (27) *converge strongly in the Fock space.*

The above assumption of convergence implies certain restrictions on the form of the vertex functions used in constructing the potential densities. These restrictions may be looked for for each model on its own, and we do not present here a general solution of this problem.

IV. DISCRETE SYMMETRIES

We discuss briefly the effect of charge conjugation, space inversion and time reversal on our theory.

The Δ distributions are invariant under charge conjugation because they are scalars. Since the vertex functions are Lorentz scalars, they are also invariant under space inversion. Hence, the same property is shared by the Δ distributions themselves. Thus

$$\Delta_n(x^P; y_1^P, y_2^P, \ldots, y_n^P) = \Delta_n(x; y_1, y_2, \ldots, y_n), \tag{56}$$

where $x^P = (x^0, -x)$.

The time reversal transformation is antiunitary. So, to prove the invariance of the Δ distributions under time reversal, we need to prove the invariance of the vertex functions under time reversal and complex conjugation. Both conditions follow from the fact that the vertex functions are Lorentz scalars, taking into account the reality condition (40). Thus

$$\Delta_n(x^T; y_1^T, y_2^T, \ldots, y_n^T) = \Delta_n(x; y_1, y_2, \ldots, y_n), \tag{57}$$

where $x^T = (-x^0, x)$.

We conclude from the above discussion that the invariance of the potential density under a discrete transformation is determined by the parity of the field and the number of field variables in each term, and has nothing to do with the form of the vertex function used.

It is trivial to show that the invariance of the potential density under charge conjugation implies the invariance of the U matrix under the same transformation. A similar result holds for space inversion: If $\mathscr{H}(x) \to \mathscr{H}(x^P)$, then $U(x) \to U(x^P)$, since $\eta(x^P) = \eta(x)$, while the S matrix remains invariant.

Time reversal needs some care. If $\mathscr{T}$ denotes the antiunitary time reversal operator, we adopt the following definition[10]:

$$\text{If } |\text{out}\rangle = S\,|\text{in}\rangle, \text{ then } \mathscr{T}|\text{in}\rangle = S\mathscr{T}|\text{out}\rangle \tag{58}$$

provided that S is invariant under time reversal. (58) can be solved to give

$$\mathscr{T} S \mathscr{T}^{-1} = S^{-1}. \tag{59}$$

Assume now that $\mathscr{T}\mathscr{H}(x)\mathscr{T}^{-1} = \mathscr{H}(x^T)$. Then from (36) we get

$$\mathscr{T} S \mathscr{T}^{-1} = I + \sum_{n=1}^{\infty} i^n \int dx_1 \int dx_2 \cdots \int dx_n\, J_n(x_1, x_2, \ldots, x_n) \times \mathscr{H}(x_1^T)\mathscr{H}(x_2^T)\cdots\mathscr{H}(x_n^T)$$

$$= I + \sum_{n=1}^{\infty} i^n \int dx_1 \int dx_2 \cdots \times \int dx_n\, J_n(-x_1, -x_2, \ldots, -x_n) \times \mathscr{H}(x_1)\mathscr{H}(x_2)\cdots\mathscr{H}(x_n). \tag{60}$$

To derive (60), we have used the relation $\eta(x^T) = \eta(-x)$. Now, from (34) and (60) we get $\mathscr{T} S \mathscr{T}^{-1} = S^{\dagger}$. Since S is unitary, (59) follows.

V. CONCLUSION AND DISCUSSION

The theory presented in the previous sections differs from conventional theories in that, instead of proposing covariant equivalents of nonrelativistic differential equations, we propose covariant equivalents of the general solutions of those equations. This approach may look unfamiliar, but it is a way of escaping the difficulties which face conventional quantum field theory.

An important feature of our approach is the independence of causality and locality of each other. The causality functions in this approach are unique, while there is a great deal of arbitrariness in the concept of locality. The restriction we impose on the vertex function of being analytic in a region containing the whole real line is necessary to have a unique extension of the function $\bar{V}(z)$ as defined in (45) when we pass from the nonrelativistic case, where $\bar{V}(z)$ is defined for $z \leqslant 0$, to the relativistic case, where $\bar{V}(z)$ should be defined for all $z \in \mathscr{R}$.

In the literature,[11] the divergencies of quantum field theory are classified in four categories: infinite-volume, infrared, ultraviolet, and infinite-number divergencies. The first three divergencies have the same origin. If $\Delta \subseteq \mathscr{R}^4$ is a measurable set, then the definability of (35) and (36) requires that $\int_\Delta \mathscr{H}(x)\,dx$ be an operator. If this is true, then all the terms in the expansion are finite. If this is not true, then $\int_\Delta \mathscr{H}(x)\,dx$ is a sesquilinear form only, and $\int_{\Delta_1} dx \int_{\Delta_2} dy\, \mathscr{H}(x)\mathscr{H}(y)$ is not defined. Thus, either all terms are finite, or divergencies start to appear at the second order of perturbation. Finiteness of the second-order terms in our theory can be achieved by the use of the $\mathscr{R}$ product, the analyticity of the vertex functions for all real values, and a suitable choice of upper bounds on the behavior of the vertex functions at large values. Hence, we can make all the terms in the perturbation series finite. However, the problem of the infinite number divergence, which deals with the convergence of the perturbation series itself, is left open.

APPENDIX
Proof of (18)

To prove Eq. (18), we introduce first the symbols Ω and Υ defined as follows:

$$\Omega(n) = \{(k, k+1);\ k = 1,2,\ldots,n-1\},$$

$$\Omega_j(n) = \{(j+k, j+k+1);\ k = 1,2,\ldots,n-1\},$$

$$\Omega(n_1, n_2, \ldots, n_l) = \Omega(n_1) \cup \Omega_{n_1}(n_2) \cup \cdots \cup \Omega_{n_1 + \cdots + n_{l-1}}(n),$$

$$\Upsilon(n_1, n_2, \ldots, n_l) = \prod_{(j,k) \in \Omega(n_1, n_2, \ldots, n_l)} \theta(t_j - t_k).$$

Thus

$$\Omega(2,3,1) = \Omega(2)\cup\Omega_2(3)\cup\Omega_5(1)$$
$$\{(1,2)\}\cup\{(3,4),(4,5)\}\cup\phi$$
$$= \{(1,2),(3,4),(4,5)\}$$

and

$$\Upsilon(2,3,1) = \theta(t_1 - t_2)\theta(t_3,t_4)\theta(t_4,t_5).$$

Finally, if $p \in \Gamma_n$ is a permutation on $(1,2,...,n)$, we then define:

$$\mathscr{S}(p) = \{(j,j+1); p_j^{-1} < p_{j+1}^{-1}; j = 1,2,...,n-1\}.$$

We proceed now to prove (18). We may write

$$Y(t) = \sum_{n=1}^{\infty} (-i)^n \int_{-\infty}^{t} dt_1 \int_{-\infty}^{t} dt_2 \cdots \int_{-\infty}^{t} dt_n \times \left\{ \sum_{j=1}^{n} (-1)^{j-1} \sum \Upsilon(n_1,n_2,...,n_j) \right\} \times V(t_1)V(t_2)\cdots V(t_n), \tag{A1}$$

where, and throughout the proof, we have

$$n_1 + n_2 + \cdots + n_j = n \text{ with } n_1,n_2,...,n_j \geqslant 1.$$

Let now

$$E_n = \sum_{j=1}^{n} \frac{(-1)^{j-1}}{j} \sum \Upsilon(n_1,n_2,...,n_j). \tag{A2}$$

We may prove by induction that

$$E_n = \frac{1}{n!} \sum_{p\in\Gamma_n} \alpha_p \prod_{j=1}^{n-1} \theta(t_{p_j} - j_{p_{j+1}}), \tag{A3}$$

where

$$\alpha_p = n! \sum_{j=1}^{n} \frac{(-1)^{j-1}}{j} \sum_{\Omega(n_1,n_2,...,n_j)\subseteq\mathscr{S}(p)} 1. \tag{A4}$$

But Card $\Omega(n_1,n_2,...,n_j) = n - j$ and, from (14), Card $\mathscr{S}(p) = \nu(p)$. Thus

$$\sum_{\Omega(n_1,n_2,...,n_j)\subseteq\mathscr{S}(p)} 1 = C_{\nu(p)}^{n-j} \tag{A5}$$

because the number of subsets of cardinality m of a set of cardinality n is C_n^m. Hence, from (A4) and (A5) we get

$$\alpha_p = n! \sum_{j=1}^{n} \frac{(-1)^{j-1}}{j} C_{\nu(p)}^{n-j} = n! \sum_{j=0}^{n-1} \frac{(-1)^{n-j-1}}{n-j} C_{\nu(p)}^{j}. \tag{A6}$$

But $\nu(p) \leqslant n-1$, and if $j > \nu(p)$, then $C_{\nu(p)}^{j} = 0$; hence (A6) can be put in the form

$$\alpha_p = -n! \sum_{j=0}^{\nu(p)} C_{\nu(p)}^{j} \frac{(-1)^{n-j}}{n-j} = (-1)^{n-1-\nu(p)}[n-1-\nu(p)]![\nu(p)]! \tag{A7}$$

Thus, $\alpha_p = C_p$, which is defined in (15), and (A3), (A7) together with (16) imply

$$E_n = (1/n!)\gamma_n(t_1,t_2,...,t_n), \tag{A8}$$

from which (18) follows.

Proof of (24)

Let $p \in \Gamma_n$ and define $\tilde{p} \in \Gamma_n$ as follows: $\tilde{p}_j = n + 1 - p_j, j = 1,2,...,n$. Noting that the correspondence $p \to \tilde{p}$ is one-to-one and that $\tilde{\tilde{p}} = p$, we get

$$C_n(x_n,x_{n-1},...,x_1) = \sum_{p\in\Gamma_n} C_p \eta(x_{\tilde{p}_1} - x_{\tilde{p}_2})\cdots\eta(x_{\tilde{p}_{n-1}} - x_{\tilde{p}_n}) = \sum_{p\in\Gamma_n} C_{\tilde{p}} \eta(x_{p_1} - x_{p_2})\cdots\eta(x_{p_{n-1}} - x_{p_n}). \tag{A9}$$

Now, $\tilde{p}_j^{-1} = p_{n+1-j}^{-1}$; hence

$$\nu(\tilde{p}) = \text{Card}\{(jj+1);\ p_{n-j+1}^{-1} < p_{n-j}^{-1};\ j = 1,2,...,n-1\} = \text{Card}\{(n-j,n-j+1);\ p_j^{-1} > p_{j+1}^{-1};\ j = 1,2,...,n-1\} = n - 1 - \nu(p). \tag{A10}$$

From (A10) and (15) we get $C_{\tilde{p}} = (-1)^{n-1} C_p$, which when applied to (A9) yields (24).

Proof of (29)

To get Eq. (29), we have to solve the equation

$$Z(x) = -i\int \eta(x-y)G(y)\,dy. \tag{A11}$$

Multiplying (A11) by $e^{-\omega x_0}e^{-ipx}$, where $\omega > 0$ and $p \in \mathscr{R}^4$, and integrating over the x variables, we get

$$\int Z(x)e^{-\omega x_0}e^{-ipx}\,dx = -i\tilde{\eta}_\omega(p)\int G(y)e^{-\omega y_0}e^{-ipy}\,dy, \tag{A12}$$

where

$$\tilde{\eta}_\omega(p) = \int \eta(x)e^{-\omega x_0}e^{-ipx}\,dx = 8\pi/[(\omega + iE)^2 + p^2]^2. \tag{A13}$$

Using (A13) and taking the inverse Fourier transform of (A12) after dividing it by $-i\tilde{\eta}_\omega(p)$, we get

$$G(y)e^{-\omega y_0} = (1/8\pi)\int [(\omega + iE)^2 + p^2]^2 iZ(x)e^{-i\omega x_0}e^{ip(y-x)}\,dx\,\bar{d}p$$

or

$$G(x) = (1/8\pi)\int [(\omega + -E)^2 + p^2]^2 \times e^{\omega(x_0 - y_0)}e^{ip(x-y)}iZ(y)\,dy\,\bar{d}p = (1/8\pi)\Box^2\int iZ(y)e^{\omega(x_0-y_0)}e^{ip(x-y)}\,dy\,\bar{d}p = (1/8\pi)\Box^2 iZ(x). \tag{A14}$$

Thus, the proof is complete.

[1] M. O. Farrukh, "Constructing quantum fields in a Fock space using a new picture of quantum mechanics," J. Math. Phys. **19**, 2332 (1978).

[2] $\hbar = 1$ throughout this paper.
[3] There is a sign difference in the definition of the effective potential and the effective potential density between this paper and Ref. 1.
[4] In this paper we use Φ and Ψ instead of Φ_0 and Φ used in Ref. 1.
[5] S. Weinberg, "Feynman rules for any spin," Phys. Rev. **133**, B1318 (1964).
[6] H. Lehmann, K. Symanzik, and W. Zimmermann, "On the Formulation of Quantized Field Theories—II," Nuovo Cimento **6**(2), (1957).
[7] R. F. Streater and A. S. Wightman, *PCT, Spin and Statistics and All That* (Benjamin, New York, Amsterdam, 1964).
[8] V. Glaser, H. Lehmann, and W. Zimmermann, "Field Operators and Retarted Functions," Nuovo Cimento **6**(5), (1957).
[9] P. Roman, *Introduction to Quantum Field Theory* (Wiley, New York, 1969).
[10] W. R. Frazer, *Elementary Particles* (Prentice-Hall, Englewood Cliffs, NJ, 1966).
[11] J. Glimm and A. Jaffe, "Quantum Field Theory Models," in *Statistical Mechanics and Quantum Field Theory*, edited by C. DeWitt and R. Stora (Les Houches, 1971) (Gordon and Breach, New York, 1971).

Nonlinear Schrödinger-type field equation for the description of dissipative systems. I. Derivation of the nonlinear field equation and one-dimensional example

D. Schuch,[a] K.-M. Chung, and H. Hartmann
Institut für Physikalische und Theoretische Chemie der Johann Wolfgang Goethe Universität, Robert-Mayer-Str. 11, D-6000 Frankfurt (Main), Federal Republic of Germany

(Received 10 August 1982; accepted for publication 3 December 1982)

Based on three principles taken from empiricism we establish a new nonlinear field theory which allows us to describe frictional effects in dissipative systems with the aid of a Schrödinger-type field equation with logarithmic nonlinearity. This nonlinear field equation corresponds to the classical Langevin equation and can be interpreted in different ways, taking the view of classical undulatory theory or probabilistic theories, respectively. Because of formal similarities, calculations can be performed independently of the accepted interpretation; only the occurring quantities have to be provided with the corresponding meaning. As an example, the nonlinear field equation for the damped harmonic oscillator is solved exactly. The solutions, a wavefunction and a wavepacketlike solution, a solution with Gaussian shape, exhibit reasonable properties, contain the correct reduced frequency $\Omega = (\omega_0^2 - \gamma^2/4)^{1/2}$ and are different from the solutions of the undamped problem. The properties of our nonlinear friction term are discussed and compared with those of similar nonlinear friction terms used by other authors. Finally, we derive a nonlinear friction term equivalent to our logarithmic nonlinearity but involving only combinations of suitably defined operators of position and linear momentum and mean values of these operators.

PACS numbers: 11.10.Lm

I. INTRODUCTION

In this work we want to develop a new theory for the description of dissipative systems. In particular, the consideration of velocity dependent frictional forces will be investigated.

In classical particle mechanics Langevin's equation

$$m\dot{\mathbf{v}} = -\nabla V - m\gamma\mathbf{v} + \mathcal{F}(t) \tag{1}$$

is assumed to give an adequate phenomenological description of a damping process due to forces which summarize collisional effects slowing down a particle during its movement.

—the first term on the right side is a conservative external force which can be derived from a potential V.

—the second term is a dissipative frictional force proportional to velocity, where γ is a friction constant. On this term we will concentrate in the following.

—finally, there is a rapidly fluctuating force $\mathcal{F}(t)$ which is purely random and allows only statistical statements. (For a classical statistical treatment of this force cf. e.g., Ref. 1.)

A field-theoretical description of dissipative systems, however, is not as easy as it might seem to be at first sight.

Especially, one has to proceed with caution if somehow quantum theoretical aspects will also be taken into account. Actually, there exist several approaches to derive a quantum theoretical description of dissipation, but most of these suffer from serious difficulties.

The two major approaches use either a linear, but time-dependent Hamiltonian, or introduce a nonlinear damping potential into the Schrödinger equation.

The first group, due to Kanai[2] and Caldirola[3] introduces an explicitly time-dependent Hamiltonian operator, which however entails some essential shortcomings discussed in a number of critical comments (see, e.g., Refs. 4 and 5), whereof the most serious shortcoming seems to be violation of the uncertainty principle. Modifications of the Kanai Hamiltonian try to get rid of these difficulties but need concepts like variable mass particles[6] for this purpose.

In order to avoid difficulties with nonphysical Hamiltonians, the representatives of the second approach try to introduce additional friction terms into the linear Schrödinger equation by various methods. As a consequence of these methods, nonlinear Schrödinger equations are obtained, where the first nonlinear approach is due to Kostin.[7,8] He discovered a nonlinear Schrödinger-type equation (NLSE) with logarithmic nonlinearity which has been rederived independently by many other authors using different methods,[9–15] or extended by additional friction terms (Cho[16]). Another type of NLSE's containing quantum mechanical expectation values has been investigated by Albrecht[17] and Hasse and Süssmann.[18,19]

However, all the NLSE's proposed hitherto suffer from the weakness that the solutions of these equations show at least in one respect unphysical behavior. So the existence of a complete set of stationary energy states, which are in most cases even identical with the solutions of the corresponding undamped problem is difficult to interpret. Furthermore, with the exception of Hasse's NLSE the solutions of the damped harmonic oscillator do not contain the correct reduced frequency.

A difficulty NLSE's have in common is that the superposition principle, which is valid in "orthodox" *linear* quan-

[a] From a dissertation submitted to the J. W. Goethe Universität Frankfurt by Dieter Schuch in partial fulfillment of the requirements for the Ph.D. degree in science.

 0022-2488/83/061652-09$02.50 © 1983 American Institute of Physics

tum *mechanics*, now in general no longer applies. But this fact by itself should not exclude that these NLSE's might be interpreted as quantum *theoretical* equations in the sense of a probabilistic interpretation of the wavefunction (if we want to restrict the term quantum *mechanics* to the *linear* theory), if such an equation is able to describe the dynamics of physical systems without inconsistencies and in accordance with empiricism, also on a microscopic level. In order to avoid terminological confusion, perhaps it might be even advantageous in this case to replace the term *quantum* theory by some other more appropriate expression.

As is known, many phenomena which within the scope of a corpuscular consideration cannot be explained until quantization can be understood within the scope of an undulatory consideration already on classical level. A prominent example is the treatment of the linear Schrödinger equation as a classical field equation, which is discussed in detail, e.g., by Hund.[20] Interpreting the Schrödinger field as a classical field of matter, physical qualities can be associated with this field with the aid of properly defined classical mean values and operators. Although there exists a striking similarity between these operators and mean values and the respective quantum mechanical operators and expectation values, this similarity is only a *formal* one, as the meaning and interpretation of these quantities is quite different in classical and quantum theories.

So, an undulatory theory for the description of dissipative systems does not necessarily have to be a quantum theoretical description. If the interpretation of a field equation (and the constituting quantities) is not compatible with fundamental principles of quantum theory, nevertheless, such an undulatory theory may yield interesting and valuable results already on the classical level.[20–22]

As already mentioned, hitherto, attempts to obtain a field equation for the description of dissipative systems by the use of methods known from the "derivation" of the linear Schrödinger equation were not yet entirely successful.

Madelung and Mrowka[23,24] established a method different from those used by the authors cited above to derive a field equation, the *linear* Schrödinger equation, where classical particle mechanics enters only in Newton's formulation.

Now, we developed a new method, similar to that of Madelung and Mrowka, to derive a nonlinear field equation corresponding to the Langevin equation, which is capable to describe dissipative systems. As there does not yet exist any pendant to our NLSE, *a priori* we have to take into account different possibilities of interpretation, classical undulatory or probabilistic ones, respectively.

1. As is discussed in detail for the linear Schrödinger equation by several authors,[20–22] also our NLSE may be interpreted as a classical field equation. Moreover, the classical field equation might be subjected to field quantization, as described by Heisenberg[25] for the linear case. However, because of the nonlinearity of our field equation an adequate quantization method should be different from the one used in the linear case. Questions arising in this connection shall not be discussed in this work.

2. A second, quite different interpretation is a probabilistic interpretation due to Fényes,[26] who applied this interpretation on the linear SE. His stochastic mechanical interpretation, which is different from the usual quantum theoretical probabilistic interpretation due to Born, regards wave mechanical processes as special Markoff processes and therefore Ψ as the probability amplitude adjoined to this Markov process and $\rho = \Psi\Psi^*$ as corresponding probability density. In principle, a corresponding stochastic mechanical interpretation, additionally taking into account the dissipative character of processes, should be considered also with respect to our NLSE. But, as Nelson[27,28] states for the linear case, at the moment it does not seem possible to perform an experiment yielding certainty as to which of the two probabilistic interpretations is the right one; the same might apply to our nonlinear case. In this work, however, this interpretation shall not be discussed in further detail.

3. As well as Madelung and Mrowka interpreted their linear field equation as quantum mechanical equation, we might try to regard our NLSE as quantum theoretical equation of motion, but now for dissipative systems.

The fact that with respect to certain concepts like superposition principle obviously differences between a nonlinear theory and usual linear quantum mechanics might exist was already mentioned. However, it can be shown that our nonlinear theory pays regard to principles like the validity of the uncertainty relation and resulting consequences like finite (positive) ground-state energies, which naturally has to be warranted within the framework of a quantum theoretical interpretation. But, these results by themselves are not necessarily a proof for the correctness of a quantum theoretical interpretation, as analogous results also can occur in the framework of a classical undulatory theory.[20]

As we could not find an absolutely convincing argument favouring one of the interpretations, and, as we do not know an *experimentum crucis* to decide with certainty which possibility is the only right one, at the moment it only remains to indicate the different possible interpretations, noting that in each case one has to investigate whether the chosen interpretation leads to inconsistencies or not.

Because of the *formal* similarity between classical undulatory theory and quantum theory, however, calculations can be performed without the necessity of deciding previously in favor of one of the possible interpretations. Only the quantities occurring have to be provided with the meaning corresponding to the accepted interpretation. Therefore, undermentioned terms like wavefunction, operator, mean value, commutator, etc., shall be used synonymously for all possible interpretations and have to be understood in each case in the framework of the chosen interpretation.

For *formal* reasons, namely to facilitate comparison of the results of our nonlinear field theory with results of other authors or results obtained in linear quantum mechanics, we will use a form where the quantity $\hbar$ ($\hbar = h/2\pi$ with h being Planck's constant) appears in our NLSE. In the course of the development of our field theory we will show that it is as well possible to ignore the existence of $\hbar$ and to avoid the appearance of quantities like m and q, which usually denote corpuscular properties like mass or charge of a particle, leading to an NLSE which can be interpreted completely within the framework of classical field theory.

Comparing the results of our nonlinear field theory with those obtained by the other authors using NLSE's, we find that the solutions of our NLSE no longer show the unphysical behavior mentioned above.

The properties of the solutions and corresponding density functions are discussed as well as the mean values of physically interesting quantities. Finally, the properties of our friction operator are investigated and compared with the most relevant operators of this kind known up till now. In the course of this comparison, it will be possible to derive also a nonlinear friction operator containing mean values of coordinate and linear momentum equivalent (at least in the case under consideration) to our logarithmic operator and similar to Hasse's friction operator.[18]

II. DERIVATION OF THE NONLINEAR FIELD EQUATION

As already mentioned, our method of deriving a field equation for the description of dissipative systems shows some similarities to a method established by Madelung[23] and Mrowka[24] applicable to conservative esystems.

Axiomatic basis of both methods are three general principles taken from empiricism:

1. Uncertainty principle or complementarity, respectively,

2. the fact that interference phenomena show up in experiments with material systems, and

3. correspondence principle or Ehrenfest's theorem, respectively.

The mathematical form of a theory taking into account this empirical knowledge follows from the structure of these principles. Due to the uncertainty principle, exact initial conditions in the sense of classical particle mechanics cannot be given. Therefore, it is only possible to develop a theory where mean values $\bar{Q}$ of quantities Q are determined with the aid of a distribution function ρ. According to the interpretation accepted, ρ is directly connected with the material density of the field or ρ is a probability density, respectively.

Paying regard to the irreversible character of dissipative processes, ρ is assumed to fulfill the Fokker–Planck equation (FPE)

$$\dot{\rho} + \operatorname{div}(\rho \mathbf{v}) - D\Delta\rho = 0. \tag{2}$$

This is an essential difference between our theory and the theory of Madelung and Mrowka, as the distribution function ρ in their nondissipative theory is supposed to fulfill the continuity equation (CE)

$$\dot{\rho} + \operatorname{div}(\rho \mathbf{v}) = 0. \tag{3}$$

Bearing in mind our second axiom, in analogy to optics (where intensity is a quadratic function of the amplitudes) the bilinear ansatz $\rho = \alpha\cdot\beta \geqslant 0$ with the field amplitudes $\alpha(\mathbf{r},t)$ and $\beta(\mathbf{r},t)$ shall be used. With these amplitudes also the flux $\mathbf{j} = \rho\mathbf{v}$ can be defined (in absence of magnetic fields) by a bilinear form,

$$\mathbf{j} = \rho\mathbf{v} = C(\beta\nabla\alpha - \alpha\nabla\beta), \tag{4}$$

with

$$\mathbf{v} = C\nabla \ln(\alpha/\beta) = C(\nabla\alpha/\alpha - \nabla\beta/\beta),$$

where C is a constant.

Finally, the third axiom shall be taken into consideration in the form of Ehrenfest's theorem, i.e., at an average the classical equations of motion are valid:

$$\overline{m\frac{d^2}{dt^2}\mathbf{r}} = \overline{m\frac{d}{dt}\mathbf{v}} = m\frac{d}{dt}\int \rho\mathbf{v}\, d\tau = m\int \frac{\partial}{\partial t}\mathbf{j}\, d\tau = \bar{\mathbf{F}} \tag{5}$$

with $d\tau$ = element of volume. Inserting the preceding definitions into Eq. (5), we obtain

$$\bar{\mathbf{F}} = 2mC\int (\dot{\beta}\nabla\alpha - \dot{\alpha}\nabla\beta)\, d\tau. \tag{6}$$

In the nondissipative case by the help of these definitions the CE (3) can be separated into two equations containing only α or β, respectively, if a separation function f is introduced, which is independent of α and β and does not contain any differential operators.

Inserting the definitions of ρ and $\mathbf{v}$ into the FPE (2), however, in this case an additional coupling of α- and β-dependent terms occurs due to the expression

$$-D\Delta\rho = -D(\alpha\Delta\beta + \beta\Delta\alpha + 2\nabla\alpha\cdot\nabla\beta), \tag{7}$$

which does not allow separation in general.

To reach separability, the coupling term has to fulfill the condition

$$-D\Delta\rho/\rho = D\Delta(\alpha\cdot\beta)/(\alpha\cdot\beta) = G(\alpha\cdot\beta) = F_1(\alpha) + F_2(\beta). \tag{8}$$

As the logarithm is a function possessing this property, we assume the following ansatz:

$$-D\Delta\rho/\rho = \Gamma(\ln\rho + \tilde{Z}), \tag{9}$$

where Γ is constant and $\tilde{Z}$ is a function independent of α and β, with $\tilde{Z} = Z + Z^* = 2Z_R$ and $Z = Z_R + iZ_I$. The effect of Z will be discussed later.

So we obtain as condition of separability

$$-D\Delta\rho = \rho\Gamma(\ln\rho + 2Z_R). \tag{10}$$

Inserting (10) in (2), we get

$$[\dot{\alpha} + C\Delta\alpha + \Gamma(\ln\alpha + Z)\alpha]/\alpha + [\dot{\beta} - C\Delta\beta + \Gamma(\ln\beta + Z^*)\beta]/\beta = H_1(\alpha) + H_2(\beta) = 0, \tag{11}$$

and with $H_1(\alpha) = -f$ and $H_2(\beta) = f$, respectively, finally we find

$$\dot{\alpha} + C\Delta\alpha + \Gamma(\ln\alpha + Z)\alpha + f\alpha = 0 \tag{12a}$$

$$\dot{\beta} - C\Delta\beta + \Gamma(\ln\beta + Z^*)\beta - f\beta = 0 \tag{12b}$$

where f is the previously mentioned separation function.

So, we achieve separability at the expense of a restrictive condition for the permissible functions $\rho = \alpha\cdot\beta$. Before continuing with the derivation of our theory, we shall first investigate the consequences condition (10) entails for the functions α and β, where the choice $\beta = \alpha^*$ guarantees positive real ρ.

In order to formally simplify the following calculations and to make possible a comparison of our results with those of other authors (where these authors accept a quantum

theoretical interpretation of their NLSE's), we will use normalized functions, although this is not *a priori* necessary.

Using the common ansatz $\alpha = N\exp(iS + R)$, $\beta = N\exp(-iS + R)$ and inserting $\rho = N^2\exp(2R)$ in (10), we obtain a nonlinear differential equation for R,

$$\Delta R + 2(\nabla R)^2 + (\Gamma/D)R + (\Gamma/D)(\ln N + Z_R) = 0, \quad (13)$$

which implies a condition for the real part of α, independent of what the imaginary part looks like.

In one dimension the ansatz $R = a + bx + cx^2$ leads to

$$R = -(\Gamma/8D)[x - (4D/\Gamma)b]^2 + \tfrac{1}{4} - \ln N - Z_R. \quad (14)$$

Now the influence of Z_R becomes obvious. Choosing $Z_R = 0$, this leads to contradiction to the assumption of normalized functions. So Z_R should at least compensate $\ln N$. Whether we also include the constant term $\frac{1}{4}$ into Z_R or not does not effect $|\alpha| = N\exp R$, but only changes the values of N and the constant term in R, respectively. Additional constant factors which might appear in order to provide α with a concrete physical meaning (e.g., α as amplitude of a field of matter), obviously also may be taken into account by Z_R, which makes the field equation independent of factors of that kind. Choosing $Z_R = -\ln N + \frac{1}{4}$, $N\exp R$ becomes a Gaussian function with

$$R = -P\tilde{x}^2, \quad (15)$$

where

$$P = \Gamma/8D \quad \text{and} \quad \tilde{x} = x - (4D/\Gamma)/b. \quad (16)$$

In three dimensions this ansatz can be generalized to

$$R = -(P_x\tilde{x}^2 + P_y\tilde{y}^2 + P_z\tilde{z}^2) \quad (15')$$

with

$$P_{x_i} = \Gamma/8D_{x_i}, \quad \tilde{x}_i = [x_i - (4D_{x_i}/\Gamma)b_i], \quad (16')$$

where the b_i may be constants or functions depending on time only. Z_R then has the form $Z_R = -\ln N + d/4 = -\ln N + \frac{3}{4}$, with d = dimension of the problem.

It shall be stressed that besides (15) and (15'), respectively, there might be other solutions of (13) (e.g., $R = 0$, and $Z_R = -\ln N$), which could entail interesting consequences. For the discussion of the following example and the comparison with the works of other authors, however, solution (15) is sufficient. Therefore, in the following we will restrict ourselves to the type of solution functions obeying Eqs. (13) and (15).

Knowing now that it is possible to find functions α allowing to fulfill condition (10), we insert this condition into the FPE (2), getting

$$\dot{\alpha} + C\Delta\alpha + \Gamma(\ln\alpha + Z)\alpha + f\alpha = 0. \quad (17)$$

The conjugate complex function β obeys the corresponding conjugate complex equation.

Inserting these equations into the averaged Eq. (6), after some arithmetics we obtain

$$\overline{\mathbf{F}} = \int \rho[-\nabla(2mCf) - m\Gamma\nabla\ln(\alpha/\beta)]\,d\tau - \hbar\Gamma\int\rho(\nabla Z_I)\,d\tau. \quad (18)$$

Using Def. (4) of $\mathbf{v}$ and identifying $2mCf = V$, where V is the potential of a conservative force, the third axiom can be satisfied if we identity $\Gamma \equiv \gamma$ and choose $\nabla Z_I = 0$, neglecting the fluctuating stochastic force $\mathcal{F}(t)$. So we attain the averaged Langevin equation without stochastic force:

$$\overline{\mathbf{F}} = \int\rho(-\nabla V)\,d\tau - \int\rho(-m\gamma\mathbf{v})\,d\tau = \overline{-\nabla V} - \overline{m\gamma\mathbf{v}}. \quad (19)$$

In addition, also the random force $\mathcal{F}(t)$ can be taken into consideration through $\nabla Z_I = -(1/\hbar\gamma)\mathcal{F}(t)$, leading to an extra term in our NLSE. The influence of this extra term and problems in connection with the consideration of the stochastic force will be mentioned in the last section.

Replacing α by Ψ, Eq. (17) attains the form of a nonlinear Schrödinger-type equation (NLSE)

$$i\hbar\dot{\Psi} = [-(\hbar^2/2m)\Delta + V + \gamma(\hbar/i) \times(\ln\Psi + Z_R + iZ_I)]\Psi = H\Psi \quad (20)$$

if $C = \hbar/2mi$ is chosen. This choice is not the only possible one, as the constant C is not determined uniquely by the three basic assumptions, but has to be taken from experimental experience. Therefore, from the viewpoint of classical field theory also a choice $C = \frac{1}{2}\nu i$, i.e., a constant without $\hbar$ and m, is possible. With $V = q\Phi$, e.g., and the definitions $\nu = m/\hbar$ and $\zeta = q/\hbar$, in this case the NLSE would read

$$i\dot{\Psi} = [-(1/2\nu)\Delta + \zeta\Phi + (\gamma/i)(\ln\Psi + Z_R + iZ_I)]\Psi, \quad (21)$$

but, as already noticed, for formal reasons this form shall not be used in the following.

So, using normalized functions, the appearance of dissipative frictional forces is expressed by the "operator"

$$W_S = \gamma(\hbar/i)(\ln\Psi - \ln N + d/4 + iZ_I), \quad (22)$$

where Z_I still has to be determined. The determination of Z_I as well as the investigation of our NLSE shall be discussed in the next section, using a simple one-dimensional example.

III. THE ONE-DIMENSIONAL DAMPED HARMONIC OSCILLATOR AS AN EXAMPLE

Preliminary remarks (concerning nomenclature and notation): In the following considerations quantities like operator, wavefunction, mean value, uncertainty relation, commutator, etc. will appear. These quantities can be defined reasonably as well in probabilistic/quantum theoretical as in classical undulatory theories. For the *formal* comparison of our field theory with results of other authors, it is not necessary to accept the same interpretation, although this might be possible. So, hereafter a uniform notation will be chosen, where an appropriate meaning has to be associated with a certain symbol depending on the accepted interpretation. The symbol $\langle A\rangle$, for example, may stand for the quantum theoretical expectation value of an operator A, $\langle A\rangle = \int\Psi^*A\Psi\,d\tau$, as well as for the classical mean value of the quantity A, $\langle A\rangle = \int\Psi^*A\Psi\,d\tau/\int\Psi^*\Psi\,d\tau$, respectively.

In the course of the derivation of our NLSE ∇Z_I was determined for different cases, but thereby Z_I itself is not yet defined uniquely.

From a purely mathematical point of view, there are

many possibilities for Z_1. A more physical argumentation already reduces this number. We investigated the most reasonable possibilities and will discuss here only two of these.

The first and most simple assumption (magnetic fields shall not be regarded in this work) would be $Z_1 = 0$. This choice entails some results which are physically difficult to explain. So, for $t \to \infty$ the mean value $\langle H \rangle$ tends to zero. If H would be interpreted in the usual (conservative) way as energy of the system, this result would be in contradiction to uncertainty principle. But, moreover, in this case the interpretation of H is not even certain, as the mean value of our friction operator, $\langle W_S \rangle$, does not vanish; thus $\langle H \rangle \neq \langle T \rangle + \langle V \rangle$ (T = operator of kinetic energy; V = operator of potential energy).

Therefore, it seems reasonable to choose Z_1 in a way that $\langle W_S \rangle = 0$ and so $\langle i\hbar\, d/dt \rangle = \langle H \rangle = \langle T \rangle + \langle V \rangle$ is valid, i.e., the mean value of the Hamiltonian is equal at any time to the sum of kinetic and potential energies, as it is in usual linear quantum mechanics or, in metaphorical meaning, in nondissipative classical particle mechanics, respectively. This choice even entails other advantages:

(1) As in this case $Z_1 = -\langle \arg \Psi + 2k\pi \rangle = -\langle S + 2k\pi \rangle$, the fact that the function $\ln \Psi$ is not defined uniquely does not cause any problems.

(2) As for $t \to \infty$ the mean value $\langle H \rangle$ now asymptotically approaches a positive constant value, no problem with the uncertainty principle arises.

So finally our friction operator accepts the form

$$\tilde{W}_S = (\gamma\hbar/i) \ln \Psi - \langle (\gamma\hbar/i) \ln \Psi \rangle, \tag{23}$$

from which it can be taken that this dissipative term is connected with deviations from a mean value.

In this case the NLSE reads

$$i\hbar\dot{\Psi} = [-(\hbar^2/2m)\Delta + V + (\gamma\hbar/i)(\ln \Psi - \langle \ln \Psi \rangle)]\Psi. \tag{24}$$

Now the NLSE (24) with the external potential $V = \frac{1}{2} m\omega_0^2 x^2$ will be solved as an example. We will distinguish two kinds of solutions of the NLSE:

(1) wave solutions with an only explicit time dependence; (2) wavepacketlike solutions, in particular solitonlike solutions, with an additional implicit time dependence obtained by a transformation to a moving coordinate according to $\tilde{x} = x - \eta(t)$.

Inserting the general ansatz $\Psi = N \exp(iS + R)$ into (24), we get a complex differential equation which can be separated into real and imaginary parts, both containing $S(x,t)$ and $R(x,t)$.

A. Wave solutions

The functions S and R have been chosen as

$$S(x,t) = k_x + l_x x + M_x x^2,$$
$$R(x) = -P(x - Q)^2 \quad \text{with } Q = (4D/\gamma)b, \tag{25}$$

where k may contain an explicit time dependence. With this ansatz the coefficients can be determined and we find

$$k_x = -\tfrac{1}{2}(\omega_0^2/\Omega)t, \quad l_x = -(m\gamma/2\hbar)Q, \quad M_x = -m\gamma/4\hbar,$$
$$Q = 0, \quad P = (m/2\hbar)\Omega \tag{26}$$

with

$$\Omega = (\omega_0^2 - \gamma^2/4)^{1/2}. \tag{27}$$

We want to stress the fact that here the correct damped frequency (27) occurs, the analog to the classical corpuscular case.

So we can write the normalized wave solution in the form

$$\Psi(x,t) = (m\Omega/\pi\hbar)^{1/4} \exp[-i(m\gamma/4\hbar)x^2 - (m\Omega/2\hbar)x^2 - \tfrac{1}{2}(i/\hbar)\hbar\omega_0(\omega_0/\Omega)t]. \tag{28}$$

B. Solitonlike solutions

Taking into account the implicit time dependence introduced by the parameter $\eta(t)$, we use the ansatz

$$S(x,t;\eta) = K_x(t) + L_x(t)\tilde{x} + M_x\tilde{x}^2 = S(\tilde{x},t),$$
$$R(x;\eta) = -P(x - \eta(t))^2 = -P\tilde{x}^2 = R(\tilde{x},t), \tag{29}$$

where $Q = -\eta(t)$ has been set and $\tilde{x} = x - \eta(t)$. The remaining coefficients are determined as in the previous case, turning out to be

$$K_x(t) = (1/\hbar)\int_0^t \mathcal{L}(\eta,\dot{\eta},t')\,dt' - \tfrac{1}{2}(\omega_0^2/\Omega)t$$

with

$$\mathcal{L}(\eta,\dot{\eta},t) = \tfrac{1}{2}m\dot{\eta}^2(t) - \tfrac{1}{2}m\omega_0^2\eta^2(t),$$
$$L_x(t) = (m/\hbar)\dot{\eta}(t), \quad M_x = -m\gamma/4\hbar \tag{30}$$

and again

$$P = (m/2\hbar)\Omega.$$

Additionally, $\eta(t)$ has to obey the classical equation of motion (without stochastic force)

$$m\ddot{\eta} + m\gamma\dot{\eta} + m\omega_0^2\eta = 0. \tag{31}$$

The normalized solution

$$\Psi(x,t;\eta(t)) = (m\Omega/\pi\hbar)^{1/4} \exp\{(i/\hbar)[m\dot{\eta}(x - \eta) - \tfrac{1}{4}m\gamma(x - \eta)^2] - (m/2\hbar)\Omega(x - \eta)^2\} \times \exp\left\{(i/\hbar)\left[\int_0^t \mathcal{L}(\eta,\dot{\eta},t')\,dt' - \tfrac{1}{2}\hbar\omega_0(\omega_0/\Omega)t\right]\right\} \tag{32}$$

shows a form similar to a wavepacket in linear quantum mechanics; it is a solitonlike solution of Gaussian form, accepting a nomenclature used by Bialynicki-Birula and Mycielski a so-called Gausson.[29,30]

C. Properties of the solutions

The wave solution (28) as well as the solitonlike solution (32) have Gaussian shape with the constant width $x_0 = (\hbar/m\Omega)^{1/2}$. The center of the wave solution is the origin $x = 0$ whereas the center of the so-called Gausson (32) is following the classical trajectory.

The Gausson (32) also can be represented using a series expansion similar to the case of the undamped harmonic oscillator:

$$\Psi(x,t;\eta(t)) = \sum_{n=0}^{\infty} C_n v_n e^{i[\chi(t) + f(x,t;\eta)]} \tag{33}$$

where

$$v_n = (\beta^{1/2}/\pi^{1/4})(1/2^n n!)^{1/2} e^{-\beta^2 x^2/2} H_n(\beta x),$$

with

$$\beta^2 = m\Omega/\hbar, \quad H_n(\beta x) = n\text{th Hermite polynomial},$$

$$\chi(t) = (1/\hbar)\int_0^t \mathscr{L}(t')\,dt' - \tfrac{1}{2}(\omega_0^2/\Omega)t,$$

$$f(x,t;\eta(t)) = (1/\hbar)[m\dot{\eta}(x-\eta) - \tfrac{1}{4}m\gamma(x-\eta)^2].$$

The coefficients C_n turn out to be

$$C_n = (m\Omega/2\hbar)^{n/2}(n!)^{-1/2}\eta^n(t) \times \exp[-\tfrac{1}{2}(m\Omega/2\hbar)\eta^2(t)] = C_n(t).$$

Therefore, the contributions of the functions v_n to the Gausson are time dependent and determined by $\eta(t)$, the classical trajectory. As $\eta(t)$ tends to zero for $t\to\infty$ only $C_0 = 1$ remains.

So the wavefunctions which are stationary solutions of the undamped problem are now no longer solutions of the problem including friction [as our friction operator contains $\ln(\Psi\Psi^*)$]. But by an infinite superposition of these wavefunctions, now oscillating with the reduced frequency Ω, with time-dependent coefficients one can construct a stable wavepacketlike solution, a Gausson which is solution of our NLSE and for $t\to\infty$ continuously approaches a final state described by wavefunction (28).

For the limit $\gamma\to 0$ the reduced frequency Ω tends to the undamped frequency ω_0, $\lim_{\gamma\to 0}\Omega = \omega_0$, and the wavefunction (28) turns into the ground state wavefunction of the undamped harmonic oscillator in linear Schrödinger field theory

$$\lim_{\gamma\to 0}\Psi(x,t) = (m\omega_0/\pi\hbar)^{1/4}\exp[-(m\omega_0/2\hbar)x^2 - i\tfrac{1}{2}\omega_0 t]. \tag{34}$$

Inserting the coefficient of the quadratic term in R, $P = m\Omega/2\hbar$, which is the same for the wavefunction and the Gausson, into Eq. (16) we can determine the "diffusion" constant D of the FPE

$$D = \frac{\hbar}{2m}\frac{\gamma/2}{\Omega} = \frac{\hbar}{2m}\frac{\gamma/2}{\sqrt{\omega_0^2 - \gamma^2/4}}, \tag{35}$$

where D is a *real* quantity and $\lim_{\gamma\to 0} D = 0$, which means the FPE turns into the usual CE (3) when γ approaches zero.

By means of the abbreviation

$$\mathbf{J}_t = \rho\mathbf{V}_t = \rho(\mathbf{v} - D\nabla\rho/\rho)$$

we write the FPE in the form

$$\dot{\rho} + \operatorname{div}\mathbf{J}_t = 0. \tag{36}$$

The stationary solutions of (37) are defined by $\dot{\rho} = 0$. Integrating and imposing the "natural boundary condition" for ρ, which means that ρ vanishes for $r\to\pm\infty$, in our case the condition for stationary solutions of the FPE can be represented as

$$D(\nabla R) = (\hbar/2m)(\nabla S) \tag{37}$$

and, in particular with $\nabla R = -2P\mathbf{r}$,

$$\nabla S = -4D(m/\hbar)P\mathbf{r}. \tag{38}$$

Regarding the wave solution (28) of our example, it can be shown easily that condition (38) is fulfilled by (28), i.e., $\dot{\rho} = \partial\rho/\partial t = 0$. Because of $\tilde{x} = x - \eta(t)$ solution (32) is implicitly time dependent, so that $\dot{\rho} = \partial\rho/\partial t - \dot{\eta}(t)\nabla\rho = 0$ had to be fulfilled. It is straightforward to show that only $\partial\rho/\partial t = 0$ is valid, but $\dot{\rho} = -\dot{\eta}\nabla\rho \neq 0$. This time dependence of ρ due to $\eta(t)$ shows that time reversibility no longer holds.

So the probability density according to solution (32) is no stationary solution of the FPE, but, for $t\to\infty$, $\rho(t)$ asymptotically approaches the stationary solution fo the FPE, corresponding to the final state (28).

D. Uncertainty principle and energy

In order to show that our solutions are in agreement with the uncertainty principle we evaluate the mean values $\langle x\rangle$, $\langle x^2\rangle$, $\langle p\rangle$, and $\langle p^2\rangle$ with the help of solutions (28) and (32), respectively.

Calculating these mean values using the wavefunction (28), we obtain

$$\langle x\rangle_w = 0, \quad \langle x^2\rangle_w = \hbar/2m\Omega$$
$$\langle p\rangle_w = 0, \quad \langle p^2\rangle_w = m\hbar\omega_0^2/2\Omega \tag{39}$$

which yields the mean-square deviations

$$\overline{\Delta x^2} = \langle x^2\rangle_w - \langle x\rangle_w^2 = \hbar/2m\Omega,$$
$$\overline{\Delta p^2} = \langle p^2\rangle_w - \langle p\rangle_w^2 = m\hbar\omega_0^2/2\Omega. \tag{40}$$

Utilizing the Gausson (32), the corresponding mean value and mean-square deviations are

$$\langle x\rangle_G = \langle\tilde{x} + \eta(t)\rangle_G = \langle\tilde{x}\rangle_G + \langle\eta(t)\rangle_G = \eta(t)$$

with

$$\langle\tilde{x}\rangle_G = 0,$$
$$\langle x^2\rangle_G = \langle\tilde{x}^2 + \eta^2 + 2\eta\tilde{x}\rangle_G = \langle\tilde{x}^2\rangle_G + \langle\eta^2\rangle_G = \hbar/2m\Omega + \eta^2 \quad \text{with } \langle\tilde{x}^2\rangle_G = \hbar/2m\Omega, \tag{41}$$
$$\langle p\rangle_G = m\dot{\eta},$$
$$\langle p^2\rangle_G = \hbar m\omega_0^2/2\Omega + m^2\dot{\eta}^2$$

and

$$\overline{\Delta x^2} = \hbar/2m\Omega + \eta^2 - (\eta)^2 = \hbar/2m\Omega,$$
$$\overline{\Delta p^2} = \hbar m\omega_0^2/2\Omega + m^2\dot{\eta}^2 - (m\dot{\eta})^2 = \hbar m\omega_0^2/2\Omega. \tag{42}$$

Regarding the uncertainty product $\overline{\Delta p^2}\cdot\overline{\Delta x^2}$, both solutions (28) and (32) produce the same result

$$\overline{\Delta p^2}\cdot\overline{\Delta x^2} = \tfrac{1}{4}\hbar^2\omega_0^2/\Omega^2 \geqslant \tfrac{1}{4}\hbar^2, \tag{43}$$

which in the limit $\gamma\to 0$ is

$$\lim_{\gamma\to 0}\overline{\Delta p^2}\cdot\overline{\Delta x^2} = \lim_{\gamma\to 0}\tfrac{1}{4}\hbar^2\omega_0^2/\Omega^2 = \tfrac{1}{4}\hbar^2, \tag{44}$$

but for $\gamma\neq 0$ always accepts a larger value.

By the help of Eqs. (39) and (41), respectively, the mean values of the kinetic and potential energies can be determined easily. For the wave solution we obtain

$$\langle T\rangle_w = (1/2m)\langle p^2\rangle = \tfrac{1}{4}\hbar^2\omega_0^2/\Omega,$$
$$\langle V\rangle_w = \tfrac{1}{2}m\omega_0^2\langle x^2\rangle = \tfrac{1}{4}\hbar^2\omega_0^2/\Omega, \tag{45}$$

or, respectively,

$$\langle E\rangle_W = \langle T\rangle_W + \langle V\rangle_W = \tfrac{1}{2}\hbar\omega_0(\omega_0/\Omega) = \tfrac{1}{2}\hbar\Omega(1+\tfrac{1}{4}\gamma^2/\Omega^2). \tag{45'}$$

The fact that $\langle E\rangle_W$ is not $\frac{1}{2}\hbar\Omega$ as one could expect at first sight might be interpreted in a way that the interaction of the oscillating system with the surrounding also influences the state of lowest energy. So it seems to be reasonable that this state is even somewhat higher in energy than the ground state of the undamped oscillator ($\omega_0/\Omega > 1$). This is similar to the fact that the velocity of a Brownian particle is slowed down by the interaction with the surrounding, but it will never come to absolute rest because of the backtransfer of energy from the environment to the particle due to this interaction.

The corresponding energies for the Gausson are

$$\langle T\rangle_G = \tfrac{1}{4}\hbar\omega_0^2/\Omega + \tfrac{1}{2}m\dot{\eta}^2, \qquad \langle V\rangle_G = \tfrac{1}{4}\hbar\omega_0^2/\Omega + \tfrac{1}{2}m\omega_0^2\eta^2, \tag{46}$$

or, respectively,

$$\langle E\rangle_G = \tfrac{1}{2}\hbar\omega_0^2/\Omega + \tfrac{1}{2}m\dot{\eta}^2 + \tfrac{1}{2}m\omega_0^2\eta^2. \tag{46'}$$

As, in the limit $t\to\infty$, $\eta(t)$ as well as $\dot{\eta}(t)$ vanishes, there only remains the constant time-independent part of the energy $\langle E\rangle_G$, which is identical with the energy of the wave solution $\langle E\rangle_W$. This energy, on the other hand, tends to the ground state energy of the undamped harmonic oscillator when γ approaches zero.

IV. SOME REMARKS CONCERNING THE FRICTION OPERATOR $\tilde{W}_S$

A. Properties of the friction operator $\tilde{W}_S$

It is a straightforward calculation to show that the friction operator $\tilde{W}_S$ is self-adjoint with respect to Ψ if the real-part of Ψ obeys Eqs. (13) and (15), i.e.,

$$\langle \tilde{W}_S\Psi\,|\,\Psi\rangle = \langle \Psi\,|\,\tilde{W}_S\Psi\rangle = \int (\tilde{W}_S\Psi)^*\Psi\,d\tau = \int \Psi^*\tilde{W}_S\Psi\,d\tau. \tag{47}$$

Determining the commutators of $\tilde{W}_S$ with the operators of $\mathbf{r}$, r^2, $\mathbf{p}$, p^2, V, and H we can use the relation $-\nabla\tilde{W}_S = (i/\hbar)\times[\tilde{W}_S,\mathbf{p}]$ to calculate the mean value

$$\langle -\nabla\tilde{W}_S\rangle = -\gamma\langle\mathbf{p}\rangle, \tag{48}$$

and, by means of $(d/dt)\mathbf{p} = (\partial/\partial t)\mathbf{p} + (i/\hbar)[H,\mathbf{p}]$, we obtain the mean value

$$\left\langle\frac{d}{dt}\mathbf{p}\right\rangle = -\langle\nabla V\rangle - \gamma\langle\mathbf{p}\rangle. \tag{49}$$

So Ehrenfest's theorem is fulfilled, i.e., the equation of motion for the expectation values corresponds to the classical equation of motion.

In the same way the energy dissipation can be determined turning out to be

$$\langle\dot{E}\rangle = -(\gamma/m)\langle p^2\rangle, \tag{50}$$

which is in accordance with classical particle mechanics.

B. Comparison of $\tilde{W}_S$ with similar operators

Rewriting our operator $\tilde{W}_S$ (23) in the form

$$\tilde{W}_S = \frac{\gamma\hbar}{2i}\ln\frac{\Psi}{\Psi^*} - \left\langle\frac{\gamma\hbar}{2i}\ln\frac{\Psi}{\Psi^*}\right\rangle + \frac{\gamma\hbar}{2i}\ln(\Psi\Psi^*) - \left\langle\frac{\gamma\hbar}{2i}\ln(\Psi\Psi^*)\right\rangle \tag{51}$$

it can easily be compared with the friction operators of Kostin[7,8] and Cho[16] and an operator investigated by Bialynicki-Birula and Mycielski.[29,30]

Kostin's operator

$$W_K = \frac{\hbar\gamma}{2i}\ln\frac{\Psi}{\Psi^*} - \left\langle\frac{\hbar\gamma}{2i}\ln\frac{\Psi}{\Psi^*}\right\rangle$$

is identical with the first and second term on the right-hand side of (51). It is quite obvious that this operator allows all solutions of the undamped harmonic oscillator (which are real functions) to be also solutions of the problem including friction. Furthermore, Kostin's solutions contain the unshifted frequency ω_0 instead of the reduced frequency $\Omega = (\omega_0^2 - \gamma^2/4)^{1/2}$ (see, e.g., Hasse[18]).

The velocity-dependent forces derivable from the friction potentials W_K and $\tilde{W}_S$, respectively, are different,

$$\mathbf{F}_K = -(\gamma\hbar/2i)\nabla\ln(\Psi/\Psi^*) \tag{52a}$$

$$\mathbf{F}_S = -(\gamma\hbar/i)\nabla\ln\Psi, \tag{52b}$$

but as long as only mean values of these operators are regarded, this difference doesn't become evident, as $\langle\mathbf{F}_K\rangle = \langle\mathbf{F}_S\rangle$. The differences appear when comparing the solutions of the NLSE's containing the different operators W_K or $\tilde{W}_S$, respectively, as shown above.

To remove the unphysical behavior of Kostin's solutions, Cho[16] heuristically introduced an additional "diffusion velocity" $-(\hbar/2m)(\nabla\rho/\rho)$ caused by a so-called "quantum potential"

$$\sigma = -(\hbar/2m)\ln(\Psi\Psi^*). \tag{53}$$

Hence, Cho's friction potential achieves the form

$$W_C = (\gamma\hbar/2i)\ln(\Psi/\Psi^*) + \tfrac{1}{2}\gamma\hbar\ln(\Psi\Psi^*) - \gamma f(t), \tag{54}$$

where "$f(t)$ is taken so that at stationary states the friction operator does not depend on time."[16] The additional term $\frac{1}{2}\gamma\hbar\ln(\Psi\Psi^*)$ already shows some similarity with the third term of (51) but is a real quantity in contrast to the purely imaginary term in (51). Because of the additional term, Cho's operator certainly does not permit the undamped solutions any longer, but he still obtains a wrong frequency, namely $\omega = (\omega_0^2 + \gamma^2/4)^{1/2} - \gamma/2$.

Under a more formal point of view, Bialynicki-Birula and Mycielski[30] profoundly investigated an NLSE with logarithmic nonlinearity and discussed the properties of the solutions in detail. The nonlinearity under investigation has the form $-b\ln(\Psi\Psi^*)$ with b choosen to be a real coefficient as in Cho's operator, in particular $b>0$. A connection between this nonlinearity and a physically justified dissipative term in the Schrödinger equation is not given, however (see, e.g., Enz[31]).

As already shown, our friction operator (51) for the first time allows us to remove both difficulties connected with

Kostin's operator and, moreover, can be derived using physical arguments without heuristic assumptions.

In order to demonstrate a formal connection between our friction operator $\tilde{W}_S$ and the usual quantum mechanical operator of velocity, $\mathbf{v}_{qm} = (\hbar/mi)\nabla$, we operate on a wavefunction Ψ by this velocity operator $\mathbf{v}_{qm}$ to obtain

$$\mathbf{v}_{qm}\Psi = \frac{1}{m}\frac{\hbar}{i}\nabla\Psi = \frac{1}{m}\frac{\hbar}{i}\frac{\nabla\Psi}{\Psi}\cdot\Psi = \frac{1}{m}\frac{\hbar}{i}(\nabla \ln \Psi)\Psi, \tag{55}$$

which shows that

$$\mathbf{v}_{qm} \hat{=} \frac{1}{m}\frac{\hbar}{i}\nabla \ln \Psi, \tag{55'}$$

i.e., application of an operator $(\hbar/mi)\nabla \ln \Psi$ would yield the same result. A kind of "potential" ϕ_v for this last mentioned operator satisfying $-\nabla\phi_v = \mathbf{v}_{qm}$ can be given by

$$\phi_v = -(\hbar/mi)\ln \Psi + C_v \tag{56}$$

with C_v = const or at most time-dependent. Quite analogous to this, a "potential" ϕ_F of a velocity-dependent force $\mathbf{F}_F = -m\gamma\mathbf{v}_{qm}$ may be defined, where

$$\mathbf{F}_F\Psi = -m\gamma\mathbf{v}_{qm}\Psi = -(\gamma\hbar/i)(\nabla \ln \Psi)\Psi \tag{57a}$$

and

$$\phi_F = (\gamma\hbar/i)\ln \Psi + C_F \tag{57b}$$

with $\mathbf{F}_F = -\nabla\phi_F$ and C_F = const or at most time-depedent. By appropriate choice of C_F ($C_F = -\langle(\gamma\hbar/i)\ln \Psi\rangle$) it can be achieved that $\phi_F \equiv \tilde{W}_S$.

Naturally, one could think of constructing a "friction potential" operator directly as a product of the operators for position and velocity (or linear momentum $\mathbf{p}_{qm}$), noting that these operators do not commute. Attempts in this heuristical direction have been started by Albrecht[17] and Hasse and Süssmann.[18,19]

Here, we want to present the following ansatz for a heuristical potential of this kind

$$\tilde{W}_+ = W_+ + W^0{}_+ = c\gamma[x + a, p_{qm} + b]_+ + W^0{}_+ \tag{58}$$

with $[\ ,\ ]_+$ = anticommutator, where the coordinate-independent quantities a, b, c, and $W^0{}_+(t)$ will be determined so that

$$\tilde{W}_S\Psi = \tilde{W}_+\Psi \tag{59}$$

is valid with

$$a = Q = -\langle x\rangle, \quad b = \hbar L_x = \langle p_{qm}\rangle, \quad c = \tfrac{1}{4}, \quad W^0{}_+ = -\hbar\gamma M_x/4P = -\langle W_+\rangle, \tag{60}$$

we finally obtain

$$\tilde{W}_+ = \tfrac{1}{4}\gamma[x - \langle x\rangle, p_{qm} + \langle p_{qm}\rangle]_+ - \langle\tfrac{1}{4}\gamma[x - \langle x\rangle, p_{qm} + \langle p_{qm}\rangle]_+\rangle = W_+ - \langle W_+\rangle. \tag{61}$$

Comparison with Hasse's operator W_H (see Table I) shows that $W_+ = W_H$. Operating on a wavefunction Ψ with the operators $\tilde{W}_+$ and W_H, respectively, the difference between these operators becomes evident:

$$\tilde{W}_+\Psi = \tilde{W}_S\Psi = W_H\Psi - \langle W_H\rangle\Psi = W_H\Psi - \hbar\gamma(M_x/4P)\Psi. \tag{62}$$

So, usually $\langle W_H\rangle = \hbar\gamma M_x/4P \neq 0$, whereas $\langle\tilde{W}_+\rangle = \langle\tilde{W}_S\rangle = 0$, as discussed previously.

For comparison, in Table I the most important friction operators we regarded are given together with the corresponding mean values and the action on a wavefunction of the type we investigated.

V. CONCLUSION AND PROSPECT

In the present work we used a formalism based on physical arguments to derive a nonlinear field equation with logarithmic nonlinearity for the description of dissipative phenomena connected with velocity-dependent frictional forces. Although in principle there exist different possibilities of interpretation for our NLSE, calculations can be performed independent of the interpretation of our equation because of the formal similarity between classical undulatory and probabilistic theories like quantum theory, only the occurring quantities and results have to be interpreted in different ways. For this reason, comparison of our results with those of other authors using NLSE's was possible, and we could show that the solutions of our NLSE are in agreement with physical experience.

Nevertheless, we would like to stress that there are still many questions open. So the connection between the macroscopic phenomenological description of the damping pro-

TABLE I. Regarded friction operators, corresponding mean values, and action on a wavefunction of the type we investigated.

W	$\langle W\rangle$	$W\Psi$
$\tilde{W}_S = (\gamma\hbar/i)(\ln \Psi - \langle \ln \Psi\rangle)$	0	$\hbar\gamma[M_x\tilde{x}^2 + L_x\tilde{x} - M_x/4P]\Psi + i\hbar\gamma[P\tilde{x}^2 - \frac{1}{4}]\Psi$
$W_K = (\gamma\hbar/i)(\ln(\Psi/\Psi^*) - \langle\ln(\Psi/\Psi^*)\rangle)$	0	$\hbar\gamma[M_x\tilde{x}^2 + L_x\tilde{x} - M_x/4P]\Psi$
$W_c = (\gamma\hbar/2i)\ln(\Psi/\Psi^*) + \frac{1}{2}\gamma\ln(\Psi\Psi^*) - \gamma f(t) = W_{C,0} - \gamma f(t)$	0 if $\gamma f = \langle W_{C,0}\rangle$	$\hbar\gamma[(M_x - P)\tilde{x}^2 - L_x\tilde{x} - M_x/4P + \frac{1}{4}]\Psi$ (if $\gamma f = \langle W_{C,0}\rangle$)
$W_A = \gamma\langle p\rangle(x - \langle x\rangle)$[a]	0	$\hbar\gamma[L_x\tilde{x}]\Psi$
$W_{SU} = \frac{1}{2}\gamma[x - \langle x\rangle, p]_+$[b]	$-\hbar\gamma^2/4\Omega$	$2\hbar\gamma[M_x\tilde{x}^2 + L_x/2\tilde{x}]\Psi + i2\hbar\gamma[P\tilde{x}^2 - \frac{1}{4}]\Psi$
$\tilde{W}_{SU} = \frac{1}{2}\gamma[x - \langle x\rangle, p]_+ - \langle\frac{1}{4}\gamma[x - \langle x\rangle, p]_+\rangle = \frac{1}{2}\{W_{SU} - \langle W_{SU}\rangle\}$[c]	0	$\hbar\gamma[M_x\tilde{x}^2 + L_x/2\tilde{x} - M_x/4P]\Psi + i\hbar\gamma[P\tilde{x}^2 - \frac{1}{4}]\Psi$
$W_H = \frac{1}{4}\gamma[x - \langle x\rangle, p + \langle p\rangle]_+$	$-\hbar\gamma^2/8\Omega$	$\hbar\gamma[M_x\tilde{x}^2 + L_x\tilde{x}]\Psi + i\hbar\gamma[P\tilde{x}^2 - \frac{1}{4}]\Psi$
$\tilde{W}_+ = \frac{1}{4}\gamma[x - \langle x\rangle, p + \langle p\rangle]_+ - \langle\frac{1}{4}\gamma[x - \langle x\rangle, p + \langle p\rangle]_+\rangle$	0	$\hbar\gamma[M_x\tilde{x}^2 + L_x\tilde{x} - M_x/4P]\Psi + i\hbar\gamma[P\tilde{x}^2 - \frac{1}{4}]\Psi$

[a] Friction operator of Albrecht.
[b] Friction operator of Süssmann.
[c] Modification of Süssmann's friction operator.

cess and the field theoretical description is established by the friction coefficient γ, appearing in both cases. In a purely microscopic theory like "orthodox" quantum mechanics no dissipation exists due to the conservative nature of the fundamental interactions. Damping is only the result of statistical interaction with many subsystems of the surrounding. The constant γ now describes an averaged effect of all these interactions which also can be observed macroscopically. Deriving our field equation, we did not evaluate γ on a microscopic level, but accepted it as a possibly empirically determinable constant. In order to investigate not only an average effect but the dynamical process on a microscopical level in detail, one has to take into account the interaction between the system under consideration and the surrounding subsystems explicitly and submit γ to a more careful analysis. At the same time it has to be considered more precisely what kind of systems are interacting, as well as the nature of the interaction. The question whether in this connection the size of the system is a limiting factor for the applicability of our NLSE, and what is the respective meaning of "microscopic", possibly has to be regarded in each case (see also Ref. 30).

Another point that needs further clarification is the consideration of the stochastic force $\mathcal{F}(t)$. There are attempts to take into account this force by an additional "random potential" $-\mathbf{r}\mathcal{F}(t)$, but in the following treatment of the problem usually this term is neglected to avoid the complexity of calculation although it might be of importance (see, e.g., Ref. 13).

For the first time Messer[32] solved an NLSE of the damped harmonic oscillator, explicitly taking account of the fluctuating force using a method developed by Svin' in[33] for the quantum theoretical description of friction in the framework of time-dependent Hamiltonians. Messer replaces differential equation (31) by the stochastic differential equation

$$m\ddot{\eta} + m\gamma\dot{\eta} + m\omega_0^2\eta = \mathcal{F}(t), \tag{63}$$

receiving as solution the probability $W(\eta,t)$ of some value η at time t for given initial values. By a solution of the stochastic NLSE the position probability $w_x(t)$ will be understood, which is a convolution of $|\Psi(x,t;\eta)|^2$ and $W(\eta,t)$:

$$w_x(t) = \int_{-\infty}^{+\infty} |\Psi(x,t;\eta)|^2 W(\eta,t)\, d\eta. \tag{64}$$

Messer made use of Kostin's nonlinear friction potential W_K and the corresponding solutions $\Psi(x,t;\eta)$. Via Z_1 it is possible to introduce the "random potential" $-\mathbf{r}\mathcal{F}(t)$ also in our NLSE. Therefore, it should be quite simple and straightforward to use our operator $\tilde{W}_S$ and the corresponding solutions instead of Kostin's and perform the same calculations again. To what extent this procedure actually takes into account the stochastic force within the framework of a microscopic theory in a proper way will be reconsidered. Some results like the energy $\langle E \rangle_W$ of the final state seem to indicate an effect due to a stochastic force, but whether there is any connection with $\mathcal{F}(t)$, and of what kind this connection could be, still needs a more detailed analysis.

In forthcoming works we will derive the NLSE for the damped motion in a magnetic field in a similar way as shown in Sec. II B and discuss the solutions of this equation. Also several other types of potentials will be regarded. Furthermore, the interpretation of our NLSE as classical field equation and its quantization will be investigated in more detail.

ACKNOWLEDGMENTS

Two of us (K.-M.C. and D.S.) wish to thank the Akademie der Wissenschaften und der Literatur zu Mainz and the Johann Wolfgang Goethe-Universität Frankfurt for supporting this work.

[1]S. Chandrasekhar, Rev. Mod. Phys. **15**, 1 (1943).
[2]E. Kanai, Prog. Theor. Phys. **3**, 440 (1948).
[3]P. Caldirola, Nuovo Cimento **18**, 393 (1941).
[4]W. E. Brittin, Phys. Rev. **77**, 396 (1950).
[5]J. R. Ray, Lett. Nuovo Cimento **25**, 47 (1979).
[6]D. M. Greenberger, J. Math. Phys. **20**, 762 (1979).
[7]M. D. Kostin, J. Chem. Phys. **57**, 3589 (1972).
[8]M. D. Kostin, J. Stat. Phys. **12**, 145 (1975).
[9]B. K. Skagerstam, Phys. Lett. B **58**, 21 (1975).
[10]B. K. Skagerstam, J. Math. Phys. **18**, 308 (1977).
[11]K. Yasue, Phys. Lett. B **64**, 239 (1976).
[12]K. Yasue, J. Stat. Phys. **16**, 113 (1977).
[13]K. Yasue, Ann. Phys. N.Y. **114**, 479 (1978).
[14]K. K. Kan and J. J. Griffin, Phys. Lett. B **50**, 241 (1974).
[15]J. D. Immele, K. K. Kan, and J. J. Griffin, Nucl. Phys. A **241**, 47 (1975).
[16]B. R. Cho, Suhak Gwa Mulri (Korean) (Math. and Phys.) **3**, 37 (1980).
[17]K. Albrecht, Phys. Lett. B **56**, 127 (1975).
[18]R. W. Hasse, J. Math. Phys. **16**, 2005 (1975).
[19]R. W. Hasse, Rep. Prog. Phys. **41**, 1027 (1978).
[20]F. Hund, *Materie als Feld* (Springer-Verlag, Berlin 1954).
[21]E. Fick, *Einführung in die Grundlagen der Quantentheorie* (Akad. Verlagsgesellschaft, Frankfurt am Main, 1974).
[22]H. Hartmann, *Die chemische Bindung. Drei Vorlesungen für Chemiker* (Springer-Verlag, Berlin, 1971).
[23]E. Madelung, *Die mathematischen Hilfsmittel des Physikers* (Springer-Verlag, Berlin, 1950), p. 432.
[24]B. Mrowka, Z. Physik **130**, 164 (1951).
[25]W. Heisenberg, *Physikalische Prinzipien der Quantentheorie* (Bibliographisches Institut, Mannheim, 1958), p. 96.
[26]I. Fényes, Z. Physik **132**, 81 (1952).
[27]E. Nelson, *Dynamical Theories of Brownian Motion* (Princeton U.P., Princeton, N.J., 1972), p. 137.
[28]E. Nelson, Phys. Rev. **150**, 1079 (1966).
[29]J. Bialynicki-Birula and J. Mycielski, Bull. Acad. Polon. Sci. Cl. III **23**, 461 (1975).
[30]J. Bialynicki-Birula and J. Mycielski, Ann. Phys. (N.Y.) **100**, 62 (1976).
[31]C. P. Enz, Phys. Unserer Zeit **10**, 141 (1979).
[32]J. Messer, Acta Phys. Austr. **50**, 75 (1979).
[33]J. R. Svin'in, Teor. Mat. Fiz. **27**, 270 (1976).

Some Yang–Mills fields constructed from Maxwell fields

P. A. Hogan
Mathematical Physics Department, University College, Belfield, Dublin 4, Ireland and School of Theoretical Physics, Dublin Institute for Advanced Studies, 10, Burlington Road, Dublin 4, Ireland

(Received 10 February 1982; accepted for publication 25 June 1982)

We show that the natural connections on the Hopf bundles $S_{2n+1}\to\mathbb{C}P_n$ $(n\geqslant 1)$ satisfy stronger conditions than Maxwell's equations. These conditions imply that the natural connections on the U(n) Stiefel bundles SU($n+1$)$\to\mathbb{C}P_n$ satisfy the Yang–Mills equations. The use of projector-valued fields enables us to pass from the Hopf to the Stiefel bundles in a natural way.

PACS numbers: 11.15.−q

1. INTRODUCTION

Many physically interesting solutions of the Yang–Mills equations (and, in particular, of Maxwell's equations) arise as the natural connections on specific principal fiber bundles. The simplest example is the magnetic monopole solution of Maxwell's equations which arises from the natural connection on the Hopf bundle[1] $S_3\to S_2=\mathbb{C}P_1$ with structure group U(1), while the Belavin *et al.*[2] instanton is constructed from the natural connection on the Hopf bundle[1] $S_7\to S_4$ with SU(2) as structure group. The natural connection on SO(5)$\to$SO(5)/SO(4) $=S_4$ describes the instanton–anti-instanton pair, by utilizing the isomorphism SO(4) $=$ SU(2)$\times$SU(2).

Research in this area of gauge theory has proceeded with steadily increasing generality. First, Trautman[1] proved that the natural connections on the U(1) Hopf bundles $S_{2n+1}\to\mathbb{C}P_n$ $(n\geqslant 1)$ satisfy Maxwell's equations and the natural connections on the SU(2) Hopf bundles $S_{4n+3}\to\mathbb{H}P_n$ satisfy the Yang–Mills equations. This work was generalized by Nowakowski and Trautman[3] who proved that the natural connections on the Stiefel bundles with unitary or orthogonal structure groups satisfy the Yang–Mills equations. This was further generalized by Harnad, Tafel, and Schnider[4] who established a similar property of the natural connections on principal bundles over Riemannian symmetric spaces.

While studying the interrelations between the results above we have noticed that the natural connections on the U(1) Hopf bundles $S_{2n+1}\to\mathbb{C}P_n$ satisfy stronger conditions than Maxwell's equations and that these conditions imply that the natural connections on the U(n) Stiefel bundles[5] SU($n+1$)$\to\mathbb{C}P_n$ satisfy the Yang–Mills equations. This is the main result of the present paper. To establish it we use the projector-valued field description of gauge theories with a unitary gauge group[6–8] which enables us to pass from the U(1) bundles to the U(n) bundles in a natural way.

2. PROJECTOR-VALUED FIELDS FOR STIEFEL BUNDLES

Let $V_{N,k}$ denote the Stiefel manifold of unitary k frames at the origin of $\mathbb{C}^N$. This manifold can be identified with the coset space U(N)/U($N-k$) in an obvious way where U($N-k$) is identified with $I_k\times$U($N-k$)$\subset$U(N). An element of $V_{N,k}$ can be displayed as an $N\times k$ matrix $(V_{\alpha i})$, $\alpha=1,2,\ldots,N$, $i=1,2,\ldots,k$ satisfying

$$\overline{V}_{\alpha i}V_{\alpha j}=\delta_{ij}, \tag{2.1}$$

with the bar denoting complex conjugation. If $(b_{ij})\in$U(k) then U(k) acts on $V_{N,k}$ to the right via

$$V_{\alpha i}\to V_{\alpha j}b_{ji}, \tag{2.2}$$

while U(N) acts to the left in similar fashion. Under the action of U(k), $V_{N,k}$ becomes a principal fiber bundle with base space the Grassmann manifold $G_{N,k}$ of k-planes through the origin of $\mathbb{C}^N$. Thus $G_{N,k}$ can be identified with the coset space U(N)/U($N-k$)$\times$U(k). Defining the 1-forms

$$\omega_{ij}=\overline{V}_{\alpha i}dV_{\alpha j}, \tag{2.3}$$

it follows from (2.1) that $\omega_{ij}\in u(k)$, the Lie algebra of U(k), and Narasimhan and Ramanan[9] have proved that (2.3) is a connection 1-form on the U(k) bundle $V_{N,k}\to G_{N,k}$ and that these Stiefel bundles are universal bundles for the groups U(k) (see also Shlafly[10] and Dubois-Violette[8]).

If, following Nowakowski and Trautman,[3] we begin with $V_{\alpha\beta}\in$U(N) and construct the Maurer–Cartan form on U(N)

$$\omega_{\alpha\beta}=\overline{V}_{\lambda\alpha}dV_{\lambda\beta}, \tag{2.4}$$

then $\omega_{\alpha\beta}\in$u(N) and (2.3) is the $k\times k$ block in the top left-hand corner of this matrix of 1-forms. Moreover, the curvature 2-form corresponding to (2.3) can be written, using the Maurer–Cartan equations, as the matrix

$$F_{ij}=\overline{\omega}_{Ai}\wedge\omega_{Aj}, \tag{2.5}$$

where capital letters take values $k+1, k+2,\ldots,N$.

The quadratic form $\overline{\omega}_{Ai}\omega_{Ai}$ is invariant under the action of U($N-k$) and U(k) on U(N) and thus it projects to a Riemannian line element ds^2 on $G_{N,k}$ which we simply write as

$$ds^2=\overline{\omega}_{Ai}\omega_{Ai}. \tag{2.6}$$

For convenience we have not introduced further notation in (2.5) and (2.6), it being understood that F_{ij} is a field of 2-forms on $V_{N,k}$ [on account of the invariance of ω_{ij} under the right action of U($N-k$) on U(N)] while ds^2 is a quadratic form on $G_{N,k}$. Being in possession of (2.6) one can calculate the dual ${}^*F_{ij}$ of F_{ij} and Nowakowski and Trautman[3] have proved that

$$D\,{}^*F_{ij}=0, \tag{2.7}$$

i.e., F_{ij} satisfies the Yang–Mills equations where D is exterior covariant differentiation with respect to the connection 1-form ω_{ij}.

We now introduce the $N\times N$ matrix

 0022-2488/83/061661-04$02.50 © 1983 American Institute of Physics

$$P_{\alpha\beta} = V_{\alpha i}\overline{V}_{\beta i}. \tag{2.8}$$

Using (2.1) we see that $P = (P_{\alpha\beta})$ is projector-valued, i.e., $P^2 = P$. Such a field was used by Corrigan, Fairlie, Templeton, and Goddard[6] to discuss self-dual Yang–Mills fields on S_4 and its application to gauge theory in general, with a unitary gauge group, has been given by Dubois–Violette.[7,8] It is the latter theory which we apply here.

First we notice that P is gauge invariant, i.e., invariant under the transformation (2.2). From the curvature (2.5) we construct a gauge invariant $N \times N$ matrix of 2-forms

$$\Omega_{\alpha\beta} = V_{\alpha i}F_{ij}\overline{V}_{\beta j}. \tag{2.9}$$

This can be written in terms of P. Using (2.4), (2.5), and (2.8) we find, in matrix notation[7]

$$\Omega = P\,dP \wedge dP = dP \wedge dP\,P, \tag{2.10}$$

with the last equality following from $P^2 = P$. For any matrix of 2-forms F_{ij}, and $\Omega_{\alpha\beta}$ related to it by (2.9), we have

$$\begin{aligned}(P\,d\Omega\,P)_{\alpha\sigma} &= V_{\alpha i}(dF_{ij} + \omega_{ih}\wedge F_{hj} - F_{ih}\wedge\omega_{hj})\overline{V}_{\sigma j}\\ &= V_{\alpha i}DF_{ij}\overline{V}_{\sigma j},\end{aligned} \tag{2.11}$$

and thus, in particular, $DF_{ij} = 0$ if and only if $P\,d\Omega\,P = 0$. Hence with Ω given by (2.10) the Bianchi identities become

$$P\,d\Omega\,P = P\,dP\wedge dP\wedge dP\,P = 0, \tag{2.12}$$

the last equality coming from (2.10) together with the fact that $P\,dP\,P = 0$ since $P^2 = P$. In this terminology the Yang–Mills equations read[7]

$$P\{d\,{}^*(dP\wedge dP)\}P = 0. \tag{2.13}$$

The line-element (2.6) takes the form

$$ds^2 = \tfrac{1}{2}\,\text{trace}\,(dP)^2 \tag{2.14}$$

and we note from (2.8) that P is Hermitian.

3. PROJECTOR-VALUED FIELDS FOR HOPF BUNDLES

The U(1) Hopf bundles

$$S_{2n+1}\to\mathbb{C}P_n \quad (n\geqslant 1) \tag{3.1}$$

constitute a special case of the Stiefel bundles corresponding to $N = n + 1$, $k = 1$. Thus by (2.1) and (2.8) the projector-valued field can be written

$$P_{\alpha\beta} = \Psi_\alpha\overline{\Psi}_\beta, \quad \Psi_\alpha\in S_{2n+1}, \tag{3.2}$$

where we have put $V_{\alpha 1} = \Psi_\alpha$ and we will, from now on, allow Greek indices to take values 0,1,2,...,n, while Latin indices take values 1,2,3,...,n. We can parametrize Ψ_α in a standard way[1]

$$\Psi_0 = \rho e^{i\theta}, \quad \Psi_j = \rho e^{i\theta}\zeta_j \tag{3.3}$$

where θ is an arbitrary real variable and

$$\rho^{-2} = 1 + \zeta_j\bar{\zeta}_j. \tag{3.4}$$

Then (3.2) becomes

$$P_{\alpha\beta} = \rho^2\zeta_\alpha\bar{\zeta}_\beta, \quad \zeta_0 = \bar{\zeta}_0 = 1. \tag{3.5}$$

The fact that θ in (3.3) drops out of (3.5) is, of course, due to the gauge invariance of P. If we now calculate the 2-form field (2.10) on $\mathbb{C}P_n$ we find that

$$\Omega = -\rho^4(\rho^{-2}d\zeta_j\wedge d\bar{\zeta}_j - \bar{\zeta}_j\zeta_l d\zeta_j\wedge d\bar{\zeta}_l)\,P, \tag{3.6}$$

and from (2.14) the line element on $\mathbb{C}P_n$ is

$$ds^2 = \rho^4(\rho^{-2}d\bar{\zeta}_j d\zeta_j - \zeta_j\bar{\zeta}_l d\bar{\zeta}_j d\zeta_l), \tag{3.7}$$

which is the Fubini–Study line element. We note that the curvature $F_{11} = F$ of the natural connection on the bundle (3.1) pulled back to $\mathbb{C}P_n$ via a cross section, specified by (3.3) with a given value of θ, is given by

$$F = \overline{\Psi}_\alpha\Omega_{\alpha\beta}\Psi_\beta, \tag{3.8}$$

and this is the coefficient of the matrix P in (3.6).

An important property of these bundles, first proved by Trautman,[1] is that the curvature F satisfies Maxwell's equations on $\mathbb{C}P_n$,

$$d\,{}^*F = 0, \tag{3.9}$$

where ${}^*F$ is the dual of F calculated with the use of the line element (3.7). This follows from Trautman's observation that ${}^*F$ is proportional to the wedge product of F with itself $(n-1)$–times and thus (3.9) follows from the U(1) Bianchi identities $dF = 0$.

In terms of the projector-valued field P, we observe that in general a similar argument to that leading to the form (2.12) of the Bianchi identities yields the result

$$P\,dP\wedge dP\wedge\cdots\wedge dP\,P = 0, \tag{3.10}$$

where the wedge product contains an odd number of factors dP. This is the content of (3.9) since, using (3.6) and (3.7), one has

$${}^*\Omega = \frac{-P}{2^{n-2}(n-1)!}\,dP\wedge dP\wedge\cdots\wedge dP, \tag{3.11}$$

where there are $(2n-2)$ factors of dP in the wedge product (an outline of the derivation of this equation is given in Appendix A). Hence, by (2.13), Maxwell's equation (3.9) becomes (3.10) with $(2n-1)$ factors of dP in the wedge product.

We have found, however, that P given by (3.5) satisfies a stronger condition than Maxwell's equations. We first have the equation

$$d\,{}^*(dP\wedge dP) = (n+1)dP\wedge{}^*F. \tag{3.12}$$

The derivation of this equation is outlined in Appendix B. We note that it has the following properties. It is consistent, since the exterior derivative of both sides vanishes on account of (3.9), and it implies that P satisfies the equation

$$\begin{aligned}d\,{}^*(dP\wedge dP) &- \{d\,{}^*(dP\wedge dP)\}P\\ &- Pd\,{}^*(dP\wedge dP) = 0,\end{aligned} \tag{3.13}$$

since $P^2 = P$, which in turn *implies* that P satisfies Maxwell's equations (3.9) in the form

$$P\{d\,{}^*(dP\wedge dP)\}P = 0. \tag{3.14}$$

4. SOME STIEFEL BUNDLES

Using the projector-valued field we may easily pass from the U(1) bundles

$$S_{2n+1}\to\mathbb{C}P_n \tag{4.1}$$

to the U(n) bundles

$$\text{SU}(n+1)\to\mathbb{C}P_n. \tag{4.2}$$

In the latter case the bundle space SU$(n+1)$ may be identified with $V_{n+1,n}$, i.e., with the set of all n-frames $V_{\alpha i}$,

$i = 1,2,...,n$, $\alpha = 0,1,2,...,n$ in $\mathbb{C}^{n+1}$. The projector-valued field (which is a field on $\mathbb{C}P_n$) is given by

$$Q_{\alpha\beta} = V_{\alpha i}\overline{V}_{\beta i} = \delta_{\alpha\beta} - V_{\alpha 0}\overline{V}_{\beta 0}, \tag{4.3}$$

where $V_{\alpha 0}$ is orthogonal to $V_{\alpha i}$. The last equality follows from the fact that $(V_{\alpha\beta}) \in SU(n+1)$. Making use again of the parametrization $V_{\alpha 0} = \rho\zeta_\alpha$, $\zeta_0 = 1, \rho^{-2} = 1 + \zeta_j\bar{\zeta}_j$ we see that

$$Q = 1 - P, \tag{4.4}$$

with P given by (3.5).

From the general theory the line element (2.14) on $\mathbb{C}P_n$ is given by

$$ds^2 = \tfrac{1}{2}\,\text{trace}\,(dQ)^2 \tag{4.5}$$

and since $dQ = -dP$ we again obtain the Fubini-Study line element (3.7) on $\mathbb{C}P_n$. From (2.10) with P replaced by Q, and then substituting as in (4.4), we find

$$\Omega = dP \wedge dP - P\,dP \wedge dP. \tag{4.6}$$

The Bianchi identities take the form

$$Q\,d\Omega\,Q = (-dP + dP\,P + P\,dP) \wedge dP \wedge dP = 0 \tag{4.7}$$

since $P^2 = P$. Finally we have

$$\begin{aligned} Qd*\Omega Q &= d*(dP \wedge dP) - \{d*(dP \wedge dP)\}P \\ &= -Pd*(dP \wedge dP) + P\{d*(dP \wedge dP)\}P \end{aligned} \tag{4.8}$$

and the right-hand side vanishes on account of (3.13) and (3.14). Hence the natural connection on the $U(n)$ Stiefel bundle $SU(n+1) \to \mathbb{C}P_n$ satisfies the Yang–Mills equations.

5. AN EXAMPLE

If we wish to extract from the gauge invariant formalism the connection and the curvature then we must specify a cross section of the bundle. We shall do this for the simplest nonabelian example in the present context, namely, the U(2) bundle

$$SU(3) \to \mathbb{C}P_2. \tag{5.1}$$

This example is also discussed by Nowakowski and Trautman.[3]

By (3.5) and (4.6) we have on on $\mathbb{C}P_n$ (see Appendix A)

$$\begin{aligned} \Omega_{00} &= \rho^6\bar{\zeta}_k\zeta_l d\zeta_k \wedge d\bar{\zeta}_l, \\ \Omega_{0k} &= -\rho^4\bar{\zeta}_j d\zeta_j \wedge d\bar{\zeta}_k + \rho^6\bar{\zeta}_k\bar{\zeta}_j\zeta_l d\zeta_j \wedge d\bar{\zeta}_l, \\ \Omega_{k0} &= -\rho^4\zeta_j d\zeta_k \wedge d\bar{\zeta}_j + \rho^6\zeta_k\bar{\zeta}_j\zeta_l d\zeta_j \wedge d\bar{\zeta}_l, \\ \Omega_{jl} &= \rho^2 d\zeta_j \wedge d\bar{\zeta}_l - \rho^4\bar{\zeta}_l\zeta_k d\zeta_j \wedge d\bar{\zeta}_k \\ &\quad - \rho^4\bar{\zeta}_k\zeta_j d\zeta_k \wedge d\bar{\zeta}_l + \rho^6\zeta_j\zeta_h\bar{\zeta}_l\bar{\zeta}_k d\zeta_k \wedge d\bar{\zeta}_h. \end{aligned} \tag{5.2}$$

Now put $n = 2$ and as a cross section of (5.1) we take the line in $\mathbb{C}^3$ through the origin and the point $(1,\zeta_1,\zeta_2)$ and map it to the 2-frame

$$\begin{aligned} \lambda_\alpha^{(1)} &= (0, -\rho_1\bar{\zeta}_2, \rho_1\bar{\zeta}_1), \\ \lambda_\alpha^{(2)} &= (\rho\rho_1^{-1}, -\rho\rho_1\zeta_1, -\rho\rho_1\zeta_2), \end{aligned} \tag{5.3}$$

where $\rho^{-2} = 1 + |\zeta_1|^2 + |\zeta_2|^2 = 1 + \rho_1^{-2}$. The connection 1-form on (5.1) pulled back to $\mathbb{C}P_2$ using this cross section is

$$\omega_{ij} = \bar{\lambda}_\beta^{(i)} d\lambda_\beta^{(j)}. \tag{5.4}$$

A straightfoward calculation yields the components

$$\begin{aligned} \omega_{11} &= \tfrac{1}{2}\rho_1^2(\zeta_2 d\bar{\zeta}_2 + \zeta_1 d\bar{\zeta}_1 - \bar{\zeta}_2 d\zeta_2 - \bar{\zeta}_1 d\zeta_1), \\ \omega_{12} &= \rho\rho_1^2(\zeta_2 d\zeta_1 - \zeta_1 d\zeta_2), \\ \omega_{21} &= -\overline{\omega}_{12}, \\ \omega_{22} &= \tfrac{1}{2}\rho^2\rho_1^2(\bar{\zeta}_1 d\zeta_1 + \bar{\zeta}_2 d\zeta_2 - \zeta_1 d\bar{\zeta}_1 - \zeta_2 d\bar{\zeta}_2). \end{aligned} \tag{5.5}$$

The components F_{ij} of the curvature 2-form pulled back to $\mathbb{C}P_2$ via (5.3) are given by

$$F_{ij} = \bar{\lambda}_\alpha^{(i)}\Omega_{\alpha\beta}\lambda_\beta^{(j)}, \tag{5.6}$$

from which we obtain

$$\begin{aligned} F_{11} &= \rho^2\rho_1^2(|\zeta_2|^2 d\zeta_1 \wedge d\bar{\zeta}_1 + |\zeta_1|^2 d\zeta_2 \wedge d\bar{\zeta}_2 - \zeta_2\bar{\zeta}_1 d\zeta_1 \wedge d\bar{\zeta}_2 \\ &\quad - \zeta_1\bar{\zeta}_2 d\zeta_2 \wedge d\bar{\zeta}_1), \\ F_{12} &= \rho^3\rho_1^2(\zeta_2^2 d\zeta_1 \wedge d\bar{\zeta}_2 - \zeta_1^2 d\zeta_2 \wedge d\bar{\zeta}_1 + \zeta_1\zeta_2 d\zeta_1 \wedge d\bar{\zeta}_1 \\ &\quad - \zeta_1\zeta_2 d\zeta_2 \wedge d\bar{\zeta}_2), \\ F_{21} &= -\overline{F}_{12}, \\ F_{22} &= \rho^4\rho_1^2(|\zeta_1|^2 d\zeta_1 \wedge d\bar{\zeta}_1 + |\zeta_2|^2 d\zeta_2 \wedge d\bar{\zeta}_2 + \bar{\zeta}_1\zeta_2 d\zeta_1 \wedge d\bar{\zeta}_2 \\ &\quad + \bar{\zeta}_2\zeta_1 d\zeta_2 \wedge d\bar{\zeta}_1). \end{aligned} \tag{5.7}$$

Since the Lie algebra of the structure group is the direct sum of u(1) and su(2), the connection components (5.5) and the curvature components (5.7) naturally decompose into a u(1)-part, which is Trautman's[1] electromagnetic instanton, and an su(2)-part attributed to Yang (private communication to the authors in Ref. 3). One could, of course, obtain (5.7) from the formalism of Nowakowski and Trautman.[3]

ACKNOWLEDGMENTS

I wish to thank Professor A. Trautman and Dr. J. Tafel, Dr. J. D. Mc Crea, and Dr. M. Clancy for helpful and stimulating discussions.

APPENDIX A

Since the details of the calculation of $*F$ appearing in (3.9) are not given by Trautman[1] it may be helpful to have an outline of the derivation of $*\Omega$ in (3.11) here.

We begin by defining the basis 1-forms for $\mathbb{C}P_n$

$$\sigma^{2k-1} = (1/\sqrt{2})\rho^2(\rho^{-2}\delta_{kj} - \bar{\zeta}_k\zeta_j)d\bar{\zeta}_j, \tag{A1}$$

$$\sigma^{2k} = (1/\sqrt{2})\rho^2 d\zeta_k, \tag{A2}$$

where $k = 1,2,3,...,n$. Then (3.7) reads

$$ds^2 = 2(\sigma^1\sigma^2 + \sigma^3\sigma^4 + ... + \sigma^{2n-1}\sigma^{2n}). \tag{A3}$$

In terms of this basis we calculate $dP_{\alpha\beta}$ using (3.5) and obtain

$$\begin{aligned} dP_{00} &= -\sqrt{2}\rho^2(\zeta_k\sigma^{2k-1} + \bar{\zeta}_k\sigma^{2k}), \\ dP_{0j} &= \sqrt{2}\rho^2(\sigma^{2j-1} - \bar{\zeta}_j\bar{\zeta}_k\sigma^{2k}), \\ dP_{j0} &= \sqrt{2}\rho^2\{(\rho^{-2}\delta_{jk} - \zeta_j\bar{\zeta}_k)\sigma^{2k} - \zeta_j\zeta_k\sigma^{2k-1}\}, \\ dP_{jl} &= \sqrt{2}\rho^2\zeta_j\sigma^{2l-1} + \sqrt{2}\rho^2\bar{\zeta}_l(\rho^{-2}\delta_{jk} - \zeta_j\bar{\zeta}_k)\sigma^{2k}. \end{aligned} \tag{A4}$$

Next we require

$$W_{\alpha\beta} = dP_{\alpha\lambda} \wedge dP_{\lambda\beta}, \tag{A5}$$

and for this we find

$$\begin{aligned} W_{00} &= -2\rho^2\sigma^{2j} \wedge \sigma^{2j-1} + 2\rho^2\zeta_j\bar{\zeta}_k\sigma^{2k} \wedge \sigma^{2j-1}, \\ W_{0k} &= -2\rho^2\bar{\zeta}_k\sigma^{2j} \wedge \sigma^{2j-1} - 2\rho^2\bar{\zeta}_j\sigma^{2j} \wedge \sigma^{2k-1}, \\ W_{k0} &= -2\zeta_j\sigma^{2k} \wedge \sigma^{2j-1} - 2\rho^2\zeta_k\sigma^{2j} \wedge \sigma^{2j-1} \\ &\quad + 2\rho^2\zeta_k\zeta_j\bar{\zeta}_h\sigma^{2h} \wedge \sigma^{2j-1}, \end{aligned}$$

$$W_{jl} = 2\sigma^{2j} \wedge \sigma^{2l-1} - 2\rho^2 \xi_j \bar{\xi}_l \sigma^{2k} \wedge \sigma^{2k-1} - 2\rho^2 \xi_j \bar{\xi}_k \sigma^{2k} \wedge \sigma^{2l-1}. \tag{A6}$$

and finally we have [cf. Eq. (2.10)]

$$\Omega = P\,dP \wedge dP = PW = -2P\sigma^{2j} \wedge \sigma^{2j-1} \tag{A7}$$

and so

$$*\Omega = -2P *(\sigma^{2j} \wedge \sigma^{2j-1}). \tag{A8}$$

Taking $\sigma^1 \wedge \sigma^2 \wedge \ldots \wedge \sigma^{2n-1} \wedge \sigma^{2n}$ as the volume element on $\mathbb{C}P_n$, we obtain, for each $j = 1,2,3\ldots,n$ (see, for example, Sternberg,[11] p. 23)

$$*(\sigma^{2j} \wedge \sigma^{2j-1}) = \sigma^1 \wedge \sigma^2 \wedge \cdots \wedge \widetilde{\sigma^{2j-1} \wedge \sigma^{2j}} \wedge \cdots \wedge \sigma^{2n}, \tag{A9}$$

$j = 1,2,3,\ldots,n$ where the tilde means "with this term missing." Hence returning to (A8), where the summation convention applies, and using (A7) and the result (easily proved by induction) that if

$$X = \sum_{j=1}^{n} \sigma^{2j-1} \wedge \sigma^{2j}, \tag{A10}$$

then

$$X \wedge X \wedge \cdots \wedge X = (n-1)! \sum_{j=1}^{n} \sigma^1 \wedge \sigma^2 \wedge \cdots \wedge \widetilde{\sigma^{2j-1} \wedge \sigma^{2j}} \wedge \cdots \wedge \sigma^{2n}, \tag{A11}$$

where there are $(n-1)$–factors of X on the left-hand side, we arrive at

$$*\Omega = \frac{-P}{2^{n-2}(n-1)!}\, dP \wedge dP \wedge \cdots \wedge dP, \tag{A12}$$

with $(2n-2)$–factors of dP as in (3.11).

APPENDIX B

We outline here the derivation of (3.12). In the notation introduced in (A5) we wish to show that

$$d *W = (n+1) dP \wedge *F. \tag{B1}$$

This calculation requires us to evaluate terms of the form $*(\sigma^{2k} \wedge \sigma^{2j-1})$ and $d *(\sigma^{2k} \wedge \sigma^{2j-1})$ where $k,j = 1,2,3,\ldots,n$. To do this we make use of the completely skew-symmetric Levi–Civita pseudo-tensor $\eta_{\alpha_1\alpha_2\cdots\alpha_{2n}} = \eta_{[\alpha_1\alpha_2\cdots\alpha_{2n}]}$ on $\mathbb{C}P_n$, where in this Appendix Greek letters take values 1,2,3,...,2n, with

$$\eta_{123\cdots 2n} = |\det g_{\mu\nu}|^{1/2} = 1, \tag{B2}$$

where $g_{\mu\nu}$ is the $2n \times 2n$ metric tensor on the basis (A1) with blocks of

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

down the diagonal and zeros everywhere else. With $\theta^\mu = (\sigma^1,\sigma^2,\sigma^3,\ldots,\sigma^{2n})$ we use the quantities[12,13]

$$\eta = *1 = \frac{1}{(2n)!}\, \eta_{\alpha_1\alpha_2\cdots\alpha_{2n}} \theta^{\alpha_1} \wedge \theta^{\alpha_2} \wedge \cdots \wedge \theta^{\alpha_{2n}},$$

$$\eta^\alpha = *\theta^\alpha = \frac{1}{(2n-1)!}\, \eta_{\alpha\alpha_1\cdots\alpha_{2n-1}} \theta^{\alpha_1} \wedge \theta^{\alpha_2} \wedge \cdots \wedge \theta^{\alpha_{2n-1}},$$

$$\eta^{\alpha\beta} = *(\theta^\alpha \wedge \theta^\beta) = \frac{1}{(2n-2)!}\, \eta^{\alpha\beta}{}_{\alpha_1\alpha_2\cdots\alpha_{2n-2}} \theta^{\alpha_1} \wedge \theta^{\alpha_2} \wedge \cdots \wedge \theta^{\alpha_{2n-2}}, \tag{B3}$$

and so on, together with the following consequences:

$$\theta^\mu \wedge \eta_{\ldots\alpha\beta} = \delta^\mu_\beta \eta_{\ldots\alpha} - \delta^\mu_\alpha \eta_{\ldots\beta} + \ldots,$$

$$d\eta_{\alpha_1\alpha_2\cdots\alpha_{2n}} = 0,$$

$$d\eta_{\alpha_1\alpha_2\cdots\alpha_k} = d\theta^\beta \wedge \eta_{\alpha_1\alpha_2\cdots\alpha_k\beta}. \tag{B4}$$

We also require the following, easily obtained from (A1),

$$d\sigma^{2k-1} = -\sqrt{2}\bar{\xi}_k \sigma^{2j} \wedge \sigma^{2j-1} - \sqrt{2}\xi_j \sigma^{2k-1} \wedge \sigma^{2j-1},$$

$$d\sigma^{2k} = -\sqrt{2}\xi_j \sigma^{2j-1} \wedge \sigma^{2k} - \sqrt{2}\bar{\xi}_j \sigma^{2j} \wedge \sigma^{2k}. \tag{B5}$$

Using this and (B3) and (B4) we obtain

$$d *(\sigma^{2k} \wedge \sigma^{2j-1}) = -\sqrt{2}n\bar{\xi}_j \eta^{2k} + \sqrt{2}\delta_{kj}\bar{\xi}_h \eta^{2h}. \tag{B6}$$

We also make use of

$$2\xi_k \eta^{2k-1} = \xi_k \sigma^{2k-1} \wedge *F, \quad -2\bar{\xi}_k \eta^{2k} = \bar{\xi}_k \sigma^{2k} \wedge *F. \tag{B7}$$

Hence, for example,

$$\begin{aligned} d *W_{00} &= dP_{00} \wedge *F - 2\sqrt{2}\rho^2 n \xi_j \eta^{2j-1} + 2\sqrt{2}\rho^2 n \bar{\xi}_j \eta^{2j} \\ &= dP_{00} \wedge *F - n\sqrt{2}\rho^2 (\bar{\xi}_j \sigma^{2j} + \xi_j \sigma^{2j-1}) \wedge *F \\ &= dP_{00} \wedge *F + n dP_{00} \wedge *F. \end{aligned} \tag{B8}$$

Following this procedure for the remaining components of $d *W$ we eventually arrive at (B1).

[1] A. Trautman, Int. J. Theor. Phys. **16**, 561 (1977).
[2] A. A. Belavin, A. M. Polyakov, A. S. Schwartz, and Yu. S. Tyupkin, Phys. Lett. B **59**, 85 (1975).
[3] J. Nowakowski and A. Trautman, J. Math. Phys. **19**, 1100 (1978).
[4] J. Harnad, J. Tafel, and S. Shnider, J. Math. Phys. **21**, 2236 (1980).
[5] These bundles have recently been used to obtain the magnetic monopole solution of Maxwell's equations with strength n times the strength of the weakest monopole, by P. A. Hogan, J. Math. Phys., to appear; there is a concomitant result[1] for the bundles $S_{2n+1} \to \mathbb{C}P_n$.
[6] E. F. Corrigan, D. B. Fairlie, S. Templeton, and P. Goddard, Nucl. Phys. B **140**, 31 (1978).
[7] M. Dubois-Violette, Phys. Lett. B **82**, 251 (1979).
[8] M. Dubois-Violette, in *Regards sur la Physique Contemporaine* (Éditions du Centre National de la Recherche Scientifique, Paris, 1979), p. 47.
[9] M. S. Narasimhan and S. Ramanan, Am. J. Math. **83**, 563 (1961).
[10] R. Schlafly, Invent. Math. **59**, 59 (1980).
[11] S. Sternberg, *Lectures on Differential Geometry* (Prentice–Hall, Englewood Cliffs, NJ, 1964).
[12] A. Trautman, Bull. Acad. Pol. Sci. **20**, 185 (1972).
[13] J. Tafel, "Lecture Notes on Gauge Theories," Institute of Theoretical Physics, University of Warsaw, unpublished, 1981.

The problem of physical coordinates in predictive Hamiltonian systems

V. Iranzo, J. Llosa, F. Marqués, and A. Molina
Departament de Física Teòrica, Universitat de Barcelona, Barcelona-28, Spain

(Received 1 December 1981; accepted for publication 11 August 1982)

In the Hamiltonian formulation of predictive relativistic systems, the canonical coordinates cannot be the physical positions. The relation between them is given by the individuality differential equations. However, due to the arbitrariness in the choice of Cauchy data, there is a wide family of solutions for these equations. In general, those solutions do not satisfy the condition of constancy of velocities' moduli, and therefore we have to reparametrize the world lines into the proper time. We derive here a condition on the Cauchy data for the individuality equations which ensures the constancy of the velocities' moduli and makes the reparametrization unnecessary.

PACS numbers: 11.80. – m, 02.30.Jr, 11.30.Cp

I. INTRODUCTION

The description of the relativistic interaction at a distance between particles has been a difficult task both from the mathematical and the conceptual viewpoints. There are some different ways to face these problems in the literature.[1–4]

In this paper we consider two versions of an action-at-a-distance theory called predictive relativistic mechanics (PRM). The main hypothesis of this theory are: (i) Poincaré invariance and (ii) the dynamics of the interacting particles is given by a second-order differential system and, therefore, the motion is completely determined by the initial positions and velocities. The versions of the PRM which we consider here are the so called "multitime formalism"[5] and "*a priori* Hamiltonian formulation,"[6] respectively.

In the multitime formalism, Hamiltonian descriptions have been derived by Bel and co-workers[7] for the electromagnetic interaction and for the scalar and vector short range ones. However, this method leads inevitably from the beginning to power expansions in some constant and, although there are iterative algorithms to obtain all interesting functions order by order, their convergence has not been proved yet. Also, it is hard to go beyond order two.

In the *a priori* Hamiltonian framework[6] it is possible to consider a wide class of models which do not need perturbative methods, at least in the first steps. However, this formalism is objected to in that it does not allow us to describe physical interactions like those mentioned above. Anyhow, the *a priori* Hamiltonian formalism will provide exact simple models of interacting relativistic particles whose study will permit us a better understanding of quantization.

Both frameworks can be related to one another by integrating the so-called position equations–see Sec. II. As we shall see later, a Poincaré-invariant predictive system (multitime formalism) can be associated with each "*a priori*" Hamiltonian one plus a good set of Cauchy data for the position equations. However, this correspondence usually implies reparametrization of the world lines of each particle into its individual proper time. This is studied in detail in Secs. II and III.

Because of the necessity of reparametrizing the world lines, the associated Poincaré-invariant predictive system does not coincide with the starting Hamiltonian one, but they are proportional, as we shall see in Sec. III. Therefore, the former does not leave the symplectic form invariant whereas the Hamiltonian system does.

As we have mentioned above, an "*a priori*" Hamiltonian system is in correspondence with a wide class of Poincaré-invariant predictive ones, depending on the prescribed Cauchy data. In Secs. IV and V we study how the Cauchy data can be restricted in order to ensure the coincidence between the "*a priori*" Hamiltonian system and the associated Poincaré-invariant predictive one.

II. PREDICTIVE HAMILTONIAN SYSTEMS

In predictive relativistic mechanics the basic equations of motion consist of N coupled second-order differential systems[8]

$$\frac{\partial x_a^\mu}{\partial \sigma_b} = \delta_{ab}\pi_a^\mu, \quad \frac{\partial \pi_a^\mu}{\partial \sigma_b} = \delta_{ab}\theta_a^\mu(x,\pi),$$
$$a, b, c, \cdots = 1,\ldots,N, \quad \mu,\nu,\cdots = 0,1,2,3. \tag{2.1}$$

Each one defines a vector field on the phase space TM^{4N}

$$\mathbf{H}_a = \pi_a^\mu \cdot \frac{\partial}{\partial x^{a\mu}} + \theta_a^\mu(x,\pi)\cdot\frac{\partial}{\partial \pi^{a\mu}}, \quad a = 1,\ldots,N. \tag{2.2}$$

The integrability conditions of (2.1) read

$$[\mathbf{H}_a, \mathbf{H}_{a'}] = 0, \quad \forall\, a' \neq a, \tag{2.3}$$

meaning that if, starting from a point P in the phase space, some among the particles are transported along their respective world lines to a new point $Q \in TM^{4N}$, the same integral of (2.1) is obtained irrespectively that you take either P or Q as initial conditions.

It is easy to proof the following[9,10]:

Theorem 2.1: Let $\mathbf{H}_1,\ldots,\mathbf{H}_N$ be vector fields of the type (2.2) such that $[\mathbf{H}_a, \mathbf{H}_{a'}] = 0$, $\forall\, a' \neq a$. Then, there exists an open neighborhood $\mathcal{U}$ of $\{0\} \times TM^{4N} \subset \mathbb{R}^N \times TM^{4N}$ and a continuously differentiable mapping

$$\varphi: \mathcal{U} \to TM^{4N},$$
$$(\sigma_1,\ldots,\sigma_N; P) \to (\varphi_a^\mu(\sigma_a, x_b^\nu, \pi_c^\rho), \dot{\varphi}_a^\mu(\sigma_a, x_b^\nu, \pi_c^\rho)), \tag{2.4}$$
$$P = (x_b^\nu, \pi_c^\rho),$$

such that:

 0022-2488/83/061665-07$02.50 © 1983 American Institute of Physics

(i) $\varphi(0,P)=P, \quad \forall\, P\in TM^{4N}, \quad 0=(0,...,0)\in\mathbb{R}^N,$

(ii) $(\sigma^2,P)\in\mathcal{U}, \quad (\sigma^1,\varphi(\sigma^2,P))\in\mathcal{U}, \quad (\sigma^1+\sigma^2,P)\in\mathcal{U}$

$\Rightarrow \varphi(\sigma^1,\varphi(\sigma^2,P))=\varphi(\sigma^1+\sigma^2,P);$ (2.5)

(iii) for any function f on TM^{4N},

$$\left[\frac{\partial}{\partial\sigma_a}f(\varphi(\sigma_1,...,\sigma_N;P))\right]_{\sigma_i=0}=(\mathbf{H}_a f)_P \tag{2.6}$$

(this means that $\mathbf{H}_a$ is the tangent vector to the trajectory of the particle a in the phase space);

(iv) uniqueness.

A very important feature due to the Kronecker δ's appearing in (2.1) is that φ^μ_a depends only on the parameter σ_a. This property is called the "individuality condition."

Theorem 2.2: Let $\mathcal{U}$ be a neighborhood of $\{0\}\times TM^{4N}\subset\mathbb{R}^N\times TM^{4N}$ and let φ be a continuously differentiable mapping:

$$\varphi:\ \mathcal{U}\to TM^{4N}$$

$$(\sigma_1,...,\sigma_N;P)\to(\varphi_a(\sigma_a,P),\dot{\varphi}_a(\sigma_a,P))$$

with the properties (i) and (ii).

Then the vector fields $\{\mathbf{H}_a\}_{a=1...N}$ defined by (2.6) are of the type (2.2) and fulfill the conditions (2.3).

The Poincaré invariance of the differential system (2.1) is assured iff θ^μ_a are translation-invariant 4-vectors. Then

$$[\mathbf{P}_\mu,\mathbf{H}_a]=0, \quad [\mathbf{J}_{\mu\nu},\mathbf{H}_a]=0, \tag{2.7}$$

where $\mathbf{P}_\mu$ and $\mathbf{J}_{\mu\nu}$ are the generators of the Poincaré group.

Equations (2.3) and (2.7) exhibit that $\mathbf{P}_\mu$, $\mathbf{J}_{\mu\nu}$, and $\mathbf{H}_a$ span a realization of an abelian extension of the Poincaré algebra.[7] The associated transformation group $\mathcal{P}\otimes\mathcal{A}_N$ is the direct product of the Poincaré group and the abelian one $\mathcal{A}_N\simeq\mathbb{R}^N$.

Hereafter will be called a Poincaré-invariant abelian system (PIAS) any family of differential systems $\{\mathbf{H}_a,\ a=1,...,N\}$ of the type (2.2) fulfilling (2.3) and (2.7).

In predictive relativistic mechanics it is also required that π_a^2 be an integral of motion for any system $\{\mathbf{H}_b,\ b=1,...,N\}$. This means that the parameter σ_a is the proper time of particle a, apart from a multiplicative constant. This implies the following relation:

$$\pi_{a\mu}\cdot\theta^\mu_a(x,\pi)=0. \tag{2.8}$$

A PIAS satisfying the conditions (2.8) will be called hereafter a Poincaré-invariant predictive system (PIPS).

A Hamiltonian formalism for the system (2.1) will be determined by picking up a simplectic Ω form on TM^{4N} invariant under $\mathcal{P}\otimes\mathcal{A}_N$.

As is well known, a function f can be associated to each field Λ, which leaves Ω invariant. This function is defined in TM^{4N}, at least locally, by

$$df=i_\Lambda\,\Omega,$$

and it is unique, apart from an arbitrary additive constant [in Eq. (2.8) i means the interior product[9]].

In particular, the functions associated with $\mathbf{P}_\mu$, $\mathbf{J}_{\mu\nu}$, and $\mathbf{H}_a$ are, respectively, the linear momentum, the angular momentum, and the Hamiltonians.

It is important to realize that in all cases of interest: $\Omega\neq dx^\mu_a\wedge d\pi^a_\mu$ in order to circumvent the no-interaction theorem.[3]

The preceding are the main features of the natural approach to obtain a Hamiltonian description [as a first step to quantization of a classical system like (2.1)]. A very important fact of this approach is that it usually involves perturbative methods, as it can be seen in the cases of electromagnetic and of short range scalar and vector interactions.[7,11]

Another approach to predictive relativistic mechanics is given by the so-called "*a priori* Hamiltonian predictive systems." It consists in giving N scalar functions $H_a(q,p)$ on the phase space TM^{4N} endowed with a symplectic form:

$$\Omega=dq^\mu_a\wedge dp^a_\mu \tag{2.9}$$

where, in order to assure the Poincaré invariance, the canonical coordinates q^μ_a behave like x^μ_a under Poincaré transformations and the momenta p^a_μ are translation invariant 4-vectors. This implies

$$\mathbf{P}_\mu=\epsilon_a\frac{\partial}{\partial q^\mu_a}$$

$$\mathbf{J}_{\mu\nu}=q_{a\mu}\cdot\frac{\partial}{\partial q^\nu_a}-q_{a\nu}\cdot\frac{\partial}{\partial q^\mu_a}+p^a_\mu\cdot\frac{\partial}{\partial p^{a\nu}}-p^a_\nu\cdot\frac{\partial}{\partial p^{a\mu}} \tag{2.10}$$

and the corresponding functions:

$$P_\mu=\epsilon_a p^a_\mu, \quad J_{\mu\nu}=q_{a\mu}\cdot p^a_\nu-q_{a\nu}\cdot p^a_\mu. \tag{2.10'}$$

The simplectic form (2.9) defines the Poisson bracket:

$$\{q^\mu_a,p^b_\nu\}=\delta^b_a\cdot\eta^\mu_\nu, \quad \{q^\mu_a,q^\nu_b\}=\{p^a_\mu,p^b_\nu\}=0, \tag{2.11}$$

in terms of which the Hamilton equations read

$$\frac{\partial q^\mu_a}{\partial\tau_b}=\{H_b,q^\mu_a\}, \quad \frac{\partial p^a_\mu}{\partial\tau_b}=\{H_b,p^a_\mu\}, \quad a,b=1,2,...,N, \tag{2.12}$$

and their integrability conditions:

$$\{H_a,H_{a'}\}=0, \quad \forall\, a'\neq a. \tag{2.13}$$

Assuming that the conditions (2.13) are fulfilled by the Hamiltonians, the Hamilton equations (2.12) are integrable, i.e., there exist functions $f^\mu_a(\tau_1,...,\tau_N;q^\nu_b,p^c_\rho)$ and $g^a_\mu(\tau_1,...,\tau_N;q^\nu_b,p^c_\rho)$ which are solutions of (2.12) and

$$f^\mu_a(0,...,0;q^\nu_b,p^c_\rho)=q^\mu_a, \quad g^a_\mu(0,...,0;q^\nu_b,p^c_\rho)=p^a_\mu.$$

However, there is still one important difference between these solutions and the integral (2.4). In general, f^μ_a and g^a_μ do not accomplish the individuality condition and depend essentially on more than one of the parameters $(\tau_1,...,\tau_N)$. Furthermore, if we require f^μ_a and g^a_μ to depend only on τ_a ($\forall\, a=1,...,N$), then it turns out—no-interaction theorem[3]—that only free particle Hamiltonian systems are allowed.

It is for this reason that the *a priori* canonical coordinates (q,p), have no physical meaning; therefore, we cannot identify q^μ_a with the physical position x^μ_a and, henceforth $\Omega\neq dx^\mu_a\wedge dp^a_\mu$.

In order to give some physical content to these "*a priori* Hamiltonian predictive systems," we must introduce the physical positions $x^\mu_a(q,p)$, which we define by the individuality condition:

$$\{H_{a'},x^\mu_a\}=0, \quad \forall\, a'\neq a, \tag{2.14}$$

which are $N-1$ partial differential equations on x_a^μ.

A system of N scalar functions $H_a(q,p)$ fulfilling the integrability conditions (2.13) and the individuality condition (2.14) will be called hereafter an "*a priori* Hamiltonian predictive system."

As far as we have assumed that (2.13) is fulfilled, the partial differential system (2.14) is integrable. Therefore, a good set of Cauchy data determines a unique solution of (2.14): $x_a^\mu(q,p)$.

We can then define the velocities

$$\dot{x}_a^\mu(q,p) \equiv \{H_a, x_a^\mu\}. \tag{2.15}$$

From (2.13) and the Jacobi identity for Poisson brackets it is obvious that

$$\{H_{a'}, \dot{x}_a^\mu\} = 0, \quad \forall\, a' \neq a. \tag{2.16}$$

If the Jacobian $|\partial(x_a^\mu, \dot{x}_b^\nu)/\partial(q_a^\mu, p_\nu^b)|$ does not vanish, the functions $x_a^\mu(q,p)$ and $\dot{x}_b^\nu(q,p)$ define a change of variables. Then, the vector field $\mathbf{H}_a^*$ associated with $H_a(q,p)$ by Ω can be written as

$$\mathbf{H}_a^* = \dot{x}_a^\mu \frac{\partial}{\partial x^{a\mu}} + \xi_a^\mu(x,\dot{x}) \cdot \frac{\partial}{\partial \dot{x}^{a\mu}}, \quad a = 1,\ldots,N, \tag{2.17}$$

where $\xi_a^\mu(x(q,p), \dot{x}(q,p)) \equiv \{H_a(q,p), \dot{x}_a(q,p)\}$.

These systems are of the same type as (2.2) and, as a consequence of (2.13), they obviously satisfy the conditions (2.3)[9]:

$$[\mathbf{H}_a^*, \mathbf{H}_{a'}^*] = 0.$$

Since $H_a(q,p)$ are scalar functions, $\dot{x}_a^\mu$ and $\xi_a^\mu(x,\dot{x})$ are translation-invariant 4-vectors.

What has been achieved is that any "*a priori* Hamiltonian predictive system" (HPS) with a good set of Cauchy data for (2.14) determines a unique PIAS, provided that the positions and velocities $(x_a^\mu, \dot{x}_b^\nu)$ are a complete set of coordinates on phase space.

We have to realize that the PIAS associated with the HPS depends crucially on the Cauchy data for the position equations. In general, two different sets of Cauchy data for (2.14) associate to the same HPS two different PIAS.

In general, the PIAS (2.17) is not affine parametrized, i.e., it is not a PIPS, because the condition (2.8) $\dot{x}_{a\mu} \cdot \xi_a^\mu(x,\dot{x}) = 0$ need not be fulfilled.

One question now arises: Is it possible to obtain a PIPS equivalent to (2.17) via a reparametrization?

III. REPARAMETRIZATION OF A PIAS

Let us start from a PIAS:

$$\mathbf{H}_a^* = \dot{x}_a^\mu \cdot \frac{\partial}{\partial x^{a\mu}} + \xi_a^\mu(x,\dot{x}) \cdot \frac{\partial}{\partial \dot{x}^{a\mu}}, \quad a = 1,\ldots,N.$$

We introduce now the following change of variables in TM^{4N}: $(x_b, \dot{x}_c) \to (x_b, \pi_c)$ defined by

$$\pi_a^\mu \equiv \beta_a(x_b, \dot{x}_c) \cdot \dot{x}_a^\mu, \tag{3.1}$$

where $\beta_a(x_b, \dot{x}_c)$ are N given scalar functions.

This transformation leaves the position invariant and only modifies the velocities' moduli. The inverse transformation is given by

$$\dot{x}_a^\nu = \alpha_a(x_b, \pi_c) \cdot \pi_a^\nu, \tag{3.2}$$

where the N functions $\alpha_a(x_b, \pi_c)$ can be obtained from

$$\alpha_a \cdot \beta_a(x_b, \alpha_c \pi_c) = 1, \quad a = 1,\ldots,N, \tag{3.3}$$

provided that the conditions of the implicit function theorem[12] are fulfilled. Moreover, it is obvious that this occurs if and only if $\{\beta_a(x_b, \dot{x}_c)\sqrt{-\dot{x}_a^2}\}_{a=1,\ldots,N}$ are functionally independent.

We then define the following second-order differential systems:

$$\mathbf{H}_a = \pi_a^\mu \cdot \frac{\partial}{\partial x^{a\mu}} + \theta_a^\mu(x,\pi) \cdot \frac{\partial}{\partial \pi^{a\mu}}, \quad a = 1,\ldots,N,$$

with

$$\theta_a^\mu(x,\pi) \equiv (\mathbf{H}_a^* \beta_a)(x,\alpha\pi) \cdot \pi_a^\mu + \beta_a^2(x,\alpha\pi) \cdot \xi_a^\mu(x,\alpha\pi). \tag{3.4}$$

The following two theorems are referred to the relations between PIAS and PIPS defined by reparametrizations.

Theorem 3.1: If the functions $\beta_a(x,\dot{x})$ introduced in (3.1) satisfy

$$\mathbf{H}_{a'}^* \beta_a = 0, \quad \forall\, a' \neq a, \tag{3.5}$$

then:

(i) The family of second-order differential systems $\{\mathbf{H}_a\}_{a=1,\ldots,N}$ defined in (3.4) is a PIAS.

(ii) Any integral of $\{\mathbf{H}_a^*\}_{a=1,\ldots,N}$ can be reparametrized into an integral of $\{\mathbf{H}_a\}_{a=1,\ldots,N}$.

Theorem 3.2:

Let

$\{\mathbf{H}_a^* = \dot{x}_a^\mu \cdot \partial/\partial x^{a\mu} + \xi_a^\mu(x,\dot{x}) \cdot \partial/\partial \dot{x}^{a\mu};\ a = 1,\ldots,N\}$ be a PIAS and

$\{\mathbf{H}_a = \pi_a^\mu \cdot \partial/\partial x^{a\mu} + \theta_a^\mu(x,\pi) \cdot \partial/\partial \pi^{a\mu};\ a = 1,\ldots,N\}$ be a PIPS.

Assume that any integral $\varphi_a^\mu(\sigma_a, x, \pi)$ of the second can be reparametrized into an integral $\psi_a^\mu(\tau_a, x, \dot{x})$ of the first. In other words, there exist $2N$ functions $f_a(\tau_b, x, \dot{x})$ and $\beta_b(x, \dot{x})$ as much differenciable as necessary, such that

$$\psi_a^\mu(\tau_a, x_b, \dot{x}_c) = \varphi_a^\mu(f_a(\tau_a, x, \dot{x}), x_b^\nu, \beta_c(x,\dot{x}) \cdot \dot{x}_c^\rho). \tag{3.6}$$

Then:

$$\text{(i)} \quad f_a(0, x, \dot{x}) = 0, \quad \beta_a(x,\dot{x}) = \left(\frac{\partial f_a}{\partial \tau_a}\right)_{(0,x,\dot{x})}^{-1}; \tag{3.7}$$

$$\text{(ii)} \quad \left(\frac{\partial f_a}{\partial \tau_a}\right)_{(\tau_a, x, \dot{x})}^{-1} = \beta_a(\psi(\tau, x, \dot{x}), \dot{\psi}(\tau, x, \dot{x})); \tag{3.8}$$

$$\text{(iii)} \quad \mathbf{H}_{a'}^* \beta_a = 0, \quad \forall\, a' \neq a; \tag{3.9}$$

$$\text{(iv)} \quad \xi_a^\mu(x,\dot{x}) = \beta_a^{-2}(x,\dot{x}) \cdot \theta_a^\mu(x, \beta\dot{x}) + \left(\frac{\partial^2 f_a}{\partial \tau_a^2}\right)_{(0,x,\dot{x})} \cdot \beta_a \cdot \dot{x}_a^\mu. \tag{3.10}$$

We have to realize that the reparametrization functions f_a depend not only on τ_a but also on x_b^μ and $\dot{x}_c^\nu$. This is due to the fact that each integral of the PIAS must be reparametrized independently of the others. This makes the problem much more involved than it seems at first sight. The foregoing two theorems establish which conditions must be imposed on the functions $f_a(\tau_a, x_b^\mu, \dot{x}_c^\nu)$ in order to preserve the predictive abelian character of the system after reparametri-

zation. This complication has not been noticed in the least in previous works.[6] It turns out, however, that all reparametrizations of PIAS existing in the literature happen to fulfill Eqs. (3.5) and hence keep the abelian character of the system.

IV. A NATURAL RESTRICTION ON THE CAUCHY DATA

We have concluded in Sec. II that the PIAS associated with a given HPS depends very strongly on the Cauchy data taken to solve the position equations (2.14).

The big variety of Cauchy data for Eq. (2.14) lies in two facts:

(i) The Cauchy surface Σ is any $(7N+1)$ submanifold of the phase space, with the only restrictions of being Poincaré invariant and of not being a characteristic of (2.14). (ii) The prescription

$$[x_a^\mu - f_a^\mu(q,p)]_\Sigma = 0 \tag{4.1}$$

is only restricted by good behavior under the Poincaré group.

In order to reduce somehow the wide variety of solutions of the position equations, we can furthermore introduce the orthogonality conditions (2.8), which in Hamiltonian formalism are written as

$$\{H_a, x_{a\mu}\} \cdot \{H_a, \{H_a, x_a^\mu\}\} = 0. \tag{4.2}$$

A first integration of this equation can be worked out immediately and leads to

$$\{H_a, x_{a\mu}\} \cdot \{H_a, x_a^\mu\} = -m_a^2(q,p), \tag{4.3}$$

where $m_a^2(q,p)$ is a positive integral of motion which is usually taken to be $2H_a(q,p)$.[6] We shall also do it hereafter.

Although (4.3) is apparently a new differential equation to add to Eq. (2.14), we shall see in the following that it can be reduced to a condition on the Cauchy prescription (4.1).

Let us first consider a Cauchy surface satisfying the above conditions (i). It will be defined by $N-1$ scalar functions:

$$\phi_B(q,p) = 0, \quad B = 2,\ldots, N,$$

and no linear combination of the vectors $\mathbf{H}_b^*$, $b \neq a$, can be tangent to Σ. That is to say:

$$\left[\sum_{c\neq a} b^c(q,p) \cdot \{H_c, \phi_B\}\right]_\Sigma = 0, \quad \forall\ B = 2,\ldots, N$$
$$\Rightarrow [b^c(q,p)]_\Sigma = 0, \quad \forall\ c \neq a,$$

which is satisfied if and only if

$$\det(\{\phi_B, H_c\}_{|\Sigma}) \neq 0, \quad B = 2,\ldots, N, \quad c \neq a. \tag{4.4}$$

If we want Σ to be a good Cauchy surface for the position equations of any particle, Eq. (4.4) need to be true for any $a = 1,\ldots, N$.

Let us now consider a vector field of the form

$$\mathbf{D}_a = \mathbf{H}_a^* + \sum_{c\neq a} \psi_{(a)}^c \mathbf{H}_c^*. \tag{4.5}$$

The functions $\psi_{(a)}^c$ can always be chosen in such a way that $\mathbf{D}_{a|\Sigma}$ is tangent to Σ, that is to say,

$$[\mathbf{D}_a \phi_B]_\Sigma = 0, \quad \forall\ B = 2,\ldots, N, \tag{4.6}$$

which can also be written as

$$\left[\{\phi_B, H_a\} + \sum_{c\neq a} \psi_{(a)}^c \{\phi_B, H_c\}\right]_\Sigma = 0. \tag{4.7}$$

But from (4.4) the linear system (4.7) is Cramer's and the restriction to Σ of $\psi_{(a)}^c(q,p)$ is uniquely determined. Substituting then these suitable values of $\psi_{(a)}^c$ into (4.5), we have for any $g(q,p)$

$$[\mathbf{D}_a\, g]_\Sigma = \mathbf{D}_{a|\Sigma}(g_{|\Sigma}). \tag{4.8}$$

Therefore, taking (2.14) and (4.1) into account and using (4.8), we can write

$$[\mathbf{H}_a^* x_a^\mu]_\Sigma = [\mathbf{D}^a x_a^\mu]_\Sigma = \mathbf{D}_{a|\Sigma} f_a^\mu. \tag{4.9}$$

Theorem 4.1: The solution of Eq. (2.14) with the set of Cauchy data $[x_a^\mu - f_a^\mu(q,p)]_\Sigma = 0$ satisfy also the condition (4.3) if and only if

$$(\mathbf{D}_{a|\Sigma} f_{a\mu})(\mathbf{D}_{a|\Sigma} f_a^\mu) = -m_a^2(q,p). \tag{4.10}$$

Proof: (a) (if) Let $x_a^\mu(q,p)$ be the solution of (2.14) with these Cauchy data, and let us define $g_a(q,p) \equiv \{H_a, x_a^\mu\} \cdot \{H_a, x_{a\mu}\}$. From (4.9) and (4.10) it follows that

$$g_{a|\Sigma} = -m_{a|\Sigma}^2. \tag{4.11}$$

Then, using (2.13), (2.16), and the Jacobi identity for Poisson brackets, we have

$$\{H_{a'}, g_a\} = 0, \quad a' \neq a. \tag{4.12}$$

Therefore, g_a is a solution of (4.12) with the Cauchy data (4.11). But the uniqueness of solutions implies

$$g_a(q,p) = -m_a^2(q,p),$$

and the first part of the proof (if) is finished.

(b) The second part of the proof (only if) is immediate.

V. APPLICATION

We shall here apply the results obtained in the last section to the family of models presented in Ref. 13 and given by the Hamiltonian functions:

$$H_b = \alpha(P \cdot y_b) - \beta P^2 - \gamma\left[\varepsilon^A y_A^2 + \sum_{A<A'} (y_A \cdot y_{A'})\right] + V(\tilde{y}_B, \tilde{z}_C) \tag{5.1}$$

where

$$P^\mu \equiv \varepsilon^a p_a^\mu, \quad y_B^\mu \equiv (1/N)P^\mu - p_B^\mu, \quad z_c^\nu \equiv q_c^\nu - q_1^\nu, \tag{5.2a}$$

$$y_1^\mu \equiv -\varepsilon^A y_A^\mu, \tag{5.2b}$$

$$\tilde{a}^\mu \equiv \pi_\nu^\mu \cdot a^\nu, \quad \pi_\nu^\mu \equiv \eta_\nu^\mu - P^\mu P_\nu / P^2, \tag{5.2c}$$

$\varepsilon^A = 1$, $a,b,c = 1,2,\ldots, N$, $A,B,C = 2,3,\ldots, N$, and α,β,γ are arbitrary real constants.

In these models, the integration of the position equations (2.14) has been worked out taking the Cauchy data:

$$[x_a^\mu - q_a^\mu]_\Sigma = 0, \quad \Sigma \equiv \{(q,p) \in TM^{4N} \mid (P \cdot z_A) = 0\}. \tag{5.3}$$

However, since the 4-velocities $\{H_a, x_a^\mu\}$ have not constant moduli the parameters τ_a, $a = 1,\ldots, N$, along the world lines of the particles, are not the proper times.

We shall here relax the prescription (5.3) and take

$$[x_a^\mu - (\tilde{q}_a^\mu + h_a(q,p) \cdot P^\mu)]_\Sigma = 0, \tag{5.4}$$

where $h_a(q,p)$ is an arbitrary scalar function which will be

fixed by requiring (4.10).

Since we are only interested in the values of $h_a(q,p)$ on $\Sigma \equiv \{(P\cdot z_B)=0\}$, it will only depend on the variables $P^2, P\cdot X, P\cdot y_B, \tilde{X}^\mu, \tilde{z}^\mu_B, \tilde{y}^\mu_A$. Also, since Σ is translation-invariant and x^μ_a and q^μ_a change like $x^\mu_a \to x^\mu_a + \epsilon^\mu$, $q^\mu_a \to q^\mu_a + \epsilon^\mu$ under a translation ϵ^μ, we have that $h_a(q,p)$ must change like

$$h_a \to h_a + \epsilon\cdot P/P^2.$$

Hence, we can write

$$h_a = P\cdot X/P^2 + \kappa_a, \tag{5.5}$$

where κ_a is translation-invariant and, therefore, it only depends on $P^2, P\cdot y_B, \tilde{z}^\mu_C, \tilde{y}^\nu_B$.

We shall here study the case $a=1$. Then we can obtain the others $x^\mu_a(q,p)$, $a\neq 1$, by a simple permutation of indices.

The vector field $\mathbf{D}_1$ introduced in (4.5) is given in this particular case by

$$\mathbf{D}_1 = \frac{1}{\alpha P^2 + N\gamma(P\cdot y_1)} \times \left[\alpha P^2\cdot\frac{\partial}{\partial\lambda} + \sum_{B=2}^{N} N\gamma P(y_1 - y_B)\cdot\frac{\partial}{\partial\lambda_B}\right], \tag{5.6}$$

where use has been made of the new parameters $\{\lambda, \lambda_B, B=2,\ldots,N\}$, which are related to the $\{\tau_a\}_{a=1,\ldots,N}$ by[13]

$$\lambda = (1/N)\varepsilon^a\tau_a, \quad \lambda_B = \tau_1 - \tau_B \tag{5.7a}$$

and

$$\frac{\partial}{\partial\lambda} = \varepsilon^a\frac{\partial}{\partial\tau_a} = \varepsilon^a\{H_a,-\},$$
$$\frac{\partial}{\partial\lambda_B} = \frac{1}{N}\frac{\partial}{\partial\lambda} - \frac{\partial}{\partial\tau_B} = -\alpha\{(P\cdot y_B),-\}. \tag{5.7b}$$

Then, the condition (4.10) on the Cauchy data (5.4) implies that, on Σ,

$$\dot{h}_1 = [(1/-P^2)\cdot(m_1^2 + \dot{\tilde{q}}_1^2)]^{1/2}, \tag{5.8}$$

where the dot means derivation by $\mathbf{D}_1$. The sign + have been chosen in order to ensure the future orientation of $\dot{x}^\mu_1$.

Working out the necessary derivatives, we obtain

$$\dot{\tilde{q}}^\mu_1 = -\frac{N\cdot\gamma\cdot\alpha\cdot P^2}{\alpha\cdot P^2 + N\cdot\gamma\cdot(Py_1)} \times [\tilde{y}^\mu_1 + (P^2)^{-1}\cdot\varepsilon^a(P\cdot y_a)\cdot\tilde{y}^\mu_a], \tag{5.9}$$

where, as was pointed out in the detailed study of these models,[13] P^2 and $(P\cdot y_B)$ are integrals of motion and the $\tilde{y}^\mu_a$ only depend on λ. Therefore, the right-hand side of (5.9) can be expressed as

$$\tilde{\psi}^\mu = \alpha^A\cdot\tilde{y}^\mu_A, \tag{5.9'}$$

where $\alpha^A(P^2, Py_B)$ are integrals of motion.

The evolution of $\tilde{y}^\mu_A$ is described by[13]

$$\frac{d\tilde{y}^\mu_A}{d\lambda} = N\cdot\{V,\tilde{y}^\mu_A\},$$
$$\frac{d\tilde{z}^\mu_A}{d\lambda} = N\cdot\{V,\tilde{z}^\mu_A\} + N\cdot\gamma\cdot(\tilde{y}^\mu_A + \varepsilon^B\tilde{y}^\mu_B). \tag{5.10}$$

Finally, substituting the solution of (5.10) in (5.9) and introducing it in (5.8), we obtain $h_1(q,p)_{|\Sigma}$ by a simple integration.

In the following we shall study in detail the particular case of an oscillatorlike potential:

$$V = -\tfrac{1}{2}k\sum_{a<a'}(\tilde{q}_a - \tilde{q}_{a'})^2 = -\tfrac{1}{2}k\left[(N-1)\cdot\epsilon^A\tilde{z}_A^2 - 2\cdot\sum_{B<B'}(\tilde{z}_B\cdot\tilde{z}_{B'})\right]. \tag{5.11}$$

Equation (5.10) becomes

$$\frac{d\tilde{y}^\mu_A}{d\lambda} = N^2k\tilde{r}^\mu_A, \quad \frac{d\tilde{r}^\mu_A}{d\lambda} = -N\cdot\gamma\cdot\tilde{y}^\mu_A \tag{5.12}$$

with

$$\tilde{r}^\mu_A \equiv \tilde{z}^\mu_A - (1/N)\varepsilon^B\tilde{z}^\mu_B, \quad A,B = 2,\ldots,N,$$
$$\tilde{r}^\mu_1 \equiv -(1/N)\varepsilon^B\tilde{z}^\mu_B. \tag{5.13}$$

If we now define $\tilde{\varphi}^\mu \equiv \alpha^B\tilde{r}^\mu_B$, we can write

$$\frac{d\tilde{\psi}^\mu}{d\lambda} = N^2\cdot k\cdot\tilde{\varphi}^\mu, \quad \frac{d\tilde{\varphi}^\mu}{d\lambda} = -N\cdot\gamma\cdot\tilde{\psi}^\mu, \tag{5.14}$$

whose general solution is

$$\tilde{\psi}^\mu(\lambda) = \tilde{R}^\mu\cdot\sin\omega\lambda + \tilde{S}^\mu\cdot\cos\omega\lambda, \quad \omega\equiv(N^3k\gamma)^{1/2}$$
$$\tilde{\varphi}^\mu(\lambda) = \frac{\omega}{N^2\cdot k}[\tilde{R}^\mu\cos\omega\lambda - \tilde{S}^\mu\sin\omega\lambda], \tag{5.15}$$

where $\tilde{R}^\mu$ and $\tilde{S}^\mu$ depend on the initial conditions.

Substituting (5.15) into (5.8) and (5.9), it yields

$$\dot{h}_1 = (1/\sqrt{-P^2})[m_1^2 + \tilde{R}^2\cdot\sin^2\omega\lambda + \tilde{S}^2\cdot\cos^2\omega\lambda + 2(\tilde{R}\cdot\tilde{S})\cos\omega\lambda\cdot\sin\omega\lambda]. \tag{5.16}$$

Since P^2 and $(P\cdot y_B)$ do not depend either on λ nor on λ_B and $\tilde{y}^\mu_A$ only depends on λ, we can write, taking (5.9') into account,

$$\tilde{\psi}^\mu(\lambda) = \tilde{\psi}^\mu(b\sigma), \tag{5.17}$$

where b is a constant of motion which we derive in Appendix A and σ is the parameter along the integral curves of $\mathbf{D}_1$:

$$\mathbf{D}_1 = \frac{d}{d\sigma}.$$

Hence, introducing this in (5.16) yields

$$\frac{dh_1}{d\sigma} = \frac{1}{\sqrt{-P^2}}\cdot[m_1^2 + \tilde{R}^2\cdot\sin^2(\omega b\sigma) + \tilde{S}^2\cdot\cos^2(\omega b\sigma) + 2(\tilde{R}\cdot\tilde{S})\sin(\omega b\sigma)\cdot\cos(\omega b\sigma)], \tag{5.18}$$

which, after some manipulations, leads to

$$\frac{dh_1}{d\sigma} = \sqrt{A^2 - B^2\sin^2(\omega b\sigma - \theta)}, \tag{5.19}$$

with

$$A^2 = (1/-P^2)[m_1^2 + \tfrac{1}{2}(\tilde{R}^2 + \tilde{S}^2)] + \tfrac{1}{2}B^2 \geqslant 0, \tag{5.20a}$$

$$B^2 = \frac{2}{-P^2}\sqrt{(\tilde{R}\tilde{S})^2 + \tfrac{1}{4}(\tilde{R}^2 - \tilde{S}^2)} \geqslant 0, \tag{5.20b}$$

$$\tan 2\theta = 2(\tilde{R}\cdot\tilde{S})/(\tilde{S}^2 - \tilde{R}^2). \tag{5.20c}$$

Since $B^2 \leqslant A^2$, the solution of (5.19) can be expressed in

terms of an elliptic integral of the first kind[14]:

$$h_1(\sigma) = \frac{A}{\omega b} \cdot \mathrm{E}\left(\frac{B}{A}, \omega b\sigma - \theta\right) + \Gamma, \tag{5.21}$$

where Γ satisfies $\mathbf{D}_1\Gamma = 0$.

We now have to explain what does this result mean. We are looking for a function $h_1(P^2, PX, Py_B, \tilde{X}^\mu, \tilde{y}^\nu_B, \tilde{z}^\rho_c)$ on Σ satisfying the ordinary differential equation (5.19). A unique solution of it is determined giving a suitable set of initial conditions, i.e., a $7N$-submanifold Σ^* of Σ such that it is not a characteristic of $\mathbf{D}_1$, and the prescribed value of h_1 on Σ^*. The situation is visualized in Fig. 1.

Equation (5.21) gives us $h_1(Q)$ in terms of Q_0 and σ [as is obvious A^2, B^2, θ are given in terms of the variables in Q_0, the initial point for the trajectory of (5.14)]. However, we want our final result in terms of the coordinates of Q.

In Appendix B the expressions of A^2 and B^2 in terms of constants of motion are given (B4). We also give there an expression (B7) for the phase θ in terms of Q. Substituting (B7) into (5.21), it yields

$$h_1(Q) = \frac{A}{\omega b} \mathrm{E}\left(\frac{B}{A}, -\rho\right) + \Gamma, \tag{5.22}$$

with

$$\rho \equiv \tfrac{1}{2} \arctan \frac{2\sqrt{N\gamma}(\tilde{\varphi} \cdot \tilde{\psi})}{\tilde{\psi}^2 - N\gamma\tilde{\varphi}^2} \tag{B7}$$

where

$$\tilde{\varphi}^\mu = -\frac{N\gamma\alpha P^2}{\alpha P^2 + N\gamma(P \cdot y_1)}\left[\tilde{r}^\mu_1 + \frac{1}{P^2}\varepsilon^a(Py_a)\tilde{r}^\mu_a\right], \tag{5.23a}$$

$$\tilde{\psi}^\mu = -\frac{N\gamma\alpha P^2}{\alpha P^2 + N\gamma(Py_1)}\left[\tilde{y}^\mu_1 + \frac{1}{P^2}\varepsilon^a(Py_a) \cdot \tilde{y}^\mu_a\right]. \tag{5.23b}$$

The integral of motion Γ is determined by fixing the initial conditions. If, for instance, we take

$$\Sigma^* \equiv \{P \cdot X = 0\}, \quad h_{1|\Sigma^*} = 0, \tag{5.24}$$

we have, taking (5.22) into account, that

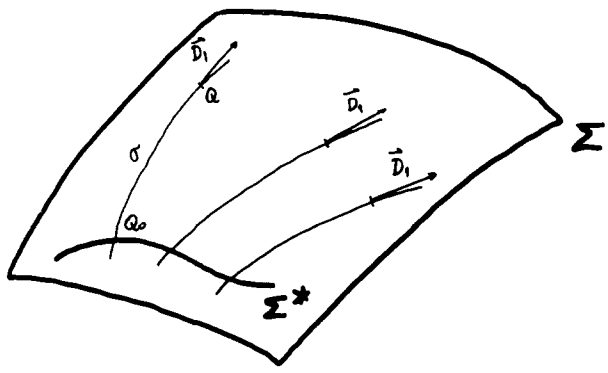
$$\Gamma = -\frac{A}{\omega b} \cdot \mathrm{E}\left(\frac{B}{A}, -\rho - \omega\lambda\right) \tag{5.25}$$

and

$$h_1(Q) = \frac{A}{\omega b}\left[\mathrm{E}\left(\frac{B}{A}, -\rho\right) - \mathrm{E}\left(\frac{B}{A}, -\rho - \omega\lambda\right)\right], \tag{5.26}$$

where λ is obtained from (A2):

$$\lambda = [2N\beta P^2 - (1/P^2)N\gamma\varepsilon^a(Py_a)^2]^{-1} \cdot (P \cdot X). \tag{5.27}$$



$Q \equiv (P^2, P_{y_B}; PX, \tilde{X}^\mu, \tilde{y}^\rho_B, \tilde{z}^\nu_c)$

$Q_0 \equiv (P^2, P_{y_B}; PX_0, \tilde{X}^\mu_0, \tilde{y}^\rho_{B_0}, \tilde{z}^\nu_{c_0})$

FIG. 1. σ = variation of the parameter between Q_0 and Q.

VI. CONCLUSIONS

This paper is a study of the equivalence between two different versions of predictive relativistic mechanics.

An N-particle system is described in the "*a priori*" Hamiltonian formalism[6] by N Poincaré-invariant functions H_a such that their mutual Poisson brackets vanish. The physical position are related to the canonical variables through the individuality conditions (2.14), which admit a wide class of solutions depending on the prescribed Cauchy data. This is the worst feature of this formalism because fixing only the Hamiltonian has not any physical meaning.

Hence, in this framework the physical model must be established in two steps: (i) the Hamiltonian functions and (ii) the Cauchy data for the position equations. Then, integrating these equations we can associate a Poincaré-invariant abelian system (PIAS) to the given Hamiltonian model (HPS).

In general, the PIAS obtained in this way does not fulfill the condition of constancy of the velocities' moduli, and therefore it is not a Poincaré-invariant predictive system (PIPS). We have shown in Sec. III that by means of a reparametrization a PIPS can be associated with a given PIAS. This reparametrization is not unique. It involves a change of coordinates of the form:

$$(x^\mu_a, \dot{x}^\nu_b) \to (x^\mu_a, \pi^\nu_b), \quad \pi^\nu_b \equiv \beta_b(x, \dot{x}) \cdot \dot{x}^\nu_b.$$

We have seen in Sec. III what conditions must be fulfilled by the proportionality factors $\beta_b(x, \dot{x})$ in order to preserve the commutativity conditions (2.3).

If a given reparametrization transforms a PIAS into a PIPS, the new variables (x^μ_a, π^ν_b) must verify: $\pi^\mu_a \pi_{a\mu} = -m_a^2$, $a = 1,\ldots, N$, where $\{m_a^2,\ a = 1 \cdots N\}$ are N arbitrary functionally independent constants of motion. Therefore, a given PIAS is reparametrized into a PIPS by as much reparametrizations as possible choices of the set $m_a^2,\ a = 1 \cdots N\}$. The most natural choice of these constants are the N Hamiltonians.

However, the Lie derivatives of the symplectic form Ω with respect to the proper time along the individual world lines (i.e., the derivatives $\mathscr{L}(\mathbf{H}_a)\Omega$ being $\mathbf{H}_a$, the tangent vector fields of the PIPS, $a = 1 \cdots N$) do not vanish and hence the PIPS is not Hamiltonian, with respect to the symplectic form Ω.

In Sec. IV we have considered the special situation in which the Cauchy data for the position equations lead directly to a PIPS and no reparametrization is necessary. In this case the PIPS is Hamiltonian. In Sec. V we have applied these results to the family of models presented in Ref. 6.

APPENDIX A

The evolution of the functions (PX), (Pz_A), $A = 2,..., N$ is given by[13]

$$(PX) = (PX)_0 + 2N\cdot\beta\cdot P^2\lambda + \alpha\varepsilon^A(Py_A)\cdot\lambda_A,$$
$$(Pz_A) = (Pz_A)_0 + N\cdot\gamma[(Py_A) + \varepsilon^B(Py_B)]\cdot\lambda + \alpha P^2\lambda_B. \tag{A1}$$

If we constrain the evolution to remain on $\Sigma \equiv \{Pz_A = 0, A = 2,..., N\}$, we have

$$\lambda_B = -\frac{N\gamma}{\alpha P^2}[(Py_B) + \varepsilon^A(Py_A)]\cdot\lambda, \tag{A2}$$
$$(PX)_\Sigma = (PX)_0 + \lambda\cdot\left[2N\cdot\beta\cdot P^2 - \frac{1}{P^2}\cdot N\cdot\gamma\cdot\varepsilon^a(Py_a)^2\right].$$

Therefore, we can write

$$\lambda = \frac{(PX)_\Sigma - (PX)_0}{2N\cdot\beta\cdot P^2 - (1/P^2)\cdot N\cdot\gamma\cdot\varepsilon^a(Py_a)^2}. \tag{A3}$$

We can also relate the parameters λ and σ

$$\frac{d(PX)_\Sigma}{d\lambda} = 2N\beta P^2 - \frac{1}{P^2}\cdot N\cdot\gamma\cdot\varepsilon^a(Py_a)^2,$$
$$\frac{d(PX)_\Sigma}{d\sigma} = \mathbf{D}_1(PX)_\Sigma = \frac{\alpha P^2}{\alpha P^2 + N\cdot\gamma\cdot(Py_1)} \times\left[2N\beta P^2 - \frac{1}{P^2}N\cdot\gamma\cdot\varepsilon^a(Py_a)^2\right]; \tag{A4}$$

hence

$$b = \frac{d\lambda}{d\sigma} = \frac{\alpha P^2}{\alpha P^2 + N\cdot\gamma\cdot(Py_1)}. \tag{A5}$$

APPENDIX B

We shall here derive the expressions of A^2, B^2, θ in terms of integrals of motion.

As is well known,[15] the system (5.14) has the tensor integral of motion:

$$\tilde{H}^{\mu\nu} \equiv N^2\cdot k\cdot\tilde{\varphi}^\mu\cdot\tilde{\varphi}^\nu + N\cdot\gamma\cdot\tilde{\psi}^\mu\cdot\tilde{\psi}^\nu. \tag{B1}$$

Substituting (5.15) into (B1), it yields

$$\tilde{H}^{\mu\nu} = N\cdot\gamma\cdot(\tilde{R}^\mu\cdot\tilde{R}^\nu + \tilde{S}^\mu\cdot\tilde{S}^\nu). \tag{B2}$$

We then define

$$\tilde{M}^{\mu\nu} \equiv \frac{1}{N\gamma}\tilde{H}^{\mu\nu} = \frac{N\cdot k}{\gamma}\tilde{\varphi}^\mu\cdot\tilde{\varphi}^\nu + \tilde{\psi}^\mu\cdot\tilde{\psi}^\nu, \tag{B3}$$

which is also a tensor integral of motion.

Using (B3), we can write

$$B^2 = [2/(-P^2)]\cdot\left[-\tfrac{1}{4}(\tilde{M}^\mu_\mu)^2 + \tfrac{1}{2}\tilde{M}^{\mu\nu}\tilde{M}_{\mu\nu}\right]^{1/2}$$
$$= [2/(-P^2)]\{\tfrac{1}{4}[(Nk/\gamma)\tilde{\varphi}^2 - \tilde{\psi}^2]^2 + (Nk/\gamma)(\tilde{\varphi}\tilde{\psi})^2\}^{1/2}, \tag{B4}$$

$$A^2 = [1/(-P^2)](m_1^2 + \tfrac{1}{2}\tilde{M}^\mu_\mu) + \tfrac{1}{2}B^2$$
$$= [1/(-P^2)]\left[m_1^2 + (Nk/2\gamma)\tilde{\varphi}^2 + \tfrac{1}{2}\tilde{\psi}^2\right] + \tfrac{1}{2}B^2,$$

where $\tilde{\varphi}^\mu$ and $\tilde{\psi}^\nu$ are given in terms of $\tilde{y}^\mu_A$ and $\tilde{z}^\nu_B$ by (5.9), (5.9′), and (5.13).

From (5.15) we have

$$\tilde{R}^\mu = \tilde{\psi}^\mu\cdot\sin\omega\lambda + \sqrt{N\gamma}\cdot\tilde{\varphi}^\mu\cdot\cos\omega\lambda,$$
$$\tilde{S}^\mu = \tilde{\psi}^\mu\cdot\cos\omega\lambda - \sqrt{N\gamma}\cdot\tilde{\varphi}^\mu\cdot\sin\omega\lambda, \tag{B5}$$

and, substituting this in (5.20c), it yields

$$\tan 2\theta = \frac{(\tilde{\psi}^2 - N\gamma\tilde{\varphi}^2)\cdot\sin(2\omega\lambda) + 2\sqrt{N\gamma}(\tilde{\varphi}\cdot\tilde{\psi})\cdot\cos(2\omega\lambda)}{(\tilde{\psi}^2 - N\gamma\tilde{\varphi}^2)\cdot\cos(2\omega\lambda) - 2\sqrt{N\gamma}(\tilde{\varphi}\tilde{\psi})\cdot\sin(2\omega\lambda)}, \tag{B6}$$

or, equivalently,

$$\theta = \omega\lambda + \rho, \quad \rho \equiv \tfrac{1}{2}\arctan[2\sqrt{N\gamma}(\tilde{\varphi}\tilde{\psi})/(\tilde{\psi}^2 - N\gamma\cdot\tilde{\varphi}^2)]. \tag{B7}$$

[1] J. A. Wheeler and R. P. Feynman, Rev. Mod. Phys. **17**, 157 (1945); J. A. Wheeler and R. P. Feynman, Rev. Mod. Phys. **21**, 425 (1949); H. Van Dam and E. P. Wigner, Phys. Rev. **192**, 838 (1966).
[2] P. A. M. Dirac, Rev. Mod. Phys. **21**, 392 (1949); B. Bakampian and L. H. Thomas, Phys. Rev. **92**, 1300 (1953); L. L. Foldy, Phys. Rev. **122**, 275 (1961).
[3] D. G. Currie, T. F. Jordan, and E. C. G. Sudarshan, Rev. Mod. Phys. **35**, 350 (1963).
[4] D. G. Currie, Phys. Rev. **142**, 817 (1966); R. N. Hill, J. Math. Phys. **8**, 201 (1967); E. H. Kerner, J. Math. Phys. **9**, 222 (1968); T. F. Jordan, Phys. Rev. **166**, 1309 (1968); P. Droz Vincent, Phys. Scripta **2**, 129 (1970); L. Bel, Ann. Inst. H. Poincaré **14**, 189 (1971); R. Arens, Arch. Rat. Mech. Anal. **47**, 255 (1972).
[5] P. Droz Vincent, Phys. Scripta **2**, 129 (1970).
[6] P. Droz Vincent, Rep. Math. Phys. **8**, 79 (1975); P. Droz Vincent, Ann. Inst. H. Poincaré **27**, 407 (1977); V. Iranzo, J. Llosa, F. Marqués, and A. Molina, Ann. Inst. H. Poincaré **35**, 1 (1981).
[7] L. Bel and J. Martin, Ann. Inst. H. Poincaré **22**, 173 (1975); L. Bel and X. Fustero, Ann. Inst. H. Poincaré A **25**, 411 (1976).
[8] P. Droz Vincent, Phys. Scripta **2**, 129 (1970).
[9] C. Godbillon, *Geometrie différentielle et mécanique analytique* (Hermann, Paris, 1969).
[10] J. Llosa, doctoral thesis, Universitat de Barcelona, 1978.
[11] C. Bona, doctoral thesis, Universitat Autònoma de Barcelona, 1980.
[12] M. Spivak, *Calculus on Manifolds* (Benjamin, New York, 1965).
[13] V. Iranzo, J. Llosa, F. Marqués, and A. Molina, Ann. Inst. H. Poincaré **35**, 1 (1981).
[14] M. R. Spiegel, *Handbook of Mathematical Formulas and Tables* (McGraw-Hill, New York, 1970).
[15] J. P. Elliott, Proc. R. Soc. London Ser. A **245**, 128 (1958); P. Droz-Vincent, Ann. Inst. H. Poincaré A **27**, 407 (1977).

A Hamiltonian–Lagrangian system of vortex boundary curves

D. J. Hebert
Department of Mathematics and Statistics, The University of Pittsburgh, Pittsburgh, Pennsylvania 15261

(Received 8 March 1982; accepted for publication 23 April 1982)

The equations of motion for the boundary of a region of constant vorticity in an inviscid, incompressible fluid are expressed as an infinite dimensional Hamiltonian system and as a Lagrangian system based upon a variational principle. In this formulation the m-fold symmetric, stationary, steady, rotating vortices are analogous to the travelling cnoidal waves in the well-known KdV system.

PACS numbers: 47.30. + s, 47.10. + g

1. INTRODUCTION

The purpose of this paper is to introduce an infinite dimensional Lagrangian and Hamiltonian system which describes the motion of a region of constant vorticity in an inviscid, incompressible fluid flow in the plane. The boundary of the region is considered as a point on a manifold, and a tangent vector at this point is determined by a divergence-free vector field whose flow moves the boundary.

The fact that the equations of motion for an ideal fluid in the plane can be described as a Hamiltonian system is well known (cf. Arnold,[1] Marsden,[2] Abraham and Marsden,[3] and Olver[4]). The present formulation differs from previous ones in that the restriction to a vorticity function which is constant on its support allows all quantities of interest to be expressed in terms of boundary integrals. Consequently, the partial differential equation of motion, in an appropriate local coordinate system, involves only one space variable, the boundary parameter. This results in direct analogies with well-studied Hamiltonian systems such as the one associated with the Korteweg–de Vries (KdV) equation.

An important class of periodic trajectories in the study of these vortex systems consists of the m-fold symmetric rotating vortices (V-states) discovered numerically by Deem and Zabusky,[5] and made precise analytically by Burbea.[6] These m-fold symmetric vortices are bifurcations of the trivial circular solution to a nonlinear travelling wave equation of boundary motion. The bifurcations occur exactly at the angular velocities $\Omega = \frac{1}{2}\omega_0(m-1)/m$ of perturbations of the unit circle of the form $\epsilon \cos m\theta$ discussed by Kelvin in 1880 (see Lamb, Ref. 7, p. 231). The elliptical vortex of Kirchoff (see Ref. 7, p. 232) is the special case $m = 2$. Further studies on the stability of these m-fold symmetric solutions (see Burbea and Landau[8]) and on their bifurcation properties (cf. McKee[9]) have shown that classical results known previously only for the circle and the ellipse ($m = 2$) have counterparts for all the m-fold symmetric regions ($m > 2$). In the analogy with the KdV system the m-fold symmetric rotating vortex plays the same role as the travelling cnoidal wave solution.

2. REGIONS OF CONSTANT VORTICITY IN AN IDEAL FLUID

The Euler equations for an inviscid, incompressible fluid in the plane may be written in vorticity-stream function form as

$$\partial_t + \partial_x\omega\partial_y\psi - \partial_x\psi\partial_y\omega = 0, \tag{1}$$

$$\Delta\psi = \omega, \tag{2}$$

where ψ is the stream function and ω is the vorticity. The velocity field of the classical fluid is the divergence-free vector field

$$\text{rot}\ \psi = (\partial_y\psi, -\partial_x\psi).$$

If this velocity field is continuously differentiable then the position of a "fluid particle" at time t is the image of an initial position under the flow of the vector field. Equation (1) says that the vorticity is convected with the flow.

The present paper is restricted to the study of a fluid determined by a vorticity function which has a constant value ω_0 on its support D, a simply connected region bounded by a Jordan curve C. It is assumed that the stream function is determined by Eq. (2) and that (1) holds in the sense of distributions. The stream function is harmonic, therefore real-analytic, outside of C and continuously differentiable in a neighborhood of C. An equation equivalent to (1) states that the boundary C is convected with the flow; thus

$$n\cdot\partial_t z = n\cdot\text{rot}\ \psi,$$

where $z(\cdot,t)$ is a parametrization of C at time t and n is the unit normal vector on the boundary. In terms of the parametrization $z(\cdot,t)$ this equation may be rewritten as

$$\partial_t x\,\partial_\theta y - \partial_t y\,\partial_\theta x = \partial_\theta\psi \tag{3}$$

or, in complex notation

$$\text{Im}\partial_t\bar{z}\,\partial_\theta z = \partial_\theta\psi.$$

The fact that the velocity field is divergence-free implies that the flow is area preserving; hence the area of the region of nonzero vorticity remains constant in time.

3. THE STREAM FUNCTION AND ENERGY AS BOUNDARY INTEGRALS

When the vorticity is constant on its support, the important quantities of hydrodynamics may be expressed in terms of the boundary only. The area of the region D is determined by Green's theorem to be

$$A = \tfrac{1}{2}\int_C x\,dy - y\,dx.$$

Suppose that the boundary C of the region D is parametrized by a function $z(\cdot)$ defined on the unit circle Γ. It is convenient

 0022-2488/83/061672-07$02.50 © 1983 American Institute of Physics

to introduce an "area kernel" determined by the parametrization $z(\cdot)$, namely,

$$a_z(\alpha,\beta) = \tfrac{1}{2}\langle (x(\alpha) - x(\beta))(dy(\alpha) - dy(\beta)) - (y(\alpha) - y(\beta))(dx(\alpha) - dx(\beta))\rangle.$$

The area A may be written then as

$$A = \int_\Gamma a_z(\alpha,\beta)d\alpha.$$

independent of β. The stream function ψ, determined by Eq. (2), is computed as the integral

$$\psi(z) = \frac{\omega_0}{2\pi}\int\int_D \ln|z-\zeta|\,dA(\zeta),$$
$$z = x + iy, \quad \zeta = \xi + i\eta.$$

Green's theorem applied to the vector field whose components are

$$H_1 = (\xi - x)\ln|z-\zeta|^2$$

and

$$H_2 = (\eta - y)\ln|z-\zeta|^2$$

yields a formula for the stream function as a boundary integral:

$$\psi(z) = \frac{\omega_0}{8\pi}\int_c ((\xi - x)\,d\eta - (\eta - y)\,d\xi)(\ln|z-\zeta|^2 - 1).$$

In terms of a parametrization $z(\cdot)$ of the boundary, this formula evaluated on the boundary will be written in the form

$$\psi(z(\alpha)) = \frac{\omega_0}{4\pi}\int_\Gamma a_z(\alpha,\beta)l_z(\alpha,\beta)d\beta - \frac{\omega_0}{4\pi}A, \tag{4}$$

where l_z is the logarithm kernel determined by the parametrization

$$l_z(\alpha,\beta) = \ln|z(\alpha) - z(\beta)|^2. \tag{5}$$

The energy of a fluid is usually defined to be the total kinetic energy

$$E = \frac{1}{2}\int\int((\partial_x\psi)^2 + (\partial_y\psi)^2)\,dA.$$

However, in the study of regions of constant vorticity, a more appropriate quantity is

$$V = -\frac{1}{2}\int\int_D \omega_0\psi\,dA,$$

a classical invariant of the fluid motion. This quantity is analogous to the potential energy in electrostatic theory for charges of density ω_0 in D and electrostatic potential ψ, or, in magnetostatics, the total energy of a steady current of density ω_0 in cylinder of cross section D, where ψ is the vector potential associated with the magnetic field induced by the current. (For a discussion of these elementary theories see Ref. 10.) In the formulation of a Hamiltonian or Lagrangian system of vortex boundary curves, V will be the potential and its negative will be the Hamiltonian. The potential V may be expressed as a boundary integral by applying Green's theorem to the vector field whose components are

$$H_1 = \frac{1}{4}\int_c((\xi - x)^2\,d\eta - (\eta - y)^2\,d\eta - 2(\xi - x)(\eta - y)\,d\xi) \times(\ln R^2 - \tfrac{3}{2}),$$

$$H_2 = \frac{1}{4}\int_c((\xi - x)^2\,d\xi - (\eta - y)^2\,d\xi + 2(\xi - x)(\eta - y)\,d\eta) \times(\ln R^2 - \tfrac{3}{2}),$$

where

$$R^2 = (x-\xi)^2 + (y-\eta)^2.$$

The result is

$$V = -(\omega_0^2)(1/16\pi)\int\int_{C\times C} a_z^2(\alpha,\beta)(l_z(\alpha,\beta) - \tfrac{3}{2})\,d\alpha\,d\beta. \tag{6}$$

4. THE MANIFOLD AND TANGENT SPACE

To formulate the concepts of Hamiltonian mechanics it is appropriate to define a manifold, a tangent space, and a sympletic form. To produce analogies of classical equations it will be convenient, in addition, to provide local coordinate systems. A natural setting in which vortex boundary curves may be considered as position vectors is the manifold P of all Jordan curves of interior area π. The image of a member of P under the flow of a divergence-free vector field remains in P. With this in mind one might choose, as a tangent vector at a point on this manifold, a divergence-free vector field (or some generalization thereof) which would play the role of the derivative at a point on a curve in the manifold. In this case, two vector fields F and G would be considered equivalent if they produced the same infinitesimal motion, i.e., if

$$F\cdot n = G\cdot n$$

on the boundary. In such a general setting, however, one can not be certain that singularities in the second derivative of the analytically continued interior or exterior stream function do not occur on the boundary itself. This situation causes difficulties concerning uniqueness of particle motion on the boundary. In the present paper, all concepts will be formulated in terms of a dense subset M of P (when P is embedded in an appropriate function space) and a linear space T which is a dense subspace of the Hilbert or Banach space which would naturally occur as the tangent space at a point. This will allow analytic extension of the interior or exterior stream function across the boundary, thus providing a smooth vector field whose flow gives a unique boundary motion. Questions of existence and uniqueness of trajectories will not be addressed; results will be obtained under the assumption that the relevant classical time-dependent solutions are known to exist.

Let M be the set of all analytic Jordan curves with interior area π and with the origin as an interior point. By the Riemann mapping theorem, each element C of M is the image of the unit circle Γ under a one-to-one analytic mapping z of the unit disc such that $z(0) = 0$ and $z'(0) > 0$. The function z may be represented as a power series

$$z(w) = \sum a_n w^n$$

whose radius of convergence is greater than 1, and the area restriction implies that

$$\pi = \sum n|a_n|^2.$$

The set M is thus identified with dense subset of manifold in

the Hilbert space $H^2 = H^2(\Gamma)$ of holomorphic functions on the unit disc with the inner product

$$\langle f,g\rangle = \int_\Gamma f\bar{g}.$$

For each C in M with representative z in H^2, define T_z to be the set of all real-valued analytic functions on C of the form $\partial_s v$. An element of T_z will be called a tangent vector at the point z. The terminology is justified by the following remarks. If $v(\cdot,t)$ is a real-valued analytic function defined on a neighborhood of C for small t, then the image C_t under the flow of the divergence-free vector field rot v is also a member of M. Two t-dependent vector fields rot $v(\cdot,t)$ and rot $w(\cdot,t)$ generate the same curve C_t if and only if they have the same normal component on C_t, i.e., if

$$\partial_s v(\cdot,t) = \partial_s w(\cdot,t). \tag{7}$$

If rot v and rot w generate different curves then the two curves are tangent at the point C_0 provided that (7) holds when $t = 0$.

In this paper, most results can be formulated in terms of the above expression for a tangent vector; however, in certain calculations and in certain other developments of the theory it is necessary to use other related functions to take the place of $\partial_s v$. Suppose that C_t represents a differentiable curve in M with representative $z(\cdot,t)$ in H^2 such that the derivative

$$\partial_t z(\cdot,0) = h(\cdot) = h_1 + ih_2$$

is an analytic member of H^2. The constancy of area implies that

$$0 = \int_c h_1\,dy - h_2\,dx = \operatorname{Im}\int_c \bar{h}\,dz \tag{8}$$

which means that the path integral on C_t of the normal component h_n of the analytic vector field $h\circ z^{-1}$ is zero. Thus, there is a real-valued analytic function v defined on the unit circle such that

$$\partial_\theta v(\theta) = h_1\partial_\theta y - h_2\partial_\theta x$$

on C_t. Conversely, by solution of the Neumann problem, each real-valued analytic v yields a corresponding analytic function h in H^2. It is worth noting also that the function v may be extended as the imaginary part of an analytic function g on a neighborhood of C such that

$$h = \partial_z\bar{g} = \partial_y v - i\partial_x v$$

on C.

In summary, the tangent space T_z [for a point C in M parametrized by $z(\cdot)$] may be identified with the linear manifold in H^2 consisting of functions which satisfy (8), with the set of real-valued analytic functions u on C such that

$$\int_C u(s)\,ds = 0$$

(which means that $u = \partial_s v$ for some v), and finally with the set of complex-valued functions of the form $\partial_z\bar{g}$, where g is analytic on a neighborhood of the region bounded by C.

5. REAL LOCAL COORDINATES

Suppose that C_0 is an element of M parametrized by z_0 in H^2. Let U be the set of all C in M such that $z_0^{-1}(C)$ is star-shaped at 0. This means that for C in U there is a real-valued function r such that C may be parametrized as $z_0(r(\theta)\exp(i\theta))$, $0\leqslant\theta\leqslant 2\pi$. The function u defined by

$$u(\theta) = \int_0^{\rho(\theta)} |\partial_\theta z_0(e^\zeta)|^2 d\rho, \tag{9}$$

where

$$\zeta = \rho + i\theta \quad\text{and}\quad \rho(\theta) = \ln r(\theta)$$

represents a (squared) distance between C and C_0 along a curvilinear coordinate curve normal to C_0. The function u will be called a local coordinate representation of C.

Let T be the set of all real-valued analytic functions u on Γ such that

$$\int_\Gamma u(\theta)\,d\theta = 0.$$

Suppose that v is a real-valued analytic function on Γ, so that $\partial_\theta v$ is in T. The following argument will show that for all t in an interval containing 0, the function $t\,\partial_\theta v$ is the coordinate function for some C in U. If the function v is extended along rays in the w plane so that $v(z_0(w)) = v(\theta)$ when $w = r\exp(i\theta)$, then

$$\text{rot } v = ((\partial_\theta v)/|\partial_\theta z_0|^2)(\partial_\theta y_0 - i\partial_\theta x_0).$$

If $z(\cdot,t)$ satisfies

$$\partial_t z(\theta,t) = \text{rot } v(z(\theta,t))$$

and

$$z(\theta,0) = z_0(\theta),$$

then since the direction of rot v is that of constant θ in the curvilinear coordinate system induced by the map z_0,

$$z(\theta,t) = z_0(\exp(\zeta(\theta,t))), \tag{10}$$

where

$$\zeta(\theta,t) = \rho(\theta,t) + i\theta.$$

A simple calculation shows that

$$\partial_t\,\rho(\theta,t)\,|\partial_\theta z(\theta,t)|^2 = \partial_\theta v(z(\theta,t)), \tag{11}$$

and

$$u(\theta,t) = \int_0^t |\partial_\theta z_0|^2\partial_t\rho\,dt = t\,\partial_\theta v$$

is in fact the local coordinate representation for $z(\cdot,t)$. Thus, the function which maps C to u, defined by Eq. (9), plays the role of a local chart, coordinatizing the set U by a subset of T containing 0. If z_0 is the identity map, then the formula (9) becomes

$$u(\theta) = \int_0^{r(\theta)} |r^2|\,\frac{1}{r}\,dr = \frac{1}{2}(r^2(\theta) - 1),$$

a polar coordinate representation for z.

Suppose that X is a vector field on M, i.e., a mapping which assigns to each point C in M a tangent vector $X(z) = \partial_\theta v_z$, where z is a parametrization of C belonging to some function space. A trajectory of the flow of X is a curve $z(\cdot,t)$ in M such that

$$\partial_t x\,\partial_\theta y - \partial_t y\,\partial_\theta x = \partial_\theta v_z. \tag{12}$$

If the parametrization of this trajectory has a representation as

$$z(\theta,t) = z_0(r(\theta,t)\exp i\theta)$$
$$= z_0(\rho(\theta,t) + i\theta),$$

and if $u(\theta,t)$ is defined by

$$u(\theta,t) = \int_0^t \partial_\theta v(z(\theta,\tau))\, d\tau \quad (v = v_z) \tag{13}$$

then Eq. (11) holds and u is a local coordinate representation for the trajectory. If instead $z(\theta,t)$ satisfies

$$\partial_t z = \text{rot } v_z,$$

and u is defined by Eq. (13), then the same equations are obtained for a different local coordinate representation of the trajectory. This will be called the representation of $z(\cdot,t)$ in local-flow coordinates for the vector field X.

6. DERIVATIVES OF FUNCTIONS ON M

Suppose that F is a mapping of M into a linear space B. The derivative of F at the point z, if it exists, is the linear mapping $d_z F$ of T into B defined by the formula

$$d_z F(\partial_\theta v) = D_\epsilon|_{\epsilon=0} F(z(\cdot,\epsilon)), \tag{14}$$

where $z(\cdot,\epsilon)$ is a curve in M such that $z(\cdot,0) = z(\cdot)$ and

$$\partial_\epsilon x\, \partial_\theta y - \partial_\epsilon y\, \partial_\theta x|_{t=0} = \partial_\theta v.$$

The second derivative of F at z is a bilinear mapping

$$\partial_z{}^2 F(\partial_\theta v, \partial_\theta w) = D_\epsilon|_{\epsilon=0} d_{z(\cdot,\epsilon)} F(\partial_\theta v)$$

defined on $T \times T$. If F is a real-valued function on M then the formula (14) gives the directional derivative at the point z in the direction of the tangent vector, and $d_z F$ is called the gradient of F at z. The functions and mappings of interest in this paper will be Frechet differentiable with the derivative representable as an integral operator, a continuous linear mapping of T into a Hilbert or Banach space B. When the derivative of F is

$$d_z F(\partial_\theta v) = \int_\Gamma f(z(\theta))\, \partial_\theta v(\theta)\, d\theta \tag{15}$$

the function f (a contangent vector in the terminology of Sec. 4 will itself be called the derivative or gradient of F.

In local coordinates there is a notation which suggests the concept of partial derivative: If $u(\cdot,\epsilon)$ is a local coordinate representation of $z(\cdot,\epsilon)$ in the sense of Sec. 5 so that

$$\partial_\epsilon u(\theta,\epsilon) = \partial_\theta v(\theta,\epsilon),$$

where $\partial_\theta v(\cdot,\epsilon)$ is the tangent vector to the curve $z(\cdot,\epsilon)$, and $u_\theta = u(\theta,0)$, then Eq. (15) will be written simply as

$$\partial_{u_\theta} F(z) = f(\theta)$$

or

$$\partial_u F(z) = f.$$

The second directional derivative in the given direction $\partial_\theta v$ is then

$$d_z^2 F(\partial_\theta v, \partial_\theta v).$$

The first and second directional derivatives of the potential V are essential in the coming sections. The computation of the first derivative is straightforward but lengthy, requiring integration by parts and the use of the symmetry of the double integral in Eq. (6). The result is

$$d_z V(\partial_\theta v) = -\int_\Gamma \psi(z(\theta))\partial_\theta v(\theta)\, d\theta. \tag{16}$$

The second derivative requires less work to compute but the result is more complicated, namely

$$d_z^2 V(\partial_\theta v, \partial_\theta w) = -\omega_0 \int_\Gamma \psi\partial(v,w) - \omega_0 \int_\Gamma \partial_\theta v\, d_z\psi(\partial_\theta w), \tag{17}$$

where

$$d_z\psi(\partial_\theta w) = \frac{\omega_0}{4\pi}\int_\Gamma (\partial_\theta w) l_z(\theta,\cdot)\, d\theta + \partial_\epsilon y\, \partial_y\psi + \partial_\epsilon x\, \partial_x\psi. \tag{18}$$

The first term on the right side of the preceding equation is an integral operator with particularly nice properties; the second term takes on different forms in different coordinate systems. In the local coordinate representation of Sec. 5 this term takes on the form $\partial_\theta w\, \partial_n \psi$ where $\partial_n \psi$ is the normal component of the stream function, or equivalently the tangential component of velocity. Let $L = L_z$ denote the linear operator defined by the first term; indeed let

$$L_z f(\theta) = \frac{1}{4\pi}\int_\Gamma f(\alpha) l_z(\alpha,\theta)\, d\alpha.$$

Since the logarithm kernel is symmetric and square integrable, the operator L is a Hilbert–Schmidt operator, which implies that L is bounded and compact and that there is an infinite sequence of eigenvalues λ_n^z tending to zero and an orthonormal basis of eigenfunctions such that

$$l_z(\alpha,\beta) = \sum \lambda_n^z f_n^z(\alpha) f_n^z(\beta).$$

This operator may be used in the bifurcation analysis which produces the m-fold symmetric vortices branching from the unit circle (see Sec. 8). A nonsymmetric version of this operator has been used by Burbea, cf. Ref. 6 in his extensive analysis of steady rotating vortices. The operator may also be seen as an integrated form of the skew-symmetric operator used by Mckee[9] in the bifurcation analysis of the same situation. In addition to the above properties the present form of the operator has the advantage of clarifying the Hamiltonian structure of the problem. Equation (16) states that the stream function is the gradient of the potential V. The modern study of infinite dimensional Hamiltonian systems such as the Korteweg–de Vries (KdV) system suggests immediately a Hamiltonian vector field. The Hamiltonian H is defined to be $-V$ and a Hamiltonian vector field is obtained by applying the skew-symmetric derivative operator ∂_θ to the gradient of H; thus the Hamiltonian vector field is

$$\partial_\theta(d_z H) = \partial_\theta\psi.$$

If $z(\cdot,t)$ is a trajectory of the flow of this vector field and $u(\cdot,t)$ is a local coordinate representation of the trajectory, then a simple form of Hamiltonian's equation results:

$$\partial_t u = \partial_\theta\psi.$$

By analogy with the KdV studies, cf. Ref. 3, one may now introduce a natural symplectic form, Poisson brackets, and other concepts of the Hamiltonian theory. This will be done in the next section in a convenient form suggested by a variational principle for a related Lagrangian.

7. A LAGRANGIAN–HAMILTONIAN SYSTEM

Suppose that $z(\cdot,t)$ is a continuously differentiable curve in M and that $\partial_\theta v(\theta,t)$ is tangent to the curve for $0 \leqslant t \leqslant a$, so that Eq. (12) holds. If $u(\theta,t)$ is defined by Eq. (13), then since $\partial_t u = \partial_\theta v$ for t in $[0,a]$ and θ in $[0,2\pi]$, there is a function q associated with u and v such that

$$\partial_t q = v \quad \text{and} \quad \partial_\theta q = u. \tag{19}$$

In the following development the function q will play the role of a position vector while v will act as the momentum vector. The action A is defined by

$$A(q,v,t) = -\frac{1}{2}\int_C v\, dq.$$

The Lagrangian is then defined to be

$$L = A - V,$$

where V is given by Eq. (6). Following the classical variational approach to Lagrangian theory, consider smooth perturbations $v(\theta,t,\epsilon)$, and $z(\theta,t,\epsilon)$ parametrized by ϵ leaving the end points ($t = 0$ and $t = a$) fixed. Let w be the tangent vector for motion of q along the ϵ curve in the sense that

$$\partial_\epsilon q = w \quad \text{with } w(\theta,0,\epsilon) = w(\theta,a,\epsilon) = 0.$$

It follows from (19) that $\partial_\epsilon u = \partial_\theta w$. It is assumed that the corresponding perturbation of z satisfies

$$\partial_\epsilon x\, \partial_\theta y - \partial_\epsilon y\, \partial_\theta x = \partial_\theta w$$

so that u is a local coordinate representation of z. The classical problem in the calculus of variations is to find extremals of the functional

$$L = \int_0^a L\, dt,$$

i.e., to find an equation satisfied by q, v, and z when

$$\partial_\epsilon L\,|_{\epsilon=0} = 0 \tag{20}$$

for each w. The contribution due to the variation of A is computed as

$$\partial_\epsilon \int_0^a A\, dt = -\int_0^a dt \int_\Gamma (\partial_t q)(\partial_\theta w)\, d\theta.$$

The variation of V is just its derivative as given in Sec. 6:

$$\partial_\epsilon V = -\int_\Gamma \psi(\partial_\theta w)\, d\theta.$$

Thus, Eq. (20) implies that

$$\int_0^a dt \int_\Gamma (\partial_t q - \psi)(\partial_\theta w)\, d\theta = 0$$

for all w, or equivalently,

$$\int_0^a dt \int_\Gamma (\partial_t u - \partial_\theta \psi) w\, d\theta = 0.$$

Therefore, as the Euler–Lagrange equations for this variational problem one obtains

$$\partial_t u = \partial_\theta \psi \tag{21}$$

which agrees with previous equations of motion for the vortex boundary. An equivalent equation is, of course,

$$\partial_t q = \psi + \text{const.} \tag{22}$$

Continuing as in a classical development, define the Hamiltonian as

$$H(q,v) = L - A = -V.$$

The partial derivative notation of Sec. 6 yields the equations

$$\partial_u H = \psi,$$

$$\partial_q H = -\partial_\theta \psi,$$

which transform (21) and (22) into Hamilton's equations in the most familiar form:

$$\partial_t q = \partial_u H,$$

$$\partial_t u = -\partial_q H.$$

These local coordinate expression suggest immediately the following form for the Poisson brackets:

$$\{F,G\} = \frac{1}{2}\int_\Gamma (\partial_u F \partial_q G - \partial_p F \partial_u G)\, d\theta.$$

A functional F is therefore in involution with the Hamiltonian H whenever

$$0 = \{F,H\} = \frac{1}{2}\int (d_u F\, \partial_\theta \psi - \partial_p F\, \psi)\, d\theta. \tag{23}$$

In the case that

$$\partial_u F = f \quad \text{and} \quad \partial_p F = -\partial_\theta f,$$

Eq. (23) holds whenever

$$\int_\Gamma f\, d\psi = 0, \tag{24}$$

for example, when f is any function which agrees with ψ on the boundary C. The classical conserved quantities for a two-dimensional ideal fluid motion are the potential V, the area A, the centroid, (m_x,m_y), and the polar moment of inertia m_r, where

$$m_x = \frac{1}{2}\int x^2\, dy,$$

$$m_y = \frac{1}{2}\int y^2 dx,$$

and

$$m_r = \frac{1}{3}\int (x^3\, dy - y^3\, dx).$$

The corresponding Poisson brackets are obtained by substituting the following derivatives for the function f in Eq. (24):

$$\partial_u A = 0, \quad \partial_u V = -\psi,$$

$$\partial_u m_x = x, \quad \partial_u m_y = y,$$

and

$$\partial_u m_r = r^2.$$

The natural symplectic form is

$$\omega(\partial_\theta v_1, \partial_\theta v_2) = \int_\Gamma v_1\, dv_2.$$

8. THE m-FOLD SYMMETRIC VORTICES

If the regions of constant vorticity are small perturbations of the unit circle, then the theory is simple, and several results are known. The Hamiltonian structure of Sec. 7 sheds light on known results and raises some new questions for

investigation. First, as was mentioned in Sec. 5, a natural local coordinate system for star-shaped boundaries in a neighborhood of the unit circle is determined by the square of the radial distance from the origin. The equation of motion of the boundary is then

$$\partial_t r^2 = 2\partial_\theta \psi \tag{25}$$

which is nothing more than an equation describing the radial component of fluid velocity. The travelling wave solutions predicted by Kelvin would have the form

$$r^2(\theta,t) = r^2(\theta + \Omega t),$$

so that Eq. (25) becomes

$$\Omega\, \partial_\theta r^2 = 2\, \partial_\theta \psi, \tag{26}$$

the equation originally studied by Deem and Zabusky.[5] Integration of (26) yields

$$\psi - (\Omega/2) r^2 = \text{const.} \tag{27}$$

This is a nonlinear eigenvalue problem whose solution is obtained from standard bifurcation theory as follows: By taking the Frechet derivatives of both sides using the local coordinates $u = \frac{1}{2}(r^2 - 1)$ so that

$$\partial_\epsilon r^2 = 2\partial_\epsilon u = 2\partial_\theta v,$$

a linear eigenvalue problem is obtained, namely,

$$d_z \psi(\partial_\theta v) - \Omega\,(\partial_\theta v) = 0, \tag{28}$$

where z is the unit circle, a trivial solution to (27) for which the stream function is constant on the unit circle. The tangential speed $\partial_n \psi$ for the unit circle solution has the constant value $\frac{1}{2}\omega_0$; hence an application of Eq. (18) reduces (28) to

$$L_z(\partial_\theta v) = \lambda\,(\partial_\theta v),$$

where $\lambda = (2\Omega/\omega_0) - 1$. The eigenvalues and eigenfunctions for this equation, in which the logarithm kernel is evaluated only on the unit circle, are well known, namely,

$$\lambda_m = -(1/m)$$

and

$$f_m(\theta) = a_m \cos m\theta + b_m \sin m\theta.$$

Since the operator L_z is bounded and since the eigenvalues have multiplicity 1, the standard bifurcation theory (cf. Ref. 11) implies that for each n there is a one-parameter family of curves $z_n(\cdot,\epsilon)$ containing the unit circle as $z_n(\cdot,0)$ which solve Eq. (27):

$$\psi_{z(\cdot,\epsilon)} - \tfrac{1}{2}\Omega_n(\epsilon) r_n^2(\cdot,\epsilon) = \text{const.}$$

where

$$r_n^2 = z_n \bar{z}_n,$$

$$\Omega_n(0) = (\omega_0/2)(1 - (1/m)), \quad z_n(\theta,0) = e^{i\theta}.$$

These solutions to the nonlinear eigenvalue problem are the m-fold symmetric vortices obtained previously by Burbea[12] and by McKee[9] using similar arguments.

These results for the unit circle expressed in terms of local coordinates suggest a much more general structure on the manifold set M. For example, associated with a trajectory $z(\cdot,t)$ of the Hamiltonian vector field $\partial_\theta \psi$ is the linearized Hamiltonian system whose equation is

$$\partial_t v = d_{z(\cdot,t)} \psi\,(\partial_\theta v). \tag{29}$$

If the curve $z(\cdot,t)$ is parametrized by the "fluid coordinates" so that

$$\partial_t z = \text{rot}\ \psi$$

and if $v(\theta,t) = v(z(\theta,t),t)$ for some function v_0 defined on a neighborhood of the curve $z(\cdot,t)$, then

$$\partial_t v = D_t v_0 = \partial_t v_0 + \partial_x v\, \partial_y \psi - \partial_y v\, \partial_x \psi$$

and, by comparison with Eq. (18), Eq. (29) may be rewritten as

$$\partial_t v_0 = L_{z(\cdot,t)}(\partial_\theta v_0). \tag{30}$$

It is worth noting that the stream function itself is a solution of this equation, i.e.,

$$\partial_t \psi = L_{z(\cdot,t)}(\partial_\theta \psi)$$

which may be shown by a simple calculation. Associated with the linearized Hamiltonian system along a trajectory in M is a linear eigenvalue problem:

$$\lambda\, \partial_\theta v = L_{z(\cdot,t)}(\partial_\theta v) \tag{31}$$

suggested by the m-fold symmetry results for the unit circle. This equation is the linearization of the equation

$$\lambda u(\cdot,t,\epsilon) = \psi_{z(\cdot,t,\epsilon)} - \psi_{z(\cdot,t)}, \tag{32}$$

where u is a local coordinate representation of $z(\cdot,t,\epsilon), z(\cdot,t,0) = z(\cdot,t)$, and

$$\partial_\epsilon u|_{\epsilon=0} = \partial_\theta v.$$

Bifurcation theory implies the existence of ϵ-families of solutions $\lambda(t,\epsilon,n), u(\cdot,t,\epsilon,n), z(\cdot,t,\epsilon,n)$ for Eq. (32). The constancy of the functions $\lambda(t,\epsilon,n)u(\cdot,t,\epsilon,n) - \psi_{z(\cdot,t,\epsilon,n)}$ suggest local symmetry properties for the manifold set M arising from classical trajectories of the Hamiltonian system. Details and consequences of this arguement will be developed elsewhere.

9. ANALOGY WITH THE KdV EQUATION

It is interesting to note the parallels which occur when the Korteweg–de Vries equation (KdV) is expressed as a Hamiltonian system in notation similar to that used for the vortex boundary system. The set M may be taken to be the linear space of infinitely differentiable functions on the unit circle whose integral is 0; the tangent space T is the same linear space. If the KdV is normalized in the form

$$\partial_t u = u\partial_\theta u + \partial_{\theta\theta\theta}, \tag{33}$$

then an appropriate Hamiltonian is

$$H = \int_\Gamma \left(\frac{1}{6} u^3 - \frac{1}{2}(\partial_\theta u)^2\right) d\theta.$$

The derivative of the Hamiltonian is calculated as

$$d_u H(\partial_\theta w) = \int_\Gamma \left(\frac{1}{2} u^2 + \partial_{\theta\theta} u\right)(\partial_\theta w)\, \partial\theta, \tag{34}$$

where u varies according to

$$\partial_\epsilon u = \partial_\theta w.$$

If q is defined by the formulae

$$\partial_\theta q = u, \quad \partial_\epsilon q = w,$$

then integration by parts gives

$$d_q H(w) = -\int (u\, \partial_\theta u + \partial_{\theta\theta\theta} u) \cdot w \tag{35}$$

and Eqs. (34) and (35) can be written in partial derivative form as

$$\partial_u H = \tfrac{1}{2}(\partial_\theta q)^2 + \partial_{\theta\theta\theta} q,$$
$$\partial_q H = -u\,\partial_\theta u - \partial_{\theta\theta\theta} u.$$

Integration of the KdV gives

$$\partial_t q = \tfrac{1}{2}(\partial_\theta q)^2 + \partial_{\theta\theta\theta} q. \tag{36}$$

Equations (33) and (36) can be rewritten as Hamilton's equations

$$\partial_t q = \partial_u H,$$
$$\partial_t u = -\partial_q H.$$

A Lagrangian approach like that of Sec. 7 will give the same results. The KdV was derived previously as a Lagrangian system by Whitham[13] and somewhat differently by Hermann.[14]

The parallels with the vortex boundary system may be noted now in a systematic manner. The second derivative of the Hamiltonian is

$$d_u^2 H(\partial_\theta v, \partial_\theta w) = \int_\Gamma \partial_\theta v(\partial_{\theta\theta} + u)\,\partial_\theta w.$$

an equation similar to (17). The linearization of the integrated equation (36) is

$$\partial_t w = (\partial_{\theta\theta} + u)\partial_\theta w.$$

The travelling wave equation for the KdV may be expressed as a nonlinear eigenvalue problem for which the bifurcation methods of Sec. 8 are applicable, indeed the equation is

$$\lambda\,\partial_\theta u = u\,\partial_\theta u + \partial_{\theta\theta\theta} u$$

or, in integrated form,

$$\lambda u = \tfrac{1}{2}u^2 + \partial_{\theta\theta} u,$$

an equation analogous to (27). The linearization of this equation is

$$\lambda w = uw + \partial_{\theta\theta} w,$$

analogous to (28). The solution of the linearized eigenvalue problem about the trivial solution $u = 0$ is simply

$$w = a\sin(\gamma t) + b\cos(\gamma t),$$

where

$$\gamma = 2n\pi \quad \text{and} \quad -\gamma^2 = \lambda.$$

The bifurcation methods apply as in Sec. 8, and the result is a well-known one-parameter family of solutions to the nonlinear travelling wave equation, namely the cnoidal waves (cf. Leibovich and Seebass).[15]

[1] V. I. Arnold, Usp. Mat. Nauk. **24**, 225 (1969).
[2] J. E. Marsden, *Applications of Global Analysis in Mathematical Physics* (Publish or Perish, Berkeley, CA, 1974).
[3] R. Abraham and J. E. Marsden, *Foundations of Mechanics* (Benjamin/Cummings, Reading, MA, 1978).
[4] P. J. Olver, "A Nonlinear Hamiltonian Structure for the Euler Equations," preprint (1981).
[5] G. S. Deem and N. J. Zabusky, Phys. Rev. Lett. **40**, 859 (1978).
[6] J. Burbea, "Motions of Vortex Patches," Lett. Math. Phys. (to appear).
[7] H. Lamb, *Hydrodynamics* (Dover, New York, 1932).
[8] J. Burbea and M. Landau, "Kelvin Waves in Vortex Dynamics and Their Stability," J. Comp. Phys. (to appear)
[9] T. G. McKee, Ph.D. Thesis, Brown University, 1981.
[10] R. P. Feynman, R. B. Leighton, and M. Sands, *The Feynman Lecture Notes in Physics*, V.II (Addison-Wesley, Reading, MA, 1966).
[11] M. G. Crandall and P. H. Rabinowitz, J. Funct. Anal., **8**, 321 (1971).
[12] J. Burbea, *Proceedings Nonlinear Evolution Equations and Dynamical Systems, Leece 1979*, Springer Lecture Notes in Physics, Vol. 120 (Springer, New York, 1979), p. 276.
[13] G. B. Whitham, *Linear and Nonlinear Waves* (Wiley, New York, 1974).
[14] R. Hermann, in *Geometric Methods in Mathematical Physics*, edited by G. Kaiser and J. E. Marsden, Lecture Notes in Mathematics, Vol. 775 (Springer, New York, 1980), p. 145.
[15] S. Leibovich and A. R. Seebass, in *Nonlinear Waves*, edited by Leibovich and Seebass (Cornell U. P., Ithaca, NY, 1974).