Low-dimensional real Lie algebras

P. Turkowski
Academy of Agriculture, Physics Department, Mickiewicza 21, 31-120 Kraków, Poland

(Received 9 December 1987; accepted for publication 25 May 1988)

All real Lie algebras of dimension up to 8 that admit a nontrivial Levi decomposition are found.

I. INTRODUCTION

The central problem arising in the theory of real Lie algebras is determination of all nonisomorphic algebras. The class of semisimple algebras has been determined long ago by Cartan. Up to now, only some results on solvable algebras are known.[1-7] The aim of the present paper is to investigate algebras of a third group, namely the semidirect sums of semisimple and solvable algebras.

Our motivation is applications to cosmology. We hope, however, that the usefulness of the knowledge of algebras in other physical contexts is evident.[8] As far as the $(1+3)$-dimensional cosmological problem is concerned, algebras of nine Bianchi types that classify homogeneous space-times that are solutions to the Einstein field equations[9] are important. Recently, multidimensional cosmologies have also been proposed.[10] Therefore, it becomes necessary to extend the Bianchi classification of the real three-dimensional Lie algebras to a more general case of larger dimension and to face the problem of evaluating the structure constants for all nonisomorphic algebras. The knowledge of the algebras implies a classification of the homogeneous cosmological models. Furthermore, the field equations can be written down directly in terms of scale factors and in terms of the structure constants of the Killing vector algebras.[11]

In the following, we sketch the method of computing algebras that are the semidirect sums of the solvable and semisimple algebras. We apply this method to algebras of dimensions up to 8. There is one five-dimensional algebra and four six-dimensional algebras, summarized in Table I. Tables II and III provide seven algebras of dimension 7 and 22 algebras of dimension 8, respectively. These algebras lead naturally to a class of cosmological models that we discuss elsewhere.[12]

II. THE SEMIDIRECT SUMS

As far as the notation and method is concerned we follow our previous work.[13] We shall use the symbol $N \dot{+} S$ for a semidirect sum, writing the ideal N first, and the subalgebra S second.

We endow the semidirect sum with a Lie algebra structure by using $[\ ,\ ]_N$ and $[\ ,\ ]_S$ in each of these subalgebras. For Lie brackets between these two subalgebras, we set

$$[e_i,e_J] = R(e_i)*e_J, \quad e_i \in S, \quad e_J \in N, \tag{1}$$

where R is a linear mapping, $R(e_i)$: $N \ni e_J \to R(e_i)*e_J \in N$. For the linear subspaces N and S of the algebra $L = N \dot{+} S$, the following relations hold:

$$[N,N] \subset N, \quad [S,S] \subset S, \quad [N,S] \subset N, \tag{2}$$

$$\dim L = \dim N + \dim S. \tag{3}$$

Furthermore, from the Jacobi identity, it follows that $R(e_i)$ is a derivation of N:

$$R(e_i)*[e_J,e_K] = [R(e_i)*e_J,e_K] + [e_J,R(e_i)*e_K]. \tag{4}$$

The set $\{R(e_i)\}$ forms a Lie algebra itself, called the *derivation algebra*, and the homomorphism of S into the derivation algebra, $S \ni e_i \to R(e_i)$, must be a representation of the semisimple algebra S by real matrices. It is clear that the zero-matrix representation of S, acting in N, is a derivation of N. This representation reduces the semidirect sum $N \dot{+} S$ to the ordinary direct sum $N \oplus S$.

The fundamental Levi–Malcev theorem[14] says that, for an arbitrary Lie algebra L with the radical N, a semisimple subalgebra S exists such that

$$L = N \dot{+} S. \tag{5}$$

The semisimple subalgebra S is called the *Levi factor*. Consequently, Lie algebras fall into the following three categories: the semisimple algebras, the solvable algebras, and the semidirect sums of solvable and semisimple algebras.

III. A CLASS OF SEMIDIRECT SUMS

Now, we shall determine all real Lie algebras $L = N \dot{+} S$ such that $\dim L \leqslant 8$ and the semidirect sum is nontrivial: $N \neq 0$, $S \neq 0$, and $\dot{+} \neq \oplus$.

We start with a semisimple algebra S, $\dim S < 8$. To extend this algebra we proceed in the following way. The commutation rules of $[S,S]$ type are known:

$$[e_i,e_j] = \sum_k C^k{}_{ij} e_k, \quad e_i,e_j,e_k \in S. \tag{6}$$

Let $n = \dim N$. One is looking only for n-dimensional representations of S. These representations are $n \times n$ matrices that determine the commutators of $[N,S]$ type, as defined in Eq. (1). Then, we make use of (4) and we calculate the radicals N. Finally, to simplify the form of the radical, we perform suitable transformations of its basis elements.

There are three simple algebras of dimension less than 8: three-dimensional so(3) and sl(2,R), and six-dimensional so(3,1). To show how the method works let us consider the algebra $L = N \dot{+} \mathrm{so}(3)$, $\dim N = 4$. The algebra so(3) with the basis $\{e_1,e_2,e_3\}$ is defined by the nonvanishing structure constants

$$C^3{}_{12} = 1, \quad C^2{}_{31} = 1, \quad C^1{}_{23} = 1. \tag{7}$$

Since so(3) has the adjoint representation by 3×3 matrices,

 0022-2488/88/102139-06$02.50

TABLE I. Real Lie algebras of dimensions 5 and 6 that admit a nontrivial Levi decomposition.

Name	Levi decomposition	The representation of the Levi factor that defines the semidirect sum	Nonzero structure constants
$L_{5,1}$	$2L_1 \oplus \mathrm{sl}(2,R)$	$D_{1/2}$	$C^2{}_{12}=2,\ C^3{}_{13}=-2,\ C^1{}_{23}=1,$ $C^4{}_{14}=1,\ C^4{}_{25}=1,\ C^5{}_{34}=1,$ $C^5{}_{15}=-1$
$L_{6,1}$	$3L_1 \oplus \mathrm{so}(3)$	ad so(3)	$C^3{}_{12}=1,\ C^2{}_{31}=1,\ C^1{}_{23}=1,$ $C^6{}_{15}=1,\ C^6{}_{24}=-1,\ C^5{}_{34}=1,$ $C^5{}_{16}=-1,\ C^4{}_{26}=1,\ C^4{}_{35}=-1$
$L_{6,2}$	$A_{3,1} \oplus \mathrm{sl}(2,R)$	$D_{1/2} \oplus D_0$	$C^2{}_{12}=2,\ C^3{}_{13}=-2,\ C^1{}_{23}=1,$ $C^4{}_{14}=1,\ C^4{}_{25}=1,\ C^5{}_{34}=1,$ $C^5{}_{15}=-1,$ $C^6{}_{45}=1$
$L_{6,3}$	$A_{3,3} \oplus \mathrm{sl}(2,R)$	$D_{1/2} \oplus D_0$	$C^2{}_{12}=2,\ C^3{}_{13}=-2,\ C^1{}_{23}=1,$ $C^4{}_{14}=1,\ C^4{}_{25}=1,\ C^5{}_{34}=1,$ $C^5{}_{15}=-1,$ $C^4{}_{46}=1,\ C^5{}_{56}=1$
$L_{6,4}$	$3L_1 \oplus \mathrm{sl}(2,R)$	D_1	$C^2{}_{12}=2,\ C^3{}_{13}=-2,\ C^1{}_{23}=1,$ $C^4{}_{14}=2,\ C^4{}_{25}=2,\ C^5{}_{34}=1,$ $C^6{}_{16}=-2,\ C^5{}_{26}=1,\ C^6{}_{35}=2$

TABLE II. Real Lie algebras of dimension 7 that admit a nontrivial Levi decomposition.

Name	Levi decomposition	The representation of the Levi factor that defines the semidirect sum	Nonzero structure constants
$L_{7,1}$	$A^{1,1}_{4,5} \oplus \mathrm{so}(3)$	ad so(3) $\oplus D_0$	$C^3{}_{12}=1,\ C^2{}_{31}=1,\ C^1{}_{23}=1,$ $C^6{}_{15}=1,\ C^6{}_{24}=-1,\ C^5{}_{34}=1,$ $C^5{}_{16}=-1,\ C^4{}_{26}=1,\ C^4{}_{35}=-1,$ $C^4{}_{47}=1,\ C^5{}_{57}=1,\ C^6{}_{67}=1$
$L_{7,2}$	$4L_1 \oplus \mathrm{so}(3)$	R_4	$C^3{}_{12}=1,\ C^2{}_{31}=1,\ C^1{}_{23}=1,$ $C^7{}_{14}=\frac{1}{2},\ C^5{}_{24}=\frac{1}{2},\ C^6{}_{34}=\frac{1}{2},$ $C^6{}_{15}=\frac{1}{2},\ C^4{}_{25}=-\frac{1}{2},\ C^7{}_{35}=-\frac{1}{2},$ $C^5{}_{16}=-\frac{1}{2},\ C^7{}_{26}=\frac{1}{2},\ C^4{}_{36}=-\frac{1}{2},$ $C^4{}_{17}=-\frac{1}{2},\ C^6{}_{27}=-\frac{1}{2},\ C^5{}_{37}=\frac{1}{2}$
$L^{p}_{7,3}$	$A^{1,p}_{4,5} \oplus \mathrm{sl}(2,R)$ $p \neq 0$	$D_{1/2} \oplus 2D_0$	$C^2{}_{12}=2,\ C^3{}_{13}=-2,\ C^1{}_{23}=1,$ $C^4{}_{14}=1,\ C^4{}_{25}=1,\ C^5{}_{34}=1,$ $C^5{}_{15}=-1,$ $C^4{}_{47}=1,\ C^5{}_{57}=1,\ C^6{}_{67}=p$
$L_{7,4}$	$A^{1}_{4,9} \oplus \mathrm{sl}(2,R)$	$D_{1/2} \oplus 2D_0$	$C^2{}_{12}=2,\ C^3{}_{13}=-2,\ C^1{}_{23}=1,$ $C^4{}_{14}=1,\ C^4{}_{25}=1,\ C^5{}_{34}=1,$ $C^5{}_{15}=-1,$ $C^6{}_{45}=1,$ $C^4{}_{47}=1,\ C^5{}_{57}=1,\ C^6{}_{67}=2$
$L_{7,5}$	$A^{1,1}_{4,5} \oplus \mathrm{sl}(2,R)$	$D_1 \oplus D_0$	$C^2{}_{12}=2,\ C^3{}_{13}=-2,\ C^1{}_{23}=1,$ $C^4{}_{14}=2,\ C^4{}_{25}=2,\ C^5{}_{34}=1,$ $C^6{}_{16}=-2,\ C^5{}_{26}=1,\ C^6{}_{35}=2,$ $C^4{}_{47}=1,\ C^5{}_{57}=1,\ C^6{}_{67}=1$
$L_{7,6}$	$4L_1 \oplus \mathrm{sl}(2,R)$	$D_{3/2}$	$C^2{}_{12}=2,\ C^3{}_{13}=-2,\ C^1{}_{23}=1,$ $C^4{}_{14}=3,\ C^4{}_{25}=3,\ C^5{}_{34}=1,$ $C^5{}_{15}=1,\ C^5{}_{26}=2,\ C^6{}_{35}=2,$ $C^6{}_{16}=-1,\ C^6{}_{27}=1,\ C^7{}_{36}=3,$ $C^7{}_{17}=-3.$
$L_{7,7}$	$4L_1 \oplus \mathrm{sl}(2,R)$	$2D_{1/2}$	$C^2{}_{12}=2,\ C^3{}_{13}=-2,\ C^1{}_{23}=1,$ $C^4{}_{14}=1,\ C^4{}_{25}=1,\ C^5{}_{34}=1,$ $C^5{}_{15}=-1,$ $C^6{}_{16}=1,\ C^6{}_{27}=1,\ C^7{}_{36}=1,$ $C^7{}_{17}=-1.$

TABLE III. Real Lie algebras of dimension 8 that admit a nontrivial Levi decomposition.

Name	Levi decomposition	The representation of the Levi factor that defines the semidirect sum	Nonzero structure constants
$L^{p}_{8,1}$	$A^{1,1,p}_{5,7} \oplus so(3)$ $p \neq 0$	$ad\, so(3) \oplus 2D_0$	$C^3{}_{12}=1, C^2{}_{31}=1, C^1{}_{23}=1,$ $C^6{}_{15}=1, C^6{}_{24}=-1, C^5{}_{34}=1,$ $C^5{}_{16}=-1, C^4{}_{26}=1, C^4{}_{35}=-1,$ $C^4{}_{48}=1, C^5{}_{58}=1, C^6{}_{68}=1,$ $C^7{}_{78}=p$
$L_{8,2}$	$A_{5,4} \oplus so(3)$	$R_4 \oplus D_0$	$C^3{}_{12}=1, C^2{}_{31}=1, C^1{}_{23}=1,$ $C^7{}_{14}=\frac{1}{2}, C^5{}_{24}=\frac{1}{2}, C^6{}_{34}=\frac{1}{2},$ $C^6{}_{15}=\frac{1}{2}, C^4{}_{25}=-\frac{1}{2}, C^7{}_{35}=-\frac{1}{2},$ $C^5{}_{16}=-\frac{1}{2}, C^7{}_{26}=\frac{1}{2}, C^4{}_{36}=-\frac{1}{2},$ $C^4{}_{17}=-\frac{1}{2}, C^6{}_{27}=-\frac{1}{2}, C^5{}_{37}=\frac{1}{2},$ $C^8{}_{45}=1, C^8{}_{67}=-1$
$L_{8,3}$	$A^{1,1,1}_{5,7} \oplus so(3)$	$R_4 \oplus D_0$	$C^3{}_{12}=1, C^2{}_{31}=1, C^1{}_{23}=1,$ $C^7{}_{14}=\frac{1}{2}, C^5{}_{24}=\frac{1}{2}, C^6{}_{34}=\frac{1}{2},$ $C^6{}_{15}=\frac{1}{2}, C^4{}_{25}=-\frac{1}{2}, C^7{}_{35}=-\frac{1}{2},$ $C^5{}_{16}=-\frac{1}{2}, C^7{}_{26}=\frac{1}{2}, C^4{}_{36}=-\frac{1}{2},$ $C^4{}_{17}=-\frac{1}{2}, C^6{}_{27}=-\frac{1}{2}, C^5{}_{37}=\frac{1}{2},$ $C^4{}_{48}=1, C^5{}_{58}=1, C^6{}_{68}=1,$ $C^7{}_{78}=1$
$L^{p}_{8,4}$	$A^{1,p,p}_{5,17} \oplus so(3)$	$R_4 \oplus D_0$	$C^3{}_{12}=1, C^2{}_{31}=1, C^1{}_{23}=1,$ $C^7{}_{14}=\frac{1}{2}, C^5{}_{24}=\frac{1}{2}, C^6{}_{34}=\frac{1}{2},$ $C^6{}_{15}=\frac{1}{2}, C^4{}_{25}=-\frac{1}{2}, C^7{}_{35}=-\frac{1}{2},$ $C^5{}_{16}=-\frac{1}{2}, C^7{}_{26}=\frac{1}{2}, C^4{}_{36}=-\frac{1}{2},$ $C^4{}_{17}=-\frac{1}{2}, C^6{}_{27}=-\frac{1}{2}, C^5{}_{37}=\frac{1}{2}$ $C^4{}_{48}=p, C^6{}_{48}=-1,$ $C^5{}_{58}=p, C^7{}_{58}=-1,$ $C^4{}_{68}=1, C^6{}_{68}=p,$ $C^5{}_{78}=1, C^7{}_{78}=p$
$L_{8,5}$	$5L_1 \oplus so(3)$	R_5	$C^3{}_{12}=1, C^2{}_{31}=1, C^1{}_{23}=1,$ $C^7{}_{14}=\frac{1}{2}, C^6{}_{24}=\frac{1}{2}, C^5{}_{34}=2,$ $C^6{}_{15}=-\frac{1}{2}, C^7{}_{25}=\frac{1}{2}, C^4{}_{35}=-2,$ $C^5{}_{16}=2, C^4{}_{26}=-2, C^7{}_{36}=1,$ $C^8{}_{16}=-1, C^5{}_{27}=-2, C^6{}_{37}=-1,$ $C^4{}_{17}=-2, C^8{}_{27}=-1,$ $C^6{}_{18}=3, C^7{}_{28}=3$
$L_{8,6}$	$A_{5,4} \oplus sl(2,R)$	$D_{1/2} \oplus 3D_0$	$C^2{}_{12}=2, C^3{}_{13}=-2, C^1{}_{23}=1,$ $C^4{}_{14}=1, C^4{}_{25}=1, C^5{}_{34}=1,$ $C^5{}_{15}=-1,$ $C^8{}_{45}=1, C^8{}_{67}=1$
$L^{p,q}_{8,7}$	$A^{1,p,q}_{5,7} \oplus sl(2,R)$ $pq \neq 0$	$D_{1/2} \oplus 3D_0$	$C^2{}_{12}=2, C^3{}_{13}=-2, C^1{}_{23}=1,$ $C^4{}_{14}=1, C^4{}_{25}=1, C^5{}_{34}=1,$ $C^5{}_{15}=-1,$ $C^4{}_{48}=1, C^5{}_{58}=1, C^6{}_{68}=p,$ $C^7{}_{78}=q$
$L^{p}_{8,8}$	$A^{1/p,1/p}_{5,9} \oplus sl(2,R)$ $p \neq 0$ $A^{1}_{5,8} \oplus sl(2,R)$ $p=0$	$D_{1/2} \oplus 3D_0$	$C^2{}_{12}=2, C^3{}_{13}=-2, C^1{}_{23}=1,$ $C^4{}_{14}=1, C^4{}_{25}=1, C^5{}_{34}=1,$ $C^5{}_{15}=-1,$ $C^4{}_{48}=1, C^5{}_{58}=1, C^6{}_{68}=p,$ $C^6{}_{78}=1, C^7{}_{78}=p$
$L^{p,q}_{8,9}$	$A^{1,p,q}_{5,13} \oplus sl(2,R)$ $q \neq 0$	$D_{1/2} \oplus 3D_0$	$C^2{}_{12}=2, C^3{}_{13}=-2, C^1{}_{23}=1,$ $C^4{}_{14}=1, C^4{}_{25}=1, C^5{}_{34}=1,$ $C^5{}_{15}=-1,$ $C^4{}_{48}=1, C^5{}_{58}=1, C^6{}_{68}=p,$ $C^7{}_{68}=-q, C^6{}_{78}=q, C^7{}_{78}=p$
$L^{p}_{8,10}$	$A^{2,p}_{5,19} \oplus sl(2,R)$ $p \neq 0$	$D_{1/2} \oplus 3D_0$	$C^2{}_{12}=2, C^3{}_{13}=-2, C^1{}_{23}=1,$ $C^4{}_{14}=1, C^4{}_{25}=1, C^5{}_{34}=1,$ $C^5{}_{15}=-1,$ $C^4{}_{48}=1, C^5{}_{58}=1, C^6{}_{68}=2,$ $C^7{}_{78}=p, C^6{}_{45}=1,$

TABLE III. (*Continued.*)

Name	Levi decomposition	The representation of the Levi factor that defines the semidirect sum	Nonzero structure constants
$L_{8,11}$	$A^{2}_{5,20} \oplus \mathrm{sl}(2,R)$	$D_{1/2} \oplus 3D_0$	$C^{2}_{12}=2, C^{3}_{13}=-2, C^{1}_{23}=1,$ $C^{4}_{14}=1, C^{4}_{25}=1, C^{5}_{34}=1,$ $C^{5}_{15}=-1,$ $C^{4}_{48}=1, C^{5}_{58}=1, C^{6}_{68}=2,$ $C^{6}_{78}=1, C^{7}_{78}=2, C^{6}_{45}=1$
$L^{p}_{8,12}$	$A^{1,1,p}_{5,7} \oplus \mathrm{sl}(2,R)$ $p \neq 0$	$D_1 \oplus 2D_0$	$C^{2}_{12}=2, C^{3}_{13}=-2, C^{1}_{23}=1,$ $C^{4}_{14}=2, C^{4}_{25}=2, C^{5}_{34}=1,$ $C^{6}_{16}=-2, C^{5}_{26}=1, C^{6}_{35}=2,$ $C^{4}_{48}=1, C^{5}_{58}=1, C^{6}_{68}=1,$ $C^{7}_{78}=p$
$L^{\varepsilon}_{8,13}$	$A_{5,4} \oplus \mathrm{sl}(2,R)$ $\varepsilon = \pm 1$	$2D_{1/2} \oplus D_0$	$C^{2}_{12}=2, C^{3}_{13}=-2, C^{1}_{23}=1,$ $C^{4}_{14}=1, C^{4}_{25}=1, C^{5}_{34}=1,$ $C^{5}_{15}=-1,$ $C^{6}_{16}=1, C^{6}_{27}=1, C^{7}_{36}=1,$ $C^{7}_{17}=-1,$ $C^{8}_{45}=1, C^{8}_{67}=\varepsilon$
$L_{8,14}$	$A_{5,1} \oplus \mathrm{sl}(2,R)$	$2D_{1/2} \oplus D_0$	$C^{2}_{12}=2, C^{3}_{13}=-2, C^{1}_{23}=1,$ $C^{4}_{14}=1, C^{4}_{25}=1, C^{5}_{34}=1,$ $C^{5}_{15}=-1,$ $C^{6}_{16}=1, C^{6}_{27}=1, C^{7}_{36}=1,$ $C^{7}_{17}=-1,$ $C^{4}_{68}=1, C^{5}_{78}=1$
$L_{8,15}$	$A_{5,3} \oplus \mathrm{sl}(2,R)$	$2D_{1/2} \oplus D_0$	$C^{2}_{12}=2, C^{3}_{13}=-2, C^{1}_{23}=1,$ $C^{4}_{14}=1, C^{4}_{25}=1, C^{5}_{34}=1,$ $C^{5}_{15}=-1,$ $C^{6}_{16}=1, C^{6}_{27}=1, C^{7}_{36}=1,$ $C^{7}_{17}=-1,$ $C^{4}_{68}=1, C^{5}_{78}=1, C^{8}_{67}=1$
$L_{8,16}$	$A^{1}_{5,15} \oplus \mathrm{sl}(2,R)$	$2D_{1/2} \oplus D_0$	$C^{2}_{12}=2, C^{3}_{13}=-2, C^{1}_{23}=1,$ $C^{4}_{14}=1, C^{4}_{25}=1, C^{5}_{34}=1,$ $C^{5}_{15}=-1,$ $C^{6}_{16}=1, C^{6}_{27}=1, C^{7}_{36}=1,$ $C^{7}_{17}=-1,$ $C^{4}_{48}=1, C^{5}_{58}=1, C^{4}_{68}=1,$ $C^{6}_{68}=1, C^{5}_{78}=1, C^{7}_{78}=1,$
$L^{p}_{8,17}$	$A^{1,p,p}_{5,7} \oplus \mathrm{sl}(2,R)$ $-1 \leqslant p \leqslant 1$	$2D_{1/2} \oplus D_0$	$C^{2}_{12}=2, C^{3}_{13}=-2, C^{1}_{23}=1,$ $C^{4}_{14}=1, C^{4}_{25}=1, C^{5}_{34}=1,$ $C^{5}_{15}=-1,$ $C^{6}_{16}=1, C^{6}_{27}=1, C^{7}_{36}=1,$ $C^{7}_{17}=-1,$ $C^{4}_{48}=1, C^{5}_{58}=1, C^{6}_{68}=p,$ $C^{7}_{78}=p$
$L^{p}_{8,18}$	$A^{1,p,p}_{5,17} \oplus \mathrm{sl}(2,R)$ $p \geqslant 0$	$2D_{1/2} \oplus D_0$	$C^{2}_{12}=2, C^{3}_{13}=-2, C^{1}_{23}=1,$ $C^{4}_{14}=1, C^{4}_{25}=1, C^{5}_{34}=1,$ $C^{5}_{15}=-1,$ $C^{6}_{16}=1, C^{6}_{27}=1, C^{7}_{36}=1,$ $C^{7}_{17}=-1,$ $C^{4}_{48}=p, C^{6}_{48}=-1,$ $C^{5}_{58}=p, C^{7}_{58}=-1,$ $C^{4}_{68}=1, C^{6}_{68}=p,$ $C^{5}_{78}=1, C^{7}_{78}=p$
$L_{8,19}$	$A_{5,4} \oplus \mathrm{sl}(2,R)$	$D_{3/2} \oplus D_0$	$C^{2}_{12}=2, C^{3}_{13}=-2, C^{1}_{23}=1,$ $C^{4}_{14}=3, C^{4}_{25}=3, C^{5}_{34}=1,$ $C^{5}_{15}=1, C^{5}_{26}=2, C^{6}_{35}=2,$ $C^{6}_{16}=-1, C^{6}_{27}=1, C^{7}_{36}=3,$ $C^{7}_{17}=-3,$ $C^{8}_{47}=1, C^{8}_{56}=-3$

TABLE III. (*Continued.*)

Name	Levi decomposition	The representation of the Levi factor that defines the semidirect sum	Nonzero structure constants
$L_{8,20}$	$A_{5,7}^{1,1,1} \oplus sl(2,R)$	$D_{3/2} \oplus D_0$	$C^2_{12}=2, C^3_{13}=-2, C^1_{23}=1,$ $C^4_{14}=3, C^4_{25}=3, C^5_{34}=1,$ $C^5_{15}=1, C^5_{26}=2, C^6_{35}=2,$ $C^6_{16}=-1, C^6_{27}=1, C^7_{36}=3,$ $C^7_{17}=-3,$ $C^4_{48}=1, C^5_{58}=1, C^6_{68}=1,$ $C^7_{78}=1$
$L_{8,21}$	$5L_1 \oplus sl(2,R)$	D_2	$C^2_{12}=2, C^3_{13}=-2, C^1_{23}=1,$ $C^4_{14}=4, C^4_{25}=4, C^5_{34}=1,$ $C^5_{15}=2, C^5_{26}=3, C^6_{35}=2,$ $C^7_{17}=-2, C^6_{27}=2, C^7_{36}=3,$ $C^8_{18}=-4, C^7_{28}=1, C^8_{37}=4$
$L_{8,22}$	$5L \oplus sl(2,R)$	$D_1 \oplus D_{1/2}$	$C^2_{12}=2, C^3_{13}=-2, C^1_{23}=1,$ $C^4_{14}=2, C^4_{25}=2, C^5_{34}=1,$ $C^6_{16}=-2, C^5_{26}=1, C^6_{35}=2,$ $C^7_{17}=1, C^7_{28}=1, C^8_{37}=1,$ $C^8_{18}=-1$

we can define $R = \text{ad so}(3) \oplus [0]$,

$$R(e_1)=\begin{pmatrix}0&0&0&0\\0&0&-1&0\\0&1&0&0\\0&0&0&0\end{pmatrix},$$

$$R(e_2)=\begin{pmatrix}0&0&1&0\\0&0&0&0\\-1&0&0&0\\0&0&0&0\end{pmatrix},$$

$$R(e_3)=\begin{pmatrix}0&-1&0&0\\1&0&0&0\\0&0&0&0\\0&0&0&0\end{pmatrix}.$$

This, via relation (1), implies the following commutation relations:

$$\begin{aligned}&[e_1,e_5]=+e_6, \quad [e_2,e_4]=-e_6, \quad [e_3,e_4]=+e_5,\\ &[e_1,e_6]=-e_5, \quad [e_2,e_6]=+e_4, \quad [e_3,e_5]=-e_4,\end{aligned} \tag{8}$$

where $\{e_4,e_5,e_6,e_7\}$ are the basis elements of N. With the help of Jacobi identity (4) we find that the four-dimensional solvable algebra N is either Abelian or a solvable algebra given by the following nonzero commutators:

$$[e_J,e_7]=e_J, \quad J=4,5,6. \tag{9}$$

This is the algebra $A_{4,5}^{1,1}$ in the list by Patera *et al.*[6] Hence we distinguish two types of algebras defined by Eqs. (7), (8) and (7)–(9), respectively,

$$L_7 = L_1 \oplus (3L_1 \oplus \text{so}(3)) = L_1 \oplus L_{6,1}$$

and

$$L_{7,1} = A_{4,5}^{1,1} \oplus \text{so}(3).$$

The term $L_{r,j}$ denotes an r-dimensional algebra of jth type, and nL_i an Abelian n-dimensional algebra.

In all other cases we proceed analogously. The results are given in Tables I–III in which we present radicals, Levi factors, representations of Levi's factors determining semidirect sums, and all nonzero structure constants. The following notation is used. The term $A_{r,j}$ denotes an r-dimensional solvable algebra of the jth type; for commutation relations see Ref. 6. The term D_J denotes the real representation of $sl(2,R)$ and is taken in its standard form,

$$e_1 \to \begin{pmatrix}2J&&&\\&2J-2&&\\&&\ddots&\\&&&-2J\end{pmatrix}, \quad e_2 \to \begin{pmatrix}0&2J&&&\\&0&2J-1&&\\&&0&\cdot&\\&&&\cdot&1\\&&&&0\end{pmatrix}, \quad e_3 \to \begin{pmatrix}0&&&&\\1&0&&&\\&2&0&&\\&&\cdot&\cdot&\\&&&2J&0\end{pmatrix},$$

where $\{e_1,e_2,e_3\}$ with the following multiplication table: $[e_1,e_2]=2e_2$, $[e_1,e_3]=-2e_3$, $[e_2,e_3]=e_1$, form a basis of $sl(2,R)$.

The terms R_4 and R_5 denote the four-dimensional and five-dimensional (respectively) real irreducible representations of so(3),

$$\bar{e}_1 \to \begin{pmatrix} 0 & 0 & 0 & -\frac{1}{2} \\ 0 & 0 & -\frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & 0 & 0 & 0 \end{pmatrix}, \quad \bar{e}_2 \to \begin{pmatrix} 0 & -\frac{1}{2} & 0 & 0 \\ \frac{1}{2} & 0 & 0 & 0 \\ 0 & 0 & 0 & -\frac{1}{2} \\ 0 & 0 & \frac{1}{2} & 0 \end{pmatrix}, \quad \bar{e}_3 \to \begin{pmatrix} 0 & 0 & -\frac{1}{2} & 0 \\ 0 & 0 & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & 0 & 0 \\ 0 & -\frac{1}{2} & 0 & 0 \end{pmatrix};$$

$$\bar{e}_1 \to \begin{pmatrix} 0 & 0 & 0 & -2 & 0 \\ 0 & 0 & 2 & 0 & 0 \\ 0 & -\frac{1}{2} & 0 & 0 & 3 \\ \frac{1}{2} & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 \end{pmatrix}, \quad \bar{e}_2 \to \begin{pmatrix} 0 & 0 & -2 & 0 & 0 \\ 0 & 0 & 0 & -2 & 0 \\ \frac{1}{2} & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{2} & 0 & 0 & 3 \\ 0 & 0 & 0 & -1 & 0 \end{pmatrix}, \quad \bar{e}_3 \to \begin{pmatrix} 0 & -2 & 0 & 0 & 0 \\ 2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix},$$

where $\{\bar{e}_1, \bar{e}_2, \bar{e}_3\}$ forms a basis for so(3). In the names of some algebras the superscripts are given to specify the value(s) of continuous parameters on which the algebra depends.

ACKNOWLEDGMENTS

I am grateful to M. Heller for reading the manuscript.

This work was supported in part by the Polish Interdisciplinary Research Project CPBP 01.03.

[1]M. A. H. MacCallum, "On the Classification of the Real four-dimensional Lie Algebras," Queen Mary College (London) preprint, 1979.
[2]V. V. Morozov, Izv. Vyssh. Uchebn. Zaved. Mat. **5**, 161 (1958).
[3]G. M. Mubarakzyanov, Izv. Vyssh. Uchebn. Zaved. Mat. **32**, 114 (1963).
[4]G. M. Mubarakzyanov, Izv. Vyssh. Uchebn. Zaved. Mat. **34**, 99 (1963)
[5]G. M. Mubarakzyanov, Izv. Vyssh. Uchebn. Zaved. Mat. **35**, 104 (1963).
[6]J. Patera, R. T. Sharp, P. Winternitz, and H. Zassenhaus, J. Math. Phys. **17**, 986 (1976).
[7]L. Magnin, J. Geom. Phys. **3**, 119 (1986).
[8]See, e.g., Ref. 6.
[9]D. Kramer, H. Stephani, M. MacCallum, and E. Herlt, *Exact Solutions of Einstein's Field Equations* (Cambridge U.P., Cambridge, 1980).
[10]A. Chodos and S. Detweiler, Phys. Rev. D **21**, 2167 (1980). Later papers on classical multidimensional cosmologies are the following: D. Sahdev, Phys. Lett. B **137**, 155 (1984); R. B. Abbott, S. M. Barr, and S. D. Ellis, Phys. Rev. D **30**, 720 (1984); E. W. Kolb, D. Lindley, and D. Seckel, *ibid.* **30**, 1205 (1984); D. Sahdev, *ibid.* **30**, 2495 (1984); J. Demaret and J.-L. Hanquin, *ibid.* **31**, 258 (1985); R. B. Abbott, S. D. Ellis, and S. M. Barr, *ibid.* **31**, 673 (1985); J. D. Barrow and J. Stein-Schabes, *ibid.* **32**, 1595 (1985); H. Ishihara, Prog. Theor. Phys. **74**, 490 (1985); A. Tomimatsu and H. Ishihara, Gen. Relativ. Gravit. **18**, 161 (1986); M. Demiański, Z. Golda, M. Heller, and M. Szydłowski, Class. Quantum Gravit. **3**, 1199 (1986); D. Lorenz-Petzold, Phys. Lett. B **167**, 157 (1986); P. Turkowski and K. Maślanka, Gen. Relativ. Gravit. **19**, 611 (1987); M. Rosenbaum, M. Ryan, L. Urrita, and R. Matzner, Phys. Rev. D **36**, 1032 (1987); M. Demiański, M. Heller, and M. Szydłowski, *ibid.* **36**, 2945 (1987); P. Turkowski, Gen. Relativ. Gravit. **19**, 1267 (1987). For a systematic treatment based on the algebraic classification of cosmologies see the following: J. Demaret and J.-L. Hanquin, Phys. Rev. D **31**, 258 (1985); P. Halpern, *ibid.* **33**, 354 (1986); M. Demiański, Z. Golda, L. M. Sokołowski, M. Szydłowski, and P. Turkowski, J. Math. Phys. **28**, 171 (1987).
[11]A. H. Taub, Ann. Math. **53**, 472 (1951); O. Heckmann and E. Schücking, in *Gravitation*, edited by L. Witten (Wiley, New York 1962), p. 438.
[12]P. Turkowski, Phys. Lett. B **207**, 36 (1988).
[13]P. Turkowski, J. Geom. Phys. **4**, 119 (1987).
[14]A. O. Barut and R. Rączka, *Theory of Group Representations and Applications* (PWN–Polish Scientific, Warsaw, 1977).

Nonlinear equations with superposition formulas and exceptional group G_2. III. The superposition formulas

L. Gagnon
Centre de Recherches Mathématiques, Université de Montréal, C.P. 6128, Succ. A, Montréal, Québec H3C 3J7, Canada

V. Hussin[a)]
Physique Théorique et Mathématique, Université de Liège, Institut de Physique au Sart Tilman, B5, B-4000, Liège 1, Belgique

P. Winternitz
Centre de Recherches Mathématiques, Université de Montréal, C.P. 6128, Succ. A, Montréal, Quebec H3C 3J7, Canada

(Received 18 August 1987; accepted for publication 1 June 1988)

Superposition formulas are derived expressing the general solution of several different systems of nonlinear ordinary differential equations in terms of a fundamental set of particular solutions. The equations, as well as the superposition formulas, are induced by the action of the exceptional Lie group G_2 (complex or real) on a homogeneous space G_2/G, where $G \subset G_2$ is a maximal subgroup of G_2. When G is either parabolic, or simple, three particular solutions are needed. When G is $SL(2,\mathbb{C}) \times SL(2,\mathbb{C})$ (or one of its real forms), then two particular solutions suffice.

I. INTRODUCTION

The present paper is the third (and last) in a series[1,2] devoted to systems of nonlinear ordinary differential equations with superposition formulas, based on the Cartan exceptional Lie group G_2. The first[1] (further referred to as I) was devoted to a study of subalgebras of the real and complex forms of the algebra g_2 and their matrix realizations. The nonlinear equations were derived in the second paper[2] (further referred to as II). In this paper we present the actual nonlinear superposition formulas.

We recall that a system of n first-order ordinary differential equations (ODE's)

$$\dot{\mathbf{y}} = \boldsymbol{\eta}(\mathbf{y},t), \quad \mathbf{y},\boldsymbol{\eta} \in F^n, \quad t \in \mathbb{R}, \quad F = \mathbb{R} \text{ or } \mathbb{C}, \tag{1.1}$$

is said to allow a nonlinear superposition formula, if its general solution can be expressed functionally in terms of a finite number m of particular solutions, and n significant constants,

$$\mathbf{y}(t) = \mathbf{F}(\mathbf{y}_1(t),...,\mathbf{y}_m(t),c_1,...,c_n). \tag{1.2}$$

If such a formula exists, then the set $\{\mathbf{y}_1(t),..,\mathbf{y}_m(t)\}$ is called a fundamental set of solutions.

Lie and Scheffers[3] have characterized all such equations and proven that they must have the form

$$\dot{\mathbf{y}}(t) = \sum_{k=1}^{r} Z_k(t)\xi_k(\mathbf{y}), \tag{1.3}$$

where the coefficients $\xi_k(\mathbf{y})$ are such that the vector fields

$$X_k = \sum_{\mu=1}^{n} \xi_k^{\mu}(\mathbf{y}) \frac{\partial}{\partial y^{\mu}} \tag{1.4}$$

generate a finite-dimensional Lie algebra L.

It was recently shown[4] that indecomposable systems of nonlinear differential equations with superposition formulas are associated with transitive primitive Lie algebras.[4] Related papers were devoted to equations based on the classical Lie groups and their maximal subgroups.[5–8] For the motivation of our interest in superposition formulas and applications we refer to earlier papers.[1,2,4–11] For an extension of the approach from Lie algebras to Lie superalgebras, see Ref. 12. For applications to the study of nonlinear wave equations see Ref. 13 and references therein. The relation between nonlinear ODE's with superposition formulas and Bäcklund transformations is discussed, e.g., by Chau.[14,15] Such systems of first-order ODE's can be used to integrate interesting second-order ODE's.[16,17]

The right-hand side of Eq. (1.3) can, for t fixed, be viewed as an element of the Lie algebra L [with a basis (1.4)]. As time t varies, the right-hand side of Eq. (1.3) describes a curve in this Lie algebra.

The solutions of the equations can in turn be written in the form[6,7]

$$\mathbf{y}(t) = G(t)\mathbf{y}_0, \tag{1.5}$$

where $\mathbf{y}_0$ is a constant vector and $G(t)$ is a path in the Lie group G, corresponding to the Lie algebra L. In order to obtain the superposition formula (1.2) explicitly, we proceed as follows.

(i) Construct the model of the homogeneous space $G/G_0 \sim M$ for which the infinitesimal group action is given by the vector fields (1.4). Here G is the connected component of the Lie group corresponding to the Lie algebra L of Lie's theorem, G_0 is the subgroup leaving the origin invariant. The Lie algebra $L_0 \subset L$ corresponding to G_0 is the subalgebra of vector fields (1.4), vanishing at the origin. Choose some parametrization of the group G and write the formula for the (in general nonlinear) action of G on M. This is formula (1.5).

(ii) Reconstruct the path $G(t)$ in the group G. To do this we take m solutions of Eqs. (1.3) and write (1.5) for them;

$$\mathbf{y}_k(t) = G(t)\mathbf{y}_{k0}, \quad k = 1,...,m. \tag{1.6}$$

[a)] Present address: Centre de Recherches Mathématiques, Université de Montréal, C.P. 6128, Succ. A, Montréal, Québec H3C 3J7, Canada.

 0022-2488/88/102145-11$02.50

We must then solve Eqs. (1.6) for the matrix elements of $G(t)$ (in the chosen parametrization). The set of algebraic equations (1.6) will determine $G(t)$ uniquely if the isotropy subgroup of G leaving the initial conditions invariant is only the identity group. In other words, $\mathbf{y}_1(t),...,\mathbf{y}_m(t)$ will be a fundamental set of solutions if the equations

$$\mathbf{y}_k(0) = G(t)\mathbf{y}_{k0}, \quad k = 1,...,m, \tag{1.7}$$

imply

$$G(t) = \mathrm{Id},$$

where Id is the identity transformation on M [and $\mathbf{y}_k(0)$ are viewed as m points on M]. Once $G(t)$ is completely determined in terms of the solutions $\mathbf{y}_k(t)$, formula (1.5) is the required superposition formula.

Each solution $\mathbf{y}_k(t)$, together with its initial condition value $\mathbf{y}_k(0)$, when substituted into (1.6), provides n algebraic equations for the matrix elements $g_{ik}(t)$ of $G(t)$. Since the dimension of the Lie group G is by assumption dim $G = r$, we need at least r equations to determine $G(t)$. Hence we obtain a lower limit on the number of solutions m in a fundamental set, namely,

$$nm \geqslant r. \tag{1.8}$$

In the case of the G_2 groups we have $r = 14$ and $n = 5$ for the case of maximal parabolic subgroups, $n = 6$ for simple SL(3,C) type subgroups, and $n = 8$ for semisimple subgroups like $SL(2,\mathbb{C}) \otimes SL(2,\mathbb{C})$. We shall see that we have $m = 3$ in the first three cases and $m = 2$ in the last. Thus the inequality (1.8) comes as close as possible to being saturated and the system of algebraic equations (1.6) is only slightly overdetermined.

The actual task of reconstructing the group element can be viewed as a problem in algebraic geometry. Thus we can consider the general element of $G_2(\mathbb{C})$ to be represented by a matrix $G(t) \in \mathbb{C}^{7\times 7}$. The 49 matrix elements g_{ik} $(i,k = 1,...,7)$ are subject to 28 orthogonality relations [since we have $G_2(\mathbb{C}) \in O(7,\mathbb{C})$], seven relations expressing the invariance of three-index tensor T, and $n \cdot m$ relations obtained from (1.6). The $35 + n \cdot m$ surfaces in $\mathbb{C}^{7\times 7}$, corresponding to these relations, must intersect in precisely one point (for any fixed t).

The reconstruction procedure in all cases treated in this paper will be essentially the same. The equations to be solved are nonlinear; however, it is always possible to solve them in two steps. In the first step we solve all the equations that are linear in some of the matrix elements g_{ik} and express these elements in terms of known solutions and one or more remaining matrix elements. The second step is nonlinear and consists either of solving a nonlinear algebraic equation for one remaining matrix element, or of finding the eigenvalues and eigenvectors of some given matrix.

The final results and the calculations involved can be quite cumbersome. In these cases we shall only outline the results here and refer for details to Ref. 18.

The paper is organized as follows. In Sec. II we obtain the superposition formulas related to the two different maximal parabolic subgroups of the complex group $G_2(\mathbb{C})$ and also of the noncompact real group $G_2^{NC}(\mathbf{R})$. The simple subgroup $SL(3,\mathbb{C}) \subset G_2(\mathbb{C})$, as well as the corresponding real forms, are treated in Sec. III. Section IV is devoted to the case of the semisimple subgroup $SL(2,\mathbb{C}) \otimes SL(2,\mathbb{C})$, and also to the corresponding real forms.

II. SUPERPOSITION FORMULA RELATED TO MAXIMAL PARABOLIC SUBGROUPS

The complex Lie algebra $g_2(\mathbb{C})$ has two different maximal parabolic subalgebras, denoted p_{α_1} and p_{α_2} in I and II. Both of them are nine dimensional. The homogeneous spaces $G_2(\mathbb{C})/P_{\alpha_1}$ and $G_2(\mathbb{C})/P_{\alpha_2}$ (P_{α_i} are the corresponding maximal parabolic subgroups) are hence five dimensional.

The nonlinear equations corresponding to the action of $G_2(\mathbb{C})$ on $G_2(\mathbb{C})/P_{\alpha_1}$ are a set of five coupled complex conformal Riccati equations [see II, Eq. (2.11)]. These are special cases of the conformal Riccati equations corresponding to the action of $O(7,\mathbb{C})$ on the homogeneous space $O(7,\mathbb{C})/P_1$ (see I and II). For the $G_2(\mathbb{C})/P_{\alpha_2}$ space we have obtained a set of five coupled equations in which the right-hand sides are polynomials of order 4 in the dependent variables [see II, Eq. (2.22)].

The homogeneous spaces $G_2(\mathbb{C})/P_{\alpha_i}$ $(i = 1,2)$ in these two cases are Grassmannians of isotropic i planes in $\mathbb{C}^{7\times i}$. The group $G_2(\mathbb{C})$ is realized as a subgroup of $O(7,\mathbb{C})$, leaving a certain third rank tensor T invariant.

When discussing parabolic subgroups P_{α_i}, it is convenient to use a realization in which $G_2(\mathbb{C})$ is realized as a group of matrices $G \in \mathbb{C}^{7\times 7}$ satisfying

$$G^T J G = J, \quad g_{ab} T_{bcd} = T_{amn} g_{mc} g_{nd}, \tag{2.1a}$$

$$(J)_{ab} = \delta_{a,8-a}, \quad a,b,c,m,n = 1,...,7, \tag{2.1b}$$

$$\begin{aligned}
&T_{bcd} = -T_{bdc} \equiv (T_b)_{cd}, \\
&(T_1)_{14} = -(1/\sqrt{2})(T_1)_{56} = i, \\
&(T_2)_{24} = (1/\sqrt{2})(T_2)_{57} = i, \\
&(T_3)_{34} = -(1/\sqrt{2})(T_3)_{67} = i, \\
&(T_4)_{17} = (T_4)_{26} = (T_4)_{35} = -i, \\
&(T_5)_{45} = -(1/\sqrt{2})(T_5)_{12} = i, \\
&(T_6)_{46} = (1/\sqrt{2})(T_6)_{13} = i, \\
&(T_7)_{47} = -(1/\sqrt{2})(T_7)_{23} = i
\end{aligned} \tag{2.2}$$

(the superscript T denotes transposition).

In keeping with II we shall use both homogeneous and affine coordinates to parametrize the Grassmannians $G_2(\mathbb{C})/P_{\alpha_i}$. In homogeneous coordinates we have

$$\xi = \begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix}, \quad X_1, X_3 \in \mathbb{C}^{i\times i}, \quad X_2 \in \mathbb{C}^{(7-2i)\times i}, \quad i = 1,2, \tag{2.3}$$

$$\xi^T J \xi = 0$$

and two points ξ' and ξ coincide if

$$\xi' = \xi H, \quad H \in \mathrm{GL}(i,\mathbb{C}) . \tag{2.4}$$

In affine coordinates the redundancy (2.4) is removed and we have

$$\eta = \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix},$$

$$Z_1 = X_1 X_3^{-1} \in \mathbb{C}^{i\times i}, \quad Z_2 = X_2 X_3^{-1} \in \mathbb{C}^{(7-2i)\times i}, \tag{2.5}$$

$$Z_1^T J_i + J_i Z_1 = -Z_2^T J_{7-2i} Z_2$$

($J_\mu \in \mathbb{C}^{\mu\times\mu}$ and it has 1's on the antidiagonal and 0's elsewhere).

In homogeneous coordinates we write the superposition formula (1.5) as

$$\begin{pmatrix} X_1(t) \\ X_2(t) \\ X_3(t) \end{pmatrix} = \begin{pmatrix} G_{11}(t) & G_{12}(t) & G_{13}(t) \\ G_{21}(t) & G_{22}(t) & G_{23}(t) \\ G_{31}(t) & G_{32}(t) & G_{33}(t) \end{pmatrix} \begin{pmatrix} X_1(0) \\ X_2(0) \\ X_3(0) \end{pmatrix}, \tag{2.6}$$

where the nonlinearity is introduced by the condition (2.4). In (2.6) $G_{ab}(t)$ is a complex matrix of the appropriate dimension ($1 \leqslant a,b \leqslant 3$).

In affine coordinates the superposition formula is nonlinear, namely,

$$\begin{aligned} Z_1(t) &= [G_{11}Z_1(0) + G_{12}Z_2(0) + G_{13}] \\ &\quad \times [G_{31}Z_1(0) + G_{32}Z_2(0) + G_{33}]^{-1}, \\ Z_2(t) &= [G_{21}Z_1(0) + G_{22}Z_2(0) + G_{23}] \\ &\quad \times [G_{31}Z_1(0) + G_{32}Z_2(0) + G_{33}]^{-1}, \end{aligned} \tag{2.7}$$

where the constant matrices $Z_1(0)$, $Z_2(0)$ are related to the initial conditions.

We recall that the equations are written in affine coordinates and that knowing a solution means knowing its affine coordinates. From them we can get the homogeneous coordinates up to the ambiguity given by (2.4).

Each solution has five independent components and hence, when substituted into (2.6) or (2.7), provides five equations for the 14 independent components of the $G_2(\mathbb{C})$ group elements g_{ik}. Since we have $3\times 5 > 14$, at least three solutions are needed to obtain g_{ik} uniquely and we shall show below that indeed three generically chosen solutions are sufficient, both for the case of P_{α_1} and P_{α_2}.

A. The $G_2(\mathbb{C})/P_{\alpha_1}$ equations

The maximal parabolic subgroup P_{α_1} of $G_2(\mathbb{C})$ leaves a one-dimensional lightlike vector space invariant. The Grassmannian consists of isotropic one-planes in $\mathbb{C}^{7\times 1}$.

1. A fundamental set of solutions

It is convenient to perform most of the reasoning in homogeneous coordinates. In this case we have $X_1, X_3 \in \mathbb{C}$, $X_2 \in \mathbb{C}^{5\times 1}$. Let us assume that we know (up to a nonvanishing scalar factor) the homogeneous coordinates of three solutions $u(t)$, $v(t)$, and $w(t)$. Out of these solutions we can form four $G_2(\mathbb{C})$ invariants, namely the scalar products

$$(u,v) = u^T J v, \quad (u,w) = u^T J w, \quad (v,w) = v^T J w, \tag{2.8}$$

and the trilinear product

$$S = TJuvw = T_{abc}(Ju)_a v_b w_c . \tag{2.9}$$

We proceed to establish the conditions to be imposed on the initial conditions $u(0)$, $v(0)$, and $w(0)$ for their joint isotropy subgroup of $G_2(\mathbb{C})$ to be the identity group. Thus the conditions

$$Gu(0) = \lambda u(0), \quad Gv(0) = \mu v(0), \quad Gw(0) = \nu w(0) , \qquad \lambda,\mu,\nu \in \mathbb{C}, \quad \lambda,\mu,\nu \neq 0 , \tag{2.10}$$

should imply

$$G = \rho I, \quad \rho \in \mathbb{C}, \quad \rho \neq 0 . \tag{2.11}$$

With no loss of generality we can, in view of transitivity, choose

$$u(0) = (0\,0\,0\,0\,0\,0\,1)^T \alpha, \quad \alpha \neq 0 . \tag{2.12}$$

The first of conditions (2.10) determines all off-diagonal elements in the last column of G to be $g_{i7} = 0$, $i = 1,\ldots,6$. The orthogonality condition in (2.1) then yields $g_{1i} = 0$, $i = 2,\ldots,6$, and also $g_{11}g_{77} = 1$. Combining these results with the invariance of T in (2.1) [choosing, e.g., $c = 7$ in (2.1)], we obtain further restrictions on g_{ik}. Finally, the isotropy group of $u(0)$ in (2.12) consists of $G_2(\mathbb{C})$ matrices, satisfying

$$\begin{aligned} &g_{i7} = 0, \quad i = 1,\ldots,6, \quad g_{11}g_{77} = 1 , \\ &g_{ij} = 0, \quad j = 2,\ldots,6 , \\ &g_{42} = g_{52} = g_{62} = 0 , \quad g_{43} = g_{53} = g_{63} = 0 . \end{aligned} \tag{2.13}$$

Now let us choose the initial condition for the second solution to be

$$v(0) = (1\,0\,0\,0\,0\,0\,0)^T \beta, \quad \beta \neq 0 . \tag{2.14}$$

Two constant vectors $\tilde{u}$ and $\tilde{v}$ on the Grassmannian can be transformed into $u(0)$ and $v(0)$ as long as their scalar product satisfies

$$(\tilde{u},\tilde{v}) = (u(0),v(0)) = \alpha\beta \neq 0 . \tag{2.15}$$

Imposing the second of conditions (2.10), and using orthogonality as well as the invariance of the tensor T, we find that the isotropy group of the pair $\{u(0),v(0)\}$ consists of matrices of the form

$$G = \begin{pmatrix} g_{11} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & g_{22} & g_{23} & 0 & 0 & 0 & 0 \\ 0 & g_{32} & g_{33} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & g_{11}g_{22} & -g_{11}g_{23} & 0 \\ 0 & 0 & 0 & 0 & -g_{11}g_{32} & g_{11}g_{33} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & g_{11}^{-1} \end{pmatrix}, \tag{2.16}$$

with

$$g_{11}(g_{22}g_{33} - g_{23}g_{32}) = 1 . \tag{2.17}$$

We take the initial condition for the third solution in the general form

$$w(0) = (w_1(0),w_2(0),\ldots,w_7(0))^T \gamma, \quad \gamma \neq 0 . \tag{2.18}$$

The matrix G of (2.16) will leave (2.18) invariant (up to a factor $\lambda \neq 0$) if the remaining matrix elements g_{ik} satisfy

$$
\begin{aligned}
&(\lambda - g_{11})w_1(0) = 0\,, \quad (\lambda - 1)w_4(0) = 0,\\
&(\lambda - g_{11}^{-1})w_7(0) = 0\,,\\
&(g_{22} - \lambda)w_2(0) + g_{23}w_3(0) = 0,\\
&g_{32}w_2(0) + (g_{33} - \lambda)w_3(0) = 0\,,\\
&(g_{22}g_{11} - \lambda)w_5(0) - g_{11}g_{23}w_6(0) = 0\,,\\
&-g_{11}g_{32}w_5(0) + (g_{11}g_{33} - \lambda)w_6(0) = 0\,.
\end{aligned}
\tag{2.19}
$$

These conditions will imply that $g_{ik} = \delta_{ik}$, i.e., $G = 1$, if and only if the components of $w(0)$ satisfy

$$
\begin{aligned}
&w_4(0) \neq 0, \quad \{w_1(0), w_7(0)\} \neq \{0,0\},\\
&w_2(0)w_6(0) + w_3(0)w_5(0) \neq 0\,.
\end{aligned}
\tag{2.20}
$$

The invariants (2.8) and (2.9) for our choice of initial vectors are

$$
\begin{aligned}
&(u(0),v(0)) = \alpha\beta, \quad (u(0),w(0)) = \alpha\gamma w_1(0)\,,\\
&(v(0),w(0)) = \beta\gamma w_7(0), \quad S = i\alpha\beta\gamma w_4(0)\,.
\end{aligned}
\tag{2.21a}
$$

Moreover, the isotropy condition $w^T(0)Jw(0) = 0$ implies

$$
\begin{aligned}
&w_2(0)w_6(0) + w_3(0)w_5(0)\\
&\quad = -w_1(0)w_7(0) - \tfrac{1}{2}w_4^2(0)\\
&\quad = (1/2\alpha^2\beta^2\gamma^2)[S^2\\
&\qquad - 2(u(0),v(0))(u(0),w(0))(v(0),w(0))]\,.
\end{aligned}
\tag{2.21b}
$$

We arrive at the following result.

Theorem 1: A fundamental set of solutions of the system of nonlinear ordinary differential equations associated with the action of $G_2(\mathbb{C})$ on $G_2(\mathbb{C})/P_{\alpha_1}$ consists of any three particular solutions u, v, and w satisfying the following conditions.

(i) At least two of the scalar products (u,v), (u,w), and (v,w) are nonzero.

(ii) $S = T(Ju)vw \neq 0$.

(iii) $S^2 - 2(u,v)(u,w)(v,w) \neq 0$.

These conditions must be satisfied by the initial data at $t = 0$ to obtain a local superposition formula and for all t for a global one. ■

2. The superposition formula

In order to turn (2.7) into a superposition formula we must now express all the matrix elements $g_{ik}(t)$ in terms of three particular solutions, say $u(t)$, $v(t)$, and $w(t)$, satisfying the conditions of Theorem 1.

The construction follows the outline presented in the Introduction and parallels the scheme used in the proof of Theorem 1.

Using two solutions, say $u(t)$ and $v(t)$, orthogonality and the invariance of the tensor T, we express all elements g_{ik} linearly in terms of g_{11} and g_{ab} with $a,b = 2,3$. We also obtain one nonlinear relation between the remaining five elements. The third solution $w(t)$ is then substituted into (2.7). This provides six more equations: from four of them we express g_{ab} $(a,b = 2,3)$ in terms of g_{11}, still only solving linear equations. One more equation then provides a quartic equation for $g_{11}(t)$. This equation has four different roots, but only one of them satisfies the obligatory relation $g_{11}(0) = 1$. For all details and the final formulas see Ref. 18.

B. The $G_2(\mathbb{C})/P_{\alpha_2}$ equations

The group P_{α_2} leaves a two-dimensional completely isotropic subspace of $\mathbb{C}^7$ invariant. The corresponding homogeneous space $G_2(\mathbb{C})/P_{\alpha_2}$ is a subspace of the Grassmannian of isotropic two-planes in $\mathbb{C}^{7\times 2}$. We shall proceed as in the case of P_{α_1}, remembering, however, that, e.g., the homogeneous coordinates ξ of (2.3) parametrize a matrix in $\mathbb{C}^{7\times 2}$.

1. A fundamental set of solutions

We shall again need three particular solutions. We choose the initial conditions for the first two in the form

$$
u(0) = (0,0,1)^T\alpha(0), \quad v(0) = (1,0,0)^T\beta(0)\,,
$$

$$
\alpha(0),\beta(0) \in GL(2,\mathbb{C})\,. \tag{2.22}
$$

The requirement that a $G_2(\mathbb{C})$ matrix [as in (2.6)] should stabilize $u(0)$ and $v(0)$ implies

$$
G_{13} = 0, \quad G_{23} = 0, \quad G_{21} = 0, \quad G_{31} = 0\,. \tag{2.23}
$$

Invoking the invariance of the tensor T, we find, element by element, that

$$
G_{12} = 0, \quad G_{32} = 0\,, \tag{2.24}
$$

and that the general element of the isotropy group of the two initial condition sets $u(0)$, $v(0)$ of (2.22) is

$$
G = \begin{pmatrix}
g_{55}g_{66} & -g_{55}g_{67} & 0 & 0 & 0 & 0 & 0\\
-g_{55}g_{76} & g_{55}g_{77} & 0 & 0 & 0 & 0 & 0\\
0 & 0 & g_{33} & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 1 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & g_{55} & 0 & 0\\
0 & 0 & 0 & 0 & 0 & g_{66} & g_{67}\\
0 & 0 & 0 & 0 & 0 & g_{76} & g_{77}
\end{pmatrix}, \tag{2.25}
$$

with

$$
g_{55}(g_{66}g_{77} - g_{67}g_{76}) = 1, \quad g_{33}g_{55} = 1\,.
$$

Let us choose the initial conditions for the third solution to be

$$
w(0) = \begin{pmatrix} w_1(0)\\ w_2(0)\\ w_3(0)\end{pmatrix}\gamma(0), \quad \gamma(0) \in GL(2,\mathbb{C})\,. \tag{2.26}
$$

We shall also denote the two columns in $w(0)$ by $w_{a1}(0)$ and $w_{a2}(0)$, $a = 1,\ldots,7$.

Using the three initial conditions "bivectors" we can form the following bilinear $G_2(\mathbb{C})$ invariants

$$
\begin{aligned}
&u^T(0)J_7v(0) = \alpha^T(0)J_2\beta(0)\,,\\
&u^T(0)J_7w(0) = \alpha^T(0)J_2w_1(0)\gamma(0)\,,
\end{aligned}
\tag{2.27}
$$

$$
v^T(0)J_7w(0) = \beta^T(0)J_2w_3(0)\gamma(0)\,. \tag{2.28}
$$

From the stabilization $Gw(0) = w(0)H$, $H \in GL(2,\mathbb{C})$, we get

$$
G_{11}w_1(0) = w_1(0)H\,, \tag{2.29a}
$$

$$
\begin{pmatrix} g_{33} & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & g_{55}\end{pmatrix} w_2(0) = w_2(0)H\,, \tag{2.29b}
$$

$$
G_{33}w_3(0) = w_3(0)H\,. \tag{2.29c}
$$

This system of 14 equations will imply that the point isotropy group of the three initial conditions is $G=1$, if and only if $\det w_1(0)\neq 0$ or $\det w_3(0)\neq 0$ and

$$D_1\equiv w_{31}(0)w_{42}(0)-w_{32}(0)w_{41}(0)\neq 0,$$
$$D_2\equiv w_{31}(0)w_{52}(0)-w_{32}(0)w_{51}(0)\neq 0,$$
$$D_3\equiv w_{41}(0)w_{52}(0)-w_{42}(0)w_{51}(0)\neq 0.$$

In fact, suppose that $\det w_3(0)\neq 0$, then $H=w_3^{-1}(0)\times G_{33}w_3(0)$ and

$$\begin{pmatrix} g_{33} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & g_{55} \end{pmatrix} w_2(0)-w_2(0)w_3^{-1}(0)G_{33}w_3(0)=0, \tag{2.30}$$

which is a linear system of six equations for g_{33}, g_{55}, g_{66}, g_{67}, g_{76}, g_{77} with the unique solution

$$g_{ik}=\delta_{ik}. \tag{2.31}$$

To put all the conditions above in an invariant form that does not depend on the specific choice of u, v, and w, we note that we can construct certain invariants (not necessarily all independent), such as

$$S_{ijk}=T_{abc}J_{aa'}u_{a'i}v_{bj}w_{ck}, \tag{2.32a}$$
$$S'_{ijk}=T_{abc}J_{aa'}u_{a'i}w_{bj}w_{ck},$$
$$a,b,c=1,\ldots,7, \quad i,j,k=1,2, \tag{2.32b}$$
$$S''_{ijk}=T_{abc}J_{aa'}v_{a'i}w_{bj}w_{ck}. \tag{2.32c}$$

Let us choose u, v, and w in the simple form $u=(001)^T$, $v=(100)^T$, and $w=(1,w_2(0),1)^T$. The justification of this choice is that the nonvanishing invariants in (2.32) are given only in terms of $w_2(0)$ as

$$\begin{aligned} S_{121}&=i\omega_{41}(0), \\ S_{122}&=i\omega_{42}(0), \\ S'_{112}&=-S_{121}+\sqrt{2}i\omega_{51}(0), \\ S'_{212}&=-S_{122}+\sqrt{2}i\omega_{52}(0), \\ S''_{112}&=\sqrt{2}i\omega_{31}(0)+S_{121}, \\ S''_{212}&=\sqrt{2}i\omega_{32}(0)-S_{122}. \end{aligned} \tag{2.33}$$

We sum up all the results of this subsection as the following theorem.

Theorem 2: A fundamental set of solutions of the equations associated with the action of $G_2(\mathbb{C})$ on $G_2(\mathbb{C})/P_{\alpha_2}$ consists of three particular solutions u, v, w with initial conditions satisfying the following.

(i) At least two of the determinants $\det(u(0),w(0))$, $\det(u(0),v(0))$, $\det(v(0),w(0))$ are nonzero.

(ii) $$(S''_{112}-S_{121})S_{121}-(S''_{112}+S_{122})S_{121}\neq 0,$$
$$(S''_{112}-S_{121})(S'_{212}-S_{122})-(S''_{212}+S_{122})(S'_{112}+S_{121})\neq 0,$$
$$S_{121}(S'_{212}-S_{122})-S_{122}(S'_{112}+S_{121})\neq 0.$$

2. The superposition formula

The procedure of reconstructing $G(t)\in G_2(\mathbb{C})$ is very similar to the one described for P_{α_1}. Indeed, using two solutions $u(t)$ and $v(t)$, orthogonality, and the invariance of the tensor T, we express all the elements g_{ik} of $G(t)$ linearly in terms of g_{55} and g_{ab} with $a,b=6,7$. The third solution gives all the remaining information: the g_{ab} $(a,b=6,7)$ are then expressed linearly in terms of g_{55} and $g_{55}(t)$ satisfies a cubic equation which admits only one solution such that $g_{55}(0)=1$. For all details see again Ref. 18.

C. Discussion of real cases $G_2^{NC}(\mathbb{R})/P_{\alpha_1}$ and $G_2^{NC}(\mathbb{R})/P_{\alpha_2}$

These cases are completely analogous to the complex cases above, except that all the entries have to be real. We also have to consider a tensor $T'=-iT$ in order to have a tensor with all real entries. The fact that the field of complex numbers $\mathbb{C}$ is algebraically closed, did not play any role above, so the transition to $\mathbb{R}$ poses no difficulty. The notations are the same as in I and II, i.e., $G_2^{NC}(\mathbb{R})$ is the noncompact real form of $G_2(\mathbb{C})$.

III. SUPERPOSITION FORMULAS RELATED TO MAXIMAL SIMPLE SUBGROUPS

This section is devoted to the ODE's related to the action of $G_2(\mathbb{C})$ on the space $G_2(\mathbb{C})/SL(3,\mathbb{C})$, $G_2^C(\mathbb{R})$ on $G_2^C(\mathbb{R})/SU(3)$, $G_2^{NC}(\mathbb{R})$ on $G_2^{NC}(\mathbb{R})/SL(3,\mathbb{R})$ and $G_2^{NC}(\mathbb{R})/SU(2,1)$. The subgroups are of (complex, or real) dimension 8. We are hence dealing with six-dimensional homogeneous spaces and with systems of six equations in each case. The equations spelled out in Ref. 2 are all special cases of projective Riccati equations.

We again use a representation of $G_2(\mathbb{C})$ as a subgroup of $O(7,\mathbb{C})$; however, the matrix J in (2.1a) is chosen to be the identity matrix I_7 when studying $G_2(\mathbb{C})$ or $G_2^C(\mathbb{R})$. When considering the noncompact real form $G_2^{NC}(\mathbb{R})$, we choose $J=J_{4,3}=\mathrm{diag}(1,1,1,1,-1,-1,-1)$. The invariant tensor T is completely antisymmetric in this realization and in agreement with (2.12) of I we take

$$T_{127}=T_{154}=T_{163}=T_{235}=T_{264}=T_{374}=T_{576}=1 \tag{3.1}$$

(all components not related to these by permutations, vanish).

We can again write the superposition formula in the form (2.6) or (2.7). However, for $G_2(\mathbb{C})$ we shall now have

$$X_1,X_3\in\mathbb{C}^{3\times 1}, \quad X_2\in\mathbb{C}^{1\times 1}, \quad Z_1=X_1/X_2, \quad Z_2=X_3/X_2, \tag{3.2a}$$
$$X_1^TX_1+X_2^TX_2+X_3^TX_3=1, \tag{3.2b}$$

and the six components (Z_1,Z_2) are all independent.

Each solution of the equation provides six equations for the 14 independent components of $G(t)$. Since we have $3\times 6>14$, we shall again need at least three solutions to reconstruct $G(t)$ and again three will be sufficient.

A. The $G_2(\mathbb{C})/SL(3,\mathbb{C})$ equations

1. Fundamental set of solutions

We choose the first known solution $u(t)$ to satisfy the initial condition

$$u(0)=(0\,0\,0\,1\,0\,0\,0)^T \tag{3.3}$$

[satisfying (3.2b) for $t=0$ and hence for all t]. The isotropy group of $u(0)$ is characterized by the condition

$$Gu(0)=\lambda u(0),\quad \lambda\in\mathbb{C},\quad \lambda\neq 0. \tag{3.4}$$

This implies $g_{i4}=0$ for $i=1,2,3,5,6,7$. Orthogonality of G, i.e., $G^TG=I$ further implies $g_{4i}=0$, $i=1,2,3,5,6,7$, and $g_{44}^2=1$. Invariance of the tensor T can be used to relate columns 5, 6, and 7 of G to columns 1,2,3, and 4,

$$\begin{aligned} g_{i5} &= -T_{ijk}g_{j1}g_{k4},\\ g_{i6} &= -T_{ijk}g_{j2}g_{k4},\quad g_{i7}=-T_{ijk}g_{j3}g_{k4}. \end{aligned} \tag{3.5}$$

Explicitly we find that the general form of an element of the isotropy group of (3.3) is

$$\begin{aligned} G&=\begin{pmatrix} G_{11} & 0 & -g_{44}G_{31}\\ 0 & g_{44} & 0\\ G_{31} & 0 & g_{44}G_{11}\end{pmatrix},\\ g_{44}&=\pm 1,\quad G_{11},G_{31}\in\mathbb{C}^{3\times 3}. \end{aligned} \tag{3.6}$$

We choose the second solution to satisfy the initial condition

$$v(0)=(\sin\alpha,0,0,\cos\alpha,0,0,0)^T,\quad 0<\alpha<\pi. \tag{3.7}$$

Clearly we have $v(0)^2=1$, $(u(0),v(0))=u(0)^Tv(0)=\cos\alpha$. The requirement that G in (3.6) should belong to the isotropy group of $v(0)$, as well as of $u(0)$, imposes further restrictions. Thus $Gv(0)=\lambda v(0)$, $\lambda\in\mathbb{C}$, $\lambda\neq 0$ implies

$$\begin{aligned} &g_{11}=\lambda,\quad (g_{44}-\lambda)\cos\alpha=0,\\ &g_{21}=g_{31}=g_{51}=g_{61}=g_{71}=0. \end{aligned}$$

Orthogonality of G implies

$$g_{11}^2=1,\quad g_{1i}=0,\quad i=2,3,5,6,7.$$

Further relations are obtained from the invariance of the tensor T:

$$g_{i7}=-T_{ijk}g_{j5}g_{k6}.$$

Finally, an element of the simultaneous isotropy group of $u(0)$ and $v(0)$ has the form

$$G=\begin{pmatrix} g_{11} & 0 & 0 & 0 & 0 & 0 & 0\\ 0 & g_{22} & -g_{11}g_{44}g_{32} & 0 & 0 & -g_{44}g_{62} & -g_{11}g_{72}\\ 0 & g_{32} & g_{11}g_{44}g_{22} & 0 & 0 & -g_{44}g_{72} & g_{11}g_{62}\\ 0 & 0 & 0 & g_{44} & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & g_{11}g_{44} & 0 & 0\\ 0 & g_{62} & g_{11}g_{44}g_{72} & 0 & 0 & g_{44}g_{22} & -g_{11}g_{32}\\ 0 & g_{72} & -g_{11}g_{44}g_{62} & 0 & 0 & g_{44}g_{32} & g_{11}g_{22}\end{pmatrix}, \tag{3.8}$$

$$g_{11}^2=1,\quad g_{44}^2=1,\quad (g_{44}-g_{11})\cos\alpha=0. \tag{3.9}$$

Thus four elements in the second column, as well as the signs of g_{11} and g_{44}, remain to be pinned down. Notice that if we choose $\alpha\neq\pi/2$, we already have $g_{44}=g_{11}$.

Let us choose a third solution $w(t)$ that satisfies the general initial condition

$$\begin{aligned} &w(0)=(w_1(0),w_2(0),\ldots,w_7(0))^T,\\ &w(0)^Tw(0)=1. \end{aligned} \tag{3.10}$$

We can form four invariants out of the vectors $u(0)$, $v(0)$, and $w(0)$, namely,

$$\begin{aligned} &(u(0),v(0))=\cos\alpha,\quad (u(0),w(0))=w_4(0),\\ &(v(0),w(0))=\sin\alpha\, w_1(0)+\cos\alpha\, w_4(0) \end{aligned} \tag{3.11}$$

and

$$S=T_{ijk}u_i(0)v_j(0)w_k(0)=\sin\alpha\, w_5(0). \tag{3.12}$$

The requirement that G of (3.8) should stabilize $w(0)$, i.e., $Gw(0)=\lambda w(0)$ implies

$$\begin{aligned} &(g_{11}-\lambda)w_1(0)=0,\quad (g_{44}-\lambda)w_4(0)=0,\\ &(g_{11}g_{44}-\lambda)w_5(0)=0. \end{aligned} \tag{3.13}$$

Thus if we require that

$$w_5(0)\neq 0 \tag{3.14}$$

and that at least two of the relations

$$w_1(0)\neq 0,\quad w_4(0)\neq 0,\quad \text{or}\quad \cos\alpha\neq 0 \tag{3.15}$$

hold, we obtain

$$g_{11}=g_{44}=1=\lambda. \tag{3.16}$$

Furthermore, the condition $Gw(0)=w(0)$ yields

$$\begin{aligned} &\begin{pmatrix} w_3(0) & w_2(0) & w_7(0) & -w_6(0)\\ w_2(0) & -w_3(0) & -w_6(0) & -w_7(0)\\ w_6(0) & -w_7(0) & w_2(0) & w_3(0)\\ w_7(0) & w_6(0) & -w_3(0) & w_2(0)\end{pmatrix}\\ &\times\begin{pmatrix} g_{22}-1\\ g_{32}\\ g_{62}\\ g_{72}\end{pmatrix}=0. \end{aligned} \tag{3.17}$$

The determinant of this system is

$$D=[w_2(0)^2+w_3(0)^2+w_6(0)^2+w_7(0)^2]^2. \tag{3.18}$$

If $w(0)$ is such that

$$D\neq 0, \tag{3.19}$$

then (3.17) implies $g_{22}=1$, $g_{32}=g_{62}=g_{72}=0$. The iso-

tropy group of $u(0)$, $v(0)$, and $w(0)$ is thus the identity group.

We arrive at the following theorem.

Theorem 3: A fundamental set of solutions of the nonlinear ODE's associated with the action of $G_2(\mathbb{C})$ on $G_2(\mathbb{C})/SL(3,\mathbb{C})$ consists of three particular solutions $u(t)$, $v(t)$, and $w(t)$. When their initial conditions $u(0)$, $v(0)$, and $w(0)$ are given in homogeneous coordinates, they must, in addition to the obvious conditions $u^2 = v^2 = w^2 = 1$, satisfy the following.

(i) At least two of the scalar product (u,v), (u,w), (v,w) are nonzero.

(ii) $S = Tuvw \neq 0$.

(iii) $1 - (u,v)^2 - (u,w)^2 - (v,w)^2 + 2(u,v)(v,w)(w,u) - S^2 \neq 0$. ■

Comments: (1) Conditions (i)–(iii) are generic ones and they must be imposed on the initial conditions at $t = 0$.

(2) Conditions (i) are an invariant formulation of (3.15). Condition (ii) in view of (3.12) assures both $\sin\alpha \neq 0$ and that (3.14) holds [i.e., guarantees that $u(0)$, $v(0)$, and $w(0)$ are linearly independent]. Condition (iii) is an invariant rewriting of (3.19).

(3) The equations are written in affine coordinates and hence so are the solutions, which have six components. The seventh component in homogeneous coordinates is calculated, up to an irrelevant sign, from the normalization conditions (e.g., $u^2 = 1$).

(4) The invariance of the scalar products and S under $G_2(\mathbb{C})$ implies that these expressions are time independent, e.g., $S(t) = S(0)$.

2. The superposition formula

In this (and only this) case we present the final superposition formula explicitly, though we drop most of the derivation. We choose three particular solutions of the nonlinear ODE's in such a manner as to satisfy the conditions of Theorem 3. These solutions are given in affine coordinates and we denote them as

$$\mu(t) = (\mu_i,\mu_a)^T, \quad \nu(t) = (\nu_i,\nu_a)^T, \quad \omega(t) = (\omega_i,\omega_a)^T, \quad i = 1,2,3, \quad a = 5,6,7. \tag{3.20}$$

We define their norm squared by, e.g., $\mu^2 = \mu_1^2 + \mu_2^2 + \mu_3^2 + \mu_5^2 + \mu_6^2 + \mu_7^2$. In homogeneous coordinates we have

$$u(t) = \begin{pmatrix} \mu_i \\ 1 \\ \mu_a \end{pmatrix} \alpha(t), \quad v(t) = \begin{pmatrix} \nu_i \\ 1 \\ \nu_a \end{pmatrix} \beta(t), \quad w(t) = \begin{pmatrix} \omega_i \\ 1 \\ \omega_a \end{pmatrix} \gamma(t), \tag{3.21}$$

where α, β, and γ are not known. We make a specific choice of initial conditions, so as to minimize algebraic complications, namely,

$$u(0) = (0\,0\,0\,1\,0\,0\,0)^T, \quad v(0) = (1/\sqrt{2})(1\,0\,0\,1\,0\,0\,0)^T, \quad w(0) = \tfrac{1}{2}(1\,1\,0\,1\,1\,0\,0)^T, \tag{3.22}$$

satisfying

$$u(0)^2 = v(0)^2 = w(0)^2 = 1. \tag{3.23a}$$

Since relation (3.2a) must hold for all times t, we also have

$$(\mu^2 + 1)\alpha^2 = 1, \quad (\nu^2 + 1)\beta^2 = 1, \quad (\omega^2 + 1)\gamma^2 = 1. \tag{3.23b}$$

The superposition formula can be written in affine coordinates as

$$\xi_i = \frac{\Sigma_{j=1}^3 g_{ij}(t)\xi_j(0) + \Sigma_{a=5}^7 g_{ia}(t)\eta_a(0) + g_{i4}(t)}{\Sigma_{j=1}^3 g_{4j}(t)\xi_j(0) + \Sigma_{a=5}^7 g_{4a}(t)\eta_a(0) + g_{44}(t)}, \quad i = 1,2,3,$$
$$\eta_a = \frac{\Sigma_{j=1}^3 g_{aj}(t)\xi_j(0) + \Sigma_{b=5}^7 g_{ab}(t)\eta_b(0) + g_{a4}(t)}{\Sigma_{j=1}^3 g_{4j}(t)\xi_j(0) + \Sigma_{b=5}^7 g_{4b}(t)\eta_b(0) + g_{44}(t)}, \quad a = 5,6,7. \tag{3.24}$$

Following the outline presented in the Introduction, we obtain

$$\begin{aligned}
g_{\alpha 1} &= (\mu^2+1)^{-1/2}\left(\frac{\mu^2+1}{\mu\cdot\nu+1}\nu_\alpha - \mu_\alpha\right), \\
g_{\alpha 2} &= (\mu^2+1)^{1/2}\left(\frac{1}{\mu\cdot\omega+1}\omega_\alpha - \frac{1}{\mu\cdot\nu+1}\nu_\alpha + (\mu^2+1)^{-1/2}\frac{1}{\mu\cdot\nu+1}T_{\alpha\beta\gamma}\nu_\beta\mu_\gamma\right), \\
g_{\alpha 3} &= \frac{(\mu^2+1)^{1/2}}{\mu\cdot\nu+1}T_{\alpha\rho\sigma}T_{\alpha\beta\gamma}\nu_\beta\mu_\gamma \times\left(\frac{1}{\mu\cdot\omega+1}\omega_\rho - \frac{1}{\mu\cdot\nu+1}\nu_\rho\right) + \frac{1}{(\mu\cdot\nu+1)^2}T_{\alpha\mu\nu}T_{\nu\beta\gamma}T_{\mu\rho\sigma}\nu_\beta\nu_\rho\mu_\gamma\mu_\sigma, \\
g_{\alpha 4} &= (\mu^2+1)^{-1/2}\mu_\alpha, \\
g_{\alpha 5} &= [1/(\mu\cdot\nu+1)]T_{\alpha\beta\gamma}\mu_\beta\nu_\gamma, \\
g_{\alpha 6} &= \frac{1}{\mu\cdot\omega+1}T_{\alpha\beta\gamma}\mu_\beta\omega_\gamma + \frac{1}{\mu\cdot\nu+1}T_{\alpha\beta\gamma}\nu_\beta\mu_\gamma - (\mu^2+1)^{-1/2}\frac{1}{\mu\cdot\nu+1}T_{\alpha\mu\rho}T_{\mu\beta\gamma}\nu_\beta\mu_\gamma\mu_\rho, \\
g_{\alpha 7} &= -\frac{\mu^2+1}{(\mu\cdot\nu+1)(\mu\cdot\omega+1)}T_{\alpha\beta\gamma}\nu_\beta\omega_\gamma + \frac{1}{\mu\cdot\omega+1}T_{\alpha\beta\gamma}\mu_\beta\omega_\gamma - \frac{1}{\mu\cdot\nu+1}T_{\alpha\beta\gamma}\mu_\beta\nu_\gamma - \frac{(\mu^2+1)^{-1/2}}{(\mu\cdot\nu+1)} \times T_{\alpha\beta\gamma}T_{\gamma\rho\sigma}\left(\frac{\mu^2+1}{\mu\cdot\nu+1}\nu_\beta - \mu_\beta\right)\nu_\rho\mu_\sigma.
\end{aligned} \tag{3.25}$$

B. The $G_2^c(\mathbb{R})/SU(3)$ equations

The equations, solutions, and superposition formulas coincide with those given above for $G_2(\mathbb{C})/SL(3,\mathbb{C})$, except that all entries are real [in the superposition formulas (3.24), etc.].

C. The $G_2^{NC}(\mathbb{R})/SL(3,\mathbb{R})$ equations

Instead of the metric given by the identity matrix I_7, we use the metric $I_{4,3} = \mathrm{diag}(1,1,1,1,-1,-1,-1)$. We must hence transform the G_2 invariant tensor T. Within $G_2(\mathbb{C})$ the appropriate transformation is

$$I_{4,3} = H^T I_7 H, \quad T' = H^{-1} THH,$$
$$H = \mathrm{diag}(1,1,1,1,i,i,i)\,. \quad (3.26)$$

Introducing the matrices $T_i = -T_i^T \in \mathbb{C}^{7\times 7}$, such that $(T_i)_{jk} = T_{ijk}$, $(i,j,k = 1,\ldots,7)$, we find that (3.26) implies $T_j' = iT_j$ for $j = 1,2,3,4$ and $T_k' = -iI_{4,3}T_k$ $(k = 5,6,7)$. For $G_2^{NC}(\mathbb{R})$ it is preferable to use a real tensor, so we put $T'' = -iT'$, and obtain

$$T_j'' = T_j, \quad j = 1,\ldots,4, \quad T_k'' = -I_{4,3}T_k \quad (k = 5,6,7)\,. \quad (3.27)$$

All the results obtained for $G_2(\mathbb{C})/SL(3,\mathbb{C})$ can now be carried over to $G_2^{NC}(\mathbb{R})/SL(3,\mathbb{R})$, except that the equations, solutions, and matrix elements g_{ik} are now real. All scalar products must be interpreted in terms of the appropriate metric, e.g., $(\mu,\nu) = \mu_1\nu_1 + \mu_2\nu_2 + \mu_3\nu_3 - \mu_5\nu_5 - \mu_6\nu_6 - \mu_7\nu_7$ and the tensor T is replaced by T'' as in (3.27). The three fundamental solutions can be chosen as in (3.20) and (3.21) [i.e., we have, e.g., $u^2(t) = 1$].

D. The $G_2^{NC}(\mathbb{R})/SU(2,1)$ equations

The group $SU(2,1)$ leaves a negative length vector in $\mathbb{R}^{7\times 1}$ invariant, hence $G_2^{NC}(\mathbb{R})/SU(2,1)$ is realized, in homogeneous coordinates, in terms of vectors $x \in \mathbb{R}^{7\times 1}$, satisfying, e.g., $x^2 = x^T I_{4,3}\, x = -1$. The fundamental set of solutions must be chosen appropriately. Without proof we state the theorem that is an adaptation of Theorem 3 and is proved in a similar manner.

Theorem 4: A fundamental set of solutions of the nonlinear ODE's associated with the action of $G_2^{NC}(\mathbb{R})$ on $G_2^{NC}(\mathbb{R})/SU(2,1)$ consists of three particular solutions $u(t)$, $v(t)$, and $w(t)$. In homogeneous coordinates we choose the initial conditions to satisfy $u^2 = v^2 = w^2 = -1$, and the following invariant conditions.

(i) At least two of the scalar products (u,v), (u,w), and (v,w) are nonvanishing.

(ii) $S = T''(I_{4,3}\,u)vw \neq 0$.

(iii) $-1 + (u,v)^2 + (u,w)^2 + (v,w)^2 + 2(u,v)(u,w)(v,w) + S^2 \neq 0$. ■

An example of a suitable choice replacing (3.22) is

$$u(0) = (0\,0\,0\,0\,0\,0\,1)^T,$$
$$v(0) = (1/\sqrt{2})\,(0\,0\,0\,0\,1\,0\,1)^T, \quad (3.28)$$
$$w(0) = \tfrac{1}{2}\,(1\,0\,0\,0\,1\,1\,1)^T.$$

The superposition formula can be derived in exactly the same manner as in the complex case and will be very similar. We do not present it here.

IV. SUPERPOSITION FORMULAS RELATED TO MAXIMAL SEMISIMPLE SUBGROUPS

The spaces to be considered in this section are $G_2(\mathbb{C})/[SL(2,\mathbb{C})\times SL(2,\mathbb{C})]$, $G_2^C(\mathbb{R})/[SU(2)\times SU(2)]$, $G_2^{NC}(\mathbb{R})/[SU(2)\times SU(2)]$, and $G_2^{NC}(\mathbb{R})/[SU(1,1)\times SU(1,1)]$. We make use of the metric I_7 in the first two cases, $I_{4,3}$ in the last two. The dimension of these spaces is $d = 8$ (complex in the first case, real in the other three). We are hence dealing with eight equations and each solution has eight independent components. Since we have $2\times 8 > 14$, two solutions provide more than enough equations for $g_{ik}(t)$. We shall show that in this case a fundamental set of solutions consists of just two solutions.

In homogeneous coordinates each solution is written as a "trivector" $u \in F^{7\times 3}$ $(F = \mathbb{C}$ or $F = \mathbb{R})$:

$$u = \begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix},$$
$$u \sim \tilde{u} = uH, \quad H \in GL(3,F), \quad X_1 X_3 \in F^{3\times 3}, \quad X_2 \in F^{1\times 3}, \quad (4.1)$$

where u and $\tilde{u}$ represent the same point in the homogeneous space. The redundancy inherent in (4.1) is removed by using affine coordinates

$$W_1 = X_1 X_3^{-1}, \quad W_2 = X_2 X_3^{-1}. \quad (4.2)$$

We can choose the normalization to be such that we have $u^T u = I$, i.e.,

$$X_1^T X_1 + X_2^T X_2 + X_3^T X_3 = I_3\,. \quad (4.3)$$

A. The $G_2(\mathbb{C})/[SL(2,\mathbb{C})\times SL(2,\mathbb{C})]$ equations

1. A fundamental set of solutions

With no loss of generality we choose the first known solution $u(t)$ to satisfy the initial condition

$$u(0) = \begin{pmatrix} 0 \\ 0 \\ I_3 \end{pmatrix} \quad (4.4)$$

satisfying (4.3). We write an element of $G_2(\mathbb{C})$ in the form

$$G = \begin{pmatrix} G_{11} & G_{12} & G_{13} \\ G_{21} & G_{22} & G_{23} \\ G_{31} & G_{32} & G_{33} \end{pmatrix},$$
$$G_{11}, G_{13}, G_{31}, G_{33} \in \mathbb{C}^{3\times 3}, \quad G_{12}, G_{32}, G_{21}^T, G_{23}^T \in \mathbb{C}^{3\times 1},$$
$$G_{22} \equiv g_{44} \in \mathbb{C}. \quad (4.5)$$

We shall also use the notation

$$G = \{g_{ik}\}, \quad i,k = 1,\ldots,7\,, \quad (4.6)$$

thus, e.g., the matrix elements of G_{13} are g_{ik} with $i = 1,2,3$, $k = 5,6,7$.

The isotropy group of $u(0)$ consists of matrices $G \in G_2(\mathbb{C})$ satisfying

$$Gu(0) = u(0)H, \quad H \in GL(3,\mathbb{C})\,, \quad (4.7)$$

which implies

$$G_{13} = 0, \quad G_{23} = 0, \quad H = G_{33}\,.$$

Orthogonality $G^T G = I_7$ further implies

$$G_{31} = 0, \quad G_{32} = 0\,, \quad (4.8)$$
$$G_{11}^T G_{11} + G_{21}^T G_{21} = I_3\,, \quad G_{12}^T G_{12} + G_{22}^T G_{22} = 1\,,$$
$$G_{11}^T G_{22} + G_{12}^T G_{12} = 0, \quad G_{33}^T G_{33} = I_3\,. \quad (4.9)$$

The invariance of tensor T makes it possible to express columns 1, 2, and 3 in terms of columns 4, 5, 6, and 7:

$$g_{a1} = -T_{aef}g_{e4}g_{f5},$$
$$g_{a2} = -T_{aef}g_{e4}g_{f6}, \quad g_{a3} = -T_{aef}g_{e4}g_{f7}. \tag{4.10}$$

Using (3.1) to make (4.10) explicit, we find that the isotropy group of $u(0)$ in (4.4) consists of matrices of the form

$$G = \begin{pmatrix} (g_{44}I_3 - R)G_{33} & G_{12} & 0 \\ -G_{12}^T G_{33} & g_{44} & 0 \\ 0 & 0 & G_{33} \end{pmatrix}, \tag{4.11a}$$

with

$$R = \begin{pmatrix} 0 & g_{34} & -g_{24} \\ -g_{34} & 0 & g_{14} \\ g_{24} & -g_{14} & 0 \end{pmatrix},$$
$$G_{12}^T G_{12} + g_{44}^2 = 1, \quad G_{33}^T G_{33} = I_3. \tag{4.11b}$$

Thus G depends on six independent parameters: three in G_{33}, three in G_{12}.

We take the initial data for the second solution $v(t)$ in the form

$$v(0) = \begin{pmatrix} X \\ 0 \\ Y \end{pmatrix},$$
$$v^T(0)v(0) = I_3, \quad \det Y \neq 0,$$
$$X, Y \in \mathbb{C}^{3\times 3}, \quad \det X \neq 0. \tag{4.12}$$

The joint isotropy group of $u(0)$ and $v(0)$ consists of matrices of the form (4.11) satisfying $Gv(0) = v(0)H$, $H \in \mathrm{GL}(3,\mathbb{C})$, i.e., such that

$$(g_{44}I_3 - R)G_{33}X' = X'G_{33},$$
$$X' = XY^{-1}, \quad G_{12}^T G_{33}X' = 0. \tag{4.13}$$

From (4.13) we find $G_{12} = 0$ and thus

$$G_{12} = 0, \quad R = 0, \quad g_{44} = 1. \tag{4.14}$$

The as yet unspecified orthogonal matrix G_{33} satisfies

$$G_{33}X' = X'G_{33}. \tag{4.15}$$

Let us now decompose X' into its symmetric and antisymmetric parts,

$$X' = S + A, \quad S = S^T, \quad A = -A^T, \tag{4.16}$$

where, with no loss of generality, we can assume that S is diagonal. If all three eigenvalues of S are different and if at least two of the matrix elements a_{12}, a_{13}, a_{23} of A are nonzero, then (4.15) implies

$$G_{33} = \lambda I_3, \quad \lambda = \pm 1. \tag{4.17}$$

By continuity we can impose $\lambda = 1$, and thus the isotropy group of the two initial data trivectors $u(0)$ and $v(0)$ is the identity group.

We thus arrive at the following theorem.

Theorem 5: A fundamental set of solutions of the nonlinear ODE's associated with the action of $G_2(\mathbb{C})$ on $G_2(\mathbb{C})/[\mathrm{SL}(2,\mathbb{C}) \otimes \mathrm{SL}(2,\mathbb{C})]$ consists of two solutions $u(t)$ and $v(t)$, satisfying certain independence conditions. In homogeneous coordinates the two solutions can be chosen to correspond to the initial conditions

$$u(0) = \begin{pmatrix} 0 \\ 0 \\ I_3 \end{pmatrix}, \quad v(0) = \begin{pmatrix} X \\ 0 \\ Y \end{pmatrix},$$
$$X, Y \in \mathbb{C}^{3\times 3}, \quad \det X \neq 0, \quad \det Y \neq 0, \tag{4.18}$$
$$XY^{-1} = \begin{pmatrix} \lambda_1 & \alpha & \beta \\ -\alpha & \lambda_2 & \gamma \\ -\beta & -\gamma & \lambda_3 \end{pmatrix}, \quad X^TX + Y^TY = I_3,$$

where at least two of the numbers α, β, γ are nonzero and $\lambda_1 \neq \lambda_2 \neq \lambda_3 \neq \lambda_1$. ■

Comments: (i) The initial conditions (4.18) are not the most general ones possible for a fundamental set of solutions. A trivial generalization is obtained by applying an arbitrary constant $G_2(\mathbb{C})$ transformation to $u(0)$ and $v(0)$ as given in (4.18).

(ii) The conditions on $u(0)$ and $v(0)$ could be formulated in an invariant manner, e.g., by requiring $u^Tu = v^Tv = I$, $\det u^Tv \neq 0$ and imposing conditions on the quantity $S_{a,ik} = T_{abc}u_{bi}u_{ck}$ $(a,b,c = 1,\ldots,7;\ i,k = 1,2,3)$, but we shall not go into this here.

2. The superposition formula

The superposition formula, in terms of affine coordinates, is

$$W_1(t) = [G_{11}W_1(0) + G_{12}W_2(0) + G_{13}] \times [G_{31}W_1(0) + G_{32}W_2(0) + G_{33}]^{-1},$$
$$W_2(t) = [G_{21}W_1(0) + G_{22}W_2(0) + G_{23}] \times [G_{31}W_1(0) + G_{32}W_2(0) + G_{33}]^{-1}, \tag{4.19}$$

where the G_{ik} are as in (4.5). The constant matrices $W_1(0)$, $W_2(0)$ in (4.19) provide the initial conditions of Theorem 5.

The reconstruction of the elements of $G_{ik}(t)$ follows the outline presented in the Introduction. In fact, using the two particular solutions $W_1(t)$ and $W_2(t)$, orthogonality and invariance of the tensor T, all the submatrices G_{ik} can be expressed linearly in terms of $G_{33}(t)$. The equation satisfied by $G_{33}(t)$ is of the form

$$G_{33}W_2(0)G_{33}^{-1} = Y, \tag{4.20}$$

where Y is completely known in terms of the particular solutions and can be solved to determine uniquely G_{33} by using the condition $G_{33}(0) = I_3$. For details see again Ref. 18.

B. Discussion of the real cases

The equations related to the compact space $G_2^C(\mathbb{R})/[\mathrm{SU}(2)\times\mathrm{SU}(2)]$ coincide with those for the complex case, except that all coefficients and solutions are real. The same holds for the superposition formulas: they coincide with (4.20).

The $G_2^{NC}(\mathbb{R})/[\mathrm{SU}(2)\times\mathrm{SU}(2)]$ equations require a very slight modification with the respect to the complex case: the tensor T should be replaced by T'' as in (3.27), I_7 by $I_{4,3}$.

The $G_2^{NC}(\mathbb{R})/[\mathrm{SU}(1,1)\times\mathrm{SU}(1,1)]$ case is somewhat different, in that homogeneous coordinates are introduced as

$$u=\begin{pmatrix} X_1 \\ Y_1 \\ X_2 \\ Y_2 \end{pmatrix},$$

$$X_1, X_2, Y_1 \in \mathbb{R}^{2\times 3}, \quad Y_2 \in \mathbb{R}^{1\times 3}, \quad u^T u = I_{2,1}, \tag{4.21}$$

$$Y=\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix}.$$

Affine coordinates are

$$W_1 = X_1 Y^{-1}, \quad W_2 = X_2 Y^{-1}, \quad \det Y \neq 0. \tag{4.22}$$

Again the tensor T'' is used instead of T, the metric is $I_{4,3}$. The reconstruction of the group element $G \in G_2^{NC}(\mathbb{R})$ must take the different subdivision of u into account, but proceeds along lines analogous to the complex case.

V. CONCLUSIONS

One conclusion to be drawn from this paper is that methods developed for obtaining nonlinear ODE's with superposition formulas, as well as the superposition formulas themselves, can be generalized from the classical Lie groups, to the exceptional simple Lie groups. In the present series we have treated the exceptional group $G_2(\mathbb{C})$ as well as its two real forms, $G_2^C(\mathbb{R})$ and $G_2^{NC}(\mathbb{R})$. The ODE's correspond to the action of the group G_2 on G_2/G, where G is some maximal subgroup of G_2. If G is parabolic or simple, the superposition formulas uses three particular solutions to express the general one. If G is semisimple, but not simple, two particular solutions suffice.

In our overall discussion of nonlinear ODE's with superposition formulas, some results are basis independent, others depend on a choice of coordinates. If we are given a system of equations of the type (1.3), we can immediately read off the vector fields X_k of (1.4) and determine whether they generate a finite dimensional Lie algebra. If they do, then this algebra L is determined in a nonambiguous manner, as is the subalgebra L_0 of vector fields vanishing at some chosen origin. The Lie algebras L and L_0 then determine the Lie groups G and $G_0 \subset G$, which completely specify the manifold $M \sim G/G_0$ and the action (1.5) of G and M.

Our approach is a complementary one. We choose a pair of Lie algebras L and L_0 such that they determine a transitive primitive Lie algebra, in order to obtain an indecomposable system of equations.[4] Once the choice is made the groups G and G_0, as well as the homogeneous space $M \sim G/G_0$, are completely determined. This determines the number of equations $n = \dim G - \dim G_0$. The number of solutions m that constitute a fundamental set is also invariantly defined.

The actual form of the considered equations depends on a choice of coordinates in M, and is highly nonunique. To see this, let G be $SL(2,\mathbb{R})$ and G_0 its two-dimensional affine subgroup. The usual coordinates on G/G_0 in this case lead to the vector fields

$$\left\{\frac{d}{dy}, y\frac{d}{dy}, y^2\frac{d}{dy}\right\}$$

and hence to the Riccati equation

$$\dot{y}(t) = Z_1(t) + Z_2(t)y + Z_3(t)y^2.$$

This equation represents an infinite class of equations, associated with the same G/G_0 and obtained by putting $y = \phi(u)$, where ϕ is any locally invertible function. For instance, choosing $y = \tan(u/2)$, we obtain a "disguised" Riccati equation

$$\dot{u}(t) = Z_1 + Z_3 + (Z_1 - Z_3)\cos u + Z_2 \sin u.$$

Similarly, each of the equations treated in this series, where we put $G = G_2$ and G_0 runs through maximal subgroups of G_2, represents an infinite class of equations, each member of which corresponds to a chosen coordinate system.

A final comment is that for $G_2(\mathbb{C})$ we have discussed four nonequivalent systems of equations, namely, (2.11), (2.22), (3.9), and (4.7) of II (with constraints discussed in II). The equations corresponding to the maximal parabolic subalgebra P_{α_1} have quadratic nonlinearities and are a special case of equations associated with the space $O(7,\mathbb{C})/SIM(5,\mathbb{C})$, where $SIM(5,\mathbb{C})$ is a maximal parabolic subalgebra of $O(7,\mathbb{C})$ (see I and II). However, for the general $O(7,\mathbb{C})$ equations we would need six solutions to obtain a superposition formula, whereas for G_2 we need only three.

The $G_2(\mathbb{C})/P_{\alpha_2}$ equations have quartic nonlinearities whereas the corresponding $O(7,\mathbb{C})/OPT(5,\mathbb{C})$ equations have quadratic ones. Moreover, the number of equations is five and seven, respectively.

The $G_2(\mathbb{C})/SL(3,\mathbb{C})$ equations have quadratic nonlinearities and are a special case of $O(7,\mathbb{C})/O(6,\mathbb{C})$ equations, which are in turn a special case of projective Riccati equations for $SL(7,\mathbb{C})/Aff(n,\mathbb{C})$. The number of solutions in a fundamental set is $m = 7$, $m = 4$, and $m = 3$ for $SL(7,\mathbb{C})$, $O(7,\mathbb{C})$, and $G_2(\mathbb{C})$, respectively.

Finally, for $G_2(\mathbb{C})/[SL(2,\mathbb{C}) \otimes SL(2,\mathbb{C})]$ we have eight equations with nonpolynomial nonlinearities. The embedding into $O(7,\mathbb{C})$ gives 12 equations, associated with the space $O(7,\mathbb{C})/[O(4,\mathbb{C}) \otimes O(3,\mathbb{C})]$, and the nonlinearities are quadratic.

ACKNOWLEDGMENTS

We thank Professor J. Beckers, who participated in a part of this study, for helpful discussions.

The research of one of the authors (P.W.) is partially supported by research grants from NSERC of Canada and the "Fonds FCAR du Gouvernement du Québec." Another author (L.G.) acknowledges a doctoral fellowship from FCAR. Financial support from the "Accords culturels Belgique-Québec" is much appreciated.

[1] J. Beckers, V. Hussin, and P. Winternitz, J. Math. Phys. **27**, 2217 (1986).
[2] J. Beckers, V. Hussin, and P. Winternitz, J. Math. Phys. **28**, 520 (1987).
[3] S. Lie and G. Scheffers, *Vorlesungen über continuierlichen Gruppen mit geometrischen und anderen Anwendungen* (Teubner, Leipzig, 1893), see also (Chelsea, New York, 1967).
[4] S. Shnider and P. Winternitz, J. Math. Phys. **25**, 3155 (1984); Lett. Math. Phys. **8**, 69 (1984).
[5] R. L. Anderson, Lett. Math. Phys. **4**, 1 (1980).
[6] R. L. Anderson, J. Harnad, and P. Winternitz, Lett. Math. Phys. **5**, 143 (1981); Physica D **4**, 164 (1982).
[7] J. Harnad, P. Winternitz, and R. L. Anderson, J. Math. Phys. **24**, 1062 (1982).

[8]M. del Olmo, M. A. Rodríguez, and P. Winternitz, J. Math. Phys. **27**, 14 (1986); **28**, 530 (1987).
[9]D. Rand and P. Winternitz, Comput. Phys. Comm. **33**, 305 (1984).
[10]M. Sorine and P. Winternitz, IEEE Trans. Automat. Control **AC-30**, 266 (1985).
[11]P. Winternitz, in *Nonlinear Phenomena, Lecture Notes in Physics*, Vol. 189, edited by K. B. Wolf (Springer, Berlin, 1983), pp. 265–331.
[12]J. Beckers, L. Gagnon, V. Hussin, and P. Winternitz, Lett. Math. Phys. **13**, 113 (1987).
[13]P. J. Vassiliou, Acta Applic. Math. **8**, 107, 149 (1987).
[14]L. L. Chau, in *Nonlinear Phenomena, Lecture Notes in Physics*, Vol. 189, edited by K. B. Wolf (Springer, Berlin, 1983), pp. 110–127.
[15]A. T. Ogielski, M. K. Prasad, A. Sinha, and L. L. Chau Wang, Phys. Lett. B **91**, 387 (1980).
[16]G. C. McVittie, Ann. Inst. H. Poincaré **40** 3, 231 (1984).
[17]J. S. R. Chisholm and A. K. Common, J. Phys. A **20**, 5459 (1987).
[18]L. Gagnon, V. Hussin, and P. Winternitz, CRM-1467, Montreal, 1987.

Superspace first-order partial differential equations through the Cartan–Kähler integration theorem

Roberto Cianci
Dipartimento di Matematica, Università di Genova, Genova, Italy and Dipartimento di Matematica, Università di Milano, Milano, Italy

(Received 27 October 1987; accepted for publication 18 May 1988)

In the paper, first-order partial differential equations are studied on superspace; the key point is the use of a suitable generalization of the Cartan–Kähler integration theorem on superspace. By means of this result, it is possible to investigate the characteristic fields and the structure of the solutions.

I. INTRODUCTION

The aim of this contribution is, first of all, to show how the Cartan–Kähler integration theorem[1] can be useful in studying particular topics in the framework of supermanifold theory.

Among these, we mention the problem of partial differential equations on superspace (we say superspace since the treatment is essentially local). On the other hand, since the Cartan–Kähler integration theorem is well suited to study the *extensions* of an integral manifold to another one of greater dimension, we see applications to the so-called rheonomy problem.[2] Finally, in this framework we can study *initial data problems*, thus proceeding toward the characterization of an *evolution* problem in some generalized sense.

As a final remark, we wish to point out that the techniques studied here for first-order partial differential equations can be fruitfully used for a large class of physically interesting *systems of partial differential equations.*

The main techniques used to study such superspace problems are essentially the following.

(a) The superfield expansion in powers of the odd θ coordinates is

$$f(x,\theta)=\sum_{\mu} f_{\mu}(x)\theta^{\mu}$$

$$=f_{\phi}(x)+f_{(1)}(x)\theta^{1}+\cdots+f_{(1,\ldots,n)}\theta^{1}\cdots\theta^{n}.$$

This technique, by transforming the problem of finding a superspace function into one of determining a set of functions of even coordinates, is very powerful, but it only partially solves our problem and does not clarify the structure of the solutions.

(b) There are Banach analysis techniques that are specialized by taking into account the structure of the sheaf of supersmooth functions; in this framework we find the generalized version of the Frobenius theorem.[3] These results are applicable only in same cases.

(c) Finally, in this paper we show how a suitable generalization of the Cartan–Kähler integration theorem[4] can be used for this scope.

Let us sum up the main advantages of this last technique.

(i) The method is of practical use. The solutions are explicitly found by using the Cauchy–Kowalewsky theorem; this technique requires only the existence of algebraic operations already well defined in view of the Banach algebra structure of the ground exterior algebra of the superspace.

(ii) There is a deep linkage with the problem of extending a given integral manifold to another of greater dimension; this result thus furnishes a generalization of the Cauchy evolution problem.

(iii) Following this method one can try to extend the remarkable theory of Guillemin for an overdetermined differential system.[5] Briefly, this author was able to decompose a system of partial differential equations into (under)determined operators that satisfy suitable commutation relations; the so-called Guillemin's normal form arises.

Further results on this topic were obtained by Goldschmidt, Quillen, Spencer, and Sternberg. A very good definition of characteristic vectors for differential systems was given. Even though in this paper we shall consider only first-order partial differential equations, this fact is also very useful in supermanifold theory.[5]

The limits of this analysis are (i) only the local behavior of the solutions can be studied; (ii) we are not able to exit from the framework of superanalytic functions [however, recent developments in this direction seem to give ideas for solving this problem (Yang[5])]; (iii) it is often difficult to get an explicit formula for the complete integral; and (iv) there are cases in which the Cartan–Kähler theorem cannot be applied; these, of course, cannot be discussed.

Concerning the definition of superspace, in this paper we use an infinitely generated exterior algebra, namely a Banach Grassmann algebra Q,[6] playing the role of the basic structure to construct the superspace. This algebra is a real, Z_2-graded commutative Banach algebra Q satisfying the following properties.

(a) Given $Q=Q_0\oplus Q_1$ with $Q_0=\mathbf{R}\oplus Q_0'$; denote by σ: $Q_0\to\mathbf{R}$ the *body* map.

(b) For each continuous Q_0 linear map f: $Q_r\to Q_s$ there exists a unique element $v\in Q_{r+s}$ such that $f(u)=vu$, for all $u\in Q_r$.

These properties are satisfied by B^{∞}, the inductive limit for $L\to\infty$ of the Grassman algebras B^L over $\mathbf{R}^L$.

Now we construct the vector superspace, vss for short, as the Banach Q_0 module $Q^{m,n}=(Q_0)^m\times(Q_1)^n$. Even if we are not interested in a complete construction of supermani-

 0022-2488/88/102156-06$02.50

folds we have to specify the sheaf of functions we shall consider. A function $f: Q^{m,n} \to Q^{m,n}$ will be called *supersmooth* if it is C^∞ and if its Fréchet differential is Q_0 linear:

$$Df_x(ph) = pDf_x(h),$$

for each $x \in Q^{m,n}$, $h \in Q^{m,n}$, and $p \in Q_0$.

A supersmooth function is called superanalytic (sa for short) if it is analytic, too. A vector $x \in Q^{m,m}$ will be denoted with $x = x^A = (x^i, \theta^\alpha)$, $i = 1,...,m$, $\alpha = 1,...,n$, $A = 1,...,m+n$. In the literature there are different definitions of superspace[7] related to different choices of the algebra Q and of the sheaf of functions used. In my opinion, the results obtained here would also be true when adopting these different definitions.

II. THE MAIN THEOREMS

First of all, analyze the problem in the classical framework of real analysis. Consider the following first-order differential equation on a real n-dimensional vector space $\mathbf{V}$:

$$F\left(x^i, \phi, \frac{\partial \phi}{\partial x^i}\right) = 0, \quad i = 1,...,n. \tag{2.1}$$

Following Cartan we construct the differentiable ideal I:

$$\begin{aligned} &F(x^i, \phi, p_i) = 0, \\ &dx^i \frac{\partial F}{\partial x^i} + d\phi \frac{\partial F}{\partial \phi} + dp_i \frac{\partial F}{\partial p_i} = 0, \\ &d\phi - dx^i p_i = 0, \quad dx^i \wedge dp_i = 0. \end{aligned} \tag{2.2}$$

We have considered both zero-forms $\omega^{(0)}$, one-forms $\omega^{(1)}$, and two-forms $\omega^{(2)}$ on the vector bundle $\pi: E \to \mathbf{V}$ whose standard fiber $\pi^{-1}(x^i)$ is isomorphic to $\mathbf{R} \times \mathbf{V}^*$.

Definition 2.1: A solution of the differential ideal I is a section Γ of E satisfying $\Gamma^* I = 0$.

Let $\pi_S: S \to \mathbf{U}$ be a reduced vector subbundle of $\pi: E \to \mathbf{V}$ with the same standard fiber $\mathbf{R} \times \mathbf{V}$. Also let $u = \dim \mathbf{U}$, and let $i: S \to E$ and $i': \mathbf{U} \to \mathbf{V}$ be the related homomorphisms.

Definition 2.2: A u-dimensional integral manifold of the differentiable ideal I is a section σ of the vector subbundle S s.t. $\sigma^* i^* I = 0$.

Let us sum up the main ideas of Cartan's technique by means of the following steps.

Step 1: Choose a point $q \in E$ such that $F(q) = 0$ and where the rank s_0 of the algebraic system $\omega^{(1)}$ is locally constant. Then determine a vector $\mathbf{v} \in T_q E$,

$$\mathbf{v} = \left(a^i \frac{\partial}{\partial x^i} + b \frac{\partial}{\partial \phi} + c_i \frac{\partial}{\partial p_i}\right)_q,$$

satisfying the algebraic system $\mathbf{v} \lrcorner \omega^{(1)} = 0$. The related representative matrix A is

$$A = \begin{pmatrix} -p_i & 1 & 0 \\ \partial F/\partial x^i + p_i\, \partial F/\partial \phi & 0 & \partial F/\partial p^i \end{pmatrix}.$$

Let $s_0(q)$ be the rank of A and $s_0'(q)$ be the rank of the *incomplete* A matrix, that is, the matrix A' obtained by erasing the first n columns:

$$A' = \begin{pmatrix} 1 & 0 \\ 0 & \partial F/\partial p^i \end{pmatrix}.$$

Now, we can find the vector $\mathbf{v}$ with *arbitrary* components $\pi_* \mathbf{v} \in T_{\pi(q)} \mathbf{V}$ if $s_0(q) = s_0'(q) \leqslant \dim E - 1$. If, furthermore, such ranks are locally constant, then there exists a one-dimensional integral manifold $i_1(s)$ of I s.t. $i(0) = q$ and $(d/ds) i_1(s)|_{s=0} = \mathbf{v}$.

Step 2: Find a vector $u \in T_q E$ solving the *polar* system

$$\mathbf{u} \lrcorner \omega^{(1)} = 0, \quad \mathbf{u} \lrcorner (\mathbf{v} \lrcorner \omega^{(2)}) = 0. \tag{2.3}$$

Let $s_0(q) + s_1(q,\mathbf{v})$ and $s_0'(q) + s_1'(q,\mathbf{v})$ be the ranks of the related matrices A_1 and A_1', respectively; then

$$A_1 = \begin{pmatrix} -p_i & 1 & 0 \\ \partial F/\partial x^i + p_i\, \partial F/\partial \phi & 0 & \partial F/\partial p^i \\ c_i & 0 & -a_i \end{pmatrix},$$

$$A_1' = \begin{pmatrix} 1 & 0 \\ 0 & \partial F/\partial p^i \\ 0 & -a_i \end{pmatrix}.$$

We find the vector $\mathbf{u}$ if $s_0 + s_1 = s_0' + s_1' \leqslant \dim E - 2$. If, furthermore, the *integral element* $\mathbf{u} \wedge \mathbf{v}$ is *regular*[1] (see 2.4), then there exists a two-dimensional integral manifold $i_2(s,t)$ s.t. equations

$$i_2(s,0) = i_1(s), \quad \frac{d}{dt} i_2(s,t)|_{s=0,t=0} = \mathbf{u} \text{ hold.}$$

Definition 2.3: A point $q \in E$ is called a zero-dimensional integral element if $F(q) = 0$. A k-dimensional plane $\Sigma_k = (q, \mathbf{u}_1 \wedge \cdots \wedge \mathbf{u}_k)$, $q \in E$, $\mathbf{u}_i \in T_q E$, $1 \leqslant i \leqslant k$, is called a k-dimensional integral element if (i) $F(q) = 0$, (ii) $\mathbf{u}_i \lrcorner \omega^{(1)} = 0$, (iii) $\mathbf{u}_i \wedge \mathbf{u}_j \lrcorner \omega^{(2)} = 0$, for all $i,j = 1,...,k$.

Denote by $O^k(q)$ the set of k-dimensional integral elements.

Definition 2.4: A k-dimensional integral element Σ_k is called regular if: (i) for $k = 0$, $q = \Sigma_0$ has a neighborhood N s.t. s_0 is constant in $N \cap O^0(N)$; and (ii) for $k \geqslant 1$, $s_0(x) + s_1(x,\mathbf{v}_1) + \cdots + s_k(x,\mathbf{v}_1,...,\mathbf{v}_k)$ is locally constant in a suitable neighborhood of $(q,\mathbf{u}_1,...,\mathbf{u}_k) \cap O^k(q)$ and Σ_k contains at least one regular $(k-1)$-dimensional integral element.

A k-dimensional integral element Σ_k is called ordinary if it contains at least one $(k-1)$-dimensional regular element.

Denote $R^k(q)$ the set of k-dimensional regular integral elements in $q \in E$.

Remark[8,9]: It is easy to show that the vectors $\mathbf{u}_1,...,\mathbf{u}_k$ give rise to a regular integral element if $s(q,\mathbf{u}_1,...,\mathbf{u}_r)$, $1 \leqslant r \leqslant k$, are chosen s.t. $s_0(q)$, $s_1(q,\mathbf{u}_1),...,s_r(q,\mathbf{u}_1,...,\mathbf{u}_r)$ assuming the maximum values in $O^r(q)$.

Definition 2.5: A differential system is called involutive if it admits an ordinary integral element Σ_m for which $\pi_* \Sigma_m = T_{\pi q} M$.

In other words, the equations yielding the ordinary element $\Sigma_m = (q, \mathbf{u}_1 \wedge \cdots \wedge \mathbf{u}_m)$ do not give any constraints between the $\mathbf{u}_i \lrcorner dx^k$, $i,k = 1,...,m$, components. This is true if, by denoting with $s_0'(q)$, $s_1'(q,\mathbf{u}_1)$,... the *reduced* characters, we have $s_r(q,\mathbf{u}_1,...,\mathbf{u}_r) = s_r'(q,\mathbf{u}_1,...,\mathbf{u}_r)$, for $0 \leqslant r \leqslant n-1$.

In this paper no study will be done to understand if a

given system is involutive or not.[1,5]

The procedure of Steps 1 and 2 can be iterated.

Theorem 2.1 (Cartan–Kähler theorem): Given an analytic differential system, a k-dimensional analytic integral manifold N_k and a point $q \in N_k$ s.t. the integral element $\Sigma_k = T_q N_k$ is regular, if $s_0 = s_0', \ldots, s_k = s_k'$ and $s_0 + s_1 + \cdots + s_k \leqslant \dim E - (k+1)$, then there exists, in a neighborhood of q, a $(k+1)$-dimensional integral manifold N_{k+1} s.t. $N_k \subset N_{k+1}$.

The condition on the ranks $s_0, \ldots, s_k$ has the following meaning: if it is verified, we surely can find an ordinary integral element Σ_{k+1} containing Σ_k.

In other words, an involutive system admits m-dimensional integral manifolds.

The proof of the Cartan–Kähler theorem can be found in Refs. 1 and 9: the main idea is to use $s_0 + \cdots + s_k$ equations to reorganize the system in a Cauchy Kowalewsky form; successively, after having found the solutions of these equations, one can easily check that the remaining equations are also automatically verified. In supermanifold theory, there are further difficulties, "residual constraints," which will be clarified in the following.

If a given integral element $(q, \mathbf{u}_1 \wedge \cdots \wedge \mathbf{u}_k)$ is not regular, *singular solutions* can arise; more precisely, we have the following definition.

Definition 2.6: A section $\Gamma: \mathbf{V} \to E$ is a k-singular solution if (i) it is a solution, and (ii) every k-dimensional vector subspace of its tangent space $T_p\Gamma$ is, for each $p \in \Gamma$, a nonregular integral element.

Concerning characteristic fields in Ref. 8 we find the following definition.

Definition 2.7 A vector field $X \in \Gamma(TE)$ is called characteristic if $X \lrcorner I \subset I$. The importance of characterisic fields is suggested by the following classical result.

Theorem 2.2[1,8,9]: The characteristic fields furnish a p-dimensional completely integrable differential system; moreover, if $y^1, \ldots, y^p$ are independent first-order integrals, then there exists a differential system Σ, equivalent to the previous one (that is, with the same solutions), which is constructed only with the differentials $dy^1, \ldots, dy^p$ and with coefficients that depend only on $y^1, \ldots, y^p$.

Following this suggestion, also in superspace theory, we have the following theorem.

Theorem 2.3: The characteristic fields yield a completely integrable Frobenius system.

Proof: If X, Y are characteristic fields and $\omega \in I$, we have

$$\mathscr{L}_X \omega = X \lrcorner d\omega + d(X \lrcorner \omega) \in I.$$

The formula

$$[X,Y] \lrcorner \omega = \mathscr{L}_Y (X \lrcorner \omega) - \mathscr{L}_X (Y \lrcorner \omega)$$

completes the proof. □

Theorem 2.4: If N is an integrable manifold of I, we can *generate* a higher-dimensional integral manifold N' by drawing N along the integral lines of the characteristic fields.

To generalize the Cartan–Kähler theorem to supermanifold theory, we have to consider the following theorems.

Theorem 2.5[10]: A $\mathrm{gl}(m,n)$ matrix, that is, a matrix of the type

$$A = \begin{pmatrix} a_i^j & a_\alpha^j \\ a_i^\beta & a_\alpha^\beta \end{pmatrix},$$

where $a_i^j, a_\alpha^\beta \in Q_0$ and $a_\alpha^j, a_i^\beta \in Q_1$, is nonsingular if and only if its body matrix $B = \sigma(A)$ is also.

In the following the numbers $s_0, \ldots, s_k$ will denote the ranks of the related body matrices. In view of this theorem, the matrix B controls the invertibility of A but is not able to exclude the presence of further constraints in the algebraic system described by A, which can happen if the rank is not maximal. These constraints will be called *residual constraints of the first type.*

Theorem 2.6 (superspace Cauchy–Kowalewsky theorem): E, W, G denote some vector superspaces and let (i) x be a coordinate in E, $y \in W$ and $\alpha \in G$, and (ii) F and p be supersmooth and analytic functions (for short SA, i.e., superanalytic). The system

$$\frac{\partial \alpha(x,y)}{\partial x} = F(x,y,\alpha(x,y),D_y\alpha(x,y)),$$
$$\alpha(0,y) = p(y) \tag{2.4}$$

admits one and only one solution if x is an even coordinate; if x is an odd one it admits a solution if and only if the "total" derivative $dF/dx = 0$, that is, a "residual constraint" of the second type.

Proof: If x is even the proof can be found in Ref. 4 where x is assumed to belong to an even vss E. The main idea is to reconstruct the classical majorant method in the case in which the variable x is real; subsequently, by a suitable change of variables one gets the proof.

If x is odd, one directly gets

$$\alpha(x,y) = p(y) + xF(0,y,p(y),D_y p(y)) + 0.$$

On the other hand, substitution in Eq. (2.4) yields

$$F(0,y,\alpha(0,y),D_y\alpha(0,y)) = F(x,y,\alpha(x,y),D_y\alpha(x,y)),$$

which completes the proof.

Theorem 2.7 (the generalized Cartan–Kähler integration theorem): Let N_h be an $(h-1)$ integral supermanifold of a closed superanalytical ideal I defined on an analytic supervector bundle $\pi: E \to \mathbf{V}$. Suppose that, in $q \in N$ the space $\Sigma_{h-1} = T_q(N_h)$ is a regular integral element and that there exists an integral tangent vector $\mathbf{v}$ s.t. $\Sigma_h = \Sigma_{h-1} \oplus \mathbf{v}$ is an h-dimensional integral element. Now, if $\sigma\pi_* \mathbf{v}$ (the body of the components of $\pi_* \mathbf{v}$) independent of Σ_{h-1} and there are no residual constraints of the first type, there exists a local neighborhood U of q in E where there is an h-dimensional integral manifold N_h containing N_{h-1}. On the contrary, if $\pi_* \mathbf{v}$ is odd, N_h exists if no residual constaint is found.

In Sec. III we shall clarify the matter of residuals in the case of first-order differential equations. The presence of residual constraints does not mean that, in general, a given integration cannot be performed; it only means that the Cartan–Kähler theorem cannot be used.

III. FIRST-ORDER PARTIAL DIFFERENTIAL EQUATIONS ON SUPERSPACE

A first-order partial differential equation on a vss is of the type

$$F\left(x^A,\phi,\frac{\partial\phi}{\partial x^A}\right)=0, \tag{3.1}$$

where $A=1,\ldots,m+n$, $x^A\in Q^{m,n}$, $\phi\in Q_0$.

Following Cartan we construct the differentiable ideal I:

$$\begin{aligned} &F(x^A,\phi,p_A)=0,\\ &dx^A\frac{\partial F}{\partial x^A}+d\phi\frac{\partial F}{\partial\phi}+dp_A\frac{\partial F}{\partial p_A}=0,\\ &d\phi-dx^A p_A=0,\quad dx^A\wedge dp_A=0. \end{aligned} \tag{3.2}$$

Forms $\omega^{(0)},\omega^{(1)},\omega^{(2)}$ are defined on the sv bundle π: $E\to Q^{m,n}$ whose standard fiber $\pi^{-1}(x^A)$ is isomorphic to $Q_0\times Q^{m,n}$.

Step 1: Choose a point $q\in E$ such that $F(q)=0$ and the rank of the algebraic system $\omega^{(1)}$ is locally constant. Determine a vector $\mathbf{v}\in T_qE$,

$$\mathbf{v}=\left(a^A\frac{\partial}{\partial x^A}+b\frac{\partial}{\partial\phi}+c_A\frac{\partial}{\partial p_A}\right)_q,$$

satisfying the system $\mathbf{v}\rfloor\omega^{(1)}=0$. The representative matrix is

$$A=\begin{pmatrix} -p_i & -p_\alpha & 1 & 0 & 0\\ \dfrac{\partial F}{\partial x^i}+p_i\dfrac{\partial F}{\partial\phi} & \dfrac{\partial F}{\partial\theta^\alpha}+p_\alpha\dfrac{\partial F}{\partial\phi} & 0 & \dfrac{\partial F}{\partial p^i} & \dfrac{\partial F}{\partial p^\alpha} \end{pmatrix}.$$

If $s_0=s_0'=2$ we find a vector $\mathbf{v}$ with arbitrary components $\pi_*\mathbf{v}$ (e.g., $\pi_*\mathbf{v}=\partial/\partial x^1$) and there exists a one-dimensional integral manifold $i_1(s)$ of I s.t. $i(0)=q$ and $(d/ds)i_1(s)|_{s=0}=\mathbf{v}$.

Step 2: Find a vector $u\in T_qE$ solving the *polar* system [Eq. (2.3)]. Let s_0+s_1 be the rank of the related representative matrix $\sigma(A_1)$, where

$$A_1=\begin{pmatrix} -p_i & -p_\alpha & 1 & 0 & 0\\ \dfrac{\partial F}{\partial x^i}+p_i\dfrac{\partial F}{\partial\phi} & \dfrac{\partial F}{\partial\theta^\alpha}+p_\alpha\dfrac{\partial F}{\partial\phi} & 0 & \dfrac{\partial F}{\partial p^i} & \dfrac{\partial F}{\partial p^\alpha}\\ c_i & c_\alpha & 0 & -a_i & -a_\alpha \end{pmatrix}.$$

Suppose that $s_1=s_1'=1$, and $s_0+s_1\leqslant\dim E-2$. We can find $\mathbf{u}$ and choose $\pi_*u=\partial/\partial x^2$. If, furthermore, this *integral element is regular*, then there exists a two-dimensional integral manifold $i_2(s,t)$ s.t.

$$i_2(s,0)=i_1(s),\quad \frac{d}{dt}i_2(s,t)|_{s=0,t=0}=u.$$

This procedure can be iterated: the step number k allows us to extend a k-dimensional integral manifold i_k to a $(k+1)$-dimensional one i_{k+1} if (i) the rank

$$s_0(q)+\cdots+s_k(q,\mathbf{u}_1,\ldots,\mathbf{u}_k)=s_0'(q)+\cdots+s_K'(q,\mathbf{u}_1,\ldots,\mathbf{u}_k)$$

is locally constant, (ii) we find an integral vector $\mathbf{u}_{k+1}$ ($\mathbf{u}_{k+1}\in T_qE$ but $\pi_*\mathbf{u}_{k+1}\notin\pi_*T_qi_k$) s.t. extends T_qi_k to a $(k+1)$-dimensional regular integral element, and (iii) no residual constraints are present.

Theorem 3.1: Let $F\in Q_0$ and $\sigma(\partial F/\partial p_i)\neq 0$ for all $i=1,\ldots,m$. The Cartan–Kähler integration procedure can be used to get an m-dimensional integral manifold.

Proof: The ranks $s_1,\ldots,s_{m-1}$ are 1 and no *residual* constraints exist; this is clear if one looks at the matrix related to the $i_k\to i_{k+1}$ extension ($k<m$):

$$A_k=\begin{pmatrix} -p_1-p_2\cdots-p_m & -p_{m+1}-p_{m+2}\cdots-p_{m+n} & 1 & 0\cdots0 & 0\cdots0\\ \dfrac{\partial F}{\partial x^i}+p_i\dfrac{\partial F}{\partial\phi} & \dfrac{\partial F}{\partial x^\alpha}+p_\alpha\dfrac{\partial F}{\partial\phi} & 0 & \dfrac{\partial F}{\partial p_i} & \dfrac{\partial F}{\partial p_\alpha}\\ -c_1 0\cdots0 & 0\cdots0 & 0 & 10\cdots\cdots0 & 0\cdots0\\ 0-c_2 0\cdots0 & 0\cdots0 & 0 & 01\cdots\cdots0 & 0\cdots0\\ \cdots & \cdots & \cdot & \vdots\quad\vdots & \cdots\\ 0\cdots0-c_k 0\cdots0 & 0\cdots0 & 0 & 0\cdots010\cdots0 & 0\cdots0 \end{pmatrix},$$

where

$$c_k=-\left[\frac{\partial F}{\partial x^k}+p_k\frac{\partial F}{\partial\phi}\right]\left(\frac{\partial F}{\partial p_k}\right)^{-1},$$

k is not summed, and the last term exists since $\sigma(\partial F/\partial p_k)\neq 0$, for all $k=1,\ldots,m$.

The structure of the linear system that arises allows us to formulate the extension problem by means of a Cauchy–Kowalewsky system. Actually, other equations are present in the requirement that a $(k+1)$-dimensional manifold i_{k+1} exists; however, these are identically verified in view of the initial data and of the Cauchy–Kowalewsky system. □

In this case the problem of residual constraints could

arise only when performing extensions in the odd sector.

Theorem 3.2: The characteristic fields of Eq. (3.1), when restricted to a $q \in E$, make any regular k-dimensional integral element into a nonregular $(k+1)$-dimensional one.

Proof: An explicit calculation shows that the *characteristic* fields obtained by following Definition 2.4 are given by the expression

$$X = v^A \frac{\partial}{\partial x^A} + b \frac{\partial}{\partial \phi} + c_A \frac{\partial}{\partial p_A},$$

where, by introducing a suitable form β,

$$v_A = \frac{\partial F}{\partial p_A} \beta (-1)^A,$$

$$c_A = -\left(\frac{\partial F}{\partial x^A} + p_A \frac{\partial F}{\partial \phi}\right) \beta$$

hold. On the other hand, the *same* expression is obtained if one searches the vectors which make any regular k-dimensional regular element into a nonregular $(k+1)$-dimensional one. □

This result clarifies the link between characteristic fields and nonextendible integral manifolds.

Theorem 3.3: The differential equation (3.1) admits characteristic tangent fields if and only if F is a homogeneous element. In such a case, if $F \in Q_0$ then $\beta \in Q_0$; if $F \in Q_1$ then $\beta \in Q_1$. The motion induced by X satisfies the Hamilton–Jacobi equation written here when, e.g., $F \in Q_0$:

$$\frac{dx^A}{dt} = \frac{\partial F}{\partial p_A} (-1)^A \beta,$$

$$\frac{dp_A}{dt} = -\left(\frac{\partial F}{\partial x^A} + p_A \frac{\partial F}{\partial \phi}\right) \beta.$$

Theorem 3.4: Equation (3.1) has no singular solutions of dimension $n \geqslant 2$. It has $n = 1$ ones only if F is homogeneous and in this case the singular integral manifolds are the characteristic lines.

IV. EXAMPLES

Let

$$F(x,y,\theta,\xi,\phi,p_x,p_y,p_\theta,p_\xi) = p_x^2 - 4\phi + p_\theta + 4\theta\alpha - \alpha = 0, \tag{4.1}$$

where $x,y \in Q_0$, $\theta,\xi \in Q_1$, $\phi \in Q_0$, and α is a fixed element of Q_1.

The ideal I is

$$\begin{aligned} &d\phi - dx\, p_x - dy\, p_y - d\theta\, p_\theta - d\xi\, p_\xi = 0, \\ &2\, dp_x\, p_x - 4\, dx\, p_x - 4\, d\theta\, p_\theta + dp_\theta + 4\, d\theta\, \alpha = 0, \\ &dx \wedge dp_x + dy \wedge dp_y + d\theta \wedge dp_\theta + d\xi \wedge dp_\xi = 0. \end{aligned} \tag{4.2}$$

The equation $\mathbf{v}_1 \lrcorner \omega^{(1)} = 0$ is represented by the matrix

$$A = \begin{pmatrix} -p_x & -p_y & -p_\theta & -p_\xi & 1 & 0 & 0 & 0 & 0 \\ -4p_x & -4p_y & -4p_\theta + 4\alpha & -4p_\xi & 0 & 2p_x & 0 & 1 & 0 \end{pmatrix}.$$

By setting $q = (0,0,0,0,\frac{1}{4},1,0,\alpha,0)$ and

$$\mathbf{v}_1 = \left(\sum_{A=1}^{4} \alpha^A \frac{\partial}{\partial x^A} + b \frac{\partial}{\partial \phi} + \sum_{B=1}^{4} c^B \frac{\partial}{\partial p_B}\right)_q,$$

we have the constraints

$$b - a^1 p_x - a^2 p_y - a^3 p_\theta - a^4 p_\xi = 0, \quad 2c^1 p_x - 4a^1 p_x - 4a^3 p_\theta + c^3 + 4a^3 \alpha = 0.$$

Since $s_0 = s_0' = 2$, there exist $\infty^{\dim E - 1 - s_0} = \infty^6$ choices for $\mathbf{v}_1$. Actually, we can freely choose $a^2 = a^3 = a^4 = 0$, $c^2 = c^3 = c^4 = 0$. Moreover, by setting $a^1 = 1$ we get

$$\mathbf{v}_1 = \left(\frac{\partial}{\partial x} + p_x \frac{\partial}{\partial \phi} + 2 \frac{\partial}{\partial p_x}\right)_q$$

and the integral line

$$i_1(x) = (x,0,0,0,x^2 + x + \tfrac{1}{4}, 2x + 1, 0, \alpha, 0).$$

Now we determine a vector $\mathbf{v}_2 \in T_q E$ satisfying Eq. (2.3), where $\mathbf{v}_1 \lrcorner \omega^{(2)} = dp_x - 2\, dx$.

We have $s_1 = s_1' = 1$ and ∞^4 choices. By setting $\mathbf{v}_2 = (\partial/\partial y)_q$ we get

$$i_2(x,y) = (x,y,0,0,x^2 + x + \tfrac{1}{4}, 2x + 1, 0, \alpha, 0).$$

Now determine $\mathbf{v}_3$ satisfying Eq. (2.3) where

$$\mathbf{v}_1 \lrcorner \omega^{(2)} = dp_x - 2\, dx, \quad \mathbf{v}_2 \lrcorner \omega^{(2)} = dp_y.$$

The related matrix is

$$A_2 = \begin{pmatrix} -p_x & -p_y & -p_\theta & -p_\xi & 1 & 0 & 0 & 0 & 0 \\ -4p_x & -4p_y & -4p_\theta + 4\alpha & -4p_\xi & 0 & 2p_x & 0 & 1 & 0 \\ -2 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix}.$$

We have $s_2 + s_2' = 1$ and ∞^2 choices for v_3. Therefore, we can solve the following system for $\phi(x,y,\theta)$ by setting $\alpha^4 = c^4 = 0$:

$$\begin{aligned}
&\frac{\partial\phi}{\partial\theta} = p_\theta\,, \quad \phi(x,y,0) = x^2 + x + \tfrac{1}{4},\\
&\frac{\partial p_x}{\partial\theta} = \frac{\partial p_\theta}{\partial x}\,, \quad p_x(x,y,0) = 2x + 1,\\
&\frac{\partial p_y}{\partial\theta} = \frac{\partial p_\theta}{\partial y}\,, \quad p_y(x,y,0) = 0,\\
&\frac{\partial p_\theta}{\partial\theta} = 4p_\theta - 4\alpha - 2\left(\frac{\partial p_\theta}{\partial x}\right)p_x, \quad p_\theta(x,y,0) = \alpha.
\end{aligned} \tag{4.3}$$

We also have the further equations

$$\begin{aligned}
&\frac{\partial\phi}{\partial x} - p_x = 0,\\
&\frac{\partial\phi}{\partial y} - p_y = 0,\\
&\frac{\partial p_x}{\partial y} - \frac{\partial p_y}{\partial x} = 0,\\
&\frac{\partial p_\theta}{\partial x} - 4p_x + 2\left(\frac{\partial p_x}{\partial x}\right)p_x = 0,\\
&\frac{\partial p_\theta}{\partial y} - 4p_y + 2\left(\frac{\partial p_x}{\partial y}\right)p_x = 0.
\end{aligned} \tag{4.4}$$

The vanishing of the residual constraints is given by

$$\frac{\partial p_\theta}{\partial\theta} = 0, \quad \frac{\partial p_\theta}{\partial x}\frac{\partial p_x}{\partial\theta} = 0. \tag{4.5}$$

Now, Eqs. (4.4) are *identically* verified in view of Eqs. (4.3). An explicit integration gives $\phi(x,y,\theta) = x^2 + x + \frac{1}{4} + \theta\alpha$, which also verifies Eqs. (4.5). The last step is trivial; one gets the required solution

$$\phi(x,y,\theta,\xi) = x^2 + x + \tfrac{1}{4} + \theta\alpha.$$

Consider the equation

$$\frac{\partial f(x,\theta)}{\partial\theta} = \theta, \quad f(x,0) = a(x). \tag{4.6}$$

Use the Cauchy–Kowalewsky theorem directly with

$$F\left(x,\theta,f\frac{\partial f}{\partial x}, \frac{\partial f}{\partial\theta}\right) = \theta.$$

The residual constraints are $dF/d\theta = 1 \neq 0$ and no solution can be found. Actually, one should have found $f(x,\theta) = a(x) + \theta^2 = a(x)$, which does not solve Eq. (4.6).

Finally, consider the system

$$F(x,\theta,\phi,p_x,p_\theta) = p_x^2 - 4\phi + 4\theta p_\theta,$$

with $\phi, F \in Q_0$, $q = 0$, $i_1(x) = (x,0,x^2,2x,0)$, and therefore

$$\mathbf{u}_1 = \left(\frac{di}{dx}\right)_0 = \left(\frac{\partial}{\partial x} + p_x\frac{\partial}{\partial\phi} + 2\frac{\partial}{\partial p_x}\right)_q.$$

We have

$$A_1 = \begin{pmatrix} -p_x & -p_\theta & 1 & 0 & 0\\ -4p_x & 0 & 0 & 2p_x & -4\theta\\ -2 & 0 & 0 & 0 & 0 \end{pmatrix}$$

and $s_0 + s_1 = 3$, $s_0' + s_1' = 2$. No solutions in the sense of Definition 2.1 can be found since we cannot freely choose the components $\pi_* \mathbf{u}_2$.

ACKNOWLEDGMENTS

This paper was partially supported by the National Group for Mathematical Physics of the Italian Research Council and by the Italian Ministry of the Public Education through the research project "Geometria e Fisica."

[1] E. Cartan, *Les System Differentiels Exterieurs et leurs Applications Geometriques* (Hermann, Paris, 1975); M. Kuranishi, Am. J. Math. **79**, 1 (1957).
[2] T. Regge and Y. Ne'eman, Riv. Nuovo Cimento **1**, 1 (1978).
[3] U. Bruzzo and R. Cianci, J. Phys. A **18**, 417 (1985).
[4] R. Cianci, in *Proceedings of the 7th Italian Conference on General Relativity and Gravitational Physics* (World Scientific, Singapore, 1986), p. 163.
[5] H. Goldschmidt, Ann. Math. **86**, 246 (1967); V. Guillemin, Am. J. Math. **90**, 270 (1968); V. Guillemin and S. Sternberg, Bull. Am. Math. Soc. **70**, 16 (1964); V. Guillemin, S. Sternberg, D. G. Quillen, Commun. Pure Appl. Math. **23**, 39 (1970); D. Spencer, Bull. Am. Math. Soc. **75**, 179 (1965); D. Yang, "*Involutive Hyperbolic Differential Systems, Memoirs of the American Mathematical Society*, Vol. 68, No. 370 (Am. Math. Soc. Reading, MA, 1987).
[6] A. Jadczyk and K. Plich, Commun. Math. Phys. **78**, 373 (1981).
[7] A. Rogers, J. Math. Phys. **21**, 1352 (1980); **22**, 939 (1981); M. Rothstein, University of Washington, preprint 1985; C. P. Boyer and S. Gitler, Trans. Am. Math. Soc. **285**, 241 (1984); V. S. Vladimirov and I. V. Volovich, Theor. Math. Phys. **59**, 317 (1984); **60**, 743 (1985); J. Hoyos, M. Quiros, J. Ramirez Mittelbrunn, and F. J. De Urries, J. Math. Phys. **25**, 833 (1984). See also the following papers: B. De Witt, *Supermanifolds* (Cambridge U.P., London, 1984); J. M. Rabin and L. Crane, Commun. Math. Phys. **100**, 141 (1985).
[8] Y. Choquet-Bruhat, *Géométrie differentielle et systèmes extérieurs* (Dunod, Paris, 1968).
[9] J. Dieudonne, *Elements d'Analyse* (Gauthier-Villars, Paris, 1971).
[10] U. Bruzzo and R. Cianci, Class. Quantum Gravit. **1**, 213 (1984).

Wronskians, geometry, and some general solutions to the nonlinear Liouville-like PDE's

Jerzy F. Plebański[a)]
Centro de Investigación y de Estudios Avanzados del IPN, Apartado Postal 14-740, México, D.F., Mexico

(Received 1 December 1987; accepted for publication 25 May 1988)

The formal properties of n-dimensional Wronskians and their geometric interpretation enable one to construct explicit analytic solutions to some nonlinear partial differential equations (PDE's) that generalize the Liouville equation $\phi_{xy} = e^{2\phi}$. The studied PDE's are (a) $\mathscr{L}_m F = \text{const}$, (b) $\mathscr{L}_m F = G$, $\mathscr{L}_m G = F$, and (c) $\mathscr{L}_m(\mathscr{L}_m F) = \text{const}$, where $\mathscr{L}_m$ is the nonlinear differential operation $\mathscr{L}_m F := \det(\partial_x^k \partial_y^l F)$, with $k,l = 0,1,\ldots,m$. Some nontrivial formal properties of the composition of the $\mathscr{L}_m$'s are established.

I. INTRODUCTION

The classical nonlinear Liouville partial differential equation (PDE)[1]

$$\phi_{xy} = e^{-2\epsilon\phi} \quad \text{and} \quad \epsilon^2 = 1, \tag{1.1}$$

which can be stated in the equivalent form

$$FF_{xy} - F_x F_y = \epsilon, \quad \text{where } F := e^{\epsilon\phi}, \tag{1.2}$$

is presently considered as one of the simplest examples of a PDE yielding solutions via the Bäcklund procedure.[2]

It has been well-known for about 100 years that the PDE [(1.1) and (1.2)] has the remarkable property of possessing the solution

$$F = (\epsilon\dot{p}\dot{q})^{-1/2}(1+pq) \Leftrightarrow \phi = \epsilon \ln[(\epsilon\dot{p}\dot{q})^{-1/2}(1+pq)], \tag{1.3}$$

where the single variable functions $p = p(x)$ and $q = q(y)$ are *arbitrary*, constrained only by the condition $\epsilon\dot{p}\dot{q} > 0$, with the dots denoting the derivatives with respect to the corresponding variables. Thus the most general solution to the Liouville equation is *algebraically* constructed from *arbitrary* single variable functions and their derivatives.

Of particular interest are (i) the mechanism that assures us that the general solutions to (1.1) and (1.2) have the form of (1.3), and (ii) the existence of other PDE's with the general solutions of a similar structure, i.e., *algebraically* constructed from arbitrary functions of a single variable and their derivatives. This paper intends to offer at least a partial answer to these questions.

In earlier work with J. D. Finley on the problem of twisting N-type solutions in complexified general relativity, we encountered as an intermediate step the PDE's

$$FF_{xy} - F_x F_y = G \quad \text{and} \quad GG_{xy} - G_x G_y = F. \tag{1.4}$$

The structural similarity of these PDE's with (1.2) suggests that they be labeled as "double" Liouville equations. In fact, the general solution to these equations may be constructed in a fashion quite similar to that for the general solution to (1.2). More specifically, the solutions can be algebraically constructed from *arbitrary* single variable functions $p_i(x)$, $q_i(y)$, $i = 1,2$, and their first and second derivatives.

The geometrical interpretation of the solutions to (1.4) has resulted in the elucidation of the formal properties of the three-dimensional Wronskians of functions of a single variable. It seems natural to develop a more general n-dimensional theory that contains PDE's (1.2) and (1.4) as special cases. Section II is a summary of the formal properties of the n-dimensional Wronskians and their basic minors, accompanied by the corresponding geometric interpretation. These are essential in Sec. III, which is concerned with the formal properties of the abstract nonlinear differential operators $\mathscr{L}_m$.

Of course, in terms of $\mathscr{L}_m$'s, (1.2) and (1.4) may be stated as

$$\text{(a)}\ \mathscr{L}_1 F = \epsilon, \quad \epsilon^2 = 1 \ \text{and (b)}\ \mathscr{L}_1 F = G, \quad \mathscr{L}_1 G = F. \tag{1.5}$$

For $m \geqslant 2$, $\mathscr{L}_m$ constitutes the natural generalization of the concept of the "Liouville operator" $\mathscr{L}_1$. In Sec. IV we investigate the chains *a-b-c* of PDE's by exploiting these generalized Liouville operators. It is shown that we are able to determine the most general analytic form of their solution for the case of an *a*-chain. In the case of a *b*-chain, we are able to determine some special solutions for $m > 2$, and the most general solutions for $m = 1$. Finally, we consider a *c*-chain for $m = 1$, which reduces to the biharmonic equation for the conformal factor of a two-dimensional Riemannian space. Also, we discuss alternative formulations of the differential problem under consideration. In Sec. V we discuss some open problems related to the results of this work. Semitrivial proofs of a computational nature are abbreviated by their basic ideas only. The nontrivial proofs are in appendices, which the interested reader might find useful in further work along the same lines as that given in the text.

The fact that the most general solution to (1.1) and (1.2) has the analytic form of (1.3) is quite useful in mathematical physics. In particular, within the theory of exact solutions in general relativity, experience has shown that whenever a Liouville equation occurs at an intermediate step, the corresponding problem is integrable up to the very end. In this respect, we believe that the techniques of this paper are to some extent manageable when applied to nonlinear Liouville-like PDE's and thus may find some useful applications.

[a)] On leave of absence from the University of Warsaw, Warsaw, Poland.

 0022-2488/88/102162-13$02.50

II. WRONSKIANS AND GEOMETRY

Let $\mathbf{F}^n$ be the set of ordered n-tuples of the smooth functions of some single variable, say $t\in\mathbf{R}$. Thus a typical element $f\in F^n$ consists of $f:=\{f_i(t), i=1,...,n\}$, where $f_i:\mathbf{R}\to\mathbf{R}$, and n is a fixed integer. If the f_i's are considered as meaningful modulo the arbitrary changes of the independent variable only, then, defining the equivalence class $C: f_i(t)\equiv f_i(t(t'))$, where $t':\mathbf{R}\to\mathbf{R}$ is an arbitrary smooth bijection, the set $\mathbf{F}^n/C$ amounts to the set of smooth curves in $\mathbf{R}^n$.

The set $\mathbf{F}^n/C$ automatically carries a rich structure induced by the concepts of the Wronskian and its basic minors.[3] These concepts are understood as the mappings W: $\mathbf{F}^n\to\mathbf{F}^1$ and $*$: $\mathbf{F}^n\to\mathbf{F}^n$, which are defined explicitly by

$$f\in\mathbf{F}^n \Rightarrow \begin{cases} Wf:=\epsilon_{i_1 i_2\cdots i_n} f_{i_1} f^1_{i_2}\cdots f^{n-1}_{i_n}, \\ *f_i:=(-1)^{n-1}\epsilon_{ii_1 i_2\cdots i_{n-1}} f_{i_1} f^1_{i_2}\cdots f^{n-2}_{i_{n-1}}, \end{cases} \tag{2.1}$$

where $f^k_i:=(d/dt)^k f_i$, with $k=0,1,...$ and $i=1,2,...,n$. Here $\epsilon_{i_1\cdots i_n}$ is the totally skew Levi-Civita symbol in n dimensions, normalized by $\epsilon_{12\cdots n}=1$.[4]

The basic formal properties of the mappings W and $*$ may be stated in the form of the following theorem.

Theorem 1: The following identities are valid for an arbitrary $f\in\mathbf{F}^n$:

$$\begin{aligned} &\text{(a)}\ W(\lambda f)=\lambda^n Wf, && \text{(e)}\ *(\lambda f_i)=\lambda^{n-1}*f_i, \\ &\text{(b)}\ Wf'=\left(\frac{dt}{dt'}\right)^{\binom{n}{2}} Wf, && \text{(f)}\ *(f'_i)=\left(\frac{dt}{dt'}\right)^{\binom{n-1}{2}} *f_i, \\ &\text{(c)}\ W(Mf)=\det M\cdot Wf, && \text{(g)}\ *(M_{ij}f_j)=\det M\cdot M^{-1}_{ji}*f_j, \\ &\text{(d)}\ W(*f)=(Wf)^{n-1}, && \text{(h)}\ *f_i=(-1)^{n-1}(Wf)^{n-2}f_i. \end{aligned} \tag{2.2}$$

In the first line of Eqs. (2.2), a smooth $\lambda:\mathbf{R}\to\mathbf{R}$ is arbitrary and $\lambda f:=\{\lambda(t)f_i(t)\}\in\mathbf{F}^n$. In the second line, given $f=\{f_i(t)\}$, a smooth bijection $t:\mathbf{R}\to\mathbf{R}$ induces $f'=\{f'_i(t')\}:=\{f_i(t(t'))\}$ in the left-hand members W. The $*$ is meant as the nonlinear differential operation with respect to t', while in the right-hand members it refers to the variable t. Note that (2.2)(b) is valid for $n\geqslant 2$, while (2.2)(f) applies for $n\geqslant 3$. In the third line $M_{ij}=\text{const}$ is an arbitrary nonsingular $n\times n$ matrix, with $Mf:=\{M_{ij}f_j(t)\}$.

The first three lines of (2.2) follow directly from the definitions of the mappings W and $*$. The proofs of the identities given in the fourth line of (2.2) are nontrivial. They may be outlined in the form of a sequence of lemmas; those of interest are given in Appendix A.

W-regular curves: According to (2.2)(b), a smooth curve in $\mathbf{R}^n$, represented as $f=\{f_i(t)\}\in\mathbf{F}^n$, has the characteristic $Wf\neq 0$, independent of the choice for its parametrization.

Observe that when $\binom{n}{2}=\text{even}$, the $\text{sgn}(Wf)$ cannot be affected by the change of the parametrization. This gives rise to the classification of the curves $f,g,...\in\mathbf{F}^n/C$ into the two basic classes

$$W\text{-regular: } Wf\neq 0, \quad W\text{-singular: } Wf=0. \tag{2.3}$$

The origin of this classification is the condition $Wf\neq 0$ for $f\in\mathbf{F}^n$, which is known[5] to constitute a necessary and sufficient condition for the linear independence of n smooth f_i's, i.e., $Wf\neq 0\Leftrightarrow\{\lambda_i=\text{const}, \lambda_i f_i(t)=0\Rightarrow\lambda_i=0\}$.[6] Consequently, a W-regular curve cannot be contained in any of $(n-1)$-hyperplanes through the distinguished origin of $\mathbf{R}^n$. Correspondingly, each W-singular curve is contained in some $(n-1)$-hyperplane through the distinguished origin of $\mathbf{R}^n$.

Normal parametrization: Given a W-regular curve represented by $\{f_i(t)\}\in\mathbf{F}^n$, $Wf\neq 0$, we propose to define its normal parameter $x\in\mathbf{R}$ via

$$dx:=\text{sgn}(Wf)\cdot|Wf|^{1/\binom{n}{2}}\,dt. \tag{2.4}$$

Heuristically, this idea is somewhat analogous to the idea of using the Pythagorean length as the natural parameter of the Frenet formulas and the concept of the relativistic proper time.

Performing the quadrature in (2.4), the derived function $x=x(t)$, with $dx/dt\neq 0$, defines its inverse $t=t(x)$. Thus the curve may be considered as given in terms of its normal parameter as $f'=f'_i(x):=f_i(t(x))$. Then it follows from (2.4) that

$$\begin{aligned} &\tbinom{n}{2}=\text{odd}\Rightarrow Wf'=1 \text{ and} \\ &\tbinom{n}{2}=\text{even}\Rightarrow Wf'=\text{sgn}(Wf). \end{aligned} \tag{2.5}$$

The differentiations in the operation W are with respect to the variable x.

According to (2.2)(d), $*$ maps W-regular curves into W-regular curves. Given a curve represented as $f\in\mathbf{F}^n$, $Wf\neq 0$, we refer to $*f\in\mathbf{F}^n$, $W(*f)\neq 0$ as the *dual* curve. Considering the curve f as represented in terms of its normal parameter, $f'=\{f'_i(x)\}$, according to (2.2)(d) and (2.2)(h), we have

$$Wf'=1\Rightarrow\begin{cases} n=\text{odd}: & W(*f')=1, \quad **f'_i=f'_i, \\ n=\text{even}: & W(*f')=1, \quad **f'_i=-f'_i, \end{cases} \tag{2.6}$$

and

$$Wf'=-1\Rightarrow\begin{cases} n=\text{odd}: & W(*f')=1, \quad **f'_i=-f'_i, \\ n=\text{even}: & W(*f')=-1, \quad **f'_i=-f'_i. \end{cases} \tag{2.7}$$

Therefore, $*$ is an involution or anti-involution among the W-regular curves and their duals. Note that according to

(2.5), the formulas (2.7) are of interest only when $\binom{n}{2}$ = even $\Rightarrow n = 4,5,8,9,\ldots$.

The ▶ mapping: The construction of the normal parameter x for a W-regular curve via (2.4) involves a quadrature. There is, however, a simple process that enables us to construct the W-regular curve, as given in terms of its normal parameter, bypassing the necessity of any integrations.

Consider a W-regular curve represented as $f = \{f_i(t)\} \in \mathbb{F}^n$, $Wf \neq 0$. Then the mapping ▶, defined by

$$\blacktriangleright f_i := |Wf|^{-1/n} f_i, \tag{2.8}$$

obviously produces another W-regular curve. The formal properties of ▶ may be summarized in the form of the following theorem.

Theorem 2: The following identities hold:

$$\begin{aligned} &\text{(a)}\ W(\blacktriangleright f) = \operatorname{sgn}(Wf), \quad &&\text{(c)}\ \blacktriangleright^2 f_i = \blacktriangleright f_i, \\ &\text{(b)}\ \blacktriangleright(\lambda f_i) = \operatorname{sgn}\lambda \blacktriangleright f_i, \quad &&\text{(d)}\ *\blacktriangleright f_i = \blacktriangleright * f_i. \end{aligned} \tag{2.9}$$

Identity (2.9)(a) follows from (2.2)(a). Similarly, (2.9)(b) with arbitrary λ: $\mathbb{R} \to \mathbb{R}$, $\lambda(t) \neq 0$ follows from (2.2)(a). Equation (2.9)(c) is a trivial consequence of the definition of ▶. Equation (2.9)(d) may be established by using (2.2)(e) and (2.2)(d).

According to (2.4) and (2.9)(a), the curve $g := \blacktriangleright f$ possesses the normal parameter x, where $x = \operatorname{sgn}(Wf) \cdot t$, if we choose the integration constant for x equal to 0. Consequently, the curve

$$g' = \{g_i'(x)\} := \{\blacktriangleright f_i|_{t = \operatorname{sgn}(Wf) \cdot x}\}, \tag{2.10}$$

according to (2.5), satisfies

$$\begin{aligned} &\tbinom{n}{2} = \text{odd} \Rightarrow Wg' = 1, \\ &\tbinom{n}{2} = \text{even} \Rightarrow Wg' = \operatorname{sgn}(Wf). \end{aligned} \tag{2.11}$$

The $(n-1)$-dimensional interpretation of ▶: According to (2.9)(c), the mapping ▶ has the nature of a projective operation. This induces its $(n-1)$-dimensional interpretation. Indeed, $Wf \neq 0 \to f_i \neq 0$, and in particular $f_n \neq 0$. Thus we can represent the f_i's as $f_i = |f_n| \tilde{h}_i$, $\tilde{h}_i := f_i / |f_n|$. Then $\tilde{h} =: (\tilde{h}_a, \epsilon)$, $a = 1,\ldots,n-1$, $\epsilon^2 = 1$, is an n-dimensional concept, $\tilde{h} = \{\tilde{h}_i\} \in \mathbb{F}^n$, while $h := \{\tilde{h}_a\} \in \mathbb{F}^{n-1}$ is considered as a $(n-1)$-dimensional concept. By $\dot{h} \in \mathbb{F}^{n-1}$ we mean $\dot{h} = \{\dot{h}_a(t)\}$, with the dot denoting the derivative. Then one obtains

$$W\tilde{h} = \epsilon(-1)^{n-1} \cdot W\dot{h}, \tag{2.12}$$

where the Wronskians on the left and the right are n- and $(n-1)$-dimensional constructs, respectively. This being the case,

$$Wf = |f_n|^n W\tilde{h} = \epsilon(-1)^{n-1} |f_n|^n W\dot{h} \quad \text{[using (2.2)(a)]}. \tag{2.13}$$

Hence, $Wf \neq 0 \Rightarrow W\dot{h} \neq 0$. The last $(n-1)$-dimensional condition has a simple geometric interpretation. Having $W\dot{h} = 0$ implies that there are nontrivial $\lambda_a = \text{const}$, $a = 1,\ldots,n-1$, such that $\lambda_a \dot{h}_a(t) = 0 \Leftrightarrow \lambda_a h_a(t) = \text{const}$. It follows that $h \in \mathbb{F}^{n-1}$, with $W\dot{h} \neq 0$, is a curve in $\mathbb{R}^{n-1}$ prohibited to be contained in *any* $(n-2)$-hyperplane in $\mathbb{R}^{n-1}$, and not only those through the origin of $\mathbb{R}^{n-1}$.

Using the definition of ▶ and (2.13), one easily sees that

$$\blacktriangleright f_a = |W\dot{h}|^{-1/n} h a, \quad \blacktriangleright f_n = \epsilon |W\dot{h}|^{-1/n}. \tag{2.14}$$

Therefore the functions (2.10), which automatically fulfill (2.11), may be considered as algebraically constructed from $h_a^k(t) = \partial_t^k h_a(t)$, $a = 1,\ldots,n-1$, $k = 0,\ldots,n-1$, where the smooth $h_a(t)$'s are *arbitrary*, constrained only by the condition that the curve $h \in F^{n-1}/C$ is prohibited to be contained in any $(n-2)$-hyperplane in $\mathbb{R}^{n-1}$, equivalently $W\dot{h} \neq 0$. Inversely, one can also show that given any W-regular curve in its normal parametrization, $g' = \{g_i'(x)\} \in \mathbb{F}^n$, so that (2.11) applies, there is $h = \{h_a(t)\} \in \mathbb{F}^{n-1}$, which "injects" it according to (2.10) and (2.14).

III. THE BASIC PROPERTIES OF $\mathscr{L}_m$ OPERATORS

Let $\mathscr{F}$ be the set of smooth functions of the two variables x and y. Consider then a sequence of nonlinear differential mappings $\mathscr{L}_m: \mathscr{F} \to \mathscr{F}$, $m = 1,2,\ldots$, defined by

$$F \in \mathscr{F} \Rightarrow \mathscr{L}_m \mathscr{F} := \det(\partial_x^k \partial_y^l F), \quad k,l = 0,1,\ldots,m \geqslant 1, \tag{3.1}$$

where ∂_x^k and ∂_y^l denote iterations of the differential operators ∂_x and ∂_y. Of course, $\partial_x^0 = \partial_y^0 = 1$.

It is convenient to extend the above definition of $\mathscr{L}_m$'s to all integer m's, postulating that

$$\begin{aligned} m = 0 \Rightarrow \mathscr{L}_m F := F, \quad m = -1 \Rightarrow \mathscr{L}_m F := 1, \\ m \leqslant -2 \Rightarrow \mathscr{L}_m F = 0. \end{aligned} \tag{3.2}$$

Here, we will outline the basic formal properties of $\mathscr{L}_m$'s. One can easily see that the definition of $\mathscr{L}_m$'s implies the "homogenity property":

$$\begin{aligned} F, A(x), B(y) \in \mathscr{F} \Rightarrow \mathscr{L}_m ABF = (AB)^{m+1} \mathscr{L}_m F, \\ m \geqslant -1, \end{aligned} \tag{3.3}$$

if $AB \neq 0$, applies for $m \leqslant -2$.

Then one can show that under the change of the independent variables $x = x(x')$, $y = y(y')$, $\dot{x}\dot{y} \neq 0$, meaning by $\mathscr{L}_m'$, $\mathscr{L}_m' F := \det(\partial_{x'}^k \partial_{y'}^l F)$, for $m \geqslant 1$, while (3.2) is valid with $\mathscr{L}_m \to \mathscr{L}_m'$, the following identity holds:

$$F \in \mathscr{F} \Rightarrow \mathscr{L}_m F = (\dot{x}\dot{y})^{-\binom{m+1}{2}} \mathscr{L}_m' F. \tag{3.4}$$

Next, one easily sees that $\mathscr{L}_1$ has the "distributive" property:

$$F, G \in \mathscr{F} \Rightarrow \mathscr{L}_1 FG = G^2 \mathscr{L}_1 F + F^2 \mathscr{L}_1 G. \tag{3.5}$$

(This follows directly from $\mathscr{F} \ni F \neq 0 \Rightarrow \mathscr{L}_1 F = F^2 \partial_x \partial_y \ln F$. Also, note that $\alpha = \text{const} \Rightarrow \mathscr{L}_1 F^\alpha = F^{2(\alpha-1)} \mathscr{L}_1 F$.)

As far as the composition of the $\mathscr{L}_m$ mappings is concerned, we claim that the basic identity

$$\mathscr{L}_1(\mathscr{L}_m F) = \mathscr{L}_{m-1} F \cdot \mathscr{L}_{m+1} F \tag{3.6}$$

is valid for every integer m. The nontrivial proof of (3.6) is outlined in Appendix B, where we also discuss the general problem of the composition $\mathscr{L}_n(\mathscr{L}_m F)$.

It follows from (3.6) and (3.5) that

$$\begin{aligned} \mathscr{L}_2(\mathscr{L}_m F) = (\mathscr{L}_{m-1} F)^2 \mathscr{L}_{m+2} F \\ + (\mathscr{L}_{m+1} F)^2 \mathscr{L}_{m-2} F. \end{aligned} \tag{3.7}$$

Indeed, (3.6) for $m = 1$ reduces to $\mathscr{L}_1(\mathscr{L}_1 F) = F \mathscr{L}_2 F$. Operating on both sides of (3.6) with $\mathscr{L}_1$ and then using the result on the left and (3.5) on the right, we have

$$\mathscr{L}_m F\cdot \mathscr{L}_2(\mathscr{L}_m F)$$
$$= (\mathscr{L}_{m-1}F)^2\mathscr{L}_1(\mathscr{L}_{m+1}F)^2\mathscr{L}_1(\mathscr{L}_{m-1}F)$$
$$= (\mathscr{L}_{m-1}F)^2\mathscr{L}_m F\cdot \mathscr{L}_{m+2}F + (\mathscr{L}_{m+1}F)^2\mathscr{L}_{m-2}F\cdot \mathscr{L}_m F \quad [\text{using (3.6)}]. \tag{3.8}$$

Canceling this by $\mathscr{L}_m F$ (in general $\mathscr{L}_m F \neq 0$), we obtain (3.7). Therefore, via the continuity argument, (3.7) is true for every F and every integer m.

Crucial for our purposes, we state the properties of the $\mathscr{L}_m$'s in the form of two theorems.

Theorem 3: For every $f_i(x)$ and $g_i(y)$ (smooth), $i = 1,2,...,n$ and $m \geqslant 1$, the identities

$$\mathscr{L}_m \sum_{i=1}^{n} f_i g_i = \begin{cases} 0, & \text{if } m \geqslant n \geqslant 1, \\ Wf\cdot Wg, & \text{if } n = m+1, \\ \sum_{i=1}^{n} {}^*f_i \cdot {}^*g_i, & \text{if } n = m+2, \end{cases} \tag{3.9}$$

hold, where W and $*$ are the mappings defined in Sec. I.

Employing the summation convention over the indices $k_i = 0,1,...,m = l_i$, we have, from the definition (3.1),

$$\mathscr{L}_m F = [1/(m+1)!]\epsilon_{k_1\cdots k_{m+1}} \times \epsilon_{l_1\cdots l_{m+1}}\partial_x^{k_1}\partial_y^{l_1}F\cdots\partial_x^{k_{m+1}}\partial_y^{l_{m+1}}F, \tag{3.10}$$

where the ϵ's are $(m+1)$-dimensional Levi-Civita symbols normalized by $\epsilon_{01\cdots m} = 1$. Consequently, with $F = f_i(x)g_i(y)$, we have

$$\begin{aligned}\mathscr{L}_m f_i g_i &= [1/(m+1)!]\epsilon_{k_1\cdots k_{m+1}}\epsilon_{l_1\cdots l_{m+1}} \times f_{i_1}^{[k_1}\cdots f_{i_{m+1}}^{k_{m+1}]} g_{i_1}^{[l_1}\cdots g_{i_{m+1}}^{l_{m+1}]} \\ &= (m+1)! f_{i_1}^{[0}\cdots f_{i_{m+1}}^{m]} g_{i_1}^{[0}\cdots g_{i_{m+1}}^{m]} \\ &= (m+1)! f_{[i_1}^{0}\cdots f_{i_{m+1}}^{m} g_{[i_1}^{0}\cdots g_{i_{m+1}]}^{m}. \end{aligned} \tag{3.11}$$

In the above, $f_i^k := (d/dx)^k f_i$, $g_i^l := (d/dy)^l g_i$, and $[\cdots]$ denotes the antisymmetrization symbol of a set of indices.

When the i's have the range $i = 1,...,n < m+1$, the antisymmetrization of $m+1$ of the indices of the above type automatically leads to 0. Thus the first line of (3.9) is true. On the other hand, if the i's have the range $i = 1,...,n = m+1$, then according to (2.1)

$$f_{[i_1}^0\cdots f_{i_{m+1}]}^m = [1/(m+1)!]\epsilon_{i_1\cdots i_{m+1}}\cdot Wf.$$

Similarly,

$$g_{i_1}^0\cdots g_{i_{m+1}}^m = [1/(m+1)!]\epsilon_{i_1\cdots i_{m+1}}\cdot Wg.$$

Therefore, making the contraction of two ϵ's over $m+1$ indices, the second line of (3.9) follows from (3.11).

Finally, if the range of i's is $i = 1,...,n = m+2$, (3.11) can be rewritten employing the concept of the generalized Kronecker δ's in the form

$$\mathscr{L}_m f_i g_i = \delta_{i_1\cdots i_{m+1};j_1\cdots j_{m+1}} f_{i_1}^0\cdots f_{i_{m+1}}^m g_{j_1}^0\cdots g_{j_{m+1}}^m. \tag{3.12}$$

On the other hand, the δ's are equivalent to the contraction of two ϵ's over one index. Hence

$$\mathscr{L}_m f_i g_i = e_{ii_1\cdots i_{m+1}}\epsilon_{ij_1\cdots j_{m+1}} f_{i_1}^0\cdots f_{i_{m+1}}^m g_{j_1}^0\cdots g_{j_{m+1}}^m. \tag{3.13}$$

This, compared with the definition of $*$ in (2.1), and remembering that presently $n = m+2$, assures the veracity of the third line of (3.9).

Corollary 1: Formula (3.12) remains valid for the range of i's and j's over $1,...,n > m+2$, and can be equivalently spelled out in the form

$$\mathscr{L}_m f_i g_i = [1/(n-m-1)!] \times \epsilon_{s_1\cdots s_{n-m-1}i_1\cdots i_{m+1}} f_{i_1}^0\cdots f_{i_{m+1}}^m \times \epsilon_{s_1\cdots s_{n-m-1}j_1\cdots j_{m+1}} g_{j_1}^0\cdots g_{j_{m+1}}^m. \tag{3.14}$$

For $n > m+2$, the objects $\epsilon_{s_1\cdots s_{n-m-1}i_1\cdots i_{m+1}} f_{i_1}^0\cdots f_{i_{m+1}}^m$ are the generalized minors of the $n\times n$ matrix $|f_i^k|$, $i = 1,...,n$, $k = 0,1,...,n-1$. Within the objectives of this paper, however, this generalization of the third line of (3.9) is of little importance.

Corollary 2: Iterating the third line of (3.9) and using (2.2)(h), it follows that

$$F = \sum_{i=1}^{m+2} f_i(x)g_i(y) \Rightarrow \mathscr{L}_m(\mathscr{L}_m F) = (Wf\cdot Wg)^m F. \tag{3.15}$$

Therefore, $\mathscr{L}_m$ is an involution among the functions of two-variables of the structure

$$F = \sum_{i=1}^{m+2} f_i(x)g_i(y),$$

with f_i's and g_i's arbitrary, being constrained only by the condition $(Wf\cdot Wg)^m = 1$.

Theorem 4: For $m \geqslant 1$,

$$\begin{aligned} &\text{(a)}\ \ \mathscr{L}_m F = 0 \text{ and } \mathscr{L}_{m-1}F \neq 0 \Rightarrow F = \sum_{i=1}^{m} f_i(x)g_i(y), \\ &\text{(b)}\ \ \mathscr{L}_m F = \text{const} \neq 0 \Rightarrow F = \sum_{i=1}^{m+1} f_i(x)g_i(y), \end{aligned} \tag{3.16}$$

the implications being understood in the sense of the existence of the corresponding functions of the one variable, constrained in the case of (3.16)(a) by the condition $Wf\cdot Wg \neq 0$. Similarly, in the case of (3.16)(b) by $Wf\cdot Wg = \text{const}$.

The proof of (3.16)(a) is given in Appendix C. Once the veracity of (3.16)(a) is granted, a simple proof of (3.16)(b) follows by employing the identity (3.6).

Indeed, with $\mathscr{L}_m F = \text{const} \neq 0$, (3.6) with $m \geqslant 1$ obviously requires $\mathscr{L}_{m-1}F\cdot\mathscr{L}_{m+1}F = 0$. If this were to hold with $\mathscr{L}_{m-1}F = 0$, then according to (3.16)(a) F would have the most general form of

$$F = \sum_{i=1}^{m-1} f_i(x)g_i(x),$$

which then according to the first line of (3.9) leads to $\mathscr{L}_m F = 0$, contradicting $\mathscr{L}_m F = \text{const} \neq 0$. Therefore, $\mathscr{L}_{m+1}F = 0$, so that according to (3.16)(a) F has the most general form

$$F = \sum_{i=1}^{m+1} f_i(x)g_i(y).$$

But then according to the second line of (3.9) $\mathscr{L}_m F = Wf\cdot Wg = \text{const}$.

We conclude this section recognizing the fact that if $0 \neq F, G \in \mathcal{F}$ are related by the condition

$$\mathcal{L}_1 F = \mathcal{L}_1 G, \tag{3.17}$$

then, parametrizing equivalently these objects according to

$$F = e^{-\psi/2}\cosh(\phi/2), \quad G = e^{-\psi/2}\sinh(\phi/2), \quad \phi \neq 0, \tag{3.18}$$

condition (3.17) can be stated in the simple form of

$$\psi_{xy} = \phi_x \phi_y. \tag{3.19}$$

IV. THE LIOUVILLE-LIKE PDE's

This section examines some PDE's constructed by using the notion of the nonlinear differential operators $\mathcal{L}_m$, $m \geqslant 1$.

We define first as the basic chain of the Liouville-like PDE's for the searched $F \in \mathcal{F}$:

$$\mathcal{L}_m F = \epsilon, \quad \epsilon^2 = 1, \quad m = 1,2,\ldots . \tag{4.1}$$

These equations together with the associated "degenerate" chain

$$\mathcal{L}_m F = 0, \quad m = 1,2,\ldots , \tag{4.2}$$

include PDE's of the form

$$\mathcal{L}_m F = \text{const}, \quad m = 1,2,\ldots . \tag{4.3}$$

If F fulfills (4.3) with const $\neq 0$, then $F' = aF$, $a = \text{const}$, fulfills $\mathcal{L}_m F' = \text{const} \cdot a^{m+1}$. Therefore, choosing a properly and dropping the prime with const $\neq 0$, (4.3) reduces to (4.1). Obviously, the proper Liouville equation (1.2) constitutes the first member of the chain (4.1) for $m = 1$.

Now, we propose to consider a chain of PDE's for searched $F, G \in \mathcal{F}$,

$$\mathcal{L}_m F = G, \quad \mathcal{L}_m G = F, \quad m = 1,2,\ldots , \tag{4.4}$$

which generalize the "double" Liouville equations (1.4), equivalent to (1.5)(b).

It is also of some interest to comment from the point of view of this paper on the nature of the differential conditions for the searched $F \in \mathcal{F}$:

$$\mathcal{L}_m(\mathcal{L}_m F) = \text{const}, \quad m = 1,2,\ldots . \tag{4.5}$$

The first member of these PDE's for $m = 1$, $\mathcal{L}_1(\mathcal{L}_1 F) = \text{const}$ will be seen to be equivalent to the biharmonic equation for the conformal factor of a two-dimensional Riemannian space, with the harmonic scalar curvature.

Of course, among the PDE's proposed above, the case of Eqs. (4.2) is the simplest. According to (3.16)(a), the most general solution to $\mathcal{L}_m F = 0$—with $\mathcal{L}_{m-1} F \neq 0$—has the form of

$$F = \sum_{i=1}^{m} f_i(x) g_i(y), \quad \mathcal{L}_{m-1} F = Wf \cdot Wg \neq 0. \tag{4.6}$$

In the terminology of Sec. II, it induces and is induced by the two W-regular curves in $\mathbb{R}^m$, x and y playing the role of the *arbitrary* parameters of these curves prohibited to be contained in any $(m-1)$-hyperplanes through the origin of $\mathbb{R}^m$.

Notice that given F in the form of (4.6), the functions $f = \{f_i(x)\}$, $g = \{g_i(y)\} \in \mathbb{F}^m$ are meaningful modulo the affine transformations only:

$$f' = Mf, \quad g' = (M^{-1})^T g, \quad M = \text{const} \in \mathrm{GL}(m), \tag{4.7}$$

the matrices M being otherwise arbitrary.

Next, examining PDE's (4.1), we observe that according to (3.16)(b) the most general solution must have the form of

$$F = \sum_{i=1}^{m+1} f_i(x) g_i(y), \tag{4.8}$$

where $f = \{f_i(x)\}$, $g = \{g_i(y)\} \in \mathbb{F}^{m+1}$ [according to the second line of (3.9)] are constrained by the condition

$$Wf \cdot Wg = \epsilon, \tag{4.9}$$

and are otherwise arbitrary.

Again, given the solution to (4.1) in the form (4.8) and (4.9), one easily sees that f and g are meaningful modulo the transformations (4.7) only, but this time with the $(m+1)\times(m+1)$ matrix $M = \text{const} \in \mathrm{GL}(m+1)$. Condition (4.9) obviously requires that both Wf and Wg be constants $\neq 0$. Using then as the special case of the transformations (4.7), $f_i \to \lambda f_i$, $g_i \to \lambda^{-1} g_i$, $\lambda = \text{const} \neq 0$, one easily sees that, without losing any generality, we can always arrange that (4.9) constraining the general form of F from (4.8) is fulfilled with

$$(Wf)^2 = 1 = (Wg)^2. \tag{4.10}$$

But with the above being valid, according to the results of Sec. II, $f, g \in \mathbb{F}^{m+1}$ may be interpreted as the two W-regular curves in $\mathbb{R}^{m+1}$ given, respectively, in terms of their *normal parameters* x and y. Thus we can interpret f and g from $\mathbb{F}^{m+1}$ in (4.8) as the two *arbitrary* W-regular curves in $\mathbb{R}^{m+1}$, which are forbidden to be constrained in any m-hyperplanes through the origin of $\mathbb{R}^{m+1}$, as given in terms of their normal parameters. Notice that with this interpretation, the most general form of the solution to (4.1) as given by (4.8) constrained by (4.10), with f and g from $\mathbb{F}^{m+1}$, remain arbitrary modulo (4.7) transformations, where the $(m+1)\times(m+1)$ matrix $M = \text{const} \in \mathrm{GL}(m+1)$ is constrained by $(\det M)^2 = 1$.

With the objectives outlined in the Introduction in mind, the basic point of this section is that, according to the properties of the mapping ▶ given in Sec. II [i.e., that a W-regular curve in $\mathbb{R}^{m+1}$ (as given in terms of its normal parameter) can be *always algebraically* constructed from m smooth functions and their derivatives up to $m+1$ order], the most general solution to (4.1) can be equivalently stated in the form

$$F = (\epsilon W\dot{p} \cdot W\dot{q})^{-1/(m+1)}\left(1 + \sum_{a=1}^{m} p_a(x) q_a(y)\right), \quad m \geqslant 1. \tag{4.11}$$

The $p = \{p_a(x)\}$ and $q = \{q_a(y)\} \in \mathbb{F}^m$ in the above are two arbitrary smooth curves in $\mathbb{R}^m$, forbidden to be constrained in any $(m-1)$-hyperplanes such that $\epsilon W\dot{p} \cdot W\dot{q} > 0$. The solution is thus *algebraically* constructed from the *arbitrary* smooth functions $p_a(x)$, $q_a(y)$, $a = 1,\ldots,m$, and their derivative $\partial_x^k p$, $\partial_y^k q$, $k = 1,\ldots,m$. This most general form of the solution to (4.1) for $m = 1$ reduces precisely to (1.3), the classical result for the proper Liouville equation (1.2), *ipso facto* providing its proof, and hence

(4.11) is a natural generalization of (1.3) for the case of $\mathscr{L}_m F = \epsilon$ PDE's.

Corollary 1: The result that the most general solution to (4.1) has the form (4.11), i.e., that F from (4.11) satisfies (4.1) and inversely, given F fulfilling (4.1), there are $p,q \in \mathbb{F}^m$ such that (4.11) is true, is easily seen to admit a "complexification."

For simplicity, until now we have constrained in this text $\mathbb{F}^n$ to be the set of real valued ordered n-tuples of smooth functions of the same real variable, and $\mathscr{F}$ to be the set of real valued smooth functions of the real variables x and y.

A moment of reflection, however, convinces us that if $\mathbb{F}^n$ is interpreted as the set of the complex valued ordered n-tuples of the holomorphic functions of the same complex variable, and, correspondingly, $\mathscr{F}$ is interpreted as the set of the complex valued holomorphic functions of the complex variables x and y, then the most general solution to the "complexified" PDE (4.1), indeed has the form of (4.11), with the holomorphic functions $\{p_a(x)\},\{q_a(y)\} \in \mathbb{F}^m$ constrained by $\epsilon W\underline{p} \cdot W\underline{q} \neq 0$, being otherwise arbitrary.

Corollary 2: The result that we are in possession of the most general solution to the PDE (4.1), either in its real or complexified version, has some interesting implications from the point of view of ODE's.

Suppose that we search the solution to (4.1) in the special case of $F = F(z)$, $z := x + y$. Then (4.1) reduces to the nonlinear ODE of $2m$ differential order,

$$\begin{vmatrix} F, & \frac{d}{dz}F \cdots \left(\frac{d}{dz}\right)^m F \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ \left(\frac{d}{dz}\right)^m F, & \left(\frac{d}{dz}\right)^{m+1} F \cdots \left(\frac{d}{dz}\right)^{2m} F \end{vmatrix} = \epsilon, \quad \epsilon^2 = 1. \tag{4.12}$$

Our general result on the level of PDE (4.1) permits us to construct easily the explicit general solution to the ODE (4.12) as endowed with the $2m$ integration constants.

Indeed, if F from (4.8) depends only on $z = x + y$, clearly $f_i(x)$ and $g_i(y)$ must have the form of

$$f_i = \sqrt{\beta_i} e^{\alpha_i x}, \quad g_i = \sqrt{\beta_i} e^{\alpha_i y}, \qquad F = \sum_{i=1}^{m+1} \beta_i e^{\alpha_i z}, \tag{4.13}$$

where $\alpha_i, \beta_i, i = 1,\ldots,m+1$, are constants. With f_i's and g_i's of this form, one easily sees that

$$Wf = M(\beta_1 \cdots \beta_{m+1})^{1/2} e^{(\alpha_1 + \cdots + \alpha_{m+1})x}, \qquad Wg = M(\beta_1 \cdots \beta_{m+1})^{1/2} e^{(\alpha_1 + \cdots + \alpha_{m+1})y}, \tag{4.14}$$

where

$$M := \begin{vmatrix} 1 & \cdots & 1 \\ \alpha_1 & & \alpha_{m+1} \\ \cdots & \cdots & \cdots \\ \alpha_1^m & \cdots & \alpha_{m+1}^m \end{vmatrix} \tag{4.15}$$

is the Van der Mond determinant, so that the (4.9) condition amounts to

$$\beta_1 \cdots \beta_{m+1} M^2 e^{(\alpha_1 + \cdots + \alpha_{m+1})z} = \epsilon. \tag{4.16}$$

This necessitates for $2(m+1)$ constants α_i and β_i the two conditions

$$\alpha_1 + \cdots + \alpha_{m+1} = 0, \quad \beta_1 \beta_2 \cdots \beta_{m+1} M^2 = \epsilon, \tag{4.17}$$

the second of these requiring obviously $M \neq 0$, so that necessarily $i \neq j \Rightarrow \alpha_i - \alpha_j \neq 0$. Modulo conditions (4.17), F from (4.13) solves (4.12), contains $2(m+1) - 2 = 2m$ arbitrary constants, and hence is the most general solution of the nonlinear ODE (4.12) of the $2m$'s differential order.

Perhaps one could guess the general shape of the solution to (4.12) in the form of (4.13) with the constants constrained by (4.17) *prima facie*, but in establishing this result, our knowledge of the most general form of the solution to the PDE (4.1) was certainly useful.

Now the case of the PDE's (4.4) is much more involved than the case of PDE's of the form (4.3), where we have succeeded in establishing the most general form of their solutions as endowed with a geometric interpretation. Since we are interested only in the nontrivial solutions to these equations, $F \neq 0 \neq G$ [because (4.4) necessitates $F = 0 \Leftrightarrow G = 0$], (i) by eliminating G, we arrive at the necessary condition

$$\mathscr{L}_m(\mathscr{L}_m F) = F, \quad m \geqslant 1, \tag{4.18}$$

and (ii) if we constrain additionally the searched F and G by $\mathscr{L}_{m+2} F = 0 \Leftrightarrow \mathscr{L}_{m+2} G = 0$, then Eqs. (4.4) admit a *special solution* of the form

$$F = \sum_{i=1}^{m+2} f_i(x) g_i(y), \quad G = \sum_{i=1}^{m+2} {}^*f_i \cdot {}^*g_i, \tag{4.19}$$

where $f,g \in \mathbb{F}^{m+2}$ are constrained by

$$(Wf)^2 = 1 = (Wg)^2, \tag{4.20}$$

and are otherwise arbitrary. This statement applies for $m \geqslant 1$.

Indeed, (4.18) is a trivial consequence of (4.4). On the other hand, with F of the form of (4.19) treated as an anzatz, according to $G = \mathscr{L}_m F$ and the third line of (3.9), G must have the form of 4. But then according to (3.15), $F = \mathscr{L}_m G = \mathscr{L}_m(\mathscr{L}_m F) = (Wf \cdot Wg)^m F$ is also fulfilled iff $(Wf\, Wg)^m = 1$. This is equivalent to $Wf \cdot Wg = \epsilon$, $\epsilon^2 = 1$, with $m =$ odd, ϵ constrained to the value $\epsilon = 1$. Rescaling $f_i \to \lambda f_i, g_i \to \lambda^{-1} g_i, \lambda = \text{const} \neq 0$, we can always arrange that the last condition be fulfilled with (4.20) being valid.

The special solution to (4.4) described by (4.19) and (4.20) with $Wf \cdot Wg = \epsilon$, $\epsilon^m = 1$, has of course a parallel interpretation to that given before to the solutions to $\mathscr{L}_m F = \epsilon$, with $m \to m+1$. Thus F and G are induced by two W-regular curves in $\mathbb{R}^{m+2}$ prohibited to be contained in any $(m+1)$-hyperplane through the origin, with x and y serving as their normal parameters. From (4.19), we observe that F may also be interpreted as given in the form of (4.11) with $m \to m+1$. Of course, the corresponding $G = \mathscr{L}_m F$ can then be evaluated in terms of $\{p_a(x)\}, \{q_a(y)\} \in \mathbb{F}^{m+1}$. Observe also that F and G from (4.19), with $f,g \in \mathbb{F}^{m+2}$, have the relative symmetric structure, compatible with the symmetry $F \to G$, $G \to F$ of Eqs. (4.2), due to the involutory relations (2.6) and (2.7), which apply because of (4.20).

The question arises, "How general is the solution (4.19) and (4.20) to (4.4)?" Answering this, we claim that, for $m = 1$, the solution constructed above is the most general solution to (4.4), which arose from a problem in general

relativity and motivated our interest in the chain of PDE's (4.4).

Indeed, in the case of $m = 1$, (4.4) reduces to (1.5)(b) and the identity (3.6) reduces to

$$\mathscr{L}_1(\mathscr{L}_1 F) = F\mathscr{L}_2 F. \tag{4.21}$$

Consequently, (4.18) with $F \neq 0$ is equivalent to

$$\mathscr{L}_2 F = 1, \tag{4.22}$$

which was shown to possess the most general solution

$$F = \sum_{i=1}^{3} f_i(x) g_i(y),$$

$Wf \cdot Wg = 1$, equivalently, $Wf = \epsilon = Wg$, $\epsilon^2 = 1$. But with $n = 3$, according to (2.5), the normal parameters can be so selected that the ϵ above is constrained to the value $\epsilon = 1$. Thus the most general solution to (1.5)(b) has the form of

$$F = \sum_{i=1}^{3} f_i(x) \cdot g_i(y), \quad G = \sum_{i=1}^{3} {}^*f_i \cdot {}^*g_i, \qquad Wf = 1 = Wg. \tag{4.23}$$

For the W-regular curves $f, g \in \mathbb{F}^3$ given in terms of their normal parameters x and y, and with the first line of (2.6) being in $*$ involution to the dual curves ${}^*f, {}^*g \in \mathbb{F}^3$, ${}^{**}f = f$ and ${}^{**}g = g$.

The above is a rather nice result. Our original problem from general relativity admits the most explicit general solution endowed with a simple geometric interpretation. The solution to (1.5)(b) induces—and is induced by—the two arbitrary W-regular smooth curves in $\mathbb{R}^3$ given in terms of their normal parametrizations; F and G are constructed from these and their $*$ dual curves. According to (4.7), given F and G, f and g from $\mathbb{F}^3$ are determined, remembering that $Wf = 1 = Wg$, and, because of (2.2)(c), are arbitrary modulo (4.7) transformations with 3×3 constant $= M \in \mathrm{SL}(3)$. The geometric interpretation given above is thus meaningful modulo SL(3) transformations of $\mathbb{R}^3$. Of course, F from (4.23) can be also represented in the form of

$$F = (Wp\dot{W}q)^{-1/3}\left(1 + \sum_{a=1}^{2} p_a(x) q_a(y)\right), \tag{4.24}$$

while $G = \mathscr{L}_1 F$ can be elaborated in terms of $p, q \in \mathbb{F}^2$, accompanied by the corresponding geometric interpretation.

However, the argument considered above for the case of Eqs. (4.4) with $m = 1$ does not work in the case of these equations with $m \geqslant 2$. Consider, e.g., the case of (4.4) with $m = 2$. Condition (4.18), employing identity (3.7) specialized for $m = 2$, reduces to

$$\mathscr{L}_2(\mathscr{L}_2 F) = (\mathscr{L}_1 F)^2 \mathscr{L}_4 F + (\mathscr{L}_3 F)^2 F = F. \tag{4.25}$$

Thus, when $\mathscr{L}_{m+2} F = \mathscr{L}_4 F = 0$, indeed $F \neq 0$ constrained by $\mathscr{L}_3 F = \pm 1$ is a solution. However, there is no *a priori* reason why $\mathscr{L}_4 F$ should be equal to zero. Similarly, there is no *a priori* reason why $\mathscr{L}_{m+2} F$ has to be equal to zero for $m > 2$.

In summary, we have established the most general solution to (4.4) for $m = 1$, and a nontrivial solution for $m \geqslant 2$. In the last case, the form of the general solution remains an open question.

Corollary: With $F =: e^{\phi}$, $G =: e^{\psi}$, (1.5)(b) assume the equivalent form of

$$\phi_{xy} = e^{\psi - 2\phi}, \quad \psi_{xy} = e^{\phi - 2\psi}. \tag{4.26}$$

On the other hand, for $m = 2$, (4.4) using the identity (4.21), are equivalent to

$$\mathscr{L}_1(\mathscr{L}_1 F) = FG = \mathscr{L}_1(\mathscr{L}_1 G), \tag{4.27}$$

or, with $F := e^{\phi}$, $\mathscr{L}_1 F := e^{\Omega}$ and $G := e^{\psi}$, $\mathscr{L}_1 G := e^{\chi}$ they assume the equivalent form of

$$\phi_{xy} = e^{\Omega - 2\phi}, \quad \Omega_{xy} = e^{\phi + \psi - 2\Omega}, \qquad \psi_{xy} = e^{\chi - 2\psi}, \quad \chi_{xy} = e^{\phi + \psi - 2\chi}. \tag{4.28}$$

Similarly, (4.4) arbitrary $m > 1$ can be equivalently stated as a set of differential conditions of the second order, with the nonlinear terms involving the notion of exponentials.

Concluding this section, we should like to explain why these PDE's are of some interest in mathematical physics. Given a two-dimensional Riemannian space of signature $(+,-)$ in its conformally flat local representation in a chart $\{x, y\}$,

$$\Lambda^1 \underset{s}{\otimes} \Lambda^1 g := 2\phi^{-2}(x, y) dx \underset{s}{\otimes} dy,$$

the condition that its scalar curvature R is harmonic, $R^{;\alpha}{}_{;\alpha} = 0$, is easily seen to be equivalent to the biharmonic equation for the conformal factor, $\phi^{;\alpha}{}_{;\alpha}{}^{;\beta}{}_{;\beta} = 0$, amounting to

$$e^{2\phi} \partial_x \partial_y (e^{2\phi} \partial_x \partial_y \phi) = 0. \tag{4.29}$$

The above PDE is equivalent to the statement that the searched ϕ fulfills

$$\mathscr{L}_1 e^{\phi} \equiv e^{2\phi} \partial_x \partial_y \phi = A(x) - B(y), \tag{4.30}$$

where A and B are arbitrary smooth functions of one variable, in the general case such that $\dot{A}\dot{B} \neq 0$.

The differential problem (4.30) had emerged as relevant in general relativity in 1962,[7] and, as is well known, constitutes the key to the general nontwisting solutions of the Petrov type III of the empty space-time Einstein equations. In a somewhat different context one should also see Brans.[8]

Up to now, Eq. (4.30) resists all attempts to construct its most general analytic solution. A *special* solution to (4.30) of the form

$$e^{\phi} = \sqrt{\tfrac{2}{3}}\,(A - B)^{3/2}/\sqrt{\dot{A}\dot{B}} \tag{4.31}$$

is well-known.

From the point of view of this paper, we observe first that with $\dot{A}\dot{B} \neq 0$, introducing in (4.30) the new independent variables $x' = A(x)$, $y' = B(y)$ and defining $F := \exp[\phi + \frac{1}{2} \ln \dot{A}\dot{B}]$ after dropping out primes, the investigated PDE assumes the form of

$$\mathscr{L}_1 F = x - y. \tag{4.32}$$

It follows that $\mathscr{L}_1(\mathscr{L}_1 F) = 1$, which coincides with (4.5) for $m = 1$. This motivates our interest in the PDE's from the chain (4.5).

Of course, using identity (4.21), (4.32) implies

$$F\mathscr{L}_2 F = 1. \tag{4.33}$$

Notice that if we define $F =: \sqrt{\tfrac{2}{3}} F'^{3/2}$, then (4.32) assumes the form

$$F' \mathscr{L}_1 F' = x - y. \tag{4.34}$$

Acting on it with $\mathscr{L}_1$ and using (3.5) and (4.21), we infer the necessity of

$$F'^3\mathscr{L}_2F' + (\mathscr{L}_1F')^3 = 1. \tag{4.35}$$

This elucidates why $F': x - y \Rightarrow \mathscr{L}_1F' = 1$ is a special solution, i.e., the mechanism of the solution (4.31), as stated in a slightly more general form.

On the other hand, if, instead of committing the independent variables to $x = A, y = B$, we just execute in (4.30) the transformation $x = x(x'), y = y(y'), \dot{x}\dot{y} \neq 0$, one easily sees that by dropping out primes and with $F := e^{\phi}$ (4.30) assumes the form

$$\mathscr{L}_1F = \sum_{i=1}^{2} k_i(x)l_i(y), \tag{4.36}$$

while the condition $\dot{A}\dot{B} \neq 0$ is now equivalent to $Wk \cdot Wg \neq 0$.

It easily follows that the differential problem studied in its most general branch and in coordinates arbitrary modulo $x = x(x'), y = y(y'), \dot{x}\dot{y} \neq 0$ is equivalent to the conditions

$$\begin{aligned} &\mathscr{L}_2F \neq 0 \Rightarrow \mathscr{L}_1F \neq 0 \Rightarrow F \neq 0, \\ &\mathscr{L}_2(\mathscr{L}_1F) = F^2\mathscr{L}_3F + (\mathscr{L}_2F)^2 = 0. \end{aligned} \tag{4.37}$$

Notice that because of (4.21), with $\mathscr{L}_1F \neq 0$, $\mathscr{L}_2(\mathscr{L}_1F) = 0$ is also equivalent to

$$\mathscr{L}_1(\mathscr{L}_1(\mathscr{L}_1F)) = 0. \tag{4.38}$$

V. CONCLUDING REMARKS

The PDE's studied in this paper constructed with the help of $\mathscr{L}_m$ nonlinear differential operators are certainly of interest as they generalize in a natural manner the Liouville equation.

Using the properties of Wronskians, we were able to find the general solutions to some of these PDE's, i.e., Eqs. (4.3) and (4.4). The latter one especially is of great importance because of its role in the problem of type N spaces. We hope that further analysis of the problems presented here allows one to find solutions to much more involved cases.

The Liouville equation has been revealed as the important one in the study of the Born–Infeld massless scalar field and in the theory of relativistic strings.[9,10] Note also that this equation in three and more dimensions is of interest in connection with the soliton and field theories.[11–13]

We hope that our generalizations of the Liouville equation will find application not only in general relativity but also in many other domains of mathematical physics.

APPENDIX A: PROOFS OF (2.2)(d) AND (2.2)(h) IDENTITIES

The basic difficulty in proving (2.2)(d) and (2.2)(h) for every $n \geqslant 2$ is due to the "proliferation" of ϵ's and the order of derivatives involved in the concepts of $**$ and $W(*f)$. More specifically, there is no obvious way to initiate the inductive process with respect to $n \geqslant 2$, and the usual combinatorics of ϵ's and related Kronecker generalized δ's cannot deal with the mentioned "proliferation" in an effective manner. Our proof will rely on some facts from the theory of linear ODE's, and the formal properties of the minors of the matrix of the Wronskian $\|f_i^k\|$, $i = 1,...,n$, $k = 0,1,...,n-1$, where $\{f_i(t)\} \in \mathbb{F}^n$. Within this proof, some "tangential" formal properties of the elements of $\mathbb{F}^n$, related to the mapping $*$, will emerge as of interest as such.

The *minors* of the Wronskian are defined for every $\{f_i(t)\} \in \mathbb{F}^n$ by

$$M_i^k := (-1)^k \epsilon_{ip_1\cdots p_k q_1 \cdots q_{n-k-1}} f_{p_1} f^1_{p_2} \cdots f^{k-1}_{q_1} \cdots f^{n-1}_{q_{n-k-1}},$$
$$i = 1,...,n, \quad k = 0,1,..,n-1, \tag{A1}$$

where the summation convention over p's and q' applies. (Furthermore, the symbols M_i^{-1} and M_i^n, where the upper index exceeds the permitted range $k = 0,1,...,n-1$, are to be understood as zero.) This definition assures us that

$$f_i^k M_i^l = \delta^{kl} \cdot Wf \tag{A2}$$

and the parallel

$$\sum_{k=0}^{n-1} f_i^k M_j^k = \delta_{ij} \cdot Wf, \tag{A3}$$

with the obvious ranges for the free indices.[14]

Observe that according to (2.1)

$$M_i^{n-1} = *f_i. \tag{A4}$$

[Formulas (A1) can be interpreted for $k = 0,1,...,n-1$ as defining for $k = 0,1,...,n-1$ the nonlinear differential mappings $\mathscr{M}^k: \mathbb{F}^n \to \mathbb{F}^n$. For our purposes the most important is the mapping $\mathscr{M}^{n-1} = *$, induced by the "basic minors" of the Wronskian.]

Observe also that with $\|M_i^k\|$ being the matrix of the minors of $\|f_i^k\|$ and $\det(f_i^k) = Wf$, an elementary identity

$$\det(M_i^k) := \epsilon_{i_1 \cdots i_n} M^0_{i_1} \cdots M^{n-1}_{i_n} = (Wf)^{n-1} \tag{A5}$$

holds, consistent with equalities (A2) and (A3).[15]

After these comments concerned with the definitions and the basic properties of the minors of the Wronskian, we will now prove the following.

Lemma 1: Given any $f = \{f_i(t)\} \in \mathbb{F}^n$, there is $A := \{A_i(t)\} \in \mathbb{F}^n$ such that

$$\mathscr{O}f_i = 0, \quad \mathscr{O} := \partial_t^n - \sum_{j=1}^{n} A_j \partial_t^{j-1}, \tag{A6}$$

and, if f_i's are linearly independent ($\Leftrightarrow Wf \neq 0$), then

$$A_n = \partial_t \ln(Wf). \tag{A7}$$

■ Suppose that $Wf \neq 0$. Then $h_i := |Wf|^{-1/n} F_i = \blacktriangleright f_i$, according to (2.2)(a), has the property $Wh = \text{const} \neq 0$. This differentiated ∂_t amounts explicitly to

$$\epsilon_{i_1 \cdots i_n} h_{i_1} \cdots h^{n-2}_{i_{n-1}} h^n_{i_n} = 0, \tag{A8}$$

where $h_i^k := (d/dt)^k h_i$, $k = 0,1,...$. From the properties of ϵ it follows then that the "vector" h_i^n must be a linear combination of "vectors" h_i^k, $k = 0,1,...,n-2$. Therefore, there is $A' = \{A'_j(t)\} \in \mathbb{F}^{n-1}$ such that

$$h_i^n - \sum_{j=1}^{n-1} A'_j h_i^{j-1} = 0. \tag{A9}$$

By substituting here $h_i = |Wf|^{-1/n} f_i$ and applying the Leibnitz rule for ∂_t^k acting on a product of two functions, one easily verifies that (A9) reduces to (A6), with A_n having the form of (A7).

Let $Wf = 0$, so that the n of f_i's are linearly independent. Also, let f_i, $i = 1,...,n-1$, be linearly independent, and hence there are $\lambda_1 = \text{const} \neq 0$ such that $f_n = \Sigma_{i=1}^{n-1} \lambda_i f_i$.

Repeating the argument given above, we infer the existence of $A'' = \{A_j''(t)\} \in \mathbb{F}^{n-1}$ such that, similar to (A6),

$$f_i^{n-1} - \sum_{j=1}^{n-1} A_j'' f_i^{j-1} = 0, \quad i = 1,...,n-1. \tag{A10}$$

But then $f_n = \Sigma_{j=1}^{n-1} \lambda_i f_i$ must *a forteriori* satisfy the same ODE (A10). Then acting on (A10) with ∂_t, we conclude that there is $A \in \mathbb{F}^n$ such that (A6) is valid. A trivial descending induction implies that it does not matter how many of n f_i's are linearly independent. There is *always* $A \in \mathbb{F}^n$ such that (A6) is true for *any* $\{f_i(t)\} \in \mathbb{F}^n$. ■■

Lemma 2: We claim that, modulo the existence of $\{A_i(t)\} \in \mathbb{F}^n$ established in Lemma 1, such that (A6) is valid for every $\{f_i(t)\} \in \mathbb{F}^n$, it is true that

$$(\partial_t - A_n) M_i^k = -M_i^{k-1} - A_{k+1} M_i^{n-1}, \tag{A11}$$

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

■ Indeed, differentiating ∂_t using the definition (A1) and remembering the total skewness of ϵ, we have

$$\partial_t M_i^k = (-1)^k \epsilon_{ip_1 \cdots p_k q_1 \cdots q_{n-k-1}} \times \{ f_{p_1} \cdots f_{p_{k-1}}^{k-2} f_{p_k}^k f_{q_1}^{k+1} \cdots f_{q_{n-k-1}}^{n-1} + f_{p_1}^0 \cdots f_{p_k}^{k-1} f_{q_1}^{k+1} \cdots f_{q_{n-k-2}}^{n-2} f_{q_{n-k-1}}^n \}. \tag{A12}$$

The term in the first line of the right-hand member of this equality, according to (A1) amounts simply to "$-M_i^{k-1}$." In the term from the second line, by using (A6) and remembering the total skewness of ϵ, we can replace

$$f_{q_{n-k-1}}^n \to \sum_{j=1}^{n} A_j f_{q_{n-k-1}}^{j-1} \to A_n f_{q_{n-k-1}}^{n-1} + A_{k+1} f_{q_{n-k-1}}^k.$$

Consequently, the contribution from the second line amounts to "$A_n M_i^k - A_{k+1} M_i^{n-1}$" and hence (A11) is true. ■■

Let now $L := A_n - \partial_t$ and $g_i := {}^*f_i \equiv M_i^{n-1}$. Then (A11) assumes the form of

$$L M_i^k = M_i^{k-1} + A_{k+1} g_i. \tag{A13}$$

A trivial induction establishes then that this implies for every non-negative integer l:

$$L^l M_i^k = M_i^{k-l} + \left(\sum_{s=0}^{l-1} L^{l-1-s} A_{k+1-s} \right) g_i. \tag{A14}$$

Specializing this for $k = n-1$, we have

$$L^l g_i = M_i^{n-1-l} + \left(\sum_{s=0}^{l-1} L^{l-1-s} A_{n-s} \right) g_i. \tag{A15}$$

The last relation leads to the next lemma.

Lemma 3: There are the smooth functions $B_s^l(t)$ such that

$$g_i^l = (-1)^l M_i^{n-1-l} + \sum_{s=0}^{l-1} B_s^l g_i^s. \tag{A16}$$

■ With (A15) valid, remembering $L := A_n - \partial_t$, (A16) is obviously true for $l = 0,...,n-1$. ■■ This established, we propose, as the last lemma needed, the following.

Lemma 4: It is true that

$$**f_i = *g_i := (-1)^{n-1} \epsilon_{ii_1 \cdots i_{n-1}} g_i g_{i_2}^1 \cdots g_{i_{n-1}}^{n-2} = (-1)^{n-1} \epsilon_{ii_1 \cdots i_{n-1}} M_{i_1}^1 \cdots M_{i_{n-1}}^{n-1}. \tag{A17}$$

■ Indeed, according to (A16), and remembering the total skewness of ϵ, the last factor in the first line of (A17) can be replaced by $g_{i_{n-1}}^{n-2} \to (-1)^{n-2} M_i^1$, the contributions from B_s^{n-2} canceling out. By a parallel argument, proceeding from the right to the left, the second factor can be replaced by $g_{i_{n-2}}^{n-3} \to (-1)^{n-3} M_i^2$, the contributions from B_s^{n-3} canceling out. Proceeding inductively this way we end up with $g_{i_1} \to (-1)^0 M_{i_1}^{n-1}$. Therefore, the right-hand member of the first line of (A17) amounts to

$$**f_i = (-1)^{n-1} \cdot (-1)^{0+1+\cdots+(n-2)} e_{ii_1 \cdots i_{n-1}} \cdot M_{i_1}^{n-1} M_{i_2}^{n-2} \cdots M_{i_{n-1}}^1. \tag{A18}$$

Permuting now the factors from the last line to the opposite order, $M_{i_1}^1 \ldots M_{i_{n-1}}^{n-1}$ and remembering the total skewness of ϵ, we conclude that the equality of $**f_i$ to the second line of (A17) is true. ■■

With the established veracity of Lemmas (1–4), the proof of identities (2.2)(d) and (2.2)(h) for any $n \geqslant 2$ is now very simple. ■ Indeed, by contracting the equality of $**f_i$ to the second line of (A17) with M_i^l, l arbitrary, we have

$$M_i^l \cdot **f_i = (-1)^{n-1} \epsilon_{i_1 \cdots i_n} M_{i_1}^0 \cdots M_{i_n}^{n-1} \cdot \delta^{l0} = (-1)^{n-1} \det(M_i^k) \cdot \delta^{l0} = (-1)^{n-1} (Wf)^{n-1} \cdot \delta^{l0} \quad [\text{via (A5)}]. \tag{A19}$$

Then multiplying this equality by f_j^l and taking the sum $\Sigma_{l=0}^{n-1}$,

$$(-1)^{n-1} (Wf)^{n-1} f_j = \left(\sum_{i=0}^{n-1} f_j^l M_i^l \right) \cdot **f_i = (Wf) \cdot **f_i \quad [\text{using (A3)}]. \tag{A20}$$

Therefore, if $Wf \neq 0$, necessarily

$$**f_i = (-1)^{n-1} (Wf)^{n-2} \cdot f_i. \tag{A21}$$

Via the continuity argument, this also must hold with f such that $Wf \to 0$, so that (2.2)(h) is true for every $n \geqslant 2$.

Similarly, we now easily prove (2.2)(d). Indeed,

$$W(*f) = \epsilon_{i_1 \cdots i_n} g_{i_1}^0 \cdots g_{i_n}^{n-1} = (-1)^{(n-1)+(n-2)+\cdots+1} \epsilon_{i_1 \cdots i_n} M_{i_1}^{n-1} \times \cdots M_{i_n}^0 \quad [\text{using (A16)}] = \epsilon_{i_1 \cdots i_n} M_{i_1}^0 \cdots M_{i_n}^{n-1} = (Wf)^{n-1} \quad [\text{using (A5)}]. \tag{A22}$$

Therefore, (2.2)(d) and (2.2)(h) are identities for every integer $n \geqslant 2$. Also, Theorem 1 is true. ■■

Corollary: Equality (A15) specialized for $l = n$ yields

$$L^n g_i = \left(\sum_{s=0}^{n-1} L^{n-1-s} A_{n-s} \right) g_i = \left(\sum_{j=1}^{n} L^{j-1} A_j \right) g_i.$$

Therefore, while to an arbitrary $\{f_i(t)\} \in \mathbb{F}^n$ according to Lemma 1, there is associated a linear ODE

$$\mathcal{O} f_i = 0, \quad \mathcal{O} := \partial_t{}^n - \sum_{j=1}^{n} A_j \, \partial_t{}^{j-1}. \tag{A23}$$

the functions $g_i := *f_i$ must fulfill the somewhat conjugated linear ODE:

$$\mathcal{O}^* g_i = 0 \quad \mathcal{O}^* := L^n - \sum_{j=1}^{n} L^{j-1} A_j, \tag{A24}$$

where $L := A_n - \partial_t$. If $Wf = \text{const} \Leftrightarrow A_n = 0$, one can show that the operator $\mathcal{O}^*$ is just the *adjoint* of $\mathcal{O}$ in the standard sense, that, for *every* $f, g \in \mathbb{F}$,

$$g \cdot \mathcal{O} f - \mathcal{O}^* g \cdot f = \partial_t h, \tag{A25}$$

where

$$h = \sum_{k,l=0}^{n-1} h_{kl}(t) f^k(t) g^l(t),$$

h_{kl} independent of f and g.

Perhaps the notion of $\mathcal{O}^*$ as "conjugated" to $\mathcal{O}$ may also be of interest when $A_n \neq 0$, but we shall not investigate this point in the present text.

APPENDIX B: PROOF OF (3.6) IDENTITY

Consider a matrix M_{ij}, with the entries in a commutative field of numbers of characteristic zero, indexed by $i, j = 1,...,m =$ a positive integer. Understanding $\epsilon_{i_1 \cdots i_m} = \epsilon_{[i_1 \cdots i_m]}$ as the totally skew m-dimensional Levi-Civita symbol normalized by $\epsilon_{1 \cdots m} = 1$, and by $\delta_{i_1 \cdots i_k; j_1 \cdots j_k}$ as the generalized Kronecker δ's, and assuming the summation convention, we have the basic identity

$$\delta_{i_1 \cdots i_k; j_1 \cdots j_k} = [1/(m-k)!] \epsilon_{i_1 \cdots i_k s_{k+1} \cdots s_m} \epsilon_{j_1 \cdots j_k s_{k+1} \cdots s_k}, \quad k = 0,1,...,m. \tag{B1}$$

Moreover, the notion of the determinant of the matrix M_{ij} then has the role of a coefficient in the identities:

$$\begin{aligned} &\text{(a)} \ \epsilon_{i_1 \cdots i_m} M_{i_1 j_1} \cdots M_{i_m j_m} = \det(M_{pq}) \epsilon_{j_1 \cdots j_m}, \\ &\text{(b)} \ \epsilon_{j_1 \cdots j_m} M_{i_1 j_1} \cdots M_{i_m j_m} = \det(M_{pq}) \epsilon_{i_1 \cdots i_m}. \end{aligned} \tag{B2}$$

The *generalized minors* of the matrix M_{ij} are then defined as

$$\begin{aligned} m_{i_1 \cdots i_k; j_1 \cdots j_k} &:= [1/(m-k)!] \epsilon_{i_1 \cdots i_k r_{k+1} \cdots r_m} \epsilon_{j_1 \cdots j_k s_{k+1} \cdots s_m} \\ &\quad \cdot M_{r_{k+1} s_{k+1}} \cdots M_{r_m s_m}, \quad k = 0,1,...,m. \end{aligned} \tag{B3}$$

Notice that for $k = 0$, $m = \det(M_{pq})$, for $k = 1$, $m_{i;j}$ are the minors of matrix in the conventional sense, and for $k = m$, $m_{i_1 \cdots i_m; j_1 \cdots j_m} = \delta_{i_1 \cdots i_m; j_1 \cdots j_m}$.

Using this definition, one easily establishes with the help of (A1) and (A2) that

$$\begin{aligned} &\text{(a)} \ M_{s_1 i_1} \cdots M_{s_k i_k} m_{s_1 \cdots s_k; j_1 \cdots j_k} = \det(M_{pq}) \cdot \delta_{i_1 \cdots i_k; j_1 \cdots j_k}, \\ &\text{(b)} \ M_{i_1 s_1} \cdots M_{i_k s_k} m_{j_1 \cdots j_k; s_1 \cdots s_k} = \det(M_{pq}) \cdot \delta_{i_1 \cdots i_k; j_1 \cdots j_k}, \\ &\qquad\qquad k = 0,1,...,m. \end{aligned} \tag{B4}$$

These general rules imply that, in particular for $k = 1$,

$$M_{si} m_{s;j} = \det(M_{pq}) \cdot \delta_{ij} = M_{is} m_{j;s}, \tag{B5}$$

where $\delta_{ij} = \delta_{i;j}$ are the standard Kronecker δ's. For $k = 2$, we have

$$\begin{aligned} M_{s_1 i_1} M_{s_2 i_2} m_{s_1 s_2; j_1 j_2} &= \det(M_{pq}) \delta_{i_1 i_2; j_1 j_2} \\ &= M_{i_1 s_1} M_{i_2 s_2} m_{j_1 j_2; s_1 s_2}. \end{aligned} \tag{B6}$$

By contracting the first line of (B6) with $m_{k_1 i_1} m_{k_2 i_2}$ we obtain

$$[\det(M_{pq})]^2 m_{k_1 k_2; j_1 j_2} = \det(M_{pq}) \cdot \delta_{i_1 i_2; j_1 j_2} m_{k_1 i_1} m_{k_2 i_2}. \tag{B7}$$

This, canceled by $\det(M_{pq})$, in general $\neq 0$ via the continuity argument, and, remembering that

$$\delta_{i_1 i_2; j_1 j_2} = \delta_{i_1 j_1} \delta_{i_2 j_2} - \delta_{i_1 j_2} \delta_{i_2 j_1},$$

leads to the identity

$$\det(M_{pq}) \cdot m_{i_1 i_2; j_1 j_2} = m_{i_1; j_1} \cdot m_{i_2; j_2} - m_{i_1; j_2} \cdot m_{i_2; j_1}, \tag{B8}$$

which is essential for our purposes.

Now using the traditional notation of $|\cdots|$ for the determinant of a matrix, one easily can show that the identity

$$\begin{vmatrix} M_{ij} & A_i \\ B_j & 0 \end{vmatrix} = -m_{i;j} A_i B_j \tag{B9}$$

holds, with i enumerating the entries into the rows and j into the columns of the determinant of the $(m+1) \times (m+1)$ matrix in the left-hand member of the identity above. The M_{ij}, A_i, and B_j are *arbitrary* with $i, j = 1,...,m$, in the above.

Slightly more difficult to demonstrate using the concepts above is the identity

$$\begin{vmatrix} M_{ij} & A_i & C_i \\ B_j & 0 & 0 \\ D_j & 0 & 0 \end{vmatrix} = m_{i_1 i_2; j_1 j_2} A_{i_1} C_{i_2} B_{j_1} D_{j_2}. \tag{B10}$$

Again, the i's enumerate the rows of the $(m+2) \times (m+2)$ matrix on the left-hand side, whose determinant is to be taken. Correspondingly, the j's enumerate the columns. Of course, in (B10), M_{ij}, A_i, C_i, B_j, D_j are *arbitrary* with $i, j = 1,...,m$.

Contracting now (B8) with $A_{i_1} C_{i_2} B_{j_1} D_{j_2}$ and using (B9) and (B10), we arrive at the identity

$$\begin{aligned} |M_{ij}| \cdot \begin{vmatrix} M_{ij} & A_i & C_i \\ B_j & 0 & 0 \\ D_j & 0 & 0 \end{vmatrix} &= \begin{vmatrix} M_{ij} & A_i \\ B_i & 0 \end{vmatrix} \cdot \begin{vmatrix} M_{ij} & C_i \\ D_j & 0 \end{vmatrix} - \begin{vmatrix} M_{ij} & A_i \\ D_j & 0 \end{vmatrix} \cdot \begin{vmatrix} M_{ij} & C_i \\ B_j & 0 \end{vmatrix}. \end{aligned} \tag{B11}$$

Now we claim, using the same notation, a more general identity

$$\begin{aligned} |M_{ij}| \cdot \begin{vmatrix} M_{ij} & A_i & C_i \\ B_j & E & G \\ D_j & H & F \end{vmatrix} &= \begin{vmatrix} M_{ij} & A_i \\ B_j & E \end{vmatrix} \cdot \begin{vmatrix} M_{ij} & C_i \\ D_j & F \end{vmatrix} - \begin{vmatrix} M_{ij} & A_i \\ D_j & H \end{vmatrix} \cdot \begin{vmatrix} M_{ij} & C_i \\ B_j & G \end{vmatrix}, \end{aligned} \tag{B12}$$

with M_{ij}, A_i, C_i, B_j, D_j, E, F, G, and H arbitrary.

Indeed, (B12) is true because one easily sees via an elementary argument that the coefficients of the arbitrary E, F, G, and H are the same on both sides of (B12). It follows that for (B12) to hold it is sufficient to verify its validity with

$E = F = G = H = 0$. Of course, this amounts to the already established (B11).[16]

With the identity (B12), we can identify the entries of the $(m+2)\times(m+2)$ determinant in the left-hand member of (B12) with the entries of the determinant for $\alpha_{m+1}F$ according to the scheme

$$\begin{vmatrix} F\cdots\partial_x^{m-1}F & \partial_x^m F & \partial_x^{m+1}F \\ \cdots & \cdots & \cdots \\ \partial_y^{m-1}F\cdots\partial_x^{m-1}\partial_y^{m-1}F & \partial_x^m\partial_y^{m-1}F & \partial_x^{m+1}\partial_y^{m-1}F \\ \partial_y^m F\cdots\partial_x^{m-1}\partial_y^m F & \partial_x^m\partial_y^m F & \partial_x^{m+1}\partial_y^m F \\ \partial_y^{m+1}F\cdots\partial_x^{m-1}\partial_y^{m+1}F & \partial_x^m\partial_y^{m+1}F & \partial_x^{m+1}\partial_y^{m+1}F \end{vmatrix} =: \begin{vmatrix} M_{ij} & A_i & C_i \\ B_j & E & G \\ D_j & H & F \end{vmatrix}. \tag{B13}$$

Using the definitions above, one easily sees that

$$\begin{aligned} \mathcal{L}_1(\mathcal{L}_m F) &= \mathcal{L}_m F\cdot\partial_x\partial_y\mathcal{L}_m F - \partial_y\mathcal{L}_m F\cdot\partial_x\mathcal{L}_m F \\ &= \begin{vmatrix} M_{ij} & A_i \\ B_j & E \end{vmatrix}\cdot\begin{vmatrix} M_{ij} & C_i \\ D_j & F \end{vmatrix} \\ &\quad - \begin{vmatrix} M_{ij} & A_i \\ D_j & H \end{vmatrix}\cdot\begin{vmatrix} M_{ij} & C_i \\ B_j & G \end{vmatrix}. \end{aligned} \tag{B14}$$

By comparing (B13) and (B14) with the identity (B12), we conclude that for every $F\in\mathcal{F}$ and $m = 1,2,\ldots$ it is true that

$$\mathcal{L}_1(\mathcal{L}_m F) = \mathcal{L}_{m-1}F\mathcal{L}_{m+1}F. \tag{B15}$$

It is now convenient to formally extend the validity of this identity for all integers m. Understanding $\mathcal{L}_0 F := F$, it is natural to define

$$\mathcal{L}_{-1}F := 1, \quad \mathcal{L}_{-2}F = \mathcal{L}_{-3}F = \cdots := 0, \tag{B16}$$

which consistently assures us that (B15) is indeed valid for all integer m's.

For $m = 1$, (B15) amounts to

$$\mathcal{L}_1(\mathcal{L}_1 F) = F\mathcal{L}_2 F. \tag{B17}$$

Substituting the above $F\to\mathcal{L}_m F$, we have

$$\mathcal{L}_1(\mathcal{L}_1(\mathcal{L}_m F)) = \mathcal{L}_m F\cdot\mathcal{L}_2(\mathcal{L}_m F). \tag{B18}$$

On the other hand, using the representation of $\mathcal{L}_1 F$ in the form of $F^2\mathcal{L}_x\mathcal{L}_y \ln F$, one easily finds that the operation $\mathcal{L}_1$ has the property

$$F,G\in\mathcal{F} \Rightarrow \mathcal{L}_1(FG) = G^2\mathcal{L}_1 F + F^2\mathcal{L}_1 G. \tag{B19}$$

Consequently, acting on both sides of (B15) with $\mathcal{L}_1$ and employing (B18) and (B19), we have

$$\begin{aligned} \mathcal{L}_m F\cdot\mathcal{L}_2(\mathcal{L}_m F) &= (\mathcal{L}_{m-1}F)^2\mathcal{L}_1(\mathcal{L}_{m+1}F) \\ &\quad + (\mathcal{L}_{m+1}F)^2\cdot\mathcal{L}_1(\mathcal{L}_{m-1}F) \\ &= (\mathcal{L}_{m-1}F)^2\mathcal{L}_m F\cdot\mathcal{L}_{m+2}F \\ &\quad + (\mathcal{L}_{m+1}F)^2\cdot\mathcal{L}_{m-2}F\cdot\mathcal{L}_m F \\ &\quad [\text{using (B15)}]. \end{aligned} \tag{B20}$$

Canceling this by $\mathcal{L}_m F$ in general $\neq 0$, via the continuity argument, we arrive at the identity

$$\begin{aligned} \mathcal{L}_2(\mathcal{L}_m F) &= (\mathcal{L}_{m-1}F)^2\cdot\mathcal{L}_{m+2}F \\ &\quad + (\mathcal{L}_{m+1}F)^2\cdot\mathcal{L}_{m-2}F, \end{aligned} \tag{B21}$$

which is valid for all integers m. In particular,

$$\mathcal{L}_2(\mathcal{L}_1 F) = F^2\mathcal{L}_3 F + (\mathcal{L}_2 F)^2 \tag{B22}$$

and

$$\mathcal{L}_2(\mathcal{L}_2 F) = (\mathcal{L}_1 F)^2\mathcal{L}_4 F + (\mathcal{L}_3 F)^2 F. \tag{B23}$$

Notice that a direct proof of (B21) based only on the definition of $\mathcal{L}_m$ operations, without employing (B15) and the (B19) property of $\mathcal{L}_1$, would be highly nontrivial. In the initial stages of this paper, (B22) has been proved in particular via the direct computation, with the assistance of Dr. Alberto García-Díaz, whose help is gratefully appreciated. At this stage of trying to find some formal properties of $\mathcal{L}_k(\mathcal{L}_l F)$, with at least one of the indices being an *arbitrary* integer, this question appeared to be an extremely messy algebraical problem.

The general problem of the result of the iteration of the $\mathcal{L}_m$ operations, i.e., $\mathcal{L}_k(\mathcal{L}_l F) = ?$, to which we have now the answer for $k = 1,2$ and l arbitrary, remains as a rather nontrivial algebraic problem.

One easily sees from the definition of $\mathcal{L}_m$ that

$$\mathcal{L}_m F = \mathcal{L}_{m-1}F\cdot\partial_x^m\partial_y^m F + \cdots, \tag{B24}$$

where "$\cdots$" denotes the terms constructed from the derivatives of F of the differential order $\leqslant 2m-1$.

It follows that

$$\begin{aligned} \mathcal{L}_n(\mathcal{L}_m F) &= \mathcal{L}_{m-1}F\cdot\mathcal{L}_{n-1}(\mathcal{L}_m F) \\ &\quad \cdot\partial_x^{n+m}\partial_y^{n+m}F + \cdots, \end{aligned} \tag{B25}$$

where "$\cdots$" denotes the terms algebraically constructed from the derivatives of F of the differential order $\leqslant 2(n+m)-1$. If we wish to express this statement in terms of the $\mathcal{L}_m$ operators, multiplying (B25) by $\mathcal{L}_{n+m-1}F$ and employing (B24), we arrive at

$$\begin{aligned} &\mathcal{L}_{n+m-1}F\cdot\mathcal{L}_n(\mathcal{L}_m F) \\ &\quad = \mathcal{L}_{m-1}F\cdot\mathcal{L}_{n-1}(\mathcal{L}_m F)\cdot\mathcal{L}_{n+m}F + \cdots. \end{aligned} \tag{B26}$$

The dots denote the terms algebraically constructed from the derivatives of F of the differential order $\leqslant 2(n+m)-1$.

On the basis of intuitive arguments, we conjecture that the "$\cdots$" terms described above consist of the algebraic constructs made of $\mathcal{L}_m F$, $m = 0,1,\ldots,k\leqslant 2(m+n)-1$. We believe that it would be of interest to determine the explicit form of the "$\cdots$" terms in (B26), determining this way the "algebra" of the composition of the $\mathcal{L}_m$ operators, $(\mathcal{L}_n\circ\mathcal{L}_m)F := \mathcal{L}_n(\mathcal{L}_m F)$, which is obviously *associative*.[17]

APPENDIX C: PROOF OF (3.16a)

■ Using the traditional explicit notation for the determinants for $\mathcal{L}_m F$ as defined by (3.1), we have

$$\mathcal{L}_m F := \begin{vmatrix} F, & \partial_x F, & \ldots, & \partial_x^m F \\ \partial_y F, & \partial_y\partial_x F, & \ldots, & \partial_y\partial_x^m F \\ \cdots & \cdots & \cdots & \cdots \\ \partial_y^m F, & \partial_y^m\partial_x F, & \ldots, & \partial_y^m\partial_x^m F \end{vmatrix}. \tag{C1}$$

Since the derivative of a determinant equals the sum of

determinants with initial columns successively differentiated, we also have

$$\partial_x \mathscr{L}_m F = \begin{vmatrix} F, & \partial_x F & \partial_x^{m-1} F, & \partial_x^{m+1} F \\ \partial_y F, & \partial_y \partial_x F & \partial_y \partial_x^{m-1} F, & \partial_y \partial_x^{m+1} F \\ \cdots & \cdots & \cdots & \cdots \\ \partial_y^m F, & \partial_y^m \partial_x F & \partial_y^m \partial_x^{m-1} F, & \partial_y^m \partial_x^{m+1} F \end{vmatrix}.$$

Now $\mathscr{L}_m F = 0$ implies that the last column of (C1) must be a linear combination of the first m columns, i.e., there are $A_j(x,y), j = 1,...,m$, such that

$$\mathscr{L}_m F = 0 \Rightarrow \partial_y^k \partial_x^m F = \sum_{j=1}^m A_j \partial_y^k \partial_x^{j-1} F, \quad k = 0,...,m. \tag{C2}$$

Similarly, if $\partial_x \mathscr{L}_m F = 0$, via the same argument there are $A_j'(x,y), j = 1,...,m$, such that

$$\partial_x \mathscr{L}_m F = 0$$

$$\Rightarrow \partial_y^k \partial_x^{m+1} F = \sum_{j=1}^m A_j' \partial_y^k \partial_x^{j-1} F, \quad k = 0,...,m. \tag{C3}$$

In particular, (C2) for $k = 0$ is given by

$$\mathscr{L}_m F = 0 \Rightarrow \partial_x^m F = \sum_{j=1}^m A_j \partial_x^{j-1} F. \tag{C4}$$

Acting on the above with ∂_y^k and applying Leibnitz's rule,

$$\mathscr{L}_m F = 0$$

$$\Rightarrow \partial_y^k \partial_x^m F = \sum_{j=1}^m \sum_{l=0}^k \binom{k}{l} \partial_y^{k-l} A_j \cdot \partial_y^l \partial_x^{j-1} F,$$

$$k = 1,...,m. \tag{C5}$$

However, the term in the right-hand side cancels with the left-hand member when the summation index l equals k because of (C2). Hence

$$\mathscr{L}_m F = 0 \Rightarrow \sum_{j=1}^m \sum_{l=0}^{k-1} \binom{k}{l} \partial_y^{k-l} A_j \cdot \partial_y^l \partial_x^{j-1} F = 0,$$

$$k = 1,...,m. \tag{C6}$$

A parallel argument applies to the case of (C3). Specializing for $k = 0$, we have

$$\partial_x \mathscr{L}_m F = 0 \Rightarrow \partial_x^{m+1} F = \sum_{j=1}^m A_j' \partial_x^{j-1} F, \tag{C7}$$

and acting on it with ∂_y^k,

$$\partial_x \mathscr{L}_m F = 0$$

$$\Rightarrow \partial_y^k \partial_x^{m+1} F$$

$$= \sum_{j=1}^m \sum_{l=0}^k \binom{k}{l} \partial_y^{k-l} A_j' \cdot \partial_y^l \partial_x^{j-1} F,$$

$$k = 1,...,m. \tag{C8}$$

Similarly, the entry in the summation over l, $l = k$, because of (C3) cancels out with the left-hand member, and we are left with the conditions

$$\partial_x \mathscr{L}_m F = 0$$

$$\Rightarrow \sum_{j=1}^m \sum_{l=0}^{k-1} \binom{k}{l} \partial_y^{k-l} A_j' \cdot \partial_y^l \partial_x^{j-1} F = 0, \quad k = 1,...,m, \tag{C9}$$

which, with $A_j' \to A_j$, formally coincide with (C6).

Lemma: Conditions (C6) imply that, for every $k = 1,...,m$,

$$h_{k,l} := \sum_{j=1}^m \partial_y^{k-l} A_j \cdot \partial_y^l \partial_x^{j-1} F = 0, \quad \text{for every } l = 0,1,...,k-1. \tag{C10}$$

We prove the above by induction. For $k = 1$, (C6) reduces to

$$\sum_{j=1}^m \partial_y A_j \cdot \partial_x^{j-1} F = 0$$

and (C10) is true, $h_{1,0} = 0$. Assume then (C10) for some $k \to k_0$, $1 \leqslant k_0 < m$, i.e., $h_{k_0,l} = 0$, $l = 0,...,k_0 - 1$. However, according to the definition of $h_{k,l}$: $\partial_y h_{k,l} \equiv h_{i_0+1,l} + h_{k_0+1,l+1}$. Consequently,

$$h_{k_0+1,l} + h_{k_0+1,l+1} = 0, \quad l = 0,1,...,k_0 - 1. \tag{C11}$$

It follows that

$$h_{k_0+1,l} = (-1)^l h_{k_0+1,0}, \quad l = 0,1,...,k_0. \tag{C12}$$

On the other hand, (C6) specialized for $k \to k_0 + 1$ requires

$$\sum_{l=0}^{k_0} \binom{k_0+1}{l} h_{k_0+1,l} = 0. \tag{C13}$$

Using (C12), because

$$\sum_{l=0}^{k_0} (-1)^l \binom{k_0+1}{l} = (1-1)^{k_0+1} - (-1)^{k_0+1} = (-1)^{k_0},$$

we see that this condition implies $h_{k_0+1,0} = 0$. So that, according to (C12),

$$h_{k_0+1,l} = 0, \quad l = 0,1,...,k_0, \tag{C14}$$

which completes the inductive proof of (C10).

Now we are sufficiently prepared to demonstrate the veracity of (3.16)(a). Indeed, with $\mathscr{L}_m F = 0$, the condition (C6) according to (C10) implies

$$h_{k,k-1} = 0, \quad k = 1,...,m, \text{ i.e.,}$$

$$\sum_{j=1}^m \partial_y A_j \cdot \partial_y^{k-1} \partial_x^{j-1} F = 0, \quad k = 1,...,m. \tag{C15}$$

But the matrix $\|\partial_y^{k-1} \partial_x^{j-1} F\|$, $k, j = 1,...,m$, has determinant equal to $\mathscr{L}_{m-1} F$. Therefore, if it is assumed that $\mathscr{L}_{m-1} F \neq 0$, the matrix is invertible. Then (C15) implies $\partial_y A_j = 0 \Rightarrow A_j = A_j(x)$. It follows that F must satisfy (C4), which reduces to a linear ODE:

$$\left(\partial_x^m - \sum_{j=1}^m A_j(x) \partial_x^{j-1}\right) F = 0. \tag{C16}$$

Understanding by $f_i(x)$, $i = 1,...,m$, the linearly independent solutions to this linear ODE, F must have the form of

$$F = \sum_{i=1}^m f_i(x) g_i(y), \tag{C17}$$

where the g_i's are "integration constants." Therefore (3.16)(a) is true. ■

[1]Y. Liouville, J. Math. Pure Appl. **18**, 71 (1853); see also A. R. Forsyth, *Theory of Differential Equations* (Dover, New York, 1959), Vol. VI.

[2]F. A. E. Pirani, D. C. Robinson, and W. F. Shadwick, *Local Jet Bundle Formulation of Bäcklund Transformations* (Reidel, Boston, 1979).

[3]"Automatically" in a parallel sense, e.g., a differential manifold automatically carries the Grassman–Cartan algebra.

[4]A reader unaccustomed to work with Levi-Civita ϵ's and generalized Kronecker δ's may find the reference J. L. Synge and A. Schild, *Tensor Calculus* (Dover, New York, 1978) useful. Of course, in (2.1) the summation convention is assumed. Since we do not need to distinguish "covariant" and "contravariant" indices, the last convention is understood as applying with respect to the indices.

[5]S. Hoene-Wroński was perhaps the first who was aware of this interpretation.

[6]The condition $\lambda_i f_i(t) = 0$, with smooth f_i's and $\lambda_i = \text{const}$, $\Sigma_{i=0}^{n} |\lambda_i| \neq 0$ implies $\lambda_i f_i^k = 0$, $k = 0,1,...,m-1$. Hence $\det(f_i^k) = Wf$ must vanish. If $Wf = 0$, the implication from the text follows.

[7]I. Robinson and A. Trautman, Proc. R. Soc. London Ser. A **265**, 463 (1962).

[8]C. H. Brans, J. Math. Phys. **12**, 1616 (1971).

[9]B. M. Barbashov, V. V. Nesterenki, and A. M. Chervyakov, J. Phys. A **13**, 301 (1980).

[10]B. M. Barbashov, V. V. Nesterenko, and A. M. Chervyakov, Commun. Math. Phys. **84**, 471 (1982).

[11]G. Leibbrant, S. Wang, and N. Zamani, J. Math. Phys. **23**, 1566 (1982).

[12]G. Leibbrant, "Do Liouville's solutions possess topological charge?," in *Wave Phenomena: Modern Theory and Applications*, edited by C. Roger and T. B. Moodie (North-Holland, Amsterdam, 1984).

[13]Y. Matsuno, J. Math. Phys. **28**, 2317 (1987).

[14]The notion of $f_i^k := (d/dt)^k f_i$ forces k to have the range $k = 0,1,...,n-1$ in (A2). The summation conversion applies with respect to indices written on "covariant" level. In (A3) we have used the explicit sum symbol dealing with "contravariant" indices. The validity of (A2) is a direct consequence of (A1). Equation (A2) is in the notation specialized for the considered case, with M_i^k being the minors of $\| f_i^k \|$, an elementary property of the minors of a matrix.

[15]By taking the determinant of (A2), we have $Wf \cdot \det(M_i^k) = (Wf)^n$. Thus if $Wf \neq 0$, $\det(M_i^k) = (Wf)^{n-1}$. By the continuity argument, this must be valid also with $Wf \to 0$.

[16]Perhaps (B12) also may be of interest as an algorithm that facilitates the evaluation of $(m+2) \times (m+2)$ determinants of rank m, reducing it to the evaluation of the $(m+1) \times (m+1)$ determinants.

[17]A hint for the reader potentially interested in this point: the basic difficulty in generalizing $\mathcal{L}_n(\mathcal{L}_m F) =$ known result in terms of $\mathcal{L}_k F$, for $n = 1,2$, amounts to the fact that while $\mathcal{L}_1$ has the "distributive" property (B19), it seems that $\mathcal{L}_2(F \cdot G)$ cannot be directly expressed as algebraically constructed from $\mathcal{L}_m F$, $m = 0,1,2$.

An infinite sum of products of Jacobi polynomials

R. A. Farrell, E. P. Gray, and R. W. Hart
The Johns Hopkins University, Applied Physics Laboratory, Laurel, Maryland 20707

(Received 6 January 1987; accepted for publication 22 June 1988)

Starting with a formula by Noble [Proc. Cambridge Philos. Soc. **59**, 363 (1963), Eq. (16)] for a certain sum of products of Jacobi polynomials, another sum of this type is evaluated.

I. INTRODUCTION

In the course of our (as yet unfinished) investigation of the inviscid, incompressible flow of a rotating fluid shell confined between concentric, spherical, corotating rigid walls, we encountered a certain sum of products of Jacobi polynomials. We needed to evaluate

$$P(\mu,l,x)=\sum_{k=0}^{\infty}\frac{(2k+\mu+l)}{(k+\mu)(k+l+\mu)}\times P_k^{(\mu+l,-1)}(x)\,P_{k+l}^{(\mu-l,-1)}(x), \tag{1a}$$

where l is a non-negative integer. The $P_n^{(\alpha,\beta)}(x)$ are the Jacobi polynomials[1] given by

$$P_n^{(\alpha,\beta)}(x)=[(\alpha+1)_n/n!]\times{}_2F_1[n+\alpha+\beta+1,-n;1+\alpha;(1-x)/2]$$

$$=\frac{(\alpha+1)_n}{n!}\sum_{s=0}^{n}\frac{(n+\alpha+\beta+1)_s(-n)_s}{s!(1+\alpha)_s}\times\left(\frac{1-x}{2}\right)^s, \tag{1b}$$

where $(z)_n=\Gamma(z+n)/\Gamma(z)$. In this paper it is shown that

$$P(\mu,l,x)=[((1-x)/2)^{-\mu}/\mu]\delta_{0,l}\quad(-1\leqslant x<1,\mu\neq 0), \tag{2}$$

where $\delta_{0,l}$ is the Kronecker delta function.

To evaluate $P(\mu,l,x)$ for $l\geqslant 1$, it was convenient to evaluate first a more general sum of products of Jacobi polynomials, which can then be related to P by considering a special case. The evaluation of $P(\mu,0,x)$ requires a separate treatment.

II. GENERAL FORMULA

We shall evaluate the sum

$$Q(a,b;l,m;x)=\sum_{k=0}^{\infty}\frac{(k+m)!\,\Gamma(a+b-l+k+2m)(2k+2m-l+a+b)}{\Gamma(a+k+m+1)\Gamma(b+k+m+1)}P_{k+m}^{(a-l-1,b)}(x)P_k^{(a+m,b-l+m-1)}(x), \tag{3}$$

where l and m are non-negative integers.

We start with a formula by Noble,[2]

$$S(a,b,c;x,y)=\sum_{k=0}^{\infty}\frac{k!(c+1)_k(2k+c+1)}{(a+1)_k(b+1)_k}P_k^{(c-b,b)}(x)P_k^{(a,c-a)}(y)$$

$$=\frac{\Gamma(a+1)\Gamma(b+1)}{\Gamma(c+1)\Gamma(a+b-c)}2^{c+1}(1+x)^{-b}(1-y)^{-a}(x-y)^{a+b-c-1}\Delta(x,y), \tag{4}$$

where

$$\Delta(x,y)=\begin{cases}1, & \text{if } -1\leqslant y<x\leqslant 1,\\ 0, & \text{if } -1\leqslant x<y\leqslant 1.\end{cases}$$

Setting $a+b-c-1=l$ (a non-negative integer), and differentiating m times with respect to y, we obtain in the limit $y\to x$

$$R(a,b;l,m;x)=\lim_{y\to x}\frac{\partial^m}{\partial y^m}S(a,b,a+b-l-1;x,y)$$

$$=\begin{cases}0, & \text{if } m\leqslant l,\\ \dfrac{\Gamma(a+1)\Gamma(b+1)\Gamma(a+m-l)m!}{\Gamma(a+b-l)\Gamma(a)l!(m-l)!}(-1)^l2^{a+b-l}(1+x)^{-b}(1-x)^{-a-m+l}, & \text{if } m\geqslant l,\end{cases} \tag{5}$$

for $-1<x<1$. We now use the formula for differentiating the Jacobi polynomials,[1]

 0022-2488/88/102175-03$02.50

$$\frac{d^m}{dy^m} P_k^{(\alpha,\beta)}(y) = 2^{-m}(\alpha+\beta+k+1)_m P_{k-m}^{(\alpha+m,\beta+m)}(y), \tag{6}$$

to obtain

$$R(a,b;l,m;x) = 2^{-m} \sum_{k=m}^{\infty} \frac{k!(a+b-l)_{k+m}(2k+a+b-l)}{(a+1)_k (b+1)_k} P_k^{(a-l-1,b)}(x) P_{k-m}^{(a+m,b-l+m-1)}(x). \tag{7}$$

Here we have used the recursion relation satisfied by the $P_n^{(\alpha,\beta)}(x)$,

$$\begin{aligned} &2(n+1)(n+\alpha+\beta+1)(2n+\alpha+\beta)P_{n+1}^{(\alpha,\beta)}(x) \\ &\quad = (2n+\alpha+\beta+2)(2n+\alpha+\beta+1)(2n+\alpha+\beta)xP_n^{(\alpha,\beta)}(x) + (2n+\alpha+\beta+1)(\alpha^2-\beta^2)P_n^{(\alpha,\beta)}(x) \\ &\qquad - 2(n+\alpha)(n+\beta)(2n+\alpha+\beta+2)P_{n-1}^{(\alpha,\beta)}(x), \end{aligned} \tag{8}$$

to verify that $P_n^{(\alpha,\beta)}(x) = 0$ when n is a negative integer. Shifting the summation index in Eq. (7) and noting that R is proportional to Q, we obtain finally

$$Q(a,b;l,m;x) = \begin{cases} 0, & \text{if } m<l, \\ (-1)^l 2^{a+b+m-l}\binom{m}{l}(a)_{m-l}(1+x)^{-b}(1-x)^{-a-m+l}, & \text{if } m \geqslant l, \end{cases} \tag{9}$$

where $\binom{m}{l} = m!/l!(m-l)!$ is the binomial coefficient.

III. EVALUATION OF $P(\mu,l,x)$

A. Case $l \geqslant 1$

Since $l \geqslant 1$, the function $P(\mu,l,x)$ can be related to $Q(a,b;l,m;x)$ by setting $a = \mu+1$, $b=1$, and $m = l-1$. Then we have

$$Q(\mu+1,1;l,l-1;x) = \sum_{k=0}^{\infty} \frac{(2k+\mu+l)}{(k+l)(k+\mu+l)} \times P_k^{(\mu+l,-1)}(x) P_{k+l-1}^{(\mu-l,1)}(x) = 0, \tag{10}$$

because $m = l-1 < l$.

We can now express $P_{n-1}^{(\alpha,1)}(x)$ in terms of $P_n^{(\alpha,-1)}(x)$ as follows. Set $\beta = 0$ in the contiguous relations,[1]

$$(\alpha+\beta+2n)P_n^{(\alpha,\beta-1)}(x) = (\alpha+\beta+n)P_n^{(\alpha,\beta)}(x) + (\alpha+n)P_{n-1}^{(\alpha,\beta)}(x) \tag{11a}$$

and

$$\tfrac{1}{2}(\alpha+\beta+2n)(1+x)P_{n-1}^{(\alpha,\beta+1)}(x) = (\beta+n)P_{n-1}^{(\alpha,\beta)}(x) + nP_n^{(\alpha,\beta)}(x). \tag{11b}$$

Then for $n \geqslant 1$,

$$P_{n-1}^{(\alpha,0)}(x) + P_n^{(\alpha,0)}(x) = [(2n+\alpha)/2n](1+x)P_{n-1}^{(\alpha,1)}(x), \tag{12}$$

from Eq. (11b), and

$$P_{n-1}^{(\alpha,0)}(x) + P_n^{(\alpha,0)}(x) = [(\alpha+2n)/(\alpha+n)]P_n^{(\alpha,-1)}(x), \tag{13}$$

from Eq. (11a), so that for $n \geqslant 1$,

$$P_{n-1}^{(\alpha,1)}(x) = [2n/(\alpha+n)(1+x)]P_n^{(\alpha,-1)}(x). \tag{14}$$

Substituting this result in Eq. (10) with $n = k+l$ gives, for $l \geqslant 1$ and $\mu \neq 0$,

$$\begin{aligned} &Q(\mu+1,1,\mu+1;l,l-1;x) \\ &\quad = \frac{2}{(1+x)} \sum_{k=0}^{\infty} \frac{(2k+\mu+l)}{(k+\mu)(k+l+\mu)} \times P_k^{(\mu+l,-1)}(x) P_{k+l}^{(\mu-l,-1)}(x) \\ &\quad \equiv [2/(1+x)]P(\mu,l,x) = 0 \quad (-1<x<1). \end{aligned} \tag{15}$$

B. Case $l=0$

$P(\mu;0;x)$ can be evaluated by integrating $S(\mu+1,1,\mu+1;x,t)$ with respect to t from -1 to x. To evaluate $\int_{-1}^{x} P_k^{(\mu+1,0)}(t)dt$, we use the differentiation formula for Jacobi polynomials, Eq. (6), with $m=1$. This gives

$$\int_{-1}^{x} P_k^{(\mu+1,0)}(t)dt = \frac{2}{(k+\mu+1)} P_{k+1}^{(\mu,-1)}(x), \tag{16}$$

since $P_{k+1}^{(\mu,-1)}(-1) = 0$ from Eq. (14) for k a positive integer or zero. Thus integrating Eq. (4) yields

$$\begin{aligned} &\int_{-1}^{x} S(\mu+1,1,\mu+1;x,t)dt \\ &\quad = 2\sum_{k=0}^{\infty} \frac{(2k+\mu+2)}{(k+1)(\mu+k+1)} P_{k+1}^{(\mu,-1)}(x) P_k^{(\mu,1)}(x) \\ &\quad = \frac{4}{1+x} \sum_{k=1}^{\infty} \frac{(2k+\mu)}{(k+\mu)^2} [P_k^{(\mu,-1)}(x)]^2 \\ &\quad = 4/(1+x)[P(\mu;0;x) - (1/\mu)] \\ &\quad = [2^{\mu+2}/\mu(1+x)][(1-x)^{-\mu} - 2^{-\mu}] \quad (\mu \neq 0), \end{aligned} \tag{17}$$

where we have utilized Eq. (14) and have shifted the summation index. Finally, Eq. (17) yields the desired result,

$$P(\mu,0,x) = 2^{\mu}/\mu(1-x)^{\mu}. \tag{18}$$

While Eq. (5) requires $1<x<1$, Eqs. (15) and (18) hold also for $x=-1$ because,[1,3] as noted above, $P_n^{(\alpha,-1)}(-1) = 0$ for a positive integer, and $P_0^{(\alpha,-1)}(-1) = 1$.

ACKNOWLEDGMENTS

This work was supported in part by the Space and Naval Warfare Systems Command under Contract No. N00039-87-C-5301, and by the National Institute of Health under Grant No. EV01019.

[1]W. Magnus, F. Oberhettinger, and R. P. Soni, *Formulas and Theorems for the Special Functions of Mathematical Physics* (Springer, New York, 1966), pp. 209–216.

[2]B. Noble, Proc. Cambridge Philos. Soc. **59**, 363 (1963), Eq. (16). See also E. R. Hansen, *A Table of Series and Products* (Prentice–Hall, Englewood Cliffs, NJ, 1975), Eq. (45.5.3).

[3]We would like to thank the referee for pointing out that the result holds for $x = -1$.

Noether's theorem in symmetric stochastic calculus of variations

Tetsuya Misawa
Department of Applied Mathematics and Physics, Kyoto University, Kyoto 606, Japan

(Received 18 August 1987; accepted for publication 1 June 1988)

Quite recently, a symmetric stochastic calculus of variations was proposed to formulate canonical stochastic dynamics, which is an extension of Nelson's stochastic mechanics. In this article a "Noether's theorem" is formulated within this calculus of variations. Conservation laws of momentum, angular momentum, and energy are proved, which are related with the same laws in quantum mechanics.

I. INTRODUCTION

Several years ago, Yasue proposed the notion of stochastic calculus of variations[1-4] within Nelson's stochastic mechanics.[5-7] This calculus has been improved recently.[8] The new calculus is called "symmetric stochastic calculus of variations" (SSCV). "Symmetric" means that this stochastic calculus employs the time-symmetric semimartingale. On the stochastic least-action principle in the SSCV, Nelson's stochastic mechanics is reformulated in the Lagrangian formalism. We present a certain class of stochastic Lagrangian systems in the SSCV that are associated with solutions of Schrödinger equations,[8] as in Nelson's mechanics.

The aim of this paper is to set up a theorem similar to Noether's theorem within the SSCV on the model of an ordinary calculus of variations,[9] and to get thereby conservation laws for the stochastic Lagrangian systems mentioned above.

II. SSCV: SUMMARY

We start with a summary of the SSCV.[8] Let (Ω, B, P) be a base probability space and x a stochastic process in $\mathbf{R}^l$, i.e., a continuous mapping $t \to x(t)$ from a time interval I into the Hilbert space $H = L^2((\Omega,P) \to \mathbf{R}^l)$. We consider two filtrations indexed by I, B_t and $\tilde{B}_t$, with $B_s \subset B_t$ and $\tilde{B}_t \subset \tilde{B}_s$ for $s \leq t$, to which $x(t)$ is adapted. By hypothesis, $x(t)$ is a time-symmetric semimartingale.[2,4,10] Moreover, the process $x(t)$ has the two mean velocities

$$Dx(t) = \lim_{h \to 0+} h^{-1} E[x(t+h) - x(t) | B_t]$$

and

$$D_* x(t) = \lim_{h \to 0+} h^{-1} E[x(t) - x(t-h) | \tilde{B}_t],$$

where $E[\cdot|\beta]$ denotes the conditional expectation with respect to the σ algebra β. Let us denote the class of stochastic processes of the above-mentioned type by K. For $L \in C^2(\mathbf{R}^{2l+1} \to \mathbf{R}^1)$, a Lagrangian, and for each process $x(t)$ in K, we define the action functional J by

$$J[x] = E\left[\int_a^b L(x(t), D^\circ x(t), t)dt\right], \tag{1}$$

where $E[\cdot]$ is the absolute expectation and $a,b \in I$, $a < b$. In the expression (1), $D^\circ x(t)$ denotes the "symmetric mean derivative" defined by

$$D^\circ x(t) \equiv \tfrac{1}{2}(D + D_*)x(t).$$

We denote by Δ the totality of processes $z(t) = z(x(t),t)$, where $z = z(x,t)$ is any smooth $\mathbf{R}^l$-valued function vanishing identically for $t = a$ and b. We note that each process $z(t)$ in Δ also belongs to K. The process $x(t)$ in K is called a stationary process of the functional J, given by (1), if $\delta J[x](z)$, the first variation of the functional J in x on K, is equal to zero for any processes z in Δ. The following theorem describes the stochastic least-action principle in the SSCV.[8]

Theorem 1: A process $x(t)$ belonging to K is a stationary process if and only if, for the process $x(t)$, the following equation holds:

$$D^\circ \frac{\partial L}{\partial D^\circ x(t)} - \frac{\partial L}{\partial x(t)} = 0. \tag{2}$$

We call this equation (2) the stochastic Euler's equation. The proof of this theorem consists of computing $\delta J[x](z)$ by Taylor's expansion and using the following lemma.

Lemma 1 (Zheng–Meyer[10]): Let $x(t)$ and $y(t)$ be stochastic processes belonging to K. Then

$$\begin{aligned} &E[x(b)\cdot y(b) - x(a)\cdot y(a)] \\ &= E\left[\int_a^b (y(t)\cdot Dx(t) + x(t)\cdot D_* y(t))dt\right] \\ &= E\left[\int_a^b (x(t)\cdot Dy(t) + y(t)\cdot D_* x(t))dt\right]. \end{aligned} \tag{3}$$

III. A NOETHER'S THEOREM IN SSCV

Now we proceed to set up a theorem similar to Noether's theorem[10] in the SSCV. Suppose that one has processes $x(t)$ and $X(t)$ belonging to K and satisfying the following conditions.

(i) The process $x(t)$ is a stationary process of the functional J given by (1).

(ii) The process $X(t)$ is derived from $x(t)$ by the following transformation depending on a parameter ε:

$$X(t) = \phi(x(t),t;\varepsilon), \tag{4}$$

where $\phi(x,t;\varepsilon)$ is a smooth function differentiable with respect to ε, and for $\varepsilon = 0$ the ϕ becomes the identity transformation.

Theorem 2: Assume that both of the processes

$$\left(\frac{\partial L}{\partial D^\circ x(t)}\right)(x(t), D^\circ x(t), t)$$

and

 0022-2488/88/102178-03$02.50

$$\lambda(x(t),t)\equiv\left(\frac{d\phi}{d\varepsilon}\right)_{\varepsilon=0}(x(t),t)$$

belong to K. If, for arbitrary a and b, $\Delta J\equiv J[X]-J[x]$ [the variation of J induced by (4)] satisfies

$$\left(\frac{d\Delta J}{d\varepsilon}\right)_{\varepsilon=0}=0, \tag{5}$$

then

$$E\left[\left(\frac{\partial L}{\partial D\circ x(t)}\right)(x(t),D\circ x(t),t)\cdot\lambda(x(t),t)\right]$$

is constant along the stationary process $x(t)$.

Proof: Suppose that ε is a small quantity. Using Taylor's expansion with respect to ε and the stochastic Euler equation (2) for $x(t)$, we find that Eq. (5) turns out to be

$$E\left[\int_a^b\left(D\circ\frac{\partial L}{\partial D\circ x(t)}\lambda+\frac{\partial L}{\partial D\circ x(t)}D\circ\lambda\right)dt\right]=0. \tag{6}$$

Applying Lemma 1 to Eq. (6) we finally obtain

$$E\left[\left(\frac{\partial L}{\partial D\circ x(t)}\lambda\right)_{t=b}-\left(\frac{\partial L}{\partial D\circ x(t)}\lambda\right)_{t=a}\right]=0. \tag{7}$$

Theorem 2 now follows from the arbitrariness of a and b.

IV. CONSERVATION LAWS IN THE STOCHASTIC LAGRANGIAN DYNAMICAL SYSTEM

We now turn to conservation laws obtainable by applying Theorem 2 to a stochastic Lagrangian dynamical system in the SSCV that corresponds to a solution of the Schrödinger equation. For this purpose we first touch upon the reformulation of Nelson's mechanics on our stochastic least-action principle.[8] Let us consider diffusion processes $x(t)$ belonging to K that are governed by the stochastic differential equation and the reversed equation[5-7]:

$$dx(t)=b(x(t),t)dt+(\hbar/m)^{1/2}dw(t), \tag{8}$$

$$dx(t)=b_*(x(t),t)dt+(\hbar/m)^{1/2}dw_*(t), \tag{9}$$

where b and b_* are certain vector-valued smooth functions, $\hbar$ is Planck's constant, and m is the mass of a particle. In (8) and (9), $w(t)$ is a standard $\mathbb{R}^l$-valued Wiener process, and $w_*(t)$ has the same properties as $w(t)$ except that the increments $w_*(t)-w_*(s)$ are independent of $x(\tau)$ for $\tau\geq t>s$. We assume that $x(t)$ has a probability density function $\rho(x,t)$. For this process we have $Dx(t)=b(x(t),t)$ and $D_*x(t)=b_*(x(t),t)$. According to Nelson,[5-7] these functions b and b_* are connected with the probability density ρ by the equations

$$\frac{\partial\rho}{\partial t}+\operatorname{div}(v\cdot\rho)=0,\quad u=\left\{\frac{\hbar}{2m}\right\}\operatorname{grad}\ln\rho, \tag{10}$$

where v, the "current velocity," and u, the "osmotic velocity," are vector-valued functions defined by $v=\frac{1}{2}(b+b_*)$ and $u=\frac{1}{2}(b-b_*)$, respectively.

Now we assume that the diffusion process $x(t)$ mentioned above is an extremal of the functional J with the Lagrangian

$$L(x,D\circ x,t)=(m/2)|D\circ x|^2-V(x,t)+(m/2)|u(x,t)|^2+(\hbar/2)\operatorname{div}u(x,t), \tag{11}$$

where V is a given potential function and u is the osmotic velocity given by the second equation of (10). Then Theorem 1 and the following lemma due to Nelson are put together to show that

$$\left(\frac{\partial}{\partial t}+v\cdot\nabla\right)v(x(t),t)=-\frac{1}{m}\operatorname{grad}V(x(t),t)+\left(u\cdot\nabla+\frac{\hbar}{2m}\Delta\right)u(x(t),t). \tag{12}$$

Lemma 2 (Nelson[5-7]): Let $f(x,t)$ be a smooth function on $\mathbb{R}^l\times\mathbb{R}^1$ and $x(t)$ the diffusion process governed by (8) and (9). Then

$$Df(x(t),t)=\left[\frac{\partial}{\partial t}+b\cdot\nabla+\left\{\frac{\hbar}{2m}\right\}\Delta\right]f(x(t),t), \tag{13}$$

$$D_*f(x(t),t)=\left[\frac{\partial}{\partial t}+b_*\cdot\nabla-\left\{\frac{\hbar}{2m}\right\}\Delta\right]f(x(t),t). \tag{14}$$

Equation (12) is just the same consequence as that of Newton's equation of motion in Nelson's mechanics. Therefore in a manner analogous to Nelson's mechanics we can determine v, u, and ρ (and hence b and b_*) from Eqs. (10) and (12), so that the diffusion process $x(t)$ is determined.[5-7] Thus Nelson's mechanics is now reformulated on the stochastic least-action principle. The diffusion process $x(t)$ together with the Lagrangian (11) is called the stochastic Lagrangian dynamical system (SLDS).

In the same manner as Nelson's, one can further show that the SLDS $x(t)$ with the Lagrangian (11) corresponds to a solution of a Schrödinger equation. Indeed, the wave function defined by $\Psi(x,t)=\{\rho(x,t)\}^{1/2}\exp\{iS(x,t)\}$, where $S(x,t)$ is such that $v(x,t)=(\hbar/m)\operatorname{grad}S(x,t)$, satisfies the Schrödinger equation with the potential function $V(x,t)$[5-7]; hence $\rho=|\Psi|^2$ gives the probability density of a particle in the position space.

We are now in a position to obtain conservation laws in the SLDS with the Lagrangian (11). These conservation laws will have their correspondents in quantum mechanics on account of the correspondence between our SLDS and the Schrödinger equation. For simplicity we consider a three-dimensional system and set $m=\hbar=1$.

(i) Conservation of momentum. Let us take up $X^i=x^i+\varepsilon$ $(i=1,2,3)$ for the transformation (4). We assume that the potential function $V(x,t)$ in the Lagrangian (11) is invariant under this transformation in $\mathbb{R}^3$. Then we can verify that Eq. (5) holds in this case. Indeed, by a computation along with Lemma 2 we obtain

$$\left(\frac{d\Delta J}{d\varepsilon}\right)_{\varepsilon=0}=E\left[\int_a^b\left\{\sum_{i=1}^3\frac{1}{2}(D-D_*)u^i(x(t),t)\right\}dt\right]. \tag{15}$$

We note that the process $u^i(x(t),t)$ belongs to K, since u^i is a smooth function. Then Lemma 1 shows that

$$E\left[\int_a^b(D-D_*)u^i(x(t),t)\,dt\right]=0,$$

so that Eq. (15) is equal to zero. Moreover, the assumption of Theorem 2 also holds in this case, since $(\partial L/$

$\partial D\circ x) = D\circ x = v$ and λ are smooth functions on x and t. Therefore by Theorem 2

$$E\left[\sum_{i=1}^{3} v^i(x(t),t)\right]$$

proves to be a conserved quantity.

(ii) Conservation of angular momentum. We now assume that the potential function $V(x,t)$ is invariant under the rotation about, say, the x^3 axis. In the same manner as in (i) we obtain

$$\left(\frac{d\Delta J}{d\varepsilon}\right)_{\varepsilon=0} = E\left[\int_a^b \left\{\frac{1}{2}(D-D_*)x^2\cdot u^1(x,t) - \frac{1}{2}(D-D_*)x^1\cdot u^2(x,t)\right\} dt\right]. \quad (16)$$

By the definition of u, the right-hand side of Eq. (16) becomes equal to zero, and hence Eq. (5) holds also in the case. Moreover, as in (i) the assumption of Theorem 2 also holds. Therefore Theorem 2 shows that

$$E[D\circ x^1(t)\cdot x^2(t) - x^1(t)\cdot D\circ x^2(t)]$$

is a conserved quantity.

(iii) Conservation of energy. Let $\tau\in C^1(I\to\mathbf{R}^1)$ be a strictly monotonic function with $\tau(a)=\bar{a}$ and $\tau(b)=\bar{b}$. The inverse function is denoted by $t(\tau)$ for $\bar{a}\leqslant\tau\leqslant\bar{b}$, and $x(t(\tau))$ is denoted by $\bar{x}(\tau)$. Then we can regard $(\bar{x}(\tau),t(\tau))$ as a stochastic process on $\mathbf{R}^3\times\mathbf{R}^1$. For this process we introduce a new Lagrangian $\bar{L}\in C^2(\mathbf{R}^{2(3+1)}\to\mathbf{R}^1)$ through the relation

$$\bar{L}\left((\bar{x},t),\left(D\circ\bar{x},\frac{dt}{d\tau}\right)\right) = L\left(\bar{x},D\circ\bar{x}\left(\frac{dt}{d\tau}\right)^{-1},t\right)\cdot\left(\frac{dt}{d\tau}\right),$$

where L is the Lagrangian (11). We denote by $\bar{J}$ the functional for the Lagrangian $\bar{L}$ with the process $(\bar{x}(\tau),t(\tau))$.

We now assume that the potential function $V(x,t)$ in the original Lagrangian (11) is independent of t. Consider the transformations in $\mathbf{R}^4$, $\bar{X}(\tau)=\bar{x}(\tau)$ and $T(\tau)=t(\tau)+\epsilon$. Then after a calculation, the left-hand side of Eq. (5) for the functional $\bar{J}$ goes over into

$$\left(\frac{d\Delta\bar{J}}{d\varepsilon}\right)_{\varepsilon=0} = E\left[\int_a^b \left\{u\frac{\partial u}{\partial t}+\frac{1}{2}\operatorname{div}\left(\frac{\partial u}{\partial t}\right)\right\}dt\right]. \quad (17)$$

By both equations of (10) and Lemma 2, the integrand of the right-hand side of Eq. (17) becomes

$$-\tfrac{1}{2}(D-D_*)(v\cdot u) - \tfrac{1}{4}(D-D_*)(\operatorname{div} v),$$

so that Eq. (17) is equal to zero because of Lemma 1. Furthermore, we also see that the assumption of Theorem 2 holds in this case. Therefore from Theorem 2 it follows that $E[\{\partial\bar{L}/\partial(dt/du)\}] = \text{const}$. On inserting the relation between L and $\bar{L}$ into this equation and setting $u=t$,

$$E[\tfrac{1}{2}|D\circ x|^2 + V(x) - \tfrac{1}{2}|u(x,t)|^2 - \tfrac{1}{2}\operatorname{div} u(x,t)]$$

is a conserved quantity.

ACKNOWLEDGMENTS

The author would like to thank Dr. Toshihiro Iwai and Dr. Kunio Yasue who read the manuscript and offered helpful advice.

[1]K. Yasue, J. Funct. Anal. **41**, 327 (1981).
[2]K. Yasue, J. Math. Phys. **23**, 1577 (1982).
[3]J.-C. Zambrini and K. Yasue, Phys. Rev. Lett. **52**, 2107 (1984).
[4]J.-C. Zambrini, Int. J. Theor. Phys. **24**, 277 (1985).
[5]E. Nelson, Phys. Rev. **150**, 1079 (1966).
[6]E. Nelson, *Dynamical Theories of Brownian Motion* (Princeton U.P., Princeton, NJ, 1967).
[7]E. Nelson, *Quantum Fluctuations* (Princeton U.P., Princeton, NJ, 1985).
[8]T. Misawa and K. Yasue, J. Math. Phys. **28**, 2569 (1987).
[9]I. M. Gelfand and S. V. Fomin, *Calculus of Variations* (Prentice-Hall, Englewood Cliffs, NJ, 1963).
[10]W. A. Zheng and P.A. Meyer, *Seminaire de Probabilités XVIII, Lecture Notes in Mathematics*, Vol. 223 (Springer, Berlin, 1984).

Canonoid transformations from a geometric perspective

José F. Cariñena and Manuel F. Rañada
Departamento de Fisica Teórica, Facultad de Ciencias, Universidad de Zaragoza, 50009 Zaragoza, Spain

(Received 8 December 1987; accepted for publication 1 June 1988)

The concept of canonoid transformation for a locally Hamiltonian vector field is introduced, and its relation with the existence of non-Noether constants of the motion is shown from a geometrical viewpoint. The equations determining generating functions for such canonoid transformations are obtained and applications to some particular problems given.

I. INTRODUCTION

Canonical transformations are very often used to reduce Hamilton equations describing the time evolution of a Hamiltonian system to a simpler set of Hamilton equations. This property of preserving the form of such equations, whatever the Hamiltonian is, characterizes the canonicity of the transformation. However, as remarked by Saletan and Cromer,[1] given a concrete Hamiltonian H, the transformations preserving the Hamilton form of the motion described by H, which are said to be canonoid with respect to H, may also play an important role in the solution of the problem at hand, and this fact has motivated recent papers on the properties of this kind of transformation.[2,3]

The approaches developed up to now, however, are local and coordinate dependent—a paper by Marmo[4] refers only slightly to the global problem. One of our aims, therefore, is to develop a global theory for canonoid transformations, using the tools of modern differential geometry, that will be valid for a more general case in which the manifold is not topologically trivial.

The paper is organized as follows: In Sec. II we give a geometric definition of canonoid transformation that reduces to the well-known one for the simplest case of a topologically trivial system, or at least when considering only local expressions for the system. The geometric version of the Poisson bracket theorem[1] gives a suggestion for such a definition, which coincides with the concept of quasicanonical transformation introduced by Marmo.[4] The theory of locally Hamiltonian dynamical systems admitting alternative formulations[5,6] will be used for studying such canonoid transformations, and we will find some global results generalizing previous contributions by Leubner and Marte[2] and Negri *et al.*,[3] to which they reduce when only trivial systems are concerned. It is also well established how to generate canonical transformations starting from "generating functions."[1,7] In contrast, as far as we know, there is no available method of generating a canonoid transformation for a concrete Hamiltonian H; therefore this will be the main goal of Sec. III, where we develop the theory of generating functions for canonoid transformations. In Sec. IV, some examples given previously are shown to arise in this way as associated to some generating functions. Finally, for the sake of completeness, the local expressions have been collected in an Appendix.

II. CANONOID TRANSFORMATIONS

The concept of canonoid transformations with respect to a Hamiltonian H was introduced by Saletan and Cromer[1] in order to name those transformations preserving the form of Hamilton's equations for a fixed Hamiltonian H. The generalization for the case of a nontrivial phase space is due to Marmo[4] under the name of quasicanonical transformations. The idea of this definition may be demonstrated by considering that given a symplectic manifold (M,ω), the integral curves of a Hamiltonian vector field Γ_H, defined by $i(\Gamma_H)\omega = dH$, are determined in Darboux coordinates for ω, for which $\omega = dq^i \wedge dp_i$, by the Hamilton equations

$$\frac{dq^i}{dt} = \frac{\partial H}{\partial p_i}, \quad \frac{dp_i}{dt} = \frac{\partial H}{\partial q^i}. \tag{2.1}$$

The same assertion would be true for a locally Hamiltonian vector field Γ, the role of H being played by a local Hamiltonian for Γ, but the explicit reference to a concrete type of coordinate seems, however, to be not quite satisfactory. The relation between locally Hamiltonian vector fields and mechanical systems whose evolution is described by Hamilton's equations may also be based on the fact that such systems are characterized by the relation of the so-called Poisson bracket theorem,[1]

$$\frac{d\{F,G\}}{dt} = \left\{\frac{dF}{dt}, G\right\} + \left\{F, \frac{dG}{dt}\right\},$$

that in geometrical terms is given by

$$\Gamma\{F,G\} = \{\Gamma F, G\} + \{F, \Gamma G\}. \tag{2.2}$$

The following theorem asserts that this relation is simply the equivalent of saying that Γ is a locally Hamiltonian vector field.

Theorem: Let (M,ω) be a symplectic manifold. A vector field $\Gamma \in \mathscr{X}(M)$ is locally Hamiltonian if and only if relation (2.2) holds.

Proof: We recall that the Poisson bracket of any two functions F and G is defined by $\{F,G\} = \omega(X_F, X_G)$, where X_F denotes the vector field such that $i(X_F)\omega = dF$. Then

$$\begin{aligned}\mathscr{L}_\Gamma\{F,G\} &= \mathscr{L}_\Gamma[\omega(X_F,X_G)] \\ &= (\mathscr{L}_\Gamma\omega)(X_F,X_G) \\ &\quad + \omega([\Gamma,X_F],X_G) + \omega(X_F,[\Gamma,X_G]),\end{aligned}$$

and therefore

 0022-2488/88/102181-06$02.50

$$\mathscr{L}_\Gamma\{F,G\} = (\mathscr{L}_\Gamma\omega)(X_F,X_G) + [\Gamma,X_G]F - [\Gamma,X_F]G,$$

from which we get

$$\Gamma\{F,G\} = -(\mathscr{L}_\Gamma\omega)(X_F,X_G) + \{\Gamma F,G\} + \{F,\Gamma G\}.$$

Consequently, if Γ is locally Hamiltonian, $\mathscr{L}_\Gamma\omega = 0$, then relation (2.2) holds, and, conversely, if this relation is true, then $(\mathscr{L}_\Gamma\omega)(X_F,X_G) = 0$ for any pair of functions. Since a local basis of $\mathscr{X}(M)$ can be built up from Hamiltonian vector fields, the Lie derivative $\mathscr{L}_\Gamma\omega$ must be zero, i.e., Γ is locally Hamiltonian.

A canonoid transformation Φ (w.r.t. H) preserves the form of Hamilton's equations, and therefore condition (2.2) would also be true for the transformed field $\Phi_*\Gamma_H$. This shows that an appropriate generalization for the concept of canonoid transformation is that of quasicanonical transformation according to Marmo's[4] terminology.

Definition: Given a locally Hamiltonian vector field $\Gamma\in\mathscr{X}(M)$ in a symplectic manifold (M,ω), a transformation $\Phi\in\mathrm{Diff}(M)$ is said to be canonoid with respect to Γ if the transformed field $\Phi_*\Gamma$ is also locally Hamiltonian.

The two-form $B = \mathscr{L}_{\Phi_*\Gamma}\omega$ is exact, $B\in B^2(M)$, since $B = dP$ with $P = i(\Phi_*\Gamma)\omega$. Thus the property $B = 0$ implies, locally, the existence of a function K such that $i(\Phi_*\Gamma)\omega = dK$.

The point is that $\Phi_*\Gamma$ is locally Hamiltonian with respect to the symplectic form ω if and only if Γ is locally Hamiltonian with respect to the symplectic structure $\Phi^*\omega$, because of the relation $\Phi^* d[i(\Phi_*\Gamma)\omega] = d[i(\Gamma)\Phi^*\omega]$. This means that when Φ is a canonoid transformation for Γ, then the vector field Γ admits a new and different locally Hamiltonian structure.

The dynamical systems admitting alternative Hamiltonian formulations have received much attention in the last few years. In particular, for Lagrangian systems, it was proved by Hojman and Harleston[8] that given two s-equivalent, but not gauge-equivalent, Lagrangians L_i, $i = 1,2$, it is possible to find some constants of motion that are the traces of the powers of the product matrix $W_2^{-1}W_1$, where W_i, $i = 1,2$, denotes the Hessian matrix

$$[W_i]_{\alpha\beta} = \frac{\partial^2 L_i}{\partial v^\alpha\,\partial v^\beta}.$$

Moreover, Henneaux[9] proved that this corresponds to the vanishing of the Lie derivative of a (1,1) tensor field whose diagonal blocks are $W_2^{-1}W_1$. The more general case of Hamiltonian dynamical systems has been studied in Refs. 5, 6, 10, and 11. Now we will show that the properties of these bi-Hamiltonian dynamical systems can be used for the study of canonoid transformations.

A. Particular case: Bidimensional symplectic manifold

If the dynamical system is two dimensional, dim $M = 2$, then any two arbitrary nondegenerate two-forms are proportional. Therefore there will exist a function $f\in C^\infty(M)$ such that $\Phi^*\omega = f\omega$.

Now because of the relation

$$\mathscr{L}_\Gamma(f\omega) = (\Gamma f)\omega + f(\mathscr{L}_\Gamma\omega),$$

we obtain that, if Γ is a locally Hamiltonian field, then the transformation Φ is canonoid w.r.t. Γ if and only if f is a constant of motion.

Next we prove that this function f relates the Poisson bracket of the pullbacks $\Phi^*(F)$, $\Phi^*(G)$ of any pair of functions $F,G\in C^\infty(M)$ to the primitive Poisson bracket $\{F,G\}$. In fact, we have

$$i(X_{F\circ\Phi})\omega = d(F\circ\Phi) = \Phi^*(dF) = \Phi^*[i(X_F)\omega] = i[\Phi_*^{-1}(X_F)]\Phi^*\omega, \tag{2.3}$$

and consequently if $\Phi^*\omega = f\omega$, then we obtain $X_{F\circ\Phi} = f\Phi_*^{-1}(X_F)$. Since the definition of the Poisson bracket is

$$\{F\circ\Phi,G\circ\Phi\} = X_{G\circ\Phi}(F\circ\Phi),$$

and taking into account the value of $X_{G\circ\Phi}$ given above, we find

$$\{F\circ\Phi,G\circ\Phi\} = f\Phi_*^{-1}(X_G)(F\circ\Phi) = fX_G(F)\circ\Phi = f\{F,G\}\circ\Phi. \tag{2.4}$$

This property may be used to get a new proof of the theorem given by Leubner and Marte[2] for the case of a "one-dimensional system," according to which a mapping $(q,p)\to(Q,P)$ carries a given canonical description (p,q,H) of a one-dimensional system into another canonical description (P,Q,K) if and only if the Poisson bracket $\{P,Q\}$ is a constant of the motion. Actually, in the particular case of taking as functions F and G in (2.4) the coordinates q and p, seeing that they are canonical conjugate variables, $\{q,p\} = 1$, we find that f is given by $f = \{Q,P\}$. Note, however, that we have established the theorem in a larger context without the assumption of M being the phase space of a system with a topologically trivial configuration space.

B. The general case

The more general case of a higher-dimensional system has recently been studied by Negri *et al.*[3] in a coordinate-dependent way, and they have proved that a necessary, but not sufficient, condition for a transformation $\eta_\alpha = \eta_\alpha(\xi)$ to be canonoid w.r.t. a Hamiltonian $H(q,p) = H(\xi)$ is that $\gamma_{\alpha\beta}A_{\alpha\beta}$ must be a constant of motion. Here $A_{\alpha\beta}$ denotes the Poisson bracket $A_{\alpha\beta} = \{\eta_\alpha,\eta_\beta\}$ and $\gamma_{\alpha\beta}$ the elements of the canonical symplectic matrix.

The existence, when Φ is canonoid, of alternative Hamiltonian formulations for the vector field Γ indicates the existence of non-Noether constants of motion. In fact, we can consider[5] the pencil of admissible closed two-forms for Γ defined by $\Phi^*\omega$ and ω, i.e., the family $\Phi^*\omega - \lambda\omega$. If $2n$ is the dimension of M, then the two $2n$-forms $(\Phi^*\omega - \lambda\omega)^{\wedge n}$ and $\omega^{\wedge n}$ are volume elements and thus proportional. The function $f_\lambda\in C^\infty(M)$ defined by

$$(\Phi^*\omega - \lambda\omega)^{\wedge n} = f_\lambda\omega^{\wedge n} \tag{2.5}$$

satisfies $\mathscr{L}_\Gamma f_\lambda = 0$. It is a polynomial of degree n in the indeterminate λ, $f_\lambda = \alpha_0 + \alpha_1\lambda + \cdots + \alpha_n\lambda^n$, with coefficients α_k, defined by

$$(-1)^k\begin{bmatrix} n \\ k \end{bmatrix}(\Phi^*\omega)^{\wedge(n-k)}\wedge\omega^{\wedge k} = \alpha_k\omega^{\wedge n}, \quad k = 0,\ldots,n, \tag{2.6}$$

that are constants of the motion, $\mathscr{L}_\Gamma \alpha_k = 0$.

We have obtained n not necessarily independent, associated constants of the motion, α_k, $k = 0,...,n-1$, and we will prove that one of them, α_{n-1}, reduces in the simplest case to the constant of motion found by Negri *et al.*[3] In fact, if (q^i,p_i), $i = 1,...,n$, is a set of Darboux coordinates for ω, then the two-form $\Phi^*\omega$ can be written using these coordinates as

$$\Phi^*\omega = \tfrac{1}{2}a_{ij}\, dq^i \wedge dq^j + b_{ij}\, dq^i \wedge dp_j + \tfrac{1}{2}c_{ij} dp_i \wedge dp_j,$$

where $A = [a_{ij}]$ and $C = [c_{ij}]$ are skew-symmetric matrices. Therefore Eq. (2.6) for $k = n-1$ reads

$$(\Phi^*\omega) \wedge \omega^{\wedge(n-1)} = n!\, \mathrm{Tr}\, B\omega^{\wedge n}, \quad B = [b_{ij}],$$

thus obtaining that $\mathscr{L}_\Gamma(\mathrm{Tr}\, B) = 0$. The coefficients b_{ij} can be shown to be identified with the Poisson brackets $\{Q^i,P_j\}$ with $Q^i = q^i \circ \Phi$ and $P_j = p_j \circ \Phi$. In fact, let $\hat{\omega}$ denote the map $\hat{\omega}$: $\mathscr{X}(M) \to \wedge^1(M)$ given by contraction, $\hat{\omega}(X) = i(X)\omega$, and let Λ be the (1,1) tensor field defined[10] by $\Lambda = \hat{\omega}^{-1} \circ (\Phi^*\omega)^\wedge$ — or, in an equivalent way,

$$(\Phi^*\omega)(X,Y) = \omega(\Lambda(X),Y), \quad \forall X,Y \in \mathscr{X}(M). \tag{2.7}$$

Such a tensor is such that $\mathscr{L}_\Gamma \Lambda = 0$ and, therefore, $\mathscr{L}_\Gamma \Lambda^k = 0$, so that the coordinate expressions of these equations are Lax equations (see, for example, Refs. 10 and 11) giving rise to a set of constants of the motion that are the traces of the integer powers of the matrix representing Λ in a local basis of fields. Using this tensor, the Poisson bracket of the Φ pullback of two functions F and G is

$$\begin{aligned}\{F\circ\Phi, G\circ\Phi\} &= i(X_{G\circ\Phi})d(F\circ\Phi) \\ &= i[\Lambda(\Phi_*^{-1}(X_G))]d(F\circ\Phi) \\ &= [\Lambda(X_G)\circ F]\circ F^{-1},\end{aligned} \tag{2.8}$$

where we have used $X_{G\circ\Phi} = \Lambda(\Phi_*^{-1}(X_G))$, which follows from (2.3), and the definition (2.7) of Λ. In particular,

$$\{Q^i,P_j\} = \Lambda\left(\frac{\partial}{\partial q^j}\right) q^i,$$

and since

$$\Lambda\left(\frac{\partial}{\partial q^j}\right) = b_{kj}\left(\frac{\partial}{\partial q^k}\right) - a_{kj}\left(\frac{\partial}{\partial p_k}\right),$$

we obtain $\{Q^i,P_j\} = b_{ij}$, and therefore $\mathrm{Tr}\, B$ reduces to

$$\mathrm{Tr}\, B = \sum_{i=1}^{n} \{Q^i,P_i\}. \tag{2.9}$$

It was also pointed out in Ref. 3 that the condition of (2.9) being a constant of the motion is a necessary, but not sufficient, condition for Φ to be canonoid. Actually, it is a rather obvious consequence of the theory developed here; not only the trace of Λ but also the traces of the integer powers of Λ must be constants of the motion. The point is whether this last condition is also sufficient for Φ to be canonoid. The answer is "no" because, assuming $\mathscr{L}_\Gamma(\mathrm{Tr}\, \Lambda^k) = 0$, $k = 1,...,n$, we will then obtain $\mathscr{L}_\Gamma(\Phi^*\omega - \lambda\omega)^{\wedge n} = 0$ for any value of λ, and it does not follow from this that $\mathscr{L}_\Gamma(\Phi^*\omega) = 0$, the condition for Φ to be canonoid.

The expression $\mathscr{L}_\Gamma(\Phi^*\omega - \lambda\omega)^{\wedge n} = 0$ can also be written as a system of n equations,

$$\mathscr{L}_\Gamma[(\Phi^*\omega)^{\wedge k} \wedge \omega^{\wedge(n-k)}] = 0, \quad k = 1,...,n, \tag{2.10a}$$

or, in an equivalent way,

$$[\mathscr{L}_\Gamma(\Phi^*\omega)] \wedge (\Phi^*\omega)^{\wedge(k-1)} \wedge \omega^{\wedge(n-k)} = 0, \quad k = 1,...,n. \tag{2.10b}$$

Actually, it is possible to find a closed form ω_1 different from zero such that

$$\omega_1 \wedge (\Phi^*\omega)^{\wedge(k-1)} \wedge \omega^{\wedge(n-k)} = 0, \quad k = 1,...,n.$$

If the (1,1) tensor field N is defined by $\omega_1(X,Y) = \omega(N(X),Y)$ for $X,Y \in \mathscr{X}(M)$, then

$$\begin{aligned}&[\omega_1 \wedge (\Phi^*\omega)^{\wedge(k-1)} \wedge \omega^{\wedge(n-k)}](X_1,Y_1,...,X_n,Y_n) \\ &\quad = \omega^{\wedge n}(N(X_1),Y_1,\Lambda(X_2),Y_2,..., \\ &\qquad \Lambda(X_k),Y_k,X_{k+1},...,Y_{2k}).\end{aligned}$$

This proves that it is sufficient to take a (1,1) tensor field N such that

$$N\left(\frac{\partial}{\partial q^i}\right) = c_{ij}\left(\frac{\partial}{\partial p_j}\right)$$

for the value of the left-hand side of $(\partial/\partial q^1, \partial/\partial p_1,..., \partial/\partial q^n, \partial/\partial p_n)$ to be zero. Thus even if $\omega_1 \neq 0$, the volume form $\omega_1 \wedge (\Phi^*\omega)^{\wedge(k-1)} \wedge \omega^{\wedge(n-k)}$ can be null.

This fact proves that the stronger condition $\mathscr{L}_\Gamma(\Phi^*\omega) = 0$ cannot be replaced by "the traces of the different powers of Λ are constants of the motion."

III. GENERATING FUNCTIONS

The canonicity of a transformation is related to the existence of an associated generating function. This function is defined on the graph of the map, and when expressed in local coordinates it clearly displays the relationship between the "old" and the "new" positions and momenta. Following this idea we will reformulate the definition of canonoid transformation.

Proposition: Let (M_1,ω_1) and (M_2,ω_2) be symplectic manifolds, π_i: $M_1 \times M_2 \to M_i$ the projections onto M_i, $i = 1,2$, and $\Omega \in \wedge^2(M_1 \times M_2)$ the symplectic two-form defined by $\Omega = \pi_1^*\omega_1 - \pi_2^*\omega_2$. Then a diffeomorphism Φ: $M_1 \to M_2$ is canonoid with respect to the locally ω_1-Hamiltonian field $\Gamma \in \mathscr{X}(M_1)$ if and only if $i_\Phi^*(\mathscr{L}_V\Omega) = 0$, where $V \in \mathscr{X}(M_1 \times M_2)$ denotes the field $V = \Gamma \times \Phi_*\Gamma$, G_Φ is the graph of Φ, and i_Φ: $G_\Phi \to M_1 \times M_2$ is the inclusion map.

Proof: First, we notice that

$$\begin{aligned}\mathscr{L}_V\Omega &= \mathscr{L}_V(\pi_1^*\omega_1 - \pi_2^*\omega_2) \\ &= \pi_1^*(\mathscr{L}_{\pi_{1*}V}\omega_1) - \pi_2^*(\mathscr{L}_{\pi_{2*}V}\omega_2).\end{aligned}$$

On the other hand, Φ induces a diffeomorphism of M_1 onto G_Φ, so we can write

$$T_{(m,\Phi(m))}G_\Phi = \{(v,\Phi_* v) \mid v \in T_m M_1\},$$

and therefore

$$\begin{aligned}&[i_\Phi^*(\mathscr{L}_V\Omega)]\{(v_1,\Phi_* v_1),(v_2,\Phi_* v_2)\} \\ &\quad = (\mathscr{L}_{\pi_{1*}V}\omega_1)(v_1,v_2) - (\mathscr{L}_{\pi_{2*}V}\omega_2)(\Phi_* v_1,\Phi_* v_2) \\ &\quad = (\mathscr{L}_\Gamma\omega_1)(v_1,v_2) - \Phi^*(\mathscr{L}_{\Phi_*\Gamma}\omega_2)(v_1,v_2) \\ &\quad = [\mathscr{L}_\Gamma(\omega_1 - \Phi^*\omega_2)](v_1,v_2),\end{aligned}$$

and the proposition is proved.

If we write $\Omega = -d\Theta$ we obtain $d[i_\Phi^*(\mathscr{L}_\nu\Theta)] = 0$; that is, $i_\Phi^*(\mathscr{L}_\nu\Theta)$ being closed is equivalent to Φ being canonoid with respect to Γ. Locally, by Poincaré's lemma, there exists a function $S \in C^\infty(G_\Phi)$ such that $i_\Phi^*(\mathscr{L}_\nu\Theta) = dS$. We call the function S a generating function for the canonoid diffeomorphism Φ. If we take $\Theta = \pi_1^*\theta_1 - \pi_2^*\theta_2$, we obtain

$$i_\Phi^*[\pi_1^*(\mathscr{L}_\Gamma\theta_1)] - i_\Phi^*[\pi_2^*(\mathscr{L}_{\Phi_*\Gamma}\theta_2)] = dS. \tag{3.1}$$

Recall that for a canonical transformation, the associated generating function measures the change of the action induced by it. Now note that we have found that, for canonoid transformations, S measures the Lie derivative of this change with respect to the field Γ.

Let Φ: $M_1 \to M_2$ be canonoid, and denote the Darboux coordinates by (q^i,p_i) and (Q^i,P_i) as usual. Then, regarding S as a function of (q^i,p_i), the relation (3.1) reads

$$\frac{\partial^2 H}{\partial q^j\,\partial p_k}p_k - \frac{\partial\{Q^k,H\}}{\partial q^j}P_k - \frac{\partial H}{\partial q^j} - \{P_k,H\}\frac{\partial Q^k}{\partial q^j} = \frac{\partial S}{\partial q^j}, \tag{3.2a}$$

$$\frac{\partial^2 H}{\partial p_j\,\partial p_k}p_k - \frac{\partial\{Q^k,H\}}{\partial p_j}P_k - \{P_k,H\}\frac{\partial Q^k}{\partial p_j} = \frac{\partial S}{\partial p_j}, \tag{3.2b}$$

These equations are obviously more complicated than those found for the canonical transformations (i.e., $p_i = \partial S/\partial q^i$, $P_i = -\partial S/\partial Q^i$), but this fact is due to the presence in (3.1) of the Lie derivative consequence of the Γ dependence. If we restrict ourselves to the so-called fouling transformations[12,13]—that is to say, fiber-preserving diffeomophisms inducing the identity in the base space, $Q^i = q^i$—the expressions (3.2) reduce to

$$\frac{\partial^2 H}{\partial q^j\,\partial p_k}(p_k - P_k) - \frac{\partial H}{\partial q^j} - \{P_j,H\} = \frac{\partial S}{\partial q^j}, \tag{3.3a}$$

$$\frac{\partial^2 H}{\partial p^j\,\partial p_k}(p_k - P_k) = \frac{\partial S}{\partial p_j}. \tag{3.3b}$$

IV. EXAMPLES

(1) Saletan and Cromer proposed[1] as a pattern for canonoid transformations the following example in one degree of freedom. The equations giving the transformation are

$$Q = q, \tag{4.1a}$$

$$P = p^{1/2} - q^2, \tag{4.1b}$$

and it is proved to be canonoid for the Hamiltonian $H = p^2/2$ of a unit-mass free particle.

In fact, using (3.3) this map is obtained as the particular one associated to the function

$$S = q^2p + p^2/2 - \tfrac{2}{3}p^{3/2}, \tag{4.2}$$

and the new Hamiltonian K such that $dK = -i(\Gamma)d[\Phi^*(\theta_0)]$ is found to be given by $dK = -\{P,H\}dQ + \{Q,H\}dP$, which when integrated gives $K = \frac{1}{3}(P+Q^2)^3$.

Moreover, using (3.3) we can now obtain not only a particular one, but the set of *all* the fouling transformations for this Hamiltonian. Indeed Eqs. (3.3) particularized for $H = p^2/2$ reduce to

$$\frac{\partial S}{\partial q} = -\frac{\partial P}{\partial q}p, \tag{4.3a}$$

$$\frac{\partial S}{\partial p} = p - P, \tag{4.3b}$$

a system that can be integrated and solved for P as a function of (q,p) if and only if S verifies the following compatibility condition:

$$p\frac{\partial^2 S}{\partial q\,\partial p} - \frac{\partial S}{\partial q} = 0. \tag{4.4}$$

The general solution of this equation is given by

$$S = f(q)p + g(p), \tag{4.5}$$

with $f(q)$ and $g(p)$ being arbitrary differentiable functions. Thus any function S in the form (4.5) generates a canonoid transformation for $H = p^2/2$, given by

$$Q = q, \tag{4.6a}$$

$$P = p - f(q) - g'(p). \tag{4.6b}$$

As the system is one dimensional, the new Hamiltonian K is related to H by $dK = \{Q,P\}dH$, which for (4.6) becomes $dK = \{1 - g''(p)\}p\,dp$, and then we obtain for the function K, when expressed in terms of the set of old coordinates (q,p), the following expression:

$$K(q,p) = \tfrac{1}{2}p^2 + g(p) - g'(p)p. \tag{4.7}$$

Note that K can never be reduced to $K = \text{const}$, since $1 - g''(p) \neq 0$ is precisely the condition imposed by the implicit function theorem for solving (4.6) for p. Furthermore, note that when the function $g(p)$ is a homogeneous function of degree 1, then $K = H$ and (4.6) reduces to a canonical transformation.

(2) As a second example we will consider the two-dimensional isotropic harmonic oscillator

$$H = \tfrac{1}{2}\{(q_1)^2 + (q_2)^2 + (p_1)^2 + (p_2)^2\}. \tag{4.8}$$

Given any differentiable function $S = S(q_1,q_2,p_1,p_2)$, every solution (P_1,P_2) of the system

$$\frac{\partial S}{\partial q_1} = \{H,P_1\} - q_1, \tag{4.9a}$$

$$\frac{\partial S}{\partial q_2} = \{H,P_2\} - q_2, \tag{4.9b}$$

$$\frac{\partial S}{\partial p_1} = p_1 - P_1, \tag{4.9c}$$

$$\frac{\partial S}{\partial p_2} = p_2 - P_2, \tag{4.9d}$$

represents a fouling canonoid transformation. The only constraint for S is that it must satisfy the compatibility conditions

$$\frac{\partial S}{\partial q_1} = -\left\{H,\frac{\partial S}{\partial p_1}\right\}, \quad \frac{\partial S}{\partial q_2} = -\left\{H,\frac{\partial S}{\partial p_2}\right\}.$$

Recently,[3] two different fouling transformations have been found for (4.8). They are

$$P_1 = 2p_1 - p_2 + q_1q_2, \tag{4.10a}$$

$$P_2 = p_2 - p_1 + \tfrac{1}{2}q_1^2, \tag{4.10b}$$

and

$$P_1 = \tfrac{1}{2}q_1^2 p_1 + \tfrac{1}{6}p_1^3, \tag{4.11a}$$

$$P_2 = \tfrac{1}{2}q_2^2 p_2 + \tfrac{1}{6}p_2^3. \tag{4.11b}$$

It is not a hard task to show that they are just two particular solutions of (4.9). In fact their associated generating functions are easily found to be

$$S = \tfrac{1}{2}q_1^2 - q_1 q_2 - \tfrac{1}{2}p_1^2 + p_1 p_2 - q_1 q_2 p_1 - \tfrac{1}{2}q_1^2 p_2,$$

and

$$S = \tfrac{1}{2}(p_1^2 + p_2^2 - q_1^2 - q_2^2) - \tfrac{1}{4}(q_1^2 p_1^2 + q_2^2 p_2^2) + \tfrac{1}{8}(q_1^4 + q_2^4) - \tfrac{1}{24}(p_1^4 + p_2^4),$$

respectively.

ACKNOWLEDGMENTS

We are indebted to the authors of Ref. 3 for sending us a copy of their paper before publication, and to the referee for drawing our attention to some of the references.

Partial financial support by CICYT (Comisión de Investigación Científica y Técnica) is also acknowledged.

APPENDIX: COORDINATE EXPRESSIONS

In this appendix we give the expressions in local coordinates for different formulas used throughout the paper.

We denote by ξ_α, $\alpha = 1,...,2n$, a set of Darboux variables for ω,

$$\omega = \Lambda_{\alpha\beta} d\xi_\alpha \wedge d\xi_\beta,$$

where

$$[\Lambda_{\alpha\beta}] = \begin{bmatrix} 0 & I_n \\ 0 & 0 \end{bmatrix}.$$

We will also use $\gamma_{\alpha\beta} = \Lambda_{\alpha\beta} - \Lambda_{\alpha\beta}$ such that

$$\gamma_{\alpha\beta}\gamma_{\alpha\rho} = \delta_{\beta\rho}.$$

If $\Gamma \in \mathscr{X}(M)$ is a locally ω-Hamiltonian field, then there exists a locally defined function $H = H(\xi_\alpha)$ such that Γ takes the form

$$\Gamma = f_\alpha \left(\frac{\partial}{\partial \xi_\alpha}\right), \quad f_\alpha = \gamma_{\alpha\lambda}\left(\frac{\partial H}{\partial \xi_\lambda}\right).$$

Let F_α be the local components of the field $\Phi_* \Gamma$,

$$\Phi_* \Gamma = F_\alpha \left(\frac{\partial}{\partial \xi_\alpha}\right).$$

Then the two-form $B \in \Lambda^2(M)$ defined by $B = \mathscr{L}_{\Phi_* \Gamma}\omega$ is given by

$$B = B_{\mu\nu} d\xi_\mu \wedge d\xi_\nu,$$

where

$$B_{\mu\nu} = \gamma_{\alpha\nu}\left(\frac{\partial F_\alpha}{\partial \xi_\mu}\right) - \gamma_{\alpha\mu}\left(\frac{\partial F_\alpha}{\partial \xi_\nu}\right).$$

In this way, we obtain that the necessary and sufficient condition for the map Φ being canonoid with respect to Γ is that

$$\gamma_{\alpha\nu}\left(\frac{\partial F_\alpha}{\partial \xi_\mu}\right) - \gamma_{\alpha\mu}\left(\frac{\partial F_\alpha}{\partial \xi_\nu}\right) = 0, \quad \mu,\nu = 1,...,2n.$$

The meaning of this system of equations is the existence of a function $K = K(\xi_\alpha)$ such that

$$F_\alpha = \gamma_{\alpha\lambda}\left(\frac{\partial K}{\partial \xi_\lambda}\right).$$

Thus $\{\xi_\alpha(t)\}$ is an integral curve of $\Phi_* \Gamma$ if and only if Hamilton's equations with respect to K hold:

$$\frac{d\xi_\alpha}{dt} = \gamma_{\alpha\lambda}\left(\frac{\partial K}{\partial \xi_\lambda}\right).$$

In an equivalent way, we can use the two-form $T = \mathscr{L}_\Gamma(\Phi^*\omega)$, which will be given by

$$T = \Phi^*(B_{\mu\nu}) \frac{\partial \eta_\mu}{\partial \xi_\alpha} \frac{\partial \eta_\nu}{\partial \xi_\beta} d\xi_\alpha \wedge d\xi_\beta,$$

where η_μ denotes $\eta_\mu = \Phi^*(\xi_\mu)$.

The one-form $P^* = P_\mu^* d\xi_\mu$ such that $dP^* = T$ is given by

$$P_\mu^* = \gamma_{\alpha\beta} L_{\alpha\mu}\left(\frac{\partial H}{\partial \xi_\beta}\right),$$

where the coefficients $L_{\alpha\beta}$ are given by

$$L_{\alpha\beta} = (\Phi^*\omega)\left(\frac{\partial}{\partial \xi_\alpha}, \frac{\partial}{\partial \xi_\beta}\right) = \gamma_{\mu\nu}\left(\frac{\partial \eta_\mu}{\partial \xi_\alpha}\right)\left(\frac{\partial \eta_\nu}{\partial \xi_\beta}\right),$$

and correspond to the Lagrange brackets of the variables ξ with respect to the η's.

In the particular case of M being an exact symplectic manifold, ω is given by $\omega = -d\theta$, and then we can reflect the canonoid character of Φ using the one-form $g \in \Lambda^1(M)$ defined by $g = -\Phi^*\theta$. In this case the condition $T = 0$ means $d(\mathscr{L}_\Gamma g) = 0$, and if g is given by $g = g_\lambda d\xi_\lambda$, we obtain

$$\mathscr{L}_\Gamma g = (\dot{g}_\lambda + R_\lambda) d\xi_\lambda,$$

with

$$\dot{g}_\lambda = \frac{\partial g_\lambda}{\partial \xi_\mu}\gamma_{\mu\alpha}\left(\frac{\partial H}{\partial \xi_\alpha}\right)$$

and

$$R_\lambda = g_\mu \gamma_{\mu\alpha}\left(\frac{\partial H^2}{\partial \xi_\alpha \partial \xi_\lambda}\right).$$

In this way we obtain finally that Φ is canonoid if the $2n$ functions g_λ, $\lambda = 1,...,2n$, satisfy the equations

$$\frac{\partial \dot{g}_\lambda}{\partial \xi_\mu} - \frac{\partial \dot{g}_\mu}{\partial \xi_\lambda} + \frac{\partial R_\lambda}{\partial \xi_\mu} - \frac{\partial R_\mu}{\partial \xi_\lambda} = 0.$$

[1] E. J. Saletan and A. H. Cromer, *Theoretical Mechanics* (Wiley, New York, 1971).

[2] C. Leubner and M. A. M. Marte, Phys. Lett. A **101**, 179 (1984).

[3] L. J. Negri, L. C. Oliveira, and J. M. Teixeira, J. Math. Phys. **28**, 2369 (1987).

[4] G. Marmo, "Equivalent Lagrangians and quasicanonical transformations," in *Lecture Notes in Physics*, Vol. 50 (Springer, Berlin, 1976), pp. 568–572.

[5] J. F. Cariñena and L. A. Ibort, J. Phys. A: Math. Gen. **16**, 1 (1983).

[6]F. Giandolfi, G. Marmo, and C. Rubano, Nuovo Cimento B **66**, 34 (1981).
[7]R. Abraham and J. E. Marsden, *Foundations of Mechanics* (Benjamin, New York, 1978).
[8]S. Hojman and H. Harleston, J. Math. Phys. **22**, 1414 (1981).
[9]M. Henneaux, Hadronic J. **4**, 2137 (1981).
[10]M. Crampin, Phys. Lett. A **95**, 201 (1983).
[11]J. F. Cariñena and L. A. Ibort, Phys. Lett. A **107**, 356 (1985).
[12]D. G. Currie and E. G. Saletan, J. Math. Phys. **7**, 967 (1966).
[13]F. J. Kennedy and E. H. Kerner, Am. J. Phys. **33**, 463 (1965).

Symmetries for the super modified KdV equation

P. H. M. Kersten
Faculty of Applied Mathematics, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands

(Received 25 February 1988; accepted for publication 8 June 1988)

Higher-order or generalized symmetries for the super modified KdV equation are constructed. Moreover, by the introduction of graded potentials nonlocal symmetries are obtained, one of them leading to the recursion operator for symmetries in a straightforward way.

I. INTRODUCTION

In the study of complete integrability of classical evolution equations such as KdV, mKdV, Boussinesq, massive Thirring, and other well-known equations there was a great emphasis on Wahlquist–Estabrook prolongation and higher-order or generalized symmetry calculations.[1–3] We constructed computer-algebra programs to handle the enormous computations arising from these concepts. As a step towards supersymmetric equations we recently constructed a graded differential geometry package in REDUCE.[4] The notions of graded differential geometry are taken from Kostant,[5] while the graded jet bundle formulation is due to Hernández Ruiperez and Muñoz Masqué.[6] The present paper deals with the construction of generalized symmetries of the super modified KdV equation by the developed software. The super modified KdV equation[7] (mKdV) is given by the following system of graded partial differential equations:

$$v_t = 6v^2v_x - v_{xxx} + \tfrac{3}{4}\psi_x\psi_{xx} + \tfrac{3}{4}\psi\psi_{xxx} + \tfrac{3}{2}v_x\psi\psi_x + \tfrac{3}{2}v\psi\psi_{xx}\,,$$
$$\psi_t = (6v^2 - 6v_x)\psi_x + (6vv_x - 3v_{xx})\psi - 4\psi_{xxx}\,, \tag{1.1}$$

where subscripts denote partial derivatives; t is the time and x is the space variable; and v,x,t are even (commuting), while ψ is odd (anticommuting).

In Sec. II higher-order symmetries are constructed satisfying a similar condition as in the classical case,[8] i.e.,

$$L_V(D^\infty I) \subset D^\infty I\,,$$

where $D^\infty I$ is the infinite prolongation of the graded ideal I. Following Ref. 9, nonlocal variables are introduced in the graded case, and nonlocal x,t-dependent higher-order symmetries of the super mKdV equation are obtained in Sec. III. Finally, in Sec. IV we derive the recursion operator[10,11] for higher-order symmetries of the super mKdV equation.

II. HIGHER-ORDER SYMMETRIES OF THE SUPER mKdV EQUATION

In a previous paper[12] concerning the super KdV equation we constructed ordinary symmetries before concentrating on higher-order symmetries. Now we shall investigate higher-order symmetries only, because ordinary symmetries can be obtained from them. Classical higher-order symmetries are defined on the infinite jet bundle $J^\infty(x,t,v,\psi)$ (see Ref. 8) and satisfy the symmetry condition

$$L_V D^\infty I \subset D^\infty I\,, \tag{2.1}$$

where $D^\infty I$ is the infinite prolongation of the exterior differential system I describing the partial differential equations by means of the action of the total partial derivative vector fields D_x, D_t defined by

$$D_x = \partial_x + v_x\,\partial_v + \psi_x\,\partial_\psi + v_{xx}\,\partial_{v_x} + \cdots\,,$$
$$D_t = \partial_t + v_t\,\partial_v + \psi_t\,\partial_\psi + v_{xt}\,\partial_{v_x} + \cdots\,. \tag{2.2}$$

Due to the fact that Eqs. (2.2) satisfy (2.1) in an obvious way, the search for higher-order or generalized symmetries can be restricted to vertical vector fields; i.e., the components of ∂_x, ∂_t are taken to be zero.

The vertical vector fields are proved[8] to have the representation

$$V = f\,\partial_v + g\,\partial_\psi + (D_x f)\partial_{v_x} + (D_x g)\partial_{\psi_x} + \cdots\,, \tag{2.3}$$

so we are only interested in the defining functions f,g of the vector field. The functions f,g are assumed to depend on a finite number of independent variables of the infinite jet bundle.

In the graded case at hand we proceed in a similar way, keeping in mind the left module structure of the vector fields. We restrict our search for higher-order symmetries to *even* vector fields; moreover, our search is for vector fields V whose defining functions f,g (2.3) depend on $x,t,v,\psi,\ldots,v_{xxxxx},\psi_{xxxxx}$, the other components being obtained by prolongation (2.3). The vector field V has to satisfy the symmetry condition (2.1), which is equivalent to

$$L_V(v_t - 6v^2v_x + v_{xxx} - \tfrac{3}{4}\psi_x\psi_{xx} - \tfrac{3}{4}\psi\psi_{xxx}$$
$$- \tfrac{3}{2}v_x\psi\psi_x - \tfrac{3}{2}v\psi\psi_{xx}) \equiv 0\,, \tag{2.4}$$
$$L_V(\psi_t - 6v^2\psi_x + 6v_x\psi_x - 6vv_x\psi + 3v_{xx}\psi + 4\psi_{xxx}) \equiv 0\,,$$

where L_V stands for the Lie derivative with respect to the vector field V, and "$\equiv 0$" should be read as equal to zero on the submanifold in the infinite jet bundle $J^\infty(x,t;v,\psi)$ defined by (1.1) and its differential consequences. Condition (2.4) leads to an overdetermined system of partial differential equations for the functions f,g including the exterior algebra defined on $\psi,\psi_x,\psi_{xx},\psi_{xxx},\psi_{xxxx},\psi_{xxxxx}$.

From now on we shall write

$$v_i = \underbrace{v_{x\cdots x}}_{i \text{ times}}\,, \quad \psi_j = \underbrace{\psi_{x\cdots x}}_{j \text{ times}}\,. \tag{2.5}$$

Using the developed integration package we obtained the following result: There are four *even* vector fields satisfying (2.4) under the above-mentioned assumptions, i.e.,

 0022-2488/88/102187-03$02.50

$$\tilde{V}_1 = v_1\,\partial_v + \psi_1\,\partial_\psi + \cdots,$$
$$\tilde{V}_2 = (-v_3 + 6v^2v_1 + \tfrac{3}{2}v_1\psi\psi_1 + \tfrac{3}{2}v\psi\psi_2 + \tfrac{3}{4}\psi\psi_3 + \tfrac{3}{4}\psi_1\psi_2)\partial_v + (-4\psi_3 + 6vv_1\psi + 6v^2\psi_1 - 6v_1\psi_1 - 3v_2\psi)\partial_\psi,$$
$$\tilde{V}_3 = -2v\,\partial_v - \psi\,\partial_\psi - 2x\tilde{V}_1 - 6t\tilde{V}_2, \tag{2.6a}$$

and the vector field

$$\tilde{V}_4 = V_4^v\,\partial_v + V_4^\psi\,\partial_\psi, \tag{2.6b}$$

whereas in (2.6b) V_4^v, V_4^ψ are given by

$$\begin{aligned} V_4^v = {} & v_5 - 10v_3v^2 - 40v_2v_1v - 10v_1^3 + 30v_1v^4 \\ & - \tfrac{15}{4}\psi\psi_5 - \tfrac{25}{4}\psi_1\psi_4 - \tfrac{5}{2}\psi_2\psi_3 - \tfrac{15}{2}v\psi\psi_4 - 5v\psi_1\psi_3 \\ & - 15v_1\psi\psi_3 + \tfrac{15}{2}v^2\psi\psi_3 - 5v_1\psi_1\psi_2 + \tfrac{15}{2}v^2\psi_1\psi_2 \\ & - 15v_2\psi\psi_2 + 15v_1v\psi\psi_2 + 15v^3\psi\psi_2 \\ & + 45v_1v^2\psi\psi_1 - \tfrac{15}{2}v_3\psi\psi_1, \\ V_4^\psi = {} & 16\psi_5 + (40v_1 - 40v^2)\psi_3 \\ & + (60v_2 - 120v_1v)\psi_2 \\ & + (50v_3 - 100v_2v - 60v_1v^2 - 70v_1^2 \\ & + 30v^4)\psi_1 + (15v_4 - 30v_3v - 30v_2v^2 \\ & - 60v_2v_1 - 60v_1^2v + 60v_1v^3)\psi \end{aligned} \tag{2.7}$$

Note that the vector fields $\tilde{V}_1,\ldots,\tilde{V}_3$ are equivalent[3] to

$$V_1 = \partial_x, \quad V_2 = \partial_t, \quad V_3 = 6t\,\partial_t + 2x\,\partial_x - 2v\,\partial_v - \psi\,\partial_\psi. \tag{2.8}$$

III. NONLOCAL SYMMETRIES OF THE SUPER mKdV EQUATION

In order to construct the recursion operator for higher-order symmetries we introduce nonlocal variables. They can be introduced by prolongation of the exterior differential system I or $D^\infty I$ by means of potential forms or equivalently, by prolongation of the total partial derivatives D_x, D_t. For details the reader is referred to Refs. 3 and 9.

We first construct the potential forms P_1 and P_2,

$$P_1 = dp_1 - dx(v) - dt(2v^3 - v_2 + \tfrac{3}{4}\psi\psi_2 + \tfrac{3}{2}v\psi\psi_1),$$
$$P_2 = dp_2 - dx(p_{2x}) - dt(p_{2t}), \tag{3.1a}$$

where

$$\begin{aligned} p_{2x} = {} & v^2 + \tfrac{1}{4}\psi\psi_1, \\ p_{2t} = {} & 3v^4 - 2v_2v + v_1^2 - \psi\psi_3 + 2\psi_1\psi_2 \\ & + \tfrac{3}{2}v\psi\psi_2 - 3v_1\psi\psi_1 + \tfrac{9}{2}v^2\psi\psi_1. \end{aligned} \tag{3.1b}$$

The nonlocal variables p_1, p_2 are just

$$p_1 = \int_{-\infty}^{x} v\,dx, \quad p_2 = \int_{-\infty}^{x} \left(v^2 + \frac{1}{4}\psi\psi_1\right) dx, \tag{3.2}$$

whereas the integrals in (3.2) are to be considered as formal ones. Motivated by the results obtained for the classical (ungraded) KdV equation and super KdV equation,[12] our search is for a nonlocal vector field V of the form

$$V = \alpha_1 t\tilde{V}_4 + \alpha_2 x\tilde{V}_2 + \alpha_3 p_2\tilde{V}_1 + V^*, \tag{3.3}$$

where $\tilde{V}_4, \tilde{V}_2, \tilde{V}_1$ are defined by (2.6) and (2.7), $\alpha_1, \alpha_2, \alpha_3$ are constants, and V^* is a vector field yet to be determined.

The prolongation of the vector field V (3.4) towards the variables $\partial_{v_1}, \partial_{\psi_1}, \ldots$ is obtained by the action of the prolonged total partial derivative vector fields $\tilde{D}_x, \tilde{D}_t$, where

$$\tilde{D}_x = D_x + v\,\partial_{p_1} + (v^2 + \tfrac{1}{4}\psi\psi_1)\partial_{p_2},$$
$$\tilde{D}_t = D_t + (2v^3 - v_2 + \tfrac{3}{4}\psi\psi_2 + \tfrac{3}{2}v\psi\psi_1)\partial_{p_1} + p_{2t}\,\partial_{p_2}. \tag{3.4}$$

We now apply the symmetry condition including the nonlocal variables p_1, p_2, i.e.,

$$\begin{aligned} & L_V(v_t - 6v^2v_1 + v_3 - \tfrac{3}{4}\psi_1\psi_2 - \tfrac{3}{4}\psi\psi_3 \\ & \quad - \tfrac{3}{2}v_1\psi\psi_1 - \tfrac{3}{2}v\psi\psi_2) \equiv 0, \\ & L_V(\psi_t - 6v^2\psi_1 + v_1\psi_1 - 60v_1\psi + 3v_2\psi + 4\psi_3) \equiv 0, \end{aligned} \tag{3.5}$$

where now by "$\equiv 0$" we mean vanishing of the Lie derivative on the submanifold of

$$J(x,t;v,\psi,p_1,p_2) = \{(x,t,v,\psi,p_1,p_2,v_1\psi_1,\ldots\},$$

defined by (1.1) and its differential consequences, together with

$$\begin{aligned} & p_{1x} = v, \quad p_{1t} = 2v^3 - v_2 + \tfrac{3}{2}v\psi\psi_1, \\ & p_{2x} = v^2 + \tfrac{1}{4}\psi\psi_1, \\ & p_{2t} = 3v^4 - 2v_2v + v_1^2 - \psi\psi_3 + 2\psi_1\psi_2 \\ & \quad + \tfrac{3}{2}v\psi\psi_2 - 3v_1\psi\psi_1 + \tfrac{9}{2}v^2\psi\psi_1. \end{aligned} \tag{3.6}$$

Conditions (3.5) lead to an overdetermined system of partial differential equations for the defining functions of V^*, whose dependency on the jet variables $v_1, \psi_1, \ldots$ is induced by the standard grading of the super mKdV equation, i.e.,

$$\deg(x) = -1, \quad \deg(t) = -3,$$
$$\deg(v) = 1, \quad \deg(\psi) = \tfrac{1}{2},$$
$$\deg(p_1) = 0, \quad \deg(p_2) = 1.$$

So we are searching for a vector field V, whose ∂_v and ∂_ψ components are of degree $\leqslant 3$ and $\leqslant 2\frac{1}{2}$, respectively.

Solving the overdetermined system of partial differential equations leads to the following result.

Theorem: The vector field V defined by

$$V = -\tfrac{3}{2}t\tilde{V}_4 - \tfrac{1}{2}x\tilde{V}_2 + p_2\tilde{V}_1 + V^*, \tag{3.7a}$$

where

$$\begin{aligned} V^* = {} & (-\tfrac{3}{2}v_2 + 2v^3 + v\psi\psi_1 + \tfrac{7}{8}\psi\psi_2)\partial_v \\ & + (-5\psi_2 - v\psi_1 + 4v^2\psi - 4v_1\psi)\partial_\psi \end{aligned} \tag{3.7b}$$

is a nonlocal (higher-order) symmetry of the super mKdV equation.

Remark 1: Due to the fact that v, ψ, p_2 satisfy the super mKdV equation (2.1), including differential consequences and (3.6), which do not depend on x, t, p_2 explicitly, the coefficients of x, t, p_2 in (3.3) can be proved to be symmetries.

Remark 2: The solution of (3.5) does admit an additional nonlocal symmetry, i.e.,

$$V = e^{2p_1}\partial_v. \tag{3.8}$$

In order to compute the Lie bracket of the vector fields $\tilde{V}_1,\ldots,\tilde{V}_4$ and V, we would have to extend the results of Sec. III towards the nonlocal variables. This leads to results similar to the results in the ungraded case, obtained for the mas-

sive Thirring model.[3] This is not the direction we shall pursue in the next section; instead we shall construct the recursion operator for higher-order symmetries leading to the commuting flows, starting from the nonlocal higher-order symmetry (3.7).

IV. THE RECURSION OPERATOR FOR SYMMETRIES OF sKdV EQUATIONS

In the case of the classical KdV equation, i.e., the ungraded case, the Lenard recursion operator is obtained by a construction based on a nonlocal vector field, i.e., the ungraded analog of the vector field V of Sec. IV, and the Hamiltonian structure of the mKdV equation.[10,11] The construction of the recursion operator for symmetries of the super mKdV equation can be obtained in a similar way and is given below. A formal proof of its properties and the fact that the higher-order symmetries commute is beyond the scope of the paper.

The super mKdV equation (1.1) can be written in the following Hamiltonian form:

$$\begin{pmatrix} u_t \\ \psi_t \end{pmatrix} = \begin{pmatrix} \partial_x & 0 \\ 0 & 4 \end{pmatrix}^{\delta/\delta_v} \left(\frac{1}{2} v^4 + \frac{1}{2} v_x^2 + \frac{1}{2} \psi_x \psi_{xx} + \frac{3}{4} v^2 \psi \psi_x + \frac{3}{4} v \psi \psi_{xx} \right), \quad (4.1)$$

where all variational derivatives are taken to be left ones. In (4.1),

$$\Omega^{-1} = \begin{pmatrix} \partial_x & 0 \\ 0 & 4 \end{pmatrix}, \quad \Omega = \begin{pmatrix} D^{-1} & 0 \\ 0 & \frac{1}{4} \end{pmatrix} \quad \left(D^{-1} = \int_{-\infty}^{x} \cdot \, dx \right) \quad (4.2)$$

are analogous to the simplectic operator and its inverse. We now proceed in a way similar to the ungraded case and calculate the variational derivative of Ωv, i.e., $(\Omega V)'$ and its adjoint $(\Omega V)'^*$. Then we are led to the recursion operator for symmetries by (cf. Refs. 10 and 11)

$$T = \Omega^{-1}\{(\Omega V)' - (\Omega V)'^*\}. \quad (4.3)$$

A simple and straightforward computation starting from V (3.7) and (4.2) results in the following equation:

$$T = \begin{pmatrix} D & 0 \\ 0 & 4 \end{pmatrix} \begin{pmatrix} -D + 4v(D^{-1}v) & \frac{3}{4}\psi D + \frac{3}{2}v\psi + \frac{1}{4}\psi_1 - v(D^{-1}\psi_1) \\ -\frac{3}{4}\psi D + \frac{3}{2}v\psi - \frac{1}{2}\psi_1 + \psi_1(D^{-1}v) & -D^2 - v_1 + v^2 - \frac{3}{16}\psi\psi_1 - \frac{1}{4}\psi_1(D^{-1}\psi_1) \end{pmatrix}$$

$$= \begin{pmatrix} -D^2 + 4v_1(D^{-1}v) + 4v^2 & \frac{3}{4}\psi_1 D + \frac{3}{4}\psi D^2 + \frac{3}{2}v_1\psi + \frac{3}{2}v\psi_1 + \frac{3}{2}v\psi D + \frac{1}{4}\psi_2 + \frac{1}{4}\psi_1 D - v\psi_1 - v_1(D^{-1}\psi_1) \\ -3\psi D + 6v\psi - 2\psi_1 + 4\psi_1(D^{-1}v) & -4D^2 - 4v_1 + 4v^2 - \frac{3}{4}\psi\psi_1 - \psi_1(D^{-1}\psi_1) \end{pmatrix}. \quad (4.4)$$

A straightforward but tedious computation shows that

$$T\begin{pmatrix} v_1 \\ \psi_1 \end{pmatrix} = \begin{pmatrix} -v_3 + 6v_1v^2 + \frac{3}{4}\psi_1\psi_2 + \frac{3}{4}\psi\psi_3 + \frac{3}{2}v_1\psi\psi_1 + \frac{3}{2}v\psi\psi_2 \\ (6v^2 - 6v_1)\psi_1 + (6vv_1 - 3v_2)\psi - 4\psi_3 \end{pmatrix},$$

i.e.,

$$T(\tilde{V}_1) = \tilde{V}_2 \quad (4.5a)$$

and

$$T(\tilde{V}_2) = \tilde{V}_4. \quad (4.5b)$$

V. CONCLUSION

The use of computer programs to handle graded differential geometry calculations, together with the notion of nonlocal graded symmetries, leads in a constructive way to the recursion operator for higher-order symmetries of the super mKdV equation.

[1] A. M. Vinogradov, "Local symmetries and conservation laws," Acta Appl. Math. 2, 21 (1984).

[2] P. K. H. Gragert, Ph.D. thesis, University of Twente, Enschede, The Netherlands, 1981.

[3] P. H. M. Kersten, "Infinitesimal symmetries: A computational approach," CWI-tract 34, Center for Mathematics and Computer Science, Amsterdam, 1987.

[4] P. K. H. Gragert and P. H. M. Kersten, "Graded differential geometry in REDUCE 3," Memorandum No. 680, Faculty of Applied Mathematics, University of Twente, The Netherlands, 1988.

[5] B. Kostant, "Graded manifolds, graded Lie theory, and prequantization," in *Differential Geometrical Methods in Mathematical Physics, Bonn, 1975, Lecture Notes in Mathematics*, Vol. 570 (Springer, Berlin, 1977), pp. 177–306.

[6] D. Hernández Ruiperez and J. Muñoz Masqué, "Global variational calculus on graded manifolds I: Graded jet bundles, structure 1-form and graded infinitesimal contact transformations," J. Math. Pures Appl. 63, 283 (1984).

[7] B. A. Kupershmidt, "A super Korteweg–de Vries equation: An integrable system," Phys. Lett. A 102, 213 (1984).

[8] F. A. E. Pirani, D. C. Robinson, and W. F. Shadwick, *Local Jet Bundle Formulation of Bäcklund Transformations*, Mathematical Physics Studies, Vol. 1 (Reidel, Boston, 1979).

[9] I. S. Krasilshchik and A. M. Vinogradov, "Nonlocal symmetries and the theory of coverings: An Addendum to A. M. Vinogradov's 'Local symmetries and conservation laws'," Acta Appl. Math. 2, 79 (1984).

[10] F. Magri and G. Morosi, "A geometrical characterization of integrable Hamiltonian systems through the theory of Poisson–Nijenhuis manifolds," Dipartimento di Matematica "F. Enriques," Universita degli Studi di Milano, 1984.

[11] H. M. M. ten Eikelder, "Symmetries for dynamical and Hamiltonian systems," CWI-tract 17, Center for Mathematical Computer Science, Amsterdam, 1985.

[12] P. H. M. Kersten and P. K. H. Gragert, "Symmetries for the super KdV-equation," J. Phys. A: Math. Gen. 21, L579 (1988).

Explicit solutions of Grassmannian σ models

B. Piette[a)]
Institut de Physique Théorique, Université Catholique de Louvain, B-1348 Louvain-la-Neuve, Belgium

(Received 23 September 1987; accepted for publication 8 June 1988)

Classical solutions of the two-dimensional Grassmannian σ models of, respectively, Euclidean and Minkowskian type are compared with each other. Some explicit solutions of both types are constructed and some of their graphs are presented.

I. INTRODUCTION

Classical two-dimensional nonlinear σ models have been studied intensively since Pohlmeyer[1] discovered that the S^n model possesses the so-called dual symmetry. Their popularity is mainly due to their property[2–4] as integrable systems with infinitely many conservation laws with corresponding generators,[5] and to their striking resemblance to four-dimensional Yang–Mills systems. Moreover, some of those models have a direct application in gauge and string theories,[6–9] as well as in the theory of gravitation[10,11] and in solid-state physics.[12,13]

A Grassmannian σ model is a (classical) field theory defined on the two-dimensional Euclidean or Minkowskian space and taking its values in a complex Grassmannian manifold $G_{pq}(\mathbb{C})$ (see Ref. 14). One can explicitly describe the classical field either by an idempotent unitary matrix g or a projector P of rank p:

$$g\colon\ x,t\to g(x,t),\quad g\in U(p+q),\quad g^2=\mathbb{1}\,, \tag{1}$$

$$P\colon\ x,t\to P(x,t),\quad P^2=P,\quad P^+=P,\quad \mathrm{Tr}\,P=p\,. \tag{2}$$

These two descriptions are completely equivalent,[3,14,15] and related by the transformation $g=(1-2P)$. When formulated in light-cone coordinates $\xi=x+t$, $\eta=x-t$, or complex coordinates $\xi=x+it$, $\eta=x-it$, respectively, Minkowskian and Euclidean σ models are indistinguishable, with the same action and the same equation of motion, namely,

$$S=\int d^2x\,\mathrm{Tr}(g^{-1}\partial_\xi g)(g^{-1}\partial_\eta g) \tag{3}$$

$$=\int d^2x\,\mathrm{Tr}(\partial_\xi P\,\partial_\eta P)\,, \tag{4}$$

$$\partial_\xi(g\,\partial_\eta^{-1}g)+\partial_\eta(g\,\partial_\xi^{-1}g)=0\,, \tag{5}$$

$$[\partial_\xi\partial_\eta P,P]=0\,. \tag{6}$$

The only difference is the nature of the variables. Thus the question arises naturally whether a solution of one type of model will give a solution of the corresponding model of the other type after the substitution (analytic continuation)

$$t\to it\,. \tag{7}$$

Clearly the matrix g (or the projector P) will satisfy (1) after the substitution, but in general the matrix g will no longer be unitary (P will not be Hermitian). Indeed, the substitution (7) need not commute with the antilinear constraint $g^+g=\mathbb{1}$. So if one has, for real x, t, $g^+(x,t)=g^{-1}(x,t)$, then in general $g^+(x,it)\neq g^{-1}(x,it)$. [From the general definition of the nonlinear σ models given in Ref. 16 one can see that the substitution (7) will transform a solution of the Euclidean model into a solution of the Minkowskian one and vice versa only for the $\mathrm{SL}(n,\mathbb{C})$ and $\mathrm{GL}(n,\mathbb{C})$ models.] So if one wants to construct classical solutions of these models, it has to be done separately for both types of model.

The explicit methods for constructing solutions are of two types. The first class is the "holomorphic" method introduced by Borchers and Garber[17] and developed by Din and Zakrzewski (see Ref. 18 for a review of their work). This method works only for Euclidean models; translating it to the Minkowskian case would yield solutions depending on either ξ or η alone, thus with vanishing action. In the special case of $\mathbb{C}P^n$ [in the notations above: $G_{1n}(\mathbb{C})$], it gives all the finite action solutions.

The second type of method is the Bäcklund transformation described for any kind of Minkowskian σ model in Ref. 16. We have shown with Antoine[15] that this method can be adapted to work in the Euclidean case as well. We are now going to summarize those methods, and use them to construct explicit solutions, displaying some of their graphs. Then we will compare the classical solutions of the two types of models, as well as the methods used to construct them.

II. SOLUTIONS OF THE MINKOWSKIAN MODEL

Solving nonlinear differential equations is known to be a difficult task. There are no algorithms to do it, but all the well-known nonlinear equations are solved in similar ways, called generically Bäcklund transformations or inverse scattering methods. In our case those equations depend on two variables x and t, and when one tries to construct a simple nontrivial solution, one gets a function that looks like a solitary wave propagating at a constant speed. The shape of the bump sometimes depends on time, but is always related to the speed of propagation. Such solutions are called one-soliton solutions. For linear differential equations, it is well-known that given two different solutions, their sum is a third solution. Can something similar be done for a nonlinear equation? The answer is yes (at least sometimes). One performs what is called a Bäcklund transformation.[19] Take for example two one-soliton solutions propagating in opposite

[a)] Chercheur Institut Interuniversitaire des Sciences Nucléaires, Belgium.

 0022-2488/88/102190-07$02.50

directions. The transformation will yield a two-soliton solution, which looks like two solitary waves crossing each other. When the two bumps are far from each other, those new solutions look like the simple superposition of the two previous ones. Once they are close to each other, the two bumps start to interact, giving rise to a complicated shape. After the interaction they continue on their way as if nothing happened. Usually a short delay occurs during the crossing, so that looking at the solution in the (x,t) plane from the top will give a graph similar to the lowest-order Feynman diagram of the electron–positron scattering, rather than a simple X shape.

This problem has been completely solved for the two-dimensional Minkowskian σ model by Saint-Aubin *et al.*[16] They showed how to construct directly n-soliton solutions. The method is very powerful, but tedious, to apply. One first has to construct a very simple solution called a vacuum solution (whose action is zero), and multiply it by the so-called dressing matrix to get the new solution. The construction of the dressing matrix is the tedious part of the construction. A one-soliton solution can be constructed explicitly, but from the complexity of the result obtained (10), it is clear that the explicit construction of any multisoliton solution is hopeless. Fortunately, once the vacuum solution has been computed, the rest of the construction is purely algebraic and pointwise, ξ and η playing the role of mere parameters. So all the tedious part can be performed numerically on a computer so as to produce graphs of those solutions.

We have constructed one-, two-, and three-soliton solutions of the Minkowskian σ model on $\mathbb{C}P^1$. In all three cases our starting point (vacuum solution) is the matrix

$$g(x,t)=\begin{pmatrix}\cos x & -\sin x\\ -\sin x & -\cos x\end{pmatrix}, \tag{8}$$

which represents a static, "wavy" background.

A. One-soliton solution

A k-soliton solution is parametrized by k complex numbers (also called poles—see Ref. 16), $\lambda_k=\lambda_k'+i\lambda_k''$ ($|\lambda_k|\neq 1$), from which the speed and the shape of each soliton will be fixed. Introduce the following notation:

$$w=\frac{1}{2}\left(\frac{\xi}{1+\bar\lambda}+\frac{\eta}{1-\bar\lambda}\right),\qquad u=2\,\mathrm{Re}(w),\quad v=2\,\mathrm{Im}(w)\,. \tag{9}$$

Then the one-soliton solution we have constructed takes the form

$$\begin{aligned}\tilde g_{11}=-\tilde g_{22}=\cos x-\frac{1}{\Lambda^2}\Bigg\{&\frac{1}{\lambda''}[\lambda'\sin x\sinh 2v+\lambda''\cos x\cosh 2v]\\&+\frac{1}{|\lambda|^2-1}[\cos(2u-3x)-|\lambda|^2\cos(2u-x)]\Bigg\},\end{aligned} \tag{10a}$$

$$\begin{aligned}\tilde g_{12}=\bar{\tilde g}_{21}=-\sin x-\frac{1}{\Lambda^2}\Bigg\{&\frac{1}{\lambda''}[\lambda'\cos x\sinh 2v-\lambda''\sin x\cosh 2v]\\&+\frac{1}{|\lambda|^2-1}[\sin(2u-3x)+|\lambda|^2\sin(2u-x)]\\&+i\left[\frac{\lambda'}{\lambda''}\sin(u-x)\cosh v-\frac{|\lambda|^2+1}{|\lambda|^2-1}\cos(u-x)\sinh v\right]\Bigg\},\end{aligned} \tag{10b}$$

where we have put

$$\Lambda^2=|\lambda|^2\left[\frac{4\cos^2(u-x)}{(1-|\lambda|^2)^2}+\frac{\cosh^2 v}{\lambda''^2}\right]. \tag{11}$$

The solution (10), with $\lambda=1.1+1.1i$, is displayed in Fig. 1 in the form of a computer-generated plot similar to those given by Saint-Aubin in the real case.[6] As an independent check we have recalculated the same solution numerically. The resulting graph is indistinguishable from the one obtained by plotting the analytical solution.

The solution plotted in Fig. 1 clearly deserves its name of "one-soliton"; it describes a single irregularity (bump)

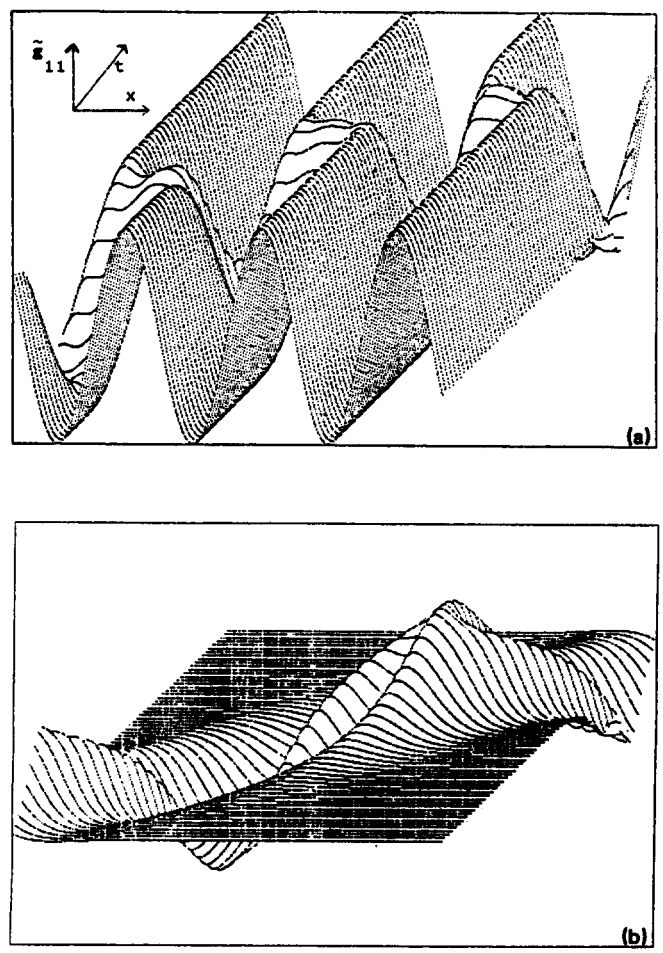


FIG. 1. The one-soliton solution (10) of the Minkowskian $\mathbb{C}P^2$ model with pole $\lambda=1.1+1.1i$: (a) $\tilde g_{11}=-\tilde g_{22}$, (b) $\mathrm{Im}\,\tilde g_{12}=-\mathrm{Im}\,\tilde g_{21}$.

propagating with constant speed $V=(1+|\lambda|^2)/2\lambda'>1$ on top of the background given by the vacuum solution g of (8) (the background is flat for the imaginary part and wavy for the real part). Note that the soliton is not simply superposed to the "wavy vacuum," it also induces a shift in it. It is clear from the conformal invariance that the model also has solutions with $V<1$. Because of this invariance, the absolute scale of the graphs is irrelevant and is therefore not indicated on Fig. 1. Because of the unitarity the size of the bump is normalized, but its shape depends on λ.

B. Multisoliton solution

For the two-soliton solution of our example shown in Fig. 2, we have chosen for the two poles $\lambda_1=1.1+1.1i$ as before and $\lambda_2=-1.1+1.1i$. This solution describes the collision of two solitons with the same asymptotic behavior but opposite speeds. One sees clearly that the main effect of the collision is a phase shift, showing that the two-soliton solution is a nonlinear superposition of two one-soliton solutions.

Exactly the same behavior is observed for the three-soliton solution plotted in Fig. 3 (the third pole here is $\lambda_3=1+4i$). Thus the situation is entirely similar to the one found for other familiar nonlinear equations, such as KdV or sine–Gordon, and also for the σ models on real Grassmann manifolds.[16]

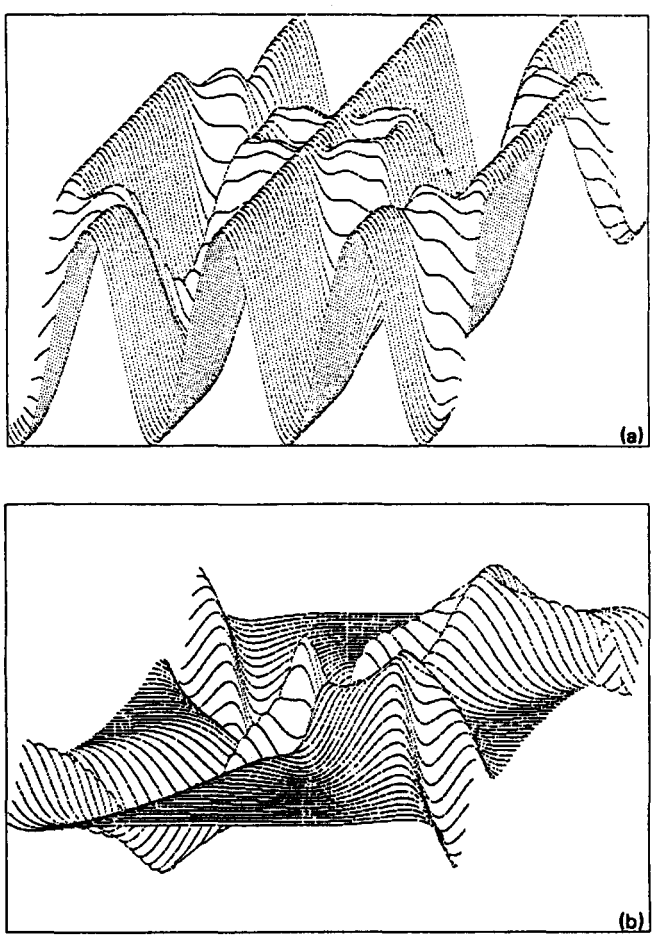


FIG. 2. The corresponding two-soliton solution, with poles $\lambda_1=1.1+1.1i$, $\lambda_2=-1.1+1.1i$: (a) $\tilde{g}_{11}$, (b) $\mathrm{Im}\,\tilde{g}_{12}$.

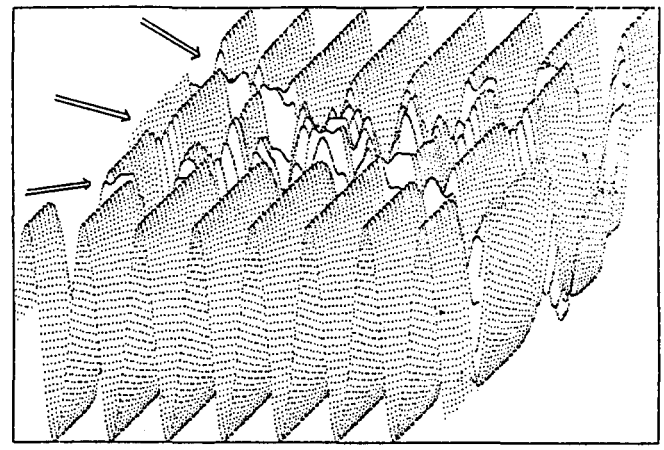


FIG. 3. The matrix element $\tilde{g}_{11}$ of the three-soliton solution with poles $\lambda_1=1.1+1.1i, \lambda_2=-1.1+1.1i, \lambda_3=1+4i$. The three solitons incoming from the left are indicated by arrows.

III. SOLUTIONS OF THE EUCLIDEAN MODEL

A peculiarity of the Euclidean models is that if a projector P satisfies

$$(1-P)\partial_\eta P=0 \quad \text{or} \quad (1-P)\partial_\xi P=0, \tag{12}$$

then it automatically satisfies Eq. (6) and is called a self-dual (anti-self-dual) solution. If its action is finite, then it is also called an instanton (anti-instanton) solution. To construct the most general self-dual solutions[12] one starts from a holomorphic $n\times p$ matrix F of maximal rank,

$$\partial_\eta F=0, \tag{13}$$

and constructs the projector

$$P=F(F^+F)^{-1}F^+. \tag{14}$$

As F is of maximal rank, P is well defined and satisfies (12a). If the entries of F are rational functions of ξ, then P is an instanton solution. Similarly, antiholomorphic functions will give the most general anti-self-dual solutions.

To construct other solutions of the Grassmannian model, one introduces the operator

$$P_\xi(F)=\partial_\xi F-F(F^+F)^{-1}F^+\partial_\xi F, \tag{15}$$

and defines its powers by

$$P_\xi^n(F)=P_\xi(P_\xi^{n-1}(F)), \tag{16}$$

where we have assumed that all the powers of $P_\xi(F)$ are of maximal rank. It can be shown[18] that

$$P_\xi^r(F)^+P_\xi^s(F)=0, \quad \forall r\neq s, \tag{17}$$

which implies that $P_\xi^s(F)=0$ for some s. It is then easy to show[18,20] that

$$P_k=P_\xi^k(F)[P_\xi^k(F)^+P_\xi^k(F)]^{-1}P_\xi^k(F)^+ \tag{18}$$

is a solution of the Grassmannian model of rank p for all k, and that if F is a rational function of ξ, then the action is finite. Moreover, the last one, P_{s-1}, is an anti-instanton solution. In the $\mathbb{C}P^n$ case, it gives all the finite action solutions. In the general Grassmannian case, many other solutions can be constructed in a similar way (see Refs. 15, 18, 20, and 21 for a detailed description), but we are not going to describe them here. The action of the above solutions reads

$$S(P_k) = \int d^2x \sum_{i=0}^{k} \mathrm{Tr}[(P_\xi^{i+1}(F)^+ P_\xi^{i+1}(F)) \times (P_\xi^i(F)^+ P_\xi^i(F))^{-1}], \quad (19)$$

which can easily be shown to be an integer in units of 2π. Let us consider an example. Take

$$F^t(\xi) = (1 + \xi + \xi^5, \xi + \xi^3, \xi^2 + \xi^4), \quad (20)$$

from which we can construct the orthonormal vector

$$Z = F/(F^+F)^{-1/2} \quad (21)$$

or, equivalently, the projector

$$P = ZZ^+. \quad (22)$$

We have shown in Fig. 4 some of the matrix elements of Z.

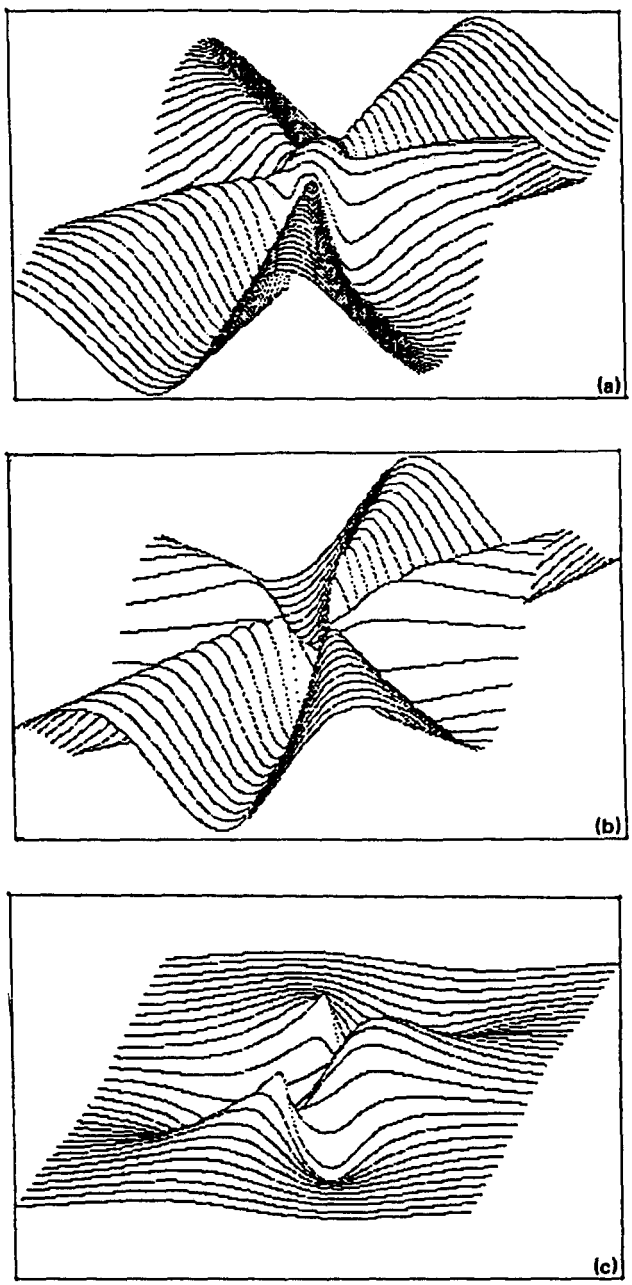


FIG. 4. The five-instanton solution [(20), (21)] of the CP^2 model: (a) Re Z_1, (b) Im Z_1, (c) Re Z_2.

One sees that the last two components of Z vanish at infinity, where the first one, being of the larger degree, looks like a starfish with five branches. It is a general feature of the instanton solutions that the highest degree in ξ of the function F from which Z is built corresponds to the topological charge of the solution, which is also called the winding number at infinity; it is also equal to the action of the solution. It is interesting to see that this number can be read directly off the graph of the solution. In our example F is of degree 5; thus the solution has an action equal to 5 (in units of 2π). It is usually called a five-instanton solution. If one takes a nonpolynomial function such as

$$F^t(\xi) = (\sin\xi, \cos\xi, \sin\xi + \cos\xi), \quad (23)$$

the vector Z will have an infinity of branches (Fig. 5) and an infinite action. The problem when plotting the entries of Z (P or g) is that there are many of them, and, moreover, they are only defined up to a gauge transformation. The Euclidean model being seen as a static model, it appears natural to plot the Lagrangian density of the solutions rather than their entries. This is a real, gauge-invariant function, and it will show where the physical things happen. Consider, for example, the instanton solutions corresponding to the following functions:

$$F_1 = \begin{pmatrix} \xi - 2 \\ \xi - 2 + i/2 \\ \xi - 2 - i/2 \end{pmatrix}, \quad F_2 = \begin{pmatrix} \xi + 2 \\ \xi + 2 - i/2 \\ \xi + 2 + i/2 \end{pmatrix},$$
$$F_3 = \begin{pmatrix} (\xi - 2)(\xi + 2) \\ (\xi - 2 + i/2)(\xi + 2 - i/2) \\ (\xi - 2 - i/2)(\xi + 2 + i/2) \end{pmatrix}. \quad (24)$$

The first two solutions are one-instanton solutions, the third one being a two-instanton solution. One sees from their Lagrangian density plots (Fig. 6) where those names come from. In general the Lagrangian density of an instanton solution of topological charge (or action) equal to r has r bumps (instantons, possibly on top of each other) localized by the zeros of the entries of F. One sees from the graphs that the first instanton is localized around 3, the second one being near -3. The two instantons of the third solution are localized at the same place. In general, the instantons will be around the point where the different entries of the holomor-

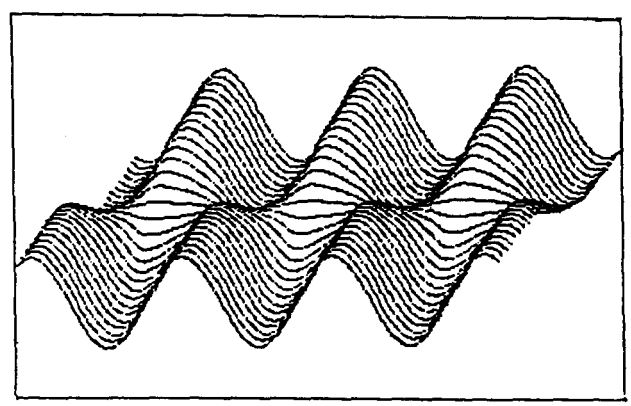
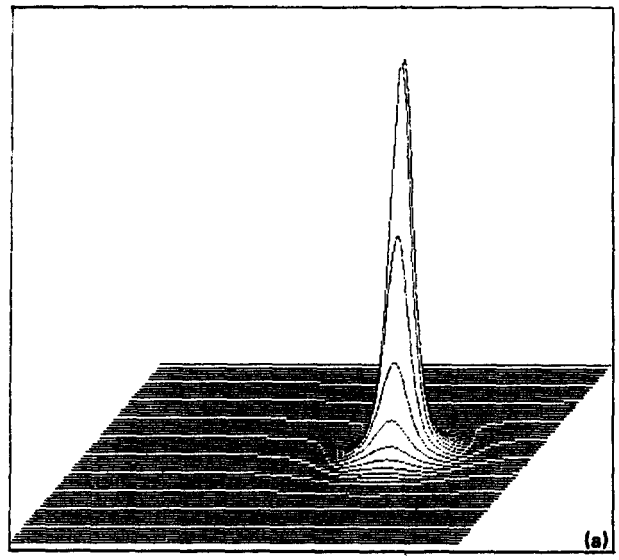
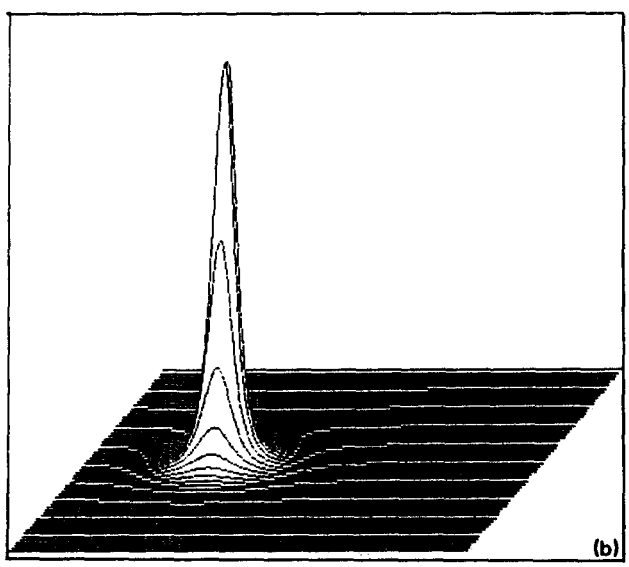


FIG. 5. The trigonometric solution [(21), (23)] of the same model: Im Z_1.





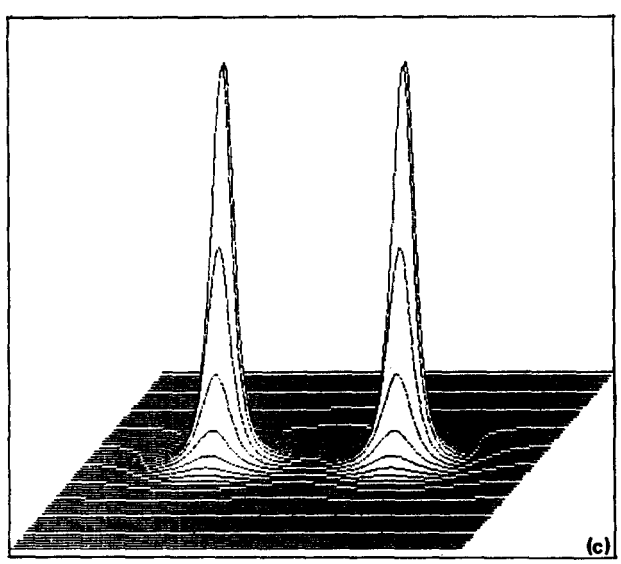


FIG. 6. The Lagrangian density [the first of Eqs. (26)] of instanton solutions (24) of the CP^2 model: (a) L for F_1, (b) L for F_2, (c) L for F_3.

phic vector F will have nearby zeros, and the farther apart those zeros will be from each other, the more spread out the instanton will be. When varying the coefficient, one can force a zero of each entry to converge to the same value. Doing that, the corresponding instanton will become more and more localized (very thin but huge), ending like a δ peak. This limit corresponds to the elimination of an instanton. Note that this limit does not commute with the computation of the action. In fact if we perform such a variation on the second instanton of our third example, doing it on the solution itself we will end up with our first one-soliton solution. If we do it in the computation of the action, the result will always be 2, even in the limit.

One can see also from our graph that our third solution can be seen as a kind of linear superposition of the two first ones. All this is true also for the anti-instanton solution, the only difference being that the topological charge is negative, and the action of those solutions is equal to the absolute value of the topological charge. Noninstanton solutions can be seen as a mixture (nonlinear superposition) of instanton and anti-instanton (see Ref. 11 for a detailed description).

Starting with F_3 in (24) one can construct the following instanton, noninstanton, and anti-instanton solutions:

$$P_0=\frac{FF^+}{F^+F},\quad P_1=\frac{P_\xi FP_\xi F^+}{P_\xi F^+P_\xi F},\quad P_2=\frac{P_\xi^2FP_\xi^2F^+}{P_\xi^2F^+P_\xi^2F}, \tag{25}$$

for which the corresponding Lagrangian densities are, respectively,

$$L_0=\frac{|P_\xi F|^2}{|F|^2},\quad L_1=\frac{|P_\xi F|^2}{|F|^2}+\frac{|P_\xi^2F|^2}{|PF|^2},\quad L_2=\frac{|P_\xi^2F|^2}{|PF|^2} \tag{26}$$

[see Figs. 6(c) and 7]. In our example, the two solitons are located on the real line around 2 and -2, the two anti-instantons being on the imaginary line around $2i$ and $-2i$. The noninstanton solution P_1 is clearly a superposition (linear for L but not for the solution itself) of the instanton and anti-instanton solutions. For the other CP^n and Grassmannian models, noninstanton solutions are somewhat more sophisticated superpositions of instanton and anti-instanton solutions.

IV. COMPARISON BETWEEN EUCLIDEAN AND MINKOWSKIAN MODELS

We have seen that the solutions of the Euclidean and Minkowskian σ models can be constructed in a very different way. So it is natural to ask if the two methods described above can be adapted to the other case. Can the holomorphic method used to construct Euclidean solutions be transformed so as to give Minkowskian ones? It is clear from the construction that one of the important points in the construction is that ξ and η are complex conjugates of each other. So, as is easy to check, when applied to construct Minkowskian solutions this method will give solutions that depend only on ξ or η, and so will be trivial. Rather than using the light-cone coordinates, one can use the hyperbolic complex variables $\xi=t+x\varepsilon$, $\eta=t-x\varepsilon$, where $\varepsilon^2=1$ and

$\bar{\varepsilon} = -\varepsilon$. This will yield nontrivial vacuum solutions of a Minkowskian σ model valued in a hyperbolic complex Grassmannian manifold (see Ref. 22 for more details).

On the other hand, the multi-Bäcklund transformation method can be modified to act on Euclidean solutions.[14,21] For example, starting from the vacuum solution

$$g_0 = \begin{pmatrix} \cos 2x & \bar{d}_1 \sin 2x & \bar{d}_2 \sin 2x \\ d_1 \sin 2x & (|d_1|^2 - |d_2|^2)\cos 2x & d_1\bar{d}_2(1-\cos 2x) \\ d_2 \sin 2x & \bar{d}_1 d_2(1-\cos 2x) & (|d_2|^2 - |d_1|^2)\cos 2x \end{pmatrix}, \tag{27}$$

where d_1 and d_2 are two complex numbers such that

$$|d_1|^2 + |d_2|^2 = 1, \tag{28}$$

one can construct the following one-soliton solution:

$$g(\lambda) = g_0 - \frac{1}{\Lambda}\left[\frac{\cos u}{2\,\mathrm{Re}\,\lambda}\{\bar{\lambda}V(y)V(w)^+ + \lambda V(w)V(y)^+\} - \frac{\cos V}{|\lambda|^2+1}\{V(w)V(w)^+ + |\lambda|^2 V(y)V(y)^+\}\right], \tag{29}$$

where

$$y = \xi/(1-\bar{\lambda}) + \eta/(1+\bar{\lambda}),$$
$$u = y - \bar{y},$$
$$v = 2x - y - \bar{y}, \tag{30}$$
$$w = 2x - y,$$

$$\Lambda = |\lambda|^2\left[\left(\frac{\cos u}{2\,\mathrm{Re}\,\lambda}\right)^2 - \left(\frac{\cos v}{|\lambda|^2+1}\right)^2\right], \tag{31}$$

$$V(z) = [\cos z, d_1 \sin z, d_2 \sin z]^t. \tag{32}$$

As can be seen from the graph (Fig. 8), this solution is quite different from the Euclidean solutions we have constructed previously. Its nonpolynomial dependence indicates that this solution has infinite action, and it is not obvious to see if one can obtain it by the holomorphic method. So rather than starting from our vacuum solution, one can try to apply the Bäcklund transformation on an instanton solution constructed before. This work has been done by Sasaki,[21] who showed that applying the multi-Bäcklund transformation on a CP^n instanton solution constructed from a holomorphic vector F gives another instanton solution, which can be constructed from another holomorphic vector $F' = \alpha F$, where α

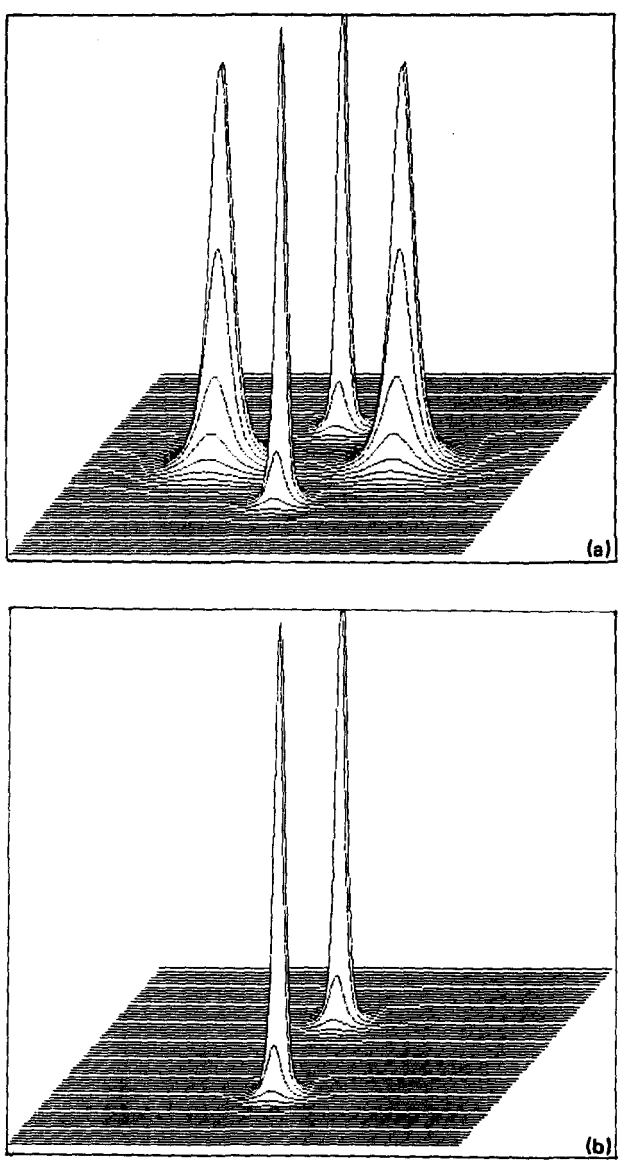


FIG. 7. The Lagrangian density (26) of (a) noninstanton and (b) anti-instanton solutions of the CP^2 models constructed from F_3 (24).

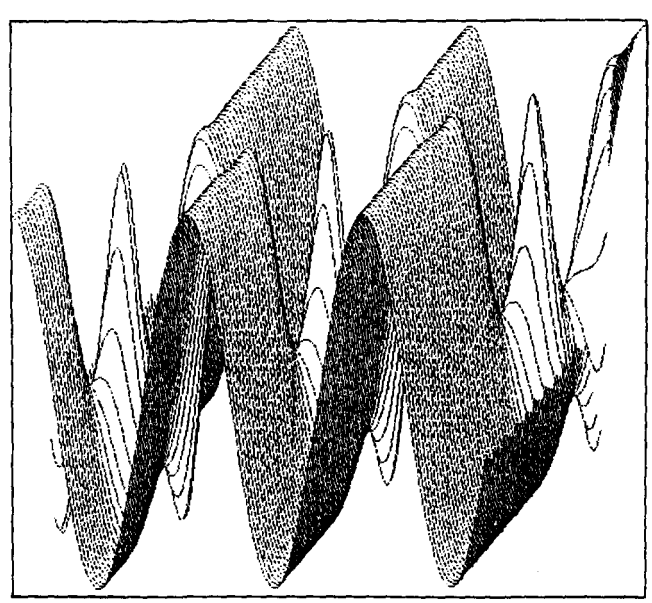


FIG. 8. The matrix element g_{11} of (29) with pole $\lambda = 2 - i$ of the Euclidean CP^2 model.

is some specified invertible matrix independent on ξ. The same applies to the noninstanton solution for which all the $P_\xi F$ will be transformed into $P_\xi F'$. This means that in general an r-instanton solution will be transformed into another r-instanton solution. The only effect of the Bäcklund transformation would be to move the position of the solitons and modify their shape. So we may say that in the $\mathbb{C}P^n$ case this transformation acts nearly trivially on finite action solutions. We suspect it is also true for Grassmannian solutions. It is quite surprising to see that this transformation is very similar to the one obtained by Arsenault *et al.* in Ref. 23, where by exponentiating the infinite-dimensional Lie algebra described in Ref. 5 on the Grassmannian solutions they showed that it reduces to a finite-dimensional $\mathrm{Sl}(N,\mathbb{C})$ action. More explicitly, the solution $P^r_\xi F$ will become a new solution $P^r_\xi F'$, where $F' = SF$, $S \in \mathrm{Sl}(N,\mathbb{C})$. It is not easy to see if α has a unit determinant or not, but if not, it cancels out when one computes the projector P. It is also not clear whether any matrix S belonging to $\mathrm{Sl}(N,\mathbb{C})$ can be obtained from a product of different α's (up to the determinant factorization). Thus the Bäcklund transformation can be considered at least as a subcase of the transformation found by Arsenault *et al.*[23]

So we have seen that Minkowskian and Euclidean Grassmann σ models are strikingly different. Each one possesses its own method for constructing classical solutions. Those methods can be adapted to the other type of model, but then one sees that it gives rather trivial results.

ACKNOWLEDGMENTS

It is a pleasure to thank Y. Saint-Aubin and W. J. Zakrzewski for many stimulating discussions and suggestions about the contents of this paper. I would like also to thank J. P. Antoine for his scientific advice and the help he provided me for the preparation of this manuscript.

[1]K. Pohlmeyer, Commun. Math. Phys. **46**, 207 (1976).
[2]H. Eichenherr and M. Forger, Nucl. Phys. B **155**, 381 (1979).
[3]H. Eichenherr and M. Forger, Nucl. Phys. B **164**, 528 (1980); **282**, 745(E) (1987).
[4]M. Forger, in *Non-linear Partial Differential Operators and Quantization Procedures (Proceedings, Clausthal, 1981), Lecture Notes in Mathematics*, Vol. 1037, edited by S. I. Anderson and H.-D. Doebner (Springer, Berlin, 1983), pp. 38–80.
[5]M. Jacques and Y. Saint-Aubin, J. Math. Phys. **28**, 2463 (1987).
[6]A. Neveu and P. Pananicolaou, Commun. Math. Phys. **53**, 31 (1978).
[7]V. V. Nesterenko, Dubna preprints JINR-E2-82-745 and JINR-E2-82-761; Lett. Math. Phys. **7**, 287 (1983).
[8]A. Kundu, J. Phys. A: Math. Gen. **19**, 1303 (1986).
[9]See the recent review of A. M. Perelomov, Phys. Rep. **146**, 135 (1987).
[10]P. O. Mazur, J. Phys. A: Math. Gen. **15**, 3173 (1982); Acta Phys. Pol. B **14**, 219 (1983); Phys. Lett. A **100**, 341 (1984).
[11]P. Breitenlohner, D. Maison, and G. Gibbons, "4-dimensional black holes from Kaluza–Klein theories," Max Planck Institute preprint MPI-PAE/PTh27/87-LPTENS 87/9.
[12]L. Schäfer and F. Wegner, Z. Phys. B **38**, 113 (1980); S. Hikami, Phys. Rev. B **24**, 2175 (1981); K. B. Efetov, Adv. Phys. **32**, 53 (1983).
[13]R. Oppermann, Nucl. Phys. B **280** [FS18], 753 (1987).
[14]J.-P. Antoine and B. Piette, J. Math. Phys. **28**, 2753 (1987); in *Recent Developments in Mathematical Physics (Proceedings, Shladming, 1987)*, edited by P. Mitter (Springer, Berlin, 1987).
[15]J.-P. Antoine and B. Piette, J. Math. Phys. **29**, 1687 (1988).
[16]Y. Saint-Aubin, Thèse, Université de Montréal, 1982 (unpublished); Y. Saint-Aubin, Lett. Math. Phys. **6**, 441 (1982); J. Harnad, Y. Saint-Aubin, and S. Shnider, Commun. Math. Phys. **92**, 329 (1984).
[17]H. J. Borchers and W. D. Garber, Math. Phys. **72**, 77 (1980).
[18]W. J. Zakrzewski, J. Geom. Phys. **1**, 39 (1984).
[19]C. Rogers and W. F. Shadwick, *Bäcklund Transformations and Their Applications* (Academic, New York, 1982).
[20]R. Sasaki, Z. Phys. C **24**, 163 (1984).
[21]R. Sasaki, J. Math. Phys. **26**, 1786 (1985).
[22]D. Lambert and B. Piette, Class. Quantum Grav. **5**, 307 (1988).
[23]G. Arsenault, M. Jacques, and Y. Saint-Aubin, Lett. Math. Phys. **15**, 65 (1988).

Integrability of restricted multiple three-wave interactions. II. Coupling constants with ratios 1 and 2

Frank Verheest
Instituut voor theoretische mechanika, Rijksuniversiteit Gent, Krijgslaan 281, B-9000 Gent, Belgium

(Received 26 January 1988; accepted for publication 8 June 1988)

The system of ordinary differential equations describing the multiple three-wave interactions with a common pump (daughter) wave is proved to be completely integrable by obtaining the necessary $2N + 1$ independent first integrals in involution for the case when the coupling constants and the frequency mismatches have ratio 1 and/or 2. This case was deemed integrable on the basis of a Painlevé analysis, but a direct proof has been lacking for some time. The first integrals are the $N + 1$ quadratic Manley–Rowe relations, the cubic Hamiltonian, $N - 1$ quartic integrals (analogous to the ones needed for complete integrability in the case of equal coupling constants and detunings in all wave triads), and a new sixth-order integral involving all wave quantities. The form of this last invariant was deduced from the recent result for the analogous interaction between five waves ($N = 2$), and essentially made possible by the proper use of irreducible forms, elementary building blocks for polynomial first integrals in involution with the Manley–Rowe invariants.

I. INTRODUCTION

The last decade has seen a rather extensive study of nonlinear wave interactions among a great number of wave triads coupled through one common wave. This is because of their many applications in different physical circumstances, such as the modeling of ocean waves or plasma turbulence.[1–8] Where the emphasis was on the integrability of the Hamiltonian system of coupled ordinary differential equations (ODE's) describing such phenomena, a real impetus has been given by Menyuk *et al.*,[4,5] who carried out the Painlevé analysis and deemed the system integrable in only two different cases (at arbitrary initial conditions).[4]

First, all N triads have equal coupling constants, but the frequency mismatches or detunings are arbitrary, though small in view of the physical applicability. In the other case, the coupling constants for some triads have to be twice as large as for the other triads, with similar ratios for the detunings.

Several authors proved the integrability of the first case by constructing or obtaining enough (here $2N + 1$) independent first integrals in involution. Menyuk *et al.*[5] did it already via Lax operators, but with a degenerate matrix yielding less than $2N + 1$ independent first integrals, so that *ad hoc* methods were needed to construct the remaining ones. Wojciechowski *et al.*[6] obtained another Lax representation giving directly all the necessary integrals. We ourselves introduced the concept of irreducible forms[7] and showed how these led to first integrals in a rather systematic way.[8] Irreducible forms can be defined in different but equivalent ways. Originally they were thought of as the simplest possible polynomial combinations of wave quantities remaining constant on the fast oscillation time scale. Thus they could serve as building blocks for polynomial first integrals on the interaction time scale.[7] Later it became clear that the same irreducible forms are obtained by looking for the simplest polynomials in the wave amplitudes that are in involution with the Manley–Rowe relations (the quadratic first integrals immediately related to the energy exchanges inside or between the wave triads).[8]

The direct proof of integrability for the case of coupling constants and detunings having ratio 1 or 2 has until very recently eluded all efforts, apparently by whatever method. We succeeded, however, in proving the integrability of the reduced system of ODE's describing two interacting triads in the absence of detunings,[9] by a proper combination of the Yoshida–Kovalevskaya approach (giving insight into the degree of possible first integrals[10,11]) and the use of irreducible forms (allowing then a direct search as advocated by Roekaerts and Schwarz,[12] but in a much simpler way). It is the aim of the present paper to generalize this recent result for arbitrary N, and to include the appropriate detunings.

II. BASIC FORMALISM FOR N TRIADS COUPLED THROUGH A COMMON PUMP WAVE

As in Ref. 8, we start from selection rules for different triads interacting through a common pump wave

$$\omega_0 = \omega_j + \Omega_j + \delta_j \quad (j = 1,...,N) . \tag{1}$$

The complex amplitudes of the waves with frequency ω_0, ω_j, and Ω_j are denoted by c, a_j, and b_j ($j = 1,...,N$). Their slow-time evolution obeys

$$\begin{aligned} \dot{a}_j &= i\lambda_j c\bar{b}_j + (i/2)\delta_j a_j , \\ \dot{b}_j &= i\lambda_j c\bar{a}_j + (i/2)\delta_j b_j , \\ \dot{c} &= i\sum_{j=1}^{N} \lambda_j a_j b_j \quad (j = 1,...,N) , \end{aligned} \tag{2}$$

plus the complex conjugate equations. The frequency mismatches or detunings δ_j were put into (2) in a symmetric way, in contrast to Ref. 8, so as to get simpler expressions later on. The dot refers to the slow-time derivative and the bar to complex conjugation.

The system (2) can be derived from the Hamiltonian

 0022-2488/88/102197-05$02.50

$$H=\sum_{j=1}^{N}\left[\lambda_j(a_j b_j \bar{c}+\bar{a}_j \bar{b}_j c)+\frac{1}{2}\delta_j(a_j\bar{a}_j+b_j\bar{b}_j)\right], \quad (3)$$

when one considers the pairs $(a_j,\bar{a}_j)$, $(b_j,\bar{b}_j)$, and $(c,\bar{c})$ as pairs of canonically conjugated variables and writes Hamilton's equations in an appropriate way.[13]

As was said already, a direct proof of the integrability of Hamiltonian systems such as (2) requires the existence of as many independent first integrals in involution as there are degrees of freedom[14] (here $2N+1$). Two first integrals A and B are in involution when their Poisson bracket vanishes, defined here as

$$\{A,B\}=\sum_{j=1}^{N}\left(\frac{\partial A}{\partial a_j}\frac{\partial B}{\partial \bar{a}_j}-\frac{\partial A}{\partial \bar{a}_j}\frac{\partial B}{\partial a_j}+\frac{\partial A}{\partial b_j}\frac{\partial B}{\partial \bar{b}_j}-\frac{\partial A}{\partial \bar{b}_j}\frac{\partial B}{\partial b_j}\right)+\frac{\partial A}{\partial c}\frac{\partial B}{\partial \bar{c}}-\frac{\partial A}{\partial \bar{c}}\frac{\partial B}{\partial c}. \quad (4)$$

Without restrictions on λ_j and δ_j, one finds already $N+1$ independent quadratic first integrals in involution, the Manley–Rowe relations:

$$E_0=\sum_{j=1}^{N}(a_j\bar{a}_j+b_j\bar{b}_j)+2c\bar{c}\,,\qquad E_j=a_j\bar{a}_j-b_j\bar{b}_j \quad (j=1,\ldots,N)\,. \quad (5)$$

In addition, H itself is a cubic first integral, giving together $N+2$ integrals out of the $2N+1$ needed for integrability.

Defining now the irreducible forms as the simplest polynomials in the wave amplitudes, in involution with the Manley–Rowe relations (5), we see that they are

$$a_j\bar{a}_j,\quad b_j\bar{b}_j,\quad c\bar{c},\quad a_jb_j\bar{c}+\bar{a}_j\bar{b}_jc,\quad a_ib_i\bar{a}_j\bar{b}_j+\bar{a}_i\bar{b}_ia_jb_j \quad (i,j=1,\ldots,N)\,. \quad (6)$$

Since the Manley–Rowe relations plus the Hamiltonian amount already to $N+2$ independent first integrals in involution, one can try to extend them to a full set of $2N+1$ integrals. If this can be done (via polynomials), the remaining first integrals have to be in involution with, among others, the Manley–Rowe relations, implying that they must be combinations of the irreducible forms[8]—hence the search for additional integrals as polynomials in the irreducible forms, rather than in the original variables themselves. This is basically very much simpler, as we shall see.

Turning to the Painlevé analysis of (2) as carried out by Menyuk *et al.*,[4] one has a first integrable case when all coupling constants λ_j have the same value ($\lambda_j=1$ after a suitable rescaling of the amplitudes). Then there are no further restrictions on the detunings δ_j. The full proof of integrability by getting all the necessary first integrals by one method or the other was given by different authors.[5,6,8]

The second case that was deemed integrable for arbitrary initial conditions is when

$$\lambda_1=\cdots=\lambda_M=2\lambda_{M+1}=\cdots=2\lambda_N\,,\qquad \delta_1=\cdots=\delta_M=2\delta_{M+1}=\cdots=2\delta_N\,. \quad (7)$$

For this case there has been no direct proof of integrability so far. Recently we succeeded in determining the missing first integral for the special case of $N=2$ and in the absence of detunings[9] ($\lambda_1=2\lambda_2=2$, $\delta_1=2\delta_2=0$). The necessary five integrals in involution are E_0, E_1, E_2, H, and

$$I=4(a_1b_1\bar{a}_2\bar{b}_2+\bar{a}_1\bar{b}_1a_2b_2)(a_2\bar{a}_2+\bar{b}_2\bar{b}_2)-2(c\bar{a}_2\bar{b}_2+\bar{c}a_2b_2)^2-[(a_2\bar{a}_2+b_2\bar{b}_2)^2+4a_2\bar{a}_2b_2\bar{b}_2](a_1\bar{a}_1+b_1\bar{b}_1)\,. \quad (8)$$

One easily recognizes its structure as a combination of irreducible forms of total degree 6. We now set out to generalize this result for arbitrary N and for detunings obeying (7). We will rescale the amplitudes in such a way that λ_j are either 1 or 2 for simplicity.

III. FIRST INTEGRALS OF DEGREE 4

For the remainder of this paper we take the subscripts m,n in the range $\{1,\ldots,M\}$, where $\lambda_m=2$ and $\delta_m=2\delta$, and p,q in the range $\{M+1,\ldots,N\}$, with $\lambda_p=1$ and $\delta_p=\delta$. Because of what we learned from the five-wave coupling, we replace the set of irreducible forms (6) by the equivalent set of building blocks:

$$K_j=a_j\bar{a}_j+b_j\bar{b}_j,\quad E_j,\quad E_0\,,\qquad H_j=\lambda_j(a_jb_j\bar{c}+\bar{a}_j\bar{b}_jc)+\tfrac{1}{2}\delta_j(a_j\bar{a}_j+b_j\bar{b}_j)\,,\qquad L_{ij}=a_ib_i\bar{a}_j\bar{b}_j+\bar{a}_i\bar{b}_ia_jb_j \quad (i,j=1,\ldots,N)\,. \quad (9)$$

The form of H_j is inspired by the structure of the Hamiltonian, so that

$$H=\sum_{j=1}^{N}H_j\,. \quad (10)$$

A full direct search is rather difficult at arbitrary N, so that we take the structure of (8) and of the derivatives of the building blocks (9) as our guides. Using (2) we find that

$$\dot{K}_j=2i\lambda_j(\bar{a}_j\bar{b}_jc-a_jb_j\bar{c})\,,\qquad \dot{H}_j=i\lambda_j\sum_{i=1}^{N}\lambda_iV_{ij}\,,\qquad \dot{L}_{ij}=(\lambda_i/2\lambda_j)K_i\dot{K}_j+(\lambda_j/2\lambda_i)K_j\dot{K}_i+i(\delta_i-\delta_j)V_{ij} \quad (i,j=1,\ldots,N)\,, \quad (11)$$

if we denote for brevity

$$V_{ij}=a_ib_i\bar{a}_j\bar{b}_j-\bar{a}_i\bar{b}_ia_jb_j \quad (i,j=1,\ldots,N)\,. \quad (12)$$

We thus immediately see that if in the last line of (11) i and j *both* refer to subscripts of the same range $\{1,\ldots,M\}$ or $\{M+1,\ldots,N\}$, then $\lambda_i=\lambda_j$ and $\delta_i=\delta_j$ imply that $L_{ij}-\frac{1}{2}K_iK_j$ ($i,j=1,\ldots,M$ or $i,j=M+1,\ldots,N$) are first integrals. These are symmetric in i and j; furthermore,

$$L_{jj}-\tfrac{1}{2}K_j^2=-\tfrac{1}{2}E_j^2 \quad (j=1,\ldots,N)\,. \quad (13)$$

There are thus $\frac{1}{2}M(M-1)+\frac{1}{2}(N-M)(N-M-1)$ additional independent first integrals. However, not all are in involution. In a way that is reminiscent of what was done in a previous paper, when all $\lambda_j=1$ and $\delta_j=\delta$ (see Ref. 8); we rearrange the new first integrals into

$$I_m=\sum_{n=m+1}^{M}\left(L_{mn}-\frac{1}{2}K_mK_n\right) \quad (m=1,\ldots,M-1)\,, \quad (14)$$

$$I_p=\sum_{q=p+1}^{N}\left(L_{pq}-\frac{1}{2}K_pK_q\right) \quad (p=M+1,\ldots,N-1)\,,$$

or something equivalent, as the choice (14) is not unique. As the subscripts on L_{ij} or K_j refer to pairs of canonically conjugate variables $(a_j,\bar{a}_j)$ and $(b_j,\bar{b}_j)$, one sees that L_{ij} and L_{kl} have no common pair of canonically conjugate variables, unless there is at least a common subscript. If we let i,j,k refer to mutually different indices, we see that the only relevant Poisson brackets involving L_{ij} and/or K_j are

$$\{L_{ij},L_{ik}\}=K_iV_{kj}\,,\quad \{L_{ii},L_{ij}\}=2K_iV_{ji}\,,$$
$$\{L_{ij},K_j\}=2V_{ji}\quad (i,j,k=1,\ldots,N)\,. \tag{15}$$

The ones not written vanish. Using (15) we can compute that

$$\begin{aligned}\{I_m,I_n\}&=\sum_{k=m+1}^{M}\sum_{l=n+1}^{M}\Big(\{L_{mk},L_{nl}\}-\frac{1}{2}\{L_{mk},K_nK_l\}\\&\quad-\frac{1}{2}\{K_mK_k,L_{nl}\}\Big)\\&=\sum_{k=n+1}^{M}\Big(\{L_{mn},L_{nk}\}+\{L_{mk},L_{nk}\}\\&\quad-\frac{1}{2}\{L_{mn},K_n\}K_k-\frac{1}{2}\{L_{mk},K_k\}K_n\\&\quad-\frac{1}{2}\{K_n,L_{nk}\}K_m-\frac{1}{2}\{K_k,L_{nk}\}K_m\Big)\\&=0\quad (m<n,\ m,n=1,\ldots,M-1)\,.\end{aligned} \tag{16}$$

Similar results hold for $\{I_p,I_q\}$ $(p,q=M+1,\ldots,N-1)$, and of course $\{I_m,I_p\}$ $(m=1,\ldots,M-1,$ $p=M+1,\ldots,N-1)$ vanish straightaway as I_m and I_p have no common pairs of canonical variables.

There are thus $(M-1)+(N-M-1)=N-2$ additional first integrals I_m,I_p $(m=1,\ldots,M-1;$ $p=M+1,\ldots,N-1)$, mutually in involution. Because they are constructed from the irreducible forms, they are automatically in involution with the Manley–Rowe relations and, as first integrals, with H. The total of independent first integrals in involution is now already $2N$, namely, E_0, $E_1,\ldots,E_N$, H, $I_1,\ldots,I_{M-1}$, $I_{M+1},\ldots,I_{N-1}$. Only one is missing, and for reasons of symmetry it will involve quantities related to all wave triads.

IV. FINAL INVARIANT OF DEGREE 6

Looking at the structure of (8), the last first integral required to prove integrability in the case $N=2$, we see that it can be written (with $\delta_1=2\delta_2=0$) as

$$I=4L_{12}K_2-2H_2^2-2K_2^2K_1+E_2^2K_1\,. \tag{17}$$

This involves a quartic irreducible form L_{12}, with quantities of a triad in the range where $\lambda_2=2$ and another with $\lambda_1=1$, multiplied by a quadratic building block K_2 from the latter range. The term H_2 belong also to this range. It thus seems logical to expect the missing first integral in the general case to include in a similar vein all possible pairings of triads from both ranges. First, however, we get from (11) that

$$\dot{H}_p=i\sum_{q=M+1}^{N}V_{qp}+2i\sum_{m=1}^{M}V_{mp}\quad (p=M+1,\ldots,N)\,,$$

$$\dot{L}_{mp}=K_m\dot{K}_p+\tfrac{1}{4}K_p\dot{K}_m+i\delta V_{mp}\quad (m=1,\ldots,M)\,, \tag{18}$$

and hence

$$\begin{aligned}&\sum_{p=M+1}^{N}\dot{H}_p=2i\sum_{m=1}^{M}\sum_{p=M+1}^{N}V_{mp}\,,\\&\sum_{m=1}^{M}\sum_{p=M+1}^{N}\dot{L}_{mp}\\&\quad=\sum_{m=1}^{M}K_m\sum_{p=M+1}^{N}\dot{K}_p+\frac{1}{4}\sum_{m=1}^{M}\dot{K}_m\sum_{p=M+1}^{N}K_p\\&\qquad+i\delta\sum_{m=1}^{M}\sum_{p=M+1}^{N}V_{mp}\\&\quad=\mathcal{K}_A\dot{\mathcal{K}}_B+\tfrac{1}{4}\dot{\mathcal{K}}_A\mathcal{K}_B+i\delta\sum_{m=1}^{M}\sum_{p=M+1}^{N}V_{mp}\,,\end{aligned} \tag{19}$$

if we call

$$\begin{aligned}\mathcal{K}_A&=\sum_{m=1}^{M}K_m=\sum_{m=1}^{M}(a_m\bar{a}_m+b_m\bar{b}_m)\,,\\\mathcal{K}_B&=\sum_{p=M+1}^{N}K_p=\sum_{p=M+1}^{N}(a_p\bar{a}_p+b_p\bar{b}_p)\,.\end{aligned} \tag{20}$$

Keeping (17) in mind we set out to compute

$$\begin{aligned}&\frac{d}{dt}\Big[\sum_{m=1}^{M}\sum_{p=M+1}^{N}4L_{mp}\sum_{q=M+1}^{N}K_q-2\Big(\sum_{p=M+1}^{N}H_p\Big)^2\Big]\\&\quad=4\mathcal{K}_A\mathcal{K}_B\dot{\mathcal{K}}_B+\mathcal{K}_B^2\dot{\mathcal{K}}_A\\&\qquad+4i\delta\mathcal{K}_B\sum_{m=1}^{M}\sum_{p=M+1}^{N}V_{mp}\\&\qquad+4\dot{\mathcal{K}}_B\sum_{m=1}^{M}\sum_{p=M+1}^{N}L_{mp}\\&\qquad-8i\sum_{p=M+1}^{N}H_p\sum_{m=1}^{M}\sum_{q=M+1}^{N}V_{mq}\,,\end{aligned} \tag{21}$$

and try to rewrite the rhs as a total derivative. Using the detailed expressions of all quantities as functions of a_j, $\bar{a}_j$, b_j, $\bar{b}_j$, c, and $\bar{c}$, we luckily find that

$$\begin{aligned}&4i\delta\mathcal{K}_B\sum_{m=1}^{M}\sum_{p=M+1}^{N}V_{mp}\\&\quad+4\dot{\mathcal{K}}_B\sum_{m=1}^{M}\sum_{p=M+1}^{N}L_{mp}\\&\quad-8i\sum_{p=M+1}^{N}H_p\sum_{m=1}^{M}\sum_{q=M+1}^{N}V_{mq}\\&\quad=2\sum_{p,q=M+1}^{N}\sum L_{pq}\dot{\mathcal{K}}_A\\&\quad=\frac{d}{dt}2\Big[\sum_{p,q=M+1}^{N}\sum L_{pq}\mathcal{K}_A\Big]-2\mathcal{K}_A\mathcal{K}_B\dot{\mathcal{K}}_B\,.\end{aligned} \tag{22}$$

We are now home, as

$$\begin{aligned}&\frac{d}{dt}\Big[4\sum_{m=1}^{M}\sum_{p=M+1}^{N}L_{mp}\mathcal{K}_B-2\Big(\sum_{p=M+1}^{N}H_p\Big)^2\Big]\\&\quad=\frac{d}{dt}\Big[2\sum_{p,q=M+1}^{N}\sum L_{pq}\mathcal{K}_A+\mathcal{K}_B^2\mathcal{K}_A\Big]\,.\end{aligned} \tag{23}$$

Thus the final invariant is

$$I_{AB}=4\sum_{m=1}^{M}\sum_{p=M+1}^{N}L_{mp}\mathcal{K}_B-2\Big(\sum_{p=M+1}^{N}H_p\Big)^2-2\sum_{p,q=M+1}^{N}\sum L_{pq}\mathcal{K}_A-\mathcal{K}_B^2\mathcal{K}_A$$

$$=\sum_{m=1}^{M}\sum_{p,q=M+1}^{N}\sum[4(a_mb_m\bar{a}_p\bar{b}_p+\bar{a}_m\bar{b}_ma_pb_p)(a_q\bar{a}_q+b_q\bar{b}_q)-2(a_pb_p\bar{a}_q\bar{b}_q+\bar{a}_p\bar{b}_pa_qb_q)(a_m\bar{a}_m+b_m\bar{b}_m)$$

$$-(a_m\bar{a}_m+b_m\bar{b}_m)(a_p\bar{a}_p+b_p\bar{b}_p)(a_q\bar{a}_q+b_q\bar{b}_q)]-2\Big[\sum_{p=M+1}^{N}\Big(a_pb_p\bar{c}+\bar{a}_p\bar{b}_pc+\frac{\delta}{2}(a_p\bar{a}_p+b_p\bar{b}_p)\Big)\Big]^2. \tag{24}$$

This first integral is automatically in involution with the Manley–Rowe relations and the Hamiltonian, so that one needs to check whether $\{I_m,I_{AB}\}$ $(m=1,...,M-1)$ and $\{I_p,I_{AB}\}$ $(p=M+1,...,N-1)$ vanish as well. Because I_{AB} is not symmetric in the quantities of all triads [as Eq. (2) themselves are also not], both sets of Poisson brackets are not analogous. So we first see that

$$\{I_m,I_{AB}\}=4\mathcal{K}_B\sum_{n=m+1}^{M}\sum_{j=1}^{M}\sum_{p=M+1}^{N}\Big\{L_{mn}-\frac{1}{2}K_mK_n,L_{jp}\Big\}-\Big(2\sum_{p,q=M+1}^{N}\sum L_{pq}+\mathcal{K}_B^2\Big)\sum_{n=m+1}^{M}\{L_{mn},\mathcal{K}_A\}$$

$$=4\mathcal{K}_B\sum_{n=m+1}^{M}\sum_{p=M+1}^{N}\Big\{L_{mn}-\frac{1}{2}K_mK_n,L_{mp}+L_{np}\Big\}=0\quad(m=1,...,M-1). \tag{25}$$

The intervening Poisson brackets are easily shown to vanish with the help of (15). Because $\{L_{mn},\mathcal{K}_A\}$ vanishes, so will $\{L_{pq},\mathcal{K}_B\}$, and we have that

$$\{I_p,I_{AB}\}=4\mathcal{K}_B\sum_{q=p+1}^{N}\sum_{m=1}^{M}\sum_{r=M+1}^{N}\Big\{L_{pq}-\frac{1}{2}K_pK_q,L_{mr}\Big\}-4\sum_{q=p+1}^{N}\sum_{r,s=M+1}^{N}\sum\Big\{L_{pq}-\frac{1}{2}K_pK_q,H_r\Big\}H_s$$

$$-2\mathcal{K}_A\sum_{q=p+1}^{N}\sum_{r,s=M+1}^{N}\sum\Big\{L_{pq}-\frac{1}{2}K_pK_q,L_{rs}\Big\}$$

$$=4\mathcal{K}_B\sum_{q=p+1}^{N}\sum_{m=1}^{M}\Big\{L_{pq}-\frac{1}{2}K_pK_q,L_{mp}+L_{mq}\Big\}-4\sum_{q=p+1}^{N}\sum_{r=M+1}^{N}\Big\{L_{pq}-\frac{1}{2}K_pK_q,H_p+H_q\Big\}H_r$$

$$-4\mathcal{K}_A\sum_{q=p+1}^{N}\sum_{r=M+1}^{N}\Big\{L_{pq}-\frac{1}{2}K_pK_q,L_{pr}+L_{qr}\Big\}+2\mathcal{K}_A\sum_{q=p+1}^{N}\Big\{L_{pq}-\frac{1}{2}K_{pq},L_{pp}+2L_{pq}+L_{qq}\Big\}$$

$$=0\quad(p=M+1,...,N-1). \tag{26}$$

Thus we have the $2N+1$ necessary first integrals in involution needed to prove complete integrability: $E_0,E_1,...,E_N,H,I_1,...,I_{M-1},I_{M+1},...,I_{N-1},I_{AB}$. That these are indeed independent can be proved via complete induction from the case $N=2$, but this is too lengthy to be included here.

V. CONCLUSIONS

The above treatment was for multiple three-wave couplings, where the one common wave is a pump wave in each triad.

If the common wave is a daughter wave in each triad, the selection rules are

$$\Omega_j=\omega_0+\omega_j+\delta_j\quad(j=1,...,N). \tag{27}$$

This case is completely analogous and will not be detailed here.

As mentioned already elsewhere,[8] the mixed case where the common wave is a pump wave in some triads and a daughter wave in others is qualitatively quite different, but was overlooked in other studies.[5,6] For $N=2$ the selection rules are

$$\omega_0=\omega_1+\Omega_1+\delta_1,\quad \Omega_2=\omega_0+\omega_2+\delta_2. \tag{28}$$

It was surmised that this case would not be integrable, and this is borne out by the Painlevé analysis that we carried out on the system of amplitude equations derived from

$$H=\lambda_1(a_1b_1\bar{c}+\bar{a}_1\bar{b}_1c)+(\delta_1/2)(a_1\bar{a}_1+b_1\bar{b}_1)$$

$$+\lambda_2(a_2\bar{b}_2c+\bar{a}_2b_2\bar{c})+(\delta_2/2)(a_2\bar{a}_2-b_2\bar{b}_2). \tag{29}$$

Because of its negative results, the analysis will not be given here. Suffice it to say that it flounders already at the first stage, where one has to determine the weights in trying to balance the most singular terms in the amplitude equations.

Multiple triads are thus integrable only if they are *all* coupled through a common pump or a common daughter wave.

We have now covered both cases proposed by Menyuk *et al.*[4] (equal coupling constants in Ref. 8 and ratios 1 and 2 at present), essentially by exploiting the structure of invariants as polynomials in the irreducible forms. This should

amply demonstrate the power of this method in providing a fairly systematic procedure for uncovering missing first integrals, especially when combined with a proper Painlevé analysis and the Yoshida–Kovalevskaya approach.

ACKNOWLEDGMENTS

I am greatly indebted to F. Cantrijn, M. Crampin, W. Hereman, andW. Sarlet for many clarifying discussions and stimulating suggestions.

[1]K. M. Watson, B. J. West, and B. I. Cohen, J. Fluid Mech. **77**, 185 (1976).
[2]J. D. Meiss, Phys. Rev. A **19**, 1780 (1979).
[3]L. Falk, *Proceedings of the 1982 International Conference on Plasma Physics* (Chalmers, Göteborg, 1982), Vol. 1, p. 96.
[4]C. R. Menyuk, H. H. Chen, and Y. C. Lee, Phys. Rev. A **27**, 1597 (1983).
[5]C. R. Menyuk, H. H. Chen, and Y. C. Lee, J. Math. Phys. **24**, 1073 (1983).
[6]S. Wojciechowski, Z. Jiang, and R. K. Bullough, Phys. Lett. A **117**, 399 (1986).
[7]F. Verheest, Proc. Egypt. Math. Phys. Soc. (to be published).
[8]F. Verheest, J. Math. Phys. **28**, 2863 (1987).
[9]F. Verheest, J. Phys. A: Math. Gen. **21**, L545 (1988).
[10]H. Yoshida, Celestial Mech. **31**, 363 (1983).
[11]H. Yoshida, Celestial Mech. **31**, 381 (1983).
[12]D. Roekaerts and F. Schwarz, J. Phys. A: Math. Gen. **20**, L127 (1987).
[13]F. Verheest, J. Phys. A: Math. Gen. **20**, 103 (1987).
[14]L. A. Pars, *A Treatise on Analytical Dynamics* (Heinemann, London, 1965), pp. 517–520.

Resonant interval action transfer between coupled harmonic oscillators

R. O. Dendy
Culham Laboratory (Euratom/UKAEA Fusion Association), Abingdon, Oxfordshire, OX14 3DB, United Kingdom

(Received 14 January 1988; accepted for publication 8 June 1988)

A formula is obtained for the action transferred between two weakly coupled harmonic oscillators, where the time-dependent frequency of one oscillator passes through resonance with the fixed frequency of the other. The analysis employs an action-angle Hamiltonian that is the classical analog of an earlier wave-mechanical description of linear mode conversion. It represents a link between generalized Hamiltonian approaches to linear mode conversion, and those based on wave dispersion relations.

I. INTRODUCTION

There is at present considerable interest in the process of linear mode conversion.[1–15] In both plasma physics and geophysical fluid dynamics, the following question arises. What is the flow of energy between the linear normal modes, when at some point x_c the inhomogeneity of the system causes the frequencies of two initially distinct normal modes with wave number k_c to become degenerate, before again diverging? This question has previously been answered from two points of view, which the present paper is intended to link. Grimshaw and Allen,[3] who were concerned with applications in geophysical fluid dynamics, considered a Hamiltonian system. Their multiple-time-scale analysis of the canonical evolution equations generated equations that were solved using parabolic cylinder functions. Cairns and Lashmore-Davies,[6,8] who were concerned with applications in plasma physics, considered the wave dispersion relations that arise from the dielectric properties of magnetized plasmas. They carried out a generalized analysis of the local dispersion relation in which wave numbers map to the operator $-i\,d/dx$. The resulting coupled first-order differential equations gave rise to a second-order system which was solved in terms of Weber's equation,[16] again using parabolic cylinder functions and reproducing where appropriate the results of earlier fourth-order calculations.[2,5] Recently, a third line of approach to linear mode conversion has been developed, using Hermitian operators.[13–15] In particular, in Ref. 15, a wave-mechanical interpretation of the class of dispersion relation considered by Cairns and Lashmore-Davies[6,8] was introduced. Using standard techniques of first-order perturbation theory, differential equations for the wave amplitudes were obtained that led in turn to Weber's equation. Once again, the standard result for energy transfer was obtained, using the asymptotic properties of parabolic cylinder functions. The success of wave-mechanical techniques represents the first of two steps that link the two basic approaches to linear mode conversion, namely, those based on wave dispersion relations and those based on Hamiltonian systems. The present paper is intended to provide the second step. We construct the classical Hamiltonian analog of the Hermitian system considered in Ref. 15, which is itself an extension of the dispersion relation approach. At the same time, the action-angle Hamiltonian that we construct and its method of solution necessarily have many features in common with the generalized Hamiltonian approach of Grimshaw and Allen.[3] In discussing these features further, we hope to show how the Hamiltonian and wave dispersion relation approaches to linear mode conversion reflect the same physical phenomenon.

II. ACTION TRANSFER IN THE HAMILTONIAN SYSTEM

Let us now construct the classical mechanical analog of the wave-mechanical system considered in Ref. 15. It consists of two weakly coupled one-dimensional harmonic oscillators: the fundamental frequency ω_1 of the first oscillator remains constant; that of the second oscillator is initially less than ω_1, but increases slowly with time. At a particular time t_c, we have

$$\omega_1 = \omega_2(t_c), \tag{1}$$

so that for a finite interval of time the fundamental frequencies are degenerate or close to degenerate. We shall calculate the action transferred between the oscillators during the resonant interval at $t \simeq t_c$ when the fundamental frequencies remain nearly degenerate. This oscillator system can be represented by the explicitly time-dependent Hamiltonian

$$H(p_1,q_1,p_2,q_2,t) = \frac{p_1^2}{2} + \frac{\omega_1^2 q_1^2}{2} + \frac{p_2^2}{2} + \frac{\omega_2^2(t) q_2^2}{2} - \eta q_1 q_2. \tag{2}$$

It is convenient to carry out a canonical transformation of H into action and angle variables using the generating function[17]

$$F(q_1,\theta_1,q_2,\theta_2,t) = \frac{\omega_1 q_1^2 \cot\theta_1}{2} + \frac{\omega_2(t) q_2^2 \cot\theta_2}{2}. \tag{3}$$

This gives

$$p_i = \frac{\partial F}{\partial q_i} = \omega_i q_i \cot\theta_i, \quad J_i = -\frac{\partial F}{\partial \theta_i} = \frac{\omega_i q_i^2}{2\sin^2\theta_i}, \tag{4}$$

and the transformed Hamiltonian $K = H + \partial F/\partial t$ becomes

$$K = \omega_1 J_1 + \omega_2(t) J_2 + \frac{\dot\omega_2}{2\omega_2} J_2 \sin 2\theta_2 - 2\eta\left(\frac{J_1 J_2}{\omega_1\omega_2}\right)^{1/2} \sin\theta_1 \sin\theta_2. \tag{5}$$

Here $\dot\omega_2$ denotes $d\omega_2/dt$. We shall assume that $\dot\omega_2$ is small in the sense that $\dot\omega_2/\omega_2 \ll \omega_1,\omega_2$, and that the coupling between the oscillators is weak in the sense that $\eta \ll \omega_1^2,\omega_2^2$. The canonical evolution equations for the actions $\dot J_i = -\partial K/\partial\theta_i$ can be written in the form

$$\frac{d}{dt}J_1^{1/2} = -\frac{\eta J_2^{1/2}}{2(\omega_1\omega_2)^{1/2}} \times[\sin(\theta_1-\theta_2)-\sin(\theta_1+\theta_2)], \tag{6}$$

$$\frac{d}{dt}J_2^{1/2} = -\frac{\dot\omega_2}{2\omega_2}J_2^{1/2}\cos 2\theta_2 + \frac{\eta J_1^{1/2}}{2(\omega_1\omega_2)^{1/2}} \times[\sin(\theta_1-\theta_2)+\sin(\theta_1+\theta_2)], \tag{7}$$

using standard trigonometric identities. The canonical evolution equations for the angles $\dot\theta_i = \partial K/\partial J_i$ become

$$\frac{d\theta_1}{dt} = \omega_1 - \frac{\eta}{(\omega_1\omega_2)^{1/2}}\left(\frac{J_2}{J_1}\right)^{1/2}\sin\theta_1\sin\theta_2, \tag{8}$$

$$\frac{d\theta_2}{dt} = \omega_2(t) + \frac{\dot\omega_2}{2\omega_2}\sin 2\theta_2 - \frac{\eta}{(\omega_1\omega_2)^{1/2}}\left(\frac{J_1}{J_2}\right)^{1/2}\sin\theta_1\sin\theta_2. \tag{9}$$

The long-time-scale consequences of the terms involving $\dot\omega_2/\omega_2$ in Eqs. (7) and (9) have been investigated, for the case of a single harmonic oscillator ($\eta = 0$), by Vandervoort.[18] Here we shall concentrate on the two-oscillator resonant interval, defined to be the interval of time when the relative phase $\theta_1 - \theta_2$ varies slowly compared to θ_1 and θ_2 themselves. By Eqs. (8) and (9), we have to leading order

$$\frac{d}{dt}(\theta_1-\theta_2) = \omega_1 - \omega_2(t). \tag{10}$$

Referring to Eq. (1), let us define a new independent variable

$$\tau = t - t_c, \tag{11}$$

and for future convenience we define

$$\mu = [\dot\omega_2/2]_{t=t_c}. \tag{12}$$

Combining Eq. (1) and Eqs. (10)–(12), it follows that to leading order during the resonant interval,

$$\theta_1 - \theta_2 = -\mu\tau^2 - \phi_0, \tag{13}$$

where ϕ_0 is a constant. This is equivalent to the wave-mechanical result given by Eq. (30) of Ref. 15. In contrast to $\theta_1 - \theta_2$, the other angular variables $2\theta_2$ and $\theta_1 + \theta_2$ that appear in Eqs. (6) and (7) oscillate rapidly during the resonant interval. On the time scale of interest, namely, the duration of the resonant interval, we assume that the changes in the actions J_i arising from these rapidly oscillating terms are negligible, since they integrate almost to zero. Let us denote the slowly varying amplitudes of the actions J_i by $\bar J_i$. Then using Eq. (13), Eqs. (6) and (7) give

$$\frac{d}{d\tau}\bar J_1^{1/2} = \frac{\eta}{2\omega_1}\bar J_2^{1/2}\sin(\mu\tau^2+\phi_0), \tag{14}$$

$$\frac{d}{d\tau}\bar J_2^{1/2} = -\frac{\eta}{2\omega_1}\bar J_1^{1/2}\sin(\mu\tau^2+\phi_0). \tag{15}$$

Here we have simplified the coupling coefficients using the fact that $\omega_2 \simeq \omega_1$ during the resonant interval. We note that the total averaged action $\bar J_1 + \bar J_2$ is conserved by Eqs. (14) and (15). This is the discrete system analog of the wave-mechanical result Eq. (22) of Ref. 15. Because we shall differentiate Eqs. (14) and (15) again with respect to τ, it is convenient to consider the complex system of which Eqs. (14) and (15) are the real part:

$$\frac{d}{d\tau}\bar J_1^{1/2} = i\frac{\eta}{2\omega_1}\exp(-i\phi_0)\bar J_2^{1/2}\exp(-i\mu\tau^2), \tag{16}$$

$$\frac{d}{d\tau}\bar J_2^{1/2} = i\frac{\eta}{2\omega_1}\exp(i\phi_0)\bar J_1^{1/2}\exp(i\mu\tau^2). \tag{17}$$

Then, defining complex variables

$$a_1 = \bar J_1^{1/2}\exp[i(\mu\tau^2+\phi_0)/2]$$

and

$$a_2 = \bar J_2^{1/2}\exp[-i(\mu\tau^2+\phi_0)/2],$$

and differentiating Eqs. (16) and (17) again with respect to τ, we obtain two uncoupled second-order differential equations:

$$\frac{d^2a_1}{d\tau^2} + \left[\left(\frac{\eta}{2\omega_1}\right)^2 + \mu^2\tau^2 - i\mu\right]a_1 = 0, \tag{18}$$

$$\frac{d^2a_2}{d\tau^2} + \left[\left(\frac{\eta}{2\omega_1}\right)^2 + \mu^2\tau^2 + i\mu\right]a_2 = 0. \tag{19}$$

Equations (18) and (19) are formally identical to Eqs. (35) and (36) of Ref. 15. A sequence of transformations,[15] some previously noted by Budden,[19] leads from these equations to Weber's equation.[16] The asymptotic properties of the roots of Weber's equation have been employed by Cairns and Lashmore-Davies[6,8] to calculate the energy transfer during linear mode conversion. It follows that the formula that describes the action transfer for coupled harmonic oscillators with a resonant interval can be obtained by relating the parameters arising in Eqs. (18) and (19) to those of Ref. 8. The fraction of the action initially possessed by the first oscillator that is transferred to the second during the resonant interval is

$$\alpha = 1 - \exp(-\pi\eta^2/4\omega_1^2\mu), \tag{20}$$

where μ is defined by Eq. (12). The combination of parameters that occurs in the exponential in Eq. (20) is physically reasonable. It is proportional to the product of two dimensionless quantities, one large and one small: $(\eta/\omega_1^2)^2(\omega_1^2/\mu)$. Here $\eta/\omega_1^2 \ll 1$ is a measure of the strength of the coupling between the oscillators, and $\omega_1^2/\mu \gg 1$ is a measure of the number of rapid oscillation periods for which the oscillators remain in approximate resonance.

Thus far we have employed an action-angle Hamiltonian because it provides the closest classical analog of the wave-mechanical system considered in Ref. 15. Let us now examine the close similarities between this action-angle description and the generalized Hamiltonian approach of Ref. 3. First, we note that the components of Eqs. (6) and (7) that are significant during the resonant interval, namely, those with argument $\theta_1 - \theta_2$, resemble Eq. (4.3) of Ref. 3, where, however, θ_1 and θ_2 are not canonical coordinates. Next, we note that during the resonant interval, the dependence of $\theta_1 - \theta_2$ on the independent variable is quadratic in Eq. (13). In Ref. 3, the corresponding dependence is more complex, but we note from Eq. (4.10) of Ref. 3 that it includes a dominant quadratic term asymptotically. Equations (14) and (15) also differ from Eq. (4.8) of Ref. 3, insofar as the coupling coefficients are themselves functions

of the independent variable. Nevertheless, it is clear from the similarity of Eqs. (18) and (19) to Eq. (4.14) of Ref. 3 that these differences are essentially minor. Both approaches lead to equations that can be solved in terms of parabolic cylinder functions, the eigenfunctions of Weber's equation.

III. CONCLUSIONS

A simple formula has been obtained for the action transferred between two weakly coupled one-dimensional harmonic oscillators, where the time-dependent frequency ω_2 of one oscillator passes through resonance with the fixed frequency ω_1 of the other. During the resonant interval, the coupled canonical evolution equations yield uncoupled second-order equations of a form that has been shown[15] to transform to Weber's equation.[16] The analysis of Weber's equation by Cairns and Lashmore-Davies,[6,8] developed during studies of energy transfer during linear mode conversion in inhomogeneous plasmas, is adapted to give the action transfer between the harmonic oscillators. The present result was obtained using an action-angle Hamiltonian system that is the classical analog of the wave-mechanical system considered in Ref. 15. The latter is itself an extension of the wave dispersion relation approach to linear mode conversion. The theory presented here has, in addition, many points in common with the generalized Hamiltonian approach to linear mode conversion developed by Grimshaw and Allen.[3] It thus represents a link between the two basic classes of description of this phenomenon.

ACKNOWLEDGMENT

The author is grateful to Dr. C. N. Lashmore-Davies for many helpful discussions of this topic.

[1]T. H. Stix, Phys. Rev. Lett. **15**, 878 (1965).
[2]Y. C. Ngan and D. G. Swanson, Phys. Fluids **20**, 1920 (1977).
[3]R. Grimshaw and J. S. Allen, Stud. Appl. Math. **61**, 55 (1979).
[4]V. Fuchs, K. Ko, and A. Bers, Phys. Fluids **24**, 1251 (1981).
[5]D. G. Swanson, Phys. Fluids **24**, 2035 (1981).
[6]R. A. Cairns and C. N. Lashmore-Davies, Phys. Fluids **25**, 1605 (1982).
[7]R. Grimshaw and J. S. Allen, Stud. Appl. Math **67**, 169 (1982).
[8]R. A. Cairns and C. N. Lashmore-Davies, Phys. Fluids **26**, 1268 (1983).
[9]L. Friedland, Phys. Fluids **28**, 3260 (1985).
[10]L. Friedland, Phys. Fluids **29**, 1105 (1986).
[11]R. Grimshaw, Stud. Appl. Math. **74**, 205 (1986).
[12]R. A. Cairns and C. N. Lashmore-Davies, Phys. Fluids **29**, 3639 (1986).
[13]R. G. Spencer, J. Math. Phys. **27**, 380 (1986).
[14]L. Friedland, G. Goldner, and A. N. Kaufman, Phys. Rev. Lett. **58**, 1392 (1987).
[15]R. O. Dendy, Phys. Fluids **31**, 298 (1988).
[16]E. T. Whittaker and G. N. Watson, *Modern Analysis* (Cambridge U.P., Cambridge, 1935), p. 347, 4th ed.
[17]H. Goldstein, *Classical Mechanics* (Addison-Wesley, Reading, MA, 1980), p. 389, 2nd ed.
[18]P. O. Vandervoort, Ann. Phys. (NY) **12**, 436 (1961).
[19]K. G. Budden, *The Propagation of Radio Waves* (Cambridge U.P., Cambridge, 1985), p. 457.

Factorization of the dissipative wave equation and inverse scattering

V. H. Weston
Department of Mathematics, Purdue University, West Lafayette, Indiana 47907

(Received 6 November 1987; accepted for publication 15 June 1988)

A decomposition of the solution of the dissipative wave equation into incoming and outgoing components across a smooth surface in a homogeneous region is presented. (The proof of the decomposition is given only for the plane surface.) This is then applied to the factorization of the dissipative wave equation into incoming and outgoing components in a planar-stratified medium. The Ricatti integral–differential equation for the reflection operator that relates the two components is obtained. It is shown how the zeroth and second moments (with respect to the transverse variable in a planar-stratified medium) of the reflection kernel can be used in the inverse problem to recover the velocity and dissipation coefficient from knowledge of the scattered field.

I. INTRODUCTION

A common technique[1,2] for the formulation of the time-dependent inverse problem, associated with the one-dimensional wave equation as well as related equations,[3] is based upon the concept of wave splitting. In one-dimensional problems this entails decomposing the wave into up- and down-going wave components. Invariant imbedding techniques[4,5] and their variations are used to get an equation for the reflection operator (the operator relating the up-going wave component to the down-going wave component). The equation for the reflection operator contains a quadratic nonlinearity and is commonly known in the terminology of inverse scattering as being a Ricatti-type equation, the reason for this being the fact that in the one-dimensional problem in the frequency domain the equation is an ordinary differential equation with the quadratic term (Ricatti equation). However, the equation for the reflection operator in the time-dependent case takes the form of an integral–differential equation with the quadratic term involving a convolution. The initial condition associated with the kernel of the reflection operator is related to the unknown coefficient of the wave equation. The importance of the whole approach is that the Ricatti equation for the reflection operator's kernel has been used successfully in numerical schemes for solving the inverse problem.[3,6–8]

Recently this approach has been generalized to the wave equation in three dimensions, first to the case where the medium is planar stratified,[9] then for the more general case of nonplanar stratifications.[10] Here the concept of up- and down-going waves is replaced by the concept of outgoing and incoming wave across a smooth surface S. The key to the analysis is the development of an incoming and outgoing wave condition on the surface S, which is expressed in terms of a linear operator relationship between the wave function u and its normal derivative $\partial u/\partial n$ on S. This is used to factor the wave equation into incoming and outgoing components. From this the form of the reflection operator and its Ricatti equation are obtained for the special case of a planar-stratified medium. At present the reflection operator and its equation are being examined for cylindrical geometry, and numerical work is applied to the associated inverse problem.[11]

Here in this paper the wave-splitting process that was developed previously[9,10] for the wave equation in $\mathbf{R}^3$ is generalized to apply to the dissipative wave equation (or telegraph equation)[12]

$$\Box u + b\frac{\partial u}{\partial t} = 0, \tag{1}$$

where

$$\Box = \frac{1}{c^2}\frac{\partial^2}{\partial t^2} - \frac{\partial^2}{\partial x_1^2} - \frac{\partial^2}{\partial x_2^2} - \frac{\partial^2}{\partial x_3^2}, \tag{2}$$

with the coefficient $b \geqslant 0$. The corresponding incoming and outgoing wave condition expressed as a linear operator relationship between u and $\partial u/\partial n$ on a smooth surface S is obtained, for the case where the coefficients b and c are constant (with the proof of the decomposition given only for the plane surface). This is then generalized to the case where the medium is stratified, with emphasis on plane stratification where b and c are functions of x_3 only. The reflection operator and its Ricatti equation are obtained for the special case of a planar-stratified medium. The associated inverse problem for a stratified half-space is treated where reflection data on the surface are used to get the coefficients b,c in the upper, or penetrable, layer of the medium. Since the waves are attenuated due to absorption, it is not practical to get the coefficients b and c deep inside the medium from reflection data on the surface.

Except for the work of Buzdin[13] or Blagoveshchenskii,[14] the time-dependent inverse problem treated here for the dissipative wave equation differs from previous work[6,7,15,16] in which the geometry that is taken is a one-dimensional slab and where reflection data on both sides are employed. A Gelfand–Levitan-type system was developed by Weston[15] and the results generalized by Krueger.[16] More recently Kristensson and Krueger[6,7] have applied the wave-splitting technique to obtain the Ricatti equation formulation for the reflection kernels.

Buzdin's[13] approach to the inverse problem for a plane-stratified medium involves directly taking two moments of the dissipative wave equation (telegraph equation) and reducing it to a system of two partial differential equations in

 0022-2488/88/102205-14$02.50

two independent variables (time t and depth parameter z). Cauchy data (reflection data) at $z=0$ are used to transfer the system into a system of integrodifferential equations, by integrating over a triangle formed by the characteristics. The resulting nonlinear system can be solved up to a certain depth of penetration.

The emphasis in this paper is on the development of the wave splitting into outgoing and incoming waves in a stratified medium, and the determination of the equation that must be satisfied by the associated reflection operator for a plane-stratified medium. This approach has the potential (as ongoing research indicates) to be applied to the more general inverse problem associated with a nonhomogeneous medium. The moment approach (applied to the reflection operator) is employed in the latter part of this paper to demonstrate the existence of the solution to the inverse problem.

II. INCOMING AND OUTGOING WAVE CONDITION FOR A HOMOGENEOUS MEDIUM

Let S be a smooth (C^2) surface enclosing an open region D_i in $\mathbf{R}^3$ that may or may not be bounded. Let the corresponding open exterior region be denoted by D_e. In this section, conditions will be derived to indicate whether or not a solution u of the dissipative wave equation (1) represents an outgoing or incoming wave across S. This condition will take the form of a linear relation between u and $\partial u/\partial n$ on S. From the physical standpoint an outgoing wave across S is produced by sources in D_i and is represented mathematically by the solution of exterior initial-value and boundary-value problems, with the initial values of u, $\partial u/\partial t$ being zero in $\overline{D_e}$. A corresponding formulation holds for incoming waves.

To develop the outgoing wave condition on S, the Kirchhoff formula for the exterior problem

$$\begin{aligned} &\Box u + bu_t = 0, \quad t>0, \quad x\in D_e, \\ &u = u_t = 0, \quad t=0, \quad x\in \overline{D_e}, \end{aligned} \tag{3}$$

where b and c are constants and $b\geqslant 0$, is needed. If $\mathscr{E}(x,t)$ is the fundamental solution of the system

$$\Box \mathscr{E} + b\mathscr{E}_t = \delta(x)\delta(t), \quad \mathscr{E}=0, \quad t<0, \tag{4}$$

then it follows in the usual manner that for $t>0$, $x_0\in D_e$,

$$u(x_0,t) = -\int_0^t \int_S \left\{ \mathscr{E}(x-x_0,t-s)\frac{\partial u}{\partial n} - u\frac{\partial}{\partial n}\mathscr{E}(x-x_0,t-s)\right\} d\sigma_x\, ds, \tag{5}$$

where $d\sigma_x$ is an element of surface area on S and $\partial/\partial n$ is the normal derivative on S taken in the direction of the normal pointing into region D_e. The precise form of $\mathscr{E}(x,t)$ is given by[12]

$$\mathscr{E}(x,t) = \left\{ \frac{cH(ct)\delta(c^2t^2-|x|^2)}{2\pi} + \frac{bc^2H(ct-|x|)I_1(bc\sqrt{c^2t^2-|x|^2}/2)}{8\pi\sqrt{c^2t^2-|x|^2}}\right\} e^{-bc^2t/2}, \tag{6}$$

where I_1 is the modified Bessel function of order 1, and $H(t)$ is the Heaviside step function.

With the variable of integration x replaced by y, and expression (6) inserted for $\mathscr{E}(x-x_0,t-s)$, relation (5) reduces to the explicit form

$$u(x_0,t) = -\frac{1}{4\pi}\int_S \left\{ \mathbf{E}u_n[x_0,y,t] - \frac{\partial r}{\partial n_y}\frac{\partial}{\partial r}(\mathbf{E}u[x_0,y,t])\right\} H\left(t-\frac{r}{c}\right) d\sigma_y, \quad \text{where } r=|x_0-y|, \tag{7}$$

and the operator $\mathbf{E}$ acting on a function $f(y,t)$, $y\in S$, is defined as follows:

$$\mathbf{E}f[x_0,y,t] = \frac{1}{r}f\left(y,t-\frac{r}{c}\right)\exp\left(-\frac{bcr}{2}\right) + \frac{b^2c^3}{4}\int_0^{t-r/c}\exp\left[\frac{bc^2(s-t)}{2}\right]\left\{\frac{I_1(\zeta)}{\zeta}\right\} \times f(y,s)\, ds, \tag{8}$$

where $\zeta = bc^2[(t-s)^2-(r/c)^2]^{1/2}/2$.

The outgoing wave condition is obtained by letting $x_0\in D_e$ approach x on S. In taking the limit one must note that the second term in the integral on the right-hand side of Eq. (7) has the property of a double-layer potential,[17] and hence has a jump discontinuity across S. Taking this into account and using the same approach that was employed for the wave equation,[10] one subsequently obtains the result

$$(\mathbf{I}+\mathbf{M})u + \mathbf{K}u_n = 0, \quad x\in S, \tag{9}$$

which is the sought-for outgoing wave condition on S. Here $\mathbf{I}$ is the identity operator, and the operators $\mathbf{M}$ and $\mathbf{K}$ are defined as follows:

$$\mathbf{K}w[x,t] = \frac{1}{2\pi}\int_S (\mathbf{E}w[x,y,t])H\left(t-\frac{r}{c}\right)d\sigma_y, \tag{10}$$

$$\mathbf{M}w[x,t] = -\frac{1}{2\pi}\int_S \frac{\partial r}{\partial n_y}\frac{\partial}{\partial r}(\mathbf{E}w[x,y,t])H\left(t-\frac{r}{c}\right)d\sigma_y, \tag{11}$$

for $x\in S$, $t\in\mathbf{R}$. The Heaviside step function is introduced in these operators to take care of the initial condition that $u=u_t=0$ for $t\leqslant 0$, $x\in S$.

For the special case where there is no dissipation, $b=0$, $\mathbf{K}$ reduces to the operator

$$\mathbf{K}w = \frac{1}{2\pi}\int_S \frac{w(y,t-r/c)}{r}H\left(t-\frac{r}{c}\right)d\sigma_y,$$

given in the previous paper.[10]

The corresponding incoming wave condition is obtained by treating the interior initial-value problem in a similar manner. The resulting incoming wave condition is given by

$$(\mathbf{I}-\mathbf{M})u - \mathbf{K}u_n = 0, \quad x\in S. \tag{12}$$

For the special case when S is a plane surface, $\partial r/\partial n = 0$ for $x,y\in S$; hence the operator $\mathbf{M}$ is identically zero. The outgoing and incoming wave conditions reduce to

$$u \pm \mathbf{K}u_n = 0, \quad x\in S \text{ (a plane surface)}. \tag{13}$$

Using the two lemmas that are proved in Appendix C, it will be shown next that the operators **K** and **M** are compact. The time interval T may have to be restricted depending upon the complexity of the surface S, due to condition (C2) of Appendix C. Hölder continuity in the time variable will be needed (see Appendix C for appropriate definitions).

Theorem 1: If S is a Lyapunov closed surface,[17] then (i) **K** is a compact operator mapping $C(S)\times C^{(0,\lambda)}[0,T]$ into $C(S)\times C_0[0,T]$, and (ii) **M** is a compact operator mapping $C(S)\times C^{(1,\lambda)}[0,T]$ into $C(S)\times C_0[0,T]$, where λ, $0<\lambda\leqslant 1$, is the Hölder index, and $C_0[0,T]$ is the space of continuous functions of t on $[0,T]$ that vanish at $t=0$.

Proof: From Eq. (10) it is seen that the operator **K** can be decomposed in the following manner:

$$\mathbf{K}u = \mathbf{G}u + \mathbf{H}u,$$

where **G** corresponds to the operator in Lemma 1 of Appendix C with kernel

$$g(x,y) = \frac{1}{2\pi r}\exp\left(-\frac{bcr}{2}\right), \quad r=|x-y|,$$

and **H** corresponds to the operator in Lemma 2 of Appendix C with kernel

$$h(x,y,t) = \frac{1}{8\pi} b^2c^3 \exp\left[-\frac{bc^2t}{2}\right]\frac{I_1(\zeta)}{\zeta},$$

with $\zeta = bc^2(t^2 - r^2/c^2)^{1/2}/2$. Because these kernels satisfy the conditions in the lemmas, the compactness of **K** follows from the compactness of **G** and **H**.

From Eq. (11) it follows that the operator **M** can be decomposed as follows:

$$\mathbf{M}u = \mathbf{G}_1 u + \mathbf{H}_1 u + \mathbf{G}_2 u_t,$$

with $\mathbf{G}_1, \mathbf{G}_2$ corresponding to the operator in Lemma 1 with respective kernels

$$g_1(x,y) = \frac{1}{2\pi}\frac{\partial r}{\partial n_y}\left[\frac{1}{r^2}+\frac{bc}{2r}+\frac{b^2c^2}{8}\right]\exp\left(-\frac{bcr}{2}\right),$$

$$g_2(x,y) = \frac{1}{2\pi rc}\frac{\partial r}{\partial n_y}\exp\left(-\frac{bcr}{2}\right),$$

and $\mathbf{H}_1$ corresponding to the operator in Lemma 2 with kernel

$$h_1(x,y) = \frac{b^4c^5 r}{32\pi}\frac{\partial r}{\partial n_y}\left\{\frac{I_2(\zeta)}{\zeta}\right\}\exp\left(-\frac{bc^2t}{2}\right),$$

with $\zeta = bc^2(t^2 - r^2/c^2)^{1/2}/2$ and I_2 the modified Bessel function of order 2. Since S is a Lyapunov surface (with Hölder smoothness in unit normal, $|n_z - n_y| \leqslant \kappa|x-y|^\alpha$, $0<\alpha\leqslant 1$), it follows from Vladimirov[17] that

$$\left|\frac{\partial r}{\partial n_y}\right| \leqslant a|x-y|^\alpha;$$

hence g_1, g_2, h_1 satisfy the requirements of the lemmas. Compactness of **M** thus follows.

III. WAVE SPLITTING IN A HOMOGENEOUS MEDIUM

With the incoming and outgoing wave conditions now established, the splitting of the solution of the dissipative equation into incoming and outgoing waves can be obtained. However, before considering this, some results on the existence and structure of the inverse operator $\mathbf{K}^{-1}$ have to be established.

Theorem 2: The null space of **K** is empty, and $\mathbf{K}^{-1}$ exists.

Proof: The proof depends upon the fact that the interior and exterior Dirichlet initial-value problem associated with Eq. (1) has a unique solution. This follows from the energy integral[18] associated with Eq. (1),

$$\iint\left\{\frac{1}{2}\frac{\partial}{\partial t}\left[\frac{1}{c^2}u_t^2 + |\nabla u|^2\right] + b(u_t)^2 - \nabla\cdot(u_t\nabla u)\right\} \times dx\,dt = 0,$$

which indicates the importance of b being non-negative.

Let $v(x,t)\in C(S)\times C[0,T]$ be a solution of $\mathbf{K}v = 0$, $x\in S$. Then set

$$u(x,t) = \frac{1}{2\pi}\int_S \mathbf{E}v[x,y,t\,]H\left(t-\frac{r}{c}\right)d\sigma_y, \tag{14}$$

where $x\in\mathbb{R}^3$. It thus follows from the Kirchhoff formula (7) (with $u_n = -2v$, $u=0$, on S) that u is a solution of Eq. (1) in D_e and D_i. Because of the presence of the Heaviside step function, it follows that $u = u_t = 0$ for $t=0$ and $x\in D_e$ or $x\in D_i$. Furthermore, since $\mathbf{K}v = 0$, it follows that $u=0$, $x\in S$. Hence from uniqueness of the solution of the Dirichlet initial-value problem it follows that $u(x,t)\equiv 0$, for $x\in D_e$ or D_i, $t\geqslant 0$. From the jump condition associated with the single-layer potential term[10,17] in expression (14),

$$\left[\frac{\partial u}{\partial n}\right]_-^+ = -2v(x,t), \quad x\in S,$$

it follows that $v\equiv 0$, $t\geqslant 0$. Thus the null space of **K** is empty, and $\mathbf{K}^{-1}$ exists.

The precise form of $\mathbf{K}^{-1}$ will now be established for the case of a plane surface. The following will be defined.

Definition: If S is the plane surface $x_3 = \text{const}$, then the transverse d'Alembertian $\Box_T$ associated with this surface is given by

$$\Box_T \equiv \frac{1}{c^2}\frac{\partial^2}{\partial t^2} - \frac{\partial^2}{\partial x_1^2} - \frac{\partial^2}{\partial x_2^2}. \tag{15}$$

Theorem 3: If S is a plane surface, then

$$\mathbf{K}^{-1} = \left(\Box_T + b\frac{\partial}{\partial t}\right)\mathbf{K}. \tag{16}$$

Proof: Let S be the surface $x_3 = 0$, and D_e the region $x_3 > 0$. Let u be a solution of Eq. (1) in the open region $x_3 > -\delta$ (where $\delta > 0$) containing the surface S, and let it satisfy the initial condition $u = u_t = 0$ there at $t=0$. Use the fact that if u is a solution of Eq. (1) in a homogeneous medium, then $\partial u/\partial x_3$ is also a solution, and apply the Kirchhoff formula (7) with u replaced by $\partial u/\partial x_3$. As before let $x_0\in D_e \to x\in S$ and, using the result $\partial u/\partial n = \partial u/\partial x_3$ on S, obtain (for the outgoing wave)

$$\frac{\partial u}{\partial x_3} + \mathbf{K}\frac{\partial^2 u}{\partial x_3^2} = 0, \quad x\in S,$$

which corresponds to Eq. (13) (upper sign) with u replaced by $\partial u/\partial x_3$. Since u satisfies Eq. (1) for $x\in S$ and $u = u_t = 0$ for $t=0$, it follows that

$$\frac{\partial u}{\partial n} + \mathbf{K}\left(\Box_T + b\frac{\partial}{\partial t}\right) u = 0,$$

or

$$\frac{\partial u}{\partial n} + \left(\Box_T + b\frac{\partial}{\partial t}\right) \mathbf{K}u = 0, \quad x \in S.$$

Comparing this last equation with Eq. (13) for the outgoing wave (upper sign), it is seen that

$$\mathbf{K}^{-1} = \left(\Box_T + b\frac{\partial}{\partial t}\right)\mathbf{K} = \mathbf{K}\left(\Box_T + b\frac{\partial}{\partial t}\right). \tag{16'}$$

■

The problem of splitting a solution of the dissipative wave equation into two wave components crossing the surface S can now be considered.

From the identity

$$u = \tfrac{1}{2}[u - \mathbf{M}u - \mathbf{K}u_n] + \tfrac{1}{2}[u + \mathbf{M}u + \mathbf{K}u_n],$$

it is seen that u can be decomposed as follows:

$$u = u^+ + u^-, \tag{17}$$

where

$$u^\pm = \tfrac{1}{2}[(\mathbf{I} \mp \mathbf{M})u \mp \mathbf{K}u_n]. \tag{18a}$$

For the case of a plane surface (since $\mathbf{M} \equiv 0$) the two components are given by

$$u^\pm = \tfrac{1}{2}[u \mp \mathbf{K}u_n], \quad S \text{ a plane.} \tag{18b}$$

For the latter case of a plane surface (say $x_3 = 0$), if u satisfies the dissipative wave equation in a region in the neighborhood of $x_3 = 0$, then since $\partial/\partial n = \partial/\partial x_3$, and $\partial/\partial x_3$ commutes with the operator $\mathbf{K}$ (its kernel does not contain x_3), it follows that

$$\begin{aligned}(u^\pm \pm \mathbf{K}u_n^\pm) &= \left(\mathbf{I} \pm \mathbf{K}\frac{\partial}{\partial x_3}\right) u^\pm \\ &= \frac{1}{2}\left(\mathbf{I} \pm \mathbf{K}\frac{\partial}{\partial x_3}\right)\left(\mathbf{I} \mp \mathbf{K}\frac{\partial}{\partial x_3}\right) u \\ &= \frac{1}{2}\left[\mathbf{I} - \mathbf{K}^2\frac{\partial^2}{\partial x_3^2}\right] u \\ &= \frac{1}{2}\left[\mathbf{I} - \mathbf{K}^2\left(\Box_T + b\frac{\partial}{\partial t}\right)\right] u = 0,\end{aligned}$$

the last result following from Eq. (16'). This implies that u^+ and u^- satisfy, respectively, the conditions for an outgoing and incoming wave on a plane surface. Hence we have the following theorem.

Theorem 4: If u satisfies the dissipative wave equation (b and c constant) in an open region D containing a plane surface S, $u = u_t = 0$ for $t \leqslant 0$, then u can be split up into two components u^+ and u^- given by Eq. (18b), which satisfy the outgoing and incoming wave condition across S.

What about the case of a general smooth surface S? Do u^+ and u^- satisfy the outgoing and incoming wave conditions on S? For the case of the nondissipative wave equation where $b = 0$, the answer is yes. The proof, which is very long and tricky, is given in the previous paper.[10] Since the major part of the analysis in the remainder of this paper pertains to the planar case, the proof of the corresponding theorem on wave splitting on a smooth surface S for the dissipative wave equation will be left to later work. However, for the remainder of this section and in the initial part of the next section we will work with the decomposition given by Eq. (18a) for a general surface, before specializing to a plane surface.

It is convenient to express the decomposition in vector form as

$$\begin{bmatrix} u^+ \\ u^- \end{bmatrix} = T\begin{bmatrix} u \\ u_n \end{bmatrix}, \tag{19}$$

where the matrix operator T is given by

$$T = \frac{1}{2}\begin{bmatrix} (\mathbf{I} - \mathbf{M}) & -\mathbf{K} \\ (\mathbf{I} + \mathbf{M}) & \mathbf{K} \end{bmatrix}. \tag{20}$$

The inverse of T is given by

$$T^{-1} = \begin{bmatrix} \mathbf{I} & \mathbf{I} \\ -\mathbf{K}^{-1}(\mathbf{I} + \mathbf{M}) & \mathbf{K}^{-1}(\mathbf{I} - \mathbf{M}) \end{bmatrix}. \tag{21}$$

IV. FACTORIZATION OF THE DISSIPATIVE WAVE EQUATION IN A PLANAR-STRATIFIED MEDIUM

The splitting developed in the previous section (and proved only for the planar case) can be used to factor the dissipative wave equation in a stratified medium in a similar manner as was done for the nondissipative case in paper.[10] The stratified medium is composed of a one-parameter family of smooth (nonintersecting) nested surfaces. With α being the parameter, the surfaces are characterized by $\alpha = \text{const}$, and for such a stratified medium the variables in Eq. (1) representing material properties will be a function of α only, i.e., $c = c(\alpha)$ and $b = b(\alpha)$. In generalizing the wave splitting to a stratified medium, the same operators $\mathbf{K}$ and $\mathbf{M}$ developed in the previous section for a homogeneous medium will be used with the surface S given by $\alpha = \text{const}$. However, these operators will be modified by replacing the material constants c and b by functions of α, namely, $c(\alpha)$ and $b(\alpha)$, respectively. Any identities involving these operators acting as a mapping from the surface S ($\alpha = \text{const}$) to the same surface will still hold. This includes the inverse relation (16) for a plane-stratified medium. With these modifications to the operators $\mathbf{K}$ and $\mathbf{M}$, the wave splitting for a stratified medium will take the same form as for the homogeneous medium, i.e., the appropriate splittings are still given by the system (20).

In order to apply the wave splitting to the factorization of Eq. (1) for a stratified medium, additional identities or relations involving the operators $\mathbf{M}$,$\mathbf{K}$ have to be developed. As in the previous paper[10] the normal derivatives $\partial\mathbf{K}/\partial n$, $\partial\mathbf{M}/\partial n$ of the operators have to be defined and their properties established.

However, instead of analyzing the general stratified medium, the remainder of this paper will concentrate on the special case of a plane-stratified medium with the purpose of establishing the usefulness of such factorization in the inverse problem.

Great simplification results for the plane-stratified medium, since the operator $\mathbf{M} \equiv 0$, and the operator $\partial\mathbf{K}/\partial n$ is easy to define and obtain.

Let the plane-stratified surfaces be given as $x_3 = \text{const}$; hence $c = c(x_3)$, $b = b(x_3)$. As in Refs. 1, 2, 9, and 10, the

factorization of the dissipative wave equation is achieved by combining the identity

$$\frac{\partial u}{\partial x_3} = \frac{\partial u}{\partial n}$$

together with the dispersive wave equation written in the form

$$\frac{\partial^2 u}{\partial x_3^2} = \Box_T u + b\frac{\partial u}{\partial t}$$

to obtain the vector formulation

$$\frac{\partial}{\partial x_3}\begin{bmatrix} u \\ u_n \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ \left(\Box_T + b\dfrac{\partial}{\partial t}\right) & 0 \end{bmatrix}\begin{bmatrix} u \\ u_n \end{bmatrix}, \tag{22}$$

where $\Box_T$ is given by Eq. (15).

Substitute the relation into Eq. (22)

$$\begin{bmatrix} u \\ u_n \end{bmatrix} = T^{-1}\begin{bmatrix} u^+ \\ u^- \end{bmatrix},$$

and premultiply the resulting system by the matrix operator T to give

$$\frac{\partial}{\partial x_3}\begin{bmatrix} u^+ \\ u^- \end{bmatrix} = W\begin{bmatrix} u^+ \\ u^- \end{bmatrix}, \tag{23}$$

where

$$W = T\begin{bmatrix} 0 & 1 \\ \left(\Box_T + b\dfrac{\partial}{\partial t}\right) & 0 \end{bmatrix}T^{-1} - T\frac{\partial T^{-1}}{\partial x_3}.$$

Using the relation

$$T\frac{\partial T^{-1}}{\partial x_3} = -\frac{\partial T}{\partial x_3}T^{-1}$$

and

$$\frac{\partial T}{\partial x_3} = \frac{1}{2}\begin{bmatrix} 0 & -1 \\ 0 & 1 \end{bmatrix}\frac{\partial \mathbf{K}}{\partial x_3},$$

together with expression (20) for T (with $\mathbf{M}\equiv 0$) as well as relation (16) for the inverse operator $\mathbf{K}^{-1}$, expression (23) can be reduced to the following:

$$\frac{\partial}{\partial x_3}\begin{bmatrix} u^+ \\ u^- \end{bmatrix} = \begin{bmatrix} -\mathbf{K}^{-1} & 0 \\ 0 & \mathbf{K}^{-1} \end{bmatrix}\begin{bmatrix} u^+ \\ u^- \end{bmatrix} + \frac{1}{2}\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}\mathbf{K}'\mathbf{K}^{-1}\begin{bmatrix} u^+ \\ u^- \end{bmatrix}, \tag{24}$$

where

$$\mathbf{K}' = \frac{\partial \mathbf{K}}{\partial x_3}. \tag{25}$$

This is the sought-for factorization of the dissipative wave equation.

V. REFLECTION OPERATOR FOR A PLANE-STRATIFIED MEDIUM

In a homogeneous region exterior to a scattering body (characterized in this paper by a stratified medium) u^+ would correspond to the scattered wave produced by the wave u^-, which is incident upon the scattering body. A linear relationship between the two waves exists in the form $u^+ = \mathbf{R}u^-$. In this section the equation that must be satisfied by the reflection (or scattering) operator $\mathbf{R}$ will be sought for the case of a plane-stratified medium. This equation plays a key role in both the direct and inverse scattering problems.

The formal equation satisfied by the operator $\mathbf{R}$ is obtained in the usual manner[2,9,10] from the factorized system (24). First, u^+ is replaced by $\mathbf{R}u^-$ in this system. Then, since the operator $\mathbf{R}$ contains x_3 as a parameter, the term $\partial(\mathbf{R}u^-)/\partial x_3$ appearing on the left-hand side of the first equation of the system is replaced by $(\partial \mathbf{R}/\partial x_3)u^- + \mathbf{R}\,\partial u^-/\partial x_3$. Finally, the second equation of system (24) is used to eliminate $\partial u^-/\partial x_3$. The resulting system is a single equation involving u^- only. It takes the form

$$\frac{\partial \mathbf{R}}{\partial x_3}u^- + (\mathbf{R}\mathbf{K}^{-1} + \mathbf{K}^{-1}\mathbf{R})u^- + \tfrac{1}{2}(\mathbf{R}\mathbf{K}'\mathbf{K}^{-1} - \mathbf{K}'\mathbf{K}^{-1}\mathbf{R})u^- = \tfrac{1}{2}\mathbf{R}\mathbf{K}'\mathbf{K}^{-1}\mathbf{R}u^- - \tfrac{1}{2}\mathbf{K}'\mathbf{K}^{-1}u^-, \tag{26}$$

which is the Ricatti integral–differential equation.

Noting that u^- is an arbitrary function vanishing at $t = 0$, the formal equation for the reflection operator [just Eq. (25) with the term u^- omitted] is easily obtained. However, for a complete analysis one needs the precise form of the operator $\mathbf{R}$. For one-dimensional cases, it can be shown from studying the initial-value problem that the reflection operator is the form of a convolution in the time variable with the spatial variable remaining a parameter. In the previous paper[9] for the nondissipative wave equation in a plane-stratified medium, it takes the form of an operator acting on a convolution (in time and transverse spatial variables), with the variable that changes in the direction of the stratifications remaining a parameter.

The precise form of $\mathbf{R}$ will not be given here. This will be done elsewhere. As will be shown in the remaining sections (following the work of Buzdin[13]) one does not need to know the precise form of the reflection operator to solve the inverse problem for a planar-stratified medium; one only needs to know the precise form of two of its moments (with regard to the transverse variables). However, the form of the reflection operator relating the outgoing wave to the incoming wave across a plane surface will be needed in the generalization of the wave-splitting concept from the case of a planar-stratified medium to that of a general nonhomogeneous medium. In fact, ongoing research has shown that for the case $b\equiv 0$ (no dissipation), Eq. (26) may be generalized to hold for a nonhomogeneous medium by suitably modifying the operator $\mathbf{K}$. Hence Eq. (26) will be important for future research in inverse scattering in a nonhomogeneous medium.

VI. EQUATIONS FOR THE MOMENTS OF THE REFLECTION OPERATOR

The moments of interest will be the zeroth and second moments. The first moment will not be used, since it will yield the same equation as the zeroth moment.

The zeroth moment of $u(x,t)$ will be defined as follows:

$$u_0(x_3,t) = \int\int_{\mathbf{R}^2} u(x,t)dx_1\,dx_2. \tag{27}$$

The corresponding zeroth moments of the operator $\mathbf{K}^{-1}$ and $\mathbf{K}'\mathbf{K}^{-1}$ (obtained in Appendix A) are given by

$$(\mathbf{K}^{-1}u)_0 = \left(\frac{\partial}{c\,\partial t} + \frac{bc}{2}\right) u_0(x_3,t) + \tilde{k}_0(x_3,t)*u_0(x_3,t), \tag{28}$$

$$(\mathbf{K}'\mathbf{K}^{-1}u)_0 = +\frac{1}{c}\frac{\partial c}{\partial x_3} u_0(x_3,t) + h_0(x_3,t)*u_0(x_3,t), \tag{29}$$

where the star indicates convolution in the time variable (with x_3 remaining as a parameter), and

$$\tilde{k}_0(x_3,t) = -(b^2c^3/4)e^{-\eta}I_1(\eta)/\eta, \tag{30}$$

$$h_0(x_3,t) = -\frac{1}{2}\frac{\partial(bc^2)}{\partial x_3} e^{-2\eta}, \tag{31}$$

with

$$\eta = bc^2t/2. \tag{32}$$

Here $I_1(\eta)$ is the modified Bessel function of order 1.

The equation for the zeroth moment of the reflection operator can be obtained in two ways. However, to get its form, the approach that must be taken is to take the zeroth moment of system (24). Using the results of Eqs. (28) and (29) for the zeroth moments of the operators involved in the system, it follows that the resulting system is a one-dimensional system involving u_0^+ and u_0^- only. Hence the zeroth moment of the reflection operator takes the form of a convolution in the time variable

$$(\mathbf{R}u^-)_0 = \mathbf{R}_0u_0^- = R_0(x_3,t)*u_0(x_3,t). \tag{33}$$

The equation for the zeroth moment can be obtained directly by taking the zeroth moment of Eq. (26) and employing operator relations of the type

$$(\mathbf{R}\mathbf{K}'\mathbf{K}^{-1}u)_0 = R_0*\left(h_0*u_0 + \frac{1}{c}\frac{\partial c}{\partial x_3}u_0\right),$$

$$(\mathbf{R}\mathbf{K}^{-1}u)_0 = R_0*\left(\frac{1}{c}\frac{\partial u_0}{\partial t} + \frac{bc}{2}u_0 + \tilde{k}_0*u_0\right),$$

$$(\mathbf{K}^{-1}\mathbf{R}u)_0 = \left(\frac{1}{c}\frac{\partial}{\partial t} + \frac{bc}{2}\right)R_0*u_0 + \tilde{k}_0*R_0*u_0.$$

In addition, if interchange of order of integration (in the convolution) and differentiation given by

$$\frac{\partial}{\partial t}(R_0*u_0) = \frac{\partial R_0}{\partial t}*u_0 + R_0(x_3,0)u_0(x_3,t),$$

$$R_0*\frac{\partial u_0}{\partial t} = \frac{\partial R_0}{\partial t}*u_0 + R_0(x_3,0)u_0(x_3,t)$$

is employed, the resulting system reduces to one of the form

$$\Gamma*u_0^- + \left[\frac{2}{c}R_0(x_3,0) + \frac{1}{2c}\frac{\partial c}{\partial x_3}\right]u_0^-(x_3,t) = 0, \tag{34}$$

where Γ is given by the left-hand side of Eq. (35). Since u_0^- is an arbitrary function vanishing at $t=0$, the differential–integral equation can be obtained for R_0 by setting $\Gamma = 0$, giving

$$\left(\frac{\partial}{\partial x_3} + \frac{2}{c}\frac{\partial}{\partial t}\right)R_0 + bcR_0 + 2\tilde{k}_0*R_0 - \frac{1}{2c}\frac{\partial c}{\partial x_3}R_0*R_0 - \frac{1}{2}R_0*h_0*R_0 + \frac{1}{2}h_0 = 0, \tag{35}$$

and the initial condition for R_0 can be obtained by setting the coefficient of $u_0^-(x_3,t)$ in Eq. (34) to zero, giving

$$R_0(x_3,0) = -\frac{1}{4}\frac{\partial c}{\partial x_3}. \tag{36}$$

The system made up of the two equations (35) and (36) is the sought-for system for determining R_0. It can be used in the direct scattering problem (as in the initial-value problem with coefficients b and c known) to determine R_0, or in the inverse problem (more to be said of this in the next section).

Note that if $b=0$, then $\tilde{k}_0 = h_0 = 0$, and system (35) and (36) reduces to the previously developed one-dimensional system for the nondissipative wave equation.[5]

The second moment, defined as

$$u_2(x_3,t) = \int\int_{\mathbf{R}^2} (x_1^2 + x_2^2)u(x,t)dx_1\,dx_2, \tag{37}$$

will be considered next. It is shown in Appendix A that the second moments of the operators $\mathbf{K}^{-1}$,$\mathbf{K}'\mathbf{K}^{-1}$ are given by

$$(\mathbf{K}^{-1}u)_2 = \left(\frac{1}{c}\frac{\partial}{\partial t} + \frac{bc}{2}\right)u_2 + \tilde{k}_0*u_2 + \tilde{k}_2*u_0, \tag{38}$$

$$(\mathbf{K}'\mathbf{K}^{-1}u)_2 = h_2*u_0 + h_0*u_2 + \frac{1}{c}\frac{\partial c}{\partial x_3}u_2, \tag{39}$$

where

$$\tilde{k}_2(x_3,t) = -2ce^{-\eta}I_0(\eta), \tag{40}$$

$$h_2(x_3,t) = 2\frac{\partial}{\partial x_3}\left(\frac{1-e^{-2\eta}}{b}\right), \tag{41}$$

with η given by Eq. (32), and I_0 the modified Bessel function of order zero.

From this the form of the second moment of $\mathbf{R}$ can be easily ascertained. Take the second moment of each equation in system (24) and employ relations (38) and (39). This will result in a one-dimensional linear system involving u_2^+, u_2^-, u_0^+, and u_0^-. Since $u_0^+ = \mathbf{R}_0u_0^-$, this implies that u_2^+ is a linear combination of u_2^- and u_0^-. When the arbitrary function u^- is chosen so that u_0^- is zero, the system corresponds to the exact system for the zeroth moments but with the pair u_0^+, u_0^- replaced by u_2^+, u_2^-. Hence it follows that

$$u_2^+ = (\mathbf{R}u^-)_2 = \mathbf{R}_2u_0^- + \mathbf{R}_0u_2^- = R_2*u_0^- + R_0*u_2^-. \tag{42}$$

The equation for the second moment $\mathbf{R}_2$ of the reflection operator can be easily obtained by taking the second moment of Eq. (26) with the arbitrary function u_2^- set equal to zero. The second moments of the particular terms in Eq. (26) are given as follows (with $u_2^- = 0$):

$$(\mathbf{RK}^{-1}u^-)_2 = R_2*\left(\frac{1}{c}\frac{\partial u_0^-}{\partial t}+\frac{bc}{2}u_0^-\right) + (R_2*\tilde{k}_0 + R_0*\tilde{k}_2)*u_0^-,$$

$$(\mathbf{K}^{-1}\mathbf{R}u^-)_2 = \left(\frac{1}{c}\frac{\partial}{\partial t}+\frac{bc}{2}\right)R_2*u_0^- + (\tilde{k}_0*R_2 + \tilde{k}_2*R_0)*u_0^-,$$

$$(\mathbf{RK'K}^{-1}u^- - \mathbf{K'K}^{-1}\mathbf{R}u^-)_2 = 0,$$

$$(\mathbf{RK'K}^{-1}\mathbf{R}u^-)_2 = \left(+\frac{2}{c}\frac{\partial c}{\partial x_3}R_0*R_2 + 2R_0*h_0*R_2 + R_0*h_2*R_0\right)*u_0^-,$$

$$(\mathbf{K'K}^{-1}u^-)_2 = h_2*u_0^-.$$

On integrating by parts the following term is reduced as follows:

$$R_2*\frac{1}{c}\frac{\partial u_0^-}{\partial t}+\frac{1}{c}\frac{\partial}{\partial t}(R_2*u_0^-) = \frac{2}{c}\frac{\partial R_2}{\partial t}*u_0^- + \frac{2}{c}R_2(x_3,0)u_0^-(x_3,t).$$

Finally, employing all these relations, the second moment of Eq. (26), with $u_2^- = 0$, has the form

$$\Gamma_1*u_0^- + (2/c)R_2(x_3,0)u_0^-(x_3,t) = 0, \tag{43}$$

where Γ_1 is given by the left-hand side of Eq. (44). Since u_0^- is an arbitrary function vanishing at $t=0$, the equation for R_2 can be obtained by setting $\Gamma_1 = 0$, yielding

$$\left(\frac{\partial}{\partial x_3}+\frac{2}{c}\frac{\partial}{\partial t}+bc\right)R_2 + \left(2\tilde{k}_0 - \frac{1}{c}\frac{\partial c}{\partial x_3}R_0 - h_0*R_0\right)*R_2 + 2\tilde{k}_2*R_0 - \frac{1}{2}h_2*R_0*R_0 + \frac{1}{2}h_2 = 0. \tag{44}$$

The initial condition is obtained by setting the coefficient of $u_0^-(x_3,t)$ in Eq. (43) to zero, giving

$$R_2(x_3,0) = 0. \tag{45}$$

VII. INVERSE SCATTERING

The results of the previous section will now be applied to the inverse problem of determining the coefficients $c(x_3)$, $b(x_3)$ in the portion $-L < x_3 < 0$ of the planar-stratified half-space $x_3 < 0$, from data on the reflected field at $x_3 = 0$. Due to the absorption of the wave as it moves into the medium, it is clear that only the leading portion of the reflected wave that is produced will be useful in the inverse problem; hence only the properties of the outer portion, or skin, of thickness L of the medium can be meaningfully determined. The depth of penetration L is the order of the e-fold distance of a wave traveling into the medium. For a homogeneous dissipative medium, L is the order of $(bc)^{-1}$. The physics of this is clearly illustrated by the exact solution (produced by a source at time t_0)

$$u(z,t) = -\left(\frac{c}{2}\right)\exp\left[-\frac{bc^2(t-t_0)}{2}\right]I_0(\zeta)H\left(t-t_0-\frac{z}{c}\right) \tag{46}$$

(where $\zeta = bc^2[(t-t_0)^2 - (z/c)^2]^{1/2}/2$) of the one-dimensional initial-value half-space problem

$$\frac{1}{c^2}\frac{\partial^2 u}{\partial t^2}+b\frac{\partial u}{\partial t}-\frac{\partial^2 u}{\partial z^2} = 0, \quad t>0, \quad z>0,$$

$$u = \frac{\partial u}{\partial t} = 0, \quad \text{for } t=0, \quad z>0,$$

$$\frac{\partial u}{\partial z} = \delta(t-t_0), \quad \text{for } z=0, \quad t>0.$$

[The solution can be obtained directly from expression (7) by integrating out the x_1, x_2 variables, substituting in the boundary conditions, and setting $x_3 = z$.] It is easily seen that the head of the pulse at time $t = t_0 - z/c + 0$ has the value

$$u\left(z,t_0-\frac{z}{c}+0\right) = -\left(\frac{c}{2}\right)\exp\left(-\frac{bcz}{2}\right). \tag{47}$$

Thus the amplitude of the leading edge of the pulse is reduced by a factor e^{-1} in a distance of $2/(bc)$.

The inverse problem is comprised of two parts. The first, or preliminary, part is a deconvolution process, which is to determine the reflection operator on $x_3 = 0$ from the measurements of the scattered field. The measurements may be made either on the surface $x_3 = 0$ or elsewhere in the homogeneous region $x_3 > 0$. In particular, for the analysis here the deconvolution process would involve the recovery of $R_0(0,t)$ and $R_2(0,t)$, for a finite time interval $0 \leqslant t < T$, from measurements of the moments of the scattered field $u_0^+(0,t)$, $u_2^+(0,t)$ produced by an arbitrary incident wave u^-. It should be noted that the process of taking moments of the scattered field would tend to smooth out the data. Also, if the incident field is produced by a point source (or source with compact support), then because of the initial-value nature of the time-dependent problem, measurements of the scattered field need only be made over a finite portion of the surface $x_3 = 0$.

The deconvolution process will not be addressed here. Instead the main problem of determining the coefficients $c(x_3)$, $b(x_3)$ from knowledge of $R_0(0,t)$ and $R_2(0,t)$, $0 \leqslant t < T$ will be analyzed. To formulate the inverse problem, a change of the dependent variable,

$$\xi = -\int_0^{x_3}\frac{1}{c(y)}dy, \quad \text{for } x_3 \leqslant 0, \tag{48}$$

will be employed. Note that $\xi \geqslant 0$, and the inverse transformation is given by

$$x_3 = -\int_0^{\xi} c\,d\xi. \tag{49}$$

In addition, it is convenient to set

$$bc^2 = B(\xi). \tag{50}$$

Equations (35) and (44) for R_0 the zeroth moment and for R_2 the second moment take the form

$$\left(\frac{\partial}{\partial \xi}-2\frac{\partial}{\partial t}-B\right)R_0 = Q_1(R_0)*R_0 + \frac{1}{4}\frac{\partial B}{\partial \xi}e^{-2\eta}, \tag{51}$$

$$\left(\frac{\partial}{\partial\xi}-2\frac{\partial}{\partial t}-B\right)R_2=Q_2(R_0)*R_2+S(R_0)\,, \tag{52}$$

where, for $m=1,2$,

$$Q_m(R_0)=-B^2e^{-\eta}\frac{I_1(\eta)}{(2\eta)}+\frac{m}{2c}\frac{\partial c}{\partial\xi}R_0-\frac{m}{4}\frac{\partial B}{\partial\xi}e^{-2\eta}*R_0\,, \tag{53}$$

$$S(R_0)=(-4c^2e^{-\eta}I_0(\eta)+P*R_0)*R_0-P\,, \tag{54}$$

$$P(\xi,t)=\frac{\partial}{\partial\xi}\left[\frac{c^2(1-e^{-2\eta})}{B}\right], \tag{55}$$

with I_0, I_1 being modified Bessel functions and $\eta=Bt/2$.

The initial conditions are given by

$$R_0(\xi,0+)=+\frac{\partial c}{\partial\xi}(4c)^{-1}\,, \tag{56}$$

$$R_2(\xi,0+)=0\,. \tag{57}$$

The inverse problem can now be stated as follows.

Inverse Problem: Given the values of $R_0(0,t)$ and $R_2(0,t)$ for $0\leqslant t<T$, and the values of c and B at $\xi=0$ denoted by c_0 and B_0, respectively, solve the nonlinear system (51), (52), (56), and (57) for $R_0(\xi,t)$, $R_2(\xi,t)$ in the triangular region $0\leqslant t\leqslant T$, $0\leqslant 2\xi\leqslant T-t$, and recover B and c as a function of ξ.

Equation (49) may then be used to obtain B and c as a function of x_3.

Unfortunately initial conditions (56) and (57) do not contain the quantity B explicitly. To obtain the initial conditions that contain B let $t\to 0+$ in Eq. (51), yielding

$$\dot R_0(\xi,0+)=\frac{1}{2}\left(\frac{\partial}{\partial\xi}-B\right)R_0(\xi,0+)-\frac{1}{8}\frac{\partial B}{\partial\xi}\,, \tag{58}$$

where the dot indicates the differentiation with respect to t. Combined with Eq. (56), initial condition (58) reduces to

$$\dot R_0(\xi,0+)=\frac{1}{8}\left(\frac{\partial}{\partial\xi}-B\right)\frac{\partial}{\partial\xi}\ln c-\frac{1}{8}\frac{\partial B}{\partial\xi}\,. \tag{59}$$

Equation (52), in the limit as $t\to 0+$, yields

$$\dot R_2(\xi,0+)=0\,. \tag{60}$$

To get a nonzero initial condition for R_2, differentiate Eq. (52) with respect to t and let $t\to 0+$ to give

$$\ddot R_2(\xi,0+)=\frac{3}{2}c\frac{\partial c}{\partial\xi}\,. \tag{61}$$

Now there may be a difficulty in using Eq. (59) due to the term $(\partial^2/\partial\xi^2)\ln c$, which may lead to an instability in the solution of the problem, since it involves the derivative of one of the unknown functions $(\partial/\partial\xi)\ln c$. To alleviate this, Eqs. (59) and (61) will be combined in an appropriate manner. However, before obtaining the appropriate initial condition, the corresponding equations for $\dot R_0$ and $\ddot R_2$ must be sought.

The equation for $\dot R_0$ is obtained by differentiating Eq. (51), giving

$$\left(\frac{\partial}{\partial\xi}-2\frac{\partial}{\partial t}-B\right)\dot R_0=Q_1(R_0)*\dot R_0+\frac{1}{4}\frac{\partial c}{c\,\partial\xi}Q_1(R_0)-\frac{B}{4}\frac{\partial B}{\partial\xi}e^{-2\eta}\,. \tag{62}$$

Noting that

$$R_0(\xi,t)=\frac{1}{4c}\frac{\partial c}{\partial\xi}+\int_0^t\dot R_0(\xi,\tau)d\tau\,, \tag{63}$$

Eq. (62) has the general form

$$\left(\frac{\partial}{\partial\xi}-2\frac{\partial}{\partial t}-B\right)\dot R_0=G_1\left(\dot R_0,\frac{1}{c}\frac{\partial c}{\partial\xi},\frac{\partial B}{\partial\xi}\right), \tag{62'}$$

where G_1 is an integrable function involving quadratures (no derivatives) of the variables $\dot R_0$, $\partial/\partial\xi\ln c$, $\partial B/\partial\xi$.

Using the fact that $R_2=\dot R_2=0$ at $t=0$, the equation for $\ddot R_2$ can be obtained by differentiating Eq. (52) twice to yield

$$\left(\frac{\partial}{\partial\xi}-2\frac{\partial}{\partial t}-B\right)\ddot R_2=Q_2(R_0)*\ddot R_2+\frac{\partial^2}{\partial t^2}S(R_0)\,, \tag{64}$$

where

$$\begin{aligned}\frac{\partial^2 S}{\partial t^2}(R_0)&=-4c^2\left[\frac{\partial^2}{\partial t^2}(e^{-\eta}I_0(\eta))\right]*R_0+2c^2BR_0-4c^2\dot R_0\\&\quad-\left[\frac{\partial}{\partial\xi}(c^2Be^{-2\eta})\right]*R_0*R_0+\frac{\partial c^2}{\partial\xi}R_0*R_0\\&\quad+\frac{\partial}{\partial\xi}(c^2Be^{-2\eta})\,.\end{aligned}$$

The equation for $\ddot R_2$ has the general form

$$\left(\frac{\partial}{\partial\xi}-2\frac{\partial}{\partial t}-B\right)\ddot R_2=G_2\left(\dot R_0,\ddot R_2,\frac{1}{c}\frac{\partial c}{\partial\xi},\frac{\partial B}{\partial\xi}\right), \tag{64'}$$

where G_2 is an integrable function involving quadratures (no derivatives) of the variables $\dot R_0$, $\ddot R_2$, $(1/c)(\partial c/\partial\xi)$, $\partial B/\partial\xi$. This will be made use of in subsequent analysis.

The remaining initial condition can now be obtained. Let $t\to 0+$ in Eq. (64), giving

$$2\dddot R_2(\xi,0+)=\left\{\left(\frac{\partial}{\partial\xi}-B\right)\ddot R_2-2c^2BR_0+4c^2\dot R_0\right\}_{t=0}-\frac{\partial}{\partial\xi}(c^2B)\,.$$

Combining this with Eqs. (56), (59), and (61) yields the resulting initial condition,

$$\frac{1}{c^2(\xi)}\ddot R_2(\xi,0+)-8\dot R_0(\xi,0+)=\frac{3}{2}\left(\frac{1}{c}\frac{\partial c}{\partial\xi}\right)^2-\frac{5B}{4c}\frac{\partial c}{\partial\xi}+\frac{1}{4}\frac{\partial B}{\partial\xi}\,. \tag{65}$$

An alternative mathematical statement of the inverse problem can now be given.

Inverse Problem (Alternative Form): Given the values of $\dot R_0(0,t)$ and $\ddot R_2(0,t)$ for $0\leqslant t<T$, and the values of c and B at $\xi=0$ denoted by c_0 and B_0, respectively, solve the nonlinear system (62), (64), (61), and (65) for $\dot R_0(\xi,t)$ and $\ddot R_2(\xi,t)$ in the triangular region $0\leqslant t<T$, $0\leqslant 2\xi\leqslant T-t$, and to recover B and c as a function of ξ.

In a numerical treatment of the alternative form of the inverse problem, one difficulty may arise due to the fact that the condition (65) requires a derivative of $\ddot R_2(\xi,t)$ to be computed at $t=0$. This may lead to an instability. However,

it is shown in Appendix B that the derivative $\ddot{R}_2(\xi,0)$ can be computed by quadratures from $\ddot{R}_2(\xi,t)$ and $\dot{R}_0(\xi,t)$.

A possible numerical approach to solving the nonlinear system could be modeled after the approach taken by Cornes and Krueger *et al.*[5,7] This approach needs to be looked into in more detail, and will be left to future analysis and computation.

However, to complete the analysis of the inverse problem in this paper, a sketch of a proof indicating that the inverse problem has a unique solution for T sufficiently small will be presented.

Essentially what has to be shown is that the system of integral–differential equations (61), (62), (64), and (65) can be transformed into a system of integral equations of the form

$$X_i = X_i^0 + F_i(X_1, X_2, X_3, X_4), \quad i = 1,...,4, \tag{66}$$

where the unknown quantities are given by

$$X_1(\xi,t) = \dot{R}_0(\xi,t), \tag{67}$$

$$X_2(\xi,t) = \ddot{R}_2(\xi,t), \tag{68}$$

$$X_3(\xi) = \frac{1}{c}\frac{\partial c}{\partial \xi}, \tag{69}$$

$$X_4(\xi) = \frac{\partial B}{\partial \xi}, \tag{70}$$

with the quantities X_i^0, F_i, $i = 1,...,4$, to be determined.

Here X_1, X_2 are continuous functions over the domain D, where

$$D = \{(\xi,t) | 0 \leqslant 2\xi \leqslant T - t,\ 0 \leqslant t \leqslant T\},$$

with norms

$$\|X_i\|_\infty = \sup_{\xi,t \in D} |X_i(\xi,t)|,$$

and X_3, X_4 are continuous functions over the interval $0 \leqslant \xi \leqslant T/2$ with norms

$$\|X_i\|_\infty = \sup_{0 \leqslant \xi \leqslant T/2} |X_i(\xi)|.$$

To transform the differential equations (62′), (64′) into a strictly integral formulation, the following solution

$$f(\xi,t) = e^{\beta(\xi)} f_0(t + 2\xi) + \int_0^\xi e^{\beta(\xi) - \beta(\xi')} g(\xi', 2\xi - 2\xi' + t) d\xi', \tag{71}$$

where

$$\beta(\xi) = \int_0^\xi B(\xi') d\xi', \tag{72}$$

of the system

$$\left(\frac{\partial}{\partial \xi} - 2\frac{\partial}{\partial t} - B\right) = g(\xi,t), \tag{73}$$

$$f(0,t) = f_0(t) \tag{74}$$

will be employed.

Let the values of $\dot{R}_0(\xi,t)$ and $\ddot{R}_2(\xi,t)$ at $\xi = 0$ be given by

$$\dot{R}_0(0,t) = R_{00}(t), \quad \ddot{R}_2(0,t) = R_{20}(t). \tag{75}$$

Then Eqs. (62′) and (64′) can be put in the form

$$\dot{R}_0(\xi,t) = e^{\beta(\xi)} R_{00}(t + 2\xi) + \int_0^\xi e^{\beta(\xi) - \beta(\xi')} G_1(\xi', 2\xi - 2\xi' + t) d\xi', \tag{76}$$

$$\ddot{R}_2(\xi,t) = e^{\beta(\xi)} R_{20}(t + 2\xi) + \int_0^\xi e^{\beta(\xi) - \beta(\xi')} G_2(\xi', 2\xi - 2\xi' + t) d\xi', \tag{77}$$

where G_1 and G_2 are given by the right-hand sides of Eqs. (62) and (64), respectively.

Note that, from Eqs. (69), (70), and (72),

$$B(\xi) = B_0 + \int_0^\xi X_4(\xi') d\xi', \tag{78}$$

$$\beta(\xi) = B_0 \xi + \int_0^\xi \int_0^{\xi'} X_4 \, d\xi'' \, d\xi', \tag{79}$$

$$c(\xi) = c_0 \exp \int_0^\xi X_3(\xi') d\xi', \tag{80}$$

where B_0 and C_0 were defined previously as the values of these quantities at $\xi = 0$. Thus it can be seen that if one sets

$$X_1^0 = R_{00}(t + 2\xi) e^{B_0 \xi}, \quad X_2^0 = R_{20}(t + 2\xi) e^{B_0 \xi}, \tag{81}$$

$$F_1 = (e^{\beta(\xi)} - e^{B_0 \xi}) R_{00}(t + 2\xi) + \int_0^\xi e^{\beta(\xi) - \beta(\xi')} G_1(\xi', 2\xi - 2\xi' + t) d\xi', \tag{82}$$

with F_2 given by expression (82) with R_{00}, G_1 replaced by R_{20}, G_2, respectively, and employs the forms of G_1, G_2 as indicated by Eqs. (62′) and (64′), then as functions of the variables X_i, $i = 1,...,4$ (noting, however, that F_1 is independent of X_3),

$$F_1 = F_1(X_1, X_2, X_4), \quad F_2 = F_2(X_1, X_2, X_3, X_4).$$

Thus Eqs. (62) and (64) can be placed in the form of Eq. (66) for $i = 1,2$, where F_1, F_2 contain no derivatives of the variables $X_i = 1,...,4$. Also, as $T \to 0+$, the domain D shrinks to zero. Since $0 \leqslant \xi \leqslant T/2$, it is easily seen that $F_1 \to O(T)$, $F_2 \to O(T)$.

The third equation of the set given by Eq. (66) is obtained by dividing both sides of Eq. (77) by $c^2(\xi)$ and letting $t \to 0$, and substituting it into condition (61). Condition (61) can then be placed in the form

$$X_3 = X_3^0 + F_3(X_1, X_2, X_3),$$

where

$$X_3^0(\xi) = 2/(3c_0^2) R_{20}(2\xi) e^{B_0 \xi}, \tag{83}$$

$$F_3 = \left\{ e^{\beta(\xi)} R_{20}(2\xi) + \int_0^\xi e^{\beta(\xi) - \beta(\xi')} G_2(\xi', 2\xi - 2\xi') d\xi' \right\} \frac{1}{c^2(\xi)} - \frac{R_{20}(2\xi)}{c_0^2}. \tag{84}$$

To transform the remaining equation [namely Eq. (65)] into the form of Eq. (66), the result in Appendix B that $\ddot{R}_2(\xi,0)$ has the form

$$\ddot{R}_2(\xi,0) = e^{\beta(\xi)}\dot{R}_{20}(2\xi) + \int_0^{\xi} e^{\beta(\xi)-\beta(\xi')}\frac{\partial G_2}{\partial t}(\xi',2\xi-2\xi')d\xi', \tag{85}$$

where the integral term is shown to be a function of $\dot{R}_0$, $\ddot{R}_2$, $(1/c)\partial c/\partial\xi$, and $\partial B/\partial\xi$, is employed. On substituting this and expression (76) with $t=0$, into Eq. (65), Eq. (65) takes the form

$$X_4 = X_4^0 + F_4(X_1, X_2, X_3, X_4),$$

where

$$X_4^0 = 5B_0X_3^0 - 6(X_3^0)^2 + 4e^{B_0\xi}[\dot{R}_{20}(2\xi)/c_0^2 - 8R_{00}(2\xi)]. \tag{86}$$

Having shown that Eqs. (61), (62), (64), and (65) can be placed in the form of Eq. (66), the inverse problem reduces to solving the system

$$X = X^0 + F(X) \tag{87}$$

for the vector valued function $X=(X_1, X_2, X_3, X_4)$ in the domain D, where $F=(F_1, F_2, F_3, F_4)$.

The form of Eq. (87) lends itself naturally to solution by successive approximations

$$X^m = X^0 + F(X^{m-1}), \tag{88}$$

starting with the initial approximation X^0. Introducing the norm

$$\|X\| = \max_{i=1,\ldots,4}\|X_i\|_\infty,$$

the iteration process (88) will converge to a unique solution[19] in the ball $U(X^0,r)$: $\|X-X^0\|<r$ if, for every $X,Y\in U(X^0,r)$, there is a $0\leqslant\theta<1$ such that

$$\|F(X)-F(Y)\|\leqslant\theta\|X-Y\|, \tag{89}$$

$$\|F(X^0)\|\leqslant(1-\theta)r. \tag{90}$$

The basic problem, then, is to show that a nonzero value of T can be chosen so that an r and $\theta<1$ can be picked so that inequalities (89) and (90) hold. The fact that this can be done for T sufficiently small follows from the result that $F_i\to O(T)$ as $T\to 0+$, which has been pointed out. This implies (leaving out the details) that

$$\|F(X)-F(Y)\|\leqslant Tp(r,T)\|X-Y\|,$$

where $p(r,T)$ is a bounded function for finite values of r and T. In addition, since $X_0=X_0(\xi)$ and $0\leqslant\xi\leqslant T/2$, the following bound can be shown:

$$\|F(X_0)\|\leqslant m(T).$$

Hence taking $\theta=\frac{1}{2}$, inequality (90) is satisfied if $r=2m(T)$, and inequality (89) is then satisfied provided that

$$Tp(2m(T),T)\leqslant\tfrac{1}{2}.$$

It is easily seen that since p is bounded, the inequality can be achieved for a sufficiently small value of T. This shows that a local solution to the inverse problem exists.

ACKNOWLEDGMENT

The research in this paper was supported by the Office of Naval Research Contract No. N-00014-86-K-0413.

APPENDIX A: ZEROTH AND SECOND MOMENTS OF K^{-1} AND K'

The operator $\mathbf{K}$ has the form on the surface S ($x_3=$ const)

$$\mathbf{K}u[x,t] = \int\int_{R^2}\int_0^t k(r,x_3,t-s)u(y_1,y_2,x_3,s)ds\,dy_1\,dy_2, \tag{A1}$$

where

$$k(r,x_3,t) = \left\{\frac{1}{r}\delta\left(t-\frac{r}{c}\right) + \frac{b^2c^3}{4\zeta}I_1(\zeta)H\left(t-\frac{r}{c}\right)\right\}\frac{\exp(-bc^2t/2)}{2\pi}, \tag{A2}$$

where $r^2=(x_1-y_1)^2+(x_2-y_2)^2$ and $\zeta=bc(c^2t^2-r^2)^{1/2}/2$. It can be shown that the zeroth and second moments of $\mathbf{K}$ are given in terms of a convolution in the time variable, as follows:

$$(\mathbf{K}u)_0 = k_0(x_3,t)*u_0(x_3,t) = \int_0^t k_0(x_3,t-s)u_0(x_3,s)ds, \tag{A3}$$

$$(\mathbf{K}u)_2 = k_2(x_3,t)*u_0(x_3,t) + k_0(x_3,t)*u_2(x_3,t), \tag{A4}$$

where

$$k_0(x_3,t) = 2\pi\int_0^\infty k(r,x_3,t)r\,dr, \tag{A5}$$

$$k_2(x_3,t) = 2\pi\int_0^\infty k(r,x_3,t)r^3\,dr. \tag{A6}$$

On changing the variable of integration from r to ζ, $k_0(x_3,t)$ is evaluated as follows:

$$k_0(x_3,t) = c\left[1+\int_0^\eta I_1(\zeta)d\zeta\right]e^{-\eta} = ce^{-\eta}I_0(\eta), \tag{A7}$$

where

$$\eta = bc^2t/2. \tag{A8}$$

The second moment term is evaluated in a similar manner, as follows:

$$k_2(x_3,t) = c^3e^{-\eta}\left\{1+\int_0^\eta\left[1-\left(\frac{\eta'}{\eta}\right)^2\right]I_1(\eta')d\eta'\right\}t^2 = c^3e^{-\eta}t^2[I_0(\eta)-I_2(\eta)] = 8\eta e^{-\eta}I_1(\eta)/(b^2c). \tag{A9}$$

The corresponding moments of the inverse operator $\mathbf{K}^{-1}$ are easily obtained. From Eq. (16) it is seen that on integrating by parts (in the transverse variables) and using Eq. (A3),

$$(\mathbf{K}^{-1}u)_0 = D(\mathbf{K}u)_0 = D(k_0*u_0), \tag{A10}$$

where

$$D = \frac{1}{c^2}\frac{\partial^2}{\partial t^2} + b\frac{\partial}{\partial t}. \tag{A11}$$

This can be further reduced to yield

$$(\mathbf{K}^{-1}u)_0 = \left(\frac{1}{c}\frac{\partial}{\partial t} + \frac{bc}{2}\right)u_0 + \tilde{k}_0 * u_0 , \tag{A12}$$

with

$$\tilde{k}_0(x_3,t) = -b^2c^3e^{-\eta}I_1(\eta)/(4\eta) . \tag{A13}$$

Similarly it follows that the second moment of $\mathbf{K}^{-1}$ can be obtained using Eqs. (A3) and (A4):

$$(\mathbf{K}^{-1}u)_2 = D(Ku)_2 - 4(Ku)_0$$

$$= D(k_2 * u_0 + k_0 * u_2) - 4k_0 * u_0 \tag{A14}$$

$$= \left(\frac{1}{c}\frac{\partial}{\partial t} + \frac{bc}{2}\right)u_2 + \tilde{k}_0 * u_2 + \tilde{k}_2 * u_0 , \tag{A15}$$

where

$$\tilde{k}_2(x_3,t) = Dk_2 - 4k_0 = -2ce^{-\eta}I_0(\eta) . \tag{A16}$$

The corresponding moments for $\mathbf{K}'$ can be easily obtained from expressions (A3) and (A4), and recalling the fact that $\mathbf{K}' = (\partial/\partial x_3)\mathbf{K}$, giving

$$(\mathbf{K}'u)_0 = \frac{\partial k_0}{\partial x_3} * u_0 , \tag{A17}$$

$$(\mathbf{K}'u)_2 = \frac{\partial k_2}{\partial x_3} * u_0 + \frac{\partial k_0}{\partial x_3} * u_2 . \tag{A18}$$

The zeroth moment of the composite operator $\mathbf{K}'\mathbf{K}^{-1}$ is obtained from (A17) and (A10), yielding

$$(\mathbf{K}'\mathbf{K}^{-1}u)_0 = \frac{\partial k_0}{\partial x_3} * (\mathbf{K}^{-1}u)_0 = \frac{\partial k_0}{\partial x_3} * D(k_0 * u_0) . \tag{A19}$$

To reduce this further, the general result for the convolution of two functions $f(t)$ and $g(t)$,

$$D(f*g) = (Df)*g + [(1/c^2)\dot{f}(0) + bf(0)]g(t) + (1/c^2)f(0)\dot{g}(t), \tag{A20}$$

has to be used. Here the dot indicates differentiation with respect to t. Using the fact that $u_0(x_3,t)$ vanishes at $t=0$, and that the convolution terms commute ($f*g = g*f$), expression (A19) is reduced as follows:

$$(\mathbf{K}'\mathbf{K}^{-1}u)_0 = D\left[\frac{\partial k_0}{\partial x_3} * (k_0 * u_0)\right] = D[\alpha * u_0]$$

$$= (D\alpha) * u_0 + (1/c^2)\dot{\alpha}(x_3,0)u_0(x_3,t) , \tag{A21}$$

where

$$\alpha(x_3,t) = \frac{1}{2}\frac{\partial}{\partial x_3}(k_0 * k_0) .$$

Inserting in expression (A7) for k_0 and changing the variable of integration, the convolution $k_0 * k_0$ can be evaluated as follows:

$$k_0 * k_0 = \left(\frac{2}{b}\right)e^{-\eta}\int_0^\eta I_0(\eta - \eta')I_0(\eta')d\eta'$$

$$= (1/b)(1 - e^{-2\eta}) ,$$

where η is given by Eq. (A8). Expression (A21) can now be simplified to

$$(\mathbf{K}'\mathbf{K}^{-1}u)_0 = +\frac{1}{c}\frac{\partial c}{\partial x_3}u_0(x_3,t) + h_0 * u_0 , \tag{A22}$$

where

$$h_0(x_3,t) = -\frac{1}{2}\frac{\partial(bc^2)}{\partial x_3}\exp(-bc^2t) . \tag{A23}$$

The second moment is obtained in a similar manner:

$$(\mathbf{K}'\mathbf{K}^{-1}u)_2 = \frac{\partial k_0}{\partial x_3} * (\mathbf{K}^{-1}u)_2 + \frac{\partial k_2}{\partial x_3} * (\mathbf{K}^{-1}u)_0$$

$$= \frac{\partial k_0}{\partial x_3} * [D(k_2 * u_0 + k_0 * u_2) - 4k_0 * u_0]$$

$$+ \frac{\partial k_2}{\partial x_3} * D(k_0 * u_0) . \tag{A24}$$

With u_0 replaced by u_2 in expression (A19) and (A22) it is easily seen that the term involving u_2 in expression (A24) is just

$$h_0 * u_2 + \frac{1}{c}\frac{\partial c}{\partial x_3}u_2 .$$

Using relation (A20) and the fact that the convolution terms commute, the term involving u_0 in expression (A24) reduces to

$$D\left[\frac{\partial}{\partial x_3}(k_0 * k_2) * u_0\right] - 2\frac{\partial}{\partial x_3}(k_0 * k_0) * u_0 .$$

The convolution $k_0 * k_2$ is easily evaluated (using the Laplace transform, or otherwise) to obtain

$$k_0 * k_2 = \left(\frac{16}{b^3c^2}\right)e^{-\eta}\int_0^\eta I_0(\eta - \eta')\eta' I_1(\eta')d\eta'$$

$$= (4/b^3c^2)\{(\eta - 1) + (\eta + 1)e^{-2\eta}\}, \tag{A25}$$

where η is given by Eq. (A8). Since it can now be shown that the time derivative of $\partial(k_0 * k_2)/\partial x_3$ vanishes at $t = 0$, it can be shown that expression (A24) reduces to

$$(\mathbf{K}'\mathbf{K}^{-1}u)_2 = h_2 * u_0 + \frac{1}{c}\frac{\partial c}{\partial x_3}u_2 + h_0 * u_2 , \tag{A26}$$

where $h_2 = D\,\partial(k_0 * k_2)/\partial x_3 - 2\partial(k_0 * k_0)/\partial x_3$, which simplifies to

$$h_2 = 2\frac{\partial}{\partial x_3}\left(\frac{1 - e^{-2\eta}}{b}\right) . \tag{A27}$$

Note: To simplify the calculation for h_2 use the following change of order of differentiation:

$$D\frac{\partial}{\partial x_3}(k_0 * k_2) = \left(\frac{1}{c^2}\frac{\partial}{\partial x_3}\frac{\partial^2}{\partial t^2} + b\frac{\partial}{\partial x_3}\frac{\partial}{\partial t}\right)k_0 * k_2 .$$

APPENDIX B: ALTERNATIVE FORM OF $\ddot{R}_2(\xi,0)$

It will be shown here how $\ddot{R}_2(\xi,0)$ may be computed by quadratures from $\dot{R}_2(\xi,t)$ and $\dot{R}_0(\xi,t)$. Differentiate Eq. (77) with respect to t, then set $t = 0$ to give

$$\ddot{R}_2(\xi,0) = e^{\beta(\xi)}\dot{R}_{20}(2\xi) + \int_0^\xi e^{\beta(\xi) - \beta(\xi')}\frac{\partial G_2}{\partial t} \times (\xi', 2\xi - 2\xi')d\xi' . \tag{B1}$$

With $G_2(\xi,t)$ given by the right-hand side of Eq. (64) it can be shown that

$$\frac{\partial}{\partial t} G_2(\xi,t) = \left(\frac{\partial}{\partial t} Q_2(R_0)\right) * \ddot{R}_2 + Q_2(R_0)|_{t=0} \ddot{R}_2 + \frac{\partial^3}{\partial t^3} [S(R_0) + 4c^2 \dot{R}_0] - 4c^2 \ddot{R}_0 .$$

It can be shown that all the terms except the last one depend upon $\ddot{R}_2$ and $\ddot{R}_0$ [noting Eq. (63)]. Thus $\partial G_2/\partial t$ has the form

$$\frac{\partial G_2}{\partial t} = T_1\left(\dot{R}_0, \ddot{R}_2, \frac{1}{c}\frac{\partial c}{\partial \xi}, \frac{\partial B}{\partial \xi}\right) - 4c^2 \ddot{R}_0 . \tag{B2}$$

To examine the critical term $\ddot{R}_0$, differentiate Eq. (76) to give

$$\ddot{R}_0(\xi,t) = e^{\beta(\xi)} \dot{R}_{00}(t + 2\xi) + \int_0^{\xi} e^{\beta(\xi) - \beta(\xi')} \frac{\partial G_1}{\partial t} (\xi', 2\xi - 2\xi' + t) d\xi' . \tag{B3}$$

With G_1 given by the right-hand side of Eq. (62), it is seen that

$$\frac{\partial G_1}{\partial t} = \frac{\partial Q_1}{\partial t} * \dot{R}_0 + Q_1(R_0)\Big|_{t=0} \dot{R}_0 + \frac{1}{4c}\frac{\partial c}{\partial \xi}\frac{\partial Q_1(R_0)}{\partial t} + \frac{B^2}{4}\frac{\partial B}{\partial \xi} e^{-2\eta};$$

hence

$$\frac{\partial G_1}{\partial t} = T_2\left(\dot{R}_0, \frac{1}{c}\frac{\partial c}{\partial \xi}, \frac{\partial B}{\partial \xi}\right).$$

Insertion of expression (B3) for $\ddot{R}_0$ into Eq. (B2) results in the fact that $\partial G_2/\partial t(\xi, 2\xi - 2\xi')$ in Eq. (B1) is a function of $\dot{R}_0$, $\ddot{R}_2$, $(1/c)\partial c/\partial \xi$, and $\partial B/\partial \xi$.

APPENDIX C: LEMMAS ON PROOF OF COMPACTNESS

Let S be a Lyapunov closed surface with total surface area A. Let $A(x_1, t_1; x_2, t_2)$ be the area of the surface of S that lies inside the sphere of center x_1 and radius ct_1 and outside the sphere of center x_2 and radius ct_2:

$$A(x_1, t_1; x_2, t_2) = \int_S H(ct_1 - |x_1 - y|) \times [1 - H(ct_2 - |x_2 - y|)] d\sigma_y . \tag{C1}$$

Since the surface is smooth and $A(x_1, t_1; x_1, t_1) = 0$, it follows that at least for small values of t, $A(x_1, t_1; x_2, t_2)$ will satisfy a Lipschitz-like condition. It will be assumed that the parameter T will be such that such a condition holds for $0 \leqslant t \leqslant T$, i.e.,

$$A(x_1, t_1; x_2, t_2) \leqslant \kappa\{|x_1 - x_2| + c|t_1 - t_2|\} , \tag{C2}$$

for $x_1, x_2 \in S$, $t_1, t_2 \in [0,T]$.

The space $C(S) \times C^{(0,\lambda)}[0,T]$ of functions $f(x,t)$, which are continuous in x and Hölder continuous in t, with Hölder exponent λ, $0 < \lambda \leqslant 1$,

$$|f(x,t_1) - f(x,t_2)| \leqslant U|t_1 - t_2|^{\lambda} ,$$

for all $x \in S$, $t_1 \neq t_2 \in [0,T]$, will be used. For this space the following norm will be employed:

$$\|f\|_{c \times c^{(0,\lambda)}} = \sup_{x \in S \atop t \in [0,T]} |f(x,t)| + \sup_{x \in S \atop t_1, t_2 \in [0,T] \atop t_1 \neq t_2} |f(x,t_1) - f(x,t_2)|/|t_1 - t_2|^{\lambda} . \tag{C3}$$

In the proof of the lemma that follows, we want to extend the definition of a function $f(x,t) \in C(S) \times C^{(0,\lambda)}[0,T]$ to a domain over negative values of t, as follows:

$$F(x,t) = \begin{cases} f(x,t) - f(x,0+), & t \geqslant 0, \\ 0, & t < 0. \end{cases} \tag{C4}$$

It can be easily shown that $F(x,t)$ is Hölder continuous on the interval $[T_1, T]$, where $T_1 < 0$, with the same Hölder coefficient, i.e.,

$$\sup_{x \in S \atop t_1, t_2 \in [T_1, T] \atop t_1 \neq t_2} |F(x,t_1) - F(x,t_2)|/|t_1 - t_2|^{\lambda} = \sup_{x \in S \atop t_1, t_2 \in [0,T] \atop t_1 \neq t_2} |f(x,t_1) - f(x,t_2)|/|t_1 - t_2|^{\lambda} . \tag{C5}$$

We will now prove the following lemma.

Lemma 1: If $\mathbf{G}f(x,t) = \int_S g(x,y) f(y,\tau) H(\tau) d\sigma_y$ has the kernel $g(x,y) = k(x,y)|x - y|^{-\alpha}$, where $\alpha < 2$ and $k(x,y)$ is continuous on $S \times S$, $\tau = t - |x - y|/c$, then

$$G: C(S) \times C^{0,\lambda}[0,T] \to C[S] \times C[0,T]$$

is compact, provided that the constant T is such that assumption (C2) holds.

Proof: Define the sequence of continuous kernels $g_n(x,y) = \min(n, g(x,y))$ to avoid the singularity at $x = y$. Since S is Lyapunov, it can be shown that $g_n \to g$ in $L_1(S)$, since there are $\varepsilon_n \to 0$ and a constant k_0 such that (Günter[20])

$$\int_S |g_n - g| \, d\sigma_y \leqslant \int_0^{\varepsilon_n} \int_0^{2\pi} \frac{k_0}{r^{\alpha}} r \, dr \, d\theta \to 0 ,$$

where (r,θ) is a local polar coordinate system centered at x, and lying in the tangent plane.

Let B be the ball in $C(S) \times C^{0,\lambda}[0,T]$ with unit norm. We shall show that $\mathbf{G}_n(B) = \{h \,|\, h = \mathbf{G}_n f, f \in B\}$ is bounded and equicontinuous, and hence, by the Arzela–Ascoli theorem, relatively compact.[21] By definition, then, $\mathbf{G}$ will be compact.[21]

Boundedness is easily shown. If $f \in B$, then

$$|\mathbf{G}_n f| \leqslant \sup_{x \in S \atop t \in [0,T]} \|\mathbf{G}_n f\|_{\infty} \leqslant \|g_n\|_{L_1(S)} \sup_{x \in S \atop t \in [0,T]} |f(x,t)| \leqslant \|g_n\|_{L_1(S)} \|f\|_{C \times C^{(0,\lambda)}} = \|g_n\|_{L_1(S)} .$$

To show equicontinuity, note that

$$|h(x_1, t_1) - h(x_2, t_2)| \leqslant \int_S |g_n(x_1, y) - g_n(x_2, y)| \, |f(y, \tau_1)| \, H(\tau_1) d\sigma_y$$

$$+\int_S |g_n(x_2,y)|\,|F(y,\tau_1)-F(y,\tau_2)|\,d\sigma_y$$

$$+\int_S |g_n(x_2,y)|\,|f(y,0+)|\,|H(\tau_1)-H(\tau_2)|\,d\sigma_y\,, \tag{C6}$$

where $\tau_i = t_i - |x_i - y|/c$ and F is defined by Eq. (C4).

Since $g_n(x,y)$ is continuous on the compact space $S\times S$, it is uniformly continuous; so for $\varepsilon>0$ there is a $\delta_1>0$ such that

$$|g_n(x_1,y)-g_n(x_2,y)|<\varepsilon/(3A), \quad \text{for } |x_1-x_2|<\delta_1\,.$$

Hence the first integral on the right-hand side of expression (C6) is bounded as follows:

$$\int_S |g_n(x_1,y)-g_n(x_2,y)|\,|f(y,\tau_1)|H(\tau_1)d\sigma_y$$

$$\leqslant\frac{\varepsilon}{3A}\int_S H(\tau_1)d\sigma_y\cdot\sup_{\substack{x\in S\\ t\in[0,T]}}|f(x,t)|$$

$$\leqslant(\varepsilon/3)\|f\|_{C\times C^{(0,\lambda)}}\,. \tag{C7}$$

To evaluate the second term, use the property of Hölder continuity, relationships (C3) and (C5), and the fact that $f\in B$ to give

$$|F(y,\tau_1)-F(y,\tau_2)|$$

$$\leqslant|\tau_1-\tau_2|^{\lambda}\,\|f\|_{C\times C^{0,\lambda}}$$

$$\leqslant\{|t_1-t_2|+(1/c)|x_1-x_2|\}^{\lambda}\,.$$

Hence given an $\varepsilon>0$, we can find a δ_2, $(2/c)\delta_2=[\varepsilon/3\|g_n\|_{L_1}]^{1/\lambda}$, such that for $c|t_1-t_2|<\delta_2$, $|x_1-x_2|<\delta_2$,

$$\int_S |g_n(x_2,y)|\,|F(y,\tau_1)-F(y,\tau_2)|\,d\sigma_y\leqslant\frac{\varepsilon}{3}\,. \tag{C8}$$

The third term in (C6) may be handled as follows:

$$\int_S |g_n(x_2,y)|\,|f(y,0+)|\,|H(\tau_1)-H(\tau_2)|\,d\sigma_y$$

$$\leqslant\|g_n\|_{\infty}\{A(x_1,t_1;x_2,t_2)$$

$$+A(x_2,t_2;x_1,t_1)\}\|f\|_{C\times C^{(0,\lambda)}}$$

$$\leqslant2\|g_n\|_{\infty}\kappa\{|x_1-x_2|+c|t_1-t_2|\}\,,$$

using relationship (C2). Hence given an $\varepsilon>0$, one can find a δ_3, where $\delta_3=\varepsilon/12\kappa\|g_n\|_{\infty}$, such that for $|x_1-x_2|<\delta_3$, $c|t_1-t_2<\delta_3$,

$$\int_S |g_n(x_2,y)|\,|f(y,0+)|\,|H(\tau_1)-H(\tau_2)|\,d\sigma_y\leqslant\frac{\varepsilon}{3}\,. \tag{C9}$$

Finally, one can combine the results of (C7)–(C9) to show that for $f\in B$,

$$|h(x_1,t_1)-h(x_2,t_2)|\leqslant\varepsilon\,,$$

whenever $|x_1-x_2|<\delta$, $c|t_1-t_2|<\delta$, where $\delta=\min[\delta_1,\delta_2,\delta_3]$. With this result we have shown equicontinuity.

Summarizing, the operator $\mathbf{G}_n$ is compact, and since $\mathbf{G}_n\to\mathbf{G}$, $\mathbf{G}$ is compact. ■

Lemma 2: If

$$\mathbf{H}f(x,t)=\int_S\int_0^{t-r/c}h(x,y;t-s)$$

$$\times H\left(t-\frac{r}{c}\right)f(y,s)ds\,d\sigma_y\,,$$

with $r=|x-y|$, has a kernel $h(x,y;t)$ that is continuous on $S\times S\times[0,T]$, then $\mathbf{H}$ is a compact operator mapping $C(S)\times C^{(0,\lambda)}[0,T]$ into $C(S)\times C[0,T]$.

Proof: Set

$$h_1(x,y;t)=\begin{cases}h(x,y;t)-h(x,y:r/c), & t>r/c,\\ 0, & t<r/c.\end{cases}$$

Then $\mathbf{H}f$ can be placed in the form

$$\mathbf{H}f=\int_S\int_0^T h_1(x,y:t-s)f(y,s)ds\,d\sigma_y$$

$$+\int_0^t\int_S h\left(x,y;\frac{r}{c}\right)f\left(y,\tau-\frac{r}{c}\right)H\left(\tau-\frac{r}{c}\right)d\sigma_y\,d\tau\,.$$

Since the kernel h_1 is continuous on the compact domain $S\times S\times[0,T]$, the first term is a compact operator. The second term involves a bounded operator (the integral in t) acting on the operator

$$\int_S h\left(x,y;\frac{r}{c}\right)f\left(y,t-\frac{r}{c}\right)H\left(t-\frac{r}{c}\right)d\sigma_y\,,$$

which by Lemma 1 is compact mapping $C(S)\times C^{0,\lambda}[0,T]$ into $C(S)\times C[0,T]$. The resulting product of operators is compact.[21] ■

[1] J. P. Corones, M. E. Davison, and R. J. Krueger, "Wave splittings, invariant imbedding and inverse scattering," *Inverse Optics*, Vol. 413, edited by A. J. Devaney (SPIE, Bellingham, WA, 1983), pp. 102–106.

[2] J. P. Corones, R. J. Krueger, and V. H. Weston, "Some recent results in inverse scattering theory," in *Inverse Problems of Acoustic and Elastic Waves*, edited by F. Santosa, Y. Poa, W. W. Symes, and C. Holland (SIAM, Philadelphia, PA, 1984), pp. 65–81.

[3] R. A. Beezley and R. J. Krueger, "An electromagnetic inverse problem for dispersive media," J. Math. Phys. **26**, 317 (1985).

[4] J. Corones and R. J. Krueger, "Obtaining scattering kernels using invariant imbedding," J. Math. Anal. Appl. **95**, 393 (1983)

[5] J. P. Corones, M. E. Davison, and R. J. Krueger, "Direct and inverse scattering in the time domain by invariant imbedding techniques," J. Acoust. Soc. Am. **74**, 1535 (1983).

[6] G. Kristensson and R. J. Krueger, "Direct and inverse scattering in the time domain for a dissipative wave operator equation. I. Scattering operators," J. Math. Phys. **27**, 1667 (1986).

[7] G. Kristensson and R. J. Krueger, "Direct and inverse scattering in the time domain for a dissipative wave operator equation. II. Simultaneous reconstruction of dissipation and phase velocity profiles," J. Math. Phys. **27**, 1683 (1986).

[8] J. P. Corones, M. E. Davison, and R. J. Krueger, "Dissipative inverse problems in the time domain," in *Inverse Methods in Electromagnetic Imaging, Part I*, edited by W. M. Boerner, H. Brand, L. A. Cram, D. T. Gjessing, W. Keydel, G. Schwierz, and M. Vogel (Reidel, Dordrecht, 1985), pp. 121–130.

[9] V. H. Weston, "Factorization of the wave equation in higher dimensions," J. Math. Phys. **28**, 1061 (1987).

[10] V. H. Weston, "Factorization of the wave equation in a non-planar stratified medium," J. Math. Phys. **29**, 36 (1988).

[11] K. L. Kreider, "Wave splitting approach to time-dependent inverse scattering for the stratified cylinder" (unpublished).

[12] P. M. Morse and H. Feshbach, *Methods of Theoretical Physics, Part I* (McGraw-Hill, New York, 1953).

[13] A. A. Buzdin, "On an inverse problem in a medium with absorption," Proc. Steklov Inst. Math. **115**, 77 (1971).

[14]A. S. Blagoveshchenskii, "On the various formulations of the one-dimensional inverse problem for the telegraph equation" Topics Math. Phys. **4**, 40 (1970).
[15]V. H. Weston, "On inverse scattering," J. Math. Phys. **15**, 209 (1974).
[16]R. J. Krueger "An inverse problem for absorbing medium with multiple discontinuities," Q. Appl. Math. **36**, 235 (1978).
[17]V. S. Vladimirov, *Equations of Mathematical Physics* (Dekker, New York, 1971).
[18]S. G. Mikhlin, *Linear Equations of Mathematical Physics* (Holt, Rinehart and Winston, New York, 1967).
[19]L. B. Rall, *Computational Solution of Nonlinear Operator Equations* (Krieger, New York, 1979).
[20]N. M. Günter, *Potential Theory and Its Applications to Basic Problems of Mathematical Physics* (Ungar, New York, 1967).
[21]C. W. Groetsch, *Elements of Applicable Functional Analysis* (Dekker, New York, 1980).

Solutions of Maxwell's equations with boundary conditions on the hyperplane $z-ct=0$

Pierre Hillion
Institut Henri Poincaré, 75231 Paris, France

(Received 30 March 1988; accepted for publication 15 June 1988)

In homogeneous free space the electromagnetic field may be represented by a second rank spinor, each component of which is a solution of the wave equation. This makes it possible to solve the boundary value problem for the electromagnetic field when the data given on the hyperplane $z-ct=0$ are entire functions. Two particular cases of boundary conditions that are not entire functions and that lead to a relativistic solution of Young's experiment are discussed.

I. INTRODUCTION

We consider the wave solutions to Maxwell's equations propagating along the $0z$ direction in homogeneous free space. We use Gaussian units, the transverse cylindrical coordinates r,θ, and as longitudinal coordinates the cone variables $\xi=z-ct$ and $\bar{\xi}=z+ct$. We introduce the complex vector $\Lambda=E+iH$ $(i=\sqrt{-1})$, where E and H are the (real) electric and magnetic fields.

From a relativistic point of view Λ is a self-dual tensor[1] having a well-known connection[1,2] with a traceless second rank spinor ψ_r^s $(r,s=1,2)$. Explicitly one has

$$\begin{aligned}&\Lambda_r-i\Lambda_\theta=e^{i\theta}\psi_1^2,\quad \Lambda_r+i\Lambda_\theta=e^{-i\theta}\psi_2^1,\\&\Lambda_z=\tfrac{1}{2}(\psi_2^2-\psi_1^1)=\psi_2^2.\end{aligned}\tag{1}$$

Let Ψ denote the matrix

$$\Psi=\begin{vmatrix}\psi_1^1 & \psi_1^2\\ \psi_2^1 & \psi_2^2\end{vmatrix};\tag{2}$$

then Ψ satisfies the Proca equation[2]

$$\begin{aligned}&\not{D}\Psi=0,\\&\not{D}=\begin{vmatrix}2\partial_{\bar{\xi}} & e^{-i\theta}(\partial_r-(i/r)\partial_\theta)\\ e^{i\theta}(\partial_r+(i/r)\partial_\theta) & -2\partial_\xi\end{vmatrix},\end{aligned}\tag{3}$$

and $\partial_r,\partial_\theta,\partial_\xi$, and $\partial_{\bar{\xi}}$ are the partial derivatives with respect to r, θ, ξ, and $\bar{\xi}$, respectively. It is easy to prove that the relations (1) and (3) imply that Λ satisfies Maxwell's equations.

When the electromagnetic field is self-conjugate,[3] Λ is a null vector $\Lambda\cdot\Lambda=0$, where the dot means the scalar product that implies

$$|E|=|H|,\quad E\cdot H=0,\quad \psi_1^2\psi_2^1-\psi_1^1\psi_2^2=0.\tag{4}$$

We are now looking for solutions of the type

$$\Lambda\equiv\Lambda(r,\theta,\xi)e^{ik\bar{\xi}},\quad \Psi\equiv\Psi(r,\theta,\xi)e^{ik\bar{\xi}}.\tag{5}$$

Taking (5) into account the derivative matrix $\not{D}$ becomes

$$\not{D}=\begin{vmatrix}2ik & e^{-i\theta}(\partial r-(i/r)\partial_\theta)\\ e^{i\theta}(\partial_r+(i/r)\partial_\theta) & -2\partial_\xi\end{vmatrix}.\tag{6}$$

Then from the equation $\not{D}\Psi=0$ and from the traceless condition, one deduces the following relations:

$$\begin{aligned}&\psi_1^1=-\psi_2^2=-(i/2k)e^{-i\theta}(\partial_r-(i/r)\partial_\theta)\psi_2^1,\\&\psi_1^2=-(i/2k)e^{-i\theta}(\partial_r-(i/r)\partial_\theta)\psi_2^2,\end{aligned}\tag{7}$$

while ψ_2^1 is a solution of the wave equation

$$\begin{aligned}&\Delta_\perp\psi_2^1+4ik\,\partial_\xi\psi_2^1=0,\\&\Delta_\perp=(1/r)\partial_r(r\,\partial_r)+(1/r^2)\partial_\theta^2.\end{aligned}\tag{8}$$

Consequently, using relations (1) and (7), solutions of Maxwell's equations can be deduced from the solution ψ_2^1 of the wave equation (8).

We discuss here the boundary value problem for the electromagnetic field with data given on the hyperplane $\xi=0$ when the boundary conditions are consistent with relations (1).

II. BOUNDARY CONDITIONS ARE ENTIRE FUNCTIONS

We start with a summary of the results we previously obtained[4] for the wave equation. One first notes that if ψ is a solution of the wave equation independent of θ, then

$$\psi_n=r^n\frac{\partial^n}{(r\,\partial_r)^n}(\psi e^{in\theta}),\quad n \text{ positive integer},\tag{9}$$

is also a solution. Consequently it is sufficient to consider the solutions $\psi\equiv\psi(r,\xi)e^{ik\bar{\xi}}$ with boundary condition $\hat{\psi}(r)$ on $\xi=0$.

These solutions are easy to obtain when $\hat{\psi}$ is an entire function[5] of order 1 and type $0<\tau<\infty$ of the dimensionless variable $v=-\lambda^2r^2$. Then one has

$$\hat{\psi}(v)=\sum_{n=0}^{\infty}\frac{\hat{\psi}_n}{n!}v^n\tag{10}$$

and $\psi(r,\xi)$ is given[4] by the relation

$$\begin{aligned}\psi(r,\xi)=\lim_{\epsilon\mapsto 0}\frac{1}{\pi}\sqrt{\frac{\alpha_\epsilon}{\pi\xi}}\int_{-\infty}^{+\infty}ds\,e^{-\alpha_\epsilon s^2/4\xi}\\ \times\int_0^{\pi}d\theta f(s+ir\cos\theta),\end{aligned}\tag{11}$$

with $\alpha_\epsilon=\epsilon+4ik$, where ϵ is a positive scalar, and

$$\begin{aligned}f(s+ir\cos\theta)=\int_0^{\infty}dt\,e^{-t}\sum_{n=0}^{\infty}(-1)^n\frac{t^n}{(2n)!}\\ \times\hat{\psi}_n(2s+2ir\cos\theta)^{2n}.\end{aligned}\tag{11'}$$

 0022-2488/88/102219-04$02.50

Let us give some examples. Starting with the entire function of type 1, $\hat{\psi}(r) = e^{-(kr^2/a)}$ ($\lambda^2 = k/a$), we get from (11) and (11')

$$\psi(r,\xi) = [1/(a+i\xi)]e^{-kr^2/(a+i\xi)}. \tag{12}$$

This expression multiplied by $e^{ik\bar{\xi}}$ is the first focus wave mode solution[6] of the wave equation.

The entire function of type $\frac{1}{2}$, $\hat{\psi}(r) = I_0(\lambda r)$, where I_0 is the modified Bessel function of the first kind of order zero, leads to

$$\psi(r,\xi) = I_0(\lambda r)e^{i\lambda^2\xi/k} \tag{13}$$

while with the entire function

$$\hat{\psi}(r) = e^{-kr^2/2a}I_0(kr^2/2a)$$

we get

$$\psi(r,\xi) = [1/(a+i\xi)^{1/2}]e^{-kr^2/2(a+i\xi)}I_0(kr^2/2(a+i\xi)). \tag{14}$$

Applying (9) to (12)–(14) supplies the corresponding solutions depending on θ with boundary condition $\hat{\psi}(r)e^{in\theta}$.

Let us now come back to Maxwell's equations. To solve the boundary value problem with data on the hyperplane $\xi = 0$ consistent with the relations (1) one just has to identify ψ_2^1 with ψ and to use relations (1) and (7).

For instance, when ψ_2^1 is identified with (12) multiplied by $e^{ik\bar{\xi}}$, that is,

$$\psi_2^1 = [1/(a+i\xi)]e^{-kr^2/(a+i\xi)}e^{ik\bar{\xi}}, \tag{15}$$

we get from (7)

$$\begin{aligned}\psi_1^1 &= -\psi_2^2 = [ir/(a+i\xi)^2]e^{-kr^2/(a+i\xi)}e^{i(k\bar{\xi}-\theta)},\\ \psi_1^2 &= [r^2/(a+i\xi)^3]e^{-kr^2/(a+i\xi)}e^{i(k\bar{\xi}-2\theta)}.\end{aligned} \tag{15'}$$

It is easy to check that the condition (4) is satisfied so that the electromagnetic field defined by (15) and (15') is self-conjugate. Substituting these expressions into (1) gives

$$\begin{aligned}\Lambda_r - i\Lambda_\theta &= [r^2/(a+i\xi)^3]e^{-kr^2/(a+i\xi)}e^{i(k\bar{\xi}-\theta)},\\ \Lambda_r + i\Lambda_\theta &= [1/(a+i\xi)]e^{-kr^2/(a+i\xi)}e^{i(k\bar{\xi}-\theta)},\\ \Lambda_z &= [ir/(a+i\xi)^2]e^{-kr^2/(a+i\xi)}e^{i(k\bar{\xi}-\theta)},\end{aligned} \tag{16}$$

which is the first focus wave mode solution of Maxwell's equations.[7] A similar calculation gives, when one uses (13) multiplied by $e^{ik\bar{\xi}}$ as an expression of ψ_2^1,

$$\begin{aligned}\Lambda_r - i\Lambda_\theta &= -(\lambda^2/4k^2)(I_0(\lambda r) - (2/\lambda r)I_1(\lambda r))\\ &\quad\times e^{i(k\bar{\xi}+\lambda^2\xi/k-\theta)},\\ \Lambda_r + i\Lambda_\theta &= I_0(\lambda r)e^{i(k\bar{\xi}+\lambda^2\xi/k-\theta)},\\ \Lambda_3 &= -(i\lambda/2k)I_1(\lambda r)e^{i(k\bar{\xi}+\lambda^2\xi/k-\theta)}.\end{aligned} \tag{17}$$

With (14) one gets similar but more intricate results. Applying (9) to the previous expressions of ψ_2^1 supplies the solutions depending on θ.

To sum up, using (10), (11), and (11') together with (1), (7), and (9) supplies a lot of solutions to Maxwell's equations among which the focus wave modes have won a special fame.[8] The behavior of solutions (16) has been discussed elsewhere[9] as has the Poynting vector and the energy density.

For the focus wave modes the consistency of the boundary conditions leads to the following relations:

$$\begin{aligned}&(\cos\theta)(\hat{E}_r + \hat{H}_\theta) + (\sin\theta)(\hat{E}_\theta - \hat{H}_r) = (r^2/a^3)e^{-kr^2/a},\\ &(\sin\theta)(\hat{E}_r + \hat{H}_\theta) - (\cos\theta)(\hat{E}_\theta - \hat{H}_r) = 0,\\ &(\cos\theta)(\hat{E}_r - \hat{H}_\theta) - (\sin\theta)(\hat{E}_\theta + \hat{H}_r) = (1/a)e^{-kr^2/a},\\ &(\sin\theta)(\hat{E}_r - \hat{H}_\theta) + (\cos\theta)(\hat{E}_\theta + \hat{H}_r) = 0,\\ &(\cos\theta)\hat{E}_z - (\sin\theta)\hat{H}_z = 0,\\ &(\sin\theta)\hat{E}_z + (\cos\theta)\hat{H}_z = (r/a^2)e^{-kr^2/a},\end{aligned}$$

where we used (16) together with the definition of Λ.

III. YOUNG'S EXPERIMENT

We ignore whether there exists a general method for solving the boundary value problem when the boundary conditions on $\xi = 0$ are not entire functions. But in some simple cases the solution is easy to guess. For instance, let us assume that one has, on $\xi = 0$,

$$\hat{\psi}(x,y) = \delta(x)\delta(y^2 - d^2), \tag{18}$$

where $\delta(x)$ is the Dirac distribution and x,y are the transverse Cartesian coordinates. The condition (18) means that $\hat{\psi}$ is zero except at the pinholes ($x = 0$, $y = \pm d$) as in Young's experiment.

In Cartesian coordinates the relations (1), (7), and (8) become, respectively,

$$\begin{aligned}&\Lambda_x - i\Lambda_y = \psi_1^2, \quad \Lambda_x + i\Lambda_y = \psi_2^1,\\ &\Lambda_z = \tfrac{1}{2}(\psi_2^2 - \psi_1^1),\end{aligned} \tag{19}$$

with

$$\begin{aligned}\psi_1^1 &= -\psi_2^2 = -(i/2k)(\partial_x - i\partial_y)\psi_2^1,\\ \psi_1^2 &= -(i/2k)(\partial_x - i\partial_y)\psi_2^2,\end{aligned} \tag{20}$$

and

$$(\partial_x^2 + \partial_y^2 + 4ik\partial_\xi)\psi_2^1 = 0. \tag{21}$$

Consequently one has to find a solution of (21) with (18) as the boundary condition.

For $\xi > 0$ the Serendip method leads at once to

$$\psi(x,y,\xi) = (1/2\xi)e^{i(k/\xi)(x^2+y^2+d^2)}\cos(2kdy/\xi), \tag{22}$$

where we assume $k = k_0 + i\epsilon$ with $\epsilon > 0$ in order to make (18) possible. Let us remark that ψ as a function of x,y,ξ is not continuous at $\xi = x = 0$, $y = \pm d$, if these points are approached along the paraboloids $\epsilon(x^2 + (y \pm d)^2) - \xi = 0$, ψ goes off to infinity while ψ assumes the value zero when they are approached along the surface $\epsilon^{3/2}(x^2 + (y \pm d)^2) - \xi^{1/2} = 0$. It is trivial to check that (22) is a solution of (21). The solution in the half space $\xi < 0$ is obtained by changing k into $-k$.

In the scalar representation of the optical fields the light intensity $I(x,y,\xi)$ is proportional to $|\psi|^2$, that is, according to (22) for $\xi > 0$,

$$\begin{aligned}I &= (1/4\xi^2)e^{-2\epsilon/\xi(x^2+y^2+d^2)}\\ &\quad\times\cos(2kdy/\xi)\cos(2k^*dy/\xi),\end{aligned} \tag{23}$$

where the asterisk denotes the complex conjugation; in the limit $\epsilon \to 0$ this expression reduces to

$$I = (1/4\xi^2)\cos^2(2k_0 dy/\xi), \tag{23'}$$

which gives the interference pattern obtained in Young's experiment[10] except that the variable ξ takes the place of $2z$ as expected since the paraxial approximation of the wave equation is obtained by changing ξ into $2z$.

Consequently (23') may be considered as the relativistic expression of Young's experiment.

Let us now identify ψ_2^1 with (22), which we write as

$$\psi_2^1 = (1/2\xi)(\cos\alpha)e^{\beta}, \quad \alpha = 2kdy/\xi,$$
$$\beta = (ik/\xi)(x^2+y^2+d^2). \tag{24}$$

Substituting this expression into (19) gives

$$\psi_1^1 = (e^{\beta}/2\xi^2)(d\sin\alpha + (x-iy)\cos\alpha,$$
$$\psi_1^2 = -(e^{\beta}/2\xi^3)[((x-iy)^2 - d^2)\cos\alpha + 2d(x-iy)\sin\alpha]. \tag{24'}$$

Since the condition (14) is not fulfilled, this field is not self-conjugate. From (19), (24), and (24') we get for the electromagnetic field

$$\Lambda_x = (e^{\beta}/4\xi^3)((\xi^2 + d^2 + (x-iy)^2)\cos\alpha - 2d(x-iy)\sin\alpha),$$
$$\Lambda_y = -(ie^{\beta}/4\xi^3)((\xi^2 - d^2 + (x-iy)^2)\cos\alpha + 2d(x-iy)\sin\alpha), \tag{25}$$
$$\Lambda_z = (e^{\beta}/2\xi^2)((x-iy)\cos\alpha + d\sin\alpha).$$

These expressions are valid for ξ in the half space $\xi > 0$. For $\xi < 0$ one has to change k into $-k$, that is, β into $-\beta$. Since k is complex $\mathrm{Re}\,\beta < 0$ and the components of Λ are bounded in the transverse direction.

Let us now use (25) to discuss Young's experiment. Because of the short wavelength of visible light the interference pattern may be observed in practice only if d is much smaller than ξ. One may also assume x and y small compared with ξ since the amplitudes Λ_j decrease as $\exp[-(\epsilon/\xi)(x^2+y^2+d^2)]$.

Consequently neglecting terms of second or higher order in $d/\xi, x/\xi, y/\xi$, and their products, the expressions (25) reduce to

$$\Lambda_x = (e^{\beta}/2\xi)\cos\alpha_0,$$
$$\Lambda_y = -(ie^{\beta}/2\xi)\cos\alpha_0, \quad \alpha_0 = 2k_0 dy/\xi, \tag{26}$$
$$\Lambda_z = (e^{\beta}/2\xi^2)(d\sin\alpha_0 + (x-iy)\cos\alpha_0,$$

since Λ_z^2 is negligible the electromagnetic field (26) is self-conjugate to second order.

We now compute the Poynting vector and the energy density given by the relations

$$S_j = -(ic/4\pi)\epsilon_{jkl}\Lambda_k\Lambda_e^*, \quad j,k,l = 1,2,3,$$
$$W = (1/4\pi)|\Lambda|^2, \tag{27}$$

where ϵ_{jkl} is the permutation tensor. Substituting (26) into (27) gives

$$S_x = -A(x\cos\alpha_0 + d\sin\alpha_0), \quad S_y = \Lambda_y\cos\alpha_0,$$
$$S_z = \Lambda\xi\cos\alpha_0 = cW, \tag{28}$$

with

$$A = (c/8\pi\xi^3)e^{-(2\epsilon/\xi)(x^2+y^2+d^2)}\cos\alpha_0. \tag{28'}$$

Introducing the variables $S_{\xi} = S_z - cW, S_{\bar{\xi}} = S_z + cW$, the previous expressions become, in the limit $\epsilon \mapsto 0$,

$$S_x = S_y = S_{\xi} = 0 \ \mathrm{mod}(\mathrm{curl}\ U),$$
$$S_{\bar{\xi}} = (c/4\pi\xi^2)\cos^2\alpha_0, \tag{29}$$

where U is the vector with components

$$U_x = U_y = 0,$$
$$U_{\xi} = -(c/8\pi\xi^3)y(\cos\alpha_0)(x\cos\alpha_0 + d\sin\alpha_0). \tag{29'}$$

As is well known,[10] from a physical point of view curl U does not play any role. So we are left with $S_{\bar{\xi}}$, which is the expression (23') of light intensity multiplied by c/π. Two comments can be made.

(i) The fact that $S_{\bar{\xi}}$ and not S_0 is the energy density of the electromagnetic field is in agreement with the asynchronous formulation of relativity.[11]

(ii) The scalar theory may be used as long as d/ξ, x/ξ, and y/ξ are small. Let us also note that the case $d = 0$ corresponds to a point source with the usual inverse square law.

As another example of the boundary value problem we consider a boundary condition of the slot type:

$$\hat{\psi}(x,y) = \delta(y). \tag{30}$$

The Serendip method still supplies the solution of (21) satisfying (30) for $\xi = 0$:

$$\psi_2^1(x,y) = (1/\sqrt{\xi})e^{iky^2/\xi}, \quad \xi > 0, \tag{31}$$

which leads to

$$\psi_1^1 = -(iy/\xi^{3/2})e^{iky^2/\xi},$$
$$\psi_1^2 = \frac{y}{\xi^{3/2}}\left(\frac{y}{\xi} - \frac{1}{2ky}\right)e^{iky^2/\xi}, \tag{31'}$$

and for the electromagnetic field to

$$\Lambda_x = \frac{1}{2\sqrt{\xi}}\left(1 + \frac{y^2}{\xi^2} - \frac{i}{2k\xi}\right)e^{iky^2/\xi},$$
$$\Lambda_y = \frac{-i}{2\sqrt{\xi}}\left(1 - \frac{y^2}{\xi^2} + \frac{i}{2k\xi}\right)e^{iky^2/\xi}, \tag{32}$$
$$\Lambda_z = -(iy/\xi^{3/2})e^{iky^2/\xi}.$$

The components of the Poynting vector are

$$S_x = \frac{cy}{8\pi k\xi^3}e^{-2\epsilon y^2/\xi}, \quad S_y = \frac{cy}{8\pi k\xi^3}e^{-2\epsilon y^2/\xi},$$
$$S_z = \frac{c}{8\pi\xi}\left(1 - \frac{y^4}{\xi^4} - \frac{1}{4k^2\xi^2}\right)e^{-2\epsilon y^2/\xi}, \tag{33}$$
$$S_0 = \frac{c}{8\pi\xi}\left[\left(1 + \frac{y^2}{\xi^2}\right)^2 + \frac{1}{4k^2\xi^2}\right]e^{-2\epsilon y^2/\xi}.$$

Let us assume ξ very large so that, taking into account the exponential factor, one may neglect the terms y/ξ of second and higher order. We further assume $k\xi \gg 1$. Then we get from (34), in the limit $\epsilon \to 0$,

$$S_x = S_y = S_{\xi} = 0, \quad S_{\bar{\xi}} = c/4\pi\xi. \tag{34}$$

One notes from (34) that $S_{\bar{\xi}}$, which is still the energy density, decreases as ξ^{-1} instead of ξ^{-2} as in (29), a behavior typical of a slot antenna.

Let us write the boundary condition $\hat{\psi}(x,y)$ on the hyperplane $\xi = 0$ as

$$\hat{\psi}(x,y) = \hat{\psi}_1(x,y) + i\hat{\psi}_2(x,y). \tag{35}$$

Then, according to (20) and (21), the boundary conditions for the electromagnetic field are

$$\begin{aligned}
&\hat{E}_x - \hat{H}_y = \hat{\psi}_1, \quad \hat{H}_x + \hat{E}_y = \hat{\psi}_2 \\
&\hat{E}_x + \hat{H}_y = -(1/4k^2)(\partial_y^2\hat{\psi}_1 - 2\,\partial_x\partial_y\hat{\psi}_2 - \partial_2^2\hat{\psi}_1), \\
&\hat{H}_x - \hat{E}_y = -(1/4k^2)(\partial_y^2\hat{\psi}_2 - 2\,\partial_x\partial_y\hat{\psi}_1 - \partial_x^2\hat{\psi}_2), \\
&\hat{E}_z = -(1/2k)(\partial_y\hat{\psi}_1 - \partial_x\hat{\psi}_2), \\
&\hat{H}_z = -(1/2k)(\partial_x\hat{\psi}_1 + \partial_y\hat{\psi}_2).
\end{aligned} \tag{36}$$

The transformation $E \mapsto H$, $H \mapsto -E$, which leaves Maxwell's equations invariant in free space, changes Λ into $-\Lambda$ and $(\hat{\psi}_1, \hat{\psi}_2)$ into $(\hat{\psi}_2, -\hat{\psi}_1)$. Consequently the Babinet principle has a simple expression leading to further results.

IV. CONCLUSION

The present work suggests a statement and a conjecture.

First the statement: the boundary conditions on the hyperplane $\xi = 0$ play the role of a source term. We shall discuss this statement elsewhere by appealing to the asynchronous formulation of relativity[11] that we already met when we proved that the component $S_{\bar{\xi}}$ is the energy density. This last result that we obtained in Sec. III is also true[9] for the focus wave modes (16).

The conjecture starts with the remark that in Young's experiment, as described here, the wave–particle duality reduces to a different choice of boundary conditions on $\xi = 0$, $d \neq 0$ versus $d = 0$. This suggests the following conjecture: the mathematical description of experiments with photons require boundary conditions on *space-time manifolds.*

[1] O. Laporte and G. E. Uhlenbeck, Phys. Rev. **37**, 1380 (1931).
[2] P. Hillion and S. Quinnez, Int. J. Theor. Phys. **25**, 727 (1986).
[3] H. Bateman, *The Mathematical Analysis of Electrical and Optical Waves* (reprint Dover, New York, 1955).
[4] P. Hillion, J. Math. Phys. **29**, 1771 (1988).
[5] R. P. Boas, *Entire Functions* (Academic, New York, 1954).
[6] A. Sezginer, J. Appl. Phys. **57**, 678 (1985).
[7] P. Hillion, J. Appl. Phys. **60**, 298 (1986).
[8] J. N. Brittingham, J. Appl. Phys. **54**, 1179 (1983).
[9] P. Hillion, "Splash wave modes in homogeneous Maxwell's equations," to appear in J. Elec. Waves Appl.
[10] M. Born and E. Wolf, *Principles of Optics* (Pergamon, New York, 1965).
[11] O. Gron, Nuovo Cimento B **17**, 14 (1973).

The dynamics of a class of quantum mean-field theories

Pavel Bóna
Department of Theoretical Physics, Faculty of Mathematics and Physics, Comenius University, 842 15 Bratislava, Czechoslovakia

(Received 13 January 1988; accepted for publication 8 June 1988)

An infinite quantum system with correctly defined dynamics τ^Q as an automorphism group of a C^*-algebra $\mathscr{E}$ of observables is determined by any continuous unitary representation $U(G)$ of a connected Lie group G, as well as by an arbitrary differentiable real function Q on the dual space $\mathfrak{g}^*$ to the Lie algebra $\mathfrak{g}$ of G with the canonically defined Poisson flow φ^Q on $\mathfrak{g}^*$. For specific choices of Q and G, the system can be obtained as the thermodynamic limit of a net of finite lattice systems with the mean-field type interaction of Hepp and Lieb [Helv. Phys. Acta **46**, 573 (1973)]. A simple nontrivial model of this type is the quasispin BCS model of superconductivity in the strong coupling limit, or a corresponding model of the Josephson junction. A peculiar feature of the considered models is τ^Q noninvariance of the usually considered C^*-algebra $\mathscr{A}$ of quasilocal observables, as well as an important role of classical dynamics φ^Q of a set of macroscopic (intensive) observables in the description of τ^Q. The work is restricted here to norm-continuous representations $U(G)$, in which case $\mathscr{E}$ is isomorphic to the tensor product $\mathscr{A}\otimes\mathscr{N}$, where $\mathscr{N}$ is the commutative algebra of classical (intensive) observables of the considered infinite quantum system.

I. INTRODUCTION

Mean-field approximations to models in statistical mechanics were introduced at the beginning of our century[1] to obtain some description of phase transitions. Quantum models leading rigorously in the thermodynamic limit to quantum mean-field dynamics as a consequence of a given net of local Hamiltonians appeared in a description of superconductivity.[2] The simplest nontrivial model of the considered type is the strong coupling limit[3,4] of the quasispin formulation[5] of the BCS model. The correctness of the description of (the local perturbations of) equilibrium states by the linearized Hamiltonian introduced by Bogoliubov[6] was demonstrated by Haag[7] by using concepts of the algebraic formulation of quantum theory. The study of the role of the Bogoliubov–Haag Hamiltonian in the description of time evolution was initiated by Thirring *et al.*[8–10] It was shown in the framework of the C^*-algebraic formalism that the description of the time evolution is representation dependent,[10] and that it can be described in equilibrium states by an automorphism group of the weak closure of the corresponding Gel'fand–Naimark–Segal (GNS) representation of the C^*-algebra $\mathscr{A}$ of quasilocal observables.[11] Studies of the thermodynamic limits of Gibbs states were performed and the existence and structure of phase transitions was determined (see, e.g., Refs. 10, 12, and 13). These quadratic models were generalized to polynomial ones and, in this general case, classical time evolution of macroscopic (intensive) quantities was established by Hepp and Lieb[14] for the restriction to the subset of "classical states" [15] of the set of all states of the infinite system; this was possible without knowing the microscopic evolution of all quasilocal quantities of the infinite quantum system due to a specific way of taking the thermodynamic limit of time evolved intensive quantities. Let us note that the results obtained in Ref. 14 coincide with those derived from the microscopic evolution τ^Q introduced below.

Further investigation of the dynamics of the mean-field models was restricted mostly to the dynamics of local perturbations of limiting Gibbs states.[16–18] Some of the conclusions of these works were rather unexpected: in some states the time evolution does not satisfy the semigroup property.[17,18] In some attempts[18,19] to define time evolution in the considered (polynomial) models the "Schrödinger picture" was used. Rieckers with his collaborators[20–24] stressed the importance of classical observables as well as a specific role of "symmetry breakdown" in the correct definition of time evolution in the GNS representations of the considered states. They proved σ-weak continuity of the obtained time evolution automorphism group of the weak closures of the representations. The necessity of enlargement of quasilocal algebras and a role of their nontrivial center in descriptions of dynamics of systems with long range interactions was studied in Refs. 25 and 26.

We shall consider in this paper a general class of the mean-field models that are generalizations of the models introduced in Sec. 2 of Ref. 14 to nonpolynomial interactions. Our language will be that of the C^*-algebraic formulation of quantum theory.[27–29] An inspiring review of the history, methodological problems as well as possible perspectives of the formalism, and its applications to systems with many degrees of freedom is contained in Primas' book.[30] We shall investigate the thermodynamic limits of local time evolutions of the considered (polynomial) models as well as generalizations of the limiting time evolutions. Since the thermodynamic limits do not exist[8] in the strong topology in the algebra of quasilocal observables $\mathscr{A}$, we shall work in the framework of the larger algebra $\mathscr{A}^{**}$, and the limits will be considered in the σ-s^*-topology generated by a subset S_0 of the normal states on $\mathscr{A}^{**}$. The set S_0 is contained in the set of classical states,[14] which contains all the permutation invariant states[15] as well as all the states for which the above mentioned considerations and conclusions[18–24] were applied. The states $\omega\in S_0$ are specified by the existence of all the

 0022-2488/88/102223-13$02.50

relevant intensive observables obtained from their local forms by strong operator limits in the strong closures $\pi_\omega(\mathscr{A})''$ of the corresponding GNS representations π_ω. $S_\mathfrak{g}$ seems to be the maximal set of states for which the time evolution of all the considered models can be defined simultaneously in a unique "natural" way.[31] The resulting time evolution of the infinite quantum system can be described as a one-parameter group of the *-automorphisms of a C^*-algebra $\mathscr{C}$ contained in $\mathscr{A}^{**}$ and containing the usually considered quasilocal algebra $\mathscr{A}$. Hence any pathological looking features of the considered models appearing in previous works are either missing or naturally explained in the framework of our approach. Let P_G be the smallest projection in the center of $\mathscr{A}^{**}$ with the property $\omega(P_G) = 1$ for all $\omega \in S_\mathfrak{g}$. Then $\mathscr{C}$ is a proper subset of $P_G\mathscr{A}^{**}$, which allows us to prove better continuity properties of the time-evolution automorphism group then were stated for limiting evolutions of weak closures of some GNS representations of $\mathscr{A}$ in Refs. 18–24. A crucial role of a kinematical group G of the systems is revealed in our definition of the dynamics.

The sets of bounded observables and states of the considered systems (i.e., the "kinematics") together with a set of their natural symmetries are introduced in Sec. II. Any of the considered systems of Secs. II and III can be regarded as a union of a countable set of equal quantum mechanical subsystems. Each of these subsystems of a given composed infinite system is determined by a weakly continuous unitary representation $U(G)$ of a connected Lie group G in a separable Hilbert space H. We shall consider in this paper only the cases of norm-continuous $U(G)$ although the general case can be dealt with with the help of more technical approaches.[25] The usually considered[15,28] C^*-algebra $\mathscr{A}$ of quasilocal observables is the infinite tensor product of a countable number of copies of the algebra $\mathscr{L}(H)$ of all bounded operators on H defined as the C^*-inductive limit[32] of W^*-tensor products of any finite numbers of copies of $\mathscr{L}(H)$. The action $U(G)$ on H induces a naturally defined automorphism group $\sigma(G)$ of $\mathscr{A}$ with a canonical extension to an automorphism group of the double dual $\mathscr{A}^{**}$ of $\mathscr{A}$. The commutative C^*-algebra $\mathscr{N}$ of considered macroscopic (intensive) observables is generated by the simplest Cesaro means (in the strong topology of $P_G\mathscr{A}^{**}$) of copies of the generators of $U(G)$. The algebra $\mathscr{N}$ belongs to the center $\mathscr{Z}$ of $\mathscr{A}^{**}$ and it is isomorphic to the algebra $C(E)$ of all continuous complex-valued functions on an $\mathrm{Ad}^*(G)$-invariant subset E of the dual space $\mathfrak{g}^*$ to the Lie algebra $\mathfrak{g}$ of G. The isomorphism is given by a $\mathscr{Z}$ valued, $\sigma(G) - \mathrm{Ad}^*(G)$-equivariant projection valued measure $E_\mathfrak{g}$ on $\mathfrak{g}^*$ (with supp $E_\mathfrak{g} = E$) via the standard functional calculus: $f[\in C(E)] \mapsto E_\mathfrak{g}(f)$ $(\in \mathscr{Z})$. One has also $P_G = E_\mathfrak{g}(\mathfrak{g}^*)$. If we consider $\mathscr{A}$ canonically embedded into $\mathscr{A}^{**}$, then the C^*-algebra $\mathscr{C}$ of observables convenient for the description of dynamics in the considered models is generated in $P_G\mathscr{A}^{**}$ by the subalgebras $P_G\mathscr{A}$ and $\mathscr{N}$. $\mathscr{C}$ is isomorphic to the tensor product $\mathscr{A} \otimes \mathscr{N}$ [in the case of norm-continuous $U(G)$]. The algebras $\mathscr{C}$, $\mathscr{A}$, and $\mathscr{N}$ are invariant with respect to (w.r.t.) the action of $\sigma(G)$, e.g., $\sigma(g)(\mathscr{N}) \subset \mathscr{N}$ for all $g \in G$.

The set of differentiable functions on $\mathfrak{g}^*$ is endowed with a natural structure of Poisson brackets (also called the "Berezin brackets") coming from the Kirillov–Kostant symplectic structure[33,34] on orbits of the coadjoint representation $\mathrm{Ad}^*(G)$ in $\mathfrak{g}^*$. This $\mathrm{Ad}^*(G)$-invariant Poisson structure[35] associates with any differentiable function Q on $\mathfrak{g}^*$ the Hamiltonian vector field λ_Q and the corresponding flow φ^Q of Poisson morphisms on $\mathfrak{g}^*$. The φ^Q invariance of Ad^*-orbits allows us to represent the action of the group φ^Q on any element of $\mathfrak{g}^*$ by an action of the Ad^*-representation of G taken on values of a cocycle g_Q: $\mathbb{R} \times \mathfrak{g}^* \to G$ (cf. Sec. II B). This fact together with the G equivariance of $E_\mathfrak{g}$ as well as with the structure of the algebra of observables $\mathscr{C}$ leads to a natural possibility to define an automorphism group τ^Q of $\mathscr{C}$ with the help of the group $\sigma(G)$: τ^Q is a kind of transfer of the classical evolution φ^Q specified by the cocycle g_Q [if $\sigma(G)$ and $E_\mathfrak{g}$ are given]. In Sec. IV the above mentioned definition of τ^Q is described, and Sec. III is devoted to the determination of the time evolution with the help of local Hamiltonians corresponding to the choice of polynomial Q's. The proof of the identity of both the definitions of τ^Q is made by comparing the infinitesimal generators (i.e., the derivations[36]) of the corresponding automorphism groups.

The resulting picture of the behavior of the infinite quantum-mechanical system corresponding to the dynamics τ^Q agrees with the common image[1] of the mean-field theories: The time evolution of any finite subsystem can be described by a Schrödinger-type equation with a time-dependent Hamiltonian (depending on the "external" mean field consisting of the classically evolved intensive observables with values in $\mathfrak{g}^*$ as well as on a given initial state of the infinite system). The mean field is, however, an exact consequence of the internal structure of the interactions in the infinite system, and no external field is put into the considered models by hand.

Let us note eventually that we have not mentioned here some important works connecting the presently discussed mean-field dynamics with other interesting problems of physics, e.g., with the question of sources of irreversibility,[14,37] and with several differently posed questions of connection of quantum and classical mechanics, cf., e.g., Ref. 25 for some citations. A discussion of these problems supplemented with corresponding citations is postponed to later work.

II. KINEMATICS OF THE SYSTEM

We shall specify here the "kinematical structure" of the considered systems by specifying, as usual, the sets of "observables" and "states" as well as some rules for their interpretation. In the framework of the C^*-algebraic description of physical systems the set of all bounded observables consists of all self-adjoint elements of a given C^*-algebra $\mathscr{C}$, and the set of all states of the system is described either by the set $S(\mathscr{C})$ of all positive normalized functionals on $\mathscr{C}$ or by a conveniently chosen subset of $S(\mathscr{C})$, e.g., for a W^*-algebra $\mathscr{C}$ the set $S_*(\mathscr{C})$ of normal states on $\mathscr{C}$ might be in some situations sufficient to describe the relevant physical situations. Since the algebra $\mathscr{C}$ of any of the presently considered systems has the tensor product structure $\mathscr{A} \otimes \mathscr{N}$, we shall discuss in Sec. II C the relation of $S(\mathscr{C})$ to $S(\mathscr{A})$ and

$S(\mathcal{N})$. The standard rule of the interpretation of the numbers $\omega(c)$ [with $\omega\in S(\mathcal{C})$, $c\in C$] as expectations for the observables c in the states ω will be complemented here by a scheme of giving physical meaning to specific $c\in\mathcal{C}$. This will be done by choosing a distinguished symmetry group G of the system: The group G might be identified with a group of transformations of a set of measuring devices. The action of the group G on the considered infinite quantum system allows us to specify a classical subsystem of a distinguished set of intensive observables with a natural Poisson structure.[35] Consideration of this classical subsystem is useful for introduction of dynamics into the considered type of models.

A. The large quantal system

Let Π be an infinite countable set and let $\mathbf{\Pi}$ denote the set of all finite subsets of Π, $|J|:=\operatorname{card} J$ for $J\in\mathbf{\Pi}$. Let $U(G)$ be a norm-continuous unitary representation in a separable Hilbert space H (for the usually considered spin systems H is finite dimensional). Let $u_p: H\to H_p$ ($p\in\Pi$) be unitary mappings onto copies H_p of H, $U_p(g):=\pi_p(U(g))$ for all $g\in G$ and $p\in\Pi$ with $\pi_p(A):=u_pAu_p^{-1}$, $A\in\mathcal{L}(H):=$ bounded operators on H. Let $\mathcal{A}$ be the C^*-inductive limit (cf. 1.23.11 in Ref. 32) of the net of the W^*-algebra $\mathcal{A}^J$ ($J\in\mathbf{\Pi}$) given by the W^*-tensor products[32,38]

$$\mathcal{A}^J:=\bigotimes_{p\in J}\pi_p(\mathcal{L}(H)).$$

The C^*-algebra $\mathcal{A}$ is simple, cf. 2.6.20 of Ref. 28. Each $\mathcal{A}^J$ ($J\in\mathbf{\Pi}$) will be considered as a ("local") subalgebra of $\mathcal{A}$ in the canonical way. Then $\mathcal{A}$ becomes a quasilocal algebra.[28] Let $\sigma(G)$ be the range of the strongly continuous homomorphism σ of G into *-aut $\mathcal{A}:=$ the automorphism group of $\mathcal{A}$. The morphism σ is determined by the action of $\sigma(G)$ on finite linear combinations of the elements $x\in\mathcal{A}$ of the form

$$x:=\bigotimes_{p\in J}\pi_p(x_p),\quad x_p\in\mathcal{L}(H),\quad J\in\mathbf{\Pi},$$

by the formula

$$\sigma(g)(x):=\bigotimes_{p\in J}\pi_p(U(g)x_pU(g^{-1})),\quad g\in G. \tag{2.1}$$

The strong continuity of σ means that the functions $g\mapsto\sigma(g)(x)$ ($x\in\mathcal{A}$) are continuous in the norm topology of $\mathcal{A}$. Let $\mathfrak{g}$ be the Lie algebra of G and let $X_\beta=X_\beta^*\in\mathcal{L}(H)$ be the generators of U:

$$\exp(-itX_\beta):=U(\exp(t\beta)),\quad \beta\in\mathfrak{g}. \tag{2.2}$$

The continuity of σ is a consequence of the assumed boundedness of X_β ($\beta\in\mathfrak{g}$). The adjoint representation $\mathrm{Ad}(G)$ on $\mathfrak{g}$ is defined by the formula

$$\mathrm{Ad}(g)\beta:=\frac{d}{dt}\Big|_{t=0}[g\exp(t\beta)g^{-1}],\quad g\in G,\quad \beta\in\mathfrak{g}.$$

The generators X_β depend linearly on $\beta\in\mathfrak{g}$, and their transformation properties under $U(G)$ are

$$X_\beta\mapsto U(g)X_\beta U(g^{-1})=X_{\mathrm{Ad}(g)\beta}. \tag{2.3}$$

Let us also write $X(\beta)$ for X_β and let

$$X^J(\beta):=|J|X_{\beta J}:=\sum_{p\in J}\pi_p(X_\beta),\quad \beta\in\mathfrak{g},\quad J\in\mathbf{\Pi}. \tag{2.4}$$

The expressions (2.4) represent elements of $\mathcal{A}^J$ (hence of $\mathcal{A}$) in a canonical way. Then we have from (2.1) and (2.3)

$$\sigma(g)(X^J(\beta))=X^J(\mathrm{Ad}(g)\beta). \tag{2.5}$$

Let $S(\mathcal{A})$ be the set of all states (i.e., positive linear normalized functionals) on $\mathcal{A}$. For any $\omega\in S(\mathcal{A})$, let $[\pi_\omega, H_\omega, \Omega_\omega]$ be the corresponding GNS triplet (the technicalities on C^*-algebras and their representations can be found in Refs. 27–29, 32, 38–40). Let $[\pi_u, H_u]$ be the universal representation of $\mathcal{A}$, i.e., the orthogonal sum of all the π_ω [$\omega\in S(\mathcal{A})$]. The double commutant $\pi_u(\mathcal{A})''$ of $\pi_u(\mathcal{A})$ in $\mathcal{L}(H_u)$ is isomorphic as a Banach space to the double topological dual $\mathcal{A}^{**}$ of $\mathcal{A}$, hence $\mathcal{A}^{**}$ is canonically endowed with the structure of a von Neumann algebra. We shall identify $\mathcal{A}^{**}$ with $\pi_u(\mathcal{A})''$. Let $\mathcal{Z}:=\pi_u(\mathcal{A})''\cap\pi_u(\mathcal{A})'$ be the center of $\mathcal{A}^{**}$. Any state $\omega\in S(\mathcal{A})$ can be uniquely extended to an (equally denoted) normal state $\omega\in S_*(\mathcal{A}^{**})\subset S(\mathcal{A}^{**})$ and the corresponding W^*-representation π_ω of $\mathcal{A}^{**}$ is the unique normal extension of $\pi_\omega(\mathcal{A})$ to $\mathcal{A}^{**}$, cf. Ref. 32, Proposition 1.21.13. We can consider π_ω as a subrepresentation of π_u, and for the extensions we have $\pi_\omega(\mathcal{A}^{**})=\pi_\omega(\mathcal{A})''$. For any representation π of $\mathcal{A}$, there is a unique projection $c(\pi)=c(\pi)^*=c(\pi)^2\in\mathcal{Z}$ called the central cover of π, Ref. 39, paragraph 3.8.1, such that $c(\pi)\mathcal{A}^{**}$ is isomorphic to $\pi(\mathcal{A})''=\pi(\mathcal{A}^{**})$, and $\pi(c(\pi))=\pi(\mathbb{1})$ [$\mathbb{1}$ is the unit element $\pi_u(\mathrm{id}_{\mathcal{A}})$ of $\mathcal{A}^{**}$]. The algebra $\mathcal{A}$ is isomorphic to its universal representation $\pi_u(\mathcal{A})$ but, due to the simplicity of $\mathcal{A}$, it is isomorphic also to any nonzero representation $\pi(\mathcal{A})$. Hence $\mathcal{A}$ can be identified either with $\pi_u(\mathcal{A})$ or with any of its nonzero subrepresentations $c(\pi)\pi_u(\mathcal{A})$.

The "local Hamiltonians" for the description of dynamics τ^J of elements in the subalgebras $\mathcal{A}^J$ ($J\in\mathbf{\Pi}$) are expressed in the considered polynomial models[14] in terms of elements from (2.4). Hence the thermodynamic limit of dynamics can be expected to exist in (and only in) the subset of states $S_{\mathfrak{g}}\subset S(\mathcal{A})$ in which limits for $J\to\Pi$ of the nets $[X_{\beta J}: J\in\mathbf{\Pi}]$ for all $\beta\in\mathfrak{g}$ exist in some suitable sense (for general interactions of the considered type; for a specific interaction, the limiting dynamics can be defined in those states in which the limits do exist for all such β's for which the $X_{\beta J}$ enter into the local Hamiltonians). Let us note that the existence of norm limit of the net $[X_{\beta J}: J\in\mathbf{\Pi}]$ is excluded for X_β different from a scalar multiple of the identity in H: Since the norm limits of all commutators $[x, X_{\beta J}]$ ($x\in\mathcal{A}$) vanish and $\mathcal{A}$ is simple, any norm limit z_β of $X_{\beta J}$ would be a scalar multiple of identity of $\mathcal{A}$. This would imply

$$\lim_J\omega(X_{\beta J})=z_\beta,\quad \text{for all } \omega\in S(\mathcal{A}).$$

But the spectrum of X_β contains at least two distinct points; hence there are normalized vectors $\varphi_j\in H$ ($j=1,2$) such that $(\varphi_1, X_\beta\varphi_1)\neq(\varphi_2, X_\beta\varphi_2)$, and for the product states $\omega_{1,2}$:

$$\omega_j\Big(\bigotimes_{p\in J}\pi_p(x_p)\Big):=\prod_{p\in J}(\varphi_j, x_p\varphi_j),\quad x_p\in\mathcal{L}(H),\quad j=1,2,$$

one obtains the desired contradiction:

$$\lim_J\omega_1(X_{\beta J})\neq\lim_J\omega_2(X_{\beta J}).$$

However, it can be easily shown that strong operator limits

$$\text{s-lim}_J\, \pi_\omega(X_{\beta J}) \in \pi_\omega(\mathscr{A})'' \tag{2.6a}$$

exist for many states $\omega \in S(\mathscr{A})$, e.g., for all permutation invariant states, i.e., the states ω for which the values

$$\omega\Big(\bigotimes_{p\in J} \pi_{j(p)}(x_p)\Big), \quad \text{for any } J\in\Pi, \quad x_p \in \mathscr{L}(H),$$

and any bijection j: $\Pi \to \Pi$ are independent of j.[17] Examples of permutation-invariant states are $\omega_{1,2}$ defined above. There are states ω, on the other hand, for which the strong limit of $\pi_\omega(X_{\beta J})$ does not exist, e.g., the product states $\omega := \otimes_{p\in\Pi} \omega^p$, in which an infinite number of the restrictions ω^p to the local subalgebras $\mathscr{A}^p := \mathscr{L}(H_p)$ coincides with φ_1 [i.e., $\omega^p(\pi_p(x)) = (\varphi_1, x\,\varphi_1)$ for all $x\in\mathscr{L}(H)$], and also an infinite number of ω^q coincides with φ_2 defined above.

The existence of the limit in (2.6a) is equivalent[41] to the existence of that limit with ω replaced by a (quasi-) equivalent state,[39,40] i.e., by a state with the same central cover $c(\pi_\omega)$ of its GNS representation. After the identification of $x\in\mathscr{A}$ with $\pi_u(x)\in\mathscr{A}^{**}$, we make a natural identification of $\pi_\omega(x)$ with $xc(\pi_\omega) := \pi_u(x)c(\pi_\omega)$, and the existence of the limits in (2.6a) is equivalent to the existence of the limits

$$\text{s-lim}_J\, X_{\beta J} P \in \mathscr{Z}, \quad \beta\in\mathfrak{g}, \tag{2.6b}$$

with $P = c(\pi_\omega)$. The limits in (2.6b), if they exist, belong to the center $\mathscr{Z}$ of $\mathscr{A}^{**}$ since the nets $J \mapsto [x, X_{\beta J}]$ converge to zero in norm for any $x\in\mathscr{A}$, where $[x,y] := xy - yx$ is the commutator. The existence of the limits in (2.6b) for some projection $P\in\mathscr{Z}$ implies the existence of the limits in (2.6a) for all such ω, for which $c(\pi_\omega) \leqslant P$. Let P_G be the largest of all such projections P for which the limits in (2.6b) exist. Let

$$X_{\beta\Pi} := \text{s-lim}_J\, X_{\beta J} P_G . \tag{2.7}$$

Due to the linear dependence of $X_{\beta\Pi}$ on $\beta\in\mathfrak{g}$, the mapping

$$\beta \mapsto \exp(i X_{\beta\Pi}) \in P_G\mathscr{Z} \tag{2.8}$$

is a norm-continuous unitary representation of the additive group $\mathfrak{g}$ in the Hilbert space $P_G H_u$. According to the Stone–Naimark–Ambrose–Godement (SNAG) theorem (see Ref. 42, § 140, p. 375, or Ref. 43, Theorem VIII. 12, or Ref. 28, Sec. 3.2.3), there is a unique projection valued measure $E_{\mathfrak{g}}$ on $\mathfrak{g}^*$ [the dual of the linear space $\mathfrak{g}$ can be identified with the dual group $\hat{\mathfrak{g}}$ of the additive group $\mathfrak{g}$ by the association with any $F\in\mathfrak{g}^*$, the character χ_F: $\beta \mapsto \chi_F(\beta) := \exp[iF(\beta)]$] with values in $P_G\mathscr{Z}$ such that

$$X_{\beta\Pi} = \int_{\mathfrak{g}^*} F(\beta)\, E_{\mathfrak{g}}(dF), \quad \beta\in\mathfrak{g}. \tag{2.9}$$

Let us denote $E := \operatorname{supp} E_{\mathfrak{g}}$ the minimal closed set $E\subset\mathfrak{g}^*$ such that $E_{\mathfrak{g}}(E) = P_G$ [$= E_{\mathfrak{g}}(\mathfrak{g}^*)$]. The set E is compact due to the boundedness of X_β's. The set $C(E)$ of complex-valued continuous functions on $\operatorname{supp} E_{\mathfrak{g}}$ is a commutative unital C^*-algebra generated (by the Weierstrass theorem) by the functions f_β ($\beta\in\mathfrak{g}$), where

$$f_\beta(F) := F(\beta), \quad F\in\mathfrak{g}^*, \tag{2.10}$$

are linear functions on $\mathfrak{g}^*$. Let $E_{\mathfrak{g}}$ denote also the mapping from $C(E)$ into $P_G\mathscr{Z}$ given by

$$E_{\mathfrak{g}} : f(\in C(E)) \mapsto E_{\mathfrak{g}}(f) := \int f(F)\, E_{\mathfrak{g}}(dF). \tag{2.11}$$

A direct consequence of the standard functional calculus of normal operators[42,43] is the following lemma.

Lemma 2.1: The mapping $E_{\mathfrak{g}}$ from (2.11) is a *-isomorphism of $C(E)$ onto the C^*-subalgebra $\mathscr{N}$ of $\mathscr{Z}$ generated by the elements

$$X_{\beta\Pi} = E_{\mathfrak{g}}(f_\beta), \quad \beta\in\mathfrak{g}. \tag{2.12}$$

Now we shall introduce two C^*-algebras suitable for description of dynamics of mean-field theories, and then we shall prove isomorphisms of these algebras.

Definition 2.2: (i) Let π_G denote the canonical embedding of the simple C^*-algebra $\mathscr{A}$ into $P_G\mathscr{A}^{**}$:

$$\pi_G : x(\in\mathscr{A}) \mapsto \pi_G(x) := \pi_u(x) P_G . \tag{2.13}$$

Let $\mathscr{C}$ be the C^*-subalgebra of $P_G\mathscr{A}^{**}$ generated by $\pi_G(\mathscr{A})$ and by $\mathscr{N} := E_{\mathfrak{g}}(C(E))$. Let us denote by $\mathscr{C}^J$ ($J\in\Pi$) the C^*-subalgebra of $\mathscr{C}$ generated by $\pi_G(\mathscr{A}^J)$ and $\mathscr{N}$. $\mathscr{C}$ is called the *algebra of observables for $\sigma(G)$-mean-field theories*, and $\mathscr{N}$ is the *algebra of $\sigma(G)$-intensive quantities.*

(ii) Let $\mathscr{A}\otimes\mathscr{N}$ be the (unique) C^*-tensor product; it is isomorphic[32] to the C^*-algebra $C(\operatorname{supp} E_{\mathfrak{g}}, \mathscr{A}) = C(E, \mathscr{A})$ of $\mathscr{A}$-valued norm-continuous functions $\hat{f}$ on E with the norm $\|\hat{f}\| := \operatorname{supp}[\|\hat{f}(F)\| : F\in E]$ and with the pointwise algebraic operations.

The algebras $\mathscr{N}$ and $\mathscr{A}$ are naturally embedded into $C(E, \mathscr{A})$ by

$$E_{\mathfrak{g}}(f)(\in\mathscr{N}) \mapsto \hat{f}, \quad \text{with } \hat{f}(F) := f(F)\,\mathrm{id}_{\mathscr{A}} \tag{2.14}$$

for all $f\in C(E)$, and

$$x(\in\mathscr{A}) \mapsto \hat{f}, \quad \text{with } \hat{f}(F) := x \ \text{(for all } F\in E). \tag{2.15}$$

The assertions of the next two lemmas are taken from Ref. 38, Proposition IV.4.7 and Lemma IV.4.18, resp. Exercise IV.4.2.

Lemma 2.3: Let $\mathscr{A}$ be a unital C^*-algebra, $\mathscr{N}$ a commutative C^*-algebra, and let τ_A: $\mathscr{A}\to\mathscr{C}$ and τ_N: $\mathscr{N}\to\mathscr{C}$ be homomorphisms into a C^*-algebra $\mathscr{C}$ with commuting ranges. Then there is a unique homomorphism τ of the C^*-tensor product $\mathscr{A}\otimes\mathscr{N}$ into $\mathscr{C}$ such that

$$\tau(x\otimes z) = \tau_A(x)\tau_N(z), \quad x\in\mathscr{A}, \quad z\in\mathscr{N},$$

and the image $\tau(\mathscr{A}\otimes\mathscr{N})$ is the C^*-subalgebra of $\mathscr{C}$ generated by $\tau_A(\mathscr{A})$ and $\tau_N(\mathscr{N})$.

Lemma 2.4: Let $\mathscr{C}$ be a unital C^*-algebra generated by commuting unital C^*-algebras $\mathscr{A}$ and $\mathscr{N}$, $\mathscr{N}$ commutative. If $xz = 0$ implies $\|x\|\,\|z\| = 0$ for any $x\in\mathscr{A}$, $z\in\mathscr{N}$, then the homomorphism

$$\sum_j x_j\otimes z_j (\in\mathscr{A}\otimes\mathscr{N}) \mapsto \sum_j x_j z_j \in \mathscr{C}$$

can be extended to an isomorphism of $\mathscr{A}\otimes\mathscr{N}$ onto $\mathscr{C}$.

With the notation of Definition 2.2, we then obtain the following proposition.

Proposition 2.5: There is a unique *-isomorphism $E_{\mathfrak{g}}$ of $C(E, \mathscr{A})$ onto $\mathscr{C}$ extending the mappings $E_{\mathfrak{g}}$ of (2.11), and π_G of (2.13). It can be expressed by the formula

$$E_{\mathfrak{g}}(\hat{f}) =: \int \hat{f}(F)\, E_{\mathfrak{g}}(dF), \quad \hat{f}\in C(E, \mathscr{A}). \tag{2.16}$$

Note: The formula (2.16) is used here to define the inte-

gral. It can be defined, however, independently and equivalently by a limit in a weaker-than-norm-topology of a sequence of integrals of step-function approximations to $\hat{f}$ [cf. Proposition 6.3.6 of Ref. 25(a)].

Proof: It suffices to show that for any nonzero $f\in C(E)$, the equality $\pi_G(x)E_{\mathfrak{g}}(f)=0$ implies $x=0$. Let $f(F)>\frac{1}{2}f(F_0)>0$ for all $F\in B_0\subset E$, with $F_0\in B_0$ and $E_{\mathfrak{g}}(B_0)\neq 0$. Then

$$0\leqslant \pi_G(x^*x)(E_{\mathfrak{g}}(f)-\tfrac{1}{2}f(F_0))E_{\mathfrak{g}}(B_0)$$
$$=-\tfrac{1}{2}f(F_0)\pi_G(x^*x)E_{\mathfrak{g}}(B_0),$$

and this implies $\pi_G(x)E_{\mathfrak{g}}(B_0)=0$. The mapping $x\to\pi_G(x)E_{\mathfrak{g}}(B_0)$ $(x\in\mathscr{A})$ is a nondegenerate representation of the simple C^*-algebra $\mathscr{A}$, hence its value is zero only for $x=0$. Q.E.D.

The isomorphism $E_{\mathfrak{g}}$ maps $C(E,\mathscr{A}^J)$ onto $\mathscr{C}^J$ $(J\in\Pi)$, endowing the algebra $\mathscr{C}$ of observables with a quasilocal structure (Ref. 28, Definition 2.6.3) coming from that of $\mathscr{A}$. In the case of $\mathscr{C}$, however, the "local algebras" $\mathscr{C}^J$ contain also intensive (i.e., "global") observables $\mathscr{N}$. Let $\sigma(G)$ also denote the extensions of the mappings from (2.1) to the corresponding *-automorphism group of $\mathscr{A}^{**}$. From (2.5) and (2.7) one has $\sigma(G)$ invariance of P_G as well as the relation

$$\sigma(g)(X_{\beta\Pi})=X_{\mathrm{Ad}(g)\beta\Pi},\quad \beta\in\mathfrak{g},\quad g\in G. \tag{2.17a}$$

This can be rewritten in the form of the G equivariance of the projection-valued measure $E_{\mathfrak{g}}$ on Borel sets $B\subset\mathfrak{g}^*$,

$$\sigma(g)(E_{\mathfrak{g}}(B))=E_{\mathfrak{g}}(\mathrm{Ad}^*(g)B),\quad g\in G. \tag{2.17b}$$

We can see from this that $\mathscr{C}$ and $\mathscr{N}$ are $\sigma(G)$-invariant subalgebras of $\mathscr{A}^{**}$. Hence we can deal with $\mathscr{N}$ as with a "kinematically independent" subsystem algebra of a classical subsystem of the large quantal system specified by the action $\sigma(G)$ of the chosen group G.

B. The classical subsystem of $\sigma(G)$-intensive quantities

We shall introduce here the canonical structure of Poisson brackets on the algebra $\mathscr{N}$, which is $\mathrm{Ad}^*(G)$ invariant. This structure will be used to determine the classical dynamics on $\mathscr{N}$ corresponding to any given differentiable $Q\in C(E)$. Then we shall define a cocycle g_Q which will be used to transfer the classical dynamics to an evolution of the large quantal system in Sec. IV. The coordinate free differential calculus on manifolds of Cartan[33,44] will be used for brevity and clarity of the expression.

Let $[\beta,\chi]$ be the Lie bracket of elements β and χ of the Lie algebra $\mathfrak{g}$ of the group G. The tangent space $T_F\mathfrak{g}^*$ to the dual $\mathfrak{g}^*$ of $\mathfrak{g}$ at the point $F\in\mathfrak{g}^*$ will be identified with the linear space $\mathfrak{g}^*$ itself by using the identity mapping on $\mathfrak{g}^*$ as a chart. Then the cotangent space $T_F^*\mathfrak{g}^*$ can be identified with the Lie algebra $\mathfrak{g}=\mathfrak{g}^{**}$; we shall also transfer the Lie-bracket structure from $\mathfrak{g}$ to $T_F^*\mathfrak{g}^*$ by this identification. Let $d_Ff\in T_F^*\mathfrak{g}^*=\mathfrak{g}$ be the value at F of the exterior differential df of $f\in C^\infty(\mathfrak{g}^*,\mathbb{R})$. A Poisson structure[35] on $\mathfrak{g}^*$ is given by the Poisson bracket

$$[f,h](F):=-F([d_Ff,d_Fh]),\quad F\in\mathfrak{g}^*,\quad f,h\in C^\infty(\mathfrak{g}^*,\mathbb{R}). \tag{2.18}$$

This bracket $[f,h]\in C^\infty(\mathfrak{g}^*,\mathbb{R})$ satisfies all the properties of the Poisson bracket on symplectic manifolds[44] (i.e., bilinearity, antisymmetry, derivation property, and Jacobi identity) except nondegeneracy. It associates a unique Hamiltonian vector field λ_Q on $\mathfrak{g}^*$ with any $Q\in C^\infty(\mathfrak{g}^*,\mathbb{R})$ by the formula

$$df(\lambda_Q):=[Q,f],\quad f\in C^\infty(\mathfrak{g}^*,\mathbb{R}). \tag{2.19}$$

Let φ be a diffeomorphism on $\mathfrak{g}^*$, and let $\varphi^*: C(\mathfrak{g}^*)\to C(\mathfrak{g}^*)$ be the "pullback" of φ, i.e., $\varphi^*f:=f\circ\varphi$. Then φ is called a Poisson automorphism iff

$$\varphi^*[f,h]=[\varphi^*f,\varphi^*h],\quad f,h\in C^\infty(\mathfrak{g}^*,\mathbb{R}).$$

Any Hamiltonian vector field generates a family of local Poisson automorphisms φ_t^Q determined by classical Hamilton equations

$$\frac{d}{dt}f_t(F)=[Q,f_t](F),\quad F\in\mathfrak{g}^*,\quad t\in r_F,\quad f\in C^\infty(\mathfrak{g}^*), \tag{2.20}$$

where $f_t:=\varphi_t^{Q^*}f$, and r_F is an open neighborhood of zero on the real axis. The vector field λ_Q is complete iff $r_F=\mathbb{R}$ for all $F\in\mathfrak{g}^*$ iff φ^Q forms a one-parameter group of Poisson automorphisms: $\varphi_{t+s}^Q=\varphi_t^Q\circ\varphi_s^Q$ for all $s,t\in\mathbb{R}$. For any compact group G each λ_Q is complete. In any case we shall choose Q such that λ_Q will be complete.

The algebra $C(E)$ is generated by restrictions to E of the linear functions f_β $(\beta\in\mathfrak{g})$ from (2.10). The Poisson bracket of such functions is (since $d_Ff_\beta=\beta$)

$$[f_\beta,f_\chi](F)=-F([\beta,\chi])=-f_{[\beta,\chi]}(F),\quad \beta,\chi\in\mathfrak{g}. \tag{2.21}$$

The adjoint action $\mathrm{Ad}(G)$ is a Lie algebra automorphism group[45] of $\mathfrak{g}$, i.e., $\mathrm{Ad}(g)$ $(g\in G)$ form a linear representation with

$$\mathrm{Ad}(g)[\beta,\chi]=[\mathrm{Ad}(g)\beta,\mathrm{Ad}(g)\chi],\quad g\in G,\quad \beta,\chi\in\mathfrak{g}. \tag{2.22}$$

This implies that the coadjoint action $\mathrm{Ad}^*(G)$ on $\mathfrak{g}^*$,

$$\mathrm{Ad}^*(g)F(\beta):=F(\mathrm{Ad}(g^{-1})\beta),\quad \beta\in\mathfrak{g},\quad g\in G, \tag{2.23}$$

consists of a group of Poisson automorphisms leaving the orbits

$$\mathrm{Ad}^*(G)F:=[\mathrm{Ad}^*(g)F:\ g\in G]\subset\mathfrak{g}^* \tag{2.24}$$

invariant. The restriction of the Poisson structure (2.18) to any nondegenerate $\mathrm{Ad}^*(G)$ orbit converts this orbit to a symplectic manifold.[33,34] Any Hamiltonian flow φ^Q leaves the Ad*-orbits invariant.[35] Due to the G equivariance (2.17b) of $E_{\mathfrak{g}}$, the set $E=\mathrm{supp}\,E_{\mathfrak{g}}$ is also left invariant by any φ^Q. Hence φ^{Q^*} is an automorphism group of $C(E)$ that determines a unique *-automorphism group of $\mathscr{N}=E_{\mathfrak{g}}(C(E))$.

Let us now introduce the cocycle g_Q. The φ^Q invariance of any Ad* orbit implies the existence of a function $g_Q:\mathbb{R}\times\mathfrak{g}^*\to G$ such that

$$\varphi_t^Q(F)=\mathrm{Ad}^*(g_Q(t,F))F,\quad t\in\mathbb{R},\quad F\in\mathfrak{g}^*. \tag{2.25}$$

We shall look for differentiable solutions of (2.25) with the "cocycle property"[46]

$$g_Q(s,\varphi_t^Q(F))g_Q(t,F)=g_Q(s+t,F),\quad g_Q(0,F)=e, \tag{2.26}$$

for all $s,t\in\mathbb{R}$ and all $F\in\mathfrak{g}^*$, with $e:=$ the identity of G. Let

$$\beta_F^Q := \left.\frac{d}{dt}\right|_{t=0} g_Q(t,F), \quad F\in\mathfrak{g}^*. \tag{2.27}$$

By differentiation of (2.25) in $t=0$, and with the help of the definition of the Ad(G)-representation as well of the commutator in $\mathfrak{g}$,[33,45] we then obtain

$$F([\beta_F^Q - d_F Q,\chi]) = 0, \quad \text{for all } \chi\in\mathfrak{g}, \quad F\in\mathfrak{g}^*. \tag{2.28}$$

This implies that a necessary condition for the validity of (2.25) is the fulfillment of the relation

$$\beta_F^Q = d_F Q + \beta_F^0, \quad F\in\mathfrak{g}^*, \tag{2.29}$$

where $\beta_F^0\in\mathfrak{g}_F$ is some of the generators of one-parameter subgroups of the stationary subgroup $G_F\subset G$ for the point F at Ad*-representation: $\mathrm{Ad}^*(\exp(t\beta_F^0))F = F$ ($t\in\mathbb{R}$), and the dependence $F\mapsto\beta_F^0$ is differentiable. By differentiation of (2.26) one gets the differential equation on the group manifold for g_Q:

$$\frac{d}{dt} g_Q(t,F) = T_e(R_{g_Q(t,F)})\,\beta_{F_t}^Q, \quad g_Q(0,F) = e, \tag{2.30}$$

for all $F\in\mathfrak{g}^*$, $t\in\mathbb{R}$; here $F_t := \varphi_t^Q F$, R_G is the right action of G onto itself: $R_g(h) := hg(g,h\in G)$, and $T_e(f)$ maps $\mathfrak{g} = T_e G$ into $T_{f(e)}G$ for any differentiable mapping f: $G\to G$ by

$$T_e(f)\,\beta := \left.\frac{d}{dt}\right|_{t=0} f(\exp(t\beta)). \tag{2.31}$$

Equation (2.30) is a finite-dimensional ordinary differential equation for g_Q. The uniqueness of the solution of (2.30),[44,47] with β_F^Q from (2.29), together with the uniqueness of the solution φ^Q of (2.20), prove the fulfillment of (2.25) as well as of (2.26) by the solution g_Q of (2.30). Let us note that (2.30) can be rewritten [equivalently for faithful $U(G)$] with the help of a unitary representation $U(G)$ as a linear (time-dependent Schrödinger) equation for the unitary family $U(g_Q(t,F))$, cf. the notations (2.2) and (2.4):

$$i\frac{d}{dt} U(g_Q(t,F)) = X(d_{F_t}Q + \beta_{F_t}^0)\,U(g_Q(t,F)), \tag{2.32}$$

with the initial condition $U(g_G(0,F)) = U(e)$. We shall see in the following sections that the thermodynamic limits of the local evolutions by the Hepp–Lieb Hamiltonians correspond to the choice $\beta^0 = 0$.

C. States and modifications

The algebra of observables $\mathscr{C}$ was defined in Definition 2.2 with the help of the limits (2.7) existing in the subspace of H_u determined by the projection P_G. We can distinguish two disjoint subsets of states on the C^*-algebra $\mathscr{A}$: The subset $S_\mathfrak{g}$ represented by nonzero vectors lying in $P_G H_u$, and the subset $S_\mathfrak{g}^\ominus$ represented by vectors in H_u orthogonal to $P_G H_u$. The restriction to $\mathscr{C}\subset P_G\mathscr{A}^{**}$ of the unique normal extension of $\omega\in S_\mathfrak{g}$ to a state on $\mathscr{A}^{**}$ is a canonically defined (and equally denoted) state $\omega\in S(\mathscr{C})$. The same procedure performed with any $\omega\in S_\mathfrak{g}^\ominus$ gives the zero function on $\mathscr{C}$: $\omega(P_G) = 0$. Let us denote by $S_\mathfrak{g}$ also the subset of states on $\mathscr{C}$ obtained from the corresponding states on $\mathscr{A}$ by the above mentioned procedure; the states $\omega\in S_\mathfrak{g}\subset S(\mathscr{C})$ will be called the *π_G-normal states* on $\mathscr{C}$.[28] It will be seen in the next section that the thermodynamic limit of an arbitrary polynomial mean-filed dynamics can be "naturally" defined in π_G-normal states only: The s^*-topology used in (3.11) and (3.13) is determined by the seminorms (3.5) defined by states $\omega\in S_\mathfrak{g}$. Any state $\omega\in S(\mathscr{A})$ can be extended, however, to a state on the C^*-algebra $\mathscr{C}$, since $\mathscr{A}$ is considered to be a C^*-subalgebra of $\mathscr{C}$ in a natural manner, cf. Definition 2.2. Although the resulting time evolution τ^Q obtained in Sec. III by a thermodynamic limit is an automorphism group of $\mathscr{C}$, it might look "unnatural" to define by $\omega_t(y) := \omega(\tau_t^Q(y))$ the evolution in the states $\omega\in S(\mathscr{C})$ obtained by an extension to $\mathscr{C}$ of those states on $\mathscr{A}$ for which the limits corresponding to (3.11) and (3.13) in the GNS representations do not exist. One of the most disturbing aspects of this unnaturality of the definition of time evolution in "improper states" is the apparent possibility of existence of such states $\omega\in S(\mathscr{C})$, in which the evolution of intensive observables $\omega_t(E_\mathfrak{g}(f_\beta))$ is not approximated by the nets $\omega_t(X_{\beta J})$ ($J\in\mathbf{II}$) of time evolved local approximations of $X_{\beta\Pi} = E_\mathfrak{g}(f_\beta)$, cf. (2.12) and (2.7). It will be shown here how to overcome this difficulty by a choice of "proper extensions" of states on $\mathscr{A}$ (including those in $S_\mathfrak{g}^\ominus$) to states on $\mathscr{C}$. We shall restrict our attention to pure states, since the w^*-limits of them (resp. of convex combinations of them) constitute the sets of all states. First we describe the structure of pure states on $\mathscr{C}$.

Proposition 2.6: Let $\mathscr{C}$ be the C^*-algebra generated by its C^*-subalgebras $\mathscr{A}$ and $\mathscr{N}$, let $\mathscr{N}$ be contained in the center of $\mathscr{C}$, and let $\mathscr{C}$ be isomorphic to the C^*-tensor product $\mathscr{A}\otimes\mathscr{N}$. Then the state $\omega\in S(\mathscr{C})$ is pure iff

$$\omega(xz) = \omega_{|\mathscr{A}}(x)\omega_{|\mathscr{N}}(z), \quad \text{for all } x\in\mathscr{A}, \quad z\in\mathscr{N}, \tag{2.33}$$

where the restrictions $\omega_{|\mathscr{A}}$ (resp. $\omega_{|\mathscr{N}}$) of ω to the subalgebra $\mathscr{A}$ (resp. to $\mathscr{N}$) are both pure. If we write $\mathscr{N} = E_\mathfrak{g}(C(E))$ for a Hausdorff compact E, cf. Lemma 2.1, then any pure state $\omega_{|\mathscr{N}}\in S(\mathscr{N})$ is of the form

$$\omega_{|\mathscr{N}}(E_\mathfrak{g}(f)) = f(F_\omega), \quad f\in C(E), \tag{2.34}$$

where $F_\omega\in E$ is a fixed point determined by the pure $\omega\in S(\mathscr{C})$ uniquely.

Proof: The first assertion is an immediate consequence of Theorem 4.4 and Lemma 4.11 of Ref. 38, Chap. IV. The second assertion is a simple consequence of the Riesz–Markov theorem (cf. Theorem IV.17 of Ref. 43): The pure states on the commutative C^*-algebra $C(E)$ are described by the Dirac measures $\delta_F: f\mapsto\delta_F(f) := f(F)$, $f\in C(E)$, on the compact E. This gives (2.34). Q.E.D.

Identification of $\mathscr{C}$ with $\mathscr{A}\otimes\mathscr{N}$ allows us to rewrite (2.33) in the form $\omega = \omega_\mathscr{A}\otimes\omega_\mathscr{N}$. Hence to obtain a pure extension $\omega\in S(\mathscr{C})$ of $\omega_{|\mathscr{A}}\in S(\mathscr{A})$, one has to choose an arbitrary $F_\omega\in E$, and the $\omega_{|\mathscr{N}}$ corresponding to F_ω by (2.34). For $\omega_{|\mathscr{A}}\in S_\mathfrak{g}$, there is, however, a unique "physically natural" choice of $\omega_{|\mathscr{N}}$ given by

$$\omega_{|\mathscr{N}}(E_\mathfrak{g}(f_\beta)) := \lim_J \omega_{|\mathscr{A}}(X_{\beta J}), \quad \beta\in\mathfrak{g}. \tag{2.35}$$

The formulas (2.33) and (2.35) determine the π_G-normal extension of $\omega_{|\mathscr{A}}$, since pure states on commutative C^*-algebras are characters. We intend to define natural extensions of pure states $\omega_{|\mathscr{A}}\in S_\mathfrak{g}^\ominus$ given formally by (2.33) and (2.35), only by reinterpreting the circuit in (2.35). Before proceeding in this way, let us mention possible modifications of the interpretation of the formalism presented in this paper.

Remark 2.7: The essential tool in the definition of classi-

cal observables $E_{\mathfrak{g}}(f_\beta) = X_{\beta\Pi}$ as well as of the algebra $\mathscr{C}$, and of the time evolution τ^Q (cf. Sec. III), is the existence of the limits in (2.6) and (2.7). Those are limits of the nets $[X_{\beta J}: J\in\mathbf{\Pi}]$, where $\mathbf{\Pi}$ was taken to be the set of *all* finite subsets of the countable set Π directed by the set inclusion: $J<K$ iff $J\subset K$, $J,K\in\mathbf{\Pi}$. One can obtain a nontrival reinterpretation of the whole formalism by taking $\mathbf{\Pi}$ to be a specific directed subset of the naturally ordered set of all finite subsets of Π. A choice of such a subset $\mathbf{\Pi}$ might be connected with some structure by which the set Π could be endowed. Let, e.g., the set Π be endowed with the structure of the m-dimensional lattice Z^m; then a natural choice of $\mathbf{\Pi}$ is the set of cubes in Z^m centered at the origin. The limits in (2.6a) then exist for a larger set $S_{\mathfrak{g}}$ of states $\omega\in S(\mathscr{A})$ than before, if J is taken there from such a conveniently chosen subset $\mathbf{\Pi}$ of the set of all finite subsets of Π; the projection P_G in (2.7) would be then larger than before. Let, e.g., $\Pi := Z$, and let $\mathbf{\Pi} := [(-m,-m+1,...,m-2,m-1,m): m\in Z_+]$; then any product state $\omega := \otimes_{p\in Z}\omega_p$ on $\mathscr{A}$ $[\omega_p\in S_*(\mathscr{L}(H_p))]$ with $\omega_{p+k}=\omega_p$ for all $p\in Z$ (here H_{p+k} is naturally identified with H_p via the unitary mapping $u_p u_{p+k}^{-1}$) belongs to the (newly defined) set $S_{\mathfrak{g}}$; hence it is represented by a vector in $P_G H_u$ (with the new P_G). Such periodic states with minimal period k larger than 1 did not belong to $S_{\mathfrak{g}}$ in the previous case of the limits (2.6) taken for the next $X_{\beta J}$ indexed by all finite subsets J of Π. All the general formalism and results of this paper remain unchanged after a redefinition of the set $\mathbf{\Pi}$ (and the corresponding redefinitions of the limits of the nets with indices $J\in\mathbf{\Pi}$) in the above mentioned sense. The only important change will be an increase of the projection P_G from (2.7); hence an enlarged domain of applicability of τ^Q in a physically natural way.

Remark 2.8: Let us mention another possible reinterpretation of limits of the nets indexed by $J\in\mathbf{\Pi}$ that could further enlarge the projection $P_G\in Z$. Let $\mathscr{I}$ be an arbitrary directed set and let $J:\mathscr{I}\to\mathbf{\Pi}$ be such a mapping that to any $K\in\mathbf{\Pi}$ there is an $i_K\in\mathscr{I}$ such that $i>i_K$ implies $J(i)\supset K$. Then the net $[X_{\beta J(i)}: i\in\mathscr{I}]$ is a subnet of the net $[X_{\beta J}: J\in\mathbf{\Pi}]$. One could require then only the existence of the limits

$$\text{s-}\lim_{i\in\mathscr{I}} X_{\beta J(i)}P\in\mathscr{Z}, \quad \text{for all } \beta\in\mathfrak{g} \tag{2.36}$$

instead of (2.6b), and define P_G as the least upper bound in $\mathscr{Z}$ of all the projections P in (2.36). This kind of limit can be used also in Sec. III in taking the thermodynamic limit of $\tau_t^{J(i)}$ defining the evolution τ^Q.

We shall now proceed to definitions of physically natural extensions of arbitrary pure states $\omega_{\mathscr{A}}\in S(\mathscr{A})$ to states on $\mathscr{C}$.

Proposition 2.9: There is a subnet $[X_{\beta J(i)}: i\in\mathscr{I}, \beta\in\mathfrak{g}]$ [$\mathscr{I}$ is a directed set and $J: i(\in\mathscr{I})\mapsto J(i)$ is as in Remark 2.8] of the net $[X_{\beta J}: J\in\mathbf{\Pi}, \beta\in\mathfrak{g}]$ such that the formula

$$\omega_{f}(E_{\mathfrak{g}}(f_\beta)) := \lim_{i\in\mathscr{I}} \omega_{\mathscr{A}}(X_{\beta J(i)}), \quad \beta\in\mathfrak{g}, \tag{2.37}$$

determines a unique pure state $\omega_{\mathscr{N}}$ on $\mathscr{N} := E_{\mathfrak{g}}(C(E))$, for an arbitrary pure state $\omega_{\mathscr{A}}$ on $\mathscr{A}$. The pure state $\omega\in S(\mathscr{C})$ determined from these $\omega_{\mathscr{A}}$ and the corresponding $\omega_{\mathscr{N}}$ by (2.33) satisfies the relation

$$\omega_{\mathscr{N}}(E_{\mathfrak{g}}(\varphi_t^{Q*}f_\beta)) = \lim_{i\in\mathscr{I}} \omega(\tau_t^Q(X_{\beta J(i)})), \quad \beta\in\mathfrak{g} \quad t\in\mathbb{R}; \tag{2.38}$$

here φ^Q is determined in (2.20), and $\tau_t^Q\in$ *-aut $\mathscr{C}$ is determined by (4.2) and (4.6) [in the present case $\hat{f}(F) := X_{\beta J(i)}$ for all $F\in E$].

Proof: For any fixed $\beta\in\mathfrak{g}$, the net $[X_{\beta J}: J\in\mathbf{\Pi}]$ is a uniformly bounded net in $\mathscr{A}^{**}$ (by the canonical inclusion of $\mathscr{A}$ into $\mathscr{A}^{**}$). Since closed balls in $\mathscr{A}^{**}$ are compact in the σ-weak topology [i.e., in $\sigma(\mathscr{A}^{**}, \mathscr{A}^*)$ topology] by the Banach–Alaoglu theorem (cf. Theorem IV.2 of Ref. 43), and the association $\beta\mapsto X_{\beta J}$ is linear, there is a subnet $[X_{\beta J(i)}: i\in\mathscr{I}, \beta\in\mathfrak{g}]$ of $[X_{\beta J}]$ convergent in w^*-topology of $\mathscr{A}^{**}$,

$$w^*\text{-}\lim_{i\in\mathscr{I}} X_{\beta J(i)} =: X_{\beta\mathscr{I}}\in\mathscr{Z}. \tag{2.39}$$

The belonging of the limits $X_{\beta\mathscr{I}}$ to the center $\mathscr{Z}$ of $\mathscr{A}^{**}$ follows as in (2.6b). Since $\omega_{\mathscr{A}}$ is pure, the right-hand side of (2.37) determines a unique element $F_\omega\in\mathfrak{g}^*$ which in turn determines the pure state ω_{f}:

$$\omega_{f}(E_{\mathfrak{g}}(f_\beta)) := F_\omega(\beta) := \omega_{\mathscr{A}}(X_{\beta\mathscr{I}}), \quad \beta\in\mathfrak{g}. \tag{2.40}$$

Let us calculate the right-hand side of (2.38) for ω given by (2.33). The pure state $\omega\in S(\mathscr{C}) = S(C(E,\mathscr{A}))$ is concentrated on the one-point subset $F_\omega\in E$ in the sense that for any $\hat{f}\in C(E,\mathscr{A})$ one has

$$\omega(E_{\mathfrak{g}}(\hat{f})) = \omega_{\mathscr{A}}(\hat{f}(F_\omega)), \quad \hat{f}\in C(E,\mathscr{A}). \tag{2.41}$$

Applying this to $\hat{f}(F) := \sigma(g_Q^{-1}(t,F))$ $(X_{\beta J(i)})$ obtained from (4.2), and by the use of (2.5), (2.37), (2.40), and (2.25), one gets

$$\begin{aligned}\lim_{i\in\mathscr{I}} \omega(\tau_t^Q(X_{\beta J(i)})) &= \lim_{i\in\mathscr{I}} \omega_{\mathscr{A}}(\sigma(g_Q^{-1}(t,F_\omega))(X_{\beta J(i)})) \\ &= \lim_{i\in\mathscr{I}} \omega_{\mathscr{A}}(X_{J(i)}(\mathrm{Ad}(g_Q^{-1}(t,F_\omega))\beta)) \\ &= F_\omega(\mathrm{Ad}(g_Q^{-1}(t,F_\omega))\beta) \\ &= \mathrm{Ad}^*(g_Q(t,F_\omega))F_\omega(\beta) \\ &= \varphi_t^Q F_\omega(\beta) = \omega_{\mathscr{N}}(E_{\mathfrak{g}}(\varphi_t^{Q*}f_\beta)),\end{aligned} \tag{2.42}$$

since $\omega_{f}\in S(C(E))$ is given by the Dirac measure on E concentrated on F_ω. Hence (2.38) is proved. Q.E.D.

One can show that the net $[X_{\beta J}: J\in\mathbf{\Pi}, \beta\in\mathfrak{g}]$ has more than one w^*-cluster point in $\mathscr{A}^{**}$ [resp. in $\mathscr{L}(\mathfrak{g},\mathscr{A}^{**})$—to be more precise] for any of the specific choices of $\mathbf{\Pi}$ mentioned in Remark 2.7. This means that the choice of the subnet $[X_{\beta J(i)}: i\in\mathscr{I}, \beta\in\mathfrak{g}]$ in the last proposition is nonunique, and there the described physically natural extension $\omega\in S(\mathscr{C})$ of $\omega_{\mathscr{A}}$ is nonunique as well. If we consider infinite systems obtained by thermodynamic limits as mere convenient approximations to large but finite physical systems, then the above mentioned ambiguity can be interpreted as a consequence of ambiguity in these kind of approximations.

III. THE THERMODYNAMIC LIMIT OF LOCAL EVOLUTIONS

Let $[\beta_j: j=1,2,...,\dim G]$ be a fixed basis of $\mathfrak{g}$, and a polynomial $Q\in C(\mathfrak{g}^*,\mathbb{R})$ in the variables $F_j := F(\beta_j)$ be given. Let

$$X_j^K := |K|X_{jK} := \sum_{p\in K} \pi_p(X_{\beta_j}),$$

$$j = 1,2,...,n := \dim G, \quad K\in\Pi. \tag{3.1}$$

Assume that an ordering of the multiplication of variables F_j in the polynomial $Q(F_1,F_2,...,F_n) := Q(F)$ is prescribed in such a way that the elements $Q^K\in\mathscr{A}$,

$$Q^K := |K|Q(X_{1K},X_{2K},...,X_{nK}), \quad K\in\Pi \tag{3.2}$$

obtained by the substitution of X_{jK} for F_j in $Q(F)$, are all self-adjoint: $Q^{K*} = Q^K$. We define the one-parameter groups $\tau^K \subset$*-aut $\mathscr{A}$ of local time evolutions by

$$\tau_t^K(x) := \exp(itQ^K)x\exp(-itQ^K), \quad x\in\mathscr{A}, \quad t\in\mathbf{R}, \tag{3.3}$$

for any finite $K\in\Pi$. We are interested in a proof of the existence of suitably defined (thermodynamic) limits

$$\tau_t^Q(x) := (\text{some topology})\text{-}\lim_K \tau_t^K(x), \quad x\in\mathscr{A}, \tag{3.4}$$

for $K\to\Pi$, with a general polynomial Q. A necessary prerequisite for the existence of τ^Q for any Q is the existence of the limits of $X_{\beta K}$ ($\beta\in\mathfrak{g}$). We have seen, however, in Sec. II A that $X_{\beta K}$ cannot have a limit in $\mathscr{A}$, and weak limits for all the $X_{\beta K}$'s in the algebra $\mathscr{A}^{**}$ exist only in the strong topology generated by the states $\omega\in S_\mathfrak{g} \subset S_*(\mathscr{A}^{**}) :=$ the set of normal states on $\mathscr{A}^{**}$, for which the central covers $c(\pi_\omega)$ are majorized by $P_G\in\mathscr{Z}$, cf. (2.7). This topology (called here also the s^*-topology[32]) is determined by the family of seminorms p_ω and p_ω^* on $\mathscr{A}^{**}$, $\omega\in S_\mathfrak{g} := P_\mathfrak{g}S_*(\mathscr{A}^{**})$,

$$p_\omega(x) := \sqrt{\omega(x^*x)}, \quad p_\omega^*(x) := \sqrt{\omega(xx^*)}, \quad x\in\mathscr{A}^{**}. \tag{3.5}$$

It is clear that on the subset of self-adjoint elements of $\mathscr{A}^{**}$ the s^*-topology coincides with the s-topology determined by the seminorms p_ω only. These topologies are Hausdorff on the subalgebra $P_G\mathscr{A}^{**} = \pi_G(\mathscr{A})''$, and we shall work in the framework of this von Neumann subalgebra of $\mathscr{A}^{**}$ identifying $\mathscr{A}$ with $\pi_G(\mathscr{A})$, cf. Definition 2.2.

Notation 3.1: Let $[\beta_j: j = 1,2,...,n := \dim G]$ be a fixed basis of $\mathfrak{g}$, and let c_{jk}^m be the structure constants of $\mathfrak{g}$:

$$[\beta_j,\beta_k] = c_{jk}^m\beta_m. \tag{3.6}$$

Let $X(\beta) := X_\beta$ ($\beta\in\mathfrak{g}$) with the norm $\|X(\beta)\|$ from $\mathscr{L}(H)$. Let the polynomial Q be written in the form of linear combination of p monomials of the maximal degree q with the upper bound $M\geqslant 1$ of the absolute values of the coefficients. Let us denote

(i) $b := \max[1 + \|X(\beta_j)\|: j = 1,2,...,n]$;
(ii) $c := \max[|c_{jk}^m|: j,k,m = 1,2,...,n]$;
(iii) $a_K := \max[nc, 2|K|b], K\in\Pi$;
(iv) $b(x) := \max[b,\|x\|], x\in\mathscr{A}$;
(v) $B^K := \mathscr{A}^K \cup [X_{\beta L}: \beta\in\mathfrak{g}, L\in\Pi]$.

We shall also use for multiple commutators in $\mathscr{A}$,

$$[y,x]^{(m+1)} := [y,[y,x]]^{(m)}, \qquad [y,x]^{(0)} := x, \quad [y,x] := yx - xy. \tag{3.7}$$

For the generators X_j^K we then have the relation,

$$[X_j^K, X_k^K] = ic_{jk}^m X_m^K, \quad K\in\Pi, \quad j,k,(m) = 1,2,...,n. \tag{3.8}$$

By multiple use of this formula, and by recursive calculations of degrees and numbers of monomials in variables X_j^K, and variables $[X_{j_1}^J,[X_{j_2}^J,...,[X_{j_r}^J,x]\cdots]]$ ($x\in\mathscr{A}^J$) and $[X_{j_1}^K,[X_{j_2}^K,...,[X_{j_r}^K,X_{kL}]\cdots]]$ ($L\in\Pi$) occurring in the multiple commutator $[Q^K,x]^{(m)}$, one obtains from elementary properties of the norm [cf. Lemma 6.2.4 in Ref. 25(a)] the following lemma.

Lemma 3.2: Let $J\in\Pi$, $x\in B^J$, $K\in\Pi$, and let m be any positive integer. Then the following estimate is valid:

$$\|[Q^K,x]^{(m)}\| < [b(x)/q](m-1)!(Mpq^2b^{q-1}a_J)^m, \tag{3.9}$$

where Notation 3.1 was used.

Now we can prove existence of the limits in (3.4) for small t depending on the choice of $x\in\mathscr{A}$.

Lemma 3.3: Let $r_K := (Mpq^2b^qa_K)^{-1}$ ($K\in\Pi$), and let $|t|\leqslant r_J$, $x\in B^J$ for any fixed $j\in\Pi$.

(i) The sum

$$\tau_t^K(x) = \sum_{m=0}^\infty \frac{(it)^m}{m!}[Q^K,x]^{(m)} \tag{3.10}$$

is convergent in norm in $\mathscr{A}$, and the convergence is uniform on the Cartesian product of the sets $[K: K\in\Pi]$, $[t: |t|\leqslant r_J]$, and $[x\in B^J: \|x\|\leqslant a]$ with any $a\in\mathbf{R}_+$.

(ii) The following limit exists in $P_G\mathscr{A}^{**}$:

$$\tau_t^Q(x) := s^*\text{-}\lim_K \tau_t^K(x). \tag{3.11}$$

Proof: (i) is a consequence of the estimate (3.9) which is independent of $K\in\Pi$ and the corresponding majorizing power series is uniformly convergent on the Cartesian product of the three sets as stated in the assertion.

The uniform boundedness in $K\in\Pi$ of the multiple commutators in the right-hand side of (3.10) together with (2.7) and (3.1) imply the existence of

$$s^*\text{-}\lim_K [Q^K,x]^{(m)}\in P_G\mathscr{A}^{**}.$$

This fact combined with assertion (i) leads to (ii). Q.E.D.

Proposition 3.4: The restriction to $\mathscr{A}^J$ of the mappings τ_t^Q from (3.11) for real t, $|t|\leqslant r_J$, are *-homomorphisms of $\mathscr{A}^J$ into $\mathscr{C}^J$ [cf. Definition 2.2(i)].

Proof: The mappings τ_t^K are inner automorphisms of $\mathscr{A}$, and their canonical extensions to $\mathscr{A}^{**}$ leave elements of $\mathscr{Z}$ (hence also P_G) invariant, so that τ_t^K can be considered as inner automorphisms of $P_G\mathscr{A}^{**}$. The properties of the s^*-convergence[32,38,39] imply then that τ_t^Q are *-homomorphisms of $\mathscr{A}^J = \pi_G(\mathscr{A}^J)$ into $P_G\mathscr{A}^{**}$. Each multiple commutator $[Q^K,x]^{(m)}$ ($x\in\mathscr{A}^J$, $K\supset J$) is a polynomial in the variables $X_{\beta K}$, and in some of the variables of the form $[X_{j_1}^J,[X_{j_2}^J,...,[X_{j_s}^J,x]\cdots]]\in\mathscr{A}^J$ which are independent of K. The strong limits of $X_{\beta K}$ are elements of $\mathscr{N}$, cf. (2.7) and Definition 2.2 (i). Hence the sums of the norm-convergent series

$$\tau_t^Q(x) = \sum_{m=0}^\infty \frac{(it)^m}{m!} s^*\text{-}\lim_K [Q^K,x]^{(m)}, \quad x\in\mathscr{A}^J, \tag{3.12}$$

are elements of $\mathscr{C}^J$. Q.E.D.

Proposition 3.5: Let $|t|\leqslant r_1$ ($:= r_J$ with $|J| = 1$), $t\in\mathbf{R}$. The limits

$$\tau_t^Q(E_\mathfrak{g}(f_\beta)) := s^*\text{-}\lim_L \tau_t^Q(X_{\beta L}), \quad \beta\in\mathfrak{g}, \tag{3.13}$$

exist in $\mathcal{N}$ and have a unique extension to the *-automorphism group τ^Q of $\mathcal{N}$ given by the formula

$$\tau_t^Q(E_{\mathfrak{g}}(f)) = E_{\mathfrak{g}}(\varphi_t^{Q*} f), \quad f \in C(E), \tag{3.14}$$

where φ^Q is the Hamiltonian flow on $\mathfrak{g}^*$ generated by the Hamiltonian function Q. τ^Q is strongly continuous on $\mathcal{N}$.

Proof: Let us calculate the right-hand side of (3.13) explicitly. Let $[Q,f]$ be the Poisson bracket on $\mathfrak{g}^*$ (2.18), and let

$$[Q,f]^{(m+1)} := [Q,[Q,f]^{(m)}], \quad [Q,f]^{(0)} := f.$$

First, we intend to prove the equality

$$i^m s^*\text{-}\lim_L s^*\text{-}\lim_K [Q^K, X_{\beta L}]^{(m)} = E_{\mathfrak{g}}([Q,f_\beta]^{(m)}), \tag{3.15}$$

for all $m \in \mathbb{Z}_+$, $\beta \in \mathfrak{g}$. Using the Lie algebra representation property (with $L \subset K$),

$$[X_\beta^K, X_{\chi L}] = iX_{[\beta,\chi]L}, \quad \beta,\chi \in \mathfrak{g}, \tag{3.16}$$

and the relation (2.21) as well as (2.7) and (2.9), we obtain

$$s^*\text{-}\lim_L s^*\text{-}\lim_K i[X_\beta^K, X_{\chi L}] = E_{\mathfrak{g}}([f_\beta, f_\chi]), \quad \beta,\chi \in \mathfrak{g}. \tag{3.17}$$

Since Q is a polynomial in f_β ($\beta \in \mathfrak{g}$), one obtains (3.15) with a help of the algebraic properties (i.e., bilinearity, antisymmetry, derivation property, and fulfillment of the Jacobi identity) of both the commutators and the Poisson brackets as well as due to the morphism properties of the mapping $E_{\mathfrak{g}}$ from (2.11). Then the norm convergence in (3.10), the properties of the s^*-convergence, the equality (3.15), and the norm continuity of the mapping $E_{\mathfrak{g}}$ give, for small $t \in \mathbb{R}$,

$$\begin{aligned}\tau_t^Q(E_{\mathfrak{g}}(f_\beta)) &= \sum_{m=0}^{\infty} \frac{t^m}{m!} E_{\mathfrak{g}}([Q,f_\beta]^{(m)}) \\ &= E_{\mathfrak{g}}\left(\sum_{m=0}^{\infty} \frac{t^m}{m!} [Q,f_\beta]^{(m)}\right) = E_{\mathfrak{g}}(f_{\beta t}),\end{aligned} \tag{3.18}$$

where

$$f_{\beta t}(F) := \sum_{m=0}^{\infty} \frac{t^m}{m!} [Q,f_\beta]^{(m)}(F). \tag{3.19}$$

It is easily seen that $f_{\beta t}$ satisfies (2.20), hence $f_{\beta t} = \varphi_t^{Q*} f_\beta$. The morphism and continuity properties of the mappings φ_t^{Q*} and $E_{\mathfrak{g}}$ on $C(E)$ generated by f_β ($\beta \in \mathfrak{g}$) imply the extension of (3.18) to (3.14) for short times t. The group property of $\varphi^{Q*} \subset$ *-aut $C(E)$ together with the fact that $E_{\mathfrak{g}}$: $C(E) \to \mathcal{N}$ is an isomorphism, prove that there is a unique one-parameter group τ^Q such that (3.14) is valid for all $t \in \mathbb{R}$. The strong continuity of τ^Q on $\mathcal{N}$ is a consequence of the uniform continuity on compacts of the flow φ^Q,[44b,47] since $E = \operatorname{supp} E_{\mathfrak{g}}$ is compact. Q.E.D.

The proposition gives a natural extension of the family of mappings τ_t^Q: $\mathcal{A}^J \to \mathcal{C}^J$ of Proposition 3.4 to an evolution group of the algebra $\mathcal{N}$ of classical observables. Further extension of τ^Q leads to the following theorem, which is the main result of this section.

Theorem 3.6: There is a unique C^*-automorphism group $\tau^Q = [\tau_t^Q: t \in \mathbb{R}]$ of $\mathcal{C}$ extending the mappings τ_t^Q of Proposition 3.4 as well as of Proposition 3.5. The group τ^Q is strongly continuous on $\mathcal{C}$, and the restrictions of τ^Q to $\mathcal{C}^J$ ($J \in \mathbf{\Pi}$) are one-parameter automorphism groups of the C^*-subalgebras $\mathcal{C}^J$ of $\mathcal{C}$.

Proof: Let us choose any $J \in \mathbf{\Pi}$, and let $|t| \leqslant r_J$ for a given $t \in \mathbb{R}$. Then we can apply Lemma 2.3 with $\mathcal{A}$ replaced by $\mathcal{A}^J$, τ_A replaced by τ_t^Q of Proposition 3.4, and τ_N substituted by τ_t^Q of Proposition 3.5. By a simultaneous use of the structure of $\mathcal{C}^J$ according to definition 2.2 as well as of the mapping from Lemma 2.4, we obtain a unique homomorphism τ_t^Q: $\mathcal{C}^J \to \mathcal{C}^J$ extending those of Propositions 3.4 and 3.5. We shall now prove the group property of these *-homomorphisms τ_t^Q of $\mathcal{C}^J$ for small real t (invertibility of τ_t^Q will then also be proved in this way), i.e.,

$$\tau_{t_1+t_2}^Q(y) = \tau_{t_1}^Q(\tau_{t_2}^Q(y)), \quad y \in \mathcal{C}^J, \quad \max(|t_1|,|t_2|) < \tfrac{1}{2} r_J. \tag{3.20}$$

Due to the structure of $\mathcal{C}^J$, it is sufficient to prove (3.20) for the elements of the form $y = xz$, $x \in \mathcal{A}^J$, $z \in \mathcal{N}$. But $\tau_t^Q(xz) = \tau_t^Q(x)\tau_t^Q(z)$, and the restriction of τ^Q to $\mathcal{N}$ satisfies (3.20). Hence it is sufficient to prove (3.20) for $y = x \in \mathcal{A}^J$.

We shall first prove the equality

$$\tau_{t_1}^Q\left(s^*\text{-}\lim_K \tau_{t_2}^K(x)\right) = s^*\text{-}\lim_K \tau_{t_1}^Q(\tau_{t_2}^K(x)). \tag{3.21}$$

(Remember that we have not proved the s^*-s^*-continuity of the morphism τ_t^Q: $\mathcal{C}^J \to \mathcal{C}^J$.) By considering that Q is a polynomial, that the product in $P_G\mathcal{A}^{**}$ is s^*-continuous on bounded sets, and by repeated use of (2.7), (3.13), (3.14), as well as of the morphism property of $\tau_{t_1}^Q$, one obtains

$$\tau_{t_1}^Q\left(s^*\text{-}\lim_K [Q^K,x]^{(m)}\right) = s^*\text{-}\lim_K \tau_{t_1}^Q([Q^K,x]^{(m)}). \tag{3.22}$$

The equality (3.21) is obtained from (3.22) by the uniform norm convergence in (3.10) as well as by the norm continuity of $\tau_{t_1}^Q$.

Since $\tau_{t_2}^K(x) \in B^J$ for all $x \in \mathcal{A}^J$ and all $K \in \mathbf{\Pi}$, we can apply Lemma 3.3 to obtain

$$\begin{aligned}&s^*\text{-}\lim_K \tau_{t_1}^Q(\tau_{t_2}^K(x)) \\ &\quad = s^*\text{-}\lim_K s^*\text{-}\lim_L \tau_{t_1}^L(\tau_{t_2}^K(x)) \\ &\quad = s^*\text{-}\lim_K s^*\text{-}\lim_L \sum_{k,m=0}^{\infty} \frac{(it_1)^k}{k!} \frac{(it_2)^m}{m!} \\ &\qquad \times [Q^L,[Q^K,x]^{(m)}]^{(k)}.\end{aligned} \tag{3.23}$$

The estimate of the form (3.9) with m replaced by $m + k$ on the right-hand side can be proved for the multiple commutators in (3.23) in the same way as in Lemma 3.2. Hence the sum in (3.23) converges in norm uniformly in $L, K \in \mathbf{\Pi}$. Considering again the specific structure of Q^L and Q^K and the s^*-continuity of the algebraic product on bounded sets, we can set $L = K$ in (3.23) and obtain

$$\tau^Q_{t_1}(\tau^Q_{t_2}(x)) = s^*\text{-}\lim_K \tau^Q_{t_1}(\tau^K_{t_2}(x))$$

$$= s^*\text{-}\lim_K \sum_{k,m=0}^{\infty} \frac{(it_1)^k}{k!}\frac{(it_2)^m}{m!}\,[Q^K,x]^{(m+k)}$$

$$= s^*\text{-}\lim_K \sum_{k=0}^{\infty} \frac{(t_1+t_2)^k}{k!}[iQ^K,x]^{(k)}$$

$$= \tau^Q_{t_1+t_2}(x)\,, \tag{3.24}$$

where the last equality is valid according to Lemma 3.3. This proves the group property (3.20) for small $t\in\mathbb{R}$, hence, since such τ^Q_t are defined on the whole $\mathscr{C}^J$, τ^Q_t can be uniquely extended to an automorphism group $\tau^Q \subset {}^*\text{-aut}\mathscr{C}^J$ for any $J\in\mathbf{\Pi}$, and these in turn have a unique extension to $\tau^Q\subset{}^*\text{-aut}\,\mathscr{C}$ due to norm continuity of any τ^Q_t (determined now on the union of all $\mathscr{C}^J$, $J\in\mathbf{\Pi}$, for any fixed $t\in\mathbb{R}$).

The strong continuity of the group τ^Q means

$$\lim_{t\to 0}\|\tau^Q_t(y)-y\| = 0, \quad \text{for all } y\in\mathscr{C}\,. \tag{3.25}$$

For $y\in\mathscr{A}^J$ ($J\in\mathbf{\Pi}$) the validity of (3.25) follows from the uniform convergence in (3.12), and this together with norm continuity of any morphism τ^Q_t implies (3.25) for all $y\in\mathscr{A}$. The strong continuity of the restriction of τ^Q to $\mathscr{N}$ was proved in Proposition 3.5, and the continuity (3.25) for general elements $y\in\mathscr{C}$ can be obtained easily from that for $y=xz$, $x\in\mathscr{A}$, $z\in\mathscr{N}$. Q.E.D.

Let us calculate the infinitesimal generator of $\tau^Q\subset{}^*\text{-aut}\,\mathscr{C}$,[48] i.e., the derivation[28,39] δ_Q of τ^Q,

$$\delta_Q(y) := \text{n-}\lim_{t\to 0}\frac{1}{t}(\tau^Q_t(y)-y), \quad y\in D(\delta_Q)\,, \tag{3.26}$$

where $D(\delta_Q)$ is the domain of δ_Q, and the limit is taken in the norm of $\mathscr{C}$, cf. Ref. 28, Consequence 3.1.8. The explicit form of δ_Q obtained in the next proposition will enable us to prove in Sec. IV an explicit expression for the time-evolved element $\tau^Q_t(y)$ for an arbitrary $y\in\mathscr{C}$.

Proposition 3.7: Let δ_Q be the infinitesimal generator of τ^Q, and let $y\in D(\delta_Q)$ be of the form $y=E_{\mathfrak{g}}(\hat{f})$ for some $\hat{f}\in C(E,\mathscr{A}^J)$, $J\in\mathbf{\Pi}$. Then the element $\delta_Q(y)\in\mathscr{C}$ is expressed by the formula

$$\delta_Q(E_{\mathfrak{g}}(\hat{f})) = \int\Big(i[X^J(d_FQ),\hat{f}(F)] + \sum_{j=1}^{n}\delta_j\hat{f}(F)\,[Q,F_j](F)\Big)E_{\mathfrak{g}}(dF)\,, \tag{3.27}$$

where the first square bracket is the commutator in $\mathscr{A}^J$, the symbol $[Q,F_j](F)$ denotes the Poisson bracket of the functions Q and F_j: $F\to F(\beta_j)$ on $\mathfrak{g}^*$ in the point $F\in\mathfrak{g}^*$, and $\delta_j\hat{f}(F)$ is the value of the partial derivative of $\hat{f}$ with respect to the component F_j of F in the point F.

Proof: The formula (3.12) is valid for all $x\in B^J$; by its differentiation at $t=0$ one obtains

$$\delta_Q(x) = s^*\text{-}\lim_K [iQ^K,x], \quad x\in B^J, \quad J\in\mathbf{\Pi}\,. \tag{3.28}$$

The derivation property of the commutator, the polynomial form of Q as well as the formulas (2.7) and (2.12), lead for $x\in\mathscr{A}^J$ to

$$\delta_Q(x) = i\sum_{j=1}^{n} E_{\mathfrak{g}}(\delta_jQ)\,[X^J_j,x]$$

$$= i\int [X^J(d_FQ),x]E_{\mathfrak{g}}(dF)\,, \tag{3.29}$$

where δ_jQ is the partial derivative with respect to F_j.

The vector field λ_Q is the derivation of the strongly continuous (cf. Proposition 3.5) automorphism group φ^{Q*} of $C(E)$ introduced by (2.20). Let $D(\lambda_Q)$ be the domain of λ_Q. By the norm continuity of the mapping $E_{\mathfrak{g}}$ from (2.11), we obtain from (2.19)

$$\delta_Q(E_{\mathfrak{g}}(f)) = E_{\mathfrak{g}}([Q,f]), \quad f\in D(\lambda_Q)\,. \tag{3.30}$$

Combining (3.29) with (3.30), one obtains (3.27) for elements $E_{\mathfrak{g}}(\hat{f})$ with $\hat{f}:=\Sigma\, x_kf_k$ for any finite number of $x_k\in\mathscr{A}^J, f_k\in D(\lambda_Q)$. Let $\mathscr{P}(E)\subset C(E)$ be the set of polynomials in f_β ($\beta\in\mathfrak{g}$). The considerations leading to (3.18) show that f_β's are analytic elements[28] for λ_Q. Since analytic elements of any derivation on an algebra form a subalgebra and δ_Q is bounded on $\mathscr{A}^J$ accoridng to (3.29), cf. also Lemma 3.3, the set $\mathscr{C}^J_P$ of elements $E_{\mathfrak{g}}(\hat{f})$ of $\mathscr{C}^J$ with $\hat{f}:=\Sigma\, x_kf_k\,[x_k\in\mathscr{A}^J, f_k\in\mathscr{P}(E)]$ forms a norm-dense subset of $\mathscr{C}^J$ of analytic elements for δ_Q which is δ_Q invariant: $\delta_Q\mathscr{C}^J_P\subset\mathscr{C}^J_P$. The union of all the $\mathscr{C}^J_P$ ($J\in\mathbf{\Pi}$) is then a δ_Q-invariant norm-dense subset of $\mathscr{C}$ consisting of analytic elements for δ_Q. Let this union be denoted by $\mathscr{C}_P$. According to 3.1.20 of Ref. 28, $\mathscr{C}_P$ is a core for δ_Q. Then the expression (3.27) is valid for all $E_{\mathfrak{g}}(\hat{f})\in\mathscr{C}_P$,

$$\delta_Q(E_{\mathfrak{g}}(\Sigma x_kf_k))$$

$$= \sum_k(\delta_Q(x_k)E_{\mathfrak{g}}(f_k) + x_kE_{\mathfrak{g}}([Q,f_k]))$$

$$= \sum_k E_{\mathfrak{g}}\Big(\sum_{j=1}^{n}(i\delta_jQ\,[X^J_j,x_k]f_k + x_k\delta_jf_k\,[Q,F_j])\Big),$$

$$x_k\in\mathscr{A}^J\,. \tag{3.31}$$

The closedness of δ_Q now gives the result, since the operator in (3.27) of the form

$$\sum_j[Q,F_j]\delta_j\colon D(\lambda_Q)\to C(E,\mathscr{A}^J)$$

is just an alternative form of the derivation λ_Q of φ^{Q*} on $C(E,\mathscr{A}^J)$. Q.E.D.

Remark 3.8: The notation $\Sigma_j[Q,F_j]\delta_j$ for λ_Q is unambiguous on the set of continuously differentiable functions $C^1(E)$. For a general $f\in D(\lambda_Q)$ the symbol $\Sigma_j[Q,F_j]\delta_j$ should be understood as the directional derivative in the direction of λ_Q (given by the vector components $[Q,F_j]$ in the basis $[\delta_j: j=1,2,\ldots,n]$ of $T_F\mathfrak{g}^*$) at any noncritical point F, i.e., $F\in\mathfrak{g}^*$ in which $d_FQ\neq 0$. For any critical F, $d_FQ=0$, the value of $\lambda_Qf=[Q,f]\in C(E)$ is taken to be zero, $[Q,f](F):=0$, for all $f\in D(\lambda_Q)$. The domain $D(\lambda_Q)$ consists of such $f\in C(E)$, for which the mentioned directional derivative can be continuously extended to the whole E with zero values at critical points.

IV. THE DYNAMICS OF GENERALIZED MEAN-FIELD MODELS

Let $\mathscr{A}$ be an arbitrary C^*-algebra, and let $\sigma(G)\subset{}^*\text{-aut}\,\mathscr{A}$ be a strongly continuous representation of a connected Lie group G. Let $E_{\mathfrak{g}}$ be a projection-valued measure on $\mathfrak{g}^*$ with values in the center $\mathscr{Z}$ of $\mathscr{A}^{**}$ [identified with $\pi_u(\mathscr{A})''$ as before]. Assume that $E_{\mathfrak{g}}$ is G equivariant, i.e.,

$$\sigma(g)(E_{\mathfrak{g}}(B))=E_{\mathfrak{g}}(\mathrm{Ad}^*(g)B),\quad g\in G,\quad \text{Borel } B\subset\mathfrak{g}^*, \tag{4.1}$$

and that the representation $\pi_u(\mathscr{A})E_{\mathfrak{g}}(B)$ of $\mathscr{A}$ is faithful for any open $B\subset\operatorname{supp}E_{\mathfrak{g}}$. Set $P_G:=E_{\mathfrak{g}}(\mathfrak{g}^*)$. Let us assume, moreover, that $E:=\operatorname{supp}E_{\mathfrak{g}}$ is compact and contains at least one nontrivial (i.e., of a positive dimension) orbit of the Ad*-representation of G. We shall identify $\mathscr{A}$ with $\pi_G(\mathscr{A})$ as in preceding sections.

Let $Q\in C^\infty(\mathfrak{g}^*,\mathbf{R})$, and g_Q be a differentiable function satisfying (2.25) and (2.26), or equivalently, g_Q is the solution of (2.30) with β_F^Q from (2.29). Let $C(E,\mathscr{A})$ and $\mathscr{N}:=E_{\mathfrak{g}}(C(E))$ be as in Definition 2.2.

Proposition 4.1: Let $\hat f\in C(E,\mathscr{A})$, $t\in\mathbf{R}$. The function $\hat f_t: E\to\mathscr{A}$ determined by

$$\hat f_t(F):=\sigma(g_Q^{-1}(t,F))(\hat f(\varphi_t^Q F)),\quad F\in\mathfrak{g}^*, \tag{4.2}$$

is norm continuous: $\hat f_t\in C(E,\mathscr{A})$. The mappings $\hat f\to\hat f_t$ ($t\in\mathbf{R}$) form a one-parameter group of *-automorphisms of $C(E,\mathscr{A})$,

$$(\hat f_t)_s=\hat f_{t+s},\quad \text{for all } t,s\in\mathbf{R}, \tag{4.3}$$

and this group is strongly continuous,

$$\lim_{t\to0}\|\hat f_t-\hat f\|=0,\quad \hat f\in C(E,\mathscr{A}). \tag{4.4}$$

Proof: The support $E\subset\mathfrak{g}^*$ of $E_{\mathfrak{g}}$ is left invariant by $\mathrm{Ad}^*(G)$ transformations due to (4.1), hence E is also φ^Q invariant. The differentiability of φ^Q and g_Q on $\mathbf{R}\times\mathfrak{g}^*$, and the strong continuity of $\sigma(G)$ together with the norm continuity of each $\sigma(g)\in{}^*\text{-aut}\,\mathscr{A}$, lead to the continuity of $\hat f_t$, i.e., $\hat f_t\in C(E,\mathscr{A})$. If we consider, in addition to the listed continuity properties, the compactness of $\operatorname{supp}E_G$, we obtain (4.4).

The morphism properties of the mapping $\hat f\to\hat f_t$ of the C^*-algebra $C(E,\mathscr{A})$ into itself are due to morphism properties of $\varphi^Q\in{}^*\text{-aut}\,C(E,\mathscr{A})$, the morphism properties of each $\sigma(g)\in{}^*\text{-aut}\,\mathscr{A}$, as well as the pointwise character of algebraic operations in $C(E,\mathscr{A})$, e.g., $(\hat f_1\hat f_2)(F):=\hat f_1(F)\hat f_2(F)$.

The group property (4.3) is obtained from the group property of the flow φ^Q, from the cocycle property (2.26), as well as from the group representation property of $\sigma(g)$,

$$\sigma(g_1g_2)=\sigma(g_1)\circ\sigma(g_2),\quad g_1,g_2\in G. \tag{4.5}$$

The group property implies invertibility. This shows that the considered mappings $\hat f\to\hat f_t$ form a strongly continuous one-parameter group of *-automorphisms of $C(E,\mathscr{A})$. Q.E.D.

The assumptions of this section left the conditions of validity of Proposition 2.5 unchanged. Let $\mathscr{C}$ be given as in definition 2.2. The following theorem is an immediate consequence of Propositions 4.1 and 2.5.

Theorem 4.2: the mappings $\tau_t^Q:\mathscr{C}\to\mathscr{C}$ ($t\in\mathbf{R}$) determined by

$$\tau_t^Q(E_{\mathfrak{g}}(\hat f)):=E_{\mathfrak{g}}(\hat f_t),\quad \hat f\in C(E,\mathscr{A}),\quad t\in\mathbf{R}, \tag{4.6}$$

form a strongly continuous one-parameter group of *-automorphisms of the C^*-algebra $\mathscr{C}$.

Proof: The morphism and continuity properties of the automorphism group of Proposition 4.1 are conserved by the *-isomorphism $E_{\mathfrak{g}}: C(E,\mathscr{A})\to\mathscr{C}$ of Proposition 2.5. Q.E.D.

Proposition 4.3: Let $\sigma(G)$ be locally faithful, i.e., the kernel $[h\in G:\sigma(h)=\sigma(e)]$ is discrete. Assume that g_Q satisfies (2.26). If the function

$$F(\in E)\mapsto\sigma(g_Q^{-1}(t,F))(x),\quad t\in\mathbf{R},\quad x\in\mathscr{A}, \tag{4.7}$$

are all constant, then $\beta_Q:=\beta_F^Q$ from (2.27) is a constant in $\mathfrak{g}$. Hence the restriction of Q to any Ad*-orbit in E is identical to the restriction of the linear function $\tilde Q(F):=F(\beta_Q)$, $F\in\mathfrak{g}^*$, up to an additive constant.

Proof: The constancy of (4.7) together with (2.26) imply that the function $t\mapsto\sigma(g_Q(t,F))\in{}^*\text{-aut}\,\mathscr{A}$ is a one-parameter group independent of F. There is an isomorphism of $\mathfrak{g}$ onto one-parameter subgroups of $\sigma(G)$ due to the local faithfulness; hence $\sigma(g_Q(t,F))=\sigma(\exp(t\beta_Q))$ for some $\beta_Q\in\mathfrak{g}$. The continuity of g_Q in F, the continuity of $\sigma(G)$, and the boundary condition $g_Q(0,F)=e$ imply $\beta_F^Q=\beta_Q$ (for all $F\in E$), so that $d_FQ=\beta_Q-\beta_F^0$, according to (2.29). The differentials d_Ff_χ ($\chi\in\mathfrak{g}$) of the functions $f_\chi(F):=F(\chi)$ contain a basis of the cotangent space to the Ad*-orbit through any $F\in E$. Hence any function Q^0 on the orbit with zero Poisson brackets $[Q^0,f_\chi]=0$ for all $\chi\in\mathfrak{g}$ on the orbit equals a constant function on the orbit,

$$Q^0(\mathrm{Ad}^*(g)F)=Q^0(F),\quad \text{for all } g\in G.$$

We have for the Q used in the definition of g_Q in (4.7),

$$-[Q,f_\chi](F)=F([d_FQ,d_Ff_\chi])$$
$$=F([\beta_Q-\beta_F^0,\chi])=F([\beta_Q,\chi]),\quad \chi\in\mathfrak{g},$$

since $F([\beta_F^0,\chi])=0$ ($F\in\mathfrak{g}^*$, $\chi\in\mathfrak{g}$) due to the definition of β_F^0, cf. the text below (2.29). Hence $[Q,f_\chi]=[\tilde Q,f_\chi]$ for all $\chi\in\mathfrak{g}$, and the restriction of $Q^0:=Q-\tilde Q$ to any Ad*-orbit in E is a constant function. Q.E.D.

Corollary 4.4: Let $\sigma(G)$ be locally faithful, and let g_Q be as above. The one-parameter group τ^Q from (4.6) leaves the C^*-algebra $\mathscr{A}$ invariant: $\tau_{\mathbf{R}}^Q(\mathscr{A})=\mathscr{A}$, iff

$$Q(\mathrm{Ad}^*(g)F)=\mathrm{Ad}^*(g)F(\beta_Q)+Q^0(F),\quad \text{for all } g\in G, \tag{4.8}$$

for some $\beta_Q\in\mathfrak{g}$, and some $Q^0\in C^\infty(E)$ constant on the orbits of $\mathrm{Ad}^*(G)$ lying in E, and $d_FQ^0=-\beta_F^0$ ($F\in E$), cf. (2.29).

Proof: The τ^Q invariance of $\mathscr{A}$ means the constancy of all the functions in (4.7), as is seen from (4.2) and (4.6), and from the identity of $\mathscr{A}$ with the $E_{\mathfrak{g}}$ image of constant functions in $C(E,\mathscr{A})$. The necessity of (4.8) for this invariance was proved in Proposition 4.3. Assuming (4.8), one obtains from (2.29): $\beta_F^Q=\beta_Q$, and the unique solution g_Q of (2.30) is independent on the parameter $F\in\mathfrak{g}^*$: $g_Q(t,F)=\exp(t\beta_Q)$. The constancy in (4.7) is now clear. Q.E.D.

Let us now derive the explicit form of the derivation δ_Q of the automorphism group τ^Q introduced in this section.

Proposition 4.5: Let $\delta_\sigma(\beta)$ ($\beta\in\mathfrak{g}$) be the derivation of the strongly continuous one-parameter group $t\mapsto\sigma(\exp(t\beta))$ of automorphisms of $\mathscr{A}$, and let λ_Q be the derivation of the strongly continuous one-parameter group φ^{Q*} of automorphisms of the C^*-algebra $C(E,\mathscr{A})$, in the notation of Sec. III. Let $\tau^Q\subset C^*\text{-aut}\,\mathscr{C}$ be given by (4.6). Then the infinitesimal generator (i.e., the derivation) of τ^Q is expressed by the formula

$$\delta_Q(E_\mathfrak{g}(\hat{f}))=\int(\lambda_Q\hat{f}(F)-\delta_\sigma(\beta_F^Q)(\hat{f}(F)))E_\mathfrak{g}(dF)\,,$$

$$E_\mathfrak{g}(\hat{f})\in D(\delta_Q)\,, \tag{4.9}$$

where β_F^Q ($F\in E$) is given in (2.29), and $D(\delta_Q)$ is the domain of the derivation δ_Q.

Proof: Due to norm continuity of the mapping $E_\mathfrak{g}$, it suffices to prove that for $\hat{f}\in E_\mathfrak{g}^{-1}(D(\delta_Q))$ one has

$$\left.\frac{d}{dt}\right|_{t=0}\hat{f}_t(F)=\lambda_Q\hat{f}(F)-\delta_\sigma(\beta_F^Q)(\hat{f}(F)),\quad\text{for } F\in E, \tag{4.10}$$

where the derivative should be taken in the norm topology of $\mathscr{A}$ (and uniformly in $F\in E$). By differentiation of (4.2) at $t=0$, we obtain

$$\left.\frac{d}{dt}\right|_{t=0}\hat{f}_t(F)=\left.\frac{d}{dt}\right|_{t=0}\hat{f}(\varphi_t^Q F)+\left.\frac{d}{dt}\right|_{t=0}\sigma(g_Q^{-1}(t,F))(\hat{f}(F))\,. \tag{4.11}$$

The first term on the right-hand-side of (4.11) gives the first term on the right-hand side of (4.10), which is an immediate consequence of the definition of λ_Q. We can write according to Remark 3.8,

$$\lambda_Q\hat{f}(F)=\sum_{j=1}^{n}[Q,F_j](F)\,\delta_j\hat{f}(F)\,. \tag{4.12}$$

The second term on the right-hand side of (4.11) is the derivative of the composite function $t\mapsto\sigma(g_0(t))$ with $g_0:=g_Q^{-1}$ at any fixed $F\in E$. The derivative of $t\mapsto g_Q^{-1}(t,F)$ is, according to (2.27), equal to $-\beta_F^Q\in\mathfrak{g}$. The derivative of $g\mapsto\sigma(g)$ at $g=e$ is the linear mapping δ_σ from the tangent space $\mathfrak{g}$ to G at the identity e to (in general unbounded) derivations $\delta_\sigma(\beta)$ ($\beta\in\mathfrak{g}$) on the C^*-algebra $\mathscr{A}$. The composition of these two differentiations gives the second term on the right-hand side of (4.10). This result can be obtained in a more explicit way by introduction of normal coordinates on a neighborhood of the identity of G. Q.E.D.

Corollary 4.6: Let $\mathscr{A}$ and $\sigma(G)$ be defined as in Sec. II. Let τ^Q of (4.2) and (4.6) correspond to $\beta_F^Q=d_FQ$ ($F\in\mathfrak{g}^*$) for any differentiable $Q\in C(E)$. Then the value of the derivation δ_Q from (4.9) taken for $\hat{f}\in C(E,\mathscr{A}^J)\cap E_\mathfrak{g}^{-1}(D(\delta_Q))$ ($J\in\Pi$) can be written in the form (3.27).

Proof: The first term in the right-hand side of (4.9) can be written in the form of the second term on the right-hand side of (3.27), in accordance with (4.12). The linearity of δ_σ as well as an explicit expression of β_F^Q give us

$$\delta_\sigma(\beta_F^Q)(x)=\sum_{j=1}^{n}\delta_jQ(F)\,\delta_\sigma(\beta_j)(x),\quad x\in\mathscr{A}^J. \tag{4.13}$$

The derivations $\delta_\sigma(\beta)$ ($\beta\in\mathfrak{g}$) are easily calculated from (2.1) and (2.2). For $x\in\mathscr{A}^J$ one has

$$\delta_\sigma(\beta)(x)=-i[X^J(\beta),x]\,. \tag{4.14}$$

Insertion of the obtained expressions into (4.9) gives (3.27). Q.E.D.

According to the Hille–Yosida theorem,[28] a continuous one-parameter group τ^Q of automorphisms of a C^*-algebra $\mathscr{C}$ is determined uniquely by determination of its generator δ_Q on some of its cores in $\mathscr{C}$. Hence the preceding corollary shows that the set of evolution groups τ^Q on $\mathscr{C}$ defined in this section contains the subset of evolutions obtained in Sec. III as thermodynamic limits of local Hamiltonian evolutions. The specification of the general case to the models considered in Sec. III consists of (i) substitution for $\mathscr{A}$ of the infinite tensor product C^*-algebra $\otimes_{p\in\Pi}\mathscr{L}(H_p)$ determined in Sec. II; (ii) choice of $\sigma(G)$ in the form (2.1) with a norm-continuous unitary representation $U(G)$; (iii) taking $\beta_F^0:=0$ (for all $F\in E$) in (2.29); and (iv) the choice of $Q\in C(E)$ to be a polynomial in variables $F_j:=F(\beta_j)$, $j=1,2,\ldots,n=\dim G$.

A general discussion of equilibrium thermodynamics of the considered models, and an analysis of specific simple examples is supposed to be published in forthcoming papers, cf. also Ref. 25(a).

ACKNOWLEDGMENTS

The author is indebted to Professor E. Brünning, Professor H. D. Doebner, Professor E. H. Lieb, Professor W. Thirring, and Professor J. Tolar for their interest in his work, to Professor K. Hepp for a remark stimulating the work on details of the subject of this paper, and to Professor J. T. Lewis for the encouragement to write this paper.

[1] P. Weiss, J. Phys. Rad. **6**, 667 (1907); H. E. Stanley, *Introduction to Phase Transitions and Critical Phenomena* (Clarendon, Oxford, 1971).
[2] J. Bardeen, L. U. Cooper, and J. R. Schrieffer, Phys. Rev. **106**, 162 (1955); **108**, 1175 (1957).
[3] D. J. Thouless, Phys. Rev. **117**, 1256 (1960).
[4] K. Baumann, G. Eder, R. Sexl, and W. Thirring, Ann. Phys. (NY) **16**, 14 (1961).
[5] P. W. Anderson, Phys. Rev. **110**, 827 (1958); **112**, 1900 (1958).
[6] N. N. Bogoliubov, Sov. Phys. JETP **34**, 58, 73 (1958); V. V. Tolmačev and S. V. Tiablikov, *ibid.* **34**, 66 (1958).
[7] R. Haag, Nuovo Cimento **25**, 287 (1962).
[8] W. Thirring and A. Wehrl, Commun. Math. Phys. **4**, 303 (1967); W. Thirring, *ibid.* **7**, 181 (1968).
[9] A. Wehrl, Commun. Math. Phys. **23**, 319 (1971).
[10] F. Jelinek, Commun. Math. Phys. **9**, 169 (1968).
[11] D. A. Dubin and G. L. Sewell, J. Math. Phys. **11**, 2990 (1970).
[12] M. Fannes, H. Spohn, and A. Verbeure, J. Math. Phys. **21**, 355 (1980); J. Quaegebeur and A. Verbeure, Ann. Inst. H. Poincarè A **XXXII**, 343 (1980).
[13] G. G. Emch and H. J. F. Knops, J. Math. Phys. **11**, 3008 (1970); G. G. Emch, "The C^*-algebraic approach to phase transitions," in *Phase Transitions and Critical Phenomena*, edited by C. Domb and M. S. Green (Academic, New York, 1972), Vol. 1.
[14] K. Hepp and E. H. Lieb, Helv. Phys. Acta **46**, 573 (1973).
[15] D. Ruelle, Commun. Math. Phys. **3**, 133 (1966).
[16] (a) O. Bratteli and D. W. Robinson, Commun. Math. Phys. **50**, 133 (1976); (b) H. Narnhofer, Acta Phys. Austriaca **49**, 207 (1978); (c) G. Lassner, in *Mathematical Physics VII*, Proceedings of the VIIth International Congress on Mathematical Physics, edited by W. E. Brittin, K. E. Gustafson, and W. Wyss (North-Holland, Amsterdam, 1984).
[17] L. van Hemmen, Fortschr. Phys. **26**, 397 (1978).

[18]E. Duffner, Physica A **133**, 187 (1985).
[19]H. Roos, Physica A **100**, 183 (1980).
[20]A. Rieckers, J. Math. Phys. **25**, 2593 (1984).
[21]W. Fleig, Acta Phys. Austriaca **55**, 135 (1983).
[22]A. Rieckers and M. Ullrich, Acta Phys. Austriaca **56**, 131, 259 (1985).
[23]A. Rieckers and M. Ullrich, J. Math. Phys. **27**, 1082 (1986).
[24]E. Duffner, Z. Phys. B **63**, 37 (1986).
[25](a) P. Bóna, *Classical Projections and Macroscopic Limits of Quantum Mechanical Systems*, preprint 1984 (a completed version 1986, submitted for publication); (b) Czech. J. Phys. **37**, 482 (1987).
[26]G. Morchio and F. Strocchi, J. Math. Phys. **28**, 622 (1987); F. Strocchi, "Long range dynamics and spontaneous symmetry breaking in many-body systems," preprint ISAS-Trieste, 1987.
[27]R. Haag and D. Kastler, J. Math. Phys. **5**, 848 (1964).
[28]O. Bratteli and D. W. Robinson, *Operator Algebras and Quantum Statistical Mechanics* (Springer, New York, 1979, 1981), Vols. I and II.
[29]G. G. Emch, *Algebraic Methods in Statistical Mechanics and Quantum Field Theory* (Wiley, New York, 1972).
[30]H. Primas, *Chemistry, Quantum Mechanics and Reductionism, Lecture Notes in Chemistry*, Vol. 24 (Springer, Berlin, 1981).
[31]Cf., however, Sec. II C, esp. Remarks 2.7 and 2.8.
[32]S. Sakai, *C^*-Algebras and W^*-Algebras* (Springer, Berlin, 1971).
[33]A. A. Kirillov, *Elementy teorii predstavlenii* (Nauka, Moscow, 1978).
[34]B. Kostant, *Quantization and Unitary Representations, Lecture Notes in Mathematics*, Vol. 170 (Springer, New York, 1970).
[35]C. M. Marle, "Poisson manifolds in mechanics," in *Bifurcation Theory, Mechanics and Physics*, edited by C. P. Bruter, A. Aragnol, and A. Lichnerowicz (Reidel, Dordrecht, 1983).
[36]O. Bratteli and D. W. Robinson, Commun. Math. Phys. **46**, 11 (1976).
[37]D. Goderis, A. Verbeure, and P. Vets, J. Math. Phys. **28**, 2250 (1987).
[38]M. Takesaki, *Theory of Operator Algebras I* (Springer, New York, 1979).
[39]G. K. Pedersen, *C^*-Algebras and their Automorphism Groups* (Academic, London, 1979).
[40]J. Dixmier, *Les C^*-Algèbres et leurs Representations* (Gauthier–Villars, Paris, 1969).
[41]See the Theorem in Sec. 5.2 of P. Bóna, "*Interaction of macrosystems with microobjects in quantum theory*," in Slovak, dissertation, Bratislava, 1974.
[42]F. Riesz and B. Sz. Nagy, *Vorlesungen über Funktionalanalysis* (DVW, Berlin, 1956).
[43]M. Reed and B. Simon, *Methods of Modern Mathematical Physics I*, Functional Analysis (Academic, New York, 1972).
[44](a) V. J. Arnol'd, *Matematičeskie osnovy klassičeskoi mechaniki* (Nauka, Moscow, 1979). (b) R. Abraham and J. E. Marsden, *Foundations of Mechanics* (Benjamin–Cummings, Reading, MA, 1978), 2nd ed. (c) Y. Choquet-Bruhat, C. DeWitt-Morette, and M. Dillard-Bleick, *Analysis, Manifolds and Physics* (North Holland, Amsterdam, 1982).
[45]A. O. Barut and R. Rączka, *Theory of Group Representations and Applications* (PWN–Polish Scientific, Warsaw, 1977).
[46]V. S. Varadarajan, *Geometry of Quantum Theory II* (Van Nostrand Reinhold, New York, 1970).
[47]L. Schwartz, *Analyse Mathématique II* (Hermann, Paris, 1967).
[48]See Sec. X.8 in M. Reed and B. Simon, *Methods of Modern Mathematical Physics II, Fourier Analysis, Self-adjointness* (Academic, New York, 1975).

Location of essential spectrum of intermediate Hamiltonians restricted to symmetry subspaces

Christopher Beattie
Department of Mathematics, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061

Mary Beth Ruskai
Department of Mathematics, University of Lowell, Lowell, Massachusetts 01854

(Received 5 February 1988; accepted for publication 25 May 1988)

A theorem is presented on the location of the essential spectrum of certain intermediate Hamiltonians used to construct lower bounds to bound-state energies of multiparticle atomic and molecular systems. This result is an analog of the Hunziker–Van Winter–Zhislin theorem for exact Hamiltonians, which implies that the continuum of an N-electron system begins at the ground-state energy for the corresponding system with $N-1$ electrons. The work presented here strengthens earlier results of Beattie [SIAM J. Math. Anal. **16**, 492 (1985)] in that one may now consider Hamiltonians restricted to the symmetry subspaces appropriate to the permutational symmetry required by the Pauli exclusion principle, or to other physically relevant symmetry subspaces. The associated convergence theory is also given, guaranteeing that all bound-state energies can be approximated from below with arbitrary accuracy.

I. INTRODUCTION

Variational techniques for obtaining upper bounds to eigenvalues of a multiparticle Hamiltonian H are well developed and can often yield quite accurate estimates to eigenvalues of interest. However, upper bounds alone cannot provide complete estimates of the error in the approximations to the eigenvalues. To do this one must bracket the eigenvalues of interest by also computing complementary lower bounds.

In general, the computational effort is greater for lower-bound estimation, and the related analysis more subtle, than that required in standard approaches for upper-bound estimation (such as Hartree–Fock and configuration interaction methods). Furthermore, lower-bound procedures usually require some form of additional *a priori* spectral information. For example, Temple's inequality[1,2] can often yield a reasonably tight lower bound to a particular eigenvalue provided that the eigenvalue of interest can be explicitly isolated from the next larger eigenvalue, which requires a good estimate on the next eigenvalue. Such needs for *a priori* spectral information often become problematic in practical circumstance. The method that we consider here, the method of intermediate Hamiltonians, has by contrast fairly relaxed requirements for *a priori* information, though effective use of this information may offer distinct computational challenges.

The method of intermediate Hamiltonians was used with great success by Bazley and Fox[3–6] to obtain lower bounds to He, and by Hill[7] to prove that H^- has only one bound state. Extensions to three-electron problems proved more difficult, although some results have been obtained.[8–11] This method requires a decomposition of the self-adjoint operator H as $H_0+\hat{H}$, where information on the discrete spectrum of H_0 is explicitly available and $\hat{H}$ is a symmetric positive-definite operator (i.e., $\hat{H}\geqslant 0$). Those eigenvalues of H_0 that lie below the infimum of the essential spectrum of H [i.e., the bottom of the continuum, denoted here as $\lambda_*(H)$] are lower bounds to the corresponding eigenvalues of H. Because these bounds invariably tend to be quite crude, one seeks improved bounds by carefully approximating $\hat{H}$ from below (in the sense of quadratic forms). As originally conceived, this was done with an increasing chain of positive semidefinite finite-rank operators. The resulting problem was equivalent to the evaluation of the spectrum of a degenerately perturbed operator with known spectrum.[12] A detailed discussion of intermediate operator methods can be found in Refs. 13–15.

The principal difficulty in applying standard intermediate operator techniques to multiparticle Hamiltonians is that the lowest point of the essential spectrum of the base operator, $\lambda_*(H_0)$, often lies below the lowest eigenvalue of H. Since finite-rank approximations to $\hat{H}$ produce compact perturbations of H_0 that leave the essential spectrum of H_0 unperturbed, the method as originally developed in Refs. 4 and 5 cannot yield convergent lower bounds. Fox[16] developed a modification of the standard intermediate operator approach utilizing *noncompact* perturbations of the base operator H_0, yet retaining the critical property of producing computationally resolvable intermediate operators. Recently Beattie[17] showed that a variant of Fox's construction yields intermediate Hamiltonians for which $\lambda_*(H_0)$ can be made arbitrarily close to $\lambda_*(H)$, the lowest point of the essential spectrum of the exact Hamiltonian. This allows, at least in principle, tight lower bounds to all eigenvalues of the Hamiltonian. These results were obtained for the full Hamiltonian operator without considering the permutational symmetries of the system. In calculations for real atomic and molecular systems, one wants to consider Hamiltonians that are restricted to appropriate symmetry subspaces so that the Pauli exclusion principle is satisfied. In this paper we show that the results of Ref. 17 can be extended to such symmetry-restricted Hamiltonians. The crucial point is to extend Beattie's analog of the Hunziker–Van Winter–Zhislin (HVZ)

 0022-2488/88/102236-05$02.50

theorem to intermediate Hamiltonians restricted to appropriate symmetry subspaces.

In addition, we discuss extensions to systems containing several species of identical particles, and to molecular systems. Finally, we show that the lower bounds obtained via this construction converge to the exact eigenvalues of H. Thus all bound states of H can, at least in principle, be approximated from below with arbitrary accuracy.

The behavior of a single particle with spin s is described by a Hamiltonian operator acting on a suitable dense subset of the Hilbert space $\mathscr{H} = L^2(\mathbb{R}^3) \otimes \mathbb{C}^{2s+1}$. For real electrons, which have two spin states, we have $\mathscr{H} = L^2(\mathbb{R}^3) \otimes \mathbb{C}^2$. The Hamiltonian for N identical particles then acts on an appropriate subspace of

$$\mathscr{H}^N = \mathscr{H} \otimes \mathscr{H} \otimes \mathscr{H} \otimes \mathscr{H} \otimes \cdots \otimes \mathscr{H} ,$$

i.e., the tensor product of N copies of $\mathscr{H}$. If the particles are bosons (s = integer), the N-particle Hamiltonian must be restricted to the symmetric subspace of $\mathscr{H}^N$, which we denote as $\mathscr{H}^N_+$. If the particles are fermions (s = half-integer), the N-particle Hamiltonian must be restricted to the antisymmetric subspace of $\mathscr{H}^N$, which we denote as $\mathscr{H}^N_-$.

The multiparticle Hamiltonians that we consider have the form

$$H_N = \sum_{i=1}^{N} [-\Delta_i + W(\mathbf{r}_i)] + \sum_{i<j} V(\mathbf{r}_i - \mathbf{r}_j) , \qquad (1)$$

where Δ_i is the three-dimensional Laplacian acting on coordinates of the ith particle, and W and V are suitable potential functions. The restriction of H_N to $\mathscr{H}^N_+$ or $\mathscr{H}^N_-$ will be denoted $H_{N,+}$ or $H_{N,-}$, respectively. For N electrons in the field of a fixed nucleus of charge Z, $W(\mathbf{r}) = -Z/r$ and $V(\mathbf{r}) = 1/r$, where $r = |\mathbf{r}|$. For molecular systems with M fixed nuclei of charge $Z_1,\ldots,Z_M$ at positions $\mathbf{R}_1,\ldots,\mathbf{R}_M$,

$$W(\mathbf{r}) = \sum_{J=1}^{M} \frac{-Z_J}{|\mathbf{r} - \mathbf{R}_J|} \quad \text{and} \quad V(\mathbf{r}) = \frac{1}{r} .$$

In the molecular case, tractable intermediate Hamiltonians appear to exist only for homonuclear diatomic molecules, in which case the Schrödinger equation for the Hamiltonian

$$h = -\Delta - Z/|\mathbf{r} - \mathbf{R}_1| - Z/|\mathbf{r} - \mathbf{R}_2|$$

can be solved exactly.

We require that the potentials in both the atomic and fixed-nuclei molecular cases satisfy conditions sufficient to assure convergent spectral approximations, as follows:

(a) W and $V \in L^2(\mathbb{R}^3) + [L^\infty(\mathbb{R}^3)]_\varepsilon$,

(b) $V \geqslant 0$ almost everywhere in $\mathbb{R}^3$,

(c) the self-adjoint operator corresponding to $h = -\Delta + W$ is bounded below and has as its spectrum negative eigenvalues of finite multiplicity, and essential spectrum $[0, \infty)$.

Although practical applications generally require that h actually have some negative eigenvalues, the analysis remains valid if some h_k possesses only essential spectrum $[0, \infty)$.

(In the case of diatomic molecules with finite nuclear mass, for example, it might actually be useful to consider $h_J = -\Delta_J \geqslant 0$; unfortunately, other difficulties prevent us from extending these techniques to molecular systems with finite nuclear mass at present. In the extension to several species of identical particles discussed at the end of Sec. IV, the charge of all particles must have the same sign.)

II. INTERMEDIATE HAMILTONIANS

Following earlier work, we consider intermediate Hamiltonians that can be expressed in the form

$$H_N^{k\alpha\beta} = \sum_{i=1}^{N} h_i^{k\alpha} + \sum_{i<j} v_{ij}^{\beta} , \qquad (2)$$

where $h_i^{k\alpha}$ and v_{ij}^{β} denote, respectively, approximations to $h_i = -\Delta_i + W(\mathbf{r}_i)$ and $V(\mathbf{r}_i - \mathbf{r}_j)$ having the following key properties.

(i) The approximation $h_i^{k\alpha}$ has the form $T_{k\alpha} + \lambda_{k+1}(h) E_k^{\perp}$, where $\lambda_{k+1}(h)$ is the kth eigenvalue of $h = -\Delta + W(\mathbf{r})$, E_k is the orthogonal projection onto the span of the eigenspaces corresponding to eigenvalues $\lambda_1,\ldots,\lambda_k$ of h, $E_k^{\perp} = I - E_k$, $T_{k\alpha}$ is symmetric and has finite rank $(k+\alpha)$ with range denoted $\Gamma_{k\alpha}$, and $T_{k\alpha} + \lambda_{k+1}(h) E_k^{\perp} < h$ in the sense of quadratic forms.

(ii) The approximation v_{ij}^{β} has finite rank with range $\Pi_\beta \otimes \Pi_\beta$ for some explicitly known β-dimensional space $\Pi_\beta \subset \mathscr{H}$, and $v_{ij}^{\beta} < V(\mathbf{r}_i - \mathbf{r}_j)$ in the sense of quadratic forms.

[Although the analysis in Ref. 17 is given for operators on $L^2(\mathbb{R}^3)$, it can readily be extended to operators on $\mathscr{H} = L^2(\mathbb{R}^3) \otimes \mathbb{C}^{2s+1}$. However, we note the following changes in notation:

$n \to k$, $A_0^n \to hE_k + \lambda_{k+1}(h) E_k^{\perp}$,

$A_0 \to h$, $\tilde{A}_0^n \to [h - \lambda_{k+1}(h)] E_k^{\perp}$,

$\lambda_0^k \to \lambda_k(h)$, $A_0^n + \tilde{A}_0^n Q^\alpha \to T_{k\alpha} + \lambda_{k+1}(h) I$,

$\mathscr{U}_k \to E_k \mathscr{H}$, $\operatorname{Span}_\alpha \{\tilde{A}_0^n q_\nu\} \to [h - \lambda_{k+1}(h)] E_k^{\perp} \Omega_\alpha$,

$A_{ij} \to V(\mathbf{r}_i - \mathbf{r}_j)$, $A_{ij} P^\beta \to v_{ij}^{\beta}$,

$\Pi_\beta \to \Pi_\beta$, $\mathscr{U}_k \vee \operatorname{span}_\alpha \{\tilde{A}_0^n q_\nu\} \to \Gamma_{k\alpha}$.]

The construction of $T_{k\alpha}$ depends on the spectral resolution of h and the choice of a finite-dimensional subspace $\Omega_\alpha \subset \mathscr{H}$:

$$T_{k\alpha} = hE_k + [h - \lambda_{k+1}(h)] E_k^{\perp} Q^\alpha ,$$

where Q^α is a nonorthogonal projection operator with range Q_α and kernel $\{[h - \lambda_{k+1}(h)] E_k^{\perp} \Omega^\alpha\}^{\perp}$. The approximation v_{ij}^{β} is defined similarly as $V(\mathbf{r}_i - \mathbf{r}_j) R_{ij}^{\beta}$, where R_{ij}^{β} is a nonorthogonal projection with range Λ_{ij}^{β} and kernel

$\{V(\mathbf{r}_i-\mathbf{r}_j)\Lambda_{ij}^{\beta}\}^{\perp}$. A more complete description and analysis may be found in Refs. 16–18.

Define the subspace $\mathscr{M}^{k\alpha\beta}=\Gamma_{k\alpha}\vee\Pi_{\beta}$. Notice that for a simple two-particle intermediate Hamiltonian given by $H_2^{k\alpha\beta}=h_i^{k\alpha}+h_j^{k\alpha}+v_{ij}^{\beta}$, a full set of explicit reducing spaces may be constructed as $\mathscr{M}^{k\alpha\beta}\otimes\mathscr{M}^{k\alpha\beta}$, $\mathscr{M}^{k\alpha\beta}\otimes[\mathscr{M}^{k\alpha\beta}]^{\perp}$, $[\mathscr{M}^{k\alpha\beta}]^{\perp}\otimes\mathscr{M}^{k\alpha\beta}$, and $[\mathscr{M}^{k\alpha\beta}]^{\perp}\otimes[\mathscr{M}^{k\alpha\beta}]^{\perp}$. It is not hard to see that on the last subspace, $H_2^{k\alpha\beta}$ reduces to a scalar multiple of the identity I, while on the first subspace it is essentially a matrix operator. On the remaining two subspaces, $H_2^{k\alpha\beta}$ is effectively a direct product of a matrix operator and a scalar multiple of I. This means that the spectrum of $H_2^{k\alpha\beta}$ can be computed explicitly through a matrix diagonalization. However, these subspaces will not be reducing subspaces for $H_{2,\pm}^{k\alpha\beta}$. To obtain reducing subspaces with the correct permutational symmetry, the first and last subspaces above must be replaced by their symmetric or antisymmetric components, while the middle two spaces must be replaced by

$$\mathscr{K}_{2,1,\pm}^{k\alpha\beta}=\mathscr{M}^{k\alpha\beta}\otimes[\mathscr{M}^{k\alpha\beta}]^{\perp}\pm[\mathscr{M}^{k\alpha\beta}]^{\perp}\otimes\mathscr{M}^{k\alpha\beta}.$$

The construction of reducing subspaces in the N-particle case follows a similar pattern. To construct reducing subspaces for $H_{N,\pm}^{k\alpha\beta}$ we let

$$\mathscr{W}_{\xi_1,\ldots,\xi_r}^{k\alpha\beta}=\mathscr{N}_1\otimes\mathscr{N}_2\otimes\cdots\otimes\mathscr{N}_N \tag{3}$$

describe a subspace of $\mathscr{H}^N$ with

$$\mathscr{N}_i=\mathscr{M}^{k\alpha\beta},\quad \text{if } i\in\{\xi_1,\ldots,\xi_r\},$$

and

$$\mathscr{N}_i=[\mathscr{M}^{k\alpha\beta}]^{\perp},\quad \text{if } i\notin\{\xi_1,\ldots,\xi_r\}.$$

If permutational symmetry is not considered, then $\mathscr{W}_{\xi_1,\ldots,\xi_r}^{k\alpha\beta}$ will be a reducing subspace for $H_N^{k\alpha\beta}$. However, $\mathscr{W}_{\xi_1,\ldots,\xi_r}^{k\alpha\beta}$ will not be a reducing subspace for the symmetry-restricted intermediate Hamiltonians $H_{N,\pm}^{k\alpha\beta}$. Therefore we now define

$$\mathscr{K}_{N,r}^{k\alpha\beta}=\bigoplus_{\xi_1,\ldots,\xi_r}\mathscr{W}_{\xi_1,\ldots,\xi_r}^{k\alpha\beta},$$

i.e., $\mathscr{K}_{N,r}^{k\alpha\beta}$ is the span of the unions of all subspaces of the form (3) with exactly r copies of $\mathscr{M}^{k\alpha\beta}$ and $N-r$ copies of $[\mathscr{M}^{k\alpha\beta}]^{\perp}$. Let $\mathscr{K}_{N,r,\pm}^{k\alpha\beta}$ denote the symmetric and antisymmetric subspaces of $\mathscr{K}_{N,r}^{k\alpha\beta}$. Then $\mathscr{K}_{N,r,\pm}^{k\alpha\beta}$ is a reducing subspace for $H_{N,\pm}^{k\alpha\beta}$ for all r, and

$$\mathscr{H}_{\pm}^{N}=\bigoplus_{r=0}^{N}\mathscr{K}_{N,r,\pm}^{k\alpha\beta}. \tag{4}$$

III. LOCATION OF ESSENTIAL SPECTRUM

Let $\lambda_*(A)=\inf\sigma_{\mathrm{ess}}(A)$ and $\lambda_1(A)=\inf\sigma(A)$. For Hamiltonians of the form (1), the celebrated HVZ theorem[19] is equivalent to the statement

$$\lambda_*(H_N)=\lambda_1(H_{N-1}),$$

i.e., the essential spectrum of the Hamiltonian of N particles in the field of a fixed nucleus (or several fixed nuclei) begins at the lowest eigenvalue of the corresponding system with $N-1$ particles. In the case of symmetry-restricted Hamiltonians, $H_{N,\pm}$, this result becomes

$$\lambda_*(H_{N,\pm})=\lambda_1(H_{N-1,\pm}).$$

Beattie[17] gave an analogous formula for intermediate Hamiltonians without symmetry restrictions. The extension of Beattie's formula to symmetry-restricted intermediate Hamiltonians is the following theorem.

Theorem 1: $\lambda_*(H_{N,\pm}^{k\alpha\beta})=\lambda_1(H_{N-1,\pm}^{k\alpha\beta})+\lambda_{k+1}(h)$.

Before proving this result, it will be useful to make some observations about the spectral properties of $H_{N,r,\pm}^{k\alpha\beta}$, by which we denote the restriction of $H_N^{k\alpha\beta}$ to $\mathscr{K}_{N,r,\pm}^{k\alpha\beta}$. All eigenfunctions of $H_{N,r,\pm}^{k\alpha\beta}$ have the form

$$\wp_{\pm}\{G(x_1,\ldots,x_r)g_1(x_{r+1})\cdots g_{N-r}(x_N)\}, \tag{5}$$

where $x_i=(\mathbf{r}_i,s_i)$ represents the space and spin coordinates of the ith particle, G is an eigenfunction of $H_{r,r,\pm}^{k\alpha\beta}$, the g_i are in $[\mathscr{M}^{k\alpha\beta}]^{\perp}$, and $\wp_{\pm}$ projects onto $\mathscr{H}_{\pm}^{N}$. Thus $H_{N,r,\pm}^{k\alpha\beta}$ has pure point spectrum consisting of eigenvalues of the form

$$\text{e-val}(H_{N,r,\pm}^{k\alpha\beta})=\text{e-val}(H_{r,\pm}^{k\alpha\beta})+(N-r)\lambda_{k+1}(h). \tag{6}$$

However, because $h^{k\alpha}$ is a multiple of the identity on the infinite-dimensional space $[\mathscr{M}^{k\alpha\beta}]^{\perp}$, the eigenvalues of $H_{N,r,\pm}^{k\alpha\beta}$ have (i) finite multiplicity, if $r=N$; and (ii) infinite multiplicity, if $r<N$. Thus $H_{N,N,\pm}^{k\alpha\beta}$ has only discrete spectrum, while $H_{N,r,\pm}^{k\alpha\beta}$ has only essential spectrum when $N>r$.

To prove Theorem 1, we first note that (4) implies

$$\lambda_1(H_{n,\pm}^{k\alpha\beta})=\min_{r=0,\ldots,n}\lambda_1(H_{n,r,\pm}^{k\alpha\beta}). \tag{7}$$

Applying a similar analysis to $\lambda_*(H_{N,\pm}^{k\alpha\beta})$ and using (7) along with the observations above, we find that

$$\begin{aligned}
\lambda_*(H_{N,\pm}^{k\alpha\beta})&=\min_{r=0,\ldots,N}\lambda_*(H_{N,r,\pm}^{k\alpha\beta})\\
&=\min_{r=0,\ldots,N-1}\lambda_*(H_{N,r,\pm}^{k\alpha\beta})\\
&=\min_{r=0,\ldots,N-1}\{\lambda_1(H_{r,r,\pm}^{k\alpha\beta})\\
&\quad+(N-r)\lambda_{k+1}(h)\}\\
&=\min_{r=0,\ldots,N-1}\{\lambda_1(H_{r,r,\pm}^{k\alpha\beta})\\
&\quad+[(N-1)-r]\lambda_{k+1}(h)\}+\lambda_{k+1}(h)\\
&=\min_{r=0,\ldots,N-1}\{\lambda_1(H_{N-1,r,\pm}^{k\alpha\beta})\}+\lambda_{k+1}(h)\\
&=\lambda_1(H_{N-1,\pm}^{k\alpha\beta})+\lambda_{k+1}(h).
\end{aligned}$$

IV. EXTENSION TO SYMMETRY SUBSPACES

Our analysis thus far has not explicitly considered spin. The spin was present implicitly by the inclusion of $\mathbf{C}^{2s+1}$ in $\mathscr{H}=L^2(\mathbf{R}^3)\otimes\mathbf{C}^{2s+1}$. When spin is explicitly considered, it suffices to use trial functions Ψ in $\mathscr{H}_{\pm}^{N}$ that have the form

$$\Psi=\sum_{t=1}^{\tau}\Phi_t(r_1,\ldots,r_N)\phi_t(s_1,\ldots,s_N), \tag{8}$$

where $\Phi_t\in[L^2(\mathbf{R}^3)]^N$ and $\phi_t\in[\mathbf{C}^{2s+1}]^N$, and for which $\{\Phi_t\}$ and $\{\phi_t\}$ are bases of irreducible representations of the symmetric group S_N. It is then natural to ask if Theorem 1 can be extended to this situation, i.e., to $H_{N,\sigma}$ defined as the operator H_N restricted to the subspace of $[L^2(\mathbf{R}^3)]^N$ corresponding to the irreducible representation σ of S_N. Before showing that such an extension is possible, we point out that Theorem 1 actually suffices for most practical calculations. If one is interested in bound states belonging to a particular subspace σ, Theorem 1 implies that it suffices to consider

intermediate Hamiltonians restricted to the subspace $\mathscr{H}_\sigma^N$, provided that the eigenvalues of interest lie below $\lambda_*(H_{N,\pm})$. However, the extension to other irreducible representations, which we give here, is useful for several reasons.

(a) One is occasionally interested in bound states embedded in the continuum, i.e., in energies E_σ, which lie in the region $\lambda_*(H_{N,\pm}) < E_\sigma < \lambda_*(H_{N,\sigma})$. (Hill's proof[9] that H^{--} has no bound states in the quartet sector is an example of such a situation.)

(b) A similar analysis can be applied to other physically relevant symmetries besides permutational symmetry.

(c) This analysis can be extended to consider several species of particles, as described below.

(d) The spectrum of $\mathscr{H}_\sigma^N$ is a subset of that for $\mathscr{H}_\pm^N$. Hence restriction to $\mathscr{H}_\sigma^N$ can reduce the density of eigenvalue clusters and increase the gap between computed eigenvalues, as compared to that when $\mathscr{H}_\pm^N$ is used. This improves both the conditioning and convergence rate of computational algorithms used ultimately to resolve the final matrix eigenvalue problem.[20]

Sigalov and Sigal[21,22] have shown how to extend the HVZ theorem to Hamiltonians restricted to symmetry subspaces. We summarize their analysis for Hamiltonians of the form $H_{N,\sigma}$. Let σ and ω denote irreducible representations of S_N and S_n, respectively, with $n < N$, so that S_n is isomorphic to a subgroup of S_N. Let $\omega < \sigma$ indicate that the irreducible representation ω is present in the decomposition of σ restricted to S_n. Sigalov and Sigal[21,22] showed that, in the case of Hamiltonians of type (1),

$$\lambda_*(H_{N,\sigma}) = \min_{\omega<\sigma} \lambda_1(H_{N-1,\omega}) . \tag{9}$$

The corresponding generalization to intermediate Hamiltonians is the next theorem.

Theorem 2:

$$\lambda_*(H_{N,\sigma}^{k\alpha\beta}) = \min_{\omega<\sigma} \lambda_1(H_{N-1,\omega}^{k\alpha\beta}) + \lambda_{k+1}(h) . \tag{10}$$

Proof: We first note that eigenfunctions of $H_{N,\sigma}^{k\alpha\beta}$ have the form

$$\wp_\sigma\{G(x_1,\ldots,x_r) g_1(x_{r+1}) \cdots g_{N-r}(x_N)\} ,$$

where the notation is as in Theorem 1, except that σ denotes the restriction to the subspace $\mathscr{H}_\sigma^N$ corresponding to σ, and G is an eigenfunction of $H_{r,r,\omega}^{k\alpha\beta}$ with $\omega < \sigma$. The proof of Theorem 1 can then be easily extended to this more general case. We omit the details.

It should be clear that our analysis could easily be extended to symmetry subspaces corresponding to several species of identical particles instead of N electrons, e.g., N_1 electrons and N_2 muons with $N_1 + N_2 = N$. In this case, $-\Delta_i$ would be replaced by $-\Delta_i/m_i$ in (1), where m_i is the mass of particle i, and H_N would act on $\mathscr{H}_-^{N_1} \otimes \mathscr{H}_-^{N_2}$.

V. CONVERGENCE

The question of what conditions on the approximating subspaces Λ_{ij}^β and Ω_α are sufficient to guarantee convergent estimates has been addressed in more general settings by Beattie,[18] Greenlee,[23] Beattie and Greenlee,[24] and Brown.[25] In our setting, the derived density criteria sufficient to guarantee convergence may be succinctly stated:

(1) $\lim \Lambda_{ij}^\beta$ is dense in $[W^{2,2}(\mathbf{R}^3) \otimes \mathbf{C}^{2s+1}] \otimes [W^{2,2}(\mathbf{R}^3) \otimes \mathbf{C}^{2s+1}]$, (11)

(2) $\lim \Omega_\alpha$ is dense in $W^{2,2}(\mathbf{R}^3) \otimes \mathbf{C}^{2s+1}$, (12)

where $W^{2,2}(\mathbf{R}^3)$ denotes the second Sobolev space embedded in $L^2(\mathbf{R}^3)$. In particular, these density conditions induce core conditions within every symmetry subspace $\mathscr{H}_\sigma^N$ sufficient to produce convergent spectral estimates for $H_{N,\sigma}^{k\alpha\beta}$, provided one additional hypothesis holds. The convergence results cited above all carry the proviso that the lowest point of the essential spectrum of the intermediate operators must move up sufficiently to expose the eigenvalues of $H_{N,\sigma}$ to convergent estimates. That such movement can be guaranteed for the full unrestricted Hamiltonian operator was shown in Refs. 18 and 24. We show here that this is also the case for Hamiltonians restricted to symmetry subspaces.

Theorem 3: Under the density hypotheses (11) and (12), the family of intermediate Hamiltonians $\{H_{N,\sigma}^{k\alpha\beta}\}_{k\alpha\beta}$ provides convergent lower-bound estimates to the lowest point of the essential spectrum of $H_{N,\sigma}$:

$$\lim_{k\alpha\beta} \lambda_*(H_{N,\sigma}^{k\alpha\beta}) = \lambda_*(H_{N,\sigma}) . \tag{13}$$

As a consequence, every lower eigenvalue of $H_{N,\sigma}$ is accessible to convergent estimates, and

$$\lim_{k\alpha\beta} \lambda_i(H_{N,\sigma}^{k\alpha\beta}) = \lambda_i(H_{N,\sigma}) , \tag{14}$$

for every i such that $\lambda_i(H_{N,\sigma}) < \lambda_*(H_{N,\sigma})$.

Proof: Consider first the case $N = 2$. We have from (10)

$$\begin{aligned}\lambda_*(H_{2,\sigma}^{k\alpha\beta}) &= \min_{\omega<\sigma} \lambda_1(H_{1,\omega}^{k\alpha\beta}) + \lambda_{k+1}(h) \\ &= \lambda_1(H_1^{k\alpha\beta}) + \lambda_{k+1}(h) \\ &= \lambda_1(h_1^{k\alpha}) + \lambda_{k+1}(h) \\ &= \lambda_1(h) + \lambda_{k+1}(h) = \lambda_*(H_{2,\sigma}) + \lambda_{k+1}(h) .\end{aligned}$$

Since $\lambda_{k+1}(h) \to 0$ as $k \to \infty$, (13) holds for $N = 2$, and the density conditions (11) and (12) guarantee that (14) holds as well. Now make an induction hypothesis and suppose that (13) and (14) hold for all $N < M$, for some $M > 1$. We may then deduce that

$$\begin{aligned}\lim_{k\alpha\beta} \lambda_*(H_{M,\sigma}^{k\alpha\beta}) &= \lim_{k\alpha\beta} \min_{\omega<\sigma} \lambda_1(H_{M-1,\omega}^{k\alpha\beta}) \\ &= \min_{\omega<\sigma} \lim_{k\alpha\beta} \lambda_1(H_{M-1,\omega}^{k\alpha\beta}) \\ &= \min_{\omega<\sigma} \lambda_1(H_{M-1,\omega}) = \lambda_*(H_{M,\sigma}) .\end{aligned}$$

Hence (13) holds for $N = M$, the density criteria again imply (14) for $N = M$, and the induction step is completed.

ACKNOWLEDGMENTS

This work was done in part while one of the authors (C.B.) was a Visiting Fellow with the Minnesota Supercomputer Institute and the Department of Computer Science at the University of Minnesota, Minneapolis. Much of this work was done during the Workshop on Atomic and Molecular Dynamics held at the Institute for Mathematics and Its Applications at the University of Minnesota from June 16 to July 24, 1987.

The research of M.B.R. was supported in part by National Science Foundation Grant No. DMS-8708905.

[1]G. Temple, Proc. R. Soc. London Ser. A **119**, 276 (1928). See also Refs. 13, 14, and 19 below.
[2]For a summary and comparison of several lower-bound techniques, see R. N. Hill, J. Math. Phys. **21**, 2182 (1980).
[3]N. W. Bazley, Phys. Rev. **120**, 144 (1960).
[4]N. W. Bazley and D. W. Fox, J. Res. Natl. Bur. Stand. B **65**(2), 105 (1961).
[5]N. W. Bazley and D. W. Fox, Phys. Rev. **124**, 483 (1961).
[6]N. W. Bazley and D. W. Fox, Arch. Rational Mech. Anal. **10**, 352 (1962).
[7]R. N. Hill, Phys. Rev. Lett. **38**, 643 (1977); J. Math Phys. **18**, 2316 (1977).
[8]D. W. Fox and V. G. Sigillito, Z. Angew. Math. Phys. **23**, 392 (1972).
[9]R. N. Hill, *Mathematical Problems in Theoretical Physics, Lecture Notes in Physics*, Vol. 116, edited by K. Osterwalder (Springer, Berlin, 1979), pp. 52–56.
[10]D. Freund, Ph.D. thesis, University of Delaware, Newark, 1982.
[11]D. Russell and W. M. Greenlee, Phys. Rev. Lett. **54**, 665 (1985).
[12]T. Kato, *Perturbation Theory for Linear Operators* (Springer, Berlin, 1976), 2nd ed.
[13]A. Weinstein and W. Stenger, *Methods of Intermediate Problems for Eigenvalues: Theory and Ramifications* (Academic, New York, 1972).
[14]H. F. Weinberger, *Variational Methods for Eigenvalue Problems* (SIAM, Philadelphia, 1974).
[15]W. Stenger, "Intermediate Problems for Eigenvalues in Quantum Theory," Chap. 14 of *The Uncertainty Principle and Foundations of Quantum Mechanics*, edited by W. C. Price and S. S. Chissick (Wiley, New York, 1977).
[16]D. W. Fox, SIAM J. Math. Anal. **3**, 617 (1972).
[17]C. Beattie, SIAM J. Math. Anal. **16**, 492 (1985).
[18]C. Beattie, "Some convergence results for intermediate problems that displace essential spectra," Applied Physics Laboratory Report No. 65, July 1982, Johns Hopkins University, Laurel.
[19]See, for example, M. Reed and B. Simon, *Methods of Mathematical Physics IV: Analysis of Operators* (Academic, New York, 1978) or W. Thirring, *A Course in Mathematical Physics III: Quantum Mechanics of Atoms and Molecules* (Springer, New York, 1979).
[20]B. Parlett, *The Symmetric Eigenvalue Problem* (Prentice–Hall, Englewood Cliffs, NJ, 1980).
[21]A. G. Sigalov and I. M. Sigal. Theor. Math. Phys. **5**, 990 (1970).
[22]I. M. Sigal, Commun. Math. Phys. **85**, 309 (1982).
[23]W. M. Greenlee, Arch. Rational Mech. Anal. **81**, 279 (1983).
[24]C. Beattie and W. M. Greenlee, Proc. R. Soc. Edinburgh A **100**, 107 (1985); **104** 349 (1986).
[25]R. D. Brown, "Convergence criteria for Aronszajn's method and for the Bazley–Fox method," Proc. R. Soc. Edinburgh (to be published).

Finitely many sphere interactions in quantum mechanics: Nonseparated boundary conditions

L. Dabrowski[a)]
International School for Advanced Studies, 34100 Trieste, Italy

J. Shabani[b)]
International Centre for Theoretical Physics, 34100 Trieste, Italy

(Received 19 February 1988; accepted for publication 11 May 1988)

Let $\dot{H}$ be the closure of the restriction of the three-dimensional Laplacian $-\Delta$ on the domain $C_0^\infty(R^3\setminus\Sigma)$, where $\Sigma=\cup_{j=1}^N \partial\overline{K(0,R_j)}$ and $\overline{K(0,R_j)}$ is a closed ball of radius R_j centered at the origin in R^3. It is well known that $\dot{H}$ is a closed symmetric operator with deficiency indices (∞,∞). In this paper all self-adjoint (s.a.) extensions of $\dot{H}$ are constructed; these extensions contain as particular cases the quantum Hamiltonian describing concentric δ- and δ'-sphere interactions. It is also shown that the s.a. extensions of $\dot{H}$ may be obtained as norm-resolvent limits of momentum cutoff and scaled separable potentials.

I. INTRODUCTION

In recent years there has been a lot of interest in studying sphere interactions in quantum mechanics; see Refs. 1–3 and the references therein.

Consider in $L^2(R^3)$ the Laplacian $-\Delta$ on the domain $D=C_0^\infty(R^3\setminus\cup_{j=1}^N\partial\overline{K(0,R_j)})$, where $K(0,R_j)$ is a closed ball of radius R_j centered at the origin in R^3, and denote by $\dot{H}$ the operator $\dot{H}=-\overline{\Delta_{|D}}$. In Ref. 2 it is shown that $\dot{H}$ is a symmetric operator with deficiency indices (∞,∞), and that $\dot{H}$ may be written in the form

$$\dot{H}=\bigoplus_{l=0}^{\infty}\tilde{U}^{-1}\dot{h}_{l,\{R\}}\tilde{U}\otimes 1, \tag{1.1}$$

where the transformation $\tilde{U}$ is given by (2.3), and, for fixed $l\in\mathbf{N}_0$, the operator $\dot{h}_{l,\{R\}}$ defined by (2.6) and (2.7) is a symmetric operator with deficiency indices (N,N). Consequently $\dot{h}_{l,\{R\}}$ admits an N^2-parameter family of self-adjoint (s.a.) extensions. A particular N-parameter subfamily of s.a. extensions corresponding to concentric δ-sphere interactions is discussed in Ref. 2.

In this paper we study the general N^2-parameter family of s.a. extensions of $\dot{h}_{l,\{R\}}$. In Sec. II, using the general theory of s.a. extensions of symmetric operators[4] and the decomposition (1.2), we obtain all s.a. extensions of $\dot{h}_{l,\{R\}}$ and $\dot{H}$, respectively. Furthermore we show that these extensions may be obtained as norm-resolvent limits of momentum cutoff (Sec. III) and scaled separable potentials (Sec. IV).

II. CONSTRUCTION OF SELF-ADJOINT EXTENSIONS

Consider in $L^2(R^3)$ the closed, non-negative minimal operator

$$\dot{H}=-\overline{\Delta C_0^\infty(R^3\setminus\cup_{j=1}^N\partial\overline{K(0,R_j)})},\quad 1\leqslant j\leqslant N, \tag{2.1}$$

where $\overline{K(0,R_j)}$ is a closed ball of radius R_j centered at the origin in R^3, and

$$\Delta=\frac{\partial^2}{\partial x_1^2}+\frac{\partial^2}{\partial x_2^2}+\frac{\partial^2}{\partial x_3^2}$$

is the Laplacian.

Following, e.g., Ref. 5, p. 160, one can decompose $L^2(R^3)$ with respect to angular momenta,

$$L^2(R^3)=L^2((0,\infty);r^2\,dr)\otimes L^2(S^2) \tag{2.2}$$

(S^2 is the unit sphere in R^3), and introduce the unitary transformation

$$\tilde{U}:\begin{cases}L^2((0,\infty);r^2\,dr)\to L^2((0,\infty)),\\ f\to(\tilde{U}f)(r)=rf(r),\quad r>0,\end{cases} \tag{2.3}$$

in order to get the following decomposition of $L^2(R^3)$:

$$L^2(R^3)=\bigoplus_{l=0}^{\infty}\tilde{U}^{-1}L^2((0,\infty);dr)\otimes[Y_l^m],\qquad l\in\mathbf{N}_0,\quad -l\leqslant m\leqslant l, \tag{2.4}$$

where $[Y_l^m]$ denotes the linear span of the spherical harmonics. With respect to the decomposition (2.4), $\dot{H}$ reads

$$\dot{H}=\bigoplus_{l=0}^{\infty}\tilde{U}^{-1}\dot{h}_{l,\{R\}}\tilde{U}\otimes 1, \tag{2.5}$$

where

$$\dot{h}_{l,\{R\}}=-\frac{d}{dr^2}+\frac{l(l+1)}{r^2}, \tag{2.6}$$

$$\begin{aligned}\mathscr{D}(\dot{h}_{l,\{R\}})=\{&f\in L^2((0,\infty))\,|\,f,f'\in AC_{\mathrm{loc}}((0,\infty));\\ &f(0_+)=0\ \text{ if } l=0;\quad f(R_{j\pm})=0;\\ &-f''+l(l+1)r^{-2}f\in L^2((0,\infty))\},\\ &l\in\mathbf{N}_0,\quad 1\leqslant j\leqslant N,\quad \{R\}=\{R_1,\ldots,R_N\}.\end{aligned} \tag{2.7}$$

Following Ref. 2 one can show that $\dot{h}_{l,\{R\}}$ has deficiency indices (N,N), and that the deficiency subspace $N_{-\bar{k}}$ is spanned by the N linearly independent functions

$$\phi_{l,j}(k,r)=\begin{cases}(i\pi/2)R_j^{1/2}H^{(1)}_{l+1/2}(kR_j)r^{1/2}J_{l+1/2}(kr), & r\leqslant R_j,\\ (i\pi/2)R_j^{1/2}J_{l+1/2}(kR_j)r^{1/2}H^{(1)}_{l+1/2}(kr), & r\geqslant R_j,\end{cases}\quad \operatorname{Im}k>0,\quad 1\leqslant j\leqslant N, \tag{2.8}$$

[a)] On leave of absence from IFT, Uniwersytet Wroctawski, Poland.
[b)] Permanent address: Department of Mathematics, University of Burundi, B. P. 2700 Bujumbura, Burundi.

 0022-2488/88/102241-04$02.50

where $J_\nu(z)$ and $H_\nu^{(1)}(z)$ are, respectively, Bessel and Hankel functions of order ν (see Ref. 6). Therefore all self-adjoint extensions of $\dot{h}_{l,\{R\}}$ are given by an N^2-parameter family of s.a. operators.

A particular N-parameter family of s.a. extensions of $\dot{h}_{l,\{R\}}$ corresponding to the quantum Hamiltonian describing N δ-interactions with supports on concentric spheres of radii $0<R_1<R_2\cdots<R_N$ has been extensively studied in Ref. 2. In this paper we investigate to which situation the other s.a. extensions of $\dot{h}_{l,\{R\}}$ correspond. We will follow the strategy of Ref. 7 in the analogous treatment of point interactions (cf. also Ref. 8).

From the general theory of s.a. extensions of symmetric operators[4] it follows that the s.a. extensions $h_{l,U,\{R\}}$ of $\dot{h}_{l,\{R\}}$ are given by

$$\mathscr{D}(h_{l,U,\{R\}})=\left\{g+\sum_{j=1}^{N}C_j\left[\phi_{l,j+}+\sum_{f=1}^{N}U_{jf}\phi_{l,f-}\right]\;\middle|\; g\in\mathscr{D}(\dot{h}_{l,\{R\}}),\quad C_j\in C\right\},\tag{2.9}$$

$$h_{l,U,\{R\}}\left\{g+\sum_{j=1}^{N}C_j\left[\phi_{l,j+}+\sum_{f=1}^{N}U_{jf}\phi_{l,f-}\right]\right\}=\dot{h}_{l,\{R\}}\,g+i\sum_{j=1}^{N}C_j\left[\phi_{l,j+}-\sum_{f=1}^{N}U_{jf}\phi_{l,f-}\right],\tag{2.10}$$

where U_{jf}, $1\leqslant j,j'\leqslant N$, denotes a unitary matrix in C^N, and $\phi_{l,j\pm}=\phi_{l,j}(\sqrt{\pm i},r)$, $\mathrm{Im}\sqrt{\pm i}>0$ provide a basis of $\ker[\dot{h}_{l,\{R\}}^{*}\mp i]$, respectively.

The case $U=-1$, i.e., $U_{jf}=-\delta_{jf}$, gives the free kinetic-energy Hamiltonian for fixed angular momentum l:

$$h_{l,-1,\{R\}}=h_{l,0}=-\frac{d^2}{dr^2}+\frac{l(l+1)}{r^2},\tag{2.11}$$

$$\mathscr{D}(h_{l,0})=\{f\in L^2((0,\infty))\,|\,f,f'\in AC_{\mathrm{loc}}((0,\infty));\ f(0_+)=0 \text{ if } l=0,\ -f''+l(l+1)r^{-2}f\in L^2((0,\infty))\},\quad l\in N_0.\tag{2.12}$$

From the above analysis we obtain a family of s.a. extensions of $\dot{H}$ given by

$$H_{U,\{R\}}=\bigoplus_{l=0}^{\infty}\tilde{U}^{-1}h_{l,U,\{R\}}\tilde{U}\otimes 1.\tag{2.13}$$

The particular case $U=-1$ yields the kinetic-energy operator

$$H_{-1,\{R\}}=-\Delta,\quad \mathscr{D}(-\Delta)=H^{2,2}(R^3)\tag{2.14}$$

[$H^{m,n}(R^3)$ being the standard Sobolev space[9]].

We note that the above treatment trivially generalizes to $n\geqslant 2$ dimensions using, e.g., Remark 2.1 in Ref. 1. Applying now Krein's formula[4] we obtain

$$(h_{l,U,\{R\}}-k^2)^{-1}=(h_{l,0}-k^2)^{-1}+\sum_{j,f=1}^{N}[M_l(k)]_{jf}\times(\phi_{l,f}(-\bar{k}),\cdot)\phi_{l,j}(k),\quad k^2\in\rho(h_{l,U,\{R\}}),\ \mathrm{Im}\,k>0,\ U\neq -1,\ l\in N_0,\tag{2.15}$$

where $\phi_{l,j}(k,r)$ is defined by (2.8) and

$$[M_l(k)]_{jf}^{-1}-[M_l(k')]_{jf}^{-1}=-(k^2-k'^2)(\phi_{l,f}(-\bar{k}),\phi_{l,j}(k'))=-(k^2-k'^2)\tilde{S}_l(k,k'),\quad k^2,k'^2\in\rho(h_{l,U,\{R\}}),\ \mathrm{Im}\,k>0,\ \mathrm{Im}\,k'>0.\tag{2.16}$$

Using the first resolvent formula

$$(k^2-k'^2)g_{l,k}g_{l,k'}=g_{l,k}-g_{l,k'},\quad \mathrm{Im}\,k>0,\ \mathrm{Im}\,k'>0,\tag{2.17}$$

where

$$g_{l,k}=(h_{l,0}-k^2)^{-1},\quad \mathrm{Im}\,k>0,\tag{2.18}$$

denotes the free resolvent with kernel

$$g_{l,k}(r,r')=\begin{cases}(i\pi/2)r^{1/2}H^{(1)}_{l+1/2}(kr)r'^{1/2}J_{l+1/2}(kr'), & r'\leqslant r,\\ (i\pi/2)r'^{1/2}H^{(1)}_{l+1/2}(kr')r^{1/2}J_{l+1/2}(kr), & r'\geqslant r\end{cases},\quad \mathrm{Im}\,k\geqslant 0,\tag{2.19}$$

we can rewrite (2.16) in the form

$$[M_l(k)]_{jf}^{-1}-[M_l(k')]_{jf}^{-1}=[g_l(k')]_{jf}-[g_l(k)]_{jf},\tag{2.20}$$

where we have used the notation

$$[g_l(k)]_{jf}=\phi_{l,j}(k,R_f)=g_{l,k}(R_j,R_f).\tag{2.21}$$

As a consequence of Eqs. (2.13) and (2.15) we infer that

$$(H_{U,\{R\}}-k^2)^{-1}=(H_{-1,\{R\}}-k^2)^{-1}+\bigoplus_{l=0}^{\infty}\bigoplus_{m=-l}^{l}\sum_{j,f=1}^{N}[M_l(k)]_{jf}(|\cdot|^{-1}\phi_{l,f}(-\bar{k})Y_l^m,\cdot)|\cdot|^{-1}\phi_{l,j}(k)Y_l^m,\quad k^2\in\rho(H_{U,\{R\}}),\ \mathrm{Im}\,k>0.\tag{2.22}$$

From the relation

$$(h_{l,U,\{R\}}-k^2)\left[\phi_{l,j+}+\sum_{f=1}^{N}U_{jf}\phi_{l,f-}\right]=(i-k^2)\phi_{l,j+}-(i+k^2)\sum_{f=1}^{N}U_{jf}\phi_{l,f-},\quad 1\leqslant j\leqslant N,\tag{2.23}$$

it follows that

$$(h_{l,U,\{R\}}+i)^{-1}\phi_{l,j+} = (2i)^{-1}\Big[\phi_{l,j+} + \sum_{j'=1}^{N} U_{jj'}\phi_{l,j'-}\Big]$$

$$= (h_{l,0}+i)^{-1}\phi_{l,j+} + \sum_{j,j'=1}^{N} \big[M_l(\sqrt{-i})\big]_{j,j'}(\phi_{l,j'+},\phi_{l,j+})\phi_{l,j-}$$

$$= (2i)^{-1}(\phi_{l,j+} - \phi_{l,j-}) + \sum_{j,j'=1}^{N} \big[M_l(\sqrt{-i})\big]_{jj'}(\phi_{l,j'+},\phi_{l,j+})\phi_{l,j'-}, \quad 1\leqslant j\leqslant N. \tag{2.24}$$

From (2.24) a straightforward computation yields

$$U_{jj'} + \delta_{jj'} = 2i\sum_{j''=1}^{N} (\phi_{l,j+},\phi_{l,j''+})^T \big[M_l(\sqrt{-i})\big]^T_{j''j'}, \quad 1\leqslant j,j'\leqslant N, \tag{2.25}$$

since the $\phi_{l,j-}$ are linearly independent. (Here $M^T_{jj'} = M_{j'j}$ denotes the transposed matrix in C^N.) We note that (2.25) may be rewritten in the form

$$\big[M_l(\sqrt{-i})\big]^{-1} = 2i(U^T+1)^{-1}\tilde{S}_l(\sqrt{-i},\sqrt{i}). \tag{2.26}$$

Therefore, using Eqs. (2.20) and (2.26), we infer

$$[(M_l(k)]^{-1} = (U^T+1)^{-1}\big[g_l(\sqrt{i}) - g_l(\sqrt{-i})\big] + g_l(\sqrt{-i}) - g_l(k). \tag{2.27}$$

Equations (2.15) and (2.27) imply that the spectrum of $h_{l,U,\{R\}}$ is given by the absolutely continuous spectrum of $h_{l,0}$, and that the point spectrum is determined by

$$\sigma_p(h_{l,U,\{R\}}) = \{k\in C \mid \det[M_l(k)]^{-1} = 0\}. \tag{2.28}$$

The particular case

$$[M_l(k)]^{-1}_{jj'} = [\alpha_{jl}^{-1}\delta_{jj'} + (g_l(k))_{jj'}]^N_{j'=1}, \quad -\infty<\alpha_{jl}\leqslant\infty, \tag{2.29}$$

gives rise to the Hamiltonian describing a finite number of δ-interactions with support on concentric spheres,[2] whereas

$$[M_l(k)]^{-1} = [\beta_{jl}^{-1}\delta_{jj'} - \tilde{\phi}'_{l,j}(k,R_{j'})]^N_{j'=1}, \quad -\infty<\beta_{jl}\leqslant\infty, \tag{2.30}$$

$$\tilde{\phi}_{l,j}(k,r) = \begin{cases} (i\pi/2)[r^{1/2}H^{(1)}_{l+1/2}(kr)]'|_{r=R_j} r^{1/2}J_{l+1/2}(kr), & r<R_j, \\ (i\pi/2)[r^{1/2}J_{l+1/2}(kr)]'|_{r=R_j} r^{1/2}H^{(1)}_{l+1/2}(kr), & r>R_j, \quad \mathrm{Im}\,k>0, \end{cases} \tag{2.31}$$

yields the Hamiltonian of $N\delta'$-interactions supported by concentric spheres.[3]

III. APPROXIMATION OF $h_{l,U,\{R\}}$ BY MOMENTUM CUT OF HAMILTONIANS

Here we define a family of operators $\hat{h}_{l,\omega}$ and show that $\hat{h}_{l,U,\{R\}}$ (the operator $h_{l,U,\{R\}}$ in momentum representation) can be obtained as a norm-resolvent limit of $\hat{h}_{l,\omega}$ as $\omega\to\infty$. Let

$$\hat{h}_{l,\omega} = p^2 + \sum_{j,j'=1}^{N} [\Lambda^\omega_l]_{jj'}\varphi^\omega_{l,j}(p)(\varphi^\omega_{l,j'},\cdot), \tag{3.1}$$

$$\varphi^\omega_{l,j}(p) = (R_j/2\pi)^{1/2}\chi_\omega(p)j_l(pR_j) = \chi_\omega(p)\varphi_{l,j}(p), \tag{3.2}$$

where $j_l(z)$ is the spherical Bessel function and χ_ω is defined by

$$\chi_\omega(p) = \begin{cases} 1, & \text{if } |p|\leqslant\omega, \\ 0, & \text{if } |p|>\omega. \end{cases} \tag{3.3}$$

The resolvent of $\hat{h}_{l,\omega}$ is given by

$$(\hat{h}_{l,\omega}-k^2)^{-1} = (p^2-k^2)^{-1} + (p^2-k^2)^{-1} \times \sum_{j,j'=1}^{N} [M^\omega_l(k)]_{jj'}|\varphi^\omega_{l,j}\rangle\langle(p^2-\bar{k}^2)^{-1}\varphi^\omega_{l,j'}|$$

$$= (p^2-k^2)^{-1} + \sum_{j,j'=1}^{N} [M^\omega_l(k)]_{jj'} \times |\chi_\omega F_{l,k,j}\rangle\langle\chi_\omega F_{l,-\bar{k},j}|, \tag{3.4}$$

where

$$F_{l,k,j}(p) = [(R_j/2\pi)^{1/2}j_l(pR_j)]/(p^2-k^2), \tag{3.5}$$

$$-[M^\omega_l(k)]^{-1}_{jj'} = [\Lambda^\omega_l]^{-1}_{jj'} + (g^\omega_l(k))_{jj'}, \tag{3.6}$$

$$(g^\omega_l(k))_{jj'} = (\varphi^\omega_{l,j},(p^2-k^2)\varphi^\omega_{l,j'}) = \frac{1}{2\pi}(R_jR_{j'})^{1/2}\int_{|p|<\omega}\frac{p^2\,dp\,j_l(pR_j)\,j_l(pR_{j'})}{p^2-k^2}, \quad \mathrm{Im}\,k>0. \tag{3.7}$$

Let $g^\omega_l(k)$ be the $N\times N$ matrix with elements $(g^\omega_l(k))_{jj'}$. Then as $\omega\to\infty$, $g^\omega_l(k)$ converges to the matrix $g_l(k)$ with elements

$$(g_l(k))_{jj'} = \frac{1}{2\pi}(R_jR_{j'})^{1/2}\int_0^\infty \frac{p^2\,dp\,j_l(pR_j)j_l(pR_{j'})}{p^2-k^2}. \tag{3.8}$$

The integral in (3.8) can be performed explicitly, and indeed one gets (see Ref. 10, p. 119)

$$\frac{1}{2\pi}(R_jR_{j'})^{1/2}\int_0^\infty \frac{p^2\,dp\,j_l(pR_j)j_l(pR_{j'})}{p^2-k^2} = g_{l,k}(R_j,R_{j'}).$$

We note that in momentum representation the resolvent of $h_{l,U,\{R\}}$ [cf. Eq. (2.15)] reads

$$(\hat{h}_{l,U,\{R\}} - k^2)^{-1}$$
$$= (p^2 - k^2)^{-1} + (p^2 - k^2)^{-1}$$
$$\times \sum_{j,j'=1}^{N} [M_l(k)]_{jj'} |\varphi_{l,j}\rangle\langle p^2 - \bar{k}^2)\varphi_{l,j'}|$$
$$= (p^2 - k^2)^{-1} + \sum_{j,j'=1}^{N} [M_l(k)]_{jj'} |F_{l,k,j}\rangle\langle F_{l,-\bar{k},j'}|. \tag{3.9}$$

A short computation shows that the rank-1 operator $|\chi_\omega F_{l,k,j}\rangle\langle\chi_\omega F_{l,-\bar{k},j'}|$ converges in Hilbert–Schmidt norm to the operator $|F_{l,k,j}\rangle\langle F_{l,-\bar{k},j'}|$ as $\omega\to\infty$ when $\operatorname{Im} k>0$, i.e.,

$$\lim_{\omega\to\infty} \| |\chi_\omega F_{l,k,j}\rangle\langle\chi_\omega F_{l,-\bar{k},j'}| - |F_{l,k,j}\rangle\langle F_{l,-\bar{k},j'}| \|_2 = 0, \quad \operatorname{Im} k>0. \tag{3.10}$$

Consequently $\hat{h}_{l,\omega}$ converges to $\hat{h}_{l,U,\{R\}}$ in the norm-resolvent sense iff

$$\lim_{\omega\to\infty} [M_l^\omega(k)]_{jj'} = [M_l(k)]_{jj'}. \tag{3.11}$$

Therefore we have to choose the matrices Λ_l^ω in such a way that (3.11) is fulfilled. Thus if we choose Λ_l to be independent of ω and equal to

$$-\Lambda_l^{-1} = -[\Lambda_l^\omega]^{-1}$$
$$= [g_l(\sqrt{i}) - g_l(\sqrt{-i})](U^T+1)^{-1}$$
$$+ g_l(\sqrt{-i}), \tag{3.12}$$

then (3.11) is trivially satisfied and we obtain the relationship between the Hermitian matrix Λ_l and the unitary matrix U. We note that there is an interesting characterization of elements of the domain $\mathscr{D}(h_{l,U,\{R\}})$ in terms of boundary conditions, namely, for every function $f_l\in\mathscr{D}(h_{l,U,\{R\}})$ the jump of the derivative of f_l at a fixed point $R_j\in[0,\infty)$ is given by

$$f_l'(R_{j+}) - f_l'(R_{j-}) = \sum_{j'=1}^{N} [\Lambda_l]_{jj'} f_l(R_{j'}), \tag{3.13}$$

where Λ_l represents coupling constants in front of δ potentials with nonseparated boundary conditions (called "nonlocal" δ potentials in Ref. 7). Locality of these potentials has been proved in Ref. 11.

IV. APPROXIMATION OF $h_{l,U,\{R\}}$ BY SCALED SEPARABLE HAMILTONIANS

In this section we show that $h_{l,U,\{R\}}$ can be obtained as a norm-resolvent limit of a family of scaled separable Hamiltonians. Let us define in $L^2((0,\infty))$

$$h_l^\varepsilon = h_{l,0} + \varepsilon^{-2} \sum_{j,j'=1}^{N} [C_l(\varepsilon)]_{jj'} |\psi_{l,j}^\varepsilon\rangle\langle\psi_{l,j'}^\varepsilon|, \quad \varepsilon>0, \tag{4.1}$$

where $C_l(\varepsilon)$ denotes a Hermitian matrix and $\psi_{l,j}^\varepsilon$ is centered around R_j, e.g.,

$$\psi_{l,j}^\varepsilon(r) = \varepsilon^{-1/2}\psi_{l,j}((r-R_j)/\varepsilon), \tag{4.2}$$

for some function $\psi_{l,j}\in L^1((0,\infty))$. The resolvent $R_{k^2}^\varepsilon$ of h_l^ε is given by

$$(h_l^\varepsilon - k^2)^{-1} = (h_{l,0} - k^2)^{-1} + \delta R_{k^2}^\varepsilon,$$
$$\delta R_{k^2}^\varepsilon = \sum_{j,j'=1}^{N} (h_{l,0} - k^2)^{-1} [D_l(\varepsilon)]_{jj'}$$
$$\times |\psi_{l,j}^\varepsilon\rangle\langle(h_{l,0} - k^2)^{-1}\psi_{l,j'}^\varepsilon|, \tag{4.3}$$

with

$$-[D_l(\varepsilon)]_{jj'}^{-1} = \varepsilon^2[C_l(\varepsilon)]_{jj'}^{-1}$$
$$+ \langle\psi_{l,j}^\varepsilon, (h_{l,0} - k^2)^{-1}\psi_{l,j'}^\varepsilon\rangle. \tag{4.4}$$

We observe that there is norm convergence:

$$\delta R_{k^2}^\varepsilon \xrightarrow[\varepsilon\to 0]{\|\cdot\|} \sum_{j,j'}^{N} [M_l(k)]_{jj'} |\phi_{l,j}(k)\rangle\langle\phi_{l,j'}(-\bar{k})|, \quad \operatorname{Im} k>0, \tag{4.5}$$

if $\varepsilon^{-1}[D_l(\varepsilon)]^{-1}$ converges to $[M_l(k)]^{-1}$ and $\int_0^\infty \psi_{l,j}(r)dr = 1\neq 0$. This allows us to adjust the ε dependence of $D_l(\varepsilon)$. We may take the following dependence of $[C_l(\varepsilon)]_{jj'}$:

$$[C_l(\varepsilon)]_{jj'}^{-1} = (\varepsilon^{-1} + O(1))(\tilde{C}_l)_{jj'}^{-1}. \tag{4.6}$$

Then the study of the limit

$$\lim_{\varepsilon\to 0_+} \varepsilon^{-1}[D_l(\varepsilon)]^{-1} = [M_l(k)]^{-1}$$

gives

$$-\tilde{C}_l^{-1} = [g_l(\sqrt{i}) - g_l(\sqrt{-i})](U^T+1)^{-1}$$
$$+ g_l(\sqrt{-i}).$$

Again $\tilde{C}_l$ corresponds to coupling constants of δ potentials with nonseparated boundary conditions.

ACKNOWLEDGMENTS

J. S. would like to thank Professor Abdus Salam, the International Atomic Energy Agency, and UNESCO for hospitality at the International Center for Theoretical Physics, Trieste, where this work was done.

[1] J.-P. Antoine, F. Gesztesy, and J. Shabani, J. Phys. A: Math. Gen. **20**, 3687 (1987).
[2] J. Shabani, J. Math. Phys. **29**, 660 (1988).
[3] J. Shabani, "Some properties of the Hamiltonian describing a finite number of δ'-interactions with support on concentric spheres," to be published in Nuovo Cimento.
[4] N. I. Akhiezer and I. M. Glazman, *Theory of Linear Operators in Hilbert Space*, Vol. 2 (Pitman, Boston, 1981).
[5] M. Reed and B. Simon, *Fourier Analysis, Self-adjointness*, Vol. 2 of *Methods of Modern Mathematical Physics* (Academic, New York, 1975).
[6] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1972).
[7] L. Dabrowski and H. Grosse, J. Math. Phys. **26**, 2777 (1985).
[8] S. Albeverio, F. Gestesy, R. Hoegh-Krohn, and H. Holden, *Solvable Models in Quantum Mechanics* (Springer, Berlin, 1988).
[9] R. A. Adams, *Sobolev Spaces* (Academic, New York, 1975).
[10] K. Gottfried, *Fundamentals*, Vol. 1 of *Quantum Mechanics* (Benjamin, Amsterdam, 1966).
[11] A. Grossmann, R. Hoegh-Krohn, and M. Mebkhout, J. Math. Phys. **21**, 2376 (1980).

The perturbed cubic Schrödinger equation: Selection mechanism, resonant limits, and spatial chaos

Paul K. Newton
Department of Mathematics, University of Illinois, 1409 West Green Street, Urbana, Illinois 61801

(Received 26 January 1988; accepted for publication 25 May 1988)

Separable solutions of the perturbed cubic Schrödinger equation that are sinusoidal in time with frequency Ω_ϵ are considered. Two processes associated with the integrable limit $\epsilon \to 0$ are demonstrated. The first will be called a selection mechanism and is demonstrated for Ginzburg–Landau-type perturbations. The second is a resonant limit process and will be demonstrated for perturbations containing spatial driving terms of wave number q. Spatially chaotic behavior is studied in the limits $\Omega_\epsilon \to 0$, $q \to 0$ and $\Omega_\epsilon \to \infty$, $q \to \infty$ such that $q/\Omega_\epsilon^\alpha = \text{const}$.

I. INTRODUCTION

We consider separable solutions of the perturbed cubic Schrödinger equation,

$$i\psi_t + \psi_{xx} + |\psi|^2\psi = \epsilon f(x,t;\psi,\psi_x,\psi_t,\ldots) \quad (\epsilon \ll 1), \tag{1.1}$$

which are of the form

$$\psi_\epsilon(x,t) = \exp(i\Omega_\epsilon t)P_\epsilon(x). \tag{1.2}$$

The periodic standing-wave solutions of the unperturbed problem, $\epsilon = 0$, are well known and have been studied previously. The perturbed nonlinear Schrödinger equation has been studied recently by a number of authors who, for the most part, focus on the fate of solitons under small perturbations in their evolution equations.[1–3] Generally speaking, these questions are important in discerning how sensitive certain solutions are to small changes in the model equations, a concept that is commonly termed structural stability. Our purpose is to demonstrate two processes that are associated with the integrable limit $\epsilon \to 0$. The first will be called a *selection mechanism* and is demonstrated for Ginzburg–Landau-type perturbations,

$$f(\psi,\psi_{xx}) = i[\psi + \psi_{xx} - |\psi|^2\psi]. \tag{1.3}$$

This form of perturbation has the effect of complexifying the coefficients of the nonlinear Schrödinger equation, which introduces important new phenomena.[4–12] The Ginzburg–Landau (GL) equation in this form arises in the study of hydrodynamic stability theory (see references listed in Ref. 5). The second process we call a *resonant limit process*, which will be demonstrated for simple linear functions,

$$f(\psi,\psi_x) = \gamma_1\psi\cos(qx) + \gamma_2\cos(qx)\exp(i\Omega_\epsilon t) - \gamma_3\psi_x \quad (\gamma_1,\gamma_2,\gamma_3 \in R \geqslant 0). \tag{1.4}$$

Section II contains a brief review of known facts about the standing-wave solutions of the unperturbed and perturbed problems. We start by studying the separable solutions (1.2) of the unperturbed problem. This leads to a second-order nonlinear ordinary differential equation (ODE) for the spatial function P_0 whose solutions are given by periodic Jacobian elliptic functions. After fixing the two constants of integration, the solutions form a continuous family which can be parametrized by a similarity parameter λ or, equivalently, by the temporal frequency Ω_0. Thus it is shown that an extra mechanism is needed in order to fix the parameter λ, thereby uniquely determining a solution of the unperturbed problem. We then study the separable solutions of the Ginzburg–Landau equation (1.1), (1.3). Results from Refs. 4–6 are mentioned in which the periodic spatial part of the solution is given both asymptotically and numerically. Contrary to the unperturbed case, once the constants of integration are chosen, there are no remaining free parameters. In Sec. III we describe a selection mechanism by which distinguished values for the similarity parameter λ and the frequency Ω_0 are picked out in the limit $\epsilon \to 0$. Explicit examples of how this selection mechanism works are given. In addition, it is mentioned that the selection process can also be viewed as a bifurcation problem in ϵ. In Sec. IV we study perturbations of the form (1.4). Using a version of the Melnikov technique[13,14] we discuss the existence of spatially chaotic behavior in special regions of the parameter space $(\gamma_1,\gamma_2,\gamma_3)$. The results are summarized in Theorem 1. The chaotic regions are then studied as functions of frequency Ω_ϵ and wave number q. In particular, a resonant limit process is introduced and summarized in Lemma 2, which describes the *size* of the chaotic regions as $\Omega_\epsilon \to 0$, $q \to 0$ and $\Omega_\epsilon \to \infty$, $q \to \infty$ such that $q/\Omega_\epsilon^\alpha = \text{const}$.

II. SPATIALLY PERIODIC STANDING WAVES

Consider first the unperturbed cubic Schrödinger equation,

$$i\psi_t + \psi_{xx} + |\psi|^2\psi = 0. \tag{2.1}$$

When (1.2) is substituted into (2.1), this yields an ordinary differential equation for the spatial function P_0:

$$\ddot{P}_0 = \Omega_0 P_0 - P_0^3. \tag{2.2}$$

It is well known that this equation can be put in the form of a Hamiltonian system with one degree of freedom.[15] For present purposes we will not pursue this approach, but instead integrate (2.2) directly. Multiplying (2.2) by P_0 and integrating once yields

$$(\dot{P}_0)^2 = E - V(P_0), \tag{2.3}$$

where E is an arbitrary constant of integration. The potential V is given by

$$V(P_0) = \tfrac{1}{2}P_0^4 - \Omega_0 P_0^2. \tag{2.4}$$

 0022-2488/88/102245-05$02.50

There are three distinct solutions to (2.3) that depend on the constant of integration (energy level) E.

(a) $E<0$:

$$P_0(x) = \pm\sqrt{2}\lambda\,\mathrm{dn}(\lambda(x-x_0),k)\,, \tag{2.5}$$

$$k^2 = (2\lambda^2-\Omega_0)/\lambda^2\,. \tag{2.6}$$

Here x_0 is the second constant of integration, and k is the modulus of the even Jacobian elliptic function dn. The similarity parameter λ is related to k and Ω_0 via (2.6). The energy level E is related to λ and Ω_0 by the formula

$$E = 2\lambda^2(\lambda^2-\Omega_0) < 0\,. \tag{2.7}$$

(b) $E>0$:

$$P_0(x) = \pm\sqrt{\Omega_0+\lambda^2}\,\mathrm{cn}(\lambda(x-x_0),k)\,, \tag{2.8}$$

$$k^2 = (\Omega_0+\lambda^2)/2\lambda^2\,. \tag{2.9}$$

Here k is the modulus of the Jacobian elliptic function cn. The energy level E is related to Ω_0 and λ via the formula

$$E = \tfrac{1}{2}(\lambda^4-\Omega_0^2)\,. \tag{2.10}$$

(c) $E=0$:

$$P_0(x) = \pm\sqrt{2}\lambda\,\mathrm{sech}(\lambda(x-x_0))\,. \tag{2.11}$$

This is the limiting case of (a), (b) in which the modulus $k\to 1$.

We now focus our attention on the even solutions (2.5). Once the two constants of integration E and x_0 are fixed, the solutions (2.5) form a continuous family parametrized by the similarity parameter λ or the frequency Ω_0. This can be seen by solving (2.7) for Ω_0:

$$\Omega_0(\lambda,E) = \lambda^2 - E/2\lambda^2\,. \tag{2.12}$$

Substituting (2.12) into (2.6) yields

$$k^2 = 1 + E/2\lambda^4\,. \tag{2.13}$$

Thus for a given value of E the parameter λ can be chosen freely, which then fixes the modulus (2.13) and the frequency (2.12). Without loss of generality we set $x_0=0$ in future considerations, and for definiteness we only consider the positive solutions.

Consider now standing-wave solutions (1.2) of the Ginzburg–Landau equation (1.1), (1.3) with $0<\epsilon\ll 1$. Upon substitution this yields a complex spatial Duffing-type equation, which we write

$$\frac{\ddot{P}_\epsilon}{P_\epsilon} = \frac{(\Omega_\epsilon+i\epsilon)}{(1-i\epsilon)} - \frac{(1+i\epsilon)}{(1-i\epsilon)}|P_\epsilon|^2\,. \tag{2.14}$$

Now introduce polar coordinates $P_\epsilon(x) = R(x)\exp(i\theta(x))$ and use the transformation

$$W \equiv U+iV = \dot{P}_\epsilon/P_\epsilon = \dot{R}/R + i\dot{\theta}\,. \tag{2.15}$$

Note that by differentiating W once we see that W satisfies a complex Riccati equation,

$$\dot{W} + W^2 = \ddot{P}_\epsilon/P_\epsilon\,. \tag{2.16}$$

Equating the real and imaginary parts of (2.14), (2.16) yields

$$\begin{aligned} \dot{U} &= \alpha_1 - U^2 + V^2 - \alpha_2 R^2\,,\\ \dot{V} &= \beta_1 - 2UV - \beta_2 R^2\,, \end{aligned} \tag{2.17}$$

where

$$\begin{aligned} \alpha_1 &= (\Omega_\epsilon-\epsilon^2)/(1+\epsilon^2)\,,\\ \alpha_2 &= (1-\epsilon^2)/(1+\epsilon^2)\,,\\ \beta_1 &= \epsilon(1+\Omega_\epsilon)/(1+\epsilon^2)\,,\\ \beta_2 &= 2\epsilon/(1+\epsilon^2)\,. \end{aligned} \tag{2.18}$$

Now equating the real and imaginary parts of (2.15) gives

$$\dot{R} = UR\,, \tag{2.19}$$

$$\dot{\theta} = V\,. \tag{2.20}$$

The system of equations (2.17), (2.19) forms a closed system of three equations in the three unknowns (U,V,R) that determines the spatial structure of standing-wave solutions (1.2) to the Ginzburg–Landau equation (1.1), (1.3). It has been shown in Refs. 5 and 6 that this system supports a rich variety of solutions (see also Refs. 7–9). For our purposes it is sufficient to note that for $\epsilon>0$, there is a range of frequencies Ω_ϵ for which periodic solutions to (2.17), (2.19) of wave number q exist and are stable. In addition, a dispersion relation has been computed relating the frequency Ω_ϵ to the spatial wave number q. We refer the reader to Figs. 1 and 2 of Ref. 5 for plots of the periodic solutions and the nonlinear dispersion relation. Contrary to the unperturbed $\epsilon=0$ case, there is no extra parameter in the problem; thus, once Ω_ϵ and ϵ are fixed, a unique periodic solution is determined.

III. SELECTION MECHANISM

Our purpose here is to show how the limit $\epsilon\to 0$ in (1.1), (1.3) selects a particular value for the similarity parameter λ in the solution (2.5) of the unperturbed problem (1.1). We will focus only on the spatially even solutions (2.5). For this, an extra relationship is needed for the $\epsilon\neq 0$ problem relating the frequency Ω_ϵ to the spatial function P_ϵ. Thus we prove the following lemma.

Lemma 1: For spatially periodic solutions of the GL equation (1.1), (1.3) of the form (1.2) (P_ϵ complex) we have

$$(1+\Omega_\epsilon) = 2\|P_\epsilon\|_4^4/\|P_\epsilon\|_2^2\,, \tag{3.1}$$

$$\|P_{\epsilon_x}\|^2 = \frac{(1-\epsilon^2)}{(1+\epsilon^2)}\|P_\epsilon\|_4^4 - \frac{(\Omega_\epsilon-\epsilon^2)}{(1+\epsilon^2)}\|P_\epsilon\|_2^2\,, \tag{3.2}$$

where

$$\|P_\epsilon\|_n^n = \int_0^L |P_\epsilon|^n\,dx, \quad L = \frac{2\pi}{q} = \text{spatial period}\,. \tag{3.3}$$

Proof: Substitute (1.2) into (1.1), (1.3) to get an ODE for the spatial function

$$\ddot{P}_\epsilon = \frac{(\Omega_\epsilon+i\epsilon)}{(1-i\epsilon)}P_\epsilon - \frac{(1+i\epsilon)}{(1-i\epsilon)}|P_\epsilon|^2 P_\epsilon\,. \tag{3.4}$$

Multiply (3.4) by the complex conjugate of P_ϵ and integrate over a period. Separating real and imaginary parts gives the result.

We now show how the condition (3.1) is used in the limit $\epsilon\to 0$ to provide an extra constraint on the relationship (2.12), (2.13), thus picking out a limiting frequency Ω_0 and amplitude λ for the solution (2.5). The limit $\epsilon\to 0$ requires us to impose two compatibility conditions relating the separable solutions of the unperturbed problem to those of the perturbed problem. These two conditions are as follows.

(1) *Periodicity condition:*

$$2K/\lambda = 2\pi/q\,, \tag{3.5}$$

where K is the complete elliptic integral. This condition guarantees that the period of the Jacobian elliptic function solution (2.5) matches the period of the $\epsilon \neq 0$ problem in the limit $\epsilon \to 0$.

(2) *Selection mechanism:*

$$(1+\Omega_\epsilon) = 2\|P_\epsilon\|_4^4/\|P_\epsilon\|_2^2\,. \tag{3.6}$$

This condition holds for all separable solutions (1.2) of the GL equation for $\epsilon \neq 0$; hence we apply the condition in the limit $\epsilon \to 0$.

We now carry out this selection process for two specific cases.

Example 1: Fix the wave number q of the $\epsilon \neq 0$ problem (1.1), (1.3) such that $q^2 = 2$. Then the limiting solution of the unperturbed cubic Schrödinger equation (1.1), for $\epsilon = 0$, is the spatially independent solution

$$\psi_0 = e^{it}\,. \tag{3.7}$$

To see this we must show that the limiting frequency $\Omega_0 = 1$, and that the spatial part of the solution (2.5) reduces to unity. This is most easily seen by fixing the modulus of (2.5):

$$k^2 = 0\,. \tag{3.8}$$

From a standard table of elliptic integrals,[16] (3.8) $\Rightarrow$

$$K = \pi/2\,. \tag{3.9}$$

Using the periodicity condition (3.5) and (3.9) we get

$$2\lambda = q\,. \tag{3.10}$$

Now using (3.8) and (2.6) gives

$$2\lambda^2 = \Omega_0\,. \tag{3.11}$$

Using the selection criterion (3.6) and the fact that $\mathrm{dn}(x,0) = 1$ gives

$$\tfrac{1}{2}(1+\Omega_0) = 2\lambda^2\,. \tag{3.12}$$

Thus (3.12) and (3.11) $\Rightarrow$

$$\Omega_0 = 1\,. \tag{3.13}$$

Furthermore, (3.13) and (3.11) $\Rightarrow$

$$\lambda^2 = \tfrac{1}{2}\,. \tag{3.14}$$

Equation (3.14) combined with (3.10) $\Rightarrow$

$$q^2 = 2\,. \tag{3.15}$$

Finally, using (3.13)–(3.15) with (3.8) and (1.2) $\Rightarrow$ (3.7).

Example 2: Fix the wave number q of the $\epsilon \neq 0$ problem (1.1), (1.3) such that $q = 0$. Then the limiting solution of the unperturbed cubic Schrödinger equation (1.1) for $\epsilon = 0$ is

$$\psi_0 = \sqrt{\tfrac{6}{5}}\exp(\tfrac{3}{5}it)\,\mathrm{sech}(\sqrt{\tfrac{3}{5}}x)\,. \tag{3.16}$$

To see this we must show that the limiting frequency $\Omega_0 = \frac{3}{5}$, and that the spatial part of the solution reduces to $\sqrt{\frac{6}{5}}\,\mathrm{sech}(\sqrt{\frac{3}{5}}\,x)$. This is most easily seen by fixing the modulus of (2.5):

$$k^2 = 1\,. \tag{3.17}$$

From a standard table of elliptic integrals,[16] (3.17) $\Rightarrow$

$$K = \infty\,. \tag{3.18}$$

Using the periodicity condition (3.5) and (3.18) $\Rightarrow$

$$q = 0\,. \tag{3.19}$$

Now using (3.17) and (2.6) $\Rightarrow$

$$\Omega_0 = \lambda^2\,. \tag{3.20}$$

Using the selection criterion (3.6) and the fact that $\mathrm{dn}(x,1) = \mathrm{sech}(x)$ gives

$$\tfrac{1}{2}(1+\Omega_0) = 4\lambda^2/3\,. \tag{3.21}$$

Here we have used the identities

$$\int_0^\infty \mathrm{sech}^2(x)\,dx = 1\,, \tag{3.22}$$

$$\int_0^\infty \mathrm{sech}^4(x)\,dx = \frac{2}{3}\,. \tag{3.23}$$

Thus (3.21) and (3.20) $\Rightarrow$

$$\lambda^2 = \Omega_0 = \tfrac{3}{5}\,. \tag{3.24}$$

Finally, using (3.17), (3.19), and (3.24) in (1.2) and (2.5) $\Rightarrow$ (3.16).

We remark that, equivalently, one can view this selection mechanism as a bifurcation problem in the parameter ϵ. The limiting frequencies Ω_0 of the unperturbed problem that are "selected" in the limit $\epsilon \to 0$ are the bifurcation points along the Ω_ϵ axis. The periodic standing-wave solutions to the system (2.17), (2.19) are the bifurcated Jacobian elliptic functions of the unperturbed problem.

IV. SPATIALLY CHAOTIC REGIONS

We now turn to perturbations of the form (1.4) and discuss the existence of spatially chaotic behavior in certain parameter regions. When the separable solution (1.2) is substituted into (1.1), (1.4) one gets

$$\ddot{P}_\epsilon = \Omega_\epsilon P_\epsilon - P_\epsilon^3 + \epsilon(\gamma_1 P_\epsilon\cos(qx) + \gamma_2\cos(qx) - \gamma_3\dot{P}_\epsilon). \tag{4.1}$$

An equation similar to this but with different perturbation terms is studied in Ref. 17, where it is shown that chaotic (temporal) behavior occurs for certain parameter regions. We first follow a similar analysis and view (4.1) as a perturbation from a Hamiltonian system for $\epsilon \ll 1$. This allows us to apply methods due to Melnikov[13,14] to show under what parameter restrictions $(\gamma_1,\gamma_2,\gamma_3)$ chaotic behavior occurs.

After rescaling (4.1) such that

$$P_\epsilon = \sqrt{\Omega_\epsilon}\,F, \quad x = \sqrt{1/\Omega_\epsilon}\,X\,, \tag{4.2}$$

one arrives at the first-order perturbed Hamiltonian system

$$\begin{aligned}\dot{F} &= G,\\ \dot{G} &= F - F^3 + \epsilon\left(\frac{\gamma_1}{\Omega_\epsilon}F\cos\left(\frac{qX}{\sqrt{\Omega_\epsilon}}\right)\right.\\ &\quad\left. + \frac{\gamma_2}{\Omega_\epsilon^{3/2}}\cos\left(\frac{qX}{\sqrt{\Omega_\epsilon}}\right) - \frac{\gamma_3}{\sqrt{\Omega_\epsilon}}G\right).\end{aligned} \tag{4.3}$$

We study the phase space of the perturbed system (4.3), $\epsilon \ll 1$, near the unperturbed separatrix solution:

$$\begin{aligned}F_0(X) &= \sqrt{2}\,\mathrm{sech}(X)\,,\\ G_0(X) &= -\sqrt{2}\,\mathrm{sech}(X)\tanh(X)\,.\end{aligned} \tag{4.4}$$

The phase space of the unperturbed problem, $\epsilon = 0$, possesses three fixed points: $(F,G) = (0,0),(\pm 1,0)$, where the origin is a saddle with a smooth separatrix orbit (4. 4) describing the stable and unstable manifolds. We briefly review the ideas described in Refs. 13, 14, and 17–19. To study the perturbed problem, the phase space can be extended to three dimensions (F,G,X) and the motion viewed in the Poincaré section $X = \text{const} \pmod{2\pi\sqrt{\Omega_\epsilon}/q}$. The perturbed stable and unstable manifolds can still be identified in the surface of section but are made up of a sequence of distinct points. Generally speaking, they no longer join smoothly. If γ_3 is large enough, they will not intersect at all. However, if the ratio of the parameters γ_3/γ_2 ($\gamma_1 = 0$) or γ_3/γ_1 ($\gamma_2 = 0$) is sufficiently small, the stable and unstable manifolds will intersect transversally. This process is called a homoclinic bifurcation and signals the onset of chaotic behavior. To check when a transverse crossing occurs, Melnikov introduced a function (now known as the Melnikov function) that measures the distance between the perturbed stable and unstable manifolds in the Poincaré section. If this distance function has a simple zero, the manifolds intersect transversally and chaotic behavior results. This method and its application to equations of the form (4.1) is by now relatively standard, and we refer the reader to Refs. 14 and 19 for a more thorough introduction to these ideas.

We first show that for sufficiently small ϵ, the condition for a homoclinic bifurcation to occur in the perturbed system (4.3) is

$$\tfrac{4}{3}\gamma_3 < H(q,\Omega_\epsilon)\,, \tag{4.5}$$

where

$$H(q,\Omega_\epsilon) = \frac{\gamma_1 \pi q^2}{\Omega_\epsilon^{3/2}} \operatorname{csch}\left(\frac{\pi q}{2\sqrt{\Omega_\epsilon}}\right) + \frac{\sqrt{2}\gamma_2 \pi q}{\Omega_\epsilon^{3/2}} \operatorname{sech}\left(\frac{\pi q}{2\sqrt{\Omega_\epsilon}}\right). \tag{4.6}$$

Since the analysis is standard, we only show part of the details. The Melnikov function measuring the distance between the perturbed stable and unstable manifolds in the Poincaré section at the point t_0 is computed from

$$\begin{aligned} D(t_0) = -\int_{-\infty}^{\infty} dt \Bigg(& \frac{\gamma_1}{\Omega_\epsilon} F_0(t-t_0) G_0(t-t_0) \cos\left(\frac{qt}{\sqrt{\Omega_\epsilon}}\right) \\ & + \frac{\gamma_2}{\Omega_\epsilon^{3/2}} G_0(t-t_0) \cos\left(\frac{qt}{\sqrt{\Omega_\epsilon}}\right) \\ & - \frac{\gamma_3}{\sqrt{\Omega_\epsilon}} G_0^2(t-t_0) \Bigg). \end{aligned} \tag{4.7}$$

After substituting (4.4) into (4.7) and simplifying, we obtain

$$\begin{aligned} D(t_0) = & -2\frac{\gamma_1}{\Omega_\epsilon} \sin\left(\frac{q}{\sqrt{\Omega_\epsilon}} t_0\right) \int_{-\infty}^{\infty} d\tau \operatorname{sech}^2(\tau) \\ & \times \tanh(\tau) \sin\left(\frac{q}{\sqrt{\Omega_\epsilon}} \tau\right) \\ & - \sqrt{2}\frac{\gamma_2}{\Omega_\epsilon^{3/2}} \sin\left(\frac{q}{\sqrt{\Omega_\epsilon}} t_0\right) \int_{-\infty}^{\infty} d\tau \operatorname{sech}(\tau) \\ & \times \tanh(\tau) \sin\left(\frac{q}{\sqrt{\Omega_\epsilon}} \tau\right) \\ & + 2\frac{\gamma_3}{\sqrt{\Omega_\epsilon}} \int_{-\infty}^{\infty} d\tau \operatorname{sech}^2(\tau) \tanh^2(\tau) \\ & (\tau = t - t_0)\,. \end{aligned} \tag{4.8}$$

Performing the remaining integrations yields the condition (4.5) for $D(t_0)$ to have a simple zero. The results of this section are summarized in the following theorem.

Theorem 1: The perturbed cubic Schrödinger equation (1.1), (1.4) supports a family of temporally sinusoidal, spatially chaotic solutions of the form (1.2) for ϵ sufficiently small, provided (4.5) is satisfied. The chaotic behavior is caused by the presence of transverse *spatial* homoclinic orbits in the associated Poincaré map, which implies that the map contains Smale horseshoes.[20]

We call the function H in (4.6) the threshold function, and study its behavior in the following section.

V. RESONANT LIMITS

First consider the case in which $\gamma_1 = 0$; thus the threshold condition (4.5) becomes

$$\frac{\gamma_3}{\gamma_2} < \frac{3\sqrt{2}}{4} \frac{\pi q}{\Omega_\epsilon^{3/2}} \operatorname{sech}\left(\frac{\pi}{2} \frac{q}{\Omega_\epsilon^{1/2}}\right) \equiv H_2(q,\Omega_\epsilon)\,. \tag{5.1}$$

The threshold function H_2 is shown in Fig. 1, where we plot H_2 as a function of Ω_ϵ for fixed q. It can be seen from this diagram, and is straightforward to prove, that $\lim_{\Omega \to 0,\infty} H_2 \to 0$. Thus these limits reduce the size of the chaotic region in the parameter space. Furthermore, there exists a value $\Omega_{\max}$ that we call the *most chaotic* frequency at which H_2 achieves a maximum. This value can be computed by solving

$$\frac{\partial H_2}{\partial \Omega_\epsilon} = 0 \tag{5.2}$$

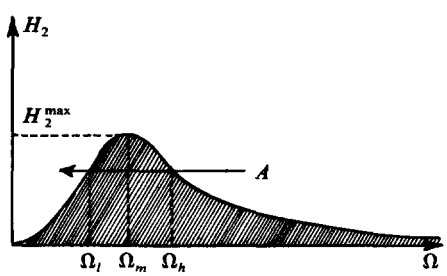


FIG. 1. Threshold function $H_2(\Omega)$ for fixed q, with chaotic region shaded and Ω_m the most chaotic frequency. The Ω_l and Ω_h are the lowest and highest frequencies giving rise to chaotic behavior for $H_2 = A$.

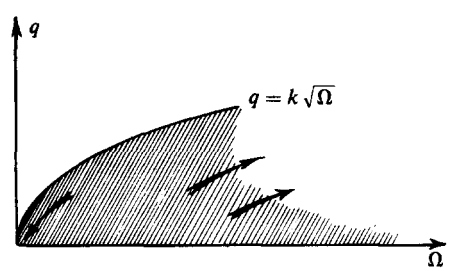


FIG. 2. The wave number q versus Ω plane showing resonant limit sector for H_1. For $q \to 0$, $\Omega \to 0$ or $q \to \infty$, $\Omega \to \infty$ in the shaded region, H_1 blows up.

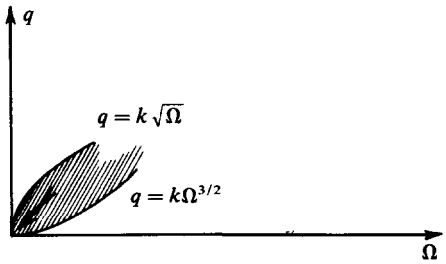


FIG. 3. The wave number q versus Ω plane showing resonant limit sector for H_2. For $q \to 0$, $\Omega \to 0$ in the shaded region, H_2 blows up.

for Ω_ϵ, holding q fixed. It is worth noting that for a fixed value of the ratio γ_3/γ_2 lying in the unshaded region marked "A" in Fig. 1, one can decrease the frequency Ω_ϵ and observe a *window* of chaotic behavior for $\Omega_\epsilon \in [\Omega_{\text{low}}, \Omega_{\text{high}}]$. The values Ω_{low} and Ω_{high} can be explicitly computed given the other parameters in the problem. Similarly, holding Ω_ϵ fixed one can study H_2 as a function of q. It is easily seen that the behavior is similar to that shown in Fig. 1. Note that there exists a value $q_{\max}$ that we call the *most chaotic* wave number, which can be computed by solving

$$\frac{\partial H_2}{\partial q} = 0 \tag{5.3}$$

for q, holding Ω_ϵ fixed.

For the case $\gamma_2 = 0$, the threshold condition (4.5) becomes

$$\frac{\gamma_3}{\gamma_1} < \frac{3\pi}{4}\frac{q^2}{\Omega_\epsilon^{3/2}}\operatorname{csch}\left(\frac{\pi}{2}\frac{q}{\Omega_\epsilon^{1/2}}\right) \equiv H_1(q, \Omega_\epsilon)\,. \tag{5.4}$$

The behavior of H_1 as a function of each variable is similar to that of H_2; hence we do not show a separate diagram. We remark only that $\Omega_{\max}$ and $q_{\max}$ can be computed as in (5.2), (5.3) with the obvious change of H_2 to H_1.

We examine now the limits of H_j ($j = 1,2$) as a function of both the variables q, Ω. This behavior is summarized in the following lemma.

Lemma 2: The following limits are taken holding $q/\Omega_\epsilon^\alpha = \text{const}$.

(a) For $q \to 0$, $\Omega_\epsilon \to 0$,

(i) $\lim H_1 = \begin{cases} \infty, & \alpha \geqslant \frac{1}{2}, \\ 0, & \alpha < \frac{1}{2}, \end{cases}$

(ii) $\lim H_2 = \begin{cases} 0, & \alpha < \frac{1}{2},\ \alpha > \frac{3}{2}, \\ \infty, & \frac{1}{2} \leqslant \alpha < \frac{3}{2}, \\ \text{const}, & \alpha = \frac{3}{2}\,. \end{cases}$

(b) For $q \to \infty$, $\Omega_\epsilon \to \infty$,

(i) $\lim H_1 = \begin{cases} \infty, & \alpha < \frac{1}{2}, \\ 0, & \alpha \geqslant \frac{1}{2}, \end{cases}$

(ii) $\lim H_2 = 0$, for all α.

The shaded regions in Figs. 2 and 3 show the sectors in the (q,Ω_ϵ) plane in which $H_j \to \infty$ as $q \to 0$, $\Omega_\epsilon \to 0$ and $q \to \infty$, $\Omega_\epsilon \to \infty$ for $j = 1,2$, respectively. Because the threshold function blows up in these limits and thus provides no upper bounds on the ratios in (5.1), (5.4), we call these limits *resonant* limits and the corresponding exponent α the *resonant* rate. The resonant limits demonstrate a cooperation between the frequency Ω_ϵ and the wave number q, increasing the size of the chaotic region in the given parameter space $(\gamma_1,\gamma_2,\gamma_3)$, and should be contrasted with the individual limits mentioned earlier and shown in Fig. 1. The proof is a relatively straightforward analysis of the functions H_j, and involves competing effects of exponential versus algebraic behavior; we therefore do not include it.

ACKNOWLEDGMENTS

I would like to thank B. Birnir and L. Sirovich for useful conversations on various aspects of this work. I also thank J. B. Keller for helpful comments on an earlier version of this manuscript and G. Francis for his help on the graphics. The present work was completed while the author was in the applied mathematics group at the Department of Mathematics, Stanford University. Their support is greatly appreciated.

This research was supported by the Office of Naval Research, the Air Force Office of Scientific Research, the Army Research Office, and the National Science Foundation.

[1] D. J. Kaup and A. C. Newell, Proc. R. Soc. London Ser. A **361**, 413 (1978).
[2] D. J. Kaup and P. J. Hansen, "The forced nonlinear Schrödinger equation," Physica D **18**, 77 (1986).
[3] R. Grimshaw, "Slowly varying solitary waves II. Nonlinear Schrödinger equation," Proc. R. Soc. London Ser. A **368**, 377 (1979).
[4] P. K. Newton and L. Sirovich, "Instabilities of the Ginzburg–Landau equation: Periodic solutions," Q. Appl. Math. **44**, 49 (1986); "Secondary bifurcation," *ibid.* **44**, 367 (1986).
[5] L. Sirovich and P. K. Newton, "Periodic solutions of the Ginzburg–Landau equation," Physica D **21**, 115 (1986).
[6] P. K. Newton, Ph.D. thesis, Brown University, 1985.
[7] P. J. Holmes, "Spatial structure of time-periodic solutions of the Ginzburg–Landau equation," Physica D **23**, 84 (1986).
[8] C. A. Holmes and D. Wood, "Studies of a complex Duffing equation in nonlinear waves on plane Poiseuille flow," in *Nonlinear Phenomena and Chaos*, edited by S. Sarkar, Malvern Physics Series (Adam Hilger, Bristol, 1986).
[9] M. J. Landman, "Solutions of the Ginzburg–Landau equation of interest in shear flow transition," Stud. Appl. Math. **76**, 187 (1987).
[10] L. Keefe, "Dynamics of perturbed wavetrain solutions to the Ginzburg–Landau equation," Stud. Appl. Math. **73**, 91 (1985).
[11] H. Moon, P. Huerre, and L. G. Redekopp, "Three-frequency motion and chaos in the Ginzburg–Landau equation," Phys. Rev. Lett. **49**, 458 (1982).
[12] L. Sirovich and J. D. Rodriguez, "Coherent structures and chaos: A model problem," Phys. Lett. A **120**, 211 (1987).
[13] V. K. Melnikov, "On the stability of the center for time periodic perturbations," Trans. Moscow Mat. Soc. **12**, 1 (1963).
[14] J. Guckenheimer and P. Holmes, *Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields* (Springer, New York, 1983).
[15] H. Goldstein, *Classical Mechanics* (Addison–Wesley, Reading, MA, 1980), 2nd ed.
[16] *Handbook of Mathematical Functions*, edited by M. Abramowitz and I. A. Stegun (Dover, New York, 1965).
[17] P. J. Holmes, "A nonlinear oscillator with a strange attractor," Philos. Trans. R. Soc. London Ser. A **292**, 419 (1979).
[18] P. J. Holmes, "Averaging and chaotic motions in forced oscillations," SIAM J. Appl. Math. **38**, 65 (1980); Errata and addenda, *ibid.* **40**, 167 (1981).
[19] A. J. Lichtenberg and M. A. Lieberman, *Regular and Stochastic Motion* (Springer, New York, 1982).
[20] S. Smale, "Differentiable dynamical systems," Bull. Am. Math. Soc. **73**, 747 (1967).

Degenerate Siegert states

David L. Huestis
Chemical Physics Laboratory, SRI International, Menlo Park, California 94025

(Received 29 September 1987; accepted for publication 1 June 1988)

In a previous paper [D. L. Huestis, J. Math. Phys. **16**, 2148 (1975)] a superposition principle was developed that allows the representation of an arbitrary wave function in an explicit uniformly convergent expansion over the discrete Siergert states for finite-range potentials. Possible difficulties were identified that could arise for special values of the potential strength due to degeneracy of the complex Siegert eigenvalues or vanishing of the norm of the Siegert state. In this paper these difficulties are addressed. By generalizing the Siegert eigenvalue problem, distinct orthogonal eigenfunctions with nonvanishing norm are obtained, recompleting the Siegert basis.

I. INTRODUCTION

For potentials of finite range [$V(r)=0$ for $r>r_0$], Siegert[1] completed the identification of resonance wave functions with the poles of the S matrix through his eigenvalue problem,

$$-\psi''+U(r)\psi=k^2\psi,\quad \psi(0)=0,\quad \psi'(r_0)=ik\psi(r_0),$$

where

$$U(r)=(2m/\hbar^2)V(r)\quad\text{and}\quad k^2=(2m/\hbar^2)E.$$

This is simply Schrödinger's s-wave eigenvalue problem with an outgoing right-hand boundary condition. If we consider the usual scattering boundary condition,

$$\psi(r)\to A\exp(ikr)+B\exp(-ikr),\quad \text{as } r\to\infty,$$

we see that the requirement that $\psi'(r)=ik\psi(r)$ for $r>r_0$ implies that B must vanish and that the S matrix defined by

$$S=-A/B$$

must have a pole at the resonance energy.

Humblet[2,3] divided the Siegert poles into three groups, called a, b, and c (see also Refs. 4–6). The b poles are the bound states, with k_b along the positive imaginary axis, $k_b=i|k_b|$. The a poles, sometimes called the antibound or virtual states, lie along the negative imaginary axis, $k_a=-i|k_a|$. The c poles are the ordinary complex resonances, sometimes called radioactive or decaying states. They are distributed in the lower half of the complex k plane, symmetrically about the imaginary axis. Humblet showed that, for potentials of finite strength and finite range, only a finite number of a and b poles exist, whereas the c poles constitute a denumerably infinite set. As the potential is made more attractive, the c poles approach from both sides of the negative imaginary k axis, eventually merging to form two a poles. As the potential is further strengthened, one of this pair moves up the imaginary k axis, eventually becoming a bound state, while the other moves down toward more negative imaginary k values, remaining an antibound state. This motion is illustrated in Fig. 1.

In our previous study,[7] we recast the Siegert eigenvalue problem in regular and dual two-component bases,

$$\hat{\Phi}_n=\begin{vmatrix}\phi_n^{(1)}(r)\\ \phi_n^{(2)}(r)\end{vmatrix},\quad \hat{\Phi}_n^\dagger=\begin{vmatrix}\phi_n^{\dagger(1)}(r)\\ \phi_n^{\dagger(2)}(r)\end{vmatrix},$$

with boundary conditions

$$\phi_n^{(1)}(0)=0,\quad \phi_n^{(1)}(r_0)=\phi_n^{(2)}(r_0),$$
$$\phi_n^{\dagger(1)}(0)=0,\quad \phi_n^{\dagger(1)}(r_0)=\phi_n^{\dagger(2)}(r_0),$$

and functional

$$\langle\hat{\Phi}_n^\dagger,\hat{\Phi}_m\rangle=\frac{1}{2}\int_0^{r_0}(\phi_n^{\dagger(1)}\phi_m^{(1)}-\phi_n^{\dagger(2)}\phi_m^{(2)}),$$

where we have added a factor of $\frac{1}{2}$ to match the usual Schrödinger normalization for bound states (see also Ref. 8). The regular and dual eigenvalue problems are then written as

$$\hat{L}\hat{\Phi}_n=\begin{vmatrix}\phi_n^{(2)\prime}(r)\\ \phi_n^{(1)\prime}(r)+\int_r^{r_0}U(x)\phi_n^{(1)}(x)dx\end{vmatrix}=ik_n\hat{\Phi}_n$$

and

$$\hat{L}^\dagger\hat{\Phi}_n^\dagger=\begin{vmatrix}\phi_n^{\dagger(2)\prime}(r)-U(r)\int_0^r\phi_n^{\dagger(2)}(x)dx\\ \phi_n^{\dagger(1)\prime}(r)\end{vmatrix}=ik_n\hat{\Phi}_n^\dagger.$$

For eigenvectors defined as above, we can easily show that

$$i(k_n-k_m)\langle\hat{\Phi}_n^\dagger,\hat{\Phi}_m\rangle=0,$$

and thus as long as $k_n\neq k_m$, we can define normalization constants such that

$$\langle\hat{\Phi}_n^\dagger,\hat{\Phi}_m\rangle=N_m\delta_{nm}.$$

This orthogonality allows us to expand arbitrary functions in the eigenbasis,

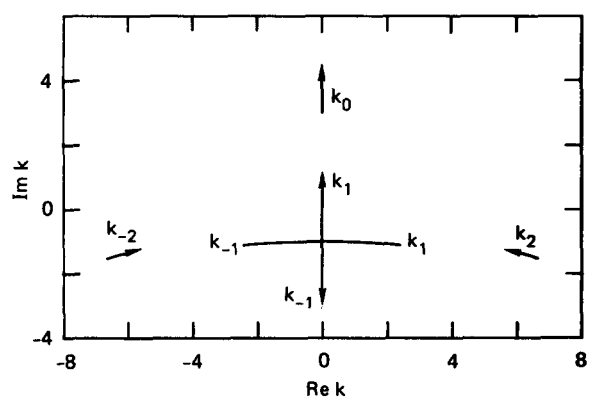


FIG. 1. Trajectories of Siegert eigenvalues for the s-wave finite square well. Arrows indicate the direction of motion from $U=-15$ to $U=-25$; k_1 and k_{-1} collide at $-i$ for $U\simeq-21.1907$.

 0022-2488/88/102250-04$02.50

$$\hat{F}=\sum_n \alpha_n \hat{\Phi}_n,$$

where

$$\alpha_n=(1/N_n)\langle\hat{\Phi}_n^{\dagger},\hat{F}\rangle.$$

In our previous investigation, we explored the restrictions on the form of the function $\hat{F}$ necessary to obtain a convergent expansion, which imposed relations between the upper and lower components of $\hat{F}$. Further difficulties arise from either degenerate eigenvalues, $k_n=k_m$ for $n\neq m$, or vanishing normalization, $N_m=0$ for $\hat{\Phi}_m\neq 0$. As we noted, Dobson[9] suggested that these two difficulties are expected to occur simultaneously. In our brief description above of the trajectories of the Siegert eigenvalues as the potential strength is increased, we indicated that the collision of pairs of c poles on the negative imaginary k axis is a normal result that is necessary for the creation of bound states. Thus we must modify the two-component formalism to account for these degeneracies.

In addition to difficulties in orthogonality and normalization, the confluence of eigenvalues leads to only a single eigenvector, while for nearby values of the potential strength we have two. As suggested by Friedman[10] and elaborated by Dobson,[9] we might attempt to recomplete the vector space by generalizing the eigenvalue problem. In the present case, this generalization corresponds to considering solutions of $(\hat{L}-ik)^2\hat{\Phi}=0$ and $(\hat{L}^{\dagger}-ik)^2\hat{\Phi}^{\dagger}=0$. For nondegenerate eigenvalues, the "squaring" of the eigenvalue problem leads only to the eigensolution previously found [if we require that $(\hat{L}-ik)\hat{\Phi}$ and $(\hat{L}^{\dagger}-ik)\hat{\Phi}^{\dagger}$ satisfy the boundary conditions]. In the degenerate case, we get two distinct eigenvectors, of which orthogonal linear combinations can be chosen having nonvanishing normalization.

II. DEGENERATE SIEGERT STATES FOR THE SQUARE-WELL POTENTIAL

Although degenerate antibound states are expected for all types of attractive finite-range potentials, we will develop our approach based on the specific example of the finite square well,

$$U(r)=-U_0,\quad 0<r<r_0=1,$$
$$U(r)=0,\quad r>r_0.$$

The regular and dual Siegert wave vectors are (see Ref. 7)

$$\hat{\Phi}_n=\begin{vmatrix}\sin(k_n'r)\\ -i\dfrac{k_n}{k_n'}\cos(k_n'r)+i\left(\dfrac{k_n}{k_n'}-\dfrac{k_n'}{k_n}\right)\cos(k_n')\end{vmatrix}$$

and

$$\hat{\Phi}_n^{\dagger}=\begin{vmatrix}\sin(k_n'r)\\ -i(k_n'/k_n)\cos(k_n'r)\end{vmatrix},$$

where k_n' is the wave number within the well, $k_n'^2=k_n^2+U_0$, and $k_n'\cos(k_n')=ik_n\sin(k_n')$ is the eigenvalue condition. The normalization constants are $N_n=(1+i/k_n)/2$, which vanish for $k_n=-i$ and diverge for $k_n=0$.

TABLE I. Some Siegert s-wave eigenvalues for the finite square well, depth $U=-21.1907$, radius $r_0=1$.

n[a]	Re k_n	Im k_n	Re k_n'	Im k_n'
0	0.0000	3.8301	2.5536	0.0000
1	0.0000	−1.0000	4.4934	0.0000
2	6.1959	−1.3863	7.6753	−1.1191
3	9.8422	−1.6683	10.8429	−1.5143
4	13.2390	−1.8901	14.0030	−1.7870
5	16.5387	−2.0728	17.1585	−1.9979
6	19.7889	−2.2278	20.3111	−2.1706
7	23.0102	−2.3626	23.4616	−2.3171

[a] The corresponding resonances with $n<0$ satisfy $k_n=-k_{-n}^{*}$.

We will first deal with the case of a bound state at zero energy, $k_n=0$, which does not result from collision of eigenvalues, but which prevents the choosing of $\phi_n^{\dagger(1)}(r)=\phi_n^{(1)}(r)$ as we previously supposed.[7] In this case, which occurs for $U_0=k_n'^2$ with $k_n'=(n+\frac{1}{2})\pi$, the Siegert eigenvectors are (with $N_n=-\frac{1}{2}$)

$$\hat{\Phi}_n=\begin{vmatrix}\sin(k_n'r)\\ 1\end{vmatrix}\quad\text{and}\quad\hat{\Phi}_n^{\dagger}=\begin{vmatrix}0\\ k_n'\cos(k_n'r)\end{vmatrix}.$$

The second special case is the first occurrence of $k_n=-i$, which happens for $U_0=1$ and which also does not correspond to a degeneracy but merely the progression of the first antibound state on its way toward becoming a bound state. In this case we divide both the regular and dual eigenvectors by k_0' and take the limit as $k_0'\to 0$, obtaining (with $N_0=-\frac{1}{6}$)

$$\hat{\Phi}_0=\begin{vmatrix}r\\ \frac{1}{2}(1+r^2)\end{vmatrix}\quad\text{and}\quad\hat{\Phi}_0^{\dagger}=\begin{vmatrix}r\\ 1\end{vmatrix}.$$

True degeneracies result for potential strengths $U_0=1+k_n'^2$, satisfying $k_n'\cot(k_n')=ik_n=1$ for $k_n'>0$. The first occurrence of this condition is for $U_0\simeq 21.1907$, resulting in the Siegert eigenvalues listed in Table I. In such a case we already have one eigensolution with a vanishing norm,

$$\hat{\Upsilon}_n=\begin{vmatrix}\sin(k_n'r)\\ (-1/k_n')\cos(k_n'r)+((1/k_n')+k_n')\cos(k_n')\end{vmatrix},$$

and

$$\hat{\Upsilon}_n^{\dagger}=\begin{vmatrix}\sin(k_n'r)\\ k_n'\cos(k_n'r)\end{vmatrix}.$$

To find additional solutions of $(\hat{L}-ik)^2\hat{\Phi}=0$ and $(\hat{L}^{\dagger}-ik)^2\hat{\Phi}^{\dagger}=0$ we consider

$$\hat{\Theta}_n=\frac{\partial}{\partial k_n'}\hat{\Phi}_n\quad\text{and}\quad\hat{\Theta}_n^{\dagger}=\frac{\partial}{\partial k_n'}\hat{\Phi}_n^{\dagger}.$$

These derivatives give

$$\hat{\Theta}_n=\begin{vmatrix}r\cos(k_n'r)\\ (1+1/k_n'^2)(\cos(k_n'r)-\cos(k_n'))+(r/k_n')\sin(k_n'r)\end{vmatrix},$$

and

$$\hat{\Theta}_n^\dagger = \begin{vmatrix} r\cos(k_n' r) \\ (1+k_n'^2)\cos(k_n' r) - k_n' r \sin(k_n' r) \end{vmatrix}.$$

With some algebra we can confirm that

$$(\hat{L} - ik_n)\hat{\Theta}_n = -k_n'\hat{\Upsilon}_n$$

and

$$(\hat{L}^\dagger - ik_n)\hat{\Theta}_n^\dagger = -k_n'\hat{\Upsilon}_n^\dagger,$$

and thus

$$(\hat{L} - ik_n)^2\hat{\Theta}_n = 0 \quad \text{and} \quad (\hat{L}^\dagger - ik_n)^2\hat{\Theta}_n^\dagger = 0.$$

We can also evaluate the functionals

$$\langle \hat{\Upsilon}_n^\dagger, \Upsilon_n \rangle = 0, \quad \langle \hat{\Theta}_n^\dagger, \Theta_n \rangle = -(1+3k_n'^2)/12,$$

and

$$\langle \hat{\Upsilon}_n^\dagger, \Theta_n \rangle = \langle \hat{\Theta}_n^\dagger, \Upsilon_n \rangle = -k_n'/4.$$

At this point can we consider choosing linear combinations of the pairs of regular and dual vectors to obtain the desired orthogonality and nonvanishing norm. One such choice is

$$\hat{\Omega}_n = (1+3k_n'^2)/(3k_n')\hat{\Upsilon}_n - \hat{\Theta}_n$$

and

$$\hat{\Omega}_n^\dagger = (1+3k_n'^2)/(3k_n')\hat{\Upsilon}_n^\dagger - \hat{\Theta}_n^\dagger,$$

which gives

$$\langle \hat{\Omega}_n^\dagger, \hat{\Theta}_n \rangle = \langle \hat{\Theta}_n^\dagger, \hat{\Omega}_n \rangle = 0$$

and

$$\langle \hat{\Omega}_n^\dagger, \hat{\Omega}_n \rangle = (1+3k_n'^2)/12 = -\langle \hat{\Theta}_n^\dagger, \hat{\Theta}_n \rangle.$$

Below we will use $\hat{\Phi}_{+n}^{(\dagger)} \equiv \hat{\Theta}_n^{(\dagger)}$ and $\hat{\Phi}_{-n}^{(\dagger)} \equiv \hat{\Omega}_n^{(\dagger)}$.

Another reasonable choice of linear combinations is

$$\hat{\Phi}_{+n} = \hat{\Upsilon}_n, \quad \hat{\Phi}_{+n}^\dagger = (1+3k_n'^2)/(6k_n')\hat{\Upsilon}_n^\dagger - \hat{\Theta}_n^\dagger,$$

and

$$\hat{\Phi}_{-n} = (1+3k_n'^2)/(6k_n')\hat{\Upsilon}_n - \hat{\Theta}_n, \quad \hat{\Phi}_{-n}^\dagger = \hat{\Upsilon}_n^\dagger,$$

with

$$\langle \hat{\Phi}_{+n}^\dagger, \hat{\Phi}_{-n} \rangle = \langle \hat{\Phi}_{-n}^\dagger, \hat{\Phi}_{+n} \rangle = 0,$$

and

$$\langle \hat{\Phi}_{+n}^\dagger, \hat{\Phi}_{+n} \rangle = \langle \hat{\Phi}_{-n}^\dagger, \hat{\Phi}_{-n} \rangle = k_n'/4.$$

This choice has the advantage that $\hat{\Phi}_{+n}$ and $\hat{\Phi}_{-n}^\dagger$ are the continuous limits of $\hat{\Phi}_n$ and $\hat{\Phi}_n^\dagger$ from nearby values of the potential strength. However, it removes the presumably natural symmetry between the upper, or physical, components of the regular and dual Siegert eigenvectors, $\phi_n^{(1)}(r) = \phi_n^{\dagger(1)}(r)$.

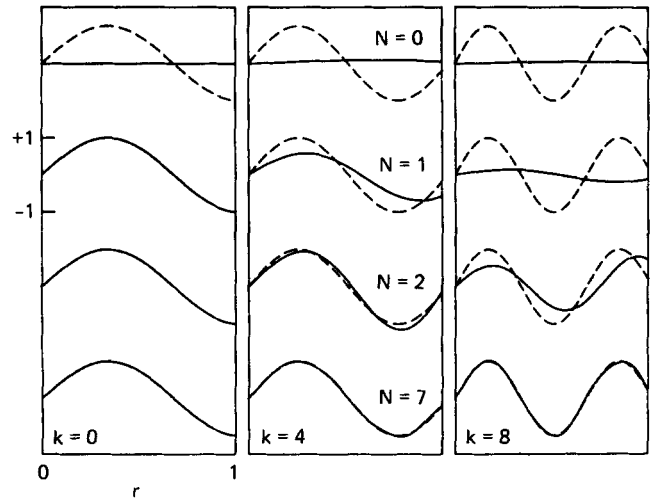


FIG. 2. Expansion $\psi_k(r) = \Sigma_{-N}^{N} \alpha_n(k)\phi_n^{(1)}(r)$ over square-well Siegert states for the first degenerate case, $U \simeq -21.1907$, for $k = 0,4,8$. Dashed lines indicate the exact solutions; solid lines represent the partial expansions to $\pm N$. The imaginary parts of the partial expansions cancel when the terms $\alpha_n\phi_n$ and $\alpha_{-n}\phi_{-n}$ are included together.

III. GENERALIZATION OF THE SUPERPOSITION PRINCIPLE

As in our previous investigation,[7] we wish to expand in the Siegert basis the vector corresponding to the exact scattering wave function:

$$\hat{\Psi}_k = \sum_n \alpha_n(k)\hat{\Phi}_n.$$

With the same assumptions as previously,

$$\hat{L}^2\hat{\Psi}_k = -k^2\hat{\Psi}_k, \quad \psi_k^{(1)}(0) = 0,$$

$$\psi_k^{(1)}(r_0) = \psi_k^{(2)}(r_0), \quad \psi_k^{(1)\prime}(r_0) = \psi_k^{(2)\prime}(r_0),$$

we can evaluate $\alpha_n(k) = \langle \hat{\Phi}_n^\dagger, \hat{\Psi}_k \rangle / N_n$ through the equality

$$(k^2 - k_n^2)^2\langle \hat{\Phi}_n^\dagger, \hat{\Psi}_k \rangle = \langle \hat{\Phi}_n^\dagger, (-\hat{L}^2 - k_n^2)^2\hat{\Psi}_k \rangle.$$

Using similar manipulations as previously,[7] we obtain

$$\alpha_n(k) = \left[-\frac{\phi_n^{\dagger(2)}(0)}{(k^2 - k_n^2)} + \frac{\phi_n^{\dagger(2)\prime\prime}(0) - (U(0) - k_n^2)\phi_n^{\dagger(2)}(0)}{(k^2 - k_n^2)^2} \right] \frac{\psi_k^{(2)\prime}(0)}{2N_n}.$$

For nondegenerate eigenvalues, we have $\phi_n^{\dagger(2)\prime\prime}(0) = (U(0) - k_n^2)\phi_n^{\dagger(2)}(0)$ [see Ref. 7, Eq. (17)] and obtain the same expansion coefficients as previously (with an additional factor of $\frac{1}{2}$ due to the present change in the definition of the "inner product" functional). In the degenerate case the expansion coefficients explicitly display the repeated nature of the eigenvalue. As we discussed above, the degenerate eigenvectors can be constructed in various ways. All consistent choices of linear combinations will lead to the same expansions, but with different expansion coefficients.

Figure 2 compares partial expansions with the exact wave functions for the first degenerate case of the finite square well, as described in Table I. The rate of convergence is much more rapid than for the much weaker potential studied previously ($U_0 = -0.5$),[7] largely because of the smaller imaginary parts of the Siegert eigenvalues. The curves la-

beled $N = 1$ include the small contribution due to $\phi_0^{(1)}(r)$ and the sum of the contributions of the degenerate $\phi_{+1}^{(1)}(r)$ and $\phi_{-1}^{(1)}(r)$. We can see that not only is the present expansion formula correct, but also that failure to generalize the Siegert eigenvalue problem would lead to significant errors.

[1]A. J. F. Siegert, Phys. Rev. **56**, 750 (1939).
[2]J. Humblet, Mém. Soc. Roy. Sci. Liège **12**, Part 4 (1952).
[3]J. Humblet and L. Rosenfeld, Nucl. Phys. **26**, 529 (1961).
[4]T. Regge, Nuovo Cimento **8**, 671 (1958).
[5]H. M. Nussenzvieg, Nucl. Phys. **11**, 499 (1959).
[6]R. G. Newton, *Scattering Theory of Waves and Particles* (McGraw-Hill, New York, 1966).
[7]D. L. Huestis, J. Math. Phys. **16**, 2148 (1975).
[8]N. Hokkyo, Prog. Theor. Phys. **33**, 1116 (1965).
[9]P. N. Dobson, Jr., J. Math. Phys. **12**, 1207 (1971).
[10]B. Friedman, *Principles and Techniques of Applied Mathematics* (Wiley, New York, 1956).

Inverse scattering for the mixed spectrum of δ potentials

Lin Jian-cheng[a)]
International Center for Theoretical Physics, 34100 Trieste, Italy

(Received 6 October 1987; accepted for publication 10 February 1988)

The inverse problem is studied in a system of mixed spectrum of which the continuous part coincides with that of a repulsive δ potential and the discrete part coincides with that of an attractive δ potential.

From the standard analysis of the one-dimensional Schrödinger equation we know that in a repulsive δ potential case, a reflection coefficient arises without any bound state, while in an attractive δ-potential case, a reflection coefficient arises with a bound state. The inverse problem for the spectrum containing only a continuous part that is the same as that of the attractive δ potential has been studied in a recent paper.[1] In that paper they find that the corresponding potential still has the same δ distribution, but besides there is one more repulsive piece to the right of the δ distribution that is just enough to prevent the formation of a bound state. In this paper, we extend the inverse problem to a mixed spectrum case in which the reflection coefficient comes from that of the repulsive δ potential and the discrete spectrum comes from that of the attractive one. We find that the factorization method can still be extended to solve this inverse problem. For the corresponding potential we find three terms: the first is the same repulsive δ distribution; the second is an attractive piece to the right of the δ distribution that is the same as that in Ref. 1 except for a negative sign (because of this, we think that this piece is responsible for creating such a bound state); the third, as shown in Fig. 1, is to the left of the δ distribution. This term is to ensure that the potential also has the same reflection coefficient in the presence of the bound state. We are now going to give the details.

For one-dimensional Schrödinger equation

$$\left[-\frac{d^2}{dx^2}+V(x)\right]\phi(x)=k^2\phi(x) \tag{1}$$

with potential

$$V_-(x)=-2\alpha\delta(x),\quad \alpha>0, \tag{2}$$

it is easy to obtain a reflection coefficient

$$R_-(k)=i\alpha/(k-i\alpha) \tag{3}$$

and a single bound state with a discrete level $p=\alpha$ and the wave function

$$\phi_-(x)=e^{-\alpha|x|} \tag{4}$$

from which one can deduce the normalization coefficient

$$\rho_-=\left[\int_{-\infty}^{\infty}|\phi(x)|^2\,dx\right]^{-1}=\alpha. \tag{5}$$

The spectral transform, thus, is

$$S_-[V_-(x)]=\{i\alpha/(k-i\alpha),\ -\infty<k<\infty,\ \alpha,\alpha\}. \tag{6}$$

For the potential

$$V_+(x)=2\alpha\delta(x),\quad \alpha>0 \tag{7}$$

it yields the following reflection coefficient:

$$R_+(k)=-i\alpha/(k+i\alpha) \tag{8}$$

and there is no longer any bound state. The corresponding spectral transform is

$$S_+[V_+(x)]=\{-i\alpha/(k+i\alpha),\ -\infty<k<\infty\}. \tag{9}$$

Now we select a mixed spectrum

$$S[V(x)]=\{-i\alpha/(k+i\alpha),\ -\infty<k<\infty;\ \alpha,\alpha\} \tag{10}$$

and ask what the form of the potential $V(x)$ is.

To solve this inverse problem we begin from the Gel'fand–Levitan–Marchenko equation[2]

$$K(x,y)+M(x,y)+\int_x^{\infty}K(x,z)M(y,z)dz=0,\quad y>x, \tag{11}$$

where the kernel $M(x)$ is defined by

$$M(x)=\frac{1}{2\pi}\int_{-\infty}^{\infty}\exp(ikx)R(k)dk+\rho\exp(-px)$$
$$=-\alpha\theta(-x)\exp(\alpha x)+\alpha\exp(-\alpha x). \tag{12}$$

Because of the existence of the $\theta(x)$ function, one has to separate the function $K(x,y)$ into three branches $P(x,y)$, $Q(x,y)$, and $T(x,y)$ (see Refs. 1 and 3) as shown in Fig. 2,

$$K(x,y)=\begin{cases}P(x,y), & \text{if } x+y>0, \quad x>0;\\ Q(x,y), & \text{if } x+y>0, \quad x<0;\\ T(x,y), & \text{if } x+y<0, \quad x<0.\end{cases} \tag{13}$$

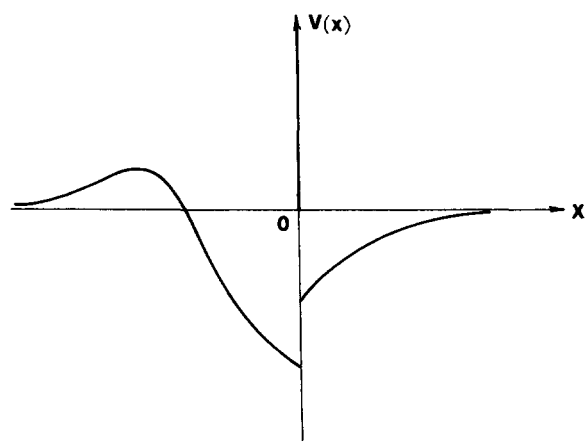


FIG. 1. The inverse potential for the spectrum given by Eq. (7).

[a)] On leave of absence from the Institute of Theoretical Physics, Academia, Sinica, Peking, China.

 0022-2488/88/102254-02$02.50

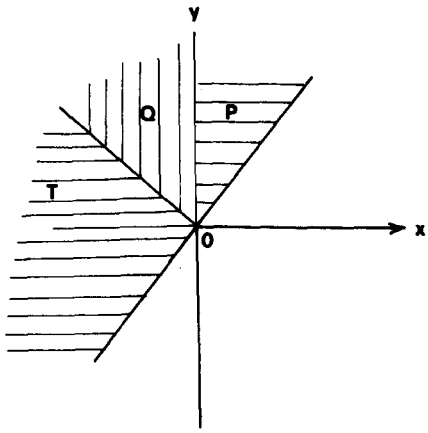


FIG. 2. The branch regions for the function $K(x,y)$.

The equations for these three functions are

$$P(x,y) + \alpha e^{-\alpha(x+y)} + \alpha \int_x^{\infty} P(x,z) e^{-\alpha(y+z)}\, dz = 0,$$

$$Q(x,y) + \alpha e^{\alpha(x+y)} + \alpha \int_x^{-x} T(x,z) e^{-\alpha(y+z)}\, dz + \alpha \int_{-x}^{\infty} Q(x,z) e^{-\alpha(y+z)} dz = 0,$$

$$T(x,y) - \alpha e^{\alpha(x+y)} + \alpha e^{-\alpha(x+y)} - \alpha \int_x^{-y} T(y,z) e^{\alpha(y+z)}\, dz + \alpha \int_x^{-x} T(x,z) e^{-\alpha(y+z)}\, dz + \alpha \int_{-x}^{\infty} Q(x,z) e^{-\alpha(y+z)}\, dz = 0. \tag{14}$$

The positive exponent term appearing in $M(x)$ makes the problem a bit more complex. However, it is still solvable and the solutions are

$$P(x,y) = \frac{-\alpha e^{-\alpha y}}{\sqrt{2}\cosh(\alpha x + \phi)}, \quad \tanh(\phi) = \frac{1}{3},$$

$$Q(x,y) = \alpha e^{-\alpha y} \frac{e^{\alpha x}[1 - 2e^{-2\alpha x}]}{4\alpha x + 2e^{-2\alpha x} - \frac{1}{2}e^{2\alpha x}}, \tag{15}$$

$$T(x,y) = -2\alpha e^{\alpha y} \frac{e^{\alpha x}[1 - 2e^{-2\alpha x}]}{4\alpha x + 2e^{-2\alpha x} - \frac{1}{2}e^{2\alpha x}} + \alpha \frac{4\alpha x + \frac{1}{2}e^{2\alpha x} - 2e^{-2\alpha x}}{4\alpha x + 2e^{-2\alpha x} - \frac{1}{2}e^{2\alpha x}}.$$

Hence

$$K(x,x+0^+) = -2\alpha[e^{-2\alpha x}/(2 + e^{-2\alpha x})]\theta(x) + \alpha \frac{4\alpha x - \frac{3}{2}e^{2\alpha x} - 2e^{-2\alpha x} + 4}{4\alpha x + 2e^{-2\alpha x} - \frac{1}{2}e^{2\alpha x}} \theta(-x). \tag{16}$$

The above equation yields the following potential:

$$V(x) = -2K_x(x,x+0^+) = 2\alpha\delta(x) - 2\alpha^2\theta(x)\operatorname{csch}^2(\alpha x + \phi) - \theta(-x) \frac{8\alpha^2 \sinh(\phi - \alpha x)[\alpha x \cosh(\phi - \alpha x) + \sinh(\phi - \alpha x)]}{[\alpha x + \sinh(\phi - \alpha x)\cosh(\phi - \alpha x)]^2}. \tag{17}$$

As seen from Eq. (17), besides the original δ distribution, there are two more pieces added to the potential, with an attractive one to the right of the δ distribution and the other one to the left of the δ distribution, which is attractive in the region near the origin and repulsive in the rest region.

In conclusion, we have extended the inverse problem for the spectrum coming from a single δ-distribution potential to the case of a mixed spectrum. The factorization method can still be generalized to solve the problem. By comparing the result with that in Ref. 1, we see that the potential form $\pm 2\alpha^2\theta(x)\operatorname{csch}^2(\alpha x + \phi)$ is responsible for creating (with negative sign) or annihilating (with a positive sign) a bound state with $p = \alpha$ and the corresponding normalization factor $\rho = \alpha$ of the wave function.

ACKNOWLEDGMENTS

The author would like to thank Professor Yu Lu and Professor R. Mohan for helpful discussions and to Professor A. Salam for his hospitality at the International Center for Theoretical Physics at Trieste, in Italy.

[1] D. Atkinson and I. C. van den Born, "Inverse scattering: Attractive δ-distribution" a recent preprint from Institute for Theoretical Physics of Groningen of Netherlands, January, 1987.

[2] F. Calogero and A. Degasperis, *Spectral Transform and Solitons* (North-Holland, Amsterdam, 1982).

[3] I. Kay and H. E. Moses, Nuovo Cimento **5**, Suppl. 1, 230 (1957).

A quantum WKB approximation without classical trajectories

F. H. Molzahn[a)]
Department of Physics, University of California, Berkeley, California 94720

(Received 22 March 1988; accepted for publication 15 June 1988)

The time evolution of a coherent state is studied in the limit $\hbar \to 0$, for a spinless nonrelativistic quantum system in Euclidean space with smooth bounded time-dependent scalar interaction $v(x,t)$. The lowest-order WKB approximation is derived rigorously, complete with an explicit bound on its L^2-norm error over a specified finite time interval. The method is based on a constructive series representation of the classical action, thereby avoiding the need to analyze the behavior of classical trajectories. The amplitude of the initial states considered is neither assumed to be of rapid decrease, nor to depend on $\hbar$.

I. INTRODUCTION

In this paper, the Wentzel–Kramers–Brillouin (WKB) approximation of the solution to the time-dependent Schrödinger equation will be rigorously derived, without the use of classical trajectories. For initial Cauchy data which belongs to a large class of generalized coherent states, the leading-order asymptotics as $\hbar \to 0$ will be obtained by employing a constructive series representation of the relevant classical action. This rather direct approach to the problem allows one to obtain a simple bound on the error in the WKB approximation (in the sense of the Hilbert space norm), complete with explicit formulas for the constant coefficient appearing in the error estimate, as well as for the size of the time interval on which this approximation is valid. This interval is short enough to prevent the occurrence of caustics.

Consider a nonrelativistic quantum system of spinless particles having mass m, moving in Euclidean space, and interacting via smooth bounded time-dependent scalar potentials. The state space of such a system is the Hilbert space $\mathscr{H} \equiv L^2(\mathbb{R}^d;\mathbb{C})$, and the Hamiltonian is (for each time $t \in \mathbb{R}$) a self-adjoint operator $H(t): \mathfrak{D}H(t) \subset \mathscr{H} \to \mathscr{H}$, induced by the differential expression

$$H\left(x, \frac{\hbar}{i}\nabla_x, t\right) = \frac{1}{2m}\left(\frac{\hbar}{i}\nabla_x\right)^2 + v(x,t)\,. \tag{1.1}$$

Here $x \in \mathbb{R}^d$ is the variable of the system configuration space; if there are N particles moving in three dimensions, $d = 3N$. The function v: $\mathbb{R}^d \times \mathbb{R} \to \mathbb{R}$ describes the total potential energy of configuration x at time t, and $h = 2\pi\hbar$ is Planck's constant. The assumption of a common mass m for all particles is not restrictive because a change of particle coordinates can always cast a Hamiltonian with distinct particle masses into the form (1.1).

Time evolution of the quantum system is described by Schrödinger's equation

$$i\hbar \frac{d}{dt}\Psi_t = H(t)\Psi_t \quad (t \in \mathbb{R})\,, \tag{1.2}$$

where d/dt denotes the strong derivative in $\mathscr{H}$, and $\Psi_t \in \mathscr{H}$ is the state of the system at time t. In this work, I will consider the Cauchy problem for (1.2) obtained by specifying, at some fixed initial time $s \in \mathbb{R}$, initial data of the form

$$\Psi_s(x) = \varphi(x)\exp[(i/\hbar)k \cdot (x-y)]\,. \tag{1.3}$$

Here $k, y \in \mathbb{R}^d$ are parameters of the initial data, and φ is an $\hbar$-independent function chosen from a suitable dense linear subset of $\mathscr{H}$. It will not be assumed that φ has compact support.

The WKB method has as its goal the determination of an approximate wave function, $\Upsilon(x,t)$, which is asymptotic to $\Psi_t(x)$ in the limit $\hbar \to 0^+$. To say the least, this method has been widely studied.[1–16] Let us briefly recall the standard WKB approach to the above Cauchy problem, in order to facilitate a comparison with the present approach. To avoid unnecessary technicalities, assume temporarily that the potential $v(x)$ is static, and choose $s = 0$.

It is shown in Ref. 1 that during a sufficiently small time interval, say $|t| < T$, the lowest-order WKB approximation has the form

$$\Psi_t(x) = \Upsilon(x,t) + R_1(t,x)\,, \tag{1.4}$$

where

$$\Upsilon(x,t) = J(t, \mathrm{y}(t,x))^{-1/2}\varphi(y(t,x))\exp[(i/\hbar)S(t,x)]\,, \tag{1.5}$$

and the error term R_1 has the L^2-norm bound

$$\sup_{|t|<T} \|R_1(t,\cdot)\| < C\hbar\,. \tag{1.6}$$

The functions J, y, and S appearing in (1.5) are determined by solving a two-time boundary value problem for the trajectories of the classical-mechanical system associated with (1.1).

Let $q \equiv q(\,\cdot\,;q_0,k)$: $(-T,T) \to \mathbb{R}^d$ be the unique classical path satisfying Newton's equation

$$\frac{d^2}{d\tau^2}q(\tau) = -m^{-1}\nabla v(q(\tau))\,, \tag{1.7a}$$

subject to the initial conditions

$$q(0) = q_0\,, \quad \frac{\partial q}{\partial \tau}(0) = \frac{k}{m}\,. \tag{1.7b}$$

Then for $|\tau| < T$, one must show that the map $q(\tau;\,\cdot\,,k)$ is a diffeomorphism of an open set containing supp φ, onto another open set in $\mathbb{R}^d_x$. Next, J is defined as the Jacobian determinant of this map

$$J(t,q_0) \equiv \det \frac{\partial q}{\partial q_0}(t;q_0,k)\,, \tag{1.8}$$

while y is the inverse map

$$\mathrm{y}(t,x) \equiv q(t;\,\cdot\,,k)^{-1}(x)\,. \tag{1.9}$$

[a)] Present address: Department of Applied Mathematics, University of Manitoba, Winnipeg, Manitoba R3T 2N2, Canada.

 0022-2488/88/102256-12$02.50

From these quantities one may obtain a classical path

$$\tau \mapsto q^*(\tau) \equiv q(\tau; y(t,x), k) \tag{1.10}$$

which satisfies the boundary conditions

$$q^*(t) = x\,, \quad \frac{\partial q^*}{\partial \tau}(0) = \frac{k}{m}\,, \tag{1.11}$$

imposed at the times 0 and $t \in (-T, T)$. Finally, the so-called action S is defined for this problem by

$$S(t,x) \equiv k \cdot (y(t,x) - y) + \int_0^t d\tau \left[\frac{m}{2} \left(\frac{\partial q^*}{\partial \tau}(\tau) \right)^2 - v(q^*(\tau)) \right]. \tag{1.12}$$

Clearly, the process of making the standard constructions just outlined mathematically rigorous involves an elaborate analysis of the difficult two-time boundary value problem for the classical trajectories. This problem is further complicated when it is not assumed (unlike Ref. 1) that supp φ be compact. Essentially this is because the constructions leading to formula (1.12) for S arise from using the method of characteristics—generally a local method—to solve the classical time-dependent Hamilton–Jacobi equation

$$\frac{\partial S}{\partial t} + \frac{1}{2m} (\nabla_x S)^2 + v(x) = 0\,, \tag{1.13}$$

subject to the initial Cauchy data $S(0,x) = k \cdot (x - y)$.

This brings us to the central idea of this paper: it is possible to entirely circumvent the introduction and analysis of classical trajectories in deriving the $\hbar \to 0$ WKB asymptotics. This is based on the observation that one really only requires a complete integral S of the Hamilton–Jacobi equation, having the correct initial condition. The functions $y(t,x)$ and $1/J(t, y(t,x))$ may then be constructed from S by straightforward differentiation.

Of course, the Hamilton–Jacobi equation (1.13) is a nonlinear first-order partial differential equation in $d+1$ variables, and so finding explicit solutions is a nontrivial task. Nevertheless, it will be shown that the complete integral $S(x,t;k,s)$, appropriate for defining $\Upsilon(x,t)$, may be obtained constructively as a convergent infinite series. The terms in this series involve parametric integrals of the derivatives of potential v over linear paths, the derivative structures being determined by tree graphs.

One advantage of using the constructive series solution S is that it eliminates the need to use a local method to solve (1.13); specifically, it avoids the implicit function theorem used to obtain the map y in (1.9). Moreover, all the properties and estimates of S and its derivatives required to prove the desired $\hbar \to 0$ asymptotics of Ψ_t may be obtained using elementary analytical methods. For example, an (x,k)-uniform time interval on which $\Upsilon(x,t)$ exists will be found, and an explicit formula for the coefficient C appearing in the error bound (1.6) will be computed. With C known, the statement "$\hbar$ is small" may be interpreted more precisely.

The plan of the paper is as follows. In Sec. II the action $S(x,t;k,s)$ will be constructed and shown to satisfy the Hamilton–Jacobi equation. The properties of its derivatives required in the WKB analysis will be derived. Section III gives a brief summary of those aspects of the exact quantum-mechanical system that are of interest. In Sec. IV the WKB wave function Υ is defined and studied, and its asymptoticity to the exact evolution is proved in Theorem 2 (the main result of the paper). The proof employs the elegant method of Maslov and Fedoriuk[1] for bounding the total error. In Sec. V this result is compared with related ones in the literature, and some concluding remarks are made. The Appendix contains a proof of the form taken by the terms in the series defining S.

It is worth remarking on the motivation for the form of this series, which is closely related to the representation of Hamilton's principal function obtained in Ref. 17. Its formal origin derives from comparing the WKB approximation and the connected graph representation[18] of the mixed coordinate-momentum representation propagator $\langle x|U(t,s)|k\rangle$, where $U(t,s)$ is the quantum time evolution operator. The details of such a comparison are given in Sec. 4.1 of Ref. 18. In terms of the function S^* discussed there, one has $S = S^* - k \cdot y$.

II. CONSTRUCTION AND ANALYSIS OF CLASSICAL ACTION

The action $S(x,t;k,s)$, which is the essential ingredient of the WKB approximation for (1.1)–(1.3), is studied in this section. Here S is shown to be a solution of the classical Hamilton–Jacobi equation, and its associated Jacobian determinant D is shown to be positive on a sufficiently small time interval. The estimates of derivatives of S and D required for the WKB error estimate are also made.

I begin by defining the class of potentials v that will be employed throughout the remainder of the paper. Here $\mathbb{N}$ denotes the positive and $\mathbb{W}$ the non-negative integers. The gradient of a function with respect to a vector argument is denoted by ∇, while ∂ denotes the derivative with respect to a scalar argument. Thus if $\alpha \in \mathbb{W}^d$ is a multi-index, and $l \in \mathbb{W}$, then

$$(\nabla^\alpha \partial^l v)(x,t) = \nabla_x^\alpha \left(\frac{\partial}{\partial t} \right)^l v(x,t)\,.$$

Definition 1: A potential energy function v is said to be in the class $\mathscr{V}$ if

(i) $v \in C^\infty(\mathbb{R}^d \times \mathbb{R}; \mathbb{R})$;

(ii) v is bounded,

$$\|v\|_\infty = \sup\{|v(x,t)| \mid (x,t) \in \mathbb{R}^d \times \mathbb{R}\} < \infty\,;$$

and (iii) there exist $U,B,K \in (0,\infty)$ such that if $\alpha \in \mathbb{W}^d$, $l \in \mathbb{W}$, and $|\alpha| + l > 0$, then

$$\|\nabla^\alpha \partial^l v\|_\infty \leqslant U B^l (K/\sqrt{d})^{|\alpha|}\,.$$

It is simple to show that $\mathscr{V}$ is a real vector space, and that if $v \in \mathscr{V}$, then $v(\,\cdot\,, t)$ is a real-analytic function for each fixed t.

As was mentioned in the Introduction, the terms of the series used to define S will be specified with the aid of certain gradient structures associated with tree graphs. It is useful to review the notation used to describe these objects.

A (labeled) tree graph[19] T on $n \in \mathbb{N}$ vertices is an ordered pair $T = (VT, ET)$. The vertex set VT of T consists of n natural numbers—the "vertex labels." The edge set ET of T consists of $n-1$ unordered pairs of distinct elements of VT; thus if $\beta \in ET$ is an "edge," we may write $\beta = \{i_\beta, j_\beta\} \subset VT$,

$i_\beta < j_\beta$. Of course if there is only one vertex, ET is empty. The tree T is a connected graph: T cannot be expressed as the union of two graphs on disjoint vertex sets. Hence each vertex label $i \in VT$ must appear in at least one edge $\beta \in ET$. Vertex sets consisting of the first n natural numbers are usually encountered, so the following abbreviation[20] will be used:

$$\bar{n} = \{1,2,...,n\} \quad (n \in \mathbb{N}) .$$

The symbol $\mathscr{T}V$ denotes the set of all trees having vertex set V. According to Cayley's theorem,[19] if the cardinality $|V| = n$, then $|\mathscr{T}V| = n^{n-2}$.

In order to explain the gradient structures to be associated with a given tree, some preliminary definitions are needed. Denote the unit interval by $I = [0,1]$, and let $g: I \times I \to \mathbb{R}$ be the unit interval Green's function

$$g(\xi,\xi') \equiv \max\{\xi,\xi'\} - 1 , \tag{2.1}$$

which is clearly a continuous symmetric function of its two arguments.

Next, if $x,k \in \mathbb{R}^d$, $t,s \in \mathbb{R}$, and $m > 0$, define a space-time linear path $w: I \to \mathbb{R}^d \times \mathbb{R}$,

$$w(\xi) \equiv (x + (\xi - 1)(t-s)k/m, s + \xi(t-s)) . \tag{2.2}$$

Now we introduce a product of n potentials $v \in \mathscr{V}$ evaluated at different arguments. Specifically, V_n: $(\mathbb{R}^d \times \mathbb{R})^n \to \mathbb{R}$,

$$V_n((x^1,\tau^1),...,(x^n,\tau^n)) \equiv V_n\left(\bigtimes_{p=1}^{n} (x^p,\tau^p)\right) \equiv \prod_{p \in \bar{n}} v(x^p,\tau^p) , \tag{2.3}$$

so that V_n is C^∞, and a symmetric function of its n $(\mathbb{R}^d \times \mathbb{R})$-valued arguments. Let ∇_i $(i \in \bar{n})$ denote the gradient with respect to the ith vector argument; in particular,

$$\nabla_i V_n\left(\bigtimes_{p=1}^{n} (x^p,\tau^p)\right) = v(x^1,\tau^1) \cdots \nabla v(x^i,\tau^i) \cdots v(x^n,\tau^n) .$$

Finally consider a given tree $T \in \mathscr{T}\bar{n}$. With each edge $\beta = \{i_\beta, j_\beta\}$ in ET, we associate a differential operator b_β, which acts on potential products V_n of the form (2.3). This b_β is a function of an n-tuple $\xi = (\xi_1,...,\xi_n) \in I^n$ and is defined by

$$b_\beta \equiv b_\beta(\xi) \equiv g(\xi_{i_\beta},\xi_{j_\beta}) \nabla_{i_\beta} \cdot \nabla_{j_\beta} . \tag{2.4}$$

Notice, since g is symmetric and the gradients commute, that b_β is indeed a well-defined function of the unordered pair β. With these notations in place, the coefficient functions that play the key role in the series expansion of S can be defined.

Definition 2: For $n \in \mathbb{N}$ define the "tree sums" a_n: $(\mathbb{R}^d \times \mathbb{R})^2 \to \mathbb{R}$ by

$$a_n(x,t;k,s) \equiv \int_{I^n} d^n\xi \sum_{T \in \mathscr{T}\bar{n}} \left[\prod_{\beta \in ET} b_\beta(\xi)\right] \times V_n\left(\bigtimes_{p=1}^{n} w(\xi_p)\right) , \tag{2.5}$$

where if $n = 1$ the empty product is taken as unity.

Lemma 1: For all n, $a_n \in C^\infty$, and for $n \geqslant 2$,

$$|a_n| \leqslant n^{n-2} U^n K^{2(n-1)} . \tag{2.6}$$

Further, if $n \in \mathbb{N}$, $\alpha,\beta \in \mathbb{W}^d$, $l \in \mathbb{W}$, then for $|\alpha + \beta| > 0$,

$$|\nabla_1^\alpha \nabla_2^\beta a_n| \leqslant n^{n-2} U^n K^{2(n-1)} n^{|\alpha+\beta|} \times |\Delta t/m|^{|\beta|} (K/\sqrt{d})^{|\alpha+\beta|} , \tag{2.7}$$

while for $|\alpha| + l > 0$,

$$|\nabla_1^\alpha \partial_1^l a_n| \leqslant n^{n-2} U^n K^{2(n-1)} n^{|\alpha|+l} \times (B + (K|k|_1/\sqrt{d}\,m))^l (K/\sqrt{d})^{|\alpha|} . \tag{2.8}$$

In these equations $|k|_1 = \Sigma_{i=1}^d |k_i|$, the time displacement is denoted by $\Delta t \equiv t - s$, and ∇_i (resp. ∂_i) denotes the derivative with respect to the ith vector (resp. scalar) argument.

Proof: These results are straightforward consequences of Definition 1, and the fact $v \in \mathscr{U}$. For example, (2.6) results from Cayley's theorem and the fact that $2(n-1)$ gradients occur in $\Pi_{\beta \in ET} b_\beta$ if $T \in \mathscr{T}\bar{n}$. In (2.7) and (2.8) the interchange of differentiation and ξ integration $\int_{I^n} d^n\xi$ is justified by Theorem B3 of Ref. 21. ■

The constructive series representation of action S can now be defined using the tree sums a_n. It is convenient at this point to fix an initial time $s \in \mathbb{R}$, a $y \in \mathbb{R}^d$, and a mass m for the remainder of the discussion.

Definition 3: In terms of $T_H \equiv (1/K)(m/eU)^{1/2}$, let Ω_s be the time segment $(s - T_H, s + T_H)$, and define a function $\Phi \equiv \Phi(\,\cdot\,,\cdot\,;\cdot\,,s): \mathbb{R}^d \times \Omega_s \times \mathbb{R}^d \to \mathbb{R}$ by the series

$$\Phi(x,t;k,s) \equiv \sum_{n=1}^{\infty} \frac{(-)^{n-1} \Delta t^{2n-1}}{n! m^{n-1}} a_n(x,t;k,s) . \tag{2.9}$$

Also define $S \equiv S(\,\cdot\,,\cdot\,;\cdot\,,s)$: $\mathbb{R}^d \times \Omega_s \times \mathbb{R}^d \to \mathbb{R}$ by

$$S(x,t;k,s) \equiv k \cdot (x - y) - \Delta t(k^2/2m) - \Phi(x,t;k,s) . \tag{2.10}$$

Notice that the first two terms on the right-hand side of (2.10) are a solution of the free Hamilton–Jacobi equation [(1.13) with $v = 0$] with initial value $k \cdot (x - y)$. Since $\Phi = 0$ if $\Delta t = 0$, or if $v = 0$, then clearly Φ should determine the modification to this solution produced when v is present. The following lemma studies the convergence properties of the series Φ, thereby showing S is well defined.

Lemma 2: (a) For each fixed $t \in \Omega_s$, series (2.9) converges absolutely, and uniformly for $x,k \in \mathbb{R}^d$. The convergence is also uniform for t in compact subsets of Ω_s.

(b) Function $\Phi \in C^\infty$. For all $\alpha,\beta \in \mathbb{W}^d$, $l \in \mathbb{W}$, and $(x,t,k) \in \mathfrak{D}\Phi$ (the domain of Φ),

$$\nabla_1^\alpha \nabla_2^\beta \partial_1^l \Phi(x,t;k,s) = \sum_{n=1}^{\infty} \frac{(-)^{n-1}}{n! m^{n-1}} \left(\frac{d}{dt}\right)^l \times [\Delta t^{2n-1} \nabla_1^\alpha \nabla_2^\beta a_n(x,t;k,s)] . \tag{2.11}$$

This series converges absolutely, and uniformly for $x \in \mathbb{R}^d$ and (t,k) in compact subsets of $\Omega_s \times \mathbb{R}^d$. If $l = 0$, the convergence is uniform for $k \in \mathbb{R}^d$.

Proof: Both (a) and (b) are straightforward consequences of Lemma 1, standard tests for series convergence, and Theorem B3 of Ref. 21. Since the results of (b) are especially important in the remaining analysis, its proof will be sketched in the most tedious case where $|\beta|, l > 0$.

Write $z = (x,t,k) \in \mathfrak{D}\Phi$ and let

$$f_n(z) = [(-)^{n-1}/n!m^{n-1}]\Delta t^{2n-1}a_n(z,s)$$

be the summand in (2.9). According to Theorem B3 mentioned above, the following three conditions must be verified to prove (2.11): (i) for every n, $f_n \in C^\infty$; (ii) $\nabla^\gamma f_n(z)$ is absolutely summable over $n \in \mathbb{N}$, for all $\gamma \in \mathbb{W}^{2d+1}$, $z \in \mathfrak{D}\Phi$; and (iii) for every compact $Z \subset \mathfrak{D}\Phi$ and $\gamma \in \mathbb{W}^{2d+1}$ there is a summable sequence $\{b_n = b_n(Z,\gamma)\}$ such that

$$|\nabla^\gamma f_n(z)| \leqslant b_n \quad (z \in Z, \text{ all } n) .$$

In view of Lemma 1, condition (i) holds, and it suffices to verify the summability in (ii) and (iii) for $n \geqslant N$, where $N \in \mathbb{N}$ will be conveniently chosen. Turning to (ii), let $\gamma = (\alpha,l,\beta)$; since $f_n \in C^\infty$ we may differentiate in the following order

$$\nabla^\gamma f_n(z) = \left(\frac{d}{dt}\right)^l \nabla_k^\beta \nabla_x^\alpha f_n(x,t,k) .$$

As in the proof of Lemma 1, we have

$$\nabla_k^\beta \nabla_x^\alpha a_n = \int_{I^n} d^n\xi \sum_{T \in \mathscr{T}_{\bar{n}}} \left[\prod_{\beta \in ET} b_\beta\right] \nabla_k^\beta \nabla_x^\alpha V_n\left(\bigtimes_{p=1}^n w(\xi_p)\right).$$

From (2.2) and (2.3) one sees that each $\partial/\partial x_j$ in ∇_x^α becomes $\Sigma_{\rho=1}^n \nabla_\rho^{\delta_j}$, while each $\partial/\partial k_j$ in ∇_k^β becomes $\Sigma_{\rho=1}^n (\Delta t/m)(\xi_\rho - 1)\nabla_\rho^{\delta_j}$. Here δ is the Kronecker delta, so $\delta_j \in \mathbb{W}^d$ has value 1 at j, and 0 elsewhere.

Choose N large enough so that $2N - 1 + |\beta| > l$. Then

$$\begin{aligned}
&\sum_{n \geqslant N} |\nabla^\gamma f_n(z)| \\
&\quad = \sum_{n \geqslant N} (n!m^{n-1+|\beta|})^{-1} \left| \left(\frac{d}{dt}\right)^l \Delta t^{2n-1+|\beta|} \int_{I^n} d^n\xi \right. \\
&\qquad \times \sum_{T \in \mathscr{T}_{\bar{n}}} \left[\prod_{\beta \in ET} b_\beta\right] \prod_{j=1}^d \left[\left(\sum_{\rho=1}^n (\xi_\rho - 1)\nabla_\rho^{\delta_j}\right)^{\beta_j}\right. \\
&\qquad \left.\left. \times \left(\sum_{\rho=1}^n \nabla_\rho^{\delta_j}\right)^{\alpha_j}\right] V_n\left(\bigtimes_{p=1}^n w(\xi_p)\right)\right|
\end{aligned}$$

(which is obviously convergent if $\Delta t = 0$). Apply Leibnitz' rule to the action of $(d/dt)^l$; those powers of d/dt acting on $\nabla_k^\beta \nabla_x^\alpha a_n$ may be taken inside the ξ integral (by a proof similar to that of Lemma 1), and then replaced by

$$\sum_{\rho=1}^n \left\{\sum_{\sigma=1}^d (\xi_\rho - 1)\frac{k_\sigma}{m}\nabla_\rho^{\delta_\sigma} + \xi_\rho \partial_\rho\right\}.$$

Upon applying the bounds in Lemma 1 to the resulting derivatives of v, and using $|\xi_\rho - 1| \leqslant 1$, one finds after some rearrangement the majorizing series

$$\begin{aligned}
&\sum_{n \geqslant N} |\nabla^\gamma f_n(z)| \\
&\quad \leqslant \sum_{n \geqslant N} \left(\frac{K^2\Delta t^2 U}{m}\right)^n \frac{n^n}{n!} \frac{m^{1-|\beta|}}{K^2} \left(\frac{nK}{\sqrt{d}}\right)^{|\alpha+\beta|} |\Delta t|^{|\beta|-1} \\
&\qquad \times \sum_{r=0}^l \binom{l}{r} \frac{(2n-1+|\beta|)!}{(2n-1+|\beta|-r)!} |\Delta t|^{-r} \\
&\qquad \times \left\{B + \frac{K|k|_1}{\sqrt{d}\,m}\right\}^{l-r} n^{l-r-2} .
\end{aligned}$$

On the right-hand side here, employ the following bounds (for $0 \leqslant r \leqslant l$, $\Delta t \neq 0$):

$$\binom{l}{r} \leqslant l!, \quad |\Delta t|^{-r} \leqslant \max\{1, |\Delta t|^{-l}\},$$

$$(B + K|k|_1/(\sqrt{d}\,m))^{l-r} \leqslant (1 + B + K|k|_1/(\sqrt{d}\,m))^l .$$

Also note that

$$\begin{aligned}
&\sum_{r=0}^l n^{l-r-2}(2n + |\beta| - 1)(2n + |\beta| - 2) \\
&\qquad \times \cdots (2n + |\beta| - r) = n^{-2}\tilde{P}_l(n) ,
\end{aligned}$$

where $\tilde{P}_l(n)$ is a polynomial in n, of degree l, and is positive for $n \geqslant N$. These facts lead to the final estimate

$$\begin{aligned}
\sum_{n \geqslant N} |\nabla^\gamma f_n(z)| &\leqslant \sum_{n=N}^\infty \left(\frac{K^2\Delta t^2 U}{m}\right)^n \frac{n^n}{n!} \left(\frac{nK}{\sqrt{d}}\right)^{|\alpha+\beta|} K^{-2} \\
&\quad \times \left(\frac{|\Delta t|}{m}\right)^{|\beta|-1} \tilde{C} n^{l-2} , \qquad (2.12)
\end{aligned}$$

where $\tilde{C} = \tilde{C}(|\Delta t|, |k|_1, l, \beta, K, B, d)$ is independent of n. Employing the bound $n^n/n! \leqslant e^n/\sqrt{2\pi}$, and the fact $|\Delta t| < T_H$, then (2.12) is seen to be convergent by the ratio test, verifying (ii).

Finally (iii) follows from (2.12) noting that the summand is monotonically increasing in $|\Delta t|$, behaving near $\Delta t = 0$ like $|\Delta t|^{2n+|\beta|-1-l}$, which is nonsingular since $n \geqslant N$. ■

The next goal is to establish that S is indeed a solution of the Hamilton–Jacobi equation. This is the content of the following proposition and theorem.

Proposition 1: The family of functions $\{a_n\}_{n \geqslant 1}$ of Definition 2 satisfies the following integral identities for all $(x,t,k) \in \mathbb{R}^{2d+1}$:

$$a_1(x,t;k,s) = \int_I d\xi\, v(w(\xi)) , \qquad (2.13a)$$

$$\begin{aligned}
a_n(x,t;k,s) = &-\frac{1}{2}\sum_{l=1}^{n-1} \binom{n}{l} \int_I d\xi\, \xi^{2n-2} \\
&\times (\nabla_1 a_l \cdot \nabla_1 a_{n-l})(w(\xi);k,s) \quad (n \geqslant 2) .
\end{aligned} \qquad (2.13b)$$

Proof: See the Appendix. ■

Let $H_c: \mathbb{R}^{2d+1} \to \mathbb{R}$ be the classical Hamiltonian associated with (1.1), viz.,

$$H_c(x,p,t) = (1/2m)\,p^2 + v(x,t) . \qquad (2.14)$$

Theorem 1: Let $v \in \mathscr{V}$. For each $(k,s) \in \mathbb{R}^d \times \mathbb{R}$, the function $S(\cdot,\cdot;k,s)$: $\mathbb{R}^d \times \Omega_s \to \mathbb{R}$ is a solution of the Hamilton–Jacobi equation

$$\partial_1 S(x,t;k,s) + H_c(x,\nabla_1 S(x,t;k,s),t) = 0 . \qquad (HJ)$$

Proof: From (2.10) and Lemma 2 it is clear $S(\cdot,\cdot;k,s)$ is C^1 so it must be shown that (HJ) holds identically for $(x,t) \in \mathbb{R}^d \times \Omega_s$. A direct computation of the left-hand side of (HJ) using (2.10) results in

$$\begin{aligned}
\partial_1 S + H_c(x,\nabla_1 S,t) = &-\partial_1\Phi(x,t;k,s) - m^{-1}k \cdot \nabla_1\Phi \\
&+ (1/2m)(\nabla_1\Phi)^2 + v(x,t) ,
\end{aligned} \qquad (2.15)$$

where omitted function arguments are the same as their last appearance. Appealing to Lemma 2, substitute series (2.9) for Φ, differentiate term-by-term, and evaluate $(\nabla_1\Phi)^2$ by

the Cauchy product; the right-hand side of (2.15) becomes

$$\sum_{n=1}^{\infty}\frac{(-)^n}{n!}\left(\frac{\Delta t^2}{m}\right)^{n-1}\times\left\{\left[(2n-1)+\Delta t\frac{\partial}{\partial t}+\frac{\Delta t}{m}k\cdot\nabla_x\right]a_n(x,t;k,s)-\delta_{n,1}v(x,t)+\frac{1}{2}\sum_{l=1}^{n-1}\binom{n}{l}(\nabla_1 a_l\cdot\nabla_1 a_{n-l})(x,t;k,s)\right\}. \tag{2.16}$$

The $\Sigma_{l=1}^{n-1}$ is absent if $n=1$. It will now be shown that the $\{\cdots\}$ in (2.16) vanishes for all $n\in\mathbb{N}$. Consider the $n=1$ case (similar arguments apply if $n\geqslant 2$). Replace (x,t) with $w(\lambda)$ in (2.13a), where $\lambda\in I$. In doing this the following composition law is used:

$$w(\xi)|_{(x,t)=w(\lambda)}=([x+(\lambda-1)\Delta tk/m]+(\xi-1)[\lambda(t-s)]k/m,\ s+\xi[\lambda(t-s)])=w(\xi\lambda).$$

Therefore (2.13a) has become

$$a_1(w(\lambda);k,s)=\int_I d\xi\, v(w(\xi\lambda)).$$

Now multiply this by λ, change the integration variable to $\gamma=\xi\lambda$, differentiate with respect to λ, and set $\lambda=1$ to find

$$\left[1+\Delta t\frac{\partial}{\partial t}+\frac{\Delta t}{m}k\cdot\nabla_x\right]a_1(x,t;k,s)=v(x,t).$$

Hence the $n=1$ term in (2.16) is zero. ■

At this point, it is worth noticing two useful consequences of the Hamilton–Jacobi equation and the smoothness of S. First, if (HJ) is differentiated with respect to the parameter k, one obtains

$$\partial_1\nabla_2 S(x,t;k,s)+(1/m)\nabla_2\nabla_1 S\,\nabla_1 S=0; \tag{2.17}$$

in the second term the $d\times d$ matrix $\nabla_2\nabla_1 S$ [with elements $(\nabla_2\nabla_1 S)_{i,j}\equiv\nabla_2^i\nabla_1^j S$] multiplies the vector $\nabla_1 S$. The second consequence concerns the determinant $D\equiv D(\cdot,\cdot;\cdot,s):\mathbb{R}^d\times\Omega_s\times\mathbb{R}^d\to\mathbb{R}$ defined by

$$D(x,t;k,s)\equiv\det\nabla_1\nabla_2 S(x,t;k,s)=\det[\delta-\nabla_1\nabla_2\Phi(x,t;k,s)], \tag{2.18}$$

where δ is the unit matrix (i.e., Kronecker delta).

Lemma 3: For all $(x,t,k)\in\mathcal{D}D$, the following continuity equation is obeyed:

$$\partial_1 D(x,t;k,s)+\nabla_x\cdot\{D(x,t;k,s)\nabla_2 H_c(x,\nabla_1 S(x,t;k,s),t)\}=0. \tag{2.19}$$

Proof: This follows from differentiating (HJ) with respect to x_i and k_j, and performing a number of algebraic manipulations. ■

I now introduce a slightly smaller time domain than Ω_s, on which the remaining analysis will be done. The reason for this is that on this smaller domain the useful quantities derived from S are particularly neatly characterized.

Definition 4: Given $\sigma\in(0,1)$, define $t(\sigma)\equiv[1+1/\sigma\sqrt{2\pi}]^{-1/2}T_H$, and introduce the time segment

$$\Omega_s(\sigma)=(s-t(\sigma),s+t(\sigma)). \tag{2.20}$$

Note that $t(\sigma)\to 0$ as $\sigma\to 0$; it will always be assumed that $\sigma\in(0,1)$. The family of convergent series $\gamma:[0,1)\times\mathbb{Z}\to\mathbb{R}$ (where $\mathbb{Z}$ denotes the integers),

$$\gamma(\lambda_\sigma,l)\equiv\sum_{n=1}^{\infty}\lambda_\sigma{}^n n^l, \tag{2.21}$$

$$\lambda_\sigma\equiv(1+(\sigma\sqrt{2\pi})^{-1})^{-1}\in(0,1), \tag{2.22}$$

will occur frequently in subsequent estimates.

Lemma 4: For $x,k\in\mathbb{R}^d$, $t\in\Omega_s(\sigma)$ and $\alpha,\beta\in\mathbb{W}^d$, $|\alpha+\beta|>0$ one has the estimate

$$|\nabla_1^\alpha\nabla_2^\beta\Phi(x,t;k,s)|<\frac{K^{|\alpha+\beta|-2}}{\sqrt{2\pi}\sqrt{d}^{|\alpha+\beta|}}\left(\frac{t(\sigma)}{m}\right)^{|\beta|-1}\gamma(\lambda_\sigma,|\alpha+\beta|-2). \tag{2.23}$$

In particular if $\alpha=\delta_i,\beta=\delta_j$ $(i,j\in\bar{d})$,

$$|\nabla_1^i\nabla_2^j\Phi(x,t;k,s)|<\sigma/d, \tag{2.24}$$

where $\nabla_l^i\equiv\nabla_l^{\delta_l^i}$ is the ith component of the gradient on the lth vector argument.

Proof: Estimate (2.23) follows at once from (2.11) and (2.7). ■

Based on these explicit estimates, many of the critical properties of the relevant quantities derived from S will now be obtained.

Proposition 2: For any $t\in\Omega_s(\sigma)$, $k\in\mathbb{R}^d$ the following map is a C^∞ diffeomorphism:

$$f\equiv\nabla_2 S(\cdot,t;k,s):\ \mathbb{R}^d\to\mathbb{R}^d.$$

Remark: For comparative purposes, I point out that if a study of the classical trajectories were to be made, then this diffeomorphism would turn out to be the map $x\mapsto y(t,x)$ in (1.9). However, no use will be made of this fact.

Proof: Lemma 2 implies f is C^∞. Let $u\in\mathbb{R}^d$ be arbitrary and consider the map $F:\ \mathbb{R}^d\to\mathbb{R}^d$ defined by

$$F(x)\equiv(\Delta t/m)k+u+\nabla_2\Phi(x,t;k,s). \tag{2.25}$$

If $x,x'\in\mathbb{R}^d$, applying Taylor's formula and Schwarz' inequality to the jth component of F shows

$$|F_j(x)-F_j(x')|\leqslant|x-x'|\,|\nabla F_j(x'+\lambda(x-x'))|,$$

for some $\lambda\in(0,1)$. But (2.25) implies that for any $z\in\mathbb{R}^d$,

$$|\nabla F_j(z)|=|\nabla_1\nabla_2^j\Phi(z,t;k,s)|<\left(\sum_{i=1}^{d}\left(\frac{\sigma}{d}\right)^2\right)^{1/2}=\frac{\sigma}{\sqrt{d}},$$

where (2.24) was used. Hence

$$|F(x)-F(x')|=\left(\sum_{j=1}^{d}(F_j(x)-F_j(x'))^2\right)^{1/2}<\sigma|x-x'|.$$

Since $\sigma<1$, F is a contraction mapping and therefore possesses a unique fixed point, say $x_*=F(x_*)$. From the definitions of f and F, this is equivalent to the statement $f(x_*)=u$. This shows f is surjective, while the uniqueness of x_* implies that f is injective. Finally, $f^{-1}\in C^\infty$ by the inverse function theorem. ■

The WKB error estimate to be found in Sec. IV requires some bounds on the determinant D, and its derivatives. This section concludes with a study of these. It is useful to begin by establishing some notation for matrix norms, and recalling a few basic facts from matrix analysis.

Let $A,B\in\mathbb{C}^{d\times d}$ be $d\times d$ complex-valued matrices. The operator norm of the linear transformation associated with A is

$$\|A\|\equiv\sup\{|Au|\ |u\in\mathbb{C}^d,\ |u|=1\},$$

where $|u|^2=\Sigma_{i=1}^d|u_i|^2$. This norm has two well-known properties of interest. The first is that

$$|\det A|\leqslant\|A\|^d, \tag{2.26}$$

while the second[22] (Neumann series) shows that if $\|B\|<1$, then $\delta+B$ is invertible and

$$\|(\delta+B)^{-1}\|\leqslant(1-\|B\|)^{-1}. \tag{2.27}$$

Another useful matrix norm is the Hilbert–Schmidt norm

$$\|A\|_2\equiv\left(\sum_{i,j=1}^d|A_{i,j}|^2\right)^{1/2}.$$

It may be used to estimate a trace via

$$|\mathrm{Tr}(AB)|\leqslant\|A\|_2\|B\|_2, \tag{2.28}$$

which is seen by regarding $\mathrm{Tr}(AB)$ as a Hermitian inner product of two vectors $A^\dagger,B$ on the index set $\bar{d}\times\bar{d}$, and applying Schwarz' inequality. The matrix norms $\|\cdot\|$ and $\|\cdot\|_2$ are both algebra norms, viz.,

$$\|AB\|\leqslant\|A\|\,\|B\|,\quad \|AB\|_2\leqslant\|A\|_2\|B\|_2, \tag{2.29}$$

and are numerically related by

$$\|A\|\leqslant\|A\|_2\leqslant\sqrt{d}\,\|A\|. \tag{2.30}$$

The derivative of a determinant will be treated as follows. If A: $(a,b)\subset\mathbb{R}\to\mathbb{C}^{d\times d}$ is C^1 and everywhere invertible, then one has the well-known result

$$\frac{d}{d\lambda}\det A(\lambda)=(\det A(\lambda))\mathrm{Tr}\left[A(\lambda)^{-1}\frac{d}{d\lambda}A(\lambda)\right], \tag{2.31}$$

which follows from differentiating the permutation expansion of the determinant and manipulating the resulting cofactors. The inverse matrix appearing in (2.31) may be estimated by[23]

$$\|A^{-1}\|\leqslant\|A\|^{d-1}/|\det A|. \tag{2.32}$$

Lemma 5: Let $x,k\in\mathbb{R}^d$, $t\in\Omega_s(\sigma)$, and $i,j\in\bar{d}$. Then

$$(1-\sigma)^d<D(x,t;k,s)<(1+\sigma)^d, \tag{2.33}$$

$$|\nabla_1^i D(x,t;k,s)|\leqslant(K/\sqrt{2\pi})\gamma(\lambda_\sigma,1)(1+\sigma)^{d-1}, \tag{2.34}$$

$$\begin{aligned}|\nabla_1^i\nabla_1^j D(x,t;k,s)|&\leqslant\left[(K/\sqrt{2\pi})\gamma(\lambda_\sigma,1)(1+\sigma)^{d-1}\right]^2\\&\quad\times(1-\sigma)^{-d}(1+d^{-1/2})\\&\quad+(K^2/\sqrt{2\pi})\gamma(\lambda_\sigma,2)(1+\sigma)^{d-1}d^{-1/2}.\end{aligned} \tag{2.35}$$

Remarks: (a) Bounds for $|\nabla_1 D|$ (and the Laplacian $|\Delta_1 D|$) are obtained by multiplying the constants on the right-hand side of (2.34) [and (2.35)] by $\sqrt{d}$ (and d).

(b) Result (2.33) together with Theorem 1 implies that S is a complete integral (in the sense of Jacobi's theorem[24]) of the Hamilton–Jacobi equation, on the domain $\mathbb{R}^d\times\Omega_s(\sigma)\times\mathbb{R}^d$.

Proof: Begin with (2.33). Let $A=\delta-\nabla_1\nabla_2\Phi$ be the matrix whose determinant is D. Using (2.24) and (2.30) we have $\|-\nabla_1\nabla_2\Phi\|<\sigma<1$. It follows that

$$|D|\leqslant\|A\|^d<(1+\sigma)^d,$$

which proves the right inequality in (2.33). Applying the Neumann series result to A gives

$$\|A^{-1}\|\leqslant(1-\|\nabla_1\nabla_2\Phi\|)^{-1}<(1-\sigma)^{-1}.$$

Consequently (2.26) implies

$$|D|=|\det A^{-1}|^{-1}\geqslant\|A^{-1}\|^{-d}>(1-\sigma)^d.$$

This proves (2.33) once it is seen that $D>0$. But this is true because $\delta-\xi\nabla_1\nabla_2\Phi$ is invertible for $\xi\in[0,1]$, and $[0,1]\ni\xi\mapsto\det(\delta-\xi\nabla_1\nabla_2\Phi)\in\mathbb{R}$ is continuous, with value 1 at $\xi=0$.

Turning to (2.34), we employ (2.31) with $\lambda=x_i$ and find

$$|\nabla_1^i D|=D\,|\mathrm{Tr}(A^{-1}\nabla_1^i A)|\leqslant D\,\|A^{-1}\|_2\|\nabla_1^i A\|_2. \tag{2.36}$$

But $D\,\|A^{-1}\|_2\leqslant D\sqrt{d}\,\|A^{-1}\|\leqslant\sqrt{d}\,\|A\|^{d-1}\leqslant\sqrt{d}\,(1+\sigma)^{d-1}$, and from (2.23) one easily gets

$$\|\nabla_1^i A\|_2=\|\nabla_1^i\nabla_1\nabla_2\Phi\|_2\leqslant(K/\sqrt{2\pi})\gamma(\lambda_\sigma,1)d^{-1/2}.$$

Inserting these results into (2.36) establishes (2.34). A similar pattern of estimates applied to the identity

$$\begin{aligned}\nabla_1^i\nabla_1^j D&=\nabla_1^j D\,\mathrm{Tr}(A^{-1}\nabla_1^i A)\\&\quad+D\,\mathrm{Tr}(A^{-1}[\nabla_1^j\nabla_1^i A-\nabla_1^j A A^{-1}\nabla_1^i A])\end{aligned}$$

leads to (2.35). ■

III. THE QUANTUM SYSTEM

In this brief section, the quantum system associated with the wave-mechanical Hamiltonian (1.1) is defined. The relevant properties of the time evolution governed by Schrödinger's equation are also recalled. Finally, the initial value problem, whose WKB approximation is sought, will be formulated.

Definition 5: Let K_0: $D_0\to\mathscr{H}=L^2(\mathbb{R}^d)$ denote the self-adjoint extension of the Laplacian on $\mathbb{R}^d$; specifically

$$D_0\equiv\{\psi\in\mathscr{H}|\alpha^2\mathscr{F}\psi(\alpha)\in L^2(\mathbb{R}^d_\alpha)\}, \tag{3.1a}$$

$$\mathscr{F}K_0\psi(\alpha)\equiv\alpha^2\mathscr{F}\psi(\alpha), \tag{3.1b}$$

where $\mathscr{F}\in\mathscr{B}(\mathscr{H})$ (the bounded operators on $\mathscr{H}\to\mathscr{H}$) denotes the unitary Fourier transform,

$$\mathscr{F}\psi(\alpha)=(2\pi)^{-d/2}\int dx\,e^{-i\alpha\cdot x}\psi(x)\quad(\psi\in L^1\cap L^2).$$

For $\hbar,m>0$ the free Hamiltonian is defined by $H_0=-(\hbar^2/2m)K_0$: $D_0\to\mathscr{H}$. For $v\in\mathscr{V}$ and $t\in\mathbb{R}$, one has the potential energy operator $V(t)\in\mathscr{B}(\mathscr{H})$, $V(t)\psi(x)\equiv v(x,t)\psi(x)$. Finally, the total Hamiltonian is defined as the self-adjoint operator

$$H(t)=H_0+V(t):\ D_0\to\mathscr{H}. \tag{3.2}$$

Proposition 3: There exists a two-parameter family of "Schrödinger evolution" operators $\{U(t,s)\in\mathscr{B}(\mathscr{H})|t,s\in\mathbb{R}\}$ with the properties (for all t,s):

(a) $U(t,s)(D_0)\subset D_0$;

(b) $\|U(t,s)\|=1$;

(c) $U(t,\cdot)$ is strongly continuous;

(d) $U(s,s)=I$ (the identity on $\mathscr{H}$);

and

(e) on the domain D_0, $U(t,s)$ is strongly differentiable with respect to t and s (separately), and satisfies

$$i\hbar\frac{d}{dt}U(t,s)\psi = H(t)U(t,s)\psi\,, \tag{3.3a}$$

$$-i\hbar\frac{d}{ds}U(t,s)\psi = U(t,s)H(s)\psi\,, \tag{3.3b}$$

for all $\psi\in D_0$.

Remark: Of course the operators $U(t,s)$ have a number of additional properties, such as unitarity. Those listed above are sufficient for the subsequent development.

Proof: The results follow readily from the general theory[25] of linear evolution equations in Banach space, with time-dependent unbounded operator coefficients, together with the facts that $H(t)$ is self-adjoint with t-invariant domain D_0, and $t\mapsto V(t)$ is strongly continuously differentiable. ■

The next result is required in order to compute the action of $H(t)$ on a smooth WKB wave function.

Lemma 6: Let $D_1\equiv\{\psi\in C^2\cap L^2(\mathbb{R}^d)\,|\,\Delta\psi\in L^2\}$, where $\Delta\psi$ denotes the ordinary Laplacian of the C^2 function ψ: $\mathbb{R}^d\to\mathbb{C}$. Then $D_1\subset D_0$ and $K_0\psi=\Delta\psi$ for $\psi\in D_1$.

Proof: This follows from a standard functional-analytic argument proving that D_1 is a core for K_0. (See Ref. 22, pp. 298 and 299.) ■

The preceding results in this section, even though well known, would require a considerable amount of work to prove from first principles. However, since such results are ultimately an essential part of any serious discussion of quantum evolution (and its approximation), the work required to prove them should not be regarded as contributing to the level of difficulty of the WKB method being presented here.

Let us now consider the Cauchy initial value problem (IVP) for Schrödinger's equation. Suppose $\psi\in D_0$ is given. Then Proposition 3 implies that

$$\mathbf{R}\ni t\mapsto\Psi_t\equiv U(t,s)\psi\in\mathscr{H} \tag{3.4}$$

is a strict solution of Schrödinger's equation with initial value ψ. That is, $\Psi_t\in\mathscr{D}H(t)$ for all t, (1.2) holds, and $\Psi_s=\psi$. More generally, if $\psi\in\mathscr{H}\setminus D_0$, then (3.4) defines a generalized solution of the IVP.

In the next section, initial data ψ of the form (1.3) will be considered. Specifically, I will assume that

$$\psi(x)=\varphi(x)\exp[(i/\hbar)k\cdot(x-y)]\,, \tag{3.5}$$

where $\varphi\in C^3(\mathbb{R}^d;\mathbb{C})$ and $\nabla^\alpha\varphi\in\mathscr{H}$ for all $\alpha\in\mathbb{W}^d$ with $|\alpha|\leqslant 3$. A simple calculation then shows $\psi\in D_1\subset D_0$, so the resultant exact solution (3.4) of the Schrödinger IVP (1.2) and (1.3) is in fact a strict solution. Moreover, the initial states of the form (3.5) form a dense linear manifold in $\mathscr{H}$.

IV. THE WKB WAVE FUNCTION AND ERROR ESTIMATE

This section contains the analysis of the lowest order $\hbar\to 0$ asymptotics of the Schrödinger IVP (1.2) and (1.3), for times in $\Omega_s(\sigma)$. The constructive series solution S of the Hamilton–Jacobi equation is used to define the WKB-approximation wave function Υ. A few simple properties of Υ are derived, and used to show that Υ is asymptotic, as $\hbar\to 0$, to the exact solution of the IVP, with respect to the norm of $\mathscr{H}$.

Definition 6: Denote by f: $\mathbb{R}^d\times\Omega_s(\sigma)\times\mathbb{R}^d\to\mathbb{R}^d$ the map

$$f(x)\equiv f(x;t,k,s)=\nabla_2 S(x,t;k,s)\,. \tag{4.1}$$

In terms of f, S, and D define the "lowest order WKB wave function" Υ: $\mathbb{R}^d\times\Omega_s(\sigma)\to\mathbb{C}$ by

$$\Upsilon(x,t)\equiv\sqrt{D(x,t;k,s)}\,\varphi(f(x))\exp[(i/\hbar)S(x,t;k,s)]\,, \tag{4.2}$$

where $k,y\in\mathbb{R}^d$ are the parameters appearing in the initial data (1.3), and $\varphi\in\mathscr{H}$ has the properties assumed at the end of Sec. III.

As indicated in (4.1), the arguments t,k,s of f will be notationally suppressed, the reason being that it is useful to regard $x\mapsto f(x)$ as a diffeomorphism of $\mathbb{R}^d$, according to Proposition 2. Regarding (4.2) recall that $D>0$ [Eq. (2.33)]. From the initial $t=s$ values $S=k\cdot(x-y)$, $f=x$, $D=1$, it is seen that Υ satisfies the initial condition (1.3) exactly,

$$\Upsilon(x,s)=\varphi(x)\exp[(i/\hbar)k\cdot(x-y)]\,. \tag{4.3}$$

Two technical properties of Υ which allow the Hilbert space operations H_0 and d/dt to be applied to Υ via partial differentiation are obtained next.

Lemma 7: (a) For each $t\in\Omega_s(\sigma)$,

$$\Upsilon(\,\cdot\,,t)\in D_1=\{\psi\in C^2\cap L^2\,|\,\Delta\psi\in L^2\}\,.$$

(b) The map $\Omega_s(\sigma)\ni t\mapsto\Upsilon(\,\cdot\,,t)\in L^2$ is strongly continuously differentiable, with

$$\left[\frac{d}{dt}\Upsilon(\,\cdot\,,t)\right](x)=\partial\Upsilon(x,t)\quad[t\in\Omega_s(\sigma),\ \text{a.a. } x]\,.$$

Proof: (a) From Lemma 2 and the hypothesis on φ, it is clear $\Upsilon(\,\cdot\,,t)\in C^2$. Also, upon noting that D is just the Jacobian determinant of f, one employs the change of integration variables $x'=f(x)$ to show $\Upsilon\in L^2$,

$$\int dx|\Upsilon(x,t)|^2=\int dx\,D(x,t;k,s)|\varphi(f(x))|^2$$
$$=\int dx'|\varphi(x')|^2=\|\varphi\|^2<\infty\,.$$

It remains to show $\Delta\Upsilon(\,\cdot\,,t)\in L^2$. A straightforward differentiation of (4.2) with respect to x yields the estimate

$$|\Delta\Upsilon(x,t)|$$
$$\leqslant D^{1/2}\Bigg\{\Bigg[\left(\frac{\nabla_1 D}{2D}\right)^2+\left(\frac{1}{\hbar}\nabla_1 S\right)^2+\frac{|\Delta_1 D|}{2D}+\frac{1}{\hbar}|\Delta_1 S|\Bigg]|\varphi(f(x))|$$
$$+\left[|\Delta f(x)|+\left(\frac{1}{D}|\nabla_1 D|+\frac{2}{\hbar}|\nabla_1 S|\right)\|\nabla f(x)\|_2\right]$$
$$\times|\nabla\varphi(f(x))|+\|\nabla f(x)\|_2^2\|\nabla\nabla\varphi(f(x))\|_2\Bigg\}\,. \tag{4.4}$$

The next step is to observe that each φ-independent factor in the $\{\cdots\}$ of (4.4) has a finite x-independent bound. For example, in Lemma 5 it was shown that $D^{-1}<(1-\sigma)^{-d}$ and $|\nabla_1 D|$, $|\Delta_1 D|<\text{const}$. In a similar fashion it is possible to

derive x-independent bounds for the other quantities $|\nabla_1 S|$, $|\Delta_1 S|$, $\|\nabla f(x)\|_2$, $|\Delta f|$ appearing in (4.4). These bounds will not be exhibited here because their precise form is not required in the remaining analysis; it is only their existence at this stage which is of importance.

The result of these observations is that

$$\int dx|\Delta\Upsilon(x,t)|^2 \leqslant \int dx\, D(x,t;k,s)(\tilde{C}_0|\varphi(f(x))| + \tilde{C}_1|\nabla\varphi(f(x))| + \tilde{C}_2\|\nabla\nabla\varphi(f(x))\|_2)^2,$$

where $\tilde{C}_i$ depend on k and σ but not on x. After changing the variable of integration again, and noting that $|\varphi|, |\nabla\varphi|, \|\nabla\nabla\varphi\|_2 \in L^2$ by hypothesis, we see $\Delta\Upsilon(\cdot,t)\in L^2$.

(b) Here the proof relies on a bound study of the first and second partials, $\partial\Upsilon$ and $\partial^2\Upsilon$, of Υ with respect to the time argument. Consider the more difficult case of $\partial^2\Upsilon$. After differentiating (4.2) twice with respect to t, one obtains an estimate quite analogous to (4.4) except that the derivatives of Υ, D, S, and f with respect to x are now with respect to t. Again, it is sought to bound the φ-independent coefficient functions by x-independent constants. This can be accomplished by first eliminating the occurrence of any t derivatives of S or D through repeated use of the Hamilton–Jacobi (HJ) and continuity (2.19) equations. Once this is done, the resulting coefficient functions only contain derivatives of S or D, with respect to x and k, and they may be estimated analogously to (a) above. These steps result in

$$|\partial^2\Upsilon(x,t)| \leqslant D(x,t;k,s)^{1/2}\{\hat{C}_0|\varphi(f(x))| + \hat{C}_1|\nabla\varphi(f(x))| + \hat{C}_2\|\nabla\nabla\varphi(f(x))\|_2\}, \tag{4.5}$$

with $\hat{C}_i$ independent of x. This implies $\partial^2\Upsilon(\cdot,t)\in L^2$, and a similar (but simpler) calculation shows $\partial\Upsilon(\cdot,t)\in L^2$ as well.

With these results, the strong derivative $(d/dt)\Upsilon(\cdot,t)$, $t\in\Omega_s(\sigma)$, can be shown to exist and equal $\partial\Upsilon(\cdot,t)$. In fact, by employing Taylor's formula, one has

$$\left\|\frac{1}{t'-t}[\Upsilon(\cdot,t')-\Upsilon(\cdot,t)]-\partial\Upsilon(\cdot,t)\right\|^2 = \left(\frac{t'-t}{2}\right)^2\int dx|\partial^2\Upsilon(x,\tau)|^2, \tag{4.6}$$

where $|\tau - t| < |t' - t|$. Since the constants $\hat{C}_i$ in (4.5) are independent of $t\in\Omega_s(\sigma)$, it is seen that the right-hand side of (4.6) tends to zero as $t'\to t$. A similar argument shows that $t\mapsto\partial\Upsilon(\cdot,t)$ is strongly continuous. ■

Another ingredient of the error analysis is the "approximate Schrödinger equation" satisfied by Υ; it will be studied next.

Lemma 8: For $t\in\Omega_s(\sigma)$ the following identity holds:

$$i\hbar\frac{d}{dt}\Upsilon(\cdot,t) = H(t)\Upsilon(\cdot,t) + \frac{\hbar^2}{2m}g_t, \tag{4.7}$$

where $g_t\in\mathscr{H}$ has the explicit form

$$g_t(x) = D(x,t;k,s)^{1/2}\exp\left[\frac{i}{\hbar}S(x,t;k,s)\right] \times\left\{\left[\frac{\Delta_1 D}{2D}-\left(\frac{\nabla_1 D}{2D}\right)^2\right]\varphi(f(x)) + \frac{1}{D}\nabla_1 D\cdot\nabla(\varphi\circ f)(x) + \Delta(\varphi\circ f)(x)\right\}. \tag{4.8}$$

Furthermore the map $t\mapsto g_t$ is strongly continuous.

Proof: Thanks to Lemmas 7 and 6 we may compute as follows:

$$\begin{aligned}\left[\left(i\hbar\frac{d}{dt}-H(t)\right)\Upsilon(\cdot,t)\right](x) &= i\hbar\,\partial\Upsilon(x,t)+\frac{\hbar^2}{2m}\Delta\Upsilon(x,t)-v(x,t)\Upsilon(x,t)\\ &= D^{1/2}\exp\left[\frac{i}{\hbar}S\right]\left\{-\varphi(f(x))\left[\partial_1 S+\frac{1}{2m}(\nabla_1 S)^2+v\right]\right.\\ &\quad + i\hbar\varphi(f(x))(2D)^{-1}\\ &\quad\times\left[\partial_1 D+\frac{1}{m}(\nabla_1 D\cdot\nabla_1 S+D\Delta_1 S)\right]\\ &\quad\left.+i\hbar\nabla\varphi(f(x))\cdot\left[\frac{\partial f}{\partial t}(x)+\frac{1}{m}\widetilde{\nabla f}(x)\nabla_1 S\right]\right\}\\ &\quad+(\hbar^2/2m)\,g_t(x),\end{aligned} \tag{4.9}$$

where $\partial f/\partial t\equiv\partial_1\nabla_2 S$ and $g_t(x)$ is given by (4.8). Now the entire $\{\cdots\}$ in (4.9) vanishes due to (HJ), (2.19), and (2.17). This implies $g_t\in\mathscr{H}$ and that (4.7) holds.

In order to prove the strong continuity of $t\mapsto g_t$, one cannot rely on Lemma 7 because H_0 is an unbounded operator. Instead one relies on the by now familiar type of argument used in the proof of Lemma 7. Such an analysis leads to the bound

$$\left|\frac{\partial}{\partial t}g_t(x)\right| \leqslant D(x,t;k,s)^{1/2}\sum_{\substack{\alpha\in\mathbb{W}^d\\|\alpha|<3}}C'_\alpha(\sigma)|\nabla^\alpha\varphi(f(x))| \qquad [t\in\Omega_s(\sigma)],$$

with $C'_\alpha(\sigma)$ independent of x. With this estimate g_t is easily shown to be strongly continuous in t. It is this part of the analysis that requires the hypothesis that the third partials of φ remain in L^2. ■

So far, Eq. (4.7) only says that Υ satisfies Schrödinger's equation approximately—to within a term $(\hbar^2/2m)g_t$ which tends (in L^2 norm, and in modulus) to zero as $\hbar\to 0$. But in quantum mechanics, Schrödinger's equation is an evolution equation in Hilbert space, and so one really wants to show that $\Upsilon(\cdot,t)$ is close to the exact solution Ψ_t in the sense of the norm of $\mathscr{H}$. This brings us to the principal result of the WKB analysis.

Theorem 2: Let $v\in\mathscr{V}$, $\hbar, m>0$, $s\in\mathbb{R}$, and $\sigma\in(0,1)$. Let $H(t)$ be as in Definition 5, and $\Psi_\cdot:\mathbb{R}\to\mathscr{H}$ be the solution of the Schrödinger initial value problem,

$$i\hbar\frac{d}{dt}\Psi_t = H(t)\Psi_t \quad (t\in\mathbb{R}), \tag{1.2}$$

$$\Psi_s(x) = \varphi(x)\exp[(i/\hbar)k\cdot(x-y)] \quad (x\in\mathbb{R}^d), \tag{1.3}$$

where $\varphi\in C^3(\mathbb{R}^d;\mathbb{C})$ and $\nabla^\alpha\varphi\in\mathscr{H}$ for all $\alpha\in\mathbb{W}^d$ with $|\alpha|\leqslant 3$.

Let Υ be constructed via Definition 6. Then for all $t\in\Omega_s(\sigma)$,

$$\|\Psi_t - \Upsilon(\,\cdot\,,t)\| \leqslant C_\varphi(\sigma)|t-s|\hbar/m\,, \tag{4.10}$$

where the finite constant is

$$C_\varphi(\sigma) \equiv \tfrac{1}{2}\|C_0(\sigma)|\varphi| + C_1(\sigma)|\nabla\varphi| + C_2(\sigma)\|\nabla\nabla\varphi\|_2\|\,, \tag{4.11}$$

and $C_i(\sigma)$ are given in (4.16) below.

Proof: The first stage of the proof is to obtain the following integrated form of (4.7):

$$\Upsilon(\,\cdot\,,t) = U(t,s)\Upsilon(\,\cdot\,,s) - \frac{i\hbar}{2m}\int_s^t d\tau\, U(t,\tau)\, g_\tau\,. \tag{4.12}$$

Here the strong Riemann integral[26] of a continuous function of τ is used. Equation (4.12) is trivial if $t=s$, and more generally follows by integrating the identity

$$\frac{d}{d\tau}\,[U(t,\tau)\Upsilon(\,\cdot\,,\tau)] = -\frac{i\hbar}{2m}\,U(t,\tau)g_\tau\,. \tag{4.13}$$

This identity (which is simple to obtain formally) is a consequence of (4.7), the strong differentiability of $U(t,\,\cdot\,)$ on $D_0 \ni \Upsilon(\,\cdot\,,\tau)$, the backward equation of motion (3.3b), and the strong continuity of $U(t,\,\cdot\,)$. Since also $\tau\mapsto g_\tau$ is strongly continuous, (4.13) may be integrated to yield (4.12).

The exact solution (3.4) obeys a similar equation,

$$\Psi_t = U(t,s)\Psi_s = U(t,s)\Upsilon(\,\cdot\,,s)\,, \tag{4.14}$$

by (1.3) and (4.3). Upon subtracting (4.12) from (4.14), taking the L^2 norm of the result, and employing the unit norm of $U(t,s)$, one has

$$\|\Psi_t - \Upsilon(\,\cdot\,,t)\| \leqslant (\hbar/2m)|t-s| \sup_{\tau\in\Omega_s(\sigma)}\|g_\tau\|\,.$$

This is essentially (4.10). Using the formula (4.8) for $g_t(x)$ yields the estimate

$$\begin{aligned}\|g_t\|^2 \leqslant \int dx\, D(x,t;k,s)\Bigg[&\left\{\frac{|\Delta_1 D|}{2D} + \left|\frac{\nabla_1 D}{2D}\right|^2\right\}|\varphi(f(x))| \\ &+ \{\|\nabla f(x)\|D^{-1}|\nabla_1 D| + |\Delta f(x)|\}|\nabla\varphi(f(x))| \\ &+ \{\|\nabla f(x)\|_2^2\}\|\nabla\nabla\varphi(f(x))\|_2\Bigg]^2\,.\end{aligned} \tag{4.15}$$

The quantities in the three $\{\cdots\}$ in (4.15) all have x-independent bounds. The bounds on D and its derivatives, and the result $\|\nabla f(x)\| = \|\delta - \nabla_1\nabla_2\Phi\| \leqslant 1+\sigma$, were obtained in proving Lemma 5. Similarly, $\|\nabla f(x)\|_2 \leqslant \sqrt{d}+\sigma$. Using the results of Lemma 2 it is easy to show

$$|\Delta f(x)| < (K/\sqrt{2\pi})\gamma(\lambda_\sigma,1)\,.$$

Upon using these bounds, and changing integration variables in (4.15), one arrives at (4.11) with the constants

$$\begin{aligned}C_0(\sigma) = (1-\sigma)^{-2d}&\left[(K/\sqrt{2\pi})\gamma(\lambda_\sigma,1)(1+\sigma)^{d-1}\right]^2 \\ &\times(\tfrac{3}{4}d + \tfrac{1}{2}\sqrt{d}) + \tfrac{1}{2}(1-\sigma)^{-d}(K^2/\sqrt{2\pi}) \\ &\times\gamma(\lambda_\sigma,2)(1+\sigma)^{d-1}\sqrt{d}\,,\end{aligned} \tag{4.16a}$$

$$C_1(\sigma) = (K/\sqrt{2\pi})\gamma(\lambda_\sigma,1)\left[((1+\sigma)/(1-\sigma))^d\sqrt{d}+1\right], \tag{4.16b}$$

$$C_2(\sigma) = (\sqrt{d}+\sigma)^2\,. \tag{4.16c}$$

The $\gamma(\lambda_\sigma,l)$ were defined in (2.21) and (2.22). ■

According to (4.10), the error in the WKB state $\Upsilon(\,\cdot\,,t)$ is also small if the time displacement is small, or if the mass is large. In the case of small $|\Delta t\,|$, one can choose σ to be close to 0. Since $\lim_{\sigma\to 0}C_j(\sigma)\propto\delta_{j,2}$, one sees that the error for short times is determined by the second partial derivatives of the amplitude φ. In the case of large mass, it seems reasonable to expect that an $m\to\infty$ asymptotic approximation, simpler than (4.2), could be obtained by truncating the series expansion (2.9) for S at an appropriate order of m^{-1}. A similar large mass asymptotic expansion for the propagator [the integral kernel of $U(t,s)$] has recently been derived in Ref. 27.

The result of Theorem 2 may be trivially extended to estimating the error in the expectation value of any bounded observable $A\in\mathscr{B}(\mathscr{H})$, in the state Υ. A simple calculation [assuming $\|\Psi_s\| = 1$, and using the unitarity of $U(t,s)$] results in

$$\begin{aligned}&\left|\langle\Psi_t, A\Psi_t\rangle - \frac{\langle\Upsilon(\,\cdot\,,t), A\Upsilon(\,\cdot\,,t)\rangle}{\|\Upsilon(\,\cdot\,,t)\|^2}\right| \\ &\quad\leqslant\|A\|\cdot\|\Psi_t - \Upsilon(\,\cdot\,,t)\|\cdot\text{ const}\,,\end{aligned}$$

for small enough $|t-s|\hbar/m$, and with the constant of order 10.

V. CONCLUDING REMARKS

In order to evaluate the merits of the above result, it is useful to compare it with other available results in the literature, for similar problems.

(1) Maslov and Fedoriuk[1] have developed a very general rigorous theory of the WKB approximation applicable to a broad class of wave equations. The equations they consider are in general defined in terms of pseudodifferential operators, and might well be said to include the majority of cases arising in physics. In its most general form, of course, their theory is technically rather difficult.

It is of interest to compare with the results of their theory when it is specialized to the case of quantum time evolution. Such a result is provided by Theorem 12.3 of Ref. 1. The problem considered there is in one way more general in that the phase function $k\cdot(x-y)$ in the initial data (1.3) is replaced by a more general function $S_0\in C^\infty(\mathbb{R}^d_x)$. Their Theorem 12.3 also obtains the higher-order asymptotics in powers of $\hbar$.

On the other hand, this is compensated by several rather restrictive assumptions. First of all, the potential is assumed to be time independent (this is not a limitation in their method), and a function of rapid decrease, $v\in\mathscr{S}(\mathbb{R}^d;\mathbb{R})$ (the Schwartz space). The consequent decay of $v(x)$ as $|x|\to\infty$ excludes the use of their result for many-body problems, where typically the total potential v will not decay for large $|x|$.

Second, Maslov and Fedoriuk assume the initial amplitude function $\varphi(x)$ in (1.3) has compact support. This assumption, ubiquitous in their general theory, is needed because of the reliance of their method on an analysis of classical trajectories. It allows them to establish the existence of a local diffeomorphism $y(t,x)$, Eq. (1.9). Finally, it is worth mentioning that Theorem 12.3 requires $\hbar$ and

$|t-s|$ to be "sufficiently small," with no estimate given on the size of the allowed regions of these parameters. (In a pioneering paper, Maslov[2] gives an estimate for the allowed time region, which is dimensionally incorrect, even after the mass is restored from the value $m=1$ he used.)

In contrast to these assumptions, note that for the present results, it was neither assumed that v decay [for example, class $\mathscr{V}$ contains potentials of the form $\sin(\alpha\cdot x)$], nor that $\varphi\in C_c^\infty$, nor that $\hbar$ be "small enough." Furthermore the size of the allowed time region is known, as is the constant C_φ in the error estimate (4.10).

(2) In a series of interesting papers,[28-30] Hagedorn has rigorously studied the semiclassical limits $\hbar\to 0$, and $m\to\infty$, of quantum evolution, using a particular kind of coherent states. The quantum system he considered is similar to the one here, except that $v(x)$ was assumed time independent, only to be C^2, and was allowed to have a quadratic growth in x. (Of course from a physical standpoint such behavior for large $|x|$ is rather artificial.) The class of initial states Ψ_s that he employed were roughly of the form (1.3), but with $\varphi(x)$ taken to be an *$\hbar$-dependent* Gaussian. Notably, this Gaussian amplitude becomes very sharply peaked as $\hbar\to 0$.

The method used by Hagedorn to obtain an approximate wave function at time t is not the WKB method, but might better be described as a "wave packet method."[31] In this method the parameters appearing in the Gaussian wave packet are evolved via a set of ordinary differential equations. This has the undesirable effect of forcing the wave packet to keep the same "shape" as it evolves, i.e., it maintains the same form as a function of x—in this instance a Gaussian form. On the other hand, from a practical standpoint the wave packet approach is easier to implement, because it only involves integrating a system of ordinary differential equations, rather than computing a classical action $S(x,t;k,s)$ by some means.

Let us briefly consider the theoretical implications now. With his approach, Hagedorn[28] obtained a lowest-order error that only scales like $\hbar^{1/2}$, in contrast to $\hbar^1$ in (4.10). Apparently this is because the shape of the approximate wave function was rather constrained, since in a later paper,[30] more general approximate wave functions were shown to exist that would yield errors of order $\hbar^{l/2}$, for a fixed choice of $l\in\mathbb{N}$. It was also shown that for more general $\hbar$-dependent initial amplitudes $\varphi\in\mathscr{S}$, an approximate wave function with error $\mathscr{O}(\hbar^{1/2})$ existed. It is important to note that these initial amplitudes also become highly localized as $\hbar\to 0$.

In these results there is again no explicit estimate given of the constant C in the error estimate, nor for the allowed time region of validity. Theorem 1.1 of Ref. 28 appears at first sight to be valid for arbitrarily large times. However, this can only be the case if $\hbar$ is assumed so small that the initial state becomes so peaked that it will not spread significantly (relative to distances over which the potential oscillates) during the chosen long time interval. (Theorem 1.1 in Ref. 30 is formulated in a way which avoids giving this impression.)

A recent extension of Hagedorn's methods by Robinson[32] has resulted in an $\mathscr{O}(\hbar^{1/2-\epsilon})$, $\epsilon>0$, WKB approximation valid for longer times (away from caustics). In this result, C is not known in detail, and it is required that $\hbar$ be "small enough" and that the initial amplitude have compact support.

(3) Recently, Klauder has been promoting[33] a so-called "global, uniform, semiclassical" approximation to wave-equation solutions. In the case of quantum mechanics, this approach basically reduces to a study of the evolutuion of Gaussian coherent states having the initial form

$$\psi_s(x)=(\Omega/\pi\hbar)^{d/4}\exp[-(\Omega/2\hbar)(x-y)^2]\times\exp[(i/\hbar)k\cdot(x-y)]\,, \tag{5.1}$$

where $\Omega>0$ is a parameter, in addition to $k,y\in\mathbb{R}^d$. These states, like Hagedorn's, are strongly peaked about $x=y$ for small $\hbar$. Klauder's approach is to apply a standard WKB approximation to describe the evolution of (5.1). However, because of the $\hbar$-dependent amplitude in (5.1), the initial condition for the relevant action in this approximation is seen to be complex: $k\cdot(x-y)+i(\Omega/2)(x-y)^2$. The method of characteristics then implies a need for complex-valued trajectories.

Unfortunately, this leads to problems with this approach. One problem stems from the fact that for a general real-analytic Hamiltonian H_c: $\mathbb{C}^{2d}\times\mathbb{R}\to\mathbb{C}$, the trajectories may fail to be defined for all $t\in\mathbb{R}$. This can occur even when the real-valued trajectories exist for all t. Hence with appropriate initial conditions a trajectory may "reach infinity" in an arbitrarily short time. As a result, it may not be possible to define the required action function for all the values of x,k,y that are necessary, for any fixed t,s. It is not appropriate to delve further into these matters here, except to remark that it seems many of the suggestive claims important to Klauder's theory have not been proved.

For the reasons discussed in (2) and (3), it was decided to use an initial amplitude function φ in (1.3) that (as in Maslov's study) is $\hbar$ independent, and is further as general as possible. It is hoped that the WKB derivation presented here for this system is found to be relatively simple, self-contained, and detailed in its result.

A few closing remarks about possible extensions of the present work may be of interest. First of all, there is some truth in the statement that the use of the constructive series solution for S, with coefficient functions determined by a sum over tree graphs, is a kind of "trick" valid only for Hamiltonians similar to (2.14). However, the following observation indicates a possible means for generalization. Series (2.9) can be viewed as a perturbative series, in "powers" of the potential v, about the action associated with the free Hamiltonian $p^2/2m$. This suggests the modifications necessary for a more general Hamiltonian, decomposable into a solvable part plus a perturbation. The vector component of linear path $w(\xi)$ should be replaced by an unperturbed trajectory, and the resulting analog of (2.13) should become a recursive definition of the action's expansion coefficients. Their analytical properties must then be obtained by induction from these recurrence relations. In general, a closed form expression like (2.5) will not be available.

The result of Theorem 2 just gives the lowest-order $\hbar\to 0$ asymptotics. It is evident, however, that in obtaining this

result, all the ingredients necessary for constructing higher-order corrections, by methods similar to those presented in Ref. 1, have been constructed. The key item in these higher corrections is the diffeomorphism f of (4.1). Only the lowest-order asymptotics were derived here because this is by far the most interesting result, and it avoids unnecessary complications in the presentation.

Finally, let us notice where the classical trajectories are lurking. As was mentioned in a remark following Lemma 5, function $S(x,t;k,s)$ is a complete integral of the Hamilton–Jacobi equation, with independent parameters $k\in\mathbb{R}^d$. Thus S induces classical trajectories via Jacobi's theorem. They may be shown to have initial momentum $p(s)=k$, and final configuration $q(t)=x$ [cf. (1.11)]. Such an analysis is similar to the one found in Sec. IV of Ref. 17 for trajectories having two fixed end points.

ACKNOWLEDGMENTS

I would like to thank Professor Robert Littlejohn's group at Berkeley for their hospitality, and also Professor Tom Osborn (Manitoba) for encouragement.

This work was supported by a NATO postdoctoral fellowship awarded by the Natural Sciences and Engineering Research Council of Canada.

APPENDIX: ARBOREAL COMBINATORICS

In this Appendix, Eqs. (2.13) will be proved. This proof is in many respects similar to the corresponding proof that arises in the study of Hamilton's principal function.[17] However, in the present proof there are several computational differences, and a significant simplification, relative to the analogous proof presented in Ref. 17. Moreover, (2.13) is of fundamental importance in the construction of the solution S to (HJ), so its derivation will be presented with an emphasis on the new features.

Let us begin with a few additional notations for trees and their combinatorics. If $n\in\mathbb{N}$ and $l\in\bar{n}$, let J_l be any of the $\binom{n+1}{l}$ distinct l-element subsets of $\overline{n+1}$, and denote its complement by $J_l^c=\overline{n+1}\setminus J_l$. The sum over all such $J_l\subset\overline{n+1}$ for a given l is denoted Σ_{J_l}.

Suppose $T\in\mathscr{T}V$ and $\beta\in ET$. If β is deleted from ET, then T is broken up into two disjoint subtrees T_1^β, T_2^β satisfying

$$V=VT_1^\beta\cup VT_2^\beta,\quad ET=ET_1^\beta\cup ET_2^\beta\cup\{\beta\},$$
$$i_\beta\in VT_1^\beta,\quad j_\beta\in VT_2^\beta.$$

Finally, let $\mathbb{F}$ denote the set of all tree graphs whose vertex sets are finite subsets of $\mathbb{N}$. The following combinatoric identity is the key to proving (2.13).

Lemma 9 (tree grafting): Let $n\in\mathbb{N}$ and suppose $f:\mathbb{F}^2\times\mathbb{N}^2\to\mathbb{R}$ is a function of an unordered pair of trees and an unordered pair of vertex labels, i.e.,

$$f(T,T';r,q)=f(T',T;r,q)=f(T,T';q,r).$$

Then

$$\frac{1}{2}\sum_{l=1}^{n}\sum_{J_l}\sum_{T\in\mathscr{T}J_l}\sum_{T'\in\mathscr{T}J_l^c}\sum_{r\in J_l}\sum_{q\in J_l^c}f(T,T';r,q)$$
$$=\sum_{T\in\mathscr{T}\overline{n+1}}\sum_{\beta\in ET}f(T_1^\beta,T_2^\beta;i_\beta,j_\beta)\,. \tag{A1}$$

Proof: See Lemma 10 of Ref. 17. ∎

Turning now to the proof of (2.13), one sees (2.13a) is just the definition of a_1. Next, replace n by $n+1$ throughout (2.13b), where now $n\in\mathbb{N}$. Let W_n be the resulting integral on the right-hand side,

$$W_n\equiv-\frac{1}{2}\sum_{l=1}^{n}\binom{n+1}{l}\int_I d\lambda\,\lambda^{2n}$$
$$\times(\nabla_1 a_l\cdot\nabla_1 a_{n+1-l})(w(\lambda);k,s)\,. \tag{A2}$$

Taking a_l, a_{n+1-l} to be defined by the tree sums (2.5), it must then be shown that W_n is the tree sum a_{n+1}. Using (2.5), the gradients in (A2) may be computed as follows:

$$\nabla_1 a_l(x,t;k,s)=\int_{I^l}d^l\xi\sum_{T\in\mathscr{T}\bar{l}}\left[\prod_{\beta\in ET}b_\beta\right]$$
$$\times\sum_{r=1}^{l}\nabla_r V_l\left(\overset{l}{\underset{p=1}{\times}}w(\xi_p)\right).$$

A similar expression can be written for $\nabla_1 a_{n+1-l}(x,t;k,s)$, but using integration variables $\xi_{l+1},\ldots,\xi_{n+1}$. When these expressions are substituted in (A2) they must be evaluated at $(x,t)=w(\lambda)$. By use of the composition law $w(\xi_p)|_{(x,t)=w(\lambda)}=w(\xi_p\lambda)$ [see the equation after (2.16)], this results in

$$W_n=-\frac{1}{2}\sum_{l=1}^{n}\binom{n+1}{l}\int_I d\lambda\,\lambda^{2n}\int_{I^{n+1}}d^{n+1}\xi$$
$$\times\sum_{T\in\mathscr{T}\bar{l}}\sum_{T'\in\mathscr{T}\overline{n+1}\setminus\bar{l}}\prod_{\beta\in ET}b_\beta\prod_{\beta'\in ET'}b_{\beta'}$$
$$\times\sum_{r\in VT}\sum_{q\in VT'}\nabla_r\cdot\nabla_q V_{n+1}\left(\overset{n+1}{\underset{p=1}{\times}}w(\xi_p\lambda)\right). \tag{A3}$$

The factor $\binom{n+1}{l}$ may be replaced by a sum over all l-element vertex sets $J_l\subset\overline{n+1}$, which play the role of $\bar{l}$ in (A3). Following this by a change of integration variables $\xi_p\to\xi_p/\lambda$ yields

$$W_n=-\int_{I^{n+1}}d^{n+1}\xi\int_M d\lambda\,\lambda^{n-1}$$
$$\times\left\{\frac{1}{2}\sum_{l=1}^{n}\sum_{J_l}\sum_{T\in\mathscr{T}J_l}\sum_{T'\in\mathscr{T}J_l^c}\left[\prod_{\beta\in ET}\left(\frac{\xi_{\beta}^{>}}{\lambda}-1\right)\nabla_{i_\beta}\cdot\nabla_{j_\beta}\right]\right.$$
$$\times\left[\prod_{\beta'\in ET'}\left(\frac{\xi_{\beta'}^{>}}{\lambda}-1\right)\nabla_{i_{\beta'}}\cdot\nabla_{j_{\beta'}}\right]$$
$$\left.\times\sum_{r\in J_l}\sum_{q\in J_l^c}\nabla_r\cdot\nabla_q V_{n+1}\left(\overset{n+1}{\underset{p=1}{\times}}w(\xi_p)\right)\right\}. \tag{A4}$$

In (A4), I use the notation $M\equiv\max\{\xi_1,\ldots,\xi_{n+1}\}$, and $\xi_\beta^{>}\equiv\max\{\xi_{i_\beta},\xi_{j_\beta}\}$ arises from the Green's function g, in b_β. The $\{\cdots\}$ in (A4) may now be simplified by use of the tree grafting identity (A1), which after a few manipulations results in

$$W_n = \int_{I^{n+1}} d^{n+1}\xi \sum_{T\in\mathscr{T}\,\overline{n+1}} \left(\prod_{\alpha\in ET} \nabla_{i_\alpha}\cdot\nabla_{j_\alpha}\right) \times V_{n+1}\left(\bigtimes_{p=1}^{n+1} w(\xi_p)\right) \times\left[(-)^n \sum_{\beta\in ET}\int_M^1 d\lambda \prod_{\gamma\in ET\setminus\{\beta\}} (\lambda-\xi_\gamma^>)\right]. \quad \text{(A5)}$$

It remains to show that the $[\cdots]$ in (A5) equals $\Pi_{\gamma\in ET}(\xi_\gamma^> - 1)$, since this is the product of $g(\xi_{i_\gamma},\xi_{j_\gamma})$ over the edge set ET. Evaluating the λ integral one finds

$$[\cdots] = \sum_{\mu=0}^{n-1} (-)^{\mu+n} \frac{1-M^{n-\mu}}{n-\mu} \times\left(\sum_{\beta\in ET}\sum_{A\subset ET\setminus\{\beta\}}^{(\mu)} \xi_{A(1)}^> \cdots \xi_{A(\mu)}^>\right). \quad \text{(A6)}$$

The sum $\Sigma_A^{(\mu)}$ requires explanation. First of all, when $\mu = 0$, $\Sigma^{(0)} = 1$. Otherwise $A\equiv\{A(1),\ldots,A(\mu)\}\subset ET\setminus\{\beta\}$ denotes a set of μ distinct edges in $ET\setminus\{\beta\}$, and $\Sigma^{(\mu)}$ specifies a sum over all possible distinct sets A of this type. Performing the summation in (A6) yields

$$[\cdots] = \sum_{\mu=0}^{n-1} (-)^{\mu+n}(1-M^{n-\mu}) \times \sum_{A\subset ET}^{(\mu)} \xi_{A(1)}^> \cdots \xi_{A(\mu)}^> .$$

The upper limit of the μ summation may now be replaced by n, because the summand vanishes for $\mu = n$. Doing this and recalling that $|ET| = n$ for $T\in\mathscr{T}\,\overline{n+1}$ gives

$$[\cdots] = \prod_{\gamma\in ET}(\xi_\gamma^> - 1) - \prod_{\gamma\in ET}(\xi_\gamma^> - M). \quad \text{(A7)}$$

The second product in (A7) vanishes: for suppose $M = \xi_l$, $l\in\overline{n+1}$; since T is connected, $l\in\gamma$ for some $\gamma\in ET$, and $\xi_l = \xi_\gamma^>$ because $\xi_l = M\geqslant\xi_i$ for all i. This proves the result desired in (A5), which then becomes $W_n = a_{n+1}$ as was to be shown. ■

[1] V. P. Maslov and M. V. Fedoriuk, *Semiclassical Approximation in Quantum Mechanics* (Reidel, Dordrecht, 1981).
[2] V. P. Maslov, USSR Comput. Math. Math. Phys. **1**, 123 (1962).
[3] M. V. Berry and K. E. Mount, Rep. Prog. Phys. **35**, 315 (1972).
[4] G. D. Birkhoff, Bull. Am. Math. Soc. **39**, 681 (1933); Proc. Natl. Acad. Sci. USA **19**, 339 (1933).
[5] D. Bollé and D. Roekaerts, Phys. Rev. A **30**, 2024 (1984).
[6] Ph. Choquard, Helv. Phys. Acta **28**, 89 (1955).
[7] C. DeWitt-Morette, Ann. Phys. (NY) **97**, 367 (1976).
[8] D. Elworthy, A. Truman, and K. Watling, J. Math. Phys. **26**, 984 (1985); A Truman, J. Math. Phys. **18**, 2308 (1977).
[9] D. Fujiwara, J. Anal. Math. **35**, 41 (1979); Duke Math. J. **47**, 559 (1980).
[10] M. C. Gutzwiller, J. Math. Phys. **8**, 1979 (1967); **10**, 1004 (1969); **11**, 1791 (1970); **12**, 343 (1971).
[11] J. B. Keller, Ann. Phys. (NY) **4**, 180 (1958).
[12] H. J. Korsch and R. Möhlenkamp, Phys. Lett. A **67**, 110 (1978).
[13] M. M. Mizrahi, J. Math. Phys. **22**, 102 (1981).
[14] T. A. Osborn and F. H . Molzahn, Phys. Rev. A **34**, 1696 (1986).
[15] L. S. Schulman, *Techniques and Applications of Path Integration* (Wiley, New York, 1981).
[16] A. Voros, Ann. Inst. H. Poincaré **24**, 31 (1976).
[17] F. H. Molzahn and T. A. Osborn, J. Math. Phys. **27**, 88 (1986).
[18] F. H. Molzahn and T. A. Osborn, J. Phys. A: Math. Gen. **20**, 3073 (1987).
[19] R. J. Wilson, *Introduction to Graph Theory* (Academic, New York, 1975).
[20] L. H. Loomis and S. Sternberg, *Advanced Calculus* (Addison–Wesley, Reading, MA, 1968).
[21] B. C. Carlson, *Special Functions of Applied Mathematics* (Academic, New York, 1977).
[22] T. Kato, *Perturbation Theory for Linear Operators* (Springer, Berlin, 1984), 2nd ed.
[23] T. Kato, Numer. Math. **2**, 22 (1960).
[24] I. M. Gelfand and M. V. Fomin, *Calculus of Variations* (Prentice–Hall, Englewood Cliffs, NJ, 1963).
[25] S. G. Krein, *Linear Differential Equations in Banach Space* (Am. Math. Soc., Providence, RI, 1971).
[26] W. O. Amrein, J. M. Jauch, and K. B. Sinha, *Scattering Theory in Quantum Mechanics* (Benjamin, Reading, MA, 1977).
[27] L. Papiez, T. A. Osborn, and F. H. Molzahn, J. Math. Phys. **29**, 642 (1988).
[28] G. A. Hagedorn, Commun. Math. Phys. **71**, 77 (1980).
[29] G. A. Hagedorn, in *Trends in Applications of Pure Mathematics to Mechanics*, edited by R. T. Knops (Pitman, London, 1981), Vol. III, p. 111.
[30] G. A. Hagedorn, Ann. Phys. (NY) **135**, 58 (1981).
[31] R. G. Littlejohn, Phys. Rep. **138**, 193 (1986).
[32] S. L. Robinson, J. Math. Phys. **29**, 412 (1988).
[33] J. R. Klauder, "Global, uniform, asymptotic wave-equation solutions for large wave numbers," AT&T Bell Laboratory preprint, 1987; Phys. Rev. Lett. **56**, 897 (1986).

The brachistochrone in almost flat space

S. G. Kamath
Department of Mathematics, Indian Institute of Technology, Madras 600 036, India

(Received 29 January 1988; accepted for publication 25 May 1988)

This paper is an extension, within the framework of general relativity, of the relativistic brachistochrone discussed recently by Goldstein and Bender [J. Math. Phys. **27**, 507 (1986)]. Assuming that the gravitational field due to a spherically symmetric source with mass M at equilibrium is weak, it is found that the brachistochrone, for a falling particle of mass m, described by $\Theta(r)$, with Θ an angle and r a distance measured from the center of symmetry, is in general a hyperelliptic integral. The latter integral can in one case be calculated exactly in terms of the normal elliptic integrals of the first and third kinds and the elementary transcendental functions. It is shown via a numerical computation using the sun's gravitational field as a reference that one can recast this exact version into a simple form, viz., $\sqrt{r}\Theta = a$, where a is a constant.

I. INTRODUCTION

In a recent paper Goldstein and Bender[1] have presented a relativistic generalization of the classic brachistochrone problem for a particle falling from rest in a uniform gravitational field. As is well known, the brachistochrone is the trajectory joining an initial position A to a final position B along which the time of transit of the falling particle is a minimum. While the classical nonrelativistic trajectory is known (with A as the origin of the coordinate system) to be a cycloid of the form $x = b(1 - \cos t), y = b(t - \sin t)$, with the parameter b determined by the end point B, Goldstein and Bender (GB) show that in the relativistic case, this is just one of three possible curves. In particular, the two new solutions are very different from the nonrelativistic case in that, for both of them, $y(x)$ increases without bound as x increases. It should be noted here that the motion of the particle of rest mass m in Ref. 1 is still assumed to take place in a uniform gravitational field with the force law $\mathbf{F} = \tilde{m}\mathbf{g}$, $\mathbf{g}$ being a constant, and $\tilde{m}(1 - v^2/c^2)^{1/2} = m$.

In this paper we have relaxed the above-mentioned restriction to a uniform gravitational field. Specifically, we consider the motion of the particle in a weak gravitational field in the sense of general relativity. As is well known in the literature[2,3] such a weak field will be represented by a metric $g_{\alpha\beta}$ that differs very little from the Minkowski metric $\eta_{\alpha\beta}$. Thus with $\eta_{\alpha\beta} = \eta^{\alpha\beta} = \mathrm{diag}(1, -1, -1, -1)$, the gravitational field is said to be weak when $|g_{\alpha\beta} - \eta_{\alpha\beta}| \ll 1$. More precisely, with the assumption that $g_{\alpha\beta}$ can be expanded as an infinite series

$$g_{\alpha\beta} = \eta_{\alpha\beta} + g^{(1)}_{\alpha\beta} + g^{(2)}_{\alpha\beta} + \cdots , \tag{1}$$

we limit ourselves in the first (linear) approximation, writing $g_{\alpha\beta} = \eta_{\alpha\beta} + g^{(1)}_{\alpha\beta}$ instead of Eq. (1). As described in detail in Refs. 2 and 3, the linearized Einstein field equations can then be solved for $g_{\alpha\beta}$ once the source of the gravitational field, viz., the energy momentum tensor $T_{\alpha\beta}$, is given.

In this paper, we consider a distribution of matter at equilibrium described by $T_{00} = \rho c^2$, $T_{\alpha\beta} = 0$ for $\alpha\beta \neq 00$, with the density ρ being time independent and spherically symmetric, so that $\rho = \rho(r)$, r being the distance from the center of symmetry. The resulting $g_{\alpha\beta}$ is then represented[2,3] in the interval

$$ds^2 = (1 - 2\Phi/c^2)c^2\,dt^2 - (1 + 2\Phi/c^2)(dx^2 + dy^2 + dz^2)\,. \tag{2}$$

Here $\Phi = GM/r$, outside the material distribution, with $M = \int \rho(r) d^3x$ and G the familiar gravitational constant.

Since the problem is now posed within the framework of general relativity, there will be two important qualitative differences from the work of Ref. 1 which we shall merely mention below, relegating the details to subsequent sections in the paper.

First, the motion of a material particle of rest mass m in the gravitational field whose metric is represented in Eq. (2) will now be governed via Hamilton's principle by the Lagrangian $L = -mc\,ds/dt$. This will replace the flat-space Lagrangian discussed in GB and given there by $L = -mc^2\gamma^{-1} + \mathscr{E}$, with $\gamma^{-2} = 1 - v^2/c^2$, $\mathscr{E} = mc^2 \times (\exp(gx/c^2) - 1)$, and $\mathbf{v}^2 = (d\mathbf{r}/dt)^2$. Second, since we are interested in the path of minimum time rather than the geodesic (which is the path of extremal action), there will be one more point of departure from Ref. 1; namely, the element of spatial distance will now be given, following Eq. (2), by

$$dl = \{(1 + 2\Phi/c^2)(dx^2 + dy^2 + dz^2)\}^{1/2}\,,$$

instead of the familiar $dl = (dx^2 + dy^2 + dz^2)^{1/2}$ as in Ref. 1. Thus the line integral representation for the time of fall will be given by

$$T = \int_A^B \frac{dl}{v}\,, \tag{3}$$

with dl as given above, v being the particle velocity. In Eq. (3), we shall regard A as lying outside the range of the weak gravitational field, namely, at infinity, so that Eq. (3) can be rewritten

$$T = \int_\infty^B \frac{dl}{v}\,, \tag{3'}$$

with the form of v being obtained from the principle of conservation of energy. Because of the assumption of spherical

 0022-2488/88/102268-05$02.50

symmetry, so that $\Phi = \Phi(r)$ only in (2), the Lagrangian L when written out in spherical polar coordinates should be cyclic in one of the angles, so that the motion of the particle can be taken, as in case of central force motion, to lie in a plane with its location specified by the coordinates (r,Θ), r being measured as mentioned earlier, from the center of the material distribution whose gravitational field is described by the metric $g_{\alpha\beta}$ represented in Eq. (2). The brachistochrone will then be obtained by solving the Euler equation associated with Eq. (3′) with the initial conditions $\Theta = 0$, $u(=1/r) = 0$. This program is carried through in the sections detailed below.

The plan of this paper is as follows. In Sec. II we obtain the expression for the velocity v appearing in (3′). Then by using this expression, we recover via the Euler equation in Sec. III the equation of the brachistochrone in the form usual to central force motion, viz., $\Theta = \Theta(u)$. This relation cannot in general be written out in a closed form. More precisely, it turns out, as we shall see in Sec. III, that one has to evaluate a hyperelliptic integral of the form

$$\Theta(u) = \int_0^u \frac{R(t)dt}{(P(t))^{1/2}},$$

with $P(t)$ being a fifth-degree polynomial in t. As is well known, for evaluating such hyperelliptic integrals[4] one usually has to resort to direct numerical integration, or to the use of complicated series expansions.

However, as we shall see in Sec. IV, it is possible to fudge a certain constant k (analogous to the constant k in Ref. 1) so as to obtain an exact dependence of $\Theta(u)$ on u; this is, unfortunately, possible only for one value of k. Still, this exact result given in Sec. IV is far too complicated, as it contains a linear combination of the normal elliptic integral of the first and third kinds, besides elementary functions like the natural logarithm whose arguments involve the Jacobian elliptic functions sn u, dn u. For our purposes it is useful to resort to an approximation whereby the complicated terms are rendered harmless by being extremely small, in fact, almost zero. The gravitational field outside the sun, which is regarded in the literature[3] as weak, turns out to be handy in this connection. As discussed in Sec. V it affords a more accessible version of the brachistochrone given in Sec. IV.

Finally, in Sec. VI we conclude with some comments on our results. Herein we offer, besides an exact analytic expression for the brachistochrone obtained by GB, a discussion of the more difficult problem, technically speaking, of the brachistochrone associated with the Schwarzschild metric. We hope to return to the latter in a subsequent publication.

II. ENERGY CONSIDERATIONS

Following the discussion in Sec. I, the Lagrangian for the material particle of rest mass m in the gravitational field is given by

$$L = -mc\frac{ds}{dt} = -mc^2\gamma^{-1}\left(1 - \frac{2\Phi}{c^2}\psi\right)^{1/2}, \tag{4}$$

with $\psi = \gamma^2(1 + v^2/c^2)$. It is easy to obtain the Hamiltonian from (4). It is given by

$$H = mc^2\gamma(1 - (2\Phi/c^2)\psi)^{-1/2}g_-, \tag{5}$$

with g_- defined below. As H in (5) is cyclic in time t by the principle of conservation of energy, we must also have

$$H = mc^2, \tag{6}$$

since the particle is assumed to fall from rest. It is easy to obtain an expression for v from Eqs. (5) and (6) through simple manipulations. We obtain

$$g_+v^2 = 2g_-\Phi, \tag{7}$$

where $g_\pm = (1 \pm (2/c^2)\Phi)$. With $\Phi = GM/r$, we see that g_-, for example, can be written as $g_- = (1 - uR)$ with $ru = 1$ and $c^2R = 2GM$, defining the Schwarzschild radius R for the material distribution of mass M. Since for macroscopic bodies for which the gravitational field can be regarded as weak (e.g., the sun) the Schwarzschild radius is very small compared to their actual radius, and since we are studying the motion of the material particle of rest mass m outside the material distribution causing the gravitational field, the case when $r = R$ can be safely ignored in (7).

III. THE EULER EQUATION

In terms of the polar coordinates (r,Θ) the element of spatial distance, following (2), is given by

$$dl = g_+^{1/2}(dr^2 + r^2\,d\Theta^2)^{1/2},$$

so that, using (7), Eq. (3′) can be rewritten

$$T = \int_\infty^B g_+(dr^2 + r^2\,d\Theta^2)^{1/2}D^{-1/2}$$
$$= \int_\infty^B g_+dr\left(1 + r^2\left(\frac{d\Theta}{dr}\right)^2\right)^{1/2}D^{-1/2}, \tag{8}$$

with $D = 2g_-\Phi$. The time taken T is a functional of the path $\Theta(r)$ so that T will be minimum when the Euler equation

$$\frac{d}{dr}\left(\frac{\partial I}{\partial\Theta'}\right) = 0, \quad \Theta' = \frac{d\Theta}{dr}, \tag{9}$$

is satisfied. Here I denotes the integrand in Eq. (8); thus one infers from (9) that

$$\frac{\partial I}{\partial\Theta'} = k, \tag{9′}$$

where k is a constant. Equation (9) is the analog of the Euler equation in GB. Substituting for I, it is easy to obtain

$$r^2g_+\Theta'(D(1 + r^2\Theta'^2))^{-1/2} = k. \tag{9″}$$

In terms of u, (9″) can be rewritten

$$g_+\Theta' = -k(D(1 + u^2\Theta'^2))^{1/2}, \tag{9‴}$$

with Θ' being the first derivative of Θ with respect to u. Note the negative sign in (9‴). Since the particle of mass m is assumed to fall from rest at infinity (where $u = 0$) to the point B where $u > 0$, it follows that, for $k > 0$ (this choice can be made without loss of generality), the solution of (9‴), namely, $\Theta(u)$, will be a decreasing function of u.

Equation (9‴) immediately leads to the solution [with $\Theta(0) = 0$]

$$\Theta(u) = \int_0^u dt(g_+^2 - Du^2k^2)^{-1/2}(Dk^2)^{1/2}, \tag{10}$$

which is the equation of the brachistochrone. Substituting for $g_\pm$ and Φ one can rework (10) as

$$\Theta(u) = \int_0^u dt\, t(1-tR)(P(t))^{-1/2}, \tag{10'}$$

with $P(t) = t(1-tR)((1+tR)^2 - \alpha^2 t^3(1-tR))$ and $\alpha^2 = k^2 c^2 R$. Note that for the integral to be real we require that (a) $uR \leqslant 1$ and (b) $(1+uR)^2/(1-uR)u^3 \geqslant \alpha^2$. The former requirement is easily met since the particle of mass m as mentioned in Sec. II is moving in the gravitational field outside a macroscopic object whose Schwarzschild radius R is very small compared to its actual radius. The latter bound, however, is an upper limit on α^2 and hence on k^2 for fixed R.

The integral in the above equation is a hyperelliptic integral, which is further manipulated as follows. Using a change of variable $s = tu$ we obtain, with $\sigma = uR$,

$$\Theta(u) = \alpha u^{3/2} \int_0^1 ds\, s(1-s\sigma)(P(s))^{-1/2}.$$

Although $P(s)$ is a sixth-order polynomial in s, it can be rewritten as a fifth-order polynomial through the use of $z = 1/s$, so that

$$\Theta(u) = \beta^{1/2} \int_1^\infty dz \times \frac{z-\sigma}{z[(z-\sigma)(z^2(z+\sigma)^2 - \beta(z-\sigma))]^{1/2}}, \tag{11}$$

where $\beta = \alpha^2 u^3$, and we have now displayed the polynomial $P(z)$ in full in (11). In the following two sections we shall study the dependence of Θ on u.

IV. AN EXACT RESULT

We shall show herein that the polynomial $P(z) = z^2(z+\sigma)^2 - \beta(z-\sigma)$ has a double root at various values of β. The utility of this double root lies in that the hyperelliptic integral in (11) can be evaluated exactly without recourse to approximation. Naturally, for general β, Eq. (11) will represent the brachistochrone associated with a weak gravitational field. From $P(z)$ we obtain

$$\frac{dP}{dz} = 4z^3 + 6\sigma z^2 + 2z\sigma^2 - \beta.$$

Let $z = z_0$ denote the double root of $P(z)$. Then $P(z_0)$ and

$$\left.\frac{dP}{dz}\right|_{z=z_0}$$

should be zero. This yields

$$\beta = 4z_0^3 + 6\sigma z_0^2 + 2z_0\sigma^2. \tag{12}$$

Clearly β is zero when $z_0 = 0$ and $z_0 = -\sigma$. Ignoring these cases, since $\beta = 0$ implies $k = 0$, we find the remaining solutions for z_0 using (12) and $P(z_0) = 0$. Thus

$$P(z_0) = -3z_0^4 + 5\sigma^2 z_0^2 + 2z_0\sigma^3 \equiv z_0 \tilde{P}(z_0). \tag{13}$$

To obtain values of z_0 other than 0 and $-\sigma$, we should solve $\tilde{P}(z_0) = 0$. This can be done by standard methods.[5] We obtain with the notation $c = \cos\mu/3$, $s = \sin\mu/3$ the following three real roots:

$$3z_0^{(1)} = 2\sigma c\sqrt{5}, \tag{14a}$$

$$3z_0^{(2)} = -\sigma(c + s\sqrt{3})\sqrt{5}, \tag{14b}$$

$$3z_0^{(3)} = -\sigma(c - s\sqrt{3})\sqrt{5}, \tag{14c}$$

with $9\tan\mu = 2\sqrt{11}$.

For each of these z_0 there is a corresponding β. However, as can be easily checked, only $z_0^{(1)}$ leads to a positive value of β, viz.,

$$3\beta^{(1)} = 8\sigma^3(1 + 5c^2 + 13c\sqrt{5}/6); \tag{15}$$

$z_0^{(2)}$ and $z_0^{(3)}$ lead to $\beta^{(2)} = O(10^{-6}\sigma^3)$ and $\beta^{(3)} < 0$, respectively. We are not reproducing these tedious though straightforward calculations here. These latter values of z_0 are therefore ignored, leading to a unique value for β. Since the double root for the polynomial $P(z)$ is now given by (14a), the remaining roots are easily found by standard methods.[5] They are complex conjugates of each other and denoted by

$$z_\pm = -p \pm i(2z_0^{(1)}p)^{1/2}, \tag{16a}$$

with $p = z_0^{(1)} + \sigma$. Thus

$$\Theta(u) = \beta^{1/2} \int_1^\infty \frac{dz(z-\sigma)}{z(z-z_0)[(z-\sigma)(z-z_+)(z-z_-)]^{1/2}}, \tag{16b}$$

where we have now dropped the superscript on $z_0^{(1)}$ and $\beta^{(1)}$. The integral in (16b) can be exactly evaluated in two steps with the help of Ref. 6. Thus with $Q(z)$ denoting the polynomial under the square root in (16b) we find for the integral

$$I_1 = \int_1^\infty dz(z-z_0)^{-1}(Q(z))^{-1/2}$$

the result

$$I_1 = g\eta^{-1} \int_0^{u_1} \frac{1-\operatorname{cn} u}{1+q\operatorname{cn} u}\, du \tag{17a}$$

$$= \frac{g}{\eta q}[-u_1 + (1-q)^{-1} \times (\pi(\phi, q^2/(q^2-1), j) - q f_1(u_1))], \tag{17b}$$

where

$$\begin{aligned} &g = A^{-1/2}, \quad A^2 = (b_1-\sigma)^2 + a_1^2, \\ &2j^2 A = A + b_1 - \sigma, \quad 2b_1 = z_+ + z_-, \\ &4a_1^2 = -(z_+ - z_-)^2, \quad \eta q = (A - \sigma + z_0), \\ &\eta = A + \sigma - z_0, \quad u_1 = \operatorname{cn}^{-1}(\cos\phi, j) = F(\phi, j), \\ &\phi = \operatorname{am} u_1 = \cos^{-1}\omega, \quad \omega(1-\sigma+A) = 1-\sigma-A. \end{aligned} \tag{18}$$

In (17) it is understood that $q^2 \neq 1$, and $F(\phi, j)$ and $\pi(\phi, q^2/(q^2-1), j)$ are the normal elliptic integrals[4] of the first and third kinds, respectively. Both are zero when $\phi = 0$. The last term in (17b) is obtained from[6]

$$f_1(u_1) = \begin{cases} h \arctan(\operatorname{sd} u_1/h), & \text{if } q^2/(q^2-1) < j^2, \\ \tilde{h} \ln\left(\dfrac{\operatorname{dn} u_1 + \tilde{h}\operatorname{sn} u_1}{\operatorname{dn} u_1 - \tilde{h}\operatorname{sn} u_1}\right), & \text{if } q^2/(q^2-1) > j^2, \end{cases} \tag{19a, 19b}$$

with $h^2(j^2 + (1-j^2)q^2) = 1 - q^2$, $\tilde{h}^2 = -h^2$.

In (17a) and (19), cn u, sn u, and dn u are the Jacobian elliptic functions, with sd u = sn u/dn u. It is of course known that cn 0 = 1 = dn 0 and sn 0 = 0; thus the lower limit in (17a) is not displayed in the various terms in (17b). There is one more integral denoted by I_2 below that we need in order to evaluate (16b), and it is given by

$$I_2 = \int_1^\infty dz\, z^{-1}(Q(z))^{-1/2}. \tag{20}$$

Clearly I_2 is just I_1 evaluated at $z_0 = 0$; it can be reached via (17b) *et seq.* through appropriate replacements. Thus for the value of β given by (15) we have the exact dependence of $\Theta(u)$ on u. It works out to

$$\Theta(u) = \beta^{1/2}(1 - \sigma/z_0)I_1 + \beta^{1/2}(\sigma/z_0)I_2. \tag{21}$$

V. NUMERICAL COMPUTATION

Equation (21) is clearly far too complicated to be of immediate physical interest. It would be ideal if it could be simplified considerably by application to a physical situation so that the equation for the brachistochrone would be more accessible. Happily, the gravitational field outside the sun, which is generally considered[3] as weak ($R/r = 10^{-5}$ at the surface of the sun), comes in handy in this connection. It needs to be pointed out here that it is our unfamiliarity with astrophysics that prevents us from extending the considerations of this section to other stellar objects whose gravitational fields are also deemed to be weak. Still, we believe that for such stellar objects the brachistochrone should not be vastly different from the expression derived here.

For the sun, the Schwarzschild radius R and the actual radius are about 3 km and 10^6 km, respectively (it is not necessary to use precise numbers here). Thus, as the particle of mass m falls from infinity in the sun's gravitational field, u increases from zero to about 10^{-6}, so that $uR \ll 1$ throughout the motion of the particle.

We now turn to a numerical estimation of the various entities in Eq. (18). With $9\tan\mu = 2\sqrt{11}$ we obtain $\mu = 36.39$ deg. Thus (15) leads to $\beta = 28.042\,264\sigma^3$; clearly, as $\sigma \ll 1, \beta < 1$. We shall choose a positive square root of β in our calculations below; clearly this corresponds to a $k > 0$. Inserting the value of μ into (14a) one can calculate z_0 in units of σ, and hence b_1, a_1, and A as defined by (18). We obtained $A^2 = 19.116\,844\sigma^2$, $\eta = 3.914\,854\sigma$, and $\eta q = 4.829\,708\sigma$. Thus q, and hence q^2, is obtained; we find that $q^2/(q^2-1) > j^2$ as $q^2 = 1.521\,986$ and $j^2 = 0.104\,62$. We therefore get the term in Eq. (21) denoted by Θ_1 as

$$\Theta_1 = 0.164\,576[-u_1 - 4.279\,218(\pi(\phi, 2.915\,76, 0.323\,45) - 1.233\,69 f_1(u_1))], \tag{22a}$$

where $f_1(u_1)$ will be reached via (19b) since $q^2/(q^2-1) > j^2$. It works out to

$$f_1(u_1) = c\ln\left(\frac{\mathrm{dn}\,u_1 + c\,\mathrm{sn}\,u_1}{\mathrm{dn}\,u_1 - c\,\mathrm{sn}\,u_1}\right), \tag{22b}$$

with $c = 0.596\,429$.

The second term in (21) denoted below by Θ_2 can also be written down in an analogous fashion. We find

$$\Theta_2 = 0.515\,278[-u_1 + 2.686\,14 \times(\pi(\phi, -0.650\,249, 0.323\,45) - 0.627\,719 f_1(u_1))], \tag{23a}$$

where $f_1(u_1)$ is now given by (19a) since $q^2/(q^2-1) < j^2$. To reassure the reader, we find, for I_2, $\eta = (A+1)\sigma$, $\eta q = (A-1)\sigma$, with A^2 quoted above and $q^2 = 0.394\,031$.

For $f_1(u_1)$ we thus obtain

$$f_1(u_1) = 1.150\,971 \arctan(0.868\,832\,\mathrm{sd}\,u_1). \tag{23b}$$

Equations (22) and (23) can now be examined in the light of our observation that $\sigma \ll 1$ throughout the motion of the particle. From (18) we note that, as $\sigma, A \ll 1$, $\omega \simeq 1$, or $\phi \simeq \cos^{-1}1$, or $\phi \simeq 0$. Thus as u increases from zero to 10^{-6}, ϕ remains almost constant at zero.

Let us now estimate Θ_1 as given by (22). The value of u_1 is reached via $u_1 = F(\phi, j)$; since ϕ is nearly zero, and[4]

$$\lim_{\phi\to 0}[F(\phi, j)/\sin\phi] = 1,$$

we infer that $u_1 \simeq \phi$. Again, as

$$\lim_{\phi\to 0}[\pi(\phi, \alpha^2, j)/\sin\phi] = 1,$$

we infer that $\pi(\phi, q^2/(q^2-1), j) \simeq \phi$. The elliptic functions work out as

$$\mathrm{sn}\,u_1 = (1 - \cos^2\phi)^{1/2} = 2\{A(1-\sigma)\}^{1/2}(1 - \sigma + A)^{-1} \simeq 2A^{1/2}$$

and

$$\mathrm{dn}\,u_1 = (1 - j^2\sin^2\phi)^{1/2} \simeq 1.$$

Thus the logarithm in (22b) becomes

$$c\log\left(\frac{1 + c\,\mathrm{sn}\,u_1}{1 - c\,\mathrm{sn}\,u_1}\right) \simeq 2c^2\,\mathrm{sn}\,u_1.$$

We thus obtain

$$\Theta_1 = -0.868\,832\phi + 0.618\,136\,\mathrm{sn}\,u_1. \tag{24a}$$

Similar remarks apply to Θ_2 given by (23). We get, with $b = 0.868\,831$,

$$\Theta_2 = b\phi - \arctan(b\,\mathrm{sn}\,u_1). \tag{24b}$$

Thus, adding (24a) and (24b), we get

$$\Theta = 0.618\,136\,\mathrm{sn}\,u_1 - \arctan(b\,\mathrm{sn}\,u_1). \tag{25}$$

We note that the term proportional to ϕ has now canceled almost exactly. Since the argument of the arctan function is very small, we now replace the last term by its argument and get

$$4\Theta = -\mathrm{sn}\,u_1 = -4.182\sigma^{1/2}. \tag{26}$$

Equation (26) leads (with $R = 3$ km) to

$$r^{1/2}\Theta = -1.816, \tag{27}$$

as the equation for the brachistochrone with reference to the gravitational field outside the sun. More importantly, we note that as u increases from zero, Θ decreases from zero and remains negative. This is just a reflection [as mentioned in connection with (9″)] of the decrease of $\Theta(u)$ with u. As the reader will have noticed, the considerations of this section depend on the smallness of the variable denoted by σ. Being infinitesimally different from zero, the doubly periodic property of the Jacobian elliptic functions $\mathrm{sn}\,u_1$, for example [see Eq. (26)], has not found any place in our calculation. Indeed, with $j^2 = 0.104\,62$, the standard definitions[5] of the symbols K, K' contained in the periods $(4K, 2iK')$ of $\mathrm{sn}(u, j)$ lead to $K \sim 1.612$ and $K' \sim 2.578$. Clearly the value of K is too large to be of interest as far as u_1 is concerned.

It is appropriate before concluding this paper to recall

for the reader's benefit the equation for the brachistochrone obtained by GB; this will be taken up in the following section.

VI. SOME COMMENTS

There are two parts to this section. First, we obtain an exact analytic form for the brachistochrone obtained by GB. This has been done in Ref. 1 for the case $k^2 = 1$ only; for $k^2 < 1$ and $k^2 > 1$ only a graphical plot of the brachistochrone has been given. Second, we comment (quite concisely, however) on the task of obtaining the brachistochrone for the full Schwarzschild metric.[2,3] Naturally this means that, while retaining therein the assumption of spherical symmetry, we are giving up the restriction to weak gravitational fields used in Secs. III–V.

To take up the GB calculation first, we consider the integral [Eq. (19) in Ref. 1]

$$y(x) = \int_0^x dt \left[\frac{k^2(1 - \exp(-2\alpha t))}{1 - k^2 + k^2 \exp(-2\alpha t)} \right]^{1/2}. \tag{28}$$

For $k^2 > 1$ this can be reworked as

$$y(x) = \xi \int_0^x dt \left[\frac{\exp(2\alpha t) - 1}{\xi^2 - \exp(2\alpha t)} \right]^{1/2}, \tag{28'}$$

with $\xi^2(k^2 - 1) = k^2$ and $c^2\alpha = g$. Using the substitution $\exp \alpha t = u$ we obtain from (28'), with $\psi = \exp \alpha x$,

$$\alpha y(x) = \xi \int_1^{\psi} u^{-1}\, du \left[\frac{u^2 - 1}{\xi^2 - u^2} \right]^{1/2},$$

which can now be easily evaluated in terms of elementary functions. We obtain

$$2\alpha y(x) = \xi \sin^{-1}(1 - w_1^2)^{1/2} + \sin^{-1}[-(1 - w_2^2)^{1/2}], \tag{29}$$

where $(\xi^2 - 1)w_1 = \xi^2 + 1 - 2\psi^2$, and $(\xi^2 - 1)w_2 = \xi^2 + 1 - 2\xi^2\psi^{-2}$. Note that in (28') we must have $\xi^2 \geqslant \psi^2$ for the integral to be real; in terms of k^2 this is rewritten $k^2 \leqslant \psi^2(\psi^2 - 1)^{-1}$. We recall here that Eq. (10') requires that $k^2 \leqslant (1 + Ru)^2/Rc^2u^3(1 - Ru)$ for the integral to be real. The counterpart of (29) when $k^2 < 1$ can also be obtained easily. We shall merely quote the result below:

$$2\alpha y(x) = \xi \ln R(x) + \sin^{-1}[-(1 - w_3^2)^{1/2}], \tag{30}$$

with

$$(\xi^2 + 1)R(x) = 2\psi^2 + \xi^2 - 1 + 2(\psi^4 + (\xi^2 - 1)\psi^2 - \xi^2)^{1/2},$$

and

$$(\xi^2 + 1)w_3 = \xi^2 - 1 - 2\xi^2\psi^{-2}.$$

As mentioned earlier, a graphical plot of (29) and (30) has been given in Ref. 1, but the numerical values given therein to α, defined by $c^2\alpha = g$, are in fact too large ($\alpha \sim 1$)—perhaps so large (in fact by several orders of magnitude) as to render invalid the Newtonian approximation for the gravitational field used in Ref. 1. It seems to us that had Goldstein and Bender used values of α that are consistent with the Newtonian approximation, typically $\alpha \sim 10^{-12}$ or even less, they would then have obtained, with the range of values for x used in their calculation, a plot for $y(x)$ that was quite insensitive to changes in x. In particular, the cycloidlike (periodic) behavior of the $k^2 > 1$ brachistochrone would not have been uncovered (see Figs. 7 and 8 in GB). We remind the reader here that we were also faced with a parallel situation of not finding a suitable role for the periodic behavior of the Jacobian elliptic functions $\mathrm{sn}(u, j)$ in Sec. V.

We shall now conclude this paper with a brief comment on the Schwarzschild brachistochrone. Our objective herein is merely to highlight for the reader the utility, technically speaking, of the weak-field approximation given by $g_{\alpha\beta} \simeq \eta_{\alpha\beta} + g^{(1)}_{\alpha\beta}$ that was used in the calculations in this paper. For the Schwarzschild metric,[2,3] the analog of Eq. (2) is given by

$$ds^2 = g_{00}c^2\, dt^2 + g_{ij}\, dx^i\, dx^j, \tag{31}$$

with $g_{00} = g_-$, $g_{ij} = -\delta_{ij} - \sigma(1 - \sigma)x_i x_j/r^2$.

Correspondingly, the Hamiltonian for the material particle of mass m, following Sec. II, will work out to

$$H = mc^2 g_- (g_- + g_{ij}\dot{x}^i\dot{x}^j)^{-1/2}.$$

However, because g_{ij} now has off-diagonal components, the counterpart of Eq. (7) for the velocity will have the form

$$v^2 = 2\Phi g_-^2 \left(g_- + \sigma\left(1 + u^2\left(\frac{d\Theta}{du}\right)^2\right)^{-1}\right)^{-1}, \tag{32}$$

thus making v^2 a function of the differential of Θ with respect to u; this feature is, however, absent in Eq. (7) and makes the problem more tractable there. But with v^2 as given by (32), one finds that the counterpart of (9") is now a fifth-order polynomial in $(d\Theta/du)^2$, thus making for a numerical, rather than exact, solution for $(d\Theta/du)^2$. The possibility of an exact solution for $(d\Theta/du)$ using the simple quadratic in (9") is thus an attractive feature of the weak gravitational field approximation.

ACKNOWLEDGMENTS

I thank my colleagues S. H. Kulkarni and D. S. Subramanyam for several useful discussions on the mathematical aspects of this paper.

[1] H. F. Goldstein and C. M. Bender, J. Math. Phys. **27**, 507 (1986): also referred to as GB in the text.

[2] M. Carmeli, *Classical Fields: General Relativity and Gauge Theory* (Wiley, New York, 1982), pp. 205–210.

[3] A. Papapetrou, *Lectures on General Relativity* (Reidel, Dordrecht, 1974), pp. 108–114.

[4] P. F. Byrd and M. D. Friedman, *Handbook of Elliptic Integrals for Engineers and Physicists* (Springer, Berlin, 1954), pp. 252ff.

[5] *Handbook of Mathematical Functions*, edited by M. Abramowitz and I. A. Stegun (Dover, New York, 1968).

[6] See Ref. 4, Eqs. 241.04 and 361.60.

Neutrino spectrum of Einstein universes

Christian Holm[a)]
School of Physics, Georgia Institute of Technology, Atlanta, Georgia 30322

(Received 17 November 1987; accepted for publication 15 June 1988)

The study of higher-dimensional Kaluza–Klein universes constructed from the unitary groups $\mathbf{U}_N$ is continued. They form Bergmann manifolds of dimension N^2 with Finslerian geometry induced by their hyperspin structure. In this paper Lagrangians for relativistic wave equations, which are generalizations of the Klein–Gordan, Dirac, and Weyl neutrino equations, are formulated. The wave equations are in general of differential order N. The hyperneutrino equation is examined in detail as the simplest example and its discrete symmetries are discussed. It is found that for $N=3$ and $N>4$ *TCP* and all its constituent symmetries are violated. The boson calculus is used to solve the linear neutrino equation exactly on $\mathbf{U}_N$ and the energy spectra of the neutrino and antineutrino are presented. It is found that the density ratio of negative to positive energy states is unity only for $N=2$, producing asymmetry for all higher-dimensional $\mathbf{U}_N$. The neutrino acquires a negligible rest mass of $O(10^{-31}\,\mathrm{eV})$ due to the global curvature of our manifold.

I. INTRODUCTION

In an earlier paper[1] we provided the unitary groups $\mathbf{U}(N,C) := \mathbf{U}_N$ with a hyperspin structure, turning them into N^2-dimensional Bergmann manifolds $\mathbf{B}_N$. A $\mathbf{B}_N$ with $N>2$ leads to a Kaluza–Klein space with Finsler geometry and an underlying $\mathbf{SL}_N$ tangent space group. The Bergmann manifolds are an alternative road to a Kaluza–Klein type of approach to unification.

A reason for considering $\mathbf{U}_N$ is its great symmetry, which makes calculations easier, and also the fact that we have unitary subgroups acting on the internal dimensions, which could account for the unitary gauge groups. We also found that all $\mathbf{U}_N$ with $N\geqslant 2$ turned out to be cosmological solutions to the hypergravity equations.[2]

In this paper we examine in detail the simplest generalized wave equation for $\mathbf{B}_N$, which is the linear hyperneutrino equation. We assume the reader to be familiar with the notation and content of Ref. 1.

Section II introduces the generalized relativistic wave equations of Ref. 3 and presents the Lagrangians they can be derived from. We also take a look at the discrete symmetries *T*, *C*, *P* and their products for the flat space hyperneutrino equation and show that it has a *TCP* invariance only for $N=2$ and $N=4$.

Section III discusses the hyperneutrino equation on $\mathbf{U}_N$ in detail. The solutions are irreducible representations of $\mathbf{U}_N$ that are obtained by means of the boson calculus. The solutions lead to an energy spectrum, which has a particle–antiparticle asymmetry for $N>2$.

Section IV concludes the work with a summary and discussion of the results.

II. WAVE EQUATIONS IN HYPERSPIN MANIFOLDS

A. Lagrangians

In Ref. 3 relativistic wave equations in $\mathbf{B}_N$ and their plane wave spectrum were introduced. We recapitulate the important points here and present in addition the Lagrangians they can be derived from. Relativistic invariance means for the N-ary theory an invariance under $\mathbf{SL}_N$, which contains the Lorentz group as a subgroup. Except for $N=2$ $\mathbf{SL}_N$ does not respect any quadratic form. The only invariant is the N-ic determinantal form $g_{\alpha\beta\cdots\nu}$, so that the scalar wave equation is of Nth differential order, although only second order in the external coordinates due to the determinantal constraint.

The equations under consideration here are the Klein–Gordon, the Dirac, and the Weyl neutrino equation, which were each generalized such that for $N=2$ the original ones emerge. For notational convenience we will write $\partial_{\dot\Sigma\Sigma}$ for $\sigma^\lambda{}_{\dot\Sigma\Sigma}\,\partial_\lambda$, where σ is the Hermitian spin vector, and define the dual ∂^D to the derivative operator via the metric as

$$(\partial^D)^\lambda := g^{\{\mu'\}\lambda}\partial_{\{\mu'\}}.$$

We make use of a collective index notation here, where $\{\mu\}$ stands for N antisymmetrized indices $\mu_1,\ldots,\mu_N$. The number of primes indicate the number of omitted indices. For details of this notation see Ref. 1.

The dual to ∂ is of differential order $N-1$. For what follows we omit the D for dual and understand it implicitly there whenever we write superscripts on ∂. We observe also one important property of $\partial_{\dot\Sigma\Sigma}$ and $\partial^{\Sigma\dot\Sigma}$, the following lemma.

Lemma:

$$\partial_{\dot\Sigma\Sigma}\partial^{\Sigma\dot\Sigma'} = g^{\{\mu\}}\,\partial_{\{\mu\}}\,(\delta^{\dot\Sigma'}{}_{\dot\Sigma}/N). \tag{2.1}$$

Proof: Write the left-hand side out with explicit spinor indices:

$$\partial_{\dot\Sigma\Sigma_1}\partial^{\Sigma_1\dot\Sigma'} = [(N-1)!]^{-1}\epsilon^{\Sigma_1\Sigma_2\cdots\Sigma_N}\epsilon^{\dot\Sigma'\dot\Sigma_2\cdots\dot\Sigma_N} \times\partial_{\dot\Sigma\Sigma_1}\partial_{\dot\Sigma_2\Sigma_2}\cdots\partial_{\dot\Sigma_N\Sigma_N}.$$

(1) For $\dot\Sigma'=\dot\Sigma$ the assertion is obviously true.

(2) If $\dot\Sigma'\neq\dot\Sigma$, then in any nonzero term we must have $\dot\Sigma'=\dot\Sigma_{N'}$ for some $N'\neq 1$. Interchange $\Sigma_{N'}$ with Σ_1. This gives a sign change from the ϵ, whereas the product of the ∂'s is symmetric. ■

Because we are dealing with higher derivative theories, we recall the relevant Euler–Lagrange equations.[4] For

[a)] Present address: Institut für Theoretische Physik, TU Clausthal, D-3392 Clausthal-Zellerfeld, West Germany.

 0022-2488/88/102273-07$02.50

$L(\varphi,\varphi^{(1)},\varphi^{(2)},\ldots,\varphi^{(N)})$, where $\varphi^{(1)}$ means $\partial_\mu\varphi(x^\lambda)$, etc., we obtain

$$\frac{\partial L}{\partial\varphi}-\partial_\mu\frac{\partial L}{\partial\varphi^{(1)}}+\partial_{(2)}\frac{\partial L}{\partial\varphi^{(2)}}\cdots+(-1)^N\partial_{(N)}\frac{\partial L}{\partial\varphi^{(N)}}=0.$$

We have assumed here a convention that the indices in terms like $g^{\mu\lambda}\varphi_\mu\varphi_\lambda$ have a definite order, say $\mu<\lambda$, for example. This ensures that no terms are counted more than once and we do not have to worry about extra correction factors.

The essentially unique scalar wave equation of least differential order has determinantal form. Equating it to a mass term M gives

$$N\det(\partial_{\dot\Sigma\Sigma})\varphi(u)=g^{\{\mu\}}\partial_{\{\mu\}}\varphi(u)=(-iM)^N\varphi(u), \tag{2.2}$$

where φ is a real or complex valued scalar field on the manifold and the Nth power of M is necessary on dimensional grounds. This is the natural generalization of the Klein–Gordon equation ($N=2$). The case of $M=0$ corresponds to the generalized d'Alembert equation. The Lagrangian which gives (2.2) is

$$L=(i)^N\bar\varphi(u)g^{\{\mu\}}\partial_{\{\mu\}}\varphi(u)-M^N\bar\varphi(u)\varphi(u)+\text{c.c.}, \tag{2.3}$$

where the bar stands for complex conjugation.

The generalized Weyl neutrino obeys the linear first-order equation

$$\partial_{\dot\Sigma\Sigma}\psi^\Sigma(u)=0. \tag{2.4}$$

Here ψ is an N-component spinor field depending on the time space variable u. The Lagrangian is

$$L=i\bar\psi^{\dot\Sigma}\partial_{\dot\Sigma\Sigma}\psi^\Sigma+\text{c.c.}$$

Equating the neutrino spinor to a multiple of a dual antispinor $\bar\mu_{\dot\Sigma}$ and closing the system with a second equation for the dual equation leads to the generalized Dirac equation

$$\partial_{\dot\Sigma\Sigma}\psi^\Sigma=-iM\bar\mu_{\dot\Sigma},\quad N\partial^{\Sigma\dot\Sigma}\bar\mu_{\dot\Sigma}=(-iM)^{N-1}\psi^\Sigma. \tag{2.5}$$

The system is chosen such that every component of the two spinors obeys the scalar wave equation (2.2). The proof is a straightforward application of (2.1). The Lagrangian is

$$L=+i\bar\psi^{\dot\Sigma}\partial_{\dot\Sigma\Sigma}\psi^\Sigma+[iN/(iM)^{N-2}]\mu_\Sigma\partial^{\Sigma\dot\Sigma}\bar\mu_{\dot\Sigma}-2M\psi^\Sigma\mu_\Sigma+\text{c.c.}$$

The two equations (2.5) are of differential order 1 and $N-1$, respectively. The system is symmetric only for $N=2$, which leads to the familiar chiral symmetry of the Dirac equation, that is, that the Weyl spinors are eigenvalues of γ_5. For $N>2$ we therefore expect a parity violation. At first it seems that the mass term for $N>2$ enters in two different ways into the Lagrangian, which gives us the choice to use two different mass parameters. But this is only apparent due to the freedom to rescale μ.

The Dirac system can also be written as a system of N coupled first-order equations for polyspinors as well as spinors. Take as an example $N=3$, where the Dirac equation (2.5) can be written as the three equations

$$\partial_{\dot\Sigma\Sigma}\psi^\Sigma=-iM\bar\mu_{\dot\Sigma},$$
$$\tfrac{1}{2}\epsilon^{\dot\Sigma\dot\Omega\dot\Lambda}\partial_{\dot\Omega\Omega}\bar\mu_{\dot\Sigma}=-iM\phi^{\dot\Lambda}{}_\Omega,$$
$$\tfrac{1}{2}\epsilon^{\Sigma\Omega\Lambda}\partial_{\dot\Lambda\Lambda}\phi^{\dot\Lambda}{}_\Omega=-iM\psi^\Sigma.$$

Here $\phi^{\dot\Lambda}{}_\Omega$ has the index structure not of a vector but of a tensor product of a dotted spinor with an undotted cospinor.

The minimal-coupling generalization of the wave equations to curved time space is simply done by replacing the ordinary derivative ∂ by the covariant derivative D. Due to the noncommutativity of D the scalar wave equation is no longer the root of the Dirac equation.

B. Symmetries of the neutrino equation

We examine now the discrete symmetries: T: = time reversal; C: = particle–antiparticle conjugation; and P: = parity of the flat space neutrino equation (2.4). Here C for the neutrino is defined as usual by $C\psi(x,t)=:\bar\psi(x,t)$. We first look at the case $N=2$.

Taking the standard Pauli matrices $\sigma^\mu=(\sigma^0,\sigma^i)$, $i=1,2,3$ and $\sigma^0=-\mathbf{1}$ as basis for the spin vector and making the usual operator associations $E=i\partial_t=i\partial_0$ and $p_i=-i\partial_i$, we obtain

$$i\sigma^\mu{}_{\dot\Sigma\Sigma}\partial_\mu\psi^\Sigma=(i\sigma^0{}_{\dot\Sigma\Sigma}\partial_0+i\sigma^i{}_{\dot\Sigma\Sigma}\partial_i)\psi^\Sigma$$
$$=(\mathbf{1}E+\sigma^i\cdot p_i)\psi=0. \tag{2.6}$$

The above equation is invariant under the discrete symmetries T and PC, but not under P and C alone. To examine T in detail take $T\psi(t,x)=U_T\bar\psi(-t,x)$, where U_T is a unitary operator. The time reversed and complex conjugated equation (2.6) appears now as

$$(\mathbf{1}E-\bar\sigma^i\cdot p_i)\bar\psi(-t,x)=0.$$

Multiplying this equation on the left by U_T and requiring that $U_T\bar\psi$ satisfies (2.6) gives the condition that $U_T\bar\sigma^i\tilde U_T=-\sigma^i$ ($\tilde U_T:=U^{-1}{}_T$). This is achieved by $U_T=\sigma^2$, because it anticommutes with the real σ^1 and σ^3, and $\sigma^2=-\bar\sigma^2$. By similar kinds of calculations one can also show that PC, and therefore PCT, is conserved.

The existence of the operator U_T is ensured by the fact that there is only one inequivalent two-dimensional representation of $d\,\mathbf{SU}_2$, which implies that the respresentation $\mathfrak{R}(u)$ is equivalent to the complex conjugate representation $\overline{\mathfrak{R}}(u)$.

For $N>2$ this is no longer true because $\mathbf{SU}_N$ has two inequivalent N-dimensional representations, and $\mathfrak{R}(u)$ cannot be transformed into $\overline{\mathfrak{R}}(u)$ by an inner automorphism.[5] Therefore an operator U_T with the desired properties does not exist. The same arguments hold for PC, where we define the parity operator P as $P\psi(x^e,x^i,t):=\psi(-x^e,x^i,t)$. The x^e are the three spatial external coordinates and x^i stands for the extra internal dimensions. This definition of P is justified because the x^i are regarded as internal degrees of freedom and are therefore left invariant.

With this most natural definition of P even PCT is violated. This can be seen as follows.

Under θ: = PCT of (2.4) we get

$$(-\sigma^0\partial_0-\sigma^e\partial_e+\sigma^i\partial_i)\psi(-t,-x^e,x^i)=0.$$

To have invariance under θ we need

$$\theta\sigma^0\theta^{-1} = \sigma^0,$$
$$\theta\sigma^e\theta^{-1} = \sigma^e,$$
$$\theta\sigma^i\theta^{-1} = -\sigma^i.$$

For flat space σ^e and σ^i form a basis for $d\,\mathbf{SU}_N$, so that θ is an involutive automorphism of $d\,\mathbf{SU}_N$ with three fixed points given by σ^e. A theorem in Ref. 6 states that θ is an inner automorphism if and only if the rank of $d\,\mathbf{SU}_N$ is equal to the rank of the set of fixed points. This is the case for $N=4$, which is the only other case besides $N=2$ where the *PCT* theorem holds. The emergence of $\mathbf{SU}_4$ as a special case is peculiar because of the homomorphism of $\mathbf{SU}_4$ onto SO(6), so that *PCT* is only conserved in those cases where we have the group isomorphisms $\mathbf{SU}_2 = \mathrm{Spin}(3)$ and $\mathbf{SU}_4 = \mathrm{Spin}(6)$, and therefore a direct relation to the orthogonal groups. This, of course, may be purely accidental.

The violation of *PCT* fits well into the scheme of second quantization of the hyperspin theory. Due to the $\mathbf{SU}_N$ symmetry, one would quantize all hyperspinor components with Fermi statistics so that $N=3$, for example, describes a Fermi spin-$\frac{1}{2}$ field and a Fermi spin-0 field. We therefore have a forced breakdown of the spin-statistics connection for $N>2$.

We note that the statement of the usual *PCT* theorem is still valid. A symmetry in accordance with the *PCT* theorem can be generated if one defines a $P^*\psi(x^e,x^i,t) := \psi(-x^e,-x^i,t)$. Obviously P^*CT remains a good symmetry of (2.4) for all N. We argue that P is a more physical symmetry than P^*, because the parity transformation should not affect the inner gauge degrees of freedom.

In all our discussions about the discrete symmetries we stress that we still work in first quantization and that we only consider inner automorphisms of the underlying algebra. The adherence to inner automorphisms is due to the fact that we copy simply the usual treatment of the Weyl neutrino equation.

One should note that there exists involutive outer automorphisms which conserve *PCT* for all N and relate $\mathfrak{R}(u)$ and $\overline{\mathfrak{R}}(u)$ for $N>2$. An outer automorphism is defined as an automorphism which is not an inner one. It consists of a rotation of the weight diagram which is not an element of the Weyl group. The significance of these outer automorphisms for the discussion of time space symmetries of the neutrino equation is not clear, because the neutrino appears in only one helicity state, so there is no room for a natural outer automorphism as is in the Dirac equation (2.5), where one can interchange the ψ^Σ and μ_Σ fields. But even this is only possible for $N=2$, because only there does a quadratic spinor metric exist that is induced by the Levi-Civita ϵ spinor.

III. THE NEUTRINO EQUATION ON $\mathbf{U}_N$

We first formulate the neutrino equation with a right invariant spin map. This switch from using the left invariant spin map of Ref. 1 to using a right invariant σ in the present work is done merely for convenience, because the resulting equation has a nicer form. The physics is of course unaffected. We recall from Ref. 1 that the right invariant spin vector σ and the spin connection Γ are given by

$$e_\lambda{}^\Omega{}_{\Omega'}\sigma^\lambda{}_{\dot\Sigma\Sigma} = i\delta_\Sigma{}^\Omega\mu_{\dot\Sigma\Sigma'}u^{\Sigma'}{}_{\Omega'}$$

and

$$\tilde e^{\lambda\Omega'}{}_\Omega\Gamma_\lambda{}^\Sigma{}_{\Sigma'} = k\{\delta^\Sigma{}_\Omega\tilde u^{\Omega'}{}_{\Sigma'} - (1/N)\delta^\Sigma{}_{\Sigma'}\tilde u^{\Omega'}{}_\Omega\},$$

where $k=-\frac{1}{2}$ for the torsion-free case. The $e_\lambda{}^\Omega{}_{\Omega'}(u)$ are a Hermitian basis for the Lie algebra of $\mathbf{U}_N$ at $u\,[=:d\,\mathbf{U}_N(u)]$, and $\mu_{\dot\Sigma\Sigma}$ is the nondegenerate group metric on $\mathbf{U}_N$. By minimal coupling of (2.6) to the curvature of $\mathbf{U}_N$ we obtain

$$\sigma^\lambda{}_{\dot\Sigma\Sigma}D_\lambda\psi^\Sigma = 0 \Rightarrow u^{\Sigma'}{}_{\Omega'}\,\partial^{\Omega'}{}_\Sigma\psi^\Sigma = m\psi^{\Sigma'}, \tag{3.1}$$

where we defined $m := -k(N-1/N)$ and $\partial^\Sigma{}_{\Sigma'} := \partial/\partial u^{\Sigma'}{}_\Sigma$.

Every element $u\in\mathbf{U}_N$ can be written as $\exp(-i(t/N))s$, where $s\in\mathbf{SU}_N$. We therefore have that

$$t = i\ln(\det u)$$

and

$$\partial_t = (-i/N)u{:}\partial_u.$$

Here the colon stands for product and trace. The operator $u\cdot\partial_u$ decomposes into its trace $i\partial_t$, and $s\cdot\partial_s$, which is a traceless operator. In this way we separate the spatial derivatives from the time derivative. The neutrino equation now looks like a Schrödinger equation:

$$i\,\partial_t\psi = m\psi - s\cdot\partial_s\cdot\psi.$$

This equation has the obvious solution $e^{-imt}\psi(0)$, where $\psi(0)$ is a spatially constant spinor, or viewed in momentum space, a neutrino at rest. The neutrino has a rest mass given by m due to the compactness of the space, which admits standing waves. The only relevant length scale for $\mathbf{U}_N$ is the present radius ρ of the universe, which we take to be of the order 10^{28} cm. We get that m is of $O(\rho^{-1}) \approx 10^{-31}$ eV, which is negligible.

To solve this equation in general, we analytically continue the equation to $\mathbf{GL}_N$ and let $z\in\mathbf{GL}_N$ (compare Ref. 1). What actually has to be solved is the eigenvalue spectrum of the operator $\mathfrak{D} := z\cdot\partial_z$, which is the Dirac operator in the theory. We write this as

$$\mathfrak{D}\cdot\psi = \lambda\psi. \tag{3.2}$$

Because of the aforementioned decomposition in time and space derivatives, a time phase factor e^{-iwt} of ψ will change λ to $\lambda + w$. By choosing $e^{-i(m-\lambda)t}$ as the phase factor for the solutions of (3.2), we obtain solutions which will satisfy (3.1).

Next we examine the invariance properties of the neutrino equation. We can act on z by left or right multiplication with $G\in\mathbf{GL}_N$ and require form invariance of the equation. Under left multiplication z and ∂_z transform contragradiently as

$$z' = G\cdot z$$

and

$$\partial_{z'} = \partial_z\cdot\tilde G.$$

The neutrino equation changes into

$$\mathfrak{D}\cdot\tilde G\cdot\psi'(G\cdot z) = m\psi'(G\cdot z).$$

Using infinitesimals $G\approx 1+g$ we obtain to first order

$$\psi'(z+g\cdot z) = \psi(z) + g\cdot\psi(z).$$

A Taylor expansion of ψ' gives to first order the change of ψ as

$$\delta\psi = (g - \mathrm{Tr}(z\cdot\partial_z\cdot g))\cdot\psi.$$

Because g is an element of $d\,\mathbf{GL}_N$, a basis for g is given by the matrix units $E_i{}^A{}_B = \delta^A{}_I\delta^{I'}{}_B$. We use now the notation that small Latin letters run from $1,\ldots,n = N^2$ and capital Latin letters go from $1,\ldots N$. Inserting this in the above equation leads to the following N infinitesimal generators:

$$g_i := g^{I'}{}_I{}^A{}_B = \delta^A{}_I\delta^{I'}{}_B - z^{I'}{}_S\,\partial^S{}_I\delta^A{}_B. \tag{3.3}$$

As is easily checked, they obey the commutation relations for $d\,\mathbf{GL}_N$.

Next we consider the invariance under right multiplication by G.

The Dirac operator $\mathfrak{D}$ was already constructed to be right invariant, therefore, not surprisingly, we find that $\psi'(z') = \psi(z)$, which by similar calculations leads to the generators g'_i of the right invariance group:

$$g'_i := g'^{I'}{}_I{}^A{}_B = -(z^S{}_I\,\partial^{I'}{}_S)\delta^A{}_B. \tag{3.4}$$

These N generators form also a $\mathbf{GL}_N$ Lie algebra, and moreover, g and g' commute. The number of diagonal operators that simultaneously commute with the spin Hamiltonian is therefore at least $2N-1$, the minus one stemming from the fact that

$$\sum_{I'=I} g^{I'}{}_I = \mathbf{1} + \sum_{I'=I} g^{I'}{}_I.$$

For $N=2$ the constants of the motion are the energy, the total angular momentum (including spin), and the "linear" momentum. The last has the form of an angular momentum, because the space part of $\mathbf{U}_2$ is S^3.

For the representations of $\mathbf{U}_N$ we follow the approach of Bargmann and Schwinger in Ref. 7. By considering homogeneous polynomials in $N^2 = n$ complex variables and using boson creation and annihilation operator methods, we will find the irreducible representations (irreps) of $\mathbf{U}_N$. This method is called the boson calculus in the literature[8] and is based on the Jordan map.[9] The following notation is mainly due to Bargmann.[7]

Consider $\mathfrak{F}_n$, the space of entire analytic functions $f(z)$, where $z = z^I{}_J$, $I, J = 1,\ldots,N$. Here z can be thought of as a point in $\mathbf{C}^n$. $\mathfrak{F}_n$ is made into a Hilbert space by defining the following scalar product:

$$\langle f|f'\rangle = \int \bar{f}(z)f'(z)d\mu_n(z),$$

where

$$d\mu_n(z) = \pi^{-n}\exp(-\bar{z}{:}z)\prod_{k=1}^{n} dx_k\,dy_k$$

$$(z^I{}_J = x^I{}_J + iy^I{}_J =: x_k + iy_k).$$

With this definition of a Hilbert space z and ∂_z are adjoint operators on $\mathfrak{F}_n$ with respect to the Hilbert space metric. Moreover, they satisfy the usual boson commutation relations for creation and annihilation operators:

$$[z^I{}_J, z^{I'}{}_{J'}] = [\partial^I{}_J, \partial^{I'}{}_{J'}] = 0, \quad [z^I{}_J, \partial^{I'}{}_{J'}] = \delta^I{}_{J'}\delta^{I'}{}_J.$$

As is well known,[10] from these operators one can obtain all the totally symmetric irreps of $\mathbf{U}_n$, all of the irreps of $\mathbf{U}_N$ and representations for $\mathbf{U}_N \otimes \mathbf{U}_N$, for which the Gelfand invariants of the two $\mathbf{U}_N$ algebras are identical.[11]

The simplest orthonormal set in $\mathfrak{F}_n$ is given by

$$\phi_n := z^{\{n\}}/\sqrt{\{n!\}} =: |n^1{}_1, n^1{}_2, \ldots, n^N{}_N\rangle,$$

where $z^{\{n\}} := (z^1{}_1)^{n^1{}_1}(z^1{}_2)^{n^1{}_2}\cdots(z^N{}_N)^{n^N{}_N}$ and $\{n!\} := n^1{}_1!n^1{}_2!\cdots n^N{}_N!$ The ϕ_n obey

$$\langle\phi_n|\phi_{n'}\rangle = \delta_{nn'}.$$

The action of the operators $z^I{}_J$ and $\partial^J{}_I$ on this set is of course the same as in the standard boson calculus of creation and annihilation operators,

$$z^I{}_J|n^1{}_1,\ldots,n^I{}_J,\ldots\rangle\sqrt{n^I{}_J + 1}\,|n^1{}_1,\ldots,n^I{}_J + 1,\ldots\rangle$$

and

$$\partial^J{}_I|n^1{}_1,\ldots,n^I{}_J,\ldots\rangle = \sqrt{n^I{}_J}\,|n^1{}_1,\ldots,n^I{}_J - 1,\ldots\rangle\,.$$

We will solve the neutrino equation for $N=2$ as a warm-up exercise. In component form the neutrino equation is

$$z^1{}_s\,\partial^s{}_1\psi_1{}^1 + z^1{}_s\,\partial^s{}_2\psi_1{}^2 = \lambda\psi_1{}^1,$$

$$z^2{}_s\,\partial^s{}_1\psi_1{}^1 + z^2{}_s\,\partial^s{}_2\psi_1{}^2 = \lambda\psi_1{}^2.$$

Assume $\psi_1{}^1 = v^1(n_i)|0,n^1{}_2,0,n^2{}_2\rangle$, $\psi_1{}^2 = v^2(n_i)|0, n^1{}_2 - 1,0,n^2{}_2 + 1\rangle$. The v^Σ are functions of the occupation numbers $n^{I'}{}_I =: n_i$. Using the boson calculus leads to the following consistency requirement:

$$\begin{pmatrix} n^1{}_2 & \sqrt{n^1{}_2(n^2{}_2+1)} \\ \sqrt{n^1{}_2(n^2{}_2+1)} & n^2{}_2+1 \end{pmatrix}\begin{pmatrix} v^1 \\ v^2 \end{pmatrix} = \lambda\begin{pmatrix} v^1 \\ v^2 \end{pmatrix}.$$

We abbreviate this equation as $B\cdot v = \lambda v$. This is an eigenvalue equation for B, which we call the dynamical matrix. Solving it in the normal fashion gives two distinct eigenvalues:

$$\lambda_- = n^1{}_2 + n^2{}_2 + 1, \quad \lambda_+ = 0. \tag{3.5}$$

The corresponding eigenvectors v are

$$v_- = \begin{pmatrix} \sqrt{n^1{}_2} \\ \sqrt{n^2{}_2+1} \end{pmatrix}$$

and

$$v_+ = \begin{pmatrix} \sqrt{n^1{}_2} \\ -\sqrt{n^2{}_2+1} \end{pmatrix}.$$

Multiplying ψ_1 by a phase factor $e^{-i(m-\lambda)t}$ gives us a solution to (3.1). Defining the energy operator as usual as $i\,\partial_t$ we obtain the following eigenvalue spectrum for E:

$$E = m - \lambda + (1/N)s,$$

where s is defined as

$$s = \sum_i n_i.$$

This s describes mathematically the degree of the polynomial in z or more physically the total number of elementary bosonic "spins." It is clearly integer valued. Inserting (3.5) gives a positive and negative energy spectrum. According to our definition in (3.1), we have $m = \frac{1}{4}$ for $N=2$,

$$E_+ = \tfrac{3}{4} + \tfrac{1}{2}s, \quad E_- = \tfrac{3}{4} - \tfrac{1}{2}(s+2).$$

Due to our assumption of the form of the state vector ψ_1, we have to treat ${n^1}_2 = 0$ as a special case. We find that for E_+, $s = 0,1,...$, while $s = 0$ is excluded for E_-. This restriction gives an energy spectrum where the positive and negative ground state is of equal magnitude. Moreover, Ref. 12 showed by a different method that the number of positive energy states equals the number of negative energy states for a given energy. We also observe that the introduction of torsion would have destroyed this symmetry.

We turn now to the question of the completeness of the solution set. We first note that we can repeat the same calculations with ${\psi_2}^1 = {v_2}^1(n)|{n^1}_1,0,{n^2}_1,0\rangle$ and ${\psi_2}^2 = {v_2}^2(n)|{n^1}_1 - 1,0,{n^2}_1 + 1,0\rangle$ and obtain the same result. By applying the raising and lowering operators (3.3) and (3.4) of the invariance group, we can obtain even more solutions. We note that these operators form two commuting $\mathbf{U}_2$ algebras as follows. Define

$$L'_- := -{g'^1}_2; \quad L'_+ := -{g'^2}_1;$$
$$L'_z := -\tfrac{1}{2}({g'^1}_1 - {g'^2}_2);$$
$$L_- := -{g^1}_2; \quad L_+ := -{g^2}_1; \quad L_z := -\tfrac{1}{2}({g^1}_1 - {g^2}_2);$$
$$L_0 := \tfrac{1}{2}({z^T}_S\,\partial^S{}_T); \quad L'_0 := \tfrac{1}{2}({z^T}_S\,\partial^S{}_T - 1)\mathbf{1}_2.$$

Here L'_0 and L_0 are the $\mathbf{U}_1$ generators. The other generators satisfy the usual angular momentum commutation relations:

$$[L_z,L_+] = L_+; \quad [L_z,L_-] = -L_-; \quad [L_+,L_-] = 2L_z;$$
$$[L'_z,L'_+] = L'_+; \quad [L'_z,L'_-] = -L'_-;$$
$$[L'_+,L'_-] = 2L'_z;$$

and

$$[L',L] = 0.$$

Let us call the eigenvalues of L'_z and L_z m_r and m_l, respectively. The eigenvalues of our two solutions $\psi_1 := \begin{pmatrix} {\psi_1}^1 \\ {\psi_1}^2 \end{pmatrix}$ and $\psi_2 := \begin{pmatrix} {\psi_2}^1 \\ {\psi_2}^2 \end{pmatrix}$ under these operators are

$$L_z\psi_1 = \tfrac{1}{2}({n^1}_2 - {n^2}_2 - 1)\psi_1; \quad L'_z\psi_1 = -\tfrac{1}{2}({n^1}_2 + {n^2}_2)\psi_1,$$
$$L_z\psi_2 = \tfrac{1}{2}({n^1}_1 - {n^2}_1 - 1)\psi_2; \quad L'_z\psi_2 = \tfrac{1}{2}({n^1}_1 + {n^2}_1)\psi_2.$$

Moreover one can show that

$$\psi_m := v_+\begin{pmatrix} |1,0,{n^2}_1,0\rangle \\ |0,0,{n^2}_1 + 1,0\rangle \end{pmatrix}$$

satisfies the requirement for the highest weight state, namely,

$$L_+\psi_m = L'_+\psi_m = 0.$$

Applying the lowering operators L_- and L'_- will generate a whole ladder of states. One expects this to produce a complete set, but the completeness proof is not available. Sen[12] proved a completeness relation for his solutions, based on the completeness of $\mathbf{SU}_2$ scalar functions. Our operator methods do not seem to work that easily, and already in Ref. 7 Bargmann stated that the operator methods do not seem to help in proving completeness of the representations.

We generalize now our approach to the case of arbitrary N. We work with the convention that the occupation numbers in ψ^Σ which are not written out are assumed to be zero. There are N solutions ψ which can be found immediately by generalization of the $\mathbf{U}_2$ solutions. They are labeled by the subscript S. The components of ψ_S written out in ket form look like

$$\psi^1{}_S = v^1{}_S(n_i)|{n^1}_S,{n^2}_S,{n^3}_S,...,{n^N}_S\rangle,$$
$$\psi^2{}_S = v^2{}_S(n_i)|{n^1}_S - 1,{n^2}_S + 1,{n^3}_S,...,{n^N}_S\rangle,$$
$$\psi^\Sigma{}_S = v^\Sigma{}_S(n_i)|{n^1}_S - 1,{n^2}_S,...,{n^\Sigma}_S + 1,$$
$${n^{\Sigma+1}}_S,...,{n^N}_S\rangle, \quad \Sigma = 3,...,N.$$

As can be easily checked, the dynamical matrix for the general case is symmetric and has the form

$$B = \begin{pmatrix} {n^1}_S & \sqrt{{n^1}_S({n^2}_S + 1)} & \cdots & \sqrt{{n^1}_S({n^N}_S + 1)} \\ \sqrt{{n^1}_S({n^2}_S + 1)} & {n^2}_S + 1 & \cdots & \vdots \\ \vdots & \vdots & \cdots & \vdots \\ \sqrt{{n^1}_S({n^N}_S + 1)} & \cdots & \cdots & {n^N}_S + 1 \end{pmatrix}.$$

Here B can be thought of as a dyadic product bb of a vector b with itself, where $b^A = \sqrt{n^A{}_S + 1 - \delta^A{}_1}$. In dyadic notation $B\cdot v$ can be written as $b(b\cdot v)$. Because b spans a one-dimensional subspace of an N-dimensional space, there are $N-1$ other eigenvectors v that are perpendicular to b with eigenvalue 0. Thus $v = b$ is the only eigenvector of B with a nonzero eigenvalue, which is $b\cdot b$ or equivalently $\mathrm{Tr}(B)$. The two eigenvalues, 0 and $\mathrm{Tr}(B)$, give rise to the two energy spectra,

$$E_+ = (N^2 - 1)/2N + (1/N)s, \quad s = 0,1,...,$$
$$E_- = (N^2 - 1)/2N - [(N-1)/N](s+N),$$
$$s = 1,2,...,$$

where we make use of the definition of m in (3.1). Here again $s = 0$ is not possible for E_-. Once more this makes the lowest positive and highest negative energy state have equal magnitude. The positive and negative energies are linear in s, but they have different slopes, depending on the value of N (see Fig. 1). This means that the negative energy spectrum is less dense, some positive energy modes do not have a corresponding antipartner. The density ratio is $N-1$, which is unity for $N = 2$.

The question arises if the asymmetry in the behavior of positive and negative energy solutions is real or just an artifact of an incomplete solution set. As before for $N = 2$ we can use the shift operators of our invariance group to construct new solutions, but do we get all the solutions? Due to the complexity of the higher $\mathbf{U}_N$ groups it is even harder to get a handle on this problem and the question is still unsolved.

A possible way to improve the calculations is to use the Gelfand states $|(m)\rangle$ (Ref. 10) of $\mathbf{U}_N \otimes \mathbf{U}_N$, which are gen-

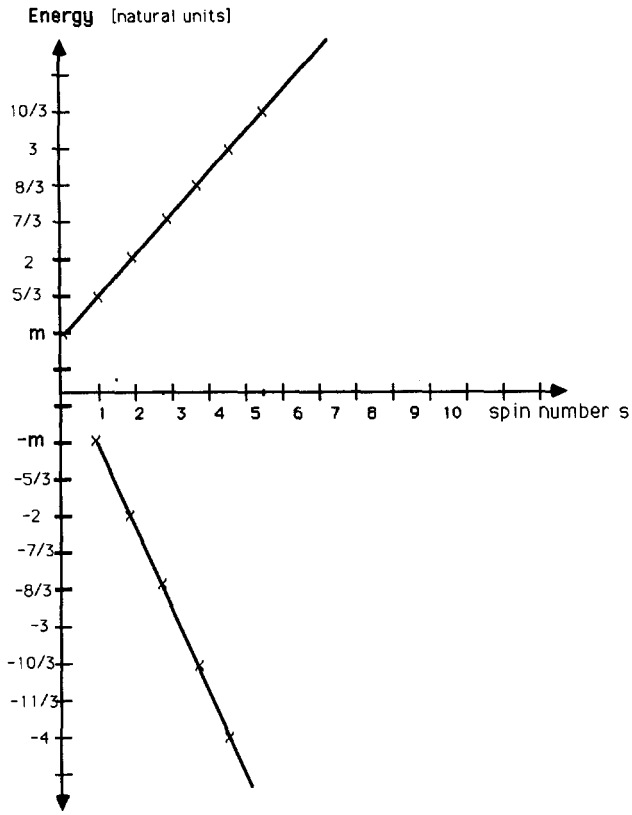


FIG. 1. Energy spectrum of the neutrino (upper graph) and antineutrino (lower graph) for $N = 3$. The cosmological rest mass is $\frac{4}{3}$ in natural units.

erated by the scalar parts of our generators in (3.3) and (3.4), i.e.,

$$ {}^{l}g^{I'}{}_{I} := z^{I'}{}_{S}\,\partial^{S}{}_{I} \quad \text{and} \quad {}^{r}g^{I'}{}_{I} := z^{S}{}_{I}\,\partial^{I'}{}_{S}. $$

As remarked earlier, ${}^{r}g$ and ${}^{l}g$ generate all the irreps of $\mathbf{U}_N \otimes \mathbf{U}_N$, but with the restriction that the Gelfand labels (m) of each state in the product are identical.

The $|(m)\rangle$ are orthonormal and the transformations induced by ${}^{r}g$ and ${}^{l}g$ are in principal well known. In fact, ${}^{l}g$ are just the matrix elements of $\mathfrak{D}$, and by using the Gelfand states as components of ψ one should be able to find the general solution. The problem is that the induced transformations on $|(m)\rangle$ consist of linear superpositions of different states, which makes this approach algebraically difficult. Using a computer with an algebraic programming language could solve this problem.

IV. CONCLUSIONS

We gave the Lagrangians for the generalized Klein–Gordon, Dirac, and Weyl-neutrino equations and discussed the symmetry properties of the flat space neutrino equation. We showed that for $N = 3$ and $N > 4$ all the discrete symmetries are broken, and even PCT is violated.

The hyperneutrino equation on $\mathbf{U}_N$ was solved by means of the boson calculus. The solutions gave rise to an energy spectrum that has symmetry between the negative and positive energy solutions only for $N = 2$. We found that the density ratio of positive to negative energy states was $N - 1$. The hyperneutrino acquired a negligible rest mass (of the order 10^{-31} eV) due to the global constant curvature of $\mathbf{U}_N$.

So far only the hyperneutrino equation was studied on $\mathbf{U}_N$. Finding solutions to the generalized Dirac (or Klein–Gordon) equation is another task. The application of index theory[13] to the group manifold is possible and will give important information about the behavior of the solutions to $\mathfrak{D}$. This would resolve any doubts about the neutrino spectrum.

The difference in density of positive and negative energy solutions for $N > 2$ is quite surprising for a maximal symmetric group like $\mathbf{U}_N$. If one assumes that the very early universe can be approximated by an $\mathbf{U}_N$, then the result could explain the matter–antimatter asymmetry in our universe.

A good dimensional reduction procedure has to be found in order to make the model more physical and to study the behavior of the energy spectrum.

What is the relevance of supersymmetry to hyperspinors? For example, in the $N = 3$ theory a three-component hyperspinor can be thought of as a supermultiplet consisting of a spin-$\frac{1}{2}$ ψ^{E} and a scalar ψ^{I}. Nevertheless we have assumed that the transformation group is $\mathbf{SL}_3$ and not one of the supergroups. From symmetry arguments (ψ is a spinor) we would treat all three components of ψ as fermions and second quantize them with anticommuting operators. But because ψ^{I} is a scalar under the Lorentz group, the spin-statistics theorem is violated.

If we wish to respect the usual spin-statistics connection, we should treat the external components as fermions and the internal components as bosons. It is even possible that all components are elements of an underlying Grassmann algebra, and then fermions are the odd elements and bosons are the even ones. The different treatment of external and internal spin components would have important consequences for the spin vector. The components $\sigma^{t}{}_{EE}$ and $\sigma^{t}{}_{II}$ would result in commuting manifold coordinates, while the mixed components $\sigma^{t}{}_{EI}$ and $\sigma^{t}{}_{IE}$ give Grassmann (anticommuting) coordinates. Is this a possible link to supermanifolds? Accidentally a $\mathbf{B}_3$ with five commuting coordinates would be much closer to the spirit of the original five-dimensional Kaluza–Klein theory.

ACKNOWLEDGMENTS

I am indebted to D. Finkelstein for many discussions, and to R. F. Fox and L. C. Biedenharn for helpful hints and criticism. I also would like to thank my colleagues at the TU Clausthal for discussions on the revised version.

[1] C. Holm, "The hyperspin structure of unitary groups," J. Math. Phys. **29**, 978 (1988); also Ph.D. thesis, Georgia Institute of Technology, 1987 (unpublished).
[2] D. Finkelstein, S. R. Finkelstein, and C. Holm, Phys. Rev. Lett. **59**, 1265 (1987); S. R. Finkelstein, Ph.D. thesis, Georgia Institute of Technology, 1987 (unpublished).
[3] D. Finkelstein, Phys. Rev. Lett. **56**, 1532 (1986); D. Finkelstein, S. R. Finkelstein, and C. Holm, Int. J. Theor. Phys. **25**, 441 (1986).
[4] E. T. Whittaker, *A Treatise on the Analytical Dynamics of Particles and Rigid Bodies* (Dover, New York, 1944).

[5]J. F. Cornwell, *Group Theory in Physics* (Academic, London, 1984), Vol. II.
[6]S. Helgason, *Differential Geometry, Lie Groups and Symmetric Spaces* (Academic, New York, 1978), p. 424, Theorem 5.6.
[7]V. Bargmann, Rev. Mod. Phys. **34**, 829 (1962); J. Schwinger, "On angular momentum," U.S. Atomic Energy Commission, NYO-3071, 1952.
[8]L. C. Biedenharn and J. D. Louck, *Angular Momentum in Quantum Physics, Encyclopedia of Mathematics and its Applications*, Vol. 8 (Addison–Wesley, Reading, MA, 1981).
[9]P. Jordan. Z. Phys. **94**, 531 (1935).
[10]W. J. Hollmann and L. C. Biedenharn, "The representations and tensor operators of the unitary groups U(n)," in *Group Theory and Its Applications*, edited by E. M. Loeble (Academic, New York, 1971), Vol. 2.
[11]J. D. Louck, J. Math. Phys. **6**, 1786 (1965).
[12]D. Sen, J. Math. Phys. **27**, 472 (1986).
[13]T. Eguchi, P. B. Gilkey, and A. J. Hanson, Phys. Rep. **66**, 213 (1980).

Equivalence of transfer matrices

M. A. Novotny
IBM Bergen Scientific Center, Allégaten 36, N-5007 Bergen, Norway

(Received 26 January 1988; accepted for publication 8 June 1988)

Three theorems dealing with transfer matrices in statistical mechanical systems are proved. The theorems state that the nonzero eigenvalues of transfer matrices formed through various prescriptions are identical. Hence it is possible to ascribe a physical meaning to all the eigenvalues of a transfer matrix, not just to the few largest eigenvalues. The first theorem states that the transfer matrix formed by building a system M layers at a time has as its only nonzero eigenvalues the eigenvalues of the transfer matrix formed by building the M layers of the system one at a time. This theorem relates the product of two $n^M \times n^M$ M-layer transfer matrices to the product of M one-layer $M \times M$ transfer matrices. The second theorem states that one of the $n^M \times n^M$ M-layer transfer matrices (for $M > 1$) has only one nonzero eigenvalue. A procedure for finding this eigenvalue and all eigenvectors is given. The third theorem generalizes the first to the case where the chosen layering is not an integer multiple of the interaction length.

I. INTRODUCTION

The transfer matrix formalism may be used to obtain the partition function of any statistical mechanical system when it has short-range interactions in at least one direction.[1–4] This allows the system to be built one layer at a time in a direction with short-range interactions. The size of the transfer matrix must be chosen to be equal to or greater than the size required to make separate layers, i.e., layers that interact only with at most two other layers. This property allows the analysis of the system in terms of a Markov process, since the interaction of one layer depends only on the preceding and following layers. For a statistical mechanical system, the partition function is given by

$$Z = \mathrm{Tr}(\mathsf{A}_1\mathsf{A}_2\mathsf{A}_3\cdots\mathsf{A}_N\mathsf{B}) \tag{1}$$

when the system has N layers. Here A_i is a one-layer transfer matrix, i.e., a matrix that adds the single layer i to the system. The matrix B describes the interactions at the boundary.

The question that naturally arises is the relationship between the eigenvalues of the different transfer matrices formed when the transfer matrix is built up with more than one layer at a time. In particular, if each layer has n states, the one-layer transfer matrices that enter Eq. (1) are of size $n \times n$. If one decides to build up the transfer matrix using M of these layers at a time, the M-layer transfer matrices one then uses are $n^M \times n^M$. Physically, one would expect that at least the largest eigenvalue of the different ways of forming the transfer matrices must be invariant. This is because the largest eigenvalue is related to the partition function Z of the system, and physical quantities such as the free energy can be derived from the partition function. In the thermodynamic limit, $N \to \infty$ in Eq. (1), only the largest eigenvalue comes into the calculation of the partition function if all A_i are identical. Similarly, in this limit a correlation length is related to the ratio of the largest and next-largest eigenvalues of the transfer matrix. Thus this ratio must also be invariant using various prescriptions to build the transfer matrices. But what about the other eigenvalues? This paper will relate the eigenvalues of the $n \times n$ one-layer transfer matrix to the $n^M \times n^M$ M-layer transfer matrix. In fact, it will be shown that quite generally the M-layer transfer matrix has at least $n^M - n$ zero eigenvalues, and the remaining n eigenvalues are just the eigenvalues of the product of M one-layer transfer matrices.

The next section contains a number of lemmas and three theorems. The first theorem relates the eigenvalues of the system when it is built up M layers at a time (an M-layer transfer matrix) to the eigenvalues of a product of M one-layer transfer matrices. The second theorem describes the structure of a portion of the M-layer transfer matrix. The third theorem generalizes the first to the case when the chosen layering is not an integer multiple of the interaction length. Finally, Sec. III contains a discussion of the physical importance of the theorems. All discussion of the physical and calculational properties of the three theorems and associated lemmas of Sec. II will be deferred to Sec. III. Consequently, the reader may initially skip over the proofs in Sec. II and concentrate on the discussion in Sec. III.

II. THEOREMS AND PROOFS

Throughout the manuscript, the mathematical properties of the row and column matrix products introduced in Ref. 5 will be used—the notation 5(x.x) will be used to refer to Eq. (x.x) of Ref. 5. In particular, the reader should note that curly brackets, { }, denote the row and column products while parentheses, ( ), denote a regular matrix. The Hadamard (element-by-element) product will be denoted by $\odot$, and the regular matrix product will be denoted by the juxtaposition of the matrices. In most instances, the dimension of the matrices will not be explicitly given—they are assumed to be such that the matrix products are defined (this assumption can always be met by having all matrices square of size $n \times n$). All matrices are assumed to have elements from the field of complex numbers. The matrix I will be reserved for the identity matrix for regular matrix multiplication, and the matrix J will always stand for the identity for Hadamard matrix multiplication (hence J has all elements equal to 1).

 0022-2488/88/102280-08$02.50

Assume each of the N one-layers is numbered with the numbering performed to minimize the difference between the numbers of interacting one-layers. Then the range of interaction is taken to be the largest difference between any two interacting one-layers.

A. Interaction range 1

The first two theorems deal with the trace of a matrix with N one-layers, with one-layer i interacting only with one-layers $i-1$ and $i+1$. The general matrix has the form[5]

$$Z=\mathrm{Tr}\left(\begin{Bmatrix} \mathsf{I} & \mathsf{A}_1 & \mathsf{J} & \cdots & \mathsf{J} & \mathsf{J} \\ \mathsf{J} & \mathsf{I} & \mathsf{A}_2 & & \mathsf{J} & \mathsf{J} \\ \mathsf{J} & \mathsf{J} & \mathsf{I} & \ddots & \mathsf{J} & \mathsf{J} \\ \vdots & \vdots & & \ddots & & \vdots \\ \mathsf{J} & \mathsf{J} & \mathsf{J} & & \mathsf{I} & \mathsf{A}_{N-1} \\ \mathsf{A}_N & \mathsf{J} & \mathsf{J} & \cdots & \mathsf{J} & \mathsf{I} \end{Bmatrix}\right), \tag{2}$$

and from 5(3.13) [without loss of generality set $\mathsf{D}_i=\mathsf{I}$ in 5(3.13)]

$$Z=\mathrm{Tr}\,(\mathsf{A}_1\mathsf{A}_2\cdots\mathsf{A}_N). \tag{3}$$

Call A_i a one-layer transfer matrix, i.e., it is the transfer matrix that adds layer i (and includes all interactions within layer i and between one-layers i and $i+1$). From 5(4.3) (with all $\mathsf{D}=\mathsf{I}$ and $\mathsf{B}=\mathsf{J}$) it is possible (when $N/2$ is an integer) to also write

$$Z=\mathrm{Tr}\left(\begin{Bmatrix} \mathsf{I} & \mathsf{A}_1 \\ \mathsf{J} & \mathsf{I} \end{Bmatrix}\begin{Bmatrix} \mathsf{J} & \mathsf{J} \\ \mathsf{A}_2 & \mathsf{J} \end{Bmatrix}\cdots\begin{Bmatrix} \mathsf{I} & \mathsf{A}_{N-1} \\ \mathsf{J} & \mathsf{I} \end{Bmatrix}\begin{bmatrix} \mathsf{J} & \mathsf{J} \\ \mathsf{A}_N & \mathsf{J} \end{bmatrix}\right). \tag{4}$$

Equation (4) gives a transfer matrix that adds two layers, e.g., layers 1 and 2, at a time. This transfer matrix has two parts. A diagonal matrix, an intra-two-layer matrix, adds the interactions within an added two-layer; e.g., the matrix containing A_1. The inter-two-layer matrix also has a special form. For example, the inter-two-layer matrix containing A_2 takes into account the interactions between the two-layer formed from one-layers 1 and 2 and the two-layer formed from one-layers 3 and 4. The question that will be addressed in this subsection is the relationship between the eigenvalues of the component one-layer transfer matrices and an M-layer transfer matrix. First, two lemmas that deal with the two-layer transfer matrix are proved.

Lemma 1: An inter-M-layer transfer matrix has the two equivalent forms

$$\begin{Bmatrix} \mathsf{J}_{n\times p} & \mathsf{J}_{n\times q} \\ \mathsf{A}_{m\times p} & \mathsf{J}_{m\times q} \end{Bmatrix}=\begin{Bmatrix} \mathsf{J}_{n\times p} \\ \mathsf{A}_{m\times p} \end{Bmatrix}\{\mathsf{I}_{p\times p} \quad \mathsf{J}_{p\times q}\}, \tag{5}$$

where the subscripts are used to show the size of the matrices.

Proof: The proof follows immediately from 5(3.8) and use of the associative law [5(3.3)]. ■

Lemma 2: The two-layer transfer matrix

$$\begin{Bmatrix} \mathsf{I} & \mathsf{B} \\ \mathsf{J} & \mathsf{I} \end{Bmatrix}\begin{Bmatrix} \mathsf{J} & \mathsf{J} \\ \mathsf{C} & \mathsf{J} \end{Bmatrix} \tag{6}$$

has the same nonzero eigenvalues as the matrix BC.

Proof: Using Lemma 1, the associative law for the row and column products [5(3.3)], and 5(3.6) gives

$$\begin{aligned}\begin{Bmatrix} \mathsf{I} & \mathsf{B} \\ \mathsf{J} & \mathsf{I} \end{Bmatrix}\begin{Bmatrix} \mathsf{J} & \mathsf{J} \\ \mathsf{C} & \mathsf{J} \end{Bmatrix}&=\left\{\begin{Bmatrix} \mathsf{I} \\ \mathsf{J} \end{Bmatrix}\begin{Bmatrix} \mathsf{B} \\ \mathsf{I} \end{Bmatrix}\right\}\begin{Bmatrix} \mathsf{J} \\ \mathsf{C} \end{Bmatrix}\{\mathsf{I} \quad \mathsf{J}\} \\ &=\left(\left(\begin{Bmatrix} \mathsf{I} \\ \mathsf{J} \end{Bmatrix}\mathsf{J}\right)\odot\left(\begin{Bmatrix} \mathsf{B} \\ \mathsf{I} \end{Bmatrix}\mathsf{C}\right)\right)\{\mathsf{I} \quad \mathsf{J}\} \\ &=\begin{Bmatrix} \mathsf{B} \\ \mathsf{I} \end{Bmatrix}\mathsf{C}\{\mathsf{I} \quad \mathsf{J}\}. \end{aligned} \tag{7}$$

However, for any two complex matrices $\mathsf{K}_{n\times m}$ and $\mathsf{L}_{m\times n}$ the nonzero eigenvalues of KL and LK are the same.[6] Thus the two-layer transfer matrix given by Eq. (6) has the same nonzero eigenvalues as

$$\{\mathsf{I} \quad \mathsf{J}\}\begin{Bmatrix} \mathsf{B} \\ \mathsf{I} \end{Bmatrix}\mathsf{C}=((\mathsf{IB})\odot(\mathsf{JI}))\mathsf{C}=\mathsf{BC}, \tag{8}$$

where 5(3.6) has been used. ■

If both B and C are $n\times n$, then the two-layer $n^2\times n^2$ transfer matrix given by Eq. (6) has at least n^2-n zero eigenvalues, with the remaining n eigenvalues the same as the product of the two one-layer transfer matrices BC. Theorem 1 generalizes Lemma 2 to the case of M-layer transfer matrices. The form of the M-layer transfer matrix given by Eq. (9) follows directly from the use of the associative laws for the row and column products and 5(3.13).

Theorem 1: The M-layer transfer matrix

$$\begin{Bmatrix} \mathsf{I} & \mathsf{A}_1 & \mathsf{J} & \cdots & \mathsf{J} & \mathsf{J} \\ \mathsf{J} & \mathsf{I} & \mathsf{A}_2 & \cdots & \mathsf{J} & \mathsf{J} \\ \mathsf{J} & \mathsf{J} & \mathsf{I} & \ddots & \mathsf{J} & \mathsf{J} \\ \vdots & \vdots & & \ddots & \ddots & \\ \mathsf{J} & \mathsf{J} & \mathsf{J} & \cdots & \mathsf{I} & \mathsf{A}_{M-1} \\ \mathsf{J} & \mathsf{J} & \mathsf{J} & \cdots & \mathsf{J} & \mathsf{I} \end{Bmatrix}\times\begin{Bmatrix} \mathsf{J} & \mathsf{J} & \cdots & \mathsf{J} & \mathsf{J} \\ \mathsf{J} & \mathsf{J} & \cdots & \mathsf{J} & \mathsf{J} \\ \mathsf{J} & \mathsf{J} & \cdots & \mathsf{J} & \mathsf{J} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathsf{J} & \mathsf{J} & \cdots & \mathsf{J} & \mathsf{J} \\ \mathsf{A}_M & \mathsf{J} & \cdots & \mathsf{J} & \mathsf{J} \end{Bmatrix} \tag{9}$$

has the same nonzero eigenvalues as $\mathsf{A}_1\mathsf{A}_2\cdots\mathsf{A}_{M-1}\mathsf{A}_M$.

Proof: For any diagonal matrices D_1 and D_2 the relationship $\mathsf{D}_1\mathsf{D}_2=\mathsf{D}_1\odot\mathsf{D}_2$ holds. Thus the intra-M-layer transfer matrix in Eq. (9) can be decomposed into $M-1$ diagonal matrices; one for each A_i. First multiply the diagonal matrix with A_{M-1} by the inter-M-layer transfer matrix, after using Lemma 1 to rewrite the inter-M-layer transfer matrix. This gives

$$\begin{aligned}&\begin{Bmatrix} \mathsf{I} & \mathsf{J} & \cdots & \mathsf{J} & \mathsf{J} \\ \mathsf{J} & \mathsf{I} & \cdots & \mathsf{J} & \mathsf{J} \\ \vdots & & \ddots & \vdots & \vdots \\ \mathsf{J} & \mathsf{J} & \cdots & \mathsf{I} & \mathsf{A}_{M-1} \\ \mathsf{J} & \mathsf{J} & \cdots & \mathsf{J} & \mathsf{I} \end{Bmatrix}\begin{Bmatrix} \mathsf{J} \\ \mathsf{J} \\ \vdots \\ \mathsf{J} \\ \mathsf{A}_M \end{Bmatrix}\{\mathsf{I} \quad \mathsf{J} \quad \cdots \quad \mathsf{J}\} \\ &=\begin{Bmatrix} \mathsf{J} \\ \vdots \\ \mathsf{J} \\ \mathsf{A}_{M-1} \\ \mathsf{I} \end{Bmatrix}\mathsf{A}_M\{\mathsf{I} \quad \mathsf{J} \quad \cdots \quad \mathsf{J}\}, \end{aligned} \tag{10}$$

where use has been made of 5(3.6) and the fact that the regular matrix product of a J matrix and a row product made up of one I and the rest J matrices gives back a J matrix. Next multiply both sides of Eq. (10) with the diagonal matrix containing A_{M-2} to give

$$\begin{Bmatrix} I & J & \cdots & J & J & J \\ J & I & \cdots & J & J & J \\ \vdots & & \ddots & \vdots & \vdots & \vdots \\ J & J & \cdots & I & A_{M-2} & J \\ J & J & \cdots & J & I & J \\ J & J & \cdots & J & J & I \end{Bmatrix} \begin{Bmatrix} J \\ J \\ \vdots \\ J \\ A_{M-1} \\ I \end{Bmatrix} \times A_M \{I \quad J \quad \cdots \quad J\}$$

$$= \begin{Bmatrix} J \\ \vdots \\ J \\ A_{M-2} \\ I \\ I \end{Bmatrix} A_{M-1} A_M \{I \quad J \quad \cdots \quad J\}. \tag{11}$$

Repeating this procedure for each of the other diagonal matrices in turn finally shows that the product of the two matrices in Eq. (9) is equal to

$$\begin{Bmatrix} A_1 \\ I \\ \vdots \\ I \end{Bmatrix} A_2 A_3 \cdots A_{M-1} A_M \{I \quad J \quad \cdots \quad J\}. \tag{12}$$

Using 5(3.6) gives

$$\{I \quad J \quad \cdots \quad J\} \begin{Bmatrix} A_1 \\ I \\ \vdots \\ I \end{Bmatrix} = A_1. \tag{13}$$

Since the nonzero eigenvalues of any two matrices $K_{n\times m} L_{m\times n}$ are the same as $L_{m\times n} K_{n\times m}$, the product of the M-layer transfer matrices in Eq. (9) has the same nonzero eigenvalues as $A_1 A_2 \cdots A_M$. ■

Theorem 2: The inter-M-layer transfer matrix

$$\begin{Bmatrix} J & J \\ C & J \end{Bmatrix} \tag{14}$$

has the same rank as the matrix C. If both C and the matrix of Eq. (14) are square, then the matrix of Eq. (14) has at most one nonzero eigenvalue.

Proof: For any nonsingular P, the rank of any matrix A satisfies[6] $\rho(PA) = \rho(A)$. Multiply the inter-M-layer transfer matrix of Eq. (14) by a particular permutation matrix (which is nonsingular) to give

$$\rho\left(\begin{Bmatrix} J_{n\times p} & J_{n\times q} \\ C_{m\times p} & J_{m\times q} \end{Bmatrix}\right) = \rho\left(\begin{Bmatrix} J_{m\times n} & I_{m\times m} \\ I_{n\times n} & J_{n\times m} \end{Bmatrix} \begin{Bmatrix} J_{n\times p} & J_{n\times q} \\ C_{m\times p} & J_{m\times q} \end{Bmatrix}\right) = \rho\left(\begin{Bmatrix} C_{m\times p} & J_{m\times q} \\ J_{n\times p} & J_{n\times q} \end{Bmatrix}\right). \tag{15}$$

The last matrix in Eq. (15) is just the Kronecker (direct) product of $C_{m\times p}$ and $J_{n\times q}$ [see 5(3.9)]. Use the well-known relationship[6] for the rank of a Kronecker product,

$$\rho\left(\begin{Bmatrix} C_{m\times p} & J_{m\times q} \\ J_{n\times p} & J_{n\times q} \end{Bmatrix}\right) = \rho(C_{m\times p})\rho(J_{n\times q}) = \rho(C_{m\times p}), \tag{16}$$

together with $\rho(J_{n\times q}) = 1$ to complete the proof of the first part of the theorem.

Use Lemma 1 and 5(3.6) to find that the nonzero eigenvalues of a square inter-M-layer matrix of Eq. (14) (with C also square) are the same as the nonzero eigenvalues of the matrix

$$\{I_{m\times m} \quad J_{m\times n}\} \begin{Bmatrix} J_{m\times n} \\ C_{n\times n} \end{Bmatrix} = (I_{m\times m} J_{m\times n}) \odot (J_{m\times n} C_{n\times n}) = J_{m\times n} C_{n\times n}. \tag{17}$$

However,[6]

$$\rho(J_{m\times n} C_{n\times n}) \leqslant \min(\rho(J_{m\times n}), \rho(C_{n\times n})) \leqslant 1,$$

since $\rho(J_{m\times n}) = 1$. Thus the square inter-M-layer matrix has at most one nonzero eigenvalue, since it has the same nonzero eigenvalues as a matrix that has a rank of at most 1. ■

The structure of the inter-M-layer square transfer matrix of size $n^M \times n^M$ of Eq. (14) (for the case where $C_{n\times n}$ is also square) can be easily illustrated. Use Lemma 1 and the properties 5(3.6) and 5(3.8) to show that the product of the matrix of Eq. (14) times itself is a matrix that has a rank equal to 1 [except in the trivial case where $\rho(CJ) = 0$]. Thus the inter-M-layer transfer matrix has a null space of dimension $n^M - n$, and has $n - 1$ generalized eigenvectors of rank 2 associated with the eigenvalue 0. To find the eigenvector associated with the nonzero eigenvalue, let e be a column vector with all n elements equal to 1, and define the vector $\mathbf{u} = C\mathbf{e}$. Then since $J = \mathbf{e}\mathbf{e}^T$, clearly $J\mathbf{e} = n\mathbf{e}$ and $J\mathbf{u} = \mathbf{e}\mathbf{e}^T C\mathbf{e} = s\mathbf{e}$, with $s = \Sigma_{ij} c_{ij}$ the sum of all the elements of C. Use Lemma 1 to give

$$\begin{Bmatrix} J \\ J \\ \vdots \\ J \\ C \end{Bmatrix} \{I \quad J \quad \cdots \quad J \quad J\} \begin{Bmatrix} \mathbf{e} \\ \mathbf{e} \\ \vdots \\ \mathbf{e} \\ \mathbf{u} \end{Bmatrix} = \begin{Bmatrix} J \\ J \\ \vdots \\ J \\ C \end{Bmatrix} s\mathbf{e} n^{M-2} = s n^{M-2} \begin{Bmatrix} \mathbf{e} \\ \mathbf{e} \\ \vdots \\ \mathbf{e} \\ \mathbf{u} \end{Bmatrix}. \tag{18}$$

Hence the n^M-dimensional vector in Eq. (18) is an unnormalized right eigenvector of the $n^M \times n^M$ inter-M-layer transfer matrix with eigenvalue sn^{M-2}. Similarly, if one defines the vector $\mathbf{v}^T = \mathbf{e}^T C$, then the left eigenvector associated with the eigenvalue sn^{M-2} is given by

$$\{\mathbf{v}^T \quad \mathbf{e}^T \quad \cdots \quad \mathbf{e}^T\}. \tag{19}$$

The generalized left eigenvectors of rank 2 are given by

$$\{\mathbf{e}^T \quad \cdots \quad \mathbf{e}^T \quad \mathbf{w}_i^T\}, \tag{20}$$

for the $n - 1$ orthogonal vectors $\mathbf{w}_i^T$ that have $\mathbf{w}_i^T C\mathbf{e} = 0$.

The generalized right eigenvectors of rank 2 are given by

$$\begin{Bmatrix} \mathbf{x}_i \\ \mathbf{e} \\ \vdots \\ \mathbf{e} \end{Bmatrix}, \tag{21}$$

where the $n-1$ orthogonal vectors $\mathbf{x}_i$ satisfy $\mathbf{e}^T C \mathbf{x}_i = 0$.

B. Longer-range interactions

This section deals with the case where layer i can interact with more than layers $i-1$ and $i+1$. However, it will be necessary to concentrate on the case of only pairwise interactions between the layers. If the system has N one-layers with one-layer i interacting with the $2R$ one-layers j with $i-R \leqslant j \leqslant i+R$, the partition function $\tilde{Z}$ is given by

$$\tilde{Z} = \mathrm{Tr}\left(\begin{Bmatrix} \mathsf{I} & \mathsf{A}_{1,1} & \mathsf{A}_{1,2} & \cdots & \mathsf{A}_{1,R} & \mathsf{J} & \mathsf{J} & \cdots & \mathsf{J} \\ \mathsf{J} & \mathsf{I} & \mathsf{A}_{2,1} & \cdots & \mathsf{A}_{2,R-1} & \mathsf{A}_{2,R} & \mathsf{J} & \cdots & \mathsf{J} \\ \mathsf{J} & \mathsf{J} & \mathsf{I} & & \mathsf{A}_{3,R-2} & \mathsf{A}_{3,R-1} & \mathsf{A}_{3,R} & & \mathsf{J} \\ & & & \ddots & & & & & \\ & & & & & \ddots & & & \\ & & & & & & & \ddots & \\ \mathsf{A}_{N,1} & \mathsf{A}_{N,2} & \mathsf{A}_{N,3} & \cdots & \mathsf{A}_{N,R} & \mathsf{J} & \cdots & \mathsf{J} & \mathsf{I} \end{Bmatrix}_{N,N} \right). \tag{22}$$

The notation $\{\cdots\}_{i,j}$ will be used to show that the matrix product has i columns of matrices and j rows of matrices. If some $\mathsf{A}_{k,l} = \mathsf{J}$, then one-layer k does not interact with one-layer l. Partition the row and column product in Eq. (22) into blocks with R matrices (when N/R is an integer and $R > 1$) to produce

$$\tilde{Z} = \mathrm{Tr}\left(\begin{Bmatrix} \mathsf{I} & \mathsf{A}_{1,1} & \mathsf{A}_{1,2} & \cdots & \mathsf{A}_{1,R-1} \\ \mathsf{J} & \mathsf{I} & \mathsf{A}_{2,1} & \cdots & \mathsf{A}_{2,R-2} \\ \mathsf{J} & \mathsf{J} & \mathsf{I} & & \mathsf{A}_{3,R-3} \\ \vdots & \vdots & & \ddots & \vdots \\ \mathsf{J} & \mathsf{J} & \mathsf{J} & \cdots & \mathsf{I} \end{Bmatrix}_{R,R} \begin{Bmatrix} \mathsf{A}_{1,R} & \mathsf{J} & \mathsf{J} & \cdots & \mathsf{J} \\ \mathsf{A}_{2,R-1} & \mathsf{A}_{2,R} & \mathsf{J} & \cdots & \mathsf{J} \\ \mathsf{A}_{3,R-2} & \mathsf{A}_{3,R-1} & \mathsf{A}_{3,R} & & \mathsf{J} \\ \vdots & \vdots & \vdots & \ddots & \\ \mathsf{A}_{R,1} & \mathsf{A}_{R,2} & \mathsf{A}_{R,3} & \cdots & \mathsf{A}_{R,R} \end{Bmatrix}_{R,R} \right.$$
$$\left. \times \cdots \begin{Bmatrix} \mathsf{I} & \mathsf{A}_{K+1,1} & \mathsf{A}_{K+1,2} & \cdots & \mathsf{A}_{K+1,R-1} \\ \mathsf{J} & \mathsf{I} & \mathsf{A}_{K+2,1} & \cdots & \mathsf{A}_{K+2,R-2} \\ \mathsf{J} & \mathsf{J} & \mathsf{I} & & \mathsf{A}_{K+3,R-3} \\ \vdots & \vdots & & \ddots & \vdots \\ \mathsf{J} & \mathsf{J} & \mathsf{J} & \cdots & \mathsf{I} \end{Bmatrix}_{R,R} \begin{Bmatrix} \mathsf{A}_{K+1,R} & \mathsf{J} & \mathsf{J} & \cdots & \mathsf{J} \\ \mathsf{A}_{K+2,R-1} & \mathsf{A}_{K+2,R} & \mathsf{J} & \cdots & \mathsf{J} \\ \mathsf{A}_{K+3,R-2} & \mathsf{A}_{K+3,R-1} & \mathsf{A}_{K+3,R} & & \mathsf{J} \\ \vdots & \vdots & \vdots & \ddots & \\ \mathsf{A}_{N,1} & \mathsf{A}_{N,2} & \mathsf{A}_{N,3} & \cdots & \mathsf{A}_{N,R} \end{Bmatrix}_{R,R} \right), \tag{23}$$

where $K = N - R$.

Lemma 3: If all the $\mathsf{A}_{i,j}$, I, and J are $n \times n$ square matrices, then

$$\begin{Bmatrix} \mathsf{I} & \mathsf{A}_{1,1} & \mathsf{A}_{1,2} & \cdots & \mathsf{A}_{1,R-1} \\ \mathsf{J} & \mathsf{I} & \mathsf{A}_{2,1} & \cdots & \mathsf{A}_{2,R-2} \\ \mathsf{J} & \mathsf{J} & \mathsf{I} & & \mathsf{A}_{3,R-3} \\ \vdots & \vdots & & \ddots & \vdots \\ \mathsf{J} & \mathsf{J} & \mathsf{J} & \cdots & \mathsf{I} \end{Bmatrix}_{R,R} \begin{Bmatrix} \mathsf{A}_{1,R} & \mathsf{J} & \mathsf{J} & \cdots & \mathsf{J} \\ \mathsf{A}_{2,R-1} & \mathsf{A}_{2,R} & \mathsf{J} & \cdots & \mathsf{J} \\ \mathsf{A}_{3,R-2} & \mathsf{A}_{3,R-1} & \mathsf{A}_{3,R} & & \mathsf{J} \\ \vdots & \vdots & \vdots & \ddots & \\ \mathsf{A}_{R,1} & \mathsf{A}_{R,2} & \mathsf{A}_{R,3} & \cdots & \mathsf{A}_{R,R} \end{Bmatrix}_{R,R}$$
$$= \mathsf{S}_R[\mathsf{A}_{1,1}\mathsf{A}_{1,2}\cdots\mathsf{A}_{1,R}]\mathsf{S}_R[\mathsf{A}_{2,1}\mathsf{A}_{2,2}\cdots\mathsf{A}_{2,R}]\cdots\mathsf{S}_R[\mathsf{A}_{R,1}\mathsf{A}_{R,2}\cdots\mathsf{A}_{R,R}], \tag{24}$$

with the special matrix product defined by

$$\mathsf{S}_R[\mathsf{A}_{i,1}\mathsf{A}_{i,2}\cdots\mathsf{A}_{i,R}] = \begin{Bmatrix} \mathsf{A}_{i,1} & \mathsf{A}_{i,2} & \cdots & \mathsf{A}_{i,R-1} & \mathsf{A}_{i,R} \\ \mathsf{I} & \mathsf{J} & \cdots & \mathsf{J} & \mathsf{J} \\ \mathsf{J} & \mathsf{I} & \cdots & \mathsf{J} & \mathsf{J} \\ \vdots & & \ddots & & \vdots \\ \mathsf{J} & \mathsf{J} & & \mathsf{I} & \mathsf{J} \end{Bmatrix}_{R,R}. \tag{25}$$

Proof: The left-hand matrix on the lhs of Eq. (24) is a diagonal matrix, and hence may be written as a product of $R-1$ diagonal matrices with diagonal matrix i made from a product of the $\mathsf{A}_{i,j}$ with $1 \leqslant j < R-i$. The right-hand matrix on the lhs of Eq. (24) can be decomposed into a product of $R-1$ diagonal matrices and R Kronecker products, which can then be multiplied together using 5(3.5). The equivalence of this decomposition follows directly from Appendix A of Ref. 7. Commuting each of the $R-1$ diagonal matrices from the decomposition of the left-hand matrix on the lhs of Eq. (24) as far as possible into the matrices from the decomposition of the right-hand matrix on the lhs of Eq. (24) gives R terms of the form

$$\begin{Bmatrix} \mathsf{I} & \mathsf{J} & \cdots & \mathsf{J} & \mathsf{J} & \cdots & \mathsf{J} \\ \mathsf{J} & \mathsf{I} & \cdots & \mathsf{J} & \mathsf{J} & \cdots & \mathsf{J} \\ & \cdots & \ddots & & & & \\ \mathsf{J} & \mathsf{J} & & \mathsf{I} & \mathsf{A}_{i,1} & \cdots & \mathsf{A}_{i,R-i} \\ \mathsf{J} & \mathsf{J} & \cdots & \mathsf{J} & \mathsf{I} & & \mathsf{J} \\ \vdots & \vdots & & & & \ddots & \\ \mathsf{J} & \mathsf{J} & \cdots & \mathsf{J} & \mathsf{J} & & \mathsf{I} \end{Bmatrix}_{R,R} \begin{Bmatrix} \mathsf{I} & & \mathsf{J} & \cdots & \mathsf{J} & \mathsf{J} & \cdots & \mathsf{J} \\ & \mathsf{J} & \mathsf{I} & \cdots & \mathsf{J} & \mathsf{J} & \cdots & \mathsf{J} \\ & & & \ddots & & & & \\ \mathsf{A}_{i,R-i+1} & \cdots & \cdots & \mathsf{A}_{i,R-1} & \mathsf{A}_{i,R} & & & \mathsf{J} \\ & \mathsf{J} & \mathsf{J} & \cdots & \mathsf{J} & \mathsf{I} & & \mathsf{J} \\ & \vdots & \vdots & & & & \ddots & \\ & \mathsf{J} & \mathsf{J} & \cdots & \mathsf{J} & \mathsf{J} & & \mathsf{I} \end{Bmatrix}_{R,R} . \tag{26}$$

Each of the R products of two matrices with the form of Eq. (26) can be multiplied together to obtain R matrices with a form given in Eq. (27). This yields that the lhs of Eq. (24) is equal to

$$\begin{Bmatrix} \mathsf{A}_{1,R} & \mathsf{A}_{1,1} & \mathsf{A}_{1,2} & \cdots & \mathsf{A}_{1,R-1} \\ \mathsf{J} & \mathsf{I} & \mathsf{J} & \cdots & \mathsf{J} \\ \mathsf{J} & \mathsf{J} & \mathsf{I} & & \mathsf{J} \\ \vdots & \vdots & & \ddots & \\ \mathsf{J} & \mathsf{J} & \mathsf{J} & & \mathsf{I} \end{Bmatrix}_{R,R} \begin{Bmatrix} \mathsf{I} & \mathsf{J} & \mathsf{J} & \cdots & \mathsf{J} \\ \mathsf{A}_{2,R-1} & \mathsf{A}_{2,R} & \mathsf{A}_{2,1} & \cdots & \mathsf{A}_{2,R-2} \\ \mathsf{J} & \mathsf{J} & \mathsf{I} & & \mathsf{J} \\ \vdots & \vdots & & \ddots & \\ \mathsf{J} & \mathsf{J} & \mathsf{J} & & \mathsf{I} \end{Bmatrix}_{R,R}$$

$$\times \cdots \begin{Bmatrix} \mathsf{I} & \mathsf{J} & \cdots & \mathsf{J} & \mathsf{J} \\ \mathsf{J} & \mathsf{I} & \cdots & \mathsf{J} & \mathsf{J} \\ & & \ddots & & \\ \mathsf{A}_{R-1,2} & \mathsf{A}_{R-1,3} & \cdots & \mathsf{A}_{R-1,R} & \mathsf{A}_{R-1,1} \\ \mathsf{J} & \mathsf{J} & \cdots & \mathsf{J} & \mathsf{I} \end{Bmatrix}_{R,R} \begin{Bmatrix} \mathsf{I} & \mathsf{J} & \cdots & \mathsf{J} & \mathsf{J} \\ \mathsf{J} & \mathsf{I} & & \mathsf{J} & \mathsf{J} \\ \vdots & & \ddots & & \vdots \\ \mathsf{J} & \mathsf{J} & & \mathsf{I} & \mathsf{J} \\ \mathsf{A}_{R,1} & \mathsf{A}_{R,2} & \cdots & \mathsf{A}_{R,R-1} & \mathsf{A}_{R,R} \end{Bmatrix}_{R,R} . \tag{27}$$

Define the permutation matrix P, which permutes matrices within the Kronecker product, i.e., which has the form

$$\mathsf{P} = \begin{Bmatrix} \mathsf{J} & \mathsf{J} & \cdots & \mathsf{J} & \mathsf{I} \\ \mathsf{I} & \mathsf{J} & \cdots & \mathsf{J} & \mathsf{J} \\ \mathsf{J} & \mathsf{I} & & \mathsf{J} & \mathsf{J} \\ \vdots & & \ddots & & \vdots \\ \mathsf{J} & \mathsf{J} & & \mathsf{I} & \mathsf{J} \end{Bmatrix}_{R,R} . \tag{28}$$

This type of permutation matrix has been extensively studied in the mathematical literature.[8,9] Insert the identity in the form $\mathsf{P}\mathsf{P}^{-1}$ (with $\mathsf{P}^{-1} = \mathsf{P}^{T}$) between each of the R matrices in Eq. (27). The leftmost matrix of Eq. (27) multiplied by P gives $\mathsf{S}_R[\mathsf{A}_{1,1}\cdots\mathsf{A}_{1,R}]$, while P^{-1} multiplied by the rightmost matrix of Eq. (27) gives $\mathsf{S}_R[\mathsf{A}_{R,1}\cdots\mathsf{A}_{R,R}]$. Similarly, P^{-1} times the matrix in the middle of Eq. (27) formed from $\mathsf{A}_{i,1}\cdots\mathsf{A}_{i,R}$ times P gives $\mathsf{S}_R[\mathsf{A}_{i,1}\cdots\mathsf{A}_{i,R}]$. ■

Lemma 4: If $\mathsf{D}_{f\times f}$ is a diagonal matrix, and the matrices A, B, and C are general matrices of the indicated size, then

$$\begin{Bmatrix} \mathsf{A}_{q\times f} & \mathsf{B}_{q\times q} \\ \mathsf{D}_{f\times f} & \mathsf{C}_{f\times q} \end{Bmatrix} = \begin{Bmatrix} \mathsf{A}_{q\times f} & \mathsf{B}_{q\times q} \\ \mathsf{I}_{f\times f} & \mathsf{J}_{f\times q} \end{Bmatrix} \begin{Bmatrix} \mathsf{D}_{f\times f} & \mathsf{C}_{f\times q} \\ \mathsf{J}_{q\times f} & \mathsf{I}_{q\times q} \end{Bmatrix} . \tag{29}$$

Proof: The last matrix on the rhs of Eq. (29) is a diagonal matrix. Use 5(3.3) and 5(3.6) to show that

$$\{\mathsf{D}_{f\times f} \quad \mathsf{C}_{f\times q}\} = \{\mathsf{I}_{f\times f} \quad \mathsf{J}_{f\times q}\} \begin{Bmatrix} \mathsf{D}_{f\times f} & \mathsf{C}_{f\times q} \\ \mathsf{J}_{q\times f} & \mathsf{I}_{q\times q} \end{Bmatrix} . \tag{30}$$

Then use the transpose of 5(3.7) (with $\mathsf{Q} = \mathsf{I}$) and the fact that $\mathsf{D}_{f\times f}$ is a diagonal matrix to produce the rhs of Eq. (29). ■

Lemma 5: If $\mathsf{A}_{i,j}$, I, and J are all square matrices of the same size, then the matrix

$$\begin{Bmatrix} \mathsf{I} & \mathsf{A}_{1,1} & \mathsf{A}_{1,2} & \cdots & \mathsf{A}_{1,R} \\ \mathsf{J} & \mathsf{I} & \mathsf{A}_{2,1} & \cdots & \mathsf{A}_{2,R-1} \\ \mathsf{J} & \mathsf{J} & \mathsf{I} & \cdots & \mathsf{A}_{3,R-2} \\ \vdots & & & \ddots & \vdots \\ \mathsf{J} & \mathsf{J} & \mathsf{J} & & \mathsf{I} \end{Bmatrix}_{R+1,R+1} \begin{Bmatrix} \mathsf{J} & \mathsf{J} & \cdots & \mathsf{J} \\ \mathsf{A}_{2,R} & \mathsf{J} & \cdots & \mathsf{J} \\ \mathsf{A}_{3,R-1} & \mathsf{A}_{3,R} & & \mathsf{J} \\ \vdots & & \ddots & \vdots \\ \mathsf{A}_{R+1,1} & \cdots & \mathsf{A}_{R+1,R} & \mathsf{J} \end{Bmatrix}_{R+1,R+1} \tag{31}$$

has the same nonzero eigenvalues as

$$\mathsf{S}_R[\mathsf{A}_{1,1}\cdots\mathsf{A}_{1,R}]\mathsf{S}_R[\mathsf{A}_{2,1}\cdots\mathsf{A}_{2,R}]\cdots\mathsf{S}_R[\mathsf{A}_{R+1,1}\cdots\mathsf{A}_{R+1,R}] . \tag{32}$$

Proof: Use 5(3.3) and 5(3.6), and the fact that the product of the leftmost column of the left-hand matrix product of Eq. (31) times the top row of the right-hand matrix product of Eq. (31) gives a J matrix, to show that this row and column may be eliminated from the products. Then use 5(3.8) to make the rightmost matrix in Eq. (31) into a matrix formed from a product of $\mathsf{A}_{i,j}$ matrices and one formed from the product of only I and J matrices. Finally, use the transpose of 5(3.7) on the rightmost matrix of Eq. (31) (with $\mathsf{Q} = \mathsf{I}$) to show that the product of the two matrices in Eq. (31) is equal to

$$\begin{Bmatrix} A_{1,1} & A_{1,2} & \cdots & A_{1,R-1} & A_{1,R} \\ I & J & \cdots & J & J \\ & \ddots & & \vdots & \vdots \\ J & J & \ddots & J & J \\ J & J & & I & J \\ J & J & \cdots & J & I \end{Bmatrix}_{R+1,R} \begin{Bmatrix} I & A_{2,1} & \cdots & A_{2,R-2} & A_{2,R-1} \\ J & I & & A_{3,R-3} & A_{3,R-2} \\ \vdots & & \ddots & & \\ J & J & & I & A_{R,1} \\ J & J & \cdots & J & I \end{Bmatrix}_{R,R}$$

$$\times \begin{Bmatrix} A_{2,R} & J & \cdots & J \\ A_{3,R-1} & A_{3,R} & & J \\ \vdots & \vdots & \ddots & \\ A_{R,2} & A_{R,3} & & J \\ A_{R+1,1} & A_{R+1,2} & \cdots & A_{R+1,R} \end{Bmatrix}_{R,R} \begin{Bmatrix} I & J & \cdots & J & J \\ J & I & & J & J \\ \vdots & & \ddots & & \vdots \\ J & J & \cdots & I & J \end{Bmatrix}_{R,R+1} . \tag{33}$$

From Lemma 3 the middle two matrices in Eq. (33) can be written as the product of the R matrices $S_R[A_{2,1}\cdots A_{2,R}]\cdots S_R[A_{R+1,1}\cdots A_{R+1,R}]$. Use 5(3.6) to see that the rightmost matrix in Eq. (33) times the leftmost matrix in Eq. (33) gives $S_R[A_{1,1}\cdots A_{1,R}]$. To complete the proof, use the fact that for any two matrices that have appropriate sizes, the nonzero eigenvalues of AB are the same as the nonzero eigenvalues of BA. ■

Next Lemma 5 is generalized to allow for the length of layering of the transfer matrix to be any integer greater than R.

Theorem 3: If all $A_{i,j}$, I, and J are square $n\times n$ matrices, then the matrix

$$\begin{Bmatrix} I & A_{1,1} & A_{1,2} & \cdots & & & & & \\ J & I & A_{2,1} & & & & & & \\ J & J & I & & & & & & \\ \vdots & \vdots & & \ddots & & & & & \\ J & J & J & & I & A_{S,1} & & A_{S,R-1} & A_{S,R} \\ J & J & J & \cdots & J & I & & A_{S+1,R-2} & A_{S+1,R-1} \\ \vdots & \vdots & \vdots & & \vdots & & \ddots & & \\ J & J & J & \cdots & J & J & & I & A_{K-1,1} \\ J & J & J & \cdots & J & J & & J & I \end{Bmatrix}_{K,K}$$

$$\times \begin{Bmatrix} J & J & \cdots & J & J & J & \cdots & J \\ J & J & \cdots & J & J & J & \cdots & J \\ J & J & \cdots & J & J & J & \cdots & J \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ J & J & \cdots & J & J & J & \cdots & J \\ A_{S+1,R} & J & \cdots & J & J & J & \cdots & J \\ \vdots & & & & \vdots & \vdots & & \vdots \\ A_{K-1,2} & A_{K-1,3} & \cdots & A_{K-1,R} & J & J & \cdots & J \\ A_{K,1} & A_{K,2} & \cdots & A_{K,R-1} & A_{K,R} & J & \cdots & J \end{Bmatrix}_{K,K} \tag{34}$$

has the same nonzero eigenvalues as

$$S_R[A_{1,1}\cdots A_{1,R}]S_R[A_{2,1}\cdots A_{2,R}]\cdots S_R[A_{K,1}\cdots A_{K,R}], \tag{35}$$

where $K = S + R$, with $S \geqslant 0$ and $R > 1$. The remaining $n^K - n^R$ eigenvalues are equal to zero.

Proof: When $S = 0$ this is Lemma 3, and when $S = 1$ this is Lemma 5. Thus assume that $S \geqslant 2$. As done in the proof of Lemma 5, eliminate the leftmost column of the matrix on the lhs of Eq. (34) and the top row of the matrix on the rhs of Eq. (34). Next use the fact that the matrices in Eq. (31) equal those in Eq. (33) to break the two matrices into four matrices with the form given in Eq. (33) [with the range in Lemma 5 set to $K - 1$, and some of the arbitrary $A_{i,r}$ matrices in Eq. (33) are equal to J matrices]. Define these four matrices to be

$$\{C\}_{K,K-1}\{D\}_{K-1,K-1}\{E\}_{K-1,K-1}\{F\}_{K-1,K}, \tag{36}$$

where the forms of the four matrices are the same as the matrices in the corresponding positions in Eq. (33). For example, this has made the definition (with $j > i$)

$$\{\mathsf{F}\}_{i,j}=\begin{Bmatrix} \mathsf{I} & \mathsf{J} & \cdots & \mathsf{J} & \mathsf{J} & \cdots & \mathsf{J} \\ \mathsf{J} & \mathsf{I} & & \mathsf{J} & \mathsf{J} & \cdots & \mathsf{J} \\ \vdots & & \ddots & & \vdots & & \vdots \\ \mathsf{J} & \mathsf{J} & & \mathsf{I} & \mathsf{J} & \cdots & \mathsf{J} \end{Bmatrix}_{i,j}. \tag{37}$$

The product $\{\mathsf{D}\}_{K-1,K-1}\{\mathsf{E}\}_{K-1,K-1}$ again has the same form as Eq. (31) (with the range equal to $K-2$), so Lemma 5 can be applied, and these two matrices can be broken into the product of four matrices with the form of Eq. (33). This process can be iterated until the center matrices are $\{\mathsf{D}\}_{R,R}\{\mathsf{E}\}_{R,R}$. This means that Eq. (34) is equal to

$$\{\mathsf{C}\}_{K,K-1}\{\mathsf{C}\}_{K-1,K-2}\cdots\{\mathsf{C}\}_{R+1,R}\{\mathsf{D}\}_{R,R}\{\mathsf{E}\}_{R,R}\{\mathsf{F}\}_{R,R+1}\{\mathsf{F}\}_{R+1,R+2}\cdots\{\mathsf{F}\}_{K-1,K}. \tag{38}$$

Use of 5(3.6) and noting which matrices are equal to J matrices shows that $\{\mathsf{F}\}_{i,j}\{\mathsf{F}\}_{j,k}=\{\mathsf{F}\}_{i,k}$, so

$$\{\mathsf{F}\}_{R,R+1}\{\mathsf{F}\}_{R+1,R+2}\cdots\{\mathsf{F}\}_{K-1,K}=\{\mathsf{F}\}_{R,K}. \tag{39}$$

Use Lemma 3 on the matrices $\{\mathsf{D}\}_{R,R}\{\mathsf{E}\}_{R,R}$, and the fact that the nonzero eigenvalues of KL for any two compatible matrices are the same as LK, to find that the nonzero eigenvalues of Eq. (34) are equal to the nonzero eigenvalues of

$$\{\mathsf{F}\}_{R,K}\{\mathsf{C}\}_{K,K-1}\{\mathsf{C}\}_{K-1,K-2}\cdots\{\mathsf{C}\}_{R+1,R}\mathsf{S}_R[\mathsf{A}_{S+1,1}\cdots\mathsf{A}_{S+1,R}]\mathsf{S}_R[\mathsf{A}_{S+2,1}\cdots\mathsf{A}_{S+2,R}]\cdots\mathsf{S}_R[\mathsf{A}_{K,1}\cdots\mathsf{A}_{K,R}]. \tag{40}$$

Make use of 5(3.3) and 5(3.6) to give (for $l>R$)

$$\{\mathsf{F}\}_{R,l}\{\mathsf{C}\}_{l,l-1}=\mathsf{S}_R[\mathsf{A}_{K-l+1,1}\cdots\mathsf{A}_{K-l+1,R}]\{\mathsf{F}\}_{R,l-1}, \tag{41}$$

where $\{\mathsf{F}\}_{R,R}=\mathsf{I}$. Finally, use Eq. (41) S times to perform the multiplication between the $\{\mathsf{F}\}_{R,K}$ matrix and the $\{\mathsf{C}\}_{l,l-1}$ matrices to complete the proof. ■

Lemma 3, Lemma 5, and Theorem 3 all require that the sizes of all the matrices in the product be the same. This was done to avoid too clumsy a notation in the proofs presented. However, the results can be generalized to remove this restriction. This generalization can be done in two ways. One way is to "pad" the $\mathsf{A}_{i,j}$ matrices with zeros so that they are all square and of the same size. This procedure will leave the physically important quantities such as the partition function invariant. The other way of generalizing the result when the matrices may have different dimensions is to use rectangular transfer matrices. In this case, Lemmas 3 and 5 and Theorem 3 can be generalized, but with the notation becoming slightly more cumbersome. For example, if layer i has n_i states and interacts with R other layers, then the corresponding matrix to Eq. (25) is

$$\begin{Bmatrix} \mathsf{A}^{(1,1)}_{n_1\times n_2} & \mathsf{A}^{(1,2)}_{n_1\times n_3} & \cdots & \mathsf{A}^{(1,R-1)}_{n_1\times n_{R-1}} & \mathsf{A}^{(1,R)}_{n_1\times n_R} \\ \mathsf{I}_{n_2\times n_2} & \mathsf{J}_{n_2\times n_3} & \cdots & \mathsf{J}_{n_2\times n_{R-1}} & \mathsf{J}_{n_2\times n_R} \\ \mathsf{J}_{n_3\times n_2} & \mathsf{I}_{n_3\times n_3} & & \mathsf{J}_{n_3\times n_{R-1}} & \mathsf{J}_{n_3\times n_R} \\ \vdots & & \ddots & & \vdots \\ \mathsf{J}_{n_R\times n_2} & \mathsf{J}_{n_R\times n_3} & & \mathsf{I}_{n_R\times n_{R-1}} & \mathsf{J}_{n_R\times n_R} \end{Bmatrix}, \tag{42}$$

where subscripts denote the matrix size and superscripts denote the matrix of pairwise interactions.

A more serious restriction is the restriction to pairwise interactions between the layers. However, in one of the most studied models in statistical mechanics, the Ising model, this restriction does not enter. This is because it has been shown by Wegner[10] that for the Ising model, multispin interactions can be rewritten into single-spin and pairwise interactions by adding additional spins.Thus if all matrices are 2×2 and symmetric (the Ising model), the restriction to pairwise interactions is not important. Unfortunately, for other models there is no general argument to overcome this restriction.

III. DISCUSSION AND CONCLUSIONS

The main result of this paper is Theorem 3, which has Theorem 1 as a special case. Theorem 3 shows that *all* the eigenvalues of transfer matrices formed by different layering prescriptions are the same, except that different prescriptions have different numbers of zero eigenvalues. This is important because it shows that assigning a physical significance to "constrained" free energies (which are proportional to the natural logarithm of nonzero eigenvalues) is a mathematically well-defined procedure. For example, the spinodal behavior of long-range Ising models was seen to correspond to the behavior of some of the smallest eigenvalues of the transfer matrix in Ref. 7. In addition, since the expectation values of operators are given by equations of the form

$$\langle O\rangle=\mathrm{Tr}(\mathsf{O}\mathsf{A}_1\mathsf{A}_2\cdots\mathsf{A}_N)/Z, \tag{43}$$

Theorem 3 also shows that the spectra for such operators is independent of the layering prescription used. These results hold for all models, but may be of particular importance for models that are conformally invariant.[11] It is worthwhile to note that Theorem 3 is easily generalized to the case where the layers have different numbers of states. Also, the restriction in Theorem 3 to layers that interact pairwise is not necessary if each layer has only two states. Thus if one regards a layer as a single spin, the restriction to pairwise interactions between the spins is not necessary for the Ising model.

Theorem 2 concerns an interesting result that may be useful in numerically computing the partition function for various models. In particular, if one uses a layering with more than one layer, then the spectral decomposition of the inter-M-layer transfer matrix is easy to perform. There is

only one nonzero eigenvalue, and its associated eigenvectors are easy to calculate [see Eqs. (18) and (19)]. The generalized eigenvectors of rank 2 used can be *any* vectors given by Eqs. (20) and (21), i.e., *any* set of vectors in $n-1$ dimensions that are orthogonal to the single eigenvector that has nonzero eigenvalue. Thus to compute the spectral decomposition of the inter-M-layer matrix it is necessary only to get an orthogonal basis that includes the single eigenvector corresponding to the nonzero eigenvalue. This is to be compared with the case where the spectral decomposition would have to contain a unique orthonormal basis if the matrix were to have all eigenvalues nondegenerate.

The matrix defined by Eq. (25) adds a single layer at a time when there are longer-range pairwise interactions. When this is implemented with each layer used for a single spin, this gives the statistical mechanical model with "screw" boundary conditions. For example, when all $\mathsf{A}_{i,j}=\mathsf{J}$ except for $j=1$ and $j=R$, multiplying N such matrices gives a square lattice of size $R\times N$. Such boundary conditions were introduced by Kramers and Wannier[4] in 1941, and were used initially in the calculation of the exact partition function of the two-dimensional Ising model.[12–14] Screw boundary conditions in two and more dimensions have been used also in numerical calculations for finite strip widths for the two- and three-dimensional uniform Ising model,[15,16] as well as for the random Ising model in two and three dimensions.[17] Equation (25) gives this sparse matrix in the general case. If all layers are of size n, then the $n^R\times n^R$ matrix of Eq. (25) has only n^{R+1} nonzero elements. This is a property that is extremely useful in numerical calculations, where the sizes of finite strips that can be studied are limited by the storage of the digital computer. Use of the properties of the row and column products also allows various relationships of the matrix of Eq. (25) to be seen. For example, it is easy to break this matrix into the product of the permutation matrix defined in Eq. (28) times the product of a diagonal matrix times the product of a direct product matrix. (For the two-dimensional Ising model, this was done in Ref. 4.) Another property that can be utilized in numerical calculations of transfer matrices comes from the relation

$$\begin{Bmatrix}\mathsf{A}&\mathsf{J}&\mathsf{B}\\ \mathsf{I}&\mathsf{J}&\mathsf{J}\\ \mathsf{J}&\mathsf{I}&\mathsf{J}\end{Bmatrix}\begin{Bmatrix}\mathsf{I}&\mathsf{J}&\mathsf{J}\\ \mathsf{J}&\mathsf{R}&\mathsf{J}\\ \mathsf{J}&\mathsf{J}&\mathsf{I}\end{Bmatrix}=\begin{Bmatrix}\mathsf{I}&\mathsf{J}&\mathsf{J}\\ \mathsf{J}&\mathsf{I}&\mathsf{J}\\ \mathsf{J}&\mathsf{J}&\mathsf{R}\end{Bmatrix}\begin{Bmatrix}\mathsf{A}&\mathsf{J}&\mathsf{B}\\ \mathsf{I}&\mathsf{J}&\mathsf{J}\\ \mathsf{J}&\mathsf{I}&\mathsf{J}\end{Bmatrix}. \tag{44}$$

Equation (44) ilustrates a property of the single-particle transfer matrix [Eq. (25)] for a square lattice with screw boundary conditions. (The property can be easily generalized to systems in different dimensions.) Equation (44) shows that whenever a layer is not interacting with an intermediate layer j, it is possible for *any* matrix R to find two matrices formed from Kronecker products of I and R that multiply the single-layer transfer matrix S to give back the same matrix S. That is, $\mathsf{ST}=\hat{\mathsf{T}}\mathsf{S}$. Here T is a Kronecker product containing R in position j, and $\hat{\mathsf{T}}$ is a Kronecker product containing R in position $j+1$. The other matrices in the Kronecker products are identity matrices, I. Multiply Eq. (44) by the right eigenvector of T, $\mathsf{T}\mathbf{v}_i=\lambda_i\mathbf{v}_i$, to give $\hat{\mathsf{T}}\mathsf{S}\mathbf{v}_i=\lambda_i\mathsf{S}\mathbf{v}_i$. Thus $\mathsf{S}\mathbf{v}_i$ is a right eigenvector of $\hat{\mathsf{T}}$ with eigenvalue λ_i. If it is easy to calculate the eigenvectors and eigenvalues of the matrices T and $\hat{\mathsf{T}}$, one computational advantage that can be used in numerical calculations is to start with a complete set of eigenvectors of T, and then multiply by S to transfer to the complete set of eigenvectors of $\hat{\mathsf{T}}$. In most cases, it is possible to make both of these sets the same. Then program the "easy" rules to multiply $\mathbf{v}_i$ by S, since if A has n states, then $\mathsf{S}\mathbf{v}_i$ is a linear combination of at most n^2 vectors. This property should be particularly useful in a random system. For example, in the random Ising model studied in Ref. 17, it is possible to multiply S by *all* $\mathbf{v}_i$ with approximately the same computational effort and memory requirements that it takes to multiply S by a single arbitrary vector. Of course this result is trivial if $\mathsf{R}=\mathsf{I}$, but the numerical advantage that should be utilized is that Eq. (44) holds for any R.

ACKNOWLEDGMENTS

The author acknowledges useful conversations with Hans Herrmann and John de Pillis.

[1] W. J. Camp and M. E. Fisher, Phys. Rev. B **6**, 946 (1972).
[2] E. H. Lieb, in *Lectures in Theoretical Physics*, edited by K. T. Mahanthappa and W. E. Brittin (Gordon and Breach, New York, 1969), Vol. XI-D.
[3] E. W. Montroll, Ann. Math. Stat. **18**, 18 (1947).
[4] H. A. Kramers and G. H. Wannier, Phys. Rev. **60**, 252 (1941).
[5] M. A. Novotny, J. Math. Phys. **20**, 1146 (1979).
[6] M. Marcus and H. Minc, *A Survey of Matrix Theory and Matrix Inequalities* (Allyn and Bacon, Boston, 1964).
[7] M. A. Novotny, W. Klein, and P. A. Rikvold, Phys. Rev. B **33**, 7729 (1986).
[8] H. V. Henderson and S. R. Searle, Linear Multilinear Algebra **9**, 271 (1981).
[9] B. Holmquist, Linear Multilinear Algebra **17**, 117 (1985).
[10] F. J. Wegner, J. Math. Phys. **12**, 2259 (1971).
[11] J. L. Cardy, in *Phase Transitions and Critical Phenomena*, Vol. 11, edited by C. Domb and J. L. Leibowitz (Academic, London, 1987), and references therein.
[12] G. F. Newell and E. W. Montroll, Rev. Mod. Phys. **25**, 353 (1953).
[13] T. Oguchi, J. Phys. Soc. Jpn. **5**, 75 (1950).
[14] T. Oguchi and Y. Taguchi, Prog. Theor. Phys. Suppl. **87**, 23 (1986).
[15] S. Gartenhaus, Phys. Rev. B **27**, 1698 (1983).
[16] N. H. Fuchs and S. Gartenhaus, Phys. Rev. B **31**, 7261 (1985).
[17] T. Oguchi and Y. Taguchi, Prog. Theor. Phys. **77**, 775 (1987).

Translating the spin-boson model into a classical system

M. Fannes[a] and B. Nachtergaele[b]
Instituut voor Theoretische Fysica, Universiteit Leuven, B-3030 Leuven, Belgium

(Received 19 April 1988; accepted for publication 15 June 1988)

It is shown rigorously how to translate the ground state energy problem for some quantum systems into an equilibrium problem for an associated one-dimensional classical system. The case of the spin-boson model for which the connection is established with a previously obtained Ising model over $\mathbb{R}$ is explicitly treated.

I. INTRODUCTION

In order to study the ground state of some specific quantum systems it can be useful to translate this problem into the computation of the partition function of an associated classical system.[1-3] The ground state is approached as a low-temperature limit of equilibrium states for the quantum system and it is this limit that is turned into a thermodynamic limit of a one-dimensional classical system at a fixed temperature.

Our contribution consists in a rigorous study of the translation to the classical system. For this purpose we start from a convergent series expansion of the free energy of the quantum system. This expansion is in terms of multitime correlation functions up to imaginary time $i\beta$ and is based on a general stability result for equilibrium states. Such a method is an alternative for the usual transformations based on path-integral techniques.[1-3] In the case of the spin-boson model the latter techniques yielded an Ising model over $\mathbb{R}$. In this paper the spin-boson model will be treated by the perturbation series approach by which we will get a classical particle model on $\mathbb{R}$. Furthermore, we will prove that this model can be transformed into a continuous Ising model as found in the stochastic approach.

In the following sections we consider successively the following points. First, in Sec. II, we study the series expansion for the free energy based on the perturbation theory for cyclic vectors representing equilibrium states. Section III is devoted to the spin-boson model. We use earlier results to compute explicitly the multitime correlation functions appearing in the series expansion of Sec. II; their properties will enable us to construct an associated classical particle model. Finally, we connect in Sec. IV this classical particle model to the Ising model over $\mathbb{R}$ which was obtained in Ref. 4.

II. A PERTURBATION EXPANSION FOR THE PARTITION FUNCTION

We describe here a quantum mechanical model as a W^*-dynamical system. A general reference for such a description is Refs. 5 and 6. The observables of the system form a von Neumann algebra $\mathcal{M}$ with normalized cyclic and separating vector Ω^0. The free dynamics $\{\alpha_t^0 | t\in\mathbb{R}\}$ is the modular automorphism group associated with the pair $(\mathcal{M},\Omega^0)$. This means that the state $x\in\mathcal{M}\mapsto\langle\Omega^0|x\Omega^0\rangle$ of $\mathcal{M}$ is a KMS state (i.e., a state satisfying the Kubo–Martin–Schwinger equilibrium condition, see Ref. 6) at inverse temperature $\beta = 1$ for the time evolution $\{\alpha_t^0 | t\in\mathbb{R}\}$.

Any self-adjoint element P in $\mathcal{M}$ defines a perturbed dynamics $\{\alpha_t^P | t\in\mathbb{R}\}$ of $\mathcal{M}$ with generator $\delta_0(\cdot) + i[P,\cdot]$, where δ_0 is the generator of $\{\alpha_t^0 | t\in\mathbb{R}\}$. It is an important stability result in the study of equilibrium states that any such perturbed dynamics allows for a perturbed KMS state. To state the result we first introduce some notation: for $n\geqslant 1$ the vectors

$$\Omega_n^P(x_1,\ldots,x_n) \equiv \alpha_{x_1}^0(P)\cdots\alpha_{x_n}^0(P)\Omega^0, \quad x_i\in\mathbb{R}, \tag{1}$$

have an analytic extension $\Omega_n^P(z_1,\ldots,z_n)$ to the domain

$$D_n = \{(z_1,\ldots,z_n) | 0 < \operatorname{Im} z_1 < \cdots < \operatorname{Im} z_n \tfrac{1}{2}\},$$

which is continuous and bounded on $\overline{D_n}$ with bound[7,8]

$$\|\Omega_n^P(z_1,\ldots,z_n)\| \leqslant \|P\|^n. \tag{2}$$

The perturbed KMS state is now given by the cyclic vector Ω^P where (Ref. 6, Theorem 5.4.4)

$$\Omega^P = \Omega^0 + \sum_{n\geqslant 1} (-1)^n \times \int_{0<s_1<\cdots<s_n<1/2} \cdots \int ds_1\cdots ds_n\, \Omega_n^P(is_1,\ldots,is_n). \tag{3}$$

In general, Ω^P will not be normalized; in fact it can easily be seen for finite systems that $-\log\|\Omega^P\|^2$ is the correction to the free energy due to the perturbation P. The aim of this section is to prove a simple expansion for $\|\Omega^P\|^2$.

Lemma 2.1: For $n\geqslant 2$, let

$$f\in\mathcal{L}^1(\{(x_1,\ldots,x_{n-1}) | 0\leqslant x_1\leqslant\cdots\leqslant x_{n-1}\leqslant 1\}, dx_1\cdots dx_{n-1})$$

then

$$\begin{aligned} A &\equiv \int_{0<s_1<\cdots<s_n<1} \cdots \int ds_1\cdots ds_n\, f(s_1,\ldots,s_{n-1}) \\ &= \int_{0<v_1<\cdots<v_n<1/2} \cdots \int dv_1\cdots dv_n\, f(v_2-v_1,\ldots,v_n-v_1) \\ &\quad + \sum_{k=1}^{n-1} \int_{0<u_1<\cdots<u_k<1/2} \cdots \int du_1\cdots du_k \\ &\quad \times \int_{0<v_1<\cdots<v_{n-k}<1/2} dv_1\cdots dv_{n-k} \\ &\quad \times f(u_k-u_{k-1},\ldots,u_k-u_1,u_k+v_1,\ldots,u_k+v_{n-k}) \end{aligned}$$

[a] Bevoegdverklaard Navorser N.F.W.O., Belgium.
[b] Onderzoeker I.I.K.W., Belgium.

 0022-2488/88/102288-06$02.50

$$+\int_{0<u_1<\cdots<u_n<1/2}\cdots\int du_1\cdots du_n$$
$$\times f(u_n-u_{n-1},\ldots,u_n-u_1)$$
$$\equiv\sum_{k=0}^{n}B_k.$$

Proof: Define the following subsets of $\mathbb{R}^{n-1}$:

$$\Lambda^0=\{(s_1,\ldots,s_{n-1})\,|\,\tfrac{1}{2}\leqslant s_1\leqslant\cdots\leqslant s_{n-1}\leqslant 1\},\tag{4}$$
$$\Lambda^n=\{(s_1,\ldots,s_{n-1})\,|\,0\leqslant s_1\leqslant\cdots\leqslant s_{n-1}\leqslant\tfrac{1}{2}\},\tag{5}$$
$$\Lambda^l=\{(s_1,\ldots,s_{n-1})\,|\,0\leqslant s_1\leqslant\cdots\leqslant s_l\leqslant\tfrac{1}{2}\leqslant s_{l+1}\leqslant\cdots\leqslant s_{n-1}\leqslant 1\},$$
$$\text{for } 1\leqslant l\leqslant n-1.\tag{6}$$

By an obvious substitution we get

$$B_0=\int_{\Lambda^n}\cdots\int ds_1\cdots ds_{n-1}\left(\frac{1}{2}-s_{n-1}\right)f(s_1,\ldots,s_{n-1}),\tag{7}$$

$$B_n=\int_{\Lambda^n}\cdots\int ds_1\cdots ds_{n-1}\left(\frac{1}{2}-s_{n-1}\right)f(s_1,\ldots,s_{n-1}),\tag{8}$$

and for $1\leqslant k\leqslant n-1$,

$$B_k=\int_{M^k}\cdots\int ds_1\cdots ds_{n-1}\left(\min\left(\frac{1}{2},s_k\right)-\max\left(s_{n-1}-\frac{1}{2},s_{k-1}\right)\right)f(s_1,\ldots,s_{n-1}),\tag{9}$$

where

$$M^k=\{(s_1,\ldots,s_{n-1})\,|\,0\leqslant s_1\leqslant\cdots\leqslant s_{k-1}\leqslant\tfrac{1}{2},\ s_{k-1}\leqslant s_k\leqslant\cdots\leqslant s_{n-1}\leqslant 1\},$$

and with the convention that $s_0=0$. Clearly all the M^k can be decomposed as unions of domains of the type Λ^l. Therefore

$$\sum_{k=0}^{n}B_k=\sum_{l=0}^{n}\int_{\Lambda^l}\cdots\int ds_1\cdots ds_{n-1}I_l(s_1,\ldots,s_{n-1})\times f(s_1,\ldots,s_{n-1}),$$

where $I_l(s_1,\ldots,s_{n-1})$ is a bounded non-negative measurable function on Λ^l independent of f. On the other hand,

$$A=\sum_{l=0}^{n}\int_{\Lambda^l}\cdots\int ds_1\cdots ds_{n-1}(1-s_{n-1})f(s_1,\ldots,s_{n-1});$$

hence the proof will be finished if we show that

$$I_l(s_1,\ldots,s_{n-1})=1-s_{n-1},\quad\text{for all } l=0,\ldots,n.\tag{10}$$

We will now prove (10) l by l. From the expressions (4)–(9) it is clear that, for a given l, the only contributions to I_l come from B_k with $k\leqslant l+1$. From (7) it is obvious that $k=0$ contributes only to $l=n$. From (8) and (9) it can also be seen that for $0\leqslant l\leqslant n-1$ all B_k with $1\leqslant k\leqslant l+1$ contribute to I_l.

We first treat the case $l=0$ which by the remarks made above consists of the single term $k=1$. On $M^1\cap\Lambda^0$ one has

$$I_0=\min(\tfrac{1}{2},s_1)-\max(s_{n-1}-\tfrac{1}{2},s_0)=1-s_{n-1}.$$

Now fix l, $1\leqslant l\leqslant n-1$, and define for $0\leqslant j\leqslant l$ (still with the convention $s_0=0$)

$$\Lambda^{l,j}=\{(s_1,\ldots,s_{n-1})\in\Lambda^l\,|\,s_j\leqslant s_{n-1}-\tfrac{1}{2}\leqslant s_{j+1}\},$$

then we have to consider contributions to I_l from M^k, $1\leqslant k\leqslant l+1$. With $\{M^k\cap\Lambda^{l,j}\,|\,0\leqslant j\leqslant k-1\}$ being a partition of $M^k\cap\Lambda^l$, we have on $M^k\cap\Lambda^{l,j}$,

$$\min(\tfrac{1}{2},s_k)-\max(s_{n-1}-\tfrac{1}{2},s_{k-1})=\begin{cases}s_k-s_{k-1}, & \text{if } k\leqslant l \text{ and } j\leqslant k-2, \quad (11)\\ s_k-s_{n-1}+\tfrac{1}{2}, & \text{if } k\leqslant l \text{ and } j=k-1, \quad (12)\\ \tfrac{1}{2}-s_{k-1}, & \text{if } k=l+1 \text{ and } j\leqslant k-2, \quad (13)\\ 1-s_{n-1}, & \text{if } k=l+1 \text{ and } j=k-1. \quad (14)\end{cases}$$

Hence for $0\leqslant j\leqslant l$,

$$I_l|_{\Lambda^{l,j}}=\sum_{k=j+1}^{l+1}\left(\min\left(\frac{1}{2},s_k\right)-\max\left(s_{n-1}-\frac{1}{2},s_{k-1}\right)\right)\Big|_{M^k\cap\Lambda^{l,j}},$$

which by (11)–(14) can be computed as follows: (i) for $j=l$ k only takes the value $l+1$ and by (14)

$$I_l|_{\Lambda^{l,l}}=1-s_{n-1};$$

(ii) for $0\leqslant j\leqslant l-1$ we have to consider the cases $k=j+1$, $j+2\leqslant k\leqslant l$, $k=l+1$ which lead by (11), (12), and (13) to

$$I_l|_{\Lambda^{l,j}}=(s_{j+1}-s_{n-1}+\tfrac{1}{2})+\sum_{k=j+2}^{l}(s_k-s_{k-1})+\left(\frac{1}{2}-s_l\right)=1-s_{n-1}.\tag{15}$$

Finally for $l=n$ an analogous computation as in (15) can be made; however, the last term $\frac{1}{2}-s_l$ (with $l=n$) is absent, but on the other hand there is an additional term arising from B_0 which precisely equals $\frac{1}{2}-s_{n-1}$. So

$$I_l(s_1,\ldots,s_{n-1})=1-s_{n-1},\quad\text{for all } 0\leqslant l\leqslant n.$$ ■

We will now apply Lemma 2.1 to compute $\|\Omega^P\|^2$, where Ω^P is the cyclic vector of a perturbed equilibrium state. Using the notation introduced in (1), the functions

$$(z_1,\ldots,z_n)\in\mathbb{C}^n\mapsto\langle\Omega_k^P(\bar{z}_k,\ldots,\bar{z}_1)\,|\,\Omega_{n-k}^P(z_{k+1},\ldots,z_n)\rangle$$

for $0\leqslant k\leqslant n$, are analytic on

$$-\tfrac{1}{2}<\operatorname{Im}z_1<\cdots<\operatorname{Im}z_k<0,$$
$$0\leqslant\operatorname{Im}z_{k+1}<\cdots<\operatorname{Im}z_n<\tfrac{1}{2}.$$

Furthermore for $z_j\in\mathbb{R}$, $j=1,\ldots,n$, we have by time translation invariance

$$\langle\Omega_k^P(z_k,\ldots,z_k)\,|\,\Omega_{n-k}^P(z_{k+1},\ldots,z_n)\rangle=\langle\Omega^0|\alpha_{z_1}^0(P)\cdots\alpha_{z_k}^0(P)\alpha_{z_{k+1}}^0(P)\cdots\alpha_{z_n}^0(P)\Omega^0\rangle=\langle\Omega^0|P\alpha_{z_2-z_1}^0(P)\cdots\alpha_{z_n-z_1}^0(P)\Omega^0\rangle.$$

Therefore by the edge of the wedge theorem[6] all these analytic functions have a common analytic extension to the domain

$$-\tfrac{1}{2}<\operatorname{Im} z_1<\cdots<\operatorname{Im} z_n<\tfrac{1}{2}$$

and

$$G^n(u_1,\ldots,u_{n-1})\equiv\langle\Omega^0|\Omega_n^P(0,iu_1,\ldots,iu_{n-1})\rangle \tag{16}$$

has an analytic extension to $0<\operatorname{Re} u_1<\cdots<\operatorname{Re} u_{n-1}<1$, which is continuous and bounded on the closure of this domain and satisfies by (2),

$$|G^n(u_1,\ldots,u_{n-1})|\leqslant\|P\|^n. \tag{17}$$

Proposition 2.2: Using the notations (3) and (16) one has

$$\|\Omega^P\|^2=1+\sum_{n\geqslant 1}(-1)^n\int_{0\leqslant u_1\leqslant\cdots\leqslant u_n\leqslant 1}\cdots\int du_1\cdots du_n \times G^n(u_1,\ldots,u_{n-1}). \tag{18}$$

Proof:

$$\|\Omega^P\|^2=1+\sum_{n\geqslant 1}(-1)^n\sum_{k=0}^{n}\int_{0\leqslant s_1\leqslant\cdots\leqslant s_k\leqslant 1/2}\cdots\int ds_1\cdots ds_k \times\int_{0\leqslant t_1\leqslant\cdots\leqslant t_{n-k}\leqslant 1/2} dt_1\cdots dt_{n-k} \times G^n(s_k-s_{k-1},\ldots,s_k-s_1,s_k+t_1,\ldots,s_k+t_{n-k}),$$

and the series is absolutely convergent by (17). For $n\geqslant 2$ we can now apply Lemma 2.1 to each order, the case $n=1$ being trivial. ■

The explicit temperature dependence can be introduced by scaling appropriately the time parameter; an explicit expansion parameter μ can also be introduced by replacing P by μP. By doing so (18) becomes

$$\|\Omega_\beta^P\|^2=1+\sum_{n\geqslant 1}(-\mu)^n\int_{0\leqslant u_1\leqslant\cdots\leqslant u_n\leqslant\beta}\cdots\int du_1\cdots du_n \times G_\beta^n(u_1,\ldots,u_{n-1}), \tag{19}$$

where

$$G_\beta^n(u_1,\ldots,u_{n-1}),\quad 0<\operatorname{Re} u_1<\cdots<\operatorname{Re} u_{n-1}<\beta, \tag{20}$$

is the analytic extension of

$$\langle\Omega_\beta^0|P\alpha_{iu_1}^0(P)\cdots\alpha_{iu_{n-1}}^0(P)\Omega_\beta^0\rangle.$$

Proposition 2.2 will be used to approach the ground state as a limit of temperature states for $\beta\to\infty$. If the functions $G_\beta^n(u_1,\ldots,u_{n-1})$ are suitable, formula (19) can be interpreted as the grand-canonical partition function of a one-dimensional classical system of size β and so the low temperature limit of the quantum system corresponds to an infinite volume limit of an associated classical system. Not all quantum systems will allow for such an interpretation; indeed a minimal requirement is that the G_β^n should be positive. We will now explicitly carry out this program for the spin-boson model.

III. A CLASSICAL PARTICLE MODEL FOR THE GROUND STATE OF THE SPIN-BOSON MODEL

We first introduce the spin-boson model that describes a two-level system interacting with a scalar Bose field. The formal Hamiltonian H is given by[9,10]

$$H=\int_{\mathbb{R}}dk\,\epsilon(k)a_k^*a_k+\int_{\mathbb{R}}dk\,\lambda(k)(a_k+a_k^*)\sigma^z+\mu\sigma^x, \tag{21}$$

where a_k, a_k^* are boson annihilation and creation operators and $(\sigma^x,\sigma^y,\sigma^z)$ are the Pauli spin matrices. The parameters μ, ϵ, λ will be assumed to satisfy the following properties: $\mu\in\mathbb{R}$; ϵ: $\mathbb{R}\to\mathbb{R}^+$ is piecewise continuous and $\forall c>0$ and $\forall\beta>0$,

$$\int_{|k|>c}dk\,\frac{1}{e^{\beta\epsilon}-1}<\infty; \tag{22}$$

λ: $\mathbb{R}\to\mathbb{R}$ is measurable and satisfies

$$\int_{\mathbb{R}}dk\,\lambda^2<\infty\quad\text{and}\quad\int_{\mathbb{R}}dk\,\frac{\lambda^2}{\epsilon}<\infty.$$

The W^*-dynamical system $(\mathcal{M},\Omega_\beta^0,\alpha_t^0)$ corresponding to the unperturbed Hamiltonian ($\mu=0$) will be defined on the Gel'fand–Naimark–Segal (GNS) representation of a state ω_β on an appropriate C^* algebra. For more details on this contruction and for a proof of the uniqueness of the equilibrium state we refer to Ref. 9.

We consider the C^* algebra $\mathcal{A}\otimes M_2$, where M_2 are the complex 2×2 matrices and where $\mathcal{A}$ is the canonical commutation relation (CCR) algebra generated by the Weyl operators

$$\{W(\phi)\,|\,\phi\in\mathcal{L}^2(\mathbb{R},(1+1/\epsilon)dk)\},$$

satisfying

$$W(\phi)W(\psi)=\exp-i\sigma(\phi,\psi)\,W(\phi+\psi),\qquad W(\phi)^*=W(-\phi),\quad\sigma(\phi,\psi)=\operatorname{Im}\langle\phi|\psi\rangle, \tag{23}$$

where

$$\langle\phi|\psi\rangle=\int_{\mathbb{R}}dk\,\bar{\phi}\psi.$$

Writing the algebra $\mathcal{A}\otimes M_2$ as $M_2(\mathcal{A})$ we represent in the basis of $\mathbb{C}^2$ that diagonalizes σ^z the equilibrium state ω_β^0 as

$$\omega_\beta^0=\frac{1}{2}\begin{pmatrix}\omega_{\beta,+} & 0\\ 0 & \omega_{\beta,-}\end{pmatrix}, \tag{24}$$

where $\omega_{\beta,\pm}$ are quasifree states of $\mathcal{A}$ given by

$$\omega_{\beta,\pm}(W(\phi))=\exp\{\pm i\operatorname{Im}\langle 2i\lambda/\epsilon|\phi\rangle-\tfrac{1}{2}\langle\phi|\coth(\beta\epsilon/2)\phi\rangle\}. \tag{25}$$

In the GNS representation of $(M_2(\mathcal{A}),\omega_\beta^0)$ we can identify $M_2(\mathcal{A})$ with its representation because $M_2(\mathcal{A})$ is simple. Then $\mathcal{M}$ is $M_2(\mathcal{A})''$ and

$$\omega_\beta^0(x)=\langle\Omega_\beta^0|x\Omega_\beta^0\rangle,\quad x\in M_2(\mathcal{A}).$$

Finally, setting $\sigma^\pm=\frac{1}{2}(\sigma^x\pm i\sigma^y)$, the unperturbed dynamics $\{\alpha_t^0\,|\,t\in\mathbb{R}\}$ is a strongly continuous group of automorphisms of $\mathcal{M}$ determined by the following relations:

$$\alpha_t^0(\sigma^\pm)=\sigma^\pm W(\pm(2i\lambda/\epsilon)(1-e^{it\epsilon})), \tag{26}$$

$$\alpha_t^0(\sigma^z)=\sigma^z, \tag{27}$$

$$\alpha_t^0(W(\phi))=\exp\{i\sigma^z\operatorname{Im}\langle 2i\lambda/\epsilon|(1-e^{it\epsilon})\phi\rangle\}W(e^{it\epsilon}\phi). \tag{28}$$

It can be checked that Ω_β^0 is a cyclic and separating vector for $\mathcal{M}$ and that $\{\alpha_t^0 | t\in\mathbf{R}\}$ is the modular automorphism group corresponding to Ω_β^0. We will now apply the perturbation theory for equilibrium states of Sec. II for a perturbation $P=\mu\sigma^x$.

Proposition 3.1: The functions G_β^n defined in (16) and (20) are given by

$$G_\beta^n(u_1,\ldots,u_{n-1})=0, \quad \text{if } n \text{ is odd},$$

$$G_\beta^{2n}(u_1,\ldots,u_{2n-1})=\exp B_\beta^{2n}(0,u_1,\ldots,u_{2n-1}), \quad n=1,2,\ldots,$$

where for $0\leqslant s_1\leqslant\cdots\leqslant s_{2n}\leqslant\beta$,

$$B_\beta^{2n}(s_1,\ldots,s_{2n}) = 8\int_{\mathbf{R}} dk\,\frac{\lambda^2}{\epsilon^2}\sum_{1<j_1<j_2\leq 2n}(-1)^{j_1+j_2} \times\frac{\sinh((\beta-s_{j_2}+s_{j_1})(\epsilon/2))\sinh((s_{j_2}-s_{j_1})(\epsilon/2))}{\sinh(\beta\epsilon/2)}. \tag{29}$$

Proof: We start by computing

$$\omega_\beta^0(\alpha_{t_1}^0(\sigma^x)\cdots\alpha_{t_n}^0(\sigma^x))\equiv F^n(t_1,\ldots,t_n)$$

and then obtain G_β^n by analytic extension and application of time translation invariance.

We first note that ω_β^0 is diagonal in the basis of $\mathbf{C}^2$ which diagonalizes σ^z; furthermore the automorphism

$$y\to\sigma^z y\sigma^z$$

of $\mathcal{M}$ commutes with $\{\alpha_t^0 | t\in\mathbf{R}\}$ and maps σ^x into $-\sigma^x$; therefore F^n vanishes for n odd.

We must now compute F^{2n}. As $\sigma^x=\sigma^+ + \sigma^-$, we obtain by (26) that

$$\alpha_{t_1}^0(\sigma^x)\cdots\alpha_{t_{2n}}^0(\sigma^x)=\sigma^+\sigma^-\prod_{j=1}^{\overrightarrow{n}}\left\{W\left(\frac{2i\lambda}{\epsilon}(1-e^{it_{2j-1}\epsilon})\right)W\left(\frac{-2i\lambda}{\epsilon}(1-e^{it_{2j}\epsilon})\right)\right\}$$
$$+\sigma^-\sigma^+\prod_{j=1}^{\overrightarrow{n}}\left\{W\left(\frac{-2i\lambda}{\epsilon}(1-e^{it_{2j-1}\epsilon})\right)W\left(\frac{2i\lambda}{\epsilon}(1-e^{it_{2j}\epsilon})\right)\right\}.$$

Observe now that ω_β^0 is invariant under the automorphism

$$y\otimes W(\phi)\to e^{i\pi\sigma^x}ye^{i\pi\sigma^x}\otimes W(-\phi), \quad y\in M_2,$$

of $\mathcal{M}$. Hence

$$F^{2n}(t_1,\ldots,t_{2n})=\omega_{\beta,+}\left(\prod_{j=1}^{\overrightarrow{n}}\left\{W\left(\frac{2i\lambda}{\epsilon}(1-e^{it_{2j-1}\epsilon})\right)W\left(\frac{-2i\lambda}{\epsilon}(1-e^{it_{2j}\epsilon})\right)\right\}\right)$$
$$=\omega_{\beta,+}\left(\prod_{j=1}^{\overrightarrow{n}}\left\{\exp i\,\mathrm{Im}\left\langle\frac{2\lambda}{\epsilon}(1-e^{it_{2j-1}\epsilon})\middle|\frac{2\lambda}{\epsilon}(1-e^{it_{2j}\epsilon})\right\rangle W\left(\frac{2i\lambda}{\epsilon}(e^{it_{2j}\epsilon}-e^{it_{2j-1}\epsilon})\right)\right\}\right)=\exp A^{2n}(t_1,\ldots,t_{2n}), \tag{30}$$

where by (23) and (25),

$$A^{2n}(t_1,\ldots,t_{2n})=i\sum_{j=1}^{n}\mathrm{Im}\left\langle\frac{2\lambda}{\epsilon}(1-e^{it_{2j-1}\epsilon})\middle|\frac{2\lambda}{\epsilon}(1-e^{it_{2j}\epsilon})\right\rangle$$
$$-i\sum_{1<j_1<j_2<n}\mathrm{Im}\left\langle\frac{2i\lambda}{\epsilon}(e^{it_{2j_1}\epsilon}-e^{it_{2j_1-1}\epsilon})\middle|\frac{2i\lambda}{\epsilon}(e^{it_{2j_2}\epsilon}-e^{it_{2j_2-1}\epsilon})\right\rangle+i\sum_{j=1}^{n}\mathrm{Im}\left\langle\frac{2\lambda}{\epsilon}\middle|\frac{2\lambda}{\epsilon}(e^{it_{2j}\epsilon}-e^{it_{2j-1}\epsilon})\right\rangle$$
$$-\frac{1}{2}\sum_{j_1=1}^{n}\sum_{j_2=1}^{n}\left\langle\frac{2\lambda}{\epsilon}(e^{it_{2j_2}\epsilon}-e^{it_{2j_2-1}\epsilon})\middle|\coth\left(\frac{\beta\epsilon}{2}\right)\frac{2\lambda}{\epsilon}(e^{it_{2j_1}\epsilon}-e^{it_{2j_1-1}\epsilon})\right\rangle$$
$$=-4\int dk\,\frac{\lambda^2}{\epsilon^2}\left[n\coth\left(\frac{\beta\epsilon}{2}\right)+\sum_{1<j_1<j_2<2n}(-1)^{j_1+j_2}\left\{i\sin((t_{j_2}-t_{j_1})\epsilon)+\coth\left(\frac{\beta\epsilon}{2}\right)\cos((t_{j_2}-t_{j_1})\epsilon)\right\}\right].$$

By the conditions (22) the function

$$(t_1,\ldots,t_{2n}) \mapsto A^{2n}(t_1,\ldots,t_{2n})$$

can be analytically extended to the domain

$$\{(z_1,\ldots,z_{2n}) | 0 < \operatorname{Im} z_1 < \cdots < \operatorname{Im} z_{2n} < \beta\}$$

and for $t_j = is_j, j = 1,\ldots,2n$,

$$A^{2n}(is_1,\ldots,is_{2n}) = 8\int dk \frac{\lambda^2}{\epsilon^2} \sum_{1\le j_1 j_2\le 2n} (-1)^{j_1+j_2} \times \frac{\sinh((\beta + s_{j_1} - s_{j_2})(\epsilon/2))\sinh((s_{j_2} - s_{j_1})(\epsilon/2))}{\sinh(\beta\epsilon/2)}. \tag{31}$$

In order to get G_β^{2n} we use time translation invariance and put in (30) $s_1 = 0$, $s_{j+1} = u_j, j = 1,\ldots,2n-1$. ■

Propositions 2.2 and 3.1 show already that $\|\Omega_\beta^p\|^2$ can be interpreted as the grand canonical partition function at fugacity $|\mu|$ of a continuous one-dimensional classical system of size β at inverse temperature 1 and with Hamiltonian

$$H_\beta^{2n-1}(u_1,\ldots,u_{2n-1}) = \sum_{j=1}^{2n-1} (-1)^j V_\beta(u_j) - \sum_{1<j_1<j_2\le 2n-1} (-1)^{j_1+j_2} V_\beta(u_{j_2} - u_{j_1}), \tag{32}$$

where

$$V_\beta(u) = 8\int_{\mathbb{R}} dk \frac{\lambda^2}{\epsilon^2} \frac{\sinh((\beta-u)(\epsilon/2))\sinh(u\epsilon/2)}{\sinh(\beta\epsilon/2)}. \tag{33}$$

It is possible to translate this classical "particle" model to an Ising model on $\mathbb{R}$. Such a model was obtained in Ref. 4 by a path integral representation of the partition function of the full spin-boson model.

IV. THE ISING MODEL OVER R

The configuration space Ω_L of a continuous Ising model on the interval $[0,L)$ is the space of all functions

$$x\in[0,M) \to S(x)\in\{-1,1\}$$

that are continuous from the right and take the value 1 at $x = 0$. The *a priori* measure on this space is determined by the jump process with transition time $1/\delta$. The jump process is a Markov process defined by the transition probability

$$P(S(x) = 1 | S(0) = 1) = \tfrac{1}{2}(1 + \epsilon^{-\delta x}).$$

On the configuration space of the Ising model this process induces up to normalization a unique measure ν_n on the configurations with exactly n jumps at the points $0 < u_1 < \cdots < u_n < L$,

$$\nu_n(du_1\cdots du_n) = \delta^n\, du_1\cdots du_n. \tag{34}$$

We will now consider Hamiltonians of the following type:

$$H_L(S) = \int_0^L ds \int_0^L dt\, U_L(t-s)S(s)S(t) \tag{35}$$

and compute their partition function at inverse temperature 1

$$Z(L,U_L) = \int_{\Omega_L^+} \nu(dS) e^{-H_L(S)}, \tag{36}$$

where Ω_L^+ is the space of configurations consisting of the paths with an even number of jumps and where ν is the (unnormalized) measure equal to ν_{2n} on the paths with exactly n jumps. We establish now the connection between such models and the classical models described in Sec. III.

Proposition 4.1: Let $t\in[-L,L]\to G_L(t)\in\mathbb{R}$ be twice continuously differentiable such that

$$G_L(t) = G_L(-t),$$
$$G_L(L-t) = G_L(t), \quad t\in[0,L],$$
$$G_L(0) = 0,$$

then

$$Z(L,G_L'') = \sum_{n=0}^{\infty} \delta^{2n} \int_{0<u_1<\cdots<u_{2n}<L} \cdots \int du_1\cdots du_{2n} \times \exp\left\{-8 \sum_{1<j_1<j_2<2n} (-1)^{j_1+j_2} G_L(u_{j_2} - u_{j_1})\right\}. \tag{37}$$

Proof: By (33)–(35) we have

$$Z(L,G_L'') = \sum_{n=0}^{\infty} \delta^{2n} \int_{0<u_1<\cdots<u_{2n}<L} \cdots \int du_1\cdots du_{2n} \times \exp - H_L(S_u),$$

where $u = (u_1,\ldots,u_{2n})$ and where S_u is the configuration

$$S_u(x) = -1, \quad u_{2k-1} \le x < u_{2k}, \quad k = 1,\ldots,n,$$
$$= +1, \quad \text{elsewhere.}$$

We now compute $H_L(S_u)$ taking into account the convention $u_0 = 0$ and $u_{2n+1} = L$. Let I_j, $0\le j\le 2n$ be the interval $[u_{2j}, u_{2j+1})$ and let χ_{I_j} be the characteristic function of I_j, then

$$S_u = \sum_{j=0}^{2n} (-1)^j \chi_{I_j},$$

and so

$$
\begin{aligned}
H_L(S_u) &= \int_0^L ds \int_0^L dt\, G_L''(t-s) \sum_{j_1=0}^{2n} \sum_{j_2=0}^{2n} (-1)^{j_1+j_2} \chi_{I_{j_1}}(s) \chi_{I_{j_2}}(t) \\
&= \sum_{j_1=0}^{2n} \sum_{j_2=0}^{2n} (-1)^{j_1+j_2} \{ -G_L(u_{j_2+1} - u_{j_1+1}) + G_L(u_{j_2+1} - u_{j_1}) + G_L(u_{j_2} - u_{j_1+1}) - G_L(u_{j_2} - u_{j_1}) \} \\
&= -4 \sum_{j_1=1}^{2n} \sum_{j_2=1}^{2n} (-1)^{j_1+j_2} G_L(u_{j_2} - u_{j_1}) + \sum_{j=1}^{2n+1} (-1)^j G_L(L-u_j) + \sum_{j=1}^{2n} (-1)^j G_L(L-u_j) \\
&\quad + \sum_{j=0}^{2n} (-1)^j G_L(L-u_j) - \sum_{j=1}^{2n} (-1)^j G_L(u_j) - \sum_{j=1}^{2n+1} (-1)^j G_L(u_j) \\
&\quad + \sum_{j=1}^{2n} (-1)^j G_L(L-u_j) - \sum_{j=1}^{2n} (-1)^j G_L(u_j) - \sum_{j=1}^{2n+1} (-1)^j G_L(u_j) \\
&\quad + \sum_{j=1}^{2n} (-1)^j G_L(L-u_j) - \sum_{j=0}^{2n} (-1)^j G_L(u_j) - \sum_{j=1}^{2n} (-1)^j G_L(u_j) \\
&= -8 \sum_{1 \le j_1 < j_2 \le 2n} (-1)^{j_1+j_2} G_L(u_{j_2} - u_{j_1}),
\end{aligned}
$$

where we have used the symmetry properties of G_L. ■

The symmetry properties of G_L are satisfied by the choice

$$G_\beta(t) = \tfrac{1}{8} V_\beta(|t|), \quad t \in [-\beta, \beta].$$

The KMS property of the equilibrium state ω_β^0 of the original unperturbed spin-boson model ensures that G_β satisfies

$$G_\beta(\beta - t) = G_\beta(t), \quad t \in [0, \beta].$$

It is now clear by Proposition 4.1 that the classical model of Sec. III is equivalent to a continuous Ising model of size $[0,\beta]$ with interaction potential

$$
\begin{aligned}
U_\beta(t) &= \frac{1}{8} \frac{d^2}{dt^2} V_\beta(t) \\
&= -\frac{1}{2} \int_{\mathbf{R}} dk\, \lambda^2 \left\{ \frac{e^{-\epsilon t} + e^{-\epsilon(\beta - t)}}{1 - e^{-\beta\epsilon}} \right\}
\end{aligned}
$$

and with $\delta = \mu$.

[1] P. W. Anderson, G. Yuval, and D. R. Hamann, Phys. Rev. B **1**, 4464 (1970).
[2] G. Yuval and P. W. Anderson, Phys. Rev. B **1**, 1522 (1970).
[3] P. W. Anderson and G. Yuval, J. Phys. C: Solid State Phys. **4**, 607 (1971).
[4] H. Spohn and R. Dümcke, J. Stat. Phys. **41**, 389 (1985).
[5] O. Bratteli and D. W. Robinson, *Operator Algebras and Quantum Statistical Mechanics* (Springer, New York, 1979).
[6] O. Bratteli and D. W. Robinson, *Operator Algebras and Quantum Statistical Mechanics II* (Springer, New York, 1981).
[7] H. Araki, Pub. R.I.M.S., Kyoto Univ. **9**, 165 (1973).
[8] H. Araki, in *C*-Algebras and Their Applications to Statistical Mechanics and Quantum Field Theory*, edited by D. Kastler (Editrice Composition, Bologna, 1975).
[9] M. Fannes, B. Nachtergaele, and A. Verbeure, Commun. Math. Phys. **114**, 537 (1988).
[10] M. Fannes, B. Nachtergaele, and A. Verbeure, J. Phys. A: Math. Gen. **21**, 1759 (1988).

The derivation of the regularized chiral Jacobian using the zeta function method

Rong-tai Wang[a)]
Physics Department, The City College of City University of New York, New York, New York 10031

(Received 30 March 1988; accepted for publication 15 June 1988)

Using the zeta function method, a general formula for the regularized chiral Jacobian to theories including non-Hermitian Dirac operators $\mathscr{D}$ defined in arbitrary even-dimensional Euclidean space is derived. The agreement of this formula with the results obtained in the differential geometric approach is also clarified.

I. INTRODUCTION

Since Fujikawa made an important observation that the path integral measure is not invariant under chiral transformation,[1] it has been clear that the associated Jacobian factor is responsible for the anomalous term in the corresponding Ward–Takahashi identities in the quantum field theory of the path integral formalism. Meanwhile, the importance of a suitable choice of the regularization scheme was also noticed by many authors. One of the regularization methods that proved to be useful is the ζ-function method[2,3] based on techniques developed by Seeley from his definition of the complex power of pseudodifferential operators,[4] and examples in some special cases and low dimensions have been worked out in Refs. 5 and 6.

In this paper, by using the ζ-function method described in Refs. 2 and 3, the derivation of an explicit formula for the chiral Jacobian in arbitrary even-dimensional Euclidean space-time is made, and the emphasis of the discussion is laid on the coincidence between this formula and the results of anomalies in the differential geometric approach (see, for example, Refs. 7–9).

The Dirac operator $\mathscr{D}$ considered in the generating function for the fermion including γ_{2n+1} coupling defined in $2n$-dimensional Euclidean space is

$$\begin{aligned}\mathscr{D} &= i(\partial\!\!\!/ + V\!\!\!/ + A\!\!\!/\gamma_{2n+1}) \\ &= i\gamma_\mu(\partial_\mu + V_\mu + A_\mu\gamma_{2n+1})\end{aligned} \tag{1.1}$$

with Clifford algebra $\{\gamma_\mu,\gamma_\nu\} = 2\delta_{\mu\nu}$, and $\gamma_{2n+1} = i^n\gamma_1\cdots\gamma_{2n}$, where the anti-Hermitian background fields V_μ or A_μ take the value in the Lie algebra of some gauge group. Under an infinitesimal chiral transformation $\Omega = 1 + \beta\gamma_{2n+1}$ specified by $\beta(x)$, which also may be Lie algebra valued, the chiral Jacobian, which we derived in this paper, can be expressed as

$$\begin{aligned}-\log J = &\frac{2}{(4\pi)^n n!}\int d^{2n}x \\ &\times \mathrm{tr}\left\{\beta\gamma_{2n+1}\sum_{m=0}^{n} B(m+1,n+1)\right. \\ &\left.\times\left[\frac{1}{(2m)!}\frac{\partial^{2m}}{\partial u^{2m}}\sum_{l_1+l_2=m+n} Q_+^{l_1}Q_-^{l_2}\right]_{u=0}\right\},\end{aligned} \tag{1.2}$$

where

$$Q_\pm = F^V_{\mu\nu}\sigma_{\mu\nu} \pm F^A_{\mu\nu}\sigma_{\mu\nu} + 2uA\!\!\!/\gamma_{2n+1}, \tag{1.3}$$

$$F^V_{\mu\nu} = \partial_\mu v_\nu - \partial_\nu V_\mu + [V_\mu,V_\nu] + [A_\mu,A_\nu], \tag{1.4a}$$

$$F^A_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu + [A_\mu,V_\nu] + [V_\mu,A_\nu], \tag{1.4b}$$

and

$$B(m+1,n+1) = m!n!/(m+n+1)!, \quad \sigma_{\mu\nu} = \tfrac{1}{4}[\gamma_\mu,\gamma_\nu]. \tag{1.5}$$

In Sec. II, the ζ-function regularization scheme for evaluating the chiral Jacobian is briefly reviewed. In Sec. III, some simplification is made for Seeley's formula of kernel $K_0(x,x)$ in the case of the Dirac operator. Section IV is the detail derivation of the formula (1.2). The applications and discussions of this formula are made in Sec. V.

II. THE ζ-FUNCTION METHOD

According to the description in Refs. 2 and 3 and the results of Seeley,[4] let A be an elliptic invertible operator of order $m > 0$, defined on some compact manifold M without boundary of dimension d; then the complex power of operator A is defined as

$$A^s = \frac{1}{2\pi i}\int_\Gamma \lambda^s(\lambda I - A)^{-1}\,d\lambda, \tag{2.1}$$

where Γ is a curve beginning at ∞, passing along the ray of minimal growth[10] to a small circle about the origin, then clockwise about the circle, and back to ∞ along the ray. The generalized ζ function formed from the eigenvalues λ_j of A is $\zeta(s,A) = \Sigma_j\lambda_j^{-s}$. Denoting $K(x,y,O)$ as the kernel of operator O, one can write

[a)] On leave of absence from Physics Department, Modern Physics Institute, Fudan University, Shanghai, China.

 0022-2488/88/102294-06$02.50

$$\zeta(s,A)=\int_M d\mu_x \operatorname{tr} K_{-s}(x,x,A) \tag{2.2}$$

with $K_s(x,x,A)=K(x,x,A^s)$.[3] Seeley's results show that the series $\Sigma_j \lambda_j^{-s}$ converges only for $\operatorname{Re} s > d/m$, and can be analytically extended to a meromorphic function of s in the whole complex plane; in particular, it is regular at $s=0$.[4] Since the derivation of the ζ function at $s=0$ is formally equal to $-\Sigma_j \log \lambda_j$, one defines the regularized determinant of the operator to be $\exp(-d\zeta/ds)|_{s=0}$. In the application of this method to the path integral approach of fermions in the presence of background fields, the generating function is regularized from the beginning,

$$Z_{\text{reg}}=\int D\bar{\psi}\, D\psi \exp\left(-\int dx\, \bar{\psi}\, \mathcal{D}\psi\right)$$

$$=\det \mathcal{D} \equiv \exp\left\{-\frac{d\zeta(s,\mathcal{D})}{ds}\right\}_{s=0}. \tag{2.3}$$

If the Dirac operator $\mathcal{D}$ is invertible, this yields a natural definition for Jacobian J associated with a fermion transformation $\psi=\Omega\psi'$, $\bar{\psi}=\bar{\psi}'\Omega$,

$$J^{-1}=\det(\Omega\,\mathcal{D}\Omega)/\det \mathcal{D}. \tag{2.4}$$

In the case of the infinitesimal chiral transformation $\Omega=1+\beta\gamma_{d+1}$, one can apply the differentiability result of the ζ function[3] and obtain

$$\log J=-2\left\{\frac{d}{ds}\left[s \operatorname{Tr}(\mathcal{D}^{-s}\gamma_{d+1}\beta)\right]\right\}_{s=0}. \tag{2.5}$$

This yields

$$J=\exp\left\{-2\int_{S^d} d\mu \operatorname{tr}(\gamma_{d+1}\beta K_0(x,x,\mathcal{D}))\right\}. \tag{2.6}$$

Thus we can evaluate the chiral Jacobian by use of Seeley's formula for $K_0(x,x,\mathcal{D})$,[4]

$$K_0(x,x\mathcal{D})=\frac{-i}{(2\pi)^d}\int_{|\xi|=1} d\xi \int_0^\infty b_{-1-d}(x,\xi,iu)\,du \tag{2.7}$$

and the relation of Seeley's coefficients b_i's becomes

$$b_{-1}(a_1-\lambda)=I,$$

$$b_{-1-l}(a_1-\lambda)+\sum_{(\alpha)=0}^{l-1}(\partial_\xi)^\alpha b_{-l+|\alpha|}\frac{(D_x)^\alpha a_0}{\alpha!} \tag{2.8}$$

$$=0 \quad (l>0).$$

Here $\sigma(\mathcal{D})=a_0+a_1$ is the symbol of the Dirac operator, $\alpha=(\alpha_1,\ldots,\alpha_d)$ is a multi-index, $|\alpha|=\Sigma_{\mu=1}^d \alpha_\mu$ is of the order of α,

$$(\partial_\xi)^\alpha=\prod_{\mu=1}^d\left(\frac{\partial}{\partial\xi_\mu}\right)^{\alpha_\mu},\quad D_x^\alpha=\prod_{\mu=1}^d\left(\frac{-i\partial}{\partial x_\mu}\right)^{\alpha_\mu}. \tag{2.9}$$

For noninvertible Dirac operator $\mathcal{D}$, one introduces the definition

$$\det'\mathcal{D}=\lim_{\alpha\to 0^+}\det(\mathcal{D}+\alpha I)/\alpha^N.$$

It has been proved that $\det'\mathcal{D}=\det(\mathcal{D}+P_{\ker})$, where N is the dimension of the null set of $\mathcal{D}$, $P_{\ker}$ is the projection onto the null set of $\mathcal{D}$, and a similar treatment to that of the invertible leads to formula (2.6).[11]

III. THE KERNEL OF DIRAC OPERATOR $K(x,x,\mathcal{D})$

The direct evaluation of the chiral Jacobian based on the relations of Seeley's coefficients (2.8) appears rather complicated and tedious when the dimension of the manifold M is more than 2. But, in the case of the Dirac operator, the principle symbol $a_1=\not{\xi}$ is independent of x, the relation of b's (2.8) can be greatly simplified.

Proposition: If the symbol of the Dirac operator defined in a D-dimensional manifold M without boundary is $\sigma(\mathcal{D})=a_1+a_0$, in which $a_1=-\not{\xi}$, $a_0=a_0(x)$, then

$$b_{-1}=(-\not{\xi}-\lambda I)^{-1}, \tag{3.1}$$

$$b_{-1-l}=\{[-b_{-1}(-\not{D}_x+a_0)]^l b_{-1}\}^{(0)} \quad (l>0), \tag{3.2}$$

where $\not{D}_x=\gamma_\mu(-i\partial/\partial x_\mu)$; $\{F(D_x)\}^{(i)}$ denotes only the terms in $F(D_x)$ with homogeneous degree i in D_x being preserved.

One obtains (3.1) from the first relation of (2.8). In order to prove (3.2) satisfies the second relation of (2.8), one should notice the following derivative property of b_{-1}:

$$\frac{\partial}{\partial\xi_\mu}b_{-1}=b_{-1}\gamma_\mu b_{-1}, \tag{3.3}$$

$$(b_{-1}\not{D}_x)^i b_{-1}a_0=\sum_{|\alpha|=i}\partial_\xi^\alpha b_{-1}\frac{D_x^\alpha a_0}{\alpha!} \quad (i\geqslant 0), \tag{3.4}$$

$$\{[b_{-1}(-\not{D}_x+a_0)]^j b_{-1}\}^{(i)}a_0$$

$$=\sum_{|\alpha|=i}\partial_\xi^\alpha\{[-b_{-1}(-\not{D}_x+a_0)]^{j-|\alpha|}b_{-1}\}^{(0)}\frac{D_x^\alpha a_0}{\alpha!}$$

$$(j>0,\quad j\geqslant i\geqslant 0). \tag{3.5}$$

Hence

$$b_{-1-l}=\{[-b_{-1}(-\not{D}_x+a_0)]^l b_{-1}\}^{(0)}$$

$$=-\sum_{i=0}^{l-1}\{[-b_{-1}(-\not{D}_x+a_0)]^{l-1}b_{-1}\}^{(i)}a_0 b_{-1}, \tag{3.6}$$

with the aid of (3.5), one can write

$$b_{-1-l}(-\not{\xi}-\lambda I)$$

$$=-\sum_{|\alpha|=0}^{l-1}\partial_\xi^\alpha\{[-b_{-1}$$

$$\times(-\not{D}_x+a_0)]^{l-|\alpha|-1}b_{-1}\}^{(0)}\frac{D_x^\alpha a_0}{\alpha!}$$

$$=-\sum_{|\alpha|=0}^{l-1}\partial_\xi^\alpha b_{-l+|\alpha|}\frac{D_x^\alpha a_0}{\alpha!}; \tag{3.7}$$

this is just the second relation in (2.8).

Using expression (3.2), Seeley's formula for $K_0(x,x,\mathcal{D})$ in (2.7) can be written as

$$K_0(x,x,\mathscr{D}) = \frac{-i}{(2\pi)^d}\int_{|\xi|=1} d\xi \times \int_0^\infty [-b_{-1}(-\not{D}_x + a_0)]^d b_{-1}\, du, \tag{3.8}$$

with $b_{-1} = (-\not{\xi} - iu)^{-1}$. We thus come to the following theorem.

Theorem: If the symbol of a Dirac operator defined in the d-dimensional manifold without boundary is $\sigma(\mathscr{D}) = -\not{\xi} + a_0(x)$, then the kernel $K_0(x,x,\mathscr{D})$ can be expressed as

$$K_0(x,x,\mathscr{D}) = -\frac{i}{(2\pi)^d d!}\left[\frac{\partial^d}{\partial\eta^d}\int_{|\xi|=1} d\xi \times \int_0^\infty du(-\not{\xi} - iu - \not{D}'_x + a'_0)^{-1}\right]_{\eta=0}, \tag{3.9}$$

in which $D'_x = \eta D_x$, $a'_0 = \eta a_0$, η is a real parameter.

IV. THE EVALUATION OF THE CHIRAL JACOBIAN

In the case of the Dirac operator as shown in (1.1), we compactify $\mathbb{R}^{2n}$ to S^{2n} by stereographic projection. Denoting $L = -\not{D}_x + a_0 = i\not{V} + i\not{A}\gamma_{2n+1}$, $L' = \eta L$, one can write

$$\begin{aligned}(-\not{\xi} + L' - iu)^{-1} &= (-\not{\xi} + L'^{\dagger} + iu)[(-\not{\xi} + L' - iu) \times (-\not{\xi} + L'^{\dagger} + iu)]^{-1} \\ &= (-\not{\xi} + L'^{\dagger} + iu)(\xi^2 + u^2 - R)^{-1} \\ &= (-\not{\xi} + L'^{\dagger} + iu)\sum_{l=0}^{\infty}(\xi^2 + u^2)^{-(l+1)}R^l, \end{aligned} \tag{4.1}$$

in which $R = R(x,\xi,u,\eta)$ is defined as

$$\begin{aligned}R(x,\xi,u,\eta) &= -L'L'^{\dagger} + iu(L'^{\dagger} - L') + (\not{\xi}L'^{\dagger} + L'\not{\xi}) \\ &= \eta^2(F^V_{\mu\nu} - F^A_{\mu\nu}\gamma_{2n+1})\sigma_{\mu\nu} \\ &\quad + 2u\eta\not{A}\gamma_{2n+1} + 2\eta\xi_\mu L^\dagger_\mu + R_c. \end{aligned} \tag{4.2}$$

Here R_c is the terms with γ's contracted and $F^V_{\mu\nu}$, $F^A_{\mu\nu}$, and $\sigma_{\mu\nu}$ are shown in (1.4) and (1.5). The kernel $K_0(x,x,\mathscr{D})$ shown in (3.9) can be written as

$$K_0(x,x,\mathscr{D}) = \sum_{i=1}^{3} K_0^{(i)}(x,x,\mathscr{D})$$

with

$$K_0^{(i)}(x,x,\mathscr{D}) = \frac{-i}{(2\pi)^{2n}}\int_{|\xi|=1} d\xi \times \int_0^\infty du\, b^{(i)}_{-1-2n}(x,\xi,iu), \tag{4.3}$$

in which

$$b^{(1)}_{-1-2n}(x,\xi,iu) = \left\{\frac{1}{(2n)!}\frac{\partial^{2n}}{\partial\eta^{2n}} \times \left[\sum_{m=0}^{n} iu(\xi^2 + u^2)^{-(n+m+1)}R^{n+m}\right]\right\}_{\eta=0}, \tag{4.4a}$$

$$b^{(2)}_{-1-2n}(x,\xi,iu) = \left\{\frac{1}{(2n)!}\frac{\partial^{2n}}{\partial\eta^{2n}} \times \left[\sum_{m=0}^{n} L'^{\dagger}(\xi^2 + u^2)^{-(n+m+1)}R^{n+m}\right]\right\}_{\eta=0}, \tag{4.4b}$$

$$b^{(3)}_{-1-2n}(x,\xi,iu) = \left\{\frac{1}{(2n)!}\frac{\partial^{2n}}{\partial\eta^{2n}} \times \left[\sum_{m=0}^{n} (-\not{\xi})(\xi^2 + u^2)^{-(n+m+1)}R^{n+m}\right]\right\}_{\eta=0}, \tag{4.4c}$$

other terms related with the expansion of (4.1) vanished under $(\partial^{2n}/\partial\eta^{2n})|_{\eta=0}$.

Now, in evaluating the chiral Jacobian as shown in (2.6), further simplification can be made under the combination of tr $\beta\gamma_{2n+1}$ (*), the ξ integral, and $(\partial^{2n}/\partial\eta^{2n})|_{\eta=0}$. If we denote $b_c^{(i)}$ as the terms in b_{-1-2n} that have no contribution under this combined operation, we can write (4.4a), (4.4b), and (4.4c) as

$$b^{(1)}_{-1-2n}(x,\xi,iu) = \sum_{m=0}^{n}\frac{iu^{2m+1}}{(\xi^2 + u^2)^{n+m+1}} \times \left[\frac{1}{(2m)!}\frac{\partial^{2m}}{\partial u^{2m}}Q_-^{n+m}\right]_{u=0} + b_c^{(1)}, \tag{4.5a}$$

$$b^{(2)}_{-1-2n}(x,\xi,iu) = \sum_{m=0}^{n}\frac{u^{2m+1}}{(\xi^2 + u^2)^{n+m+1}} \times \left[\frac{1}{(2m+1)!}\frac{\partial^{2m}}{\partial u^{2m+1}}L^\dagger Q_-^{n+m}\right]_{u=0} + b_c^{(2)}, \tag{4.5b}$$

$$b^{(3)}_{-1-2n}(x,\xi,iu) = \sum_{m=0}^{n}\frac{u^{2m+1}(-2\xi_\mu)}{(\xi^2 + u^2)^{n+m+1}}\left[\frac{1}{(2m+1)!}\frac{\partial^{2m}}{\partial u^{2m+1}} \times \sum_{l_1+l_2=m+n} Q_+^{l_1}L^\dagger Q_-^{l_2}\right]_{u=0} + b_c^{(3)}, \tag{4.5c}$$

where $Q_{\pm}$ is shown in (1.3). In deriving these expressions, we have noticed the following two facts: (a) upon using the properties of Dirac γ matrices, the terms survived under $\mathrm{tr}\,\beta\gamma_{2m+1}$ (*) must be of the form $\mathrm{tr}\,\beta\gamma_{2n+1}(\gamma_{\mu_1}\cdots\gamma_{\mu_{2n}})$ with $\mu_1\neq\mu_2\neq\cdots\neq\mu_{2n}$; (b) upon using the symmetric property of integration, all the terms without power of ξ_μ can be discarded. For example, in deriving (4.5a), the terms with one or more factor of $\xi_\mu L_\mu$ of R_c in the expansion of R^{n+m} do not contribute to $\mathrm{tr}\,\beta\gamma_{2n+1}K_0^{(1)}(x,x,\mathscr{D})$. So $\xi_\mu L_\mu$ and R_c in the expression of R in (4.2) can be deleted and $R(x,\xi,iu,\eta=1)=Q_-(x,u)+\cdots$. Now, the terms with factor η^{2n} in the expansion of R^{n+m} are accompanied with a factor u^{2m}, thus we can make the replacement of $[\partial^{2n}/(2n)!\,\partial\eta^{2n}]|_{\eta=0}$ with $[\partial^{2m}/(2m)!\,\partial u^{2m}]|_{u=0}$ as shown in (4.5a). In deriving (4.5b), $\xi_\mu L_\mu$ and R_c in R can also be deleted, the terms with the factor η^{2n} in the expansion of $L^{\dagger}R^{n+m}$ are now accompanied with u^{2m+1}. From (4.4c) to (4.5c), we first collect the terms with an even power of ξ_μ in the expansion of $-\not{\xi}(R)^{n+m}$ that contribute to $\mathrm{tr}\,\beta\gamma_{2n+1}K_0^{(3)}$ as

$$\sum_{l_1+l_2=m+n}(-\not{\xi})\bar{R}^{l_1}(2\xi_\mu L_\mu'^{\dagger})R^{l_2}=\sum_{l_1+l_2=m+n}(-2\xi_\mu^2)\bar{R}^{l_1}L'^{\dagger}R^{l_2},\tag{4.6}$$

where

$$\bar{R}(x,\xi,iu,\eta)=\eta^2(F^V_{\mu\nu}+F^A_{\mu\nu}\gamma_{2n+1})\sigma_{\mu\nu}+2u\eta\not{A}\gamma_{2n+1}+\cdots,$$

and

$$\bar{R}(\eta=1)=Q_+ +\cdots,$$

then make a similar replacement as above, and we come to the result of (4.5c).

After integrating over ξ and u, the nonvanishing part of $K_0^{(i)}(x,x,\mathscr{D})$ under $\mathrm{tr}\,\beta\gamma_{2n+1}$ (*) becomes, respectively,

$$\frac{1}{(4\pi)^n n!}\sum_{m=0}^{\infty}B(m+1,n+1)(m+n+1)\times\left[\frac{1}{(2m)!}\frac{\partial^{2m}}{\partial u^{2m}}Q_-^{n+m}(u)\right]_{u=0},\tag{4.7a}$$

$$\frac{-i}{(4\pi)^n n!}\sum_{m=0}^{\infty}B(m+1,n+1)(m+n+1)\times\left[\frac{1}{(2m+1)!}\frac{\partial^{2m+1}}{\partial u^{2m+1}}L^{\dagger}Q_-^{n+m}(u)\right]_{u=0},\tag{4.7b}$$

$$\frac{i}{(4\pi)^n n!}\sum_{m=0}^{\infty}B(m+1,n+1)\times\left[\frac{1}{(2m+1)!}\frac{\partial^{2m+1}}{\partial u^{2m+1}}\times\sum_{l_1+l_2=m+n}Q_+^{l_1}(u)L^{\dagger}Q_-^{l_2}(u)\right]_{u=0}.\tag{4.7c}$$

On the other hand, in the sense of neglecting the terms with γ matrices contracted, one can easily check

$$\not{A}Q_-=Q_+(i\not{V}-i\not{A}\gamma_{2n+1})-(i\not{V}-i\not{A}\gamma_{2n+1})Q_-+iu(Q_+-Q_-)+\cdots;\tag{4.8}$$

this leads to

$$L^{\dagger}Q'_-=Q'_+(i\not{V}-i\not{A}\gamma_{2n+1})+iu(Q'_+-Q'_-)+\cdots.\tag{4.9}$$

With the aid of the above relation, the summation over (4.7a), (4.7b), and (4.7c) can be simplified, the final result of $\mathrm{tr}\,\beta\gamma_{2n+1}K_0(x,x,\mathscr{D})$ can be written as

$$\mathrm{tr}\,\beta\gamma_{2n+1}K_0(x,x,\mathscr{D})=-\frac{1}{(4\pi)^n n!}\mathrm{tr}\left\{\beta\gamma_{2n+1}\sum_{m=0}^{n}B(m+1,n+1)\times\left[\frac{1}{(2m)!}\frac{\partial^{2m}}{\partial u^{2m}}\times\sum_{l_1+l_2=m+n}Q_+^{l_1}(u)Q_-^{l_2}(u)\right]_{u=0}\right\}.\tag{4.10}$$

We thus derive the general expression of the chiral Jacobian as shown in (1.2).

V. APPLICATIONS AND DISCUSSIONS

(1) An example. In practical applications, the formula (1.2) appears rather convenient. As an exmaple, one can easily obtain the Bardeen anomaly[12] in four-dimensional Euclidean space with the Dirac operator shown in (1.1) as

$$-\log J=\frac{1}{(4\pi)^2}\int d^4x\times\mathrm{tr}\{\beta\epsilon_{\mu\nu\rho\sigma}[(F^V_{\mu\nu}F^V_{\rho\sigma}+\tfrac{1}{3}F^A_{\mu\nu}F^A_{\rho\sigma})-\tfrac{8}{3}(F^V_{\mu\nu}A_\rho A_\sigma+A_\mu F^V_{\nu\rho}A_\sigma+A_\mu A_\nu F^V_{\rho\sigma})+\tfrac{32}{3}A_\mu A_\nu A_\rho A_\sigma]\}.\tag{5.1}$$

Here $(F^V_{\mu\nu}F^V_{\rho\sigma}+\frac{1}{3}F^A_{\mu\nu}F^A_{\rho\sigma})$ is simply read out from the terms in $\mathrm{tr}\,\beta\gamma_5B(1,3)(Q_+^2+Q_+Q_-+Q_-^2)$ with factor u^0, $-\frac{8}{3}(F^V_{\mu\nu}A_\rho A_\sigma+A_\mu F^V_{\nu\rho}A_\sigma+A_\mu A_\nu F^V_{\rho\sigma})$ is read out from the terms in $\mathrm{tr}\{\beta\gamma_5B(2,3)(Q_+^3+Q_+^2Q_-+Q_+Q_-^2+Q_-^3)\}$ with factor u^2, while $\frac{32}{3}A_\mu A_\nu A_\rho A_\sigma$ is from the terms in $\mathrm{tr}\{\beta\gamma_5B(3,3)(Q_+^4+Q_+^3Q_-+Q_+^2Q_-^2+Q_+Q_-^3+Q_-^4)\}$ with factor u^4.

(2) The case of $A=0$. In this case, the Dirac operator $\mathscr{D}=i\not{\partial}+i\not{V}$ is Hermitian, and $Q_{\pm}=F^V_{\mu\nu}$ is independent of z, and the formula (1.2) yields

$$-\log J=\frac{2}{(4\pi)^n(n+1)!}\int d^{2n}x\times\mathrm{tr}\{\beta(x)\epsilon_{\mu_1\nu_1\cdots\mu_n\nu_n}F_{\mu_1\nu_1}\cdots F_{\mu_n\nu_n}\};\tag{5.2}$$

this result coincides with that based on Fujikawa's regularization scheme.[1]

(3) The case of chiral gauge coupling. The Dirac operator in the case of chiral gauge coupling is

$$\mathscr{D} = i\partial\!\!\!/ + iA\!\!\!/^{(+)}\tfrac{1}{2}(1+\gamma_{2n+1}). \tag{5.3}$$

Since $A_\mu = V_\mu = \frac{1}{2}A_\mu^{(+)}$, $Q_\pm$ becomes

$$Q_\pm = F_{\mu\nu}^{(+)}\sigma_{\mu\nu}\hat{P}_\pm + uA\!\!\!/^{(+)}\gamma_{2n+1} \quad [\hat{P}_\pm = \tfrac{1}{2}(1+\gamma_{2n+1})], \tag{5.4}$$

where

$$F_{\mu\nu}^{(+)} = \partial_\mu A_\nu^{(+)} - \partial_\nu A_\mu^{(+)} + [A_\mu^{(+)}, A_\nu^{(+)}].$$

In the sense of preserving terms with $2n$ γ matrices, we can replace

$$\sum_{l_1+l_2=m+n} Q_+^{l_1} Q_-^{l_2}$$

by $(1-Q_+)^{-1}(1-Q)^{-1}$. By using the formulas of chiral expansions[13]

$$(1 - F_{\mu\nu}\sigma_{\mu\nu}\hat{P}_+ + A\!\!\!/\gamma_{2n+1})^{-1} = \hat{P}_- + (1-A\!\!\!/)(1 - F_{\mu\nu}\sigma_{\mu\nu} + A\!\!\!/A\!\!\!/)^{-1}A\!\!\!/\hat{P}_- + (1-A\!\!\!/)(1 - F_{\mu\nu}\sigma_{\mu\nu} + A\!\!\!/A\!\!\!/)^{-1}\hat{P}_+, \tag{5.5a}$$

$$(1 - F_{\mu\nu}\sigma_{\mu\nu}\hat{P}_- + A\!\!\!/\gamma_{2n+1})^{-1} = \hat{P}_+ + \hat{P}_+A\!\!\!/(1 - F_{\mu\nu}\sigma_{\mu\nu} + A\!\!\!/A\!\!\!/)^{-1}(1-A\!\!\!/) + \hat{P}_-(1 - F_{\mu\nu}\sigma_{\mu\nu} + A\!\!\!/A\!\!\!/)^{-1}(1-A\!\!\!/), \tag{5.5b}$$

we can expand $(1-Q_+)^{-1}(1-Q_-)^{-1}$ as

$$\begin{aligned}(1-Q_+)^{-1}(1-Q_-)^{-1} = {} & (1 - F_{\mu\nu}^{(+)}\sigma_{\mu\nu} + u^2A\!\!\!/^{(+)}A\!\!\!/^{(+)})^{-1} \\ & - u^2A\!\!\!/^{(+)}(1 - F_{\mu\nu}^{(+)}\sigma_{\mu\nu} + u^2A\!\!\!/^{(+)}A\!\!\!/^{(+)})^{-1}A\!\!\!/^{(+)}(1 - F_{\mu\nu}^{(+)}\sigma_{\mu\nu} + u^2A\!\!\!/^{(+)}A\!\!\!/^{(+)})^{-1} \\ & - u^2(1 - F_{\mu\nu}^{(+)}\sigma_{\mu\nu} + u^2A\!\!\!/^{(+)}A\!\!\!/^{(+)})^{-1}A\!\!\!/^{(+)}(1 - F_{\mu\nu}^{(+)}\sigma_{\mu\nu} + u^2A\!\!\!/^{(+)}A\!\!\!/^{(+)})^{-1}A\!\!\!/^{(+)} \\ & + (\text{terms with odd number of } \gamma\text{'s}).\end{aligned} \tag{5.6}$$

Thus $\operatorname{tr}\beta\gamma_{2n+1}(\cdots)$ in (4.10) becomes

$$\begin{aligned}&\operatorname{tr}\Big\{\beta\gamma_{2n+1}\sum_{m=0}^{n} B(m+1,n+1)\frac{1}{(2m)!}\frac{\partial^{2m}}{\partial u^{2m}}\Big[(F_{\mu\nu}^{(+)}\sigma_{\mu\nu} - u^2A\!\!\!/^{(+)}A\!\!\!/^{(+)})^n \\ &\quad - u^2\sum_{l=0}^{n-1}A\!\!\!/^{(+)}(F_{\mu\nu}^{(+)}\sigma_{\mu\nu} - u^2A\!\!\!/^{(+)}A\!\!\!/^{(+)})^lA\!\!\!/^{(+)}(F_{\mu\nu}^{(+)}\sigma_{\mu\nu} - u^2A\!\!\!/^{(+)}A\!\!\!/^{(+)})^{n-l-1} \\ &\quad - u^2\sum_{l=0}^{n-1}(F_{\mu\nu}^{(+)}\sigma_{\mu\nu} - u^2A\!\!\!/^{(+)}A\!\!\!/^{(+)})^lA\!\!\!/^{(+)}(F_{\mu\nu}^{(+)}\sigma_{\mu\nu} - u^2A\!\!\!/^{(+)}A\!\!\!/^{(+)})^{n-l-1}A\!\!\!/\Big]\Big\}_{u=0} \\ &= \operatorname{tr}\Big\{\beta\gamma_{2n+1}\int_0^1 dt\,[(\tilde{F}\!\!\!/^{(+)}\sigma_{\mu\nu})^n - t(1-t)\sum_{l=0}^{n-1}A\!\!\!/^{(+)}(\tilde{F}\!\!\!/^{(+)}\sigma_{\mu\nu})^lA\!\!\!/^{(+)}(\tilde{F}\!\!\!/^{(+)}\sigma_{\mu\nu})^{n-l-1} \\ &\quad - t(1-t)\sum_{l=0}^{n-1}(\tilde{F}\!\!\!/^{(+)}\sigma_{\mu\nu})^lA\!\!\!/^{(+)}(\tilde{F}_{\mu\nu}\sigma_{\mu\nu})^{n-l-1}A\!\!\!/^{(+)}\Big]\Big\},\end{aligned} \tag{5.7}$$

where the integral expression for the beta function,

$$B(m+1,n+1) = \int_0^1 t^n(1-t)^m\,dt,$$

is used and $\tilde{F}_{\mu\nu}^{(+)}$ is defined as

$$\tilde{F}_{\mu\nu}^{(+)} = F_{\mu\nu}^{(+)} - (1-t)[A_\mu^{(+)}, A_\nu^{(+)}]. \tag{5.8}$$

Now, with the aid of the notation of the exterior differential form $A = A_\mu\,dx^\mu$, $F = \frac{1}{2}F_{\mu\nu}\,dx^\mu\wedge dx^\nu$, $A_t = tA$, $F_t = dA_t + A_t^2$, we can express the chiral Jacobian in the case of gauge coupling as

$$-\log J = 2\frac{1}{(2\pi i)^n\cdot n!}\operatorname{Tr}\int_0^1 dt\,\{\beta(F_t^{(+)n} - nt(1-t)[A^{(+)}, P(A^{(+)}, F_t^{(+)n-1})])\}, \tag{5.9}$$

or in more compact form[7]

$$-\log J = 2\frac{1}{(2\pi i)^n(n-1)!}\int_0^1 dt(1-t)\times\operatorname{Tr}[\beta\,dp(A^{(+)}, F_t^{(+)n-1})], \tag{5.10}$$

in which $P(\Lambda_1,\ldots,\Lambda_n)$ is a symmetrized product.

(4) Gauge anomalies. The ζ-function regularization scheme can also be applied to derive the gauge anomaly. Rewrite the Dirac operator in gauge coupling as $\mathscr{D} = (i\partial\!\!\!/ + iA\!\!\!/^{(+)})\hat{P}_+ + i\partial\!\!\!/_-\hat{P}_-$ under gauge transformation $A^{(+)}\to g_+^{-1}(A^{(+)} + d)g_+$, $g_+ = \exp v_+(x)$, the Dirac transforms as $\mathscr{D}\to\mathscr{D}(g) = \mathscr{D} + \delta\mathscr{D}$ with $\delta\mathscr{D} = [\mathscr{D}\hat{P}_+, v_+]$. Here $v_+(x)$ is an infinitesimal quantity related to some gauge group $G^{(+)}$. In the path integral formalism, the Jacobian factor is related to the following equality:

$$\det\mathscr{D} = J\det\mathscr{D}(g). \tag{5.11}$$

Within the ζ-function regularization scheme, we can express $\log J$ as

$$\log J = \zeta'(0,\mathscr{D}(g)) - \zeta'(0,\mathscr{D}). \tag{5.12}$$

By applying the ζ-function differentiability results,[3] we have

$$\log J = -\frac{d}{ds}[s\,\mathrm{Tr}(\mathscr{D}^{-s-1}\delta\mathscr{D})]_{s=0}$$
$$= -\frac{d}{ds}[s\,\mathrm{Tr}(v_+\hat{P}_+\mathscr{D}^{-s}) - s\,\mathrm{Tr}(v_+\hat{P}_-\mathscr{D}^{-s})]_{s=0}$$
$$= -\frac{d}{ds}[s\,\mathrm{Tr}(v_+\gamma_{2n+1}\mathscr{D}^{-s})]_{s=0}. \tag{5.13}$$

Compare the expression above with that of (2.5), a factor 2 has disappeared. So we can easily obtain a formula in exterior differential forms for gauge anomaly from (4.10) as

$$\log J = \frac{1}{(2\pi i)^n(n-1)!} \times \int_0^1 dt(1-t)\mathrm{Tr}[\beta\, dP(A^{(+)},F_t^{(+)n-1})]. \tag{5.14}$$

(5) The chiral Jacobian in differential forms. In the form of chiral projection, the Dirac operator (1.1) can also be written as

$$\mathscr{D} = (i\partial\!\!\!/ + iA\!\!\!/^{(+)})\hat{P}_+ + (i\partial\!\!\!/ + iA\!\!\!/^{(-)})\hat{P}_-, \tag{5.15}$$

with $A^{(\pm)} = V \pm A$. Using a similar treatment as exhibited in the case of chiral gauge coupling, the formula shown in (1.2) can recast into exterior differential forms of $A^{(\pm)}$ as

$$-\log J(A^{(+)},A^{(-)}) = \frac{1}{(2\pi)^{n+1}(n+1)!}\int_{S^{2n}}\omega_{2n}^1(A^{(+)},A^{(-)}), \tag{5.16}$$

where

$$\omega_{2n}^1(A^{(+)},A^{(-)}) = (n+1)\int_0^1 dt\,\mathrm{str}\{2\beta,F^n(t)\} - n(n+1)\int_0^1 dt\,t(1-t) \times \mathrm{str}\{[2\beta,A^{(+)}-A^{(-)}], A^{(+)}-A^{(-)},F^{n-1}(t)\} \tag{5.17}$$

with

$$F(t) = tF_+ + (1-t)F_- - t(1-t)(A^{(+)}-A^{(-)})^2. \tag{5.18}$$

Furthermore, by solving the equation $d\omega_{2n}^1 = \delta_\beta\omega_{2n+1}$ under condition $\delta^2\omega_{2n+1} = 0$ we obtain

$$\omega_{2n+1}(A^{(+)},A^{(-)}) = (n+1)\int_0^1 dt\{A^{(+)}-A^{(-)},F^n(t)\}. \tag{5.19}$$

This is in agreement with the result of Lott using cohomology,[14] and therefore the topological invariant associated with the Dirac operator (5.15) must be

$$[(2\pi)^{n+1}(n+1)!]^{-1}(\mathrm{tr}\,F^{(+)n+1} - \mathrm{tr}\,F^{(-)n+1}), \tag{5.20}$$

it is related to the Atiyah–Singer index density in $2(n+1)$-dimensional space.[15]

As indicated in our previous paper,[16] we may conclude that all the differential geometric objects in the approach of anomalies, both Abelian or non-Abelian, can be traced out oppositely under the properly selected regularization scheme in the path integral formalism.

ACKNOWLEDGMENTS

I would like to thank Professor Guang-jiong Ni for valuable discussions. I am also grateful to Professor Ngee-pong Chang for helpful conversations and hospitality at the City College of New York.

The work is partly supported by the Science Foundation of National Education Committee of China under Grant No. KF 12016.

[1] K. Fujikawa, Phys. Rev. Lett. **42**, 1195 (1979); Phys. Rev. D **21**, 2848 (1980).
[2] S. W. Hawking, Commun. Math. Phys. **55**, 133 (1977).
[3] R. E. Gamboa Saraví, M. A. Muschietti, and J. E. Solomin, Commun. Math, Phys. **89**, 393 (1983).
[4] R. T. Seeley, Am. Math. Soc. Proc. Symp. Pure Math. **10**, 288 (1967).
[5] R. E. Gamboa Saraví, M. A. Muschietti, F. A. Schaponsnik, and J. E. Solomin, Ann. Phys. (NY) **157**, 360 (1984).
[6] R. E. Gamboa Saraví, M. A. Muschietti, F. A. Schaponsnik, and J. E. Solomin, J. Math. Phys. **26**, 2045 (1985).
[7] B. Zumino, Y. S. Wu, and A. Zee, Nucl. Phys. B. **239**, 477 (1984).
[8] L. Alvarez-Gaumé and P. Ginsparg, Ann. Phys. (NY) **161**, 423 (1985).
[9] M. F. Atiyah and I. M. Singer, Proc. Natl. Acad. Sci USA **81**, 2598 (1984).
[10] A. Agmon, Commun. Pure Appl. Math. **15**, 119 (1982).
[11] R. E. Gamboa-Saraví, M. A. Muschietti and J. F. Solomin, Commun. Math. Phys. **93**, 407 (1984).
[12] W. A. Bardeen, Phys. Rev. **184**, 1948 (1969).
[13] J. M. Gipson, Phys. Rev. D **33**, 1061 (1986).
[14] J. A. Lott, Commun. Math. Phys. **97**, 371 (1985).
[15] M. Atiyah and I. M. Singer, Ann. Math. **93**, 119 (1971).
[16] R-t. Wang and G-j. Ni, J. Phys. A **21**, 1811 (1988).

Hamilton's form for the Kadomtsev–Petviashvili equation

Mladen Victor Wickerhauser
Department of Mathematics, University of Georgia, Athens, Georgia 30602

(Received 17 September 1987; accepted for publication 1 June 1988)

The scattering transform for the Kadomtsev–Petviashvili equation (KPII) is a local symplectomorphism. Pulling back the Hamiltonians for the linear evolutions of scattering data gives Hamiltonians for the KPII hierarchy: they are values of the associated scattering data at distinguished points. This method yields simple proofs that KPII has infinitely many commuting flows and simplifies their calculation. It also provides a Plancherel-type theorem.

I. INTRODUCTION

Fadeev and Zakharov[1,2] have observed that the scattering map that underlies the inverse scattering method is a canonical transformation to "action-angle variables" for a number of $(1+1)$-dimensional nonlinear evolution equations: sine-Gordon, nonlinear Schrödinger, Korteweg–de Vries, and the Toda lattice. Without dwelling on the analysis, they discovered the appropriate symplectic structures, Hamiltonian functions, and Poisson brackets for these equations. The intention in this paper is to present similar results for the $(2+1)$-dimensional Kadomtsev–Petviashvili (KPII) equation $\partial_1(\dot{u}+\partial_1^3 u+u\,\partial_1 u)+\partial_2^2 u=0$, and other equations in its hierarchy. An application of some previous work will, in addition, describe the underlying symplectic manifolds in a neighborhood of their distinguished point 0. This analytic information will be used to prove that the canonical transformation is in fact a (local) symplectomorphism.

Earlier,[3] we observed that certain values of the scattering data associated to a function u were constants of all KP-type motions of u. In this paper we will show that these constants of the motion may be used as Hamiltonian functionals which, together with a natural symplectic structure on the manifold of solutions, put the KP evolution into an obvious Hamiltonian form. Such a structure has been found for certain one-dimensional evolutions by similar means.[4]

The advantages of putting the KP equation in Hamiltonian form are that symmetries and constants of motion may be very easily calculated. In particular, new quantities that are preserved under the KP flow may be explicitly written as integrals of polynomials in the potential and its derivatives with no more sophisticated a mathematical tool than the multiplication of power series. It may be of interest to consider what physical meaning may be assigned to these constants of the motion, as they are composed of polynomials in the terms of the KP equation itself.

Results cited in this paper answer a question of Lipovskii[5] who has independently discovered the Poisson bracket for the KP Hamiltonian system under the assumption that the scattering data satisfy certain boundedness properties.

These results reproduce the construction of the Korteweg–de Vries (KdV) Hamiltonian system[2] under the restriction $du/dy=0$. This implies that both $dm/dy=0$ and $d\tilde{m}/dy=0$, resulting in the simplification $\tilde{m}=m$. One well-known but important consequence of this last equation is that the scattering problem for the KdV equation is self-adjoint, and that the recursion operator which generates the KdV hierarchy has m^2 as an eigenfunction—it is the so-called square-eigenfunction operator. Zakharov and Konopelchenko have shown that no single recursion operator exists for the $(2+1)$- or higher-dimensional examples.[6] Here, their obstruction is overcome by introducing an only slightly more complicated formula for elements of the hierarchy.

Much of the novelty of the KPII equation develops from its non-self-adjoint scattering problem, and the resulting difficulty in the physical interpretation of eigenvalues. The method used in this paper departs from classical S-matrix scattering theory in that no explicit physical meaning is assigned to the function α. It is merely called scattering data for historical reasons and by analogy with the KdV and other $(1+1)$-dimensional equations. It has been shown by Beals, Coifman, and others that the classical S matrix depends upon too many variables to be in one-to-one correspondence with potentials for any example equation depending upon two or more space variables. As a consequence, there must be constraints upon S, which are in general nonlinear, and which it is difficult to show are preserved under linear evolutions. Without these constraints, the inverse map does not exist. Hence the classical scattering transform is not a symplectomorphism. It is necessary to use the "unphysical" scattering data of this paper to linearize the KP equation and write it in Hamiltonian form.

II. KP HAMILTONIANS

Let $B=\{u\in L^1\cap L^2(\mathbb{R}^2)\,|\,\int_{-\infty}^{\infty}u(x_1,x_2)dx_1=0$ a.e. $x_2\}$. Then B is a closed subspace of $L^1\cap L^2$. We shall be concerned with an open neighborhood of 0 in B: this will contain the integral curves of all KP-type evolutions from small initial conditions.

Since B is a Banach space, we can identify it with its tangent spaces at each point: for all $b\in B$, $T_bB\cong B$. Also, its tangent bundle is trivial and may be identified as $TB\cong B\times B$. Hence for convenience we shall define our two-forms on $B\times B\cong(TB/B)\times(TB/B)$, suppressing the dependence on base points in TB. We shall do likewise for the (linear Banach) manifold of scattering data and its (trivial) tangent bundle.

Introduce the bilinear form $\Omega: B\times B\to\mathbb{C}$ by

$$\Omega(u,v)=\langle\partial_1^{-1}u,v\rangle\,, \tag{1}$$

where

 0022-2488/88/102300-03$02.50

$$(\partial_1^{-1}u)(x_1,x_2) = \int_{-\infty}^{x_1} u(y,x_2)dy$$

and

$$\langle a,b\rangle = \int_{\mathbb{R}^2} ab\,.$$

It is a simple exercise to show that Ω is both skew symmetric and weakly nondegenerate, and hence is a "symplectic" form on the linear Banach manifold B.

Recall (see, for example, Ref. 3) that the KP evolution may be linearized by mapping a solution u to scattering data associated to the perturbed heat operator

$$\partial_1^2 + 2z\partial_1 - \partial_2 - u \equiv L_z - u\,. \tag{2}$$

After finding $m = m(x,z)$, the solution to $L_z m = um$ with $m \to 1$ as $|x| \to \infty$, the scattering data may be written as

$$\alpha(z) = \int_{\mathbb{R}^2} u(x)m(x,z)\exp\{x_1(\bar{z}-z) + x_2(\bar{z}^2 - z^2)\}dx\,. \tag{3}$$

Then KP-type evolutions of u correspond to evolutions of α of the form

$$\alpha(z,t) = \alpha(z)\exp t\{\phi(\bar{z}) - \phi(z)\}, \tag{4}$$

where $\phi: \mathbb{C} \to \mathbb{C}$ is suitably chosen. For example, the KP equation itself arises from $\phi(z) = z^3$.

Observe that if z is real, then

$$\frac{d}{dt}\alpha(z,t) = [\phi(\bar{z}) - \phi(z)]\alpha(z,t) = 0\,. \tag{5}$$

Hence $\alpha(z,t)$, $z \in \mathbb{R}$, is a constant of every KP-type motion of U. Fixing $z \in \mathbb{R}$, we can use $\alpha(z,t)$ as a Hamiltonian function on B: Let

$$H(u) = \alpha(z) = \int_{\mathbb{R}^2} u(x)m(x,z)dx\,. \tag{6}$$

Here $\bar{z} - z = 0$ and $\bar{z}^2 - z^2 = 0$, simplifying Eq. (3), and m solves Eq. (2). Notice that we can take

$$m(x,z) = (I - G_z M_u)^{-1}1\,, \tag{7}$$

where G_z is convolution with the Green's function of L_z [in Eq. (2)] and M_u is multiplication by $u(x)$. Such solutions m are normalized, in the sense that $\lim_{|x|\to\infty} m(x,z) = 1$. The existence and uniqueness of m in $L^\infty(\mathbb{R}^2)$ is guaranteed for u near 0 in B (see Ref. 3). Thus

$$H(u) = \int_{\mathbb{R}^2} u(I - G_z M_u)^{-1}1\,dx\,. \tag{8}$$

It is evident what the corresponding Hamiltonian vector field is. First, we calculate the gradient of H,

$$\langle dH(u),r\rangle \equiv \frac{d}{d\epsilon}H(u+\epsilon r)\Big|_{\epsilon=0}$$

$$= \int_{\mathbb{R}^2} (r(x)m(x,z) + u(x)(I - G_z M_u)^{-1} \times G_z r(x)m(x,z))dx; \tag{9}$$

then combining terms,

$$= \int_{\mathbb{R}^2} (I - M_u G_z)^{-1} r(x)m(x,z)dx;$$

and transposing,

$$= \int_{\mathbb{R}^2} [(I - G_z^t M_u)^{-1}1]r(x)m(x,z)dx\,.$$

Here we make the definition

$$\tilde{m}(x,z) = (I - G_z^t M_u)^{-1}1\,. \tag{10}$$

This yields

$$\langle dH(u),r\rangle = \int_{\mathbb{R}^2} \tilde{m}(x,z)r(x)m(x,z)dx\,,$$

and (by the nondegeneracy of the inner product)

$$dH(u) = \tilde{m}(x,z)m(x,z)\,. \tag{11}$$

Second, we calculate the Hamiltonian vector field associated to H,

$$\begin{aligned} &\Omega(X_H,r) = \langle dH,r\rangle\,,\\ &X_H(u) = \partial_1(\tilde{m}(x,z)m(x,z))\,. \end{aligned} \tag{12}$$

These natural choices for the symplectic structure and for the functional have led us directly to the fundamental evolution as Hamilton's equation of motion:

$$\frac{d}{dt}u = X_H(u) = \partial_1(\tilde{m}(x,z)m(x,z))\,. \tag{13}$$

III. COMMUTING FLOWS

Let H, K be two functionals on B. Define the Poisson bracket as usual:

$$\{H,K\} = \Omega(X_H,X_K)\,. \tag{14}$$

Theorem: If z, w are distinct reals, and we set $H(u) = \alpha(z)$, $K(u) = \alpha(w)$, then $\{H,K\} = 0$.

Proof:

$$\{H,K\} = \int_{\mathbb{R}^2} \tilde{m}(x,z)m(x,z)\partial_1(\tilde{m}(x,w)m(x,w))dx\,.$$

Now

$$\partial_1(\tilde{m}(x,w)m(x,w)) = \frac{d}{dt}u\,,$$

and for this fundamental evolution of u one has

$$\frac{d}{dt}\alpha(\zeta) = [(\bar{\zeta} - w)^{-1} - (\zeta - w)^{-1}]\alpha(\zeta)$$

as in Eq. (5).

But also, this integral is the tangent map for the scattering transform,

$$\frac{d\alpha(z)}{dt} = \int_{\mathbb{R}^2} \tilde{m}(x,z)\left[\frac{d}{dt}u(x)\right]m(x,z)dx\,.$$

So since z is real,

$$\{H,K\} = [(\bar{z} - w)^{-1} - (z - w)^{-1}]\alpha(z) = 0\,. \qquad \square$$

Remark: If $z = w$, then $H = K$ and the antisymmetry of Ω guarantees that $\{H,K\} = 0$.

This provides a simple proof that there are infinitely many commuting flows: there is a different one for each choice of $z \in \mathbb{R}$. To obtain the Hamiltonian for the k th evolution in the KPII hierarchy, one takes the appropriate combination of these fundamental flows. The fundamental evolution whose Hamiltonian function is given by $H_z(u) = \alpha(z)$ can be written

$$\dot{u} = \partial_1(m(x,z)\tilde{m}(x,z)) = \sum_{k=1}^{\infty} \frac{\partial_1 M_k(x)}{z^k},$$

where $M_k(x)$ is a moment of $m\tilde{m}$ and gives the k th evolution in the KPII hierarchy. We can exploit the smooth dependence of H_z on z to pick out the k th moment. Writing $X_H(z)$ for the Hamiltonian vector field given by the functional $u \to \alpha(z)$, then the k th moment is extracted by

$$\frac{1}{k!}\frac{d^k}{dz^k} X_H\left(\frac{1}{z}\right)\Big|_{z=0}.$$

Alternatively, one can think of this as the Hamiltonian vector field corresponding to the functional

$$u \to \frac{1}{k!}\frac{d^k}{dz^k}\alpha\left(\frac{1}{z}\right)\Big|_{z=0}$$

since operations in z commute with everything in sight. The KPII equation corresponds to $k = 3$.

IV. SYMPLECTOMORPHISM

There is also a symplectic structure on the (linear) manifold of scattering data that puts all KP-type evolutions into (linear) Hamiltonian form.

Let $C = L^2(\mathbb{C}; d\bar{z}\,dz)$.

Define ω: $C \times C \to \mathbb{C}$ by

$$\omega(\alpha,\beta) = \int_{\mathbb{C}} \operatorname{sgn}(\operatorname{Im} z)\alpha(z)\beta(\bar{z})d\bar{z}\,dz. \tag{15}$$

Evidently ω is skew-symmetric bilinear and weakly nondegenerate. By setting A equal to multiplication by $\phi(\bar{z}) - \phi(z)$, which is ω skew, one has by a familiar argument (Ref. 7, p. 459) that A satisfies the following proposition.

Proposition: A is Hamiltonian on some domain in C, and the equation $\dot{\alpha} = A\alpha = [\phi(\bar{z}) - \phi(z)]\alpha$ is the equation of motion for the functional $H_A(\alpha) = \frac{1}{2}\omega(A\alpha,\alpha)$.

These Hamiltonian systems on (B,Ω) and (C,ω) are related by the forward and inverse scattering maps. Denote these by S: $B \to C$ and S^{-1}: $C \to B$. Then as derived in Ref. 3,

$$S(u)(z) = k \operatorname{sgn}(\operatorname{Im} z) \int_{\mathbb{R}^2} u(x)m(x,z)\exp x\cdot v(z)dx, \tag{16}$$

$$S^{-1}(\alpha)(x) = k'\frac{\partial}{\partial x_1}\int_{\mathbb{C}} \operatorname{sgn}(\operatorname{Im} z)\alpha(z)m(x,\bar{z}) \times \exp x\cdot v(\bar{z})d\bar{z}\,dz, \tag{17}$$

where $v(z) = (z - \bar{z}, z^2 - \bar{z}^2) \in \mathbb{R}^2$, $z \in \mathbb{C}$, and k, k' are constants, $k = 1/2\pi$ and $k' = 1/2\pi^2$.

Using the relationship between u and m, or between α and m, it is a simple matter to compute the gradients $dS(u)$ and $dS^{-1}(\alpha)$,

$$dS(u)(x) = k \operatorname{sgn}(\operatorname{Im} z)m(x,z)\tilde{m}(x,\bar{z})\exp x\cdot v(z), \tag{18}$$

$$dS^{-1}(\alpha)(z) = k'\frac{\partial}{\partial x_1}(m(x,\bar{z})\tilde{m}(x,z)\exp x\cdot v(\bar{z})), \tag{19}$$

where $\tilde{m}$ is defined at u by Eq. (10) and at α by $\tilde{m}$ at $S^{-1}(\alpha)$. Extending both Ω and ω to the tangent bundles TB and TC in the natural way yields the following Plancherel-type result.

Proposition: If $\alpha = S(u)$, then

$$\Omega(dS^{-1}(\alpha)\cdot\dot{\alpha},\dot{u}) = c\omega(\dot{\alpha},dS(u)\cdot\dot{u}),$$

where $c = -k'/k = i/\pi$.

Proof:

$$\begin{aligned}
&\Omega(dS^{-1}(\alpha)\cdot\dot{\alpha},\dot{u}) \\
&= \int_{\mathbb{R}^2} \partial_1^{-1}(dS^{-1}(\alpha)\cdot\dot{\alpha})(x)\dot{u}(x)dx \\
&= \int_{\mathbb{R}^2} \partial_1^{-1}\Big[k'\partial_1\int_{\mathbb{C}} m(x,\bar{z})\tilde{m}(x,z) \\
&\quad \times \exp x\cdot v(\bar{z})\dot{\alpha}(z)d\bar{z}\,dz\Big]\dot{u}(x)dx \\
&= k'\int_{\mathbb{C}} \dot{\alpha}(z)\Big[\int_{\mathbb{R}^2} m(x,\bar{z})\tilde{m}(x,z) \\
&\quad \times \exp x\cdot v(\bar{z})\dot{u}(x)dx\Big]d\bar{z}\,dz \\
&= -\frac{k'}{k}\int_{\mathbb{C}} \operatorname{sgn}(\operatorname{Im} z)\dot{\alpha}(z)(dS(u)\cdot\dot{u})(\bar{z})d\bar{z}\,dz \\
&= -\frac{k'}{k}\omega(\dot{\alpha},dS(u)\cdot\dot{u}).
\end{aligned}$$

Theorem: The scattering and inverse scattering maps are symplectomorphisms between (B,Ω) and (C,ω).

Proof: That S and S^{-1} are local isomorphisms between B and C follows from Ref. 3. Both S and S^{-1} are real analytic in the sense of Coifman and Meyer,[8] being expressible as power series in a functional variable. Hence they are both local diffeomorphisms.

It remains to show that $c\omega = S^*\Omega$, where c is some constant, in fact the same one as above. But this follows from the last proposition. If $u \in B$, $p,q \in T_uB$, $\alpha \in C$, $\xi,\eta \in T_\alpha C$, and everything is related by $\alpha = S(u)$, $\xi = dS(u)\cdot p$, $\eta = dS(u)\cdot q$, then

$$\begin{aligned}
\omega(\xi,\eta) &= \omega(dS(u)\cdot p, dS(u)\cdot q) \\
&= c^{-1}\Omega(dS^{-1}(\alpha)\circ dS(u)\cdot p,q) = c^{-1}\Omega(p,q). \qquad \square
\end{aligned}$$

ACKNOWLEDGMENTS

The author is grateful to Professor R. R. Coifman for many insightful conversations, and to Professor M. Adams for his critical reading.

This work was partially supported by the University of Georgia Research Foundation.

[1] L. D. Fadeev, "A Hamiltonian interpretation of the inverse scattering method," in *Solitons*, edited by R. K. Bullough and P. J. Caudrey (Springer, New York, 1981).

[2] V. E. Zakharov and L. D. Fadeev, "Korteweg–de Vries equation: A completely integrable Hamiltonian system," Functional Anal. Appl. **5**, 18 (1971).

[3] M. V. Wickerhauser, "Inverse scattering for the heat operator and evolutions in 2 + 1 variables," Commun. Math. Phys. **108**, 67 (1987).

[4] R. Beals and R. R. Coifman, "Scattering and inverse scattering for first order systems II," Yale preprint, 1987.

[5] V. D. Lipovskii, "Hamiltonian structure of the Kadomtsev–Petviashvili equation in the class of decreasing Cauchy data," Functional Anal. Appl. **20**, 35 (1986).

[6] V. F. Zakharov and B. G. Konopelchenko, "On the theory of recursion operator," Commun. Math. Phys. **94**, 483 (1984).

[7] R. Abraham and J. E. Marsden, *Foundations of Mechanics* (Benjamin/Cummings, Reading, MA, 1978), 2nd ed.

[8] R. R. Coifman and Y. Meyer, "Nonlinear harmonic analysis and analytic dependence," Proc. Symp. Pure Math. **43**, 71 (1985).

Liouville theorem for the Yang–Mills self-duality equations

Shahn Majid
Lyman Laboratory of Physics, Harvard University, Cambridge, Massachusetts 02138

(Received 14 August 1987; accepted for publication 18 May 1988)

It is shown that under certain conditions one may associatively matrix-multiply Lie-algebra-valued matrices with a componentwise Lie bracket. Using this, a simple algebraic constraint on a Lie-algebra-valued antisymmetric $n\times n$ matrix F, which in $n=4$ is essentially self-duality or anti-self-duality, is described. Somewhat in analogy with Liouville's theorem for the Cauchy–Riemann equations in $n=2$, it is shown that, for $n>4$, the constraint implies that the Lie subalgebra generated by the matrix elements $\{F_{\mu\nu}\}$ decomposes into copies of $\underline{SO}(n)$ plus a few degenerate cases. The result may be relevant to the structure of the quantum chromodynamic vacuum.

I. INTRODUCTION

The results in this paper, Theorem 1 and Theorem 2, are purely linear algebra concerning matrix multiplication with a Lie bracket of matrices $A,B\in M(n,\underline{G})=n\times n$ matrices with values in a Lie algebra $\underline{G}$, possibly infinite dimensional. By matrix multiplication we shall always mean

$$A\cdot B_\mu{}^\nu=[A_\mu{}^\alpha,B_\alpha{}^\nu]\,,$$

where $[\ ,\ ]$ denotes the Lie bracket. Namely, we will show that if an antisymmetric Lie-algebra-valued matrix F obeys $W(F,F)=0$ in the expression (1.5) below, and if $n>4$, then all higher matrix products-with-commutators F^m are antisymmetric and obey (1.8). This result complements a standard product encountered in Yang–Mills gauge theory, namely, the exterior product-with-commutators of Lie-algebra-valued forms.

It is remarkable that the proof of this little fact will take several pages of tensor algebra, albeit index manipulations of a sort with which theoretical physicists are surely familiar. The author has not been able to come up with an abstract proof as more usual methods (for analyzing the decomposition of tensor products) do not seem to be applicable in a useful way to the case of multiplication-with-commutators. Therefore a few motivating remarks are perhaps in order. The first two have to do with the geometrical setting concerning Yang–Mills connections of high symmetry, while the third is algebraic. These remarks are not used in the paper, which will be entirely taken up with the proof of the above statement.

Remark 1 (Liouville theorem for the self-duality equations): On a Riemannian manifold of dimension n, (M,g,∇), the derivative of vector field ξ can always be decomposed under the action of $O(n)$ as

$$\nabla_\mu\xi_\nu\equiv(\!(\nabla\xi)\!)_{\mu\nu}+\tfrac{1}{2}\nabla\wedge\xi_{\mu\nu}+(g_{\mu\nu}/n)\nabla\cdot\xi\,,\qquad(1.1)$$

where $(\!(\nabla\xi)\!)$ denotes the symmetric traceless part, $\nabla\wedge\xi=d\wedge\xi$ is the curl, and $\nabla\cdot\xi$ is the divergence. The equation

$$(\!(\nabla\xi)\!)=0\qquad(1.2)$$

says that ξ is a conformal Killing vector. The Lie bracket of two conformal Killing vectors is again one; so these generate a group, the conformal group (diffeomorphisms of M that preserve g up to scale). In $n=2$ on $\mathbb{R}^2$ with the flat metric and standard coordinates x, (1.2) is just the Cauchy–Reimann equations for $\xi_1+\iota\xi_2$ as a function of $z=x_1+\iota x_2$,

$$(\!(\nabla\xi)\!)=0\Leftrightarrow\frac{\partial\xi_1}{\partial x_2}=\frac{\partial\xi_2}{\partial x_1}\quad\text{and}\quad\frac{\partial\xi_1}{\partial x_1}=\frac{1}{2}\partial\cdot\xi=\frac{\partial\xi_2}{\partial x_2}$$
$$\Leftrightarrow\xi(z)\ \text{analytic}\quad(\text{for } n=2)\,,\qquad(1.3)$$

and Liouville's theorem[1] asserts that

$$\text{conformal group}\,(\mathbb{R}^n)=O(n+1,1)\quad(\text{for } n>2)\,.\qquad(1.4)$$

For a general Riemannian manifold the corresponding assertion is

$$\dim\{\xi:(\!(\nabla\xi)\!)=0\}\leqslant\dim O(n+1,1)\quad(\text{for } n>2)\,.$$

Similarly, let F be an antisymmetric Lie-algebra matrix [or, more generally, let F be the curvature two-form of a connection on a principal bundle P over M with structure group G, $F\in\Omega^2(M)\otimes\text{ad}\,P_G$]. Such a matrix can be decomposed under the action of $O(n)$ as

$$\equiv W(F,F)_{\mu\nu\alpha\beta}+[1/(n-2)](g_{\nu\alpha}F^2_{\mu\beta}-g_{\mu\alpha}F^2_{\nu\beta}$$
$$-g_{\nu\beta}F^2_{\mu\alpha}+g_{\mu\beta}F^2_{\nu\alpha})\,,\qquad(1.5)$$

which defines the tensor $W(F,F)$, where $F^2\equiv F\cdot F$ multiplied as a matrix (using g), and with commutators in the Lie algebra $\underline{G}$. A similar decomposition is well known for the Riemann tensor, where it proved very interesting to study the case when the Weyl tensor vanishes. Therefore consider F, antisymmetric, for which

$$W(F,F)=0\,,\qquad(1.6)$$

which turns out in $n=4$ as essentially the self-duality or anti-self-duality equations

$$W(F,F)=0\Leftrightarrow[F_{+\mu\nu},F_{-\alpha\beta}]=0$$
$$\overset{G=SU(2),SU(3),U(2)}{\Leftrightarrow}\qquad F_+=0\ \text{or}\ F_-=0\ \text{or Abelian}$$
$$(\text{for } n=4)\,,\qquad(1.7)$$

where $F_\pm$ are the self-dual and anti-self-dual parts of F. The proof of this is essentially given in Lemma 4.6 and Propositions 7.1 and 7.2 of Ref. 2, and also below in Eq. (2.18). For $n>4$ our Theorem 1(ii) asserts that higher-matrix prod-

 0022-2488/88/102303-08$02.50

ucts-with-Lie-bracket F^m are also of this type, and indeed

$$W(F,F)=0 \Rightarrow F^m \text{ antisymmetric}, \quad \forall m \geqslant 1,$$
$$[F^{m_1}_{\mu\nu}, F^{m_2}_{\alpha\beta}] = [1/(n-2)](g_{\nu\alpha}F^{m_1}_{\mu\beta}{}^{+m_2} - g_{\mu\alpha}F^{m_1}_{\nu\beta}{}^{+m_2} - g_{\nu\beta}F^{m_1}_{\mu\alpha}{}^{+m_2} + g_{\mu\beta}F^{m_1}_{\nu\alpha}{}^{+m_2})$$
$$(\text{for } n>4). \tag{1.8}$$

So matrix multiplication generates a ring whose elements F^m span as a vector space the Lie algebra of the primitive holonomy group of F (which is by definition at each point in M the Lie subalgebra generated by the matrix elements $\{F_{\mu\nu}\} \subset \underline{G}$). Theorem 2 asserts that generically, for $n>4$, the primitive holonomy group Lie algebra is equal to

$$\{F_{\mu\nu}\} = \bigoplus_{i=1}^{q} \underline{SO}(n)_{\mathbb{C}}, \tag{1.9}$$

for some integer q. Here "generically" means that the roots of a certain characteristic polynomial should be nondegenerate, typical exceptions being when F is nilpotent under matrix multiplication. This is a completely local theorem and applies at each point in M independently.

A study of connections whose curvature obeys constraints such as (1.6), i.e., differential geometric aspects aiming at a structure theorem (cf. that of locally symmetric spaces[3]), has been undertaken and may be reported elsewhere. Examples of $W(F,F)=0$ in $n=4$ are provided in Ref. 2 by the curvature of connections on S^4 that are minima of the Yang–Mills action (i.e., for which the Hessian of second derivatives is $\geqslant 0$). The authors thereby show that all such connections for $G=\mathrm{SU}(2)$, $\mathrm{SU}(3)$, or $\mathrm{U}(2)$ are either self-dual or anti-self-dual. An example in $n \geqslant 4$ is provided by the one-quasi-instanton/anti-instanton whose $O(n)$ symmetry was exploited in Ref. 4. According to our Theorem 2, this is the prototypical example. This is recalled in the Appendix along with an estimate of how hard it is in general for matrices to obey (1.6).

Remark 2 (Significance for the quantum chromodynamic vacuum): With regard to the physics literature, many authors have previously studied generalizations of the self-duality equations[5] to higher dimensions.[6] These have all been concerned with algebraic constraints on the curvature such that the Yang–Mills equations or some related equation follows automatically. The present work has also been motivated by physics. In the background field method to the quantum field theory of gauge fields,[7] one wishes to minimize the effective action due to all particles in the theory moving in background A. This effective action—not at all the Yang–Mills action—is impossible to evaluate in general. However, the minimum, the expectation value of the quantized gauge field in the quantum chromodynamic (QCD) vacuum configuration, is expected to be highly symmetric. One may therefore try to guess or classify the connections with high symmetry and look for the vacuum connection by minimizing only among these. Unfortunately our Theorem 2, which would be highly restrictive, is not directly applicable to $n=4$, but one may note that the pole in our proof, $1/(n-4)$, is quite reminiscent of the uv divergences of QCD when parametrized by dimensional regularization as, for example, in Ref. 4, already referred to in remark 1. What is needed is a structure theorem for connections of high symmetry. For example, in the $G=\mathrm{U}(1)$ case, the only Poincaré-invariant connections have $F=0$.

Remark 3 [Kac–Moody algebra $\widetilde{SO(n)}$]: It is perhaps worth pointing out that Kac–Moody loop-group algebras (which were recently connected with the Yang–Mills self-duality equation[8]) provide an example of Theorem 1(ii). This is also given in the Appendix.

Returning to the paper, in order to prove Theorem 2 we shall actually be forced to prove a more general theorem to the effect that antisymmetric matrices A, B, C obeying certain conditions including $W(A,B)=0$ are closed under a matrix multiplication similar to that above which becomes commutative and associative up to certain scale terms involving $g_{\mu\nu}$. This will be the main theorem, Theorem 1. For various reasons we refer to this algebra as "conformal" algebra.

II. ON THE MULTIPLICATION OF LIE-ALGEBRA-VALUED MATRICES

In the rest of this paper we work entirely locally, mostly with antisymmetric matrices with values in $\underline{G}$, a Lie algebra over $\mathbb{R}$, possibly infinite dimensional. In fact Theorem 1, the main theorem, applies whenever there is an anticommutative bilinear bracket obeying the Jacobi identity on $\underline{G}$, a vector space over a field of suitable characteristic:

$$[a,[b,c]] = [[a,b],c] + [b,[a,c]],$$
$$[a,b] = -[b,a], \quad a,b,c \in \underline{G}. \tag{2.1}$$

For convenience one can also consider all matrices as two-forms with one index raised by a metric g a symmetric positive definite matrix. There is a natural operator generating the action of $\mathrm{SO}(n)$ preserving this g, the spin operator

$$S_{\mu\nu}(v_\alpha) \equiv (Sv)_{\mu\nu\alpha} = g_{\nu\alpha}v_\mu - g_{\mu\alpha}v_\nu, \tag{2.2}$$

which will play the role of an identity. On the space $M(n,\underline{G})$ we will always denote by $\cdot$ or by omission the product

$$\cdot : M(n,G) \otimes M(n,\underline{G}) \to M(n,\underline{G}) : A_{\mu\sigma} \otimes B_{\tau\nu} \mapsto [A_{\mu\alpha}, B^{\alpha}{}_{\nu}], \tag{2.3}$$

which is of course nonassociative in general. The definitions

$$\Lambda^2 \otimes \underline{G} = \{A \in M(n,\underline{G}) \mid A \text{ antisymmetric}\},$$
$$C^2 \otimes \underline{G} = \{A \in M(n,\underline{G}) \mid A \text{ symmetric traceless}\},$$
$$\Lambda^0 \otimes \underline{G} \equiv \underline{G}, \tag{2.4}$$

will always hold, and Tr will always denote trace in $M(n)$ (never in some enveloping algebra of $\underline{G}$),

$$\mathrm{Tr}\, A = A_\mu{}^\mu = A_{\mu\nu}g^{\mu\nu}. \tag{2.5}$$

To start with, $\otimes$ means over $\mathbb{R}$, but very soon, at Eq. (2.11), we introduce $\otimes$ to mean "with commutators."

Lemma 1 (Projections and inclusions): Corresponding to the essentially unique contractions provided by the metric g,

$$\Lambda^2 \otimes \Lambda^2 \xrightarrow{g} M(n) \xrightarrow{\mathrm{Tr}} \Lambda^0 : X_{\mu\nu\alpha\beta} \xmapsto{g} X_\mu{}^\lambda{}_{\lambda\beta} \equiv X_{\mu\beta} \xmapsto{\mathrm{Tr}} X^\lambda{}_\lambda, \tag{2.6}$$

there exist unique spin-invariant inclusions

$$\Lambda^0 \overset{i'}{\hookrightarrow} M(n) \overset{i}{\hookrightarrow} \Lambda^2 \otimes \Lambda^2 \tag{2.7}$$

such that $gi = 1$, $\mathrm{Tr}\, i' = 1$. Indeed, explicitly,

$$\phi \overset{i'}{\mapsto} \frac{g_{\mu\nu}}{n}\phi, \quad X_{\mu\nu} \overset{i}{\mapsto} (iX)_{\mu\nu\alpha\beta}$$

$$= \left(\frac{g_{\nu\alpha}X_{\mu\beta} - g_{\nu\beta}X_{\mu\alpha} + g_{\mu\beta}X_{\nu\alpha} - g_{\mu\alpha}X_{\nu\beta}}{n-2}\right)$$

$$-\left(\frac{g_{\nu\alpha}g_{\mu\beta} - g_{\mu\alpha}g_{\nu\beta}}{(n-2)(n-1)}\right)X^{\mu}{}_{\mu}\,, \tag{2.8}$$

which implies

$$\phi \overset{ii'}{\mapsto} \frac{(g_{\nu\alpha}g_{\mu\beta} - g_{\mu\alpha}g_{\nu\beta})}{n(n-1)}\phi\,.$$

These are clearly unique given only the metric g [which is invariant under (2.2)].

With these inclusions understood g and Tr are projections $g^2 = g$ and $\mathrm{Tr}^2 = \mathrm{Tr}$, and thus have eigenvalues 0 or 1. Thus

$$M(n) = [1] \oplus \ker \mathrm{Tr} = [1] \oplus \underset{\text{antisym}}{\Lambda^2} \oplus \underset{\text{sym. traceless}}{C^2}\,.$$

Similarly, thinking of $\underline{SO}(n) \cong \Lambda^2$, and of $\Lambda^2 \otimes \Lambda^2$ as tensors of rank 2 associated to $T_e\,SO(n)$ we have

$$\Lambda^2 \otimes \Lambda^2 = [1] \oplus \ker \mathrm{Tr}$$

$$= [1] \oplus \underset{\text{antisym } [\mu\nu]\leftrightarrow[\mu\beta]}{\tilde{\Lambda}^2(\Lambda^2 \otimes \Lambda^2)} \oplus \underset{\text{sym. traceless } [\mu\nu]\leftrightarrow[\mu\beta]}{\tilde{C}^2(\Lambda^2 \otimes \Lambda^2)}\,.$$

Indeed

$$\begin{array}{ccccccccc}
\Lambda^0 & \hookrightarrow & [1] & \hookrightarrow & [1], & \Lambda^0 & \overset{\mathrm{Tr}}{\leftarrow} & [1] & \overset{g}{\leftarrow} & [1], \\
 & & \oplus & & \oplus & & & \oplus & & \oplus \\
 & & \Lambda^2 & \hookrightarrow & \tilde{\Lambda}^2(\Lambda^2 \otimes \Lambda^2), & 0 & \overset{\mathrm{Tr}}{\leftarrow} & \Lambda^2 & \overset{g}{\leftarrow} & \tilde{\Lambda}^2(\Lambda^2 \otimes \Lambda^2), \\
 & & \oplus & & \oplus & & & \oplus & & \oplus \\
 & & C^2 & \hookrightarrow & \tilde{C}^2(\Lambda^2 \otimes \Lambda^2), & 0 & \overset{\mathrm{Tr}}{\leftarrow} & C^2 & \overset{g}{\leftarrow} & \tilde{C}^2(\Lambda^2 \otimes \Lambda^2).
\end{array} \tag{2.9}$$

To see this one has only to check the symmetry properties of (2.6) and (2.8). The space [1] is just Λ^0 with the inclusion understood to give a tensor of appropriate rank as the identity.

Finally, with these inclusions understood, g itself is a projection operator. So

$$\omega \equiv \ker g|_{\tilde{C}^2(\Lambda^2\otimes\Lambda^2)}\,,$$
$$\dim \omega = [(n-3)/8]n(n^2+n+2)\,,$$
$$\overline{\omega} \equiv \ker g|_{\tilde{\Lambda}^2(\Lambda^2\otimes\Lambda^2)}\,,$$
$$\dim \overline{\omega} = [(n-3)/8](n-1)n(n+2)\,,$$
$$W \equiv \ker g = \omega \oplus \overline{\omega}\,,$$
$$\dim W = [(n-3)/4]n^2(n+1)\,,$$

giving ("spin" decomposition)

$$\Lambda^2 \otimes \Lambda^2 = \omega \oplus \overline{\omega} \oplus \Lambda^2 \oplus C^2 \oplus \Lambda^0\,. \tag{2.10}$$

[In components this reads

$$X_{\mu\nu\alpha\beta} = \omega_{\mu\nu\alpha\beta}(X) + \overline{\omega}_{\mu\nu\alpha\beta}(X)$$
$$+\left(\frac{g_{\nu\alpha}X_{\mu\beta} - g_{\nu\beta}X_{\mu\alpha} - g_{\mu\sigma}X_{\nu\beta} + g_{\mu\beta}X_{\nu\alpha}}{n-2}\right)$$
$$-\left(\frac{g_{\nu\alpha}g_{\mu\beta} - g_{\mu\alpha}g_{\nu\beta}}{(n-2)(n-1)}\right)\mathrm{Tr}\, X$$

and

$$X_{\mu\beta} = X_{\mu}{}^{\lambda}{}_{\lambda\beta} = X_{[\mu\beta]} + X_{(\mu\beta)} + (\mathrm{Tr}\, X/n)g_{\mu\beta}\,.$$

For example, $X_{\mu\nu\alpha\beta} = R_{\mu\nu\alpha\beta}$, $\omega =$ Weyl tensor, $X_{\mu\nu}$ = Ricci, $\overline{\omega}$ related to torsion.]

Now consider $M(n,\underline{G})$ and define $\otimes^{\underline{G}}$, which is *not* associative, by (equality for $\underline{G}$ semisimple)

$$M(n,\underline{G}) \otimes^{\underline{G}} M(n,\underline{G})$$
$$\overset{\text{def}}{=} \{A_{\mu\nu} \otimes^{\underline{G}} B_{\alpha\beta}\} \overset{\text{def}}{=} \{[A_{\mu\nu}, B_{\alpha\beta}]\}$$
$$\subseteq (M(n) \otimes_{\mathbb{R}} M(n)) \otimes \underline{G}\,. \tag{2.11}$$

Then, with product (2.3) and $A,B \in \Lambda^2 \otimes \underline{G}$, (2.10) becomes in this case

$$A \otimes^{\underline{G}} B \overset{\text{def}}{=} \omega(A,B) + \overline{\omega}(A,B) + [AB] + (\!(AB)\!) + \mathrm{Tr}\, AB\,. \tag{2.12}$$

Then

$$[AB] = [BA],\ (\!(AB)\!) = -(\!(BA)\!),\ \mathrm{Tr}\, AB = -\mathrm{Tr}\, BA\,,$$
$$\omega(A,B) = -\omega(B,A),\ \ \overline{\omega}(A,B) = \overline{\omega}(B,A)\,. \tag{2.13}$$

Here the projectors ω: $A \otimes^{\underline{G}} B \mapsto \omega(A,B)$, etc., are just the projections onto the respective spaces in (2.10) and $W = \omega + \overline{\omega}$.

Examples (low dimensions, $A,B \in \Lambda^2 \otimes \underline{G}$, $\otimes^{\underline{G}}$ understood):

$\underline{n=2}$:

$$A \otimes B = \mathrm{Tr}\, AB \quad \text{(by inspection)}\,, \tag{2.14}$$

i.e., $\omega \equiv 0$, $\overline{\omega} \equiv 0$, $[AB] \equiv 0$, $(\!(AB)\!) \equiv 0$ in (2.12). Some intermediate inclusion maps are singular, which could lead to infinite-dimensional algebras if limits such as $[AB] \to 0$, $n \to 2$ are taken suitably. (Compare the conformal group in $n = 2$.)

$\underline{n=3}$:

$$A\otimes B=[AB]+((AB))+\operatorname{Tr}AB\quad(\dim W=0)\,. \tag{2.15}$$

$\underline{n=4}$:

$$A_+\otimes B_+=\omega(A_+,B_+)+[A_+B_+]+\operatorname{Tr}A_+B_+\,,$$
$$[A_+B_+]\in\Lambda^2_+\otimes\underline{G} \tag{2.16}$$

[cf. spin decomposition

$$(1,0)\otimes(1,0)=(2,0)\oplus(1,0)\oplus(0,0)$$
$$(*\omega\propto\operatorname{Tr}W^*=\operatorname{Tr}A_+B_+)]\,,$$
$$A_+\otimes B_-=\overline{\omega}(A_+,B_-)+((A_+B_-)) \tag{2.17}$$

[cf. $(1,0)\otimes(0,1)=(1,1)$ $(\overline{\omega}\sim W^*=((A_+B_-))$ overcounts)] .

Proof: Equation (2.17) is essentially Proposition 7.16 of Ref. 2, and Eq. (2.16) is Lemma 4.6 of Ref. 2, so I shall be brief. One may choose a basis τ^a for the Lie algebra with structure constants $f^a{}_{bc}$; then $A=A^a\tau^a$, etc. The ordinary tensor product of matrices $A^a\otimes B^b$ may then be decomposed according to the ordinary representation theory of SO(4), which was indicated below (2.16) and (2.17). (One can only lose representations when one antisymmetrizes $[A,B]^a=A^bB^cf^a{}_{bc}$.) Now

$$((AB))\in(1,1),\quad[AB]\in(1,0)\oplus(0,1)\,,$$
$$W(A,B)\in(2,0)\oplus(0,2)\oplus(1,1)\,.$$

So comparing with (2.12) we deduce $W(A_+,B_+)\in(2,0)$, $W(A_+,B_-)\sim((A_+B_-))\in(1,1)$, and (0,0) and (1,0) terms vanishing or as included in (2.16) and (2.17). Here $\sim$ denotes the identification through

$$W\mapsto W^{*\mu\nu}\equiv\tfrac{1}{2}\epsilon_{\mu\sigma\alpha\beta}W_\nu{}^{\sigma\alpha\beta}\,,$$

which is symmetric as W is contractionless, and for which the contribution from $\overline{\omega}$ in W is traceless in view of the symmetries of the totally antisymmetric tensor ϵ. This much does not use the bracket; it applies for ordinary tensor products also. Next $W\in(2,0)\oplus(0,2)$ is equivalent to saying that $W_{\mu\nu\alpha\beta}=W_{\alpha\beta\mu\nu}$ (as with the Weyl tensor), which, bearing in mind the anticommutativity, is equivalent to $W(A_+,B_+)=-W(B_+,A_+)$, i.e., $\overline{\omega}$ vanishes in (2.16). Similarly $W(A_+,B_-)=W(B_-,A_+)=\overline{\omega}\in(1,1)$, which completes the proof.

From this we see that

$$W(F,F)=2W(F_+,F_-)\sim((F_+F_-))=0\Leftrightarrow F_+\otimes F_-=0\,, \tag{2.18}$$

where we have to check that the $\sim$ is nonsingular in the present context, i.e., $W^*=0\Leftrightarrow((F_+F_-))=0$. One can in fact show by applying the Hodge operator $*$ to (2.12) that

$$((AB))=((A_+B_-))+((A_-B_+))\,,$$
$$W^*(A,B)=\operatorname{Tr}AB+((A_+B_-))-((A_-B_+))\,.$$

For $G=\mathrm{SU}(2)$, SU(3), U(2) one may easily see that the right-hand side is equivalent to $F_+=0$ or $F_-=0$, which is Lemma 7.2 of Ref. 2. A final point in $n=4$, also from (2.16) and (2.17), is that the product (2.13) acts naturally on quaternions $\otimes\underline{G}=(\Lambda^2_+\oplus\Lambda^0)\otimes\underline{G}$.

So far the only difference with ordinary matrix multiplication has been anticommutativity in $\underline{G}$. The real problem comes when we consider higher products.

Theorem 1 ["conformal algebra," with the decomposition (2.10), (2.12) understood and with $\otimes^{\underline{G}}$ understood]: (i) If A, B, C, $[A]\otimes[B]$, $[B]\otimes[C]$, $[C]\otimes[A]\in(\Lambda^2\oplus\Lambda^0)\otimes\underline{G}$, with $n>3$ and

$$\circ: A\otimes B\mapsto[A]\cdot[B]: (\Lambda^2\oplus\Lambda^0)\otimes\underline{G}$$
$$\otimes(\Lambda^2\oplus\Lambda^0)\otimes\underline{G}\to(\Lambda^2\oplus\Lambda^0)\otimes\underline{G}\,, \tag{2.19}$$

then

$$A\circ B-B\circ A=2\operatorname{Tr}A\circ B\,, \tag{2.20}$$

$$A\circ(B\circ C)-(A\circ B)\circ C=[A]\underset{\bar{n}}{\operatorname{Tr}}B\circ C+[C]\underset{\bar{n}}{\operatorname{Tr}}A\circ B$$
$$+[B]\underset{\overline{n-1}}{\operatorname{Tr}}C\circ A-\operatorname{Tr}B\circ(C\circ A), \tag{2.21}$$

and, for $n>4$ only,

$$[A]\otimes[B\circ C]\in(\Lambda^2\oplus\Lambda^\circ)\otimes\underline{G}\,, \tag{2.22}$$

closing the algebra.

(ii) if A, $A\otimes A\in\Lambda^2\otimes\underline{G}$ [i.e., $W(A,A)=0$, $A\in\Lambda^2\otimes\underline{G}$] and $n>4$, then

$$A^m\circ A^{m'}=A^m\cdot A^{m'}=A^{m+m'}$$
$$=A^m\otimes A^{m'}\in\Lambda^2\otimes\underline{G},\quad\forall m,m'\geqslant 1. \tag{2.23}$$

To prove this we shall prove a sequence of lemmas. Our proof will be computational since a number of basic algebraic properties need to be verified.

The Jacobi identity is

$$A\otimes(B\otimes C)=(A\otimes B)\otimes C+\overbrace{A\otimes(B\otimes C)}, \tag{2.24}$$

where AC indicates that the Lie bracket acts between these two first. Otherwise the order will be denoted by brackets or by emphasizing the "$\cdot$".

Lemma 2:

(i) $$\operatorname{Tr}A\cdot BC+\operatorname{Tr}B\cdot CA+\operatorname{Tr}C\cdot AB=0\,, \tag{2.25}$$

and for $A,B,C\in\Lambda^2\otimes\underline{G}$,

$$\operatorname{Tr}A\cdot BC=\operatorname{Tr}A\cdot CB. \tag{2.26}$$

(ii) (associativity) For $A,B,C\in\Lambda^2\otimes\underline{G}$ with

$$B: W(A,C)_{\mu\nu}\equiv[B^{\alpha\beta},W_{\mu\alpha\nu\beta}(A,C)]\,,$$
$$A\cdot BC-AB\cdot C$$
$$=-B:W(A,C)-\frac{B}{(n-2)(n-1)}\operatorname{Tr}AC$$
$$+\frac{1}{n-2}(B\cdot AC-AC\cdot B)-\frac{n}{n-2}\operatorname{Tr}B\cdot AC. \tag{2.27}$$

Proof: (i) With $\underline{A}^a\in M(n,\mathbb{R})$, $\{\tau^a\}=\underline{G}$, directly

$$\sum_{\text{cyclic}}\operatorname{Tr}A\cdot BC=\operatorname{Tr}\underline{A}^a\underline{B}^b\underline{C}^c\sum_{\text{cyclic}}[\tau^a,[\tau^b,\tau^c]]=0\,,$$

since Tr is cyclically invariant and

$$\operatorname{Tr}A\cdot BC=\operatorname{Tr}A\cdot[BC]=\operatorname{Tr}A\cdot CB$$

on $\Lambda^2\otimes\underline{G}$.

(ii) Contract (2.24) to obtain (2.27) and again to ob-

tain (2.26). This is straightforward and is left as a simple exercise in tensor calculus. Use (2.12) and (2.13). If t denotes the twist map as in (2.30) below, the first step is

$$A\otimes BC=(A\otimes B)\cdot C-(1\otimes t)(A\otimes C)\cdot B. \tag{2.28}$$

From this lemma we see that on projection to C^2,

$$((A\cdot BC))-((AB\cdot C))=-((B{:}W(A,C)))+[2/(n-2)]((B\cdot AC)).$$

Thus under the hypotheses of the theorem now with $[B],[C]\in\Lambda^2\otimes G$ as in the lemma,

$B\circ C\equiv[B]\cdot[C]=[[B][C]]+\operatorname{Tr}B\circ C$ (by hypothesis);

hence

$$((A\circ(B\circ C)))+((C\circ(A\circ B)))=[2/(n-2)]((B\circ(C\circ A))).$$

Subtracting cyclic permutations this implies $((A\circ(B\circ C)))=0$ for $n\neq 3$. Note in passing that

$$\tfrac{1}{2}(AB+BA)=\tfrac{1}{2}([A][B]+\operatorname{Tr}A\cdot\operatorname{Tr}B+(\operatorname{Tr}A)\cdot[B]+(\operatorname{Tr}B)\cdot[A]+B\leftrightarrow A)=[A\circ B],$$

and with

$$[B]\cdot[A][C]-[A][C]\cdot[B]=2\operatorname{Tr}B\circ(A\circ C)+(2/n)[B]\operatorname{Tr}A\circ C$$

in the lemma and with $W[A]\otimes[C]=0$, by hypothesis, we obtain (2.19)–(2.21).

To prove (2.22) with $A,B,C\in\Lambda^2\otimes\underline{G}$ (the scale parts are consistently projected out) from (2.28),

$$A\otimes BC=(A\otimes B)\cdot C-(1\otimes t)(A\otimes C)\cdot B,$$

$$A\otimes[BC]=(1\otimes\Lambda^2)(A\otimes B)\cdot C+B\leftrightarrow C.$$

Here $\Lambda^2\equiv[(1-t)/2]$ projects onto Λ^2 and $\cdot$ denotes contraction of adjacent indices with $[\ ,\ ]$. Inserting (2.12),

$$A\otimes[BC]=(1\otimes\Lambda^2)(WA\otimes B)\cdot C+(1\otimes\Lambda^2)(iAB)\cdot C+B\leftrightarrow C. \tag{2.29}$$

In $\Lambda^2\otimes\Lambda^2\hookrightarrow\Lambda^2\otimes\Lambda^2\otimes\Lambda^2\to\Lambda^2\otimes\Lambda^2$ the possible positions of the contraction and inclusion maps do not commute; in this one equation we have to write the inclusion maps explicitly. We find

$$2(1\otimes\Lambda^2)(iX)\cdot Y=i(X\cdot Y)+\frac{i'n}{n-2}\operatorname{Tr}XY-\frac{\operatorname{Tr}X}{n-1}\left(iY+\frac{i'n}{n-2}\operatorname{Tr}Y\right)+RX\otimes Y,$$

where R is defined in the following lemma.

Lemma 3: Define

$$R{:}M(n)\otimes M(n)\to\Lambda^2\otimes\Lambda^2{:}X_{\sigma\tau\mu\nu}\mapsto(X_{\tau\mu\sigma\nu}-X_{\sigma\mu\tau\nu}-X_{\tau\nu\sigma\mu}+X_{\sigma\nu\tau\mu})/(n-2),$$

$$T{:}M(n)\otimes M(n)\to M(n)\otimes M(n){:}X_{\sigma\tau\mu\nu}\mapsto X_{\mu\nu\sigma\tau}, \tag{2.30}$$

$$t{:}\Lambda^2\to\Lambda^2{:}X_{\mu\nu}\mapsto X_{\nu\mu},$$

$$Q{:}M(n)\otimes M(n)\to\Lambda^2\otimes M(n){:}X_{\sigma\tau\mu\nu}\to X_{\tau\mu\sigma\nu}-X_{\sigma\mu\tau\nu},$$

so $R=Q(1+T)/(n-2)$. Then

$$\text{(i) } TR=Rt\otimes t, \tag{2.31}$$

$$R{:}C^2\otimes C^2\to\omega, \tag{2.32}$$

$$R{:}\Lambda^2\otimes\Lambda^2\to\tilde{S}^2(\Lambda^2\otimes\Lambda^2),\quad R{:}\Lambda^2\to 0, \tag{2.33}$$

$$R{:}\,1\otimes X\mapsto\frac{1}{n}X+\frac{\operatorname{Tr}X}{n-2},\quad R{:}\,1\mapsto\frac{-2}{n-2}1 \tag{2.34}$$

(inclusions understood);

$$\text{(ii) } Q^2=\begin{cases}2+Q, & \text{on } \Lambda^2\otimes M(n),\\ -Q, & \text{on } S^2\otimes M(n)\end{cases} \tag{2.35}$$

(where $S^2\equiv C^2+\operatorname{Tr}=$ symmetric matrices),

$$R^2=\left(\frac{Q}{n-2}\right)^2(1+T)(1+t\otimes t)=\begin{cases}\left(\frac{2}{n-2}\right)^2(1+T)+\frac{2}{n-2}R, & \text{on } \Lambda^2\otimes\Lambda^2,\\ 0, & \text{on } \Lambda^2\otimes S^2 \text{ or } S^2\otimes\Lambda^2,\\ -\frac{2}{n-2}R, & \text{on } S^2\otimes S^2.\end{cases} \tag{2.36}$$

Proof: (i) By inspection and

$$(gRX)_\sigma{}^\lambda{}_{\lambda\nu}=(X^\lambda{}_{\lambda\sigma\nu}-X_\sigma{}^\lambda{}_{\lambda\nu}-X^\lambda{}_{\nu\sigma\lambda}+X_{\sigma\nu}{}^\lambda{}_\lambda)/(n-2)$$

and comparison with (2.6)–(2.8) $\Rightarrow i=nR1\otimes-i'\operatorname{Tr}/(n-2)$, $1\propto R1\otimes 1$ and use (ii) for R^2.

(ii) $Q^2{:}X_{\mu\nu\alpha\beta}\cdots\mapsto X_{\nu\alpha\mu\beta}-X_{\mu\alpha\nu\beta}\cdots\mapsto X_{\alpha\mu\nu\beta}-X_{\nu\mu\alpha\beta}-X_{\alpha\nu\mu\beta}+X_{\mu\nu\alpha\beta}$.

Then

$$R^2=\left(\frac{Q}{n-2}\right)(1+T)R=\frac{Q}{n-2}R(1+t\otimes t)=\frac{Q^2}{(n-2)^2}(1+T)(1+t\otimes t).$$

Lemma 4 ($A,B,C\in\Lambda^2\otimes\underline{G}$):

$$\text{(i) } WA\otimes[BC]=W(1\otimes\Lambda^2)(WA\otimes C)\cdot B-W(R/2)B\otimes((AC))-(R/2)WB\otimes[AC]+B\leftrightarrow C, \tag{2.37}$$

where, from the definition, $W=1-g$. So

$$W(1\otimes\Lambda^2)(WA\otimes C)\cdot B=(1\otimes\Lambda^2)(WA\otimes C)\cdot B+\tfrac{1}{2}B{:}\,WA\otimes C,$$

$$WRB\otimes((AC))=RB\otimes((AC))+2[B\cdot((AC))]/(n-2)\in\bar{\omega}. \tag{2.38}$$

$$\text{(ii) } \{\bar{\omega}A\otimes B=0\ \bar{\omega}B\otimes C=0\ \bar{\omega}C\otimes A=0\}\Rightarrow WA\otimes[BC]+WB\otimes[CA]+WC\otimes[AB]=0. \tag{2.39}$$

Proof: Applying the previous lemma to (2.29) we have

$$A\otimes[BC]=2(1\otimes\Lambda^2)(WA\otimes B)\cdot C+AB\cdot C+[n/(n-2)]\operatorname{Tr}AB\cdot C-[1/(n-1)](\operatorname{Tr}AB)C+RAB\otimes C,$$

with averaging of $B \leftrightarrow C$ left understood in this proof. We have from Lemma 2

(a) $\operatorname{Tr} B\cdot AC = \frac{1}{2}(\operatorname{Tr} B\cdot AC + \operatorname{Tr} C\cdot AB)$

$$= -\tfrac{1}{2}\operatorname{Tr} A\cdot BC = \operatorname{Tr} C\cdot AB = -\operatorname{Tr} AB\cdot C$$

and

$$B\cdot AC - AC\cdot B = 2[B((AC))] + 2B\operatorname{Tr} AC + 2((B[AC])).$$

Thus Lemma 2(ii) becomes

(b) $$AB\cdot C = A\cdot[BC] - \tfrac{1}{2}\operatorname{Tr} A\cdot BC - \frac{B}{n(n-1)} \times \operatorname{Tr} AC - \frac{2}{n-2}[B((AC))] - \frac{2}{n-2}((B[AC])) + B{:}\, WA\otimes C.$$

(Note here that $\operatorname{Tr} AC \in \Lambda^0 \otimes \underline{G}$. We could also have written $-[1/(n-1)]B\cdot\operatorname{Tr} AC$, meaning $\operatorname{Tr} AC \in S^2 \otimes \underline{G}$ by inclusion.)

Finally, using the previous lemma,

(c) $$\begin{aligned} RAB\otimes C &= -RC\otimes AB = -RWC\otimes[AB] - RC\cdot[AB] - RC\otimes((AB)) - RC\otimes\operatorname{Tr} AB \\ &= -RWC\otimes[AB] - R((C\cdot[AB])) - R\operatorname{Tr} C\cdot[AB] - RC\otimes((AB)) - C(1/n)\operatorname{Tr} AB \\ &= -RWC\otimes[AB] + [2/(n-2)]((C\cdot[AB])) - [1/(n-2)]\operatorname{Tr} A\cdot BC - RC\otimes((AB)) - C(1/n)\operatorname{Tr} AB. \end{aligned}$$

Then combining these results and comparing with the spin decomposition (2.10) we see

(d) $$\begin{aligned} WA\otimes[BC] &= 2(1\otimes\Lambda^2)(WA\otimes B)\cdot C - RWC\otimes[AB] \\ &\quad + B{:}\, WA\otimes C - RB\otimes((AC)) \\ &\quad - [2/(n-2)][B((AC))], \end{aligned}$$

i.e., we obtain the lemma. We see as a check that $WA\otimes[BC]$ is contractionless. The second part (ii) follows easily from (2.13) and (i) by adding cyclic permutations.

We are now able to complete the proof of Theorem 1(i). Under the conditions of the theorem the last lemma (i) becomes, with $A,B,C\in\Lambda^2\otimes\underline{G}$,

$$WA\otimes[BC] = -(R/2)WB\otimes[AC] - (R/2)WC\otimes[AB] \Rightarrow \bar{\omega}A\otimes[BC] = 0, \tag{2.40}$$

using Lemma 3(i). Writing $\alpha = [2/(n-2)]$ we have, from Lemma 3(ii),

$$R\omega A\otimes[BC] = -\alpha^2\omega B\otimes[AC] + \alpha\omega A\otimes[BC] - \alpha^2\omega C\otimes[AB].$$

Inserting cyclic permutations into (2.39) we have

$$2\omega A\otimes[BC] = (\alpha^2-\alpha)\omega C\otimes[BA] + (\alpha^2-\alpha)\omega B\otimes[CA] + 2\alpha^2\omega A\otimes[BC].$$

So for $n\neq 4$ we have

$$\omega A\otimes[BC] + (1/n)\omega C\otimes[BA] + (1/n)\omega B\otimes[CA] = 0 \Rightarrow (1-1/n)\omega A\otimes[BC] = 0,$$

which completes the proof of (2.22) and Theorem 1(i), where $[A],[B],[C]\in\Lambda^2\otimes\underline{G}$ as used in all our lemmas. In $n=4$ we can only conclude that $\bar{\omega}A\otimes[BC]=0$, which is true for the class of self-dual fields (2.16). One may be able to prove it only assuming as much for $A\otimes B$, $B\otimes C$, $C\otimes A$.

To prove Theorem 1(ii), since all the results of Theorem 1(i) hold we have $W(A\otimes A^2)=0$. Now $A^2\in\Lambda^2\otimes\underline{G}$ automatically by (2.13) and $A\cdot A^2\in(\Lambda^2\oplus\Lambda^0)\otimes\underline{G}$, etc. We have to show that the traces $\operatorname{Tr} A\cdot A^2$, etc., all vanish. Using Lemma 2(i),

$$\operatorname{Tr} A^{p-m}\cdot A^m \overset{\text{def}}{\equiv} \operatorname{Tr} A^{p-m}\cdot(A\cdot A^{m-1}) = -\operatorname{Tr} A\cdot(A^{m-1}\cdot A^{p-m}) - \operatorname{Tr} A^{m-1}\cdot(A^{p-m}\cdot A).$$

We *assume as an induction hypothesis* that the result holds for $p-1$, i.e.,

$$A^{m'}\cdot A^{m''} = A^{m'+m''}\in\Lambda^2\otimes G, \quad \forall 2\leqslant m'+m''\leqslant p-1 \quad (\text{true for } p-1=2). \tag{2.41}$$

Thus *for* $1<m<p$, using Lemma 2(i),

$$a_m \equiv \operatorname{Tr} A^{p-m}\cdot A^m = \operatorname{Tr} A^{p-1}\cdot F + \operatorname{Tr} A^{p-m+1}\cdot A^{m-1} = a_1 + a_{m-1} \therefore = ma_1 = -a_{p-m}$$

by (2.41) assumption; therefore $ma_1 = -(p-m)a_1 \Rightarrow a_1 = 0$.

Thus from Theorem 1(i)

$$A^{p-m}\cdot A^m = A^{p-m+1}\cdot A^{m-1}\cdots = A^p\in\Lambda^2\otimes\underline{G}.$$

This concludes the proof of Theorem 1.

We now investigate the case $W(F,F)=0$. From Theorem 1(ii) we see that now $F^mF^{m'} = F^{m+m'}\in\Lambda^2\otimes\underline{G}$ (anti-Hermitian in $\underline{G}$) and

$$[F^m_{\mu\nu}, F^{m'}_{\alpha\beta}] = [1/(n-2)](g_{\nu\alpha}F_{\mu\beta}{}^{m+m'} - g_{\nu\beta}F^{m+m'}_{\mu\alpha} - g_{\mu\alpha}F^{m+m'}_{\nu\alpha} + g_{\mu\beta}F^{m+m'}_{\nu\alpha}). \tag{2.42}$$

Theorem 2: For $\underline{G}$ finite dimensional, $n>4$ and generically,

$$W(F,F)=0 \Rightarrow \underline{H}_0 \equiv \text{primitive holonomy} = \bigoplus_{i=i}^{q} \underline{\mathrm{SO}}(n)_C, \tag{2.43}$$

for some integer q. Except for singular points this integer q should be constant over the manifold. (This will not be prov-

en here; cf. Ref. 9.) The decomposition is over $\mathbb{C}$.

Since the Lie algebra $\underline{H}_0 \overset{\text{def}}{=} \{F_{\mu\nu}\} \subseteq \underline{G}$ is finite dimensional, there exists q such that (with μ_k real for F anti-Hermitian in $\underline{G}$)

$$F^{q+1} = \mu_q F^q + \cdots + \mu_1 F.$$

$$E^{(i)}(F) = \sum_r e_r^{(i)} F^r,$$

$$E^{(i)}(F) \cdot E^{(j)}(F) = \delta^{ij}\left(q - t\frac{\partial}{\partial t}\right) P(t)\,|_{t=\lambda^{(j)}} E^{(j)}(F),$$

Proof:

$$\underline{\Lambda} = \begin{pmatrix} 0 & & & & \mu_1 \\ 1 & 0 & & & \\ & 1 & \ddots & & \vdots \\ & & \ddots & 0 & \\ & & & 1 & \mu_q \end{pmatrix}$$

represents F acting as the ring of polynomials in F, $F \cdot a_k F^k = (\underline{\Lambda} a)_k F^k$, and has eigenvalues $\lambda^{(i)}$ and eigenvectors $e^{(i)}$ given by solutions of

$$\begin{pmatrix} 0 & & & & \mu_1 \\ 1 & 0 & & & \\ & 1 & \ddots & & \vdots \\ & & \ddots & 0 & \\ & & & 1 & \mu_q \end{pmatrix} \begin{pmatrix} e_1 \\ \vdots \\ e_q \end{pmatrix} = \lambda \begin{pmatrix} e_1 \\ \vdots \\ e_q \end{pmatrix},$$

$$e_1 = \mu_1\left(\frac{1}{\lambda}\right) e_q \cdots$$

$$\Rightarrow e_q = \frac{1}{\lambda}\left(\mu_q + \frac{1}{\lambda}\left(\mu_{q-1} + \cdots + \frac{1}{\lambda}\mu_1\right)\cdots\right) e_q.$$

Thus

$$\lambda^q = \sum_{r=1}^{q} \mu_r \lambda^{r-1},$$

$$e_r = \lambda^{-r} \sum_{k=1}^{r} \mu_k \lambda^{k-1} = \lambda^{q-r} - \sum_{\iota \equiv k-r=1}^{q-r} \mu_{r+\iota} \lambda^{\iota-1}.$$

Then $E^{(i)}(F)$ acts on $E^{(j)}(F)$ as

$$e_r^{(i)} \lambda^{(j)r} = \sum_{r=1}^{q} \lambda^{(j)r} \lambda^{(i)-r} \left(\sum_{k=1}^{r} \mu_k \lambda^{(i)k-1}\right)$$

$$= \sum_{k=1}^{q} \sum_{r=k}^{q} \lambda^{(j)r} \lambda^{(i)-r} \mu_k \lambda^{(i)k-1}$$

$$= \sum_{k=1}^{q} \mu_k \lambda^{(i)k-1} \alpha^k \sum_{r=k}^{q} \alpha^{r-k}, \quad \alpha \equiv \frac{\lambda^{(j)}}{\lambda^{(i)}},$$

for

$$\lambda^{(j)} \neq \lambda^{(i)} : = \sum_{k=1}^{q} \mu_k \lambda^{(i)k-1} \alpha^k \left(\frac{1-\alpha^{q-k+1}}{1-\alpha}\right)$$

$$= \frac{\lambda^{(j)} - \lambda^{(j)q}}{\lambda^{(i)} - \lambda^{(j)}} - \frac{\alpha^{q+1}}{1-\alpha} \sum_{k=1}^{q} \mu_k \lambda^{(i)k-1} = 0,$$

Define

$$P(t) = \sum_{k=1}^{q} \mu_k t^{k-1}.$$

Explicitly in terms of the data (μ_k) and independent of the signature of g,

$$\begin{cases} e_r^{(i)} = (\lambda^{(i)})^{-r} \sum_{k=1}^{r} \mu_k \lambda^{(j)k-1} & \text{(nondegenerate case)}, \\ \text{where } \lambda^q = p(\lambda) \text{ has solutions } \lambda^{(i)} \text{ over } \mathbb{C}. \text{ Complex} \\ \text{solutions give conjugate pairs } E^{(i)}, \overline{E}^{(i)} \text{ for } \mu_k \text{ real.} \end{cases} \tag{2.44}$$

$$\lambda^{(j)} = \lambda^{(i)} : = \sum_{k=1}^{q} \mu_k \lambda^{(j)k-1}(q-k+1)$$

$$= \left(q - t\frac{\partial}{\partial r}\right) P(t)\,|_{t=\lambda^{(j)}}.$$

It remains only to normalize the generators $E^{(i)}$ to obtain generators of $\underline{SO}(n)$. (The degenerate cases $\lambda^{(i)} = \lambda^{(j)}$, $j \neq i$, $E^{(i)} = E^{(j)}$ leave open a variety of exceptional spaces, for example, $\lambda^{(i)} \equiv \mu_k \equiv 0$, i.e., $F^{q+1} = 0$.)

This concludes our purely algebraic considerations.

ACKNOWLEDGMENT

This work was supported in part by the Herchel Smith Scholarship of Emmanuel College, Cambridge, England.

APPENDIX: EXISTENCE AND SOME EXAMPLES

(1) The prototypical example of a $W(F,F) = 0$ connection is, according to Theorem 2, the one-quasi-instanton with curvature

$$F = [4\lambda^2/(\lambda^2 + x^2)^2]\,\Sigma; \tag{A1}$$

$$\Sigma_{\mu\nu} \equiv \tfrac{1}{4}[\gamma_\mu, \gamma_\nu], \quad \{\gamma_\mu, \gamma_\nu\} = 2\delta_{\mu\nu},$$

which, for $n = 4$, is an instanton and an anti-instanton in each $SO(3) \equiv SU(2)/Z_2$ factor. (For higher dimensions they still play a vital role and have been termed quasi-instantons.)[4] Here Σ in the Clifford algebra generates rotations in $SO(n)$, and the $SO(n-1)$ factors are picked out by the projection

$$\Sigma_\pm = [(1 \pm \iota^{n/2}\Gamma)/2]\Sigma, \tag{A2}$$

$$\Gamma = \gamma^0 \gamma^1 \cdots \gamma^{n-1}, \quad \Gamma^2 = (-1)^{n/2}.$$

From this it is very easy to prove that in $n = 4k$ dimensions these obey the so-called generalized self-duality[10] (chirality),

$$\Lambda^{n/2} \otimes \mathrm{Cliff}(\mathbb{R}^n) \ni *\Sigma \wedge \Sigma \wedge \cdots \wedge \Sigma$$

$$\equiv *\Sigma^{n/2} = \Gamma \Sigma^{n/2} = \pm \iota^{n/2} \Sigma^{n/2}. \tag{A3}$$

The above construction is well known for $n = 4$.[11] One may easily see that for $n \geqslant 4$ the quasi-instanton instead extremizes the action

$$\int_M \|F\|^{n/2}, \tag{A4}$$

and that in fact there is an $O(n+1)$ symmetry up to gauge transformations.[11] Typically, as checked for $n=8$ in Ref. 10, they are also related to topological lower bounds.

(2) Let $L^{(m)}_{\mu\nu}$, $m\in\mathbb{Z}$, denote the operators generating a Kac–Moody algebra $\widehat{SO}(n)$ defined by the structure relations $L^{(m)}_{\mu\nu} = -L^{(m)}_{\nu\mu}$ and relations of the form

$$\begin{aligned}[L^{(m_1)}_{\mu\nu}, L^{(m_2)}_{\alpha\beta}] = [1/(n-2)](&g_{\nu\alpha}L^{(m_1+m_2)}_{\mu\beta} - g_{\mu\alpha}L^{(m_1+m_2)}_{\nu\beta} - g_{\nu\beta}L^{(m_1+m_2)}_{\mu\alpha} \\ &+ g_{\mu\beta}L^{(m_1+m_2)}_{\nu\alpha}) \\ &+ c_n(m_1,m_2)(g_{\nu\alpha}g_{\mu\beta} - g_{\mu\alpha}g_{\nu\beta}),\end{aligned}$$

where c is the central extension. From this we see that the $L^{(m)}$ fulfill all the requirements of Theorem 1(i). In accordance with the theorem the $L^{(m)}$ under the product $\circ$ form an almost commutative algebra. In the present case it is clear that the noncommutativity is due to the presence of the central extension c, and that the algebra is in fact associative because the Tr terms are proportional to the identity. This is central so it contributes nothing in (2.21).

Again, our "Liouville theorem" asserts this is the prototypical example in the infinite-dimensional case. Namely, suppose instead we are given just three operator-valued matrices $L^{(0)}$, $L^{(1)}$, and $L^{(-1)}$, obeying a relation of the general form above—perhaps a more complicated operator in the extension— and $n>4$. Then define $L^{(\pm m)} = (L^{(\pm 1)})^m$. By Theorem 1(ii) these are all antisymmetric operator-valued matrices and obey the same relation as the Kac–Moody algebra above, except that the theorem does not predict anything about the extension term, only that such trace terms are allowed by the algebra.

(3) Finally, we give an intuitive estimate of how hard it is to satisfy $W(F,F)=0$. For $C^0 \equiv \{F_{\mu\nu}\} \subset \underline{G}$ let

$$\dim[C^0,C^0] = \lambda \dim C^0$$

as vector spaces in $\underline{G}$. Then the dimension of the solution space per $\dim C^0$ is essentially [using (2.10) with $W(F,F) = \bar{\omega}(F,F)$]

$$\begin{aligned}&\frac{\dim F - \dim W(F,F)}{\dim C^0} \\ &= \frac{n(n-1)}{2} - \frac{(n-3)(n-1)n(n+2)}{8}\lambda \\ &= \frac{n(n-1)}{2}\left(1 - \lambda\frac{(n-3)(n+2)}{4}\right).\end{aligned}$$

So the critical λ above which there are no solutions is, for example,

$$n=4,\quad \lambda = \tfrac{2}{3};$$
$$n=5,\quad \lambda = \tfrac{2}{4};$$
$$n=6,\quad \lambda = \tfrac{1}{6}.$$

So long as $\lambda > 0$ there should be a class of sparse matrices for which $W(F,F)=0$ holds identically. For example, in $n=4$ we have seen that the class of all $F^a_{\mu\nu}$, with self-dual and anti-self-dual parts decoupled, leaves behind in $[C^0,C^0]$ just $\frac{2}{3}$ of the possible cases

$$[F_{+\mu\nu},F_{+\alpha\beta}],[F_{-\mu\nu},F_{-\alpha\beta}],[F_{+\mu\nu},F_{-\alpha\beta}] = 0,$$

and is just the class for which $W(F,F)=0$ in $n=4$. This estimate shows quantitatively how $W(F,F)$ generalizes self-duality to higher dimensions where it involves $\underline{G}$ in a more detailed way.

[1] See, for example, J. A. Dieudonne, *Algèbre Linéaire et Géométrie Élémentaire* (Hermann, Paris, 1964), Appendix III.6; S. Kobayashi and K. Nomizu, *Foundations of Differential Geometry* (Wiley–Interscience, New York, 1963), Vol. 1.

[2] J.-P. Bourguignon and H. B. Lawson, "Stability and isolation phenomena for Yang–Mills fields," Commun. Math. Phys. **79**, 189 (1981).

[3] Z. I. Szbaó, "Structure theorems on Riemannian symmetric spaces satisfying $R[x,y]\cdot R = 0$. I. The local version," J. Diff. Geom. **17**, 531 (1982).

[4] G. Shore, "Dimensional regularization and instantons," Ann. Phys. (NY) **122**, 321 (1978).

[5] D. Freed and K. Uhlenbeck, *Instantons and Four Manifolds* (Springer, New York, 1984).

[6] R. Ward, "Completely solvable gauge field equations in dimension greater than four," Nucl. Phys. B **236**, 381 (1984); E. Corrigan, C. Devchaud, D. B. Fiarlie, and J. Nuyts, "First order equations for gauge fields in spaces of dimension greater than four," Nucl. Phys. B **214**, 452 (1983); E. Corrigan, P. Goddard, and A. Kent, "Some comments on the ADHM construction in 4k dimensions," Commun. Math. Phys. **100**, 1 (1985).

[7] See, for example, G. Shore, "Symmetry restoration and the background field method," Ann. Phys. (NY) **137**, 262 (1981).

[8] L. Crane, "Action of the loop group on the self-dual Yang–Mills equations," Commun. Math. Phys. **204**, 391 (1987); V. Kac, *Infinite Dimensional Lie Algebras* (Birkhauser, Boston, 1981).

[9] A. Lichnerowicz, *Théorie Globàle des connexions et des groupes d'holonomie* (English translation: Noordhoff, Leiden, 1976).

[10] B. Grossman, T. Kephart, and J. Stasheff, "Solutions to Yang–Mills equations in eight dimensions and the last Hopf map," Commun. Math. Phys. **96**, 431 (1985). But see important Erratum, Commun. Math. Phys. **100**, 311 (1985).

[11] R. Jackiw and C. Rebbi, "Conformal properties of a Yang–Mills pseudoparticle," Phys. Rev. D **14**, 517 (1976); R. Jackiw, C. Nohl, and C. Rebbi, "Conformal properties of pseudoparticle configurations," Phys. Rev. D **15**, 1642 (1977).

Minimal models of U(1) conformal current algebra

P. Furlan
International School for Advanced Studies (SISSA/ISAS), Trieste, Italy and Istituto di Fisica Teorica dell'Università di Trieste, Trieste, Italy and Istituto Nazionale di Fisica Nucleare (INFN), Sezione di Trieste, Trieste, Italy

G. M. Sotkov[a)]
International School for Advanced Studies (SISSA/ISAS), Trieste, Italy and Istituto Nazionale di Fisica Nucleare (INFN), Sezione di Trieste, Trieste, Italy

Ya. Stanev
Department of Theoretical Physics, University of Sofia, Sofia, Bulgaria

(Received 21 December 1987; accepted for publication 25 May 1988)

A detailed study of the degenerate representations of the $U(1)\times U(1)$ conformal current algebra and field-theoretical models based on it is presented. The main peculiarity of these models is that they contain an infinite set of fields with unfixed anomalous dimensions. Simple fusion rules for these fields are found, and their four-point functions are calculated explicitly. The case of Thirring-like two-dimensional models is discussed in detail.

I. INTRODUCTION

The recent applications of the methods of the short-distance operator product expansions (OPE's) to different two-dimensional, field-theoretical, critical statistical and string models[1-12] are based on the specific properties of the reducible representations of the Virasoro algebra.[2] The crucial property is that for each value of the central charge $0<c\leqslant 1$ and for special values of the conformal dimensions there exist "null vectors" in the representation space.[1,2] Then as a consequence of the conformal invariance and the null vector condition, only a finite number of conformal families contribute to the OPE of two degenerate representations (giving rise to the so-called fusion rules[1]). It turns out that because of the fusion rules, the associative OPE algebra of the fields corresponding to these degenerate representations is closed under the usual OPE multiplication. In this way, for each value of $0<c\leqslant 1$ we have a finite-dimensional associative algebra of the conformal fields, which defines a so-called minimal model. An important property of these models is that due to the existence of null vectors, the four-point functions of the fields satisfy specific differential equations and can be found explicitly[7] (the same is true in principle also for the n-point functions).

In the cases where $c\geqslant 1$ (for $c=1, \Delta\neq n^2/4, n=0,1,\ldots$), because of the absence of null vectors the infinite conformal symmetry is not enough to find the exact solutions of the corresponding models. This problem has a simple solution for models (with $c\geqslant 1$) possessing a symmetry larger than the conformal one, such as the minimal models with $N=1,2,3$ extended superconformal symmetry[4,10-13] and the conformal current-algebra models.[3,6,9] In these cases the null vectors of the larger algebra of symmetry (the conformal subalgebra has no null vectors for $c>1$) lead to corresponding new fusion rules, which in turn provide equations for the four-point functions.[4,11,12] For each fixed value of $c\geqslant 1$, the corresponding minimal model (with larger symmetry) contains only a finite set of fields whose dimensions are rational numbers given by Kac-type formulas.[14]

Since many of the well-known two-dimensional models with $c\geqslant 1$ and based on the conformal current algebras such as the Thirring model ($c=1$) or the string models ($c=D$, $D=$ positive integer) contain fields with arbitrary (non-quantized) anomalous dimensions, a natural question arises as to whether these models have all the properties of the minimal models, i.e., null vectors, fusion rules, and equations for four-point functions but without Kac[2] quantization of the dimensions. The positive answer to this question is based on the properties of the degenerate representations of the $U(1)^{\otimes D}\otimes U(1)^{\otimes D}$ chiral conformal current algebra.[15] In this case the only new property is that we do not have a finite closed OPE algebra of the fields, but nevertheless, because of the existence of null vectors in the full algebra, we have simple fusion rules for this infinite set of primary fields[1] (of the full algebra), and the calculation of their four-point functions is straightforward. The main part of these results can be seen immediately from the "bosonization rules" for the representations of the $U(1)^{\otimes D}$ conformal current algebra.[15]

In this paper we present a detailed study of minimal models based on the representations of the conformal $U(1)\times U(1)$ current algebra in two-dimensional Minkowski space-time. This algebra can be written as a direct sum of two "one-dimensional" $U(1)$ current algebras (CA's) (right: L_n, J_n; and left: $\bar{L}_n, \bar{J}_n$):

$$[L_n,L_m]=(n-m)L_{n+m}+(c/12)n(n^2-1)\delta_{n+m,0}, \tag{1.1a}$$

$$[L_n,J_m]=-mJ_{n+m}, \tag{1.1b}$$

$$[J_n,J_m]=n\delta_{n+m,0}, \quad m,n\in\mathbb{Z}, \tag{1.1c}$$

and the same for $\bar{L}_n, \bar{J}_n$. We shall study the field-theoretical models in the compactified Minkowski space-time $\bar{M}$,

[a)] On leave of absence from the Institute for Nuclear Research and Nuclear Energy, Bulgarian Academy of Sciences, Sofia 1184, Bulgaria.

 0022-2488/88/102311-09$02.50

$$\bar{M}=S^1\times S^1=\left\{(z,\bar{z}):\ \overset{(-)}{z}=\left(\overset{(-)}{\xi}+i\right)\Big/\left(1+i\overset{(-)}{\xi}\right),\right.$$
$$\left.\xi=x^1-x^0,\ \bar{\xi}=-x^1-x^0,\ \overset{(-)}{|z|}=1\right\},$$

that are invariant under local reparametrizations $z\to w=w(z)$ $[\bar{z}\to\bar{w}=\bar{w}(\bar{z})]$ and local $U(1)\times U(1)$ gauge transformations. Due to the splitting of the variables z and $\bar{z}$ we can consider only the right counterpart of these models, leaving two-dimensional constructions to the last section of this paper.

The reparametrizations are generated by the symmetric, conserved, and traceless stress-energy tensor $\Theta_{\mu\nu}$, and hence with two independent components $\Theta(\xi)$ and $\bar{\Theta}(\bar{\xi})$ only.

In the compact picture we have

$$T(z)=-[8\pi/(1-iz)^4]\Theta[(z-i)/1-iz)],$$
$$\bar{\partial}T(z)=0=\partial\bar{T}(\bar{z}),$$

and $T(z)$ transforms inhomogeneously under reparametrization:

$$T(z)\to\tilde{T}(z)=(w'(z))^2T(w(z))+\frac{c}{12}\left(\frac{w''}{w'}-\frac{3}{2}\left(\frac{w''}{w'}\right)^2\right). \tag{1.2}$$

The conserved current $j^\mu(x)$ and its dual $j_5^\mu(x)=\varepsilon^{\mu\nu}j_\nu(x)$ generate the $U_L(1)\times U_R(1)$ local gauge transformations, and in the light-cone variables we have

$$\overset{(-)(-)}{j}(\xi)=j_0(x)+\overset{(-)}{j_1}(x),\quad \bar{\partial}j(\xi)=0,$$

$$J(z)=-[4\pi i/(1-iz)^2]j[(z-i)/(1-iz)],\quad \bar{\partial}J(z)=0,$$

on the unit circle. The current $J(z)$ has anomalous gauge transformations,

$$J(z)\to\tilde{J}(z)=J(z)+u'(z), \tag{1.3}$$

and transforms homogeneously under reparametrizations:

$$J(z)\to\tilde{J}(z)=w'(z)J(w(z)). \tag{1.4}$$

According to (1.2)–(1.4), for infinitesimal transformations $w(z)=z+\varepsilon(z)$ and $u(z)=1+v(z)$ we obtain

$$\begin{aligned}&\delta_\varepsilon T(z)=(\varepsilon(z)\partial+2\varepsilon'(z))T(z)+c/12\varepsilon'''(z),\\&\delta_\varepsilon J(z)=(\varepsilon(z)\partial+\varepsilon'(z))J(z),\\&\delta_v J(z)=v'(z).\end{aligned} \tag{1.5}$$

Equations (1.5) can be rewritten in the form of operator product expansions[3]:

$$T(z_1)T(z_2)=\frac{c}{2z_{12}^4}+\frac{2}{z_{12}^2}T(z_2)+\frac{1}{z_{12}}T'(z_2)+O(1), \tag{1.6a}$$

$$T(z_1)J(z_2)=(1/z_{12}^2)J(z_2)+(1/z_{12})J'(z_2)+O(1), \tag{1.6b}$$

$$J(z_1)J(z_2)=1/z_{12}^2+2T(z_2)+O(z_{12}), \tag{1.6c}$$

where the normalization of the c-number term in Eq. (1.6c) is due to our choice

$$\langle J(z_1)J(z_2)\rangle=1/z_{12}^2.$$

Using the standard definition of the normal product,

$$:J^2(z):=\lim_{z_1\to z_2=z}\{J(z_1)J(z_2)-\langle J(z_1)J(z_2)\rangle\},$$

in Eq. (1.6c) and taking the limit, we get the well-known Sugawara formula

$$T(z)=\tfrac{1}{2}:J^2(z):. \tag{1.7}$$

Equations (1.6), written in terms of the Laurent coefficients of the fields $T(z)$ and $J(z)$,

$$\begin{aligned}&T(z)=\sum_{n\in\mathbb{Z}}\frac{L_n}{z^{n+2}},\quad L_n=\oint_c T(z)z^{n+1}\,dz,\\&J(z)=\sum_{n\in\mathbb{Z}}\frac{J_n}{z^{n+1}},\quad J_n=\oint_c J(z)z^n\,dz,\end{aligned} \tag{1.8}$$

take the form (1.1) of the $U(1)$ conformal current algebra. The Sugawara formula for the Laurent coefficients L_n, J_n,

$$L_n=\frac{1}{2}\left(\sum_{k>-n}+\sum_{k>1}\right)J_{-k}J_{n+k}, \tag{1.9}$$

demonstrates that the Virasoro algebra belongs to the enveloping algebra of the $U(1)$ current algebra.

The main building blocks in the construction of the minimal models invariant under transformations of the algebra (1.1) are the so-called *primary fields*.[1] They transform homogeneously under reparametrizations and local $U(1)$ gauge transformations,

$$\begin{aligned}&\delta_\varepsilon\phi_{\Delta,q}(z)=(\varepsilon(z)\partial+\Delta\varepsilon'(z))\phi_{\Delta,q}(z),\\&\delta_v\phi_{\Delta,q}(z)=-qv(z)\phi_{\Delta,q}(z),\end{aligned} \tag{1.10}$$

or, equivalently, in terms of L_n and J_n,

$$[L_n,\phi_{\Delta,q}(z)]=(z^{n+1}\partial+(n+1)\Delta z^n)\phi_{\Delta,q}(z), \tag{1.11a}$$

$$[J_n,\phi_{\Delta,q}(z)]=-qz^n\phi_{\Delta,q}(z),\quad n\in\mathbb{Z}, \tag{1.11b}$$

where each primary field $\phi_{\Delta,q}$ is completely determined by its dimension Δ and by its $U(1)$ charge q for fixed c.

The Hermitian conjugate field $\phi^*_{\Delta,q}(z)$ has a charge opposite to $\phi_{\Delta,q}(z)$; so we have

$$[J_n,\phi^*_{\Delta,q}(z)]=qz^n\phi^*_{\Delta,q}(z),\quad n\in\mathbb{Z}. \tag{1.11c}$$

We shall use also the following more compact form of these commutation relations:

$$\begin{aligned}&[T^{(-)}(z),\phi(z')]=\left(\frac{1}{z-z'}\frac{\partial}{\partial z'}+\frac{\Delta}{(z-z')^2}\right)\phi(z'),\\&[J^{(-)},\phi(z')]=-[q/(z-z')]\phi(z'),\end{aligned} \tag{1.12}$$

where

$$T^{(-)}(z)=\sum_{n=0}^{\infty}\frac{L_n}{z^{n+2}}$$

and

$$J^{(-)}=\sum_{n=0}^{\infty}\frac{J_n}{z^{n+1}}.$$

The properties of both the primary states[1] [corresponding to the primary fields (1.10)] and the null vectors presented in the next section are the basic tool in our examination of the $U(1)$ conformal current-algebra fusion rules. These rules determine the structure of the conformal field OPE associative algebra (infinite dimensional in our case), which defines the corresponding minimal model. Since the Sugawara formula (1.7) implies that the central charge is uniquely fixed to be $c=1$, it seems that we have only one (1-D) minimal model corresponding to this value of c. But the analogy with the minimal models of the Virasoro algebra, where for each fixed c we have only one model, does not work in our case because of the absence of quantization of

the dimensions. In fact, the only restriction we have is the equality (2.4b)

$$\Delta = q^2/2.$$

Then following the fusion rules (2.13) and considering, for example, a conformal field with integer charge, we see that it creates an infinite set of fields with integer and half-integer dimensions only. Starting with the half-integer value of the charge we obtain another infinite set of fields with dimensions and charges different from the first ones. So we shall have many different models corresponding to $c = 1$. Indeed, if we want to have a model with arbitrary (unfixed) anomalous dimension of the field, then we have only one infinite set of fields whose dimensions and charges are parametrized by the charge q of the initial (primary) field $\phi_{\Delta,q}$. The conclusion is that the symmetries of the model are not enough to fix some critical values of the charge and of the anomalous dimension.

All the four-point functions of the fields of the model are calculated in Sec. III. The fourth section of the paper is devoted to the construction of two-dimensional models by different compositions of left and right one-dimensional models. We analyze in detail the case of the massless Thirring model,[16,17] where $c = 1$ and $|\Delta - \bar{\Delta}| = \frac{1}{2}$. In the last section we present our investigation of the reduction of the U(1) conformal CA representations to the representation of the Virasoro subalgebra.

We will conclude this general discussion of the properties of $U(1)^{\otimes D}$ conformal CA minimal models with the case $c = D > 1$ (say $D = 26$). As is shown in our recent paper,[15] the structure of the representations of the $U(1)^{\otimes D}$ conformal current algebra is the same as in the case $c = 1$. Then the corresponding minimal models contain an infinite set of conformal fields with arbitrary anomalous dimensions parametrized by D arbitrary "charges" (or momenta) of the initial field. The case $D = 26$ provided us with fusion rules for the string vertex operators and simple differential equations for the string vertices four-point functions.

The generalization of these results to the case of supersymmetric minimal models of $U(1)^{\otimes D} \otimes \mathrm{CAR}^{\otimes D}$ superconformal current algebra (CAR is the canonical anticommutation relation algebra $\{\bar{\psi}^i_n, \psi^j_n\} = \delta_{n+m}\delta^{ij}$) seems to be straightforward.

II. NULL VECTORS AND FUSION RULES

The representations of the U(1) conformal current algebra (1.1) can be characterized by three numbers: dimension Δ, charge q, and central charge c. Since the energy positivity implies positivity of the L_0, we shall be interested only in the lowest-weight representations of the algebra (1.1). Each representation of this kind is generated by the *primary state* $|\Delta,q\rangle$ defined as follows (c is fixed):

$$L_0|\Delta,q\rangle = \Delta|\Delta,q\rangle, \tag{2.1a}$$

$$J_0|\Delta,q\rangle = q|\Delta,q\rangle, \tag{2.1b}$$

$$J_n|\Delta,q\rangle = 0 = L_n|\Delta,q\rangle, \quad n > 0. \tag{2.1c}$$

Then the representation space (called the conformal family of $|\Delta,q\rangle$) is spanned by the vectors

$$L_{-n_1}\ldots L_{-n_k}J_{-m_1}\ldots J_{-m_l}|\Delta,q\rangle,$$

$$m_1 \geqslant m_2 \geqslant \ldots \geqslant m_l > 0, \quad n_1 \geqslant n_2 \geqslant \ldots \geqslant n_k > 0,$$

with charge q and dimension $\Delta_{k,l} = \Delta + N_{k,l}$, where

$$N_{k,l} = \sum_{i=1}^{k} n_i + \sum_{i=1}^{l} m_i$$

is called "level" $N_{k,l}$. The correspondence between primary fields (1.11) and primary states (2.1) is given by the following almost obvious equality:

$$|\Delta,q\rangle = \phi_{\Delta,q}(0)|0\rangle. \tag{2.2}$$

The vacuum state $|0\rangle$ is a $\mathrm{SL}(2,R) \otimes \mathrm{U}(1) = \{L_{\pm 1}, L_0, J_0\}$ invariant primary state with $\Delta = 0 = q$. Therefore we have

$$L_n|0\rangle = 0, \quad n \geqslant -1, \quad J_n|0\rangle = 0, \quad n \geqslant 0.$$

The reducible (or "degenerate") representations of (1.1) are defined by the condition that there exists a state $|\Delta + N,q\rangle$ at level N that is again a primary state. Using the definitions (2.1), (1.1), and the Hermiticity conditions $L^*_n = L_{-n}$, $J^*_n = J_{-n}$, it is easy to show that $|\Delta + N,q\rangle$ is orthogonal to each vector in the representation space spanned by $|\Delta,q\rangle$, that it has zero norm, and therefore that it can be consistently put equal to zero:

$$|\Delta + N,q\rangle = 0. \tag{2.3}$$

An important problem we have to solve is the explicit construction of the null vectors (2.3). For the first level ($N = 1$), Eqs. (2.1a), (2.1b), and (1.1a)–(1.1c) imply

$$|\zeta^{(1)}\rangle = (L_{-1} + a_1 J_{-1})|\Delta,q\rangle.$$

Imposing that $|\zeta^{(1)}\rangle$ satisfies Eq. (2.1c) we get the relations

$$a_1 = -q, \tag{2.4a}$$

$$\Delta = q^2/2, \tag{2.4b}$$

and finally

$$|\zeta^{(1)}\rangle = (L_{-1} - J_0 J_{-1})|\Delta,q\rangle = 0 \tag{2.5}$$

iff $\Delta = q^2/2$. The same considerations at the second level give us

$$|\zeta^{(2)}\rangle = (L_{-2} - \tfrac{1}{2}J^2_{-1} - J_0 J_{-2})|\Delta,q\rangle = 0 \tag{2.6}$$

iff $c = 1$ and $\Delta = q^2/2$.

In principle we can continue this construction to higher levels, and at each level we find a new null vector without any new restriction on the parameters of the model: c, Δ, q. This means that the primary state $|\Delta,q\rangle$ with $c = 1$, $\Delta = q^2/2$ is degenerate at each level $N \in \mathbb{Z}_+$ (contrary to the case of the Virasoro degenerate primary states). In fact all these properties of degeneracy are coded within the Sugawara formula (1.9). Using the definition (2.1) and Eq. (1.9) we get the following general construction of the null vectors:

$$|\zeta^{(n)}\rangle = \left(L_{-n} - \frac{1}{2}\sum_{k=0}^{n} J_{-k}J_{k-n}\right)|\Delta,q\rangle, \quad n \geqslant 0 \tag{2.7}$$

(iff $\Delta = q^2/2$ and $c = 1$). The null vector condition (2.5) can be written as an equation for the primary field $\phi_{\Delta,q}(z)$ corresponding to the primary state $|\Delta,q\rangle$ via Eq. (2.2). Taking into account Eqs. (1.11) we get for $\phi_{\Delta,q}$ the Thirring-like equation

$$\frac{\partial}{\partial z}\phi_{\Delta,q}(z) = -q :J(z)\phi_{\Delta,q}(z):, \tag{2.8}$$

where the normal product : : is given by

$$:J(z)\phi(z): = J^{(+)}(z)\phi(z) + \phi(z)J^{(-)}(z), \qquad (2.9)$$

and $J^{(+)}(z) = J(z) - J^{(-)}(z)$.

The key points in our construction of the fusion rules for degenerate primary fields of the algebra (1.1) are the null vector constructions (2.5) and (2.7) together with Eq. (2.4b) and the U(1) charge conservation law. Let us consider two degenerate primary fields $\chi_1(z)$, $\chi_2(z)$, with dimensions Δ_1, Δ_2 and nonzero charges q_1, q_2 related through Eq. (2.4b):

$$\Delta_{1,2} = \tfrac{1}{2} q_{1,2}^2.$$

In principle many conformal U(1) current-algebra families can contribute to the OPE of these fields. We denote by $\theta^{(i)}(z)$, $i = 1,2,...$, the primary fields of each one of these families, with dimension $\Delta^{(i)}$ and charge $q^{(i)}$. Because of the Sugawara formula (1.9), for each primary field of the algebra (1.1) we have

$$L_0|\Delta^{(i)},q^{(i)}\rangle = \tfrac{1}{2}J_0^2|\Delta^{(i)},q^{(i)}\rangle, \qquad (2.10a)$$

or

$$\Delta^{(i)} = \tfrac{1}{2}(q^{(i)})^2, \qquad (2.10b)$$

which means that each primary field is degenerate at zero level [i.e., Eq. (2.10) presents the 0-level null vector.] Then taking only the leading term contributions in the well-known OPE formulas,[18]

$$\chi_1(z_1)\chi_2(z_2) = \langle\chi_1(z_1)\chi_2(z_2)\rangle + \sum_{i\geq 1}\frac{N_{120}^{(i)}}{N_0^{(i)}} z_{12}^{\Delta^{(i)}-\Delta_1-\Delta_2} \times[\theta^{(i)}(z_2) + O(z_{12})], \qquad (2.11)$$

we can analyze the charge conservation law, commuting both sides of Eq. (2.11) and J_0 and using Eq. (1.11b), since the $\theta^{(i)}(z)$ fields are primary fields for the full algebra. We obtain the identity

$$\sum_{i\geq 1}\frac{N_{120}^{(i)}}{N_0^{(i)}} z_{12}^{\Delta^{(i)}-\Delta_1-\Delta_2}[-(q_1+q_2)\theta^{(i)}(z_2) + O(z_{12})] = \sum_{i\geq 1}\frac{N_{120}^{(i)}}{N_0^{(i)}} z_{12}^{\Delta^{(i)}-\Delta_1-\Delta_2}[-q^{(i)}\theta^{(i)}(z_2) + O(z_{12})],$$

which proves the conservation of the U(1) charge, i.e.,

$$q^{(i)} = q_1 + q_2, \quad \text{for each } i \geq 1. \qquad (2.12)$$

Therefore only one primary field $\phi_{\Delta,q}$ with charge $q = q_1 + q_2$ and dimension $\Delta = \frac{1}{2}(q_1+q_2)^2$ contributes to the OPE (2.11) of the fields χ_1 and χ_2, and the fusion rule can be written symbolically

$$\chi_{\Delta_1}^{q_1}\chi_{\Delta_2}^{q_2} = [\chi_\Delta^{q_1+q_2}].$$

We denote with $[\chi_\Delta]$ the conformal family of the primary field χ_Δ.

Let us suppose we have a model with only two basic primary fields $\chi_{1,2}$ [together with their Hermitian conjugates $\chi_{1,2}^*$ with opposite U(1) charges]. Then we can write the fusion rules as follows:

$$\begin{aligned}
&\chi_1\chi_1^* = \chi_2\chi_2^* = [I],\\
&\chi_1\chi_1 = [\xi], \quad q_\xi = 2q_1, \quad \Delta_\xi = 2q_1^2,\\
&\chi_2\chi_2 = [\eta], \quad q_\eta = 2q_2, \quad \Delta_\eta = 2q_2^2,\\
&\chi_1\chi_2 = [\theta], \quad q_\theta = q_1+q_2, \quad \Delta_\theta = \tfrac{1}{2}(q_1+q_2)^2,\\
&\chi_1\chi_2^* = [\varphi], \quad q_\varphi = q_1-q_2, \quad \Delta_\varphi = \tfrac{1}{2}(q_1-q_2)^2.
\end{aligned} \qquad (2.13a)$$

Repeating the same procedure for the composite primary fields $\xi,\eta,...$ we have to introduce the new primary fields,

$$\begin{aligned}
&\xi\chi_1 = [\Xi_1], \quad q_{\Xi_1} = 3q_1, \quad \Delta_{\Xi_1} = \tfrac{9}{2}q_1^2,\\
&\xi\chi_2 = [\Xi_2], \quad q_{\Xi_2} = 2q_1+q_2, \quad \Delta_{\Xi_2} = \tfrac{1}{2}(2q_1+q_2)^2,\\
&\xi\chi_1^* = [\tilde{\Xi}_1], \quad q_{\tilde{\Xi}_1} = q_1, \quad \Delta_{\tilde{\Xi}_1} = \tfrac{1}{2}q_1^2,\\
&\xi\chi_2^* = [\tilde{\Xi}_2], \quad q_{\tilde{\Xi}_2} = 2q_1-q_2, \quad \Delta_{\tilde{\Xi}_2} = \tfrac{1}{2}(2q_1-q_2)^2,
\end{aligned} \qquad (2.13b)$$

and the corresponding ones for η,θ,φ (indicated by E, Θ, and Φ). We can continue this procedure, but already at this level we can completely verify the characterizing features of our model.

(i) At each step we must always introduce new composite primary fields; therefore we have an infinite set of them in the full associative algebra.

(ii) Nevertheless we perfectly control their U(1) and conformal properties.

(iii) Many of the families we find, going from level to level, are identical to other families; for example,

$$\begin{aligned}
&[\tilde{\Xi}_1] = [\tilde{\Theta}_2] = [\Phi_2] = [\chi_1], \quad [\tilde{E}_2] = [\tilde{\Theta}_1] = [\chi_2],\\
&[\Theta_1] = [\Xi_2], \quad [\Theta_2] = [E_1], \quad [\Phi_1] = [\tilde{\Xi}_2],\\
&[\tilde{\Phi}_1] = [\chi_2^*], \quad [\tilde{\Phi}_2] = [\tilde{E}_2^*].
\end{aligned} \qquad (2.14)$$

III. FOUR-POINT FUNCTIONS

In this section we shall show how conformal invariance, U(1) charge conservation, and the existence of the null vector $|\zeta^{(1)}\rangle$ of Eq. (2.5) [together with the relations (2.4)] allow us to calculate the four-point function of any four primary fields $\chi_i(z)$, with U(1) charges q_i and dimensions $\Delta_i = \frac{1}{2}q_i^2$, where $i = 1,2,3,4$.

The projective Ward identities, obtained by imposing sl(2,R) invariance of the N-point functions,[1] give us the general form of the four-point function,

$$\langle 0|\chi_1(z_1)\chi_2(z_2)\chi_3(z_3)\chi_4(z_4)|0\rangle = z_{12}^{\Delta_3+\Delta_4-\Delta_1-\Delta_2} z_{13}^{-2\Delta_3} z_{14}^{\Delta_2+\Delta_3-\Delta_1-\Delta_4} \times z_{24}^{\Delta_1-\Delta_2-\Delta_3-\Delta_4} f(\eta), \qquad (3.1a)$$

where

$$\eta := z_{13}z_{24}/z_{12}z_{34}, \qquad (3.1b)$$

and $f(\eta)$ is a function to be determined. The relation

$$\langle 0|\chi_1(z_1)\chi_2(z_2)\chi_3(z_3)\chi_4(z_4)J_0|0\rangle = 0, \qquad (3.2a)$$

with the help of Eq. (1.11b), leads to the U(1) charge conservation, i.e.,

$$\sum_{i=1}^{4} q_i = 0. \qquad (3.2b)$$

Translational invariance and the null vector condition (2.5) give us the relation

$$\langle 0|\chi_1(z_{14})\chi_2(z_{24})\chi_3(z_{34})(L_{-1}+q_4J_{-1})|\chi_4\rangle = 0, \quad (3.3a)$$

which, using Eqs. (1.11), provides us with the following first-order differential equation for the four-point function (3.1a):

$$\left(\frac{\partial}{\partial z_{14}}+\frac{\partial}{\partial z_{24}}+\frac{\partial}{\partial z_{34}}-\frac{q_1q_4}{z_{14}}-\frac{q_2q_4}{z_{24}}-\frac{q_3q_4}{z_{34}}\right)$$
$$\times\langle 0|\chi_1(z_1)\chi_2(z_2)\chi_3(z_3)\chi_4(z_4)|0\rangle = 0. \quad (3.3b)$$

Inserting the expression (3.1a) into Eq. (3.3b) and making use of Eqs. (2.4b) and (3.2b), we transform Eq. (3.3b) into a first-order differential equation for $f(\eta)$,

$$(q_2q_3/(\eta-1)-q_3q_4)f(\eta) = \eta f'(\eta), \quad (3.4)$$

which we easily solve, getting the final result for the four-point function:

$$\langle 0|\chi_1(z_1)\chi_2(z_2)\chi_3(z_3)\chi_4(z_4)|0\rangle = C(q)\prod_{i<j=1}^{4} z_{ij}^{q_iq_j}, \quad (3.5)$$

where $C(q)$ is a constant possibly dependent on q_i, $i=1,2,3,4$, and the q_i's are related to the conformal degrees Δ_i by Eq. (2.4b).

IV. TWO-DIMENSIONAL MODELS: THE THIRRING MODEL

Let us consider a particular set of two-dimensional models (i.e., models in which fields are functions of both z and $\bar{z}$) that can be "factorized" as products of two one-dimensional models of the same kind we dealt with in the previous section. To this end let us introduce two one-dimensional primary fields: $\chi(z)$ with U(1) charge q and dimension $\Delta=\frac{1}{2}q^2$, and $\bar{\chi}(\bar{z})$ with $\bar{q}$ and $\bar{\Delta}=\frac{1}{2}\bar{q}^2$. We define the following two-dimensional field (q and q are supposed to be nonzero):

$$\psi(z,\bar{z}) := \chi(z)\bar{\chi}(\bar{z}). \quad (4.1a)$$

It is immediately seen, using Eqs. (1.11) and the analogous ones for the $\bar{z}$-dependent fields, that $\psi(z,\bar{z})$ is a primary field under both the left and right current and Virasoro algebras with dimensions

$$(\Delta,\bar{\Delta}) = (\tfrac{1}{2}q^2,\tfrac{1}{2}\bar{q}^2) \quad (4.1b)$$

and U(1) charges

$$(q,\bar{q}). \quad (4.1c)$$

Therefore its dimension is given by

$$d := \Delta+\bar{\Delta} = \tfrac{1}{2}(q^2+\bar{q}^2), \quad (4.2a)$$

while its helicity is given by

$$h := \Delta-\bar{\Delta} = \tfrac{1}{2}(q^2-\bar{q}^2). \quad (4.2b)$$

It is well known[1,3,6] that, in order to have a local theory, fields must have integer or half-integer helicity (i.e., spin); therefore we shall put this constraint on Eq. (4.2b) when we shall deal with a concrete case in what follows.

The existence of the one-dimensional fusion rules of Sec. II and the complete factorization of the z and $\bar{z}$ dependence allow us to write down immediately the fusion rules for the two-dimensional models we consider here. Let us suppose we have two primary fields,

$$\psi_1(z,\bar{z}) = \chi_1(z)\bar{\chi}_1(\bar{z}), \quad (4.3a)$$

with U(1) charges $(q_1,\bar{q}_1)$, and

$$\psi_2(z,\bar{z}) = \chi_2(z)\bar{\chi}_2(\bar{z}), \quad (4.3b)$$

with U(1) charges $(q_2,\bar{q}_2)$; their dimensions are given by Eq. (4.1b). Then taking into account the results of Sec. II, we have

$$\begin{aligned}
&\psi_i(z,\bar{z})\psi_i(z,\bar{z}) = [\Psi^{(i,i)}(z,\bar{z})], \quad i=1,2,\\
&\psi_1(z,\bar{z})\psi_2(z,\bar{z}) = [\Psi^{(1,2)}(z,\bar{z})],\\
&\psi_1(z,\bar{z})\psi_2^*(z,\bar{z}) = [\tilde{\Psi}^{(1,2)}(z,\bar{z})],
\end{aligned} \quad (4.4)$$

where the composite primary fields Ψ are given by [see Eqs. (2.13a)]

$$\begin{aligned}
&\Psi^{(1,1)}(z,\bar{z}) = \xi(z)\bar{\xi}(\bar{z}),\\
&\Psi^{(2,2)}(z,\bar{z}) = \eta(z)\bar{\eta}(\bar{z}),\\
&\Psi^{(1,2)}(z,\bar{z}) = \theta(z)\bar{\theta}(\bar{z}),\\
&\tilde{\Psi}^{(1,2)}(z,\bar{z}) = \varphi(z)\bar{\varphi}(\bar{z}).
\end{aligned} \quad (4.5)$$

Their U(1) charges are given by

$$\begin{aligned}
&q_{\Psi^{(i,i)}} = 2q_i, \quad \bar{q}_{\Psi^{(i,i)}} = 2\bar{q}_i, \quad i=1,2,\\
&q_{\Psi^{(1,2)}} = q_1+q_2, \quad \bar{q}_{\Psi^{(1,2)}} = \bar{q}_1+\bar{q}_2,\\
&q_{\tilde{\Psi}^{(1,2)}} = q_1-q_2, \quad \bar{q}_{\tilde{\Psi}^{(1,2)}} = \bar{q}_1-\bar{q}_2,
\end{aligned} \quad (4.6)$$

while their dimensions are easily recovered from Eq. (4.6) with the help of relation (4.1b).

Let us write, as an example, the OPE of $\psi_1\psi_2$; we have[16]

$$\begin{aligned}
&\psi_1(z_1,\bar{z}_1)\psi_2(z_2,\bar{z}_2)\\
&= \langle\psi_1(z_1,\bar{z}_1)\psi_2(z_2,\bar{z}_2)\rangle + z_{12}^{\Delta-\Delta_1-\Delta_2}\bar{z}_{12}^{\bar{\Delta}-\bar{\Delta}_1-\bar{\Delta}_2}\\
&\times\sum_{n=0}^{\infty} z_{12}^n \sum_{m=0}^{\infty}\bar{z}_{12}^m \frac{N_{12n,m}}{N_{n,m}} B^{-1}(\delta_1+n,\delta_2+n)\\
&\times B^{-1}(\bar{\delta}_1+m,\bar{\delta}_2+m)\\
&\cdot\int_0^1 du(1-u)^{\delta_1+n-1}u^{\delta_2+n-1}\\
&\cdot\int_0^1 dv(1-v)^{\bar{\delta}_1+m-1}v^{\bar{\delta}_2+m-1}\\
&\times\Psi^{(1,2)}_{\Delta+n,\bar{\Delta}+m}(z_2+uz_{12},\bar{z}_2+v\bar{z}_{12}),
\end{aligned} \quad (4.7a)$$

where $N_{12n,m}$ is the coefficient appearing in the three-point function

$$\langle\psi_1\psi_2\Psi^{(1,2)*}_{\Delta+n,\bar{\Delta}+m}\rangle,$$

while $N_{n,m}$ is that appearing in the two-point function

$$\langle\Psi^{(1,2)}_{\Delta+n,\bar{\Delta}+m}\Psi^{(1,2)*}_{\Delta+n,\bar{\Delta}+m}\rangle,$$

and the two-dimensional quasiprimary fields $\Psi^{(1,2)}_{\Delta+n,\bar{\Delta}+m}$ are given by (for definitions and more details see Ref. 16)

$$\begin{aligned}
&n!\,m!\frac{N_{12n,m}}{N_{n,m}}\Psi^{(1,2)}_{\Delta+n,\bar{\Delta}+m}(z,\bar{z})\\
&= \lim_{z_{1,2}\to z}\lim_{\bar{z}_{1,2}\to\bar{z}} D_n^{(\delta_1,\delta_2)}(\partial_1,\partial_2)D_m^{(\bar{\delta}_1,\bar{\delta}_2)}(\bar{\partial}_1,\bar{\partial}_2)\\
&\cdot\{z_{12}^{\Delta_1+\Delta_2-\Delta}\bar{z}_{12}^{\bar{\Delta}_1+\bar{\Delta}_2-\bar{\Delta}}\psi_1(z_1,\bar{z}_1)\psi_2(z_2,\bar{z}_2)\}.
\end{aligned} \quad (4.7b)$$

Taking into account factorization, we easily find that

$$\frac{N_{12n,m}}{N_{n,m}}\Psi^{(1,2)}_{\Delta+n,\bar{\Delta}+m}(z,\bar{z}) = \frac{N_{12n}\bar{N}_{12m}}{N_n\bar{N}_m}\theta_{\Delta+n}(z)\theta_{\bar{\Delta}+m}(\bar{z}). \quad (4.8)$$

If we now adopt the quasiprimary (one-dimensional) field normalization[16]

$$N_{12n} = \Delta N_n, \quad n \geqslant 0, \tag{4.9a}$$

supplemented by the analogous one for two-dimensional fields, i.e.,

$$N_{12n,m} = \Delta\bar{\Delta} N_{n,m}, \tag{4.9b}$$

we get the complete factorization of the quasiprimary fields,

$$\Psi_{\Delta+n,\bar{\Delta}+m}(z,\bar{z}) = \theta_{\Delta+n}(z)\bar{\theta}_{\Delta+m}(\bar{z}), \tag{4.10a}$$

together with the factorization of the coefficients,

$$N_{12n,m} = N_{12n}\bar{N}_{12m} \tag{4.10b}$$

and

$$N_{n,m} = N_n \bar{N}_m. \tag{4.10c}$$

Using again factorization and Eq. (3.5), we can easily get the explicit form of the four-point function of ψ fields:

$$\begin{aligned}&\langle\psi_1(z_1,\bar{z}_1)\psi_2(z_2,\bar{z}_2)\psi_3(z_3,\bar{z}_3)\psi_4(z_4,\bar{z}_4)\rangle \\ &\quad = \langle\chi_1(z_1)\chi_2(z_2)\chi_3(z_3)\chi_4(z_4)\rangle \\ &\qquad \times\langle\bar{\chi}_1(\bar{z}_1)\bar{\chi}_2(\bar{z}_2)\bar{\chi}_3(\bar{z}_3)\bar{\chi}_4(\bar{z}_4)\rangle \\ &\quad = C(q)\bar{C}(\bar{q})\prod_{i<j=1}^{4} z_{ij}^{q_i q_j}\prod_{k<l=1}^{4}\bar{z}_{kl}^{\bar{q}_k\bar{q}_l}.\end{aligned} \tag{4.11}$$

Let us now come to a concrete example: the Thirring model.[16,17] In our framework it is a two-dimensional model with two primary fields having equal dimensions and opposite fixed helicities $\pm\frac{1}{2}$; let us denote them by

$$\psi_1(z,\bar{z}), \quad \text{with helicity } h_1 = -\tfrac{1}{2},$$
$$\psi_2(z,\bar{z}), \quad \text{with helicity } h_2 = +\tfrac{1}{2}.$$

From Eqs. (4.2) we see that the U(1) charges must obey the following relations:

$$q_1^2 + \bar{q}_1^2 = q_2^2 + \bar{q}_2^2, \quad q_1^2 - \bar{q}_1^2 = -1, \quad q_2^2 - \bar{q}_1^2 = 1. \tag{4.12}$$

Therefore we can express all four charges in terms of only one (positive) parameter λ:

$$q_1 = \bar{q}_2 = \sqrt{2\lambda}, \quad q_2 = \bar{q}_1 = \sqrt{2\lambda+1}. \tag{4.13}$$

The conformal degrees are then

$$\Delta_1 = \bar{\Delta}_2 = \lambda, \quad \Delta_2 = \bar{\Delta}_1 = \lambda + \tfrac{1}{2}. \tag{4.14}$$

The fields $\psi_{1,2}$ can be identified with the fields appearing in Eqs. (4.3) once we fix the charges with Eq. (4.13). All fusion rules (4.4) are valid again with

$$\begin{aligned}&q_{\Psi^{(1,1)}} = 2\sqrt{2\lambda}, \quad \bar{q}_{\Psi^{(1,1)}} = 2\sqrt{2\lambda+1}, \\ &q_{\Psi^{(2,2)}} = 2\sqrt{2\lambda+1}, \quad \bar{q}_{\Psi^{(2,2)}} = 2\sqrt{2\lambda}, \\ &q_{\Psi^{(1,2)}} = \sqrt{2\lambda} + \sqrt{2\lambda+1}, \quad \bar{q}_{\Psi^{(1,2)}} = \sqrt{2\lambda} + \sqrt{2\lambda+1}, \\ &q_{\tilde{\Psi}^{(1,2)}} = \sqrt{2\lambda} - \sqrt{2\lambda+1}, \quad \bar{q}_{\tilde{\Psi}^{(1,2)}} = \sqrt{2\lambda+1} - \sqrt{2\lambda}.\end{aligned} \tag{4.15}$$

We see that

$$h_{\Psi^{(1,1)}} = -h_{\Psi^{(2,2)}} = -2, \quad h_{\Psi^{(1,2)}} = h_{\tilde{\Psi}^{(1,2)}} = 0, \tag{4.16}$$

i.e., $\Psi^{(1,1)}$ together with $\Psi^{(2,2)}$ corresponds to spin 2, while $\Psi^{(1,2)}$ and $\tilde{\Psi}^{(1,2)}$ are scalar fields. From Eqs. (4.6) and (4.16) it is evident that while U(1) charge is conserved, there exists no "helicity conservation" in the fusion rules (4.4).

As already remarked in the general case, and as appears from the first level's analysis just outlined above, we have in the Thirring model an *infinite* set of *primary fields*, but we can have complete control of all fusion rules, going level by level with the same kind of analysis we have exploited above.

Also, all four-point functions of primary fields are easily made explicit using Eq. (4.11). For example,

$$\begin{aligned}&\langle\psi_1(z_1,\bar{z}_1)\psi_2^*(z_2,\bar{z}_2)\psi_2(z_3,\bar{z}_3)\psi_1^*(z_4,\bar{z}_4)\rangle \\ &\quad = c(\lambda)(z_{13}z_{24}/z_{12}z_{34})^{\sqrt{2\lambda(2\lambda+1)}} z_{14}^{-2\lambda} z_{23}^{-2\lambda-1} \\ &\qquad \times(\bar{z}_{13}\bar{z}_{24}/\bar{z}_{12}\bar{z}_{34})^{\sqrt{2\lambda(2\lambda+1)}}\bar{z}_{14}^{-2\lambda-1}\bar{z}_{23}^{-2\lambda},\end{aligned} \tag{4.17a}$$

$$\begin{aligned}&\langle\tilde{\Psi}^{(1,2)}(z_1,\bar{z}_1)\tilde{\Psi}^{(1,2)^*}(z_2,\bar{z}_2)\tilde{\Psi}^{(1,2)}(z_3,\bar{z}_3)\tilde{\Psi}^{(1,2)^*}(z_4,\bar{z}_4)\rangle \\ &\quad = b(\lambda)\left(\frac{z_{13}z_{24}}{z_{12}z_{14}z_{23}z_{34}}\times\frac{\bar{z}_{13}\bar{z}_{24}}{\bar{z}_{12}\bar{z}_{14}\bar{z}_{23}\bar{z}_{34}}\right)^{4\lambda+1-2\sqrt{2\lambda(2\lambda+1)}}.\end{aligned} \tag{4.17b}$$

Even if the factorization of the Thirring model greatly facilitates its understanding, as we have just seen, we can get the same results using the two-dimensional formulation only.

Let us start from a massless spin-$\frac{1}{2}$ field coupled to a canonical U(1) vector current, which is conserved together with its dual.[3,6,16,17] Let the fields ψ_1 and ψ_2 be the left and right components of the spin-$\frac{1}{2}$ field and q_1, q_2, $\bar{q}_1$, $\bar{q}_2$ their U(1) charges; i.e., we postulate that the following commutation relations hold (in the compact picture):

$$[J_n, \psi_{1,2}(z,\bar{z})] = -q_{1,2} z^n \psi_{1,2}(z,\bar{z}), \tag{4.18a}$$

$$[\bar{J}_n, \psi_{1,2}(z,\bar{z})] = -\bar{q}_{1,2}\bar{z}^n\psi_{1,2}(z,\bar{z}), \quad n\in\mathbb{Z}. \tag{4.18b}$$

Furthermore we suppose that the fields $\psi_{1,2}$ transform homogeneously under conformal transformations, i.e., that they are primary conformal fields with dimensions $(\Delta_1,\bar{\Delta}_1)$ and $(\Delta_2,\bar{\Delta}_2)$, satisfying the commutation relations

$$\begin{aligned}&[L_n,\psi_{1,2}(z,\bar{z})] \\ &\quad = z^n\left(z\frac{\partial}{\partial z} + (n+1)\Delta_{1,2}\right)\psi_{1,2}(z,\bar{z}),\end{aligned} \tag{4.19a}$$

$$\begin{aligned}&[\bar{L}_n,\psi_{1,2}(z,\bar{z})] \\ &\quad = \bar{z}^n\left(\bar{z}\frac{\partial}{\partial\bar{z}} + (n+1)\bar{\Delta}_{1,2}\right)\psi_{1,2}(z,\bar{z}), \quad n\in\mathbb{Z}.\end{aligned} \tag{4.19b}$$

The presence of zeroth-level and first-level null fields or, equivalently, the existence of the Sugawara formula (1.7) [or (1.9)] and the analogous one for $\bar{T}(\bar{z})$ (or $\bar{L}_n$) again constrains U(1) charges and conformal dimensions to be linked by Eq. (4.1b), while $\psi_{1,2}(z,\bar{z})$ must obey the Thirring equations

$$\begin{aligned}&\frac{\partial\psi_{1,2}(z,\bar{z})}{\partial z} + q_{1,2}:J(z)\psi_{1,2}(z,\bar{z}): \\ &\quad = 0 = \frac{\partial\psi_{1,2}(z,\bar{z})}{\partial\bar{z}} + \bar{q}_{1,2}:\bar{J}(\bar{z})\psi_{1,2}(z,\bar{z}):.\end{aligned} \tag{4.20}$$

Since ψ_1 and ψ_2 have the same dimensions and their helicities are $(-\frac{1}{2})$ and $(+\frac{1}{2})$, respectively, we can state Eqs. (4.13) and (4.14) again. We can show here also that there exists only one conformal family of the full algebra contributing to the OPE of any two among the four chiral fields ψ_1, ψ_2, ψ_1^*, ψ_2^*. Let us take the case $\psi_1\psi_2^*$ as an example,

and let us denote by $\varphi^{(i)}(z,\bar{z})$, $i\geqslant 1$, all primary fields that contribute in principle (together with their descendants) to their OPE; i.e., in analogy with Eq. (2.11),

$$\psi_1(z_1\bar{z}_1)\psi_2^*(z_2,\bar{z}_2)=\sum_{i\geqslant 1}\frac{\tilde{N}^{(i)}_{120,0}}{\tilde{N}^{(i)}_{0,0}}\,z_{12}^{\Delta^{(i)}-1/2-2\lambda}\,\bar{z}_{12}^{\bar{\Delta}^{(i)}-1/2-2\lambda}\times[\varphi^{(i)}(z_2,\bar{z}_2)+O(z_{12})+O(\bar{z}_{12})]. \tag{4.21}$$

Using (the global version of) Eqs. (4.18) and taking into account that all fields appearing in Eq. (4.21), being primary, transform homogeneously under current algebra, we get from Eq. (4.21)

$$[J^{(-)}(z),\varphi^{(i)}(z',\bar{z}')]=[(\sqrt{2\lambda+1}-\sqrt{2\lambda})/(z-z')]\varphi^{(i)}(z',\bar{z}'), \tag{4.22a}$$

$$[J^{(-)}(\bar{z}),\varphi^{(i)}(z',\bar{z}')]=[(\sqrt{2\lambda}-\sqrt{2\lambda+1})/(\bar{z}-\bar{z}')]\varphi^{(i)}(z',\bar{z}'). \tag{4.22b}$$

Then using the Thirring equations (4.20) in the three-point function $\langle\psi_1\psi_2^*\varphi^{(i)*}\rangle$ together with commutation relations (4.22), we get the following first-order differential equations:

$$\left(\frac{\partial}{\partial z_1}+\frac{\sqrt{2\lambda(2\lambda+1)}}{z_{12}}+\frac{2\lambda-\sqrt{2\lambda(2\lambda+1)}}{z_{13}}\right)\times\langle\psi_1(z_1,\bar{z}_1)\psi_2^*(z_2,\bar{z}_2)\varphi^{(i)*}(z_3,\bar{z}_3)\rangle=0, \tag{4.23a}$$

$$\left(\frac{\partial}{\partial \bar{z}_1}+\frac{\sqrt{2\lambda(2\lambda+1)}}{\bar{z}_{12}}+\frac{2\lambda+1-\sqrt{2\lambda(2\lambda+1)}}{\bar{z}_{13}}\right)\times\langle\psi_1(z_1,\bar{z}_1)\psi_2^*(z_2,\bar{z}_2)\varphi^{(i)*}(z_3,\bar{z}_3)\rangle=0,\quad i\geqslant 1. \tag{4.23b}$$

Inserting the explicit expression for the three-point function $\langle\psi_1\psi_2^*\varphi^{(i)*}\rangle$, dictated by conformal invariance, into Eqs. (4.23), we find at last that

$$\Delta^{(i)}=\bar{\Delta}^{(i)}=2\lambda+\tfrac{1}{2}-\sqrt{2\lambda(2\lambda+1)},\quad i\geqslant 1. \tag{4.24}$$

We see that the conformal degrees of all primary fields $\varphi^{(i)}$ coincide, signifying that we have only one conformal (and current-algebra) family contributing to the OPE of $\psi_1\psi_2^*$. The same demonstration holds for the OPE of any other two among our chiral fields ψ_1, ψ_2, ψ_1^*, ψ_2^*.

In reaching Eq. (4.24) we have intentionally stressed the role played by the Thirring equations (4.20). We could have started instead from the observation that relation (4.1b) holds for a generic primary field; then the uniqueness of U(1) charges for all fields $\varphi^{(i)}$, shown by Eqs. (4.22), would have implied uniqueness of all conformal degrees also.

The explicit form of the four-point functions can also be easily recovered using the Thirring equations and the relations (4.13) and (4.14). Let us see how to obtain Eq. (4.17a) again.

From conformal invariance we have that

$$\langle\psi_1(z_1,\bar{z}_1)\psi_2^*(z_2,\bar{z}_2)\psi_2(z_3,\bar{z}_3)\psi_1^*(z_4,\bar{z}_4)\rangle=z_{13}^{-2\lambda-1}z_{14}z_{24}^{-2\lambda-1}\bar{z}_{13}^{-2\lambda}\bar{z}_{14}^{-1}\bar{z}_{24}^{-2\lambda}f(\eta)\bar{f}(\bar{\eta}), \tag{4.25}$$

where f and $\bar{f}$ are unknown functions of η and $\bar{\eta}$, defined by Eq. (3.1b). Using the Thirring equation (4.20) in z_1 for $\psi_1(z_1,\bar{z}_1)$ and with the help of (the global version of) Eq. (4.19a), we get the following first-order differential equation for $f(\eta)$:

$$[(2\lambda+1)/(\eta-1)-\sqrt{2\lambda(2\lambda+1)}]f(\eta)+\eta f'(\eta)=0, \tag{4.26a}$$

which can be easily solved, giving

$$f(\eta)=\text{const}\times(\eta/(\eta-1))^{2\lambda+1}\times\eta^{\sqrt{2\lambda(2\lambda+1)}}. \tag{4.26b}$$

Proceeding in the same way for the $\bar{z}_1$ dependence of $\psi_1(z_1,\bar{z}_1)$, we finally get Eq. (4.17a) again.

V. REDUCTION TO THE VIRASORO SUBALGEBRA

Until now we have always dealt with the representations of the full algebra (1.1). Since the irreducible representations of the U(1) conformal CA are, in general, reducible representations of the Virasoro subalgebra (1.1a), an interesting problem for the description of the explicit reduction of a given conformal CA Verma module into a sum of the Virasoro algebra's representations arises.

Following the methods and the results of Refs. 19 and 20 for the character of the irreducible lowest-weight degenerate ($\Delta=q^2/2$) representation of the algebra (1.1), we get

$$\chi_q(t)=t^{q^2/2}\eta(t), \tag{5.1}$$

where the Dedekind function $\eta(t)$ is given by

$$\eta(t)=t^{1/24}\prod_{n=1}^{\infty}(1-t^n).$$

In the case $c=1$ for the representations of the Virasoro algebra (1.1a) one has to distinguish two different expressions for the characters of the corresponding representations[19,20]:

$$\chi^{\text{Vir}}_{\Delta}(t)=t^{\Delta}/\eta(t),\quad \text{for } \Delta\neq n^2/4,\quad n\in\mathbb{Z}, \tag{5.2a}$$

$$\chi^{\text{Vir}}_{n^2/4}(t)=(t^{n^2/4}-t^{(n+2)^2/4})/\eta(t),\quad \text{for } \Delta=n^2/4,\quad n\in\mathbb{Z}. \tag{5.2b}$$

Then for $q=n/\sqrt{2}$ we can represent the character (5.1) in terms of the Virasoro characters (5.2b),

$$\chi_{n/\sqrt{2}}(t)=\frac{1}{\eta(t)}\sum_{k=0}^{\infty}(t^{(\eta+2k)^2/4}-t^{(\eta+2k+2)^2/4})=\sum_{k=0}^{\infty}\chi^{\text{Vir}}_{(n+2k)^2/4}(t), \tag{5.3}$$

and consequently for these special values of the U(1) charge $q=n/\sqrt{2}$ the corresponding representations of the CA (1.1) contain infinitely many Virasoro representations with dimensions

$$\Delta_k=(n+2k)^2/4\equiv n^2/4+k(k+n),\quad k=0,1,\ldots.$$

In the field-theoretical language this observation means that the U(1) conformal current family $[\psi^{n}_{\Delta=q^2/2}]$ splits into infinite sets of Virasoro conformal families $[\psi^{\text{Vir}}_{\Delta_k}]$ generated by Virasoro primary fields (states). We shall give as examples the explicit constructions of the simplest primary states of Vir [they are not primary ones for the full algebra (1.1)]:

$q=0$	$k=1$	$J_{-1}\|0,0\rangle$
	$k=2$	$(4J_{-3}J_{-1}-3J^2_{-2}-2J^1_{-1})\|0,0\rangle$
$q=1/\sqrt{2}$	$k=1$	$(J_{-2}-\sqrt{2}J^2_{-1})\|\frac{1}{4},1/\sqrt{2}\rangle$
$q=\sqrt{2}$	$k=1$	$(J_{-3}-3/\sqrt{2}\,J_{-2}J_{-1}+J^3_{-1})\|1,\sqrt{2}\rangle$
$q=3/\sqrt{2}$	$k=1$	$(-3\sqrt{2}J_{-4}+8J_{-3}J_{-1}+3J^2_{-2}-6\sqrt{2}J_{-2}J^2_{-1}+2J^4_{-1})\|\frac{9}{4},3/\sqrt{2}\rangle$.

In the general case (when $\Delta\neq n^2/4$), because of the coincidence of the characters (5.1) and (5.2a) only one Virasoro representation corresponds to each conformal U(1) CA representation. Because of the fusion rules (2.13), even in these models ($q\neq n/\sqrt{2}$) we need the family of the unity operator $[I]$ ($q=0$), which contains infinitely many Virasoro conformal families:

$$\chi_0(t)=\sum_{k=0}^{\infty}\chi_{k^2}^{\mathrm{Vir}}(t).$$

Here $k=0$ corresponds to the family of the Virasoro unit operator, $k=1$ gives the family of the U(1) current, and $k=2$ ($\Delta=4$) describes the new "conserved current" family of the Virasoro primary field

$$V(z):=\sum_{n=0}^{\infty}V_n z^{-n-4},$$

where the Laurent coefficients V_n can be realized in terms of the U(1) current modes J_n,

$$\begin{aligned}V_n=&\Big(\sum_{k=-n}^{\infty}+\sum_{k=1}^{\infty}\Big)[(k-1)(k-2)\\&+(n+k+1)(n+k+2)\\&+3(k-1)(n+k+1)]J_{-k}J_{n+k}\\&-4\Big[\sum_{k=1}^{\infty}\sum_{l=1}^{\infty}J_{-k}J_{-l}L_{n+k+l}\\&+2\sum_{k=1}^{\infty}\sum_{l=0}^{\infty}J_{-k}L_{n+k-l}J_l\\&+\sum_{k=0}^{\infty}\sum_{l=0}^{\infty}L_{n-k-l}J_kJ_l\Big],\end{aligned}$$

and $V_n|0,0\rangle=0$, $n\geqslant -3$. This current has rather complicated transformation properties,

$$\begin{aligned}[L_m,V_n]&=(3m-n)V_{m+n},\\[J_m,V_n]&=2m[(m-1)(m-2)\\&\quad+(m+n+1)(m+n+2)\\&\quad+3(m-1)(m+n+1)]J_{m+n}\\&\quad-16m\Big[\sum_{k=1}^{\infty}J_{-k}L_{n+k}+\sum_{k=0}^{\infty}L_{n-k}J_k\Big],\end{aligned}$$

with respect to the full algebra (1.1), and it has no simple physical meaning as the U(1) current and the stress-energy tensor do. The same is true for the currents with dimensions k^2 ($k=3,4,5,\ldots$).

The simplest fusion rule for two primary fields with opposite charges (2.13),

$$\varphi^{(q)}_{q^2/2}\varphi^{(-q)}_{q^2/2}=[I]\ (\equiv[\varphi^{(0)}_0]),$$

can be rewritten (with respect to the Virasoro algebra) in the form

$$\varphi^{(q)}_{q^2/2}\varphi^{(-q)}_{q^2/2}=\sum_{k=0}^{\infty}[\varphi^{(0)}_{k^2}],$$

and, together with the other fusion rules (2.13), gives the multiplication laws of the infinite associative algebra of the Virasoro conformal fields describing the corresponding $c=1$ model.

Note added: After the completion of the manuscript we received the preprint by Bagger *et al.*[21] where results similar to ours connected with the treatment of the Thirring model appear.

ACKNOWLEDGMENTS

We thank I. T. Todorov and R. Paunov for useful discussions. G. M. S. is grateful to the International School for Advanced Studies in Trieste and to the local section of the Instituto Nazionale di Fisica Nucleare, Trieste, for hospitality and financial support.

[1](a) A. A. Belavin, A. M. Polyakov, and A. B. Zamolodkikov, Nucl. Phys. B **241**, 333 (1984); (b) D. Friedan, Z. Qiu, and S. Shenker, Phys. Rev. Lett. **52**, 1575 (1984).
[2]V. G. Kac, "Covariant form for infinite-dimensional Lie algebras and superalgebras," in *Lecture Notes in Physics*, Vol. 94 (Springer, Berlin, 1979), pp. 441–445.
[3]V. Knizhnik and A. B. Zamolodchikov, Nucl. Phys. B. **247**, 83 (1984).
[4]M. Bershadsky, V. Knizhnik, and M. Teitelman, Phys. Lett. B **151**, 37 (1985); D. Friedan, Z. Qiu, and S. Shenker, Phys. Lett. B **151**, 37 (1985); Z. Qiu, Nucl. Phys. B **270** [FS16], 205 (1986).
[5]D. Friedan, E. Martinec, and S. Shenker, Nucl. Phys. B **271**, 93 (1986); L. Dixon, D. Friedan, E. Martinec, and S. Shenker, Nucl. Phys. B **282**, 13 (1987).
[6]I. T. Todorov, Phys. Lett. B **153**, 77 (1985); Bulg. J. Phys. **12**, 3 (1985); "Infinite algebras in 2-dimensional QFT models," in *Differential Geometric Methods in Theoretical Physics*, edited by H. D. Doebner and T. D. Palev (World Scientific, Singapore, 1986), pp. 197–247.
[7]Vl. Dotsenko and V. Fateev, Nulc. Phys. B **240** [FS12], 312 (1984); **251** [FS13], 691 (1985).
[8]A. B. Zamolodchikov and V. A. Fadeev, (a) Zh. Eksp. Theor. Fiz. **89**, 380 (1985); (b) **90**, 1553 (1986).
[9]D. Gepner, "On the spectrum of the 2D conformal field theories," Princeton University preprint, June 1986; D. Gepner and E. Witten, Nucl. Phys. B **278**, 493 (1986).
[10](a) P. Di Vecchia, J. L. Petersen, and M. Yu, Phys. Lett. B **172**, 211 (1986); (b) G. Sotkov and M. Stanishkov, Phys. Lett. B **177**, 361 (1986).
[11]A. Schwimmer and N. Seiberg, Phys. Lett. B **184**, 191 (1987).
[12]G. Mussardo, G. Sotkov, and M. Stanishkov, "Fine structure of $N=1$ supersymmetric OPE algebras and critical statistical models," ISAS preprint 30/87/EP, Trieste, 1987.
[13]G. M. Sotkov, I. T. Todorov, M. S. Stanishkov, and V. Yu. Trifonov, "Higher symmetries in conformal QFT models," lecture at the *Symposium on Topological and Geometrical Methods in Field Theory*, Espoo, Finland, June 1986, edited by J. Hietarinta and J. Westerholm (World Scientific, Singapore, 1986), pp.195–217.
[14]W. Boucher, D. Friedan, and A. Kent, Phys. Lett. B **172**, 316 (1986); P. Di Vecchia, J. L. Petersen, M. Yu, and H. B. Zheng, Phys. Lett. B **174**, 280 (1986).
[15]P. Furlan and G. Sotkov, "U(1)^{*D} conformal current algebra models" (unpublished).
[16]B. Klaiber, "The Thirring model," in *Lectures in Theoretical Physics*, edited by A. O. Barut and N. E. Brittin (Gordon and Breach, New York,

1968), Vol. X-A, pp. 141–176; G. F. Dell'Antonio, V. Frishman, and D. Zwanziger, Phys. Rev. D **6**, 988 (1972).
[17]L. K. Hadjiivanov, S. C. Minkov, and D. T. Stoyanoiv, J. Phys. A **12**, 119 (1979).
[18]G. M. Sotkov, I. T. Todorov, and V. Yu. Trifonov, Lett. Math. Phys. **12**, 127 (1986); P. Furlan, G. M. Sotkov, and I. T. Todorov, "Quasiprimary composite fields and null vectors in critical Ising type models," ISAS preprint 69/86/EP, Trieste, 1986.
[19]C. Itzykson and J. B. Zuber, Nucl. Phys. B **275** [FS17], 580 (1986).
[20]P. Di Vecchia, H. Saleur, and J. B. Zuber, "Modular invariance in non minimal two-dimensional conformal theories," Saclay preprint SPhT-86/187.
[21]J. Bagger, D. Nemeshansky, N. Sieberg, and Sh. Yankielowicz, "Bosons, fermions and Thirring strings," Harvard University preprint HUTP-86.

ERRATA

Erratum: New classes of symmetries for partial differential equations [J. Math. Phys. 29, 806 (1988)]

G. W. Bluman and G. J. Reid
Department of Mathematics, University of British Columbia, Vancouver, British Columbia V6T 1Y4, Canada

Sukeyuki Kumei
Faculty of Textile Science, Shinsu University, Ueda, Nagano-ken 386, Japan

(Received 28 June 1988; accepted for publication 13 July 1988)

The following corrections should be made.

(i) In Eq. (4.6) $\partial/\partial t$ should be replaced by $\partial/\partial u$. Equation (4.9) should be replaced by

$$\tilde{L}_4 = x\frac{\partial}{\partial x} + \frac{2}{\nu}(u+\kappa)\frac{\partial}{\partial u} + \left[\left(1+\frac{2}{\nu}\right)\phi + \frac{2\kappa x}{\nu}\right]\frac{\partial}{\partial \phi}.$$

(ii) Result 3 on p. 810 should read as follows.

3. $K(u) = \lambda(u+\kappa)^{-2}$, $\{\lambda,\kappa\}$ arbitrary constants. Here G_T is an ∞-parameter group with infinitesimal generators $\tilde{L}_0$, $\tilde{L}_1$, and $\tilde{L}_3$ given by (4.8), $\tilde{L}_4$ given by (4.9), and

$$\tilde{L}_2 = -x(\phi+\kappa x)\frac{\partial}{\partial x} + (u+\kappa)[\phi + 2\kappa x + xu]\frac{\partial}{\partial u} + [2\lambda t + \kappa x(\phi+\kappa x)]\frac{\partial}{\partial \phi}, \tag{4.10a}$$

$$\tilde{L}_5 = -x[(\phi+\kappa x)^2 + 2\lambda t]\frac{\partial}{\partial x} + 4\lambda t^2\frac{\partial}{\partial t} + (u+\kappa)[6\lambda t + (\phi+\kappa x)^2 + 2x(u+\kappa)(\phi+\kappa x)]\frac{\partial}{\partial u} + [\kappa x(\phi+\kappa x)^2 + 2\lambda t(2\phi+3\kappa x)]\frac{\partial}{\partial \phi}, \tag{4.10b}$$

$$\tilde{L}_\infty = \Theta(w,t)\frac{\partial}{\partial x} - (u+\kappa)^2\frac{\partial \Theta(w,t)}{\partial w}\frac{\partial}{\partial u} - \kappa\Theta(w,t)\frac{\partial}{\partial \phi}. \tag{4.10c}$$

Here $w = \phi + \kappa x$ and $v = \Theta(w,t)$ is an arbitrary solution of the linear differential equation

$$\lambda\frac{\partial^2 v}{\partial w^2} - \frac{\partial v}{\partial t} = 0. \tag{4.11}$$

(iii) In result 4 on p. 810 condition (a) should be replaced by

(a) $p^2 - 4q - r^2 = 0$, $p^2 - 4q > 0$.

Erratum: The Vaidya–Patel solution with Robertson–Walker metric as a rotating inflationary scenario [J. Math. Phys. 29, 1514 (1988)]

Öyvind Grön
Oslo College of Engineering, Cort Adelers gt. 30, N-0254 Oslo 2, Norway and Institute of Physics, University of Oslo, P.O. Box 1048 Blindern, N-0316 Oslo 3, Norway

Harald H. Soleng
Institute of Physics, University of Oslo, P.O. Box 1048 Blindern, N-0316 Oslo 3, Norway

(Received 13 July 1988; accepted for publication 20 July 1988)

In Sec. III, p. 1516, left column, it is stated that the structure coefficients in an orthonormal basis satisfy the relations $C^\mu{}_{0\nu} = -C^\nu{}_{0\mu}$. This is not correct in general. Accordingly the general expression for the shear scalar expressed in terms of the structure coefficients of a comoving tetrad basis, Eq. (3.11), should read

$$\sigma^2 = \tfrac{1}{3}[(C^1_{01})^2 + (C^2_{02})^2 + (C^3_{03})^2 - C^1_{01}C^2_{02} - C^1_{01}C^3_{03} - C^2_{02}C^3_{03}] + \tfrac{1}{4}[(C^1_{02} + C^2_{01})^2 + (C^1_{03} + C^3_{01})^2 + (C^2_{03} + C^3_{02})^2]. \tag{3.11}$$

In the particular model under consideration, $C^\mu{}_{0\nu} = 0$, when $\mu \neq \nu$, and the cross terms of Eq. (3.11) are zero.

Hence the error does not affect the results obtained.

 0022-2488/88/102320-02$02.50