Eigenvalue equations for Krichever–Novikov algebras

D. B. Fairlie and P. Fletcher[a)]
Department of Mathematical Sciences, University of Durham, Durham DH1 3LE, England
J. Nuyts[b)]
CERN, Geneva 23, CH 1211, Switzerland

(Received 10 August 1988; accepted for publication 23 November 1988)

Nontrivial examples of Krichever–Novikov algebras are constructed. The construction involves parameters that satisfy eigenvalue conditions. In the linear case, for an r-term algebra, the eigenvalues are $0,1,\dots,r-3$.

I. INTRODUCTION

Recently, Krichever and Novikov[1,2] have proposed as an algebraic extension of conformal invariance on an arbitrary genus Riemann surface the study of algebras of the type

$$[N_m,N_n] = \sum_{k=0}^{r-1} C^k(m,n) N_{m+n-r+1+2k}, \qquad (1.1)$$

where the genus of the surface is $(r-1)/3$, for $r = 1 \bmod 3$. These will be referred to as Krichever–Novikov (KN) algebras. The work of Krichever and Novikov has been continued by Bonora *et al.*[3] and Mezincescu *et al.*[4] The simplest and best-known case of a KN algebra is the Virasoro algebra

$$[L_m,L_n] = (m-n)L_{m+n} + (m^3-m)c\delta_{m+n,0}, \qquad (1.2)$$

which describes conformal invariance on a sphere (genus 0). This algebra is said to be Z graded and the KN algebras (1.1) are said to be generalized graded. In this article we take a purely algebraic point of view and raise the question of the classification of the possible structure constants and their functional dependence on m, n as permitted by the Jacobi identities. This is a problem similar to the classification of finite Lie algebras and we must take care that any solutions are not simply isomorphic to the Virasoro algebra.

The case $r = 1$ of a KN algebra (without central extension) is in general

$$[L_m,L_n] = C(m,n)L_{m+n} \qquad (1.3)$$

and one solution for the structure constant is

$$C(m,n) = (m-n)[\alpha(m)\alpha(n)/\alpha(m+n)], \qquad (1.4)$$

where the arbitrary function $\alpha(m)$ can be renormalized away by dividing L_m by $\alpha(m)$. We believe that this is the only solution, having made the assumption of an ansatz for $C(m,n)$ as the ratio of multinomials in m, n—but we know of no general proof of (1.4). However, it turns out that there are many different possible sets of structure constants $C^k(m,n)$ for a KN algebra with $r > 1$.

Let us call such an algebra a three-, four-, etc. term algebra according to the number of terms on the rhs of (1.1). The determination of allowable structure constants that satisfy the Jacobi identities is an extremely difficult nonlinear problem: The number of such identities that arise in the r-term case is $2r-1$.

Fortunately, there is a class of allowable algebras for which the problem may be transmuted to a simpler, although still difficult form, i.e., those for which the operators N_m may be reexpressed as a finite sum of contiguous even or odd Virasoro generators L_m:

$$N_m = \sum_{k=0}^{r-1} a^k(m) L_{m-r+1+2k}. \qquad (1.5)$$

The conditions for N_m to generate a KN algebra are still $2r-1$ equations, but the r unknown functions $a^k(m)$ depend on one variable rather than two. One may conjecture that all KN algebras may be expressed as (1.5): Evidence for this viewpoint will be given in Sec. V. If the relationship (1.5) can be inverted to express the L's in terms of a sum (possibly semi-infinite) of the N's, then the algebra is equivalent to the full Virasoro algebra; otherwise, and we shall exhibit examples, it is an infinite subalgebra.

These algebras may be categorized according to the degree of m in the $a^k(m)$: The constraints for a closed algebra become more complicated as this increases. We shall use a convenient representation of the Virasoro generators:

$$L_m = z^{1-m}\frac{d}{dz}. \qquad (1.6)$$

Within the representation (1.6), Eq. (1.5) becomes a power series in z:

$$N_m = \sum_{k=0}^{r-1} a^k(m) z^{r-1-2k} z^{1-m}\frac{d}{dz}. \qquad (1.7)$$

We shall demonstrate that there are no constraints for the case of constant $a^k(m)$, obtain eigenvalue conditions for the parameters in $a^k(m)$ for the cases that are linear and quadratic in m, and present solutions of these that exhibit remarkable regular integral sequences.

II. CONSTANT COEFFICIENTS

The KN algebras for which the $a^k(m)$ are independent of m are straightforward and correspond to the well-known Virasoro case—the structure constants $C^k(m,n)$ are just $(m-n)$ up to multiplicative constants. Indeed, it is clear from the following theorem that all KN algebras of this form may be expressed as a sum of L's.

Theorem 1: If the coefficients $a^k(m)$ in (1.5) do not depend on m, then the N_m form a KN algebra satisfying

[a)] Supported by a Durham University Research Studentship.
[b)] Permanent address: Service de Physique Théorique, Université de l'Etat à Mons, 19 Avenue Maistriau, Mons, Belgium 7000.

 0022-2488/89/050957-08$02.50

(1.1) with $C^k(m,n) = (m-n)a^k$.

Proof: If the $a^k(m)$ are constants, (1.7) becomes

$$N_m = \sum_{k=0}^{r-1} a^k z^{r-1-2k} z^{1-m} \frac{d}{dz}$$
$$= g(z) z^{1-m} \frac{d}{dz}, \tag{2.1}$$

defining $g(z)$ as a power series. When forming, the commutator $[N_m, N_n]$ terms symmetric in m, n cancel, leaving only the part where z^{1-n} is differentiated,

$$[N_m, N_n] = (m-n)(g(z))^2 z^{1-m-n} \frac{d}{dz}$$
$$= (m-n) \sum_{k=0}^{r-1} a^k g(z) z^{1-(m+n-r+1+2k)} \frac{d}{dz}$$
$$= (m-n) \sum_{k=0}^{r-1} a^k N_{m+n-r+1+2k}\,. \tag{2.2}$$

Comparing (2.2) with (1.1) gives the result.

Alternatively, if $p(z)$ is an elliptic or hyperelliptic function of z satisfying $(p'(z))^2 = g(p(z))$, where $g(p)$ is a polynomial of finite degree in p, then

$$N_m = p'(z)(p(z))^{1-m-r+1} \frac{d}{dz} \tag{2.3}$$
$$= g(p) p^{1-m-r+1} \frac{d}{dp}\,. \tag{2.4}$$

Commuting (2.3) gives a sum of N's. If $g(p)$ is a polynomial in p^2, the N_m form a KN algebra, with constant coefficients. Putting $L_m = p^{1-m}(d/dp)$ makes (2.4) equivalent to (1.5).

III. LINEAR COEFFICIENTS

If the $a^k(m)$ depend on m the problem is much harder. When forming the commutator, $g(z)$ is a function of m, so there are more antisymmetric terms and the proof in Sec. II for the constant case no longer works. We need to find the constraints on the $a^k(m)$ to obtain a KN algebra. The Jacobi identities are satisfied automatically since we are making the assumption that N_m is a sum of L's. The condition for a KN algebra is that the N's satisfy (1.1) for some structure constants $C^k(m,n)$. The way to check this is to express N_m and N_n as sums of L's, commute them, and ensure that the resulting sum of L's may be expressed as a sum of N's:

$$\begin{array}{ccc} [N,N] & \overset{?}{=} & \Sigma N \\ \downarrow & & \uparrow \\ [\Sigma L, \Sigma L] & \rightarrow & \Sigma L \end{array}$$

We shall start by considering the following ansatz for KN algebras linear in m:

$$N_m = \left((m+f)z + (m-f)\frac{1}{z}\right)\left(z + \frac{1}{z}\right)^{r-2} z^{1-m} \frac{d}{dz} \tag{3.1}$$

and we shall find the possible values of the parameter f for a closed algebra. The ansatz (3.1) was chosen on the basis of experiments using the computer algebra package REDUCE, which indicate that all such algebras may be expressed as (3.1) or our later generalization. Note that (3.1) is invariant up to sign under $z \mapsto -1/z$, $m \mapsto -m$: We shall refer to this as the parity automorphism. This is the remnant of the automorphisms of the Virasoro algebra generated by $L_m \mapsto \pm (1/\lambda) L_{\lambda m}$. The original KN algebra has a different normalization for the N's. In Ref. 1 the normalization is chosen so that the end structure constants are given by

$$C^{r-1}(m,n) = (m-n)\,,$$
$$C^0(m,n) = (m-n)[\alpha(m)\alpha(n)/\alpha(m+n-r+1)]\,. \tag{3.2}$$

The conditions of parity invariance are then

$$N_{-m} = \mp \alpha(-m) N_m, \quad \alpha(m)\alpha(-m) = -1\,. \tag{3.3}$$

Equations (3.3) imply that the general structure of $\alpha(m)$ is given by

$$\alpha(m) = A^m \prod_i \frac{m - \beta_i}{m + \beta_i}, \tag{3.4}$$

where the β_i and A are constants. Our ansatz corresponds to the case where there is only one factor in the product. The merit of this ansatz is that the $2r-1$ conditions for closure reduce to $r+1$ linear equations for r unknowns, giving a single consistency requirement. The commutator may be calculated as

$$[N_m, N_n] = (m-n)\left\{\left((m+f)(n+f)z + (m-f)(n-f)\frac{1}{z}\right)\left(z+\frac{1}{z}\right) \times 4f(f-1)\right\}\left(z+\frac{1}{z}\right)^{2r-4} z^{1-m-n}\frac{d}{dz} \tag{3.5}$$

or, as a power series in z, as

$$[N_m, N_n] = (m-n) \times \sum_{k=0}^{r} \begin{pmatrix} \binom{r-1}{k}(m+f)(n+f) \\ + \binom{r-1}{k-1}(m-f)(n-f) \\ - \binom{r-2}{k-1} 4f(f-1) \end{pmatrix} \times z^{r-2k}\left(z+\frac{1}{z}\right)^{r-2} z^{1-m-n} \frac{d}{dz}\,. \tag{3.6}$$

Equation (1.1) may also be expressed as a power series in z by using the ansatz (3.1) for N_m and rearranging the summation:

$$[N_m, N_n] = \sum_{k=0}^{r} \begin{pmatrix} C^k(m,n)(m+n-r+2k+1+f) \\ + C^{k-1}(m,n)(m+n-r+2k-1-f) \end{pmatrix} \times z^{r-2k}\left(z+\frac{1}{z}\right)^{r-2} z^{1-m-n}\frac{d}{dz}, \tag{3.7}$$

where $C^{-1}(m,n) = C^r(m,n) = 0$.

The set of $r+1$ equations obtained by equating the coefficients in (3.6) and (3.7) may be written as a determinant which must be zero for all m, n for a closed algebra. We shall illustrate this for the case $r=5$, from which the general case is evident. We remove the factor of $(m-n)$ and define

$$\lambda = (m+f)(n+f)\,,$$

$$\mu = (m-f)(n-f)\,, \tag{3.8}$$

$$\nu = -4f(f-1)\,.$$

We will treat λ, μ, and ν as arbitrary parameters in the first stage of the analysis. Then for $r=5$ the determinant is

$$\begin{vmatrix} m+n-4+f & 0 & 0 & 0 & 0 & \binom{4}{0}\lambda \\ m+n-4-f & m+n-2+f & 0 & 0 & 0 & \binom{4}{1}\lambda+\binom{4}{0}\mu+\binom{3}{0}\nu \\ 0 & m+n-2-f & m+n+f & 0 & 0 & \binom{4}{2}\lambda+\binom{4}{1}\mu+\binom{3}{1}\nu \\ 0 & 0 & m+n-f & m+n+2+f & 0 & \binom{4}{3}\lambda+\binom{4}{2}\mu+\binom{3}{2}\nu \\ 0 & 0 & 0 & m+n+2-f & m+n+4+f & \binom{4}{4}\lambda+\binom{4}{3}\mu+\binom{3}{3}\nu \\ 0 & 0 & 0 & 0 & m+n+4-f & \binom{4}{4}\mu \end{vmatrix}.$$

This may be reduced by row and column operations. Put $R\,1' = R\,1 - R\,2 + R\,3 - R\,4 + \cdots$. Then it is evident that the combinatorial factors in the last column cancel and that $f=0$ is an eigenvalue. Then perform $C\,5' = C\,5 + C\,4$, $C\,4' = C\,4 + C\,3$, etc., giving

$$\begin{vmatrix} 2f & 0 & 0 & 0 & 0 & 0 \\ m+n-4-f & 2m+2n-6 & m+n-2+f & 0 & 0 & \binom{4}{1}\lambda+\binom{4}{0}\mu+\binom{3}{0}\nu \\ 0 & m+n-2-f & 2m+2n-2 & m+n+f & 0 & \binom{4}{2}\lambda+\binom{4}{1}\mu+\binom{3}{1}\nu \\ 0 & 0 & m+n-f & 2m+2n+2 & m+n+2+f & \binom{4}{3}\lambda+\binom{4}{2}\mu+\binom{3}{2}\nu \\ 0 & 0 & 0 & m+n+2-f & 2m+2n+6 & \binom{4}{4}\lambda+\binom{4}{3}\mu+\binom{3}{3}\nu \\ 0 & 0 & 0 & 0 & m+n+4-f & \binom{4}{4}\mu \end{vmatrix}.$$

Finally, $R\,5' = R\,5 - R\,6$, $R\,4' = R\,4 - R\,5'$,..., yielding

$$\begin{vmatrix} 2f & 0 & 0 & 0 & 0 & 0 \\ m+n-4-f & m+n-4+f & 0 & 0 & 0 & \binom{3}{0}\lambda \\ 0 & m+n-2-f & m+n-2+f & 0 & 0 & \binom{3}{1}\lambda+\binom{3}{0}\mu+\binom{2}{0}\nu \\ 0 & 0 & m+n-f & m+n+f & 0 & \binom{3}{2}\lambda+\binom{3}{1}\mu+\binom{2}{1}\nu \\ 0 & 0 & 0 & m+n+2-f & m+n+2+f & \binom{3}{3}\lambda+\binom{3}{2}\mu+\binom{2}{2}\nu \\ 0 & 0 & 0 & 0 & m+n+4-f & \binom{3}{3}\mu \end{vmatrix}.$$

This is then $2f$ times a 5×5 subdeterminant which is equivalent to the determinant in the four-term case with f shifted to $f+1$. Thus the five-term matrix has a zero at $f=0$ and also at zeros of the four-term case shifted by 1. It is clear that the same row and column operations will work for the r-term case, which has zeros at $f=0, f'+1$, where the f' are the zeros of the $(r-1)$-term case. The three-term determinant has a single zero at $f=0$, completing the inductive proof that the r-term case has zeros at $f=0,1,\ldots,r-3$, regardless of λ, μ, and ν. It should be noted that the $f=0$ case is not interesting since the factor m may be normalized away. Explicit calculation of the three-term determinant with λ, μ, and ν as in (3.8) shows that the case with $r=3$ has extra zeros at $f=1,2$: These zeros do not carry over in the above induction since they are removed by the shift in f.

This simple family may be extended by noting that the only properties of the binomial coefficients $\binom{n}{i}$ used in the above proof are that $\Sigma_{i=0}^{n}(-1)^i\binom{n}{i}=0$ and $\binom{n-1}{i-1}+\binom{n-1}{i}=\binom{n}{i}$. There are other sets of coefficients that satisfy these conditions—those in the expansion of $(z+1/z)^{n-q}(z-1/z)^q$, which we shall refer to as $\binom{n}{i}^{(q)}$. When $q=0$ these reduce to the ordinary binomial coefficients. Note that $\binom{n}{n-i}^{(q)}=(-1)^q\binom{n}{i}^{(q)}$, i.e., the parity automorphism still holds, and also that $\binom{n-1}{i-1}^{(q-1)}-\binom{n-1}{i}^{(q-1)}=\binom{n}{i}^{(q)}$.

There is very similar inductive proof for the family

$$N_m=\left((m+f)z+(-1)^p(m-f)\frac{1}{z}\right)\left(z+\frac{1}{z}\right)^{r-2-q}\times\left(z-\frac{1}{z}\right)^q z^{1-m}\frac{d}{dz}, \tag{3.9}$$

which has the three parameters r, q $(q\leqslant r-2)$, and p $(p=0,1)$. The algebra is closed for the following values of f:

$$f=\begin{cases}0, & \text{if } p=0,\ q=r-2,\\ 0,1,2, & \text{if } p=0,\ q=r-3,\\ 0,1,\dots,r-3-q, & \text{if } p=0,\ q>r-3,\\ 0,1,\dots,q-1, & \text{if } p=1.\end{cases}$$

The solution (3.9) has positive parity if $r-1-q+p$ is even. If we abandon the parity requirement, then we can find more general solutions with two free parameters, e.g., in the four-term case,

$$\begin{aligned}N_m&=(m+f)L_{m+3}+c(m+f+\tfrac{2}{3})L_{m+1}\\&\quad+(c^2/3)(m+f+\tfrac{4}{3})L_{m-1}\\&\quad+(c^3/27)(m+f+2)L_{m-3}\,.\end{aligned} \tag{3.10}$$

This is simply a transformation of the previous solution of the form

$$L_m\mapsto 3c^{m/2}L_{m+f+2}\,. \tag{3.11}$$

The parity requirement is tantamount to the imposition of unitarity or antiunitarity. Thus the eigenvalue condition plays a similar role in the restrictions on the c number for unitary representations of the Virasoro algebra found by Belavin *et al.*[5] and Friedan *et al.*[6]

IV. QUADRATIC COEFFICIENTS

The form of construction for linear coefficients generalizes to second order in m. We shall choose a basic ansatz which respects the parity operation:

$$N_m=\left[(m^2+am+b)z^2+2(m^2+c)+(m^2-am+b)\frac{1}{z^2}\right]\left(z+\frac{1}{z}\right)^{r-3}z^{1-m}\frac{d}{dz}. \tag{4.1}$$

In the case (4.1) there are two eigenvalue equations since there are $r+2$ linear equations for the r structure constants. For the simplest example, $c=b$; the ansatz (4.1) reduces to

$$N_m=\left[(m^2+am+b)z+(m^2-am+b)\frac{1}{z}\right]\left(z+\frac{1}{z}\right)^{r-2}z^{1-m}\frac{d}{dz} \tag{4.2}$$

and there is only one eigenvalue equation remaining since $c=b$ is a solution to the other one. Note that if a is zero the ansatz (4.2) is trivial since the $C^k(m,n)$ are constant; this is because the N_m can be renormalized by dividing by the factor (m^2+b). Also, if b is zero it reduces to the linear case by dividing N_m by m with a replaced by f, so it is no surprise that a satisfies the same conditions as f. For the general case, the linear equations that must be solved for the structure constants take the form

$$\begin{aligned}&\begin{pmatrix}F(m+n-r+1+2j)C^j(m,n)\\ +H(m+n-r-1+2j)C^{j-1}(m,n)\\ +G(m+n-r-3+2j)C^{j-2}(m,n)\end{pmatrix}\\&\quad=(m-n)\sum_{i=0}^{4}\binom{r-3}{j-i}R_i\,,\end{aligned} \tag{4.3}$$

where

$$\begin{aligned}&F(m)=m^2+am+b,\quad H(m)=2(m^2+c),\\&G(m)=m^2-am+b=F(-m),\\&R_0=F(m)F(n)\,,\\&R_1=F(m)H(n)+F(n)H(m)\\&\qquad+[2/(m-n)](H(m)F(n)-H(n)F(m))\,,\\&R_2=H(m)H(n)+F(m)G(n)+F(n)G(m)\\&\qquad+[4/(m-n)](F(n)G(m)-F(m)G(n))\,,\\&R_3=H(m)G(n)+H(n)G(m)\\&\qquad+[2/(m-n)](H(n)G(m)-H(m)G(n))\,,\\&R_4=G(m)G(n)\,,\end{aligned}$$

and where $j=0,\dots,r+1$ and $C^{-2}(m,n)=C^{-1}(m,n)=C^r(m,n)=C^{r+1}(m,n)=0$.

The conditions for these equations to admit a nontrivial solution are that the following $(r+2)\times(r+1)$ matrix is of rank r for all m, n:

$$
\begin{pmatrix}
F(s-r+1) & 0 & 0 & \cdots & 0 & \Sigma_{i=0}^{4}\binom{r-3}{0-i}R_i \\
H(s-r+1) & F(s-r+3) & 0 & \cdots & 0 & \Sigma_{i=0}^{4}\binom{r-3}{1-i}R_i \\
G(s-r+1) & H(s-r+3) & F(s-r+5) & \cdots & 0 & \Sigma_{i=0}^{4}\binom{r-3}{2-i}R_i \\
0 & G(s-r+3) & H(s-r+5) & \cdots & 0 & \Sigma_{i=0}^{4}\binom{r-3}{3-i}R_i \\
0 & 0 & G(s-r+5) & \cdots & 0 & \Sigma_{i=0}^{4}\binom{r-3}{4-i}R_i \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & 0 & \cdots & G(s+r-1) & \Sigma_{i=0}^{4}\binom{r-3}{r+1-i}R_i
\end{pmatrix}
$$

where $s = m + n$. The way we approach this problem is by using row and column operations to introduce two rows of zeroes, which puts constraints on a, b, and c.

It proves convenient to define the following functions:

$$\Delta(k,x) = kG(x-2) + H(x) - kG(x)\,, \quad \Phi(k) = F(x) - \Delta(k,x+2k) + G(x+2k)\,, \quad \Theta(k,x) = \Delta(k,x) - 2G(x)\,. \tag{4.4}$$

Note that Φ is independent of x and may be written as

$$\Phi(k) = 4k^2 - 4k - 4ak + 2b - 2c; \tag{4.5}$$

also, note that $\Delta(0,x) = H(x)$.

We now subtract each row from the one preceding it, starting at the bottom of the matrix and working upward.

$$
\begin{pmatrix}
\Phi(0) & -\Phi(0) & \cdots & \pm\Phi(0) & 0 \\
H(s-r+1)-G(s-r+1) & \Phi(0) & \cdots & \mp\Phi(0) & \Sigma_{i=0}^{4}\binom{r-4}{0-i}R_i \\
G(s-r+1) & H(s-r+3)-G(s-r+3) & \cdots & \pm\Phi(0) & \Sigma_{i=0}^{4}\binom{r-4}{1-i}R_i \\
0 & G(s-r+3) & \cdots & \mp\Phi(0) & \Sigma_{i=0}^{4}\binom{r-4}{2-i}R_i \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & \cdots & G(s+r-1) & \Sigma_{i=0}^{4}\binom{r-4}{r-i}R_i
\end{pmatrix}
$$

Next we add columns, starting with the penultimate one and adding the one before, working from right to left. We then repeat the previous operation of subtracting rows, this time stopping at the second row.

$$
\begin{pmatrix}
\Phi(0) & 0 & 0 & \cdots & 0 & 0 \\
\Theta(0,s-r+1) & \Phi(1) & -\Phi(1) & \cdots & \pm\Phi(1) & 0 \\
G(s-r+1) & \begin{matrix}\Delta(1,s-r+3)\\ -G(s-r+3)\end{matrix} & \Phi(1) & \cdots & \mp\Phi(1) & \Sigma_{i=0}^{4}\binom{r-5}{0-i}R_i \\
0 & G(s-r+3) & \begin{matrix}\Delta(1,s-r+5)\\ -G(s-r+5)\end{matrix} & \cdots & \pm\Phi(1) & \Sigma_{i=0}^{4}\binom{r-5}{1-i}R_i \\
0 & 0 & G(s-r+5) & \cdots & \mp\Phi(1) & \Sigma_{i=0}^{4}\binom{r-5}{2-i}R_i \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & 0 & \cdots & G(s+r-1) & \Sigma_{i=0}^{4}\binom{r-5}{r-1-i}R_i
\end{pmatrix}.
$$

We then repeat the previous operations r times, each time operating on one fewer row and column. The top part of the resulting matrix is then

$$\begin{pmatrix}
\Phi(0) & 0 & 0 & 0 & \cdots & 0 \\
\Theta(0, s-r+1) & \Phi(1) & 0 & 0 & \cdots & 0 \\
G(s-r+1) & \Theta(1, s-r+3) & \Phi(2) & 0 & \cdots & 0 \\
0 & G(s-r+3) & \Theta(2, s-r+5) & \Phi(3) & \cdots & 0 \\
0 & 0 & G(s-r+5) & \Theta(3, s-r+7) & \cdots & 0 \\
0 & 0 & 0 & G(s-r+7) & \cdots & 0 \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots
\end{pmatrix}$$

and the bottom part is

$$\begin{pmatrix}
\vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\
0 & \cdots & \Phi(r-3) & 0 & 0 & R_0 - R_1 + R_2 - R_3 + R_4 \\
0 & \cdots & \Theta(r-3, s+r-5) & \Phi(r-2) & 0 & R_1 - 2R_2 + 3R_3 - 4R_4 \\
0 & \cdots & G(s+r-5) & \Theta(r-2, s+r-3) & \Phi(r-1) & R_2 - 3R_3 + 6R_4 \\
0 & \cdots & 0 & G(s+r-3) & \Theta(r-1, s+r-1) & R_3 - 4R_4 \\
0 & \cdots & 0 & 0 & G(s+r-1) & R_4
\end{pmatrix}.$$

All the solutions (i.e., values of c, a, and b that give this matrix rank r, corresponding to closed algebras) that we have found are such that two of the Φ's are zero, which we shall label $\Phi(k)$ and $\Phi(k+l)$. If none of the Φ's are zero it is easy to prove that $r = 4$. For $r > 4$ the possible algebras (4.1) all fit into a parametrization of c, a, and b in terms of k, l, and j, given by

$$\left.\begin{aligned}
&c = b - 2k(k+l), \\
&a = 2k + l - 1, \\
&b = \begin{cases} (k+j)(k+l-j-1), & l \text{ even}, \\ \text{arbitrary}, \quad l \text{ odd}, \end{cases}
\end{aligned}\right\}$$

$$\text{where} \quad \begin{cases} k = 0,\dots,r-5, \\ l = 1,\dots,r-4-k, \end{cases}$$

$$\text{or} \quad \begin{cases} k = 0{:}\, l = r-3^{\mathrm{b}}, \quad r-2^{\mathrm{c}}, \\ k = 1{:}\, l = r-4^{\mathrm{b}}, \\ k = 2{:}\, l = r-4^{\mathrm{d}}, \end{cases} \tag{4.6}$$

and where $j = 0,\dots,\frac{1}{2}l - 1$ (l even). The marks [b], [c], and [d] indicate special cases which will be explained later.

If $\Phi(k) = 0$ for some $k \leqslant r-5$, then row $k+1$ is a linear combination of the previous rows, assuming that k is the smallest zero of Φ. Row $k+l+1$ is also a linear combination of previous rows, provided that $\Phi(k+l) = 0$ (where $k + l \leqslant r-4$) and that the determinant made from rows $(k+2),\dots,(k+l+1)$ and columns $(k+1),\dots,(k+l)$ is zero. For example, the top two rows are all zero if $\Phi(0) = 0$, $\Phi(1) = 0$, and $\Theta(0, s-r+1) = 0$, where the determinant in this case is just one element.

Now, if $\Phi(k) = \Phi(k+l) = 0$, then a, c, and the other Φ's are fixed:

$$\begin{aligned} &a = 2k + l - 1, \quad c = b - 2k(k+l), \\ &\Phi(k+j) = 4j(j-l). \end{aligned} \tag{4.7}$$

The size of the determinant is $l \times l$ and is of the form

$$\begin{vmatrix}
\Theta(k,x) & \Phi(k+1) & 0 & \cdots & 0 & 0 \\
G(x) & \Theta(k+1, x+2) & \Phi(k+2) & \cdots & 0 & 0 \\
0 & G(x+2) & \Theta(k+2, x+4) & \cdots & 0 & 0 \\
0 & 0 & G(x+4) & \cdots & 0 & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & 0 & \cdots & G(x+2l-4) & \Theta(k+l-1, x+2l-2)
\end{vmatrix}.$$

If l is odd then the above determinant is automatically zero (independent of b). By elementary row and column operations the determinant may be rewritten as an antisymmetric determinant of odd dimension and hence vanishes. Otherwise, for even l, the determinant vanishes if

$$b = (k+j)(k+l-j-1), \quad \text{for } j = 0,1,\dots,\tfrac{1}{2}l - 1. \tag{4.8}$$

We shall illustrate this for $l = 6$, from which the general case is evident. We shall make two further substitutions to simplify the expressions:

$$b' = b - k(k+l-1), \quad u = 2(k-x) - 3.$$

The determinant then becomes

$$
\begin{vmatrix}
5(u-3) & -20 & 0 & 0 & 0 & 0 \\
\frac{1}{4}(u-13)(u-3)+b' & 3u+17 & -32 & 0 & 0 & 0 \\
0 & \frac{1}{4}(u-9)(u+1)+b' & u+33 & -36 & 0 & 0 \\
0 & 0 & \frac{1}{4}(u-5)(u+5)+b' & -(u-33) & -32 & 0 \\
0 & 0 & 0 & \frac{1}{4}(u-1)(u+9)+b' & -(3u-17) & -20 \\
0 & 0 & 0 & 0 & \frac{1}{4}(u+3)(u+13)+b' & -5(u+3)
\end{vmatrix}
$$

All the lower diagonal elements may be made proportional to b' by column operations. When this is done the top left element turns out to be zero, so that if $b'=0$, the determinant is zero. One may then restore the determinant to a similar continuant form to the above by row operations. We repeat the column operations on the last four columns, so that the lower diagonal entries have a factor $(b'-4)$; again a zero appears on the diagonal [in the (3,3) position]. We restore the continuant form again by row operations on the last three rows. Finally, we repeat the column and row operations on the last two rows and columns, giving

$$
\begin{vmatrix}
0 & -20 & 0 & 0 & 0 & 0 \\
b' & -(u-13) & -32 & 0 & 0 & 0 \\
0 & \frac{1}{4}(u-13)(u+3)+b' & 0 & -36 & 0 & 0 \\
0 & 0 & b'-4 & -\frac{3}{2}(u-7) & -32 & 0 \\
0 & 0 & 0 & \frac{1}{4}(u-9)(u+1)+b' & 0 & -20 \\
0 & 0 & 0 & 0 & b'-6 & -\frac{15}{8}(u-1)
\end{vmatrix}.
$$

Manifestly, the value of the determinant is $14400b'(b'-4)(b'-6)$, with the zeros as given by (4.8).

There are two identities relating the R_i:

$$R_0 - R_1 + R_2 - R_3 + R_4 = \Phi(0)\Phi(1)\,, \tag{4.9}$$

$$R_1 - 2R_2 + 3R_3 - 4R_4 = 2\Phi(0)((a-2)(m+n) - \Phi(1))\,, \tag{4.10}$$

which introduce extra solutions for closed algebras. If $\Phi(0)=0$ or $\Phi(1)=0$, (4.9) implies that there is another zero in the last column, giving an additional value of l for $k=0,1$, marked [b]. If $\Phi(0)=0$ [or $\Phi(1)=0$ and $a=2$] then (4.10) implies that there are two extra zeros in the last column, so there is yet another value of l permitted if $k=0$, marked [c]. If $\Phi(2)=0$, $\Phi(r-2)=0$, and $r>5$ there is also a solution. In this case, the $\Phi(2)$ serves a dual purpose: It allows the third row to be made zero and allows us to make two extra zeros in the last column in the $(r-2)$ and $(r-1)$ rows. These rows are

$$
\begin{pmatrix}
0 & \cdots & \Phi(r-3) & 0 & 0 & R_0 - R_1 + R_2 - R_3 + R_4 \\
0 & \cdots & \Theta(r-3, s+r-5) & \Phi(r-2) & 0 & R_1 - 2R_2 + 3R_3 - 4R_4
\end{pmatrix}.
$$

However, if $\Phi(2)=0$, then $R_1 - 2R_2 + 3R_3 - 4R_4 = \frac{1}{2}\Phi(0)\Phi(1)(s-4)$ and the values of Φ and Θ may be calculated to give

$$
\begin{pmatrix}
0 & \cdots & -4(r-5) & 0 & 0 & \Phi(0)\Phi(1) \\
0 & \cdots & -2(r-5)(s-4) & 0 & 0 & \frac{1}{2}\Phi(0)\Phi(1)(s-4)
\end{pmatrix},
$$

showing that adding the required multiple of column $r-2$ to the last column makes the two entries shown zero. Thus the determinant of the same form as above with $k=2$ and $l=r-2$ is the only other condition. These solutions are marked [d].

In the same way as for the linear case, the matrices for $r=3$ and $r=4$ are special and their eigenvalues must be calculated explicitly. We classify the results in Table I.

It is possible to make a similar generalization to (3.9) by replacing some of the factors of $(z+1/z)$ in (4.1) by $(z-1/z)$.

V. DISCUSSION

In this paper we have raised the question of the classification of allowable KN algebras with a parity constraint and have found a pleasing integral regularity in the solutions for the parameters in our ansätze. The simplicity of the results belies the tortuous route to their discovery and makes us wonder whether there may be a deeper understanding behind them. Indeed, it is surprising that there are any zeros of our determinants independent of m, n.

A pertinent question arises as to whether these algebras are in fact merely transformations of the original Virasoro algebra. In those cases where (1.5) can be inverted to solve for L_m in terms of a sum, albeit infinite, of N's, then this just creates a representation for L_m in terms of pseudodifferential operators.[7] For the linear solutions the $a^k(m)$ have zeros which prevent this inversion for all L's; thus they are genuinely different. [The matrix of the set of linear equations (1.5) from $m=-\infty$ to a given m is lower triangular: If the diagonal contains any zeros, it is singular.] Similar results hold for the quadratic case, when b is such that the

TABLE I. Allowable quadratic KN algebras up to $r=7$.

	c	a	b	k	l
$r=3$	b	0	free	0	1[a]
		1	0		2[a]
		2	free		3[a]
$r=4$	b	0	free	0	1[a]
		1	0		2[a]
	$b-2a$	2	free	1	1[a]
		3	2		2[a]
		4	5		3[a]
	$b-4a+4$	4	3	2	1[a]
$r=5$	b	0	free	0	1
		1	0		2[b]
		2	free		3[c]
	$b-2a$	2	free	1	1[b]
$r=6$	b	0	free	0	1
		1	0		2
		2	free		3[b]
		3	0,2		4[c]
	$b-2a$	2	free	1	1
		3	2		2[b]
	$b-4a+4$	5	6	2	2[d]
$r=7$	b	0	free	0	1
		1	0		2
		2	free		3
		3	0,2		4[b]
		4	free		5[c]
	$b-2a$	2	free	1	1
		3	2		2
		4	free		3[b]
	$b-4a+4$	4	free	2	1
		6	free		3[d]

[a] The $r=3,4$ cases are special.
[b] Using (4.9).
[c] Using (4.9) and (4.10).
[d] See text.

quadratic $F(m)$ factorizes. Note that this always happens when b is not arbitrary.

Returning to the general problem of whether the KN generators can be expressed in terms of a finite sum of Virasoro generators, we adduce some evidence for this hypothesis from an examination of the structure constants for the second highest term. In the KN normalization the leading constant, as we have remarked, is simply $(m-n)$. At the next stage the Jacobi identity is linear in the structure constant $C^{r-2}(m,n)$ and is, explicitly,

$$[(m-n)C^{r-2}(m+n+r-1,p) + (m+n+r-3-p)C^{r-2}(m,n)] + \text{cyclic} = 0\,. \tag{5.1}$$

Putting $p=1-r$ we can solve for $C^{r-2}(m,n)$ in terms of $C^{r-2}(m,1-r)$ through

$$\begin{aligned} C^{r-2}(m,n) = \tfrac{1}{2}\{&(m-n)C^{r-2}(m+n+r-1,1-r) \\ &-(m-n-2)C^{r-2}(m,1-r) \\ &-(m-n+2)C^{r-2}(n,1-r)\}\,. \end{aligned} \tag{5.2}$$

The remarkable property of (5.2) is that it is also a solution of (5.1) for all m, n, and p. Furthermore, if one looks for a representation for N_m of the form of (1.5) the coefficients are given by

$$\begin{aligned} &a^{r-1}(m) = 1\,, \\ &a^{r-2}(m) = C^{r-1}(m,1-r) + (m+r)a^{r-2}(1-r)\,. \end{aligned} \tag{5.3}$$

Of course, this is only a first-order identification, but it holds promise that similar connections may be deduced for further terms in the expansion. This conjecture is supported by computer experiments using REDUCE.

We have not discussed the possible central terms in the algebra. Such terms will be automatically induced by whatever representation is employed for the Virasoro generators.

ACKNOWLEDGMENTS

We would like to thank Cosmas Zachos for many useful conversations. Author PF thanks the University of Durham for a Research Studentship and author JN thanks CERN for its hospitality.

[1] I. M. Krichever and S. P. Novikov, Funkcional. Anal. i Priložen. **21** (2), 46 (1987); **21** (4), 47 (1987).
[2] L. Alvarez-Gaumé, C. Gomez, G. Moore, and C. Vafa, preprint CERN-TH-4883/87.
[3] L. Bonora, M. Bregola, P. Cotta-Ramusino, and M. Martellini, Phys. Lett. B **205**, 53 (1988); L. Bonora, M. Martellini, M. Rinaldi, and J. Russo, Phys. Lett. B **206**, 444 (1988); L. Bonora, M. Rinaldi, J. Russo, and K. Wu, preprint SISSA 38/EP.
[4] L. Mezincescu, R. Nepomechie, and C. Zachos, preprint ANL-HEP-PR-88-23.
[5] A. A. Belavin, A. M. Polyakov, and A. B. Zamolodchikov, Nucl. Phys. B **241**, 333 (1984).
[6] D. Friedan, Z. Qiu, and S. H. Shenker, Phys. Rev. Lett. **52**, 1575 (1984).
[7] A. Beilinson, Yu. Manin, and V. Schechtman, preprint, Steklov Institute (1986).

Construction of indecomposable representations of inhomogeneous Lie groups

Vittorio Cantoni[a)]
Dipartimento di Matematica, Università di Milano, Via C. Saldini 50, Milano, Italy

(Received 27 April 1988; accepted for publication 7 December 1988)

Simple methods of construction of indecomposable representations of inhomogeneous Lie groups are considered and applied to the Poincaré group.

I. INTRODUCTION

In the last two decades there has been an increasing interest in the study of indecomposable representations of Lie groups and algebras. From the point of view of mathematical physics the attention is naturally focused on groups that present themselves as symmetries of physical models, and in this respect prominent roles are played by the homogeneous and the inhomogeneous Lorentz groups.[1] In particular, indecomposable representations of these groups have been considered in connection with unstable particles,[2,3] and it has been conjectured that they might prove significant, perhaps even "essential" (Dirac[4]) as a basis for new rigorous theories of interactions.

The purpose of this paper is not to discuss such conjectures, but rather to provide, for a class of Lie groups that includes the inhomogeneous Lorentz group (Poincaré group), and for their Lie algebras, simple methods of construction of indecomposable representations of a kind that might turn out to be suitable for applications.

The first part of the paper describes a method of construction that applies to any "inhomogeneous" Lie group and its Lie algebra. Although in the case of the Poincaré group the indecomposable representations so obtained essentially coincide with the ones determined by Barut,[2] Rączka,[3] and Guichardet[5] via Mackey's induction technique, we believe that the method presented here deserves to be brought to attention for its simplicity, and because it can be regarded as a particular application of the more general construction considered in Sec. III.

While each of the representations directly obtainable by the method described in Sec. II has irreducible subquotients that are mutually equivalent, the more general method described in Sec. III can also be used to construct indecomposable representations possessing preassigned *inequivalent* subquotients. In contrast with the simpler situation of Sec. II, the actual possibility of such an "assemblage" of two representations now depends, in general, on the structure of the latter.

In Sec. IV the method is adapted to the particular case of the Poincaré group $\mathscr{P}$, where certain simplifications arise. By way of illustration, it is shown how to recover the "natural representation" of $\mathscr{P}$ by 5×5 matrices, and the special representations obtainable by the simpler construction of Sec. II. Finally, examples of indecomposable representations of $\mathscr{P}$ with inequivalent *unitary* subquotients are exhibited; but a full classification of such representations has not been attempted yet.

II. INDECOMPOSABLE REPRESENTATIONS WITH EQUIVALENT SUBQUOTIENTS

By *inhomogeneous Lie group* we mean a Lie group $G = NH$ with a semidirect product structure with respect to a *commutative* normal closed subgroup N and to a closed subgroup H. (In particular, G could be any inhomogeneous orthogonal or pseudoorthogonal group, e.g., the inhomogeneous Lorentz group.)

Given an inhomogeneous Lie group G with Lie algebra $\mathfrak{g}$, we shall show that to each irreducible faithful representation ρ of $\mathfrak{g}$ on a complex (possibly infinite-dimensional) linear space V there corresponds a series of indecomposable representations $R_{\mu n}$ of $\mathfrak{g}$, parametrized by an arbitrary complex number μ and an arbitrary integer n. If ρ is finite dimensional, each $R_{\mu n}$ generates an indecomposable representation of G_0, the connected component of the identity of the group G. If ρ (not necessarily finite dimensional) generates a unitary representation of G_0, the representations $R_{\mu n}$ generate indecomposable representations of G_0 at least for real values of the parameter μ.

A. Indecomposable representations of the Lie algebra

The assumptions on the structure of G imply that its Lie algebra $\mathfrak{g}$ is a semidirect sum of the Lie algebras $\mathfrak{n}$ and $\mathfrak{h}$ of N and H, respectively, $\mathfrak{n}$ being a commutative ideal of $\mathfrak{g}$, so that

$$[\mathfrak{h},\mathfrak{h}] \subset \mathfrak{h}, \tag{1}$$

$$[\mathfrak{n},\mathfrak{h}] \subset \mathfrak{n}, \tag{2}$$

$$[\mathfrak{n},\mathfrak{n}] = 0. \tag{3}$$

For a given representation

$\rho: \gamma \to \rho(\gamma) \quad (\gamma \in \mathfrak{g})$

of the Lie algebra $\mathfrak{g}$, denote by

$\rho_{\mathfrak{h}}: \quad \chi \to \rho(\chi) \quad (\chi \in \mathfrak{h})$

the restriction of ρ to $\mathfrak{h}$. For a given integer n, consider the direct sum ${}^nR$ of n copies of $\rho_{\mathfrak{h}}$, so that the representation space ${}^nV$ is constituted by the ordered n-tuples of vectors of

[a)] Temporary member of MSRI, Berkeley, CA.

 0022-2488/89/050965-05$02.50

the representation space V of ρ. Each n-tuple will be regarded as a column matrix

$$\begin{pmatrix} v^1 \\ v^2 \\ \vdots \\ v^n \end{pmatrix} \quad (v^i \in V),$$

and briefly denoted by $\{v^i\}$. [Similarly, any square matrix with entries a^i_j will be denoted by $\{a^i_j\}$, or simply a, the upper index referring to the rows and the lower index to the columns $(i,j = 1,2,...,n)$.]

The representation ${}^nR$ is characterized by the action

$${}^nR(\chi)\ \{v^i\} = \begin{pmatrix} \rho(\chi)v^1 \\ \rho(\chi)v^2 \\ \vdots \\ \rho(\chi)v^n \end{pmatrix} \equiv \{\rho(\chi)v^i\} \quad \text{for } \chi \in \mathfrak{h}. \tag{4}$$

There are many ways to extend the representation ${}^nR$ of $\mathfrak{h}$ to a representation of $\mathfrak{g}$. Of course, one could take the direct sum R of n copies of ρ, characterized by the action

$$R(\gamma)\{v^i\} = \{\rho(\gamma)v^i\} \quad (\gamma \in \mathfrak{g}).$$

But one can also define, more generally, a representation R_α such that

$$R_\alpha(\chi) = {}^nR(\chi), \quad \text{for } \chi \in \mathfrak{h}, \tag{5}$$

with ${}^nR(\chi)$ defined as in (4), and

$$R_\alpha(\nu)\{v^i\} = \{\alpha^i_j \rho(\nu) v^j\}, \quad \text{for } \nu \in \mathfrak{n}, \tag{6}$$

where $\alpha \equiv \{\alpha^i_j\}$ is an arbitrarily fixed complex $n \times n$ matrix. [The summation convention on repeated indices is adopted. Therefore the right-hand side of (6) represents a column matrix. All indices run from 1 to n.]

It is easy to check that (4)–(6) actually define a representation of $\mathfrak{g}$. In fact, by construction ${}^nR$ is a representation of $\mathfrak{h}$, so that

$$R_\alpha(\chi_1)R_\alpha(\chi_2) - R_\alpha(\chi_2)R_\alpha(\chi_1) = R_\alpha([\chi_1,\chi_2]), \quad \text{for } \chi_1,\chi_2 \in \mathfrak{h}.$$

Moreover, since the operators $\rho(\gamma)$ are linear and for $\nu \in \mathfrak{n}$ the map $\nu \to \rho(\nu)$ is a representation of the commutative Lie algebra $\mathfrak{n}$, from (6) and for ν_1, $\nu_2 \in \mathfrak{n}$ one has

$$\begin{aligned}(R_\alpha(\nu_1)R_\alpha(\nu_2) - R_\alpha(\nu_2)R_\alpha(\nu_1))\ \{v^i\} &= \{\alpha^i_j \alpha^j_k (\rho(\nu_1)\rho(\nu_2) - \rho(\nu_2)\rho(\nu_1))v^k\} \\ &= \{\alpha^i_j \alpha^j_k \rho([\nu_1,\nu_2])v^k\} = 0,\end{aligned}$$

so that

$$R_\alpha(\nu_1)R_\alpha(\nu_2) - R_\alpha(\nu_2)R_\alpha(\nu_1) = R_\alpha([\nu_1,\nu_2]),$$

for $\nu_1,\nu_2 \in \mathfrak{n}$. Finally, for $\chi \in \mathfrak{h}$ and $\nu \in \mathfrak{n}$ one gets, from (5) and (6),

$$\begin{aligned}(R_\alpha(\nu)R_\alpha(\chi) - R_\alpha(\chi)R_\alpha(\nu))\{v^i\} &= \{\alpha^i_j(\rho(\nu)\rho(\chi) - \rho(\chi)\rho(\nu))v^j\} \\ &= \{\alpha^i_j \rho([\nu,\chi])v^j\} = R_\alpha([\nu,\chi])\{v^i\},\end{aligned}$$

where $R_\alpha([\nu,\chi])$ has the form (6), correctly on account of (2). Hence one gets the desired commutation relations for the representatives of any basis of $\mathfrak{g}$ with elements belonging either to $\mathfrak{n}$ or to $\mathfrak{h}$.

Consider now two representations R_α and R_β of $\mathfrak{g}$ constructed as above, from *the same* representation ρ but by means of *different* complex matrices α and β. *Here R_α and R_β are equivalent representations whenever α and β are equivalent matrices.*

In fact, let $m \equiv \{m^i_j\}$ be an invertible matrix such that $\alpha = m\beta m^{-1}$. On account of the linearity of the operators $\rho(\gamma)$, for the linear map M of ${}^nV$ onto itself defined by

$$M: \quad \{v^i\} \to M\{v^i\} \equiv \{m^i_j v^j\},$$

one has $\{\rho(\chi)m^i_j v^j\} = \{m^i_j \rho(\chi)v^j\}$ so that, on account of (4) and (5),

$$R_\alpha(\chi)M = {}^nR(\chi)M = M\,{}^nR(\chi) = MR_\beta(\chi), \quad \text{for } \chi \in \mathfrak{h}.$$

On the other hand, one has

$$\{\alpha^i_j \rho(\nu) m^j_k v^k\} = \{m^i_j \beta^j_k \rho(\nu) v^k\}, \tag{7}$$

so that, on account of (6),

$$R_\alpha(\nu)M = \mu R_\beta(\nu), \quad \text{for } \nu \in \mathfrak{n}. \tag{8}$$

The relations (7) and (8) exhibit the equivalence of the representations R_α and R_β.

Therefore, without loss of generality, in our construction it can be assumed that the matrix $\{\alpha^i_j\}$ has Jordan canonical form. To each Jordan block

$$\begin{pmatrix} \mu & 1 & 0 & \cdots & 0 \\ & \mu & \ddots & \ddots & \vdots \\ & & \ddots & \ddots & 0 \\ & 0 & & \ddots & 1 \\ & & & & \mu \end{pmatrix}$$

there corresponds a representation of $\mathfrak{g}$, which will be denoted by $R_{\mu n}$, parametrized by a complex number μ (the common value of the diagonal elements) and an integer n (the dimension of the block). Notice that R_{11} is just the representation ρ itself.

It is evident that if ρ is irreducible, and faithful, each representation $R_{\mu n}$ is indecomposable, with irreducible subquotients all equivalent to $R_{\mu 1}$.

We shall now consider two types of sufficient conditions in order that the representations $R_{\mu n}$ of the Lie algebra $\mathfrak{g}$ generate representations of the connected component of the identity of the group G.

B. Indecomposable representations of the group

If the representation ρ: $\gamma \to \rho(\gamma)$ of $\mathfrak{g}$ is finite dimensional, it generates a representation ρ: $g \to \rho(g)$ of G_0. In this case the representation $R_{\mu n}$: $\gamma \to R_{\mu n}(\gamma)$ of $\mathfrak{g}$ generates a representation $R_{\mu n}$: $g \to R_{\mu n}(g)$ of G_0 whose restriction to $H_0 \equiv H \cap G_0$ coincides with the direct sum of n copies of the restriction of ρ to H_0, while for any element of N of the form $\exp \nu$ $(\nu \in \mathfrak{n})$, one has

$$R_{\mu n}(\exp \nu) = \exp\begin{pmatrix} \mu\rho(\nu) & \rho(\nu) & 0 & \cdots & \cdots & 0 \\ & \mu\rho(\nu) & \rho(\nu) & 0 & \cdots & 0 \\ & & & & & \vdots \\ & 0 & & & & \rho(\nu) \\ & & & & & \mu\rho(\nu) \end{pmatrix}$$
$$= \exp(\mu\rho(\nu)) \cdot \begin{pmatrix} 1 & \rho(\nu) & (1/2!)(\rho(\nu))^2 \cdots [1/(n-1)!](\rho(\nu))^{n-1} \\ & 1 & \rho(\nu) \cdots [1/(n-2)!](\rho(\nu))^{n-2} \\ & & \vdots \\ & 0 & \rho(\nu) \\ & & 1 \end{pmatrix} \tag{9}$$

(to be interpreted as acting on $\{v^i\}$ by matrix multiplication on the left).

If the representation ρ: $\gamma \to \rho(\gamma)$ of $\mathfrak{g}$ (now not necessarily finite dimensional) generates a unitary representation $\rho: g \to \rho(g)$ of G_0, again the restriction to $\mathfrak{h}$ of the representation $R_{\mu n}$ of $\mathfrak{g}$ generates the direct sum of n copies of the restriction of ρ to H_0. In the representation ρ of G_0 the elements of N of the form $\exp \nu$ ($\nu \in \mathfrak{n}$) are represented by unitary operators of the form $\exp \rho(\nu)$, where $\rho(\nu)$ is skew adjoint. If μ is real, $\mu\rho(\nu)$ is also skew adjoint, and $R_{\mu n}(\exp \nu)$ is well defined and given by (9).

Thus for any choice of μ and n in the finite-dimensional case, and at least for real values of μ and arbitrary n when ρ gives rise to a unitary representation of G_0, the representations $R_{\mu n}$ give rise to indecomposable representations of G_0.

III. A MORE GENERAL CONSTRUCTION

Given two representations ρ and σ, we shall say that they can be *assembled* (in the given order) if there exists an indecomposable representation R with a subrepresentation equivalent to ρ and quotient equivalent to σ.

With the same assumptions on G as in Sec. II, we now consider the problem of assembling two of its representations, ρ and σ, without assuming that they be mutually equivalent. We shall also allow ρ and σ to be reducible, so that indecomposable representations of increasing length might be obtained by repeated use of the assembling process.

A. The assemblage of two representations

First, in analogy with the construction of Sec. II A, and with similar notations, we construct the direct sum R of the restrictions $\rho_{\mathfrak{h}}$ and $\sigma_{\mathfrak{h}}$ of ρ and σ to the Lie algebra $\mathfrak{h}$, acting on the linear space $V \equiv V_\rho \oplus V_\sigma$ of vectors $v \equiv \binom{V_\rho}{V_\sigma}$. In this representation the representative $R(\chi)$ of an element χ of $\mathfrak{h}$ acts on V according to the scheme

$$R(\chi):\ \begin{pmatrix} V_\rho \\ V_\sigma \end{pmatrix} \to \begin{pmatrix} \rho(\chi) & 0 \\ 0 & \sigma(\chi) \end{pmatrix} \begin{pmatrix} v_\rho \\ v_\sigma \end{pmatrix}. \tag{10}$$

The representation R could be extended to a representation of $\mathfrak{g}$ by representing the generic element of $\mathfrak{n}$ by

$$R(\nu):\ \begin{pmatrix} v_\rho \\ v_\sigma \end{pmatrix} \to \begin{pmatrix} \rho(\nu) & 0 \\ 0 & \sigma(\nu) \end{pmatrix} \begin{pmatrix} v_\rho \\ v_\sigma \end{pmatrix}, \tag{11}$$

and this would simply give the direct sum of ρ and σ. But if we wish the extended representation R to be indecomposable, with a subrepresentation ρ and quotient σ, we can try to replace the operators $R(\nu)$ given by (11) by new operators of the form

$$R(\nu):\ \begin{pmatrix} v_\rho \\ v_\sigma \end{pmatrix} \to \begin{pmatrix} \rho(\nu) & \tau(\nu) \\ 0 & \sigma(\nu) \end{pmatrix} \begin{pmatrix} v_\rho \\ v^\sigma \end{pmatrix} \quad (\nu \in \mathfrak{n}), \tag{12}$$

where, for each $\nu \in \mathfrak{n}$, $\tau(\nu)$ is a linear map from V_σ into V_ρ [i.e., $\tau(\nu) \in L(V_\sigma, V_\rho)$], and $\nu \to \tau(\nu)$ is a linear map from $\mathfrak{n}$ into $L(V_\sigma, V_\rho)$.

The choice of τ must be compatible with the commutation relations (2) and (3) of the elements of $\mathfrak{n}$ with the elements of $\mathfrak{h}$, and of the elements of $\mathfrak{n}$ with each other. In terms of the basis elements in the finite-dimensional spaces $\mathfrak{h}$ and $\mathfrak{n}$, such commutation relations give two finite sets of "assembling conditions" involving τ. The problem is to see whether for given ρ and σ there exist nonzero choices of τ satisfying such conditions, and, if so, to determine the possible choices of τ explicitly.

B. Elementary examples

Before proceeding to the application of the method to the Poincaré group, let us consider two simple examples provided by the Euclidian group of the plane. In this case the generic element of G will be denoted by $(\mathbf{t},\varphi)$, where $\mathbf{t} = a\mathbf{u} + b\mathbf{v}$ is the translation vector with components a and b with respect to axes x and y oriented as the unit vectors $\mathbf{u}$ and $\mathbf{v}$, φ is the angle of rotation, and the action on the plane xy is given by

$$(\mathbf{t},\varphi):\quad (x,y) \to (x\cos\varphi - y\sin\varphi + a, x\sin\varphi + y\cos\varphi + b). \tag{13}$$

The generators e_0, e_1, and e_2 of the one-parameter subgroups of rotations and translations along the axes satisfy the relations

$$[e_0,e_1] = e_2,$$
$$[e_0,e_2] = -e_1, \tag{14}$$
$$[e_1,e_2] = 0. \tag{15}$$

(a) Let ρ and σ be the representations of G given by

$$\rho((\mathbf{t},\varphi)) = e^{i\varphi}, \quad \sigma((\mathbf{t},\varphi)) = 1.$$

The representation spaces V_ρ and V_σ are one dimensional and one has

$$\rho(e_0) = i, \quad \rho(e_1) = 0, \quad \rho(e_2) = 0,$$
$$\sigma(e_0) = 0, \quad \sigma(e_1) = 0, \quad \sigma(e_2) = 0.$$

In R one has

$$R(e_0)=\begin{pmatrix} i & 0\\ 0 & 0\end{pmatrix}$$

and we want to represent e_1 and e_2 by

$$R(e_1)=\begin{pmatrix} 0 & \tau_1\\ 0 & 0\end{pmatrix},\quad R(e_2)=\begin{pmatrix} 0 & \tau_2\\ 0 & 0\end{pmatrix}.$$

The commutation relations (14) give $i\tau_1=\tau_2$, $i\tau_2=-\tau_1$, while (15) is automatically satisfied. So we can take $\tau_1=1$, $\tau_2=i$. Exponentiating

$$\varphi\begin{pmatrix} i & 0\\ 0 & 0\end{pmatrix},\quad a\begin{pmatrix} 0 & 1\\ 0 & 0\end{pmatrix},\text{ and } b\begin{pmatrix} 0 & i\\ 0 & 0\end{pmatrix}$$

we get

$$R((0,\varphi))=\begin{pmatrix} e^{i\varphi} & 0\\ 0 & 1\end{pmatrix},\quad R((a\mathbf{u},0))=\begin{pmatrix} 0 & a\\ 0 & 1\end{pmatrix},\quad R((b\mathbf{v},0))=\begin{pmatrix} 0 & ib\\ 0 & 1\end{pmatrix}$$

so that

$$R((\mathbf{t},\varphi))=\begin{pmatrix} e^{i\varphi} & a+ib\\ 0 & 1\end{pmatrix}.$$

The action on vectors of $V_\rho\oplus V_\sigma$ of the form $\binom{z}{1}$ is given by

$$\begin{pmatrix} z\\ 1\end{pmatrix}\to\begin{pmatrix} e^{i\varphi} & a+ib\\ 0 & 1\end{pmatrix}\begin{pmatrix} z\\ 1\end{pmatrix}=\begin{pmatrix} e^{i\varphi}z+a+ib\\ 1\end{pmatrix},$$

and, setting $z=x+iy$, from this indecomposable representation we recover the full action (13) of the group on the plane.

Thus, in this example, the assemblage of two representations of G in each of which the translations were represented trivially gives rise to a faithful representation of G.

(b) Now choose ρ and σ such that

$$\rho((\mathbf{t},\varphi))=e^{im_1\varphi},\quad \sigma((\mathbf{t},\varphi))=e^{im_2\varphi}.$$

Again V_ρ and V_σ are one dimensional. We have

$$R(e_0)=\begin{pmatrix} im_1 & 0\\ 0 & im_2\end{pmatrix},$$

and we want to represent e_1 and e_2 by

$$R(e_1)=\begin{pmatrix} 0 & \tau_1\\ 0 & 0\end{pmatrix},\quad R(e_2)=\begin{pmatrix} 0 & \tau_2\\ 0 & 0\end{pmatrix}.$$

The relations (13) and (14) now give $i(m_1-m_2)\tau_1=\tau_2$, $i(m_1-m_2)\tau_2=-\tau_1$, from which it is easily inferred that if τ does not vanish, one must have $m_2=m_1\pm 1$. No further restriction is imposed by the remaining condition (15).

This example shows that the assemblage is possible, in general, only if the given representations ρ and σ satisfy certain compatibility conditions.

IV. APPLICATION TO THE POINCARÉ GROUP

Henceforth we shall assume that G is the Poincaré group $\mathscr{P}$, N is the four-dimensional subgroup of space-time translations, and H the covering group $\mathrm{Sl}(2,C)$ of the homogeneous Lorentz group. In a suitable basis $\{m_{ab},p_c\}$ (where the indices take the values 0,1,2,3 and $a<b$) the structure of the complexified Lie algebra is given by

$$[m_{ab},m_{cd}]=-i(g_{ac}m_{bd}+g_{bd}m_{ac}-g_{ad}m_{bc}-g_{bc}m_{ad}),\tag{16}$$

$$[m_{ab},p_c]=i(g_{bc}p_a-g_{ac}p_b),\tag{17}$$

$$[p_b,p_c]=0\tag{18}$$

(where $g_{00}=-g_{11}=-g_{22}=-g_{33}=1$).

Let ρ and σ be representations of $\mathfrak{g}$. We shall denote by $\{{}^\rho M_{ab},{}^\rho P_c\}$ and $\{{}^\sigma M_{ab},{}^\sigma P_c\}$ the corresponding representatives of the basis elements $\{m_{ab},p_c\}$. They satisfy commutation relations analogous to (16)–(18).

We consider the problem of constructing a representation R by assembling ρ and σ.

A. Reduction of the assembling conditions

In the representation R (whenever it exists) the representatives of the basis elements of $\mathfrak{g}$ will be denoted simply by $\{M_{ab},P_c\}$, so that the commutation relations corresponding to (17) and (18) are

$$[M_{ab},P_c]=i(g_{bc}P_a-g_{ac}P_b),\tag{19}$$

$$[P_b,P_c]=0.\tag{20}$$

According to the construction of Sec. III A, the operators M_{ab} and P_c acting on the direct sum of the representation spaces V_ρ and V_σ should be schematically represented by matrices of the form

$$\begin{pmatrix} {}^\rho M_{ab} & 0\\ 0 & {}^\sigma M_{ab}\end{pmatrix}\begin{pmatrix} {}^\rho P_c & \tau_c\\ 0 & {}^\sigma P_c\end{pmatrix},\tag{21}$$

where the entries with a label ρ or σ are operators acting on V_ρ and V_σ, respectively, while the entries τ_c are linear maps from V_σ to V_ρ.

Setting $M_{03}=N_3$, the set of assembling conditions (19) can now be replaced by the single equation

$$[N_3,[N_3,P_0]]+P_0=0,\tag{22}$$

which is a consequence of the commutation relations (19) themselves: in fact, it can be shown[6] that whenever it is possible to determine an operator P_0 satisfying (22), there exists a unique set of operators P_1, P_2, and P_3 (determined by P_0 and by the commutation relations) such that the four P_a's satisfy all the conditions (19).

Similarly, it can be shown[7] that the operators P_a commute with each other, as required by the set of assembling conditions (20), if and only if the single equation

$$[P_0,[P_0,N_3]]=0\tag{23}$$

holds.

Thus our problem is reduced to the solution of the two equations (22) and (23), where P_0 has the form

$$P_0=\begin{pmatrix} {}^\rho P_0 & \tau_0\\ 0 & {}^\sigma P_0\end{pmatrix}$$

and the unknown in τ_0.

Since ρ and σ are representations of $\mathfrak{g}$, so that the pair of operators $({}^\rho P_0,{}^\rho N_3)$ and $({}^\sigma P_0,{}^\sigma N_3)$ already satisfy equations analogous to (22) and (23), it is easily seen from the matrix representations of P_0 and N_3 that (22) and (23) themselves are equivalent to the two equations

$$({}^\rho N_3)^2\tau_0+\tau_0({}^\sigma N_3)^2-2{}^\rho N_3\tau_0\,{}^\sigma N_3+\tau_0=0,\tag{24}$$

and

$$ {}^\rho P_0 \tau_0\, {}^\sigma N_3 + {}^\rho N_3 \tau_0^\sigma P_0 + \tau_0\, {}^\sigma P_0\, {}^\sigma N_3 + {}^\rho N_3\, {}^\rho P_0 \tau_0 - 2(\tau_0\, {}^\sigma N_3^\sigma P_0 + {}^\rho P_0\, {}^\rho N_3 \tau_0) = 0. \quad (25) $$

B. Simple examples

(a) Let ρ be the representation of the homogeneous Lorentz group on four-vectors of Minkowskian space-time, extended trivially to $\mathscr{P}$ (i.e., the translations are represented by the identity transformation). Let σ be the trivial one-dimensional representation of $\mathscr{P}$ (i.e., every element of $\mathscr{P}$ is represented by the identity). Then, with respect to an orthonormal basis $\{e_0, e_1, e_2, e_3\}$ in space-time, ${}^\rho N_3$ has the form

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

${}^\sigma N_3$, ${}^\rho P_0$, and ${}^\sigma P_0$ are zero, and it is easy to check that in this case no actual restriction is imposed on the choice of τ_0 by conditions (24) and (25). Thus ρ and σ can be assembled by means of any nonzero linear map from V_σ to V_ρ, represented on the direct sum $V_\rho \oplus V_\sigma$ by a 5×5 matrix of the form

$$ \begin{pmatrix} 0 & 0 & 0 & 0 & \tau_0^0 \\ 0 & 0 & 0 & 0 & \tau_0^1 \\ 0 & 0 & 0 & 0 & \tau_0^2 \\ 0 & 0 & 0 & 0 & \tau_0^3 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}. $$

If, in particular, we make the choice $\tau_0^0 = 1$, $\tau_0^1 = \tau_0^2 = \tau_0^3 = 0$, we get a representation of the Lie algebra which generates, by exponentiation, the representation of $\mathscr{P}$ by 5×5 matrices describing the natural action of the group on space-time,

$$ \begin{pmatrix} x^0 \\ x^1 \\ x^2 \\ x^3 \\ 1 \end{pmatrix} \to \begin{pmatrix} & & & & a^0 \\ & & & & a^1 \\ & & \Lambda & & a^2 \\ & & & & a^3 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x^0 \\ x^1 \\ x^2 \\ x^3 \\ 1 \end{pmatrix} = \begin{pmatrix} \Lambda_i^0 x^i + a^0 \\ \Lambda_i^1 x^i + a^1 \\ \Lambda_i^2 x^i + a^2 \\ \Lambda_i^3 x^i + a^3 \\ 1 \end{pmatrix}, $$

where Λ is the 4×4 matrix representing the homogeneous part of the transformation, and the a^i's are the components of the translation vector.

(b) Let ρ be any faithful representation of $\mathscr{P}$, and let σ be equivalent to ρ. Now the representation spaces V_ρ and V_σ can be identified, and $V = V_\rho \oplus V_\sigma$. Here N_3 and P_0 have the forms

$$ N_3 = \begin{pmatrix} {}^\rho N_3 & 0 \\ 0 & {}^\rho N_3 \end{pmatrix}, \quad P_0 = \begin{pmatrix} {}^\rho P_0 & \tau_0 \\ 0 & {}^\rho P_0 \end{pmatrix}, $$

and since ρ is a representation of $\mathscr{P}$, it must satisfy Eqs. (22) and (23), i.e.,

$$ ({}^\rho N_3)^2\, {}^\rho P_0 + {}^\rho P_0 ({}^\rho N_3)^2 - 2\,{}^\rho N_3\, {}^\rho P_0\, {}^\rho N_3 + {}^\rho P_0 = 0, \quad (26) $$

$$ ({}^\rho P_0)^2\, {}^\rho N_3 + {}^\rho N_3 ({}^\rho P_0)^2 - 2\,{}^\rho P_0\, {}^\rho N_3\, {}^\rho P_0 = 0. \quad (27) $$

It is immediately seen that in this case Eqs. (24) and (25) coincide with (26) and (27) if one chooses

$$ \tau_0 = {}^\rho P_0. $$

This is a special case of the construction of Sec. II, which could be entirely recovered by a slight generalization of this example and iteration.

C. Examples of representations with irreducible unitary subquotients

Consider the representation of the principal series of $\mathrm{Sl}(2,C)$ associated with the integer or half-odd integer j_0 and with the real number λ such that the operators

$$ F = -\frac{1}{2} \sum_{a<b} (M_{ab})^2 $$

and

$$ G = \frac{1}{4} \sum \epsilon^{abcd} M_{ab} M_{cd} $$

have eigenvalues $1 + \lambda^2 - j_0^2$ and $j_0 \lambda$, respectively. Denote by $H^{(j_0,\lambda)}$ its representation space.

It has been shown in Ref. 7 that for each j_0 (positive, negative, or zero) the direct integral representation T^{j_0} of $\mathrm{Sl}(2,C)$ acting on the direct integral of Hilbert spaces

$$ H^{j_0} = \int_{-\infty}^{\infty} H^{(j_0,\lambda)}\, d\lambda $$

can be extended, by means of suitably defined operators P_a, to a representation U^{j_0} of the Poincaré group $\mathscr{P}$ which is a realization of the irreducible unitary representation with zero mass and helicity j_0.

On the other hand, an equation identical to our present condition (24) was considered in Ref. 9, where the unknown was a map $P_{0\,j_0'}^{\,j_0}$ from H^{j_0} to $H^{j_0'}$ (corresponding to our present map τ_0 from V_σ to V_ρ). It was shown that the equation admits nonzero solutions provided that $j_0' = j_0 \pm 1$ or $j_0' = j_0$, and that the solutions depend on arbitrary functions. It can be shown that if our representations ρ and σ are identified with $U^{j_0'}$ and U^{j_0}, respectively, such functions can be chosen in such a way that the assembling condition (25) is also satisfied. Therefore U^{j_0} and $U^{j_0'}$ can be assembled provided that $j_0' = j_0 \pm 1$ or $j_0' = j_0$.

[1]I. M. Gel'fand and V. A. Ponomarev, Usp. Mat. Nauk. **23**, 3 (1968).
[2]A. O. Barut and R. Rączca, *Theory of Group Representations and Applications* (PWN Polish Scientific, Warsaw, 1977).
[3]R. Rączca, Ann. Inst. H. Poincaré **19**, 341 (1973).
[4]P. A. M. Dirac, Int. J. Theor. Phys. **23**, 677 (1984).
[5]A. Guichardet, Astérisque **124-125**, 213 (1985).
[6]M. A. Naimark, *Les Représentations Linéaires du Groupe de Lorentz* (Dunod, Paris, 1962).
[7]V. Cantoni, Rend. Circ. Mat. Palermo **24**, 35 (1975).
[8]H. Joos, Fortschr. Phys. **10**, 65 (1962).
[9]V. Cantoni, Ann. Mat. Pura Appl. **89**, 365 (1971).
[10]V. Cantoni, Ann. Mat. Pura Appl. **104**, 327 (1975).

Collectivity and geometry. VI. Spectra and shapes in the three-dimensional case

E. Chacón
Instituto de Física, UNAM, Apdo. Postal 20-364, México, D.F. 01000, México
P. Hess[a)]
Instituto de Ciencias Nucleares, UNAM, Apdo. Postal 70-543, México, D.F. 04510, México
M. Moshinsky[b)]
Instituto de Física, UNAM, Apdo. Postal 20-364, México, D.F. 01000, México

(Received 6 October 1987; accepted for publication 7 December 1988)

In this paper the reports on collectivity and geometry are concluded, where the microscopic description of many-body collective motions and their relation with the symplectic geometry of the n-particle system are reexamined. In the present paper it is shown that, modulo linear canonical transformations, the symplectic algebra sp$(6,R)$ admits only three maximal subalgebras sp$(2,R) \oplus$ o(3), u(3), and cm(3), which contain the rotation algebra o(3). The objective is to discuss the spectra and shapes of "pure" many-body systems for which the Hamiltonian is associated with a Casimir operator up to the second degree in the generators of a given maximal subalgebra, as well as those of "transitional" systems, where the Hamiltonian is a function of the generators and Casimir operators of several of the maximal subalgebras.

I. INTRODUCTION AND SUMMARY

In this paper we intend to conclude our reports on "Collective and Geometry"[1–5] by discussing in a systematic fashion the group theory behind a symplectic model of collective motions of an n-body system in a d-dimensional space, where d is an arbitrary integer. After determining *all* the maximal subalgebras of sp$(2d,R)$ that contain the generators of the orthogonal subalgebra o(d), we shall particularize our discussion to the physical case when $d = 3$. Our objective will be to determine the spectra and shapes of "pure" many-body systems whose states are characterized by irreducible representations (irreps) of the maximal subalgebras, as well as those where the Hamiltonians involve a mixture of Casimir operators of the different maximal subalgebras, i.e., "transitional" systems.

Our analysis will proceed along the following lines. In Sec. II we derive the maximal subalgebras of sp$(2d,R)$ that contain the generators of the orthogonal subalgebra o(d), modulo linear canonical transformations in the phase space of the n-body system. The discussion leads only to three maximal subalgebras u(d), sp$(2,R) \oplus$ o(d), and cm(d) for $d \geqslant 3$, but allows for three more when $d = 2$.

In Sec. III we discuss the spectra of the Hamiltonians for the "pure" many-body systems associated with the three maximal subalgebras mentioned above. We also introduce those operators whose expectation values can give us the shapes of the eigenstates of these Hamiltonians.

In Sec. IV we introduce the monomial basis state for irreps of sp$(6,R)$ in the positive discrete series[6]: With its help we discuss the spectra and shapes of "transitional" systems.

Finally, in Sec. V we discuss the conclusions that can be derived from the present paper, as well as from the entire whole series of this work.

[a)] Fellow of the "Deutscher Akademischer Austauschdienst" (DAAD).
[b)] Member of El Colegio Nacional.

II. MAXIMAL SUBALGEBRAS OF sp(2*d*,*R*)

We start by considering an A-body system of particles in d-dimensional space. If we eliminate the center of mass motion and designate by x_{is}, p_{is}, $i = 1,2,...d$, $s = 1,2,...,$ $n \equiv A - 1$ the Jacobi relative coordinates and momenta, the generators of sp$(2d,R)$ can be written as

$$M_{ij} = \tfrac{1}{4}(p_{is}p_{js} - x_{is}x_{js}), \tag{2.1a}$$

$$N_{ij} = \tfrac{1}{4}(x_{is}p_{js} + p_{is}x_{js}), \tag{2.1b}$$

$$R_{ij} = \tfrac{1}{4}(p_{is}p_{js} + x_{is}x_{js}), \tag{2.1c}$$

$$L_{ij} = (x_{is}p_{js} - x_{js}p_{is}), \tag{2.1d}$$

where the repeated indices s are summed from 1 to n and all generators are obviously Hermitian.

The M_{ij}, N_{ij}, R_{ij} are symmetric in the indices i, $j = 1,2,...,d$, while L_{ij} is antisymmetric; thus the total number of generators is

$$3(d/2)(d+1) + (d/2)(d-1) = d(2d+1). \tag{2.2}$$

The commutation relations are as follows:

$$\begin{aligned}[M_{ij},M_{i'j'}] &= [N_{ij},N_{i'j'}] \\ &= -[R_{ij},R_{i'j'}] \\ &= (i/16)(L_{j'i}\delta_{i'j} + L_{i'j}\delta_{ij'} \\ &\quad + L_{j'j}\delta_{ii'} + L_{i'i}\delta_{jj'}),\end{aligned} \tag{2.3a}$$

$$\begin{aligned}&[M_{ij},N_{i'j'}] \\ &\quad = (-i/4)(R_{ij'}\delta_{ji'} + R_{ji'}\delta_{ij'} + R_{ii'}\delta_{jj'} + R_{jj'}\delta_{ii'}),\end{aligned} \tag{2.3b}$$

$$\begin{aligned}&[R_{ij},M_{i'j'}] \\ &\quad = (i/4)(N_{ij'}\delta_{ji'} + N_{ji'}\delta_{ij'} + N_{ii'}\delta_{jj'} + N_{jj'}\delta_{ii'}),\end{aligned} \tag{2.3c}$$

$$\begin{aligned}&[N_{ij},R_{i'j'}] \\ &\quad = (i/4)(M_{ij'}\delta_{ji'} + M_{ji'}\delta_{ij'} + M_{ii'}\delta_{jj'} + M_{jj'}\delta_{ii'}),\end{aligned} \tag{2.3d}$$

 0022-2488/89/050970-11$02.50

$$\left[L_{ij},\begin{pmatrix}M_{i'j'}\\N_{i'j'}\\R_{i'j'}\end{pmatrix}\right]$$
$$= -i\left\{\begin{pmatrix}M_{ij'}\\N_{ij'}\\R_{ij'}\end{pmatrix}\delta_{ji'} + \begin{pmatrix}M_{ii'}\\N_{ii'}\\R_{ii'}\end{pmatrix}\delta_{jj'} - \begin{pmatrix}M_{ji'}\\N_{ji'}\\R_{ji'}\end{pmatrix}\delta_{ij'} - \begin{pmatrix}M_{jj'}\\N_{jj'}\\R_{jj'}\end{pmatrix}\delta_{ii'}\right\}, \tag{2.3e}$$

$$[L_{ij},L_{i'j'}] = -i(L_{ij'}\delta_{i'j} + L_{ji'}\delta_{ij'} + L_{j'j}\delta_{ii'} + L_{i'i}\delta_{jj'}). \tag{2.3f}$$

The L_{ij} are the generators of the orthogonal subalgebra $o(d)$ of $sp(2d,R)$. As noted by van der Jeugt and Meyer,[7] an important step in obtaining the maximal subalgebras of $sp(2d,R)$ is to decompose its generators into their irreducible parts with respect to the $o(d)$ subalgebra.

Just as x_{is}, p_{is} corresponds[8] to the irrep $[1]$ of $o(d)$, the M_{ij}, N_{ij}, R_{ij}, which are symmetric in the indices, i, j, must also correspond[8] to the irrep $[2]$ or $[0]$ of $o(d)$, while L_{ij}, which is antisymmetric in i, j, corresponds to irrep $[1^2]$ of $o(d)$. Obviously[8] the irrep $[0]$ is associated with the scalars

$$R \equiv R_{ii}, \tag{2.4a}$$

$$M \equiv M_{ii}, \tag{2.4b}$$

$$N \equiv N_{ii}, \tag{2.4c}$$

where repeated indices are summed from $i = 1$ to d, while the irrep $[2]$ is associated with the traceless tensors

$$\mathbf{R}_{ij} \equiv R_{ij} - (R/d)\delta_{ij}, \tag{2.5a}$$

$$\mathbf{M}_{ij} \equiv M_{ij} - (M/d)\delta_{ij}, \tag{2.5b}$$

$$\mathbf{N}_{ij} \equiv N_{ij} - (N/d)\delta_{ij}. \tag{2.5c}$$

The set of $d(2d+1)$ generators (2.1) can be written inside angular brackets as

$$\langle L_{ij},M_{ij},N_{ij},R_{ij}\rangle, \quad i,j = 1,2,...,d, \tag{2.6}$$

but, as suggested by van der Jeugt and Meyer,[7] they could also be expressed in terms of independent linear combinations of the scalars M, N, R and of the traceless tensors $\mathbf{M}_{ij}$, $\mathbf{N}_{ij}$, $\mathbf{R}_{ij}$ as

$$\left\langle L_{ij}; \begin{matrix}\alpha\mathbf{M}_{ij} + \beta\mathbf{N}_{ij} + \gamma\mathbf{R}_{ij}\\ aM + bN + cR\end{matrix}; \begin{matrix}\alpha'\mathbf{M}_{ij} + \beta'\mathbf{N}_{ij} + \gamma'\mathbf{R}_{ij}\\ a'M + b'N + c'R\end{matrix}; \begin{matrix}\alpha''\mathbf{M}_{ij} + \beta''\mathbf{N}_{ij} + \gamma''\mathbf{R}_{ij}\\ a''M + b''N + c''R\end{matrix}\right\rangle. \tag{2.7}$$

This form of writing the generators of $sp(2d,R)$ has the following advantages.

(i) Once we have one of the terms of the upper row of (2.7) such as $\alpha\mathbf{M}_{ij} + \beta\mathbf{N}_{ij} + \gamma\mathbf{R}_{ij}$ for a given i,j we can, by commuting the term with L_{ij} as in (2.3e) obtain this linear combination for all other i, j indices. Thus from the beginning we write the linear combination for all possible i,j in the traceless tensors of the upper row.

(ii) The lower row of (2.7) is formed by linear combinations of the scalars R, M, N with respect to $o(d)$ Lie algebra and thus they commute with L_{ij}. Furthermore, these linear combinations close under commutation since from (2.3b)–(2.3d), we obtain

$$[M,N] = -iR, \quad [R,M] = iN, \quad [N,R] = iM. \tag{2.8}$$

Thus the linear combinations are the generators of an $sp(2,R)$ algebra.[9]

(iii) The commutators of the terms in the lower row of (2.7) with those in the upper row correspond to the irrep $[2]$ of $o(d)$ and thus necessarily give combinations of the traceless tensors $\mathbf{M}_{ij}$, $\mathbf{N}_{ij}$, $\mathbf{R}_{ij}$, i.e., linear combinations of terms in the upper row.

(iv) The commutators of two terms in the upper row of (2.7) now give linear combinations of the traceless tensors $\mathbf{M}_{ij}$, $\mathbf{N}_{ij}$, $\mathbf{R}_{ij}$ and the scalars M, N, R since from (2.3) we obtain, for example,

$$[\mathbf{R}_{ij},\mathbf{M}_{i'j'}] = i\{\tfrac{1}{4}(\mathbf{N}_{ij'}\delta_{ji'} + \mathbf{N}_{ii'}\delta_{jj'} + \mathbf{N}_{jj'}\delta_{ii'} + \mathbf{N}_{ji'}\delta_{ij'}) - d^{-1}\delta_{ij}\mathbf{N}_{i'j'} - d^{-1}\delta_{i'j'}\mathbf{N}_{ij}\} - i(d^{-2}\delta_{ij}\delta_{i'j'} - (2d)^{-1}\delta_{ij'}\delta_{ji'} - (2d)^{-1}\delta_{ii'}\delta_{jj'})N. \tag{2.9}$$

It is important to notice that the term in the curly brackets in (2.9) is different from zero when $d \geqslant 3$, but vanishes for $d = 2$ since then

$$\mathbf{N}_{11} = \tfrac{1}{2}(N_{11} - N_{22}) = -\mathbf{N}_{22}. \tag{2.10}$$

We shall show in the Appendix that this difference in the behavior of (2.9) for $d \geqslant 3$ and $d = 2$ is responsible for the fact that there are only three maximal subalgebras in the former case, while there are six for the latter.

We now apply expression (2.7) for the set of generators of $sp(2d,R)$ having properties (i)–(iv) to the derivation of the maximal subalgebras of $sp(2d,R)$.

We start by suppressing the upper row of generators of $sp(2d,R)$ in (2.7) and asking whether the remaining ones form a maximal subalgebra. We note from property (i) that if we add one of the traceless tensors, say $\mathbf{M}_{ij}$ with fixed i, j, we obtain all the other components by commutation with L_{ij}. Furthermore, from (iii) and (2.3) we note that the commutation of $\mathbf{M}_{ij}$ with R, N would give $\mathbf{N}_{ij}$, $\mathbf{R}_{ij}$; thus we cannot add any generator of the upper row of (2.7) without recovering the full set of generators of $sp(2d,R)$. Thus we conclude that

$$\langle L_{ij},R,M,N\rangle \tag{2.11}$$

is already a maximal subalgebra; from (2.3f) and (2.8) it actually corresponds to the subalgebra

$$sp(2,R) \oplus o(d). \tag{2.12}$$

From the discussion of the previous paragraph we see that we do not obtain a maximal subalgebra if we suppress only one of the terms of the upper row of (2.7), so the next possibility is to suppress one term from the upper and one from the lower row. Without loss of generality we can then suppress the last column in (2.7) and ask ourselves whether the remaining terms, i.e.,

$$\left\langle L_{ij}; \begin{matrix} \alpha\mathbf{M}_{ij}+\beta\mathbf{N}_{ij}+\gamma\mathbf{R}_{ij} \\ aM+bN+cR \end{matrix}; \begin{matrix} \alpha'\mathbf{M}_{ij}+\beta'\mathbf{N}_{ij}+\gamma'\mathbf{R}_{ij} \\ a'M+b'N+c'R \end{matrix} \right\rangle \tag{2.13}$$

can give us a maximal subalgebra. For this to be the case the generators in (2.13) must close under commutation and, since we have 12 real parameters in a, b, c, etc. this is likely to occur in an infinite number of ways. We must then distinguish between different types of maximal subalgebras rather than the maximal subalgebras themselves.

Note that the type of maximal subalgebra will not change if we carry out a linear canonical transformation of the x_{is}, p_{is} to x'_{is}, p'_{is}, which implies a transformation of the bilinear expressions $L_{ij}, M_{ij}, N_{ij}, R_{ij}$ in x_{is}, p_{is} given in (2.1) to $L'_{ij}, M'_{ij}, N'_{ij}, R'_{ij}$ in x'_{is}, p'_{is}. Since the commutation rules (2.3) follow from

$$[x_{is}, p_{jt}] = i\delta_{ij}\delta_{st} \tag{2.14}$$

we have the same rules for the primed generators since a canonical transformation maintains the commutation relation (2.14) for x'_{is}, p'_{jt}. Thus the structure constants in the subalgebras remain the same in the primed and unprimed picture and the statement at the beginning of this paragraph is justified.

From the analysis of the previous paragraph we conclude that we must only discuss maximal subalgebras "modulo" linear canonical transformations: Since the ones related with rotations in d-dimensional space are irrelevant for our analysis we shall concentrate on the simple[10]

$$\begin{pmatrix} x'_{is} \\ p'_{is} \end{pmatrix} = \begin{pmatrix} \lambda & \mu \\ \nu & \tau \end{pmatrix} \begin{pmatrix} x_{is} \\ p_{is} \end{pmatrix}, \quad \lambda\tau - \mu\nu \doteq 1, \tag{2.15}$$

where λ, μ, ν, τ are real numbers independent of i, s.

We can write the most general real matrix of determinant 1 in the form

$$\begin{pmatrix} \lambda & \mu \\ \nu & \tau \end{pmatrix} = \begin{pmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{pmatrix} \times \begin{pmatrix} e^{\theta} & 0 \\ 0 & e^{-\theta} \end{pmatrix} \begin{pmatrix} \cos\chi & -\sin\chi \\ \sin\chi & \cos\chi \end{pmatrix}. \tag{2.16}$$

Now, taking into account the definitions (2.1) of M_{ij}, N_{ij}, R_{ij} in terms of x_{is}, p_{is} and the corresponding definition for $M'_{ij}, N'_{ij}, R'_{ij}$ in terms of x'_{is}, p'_{is}, we see that when the phase space variables are related by a rotation in φ we have

$$\begin{pmatrix} M'_{ij} \\ N'_{ij} \\ R'_{ij} \end{pmatrix} = \begin{pmatrix} \cos 2\varphi & \sin 2\varphi & 0 \\ -\sin 2\varphi & \cos 2\varphi & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} M_{ij} \\ N_{ij} \\ R_{ij} \end{pmatrix}; \tag{2.17a}$$

we have a similar result when the rotation is in χ. On the other hand, for the dilation in θ we obtain

$$\begin{pmatrix} M'_{ij} \\ N'_{ij} \\ R'_{ij} \end{pmatrix} = \begin{pmatrix} \cosh 2\theta & 0 & -\sinh 2\theta \\ 0 & 1 & 0 \\ -\sinh 2\theta & 0 & \cosh 2\theta \end{pmatrix} \begin{pmatrix} M_{ij} \\ N_{ij} \\ R_{ij} \end{pmatrix}. \tag{2.17b}$$

We now return to the set of generators in (2.7) to try to see whether they close under commutation "modulo" a convenient canonical transformation. For this purpose we first will express the linear combination

$$\alpha\mathbf{M}_{ij} + \beta\mathbf{N}_{ij} + \gamma\mathbf{R}_{ij} \tag{2.18}$$

in a canonical form.

If the real coefficients α, β, γ satisfy

$$(\gamma^2 - \beta^2 - \alpha^2) > 0 \tag{2.19a}$$

we can carry a canonical transformation by rotation of an angle φ to make the new $\boldsymbol{\beta} = 0$ and replace γ, α by $\boldsymbol{\gamma} = \delta\cosh\theta$, $\boldsymbol{\alpha} = \delta\sinh\theta$, which clearly satisfy (2.19a) as $\delta^2 > 0$. We then see that the transformation (2.17b) leads to the canonical form $\mathbf{R}_{ij}$ for (2.18) when (2.19a) holds.

If the coefficients satisfy

$$(\gamma^2 - \beta^2 - \alpha^2) < 0 \tag{2.19b}$$

we can, as before, carry a canonical transformation by rotation of an angle φ to make the new $\boldsymbol{\beta} = 0$ and now replace γ, α by $\boldsymbol{\gamma} = \delta\sinh\theta$, $\boldsymbol{\alpha} = \delta\cosh\theta$, which clearly satisfy (2.19b) as $-\delta^2 < 0$. We then see that the transformation (2.17b) leads to the form $\mathbf{M}_{ij}$. Now, using a rotation by an angle $\pi/4$ in χ space we see, from an expression similar to (2.17a), that $\mathbf{M}_{ij}$ can be transformed to $\mathbf{N}_{ij}$. Thus in this case we have the canonical form $\mathbf{N}_{ij}$ for (2.18) when (2.19b) holds.

Finally, if

$$(\gamma^2 - \beta^2 - \alpha^2) = 0 \tag{2.19c}$$

we again use the rotation in φ to make $\boldsymbol{\beta} = 0$ and

$$\boldsymbol{\gamma}^2 - \boldsymbol{\alpha}^2 = 0 \quad \text{or} \quad \boldsymbol{\gamma} = \pm\boldsymbol{\alpha}, \tag{2.20}$$

which leads to the two canonical forms

$$\mathbf{R}_{ij} \pm \mathbf{M}_{ij} = \begin{cases} \frac{1}{2}[p_{is}p_{js} - d^{-1}(p_{ks}p_{ks})\delta_{ij}] \\ \frac{1}{2}[x_{is}x_{js} - d^{-1}(x_{ks}x_{ks})\delta_{ij}] \end{cases}. \tag{2.21}$$

The forms (2.21) are equivalent since the canonical transformation $x'_{is} = p_{is}, p'_{is} = -x_{is}$ takes one into the other. We then retain $\mathbf{R}_{ij} - \mathbf{M}_{ij}$ as the canonical form of (2.18) when (2.19c) holds.

In studying the types of maximal subalgebras we can then restrict the first linear combination of the generators appearing in the upper row of (2.7) to one of three canonical forms depending on whether $(\gamma^2 - \beta^2 - \alpha^2) \gtreqless 0$, i.e.,

$$\alpha\mathbf{M}_{ij} + \beta\mathbf{N}_{ij} + \gamma\mathbf{R}_{ij} \to \begin{cases} \mathbf{R}_{ij}, & (\gamma^2 - \beta^2 - \alpha^2) > 0, \\ \mathbf{N}_{ij}, & (\gamma^2 - \beta^2 - \alpha^2) < 0, \\ \mathbf{R}_{ij} - \mathbf{M}_{ij}, & (\gamma^2 - \beta^2 - \alpha^2) = 0. \end{cases} \tag{2.22}$$

For example, if in (2.22) we select the first term, i.e., $\mathbf{R}_{ij}$, then the coefficients $a, b, c; \alpha', \beta', \gamma'; a', b', c'$ in the remaining terms in (2.7) must be chosen in such a way that the commutator of any pair of generators is a linear combination of them. This gives a number of relations between the coefficients mentioned and, as shown in the Appendix, we arrive, up to multiplicative constants, at a maximal subalgebra whose generators are

$$\langle L_{ij}, \mathbf{R}_{ij}, R \rangle, \quad \mathrm{u}(d). \tag{2.23}$$

We shall later identify this subalgebra with the d-dimensional unitary one, i.e., $\mathrm{u}(d)$, as indicated in (2.23).

If we begin with the second term in (2.22), i.e., $\mathbf{N}_{ij}$, we arrive, via the same analysis (also given in the Appendix), at a maximal subalgebra whose generators are

$$\langle L_{ij},\mathbf{N}_{ij},\mathbf{R}_{ij}-\mathbf{M}_{ij},N,R-M\rangle;\quad \mathrm{cm}(d). \tag{2.24}$$

We shall later identify this subalgebra with the d-dimensional collective motion one, i.e., cm(d), as indicated in (2.24).

If we begin with the last term in (2.22), i.e., $\mathbf{R}_{ij}-\mathbf{M}_{ij}$, the analysis in the Appendix leads us again to (2.24); therefore, we do not obtain a new maximal subalgebra.

All of the above results hold when $d\geqslant 3$. If $d=2$, the considerations mentioned in (iv) following Eq. (2.9) allow other maximal subalgebras besides (2.11), (2.23), and (2.24). Again, using the results of the Appendix, besides (2.12), (2.23), and (2.24) when $d=2$, we obtain the following maximal subalgebras:

$$\langle L_{12},\mathbf{N}_{ij},\mathbf{M}_{ij},R\rangle,\quad \mathrm{sp}'(2,R)\oplus\mathrm{sp}''(2,R), \tag{2.25a}$$

$$\langle L_{12},\mathbf{R}_{ij},\mathbf{M}_{ij},N\rangle,\quad \mathrm{o}(3,1), \tag{2.25b}$$

$$\langle L_{12},\mathbf{R}_{ij}-\mathbf{M}_{ij},(\mathbf{R}_{ij}+\mathbf{M}_{ij})+\epsilon\mathbf{N}_{ij},(R-M)+(2/\epsilon)N\rangle. \tag{2.25c}$$

Again, we shall later identify the subalgebras (2.25a) and (2.25b) with sp$'(2,R)\oplus$sp$''(2,R)$ and o$(3,1)$, respectively. The maximal subalgebra (2.25c) is a new one, in which we have the arbitrary real parameter ϵ.

Now, having obtained the generators of the maximal subalgebras, we proceed to justify their names.

We have shown that from (2.3f) and (2.8) the maximal subalgebra whose generators are given by (2.11) corresponds to sp$(2,R)\oplus$o(d).

The generators (2.23) can be written as

$$\langle L_{ij},R_{ij}\rangle, \tag{2.26}$$

where the L_{ij}, R_{ij} are given by (2.1c) and (2.1d); in the case $d=3$ these are precisely the generators that Elliott[11] associates with u(3). For arbitrary d these generators will be the generators of u(d), as can also be checked from some of the commutation relations in (2.3a), (2.3e), and (2.3f).

It is equally easy to see that the generators of (2.24), which we can write as

$$\langle L_{ij},N_{ij},R_{ij}-M_{ij}\rangle, \tag{2.27}$$

correspond to the cm(d) Lie algebra since the L_{ij}, N_{ij} defined by (2.1b) and (2.1d) are the generators of the general linear algebra in d dimensions, while from (2.1a) and (2.1c) we have

$$R_{ij}-M_{ij}=\tfrac{1}{2}x_{is}x_{js}, \tag{2.28}$$

which is an Abelian subalgebra associated with the mass quadrupole of the many-body system. Together, (2.27) and (2.28) correspond to the definition of collective motion Lie algebra as given by Rosensteel and Rowe[12] and Weaver *et al.*[12] for the case $d=3$.

To identify the additional maximal subalgebras (2.25a) and (2.25b) that appear in the case $d=2$, we first introduce the creation and annihilation operators

$$\eta_{is}=(1/\sqrt{2})(x_{is}-ip_{is}),\quad \xi_{is}=(1/\sqrt{2})(x_{is}+ip_{is}). \tag{2.29}$$

Inverting relations (2.29), we can express the generators appearing in (2.25a) and (2.25b) in terms of η_{is} and ξ_{is}. We then compare the generators with those of Eq. (4.7) of Ref. 5 and conclude that, as already indicated in (2.25a) and (2.25b), they correspond to the maximal subalgebras

$$\mathrm{sp}'(2,R)\oplus\mathrm{sp}''(2,R),\quad \mathrm{o}(3,1). \tag{2.30}$$

We now proceed to discuss the spectra of the Hamiltonians associated with the maximal subalgebras for $d=3$ and the shape of the corresponding eigenstates.

III. SPECTRA AND SHAPES OF "PURE" SYSTEMS

We could say that the analysis of Sec. II indicates the existence of only three "pure" cases in the symplectic model—sp$(2,R)\oplus$O(3), u(3), cm(3)—as shown by the vertices in the triangle in Fig. 1, and of "transitional" cases corresponding to any point in the perimeter of or inside the triangle (except for the vertices).

In this section we discuss the spectra of Hamiltonians associated with the three pure cases and indicate how the shapes of the corresponding eigenstates can be determined. In Sec. IV we sketch a similar analysis for the transitional cases.

A. The case of sp(2,*R*)⊕o(3)

In our discussions we deal with a definite irrep of sp$(6,R)$ in the positive discrete series given by a partition involving three non-negative integers. Following the notation of Ref. 6 we denote this irrep by

$$[h_{13}\,h_{23}\,h_{33}], \tag{3.1}$$

where the index 3 is convenient when discussing the monomial basis[6] used in Sec. IV.

The generators of the maximal subalgebra sp$(2,R)\oplus$o(3) are the L_{ij} of (2.1d) and M, N, R of (2.4); from (2.3f) and (2.8) the Casimir operators are

$$L^2=\tfrac{1}{2}L_{ij}L_{ij}, \tag{3.2a}$$

$$T^2=R^2-M^2-N^2=R(R-1)-(M+iN)(M-iN). \tag{3.2b}$$

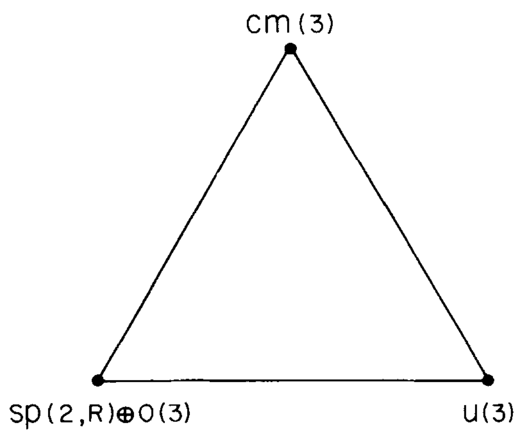

FIG. 1. Diagrammatic representations of the pure cases in the symplectic model represented by the closed circles in the vertices of a triangle indicating the maximal subalgebras sp$(2,R)\oplus$o(3), u(3), and cm(3). The transitional cases correspond to all other points in the perimeter or inside the triangle.

Of course, L^2 is the total angular momentum with the eigenvalues $l(l+1)$ with the integer l, while L_{12} is its projection along direction 3 with the eigenvalue $m = l, l-1, \ldots, -l$.

We denote by r the eigenvalue of the operator

$$R = \tfrac{1}{4}(p_{is}p_{is} + x_{is}x_{is}), \tag{3.3}$$

$2R$ is the Hamiltonian of the harmonic oscillator, and $4r$ is a non-negative integer that gives the irrep of the subalgebra $\mathscr{o}(2)$ of sp$(2,R)$. Finally, we note from (3.2b) that $M - iN$ is the lowering operator[8] of sp$(2,R)$; thus a state of lowest weight in this algebra vanishes when we apply $M - iN$ on it. If in this case we denote the eigenvalue of R by t instead of by r, we see from (3.2b) that the eigenvalue of T^2 becomes $t(t-1)$, while the possible values of r in that irrep of sp$(2,R)$ are

$$r = t, \quad t+1, \quad t+2, \ldots . \tag{3.4}$$

We can now express the basis for the irrep in the positive discrete series of the chain sp$(6,R) \supset$ sp$(2,R) \oplus$ o(3) as the ket

$$|[h_{13}h_{23}h_{33}]\alpha\; trlm\rangle, \tag{3.5}$$

where $[h_{13}h_{23}h_{33}]$, t, r, l, m are, respectively, the irreps of sp$(6,R)$, sp$(2,R)$, $\mathscr{o}(2)$, o(3), o(2) and α is the multiplicity index which distinguishes between repeated irreps of sp$(2,R) \oplus$ o(3) contained in a given irrep $[h_{13}h_{23}h_{33}]$ of sp$(6,R)$. An explicit procedure for constructing states of type (3.5) is given in Eq. (7.2) of Ref. 3, with which one can also determine the values of l compatible with $[h_{13}h_{23}h_{33}]$ and t.

Since $2r$ is the eigenvalue of the oscillator Hamiltonian $2R$, the discussion of Ref. 6, sketched also in Sec. IV, indicates that the minimum value of $2r$ is $h_{13} + h_{23} + h_{33}$. In this case we could write

$$t = \Lambda + \sigma, \quad \sigma = \tfrac{1}{2}(h_{13} + h_{23} + h_{33}), \quad \Lambda = 0,1,2,\ldots \tag{3.6}$$

and the eigenvalue E_Λ of T^2 takes the form

$$E_\Lambda = t(t-1) = \sigma(\sigma - 1) + (2\sigma - 1)\Lambda + \Lambda^2. \tag{3.7}$$

For a harmonic oscillator shell model[13] for particles satisfying Fermi statistics the σ is large even for a small number of particles, e.g., for $A = 16$, $\sigma = 12$ and for $A = 20$, $\sigma = 20$. Thus the variation of energy with Λ is given mainly by the linear term $(2\sigma - 1)\Lambda$, i.e., we have a *vibrational spectrum*.

B. The case u(3)$\supset$o(3)

The generators in the case u(3) $\supset$ o(3) are the R_{ij}, L_{ij}, of (2.1c) and (2.1d), where the latter are those of o(3). As indicated in (3.2a), the Casimir operator of o(3) is $L^2 = \tfrac{1}{2}L_{ij}L_{ij}$, while from (2.1) and (2.29) the linear and quadratic Casimir operators of u(3) are given by[14]

$$H_0 = C_{ii} = 2R, \quad \Gamma = C_{ij}C_{ji}, \tag{3.8}$$

where

$$C_{ij} = \tfrac{1}{2}(\eta_{is}\xi_{js} + \xi_{js}\eta_{is}). \tag{3.9}$$

The quadrupole–quadrupole interaction is defined by the operator[14]

$$Q^2 \equiv -4\mathbf{R}_{ij}\mathbf{R}_{ji} = -\Gamma + \tfrac{1}{3}H_0^2 + \tfrac{1}{2}L^2, \tag{3.10}$$

where the rhs is obtained with the help of (2.29), (3.8), and (3.9).

To determine the eigenvalues of Q^2 we first note that the basis for the irreps in the positive discrete series of the chain sp$(6,R) \supset$ u(3) $\supset$ o(3) can be expressed by the ket

$$|[h_{13} \quad h_{23} \quad h_{33}]\beta(k_1 \quad k_2 \quad k_3)\Omega lm\rangle, \tag{3.11}$$

where $[h_{13}\, h_{23}\, h_{33}]$, (k_1, k_2, k_3), l, m are, respectively, the irreps of sp$(6,R)$, u(3), o(3), o(2). Here β and Ω are multiplicity indices which distinguish between repeated irreps of u(3) in a given irrep of sp$(6,R)$ and between repeated irreps of o(3) in a given irrep of u(3). These states have been discussed by Rowe and Rosensteel.[12]

The operators H_0, Γ, L^2 are then clearly diagonal in the basis of states (3.11) and thus their eigenvalues are well known.[11,14] The spectrum of energy levels associated with the quadrupole–quadrupole interaction (3.10), first obtained by Elliott,[11] then takes the form[11,14]

$$E_{k_1k_2k_3,l} = -[k_1^2 + k_2^2 + k_3^2 + 2(k_1 - k_3)] + \tfrac{1}{3}(k_1 + k_2 + k_3)^2 + \tfrac{1}{2}l(l+1), \tag{3.12}$$

giving rise to *rotational bands*, each of which is characterized by the irrep $(k_1\, k_2\, k_3)$ of u(3).

The angular momentum content in each rotational band is related with the irreps l of o(3) in a given irrep $(k_1\, k_2\, k_3)$ of u(3) and is well known.[15] The irreps $(k_1\, k_2\, k_3)$ of u(3), present in a given irrep $[h_{13}\, h_{23}\, h_{33}]$ of sp$(6,R)$ has also been fully investigated.[12,16,17] Thus when $[h_{13}\, h_{23}\, h_{33}]$ is specified, all possible rotational bands $(k_1\, k_2\, k_3)$ and corresponding angular momenta l are known.

C. The case cm(3)

There remains only the case cm(3) to complete our discussion of the maximal subalgebras of sp$(6,R)$, i.e., the cases of "pure" Hamiltonians.

The Casimir operators of cm(3) are of the third and fourth degree in the generators[18] and we could try to discuss their spectra in an analytic fashion, as we have done for sp$(2,R) \oplus$ o(3). Unfortunately, cm(3) does not contain among its operators the $2R$ of (3.17), i.e., the harmonic oscillator Hamiltonian, nor

$$2(R + M) = \tfrac{1}{2}p_{is}p_{is}, \tag{3.13}$$

the kinetic energy of the many-particle system in three dimensions.

Any realistic Hamiltonian involving the Casimir operators of cm(3) must also include either $2R$ or $2(R+M)$; thus it will automatically fall in the category of a "transitional" Hamiltonian, for which only a numerical analysis is feasible. We shall discuss this type of computation in Sec. IV.

D. Shape operators

We turn now our attention to the problem of shape. The eigenvalues of the shape operators are the principal values[19] of the quadrupole matrix

$$\mathsf{q} = \|q_{ij}\| = \|x_{is}x_{js}\| = 2\|R_{ij} - M_{ij}\|, \tag{3.14}$$

i.e., its diagonal components in the frame of reference fixed in the body, given by the secular equation

$$\det\|\lambda\delta_{ij} - q_{ij}\| = 0, \quad i,j = 1,2,3. \tag{3.15}$$

Equation (3.15) leads to a cubic equation in λ, which gives rise to three real and positive eigenvalues designated $\lambda_k = \rho_k^2$, $k = 1,2,3$. The roots ρ_k are then related to the three principal axes of an ellipsoid[19] and thus give a measure of the shape.

Since rather than shape we are concerned with the deformation of the many-body system away from sphericity, it is convenient to express ρ_k^2, $k = 1,2,3$ in terms of three new parameters ρ, b, c through the relation

$$\rho_k^2 = (\rho^2/3)\{1 + 2b\cos[c - (2\pi k/3)]\}, \quad k = 1,2,3, \tag{3.16}$$

where now b^2, $b^3\cos 3c$ provide a measure of deviation from the spherical shape, as for $b = 0$, $\rho_k^2 = (\rho^2/3)$, $k = 1,2,3$.

Through the standard method of solving cubic equations with the help of trigonometric functions,[20] we immediately find that

$$\rho^2 = \mathrm{tr}\,\mathsf{q}, \tag{3.17a}$$

$$b^2 = \tfrac{3}{2}[(\mathrm{tr}\,\mathbf{q}^2)/(\mathrm{tr}\,\mathsf{q})^2], \tag{3.17b}$$

$$b^3\cos 3c = \tfrac{27}{2}[\det\mathbf{q}/(\mathrm{tr}\,\mathsf{q})^3], \tag{3.17c}$$

where $\mathbf{q}$ is the traceless part of the matrix q, i.e.,

$$\mathbf{q} = \|q_{ij} - \tfrac{1}{3}\,\mathrm{tr}\,\mathsf{q}\,\delta_{ij}\|. \tag{3.18}$$

A measure of the deformation of an eigenstate of a Hamiltonian is given by the expectation values of the operators b^2, $b^3\cos 3c$ with respect to these eigenstates. Since powers of $\rho^2 = \mathrm{tr}\,\mathsf{q}$ appears in the denominator, they are cumbersome to evaluate; thus we prefer to define deformation as the following ratio of expectation values:

$$\langle b^2\rangle = \tfrac{3}{2}[\langle\mathrm{tr}\,\mathbf{q}^2\rangle/\langle(\mathrm{tr}\,\mathsf{q})^2\rangle], \tag{3.19a}$$

$$\langle b^3\cos 3c\rangle = \tfrac{27}{2}[\langle\det\mathbf{q}^2\rangle/\langle(\mathrm{tr}\,\mathsf{q})^3\rangle], \tag{3.19b}$$

where the angular brackets represent expectation values of the operators indicated with respect to the eigenstates of the Hamiltonians under study.

In Sec. IV we shall evaluate the deformation parameters (3.19) for some specific many-body systems.

IV. SPECTRA AND SHAPES OF "TRANSITIONAL" SYSTEMS

In Sec. III we discussed the spectra of Hamiltonians associated with the three maximal subalgebras and the operator whose expectation value with respect to the corresponding eigenstates gives their shape. We now want to see how the spectra and shapes change when we go in a continuous fashion from one to another of the maximal subalgebras, i.e., the case of "transitional" systems.

We will be concerned with a set of A particles, but, as mentioned at the beginning of Sec. II, we eliminate the center of mass motion. Thus our x_{is}, p_{is}, $i = 1,2,3$; $s = 1,2,\ldots,n = A - 1$ are Jacobi relative coordinates and momenta. Our many-body systems will be characterized by $(A, [h_{13}\, h_{23}\, h_{33}])$, where the square brackets correspond to the irrep of $\mathrm{sp}(6,R)$.

By definition, a "transitional" Hamiltonian cannot be diagonal in any of the bases associated with a maximal subalgebra: Thus we would have to determine its matrix representation with respect to one definite basis and diagonalize this matrix numerically. Unfortunately, neither (3.5) nor (3.11), associated with $\mathrm{sp}(2,R)\oplus \mathrm{o}(3)$ or $\mathrm{u}(3)$, provides an orthonormal basis since they still have multiplicity indices. Furthermore, the matrix representation of the generators with respect to these bases is rather complex and in fact has to be determined through computer programs. Thus in the following analysis we decided to use what we call the "monomial basis" of Ref. 6 and whose main characteristics we briefly review.

A. Monomial basis states for irreps of sp(6,*R*) in the positive discrete series and matrix representation with respect to them of generators, Hamiltonians, and shape operators

In terms of the creation and annihilation operators (2.29) the realization of the generators of $\mathrm{sp}(6,R)$ is[19]

$$B_{ij}^{\dagger} = \eta_{is}\eta_{js}, \tag{4.1a}$$

$$C_{ij} = \tfrac{1}{2}(\eta_{is}\xi_{js} + \xi_{js}\eta_{is}), \tag{4.1b}$$

$$B_{ij} = \xi_{is}\xi_{js}, \tag{4.1c}$$

where, as before, the repeated indices s are summed from 1–n and i, j over the values 1, 2, 3. The C_{ij} are the generators of the $\mathrm{u}(3)$ subalgebra of $\mathrm{sp}(6,R)$.

The full monomial basis associated with the irrep $[h_{13}\, h_{23}\, h_{33}]$ in the positive discrete of $\mathrm{sp}(6,R)$ has the form

$$|n_{ij},h_{ij}\rangle = \prod_{i\leqslant j=1}^{3}(B_{ij}^{\dagger})^{n_{ij}}|h_{ij}\rangle, \tag{4.2}$$

where the n_{ij} are non-negative integers and

$$|h_{ij}\rangle = \left|\begin{matrix} h_{13} & & h_{23} & & h_{33} \\ & h_{12} & & h_{22} & \\ & & h_{11} & & \end{matrix}\right\rangle \tag{4.3}$$

is a Gelfand state,[14,21] with the labels satisfying the inequalities $h_{ij}\geqslant h_{i,j-1}\geqslant h_{i+1,j}$.

While the basis (4.2) has a very simple analytic form, it is clearly nonorthonormal; besides, it does not correspond to an eigenstate of the total angular momentum $L^2 = \tfrac{1}{2}L_{ij}L_{ij}$, where $L_{ij} = C_{ij} - C_{ji}$. However, the basis (4.2) can be made to have a definite projection in direction 3 of the angular momentum, i.e., to be an eigenstate of L_{12}. Furthermore, $|n_{ij},h_{ij}\rangle$ is also an eigenstate of the operator $2R = C_{ii}$, with the eigenvalue

$$2\sum_{i\leqslant j=1}^{3} n_{ij} + (h_{13}+h_{23}+h_{33}) \equiv 2(\mathcal{N}+\sigma), \tag{4.4}$$

where σ is given by (3.6) and, for brevity, we will refer to $\mathcal{N} = \Sigma_{i\leqslant j=1}^{3} n_{ij}$ as the number of quanta.

An arbitrary generator X of the $\mathrm{sp}(6,R)$ Lie algebra, when acting on the state (4.2), gives a linear combination of these states, i.e.,

$$X|n_{ij},h_{ij}\rangle = \sum_{n'_{ij},h'_{ij}} |n'_{ij},h'_{ij}\rangle(n'_{ij},h'_{ij}|X|n_{ij},h_{ij}\rangle. \tag{4.5}$$

The last term in (4.5) is not a matrix element, but a numerical coefficient, a fact we emphasize by using a round instead of an angular bracket on its lhs. Note that we stay within one representation of sp(6,R), i.e., $h'_{i3} = h_{i3}$, $i = 1, 2, 3$.

All the coefficients $(n'_{ij},h'_{ij}|X|n_{ij},h_{ij})$ that correspond to the set of generators $\{X\}$ of sp(6,R) in (4.1) were given in explicit and closed form in Ref. 6.

We now turn to Hamiltonians H in the enveloping algebra of sp(6,R), i.e., polynomial functions of the generators (2.1) or (4.1) which are Hermitian and invariant under rotations and time reflections.[5,6] If we apply H to the states (4.2) we again obtain a linear combination of these states, i.e.,

$$H|n_{ij},h_{ij}\rangle = \sum_{n'_{ij},h'_{ij}} |n'_{ij},h'_{ij}\rangle (n'_{ij},h'_{ij}|H|n_{ij},h_{ij}\rangle. \tag{4.6}$$

The coefficients on the rhs of (4.6) can be obtained immediately since we know the corresponding coefficients for $B^\dagger_{kl}, C_{kl}, B_{kl}, k,l = 1, 2, 3$ given in Ref. 6. Although the basis $|n_{ij},h_{ij}\rangle$ is not orthonormal, it was shown in Ref. 6 that eigenvalues of H, i.e., the energy levels E, are given by the secular equation

$$\det\|(n'_{ij},h'_{ij}|H|n_{ij},h_{ij}\rangle - E\prod_{i\leqslant j=1}^{3}\delta_{n'_{ij}n_{ij}}\delta_{h'_{ij}h_{ij}}\| = 0. \tag{4.7}$$

The matrix appearing in (4.7) is in general of infinite dimension, in which case we must, for calculation purposes, introduce a cutoff procedure for the n_{ij}, n'_{ij}. One possibility is to require that $\Sigma_{i\leqslant j} n_{ij} \leqslant \mathscr{N}$, where $\mathscr{N}$ is a convenient upper bound for the sum indicated.

There are some cases in which the matrix (4.7) breaks into finite blocks, as happens, for example, for the Hamiltonians associated with the maximal subalgebras sp(2,R) $\oplus$ o(3) and u(3), both of which commute with $2R = C_{ii}$. From (4.4) we then see that we have finite blocks associated with each value of $\mathscr{N}$.

Finally, we consider the shape operators (3.17) associated with the matrix $q_{ij} = x_{is}x_{js}$ which, from (2.29), can be written as

$$q_{ij} = \tfrac{1}{2}(B^\dagger_{ij} + C_{ij} + C_{ji} + B_{ij}). \tag{4.8}$$

The application of the operators in (3.17), i.e., $(\mathrm{tr}\,\mathbf{q})^2$, $(\mathrm{tr}\,\mathbf{q})^3$, $\mathrm{tr}\,\mathbf{q}^2$, $\det\mathbf{q}$, designated generically by $\mathscr{O}$, to the states $|n_{ij},h_{ij}\rangle$ again gives linear combinations of these states, with the coefficients

$$(n'_{ij},h'_{ij}|\mathscr{O}|n_{ij},h_{ij}\rangle, \tag{4.9}$$

which can be determined with the help of the coefficients in (4.5) for the generators of sp(6,R).

With the help of (4.9) it is straightforward to determine the expectation value of the operators $\mathscr{O}$ with respect to the eigenstates of the Hamiltonian H: The latter are linear combinations of the $|n_{ij},h_{ij}\rangle$, with the coefficients determined by the secular equation (4.7).

We have thus outlined the procedure we follow to obtain the spectra and shape for "transitional" cases. We proceed to illustrate the analysis in specific examples. A more complete discussion of this program will be presented elsewhere.

B. Transitional Hamiltonians involving the subalgebras sp(2R) $\oplus$ o(3) and u(3)

We shall consider the Hamiltonian

$$H = (1-x)T^2 + xQ^2, \tag{4.10}$$

where x is a real parameter in the interval $0 \leqslant x \leqslant 1$ and T^2, defined in (3.2b), is the Casimir operator of the sp(2,R) algebra, while Q^2 of (3.10) is the quadrupole–quadrupole interaction associated with su(3).

For $x = 0$ or 1 the spectrum of H for a system $(A,[h_{13}\,h_{23}\,h_{33}])$ is given by (3.7) or (3.12), while for x in the open interval $0 < x < 1$ it has to be evaluated numerically through procedures such as the one outlined in Sec. IV A. These procedures also provide the eigenstates, whose angular momenta can be determined by applying to them the operator L^2.

We note that H of (4.10) commutes with $H_0 \equiv C_{ii} = 2R$ and thus the $\mathscr{N}$ of (4.4) is a good quantum number for any x. For $x = 0$ the energy levels are characterized by the Λ of (3.7); thus we require the values of this quantum number which are compatible with $\mathscr{N}$. As discussed in Refs. 1 and 3 the total number of quanta of the state (4.2) that, as indicated in (4.4), is given by $2(\mathscr{N} + \sigma)$, characterizes the one-row irrep of the Lie algebra u($3n$). On the other hand, the one-row irrep of the subalgebra o($3n$), related to that[1,3] of sp(2,R), is given by the $2(\Lambda + \sigma)$ of (3.6). Because of the well-known[8] relations between the single-row irreps of u($3n$) and o($3n$) we conclude that $\Lambda = \mathscr{N}, \mathscr{N} - 1, \mathscr{N} - 2,\ldots,0$. If $\mathscr{N} \ll \sigma$ we have from E_Λ of (3.7) that the levels of H for $x = 0$, when $\Lambda = 0,1,\ldots,\mathscr{N}$, are almost equally spaced.

We shall discuss only the example (20, [12, 4, 4]) with $\mathscr{N} = 1$, although the conclusions we derive hold, also, for other cases $(A,[h_{13}\,h_{23}\,h_{33}])$ we have analyzed. In Fig. 2 we

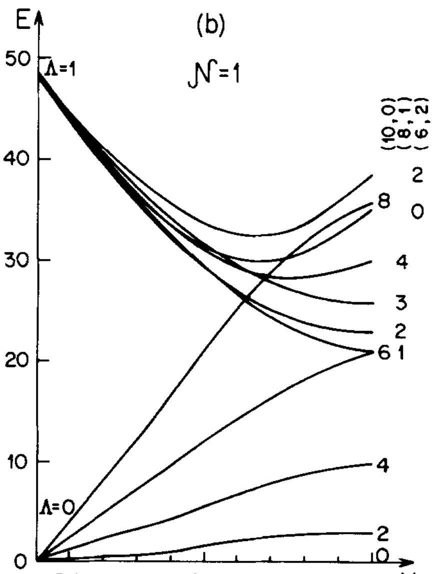

FIG. 2. Spectra of the open shell system (20, [12,4,4]) for the Hamiltonian (4.10) and one quantum of excitation $\mathscr{N} = 1$. The Λ, (λ,μ), and l denote, respectively, the irreps of sp(2,R), su(3), and o(3). The energy is given in dimensionless units. The values at $x = 0$ and $x = 1$ correspond to the vibrational or rotational limits.

show the spectra as a function of x, where we renormalized the ground state energy to zero, i.e., a straight line along the abscissa. At $x = 0$ (the vibrational limit) the values $\Lambda = 0,1$ characterize the irrep of sp$(2,R)$. At $x = 1$ (the rotational limit), $(\lambda\mu)$ in vertical parentheses give the irrep of su(3), where $\lambda = k_1 - k_2, \mu = k_2 - k_3$ for the $(k_1 k_2 k_3)$ appearing in (3.12). In the column below each $(\lambda\mu)$ we give the angular momentum of the state.

Note that the low lying collective states correspond to $\mathcal{N} = 0$, as our example is related more to the giant isoscalar resonances in ^{20}Ne. However, our main point in Fig. 2 is to illustrate the change of spectra when going from the sp$(2,R) \oplus$ o(3) to the u(3) maximal subalgebra.

From the procedure outlined at the end of Sec. IV A we can also calculate the shape of the eigenstates associated with definite eigenvalues of H. For example, for the lowest $l = 0$ state, the expectation value $\langle b^2 \rangle$ of (3.19a) is 0.105 at $x = 0$ and 0.120 at $x = 1$, a change of only about 10%; this holds for all the examples $(A,[h_{13}\ h_{23}\ h_{33}])$ we have discussed when H has the form (4.10).

We thus conclude that while there is a considerable change in the spectra when we go from the vibrational $(x = 0)$ to the rotational $(x = 1)$ limit, the deformation of the eigenstates of H does not change very much. In this respect the conclusions are similar to those obtained in Ref. 5, where the case of two space dimensions was considered.

C. Hamiltonian involving the subalgebras cm(3) and u(1)

The cm(3) algebra provides us with the interacting term

$$\operatorname{tr} \mathbf{q}^2, \tag{4.11}$$

with $\mathbf{q}$ given by (3.18), which corresponds to the full physical quadrupole interaction. The term (4.11) cannot be considered as a Hamiltonian by itself since we must add to it the kinetic energy contained in $2R$. The operator R, defined in (2.1c) and (2.4a), is the generator of a u(1) subalgebra. Together with the powers and products of the generators of cm(3), R indicates that our Hamiltonian is of the transitional type and could be written as[22]

$$H = \hbar\omega(2R) + V_{\text{coll}}, \tag{4.12a}$$

$$V_{\text{coll}} = c_1 \rho^4 b^2 + c_2 \rho^6 b^3 \cos 3c + c_3 \rho^8 b^4. \tag{4.12b}$$

The variables b^2, $b^3 \cos 3c$, and p^2 are related to tr $\mathbf{q}$, tr $\mathbf{q}^2$, and det $\mathbf{q}$ through Eqs. (3.17). The coefficient $\hbar\omega$ is introduced to specify the frequency of the oscillator and c_1, c_2, c_3 are model parameters.

The V_{coll} of (4.12b) is of the type used by Rowe[23] and is related to the potential used in the geometrical model introduced by Hess *et al.*[22] In Ref. 23 and the references given therein many calculations were performed. Therefore, we do not need to repeat these calculations, but rather summarize some of the results that are in structure similar to those obtained in Ref. 5.

The H of (4.12) can be interpreted as a vibrational or rotational Hamiltonian depending on the parameters c_1, c_2, c_3. If, for example, $c_2 = c_3 = 0$, V_{coll} as a function of b^2 is a parabola with a minimum at $b = 0$. The spectra of H will then be vibrational, with small deformations for the eigenstates.[5,23] On the other hand, if $c_2, c_3 \neq 0$, we can have a situation in which V_{coll} as a function of b^2 has a minimum quite far away from $b = 0$, in which case the H of (4.12) has a rotational spectrum with very strong deformations for the eigenstates.[5,23]

The statements of the previous paragraph are corroborated by calculations using the techniques of Sec. IV A or other methods.[12] Furthermore, the statements also agree with our results for the corresponding problem in two-dimensional space.[5]

In Sec. V we present the conclusions that follow from the present series of papers.

V. CONCLUSION

As mentioned in the Introduction to this series of papers,[1] we wished to discuss the relation between collective behavior in many-body systems and geometrical concepts. Specifically, we intended to reexamine the microscopic description of nuclear collective motions and their relation with the symplectic geometry of the A-nucleon system.

We were certainly not the only ones interested in this type of program, so we shall briefly review the work of other groups in this field and then indicate the nature of our specific contributions and the conclusions that follow.

As is well known, collective coordinates were first introduced in nuclei in the 1930's through the liquid drop model of Bohr[24]; in the early 1950's Bohr and Mottelson[25] had already correlated them with the many-body mass quadrupole matrix $\mathbf{q} = \|\mathbf{q}_{ij}\|$ of (3.8).

Since the $\mathbf{q}_{ij} = x_{is}x_{js} - (\frac{1}{3})(x_{ks}x_{ks})\delta_{ij}$ is summed over the particle indices or, more correctly, over the $s = 1,2,...,n = A - 1$ Jacobi coordinates, as from the beginning we eliminate the center of mass motion, the quadrupole matrix is an invariant of o(n). This fact made it possible, in the early 1970's, for groups led by Filippov[26] and Vanagas[27] to express the $3n$ coordinates x_{is}, $i = 1,2,3$; $s = 1,2,...,n$ in terms of three deformation parameters (the ρ_k of Sec. IV), three Euler angles ϑ_k, $k = 1,2,3$ that take us from the frame of reference fixed in space to the one fixed in the body, and $3n - 6$ remaining variables.

Only the first six parameters, i.e., ρ_k, ϑ_k, $k = 1,2,3$ played a role in collective excitations since the states were associated with a definite irrep of o(n) fixed by shell model considerations.

By the mid-1970's Rosensteel and Rowe[12] and Biedenharn *et al.*[18] initiated an approach to the problem by first identifying the desired collective motions and then determining the operators that generate these motions, as well as the Lie algebra they satisfy, which turned out to be sp$(6,R)$.

At first glance it seemed that work on a microscopic description of collective motions in the Soviet Union[26,27] and North America[12,18] were not related since the first was based on the o(n) Lie algebra, while the second used sp$(6,R)$. However, both dealt with a problem of $3n$ degrees of freedom associated with the metaplectic representation of the Lie algebra sp$(6n,R)$, which contains among its subalgebras sp$(6,R) \oplus$ o(n). In this case the irreps of sp$(6,R)$ and o(n)

are related, i.e., "complementary," as was shown by Moshinsky and Quesne[28] in the early 1970's. Thus the two approaches were equivalent, as was soon realized by all concerned.

Thus one arrived at what is now known as the symplectic model of the nucleus based on $sp(6,R)$, which besides the groups mentioned was also analyzed by Deenen and Quesne,[29] Kramer *et al.*,[30] and the present authors and their collaborators.[19]

The work of all these groups mainly stressed the chain $sp(6,R) \supset u(3) \supset o(3)$, which leads to spectra with rotational bands and where one expects to have strongly deformed nuclei. However, the Bohr and Mottelson model and the interacting boson approximation also give the possibility of vibrational bands; thus the question arose as to whether in the symplectic model there was another chain of subalgebras which could bring out this type of spectra.

It was shown by Moshinsky[1] that such a chain could be $sp(6,R) \supset sp(2,R) \oplus o(3)$ and thus it merited as careful an analysis as had been lavished on $sp(6,R) \supset u(3) \supset o(3)$.

The present series of papers then began with the purpose of obtaining the basis states for irreps in the positive discrete series of the chain $sp(6,R) \supset sp(2,R) \oplus o(3)$, as well as the matrix elements of the generators with respect to them. We quickly realized that the problem was difficult and thus we first looked at the case when the dimensionality of the space was $d = 2$ rather than $d = 3$, i.e., at the chain $sp(4,R) \supset sp(2,R) \oplus o(2)$.[2] We then considered the case $d = 3$ for closed shells,[3] i.e., the irrep $[h\ h\ h]$ of $sp(6,R)$, as well as the general case, to obtain the branching rules and the basis states in terms of elementary permissible diagrams.[4]

At this stage we realized that the problem had to be looked at in a more general context if we were going to understand its structure. The following questions arose for $sp(2d,R)$ in general, although our interest was centered only on $d = 3$ or 2.

(i) What are the maximal subalgebras of $sp(2d,R)$ which contain $O(d)$?

(ii) What are the chains of maximal subalgebras of $sp(2d,R)$ which are more convenient for labeling the states and determining the matrix elements of the generators with respect to them?

(iii) What type of Hamiltonians (particularly those of second degree in the generators) can be associated with the Casimir operators of the maximal subalgebras and what is the nature of their spectra?

(iv) Can one introduce operators whose expectation values characterize the shape of an n-body system in a d-dimensional space and are these operators in the enveloping algebra of $sp(2d,R)$?

(v) If a given chain of subalgebras gives, through its Casimir operators, a Hamiltonian whose spectrum is rotational, does this imply a large deformation? Conversely, if the spectrum is vibrational, does this imply small deformations?

In Ref. 5 we answered questions (i)–(v) for $d = 2$; we refer the reader to this paper for the detailed response. In the present paper we answer the questions for the case $d = 3$ and proceed to summarize our responses using the corresponding labels (i)–(v).

(i) The maximal subalgebras of $sp(6,R)$ "modulo" linear canonical transformations are $sp(2,R) \oplus o(3)$, $u(3)$, and $cm(3)$, as discussed in Sec. II.

(ii) None of the maximal subalgebras gives a convenient basis for determining the matrix representation of the generators of $sp(6,R)$, so we preferred the nonorthonormal monomial basis (4.2), for which explicit analytic expressions for the matrix representations were derived in Ref. 6.

(iii) Hamiltonians of second degree in the generators of $sp(6,R)$ can be associated with the Casimir operators of the maximal subalgebras $sp(2,R) \oplus o(3)$ and $u(3)$, giving, respectively, vibrational and rotational spectra, as discussed in Sec. III.

(iv) One can introduce operators whose expectation values characterize the shape, i.e., the deformation away from sphericity, of eigenstates associated with definite Hamiltonians. These shape operators are in the enveloping algebra of the maximal subalgebra $cm(3)$, as shown in Sec. III; their expectation values can be obtained with the help of the techniques outlined in Sec. II.

(v) The $u(3)$ maximal subalgebra gives rise to rotational spectra through a Casimir operator which commutes with $2R$ of (3.3), i.e., the harmonic oscillator Hamiltonian. Thus as we go to higher excitations of quanta, i.e., $\mathcal{N} = 0,1,2,3,\ldots$, we will not be obtaining the low lying spectra, but the spectrum associated with giant quadrupole and monopole resonances. This also occurs for the $sp(2,R) \oplus o(3)$ maximal subalgebra whose Casimir operator also commutes with $2R$, but gives rise to vibrational spectra. The interesting point is that at least for the levels in the high excitation region of monopole and quadrupole giant resonances, the deformation in the $sp(2,R) \oplus o(3)$ and $u(3)$ limits have almost the same values. Thus the type of spectra (vibrational or rotational) is not necessarily correlated with the type of deformation (small or large).

On the other hand, if we add to $2R$ terms of higher than second degree in the generators of $cm(3)$ that are Hermitian and invariant under rotations and time reflections, we obtain a Hamiltonian which mixes strongly the energy levels with different numbers of quanta $\mathcal{N}$ and modifies the *low lying* energy levels of nuclei, giving rotational bands for open shell nuclei accompanied by strong deformations. Thus we come to the conclusion that while rotational bands can be generated both by the $u(3)$ and $cm(3)$ maximal subalgebras, only the latter, at least in the symplectic nuclear model, seem to be accompanied by an arbitrary strong deformation.

ACKNOWLEDGMENTS

We would like to conclude this series of papers by thanking O. Castaños, M. Nicolescu, and R. T. Sharp for their collaboration. We are also grateful for extensive discussions regarding these matters with P. Kramer and C. Quesne and in this particular paper with J. van der Jeugt and H. de Meyer. It would be too lengthy to mention all those who with their remarks, both supportive and critical, clarified our ideas in this field, but we would like to thank them all.

APPENDIX: MAXIMAL SUBALGEBRAS OF sp(2*d*,*R*)

In Sec. II we proved that the set of generators

$$\langle L_{ij},R,M,N\rangle \tag{A1}$$

is a maximal subalgebra of the sp$(2d,R)$ Lie algebra. In this Appendix we present a derivation of the two other distinct sets of generators that are also, for any d, maximal subalgebras of sp$(2d,R)$ containing the o(d) generators; we also discuss the additional possibilities that occur in the exceptional case $d=2$. As shown in Sec. II, these subalgebras can be obtained by consideration of commutators among the various operators in the scheme

$$\left\langle L_{ij};\ \begin{matrix}\alpha\mathbf{M}_{ij}+\beta\mathbf{N}_{ij}+\gamma\mathbf{R}_{ij}\\ aM+bN+cR\end{matrix};\ \begin{matrix}\alpha'\mathbf{M}_{ij}+\beta'\mathbf{N}_{ij}+\gamma'\mathbf{R}_{ij}\\ a'M+b'N+c'R\end{matrix}\right\rangle. \tag{A2}$$

To begin, we have

$$\begin{aligned}&[\alpha\mathbf{M}_{ij}+\beta\mathbf{N}_{ij}+\gamma\mathbf{R}_{ij},\ \alpha\mathbf{M}_{i'j'}+\beta\mathbf{N}_{i'j'}+\gamma\mathbf{R}_{i'j'}]\\ &\quad=(i/16)(\gamma^2-\beta^2-\alpha^2)(L_{ij'}\delta_{i'j}+L_{ji'}\delta_{ij'}\\ &\qquad+L_{jj'}\delta_{ii'}+L_{ii'}\delta_{jj'}),\end{aligned} \tag{A3}$$

so that depending on $(\gamma^2-\beta^2-\alpha^2)$ being larger, equal, or smaller than zero we have different types of subalgebras. As seen in (2.22), there are three canonical forms for the operator $\alpha\mathbf{M}_{ij}+\beta\mathbf{N}_{ij}+\gamma\mathbf{R}_{ij}$ determined by the value of $(\gamma^2-\beta^2-\alpha^2)$. We proceed to discuss each of the cases separately, replacing in (A2) the operator $\alpha\mathbf{M}_{ij}+\beta\,\mathbf{N}_{ij}+\gamma\mathbf{R}_{ij}$ by its corresponding canonical form.

Case (i) $(\gamma^2-\beta^2-\alpha^2)>0$. In case (i) the scheme (A2) becomes, without loss of generality,

$$\left\langle L_{ij};\ \begin{matrix}\mathbf{R}_{ij}\\ aM+bN+cR\end{matrix};\ \begin{matrix}\alpha'\mathbf{M}_{ij}+\beta'\mathbf{N}_{ij}\\ a'M+b'N+c'R\end{matrix}\right\rangle. \tag{A4}$$

However, by means of a canonical transformation of the type (2.17a), which leaves $\mathbf{R}_{ij}$ invariant, we can replace $\alpha'\mathbf{M}_{ij}+\beta'\mathbf{N}_{ij}$ by either $\mathbf{M}_{ij}$ or $\mathbf{N}_{ij}$ alone; since the two possibilities lead to equivalent subalgebras we choose $\mathbf{M}_{ij}$ and assert that still, without losing generality, the scheme (A2) becomes

$$\left\langle L_{ij};\ \begin{matrix}\mathbf{R}_{ij}\\ aM+bN+cR\end{matrix};\ \begin{matrix}\mathbf{M}_{ij}\\ a'M+b'N+c'R\end{matrix}\right\rangle. \tag{A5}$$

Consider now, for $i\neq j$, the commutator

$$[\mathbf{R}_{ij},\mathbf{M}_{jj}]=i(\tfrac{1}{2}-1/d)\mathbf{N}_{ij}. \tag{A6}$$

Let us suppose for the present that $d\geqslant 3$. If (A5) is going to close under commutation, then from (A6) we conclude that it must be independent of $\mathbf{M}_{ij}$. Next, the commutators

$$[\mathbf{R}_{ij},aM+bN+cR]=i(a\mathbf{N}_{ij}-b\mathbf{M}_{ij}), \tag{A7a}$$

$$[\mathbf{R}_{ij},a'M+b'N+c'R]=i(a'\mathbf{N}_{ij}-b'\mathbf{M}_{ij}) \tag{A7b}$$

imply that $a=b=a'=b'=0$ if (A5) is going to close under commutation. In this way we have demonstrated that when $d\geqslant 3$, (A5) is a maximal subalgebra of sp$(2d,R)$ if it consists only of the elements

$$\langle L_{ij},\mathbf{R}_{ij},R\rangle. \tag{A8}$$

When $d=2$, the set (A8) is still a maximal subalgebra. However, if $d=2$, then (A6) does not preclude the presence of $\mathbf{M}_{ij}$; keeping $\mathbf{M}_{ij}$, then (A7a) and (A7b) tell us only that $a=a'=0$ if we want to have an algebra, while from

$$[\mathbf{M}_{ij},bN+cR]=-i(b\,\mathbf{R}_{ij}+c\mathbf{N}_{ij}), \tag{A9a}$$

$$[\mathbf{M}_{ij},b'N+c'R]=-i(b'\mathbf{R}_{ij}+c'\mathbf{N}_{ij}) \tag{A9b}$$

we deduce that (A5) can close under commutation if $c=c'=0$. Finally, from (2.9) with $d=2$, we have

$$[\mathbf{R}_{ij},\mathbf{M}_{i'j'}]=(i/4)(\delta_{ii'}\delta_{jj'}+\delta_{ij'}\delta_{ji'}-\delta_{ij}\delta_{i'j'})N. \tag{A10}$$

Thus we conclude that sp$(4,R)$ has the maximal subalgebra

$$\langle L_{12},\mathbf{R}_{ij},\mathbf{M}_{ij},N\rangle. \tag{A11a}$$

Of course, it also has the maximal subalgebra

$$\langle L_{12},\mathbf{R}_{ij},\mathbf{N}_{ij},M\rangle \tag{A11b}$$

related to (A11a) by a canonical transformation of the type (2.17a). Both (A11a) and (A11b) are realizations of the o(3,1) Lie algebra.

Case (ii) $(\gamma^2-\beta^2-\alpha^2)<0$. In case (ii) the scheme (A2) becomes

$$\left\langle L_{ij};\ \begin{matrix}\mathbf{N}_{ij}\\ aM+bN+cR\end{matrix};\ \begin{matrix}\alpha'\mathbf{M}_{ij}+\gamma'\mathbf{R}_{ij}\\ a'M+b'N+c'R\end{matrix}\right\rangle. \tag{A12}$$

By means of a canonical transformation of the type (2.17b) that leaves $\mathbf{N}_{ij}$ invariant, it is possible to replace the operator $\alpha'\mathbf{M}_{ij}+\gamma'\mathbf{R}_{ij}$ in (A12) by a standard form whose expression is determined by the relative values of α' and γ', as we proceed to discuss.

If $\gamma'^2>\alpha'^2$ the standard form is $\mathbf{R}_{ij}$; then by a rotation of type (2.17a) we can replace $\mathbf{N}_{ij}$ in (A12) by $\mathbf{M}_{ij}$, thereby leaving (A12) exactly as in (A5) and thus giving no new results.

If $\gamma'^2<\alpha'^2$ the standard form is $\mathbf{M}_{ij}$ and therefore (A12) becomes

$$\left\langle L_{ij};\ \begin{matrix}\mathbf{N}_{ij}\\ aM+bN+cR\end{matrix};\ \begin{matrix}\mathbf{M}_{ij}\\ a'M+b'N+c'R\end{matrix}\right\rangle. \tag{A13}$$

Now,

$$\begin{aligned}&[\mathbf{N}_{ij},\mathbf{M}_{ij}]\\ &\quad=i\left\{\frac{1}{4}(\mathbf{R}_{i'j}\delta_{j'i}+\mathbf{R}_{j'i}\delta_{i'j}+\mathbf{R}_{ii'}\delta_{jj'}+\mathbf{R}_{jj'}\delta_{ii'})\right.\\ &\qquad\left.-\frac{1}{d}\mathbf{R}_{ij}\delta_{i'j'}-\frac{1}{d}\mathbf{R}_{i'j'}\delta_{ij}\right\}\\ &\qquad+i\left\{\frac{1}{2d}\delta_{ii'}\delta_{jj'}+\frac{1}{2d}\delta_{ij'}\delta_{ji'}-\frac{1}{d^2}\delta_{ij}\delta_{i'j'}\right\}R\end{aligned} \tag{A14}$$

and the first curly bracket vanishes identically for $d=2$, but not for $d\geqslant 3$. We thus conclude that the set (A13) lead to no subalgebra when $d\geqslant 3$, while (A14) supplemented with (A9a), (A9b), and

$$[\mathbf{N}_{ij},aM+bN+cR]=i(a\mathbf{R}_{ij}+c\mathbf{M}_{ij}), \tag{A15a}$$

$$[\mathbf{N}_{ij},a'M+b'N+c'R]=i(a'\mathbf{R}_{ij}+c'\mathbf{M}_{ij}) \tag{A15b}$$

permit us to deduce that, when $d=2$, (A13) gives a subalgebra if $a=b=a'=b'=0$, i.e., sp$(4,R)$ has the maximal subalgebra

$$\langle L_{12},\mathbf{N}_{ij},\mathbf{M}_{ij},R\rangle, \tag{A16}$$

which is a realization of the Lie algebra sp$'(2,R)\oplus$sp$''(2,R)$.

Finally, if $\gamma'=\pm\alpha'$ the standard forms of

$\alpha'\mathbf{M}_{ij}+\gamma'\mathbf{R}_{ij}$ are $\mathbf{R}_{ij}+\mathbf{M}_{ij}$ and $\mathbf{R}_{ij}-\mathbf{M}_{ij}$, which are given in (2.21) in terms of coordinates and momenta and were shown, in the text following (2.21), to be related to each other by a linear canonical transformation which only changes the sign of $\mathbf{N}_{ij}$. Consequently, without losing generality, we can now replace (A12) by

$$\left\langle L_{ij}; \frac{\mathbf{N}_{ij}}{aM+bN+cR}; \frac{\mathbf{R}_{ij}-\mathbf{M}_{ij}}{a'M+b'N+c'R}\right\rangle. \quad \text{(A17)}$$

Then the commutators (A15a) and (A15b) allow us to deduce that $c=-a$, $c'=-a'$ if (A17) is going to give an algebra, so effectively we can replace the two linear combinations of the scalar operators in (A17) by N and $(R-M)$. It is then found that all other commutators between the operators in (A17) give linear combinations of the operators already appearing. Hence the third maximal subalgebra of sp$(2d,R)$ that we obtain is

$$\langle L_{ij},\mathbf{N}_{ij},N,\mathbf{R}_{ij}-\mathbf{M}_{ij},R-M\rangle. \quad \text{(A18)}$$

Case (iii) $(\gamma^2-\beta^2-\alpha^2)=0$. The scheme (A2) now becomes, according to (2.22),

$$\left\langle L_{ij}; \frac{\mathbf{R}_{ij}-\mathbf{M}_{ij}}{aM+bN+cR}; \frac{\delta(\mathbf{M}_{ij}+\mathbf{R}_{ij})+\epsilon\mathbf{N}_{ij}}{a'M+b'N+c'R}\right\rangle. \quad \text{(A19)}$$

Consider first, for $i\neq j$, the commutator

$$\begin{aligned}[]
[\mathbf{R}_{ij}-\mathbf{M}_{ij},\delta(\mathbf{R}_{jj}+\mathbf{M}_{jj})+\epsilon\mathbf{N}_{jj}]
&=\frac{1}{4}\,\delta L_{ij}+2i\delta\left(\frac{1}{2}-\frac{1}{d}\right)\mathbf{N}_{ij}\\
&\quad+i\epsilon\left(\frac{1}{2}-\frac{1}{d}\right)(\mathbf{R}_{ij}-\mathbf{M}_{ij}). \quad \text{(A20)}
\end{aligned}$$

If (A19) is going to close under commutation, then the fact that no term $(\mathbf{R}_{jj}+\mathbf{M}_{jj})$ appears on the rhs of (A20) implies that necessarily $\delta=0$. However, then we see that by setting $\delta=0$ in (A19) this scheme becomes identical to (A17); thus we eventually obtain in case (iii) the same maximal subalgebra (A18) which was obtained for case (ii). Of course, this conclusion remains valid as long as $d\geqslant 3$, since for $d=2$ (A20) does not preclude the presence of $\mathbf{R}_{jj}+\mathbf{M}_{jj}$.

Restricting ourselves now to $d=2$ and setting $\delta=1$ in (A19), the commutator

$$\begin{aligned}[]
[\mathbf{R}_{ij}-\mathbf{M}_{ij},aM+bN+cR\,]
&=ib(\mathbf{R}_{ij}-\mathbf{M}_{ij})+i(a+c)\mathbf{N}_{ij} \quad \text{(A21)}
\end{aligned}$$

imposes $c=-a$ if (A19) is going to close under commutation, while

$$\begin{aligned}[]
[\mathbf{R}_{ij}+\mathbf{M}_{ij}+\epsilon\mathbf{N}_{ij},aM+bN-aR\,]
&=i2a\mathbf{N}_{ij}-ib(\mathbf{R}_{ij}+\mathbf{M}_{ij})+i\epsilon a(\mathbf{R}_{ij}-\mathbf{M}_{ij}) \quad \text{(A22)}
\end{aligned}$$

requires, for the same purpose, $b=-2a/\epsilon$. The scheme (A19) thus becomes

$$\begin{aligned}
&\langle L_{12},(\mathbf{R}_{ij}-\mathbf{M}_{ij}),(\mathbf{R}_{ij}+\mathbf{M}_{ij}+\epsilon\mathbf{N}_{ij}),\\
&[R-M+(2/\epsilon)N\,]\rangle. \quad \text{(A23)}
\end{aligned}$$

Equation (A23) is another maximal subalgebra of sp$(4,R)$ since we have, when $d=2$,

$$\begin{aligned}[]
[\mathbf{R}_{ij}-\mathbf{M}_{ij},\mathbf{R}_{i'j'}+\mathbf{M}_{i'j'}+\epsilon\mathbf{N}_{i'j'}\,]
&=\tfrac{1}{4}i\kappa L_{12}+\tfrac{1}{4}i\epsilon(\delta_{ii'}\delta_{jj'}+\delta_{i'j}\delta_{ij'}\\
&\quad-\delta_{ij}\delta_{i'j'})[R-M+(2/\epsilon)N\,], \quad \text{(A24)}
\end{aligned}$$

with $\kappa=1$, 0, or -1 depending on the values of the four indices.

[1]M. Moshinsky, J. Math. Phys. **25**, 1555 (1984).
[2]E. Chacón, P. Hess, and M. Moshinsky, J. Math. Phys. **25**, 1565 (1984).
[3]O. Castaños, E. Chacón, and M. Moshinsky, J. Math. Phys. **25**, 2815 (1984).
[4]M. Moshinsky, M. Nicolescu, and R. T. Sharp, J. Math. Phys. **26**, 2995 (1985).
[5]E. Chacón, P. Hess, and M. Moshinsky, J. Math. Phys. **28**, 2223 (1987).
[6]E. Chacón and M. Moshinsky, J. Phys. A Math. Gen. **20**, 4595 (1987).
[7]J. van der Jeugt and H. de Meyer, J. Phys. A Math. Gen. **20**, L263 (1987).
[8]M. Hamermesh, *Group Theory* (Addison-Wesley, Reading, MA, 1962); G. Racah, in *Springer Tracts*, edited by G. Höhler (Springer, Berlin, 1965), Vol. 37.
[9]O. Castaños, E. Chacón, M. Moshinsky, and C. Quesne, J. Math. Phys. **26**, 2107 (1985).
[10]M. Moshinsky and C. Quesne, J. Math. Phys. **12**, 1772 (1971); M. Moshinsky and P. Winternitz, *ibid.* **21**, 1667 (1980).
[11]J. P. Elliott, Proc. R. Soc. London Ser. A **245**, 128, 562 (1958).
[12]G. Rosensteel and D. J. Rowe, Ann. Phys. (NY) **126**, 198, 343 (1980); L. Weaver, L. C. Biedenharn, and R. Y. Cusson, *ibid.* **77**, 250 (1973).
[13]P. Kramer and M. Moshinsky, in *Group Theory and Its Applications*, edited by E. M. Loebl (Academic, New York, 1968).
[14]M. Moshinsky, *Group Theory and the Many Body Problem* (Gordon & Breach, New York, 1968), pp. 26, 37, and 44.
[15]V. Bargmann and M. Moshinsky, Nucl. Phys. **23**, 177 (1961); M. Moshinsky, Rev. Mod. Phys. **34**, 813 (1962).
[16]R. Gaskell, G. Rosensteel, and R. T. Sharp, J. Math. Phys. **22**, 2732 (1981).
[17]O. Castaños, E. Chacón, A. Frank, and M. Moshinsky, J. Math. Phys. **20**, 35 (1979).
[18]O. L. Weaver, R. Y. Cusson, and L. C. Biedenharn, Ann. Phys. (NY) **102**, 493 (1976).
[19]O. Castaños, E. Chacón, P. Hess, A. Frank, and M. Moshinsky, J. Math. Phys. **23**, 2537 (1982).
[20]S. Barnard and J. M. Child, *Higher Algebra* (MacMillan, London, 1936).
[21]I. M. Gelfand and M. L. Zetlin, Dokl. Akad. Nauk SSSR **71**, 825 (1950).
[22]P. O. Hess, M. Seiwert, J. A. Maruhn, and W. Greiner, Z. Phys. A **296**, 147 (1980); P. O. Hess, J. A. Maruhn, and W. Greiner, J. Phys. G **67**, 737 (1981).
[23]P. Park, J. Carvalho, M. Vassanji, D. J. Rowe, and G. Rosensteel, Nucl. Phys. A **414**, 93 (1984); *Group Theory and its Applications in Physics, 1980*, AIP Conference Proceedings No. 71, edited by T. H. Seligman (AIP, New York, 1980), p. 177.
[24]N. Bohr and F. Kalckar, Danske Vid. Selsk. Mat.-Fys. Medd. **14**, 10 (1937).
[25]A. Bohr and B. R. Mottelson, Danske Vid. Selsk. Mat.-Fys. Medd. **27**, 16 (1953).
[26]G. F. Filippov and V. I. Ovcharenko, Yad. Fiz. **30**, 646 (1979); V. C. Vasilevsky, Ya. F. Smirnov, and G. F. Filippov, *ibid.* **32**, 987 (1980).
[27]V. Vanagas, *Lecture Notes in Physics* (Univ. of Toronto, Toronto, 1977); V. Vanagas, in *Group Theory and its Applications in Physics*, AIP Conference Proceedings No. 71, edited by T. H. Seligman (AIP, New York, 1980), p. 220.
[28]M. Moshinsky and C. Quesne, J. Math. Phys. **11**, 1631 (1970); M. Moshinsky and C. Quesne, Phys. Lett. B **29**, 482 (1971).
[29]J. Deenen and C. Quesne, J. Math. Phys. **23**, 2004 (1982); **25**, 2354 (1984); **26**, 2705 (1985).
[30]P. Kramer, Ann. Phys. (NY) **141**, 254, 269 (1982); P. Kramer, Z. Papadopolos, and W. Schweitzer, Nucl. Phys. A **441**, 461 (1984).

The integral theorem for supersymmetric invariants

F. Constantinescu and H. F. de Groote
Fachbereich Mathematik, Johann Wolfgang Goethe Universität, 6000 Frankfurt am Main, West Germany

(Received 7 September 1988; accepted for publication 14 December 1988)

A supersymmetric integral theorem that extends results of Parisi, Sourlas, Efetov, Wegner, and others is rigorously proved. In particular, arbitrary generators are allowed in the integrand (instead of canonical ones) and the invariance condition is very much relaxed. The connection with Cauchy's integral formula is made transparent. In passing, the unitary Lie supergroup is studied by using elementary methods. Applications in the theory of disordered systems are discussed.

I. INTRODUCTION

Supersymmetry provides a useful computational framework in several areas of physics. Besides field theory and gravitation, applications in condensed matter and nuclear physics, as well as in stochastic differential equations, are well known. One of the particularities of superanalysis is the existence of integral theorems first applied by Parisi and Sourlas to dimensional reduction.[1] Roughly speaking, this says that

$$\int F(V)dV = F(0), \tag{1.1}$$

where $V=(x,\Theta_1,\Theta_2)$ and F is supersymmetric with zero boundary condition at infinity. Here $x\in\mathbb{R}^d$, $d\geqslant 1$, and Θ_1, Θ_2 are canonical generators of a Grassmann algebra. The integration in (1.1) is taken in Berezin's sense. The function F is invariant with respect to "superrotations," i.e., transformations that preserve the sum $x^2+\Theta_1\Theta_2$. Several extensions of (1.1) exist. The result (1.1) has no counterparts in classical analysis or invariant theory.

Another result, which goes back to Efetov[2] and Wegner[3] (see also Ref. 4), says that

$$\int F(Q)dQ = F(0), \tag{1.2}$$

where $Q=\left(\begin{smallmatrix} a & \Theta \\ \beta & ib \end{smallmatrix}\right)$, $a,b\in\mathbb{R}$, is a 2×2 supermatrix and F is an invariant function, [i.e., $F(Q)=F(S^{-1}QS)$, S superunitary] with zero boundary conditions at infinity. Again, the integration in (1.2) is taken in the sense of Berezin (for precise definitions see Sec. II). Recently Wegner has proved a more general result that contains (1.1) and (1.2) as particular cases.[3]

In this paper, we extend the results above in several directions. In particular we allow as integration variables arbitrary generators instead of canonical ones, and make consequent use of Berezin's integration theory.[5]

On the other hand, we weaken the invariance condition on F from a classical Lie supergroup to a discrete group isomorphic to $\mathbb{Z}^n$. Our proofs are rigorous. We use a nice idea of Wegner and the complex $z,\bar{z}$ formalism introduced in Ref. 4 in order to perform induction and to deal with the matrix case. The main result is given in Theorem 4.1. The ideas of the proof develop gradually from Sec. III to Sec. IV. Section II contains preparatory material that was adapted to our needs. In Sec. V we study by elementary methods the unitary supergroups. No use of the corresponding Lie-superalgebras is necessary. Section VI contains interesting additional material.

We want to stress that our proof reveals, in particular, the intimate connection of the supersymmetric integral formulas with Cauchy's integral formula of complex analysis. The matrix case of the present supersymmetric integral theorem should play an important role in the rigorous study of the extended state region of the Anderson problem in the n-orbital model.[2,3] For details see the discussion in Ref. 4. We also expect applications to other areas of condensed matter physics.

II. FUNCTIONS ON SUPERSPACE AND SUPERSPACE INTEGRATION

In this section we follow Ref. 5 and specialize to our needs. Let Λ_q be a complex Grassmann algebra with canonical generators $\xi_1,\ldots,\xi_q$. Let $\Lambda_{p,q}$ be the algebra with elements

$$f=f(x,\xi)=\sum_{k\geqslant 0}\sum_{i_1<\cdots<i_k} f_{i_1\cdots i_k}(x)\xi_{i_1}\cdots\xi_{i_k}, \tag{2.1}$$

where $x\in\mathbb{R}^p$ and $f_{i_1\cdots i_k}$ are complex-valued infinitely differentiable functions of x. Let $x_1,\ldots,x_p$ be the usual coordinate functions on $\mathbb{R}^p$. Then $x_1,\ldots,x_p,\xi_1,\ldots,\xi_q$ are canonical generators of $\Lambda_{p,q}$.

Let us denote by f_0 the zero order ($k=0$) term in (2.1). The body $m(f)$ of f is defined by

$$m(f)(x)=f(x,0)=f_0(x). \tag{2.2}$$

If $f_1,\ldots,f_r$ are real or complex-valued functions then the range of $(m(f_1),\ldots,m(f_r))$ is denoted by $\operatorname{Spec}(f_1,\ldots,f_r)$.

Now we recall the general concept of functions depending on commutative and noncommutative variables.[5] Let $\varphi=(\varphi_1,\ldots,\varphi_m)$, $\psi=(\psi_1,\ldots,\psi_n)$, where $\varphi_1,\ldots,\varphi_m$ and $\psi_1,\ldots,\psi_n$ are even and odd elements in $\Lambda_{p,q}$ respectively. If $g=g(x_1,\ldots,x_m)$ is a complex-valued function on $\mathbb{R}^m$, the superposition $g(\varphi_1,\ldots,\varphi_m)$ is defined as follows: separate in φ_i the term of zero degree

$$\varphi_i(x,\xi)=a_i(x)+h_i(x,\xi),\deg h_i\geqslant 2$$

and define $g(\varphi_1,\ldots,\varphi_m)$ by a formal Taylor series expansion:

 0022-2488/89/050981-12$02.50

$$g(\varphi_1,\ldots,\varphi_m)$$
$$= g(a_1,\ldots,a_m) + \sum_{\Sigma k_i > 0} \frac{h_1^{k_1}\cdots h_m^{k_m}}{k_1!\cdots k_m!} \times \frac{\partial^{k_1+\cdots+k_m}}{\partial a_1^{k_1}\cdots\partial a_m^{k_m}} g(a_1,\ldots,a_m) \quad (2.3)$$

(notice that the h_i's are nilpotent, so the sum on the right-hand side is finite). Now if $F = F(y,\eta) \in \Lambda_{m,n}$

$$F(y,\eta) = \sum_{k>0} \sum_{i_1<\cdots<i_k} F_{i_1\cdots i_k}(y)\eta_{i_1}\cdots\eta_{i_k}, \quad (2.4)$$

we can define $F(\varphi,\psi) \in \Lambda_{p,q}$ by

$$F(\varphi,\psi) = \sum_{k>0} \sum_{i_1<\cdots<i_k} F_{i_1\cdots i_k}(\varphi_1,\ldots,\varphi_m)\psi_{i_1}\cdots\psi_{i_k}. \quad (2.5)$$

Here, $F(\varphi,\psi) = F(\varphi_1,\ldots,\varphi_m,\psi_1,\ldots,\psi_n)$ is a function of even and odd elements in $\Lambda_{p,q}$ with values in $\Lambda_{p,q}$. Formally it has to be understood as a function in the usual sense depending on commutative and noncommutative variables.

Let $y_i(x,\xi)$, $i = 1,\ldots,p$, be even and η_j, $j = 1,\ldots,q$, be odd elements of $\Lambda_{p,q}$. The set $\{y_1,\ldots,y_p,\eta_1,\ldots,\eta_q\}$ is called a system of generators for $\Lambda_{p,q}$ if (i) Spec $(y_1,\ldots,y_p) = \mathbb{R}^p$, (ii) every element $f \in \Lambda_{p,q}$ can be written by means of y_i,η_j in the form

$$f = \sum_k \sum_{i_1<\cdots<i_k} f_{i_1\cdots i_k}(y_1,\ldots,y_p)\eta_{i_1}\cdots\eta_{i_k}. \quad (2.6)$$

Now we define Berezin's integral and formulate the main theorem concerning the change of (superspace) variables.

Let x_i,ξ_j be a system of generators in $\Lambda_{p,q}$ (not necessarily the canonical ones!). Then we can write $x_i = x_i(s,\xi)$, where $s \in \mathbb{R}^p$. Let $t_i = x_i(s,0) \in \mathbb{R}^p$.

Definition 2.1: Let $f \in \Lambda_{p,q}$. We set

$$\int f(x,\xi)d_{x,\xi} = \int f(t,\xi)dt\,d\xi. \quad (2.7)$$

The integral w.r.t. dt ($dt = \Pi_{i=1}^p dt_i$) is understood in the usual sense whereas $d\xi = \Pi_{j=1}^q d\xi_i$ indicates integration w.r.t. the anticommuting ξ_j defined by

$$\int d\xi_j = 0, \quad \int \xi_j\, d\xi_j = 1 \quad (2.8)$$

and

$$\int ab\, d\xi_j = a \int b\, d\xi_j \quad (2.9)$$

if

$$\frac{\partial}{\partial \xi_j} a = 0.$$

From this definition it follows that for $f \in \Lambda_{p,q}$ we have

$$\int f(x,\xi)d_{x,\xi} = \int f_{1\cdots q}(t_1,\ldots,t_p)dt_1,\ldots,dt_p. \quad (2.10)$$

In classical analysis it is convenient to assume that the differentials anticommute among themselves but commute with the variables. In superanalysis we extend this property by assuming that the dx_i are anticommuting but commute with $x_i,\xi_j,d\xi_j$, which commute among themselves, i.e., besides known relations we have

$$[d\xi_i,\xi_j] = [d\xi_i,d\xi_j] = [dx_i,d\xi_j] = [d\xi_i,x_j] = \{dx_i,\xi_j\} = 0 \quad (2.11)$$

($\{\cdots\}$ denotes the anticommutator).

Now we prepare the change of variable formula for the Berezin integration. Let $x_j = x_j(y,\eta)$, $\xi_j = \xi_j(y,\eta)$ be a transformation from one system of generators in $\Lambda_{p,q}$ to another. Let

$$A_{ik} = \frac{\partial x_i}{\partial y_k}, \quad B_{ik} = x_i \frac{\overleftarrow{\partial}}{\partial \eta_k}, \quad C_{ik} = \frac{\partial \xi_i}{\partial y_k}, \quad D_{ik} = \frac{\partial \xi_i}{\partial \eta_k}, \quad (2.12)$$

where $\overleftarrow{\partial}/\partial\eta_k$ means left derivative.

Let

$$R = R(x,\xi;y,\eta) = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \quad (2.13)$$

be the supermatrix with matrix elements (2.12). The superdeterminant (Berezinean) of the supermatrix is

$$\Delta = \Delta(x,\xi;y,\eta) = S\det R = \det(A - BD^{-1}C)\det D^{-1}. \quad (2.14)$$

Note that D is invertible as a consequence of the fact that (y,η) and (x,ξ) are generator systems.[5]

Now let $f = f(x,\xi)$ be an infinitely differentiable function in $\Lambda_{p,q}$ with compact support. This means that the coefficients of f in its expansion as elements of $\Lambda_{p,q}$ are C^∞ functions of x having compact support. Then we have

$$\int f(x(y,\eta),\xi(y,\eta))\Delta(x,\xi;y,\eta)d_{y,\eta} = \int f(x,\xi)d_{x,\xi}. \quad (2.15)$$

In (2.15) $f(x(y,\eta),\xi(y,\eta))$ has to be understood in the sense of (2.6) with $m = p$ and $n = q$. The nontrivial proof of (2.15) is contained in Ref. 5. Examples are given in Ref. 5 that show that (2.15) can be wrong if f does not have compact support because in this case boundary terms may appear. We will use (2.15) for functions without compact support but assume zero boundary conditions at infinity.

Definition 2.2: A function $f \in \Lambda_{p,q}$ satisfies the zero boundary condition at infinity if $f(x,0) = O(|x|^{-(p+\epsilon)})$ for some $\epsilon > 0$.

For functions with zero boundary condition at infinity (2.15) can be applied (see Sec. III).

We close this section by considering the more general case in which the change of variables in an integral involves commutative and noncommutative *parameters*. We will need this in Sec. IV. Let $x_1,\ldots,x_p,u_1,\ldots,u_p,\xi_1,\ldots,\xi_q,\sigma_1,\ldots,\sigma_q$ and $y_1,\ldots,y_p,v_1,\ldots,v_p,\eta_1,\ldots,\eta_q,\zeta_1,\ldots,\zeta_q$ be two systems of generators in $\Lambda_{p+p',q+q'}(U)$, $U \subset \mathbb{R}^{p+p'}$. The change from the first to the second system of generators is given by

$$x_i = x_i(y,v,\eta,\zeta), \quad \xi_j = \xi_j(y,v,y,\zeta), \quad u_i = v_i, \quad \sigma_j = \zeta_j. \quad (2.16)$$

In this case (2.15) reads

$$\int f(x(y,u,\eta,\sigma)\xi(y,u,\eta,\sigma))\Delta(x,\xi;y,\eta)d_{y,\eta}$$

$$= \int f(x,\xi)d_{x,\xi}. \tag{2.17}$$

On the left-hand side of (2.17) both f and the Berezinean Δ depend generally on the parameters u_i,σ_j. In Sec. IV we will need (2.17) in a particular case in which although (2.15) depend on parameters, both f and Δ are independent of them.

III. SUPERSYMMETRIC INTEGRAL THEOREM FOR SOME PARTICULAR CASES

In this section we will be concerned with two particular cases of the integral theorem; one for two-component supervectors and the other for 2×2 supermatrices. In these particular cases the connection between the supersymmetric integral theorem and the Cauchy formula of complex analysis is explicitly worked out. The vector case (in the particular case of canonical generators—see below) is a well-known result of Parisi and Sourlas,[1] which was used to prove the dimensional reduction of RFIM (random field Ising model). The matrix variant of the Parisi–Sourlas result (again in the case of canonical generators) was used by Efetov,[2] Wegner,[3] Constantinescu,[4] and Verbaarschot, Weidenmüller, and Zirnbauer[6,7] in the study of localization problems of disordered electronic systems and the study of random matrices and compound-nucleus scattering, respectively. Following an idea in Ref. 4 we use the complex $z,\bar{z}$ formalism which makes the connection to the Cauchy theorem transparent.

Let $\Lambda_2 = \Lambda_2(\Theta,\beta)$ be a Grassmann algebra with canonical generators Θ,β. We consider first the vector case in which the vector components are the canonical generators of $\Lambda_{2,2}$:

$$U = (z,\Theta), \quad V = \begin{pmatrix} \bar{z} \\ \beta \end{pmatrix}, \tag{3.1}$$

where $z = x + iy$, $\bar{z} = x - iy$, $x, y\in\mathbb{R}$. Let $F = F(U,V)$ be a function of the pair (U,V) of supervectors in the sense of Sec. II. Let us comment a little on F. We write first

$$F(U,V) = \widetilde{F}(x,y,\Theta,\beta), \tag{3.2}$$

where $\widetilde{F}$ is an element of $\Lambda_{2,2}$. As such it defines as in Sec. II a function $\widetilde{F}(\varphi_1,\varphi_2,\psi_1,\psi_2)$ of two commutative and two noncommutative "variables" in an algebra $\Lambda_{p,q}$ which we denote by φ_1,φ_2 (even) and ψ_1,ψ_2 (odd). In fact $p = 2$ will be sufficient for our needs whereas $q > 2$ will depend on the problem circumstances. Two of these q canonical generators are supposed to be Θ and β. Certainly we can write

$$F(U,V) = F_0(x,y) + F_1(x,y)\Theta + F_2(x,y)\beta + F_3(x,y)\Theta\beta, \tag{3.3}$$

where F_i, $i = 0,1,2,3$ depend only on $x, y\in\mathbb{R}$. Without risking too much confusion (we hope!) we write (3.3) in the form

$$F(U,V) = F_0(z,\bar{z}) + F_1(z,\bar{z})\Theta + F_2(z,\bar{z})\beta + F_3(z,\bar{z})\Theta\beta. \tag{3.4}$$

As an example we give

$$F(U,V) = e^{-\mathrm{Str}(VU)} = e^{-z\bar{z}} + e^{-z\bar{z}}\Theta\beta \tag{3.5}$$

i.e., in this case $F_0(z,\bar{z}) = F_3(z,\bar{z}) = e^{-z\bar{z}}$. We say that F is *regular* if F_i are infinitely differentiable in $\mathbb{R}^2$ (this condition can be relaxed to C^n for some finite n).

Suppose that ξ is a canonical generator in Λ_q (the Grassmann algebra which serves in the construction of $\Lambda_{p,q}$—see Sec. II).

In the following we need the supermatrix

$$S = \begin{pmatrix} 1 & \xi \\ 0 & 1 \end{pmatrix}. \tag{3.6}$$

The inverse S^{-1} of S is given by

$$S^{-1} = \begin{pmatrix} 1 & -\xi \\ 0 & 1 \end{pmatrix}. \tag{3.7}$$

Now let $U' = US^{-1}$, $V' = SV$ be the transformed supervectors U and V. Using the definition of $\widetilde{F}$ as function of even and odd variables φ_1,φ_2 and ψ_1,ψ_2, respectively, we can write, by using (3.2)–(3.4),

$$F(U',V') = F(US^{-1},SV) = \widetilde{F}(\varphi_1,\varphi_2,\psi_1,\psi_2), \tag{3.8}$$

where $\varphi_1,\varphi_2,\psi_1,\psi_2$ are to be computed with the help of U, V, and S. This can be done as follows (the result obviously coincides with the formal computation). We write $U' = (z',\Theta')$, $V' = \binom{z_1}{\beta'}$.

Then we have

$$z' = z, \quad z_1 = \bar{z} + \xi\beta, \quad \Theta' = -z\xi + \Theta, \quad \beta' = \beta. \tag{3.9}$$

We remark in passing that $z_1 \neq \bar{z}'$ (for $\xi \neq \beta$). This point will be discussed below. The result follows from

$$F(U',V') = F_0(z',z_1) + F_1(z',z_1)\Theta' + F_2(z',z_1)\beta' + F_3(z',z_1)\Theta'\beta'$$

by just expanding $F_i(z',z_1) = F_i(z,\bar{z} + \beta\xi)$ in Taylor series and using (3.9) for Θ' and β'. The result is

$$F(U',V') = F(U,V) - zF_1(z,\bar{z})\xi - zF_3(z,\bar{z})\xi\beta + \frac{\partial F_0}{\partial \bar{z}}\xi\beta + \frac{\partial F_1}{\partial \bar{z}}\xi\beta\Theta. \tag{3.10}$$

Now we define *invariant* functions of supervectors.

Definition 3.1: We say that the regular function $F = F(U,V)$ of the pair of supervectors (U,V) is S invariant if

$$F(US^{-1},SV) = F(U,V) \tag{3.11}$$

for S given by (3.6) with $\xi \neq \beta$.

Remark: In Definition 3.1 the function F has to be considered in the sense defined above (see Sec. II).

Then we have the following lemma.

Lemma 3.1: Let $F = F(U,V)$ be a regular S invariant function of the pair of supervectors (U,V). Then we have

$$F_3(z,\bar{z}) = \frac{1}{z}\frac{\partial F_0}{\partial \bar{z}}. \tag{3.12}$$

Proof: Follows immediately from (3.10) by identifying the coefficients of $\xi\beta$. A further result (which we do not need) is $F_1(z,\bar{z}) = 0$.

Remark: The same result follows if we use instead of S given by (3.6) the supermatrix

$$S=\begin{pmatrix}1 & 0\\ \xi & 1\end{pmatrix} \tag{3.13}$$

and its inverse

$$S^{-1}\begin{pmatrix}1 & 0\\ -\xi & 1\end{pmatrix}. \tag{3.14}$$

In Definition 3.1 we have to assume $\xi \neq \Theta$. The reader can work out the relations (3.9) in this case. From Lemma 3.1 follows a first integral theorem.

Lemma 3.2: Let F be as in Lemma 3.1 with zero boundary conditions at infinity in the sense of Sec. II. Then

$$\int F(U,V)dV = F_0(0,0) = F(0,0), \tag{3.15}$$

where $dU\,dV = 2\pi i\, dz\, d\Theta\, d\bar{z}\, d\beta = 2\pi i\, dz\, d\bar{z}\, d\Theta\, d\beta$.

Remark: Whereas $d\Theta$ and $d\beta$ are supposed to commute (among themselves and with the Θ and β), dz and $d\bar{z}$ have to be interpreted as (complex) differential forms: $dz\,d\bar{z} \equiv dz\wedge d\bar{z} = 2i\,dy\wedge dx \equiv 2i\,dy\,dx$ and we omit $\wedge$ in the exterior product. Then $dU\,dV = 4\pi\, dx\, dy\, d\Theta\, d\beta$. In the proof of Lemma 3.2 we need the following well-known result of complex analysis.

Lemma 3.3 (Cauchy)*:* Let $f(z,\bar{z})$, where $z = x+iy$, $\bar{z} = x-iy$; $x, y\in\mathbb{R}$, be a continuously differentiable function with zero boundary conditions at infinity. Then

$$\int \frac{1}{\bar{z}}\frac{\partial f}{\partial z}dz\,d\bar{z} = \int \frac{1}{z}\frac{\partial f}{\partial \bar{z}}dz\,d\bar{z} = 2\pi i f(0,0). \tag{3.16}$$

Proof of Lemma 3.2: $F = F(U,V)$ being S invariant, we use (3.12) and (3.16) to get (3.15).

Remark: Assuming the S invariance of $F = F(U,V)$ with respect to S given by (3.6) and (3.12) we get $F_1(z,\bar{z}) = 0$ as well as $F_2(z,\bar{z}) = 0$ such that

$$\begin{aligned} F(U,V) &= F_0(z,\bar{z}) + \frac{1}{z}\frac{\partial}{\partial \bar{z}}F_0(z,\bar{z})\Theta\beta \\ &= F_0(z,\bar{z}) + \frac{1}{\bar{z}}\frac{\partial}{\partial z}F_0(z,\bar{z})\Theta\beta. \end{aligned} \tag{3.17}$$

The equality $(1/z)(\partial/\partial\bar{z})F = (1/\bar{z})(\partial/\partial z)F$ implies that $F(z,\bar{z})$ depends only on $z\bar{z} = x^2+y^2$ and therefore

$$\begin{aligned} F(U,V) &= f_0(|z|^2) + f_0'(|z|^2)\Theta\beta \\ &= f_0(|z|^2+\Theta\beta) = f_0(VU), \end{aligned} \tag{3.18}$$

where $f_0(|z|^2) = F_0(z,\bar{z})$.

Now let us generalize the result (3.15) to functions $F = F(U,V)$, where the components of U and V are generators but no longer canonical. In order to avoid inflationary notation we will continue to write $U = (z,\Theta)$, $V = (\begin{smallmatrix}\bar{z}\\ \beta\end{smallmatrix})$, $z = x+iy$, $\bar{z} = x-iy$, where now x,y,Θ,β are generators in $\Lambda_{2,2}$. We denote $x_0 = m(x)$, $y_0 = m(y)$, where $x_0,y_0\in\mathbb{R}$. Then $x = x_0+x_1$, $y = y_0+y_1$. Let $z_0 = x_0+iy_0$, $\bar{z}_0 = x_0-iy_0$. Then $z = z_0+x_1+iy_1$, $\bar{z} = \bar{z}_0+x_0-iy_1$. The function $F = F(U,V)$ is again considered in the sense of Sec. II, i.e., as a function of two commutative and two anticommutative variables in $\Lambda_{2,q}$, $q>2$. It is easy to see that taking an odd ξ in S given by (3.6) from the Grassmann algebra Λ_q (which is used to construct $\Lambda_{2,q}$) such that $\xi\beta\neq 0$ [or $\xi\Theta\neq 0$ if S has the form (3.13)] then the arguments of Lemmas 3.1 and 3.2 go through. Indeed, the transformations $U' = US^{-1}$, $V' = SV$ *preserve* the complex conjugacy of the zeroth order components (bodies) of $z,\bar{z}$ (i.e., of U and V) being in this sense formpreserving. This and the definition of Berezin's integral reduce the proof of the general case to the proof of the Lemmas 3.1 and 3.2.

We state the final result as a generalization of Lemma 3.2.

Lemma 3.4: Let $F = F(U,V)$ be a regular function of the pair of supervectors

$$U = (z,\Theta),\quad V = \begin{pmatrix}\bar{z}\\ \beta\end{pmatrix},\quad z = x+iy,\quad z = x-iy,$$

where x,y,Θ,β are generators of $\Lambda_{2,2}$. Suppose that F has zero boundary conditions at infinity and is invariant w.r.t. S given by (3.6) with $\xi\neq 0$ in Λ_q, $q>2$, which contains the two odd canonical generators of $\Lambda_{2,2}$. In formulating the S invariance F has to be considered in the sense of Sec. II as a function of commutative and noncommutative variables with values in $\Lambda_{2,q}$. Then

$$\int F(U,V)dU\,dV = F_0(0,0) = F(0,0), \tag{3.19}$$

where the integral in (3.19) is considered in the Berezin's sense and the meaning of $dU\,dV$ is obvious.

We pass now to the more interesting case of matrix invariants. We start again with the simple case of a 2×2 supermatrix

$$Q = \begin{pmatrix}a & \Theta\\ \beta & ib\end{pmatrix}, \tag{3.20}$$

where a,b,Θ,β are canonical generators of the algebra $\Lambda_{2,2}$. The imaginary unit i in front of b in Q will play an important role in what follows. As above we need also the algebra $\Lambda_{2,q}$, $q>2$ such that $\Lambda_{2,2}\subset\Lambda_{2,q}$. Let $F = F(Q)$ be a function of Q which we interpret as

$$F = F(Q) = \tilde{F}(a,b,\Theta,\beta), \tag{3.21}$$

where the arguments in $\tilde{F}$ can be replaced by commutative and anticommutative elements in $\Lambda_{2,q}$ (see discussion above concerning the vector case).

We introduce complex variables $z = a+ib$, $\bar{z} = a-ib$ and write as in (3.4)

$$F(Q) = F_0(z,\bar{z}) + F_1(z,\bar{z})\Theta + F_2(z,\bar{z})\beta + F_3(z,\bar{z})\Theta\beta. \tag{3.22}$$

As an example we mention

$$F(Q) = e^{-\mathrm{Str}\,Q^2} = e^{-z\bar{z}} - 2e^{-z\bar{z}}\Theta\beta, \tag{3.23}$$

where $F_0 = e^{-z\bar{z}}$, $F_1 = F_2 = 0$, $F_3 = -2e^{-z\bar{z}}$.

Now let $Q' = S^{-1}QS$ with

$$Q = \begin{pmatrix}a' & \Theta'\\ \beta' & ib'\end{pmatrix}.$$

Then $a' = a+\beta\xi$, $ib' = ib+\beta\xi$, $\Theta' = \Theta+\xi(a-ib)$, $\beta' = \beta$ which in the $z,\bar{z}$ notation reads: $z_1 = a'+ib' = z+2\beta\xi$, $\bar{z}' = a-ib = \bar{z}$, $\Theta' = \Theta+\bar{z}\xi$, $\beta' = \beta$.

We can write as in the vector case

$$\begin{aligned} F(Q') = F(S^{-1}QS) &= F_0(z_1,\bar{z}') + F_1(z_1,\bar{z}')\Theta' \\ &\quad + F_2(z_1,\bar{z}')\beta' + F_3(z_1,\bar{z}')\Theta'\beta' \end{aligned} \tag{3.24}$$

and get

$$F(Q') = F(Q) + \bar{z}F_1(z,\bar{z})\xi + \bar{z}F_3(z,\bar{z})\xi\beta - 2\frac{\partial F_0}{\partial z}\xi\beta - 2\frac{\partial F_1}{\partial z}\xi\eta\Theta. \qquad (3.25)$$

We introduce the following.

Definition 3.2: We say that the regular function $F = F(Q)$ of the supermatrix Q is S invariant if

$$F(S^{-1}QS) = F(Q) \qquad (3.26)$$

for S given by (3.6) with $\xi \neq \beta$. Then we have the following lemma.

Lemma 3.5: Let $F = F(Q)$ be a regular, S invariant function of the supermatrix Q. Then

$$F_3(z,\bar{z}) = \frac{2}{\bar{z}}\frac{\partial F_0}{\partial z}. \qquad (3.27)$$

Proof: Follows from (3.25) and (3.26) by identifying the coefficients of $\xi\beta$.

Remark: The S invariance implies also $F_1 = 0$. The integral theorem in the $z,\bar{z}$ formalism follows.

Lemma 3.6: Let $F = F(Q)$ be as in Lemma 3.4 and having zero boundary condition at infinity. Then

$$\int F(Q)dQ = F_0(0) = F(0), \qquad (3.28)$$

where

$$dQ = \pi i\, dz\, d\Theta\, d\bar{z}\, d\beta = \pi i\, dz\, d\bar{z}\, d\Theta\, d\beta. \qquad (3.28')$$

Proof: Follows from (3.16) and (3.27). A second proof of this Lemma appears in Sec. IV.

Remark: In $dQ = \pi i\, dz\, d\Theta\, d\bar{z}\, d\beta$ we can replace $dz\, d\bar{z} = -2i\, dx\, dy$ to get $dQ = 2\pi\, dx\, dy\, d\Theta\, d\beta$. This is to be compared with $dU\, dV = 4\pi\, dx\, dy\, d\Theta\, d\beta$ in the vector case. As in the vector case, Lemma 3.6 can be generalized for functions $F = F(Q)$, where the matrix elements of Q are generators but no longer canonical (i.e., for the case of matrices Q where a,b are commutative and Θ,β noncommutative "variables"). We do not repeat the arguments which in this case do not differ from the vector case studied above but only give the final result in form of the following.

Lemma 3.7: Let $F = F(Q)$ be a regular function of the 2×2 supermatrix $Q = \begin{pmatrix} a & \Theta \\ \beta & ib \end{pmatrix}$, where a,b,Θ,β are generators of $\Lambda_{2,2}$. Suppose that F has zero boundary conditions at infinity and is invariant w.r.t. S given by (3.6) with $\xi\beta \neq 0$ in Λ_q, $q > 2$, which contains the two canonical generators of $\Lambda_{2,2}$. In formulating the S invariance F has to be considered in the sense of Sec. II as a function of commutative and noncommutative variables with values in $\Lambda_{2,q}$. Then

$$\int F(Q)dQ = F_0(0) = F(0), \qquad (3.29)$$

where the integral in (3.29) is considered in the sense of Berezin and the meaning of dQ is obvious from the discussion above.

Remarks: (i) Again the transformation $Q' = S^{-1}QS$ preserves the reality of the zeroth-order terms in a and b (i.e., the bodies of transformed a',b' stay real). This reduces the proof of Lemma 3.7 to the proof of Lemma 3.6.

(ii) The matrix S used to formulate the S invariance can be taken of the form (3.12) with $\xi\Theta \neq 0$. It is interesting to remark that in this case the S invariance of $F(Q)$ gives the *same relation* (3.27) *in contradistinction* to the vector case in which invariance w.r.t. S given by (3.6) and (3.12) induced the special form (3.18). As an example we mention the superdeterminant $F(Q) = S\det Q$ which is S invariant w.r.t. all invertible 2×2 supermatrices S but cannot be written as a function of $|z|$. Indeed we have for $b \neq 0$

$$S\det Q = \frac{z+\bar{z}}{z-\bar{z}}\left(1 - \frac{4}{z^2-\bar{z}^2}\Theta\beta\right). \qquad (3.30)$$

Generally all we can say is that a function $F = F(Q)$ of Q, which is invariant w.r.t. S is given by (3.6) and (3.12) with $\xi \neq \beta$, Θ, has the form

$$F(Q) = F_0(z,\bar{z}) + \frac{2}{\bar{z}}\frac{\partial F_0(z,\bar{z})}{\partial z}\theta\beta. \qquad (3.31)$$

For the example (3.30) this equality is obviously satisfied.

(iii) Let us consider a supermatrix S of the form

$$S = \begin{pmatrix} 1 - \frac{1}{2}\xi\eta & -\xi \\ \eta & 1 - \frac{1}{2}\eta\xi \end{pmatrix}, \qquad (3.32)$$

where the inverse is

$$S^{-1} = \begin{pmatrix} 1 - \frac{1}{2}\xi\eta & \xi \\ -\eta & 1 - \frac{1}{2}\eta\xi \end{pmatrix} \qquad (3.32')$$

and ξ,η are canonical generators of Λ_q. Then S invariance of $F(Q)$ with S given by (3.32) implies (3.31) too, as it can be easily verified. We will use this result which is also valid for two-component supervectors in Sec. V.

Before closing this section we comment on the results obtained so far. The reader may have questioned the necessity of introducing the imaginary unit in front of b in the supermatrix Q. This is quite essential for the validity of the formulas (3.28, 29). Leaving out i in front of $b\in\mathbb{R}$ the integral in (3.28, 29) can diverge because the singular set in F_3 (the line $a = b$) is not integrable! The imaginary unit in front of b reduces the singularity to the single point $a = b$. It is amusing that (3.28, 29) follows directly from the Cauchy integral formula in the $z,\bar{z}$ formalism. This provides some hints of the intimate connections between complex and noncommutative extensions of real analysis.

IV. THE MAIN RESULT (PARISI–SOURLAS–WEGNER SUPERSYMMETRIC INTEGRAL FORMULA)

In this section we formulate and prove a general version of the supersymmetric integral formula. It was proposed by Wegner as a generalization of the Parisi–Sourlas result for the case in which the integrand depends both of supervectors *and* supermatrices. Our version is more general than that proposed by Wegner (for a discussion of this point see Sec. V). We call it the Parisi–Sourlas–Wegner supersymmetric integral formula. Let us remark that a particular case of it was used already some time ago by Efetov in his study of the localization problem.[2] We study this formula in two steps: we start with the case of canonical generators (the case used so far in physics) and then extend it to arbitrary ones. The nice point in our proof is that the connection to the Cauchy integral theorem of complex analysis becomes transparent. Before starting we remark that in contrast to the cases studied by Parisi, Sourlas, Efetov, and Wegner we very much relax the invariance condition: instead of some super Lie groups (orthogonal, unitary, etc.) as in the physical work mentioned above we only need invariance w.r.t. a *discrete* group isomorphic to the group $\mathbb{Z}^n$ of integers.

Let $U_1,V_1,U_2,V_2,...,U_m,V_m$ be m pairs of $r+r=2r$ component supervectors of the form

$$U_i = (z_1^{(i)},...,z_r^{(i)},\rho_1^{(i)},...,\rho_r^{(i)}), \quad V_i = \begin{pmatrix} \bar{z}_1^{(i)} \\ \vdots \\ \bar{z}_r^{(i)} \\ \sigma_1^{(i)} \\ \vdots \\ \sigma_r^{(i)} \end{pmatrix}, \tag{4.1}$$

where $\bar{z}_j^{(i)} = x_j^{(i)} - iy_j^{(i)}$ is the complex conjugate of $z_j^{(i)} = x_j^{(i)} + iy_j^{(i)}$, $x_j^{(i)}, y_j^{(i)} \in \mathbb{R}$, $i=1,....,m$; $j=1,...,r$ and $\rho_1^{(i)},...,\rho_r^{(i)}$, $\sigma_r^{(i)},...,\sigma_r^{(i)}$ are canonical generators of the Grassmann algebra Λ_{2r}. In Sec. III we studied the case $r=1=m$. Let $U=(U_i)$, $V=(V_i)$, and $F=F(U,V)$ be a function of (U,V) in the sense of Sec. II. For a discussion of this point the reader is referred to Sec. III where the case $r=1$ was considered in detail. No new aspects appear for $r>1$. Here, $F=F(U,V)$ is an element of $\Lambda_{2r,2r}$ that defines a function with values in $\Lambda_{2r,q}$, $q>2r$ ($\Lambda_{2r,2r} \subset \Lambda_{2r,2q}$).

Let $\xi_1,...,\xi_r$ be new canonical generators, different from $\sigma_1,...,\sigma_r$. For $i=1,...,r$ let

$$X_i = \xi_i E_{i,r-i+1},$$

i.e., $(X_i)_{jk} = \delta_{ij}\delta_{r-i+1,k}\xi_i$,

and

$$S_i = \begin{pmatrix} I_r & X_i \\ 0 & I_r \end{pmatrix}. \tag{4.2}$$

It is immediate that for all $i,j \leqslant r$,

$$S_i S_j = \begin{pmatrix} I_r & X_i + X_j \\ 0 & I_r \end{pmatrix}$$

and

$$S_i^{-1} = \begin{pmatrix} I_r & -X_i \\ 0 & I_r \end{pmatrix}. \tag{4.3}$$

Hence the group G generated by $S_1,...,S_r$ is isomorphic to $\mathbb{Z}^r$.

We are going to generalize first the results of Sec. III to the case $r>1$. We say that the regular function $F=F(U,V)$, $U=(U_i)$, $V=(V_i)$, $i=1,...,m$ is G invariant if

$$F(US^{-1},SV) = F(U,V) \tag{4.4}$$

for all $S \in G$ (for a more general case see Def. 4.1).

Here $US^{-1} = (U_i S^{-1})$, $SV = (SV_i)$. Then we have the following.

Lemma 4.1: Let $F=F(U,V)$ be a regular G invariant function of supervectors $U=(U_i)$, $V=(V_i)$, $i=1,...,m$. Suppose that F satisfies the zero boundary condition at infinity. Then

$$\int F(U,V)\,dU\,dV = F_0(0,0) = F(0,0), \tag{4.5}$$

where $dU\,dV = \Pi_{i=1}^m\, dU_i\, dV_i$ and $dU_i\, dV_i$ are as in Lemma 3.2.

Proof: In the first step we consider the case $m=1$ and r arbitrary. We write

$$\int F(U,V)\,dU\,dV = \int \left[\int F(U,V) \prod_{j=2}^{r} (2\pi i\, dz_j\, d\bar{z}_j) \times \prod_{j=1}^{r-1} (d\rho_j\, d\sigma_j) \right] 2\pi i\, dz_1\, d\bar{z}_1\, d\rho_r\, d\sigma_r. \tag{4.6}$$

We have

$$\int F(U,V) \prod_{j=2}^{r} (2\pi i\, dz_j\, d\bar{z}_j) \prod_{j=1}^{r-1} (d\rho_j\, d\sigma_j) = \int F(US_2^{-1},S_2V) \prod_{j=2}^{r} (2\pi i\, dz_j\, d\bar{z}_j) \prod_{j=1}^{r-1} (d\rho_j\, d\sigma_j), \tag{4.7}$$

where $S = S_1 S_2 = S_2 S_1$ and

$$S_1 = \begin{pmatrix} 1 & & \xi_1 \\ & \ddots & \\ & & 1 \end{pmatrix}. \tag{4.8}$$

The proof of (4.7) is omitted because a similar proof appears below.

This shows that

$$\int F(U,V) \prod_{j=2}^{r} (2\pi i\, dz_j\, d\bar{z}_j) \prod_{j=1}^{r-1} (d\rho_j\, d\sigma_j)$$

as a function of (z_1,ρ_r), $\binom{\bar{z}_1}{\sigma_r}$ is S_1 invariant and has zero boundary condition at infinity. Application of Lemma 3.2 reduces the number of components in the vectors U and V by one. Repeated application of the same argument proves the result for $m=1$ and r arbitrary.

The next step in the proof of (4.5) is to consider the general case

$$\int F(UV)\,dU\,dV = \int F(U_1,V_1,...,U_m,V_m) \prod_{i=1}^{m} dU_i\, dV_i \tag{4.9}$$

where F is regular and G invariant:

$$F(U_1S^{-1},SV_1,U_2S^{-1},SV_2,...,U_mS^{-1},SV_m) = F(U_1,V_1,U_2,V_2,...,U_m,V_m) \tag{4.10}$$

for all $S \in G$ with zero boundary condition at infinity. The proof of (4.9) goes by induction in m. Let us denote

$$H(U_1,V_1) = \int F(U_1,V_1,U_2,V_2,...,U_m,V_m) \prod_{i=2}^{m} dU_i\, dV_i. \tag{4.11}$$

We prove that the function $H(U_1,V_1)$ is G invariant. The idea is to take a change of variables on the right-hand side of (4.11) as

$$U_i' = U_i S^{-1}, \quad V_i' = SV_i, \quad i=2,...,m, \tag{4.12}$$

and to apply the Berezin's change of variables formula in the integral (4.9). This is not entirely trivial because the generators in S enter the change of variable formula as (noncommutative) parameters. But this situation was discussed in Sec. II [see (2.15) and (2.16)].

We only have to compute the superdeterminant (Berezinean) of the transformation (4.12). This can be done by elementary arguments but there is a very simple way seeing that the Berezinean Δ equals unity. Indeed, let

$$U=(z_1,...,z_r,\rho_1,...,\rho_r),\quad V=\begin{pmatrix}\bar z_1\\ \vdots\\ \bar z_r\\ \sigma_1\\ \vdots\\ \sigma_r\end{pmatrix}$$

be two supervectors. Then we have

$$\int e^{-\mathrm{Str}\,VU}\,dU\,dV=1. \tag{4.12'}$$

The Berezinean Δ of the linear transformation $U'=US^{-1}$, $V'=SV^{-1}$ does not depend on $z_i,\bar z_i,\rho_i,\sigma_i$, $i=i,...,r$. On the other hand, the integrand is obviously S invariant such that $\Delta=1$. This shows that

$$\begin{aligned}H(U_1S^{-1},SV_1)&=\int F(U_1S^{-1},SV_1,U_2,V_2,...,U_m,V_m)\\&\quad\times\prod_{i=2}^m dU_i\,dV_i\\&=\int F(U_1S^{-1},SV_1,U_2S^{-1},SV_2,...,U_mS^{-1},SV_m)\\&\quad\times\prod_{i=i}^m dU_i\,dV_i=H(U_1,V_1),\end{aligned} \tag{4.13}$$

i.e., $H(U_1,V_1)$ is G invariant (and obviously has zero boundary condition at infinity). Now the previous result and induction in m proves the result.

We will study now the case of supermatrices of the form

$$Q=\begin{pmatrix}A & B\\ C & iD\end{pmatrix} \tag{4.14}$$

where A and D are usual Hermitian $r\times r$ matrices and $B=\Theta$, $C=\beta$ have odd canonical generators Θ_{ij},β_{ij}, $i,j=1,...,r$ as matrix elements. The total number of canonical generators in (4.14) is $4r^2$, where $2r^2$ of them are even and $2r^2$ odd.

Let $F=F(Q)$ be a function of the supermatrix in the sense of Sec. II, $F\in\Lambda_{2r^2,2r^2}$, with values in $\Lambda_{2r^2,q}$ where $q>2r^2$, $\Lambda_{2r^2,2r^2}\subset\Lambda_{2r^2,q}$. We say that the regular function $F=F(Q)$ is G invariant if

$$F(S^{-1}QS)=F(Q) \tag{4.15}$$

for all $S\in G$ with $\xi_i\neq\beta_{i1}$, $i=1,2,...,r$. We will prove the integral theorem for matrices (see below) by using a nice inductive way suggested by Wegner.[3] To understand this idea let us write the supermatrix Q explicitly:

$$Q=\left(\begin{array}{c|c|c} a_{11} & \boxed{a_{12}\cdots a_{1r}\quad \Theta_{11}\cdots\Theta_{1,r-1}} & \Theta_{1r}\\ a_{21} & a_{22}\cdot\cdot\cdot\cdot\cdot\cdot\cdot\cdot\Theta_{2,r-1} & \Theta_{2r}\\ \vdots & & \vdots\\ a_{r1} & & \Theta_{rr}\\ \beta_{11} & & id_{rr}\\ \vdots & & \vdots\\ \beta_{r-1,1} & \beta_{r-1,2}\cdots id_{r-1,r-1} & id_{r-i,r}\\ \beta_{r1} & \boxed{\beta_{r2}\cdots\beta_{rr}\quad id_{r1}\cdots id_{r,r-1}} & \Theta_{1r}\end{array}\right). \tag{4.16}$$

This suggests for Q the decomposition

$$Q=Q_{r,r}=\begin{pmatrix}a_{11} & U & \Theta_{ir}\\ V & Q_{r-2,r-2} & \tilde V\\ \beta_{r1} & \tilde U & id_{rr}\end{pmatrix}, \tag{4.17}$$

where

$$U=(a_{12}\cdots a_{1r},\Theta_{11}\cdots\Theta_{1,r-1}),\quad V=\begin{pmatrix}\bar a_{12}\\ \vdots\\ \bar a_{1r}\\ \beta_{11}\\ \vdots\\ \beta_{r-11}\end{pmatrix}, \tag{4.18}$$

and

$$\tilde V=\begin{pmatrix}\Theta_{2r}\\ \vdots\\ \Theta_{rr}\\ id_{1r}\\ \vdots\\ id_{r-1,r}\end{pmatrix},\quad \tilde U=(\beta_{r2}\cdots\beta_{rr}i\bar d_{1r}\cdots i\bar d_{r-1,r}) \tag{4.19}$$

are supervectors and

$$Q_{r-2,r-2}=\begin{pmatrix}a_{22}\cdots a_{2r} & \Theta_{21}\cdots\Theta_{2,r-1}\\ \cdots & \cdots\\ a_{r2}\cdots a_{rr} & \Theta_{r1}\cdots\Theta_{r,r-1}\\ \beta_{12}\cdots\beta_{1r} & id_{11}\cdots id_{1,r-1}\\ \cdots & \cdots\\ \beta_{r-1,2}\cdots\beta_{r-1,r} & id_{r-1,1}\cdots id_{r-1,r-1}\end{pmatrix} \tag{4.20}$$

is a supermatrix. We remark that the supervectors $\tilde U,\tilde V$ in (4.19) have the even and odd elements interchanged.

Now for the function $F=F(Q)$ in (4.15) we can write

$$F=F(Q)=F(U,V,\tilde U,\tilde V,Q_1,Q_2), \tag{4.21}$$

where

$$Q_2=Q_{r-2,r-2}\quad\text{and}\quad Q_1=\begin{pmatrix}a_{11} & \Theta_{1r}\\ \beta_{r_1} & id_{rr}\end{pmatrix}.$$

In this way a function of a supermatrix Q appears as a function of supervectors and supermatrices of lower order. This decomposition will be used in the inductive procedure for proving the integral theorem for matrices. Let S_1 be given by (4.8). Then

$$F(S^{-1}QS) = F(US_2^{-1},S_2V,\tilde{U}S_2^{-1},S_2\tilde{V},\\ S_1^{-1}Q_1S_1,S_2^{-1}Q_2S_2), \tag{4.22}$$

where $S = S_1S_2$.

Now we are in the position of proving the following.

Lemma 4.2: Let $F = F(Q)$ be a regular, G-invariant function of the supermatrix Q given by (4.14). Suppose that F satisfies the zero boundary condition at infinity. Then

$$\int F(Q)dQ = F_0(0) = F(0), \tag{4.23}$$

where dQ is inductively defined by (3.28′), (4.5), and $dQ = dQ_2\, dQ_1\, d\tilde{U}\, d\tilde{V}\, dU\, dV$.

Proof: The idea of the proof is at hand. Like in the proof of Lemma 4.1 we integrate out first the vectors $U,V,\tilde{U},\tilde{V}$, then the matrix Q_1. This is the inductive step we need. Besides the result on Berezinean of the change of variables for vectors which was used in Lemma 4.1 we need also a second result concerning the Berezinean Δ of the transformation $Q^1 = S^{-1}QS$ with S given by (4.2). An easy way to see that $\Delta = 1$ is to use the relation

$$\int e^{-\operatorname{Str} Q^2} dQ = 1, \tag{4.24}$$

which can be proved directly and proceed like in the proof of Lemma 4.1.

Putting together Lemmas 4.1 and 4.2 we get an integral theorem concerning invariant, regular functions of supervectors and supermatrices with canonical generators as elements. We skip it here because we are going to formulate now a more general case in which the generators are no longer canonical. The next theorem will be the main result of this paper.

Let

$$U_i = (z_1^{(i)},...,z_r^{(i)},\rho_1^{(i)},...,\rho_r^{(i)}), \quad V = \begin{pmatrix} \bar{z}_1^{(i)} \\ \vdots \\ \bar{z}_r^{(i)} \\ \sigma_1^{(i)} \\ \vdots \\ \sigma_r^{(i)} \end{pmatrix}, \quad i = 1,...,m, \tag{4.25}$$

be m pairs of supervectors where $z_j^{(i)} = x_j^{(i)} + iy_j^{(i)}$, $\bar{z}^{(i)} = x_j^{(i)} - iy_j^{(i)}$ and let

$$Q = \begin{pmatrix} A^{(i)} & B^{(i)} \\ C^{(i)} & iD^{(i)} \end{pmatrix}, \quad i = 1,...,n \tag{4.26}$$

be n supermatrices with matrix elements $A^{(i)} = (a_{jk}^{(i)})$, $B^{(i)} = (\Theta_{jk}^{(i)})$, $C^{(i)} = (\beta_{jk}^{(i)})$, $D^{(i)} = (d_{jk}^{(i)})$, $j,k = 1,...,r$ where $a_{jk}^{(i)} = x_{jk}^{(i)} + iy_{jk}^{(i)}$, for $j<k$ and $a_{kj}^{(i)} = x_{kj}^{(i)} - iy_{kj}^{(i)}$ for $k>j$ and similar relations for $D^{(i)}$: $d_{jk}^{(i)} = u_{jk}^{(i)} + i\Theta_{jk}^{(i)}$, $d_{kj}^{(i)} = u_{kj}^{(i)} - iV_{kj}^{(i)}$ for $j\leqslant k$. Assume that the components of U_i, V_i and the matrix elements of Q_i are generators of an algebra $\Lambda_{2mr+2nr^2,2mr+2nr^2}$.

Let $F = F(U_1,V_1,...,U_m,V_m,Q_1,...,Q_n)$ be a function of the m pairs of supervectors U_i and V_i, and of n supermatrices Q_l. Now suppose that there is an algebra $\Lambda_{2mr+2nr^2,q}$ with $q>2mr+2nr^2$ (odd generators) large enough and such that $\Lambda_{2mr+2nr^2,2mr+2nr^2} \subset \Lambda_{2mr+2nr^2,q}$. Suppose that F belongs to $\Lambda_{2mr+2nr^2,2mr+2nr^2}$. According to arguments in Sec. II we can interpret F as a function of commutative and noncommutative "variables" in $\Lambda_{2mr+2nr^2,q}$. We will need this interpretation in order to define the G invariance of F.

Definition 4.1: We say that the regular function $F = F(U_i,V_i,Q_l)$, $i = 1,...,m$, $1 = 1,...,n$ is G invariant if

$$F(U_iS^{-1},SV_i,S^{-1}Q_lS) = F(U_i,V_i,Q_l) \tag{4.27}$$

for all $S\in G$, where S is given by (4.2) with matrix elements ξ_j, $j = 1,...,r$ being odd elements in $\Lambda_{2mr+2nr^2,q}$ such that $\xi_j\sigma_j \neq 0$ and $\xi_j\beta_{jj} \neq 0, j = 1,...,r$.

Following the representation (4.16) we decompose the supermatrices Q_l inductively as in (4.17). This allows us to define dQ_l, $l = 1,...,n$ as in Lemma 4.2 but now for the general case of arbitrary generators. The same remark concerns $dU_i dV_i$, $i = 1,...,m$. Now we have the following.

Theorem 4.1: Let $F = F(U_1,V,...,U_m,V_m,Q_1,...,Q_n)$ be a function of m pairs of supervectors U_i,V_i and n supermatrices Q_l with zero boundary condition at infinity. The function F has to be considered as a function of commutative and noncommutative variables in $\Lambda_{2mr+2nr^2+r^2,q} \supset \Lambda_{2mr+2nr^2,2mr+2nr^2}$ where $q \geqslant 2mr + 2nr^2 + r$ in the sense of Sec II. If F is G invariant we have the supersymmetric integral formula

$$\int F(U_1,V_1,...,U_m,V_m,Q_1,...,Q_l)\\ \times dQ_1,...,dQ_l\, dU_1\, dV_1,...,dU_m\, dV_m\\ = F(0,0,...,0,0,0,...,0) \equiv F(0). \tag{4.28}$$

Proof: The proof follows from the chain of lemmas presented in Secs. III and IV by integrating out first the supervectors and then inductively the supermatrices. In the last step the decomposition (4.18) of supermatrices, proposed by Wegner, is essential.

Remarks: (i) In the transformations $U^1 = US^{-1}$, $V^1 = SV, Q^1 = S^{-1}QS$ the zero-order terms (bodies) of the even generators preserves the complex conjugation.

(ii) By induction, the proof of Theorem 4.1 was in fact reduced to the corresponding result for a pair of two-component supervectors and a 2×2 supermatrix, respectively. In this particular case we have shown in Sec. II that the integral theorem follows via the $z,\bar{z}$ formalism from the Cauchy formula.

(iii) The G invariance used in Theorem 4.1 is a minimal invariance condition involving only the discrete group $\mathbb{Z}^r$.

(iv) Certainly we can replace S given by (4.2) in Theorem 4.1 by

$$S_1 = \begin{pmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 1 & & & & \\ & & \xi_r & 1 & & & \\ & \iddots & & & \ddots & & \\ & & & & & 1 & \\ \xi_1 & & & & & & 1 \end{pmatrix} \tag{4.29}$$

or by other variants

$$S=\begin{pmatrix}1 & \xi\\ 0 & 1\end{pmatrix},\quad S=\begin{pmatrix}1 & 0\\ \xi & 1\end{pmatrix},\quad \xi=\begin{pmatrix}\xi_1 & & \\ & \ddots & \\ & & \xi_r\end{pmatrix}. \tag{4.30}$$

The form (4.30) will be needed in Sec. V.

In the next section we study special cases involving invariance under a Lie supergroup (orthogonal, unitary, etc.) which were used in physics (see for instance Ref. 6).

V. SPECIAL CASES

The new result proved in Sec. IV shows that the invariance condition under which the supersymmetric integral formula applies is very weak. Indeed invariance under the discrete groups $\mathbb{Z}$ or $\mathbb{Z}^n$ is sufficient. Certainly some strong invariance, as for instance invariance under a Lie supergroup, can also imply integral formulas. The simplest case of this type appears in the remark (iii) after Lemma 3.7 of Sec. III. We will study here only the case of the unitary supergroups. For defining these supergroups we have to introduce first an involution in our Grassmann algebras. We follow Ref. 8.

Let $\Lambda = \Lambda_q$ be a Grassmann algebra with the usual decomposition into even and odd parts:

$$\Lambda = {}^0\Lambda \oplus {}^1\Lambda. \tag{5.1}$$

As is well known, it is possible to introduce in Λ two kinds of adjoint operations which extend the complex conjugation in $\mathbb{C}$. The adjoint of the *first kind* is an involution $\alpha \to \alpha^*$ on Λ defined by

a_1) $({}^i\Lambda)^* \subset {}^i\Lambda$, $i=0,1$,

b_1) $(\alpha\beta)^* = \beta^*\alpha^*$ for all $\alpha,\beta \in \Lambda$,

c_1) $\alpha^{**} = \alpha$ for all $\alpha \in \Lambda$.

The adjoint of the *second kind* is an antiinvolution $\alpha \to \alpha^*$ on Λ defined by

a_2) $({}^i\Lambda)^* \subset {}^i\Lambda$, $i=0,1$,

b_2) $(\alpha\beta)^* = \alpha^*\beta^*$ for all $\alpha,\beta \in \Lambda$,

c_2) $\alpha^{**} = (-1)^\gamma \alpha$ for all $\alpha \in {}^\gamma\Lambda$, $\gamma \in \{0,1\}$.

An involution of second kind can exist only if q in $\Lambda = \Lambda_q$ is an even integer. We will use in this section only the involution of the second kind although all results will be valid also for the involution of the first kind. We remark that in a pair of supervectors U,V or in a supermatrix the number of generators is even so that there is no problem in working with the second kind involution.

Let us define the unitary supergroup $\mathrm{UPL}_2(n,m)$ of supermatrices of the second kind (i.e., second kind involution is used denoted as index 2 on UPL).

Let

$$A=\begin{pmatrix} a & \xi\\ \eta & b\end{pmatrix}, \tag{5.2}$$

be a supermatrix over Λ_q, and let by definition

$$A^{\#}=\begin{pmatrix} {}^t a^* & -{}^t\eta^*\\ {}^t\xi^* & {}^t b^*\end{pmatrix}, \tag{5.3}$$

where $a \mapsto {}^t a$ is the usual transposition of matrices. Then we define

$$\mathrm{UPL}_2(n,m) = \{U \in \mathrm{PL}(n,m) \mid U^{\#} = U^{-1}\}. \tag{5.4}$$

In (5.4) $\mathrm{PL}(n,m)$ is the general linear graded Lie group of invertible supermatrices. Note that a supermatrix A as in (5.2) is invertible if a and b are invertible matrices.

In this section we will consider functions F of r supervectors and $r\times r$ supermatrices which are invariant w.r.t. the unitary supergroup $\mathrm{UPL}_2(r,r)$.

We will show that the supersymmetric integral theorem is valid for functions F that satisfy a zero boundary condition at infinity and are invariant w.r.t. the supergroup UPL_2. Certainly a result of this type is weaker than the result of Theorem 4.1 in which the invariance condition is imposed by means of a discrete group. Nevertheless we study this case because it appears in physical applications Ref. 6. A similar result is valid for the case of invariance with respect to other classical Lie supergroups (e.g., orthogonal). Before proving this result we need a general formula for the elements of UPL_2. This will be done in the following.

Lemma 5.1: For all $n,m \geqslant 1$ we have

$$\mathrm{UPL}_2(n,m) = \left\{ \begin{pmatrix} u & 0\\ 0 & v\end{pmatrix} \begin{pmatrix} (I_n+\zeta\zeta^+)^{-1/2} & -(I_n+\zeta\zeta^+)^{-1/2}\zeta\\ (I_m+\zeta^+\zeta)^{-1/2}\zeta^+ & (I_m+\zeta^+\zeta)^{-1/2}\end{pmatrix} \;\middle|\; \begin{pmatrix} u & 0\\ 0 & v\end{pmatrix}\text{-superunitary and } \zeta\text{-}(n,m) \text{ matrix with odd elements} \right\}, \tag{5.5}$$

where $a^+ = {}^t a^*$, $\zeta^+ = {}^t\zeta^*$, etc.

Proof: We give an elementary proof based only on the definition of UPL_2. Indeed from (5.4) we get

$$\begin{pmatrix} I_n & 0\\ 0 & I_m\end{pmatrix} = \begin{pmatrix} a^+ & -\eta^+\\ \xi^+ & b^+\end{pmatrix}\begin{pmatrix} a & \xi\\ \eta & b\end{pmatrix} = \begin{pmatrix} a^+a-\eta^+\eta & a^+\xi-\eta^+b\\ \xi^+a+b^+\eta & \xi^+\xi+b^+b\end{pmatrix}. \tag{5.6}$$

Equation (5.6) is equivalent to the following set of equations:

$$a^+a-\eta^+\eta = I_n, \tag{5.7}$$

$$a^+\xi-\eta^+b=0, \tag{5.8a}$$

$$\xi^+a+b^+\eta=0, \tag{5.8b}$$

$$\xi^+\xi+b^+b=I_m. \tag{5.9}$$

We get from (5.8a)

$$\eta^+ = a^+\xi b^{-1}. \qquad (5.10)$$

We use $a^{++} = a$ (for a even) an $\eta^{++} = -\eta$ (for η odd) in order to write

$$\eta = -(b^{-1})^+\xi^+ a. \qquad (5.11)$$

Introducing (5.10) and (5.11) in (5.7) gives

$$I_n = a^+a - \eta^+\eta = a^+a + a^+\xi b^{-1}(b^{-1})^+\xi^+ a$$
$$= a^+(I_n + \xi b^{-1}(\xi b^{-1})^+)a.$$

Let us use the shorthand notation

$$\zeta = \xi b^{-1}. \qquad (5.12)$$

Then

$$I_n = a^+(I_n + \zeta\zeta^+)a. \qquad (5.13)$$

Now observe that ζ has odd entries, hence in case $m\cdot n > 1$ we have

$$(\zeta\zeta^+)^+ = -\zeta^{++}\zeta^+ = \zeta\zeta^+, \qquad (5.14)$$

and similarly

$$(\zeta^+\zeta)^+ = \zeta^+\zeta. \qquad (5.15)$$

[In case $m = n = 1$ we simply have

$$(\zeta\zeta^*)^* = \zeta^*\zeta^{**} = -\zeta^*\zeta = \zeta\zeta^*$$

and $(\zeta^*\zeta)^* = (\zeta^*\zeta)$.]

Moreover we can write the square root of $I_n + \zeta\zeta^+$ as the (finite!) sum

$$(I_n + \zeta\zeta^+)^{1/2} = I_n + \frac{1}{2}\zeta\zeta^+ - \frac{1}{2\cdot 4}(\zeta\zeta^+)^2 + \cdots.$$

Due to (5.14) we obtain

$$((I_n + \zeta\zeta^+)^{1/2})^+ = (I_n + \zeta\zeta^+)^{1/2}.$$

This enables us to rewrite (5.13) as

$$I_n = a^+(I_n + \zeta\zeta^+)a$$
$$= a^+(I_n + \zeta\zeta^+)^{1/2}(I_n + \zeta\zeta^+)^{1/2}a$$
$$= ((I_n + \zeta\zeta^+)^{1/2}a)^+(I_n + \zeta\zeta^+)^{1/2}a.$$

This shows that the even matrix

$$a_\zeta = (I_n + \zeta\zeta^+)^{1/2}a \qquad (5.16)$$

is *unitary* with respect to the adjoint operation $x \mapsto x^+$. The very same reasoning shows that also

$$b_\zeta = (I_m + \zeta^+\zeta)^{1/2}b \qquad (5.17)$$

is unitary with respect to $x \mapsto x^+$. By (5.11) and the definition of ζ we have

$$\eta a^{-1} = -\zeta^+. \qquad (5.18)$$

Equations (5.16)–(5.18) show that we have the following product decomposition of the unitary supermatrix $\left(\begin{smallmatrix} a & \xi \\ \eta & b\end{smallmatrix}\right)$:

$$\begin{pmatrix} a & \xi \\ \eta & b \end{pmatrix} = \begin{pmatrix} (I_n + \zeta\zeta^+)^{-1/2} & \zeta(I_m + \zeta^+\zeta)^{-1/2} \\ -\zeta^+(I_n + \zeta\zeta^+)^{-1/2} & (I_m + \zeta^+\zeta)^{-1/2} \end{pmatrix} \times \begin{pmatrix} a_\zeta & 0 \\ 0 & b_\zeta \end{pmatrix}. \qquad (5.19)$$

Application of # gives

$$\begin{pmatrix} a & \xi \\ \eta & b \end{pmatrix}^{\#} = \begin{pmatrix} a_\zeta^+ & 0 \\ 0 & b_\zeta^+ \end{pmatrix} \begin{pmatrix} (I_n + \zeta\zeta^+)^{-1/2} & -(I_n + \zeta\zeta^+)^{-1/2}\zeta \\ (I_m + \zeta^+\zeta)^{-1/2}\zeta^+ & (I_m + \zeta^+\zeta)^{-1/2} \end{pmatrix}. \qquad (5.20)$$

It is a simple calculation to show that conversely every supermatrix of the form

$$\begin{pmatrix} u & 0 \\ 0 & v \end{pmatrix} \begin{pmatrix} (I_n + \zeta\zeta^+)^{-1/2} & -(I_n + \zeta\zeta^+)^{-1/2}\zeta \\ (I_m + \zeta^+\zeta)^{-1/2}\zeta^+ & (I_m + \zeta^+\zeta)^{-1/2} \end{pmatrix},$$

where $\left(\begin{smallmatrix} u & 0 \\ 0 & v\end{smallmatrix}\right)$ superunitary and ζ an (n,m) matrix with odd entries belongs to $\mathrm{UPL}_2(n,m)$. This proves the lemma.

Remarks: (i) Let Λ_q be the Grassmann algebra of a q dimensional complex vector space: $\Lambda = \Lambda(V)$. One can show that conjugations of the first kind are in one-to-one correspondence with semilinear mappings ψ: $V \to V$ (i.e., ψ is additive and $\psi(\lambda v) = \bar{\lambda}\psi(v)$ for all $\lambda \in \mathbb{C}$, $v \in V$) such that $\psi^2 = id_V$. Similarly conjugations of the second kind are in one-to-one correspondence with semilinear mappings φ: $V \to V$ such that $\varphi^2 = -id_V$, i.e., with quaternionic structures on V. Therefore conjugations of the first kind exist for all $q \in \mathbb{N}$, conjugations of the second kind exist for all *even* $q \in \mathbb{N}$.

(ii) Obviously one can prove in the same way the very analogue of Lemma 5.1 also for conjugations of the first kind. For unitary supergroups of the first kind a formula similar to (5.5) was obtained by Berezin using the Lie superalgebra of the corresponding supergroups (see Ref. 5, p. 276–277). The precise relation between the two kinds of unitary supergroups will be discussed in a forthcoming paper, answering a question in Ref. 8.

(iii) Now we will spell out a subgroup of UPL_2 needed in the rest of this section. Indeed let $u = v = I_r$ and

$$\zeta = \begin{pmatrix} \zeta_1 & & \\ & \ddots & \\ & & \zeta_r \end{pmatrix}, \zeta^+ = \begin{pmatrix} \zeta_1^* & & \\ & \ddots & \\ & & \zeta_r^* \end{pmatrix}.$$

The matrix S in (5.5) constructed with these u,v,ζ,ζ^+ generates a subgroup of $\mathrm{UPL}_2(r,r)$. For $r = 1$ we have

$$S = \begin{pmatrix} 1 - \frac{1}{2}\zeta_1\zeta_1^+, & -\zeta_1 \\ \zeta_1^+ & 1 + \frac{1}{2}\zeta_1^+\zeta_1 \end{pmatrix}. \qquad (5.21)$$

But this is exactly the supermatrix (3.32) where $\eta = \zeta_1^+$,

$\xi = \zeta_1$ which was used to prove one of the variants of the integral theorem. With these preparations we can state the following result which is a weak form of Theorem 4.1

Theorem 5.1: Let $F = F(U_1, V_1, ..., U_m, V_m, Q_1, ..., Q_n)$ be a function of m pairs of supervectors U_i, V_i and n supermatrices Q_l with zero boundary condition at infinity. The function F has to be considered as depending of commutative and noncommutative variables in the sense of Theorem 4.1. If F is invariant w.r.t. the unitary supergroup UPL(r,r) (of the first or second kind), then the integral formula (4.28) is valid.

Proof: The invariance with respect to the unitary supergroup implies invariance with respect to S in (5.21). The remark (iii) after Lemma 3.7 in Sec. III and the inductive arguments in Sec. IV complete the proof.

Remark: In applications to disordered systems and nuclear physics, the unitary supergroup appears as a group of transformations which preserves the super scalar product.

VI. REMARKS AND CONCLUSIONS

We have extended results by Parisi, Sourlas, Efetov, Wegner, and others concerning integral formulas for supersymmetric invariant functions. In particular we worked with arbitrary generators as variables in the integrand function using throughout Berezin's integration theory.[5] The supersymmetric invariance condition was relaxed from a classical Lie supergroup to a discrete group. In Sec. V we needed the general form of elements of unitary supergroups for which we gave proof without using Lie superalgebras.

In proving the integral formula for supermatrices we used an idea of Wegner. Before finishing we want to remark that a second proof of the integral theorem could be possible by using an ideal of Efetov. We give some details for the case $F(Q)$, where Q is a 2×2 supermatrix.

Notice that Berezin's integration is a delicate matter. This is shown in the following example which we learned from Ref. 7.

Consider the function

$$F(Q) = e^{-\operatorname{Str} Q^2}, \tag{6.1}$$

where Q is given by (3.20). We already know that

$$\textstyle\int F(Q)dQ = 1 \tag{6.2}$$

(using our norming conventions for dQ, otherwise the result is 2π). On the other hand, the supermatrix Q is diagonalizable with eigenvalues

$$\lambda_1 = a + \frac{\Theta\beta}{a - ib}, \quad \lambda_2 = ib + \frac{\Theta\beta}{a - ib}. \tag{6.3}$$

We could try to make a change of variables in (6.2) to λ_1, λ_2 keeping the odd variables unchanged. Because $F(Q) = \exp(-\lambda_1^2 - \lambda_2^2)$ does not depend on Θ or β and because the Berezinean of this transformation equals unity, the formal result we obtain zero instead of one! The contradiction is explained by the singularity at $a = b$ in (6.3). This phenomen is typical for supermatrices and seems to invalidate the idea of changing the variables to the eigenvalues. Nevertheless we can apply a slightly modified change of variables which removes the singularity at $a = b$.

We introduce new variables $\varkappa_1$, $\varkappa_2$, ξ, η by

$$\begin{aligned} a &= \varkappa_1 + \xi\eta(\varkappa_1 - i\varkappa_2), \\ ib &= i\varkappa_2 + \xi\eta(\varkappa_1 - i\varkappa_2), \\ \Theta &= -\xi(\varkappa_1 - i\varkappa_2), \\ \beta &= \eta(\varkappa_1 - i\varkappa_2). \end{aligned} \tag{6.4}$$

The Berezinean of this transformation can be explicitly computed as

$$\Delta = \left| \frac{\partial(a,b,\Theta,\beta)}{\partial(\varkappa_1,\varkappa_2,\xi,\eta)} \right| = \left(\frac{1}{\varkappa_1 - i\varkappa_2} \right)^2. \tag{6.5}$$

The change of variables formula (see Sec. II) gives

$$\int F(Q)dQ = \int d\varkappa_1\, d\varkappa_2\, d\eta\, d\xi \frac{1}{(\varkappa_1 - i\varkappa_2)^2} e^{-(\varkappa_1^2 + \varkappa_2^2)} \tag{6.6}$$

which again equals zero! The contradiction is explained by the fact that the Berezinean $(\varkappa_1 - i\varkappa_2)^{-2}$ is not integrable with respect to $d\varkappa_1\, d\varkappa_2$.

However, the formal computation above provides us with an idea to rigorously prove the integral theorem for 2×2 supermatrices. Indeed, let us consider

$$I_\mu = \int F(Q) e^{-\mu \operatorname{Str} Q^2} dQ, \quad \mu \geqslant 0. \tag{6.7}$$

We want to compute I_0. Differentiating with respect to μ gives

$$\frac{dI_\mu}{d_\mu} = \int (-\operatorname{Str} Q^2) F(Q) e^{-\mu \operatorname{Str} Q^2} dQ. \tag{6.8}$$

Performing the change of variables (6.4) we get

$$\frac{dI_\mu}{d_\mu} = 0 \tag{6.9}$$

and this result is now correct because

$$\operatorname{Str} Q^2 = \varkappa_1^2 + \varkappa_2^2 = (\varkappa_1 + i\varkappa_2)(\varkappa_1 - i\varkappa_2)$$

cancels the nonintegrable singularity of the Berezinean. It follows that

$$I_\mu = \text{const.} = I_0. \tag{6.10}$$

This constant can be found easily by applying the usual Laplace method for $\mu \to \infty$. The expected result $\int F(Q)\, dQ = F(0)$ follows.

It is very appealing to use this idea for a proof of the general case of the integral theorem but this does not seem to be easy. In any case we prefer the proof in Sec. IV giving a more general result (for the diagonalization of supermatrices the discrete group used in Sec. IV is not sufficient).

ACKNOWLEDGMENTS

We are grateful to Professor F. Wegner for a private communication concerning his supersymmetric integral formula which was the starting point of the present work. Several helpful discussions with C. de Paiva and U. Scharffenberger are also acknowledged.

The work of FC was supported in part by the Deutsche Forschungsgemeinschaft.

[1]G. Parisi and G. Sourlas, Phys. Rev. Lett. **43**, 744 (1979), see also A. Klein, "Supersymmetry and a two-dimensional reduction in random pheonomena," in *Proceedings of the Second Workshop on Quantum Proba-*

bility and Applications, Heidelberg, 1984 (Springer Lecture Notes in Mathematics, to appear).
[2]K. B. Efetov, Adv. Phys. **32**, 53 (1983).
[3]F. Wegner, Z. Phys. B **49**, 297 (1983); and (private communication, 1987).
[4]F. Constantinescu, J. Stat. Phys. **50**, 1167 (1988); see also F. Constantinescu, G. Felder, K. Gawedzki, and A. Kupiainen, *ibid.* **48**, 365 (1987).
[5]F. A. Berezin, *Introduction to Superanalysis* (Reidel, Dordrecht, 1987).
[6]J. J. Verbaarshot, H. A. Weidenmüller, and M. R. Zirnbauer, Phys. Rep. **129**, 365 (1985).
[7]M. Zirnbauer, Nucl. Phys. B. **265**, 375 (1986).
[8]V. Rittenberg and M. Scheunert, J. Math. Phys. **19**, 709 (1978).

On linear unitary transformations of two canonical variables

Jerzy F. Plebański[a)]
Centro de Investigacion y de Estudios Avanzados del IPN, Departamento de Fisica, Apdo. Postal 14-740, 07000 Mexico, D. F., Mexico
J. D. Finley, III
Physics and Astronomy Department, University of New Mexico, Albuquerque, New Mexico 87131

(Received 29 December 1987; accepted for publication 28 December 1988)

A useful parametrization of the groups SO(0,3) and SO$^{\uparrow}$(2,1) is presented that has simple, rational composition laws, albeit that it has a (quasigraded) structure. This "tangential" parametrization is also advantageous in providing a rather simple 1-1 "picture" of the elements of the groups in question. As well, the parametrization allows an explicitly finite result for the composition law needed for the (abstract) Baker–Campbell–Hausdorff formula for exponentials of objects formed from the corresponding Lie algebras. This approach, in turn, allows a useful beginning to the problem of the determination of flows on the group manifold, i.e., the determination of analytic curves given the initial direction of the curve.

I. INTRODUCTION

Motivated by a return, after more than 30 years,[1,2] to a study of linear unitary transformations, we would like to present some subtleties in the description of the three-dimensional, real Lie groups SO$(n_+,3-n_+)$ that still deserve some interest and have previously escaped much attention.

Restricting consideration to the connected component containing the identity, these groups can be conceived as the set of all 3×3, real matrices L such that

$$g_{\alpha\beta}L^{\alpha}{}_{\eta}L^{\beta}{}_{\delta}=g_{\eta\delta}, \tag{1.1a}$$

$$\det(L^{\alpha}{}_{\beta}) = +1, \tag{1.1b}$$

where $g_{\alpha\beta}$, with $\alpha,\beta = 1,\ 2,\ 3$, are the components of the metric on the underlying space. There are, however, in three dimensions, only two *essentially different* signatures for such a metric:

$$\|g_{\alpha\beta}\| = \begin{cases} \|\mathrm{diag}(1,1,-1)\|, & \text{hyperbolic,} \\ \|\mathrm{diag}(-1,-1,-1)\|, & \text{elliptic,} \end{cases} \tag{1.2}$$

where the (unusual) choice of the signature for the elliptic case will arrange in *both* cases that

$$\det(g_{\alpha\beta}) = -1. \tag{1.3}$$

This is then equivalent to saying that we study, simultaneously, the two groups,[3] SO(0,3) and SO$^{\uparrow}$(2,1) where, for the latter case, we must also make the restriction

$$L^{3}{}_{3}\geqslant 1, \quad \text{for SO}^{\uparrow}(2,1), \tag{1.4}$$

so as to maintain our discussion with the component containing the identity. It is well known that the generators γ^{α} for the Lie algebras for these two groups may be chosen so as to satisfy the commutation relations (CR)[4]

$$[\gamma^{\alpha},\gamma^{\beta}] = -\epsilon^{\alpha\beta\eta}\gamma_{\eta}, \tag{1.5}$$

where the metric is used to lower the indices

$$\gamma_{\eta} = g_{\eta\delta}\gamma^{\delta}. \tag{1.6}$$

It is the intent of this article to present a parametrization for these two groups that allows a simple formula for the group composition in terms of rational functions, involving numerator and denominator no worse than products of the original parameters. The simplicity of this representation of the composition law will also allow the explicit, finite determination of the composition law for the Baker–Campbell–Hausdorff (BCH) formula, i.e., to find z where $e^{x}e^{y}=e^{z}$, where x and y are linear and homogenous in the generators γ^{α}, taken only as abstract quantities obeying (1.5) and (1.6). A complete systematization of the results of both the group composition law and the (associated) Baker–Campbell–Hausdorff composition law is laid out in Table I, in Sec. IV.

An additional benefit will be a beginning of an analysis of analytic curves on these group manifolds, the starting point for the attack again being the simplicity of the composition law. All benefits, of course, come at some price. In this case, the price comes because the range of the particular parameter space is infinite; therefore, a boundary set must be added with a slightly different form of parametrization, obtained by taking limits. Nonetheless, as will be seen, quite complete characterizations of all the elements of both groups may be given succinctly and simply.

II. THE CANONICAL PARAMETRIZATION OF SO$^{\uparrow}$(2,1) AND SO(0,3)

We first give a particularly useful way of realizing the canonical parametrization. Let

$$x:=x_{\alpha}\gamma^{\alpha}, \quad \{x_{\alpha}\}\in\mathscr{R}^{3} \tag{2.1}$$

and consider for real λ the curve in the Lie algebra defined by

$$\gamma^{\alpha}(\lambda):=e^{\lambda x}\gamma^{\alpha}e^{-\lambda x}, \quad \lambda\in\mathscr{R}. \tag{2.2}$$

Using the commutation relations, one then calculates that

$$\frac{d}{d\lambda}\gamma^{\alpha}(\lambda) = e^{\lambda x}[x,\gamma^{\alpha}]e^{-\lambda x} = x_{\rho}\epsilon^{\alpha\rho}{}_{\beta}\gamma^{\beta}(\lambda):=M^{\alpha}{}_{\beta}\gamma^{\beta}(\lambda), \tag{2.3}$$

where, as in the spinor notation of van der Waerden and Infeld, the dot denotes the original position of the index manipulated by $g_{\alpha\beta}$. By defining

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

 0022-2488/89/050993-07$02.50

TABLE I. Parametrization of group products, viewed from $\mathscr{S}\cup\partial\mathscr{S}$.

Condition	Composition	Effective parametrization	BCH composition
$1+s_\alpha t^\alpha\neq 0$	$L(s^\rho)L(t^\rho)=L(r^\rho)$	$r^\rho=(1+s_\beta t^\beta)^{-1}(s^\rho+t^\rho\epsilon^{\rho\alpha\delta}s_\alpha t_\delta)$	$e^{x[t^\rho]}e^{x[s^\rho]}=e^{x[r^\rho]}$
$1+s_\alpha t^\alpha=0$	$L(s^\rho)L(t^\rho)=L(\bar r^\rho)$	$\bar r^\rho=s^\rho+t^\rho+\epsilon^{\rho\alpha\delta}s_\alpha t_\delta$	$e^{x[t^\rho]}e^{x[s^\rho]}=e^{x[\bar r^\rho]}$
$\bar s_\alpha t^\alpha\neq 0$	$L(\bar s^\rho)L(t^\rho)=L(r^\rho)$	$r^\rho=(\bar s_\beta t^\beta)^{-1}(\bar s^\rho+\epsilon^{\rho\alpha\delta}\bar s_\alpha t_\delta)$	$e^{x[t^\rho]}e^{x[\bar s^\rho]}=e^{x[r^\rho]}$
$\bar s_\alpha t^\alpha=0$	$L(\bar s^\rho)L(t^\rho)=L(\bar r^\rho)$	$\bar r^\rho=\bar s^\rho+\epsilon^{\rho\alpha\delta}\bar s_\alpha t_\delta$	$e^{x[t^\rho]}e^{x[\bar s^\rho]}=e^{x[\bar r^\rho]}$
$s_\alpha\bar t^\alpha\neq 0$	$L(s^\rho)L(\bar t^\rho)=L(r^\rho)$	$r^\rho=(s_\beta\bar t^\beta)^{-1}(\bar t^\rho+\epsilon^{\rho\alpha\delta}s_\alpha\bar t_\delta)$	$e^{x[\bar t^\rho]}e^{x[s^\rho]}=e^{x[r^\rho]}$
$s_\alpha\bar t^\alpha=0$	$L(s^\rho)L(\bar t^\rho)=L(\bar r^\rho)$	$\bar r^\rho=\bar t^\rho+\epsilon^{\rho\alpha\delta}s_\alpha\bar t_\delta$	$e^{x[\bar t^\rho]}e^{x[s^\rho]}=e^{x[\bar r^\rho]}$
$\bar s_\alpha\bar t^\alpha\neq 0$	$L(\bar s^\rho)L(\bar t^\rho)=L(r^\rho)$	$r^\rho=(\bar s_\beta\bar t^\beta)^{-1}\epsilon^{\rho\alpha\delta}\bar s_\alpha\bar t_\delta$	$e^{x[\bar t^\rho]}e^{x[\bar s^\rho]}=e^{x[r^\rho]}$
$\bar s_\alpha\bar t^\alpha=0$	$L(\bar s^\rho)L(\bar t^\rho)=L(\bar r^\rho)$	$\bar r^\rho=\epsilon^{\rho\alpha\delta}\bar s_\alpha\bar t_\delta$	$e^{x[\bar t^\rho]}e^{x[\bar s^\rho]}=e^{x[\bar r^\rho]}$

$$\Delta:=\sqrt{x_\rho x^\rho}=\sqrt{g_{\rho\sigma}x^\rho x^\sigma}\,, \tag{2.4}$$

one easily shows that the matrix

$$M=\|M^\alpha{}_\beta\|=\|x_\rho\epsilon^{\alpha\rho}{}_\beta\| \tag{2.5a}$$

satisfies the Hamilton–Cayley equation

$$M^3-\Delta^2M=0, \tag{2.5b}$$

and consequently has the eigenvalues $(0,\Delta,-\Delta)$.

Taking as boundary condition for (2.3) the obvious $\gamma^\alpha(0)=\gamma^\alpha$, we easily see that the solution to (2.3) has the form

$$\gamma^\alpha(\lambda)=L^\alpha{}_\beta(\lambda)\gamma^\beta, \tag{2.6}$$

where

$$L(\lambda):=\|L^\alpha{}_\beta(\lambda)\|=I+[\sinh(\Delta\lambda)/\Delta]M+[(\cosh(\Delta\lambda)-1)/\Delta^2]M^2, \tag{2.7}$$

while it is also worthwhile to write down explicitly that

$$(M^2)^\alpha{}_\beta=\Delta^2\delta^\alpha{}_\beta-x^\alpha x_\beta. \tag{2.8}$$

By setting

$$L:=L(1)=I+(\sinh\Delta/\Delta)M+[(\cosh\Delta-1)/\Delta^2]M^2, \tag{2.9}$$

or

$$L=\|L^\alpha{}_\beta\|=\|\delta^\alpha{}_\beta+(\sinh\Delta/\Delta)x_\rho\epsilon^{\alpha\rho}{}_\beta+[(\cosh\Delta-1)/\Delta^2](\Delta^2\delta^\alpha{}_\beta-x^\alpha x_\beta)\|,$$

and noting that L shares eigenvectors with M, one easily sees that the eigenvalues of L are $(1,e^\Delta,e^{-\Delta})$, from which we have that

$$\det(L)=+1, \tag{2.10a}$$

$$\mathrm{Tr}(L)=1+2\cosh(\Delta). \tag{2.10b}$$

Notice also that when $\Delta\to 0$, the limit exists and then

$$L|_{\Delta=0}=\|\delta^\alpha{}_\beta+x_\rho\epsilon^{\alpha\rho}{}_\beta-\tfrac{1}{2}x^\alpha x_\beta\|,\quad (L|_{\Delta=0}-I)^3=0. \tag{2.11}$$

As well, insertion of the components of (2.9) into (1.1) tells us that our L satisfies that equation.

We also have from (2.9) that

$$L^3{}_3=1+[(\cosh\Delta-1)/\Delta^2](x_1x^1+x_2x^2). \tag{2.12}$$

So far, all formulas apply parallelly for either the case of the hyperbolic or the elliptic signature, from (1.2). At this point, however, their interpretation bifurcates. To explicate this, we must examine the quantity Δ more closely. First, note that (2.9) defining L is insensitive to the choice of sign of Δ. We therefore choose the sign of the square root so that when

$$x_\alpha x^\alpha=\Delta^2\geqslant 0\to\Delta\geqslant 0 \tag{2.13a}$$

and when

$$x_\alpha x^\alpha=\Delta^2\leqslant 0\to\Delta=+i\theta,\quad\theta\geqslant 0. \tag{2.13b}$$

Therefore for $g_{\alpha\beta}=\mathrm{diag}(1,1,-1)$,

$$L^3{}_3=1+[(\cosh\Delta-1)/\Delta^2][(x_1)^2+(x_2)^2]\geqslant 1, \tag{2.14}$$

so that we have verified the following theorem.

Theorem 1: For $x:=x_\alpha\gamma^\alpha$, $\{x_\alpha\}\in\mathscr{R}^3$, and γ^α are arbitrary quantities satisfying $[\gamma^\alpha,\gamma^\beta]=-(\epsilon^{\alpha\beta\eta}g_{\eta\delta})\gamma^\delta$, the matrices $L=\|L^\alpha{}_\beta\|$, determined by

$$e^x\gamma^\alpha e^{-x}=L^\alpha{}_\beta\gamma^\beta, \tag{2.15}$$

form the elements of the defining (three-dimensional, matrix) representations of the group $SO^\uparrow(2,1)$ or $SO(0,3)$, depending on whether $g_{\gamma\delta}$ has hyperbolic or elliptic signature, respectively, as outlined in (1.2).

Now, the standard interpretation of $SO^\uparrow(2,1)$ transformations, as parametrized by x^α, is this: if x^α is "timelike," $x^\alpha x_\alpha<0\to\Delta=i\theta$, the transformation consists of the usual trigonometric rotation through θ in the plane perpendicular to x^α. If x^α is "null," $x_\alpha x^\alpha=0$, (2.11) applies and we have the odd case of $L-I$ being nilpotent, $(L-I)^3=0$. Of course, if (2.11) holds with $x^\alpha\to 0$, then $L=I$. Finally, if x^α is "spacelike," $x^\alpha x_\alpha>0$, the transformation consists of a hyperbolic rotation through Δ in the plane perpendicualr to x^α.

This standard parametrization has, however, a peculiar property, the manifestation of which is somewhat different for the two groups. For $SO^\uparrow(2,1)$, if the x^α's are such that

$$x_\alpha x^\alpha=x_1^2+x_2^2-x_3^2=-(2\pi n)^2,\quad n=1,2,\ldots\Rightarrow\Delta=2\pi in, \tag{2.16}$$

then $L=\|L^\alpha{}_\beta\|=I$. Thus, for this parametrization, all points of the infinite sequence of hyperboloids

$x_\alpha x^\alpha = -(2\pi n)^2$ induce the identity of the group. An alternative manner of stating this is that

$$x := x_\alpha \gamma^\alpha, \quad x_\alpha x^\alpha = -(2\pi n)^2 \Rightarrow [e^x, \gamma^\alpha] = 0, \quad (2.17)$$

and thus, e^x with x of the nature described above, must be a function of the Casimir operator

$$C := \gamma_\alpha \gamma^\alpha. \quad (2.18)$$

The elliptic case of the signature (-,-,-) is much simpler. There with $\Delta = \sqrt{x_\alpha x^\alpha} = \sqrt{-x_1^2 - x_2^2 - x_3^2}$, $\Delta = i\theta$, for $\theta \geqslant 0$ as noted in (2.13b). This time, however, $\theta = 0 \leftrightarrow x^\alpha = 0$. Notice that presently (2.12) takes the form

$$L^3{}_3 = 1 - [(1-\cos\theta)/\theta^2](x_1^2 + x_2^2), \quad (2.19)$$

which implies $L^3{}_3 \leqslant 1$. With the timelike and null cases presently absent, the transformation consists simply in the trigonometric rotations through θ in the planes perpendicular to x^α. However, these transformations have the peculiar property that

$$-x_\alpha x^\alpha = x_1^2 + x_2^2 + x_2^2 = (2\pi n)^2,$$
$$n = 0,1,2,\ldots \Rightarrow \Delta = 2\pi i n, \quad (2.20)$$

which reduces L, from (2.9), to I. Thus, with the x^α parametrization of SO(0,3) matrices, all points of the infinite sequence of spheres $-x_\alpha x^\alpha = (2\pi n)^2$ induce the identity of the group, and also in the elliptic case the implication (2.17) is valid, so that e^x must be a function of the Casimir operator of the group SO(0,3).

III. THE "TANGENTIAL" PARAMETRIZATION OF SO$^\uparrow$(2,1) AND SO(0,3) MATRICES

The canonical Lie parametrization of the SO$^\uparrow$(2,1) and SO(0,3) matrices suffers two disadvantages: (1) the points $-x_\alpha x^\alpha = (2\pi n)^2$ all induce the identity of the group; *and* (2) the explicit group composition law is rather involved. We will propose now an alternative parametrization—simultaneously for both the groups under discussion—which eliminates the difficulties mentioned above, providing a *rational* parametrization of the matrices of the defining representations of both groups.

There is a price to be paid for this: the new parametrization does not cover uniformly *all* L matrices from (2.9) as parametrized by the real x^α's.

We define new parameters $\{t^\alpha\} \in \mathscr{R}^3$ determined by

$$t^\alpha := [\tanh(\Delta/2)/\Delta] x^\alpha, \quad \Delta := \sqrt{x_\alpha x^\alpha}. \quad (3.1)$$

Problems can arise with the $\mathscr{R}^3 \to \mathscr{R}^3$ mapping when either $\tanh(\Delta/2)/\Delta = 0$, or $\tanh(\Delta/2)/\Delta = \infty$. The case $\Delta = 0$, if we consider the coefficient as $\lim_{\Delta\to 0}(\tanh \Delta/2/\Delta) = \frac{1}{2}$, causes no problems in the case of either signature, given by (1.2). For hyperbolic signature $\Delta \to 0$ reduces (3.1) to $t_\alpha = \frac{1}{2} x_\alpha$, while for the elliptic signature, where $\Delta = 0 \Leftrightarrow x_\alpha = 0$, $\Delta \to 0$ implies $t_\alpha = 0$. When $\Delta \neq 0$ but $\tanh(\Delta/2)/\Delta = 0$ obviously Δ must have the form $\Delta = 2\pi i n$, $n = 1,2,\ldots \Rightarrow -x_\alpha x^\alpha = (2\pi n)^2$, and this can occur for both signatures. This is a rather nice property of the parameters t^α: all the points that induce the identity of SO$^\uparrow$(2,1) or SO(0,3) (apart from the point $x^\alpha = 0$) are mapped by (3.1) into a *single* point $t^\alpha = 0$.

More troublesome is the case of $\Delta \neq 0$, $[\tanh(\Delta/2)/\Delta] = \infty$, where $\Delta \to 2\pi i(n + 1/2)$, $n = 0,1,2,\ldots$. In this case *finite* t^α that correspond to *finite* x^α do not exist. Observe that this situation—which can occur for both signatures—corresponds, according to (2.9), to

$$\Delta \to 2\pi i(n+1/2), \quad n = 0,1,2,\ldots:$$
$$-x_\alpha x^\alpha = (2\pi)^2(n+1/2)^2,$$
$$L \to P := \| -\delta^\alpha{}_\beta + 2(x^\alpha x_\beta / x^\rho x_\rho) \| \Rightarrow P^2 = I. \quad (3.2)$$

Thus the case of L tending to the involution P *cannot* be covered by finite t^α.

We will soon demonstrate a solution to this problem. We prefer first, however, to consider in some detail those elements of these groups that are covered by these parameters. We exclude the singular points by insisting that $\Delta \neq 2\pi i n$, $n = 1,2,\ldots$, and $\Delta \neq 2\pi i(n+1/2)$, $n = 0,1,2,\ldots$. Under these circumstances (3.1) may be inverted to give

$$x^\alpha = \Delta \coth(\Delta/2) t^\alpha, \quad \tanh^2(\Delta/2) = t_\alpha t^\alpha,$$
$$1 - t_\alpha t^\alpha > 0, \quad (3.3)$$

where we see that the set of such $\{t^\alpha\}$ fills up all of $\mathscr{R}^3$. We refer to this as the normal domain of the tangential parameters, t^α. This expression for x^α may now be inserted into (2.9) to realize our desired parametrization

$$L^\alpha{}_\beta(t^\rho) = \frac{(1 + t_\rho t^\rho)\delta^\alpha{}_\beta + 2t_\rho \epsilon^{\alpha\rho}{}_\beta - 2t^\alpha t_\beta}{1 - t_\sigma t^\sigma}. \quad (3.4)$$

We note that if such an $L^\alpha{}_\beta$ is given, then t^ρ may be determined by

$$4/(1 - t_\beta t^\beta) = \mathrm{Tr}(L) + 1, \quad (3.5a)$$

$$4t^\alpha/(1 - t_\beta t^\beta) = \epsilon^{\alpha\beta\eta} L_{\beta\eta}. \quad (3.5b)$$

We will define the set $\mathscr{L}$ as the set of all such matrices

$$\mathscr{L} \equiv \{L(t^\rho) | L^\alpha{}_\beta \text{ given by (3.4)}, \ \{t^\rho\} \in \mathscr{R}^3\}. \quad (3.6)$$

Next, we consider the set $\partial\mathscr{L}$, of limits of the matrices in the set $\mathscr{L}$, obtained via the limiting process

$$t^\alpha =: \lambda \bar{t}^\alpha, \quad \bar{t}_\alpha \bar{t}^\alpha \neq 0, \quad \lambda \in \mathscr{R}, \quad \{t^\alpha\} \in \mathscr{R}^3, \quad (3.7)$$

$$\partial\mathscr{L} \ni L(\bar{t}^\rho) = \|L^\alpha{}_\beta(\bar{t}^\rho)\|$$
$$= \| -\delta^\alpha{}_\beta + 2(\bar{t}^\alpha \bar{t}_\beta / \bar{t}^\rho \bar{t}_\rho) \|, \quad -\bar{t}^\rho \bar{t}_\rho > 0, \quad (3.8)$$

where we note that the parameters $\{\bar{t}^\rho\} \in \mathscr{R}^3$ are meaningful only modulo proportionality

$$\bar{t}^\alpha \equiv \lambda \bar{t}^\alpha, \quad \lambda \in \mathscr{R}. \quad (3.9)$$

Note also that

$$\mathrm{Tr}[L(\bar{t}^\rho)] = -1, \quad (3.10)$$

whereas one sees from (3.5a) that $\mathrm{Tr}[L(t^\rho)] \neq -1$. The eigenvalues of $L(\bar{t}^\rho)$ are $(1,-1,-1)$, while $\bar{t}^\rho$ is determined modulo proportionality by letting it correspond to the eigenvector for the eigevalue $+1$.

The matrices in $\partial\mathscr{L}$ are important in the context of this discussion, especially because they correspond to those matrices L which have x^ρ such that $\Delta = 2\pi i(n+1/2)$, i.e., exactly those matrices that cannot be covered by the form

$L(t^\rho)$. Comparing (3.8) and (3.2) we see that we can make the identification

$$\bar{t}^\alpha = x^\alpha, \quad -x_\alpha x^\alpha = [2\pi(n+1/2)]^2, \quad n = 0,1,2,..., \tag{3.11}$$

noting, of course, that

$L^2(\bar{t}^\rho) = I$, analogously to (3.2).

The complete situation is then summed up by the following which notes that:

$$\bar{\mathscr{L}} \equiv L \cup \partial\mathscr{L} \tag{3.12}$$

does indeed cover all the matrices $L^\alpha{}_\beta$ from (2.9). We can refer to the space $\bar{\mathscr{L}}$ as a quasigraded structure since the value of Tr(L) for $L \in \bar{\mathscr{L}}$ distinguishes whether $L \in \mathscr{L}$ or of $\partial\mathscr{L}$. As will be shown below, the product of two such L's may lie in either $\mathscr{L}$ or $\partial\mathscr{L}$ in *each* of the possible cases where the two members of the product are (1) both from $\mathscr{L}$, (2) one from $\mathscr{L}$ and one from $\partial\mathscr{L}$, or (3) both from $\partial\mathscr{L}$.

Theorem 2: If the 3×3, real matrix $L = \|L^\alpha{}_\beta\|$ is an element of either SO$'$(2,1) or SO(0,3), depending on the signature of $g_{\alpha\beta}$, then there are only two possible cases: (i) if Tr$(L) \neq 1$, then $L = L(t^\rho) \in \mathscr{L}$ may be considered as parametrized in the form given in (3.4), where $\{t^\rho\} \in \mathscr{R}^3$ is determined by (3.5) and the relation to the canonical parametrization is given by (3.3); (ii) if Tr$(L) = -1$, then $L = L(\bar{t}^\rho) \in \partial\mathscr{L}$, parametrized as in (3.8), with $\bar{t}^\rho$ being the eigenvector for eigenvalue $+1$, with the relation to the canonical parametrization given by (3.11).

A worthwhile picture of the (twofold) parameter space, described by the t^α's and the $\bar{t}^\alpha$'s, is obtained by envisioning the $\{t^\alpha\}$ as $\mathscr{R}^3$, and the $\{\bar{t}^\alpha\}$, which are only defined modulo a proportionality (of either sign) as a "closure" of $\mathscr{R}^3$ by the real projective sphere RP^2.

IV. COMPOSITION LAWS VIA $L(t^\rho)$ AND THE BCH PRODUCT

Consider now a product $L(s^\alpha)L(t^\beta)$ of two matrices from L. If this product results also in a matrix from L, say $L(r^\alpha)$, then we can work out from (3.6) that

$$\frac{1}{1-r_\alpha r^\alpha} = \frac{(1+s_\alpha t^\alpha)^2}{(1-s_\beta s^\beta)(1-t_\eta t^\eta)}, \tag{4.1a}$$

$$\frac{1}{1-r_\beta r^\beta} = \frac{(1+s_\rho t^\rho)}{(1-s_\beta s^\beta)(1-t_\eta t^\eta)} (s^\alpha + t^\beta + \epsilon^{\alpha\beta\eta} s_\beta t_\eta). \tag{4.1b}$$

Therefore

$$1 + s_\alpha t^\alpha \neq 0 \Rightarrow r^\alpha = \frac{s^\alpha + t^\alpha + \epsilon^{\alpha\beta\eta} s_\beta t_\eta}{1 + s_\rho t^\rho}. \tag{4.2}$$

This formula provides us with a very simple *rational* composition law of the parameters from $\mathscr{R}^3$, which, at least for sufficiently small s^α's and t^α's—such that $1 + s_\alpha t^\alpha > 0$—describes the composition laws of the two groups SO$'$(2,1) and SO(0,3) for *finite* transformations.

This result was mentioned in Ref. 5, in the case of the orthogonal group SO(3) $\equiv$ SO(0,3), in vectorial notation, and with the signature $(+,+,+)$ in the form of

$$\mathbf{r} = \frac{\mathbf{s} + \mathbf{t} + \mathbf{s}\times\mathbf{t}}{1 - \mathbf{s}\cdot\mathbf{t}}, \tag{4.3}$$

which, in the neighborhood of the identity (s and t small), would seem to be a preferable alternative to the usual SO(3) composition law spelled out in terms of Euler angles.[6]

When this formula was first established (see Ref. 6), one of the authors was satisfied with its validity in a neighborhood of the identity, only. However, in the context of this present article, one is able to do much better, establishing *rational* composition laws covering *all* elements of the groups in question.

To see this, we rewrite (4.2) in a form suitable for taking an appropriate limit:

$$r^\alpha =: \sigma^{-1}\bar{r}^\alpha, \quad \sigma := 1 + s_\alpha t^\alpha, \quad \bar{r}^\alpha := s^\alpha + t^\alpha + \epsilon^{\alpha\beta\eta} s_\beta t_\eta$$
$$\Rightarrow \bar{r}_\alpha \bar{r}^\alpha = \sigma^2 - (1 - s_\alpha s^\alpha)(1 - t_\alpha t^\alpha). \tag{4.4}$$

Inserting this form of r^α into (3.5) we find that $L^\alpha{}_\beta(r^\rho)$ admits a limit as $\sigma \to 0$, amounting to

$$\lim_{\sigma\to0} L^\alpha{}_\beta(\sigma^{-1}\bar{r}^\rho) = -\delta^\alpha{}_\beta + 2(\bar{r}^\alpha \bar{r}_\beta / \bar{r}^\rho \bar{r}_\rho) \in \partial\mathscr{L}, \tag{4.5a}$$

while

$$1 + s_\alpha t^\alpha \to 0, \quad \bar{r}^\alpha = s^\alpha + t^\alpha + \epsilon^{\alpha\beta\eta} s_\beta t_\eta, \tag{4.5b}$$

and

$$-\bar{r}_\alpha \bar{r}^\alpha = (1 - s_\rho s^\rho)(1 - t_\sigma t^\sigma) > 0.$$

If one prefers, the same result can be obtained more directly, with greater effort, by computing the product $L(s^\rho)L(t^\rho)$ under the assumption that $1 + s_\alpha t^\alpha = 0$.

Via a similar technique to that which just led to (4.5), some work allows for the calculation of the composition rules for any choices of matrices from the entire quasigraded structure $\bar{\mathscr{L}} = \mathscr{L} \cup \partial\mathscr{L}$.

Theorem 3: For the parametrization of the groups SO$'$(2,1) and SO(0,3) given by $\{t^\alpha\}$, $\{\bar{t}^\beta\} \in \mathscr{R}^3$, with $\{\bar{t}^\beta\}$ only meaningful modulo proportionality, via $L(t^\alpha) \in \mathscr{L}$ and $L(\bar{t}^\beta) \in \partial\mathscr{L}$ as described in Theorem 2, there exists a simple *rational* composition law with numerator and denominator having nothing worse than products of the parameters for any product of choices of matrices from $\bar{\mathscr{L}} = \mathscr{L} \cup \partial\mathscr{L}$, which is given completely in Table I.

Also in the table are listed the results of the next inquiry—into a Baker–Campbell–Hausdorff composition law—since this composition, in the abstract Lie group rather than with the 3×3, real matrices, also involves these same (eight) cases.

Various "interpretative" comments concerning the table are in order. The first is simply that it is clear that all possible products of matrices from $\bar{\mathscr{L}}$ are included. If we specify as a "condition" the four quantities in the first column, namely $1 + s_\alpha t^\alpha$ when both L's lie in $\mathscr{L}$, $\bar{s}_\alpha t^\alpha$ or $s_\alpha \bar{t}^\alpha$ when one is from $\mathscr{L}$ and one from $\partial\mathscr{L}$, and $\bar{s}_\alpha \bar{t}^\alpha$ when both are from $\partial\mathscr{L}$, we can see—very reasonably—that when the condition does not vanish the product is in $\mathscr{L}$, while vanishing of the condition ensures that the product is in $\partial\mathscr{L}$. [One easily verifies that each of the r^α's satisfies $1 - r_\alpha r^\alpha > 0$, $\{r^\alpha\} \in \mathscr{R}^3$, as needed for $L(r^\alpha) \in \mathscr{L}$, while each of the $\bar{r}^\alpha$'s satisfies $-\bar{r}_\alpha \bar{r}^\alpha > 0$, $\{\bar{r}^\alpha\} \in \mathscr{R}^3$, as needed for $L(\bar{r}^\alpha) \in \partial\mathscr{L}$. As well, one notes that the r^{α}'s given are independent of proportionality transformations of any $\bar{s}^{\alpha}$'s or $\bar{t}^{\alpha}$'s they may contain.]

As well, comments concerning the Baker–Campbell–Hausdorff results listed in Table I are now in order. By $x[t^{\rho}]$ or $x[\bar{t}^{\rho}]$, we mean the elements of the abstract Lie algebra referred to in (2.1), but with the x_{α} taken as functions of either t^{ρ} or $\bar{t}^{\rho}$:

$$x[t^{\rho}] := [\Delta/\tanh(\Delta/2)]t_{\alpha}\gamma^{\alpha}, \quad \{t^{\alpha}\}\in\mathscr{R}^3,$$
$$\tanh^2(\Delta/2) = t_{\alpha}t^{\alpha}, \quad 1 - t_{\alpha}t^{\alpha} > 0, \tag{4.6}$$

$$x[\bar{t}^{\rho}] := \lim_{\lambda\to\pm\infty} x[t^{\rho}] = \left[(\pi\bar{t}_{\alpha}/\sqrt{-\bar{t}_{\alpha}\bar{t}^{\alpha}})\right]\gamma^{\alpha},$$
$$\{\bar{t}^{\alpha}\}\in\mathscr{R}^3, \quad -\bar{t}_{\alpha}\bar{t}^{\alpha} > 0. \tag{4.7}$$

Since these two sorts of quantities x specify all the elements of the respective Lie algebras of the (abstractly defined) groups $SO^{\uparrow}(2,1)$ or $SO(0,3)$, their exponentials constitute the very elements of the (abstract) group, again parametrized by our "tangential" parametrization. Therefore it is not surprising that we are able to completely express the products of these exponentials in terms of the parametrized form of the group composition law. The representation should be contrasted with the matrix representation obtained via Theorem 1, (2.15). The two representations are adjoint one to another, explaining the inversion of order in the BCH composition rule relative to the other one. This property is seen by rewriting (2.15) in the current notation, giving

$$e^{x[t^{\rho}]}\gamma^{\alpha}e^{-x[t^{\rho}]} = L^{\alpha}{}_{\beta}(t^{\rho})\gamma^{\beta}, \tag{4.8}$$

with similar expressions where one or both of the $x[t^{\rho}]$ are exchanged for $x[\bar{t}^{\rho}]$. Assuming that $1 + s_{\alpha}t^{\alpha}\neq 0$, we may now write

$$e^{x[t^{\rho}]}e^{x[s^{\rho}]}\gamma^{\alpha}e^{-x[s^{\rho}]}e^{-x[t^{\rho}]} = L^{\alpha}{}_{\beta}(s^{\rho})e^{x[t^{\rho}]}\gamma^{\beta}e^{-x[t^{\rho}]}$$
$$= L^{\alpha}{}_{\beta}(s^{\rho})L^{\beta}{}_{\delta}(t^{\rho})\gamma^{\delta}, \tag{4.9}$$

demonstrating, for this case, the desired statement. The other cases follow in the same fashion.

As well, now, we may use the composition rule from Table I to obtain

$$\Lambda\gamma^{\alpha}\Lambda^{-1} = \gamma^{\alpha}, \tag{4.10a}$$

where

$$\Lambda = \Lambda(t^{\rho},s^{\rho}) := e^{-x[r^{\rho}]}e^{x[t^{\rho}]}e^{x[s^{\rho}]}\neq 0. \tag{4.10b}$$

Since Λ therefore commutes with all γ^{α} it must be a function of the Casimir operator $\gamma_{\alpha}\gamma^{\alpha}$.

In fact, we will now give a simple argument to show that Λ is identically equal to 1, which will verify the BCH entry in the first line of Table I. Because all the various lines of Table I were obtained by a sequence of limiting transitions from the first line, it is clear that a similar sequence of limiting transitions may be repeated with respect to the above argument, leading in turn to each of the other BCH entries in Table I. In order, now, to verify that Λ is equal to 1, we first recall some essential facts from combinatorial group theory.[7]

Let x and y be generators of a free associative algebra. Then, the *Magnus bracket* $\{\cdot,\cdot\}$ is defined as follows: $\{x^0,y\} = \{1,y\} := y$ and for $n = 0,1,2,\ldots$, $\{x^{n+1},y\} := [x,\{x^n,y\}]$. If $f(x)$ is a formal series with coefficients from some field of numbers of characteristic zero (in practice $\mathscr{R}$ or $\mathscr{C}$), $f(x) = \Sigma_{n=0}^{\infty}a_n x^n$ then $\{f(x),y\}$ is an abbreviation for $\Sigma_{n=0}^{\infty}a_n\{x^n,y\}$. For example, we have

$$\{e^{\lambda x},y\} = \sum_{n=0}^{\infty}\frac{\lambda^n}{n!}[x[x\cdots[x,y]\cdots] = e^{\lambda x}ye^{-\lambda x}, \tag{4.11}$$

$\lambda\in\mathscr{R}$ (the Schwinger formula). This definition allows us to state Magnus' formula:

$$\frac{d}{dt}e^{x(t)} = X(t)e^{x(t)},$$
$$X(t) := \left\{\frac{e^x - 1}{x},\dot{x}\right\} = \int_0^1 d\lambda\{e^{\lambda x},\dot{x}\}$$
$$= \int_0^1 d\lambda\, e^{\lambda x}\dot{x}e^{-\lambda x}, \quad \lambda,\ t\in\mathscr{R}. \tag{4.12}$$

Then, in the case in question where the $x(t)$ are linear and homogeneous in the generators γ^{α} the $\dot{x}(t)$ are also linear, homogeneous, and therefore so will be the corresponding $X(t)$, having the form

$$X(t) = X_{\alpha}(t)\gamma^{\alpha}, \tag{4.13}$$

where for each t, $\{X_{\alpha}(t)\}\in\mathscr{R}^3$.

Similarly, defining $Y(t)$ by

$$\frac{d}{dt}e^{x(t)} = e^{x(t)}Y^{(t)}, \tag{4.14}$$

the completely analogous argument tells us that when the $x(t)$ are linear and homogeneous in the generators γ^{α} then so will be $Y(t)$, taking the form

$$Y(t) = Y_{\alpha}(t)\gamma^{\alpha}, \tag{4.15}$$

where for each t, $\{Y_{\alpha}(t)\}\in\mathscr{R}^3$.

We now rewrite the expression (4.10b) for Λ so that

$$e^{x[t^{\rho}]}e^{x[s^{\rho}]} = \Lambda(t^{\rho},s^{\rho})e^{x[r^{\rho}]}, \tag{4.16}$$

and suppose that the t^{α}'s are arbitrary analytic functions of t. Then also the r^{α}'s are analytic functions of t, all exponents being linear in γ^{α}'s; the s^{α}'s we treat as constants.

By differentiating (4.14) with respect to t and using (4.13), it follows that

$$\Lambda X_{\alpha}\gamma^{\alpha} = \frac{d}{dt}\Lambda + \Lambda\dot{X}_{\alpha}\gamma^{\alpha}, \tag{4.17}$$

and consequently $0 = d/dt$, $\Lambda = \dot{t}^{\alpha}(\partial\Lambda/\partial t^{\alpha})$. This can be true for every $t^{(\alpha)}(t)$ iff $\partial\Lambda/\partial t^{\alpha} = 0$. Similarly, consider (4.16) with t^{α} independent of t, and with s^{α}'s arbitrary analytic functions of t. Differentiating (4.16), it follows that with Y_{α} and $\dot{Y}_{\alpha}$ being scalars

$$\Lambda Y_{\alpha}\gamma^{\alpha} = \frac{d}{dt}\Lambda + \Lambda\dot{Y}_{\alpha}\gamma^{\alpha}, \tag{4.18}$$

and consequently, $0 = d\Lambda/dt$, $\Lambda = \dot{s}^{\alpha}(\partial\Lambda/\partial s^{\alpha}) \Leftrightarrow 0 = \partial\Lambda/\partial s^{\alpha}$. Therefore Λ in (4.16) cannot depend on either t^{α} or s^{α}. Since this is true, by specializing (4.16) for

$t^\alpha = 0 = s^\alpha$, we find $\Lambda = 1$, which completes our proof of the Baker–Campbell–Hausdorff composition law,[8] which we restate in the form of the following theorem.

Theorem 4: In the parametrization of $SO^\uparrow(2,1)$ and $SO(0,3)$ $\{t^\alpha\}\in\mathscr{R}^3$ and $\{\bar{t}^\alpha\}\in\mathscr{R}^3$ modulo proportionality, the realization of the group elements via $e^{x[t^\rho]}$ and $e^{x[\bar{t}^\rho]}$ admits the Baker–Campbell–Hausdorff-type group composition law for all possible products as given explicitly in Table I.

An interesting corollary, analogous to the situation originally stated in (3.2), i.e., $L^2(\bar{t}^\rho) = I$, is now given.

Corollary:

$$e^{2x[\bar{t}^\rho]} = I. \tag{4.19}$$

The corollary follows from the seventh line of Table I, choosing $\bar{s}^\alpha$ the same as $\bar{t}^\alpha$, we have $\bar{r}^\alpha = (\bar{t}_\beta \bar{t}^\beta)^{-1}\epsilon^{\alpha\beta\delta}\bar{t}_\beta\bar{t}_\delta = 0$, independently of the precise magnitude of $\bar{t}^\alpha$, verifying that these elements are involutions in the abstract Lie group as well.

V. SOME PROPERTIES OF ANALYTIC CURVES ON THE GROUP MANIFOLDS

Generalizing a comment at the beginning of Sec. II, if we choose $x_\alpha = x_\alpha(t)$ to be arbitrary (real) analytic functions of t, then the realization via (2.15) of elements $L^\alpha{}_\beta(t)$ of the appropriate group, $SO^\uparrow(2,1)$ or $SO(0,3)$, describes an analytic curve on the manifold of the defining, three-dimensional representation of the group, while $e^{x(t)}$ similarly generates the "same" curve on the abstract group manifold.

Using the Magnus formulas (4.12) and writing $X(t)$, as was done there, for the quantity that gives the rate of change along the curve, we have

$$\begin{aligned} X(t) :&= \left\{\frac{d}{dt}e^{x(t)}\right\}e^{-x(t)} = \dot{x}_\alpha \int_0^1 d\lambda\, e^{\lambda x}\gamma^\alpha e^{-\lambda x} \\ &= \dot{x}_\alpha\left\{\int_0^1 d\lambda\, L^\alpha{}_\beta(\lambda)\right\}\gamma^\beta \equiv \dot{x}_\alpha N^\alpha{}_\beta\gamma^\beta, \end{aligned} \tag{5.1}$$

where the matrix $N = \|N^\alpha{}_\beta\|$ is easily computed from the explicit form of $L^\alpha{}_\beta$ given in (2.9):

$$\begin{aligned} N &= I + [(\cosh\Delta - 1)/\Delta^2]M + [(\sinh\Delta - \Delta)/\Delta^3]M^2 \\ &= \|\delta^\alpha{}_\beta + [(\cosh\Delta - 1)/\Delta^2]x_\rho\epsilon^{\alpha\rho}{}_\beta \\ &\quad + [(\sinh\Delta - \Delta)/\Delta^3]\,(\Delta^2\delta^2{}_\beta - x^\alpha x_\beta)\|, \end{aligned} \tag{5.2}$$

remembering that $\Delta = \sqrt{x_\alpha x^\alpha}$.

Because N shares eigenvectors with M, one easily sees that N has the eigenvalues $(1,(e^\Delta - 1)/\Delta, (e^{-\Delta} - 1)/-\Delta)$, from which we see that

$$\det(N) = (2\sinh(\Delta/2)/\Delta)^2, \tag{5.3a}$$

$$\mathrm{Tr}(N) = 1 + 2(\sinh\Delta/\Delta). \tag{5.3b}$$

Also we notice that in the limiting case $\Delta \to 0$, we have

$$\begin{aligned} &N = \|\delta^\alpha{}_\beta + \tfrac{1}{2}x_\rho\epsilon^{\alpha\rho}{}_\beta - \tfrac{1}{6}x^\alpha x_\beta\|, \\ &(N - I)^3 = 0, \quad \det(N) = 1, \quad \mathrm{Tr}(N) = 3, \end{aligned} \tag{5.4}$$

which is only interesting for $SO^\uparrow(2,1)$ since for $SO(0,3)$, $\Delta = 0 \Rightarrow x^\alpha = 0$. We also notice that $\det(N)$ is positive, and therefore N^{-1} exists, except for the singular cases of $\Delta = 2\pi i n$, $n = 1,2,\ldots$. In these singular cases, we have

$$N = \|-x^\alpha x_\beta/(2\pi n)^2\|, \quad \Delta = 2\pi i n, \quad n = 1,2,\ldots . \tag{5.5}$$

Outside of these singular cases N^{-1} exists, and is readily seen to amount to

$$\begin{aligned} N^{-1} &= I - \frac{1}{2}M - \frac{1}{\Delta^2}\left(1 - \frac{\sinh\Delta/\Delta}{2(\cosh\Delta - 1)/\Delta^2}\right)M^2 \\ &= \|\delta^\alpha{}_\beta - \frac{1}{2}x_\rho\epsilon^{\alpha\rho}{}_\beta - \frac{1}{\Delta^2}\left(1 - \frac{\Delta}{2}\coth\left(\frac{\Delta}{2}\right)\right) \\ &\quad \times (\Delta^2\delta^\alpha{}_\beta - x^\alpha{}_\beta)\|. \end{aligned} \tag{5.6a}$$

Notice that then in particular

$$\lim_{\Delta\to 0} N^{-1} = \|\delta^\alpha{}_\beta - \tfrac{1}{2}x_\rho\epsilon^{\alpha\rho}{}_\beta - \tfrac{1}{12}x^\alpha x_\beta\|. \tag{5.6b}$$

All of the results above concerned with $X(t)$ will assume much simpler form if we represent $X(t)$ in terms of the tangential parametrization; i.e., $X[\bar{t}^\rho]$ from (4.7), with correspondingly the parameters t^α or $\bar{t}^\alpha$ interpreted as arbitrary analytic functions of $t\in\mathscr{R}$. By substituting from (4.6) into (5.2), a straightforward computation leads to the rather simple result stated in the following theorem, along with the results obtained by also setting in these formulas $t^\alpha(t) = \lambda\bar{t}^{\alpha(t)}$, $\lambda = \text{const}$, $\infty > -\bar{t}_\alpha\bar{t}^\alpha > 0$, as before and executing the limit $\lambda \to \infty$ of $\dot{t}_\beta K^\beta{}_\alpha$.

Theorem 5: For the parametrization of the groups $SO^\uparrow(2,1)$ and $SO(0,3)$ given by $\{t^\alpha\}\in\mathscr{R}^3$ and $\{\bar{t}^\alpha\}\in\mathscr{R}^3$ modulo proportionality, the analytic curves on the group manifold generated by taking $t^\alpha = t^\alpha(t)$, or $\bar{t}^\alpha(t)$, $t\in\mathscr{R}$, as arbitrary, (real) analytic functions of t, and using $\exp\{x[t^\rho(t)]\}$ or $\exp\{x[\bar{t}^\rho(t)]\}$ to generate analytic curves on the group manifold, we may explicitly write the rate of change along the curve—the generating vector field, $X(t)$—in the linear, homogeneous form

$$X[t^\rho(t)] := \left\{\frac{d}{dt}e^{x[t^\rho(t)]}\right\}e^{-x[t^\rho(t)]} = \dot{t}_\beta(t)K^\beta{}_\alpha\gamma^\alpha \tag{5.7a}$$

or

$$\bar{X}[\bar{t}^\rho(t)] := \left\{\frac{d}{dt}e^{x[\bar{t}^\rho(t)]}\right\}e^{-x[\bar{t}^\rho(t)]} = \dot{\bar{t}}_\beta\bar{K}^\beta{}_\alpha\gamma^\alpha, \tag{5.7b}$$

where

$$K := \|K^\alpha{}_\beta\| = \|[2/(1 - t_\sigma t^\sigma)]\,(\delta^\alpha{}_\beta + t_\rho\epsilon^{\alpha\rho}{}_\beta)\| \tag{5.8a}$$

and

$$\bar{K} := \|\bar{K}^\alpha{}_\beta\| = \|2\bar{t}_\rho\epsilon^{\alpha\rho}{}_\beta/(-\bar{t}_\sigma\bar{t}^\sigma)\|. \tag{5.8b}$$

A few observations about the algebraic structure of the matrices K and $\bar{K}$ are now needed. Similarly to M from (2.5), the matrix $M' := \|t_\rho\epsilon^{\alpha\rho}{}_\beta\|$, fulfills the Hamilton–Cayley equation $M'^3 - t_\alpha t^\alpha M' = 0$, having thus eigenvalues $(0, \sqrt{t_\alpha t^\alpha}, -\sqrt{t_\alpha t^\alpha})$. Because K shares eigenvectors with M', it follows that the eigenvalues of K are

$$\left(\frac{2}{1 - t_\alpha t^\alpha}, \frac{2(1 + \sqrt{t_\alpha t^\alpha})}{1 - t_\alpha t^\alpha}, \frac{2(1 - \sqrt{t_\alpha t^\alpha})}{1 - t_\alpha t^\alpha}\right)$$

and therefore $\det(K) = 8/(1 - t_\alpha t^\alpha)^2$; thus, in the normal domain of $\{t^\alpha\}$, $\det(K)$ is positive, and the matrix K^{-1} exists. [On the other hand, the matrix $\overline{K}$, admitting an eigenvalue of zero, corresponding to $\bar{t}^\alpha$ serving as the eigenvector, has $\det(\overline{K}) = 0$.] One easily sees that the explicit form of K^{-1} is

$$K^{-1} = \|\tfrac{1}{2}(\delta^\alpha{}_\beta - t_\rho \epsilon^{\alpha\rho}{}_\beta - t^\alpha t_\beta)\|, \tag{5.9}$$

this quantity being simply quadratic in the functions $t^\alpha = t^\alpha(t)$. Consequently, the equations $X_\alpha(t) = \dot{t}_\beta K^\beta{}_\alpha$ can be inverted with respect to $\dot{t}_\alpha$ in the simple form of

$$\dot{t}_\alpha = X_\beta(t) K^{-1\beta}{}_\alpha = \tfrac{1}{2} X_\beta(t)(\delta^\beta{}_\alpha - t_\rho \epsilon^{\beta\rho}{}_\alpha - t^\beta t_\alpha). \tag{5.10}$$

These relations are of interest for many reasons.

Among other things, Eq. (5.10) is crucial for attempts to solve the problem: given arbitrary analytic functions $\{X_\alpha(t)\} \in \mathscr{R}^3$, can we determine $x(t) = x_\alpha(t)\gamma^\alpha$ such that

$$\frac{d}{dt} e^{x(t)} = X_\alpha(t)\gamma^\alpha e^{x(t)}, \quad e^{x(0)} = I? \tag{5.11}$$

This entire problem we will save for later reports. In the present paper we will only investigate the case of X_α and $\dot{X}_\alpha$ being *colinear*, i.e., there is some $\mu(t) \neq 0$ such that

$$\dot{X}_\alpha(t) = (\dot{\mu}/\mu) X_\alpha(t) \Leftrightarrow X_\alpha = \mu(t) Y_\alpha, \tag{5.12}$$

$\{Y_\alpha\} \in \mathscr{R}^3$, independent of t.

Under the conditions of (5.12) the solution to the problem stated in (5.11) simply amounts to

$$x(t) = \nu(t) Y_\alpha \gamma^\alpha, \quad \text{with } \nu(t) := \int_0^t dt' \mu(t'). \tag{5.13}$$

On the other hand, using the presentations in Theorem 5, we can easily establish the conditions for the colinearity expressed in (5.12). Using the expressions for K^{-1}, we find that linear independence of X_α and $\dot{X}_\beta$ is expressed by

$$Y^\alpha := \epsilon^{\alpha\beta\eta} X_\beta \dot{X}_\eta = [8/(1 - t_\kappa t^\kappa)^2] K^{-1\alpha}{}_\gamma \epsilon^{\lambda\rho\sigma} \dot{t}_\rho \ddot{t}_\alpha. \tag{5.14}$$

It follows that X_α, $\dot{X}_\alpha$ are colinear iff $\dot{t}_\alpha$, $\ddot{t}_\alpha$ are colinear. In addition, setting (5.14) $t^\alpha(t) = \lambda \bar{t}^\alpha(t)$, $-\bar{t}_\alpha \bar{t}^\alpha > 0$, $\lambda = \text{const}$, and executing the limiting transition $\lambda \to \infty$, we obtain

$$\begin{aligned} \overline{Y}^\alpha &:= \epsilon^{\alpha\beta\eta} \overline{X}_\beta \dot{\overline{X}}_\eta = (2/\bar{t}_\kappa \bar{t}^\kappa)^2 W \bar{t}^\alpha, \\ W &:= \epsilon_{\alpha\beta\gamma} \bar{t}^\alpha \dot{\bar{t}}^\beta \ddot{\bar{t}}^\gamma, \end{aligned} \tag{5.15}$$

which could also have been directly derived from (5.8b).

Therefore, $(\overline{X}_\alpha, \dot{\overline{X}}_\alpha)$ are colinear *iff* the Wronskian W vanishes, implying that the curve $\bar{t}^\alpha = \bar{t}^\alpha(t)$ in $\mathscr{R}^3$ must be contained in a plane through the origin of $\mathscr{R}^3$.

VI. CONCLUDING REMARKS

The main results of our paper can be summarized as follows. Elements of the defining representation of the group SO(0,3) or $SO^\uparrow(2,1)$ constitute the set $\overline{\mathscr{L}} = \mathscr{L} \cup \partial\mathscr{L}$, where the elements of $\mathscr{L}$ can be parametrized by $\{t^\alpha\} \in \mathscr{R}^3$, fulfilling the condition (3.3), i.e., $1 - t^\alpha t_\alpha > 0$ ("the tangential parametrization"), while the members of $\partial\mathscr{L}$, which can be obtained from $\mathscr{L}$ by a limiting process, are parametrized by all directions of straight lines in $\mathscr{R}^3$ in the case of SO(0,3), or by the directions $\lambda \bar{t}^\alpha$ such that $\bar{t}^\alpha \bar{t}_\alpha < 0$ in the case of $SO^\uparrow(2,1)$. The group composition law of our parameters is rational and very simple (Table I). By using this parametrization we are also able to express the Baker–Campbell–Hausdorff formula in a concise form. Finally, note that *some* ideas of the present paper have found generalizations in work on quaternionlike algebras.[9]

[1] L. Infeld and J. Plebański, Acta Phys. Pol. **14**, 41 (1955). I. Bialynicki-Birula has generalized the ideas of this paper in his M.S. thesis (University of Warsaw, 1956) to the case of an arbitrary number of $2n$ canonical variables, including $n \to \infty$, denumerable or nondenumerable.

[2] J. Plebański, Acta. Phys. Pol. **14**, 275 (1955), where results of Ref. 1 found some applications.

[3] Of course the group SO(0,3) is equivalent to the more usual notation SO(3); however, our choice of the sign of the elliptic metric is more convenient for our purposes and we therefore maintain this notation.

[4] The earlier work, in Refs. 1 and 2, began with the usual canonical commutation rule for the Hermitean operators q and p: $[q,p] = i\hbar$. A basis for objects quadratic in q and p is given by the operators Γ^α, $\alpha = 1,2,3$, defined by $\Gamma^1 := \frac{1}{4}(pq + qp)$, $\Gamma^2 := \frac{1}{4}(p^2 - q^2)$, $\Gamma^3 := \frac{1}{4}(p^2 + q^2)$. A realization of the commutation relations in (1.5) is then given by $\gamma^\alpha = (1/2i\hbar)\Gamma^\alpha$. Other work on linear unitary transformations, particularly in the aspect of proceeding equivalently with Poisson brackets and commutators is given in M. Moshinsky and C. Quesne, in *Proceedings of XV Solvay Conference in Physics*, Brussels, 1971 (Gordon and Breach, New York, 1974); and in K. B. Wolf, J. Math. Phys. **15**, 1295, 2102 (1974).

[5] B. Mielnik and J. Plebański, Ann. Inst. H. Poincare **A-XII**, 215 (1970).

[6] The rule (4.2), valid for both $SO^\uparrow(2,1)$ and SO(0,3), was first established, via a technique different from the present text, involving spinorial realizations of the Lie algebras in question, in a monograph "On the generators of the n-dimensional pseudo-unitary and pseudo-orthogonal groups" by J. Plebański, Centro de Investigacion y de Estudios Avanzados, 1966, p. 322, distributed in 100 copies, but for reasons of its size and the cost involved, this text was never properly published. Nevertheless, Pursey, Phys. Rev. D **29**, 1848 (1984) has inquired concerning the origins of the formula written there.

[7] W. Magnus, A. Karrass, and D. Solitar, *Combinatorial Group Theory* (Dover, New York, 1976).

[8] It should be mentioned that B. Mielnik, when studying quantum-mechanical realizations of $SO^\uparrow(2,1)$, has employed BCH compositions of this type: B. Mielnik, J. Math. Phys. **27**, 2290 (1986).

[9] J. F. Plebański and M. Przanowski, J. Math. Phys. **29**, 529 (1988).

The theory of screws: A new geometric representation for the group SU(1,1)

R. Simon[a] and N. Mukunda[b]
Center for Theoretical Studies, Indian Institute of Science, Bangalore 560 012, India
E. C. G. Sudarshan[a]
Department of Physics, University of Texas at Austin, Austin, Texas 78712

(Received 27 September 1988; accepted for publication 23 November 1988)

Hamilton's theory of turns, which gives a geometrical description of the elements and structure of the compact group SU(2), is generalized to a theory of screws for the noncompact group SU(1,1). Group elements are pictured as geometric objects in a three-dimensional Minkowski space, and the composition law is reduced to a geometric operation on them. A new classification of elements of SU(1,1), leading to an interesting structural result about the group manifold, is introduced.

I. INTRODUCTION

The unitary unimodular group SU(2) is the simplest example of a non-Abelian compact Lie group with a simple Lie algebra. It is of fundamental significance in quantum mechanical problems, being basic to the quantum theory of angular momentum as the covering group of the rotation group SO(3).[1] It also plays an important role in theories of internal symmetry for nuclear and particle physics; and it is of significance in polarization optics as well.[2] In the context of angular momentum theory, it is often convenient to parametrize the elements of SU(2) by Euler angles, which makes the irreducible representation matrices in a suitable basis easy to deal with. The group composition law, however, is rather cumbersome with this parametrization. On the other hand, the use of homogeneous Euler parameters simplifies the expressions for group multiplication to some extent, but involves the use of nonindependent parameters.

All these are of course algebraic ways of describing the elements and the composition law of SU(2). It is remarkable, but unfortunately not too well known, that as long ago as 1853 Hamilton had invented a geometrical or pictorial way of representing SU(2) elements and their multiplication, which is extremely elegant and gives one a direct and vivid grasp of the structure of SU(2). This is the so-called method of turns.[3] To appreciate Hamilton's method, let us first recall the much simpler case of the Abelian group of translations in Euclidean three-dimensional space. Each translation is representable as a vector in space, only the direction and magnitude being significant, and the location irrelevant. The composition of two translations is given by the head-to-tail parallelogram rule of vector addition; and taking the inverse amounts to reversing direction. In Hamilton's theory of turns, we have a generalization of such a picture from the Abelian translation group to the non-Abelian SU(2). Instead of vectors in space, we deal with directed great circle arcs, of length $\leqslant\pi$, on a unit sphere S^2 in a Euclidean three-dimensional space. Two such arcs are deemed equivalent if by sliding one along its great circle it can be made to coincide with the other. Equivalence classes of such arcs are called turns, and elements of SU(2) can be represented by them. Perhaps the most remarkable feature is that the noncommutative multiplication law of SU(2) can be translated into the language of turns thus: given two elements of SU(2), slide the corresponding representative great circle arcs on their respective great circles till the head of the first coincides with the tail of the second. Then the product element is represented by the turn determined by the great circle arc from the tail of the first to the head of the second arc. Inverses of elements go into reversed turns. A detailed account of Hamilton's work can be found in the monograph of Biedenharn and Louck.[4]

The purpose of this paper is to show that Hamilton's method can be generalized from the compact SU(2) to the noncompact group SU(1,1) in a very interesting way.[5] This group is the simplest non-Abelian noncompact Lie group with a simple Lie algebra, and shares with SU(2) a common complex extension. Like SU(2), SU(1,1) too is of great importance in many physical problems. Thus one may mention the theory of axially symmetric optical systems in first-order Fourier optics[6]; the group Sl(2,R) of real linear canonical transformations in one pair of canonical variables, which is isomorphic to SU(1,1); and the group of Bogoliubov transformations on one creation-annihilation operator pair, relevant for squeezed states.[7]

We shall use the term "screws" for SU(1,1) in place of the turns of SU(2). In analyzing the geometrical properties of screws and the description of the SU(1,1) group structure using them, some important differences as compared to the SU(2) case will be evident. While the carrying over of Hamilton's ideas from SU(2) to SU(1,1) is thus not trivial, it is gratifying that it can in fact be done.

The material of this paper is arranged as follows. In Sec. II we give a brief review of the method of turns for SU(2). Section III develops the method of screws for SU(1,1) in detail, paying special attention to those features that distinguish it from SU(2) but at the same time guided by the SU(2) case. The concluding section (Sec. IV) collects some pertinent remarks and points out some possible applications of our new method.

II. REVIEW OF TURNS FOR SU(2)

In this section we review very briefly the theory of turns for SU(2), in a form suitable for the intended extension to

[a] Present address: The Institute of Mathematical Sciences, C.I.T. Campus, Madras 600 113, India.
[b] Jawaharlal Nehru Fellow.

 0022-2488/89/051000-07$02.50

SU(1,1). While the content is essentially the same as in the account given in Ref. 4, it is expressed in a form convenient for our purposes.

As is well known, any matrix u in the defining representation of SU(2) [any $u \in SU(2)$] can be written in terms of homogeneous Euler parameters and Pauli matrices as

$$u = a_0 - i\mathbf{a}\cdot\boldsymbol{\sigma} \tag{2.1}$$

(the unit matrix accompanying a_0 is omitted), where a_0 and $\mathbf{a}$ are a real scalar and real Euclidean three-vector constrained by

$$a^2 + \mathbf{a}\cdot\mathbf{a} = 1. \tag{2.2}$$

In this way, elements of SU(2) correspond one-to-one to points on S^3. The constraint (2.2) suggests that we choose any two unit vectors $\mathbf{n}$, $\mathbf{n}' \in S^2$ and set

$$a_0 = \mathbf{n}\cdot\mathbf{n}', \quad \mathbf{a} = \mathbf{n}\wedge\mathbf{n}'; \tag{2.3}$$

for then, if $\theta/2$ is the angle between $\mathbf{n}$ and $\mathbf{n}'$, and $\hat{a}$ is the unit vector along $\mathbf{n}\wedge\mathbf{n}'$,

$$a_0 = \cos\theta/2, \quad \mathbf{a} = \hat{a}\sin\theta/2, \tag{2.4}$$

and condition (2.2) is obviously satisfied.

We are thus led to define, for any $\mathbf{n}$, $\mathbf{n}' \in S^2$, the following SU(2) element $A(\mathbf{n},\mathbf{n}')$:

$$A(\mathbf{n},\mathbf{n}') = \mathbf{n}\cdot\mathbf{n}' - i\mathbf{n}\wedge\mathbf{n}'\cdot\boldsymbol{\sigma}. \tag{2.5}$$

One can easily convince oneself that for any a_0, $\mathbf{a}$ obeying Eq. (2.2), choices of $\mathbf{n}$, $\mathbf{n}'$ can certainly be made so that Eq. (2.3) will be valid. Thus every $u \in SU(2)$ is obtained by making all possible choices of $\mathbf{n}$ and $\mathbf{n}'$ in $A(\mathbf{n},\mathbf{n}')$.

The geometrical meaning of the element $A(\mathbf{n},\mathbf{n}')$ is clarified by computing the element $R(A(\mathbf{n},\mathbf{n}')) \in SO(3)$ that corresponds to it under the $SU(2)\rightarrow SO(3)$ homomorphism. We find:

$$\begin{aligned} &A(\mathbf{n},\mathbf{n}')\mathbf{n}\cdot\boldsymbol{\sigma}A(\mathbf{n},\mathbf{n}')^{-1} = \mathbf{n}''\cdot\boldsymbol{\sigma}, \\ &\mathbf{n}'' = 2\mathbf{n}\cdot\mathbf{n}'\mathbf{n}' - \mathbf{n}, \\ &\mathbf{n}\cdot\mathbf{n}' = \mathbf{n}'\cdot\mathbf{n}'', \quad \mathbf{n}\wedge\mathbf{n}' = \mathbf{n}'\wedge\mathbf{n}''. \end{aligned} \tag{2.6}$$

Thus $R(A(\mathbf{n},\mathbf{n}'))$ is a right-handed rotation about $\mathbf{n}\wedge\mathbf{n}'$ as the axis with *twice* the angle ($\leqslant\pi$) enclosed between $\mathbf{n}$ and $\mathbf{n}'$. In particular, $R(A(\mathbf{n},\mathbf{n}'))$ *does not* rotate $\mathbf{n}$ into $\mathbf{n}'$ but overshoots it in the plane of $\mathbf{n}$ and $\mathbf{n}'$ to $\mathbf{n}''$. On $\mathbf{n}'$ the effect is given by

$$A(\mathbf{n},\mathbf{n}')\mathbf{n}'\cdot\boldsymbol{\sigma}A(\mathbf{n},\mathbf{n}')^{-1} = (2\mathbf{n}'\cdot\mathbf{n}''\mathbf{n}'' - \mathbf{n}')\cdot\boldsymbol{\sigma}, \tag{2.7}$$

$\mathbf{n}''$ being as in Eq. (2.6).

The construction of $A(\mathbf{n},\mathbf{n}')$ enjoys the following properties:

$$A(\mathbf{n},\mathbf{n}')^{-1} = A(\mathbf{n}',\mathbf{n}), \tag{2.8a}$$

$$A(\mathbf{n}',\mathbf{n}'')A(\mathbf{n},\mathbf{n}') = A(\mathbf{n},\mathbf{n}''). \tag{2.8b}$$

(Here $\mathbf{n}$, $\mathbf{n}'$, and $\mathbf{n}''$ are independently chosen points on S^2.) In addition, $A(\mathbf{n},\mathbf{n}')$ is unchanged if both $\mathbf{n}$ and $\mathbf{n}'$ are subjected to a common rotation about $\mathbf{n}\wedge\mathbf{n}'$. This SO(2) invariance property motivates the following equivalence relation: joining $\mathbf{n}$ and $\mathbf{n}'$ by a great circle arc of length $\leqslant\pi$, this arc is equivalent to all other arcs obtained by sliding the given arc on its great circle. An equivalence class of arcs is called a turn.

Based on Eqs. (2.8), the SU(2) group operations can immediately be given a geometrical description. Each $u \in SU(2)$ (other than $u = \pm\mathbb{1}$) corresponds to a unique turn. For $u = \mathbb{1}$ we take the null turn, i.e., any $\mathbf{n} = \mathbf{n}' \in S^2$. For $u = -\mathbb{1}$, *any* great semicircle will do, and they are all deemed equivalent. The inverse of $u = A(\mathbf{n},\mathbf{n}')$ corresponds, by Eq. (2.8a), to reversing the sense of the turn but retaining the same great circle. To compute the product $u'u$ for arbitrary u' and u, we remark that since any two great circles on S^2 definitely intersect, we can choose $\mathbf{n}$, $\mathbf{n}'$, $\mathbf{n}'' \in S^2$ such that $u = A(\mathbf{n},\mathbf{n}')$, $u' = A(\mathbf{n}',\mathbf{n}'')$; then by Eq. (2.8b),

$$u'u = A(\mathbf{n}',\mathbf{n}'')A(\mathbf{n},\mathbf{n}') = A(\mathbf{n},\mathbf{n}''). \tag{2.9}$$

Thus the turn for the product is indeed obtained by the geometrical operation described in the Introduction.

We remark that while the geometrical construction is in a three-dimensional Euclidean space (more precisely, on an S^2 therein), we are able to represent SU(2) elements, not merely those of SO(3), faithfully by turns.

To conclude this section, we return to the geometrical meaning of $A(\mathbf{n},\mathbf{n}')$ revealed in Eqs. (2.6) and ask: is there a simple expression for an element $B(\mathbf{n},\mathbf{n}') \in SU(2)$ such that, unlike with $A(\mathbf{n},\mathbf{n}')$, $R(B(\mathbf{n},\mathbf{n}'))$ will be a rotation about $\mathbf{n}\wedge\mathbf{n}'$ which takes $\mathbf{n}$ precisely to $\mathbf{n}'$ through an angle $\leqslant\pi$? The properties of $A(\mathbf{n},\mathbf{n}')$ tell us that we must take

$$\begin{aligned} B(\mathbf{n},\mathbf{n}') &= A(\mathbf{n},(\mathbf{n}+\mathbf{n}')/|\mathbf{n}+\mathbf{n}'|) \\ &= [2(1+\mathbf{n}\cdot\mathbf{n}')]^{-1/2}(\mathbb{1} + A(\mathbf{n},\mathbf{n}')), \end{aligned} \tag{2.10}$$

$$A(\mathbf{n},\mathbf{n}') = (\operatorname{sgn}\mathbf{n}\cdot\mathbf{n}')B(\mathbf{n},\mathbf{n}''),$$

where $\mathbf{n}''$ is as in Eq. (2.6). Then we find

$$\begin{aligned} &B(\mathbf{n},\mathbf{n}')\mathbf{n}\cdot\boldsymbol{\sigma}B(\mathbf{n},\mathbf{n}')^{-1} = \mathbf{n}'\cdot\boldsymbol{\sigma}, \\ &B(\mathbf{n},\mathbf{n}')\mathbf{n}\wedge\mathbf{n}'\cdot\boldsymbol{\sigma}B(\mathbf{n},\mathbf{n}')^{-1} = \mathbf{n}\wedge\mathbf{n}'\cdot\boldsymbol{\sigma}. \end{aligned} \tag{2.11}$$

As is to be expected, B is not unambiguously defined when $\mathbf{n}' = -\mathbf{n}$. We may remark that this construction of $B(\mathbf{n},\mathbf{n}') \in SU(2)$ is useful in computations of the Pancharatnam–Aharonov–Anandan phase[8] for two-level quantum systems, as shown in Ref. 2.

III. GENERALIZATION OF TURNS TO SU(1,1): THEORY OF SCREWS

We now show how a geometrical method can be developed for the noncompact group SU(1,1), similar in spirit to turns for SU(2). For convenience of exposition, this section is divided into subsections.

A. Notational preliminaries and definitions

The defining representation of SU(1,1) consists of two-dimensional complex pseudounitary unimodular matrices of the form

$$\begin{aligned} &A = \begin{pmatrix} \alpha & \beta \\ \beta^* & \alpha^* \end{pmatrix}, \\ &\det A = |\alpha|^2 - |\beta|^2 = 1, \\ &A^\dagger\sigma_3 A = \sigma_3. \end{aligned} \tag{3.1}$$

The geometrical constructions to follow will be in a three-dimensional Minkowski space $\mathscr{M}$ with vector indices $a,b,... = 0,1,2$ and diagonal metric η_{ab} with signature $(-,+,+)$. By adjoining factors of i to two of the Pauli matrices, we define the matrices ρ_a by

$$\rho_0 = \sigma_3, \quad \rho_1 = i\sigma_1, \quad \rho_2 = i\sigma_2. \tag{3.2}$$

Then

$$\rho_a \rho_b = -\eta_{ab} + i\epsilon_{ab}{}^c \rho_c, \tag{3.3}$$

where ϵ_{abc} is the Levi-Civita symbol with $\epsilon_{012} = 1$. For vectors x, y in $\mathscr{M}$ with inner product $x \cdot y = \eta_{ab} x^a y^b$, define the cross product $z = x \wedge y$ through

$$z^a = \epsilon^a{}_{bc} x^b y^c. \tag{3.4}$$

It is useful to introduce three unit vectors e_a along the coordinate directions in $\mathscr{M}$. For them we have

$$e_a \cdot e_b = \eta_{ab}, \quad e_a \wedge e_b = \epsilon_{ab}{}^c e_c. \tag{3.5}$$

The following identities are immediate consequences of these definitions:

$$x \cdot y \wedge z = x \wedge y \cdot z = y \cdot z \wedge x; \tag{3.6a}$$

$$x \wedge (y \wedge z) = x \cdot y \; z - x \cdot z \; y; \tag{3.6b}$$

$$x \cdot \rho y \cdot \rho = -x \cdot y + i \; x \wedge y \cdot \rho; \tag{3.6c}$$

$$(x \cdot y)^2 - (x \wedge y)^2 = x^2 y^2; \tag{3.6d}$$

$$w \cdot x \; y \wedge z + w \cdot y \; z \wedge x + w \cdot z \; x \wedge y = (x \cdot y \wedge z) \; w. \tag{3.6e}$$

The last identity arises by expanding $x \wedge (w \wedge (y \wedge z))$ in two different ways.

For a real scalar λ and a real vector μ in $\mathscr{M}$, consider the matrix

$$A = \lambda + i\mu \cdot \rho = \begin{pmatrix} \lambda + i\mu^0 & -\mu^1 + i\mu^2 \\ -\mu^1 - i\mu^2 & \lambda - i\mu^0 \end{pmatrix}. \tag{3.7}$$

It is clear upon comparing with Eq. (3.1) that $A \in \mathrm{SU}(1,1)$ if

$$\det A = \lambda^2 - \mu \cdot \mu = 1, \tag{3.8}$$

and, conversely, any $A \in \mathrm{SU}(1,1)$ can be expressed in the form (3.7) for unique real λ, μ satisfying (3.8). This is the analog of the use of homogeneous Euler parameters for SU(2).

The SU(1,1) to SO(2,1) homomorphism is easily set up. If $B, B' \in \mathrm{SU}(1,1)$, from Eqs. (3.3) and (3.6), it follows that

$$Bx \cdot \rho B^{-1} = x' \cdot \rho, \quad x'^a = \Lambda(B)^a{}_b x^b, \quad \Lambda(B) \in \mathrm{SO}(2,1), \quad \Lambda(B')\Lambda(B) = \Lambda(B'B). \tag{3.9}$$

Therefore

$$B(\lambda + i\mu \cdot \rho)B^{-1} = \lambda + i\mu' \cdot \rho, \quad \mu' = \Lambda(B)\mu. \tag{3.10}$$

Thus under conjugation of $A = \lambda + i\mu \cdot \rho$ by B, the scalar λ is invariant while the vector μ undergoes the SO(2,1) transformation determined by B. The explicit expression for $\Lambda(B)$ in terms of B can be easily worked out, but we do not need it in our analysis.

B. Classification of finite elements of SU(1,1): The exponential map

The representation (3.7) can be used to classify SU(1,1) elements in a convenient form. A vector $\mu \in \mathscr{M}$ is timelike (t), lightlike (l), or spacelike (s) according to whether $\mu^a \mu_a$ is negative, zero, or positive. We will say $A = \lambda + i\mu \cdot \rho \in \mathrm{SU}(1,1)$ is of type t, l, or s according to the nature of μ. This accounts for all elements except $A = \pm \mathbb{1}$ when μ vanishes. In the t and l cases, a further split into positive and negative types depending on the sign of μ^0 is possible. This classification results in five nonintersecting subsets of SU(1,1) whose union along with $\pm \mathbb{1}$ is SU(1,1).

Since $\mu^a \mu_a = \lambda^2 - 1$, the type of an element is fixed by $\lambda = \frac{1}{2} \operatorname{tr} A$. Assuming μ does not vanish identically,

$$\begin{aligned} -1 < \lambda < 1 \quad &\Leftrightarrow \quad A \text{ of type } t; \\ \lambda = \pm 1 \quad &\Leftrightarrow \quad A \text{ of type } l; \\ \lambda < -1 \text{ or } \lambda > 1 \quad &\Leftrightarrow \quad A \text{ of type } s. \end{aligned} \tag{3.11}$$

From Eq. (3.10) it follows that the type of an element is invariant under conjugation by any SU(1,1) element. For the t and l cases, the positive or negative nature is also preserved.

This classification of finite SU(1,1) elements is to be contrasted with a similar classification of elements in the Lie algebra of SU(1,1). The distinction is important because there are finite elements in the group that do not lie on any one-parameter subgroup at all. This is the meaning of the statement that SU(1,1) is not of exponential type. The exponential map takes the Lie algebra of SU(1,1) into a subset of the SU(1,1) group manifold. The complement of the range of this map can be easily characterized. The one-parameter subgroups of the types t, l, and s (with parameter τ) are

$$\begin{aligned} A_t(u;\tau) &= \cos\tau + iu \cdot \rho \sin\tau, \quad 0 \leqslant \tau \leqslant 2\pi, \quad u^2 = -1; \\ A_l(u,\tau) &= 1 + iu \cdot \rho\tau, \quad \tau \in \mathbb{R}, \quad u^2 = 0; \\ A_s(u;\tau) &= \cosh\tau + iu \cdot \rho \sinh\tau, \quad \tau \in \mathbb{R}, \quad u^2 = 1. \end{aligned} \tag{3.12}$$

Upon comparison with the classification of SU(1,1) elements, we see that the range of the exponential map consists of all elements of type t, elements of type l with $\lambda = 1$, of type s with $\lambda > 1$, and the two elements $\pm \mathbb{1}$. The complement of this range is therefore

$$\Phi = \{l \text{ type with } \lambda = -1\} \cup \{s \text{ type with } \lambda < -1\}. \tag{3.13}$$

In short, $\lambda = \frac{1}{2} \operatorname{tr} A \leqslant -1$ and $\mu \neq 0$ implies that A does not lie on any one-parameter subgroup; nevertheless A itself can be classified. We may also note that if $A \in \Phi$, then $A^{-1} \in \Phi$ as well.

In the sequel we will see the importance of this classification of SU(1,1) elements for the theory of screws.

C. Definition of the screw

Guided by the definition of turns for SU(2), for any two real three-vectors $x, y \in \mathscr{M}$ we define a complex two-dimensional matrix

$$A(x,y) = x \cdot y + i \; x \wedge y \cdot \rho. \tag{3.14}$$

From the identity (3.6d) we see that

$$\det A(x,y) = x^2 y^2. \tag{3.15}$$

Since $A(x,y)$ is of the general form (3.7), we see that it is an element of SU(1,1) if and only if $x^2 y^2 = 1$. That is, both x and y must be spacelike or both timelike, and have reciprocal norms. Henceforth we restrict our attention to spacelike vectors; the reason for this will be clear when we study the composition of screws. We next use the freedom of reciprocal scalings of x and y to arrange them both to be unit spacelike vectors.

In $\mathscr{M}$ define the single-sheeted spacelike unit hyperboloid

$$\Sigma = \{x | x\cdot x \equiv (x^1)^2 + (x^2)^2 - (x^0)^2 = 1\}. \tag{3.16}$$

The role of S^2 on which turns were defined for SU(2) will now be played by Σ. We know from Eqs. (3.14) and (3.15) that

$$x,y\in\Sigma \Rightarrow A(x,y)\in \mathrm{SU}(1,1). \tag{3.17}$$

We wish to prove the converse: any $A\in\mathrm{SU}(1,1)$ is of the form $A(x,y)$ for suitable x,y on Σ:

$$A \equiv \lambda + i\mu\cdot\rho = A(x,y) \equiv x\cdot y + ix\wedge y\cdot\rho. \tag{3.18}$$

If this is to be so, evidently, both x and y must be Lorentz orthogonal to μ, and also lie on Σ. Therefore for each μ we define

$$\mathscr{C}(\mu) = \{x | x\cdot x = 1, \quad \mu\cdot x = 0\}. \tag{3.19}$$

Clearly $\mathscr{C}(\mu)$ is the intersection of Σ with the plane through the origin orthogonal to μ. For μ of type t, l, or s, $\mathscr{C}(\mu)$ consists, respectively, of a closed ellipse, a pair of parallel disjoint infinite straight lines (generators of Σ), or a pair of disjoint infinite hyperbolas.

Theorem: Given any $A = \lambda + i\mu\cdot\rho\in\mathrm{SU}(1,1)$, we can choose $x, y\in\mathscr{C}(\mu)$ such that $A(x,y) = \lambda + i\mu\cdot\rho$. Moreover, either x or y can be chosen arbitrarily on $\mathscr{C}(\mu)$, the other being then uniquely fixed.

Proof: Since our entire treatment is manifestly SO(2,1) covariant, and under conjugation we have the behavior given in Eq. (3.10), we can with no loss of generality put μ of each type into a convenient standard configuration, and then carry out the construction.

<u>A of type t:</u> We can take

$$\mu = \xi e_0, \quad \xi = \sin\theta, \quad \lambda = \cos\theta, \quad 0<\theta<2\pi, \quad \theta\neq\pi. \tag{3.20}$$

Then

$$\begin{aligned}\mathscr{C}(\mu) &= \{x | x^0 = 0, \quad (x^1)^2 + (x^2)^2 = 1\}\\ &= \{(0,\cos\varphi, \quad \sin\varphi) | 0\leqslant\varphi<2\pi\}.\end{aligned} \tag{3.21}$$

If $x, y\in\mathscr{C}(\mu)$ correspond to parameter values φ,φ', respectively,

$$\begin{aligned}&\lambda + i\mu\cdot\rho = A(x,y) \Leftrightarrow\\ &\lambda = x\cdot y, \quad \mu = x\wedge y \Leftrightarrow\\ &\varphi - \varphi' = \theta.\end{aligned} \tag{3.22}$$

Clearly, if θ is known, either x or y can be freely chosen on $\mathscr{C}(\mu)$, the other is then uniquely fixed.

<u>A of type l:</u> We can set

$$\mu = \epsilon'(e_0 + e_2), \quad \lambda = \epsilon, \quad \epsilon,\epsilon' = \pm 1. \tag{3.23}$$

Then $\mathscr{C}(\mu)$ is a pair of generators of Σ:

$$\begin{aligned}\mathscr{C}(\mu) &= \{x | x^0 = x^2, \quad (x^1)^2 = 1\}\\ &= \{(a,\delta,a) | a\in\mathbb{R}, \quad \delta = \pm 1\}.\end{aligned} \tag{3.24}$$

By taking $x = (a,\delta,a)$ and $y = (a',\delta',a')$,

$$\begin{aligned}&\lambda + i\mu\cdot\rho = A(x,y) \Leftrightarrow,\\ &\delta\delta' = \epsilon, \quad a\delta' - a'\delta = \epsilon'.\end{aligned} \tag{3.25}$$

If ϵ and ϵ' are known, any choice of (a,δ) leads to unique (a',δ') and the converse. If $\epsilon = 1$, x and y are on the same branch of $\mathscr{C}(\mu)$; if $\epsilon = -1$, they are not.

<u>A of type s:</u> We can take

$$\begin{aligned}&\mu = \xi e_2, \quad \xi = \epsilon\sinh\zeta, \quad \lambda = \epsilon\cosh\zeta,\\ &\epsilon = \pm 1, \quad \zeta\in\mathbb{R}, \quad \zeta\neq 0.\end{aligned} \tag{3.26}$$

Then $\mathscr{C}(\mu)$ consists of two branches of a hyperbola on Σ:

$$\begin{aligned}\mathscr{C}(\mu) &= \{x | x^2 = 0, \quad (x^1)^2 - (x^0)^2 = 1\}\\ &= \{(\delta\sinh\eta, \quad \delta\cosh\eta, 0) | \eta\in\mathbb{R},\\ &\qquad \delta = \pm 1\}.\end{aligned} \tag{3.27}$$

By choosing parameters δ, η for x and δ', η' for y,

$$\lambda + i\mu\cdot\rho = A(x,y) \Leftrightarrow \delta\delta' = \epsilon, \quad \eta - \eta' = \zeta. \tag{3.28}$$

Once again, either x or y can be freely chosen on $\mathscr{C}(\mu)$, the other being then completely determined. If $\epsilon = +1$, x and y lie on the same branch of $\mathscr{C}(\mu)$, otherwise not.

This completes the proof, except for the remark that for $A = \pm \mathbb{1}$, we can set $x = \pm y$ on Σ and then choose x freely, i.e.,

$$x\in\Sigma: A(x, \pm x) = \pm \mathbb{1}. \tag{3.29}$$

We are now able to define a screw precisely. Notice that with $A = \lambda + i\mu\cdot\rho\in\mathrm{SU}(1,1)$, the SO(2,1) "rotation" $\Lambda(A)$ leaves $\mu\in\mathscr{M}$ invariant, and alters only vectors in the plane orthogonal to μ. By borrowing the familiar SO(3) language, we can say that $\Lambda(A)$ is a Lorentz rotation about μ as axis. A screw is then an equivalence class of ordered pairs of points (x,y) on Σ, the equivalence being with respect to common SO(2,1) transformations of both x and y about $x\wedge y$ as axis and, in case $x\wedge y$ is not of type t, the transformation $x\to -x$, $y\to -y$ as well. Since $A(x,y)$ is clearly invariant under such transformations, it follows from the above theorem that there is a one-to-one correspondence between screws and SU(1,1) elements, determined by Eq. (3.14). Given $x,y\in\Sigma$ with $x\neq\pm y$, there is a unique $\mathscr{C}(\mu)$ on which x and y lie, and then the equivalence is with respect to motions along $\mathscr{C}(\mu)$ induced by SO(2,1) rotations about $\mu = x\wedge y$, along with the "reflection" $x\to -x$, $y\to -y$ in case μ is not of type t. For $A = \mathbb{1}$, we have the "null screw" given by (x,x) for any $x\in\Sigma$; and for $A = -\mathbb{1}$, any pair $(x, -x)$ can be used.

D. Screws as directed arcs on Σ

We have just defined a screw as an equivalence class of ordered pairs of points (x,y) on Σ. One may wish, analogous to turns for SU(2), to define a screw as (the equivalence class of) the directed arc from x to y along $\mathscr{C}(\mu = x\wedge y)$. However, in the proof of the theorem of the previous subsection, we found that among l- and s-type elements of SU(1,1), there are situations when x and y have to be chosen on distinct branches of $\mathscr{C}(\mu)$. Now we see that these elements are precisely the ones in $\Phi\subset\mathrm{SU}(1,1)$, the complement of the range of the exponential map. So at first sight it appears that for elements in Φ, screws cannot be visualized as connected arcs on Σ. There is, however, a way of overcoming this problem, which does not exist for turns for SU(2). The clue is that A and $-A$ in SU(1,1) share the same $\mathscr{C}(\mu)$, and if $A \in \Phi$, then $-A \notin \Phi$. Hence, if $A \in \Phi$, the ordered pair of points corresponding to $-A$ can definitely be connected by a directed arc along $\mathscr{C}(\mu)$. This leads to an alternative definition of a screw.

A screw is a pair consisting of an equivalence class of connected directed arcs along a $\mathscr{C}(\mu)$, the equivalence being with respect to SO(2,1) transformations that map $\mathscr{C}(\mu)$ onto itself and with respect to reflection if μ is not of type t; and a flag that can take the values ± 1. Given $A = A(x,y) \notin \Phi$, its screw is (the equivalence class of) the directed arc from x to y along $\mathscr{C}(x \wedge y)$, with the flag $+1$; if however $A = A(x,y) \in \Phi$, then the screw is represented by (the class of) the directed arc for $-A = A(x,-y) \notin \Phi$, with the flag -1.

We shall return to this use of the flag after discussing the geometrical composition procedure for screws.

E. Composition of screws

From the representation (3.7) for a general $A \in$ SU(1,1), and the properties of the matrices ρ, we can see that passage to the inverse corresponds to reversing the sign of μ but leaving λ unchanged:

$$A = \lambda + i\mu \cdot \rho \Rightarrow A^{-1} = \lambda - i\mu \cdot \rho. \qquad (3.30)$$

It is then clear that, given the screw for A, the screw for A^{-1} is obtained by interchanging the entries in the ordered pair of vectors (x,y), x and $y \in \mathscr{C}(\mu)$, or equivalently by reversing the directed arc without altering the flag.

To develop a geometrical rule for the composition of screws, we first give an algebraic result following from the construction (3.14), the form of which is suggested by the result (2.8b) in the SU(2) case:

$$x,y,z \in \mathscr{M}: \quad A(z,y)A(x,z) = z^2 A(x,y). \qquad (3.31)$$

The proof is quite straightforward, and involves judicious use of the various identities (3.6). In fact, the precise construction of $A(x,y)$ in Eq. (3.14) was motivated by the desire to have the result (3.31). The idea now is to see if (3.31) can be exploited to convert the rule for composition of any two SU(1,1) elements into a geometrical operation on Σ.

If two elements $A, A' \in$ SU(1,1) are expressed in the form (3.7), their product can be put into the same form:

$$\begin{aligned} A'' &= A'A = (\lambda' + i\mu' \cdot \rho)(\lambda + i\mu \cdot \rho) = \lambda'' + i\mu'' \cdot \rho, \\ \lambda'' &= \lambda'\lambda + \mu' \cdot \mu, \qquad (3.32) \\ \mu'' &= \lambda'\mu + \lambda\mu' - \mu' \wedge \mu. \end{aligned}$$

On the other hand, if in Eq. (3.31) all three vectors x,y,z are chosen on Σ, then both $A(x,z)$ and $A(z,y)$ will belong to SU(1,1) and

$$x,y,z \in \Sigma: A(z,y)A(x,z) = A(x,y). \qquad (3.33)$$

We can conclude that if, given the two general elements A, $A' \in$ SU(1,1), we determine $\mathscr{C}(\mu)$ and $\mathscr{C}(\mu')$ and find that they intersect, we can then choose $z \in \mathscr{C}(\mu) \cap \mathscr{C}(\mu')$; from Sec. III C we are then assured that $x \in \mathscr{C}(\mu)$ and $y \in \mathscr{C}(\mu')$ exist uniquely such that $A = (x,z)$ and $A' = A(z,y)$. Then $A'A$ is determined by Eq. (3.33): the ordered pair for $A'A$ consists of the first member of the A pair and the second member of the A' pair, in that order. However, unlike the SU(2) case where we know that any two great circles on S^2 definitely intersect, in the Lorentzian geometry of $\mathscr{M}$ it does sometimes happen that $\mathscr{C}(\mu) \cap \mathscr{C}(\mu') = \phi$! We are thus led to the question: when is $\mathscr{C}(\mu) \cap \mathscr{C}(\mu') \neq \phi$?

In general, if two vectors $\mu, \mu' \in \mathscr{M}$ are given, assumed to be linearly independent, then the planes through the origin orthogonal, respectively, to μ and to μ' will intersect along a straight line:

$$\mu \cdot x = \mu' \cdot x = 0 \Rightarrow x = \alpha \mu \wedge \mu'. \qquad (3.34)$$

Here α is a parameter along the line. This line will cut Σ if there is a real value of α for which

$$\alpha^2 (\mu \wedge \mu')^2 \equiv \alpha^2 ((\mu \cdot \mu')^2 - \mu^2 \mu'^2) = 1, \qquad (3.35)$$

which will happen if and only if

$$(\mu \cdot \mu')^2 - \mu^2 \mu'^2 > 0. \qquad (3.36)$$

We can now systematically analyze the six possible kinematical situations, listing the nature of the pair μ, μ' at tt, tl, ts, ll, ls, and ss, and check in each whether the inequality (3.36) can be obeyed. (Naturally, in the "off-diagonal" cases, such as tl, it does not matter which of μ and μ' is of type t and which of type l.) Keeping in mind the nature of Lorentzian geometry, we find that in four situations, the inequality is uniformly obeyed:

$$tt, tl, ts, ll: (\mu' \cdot \mu)^2 - \mu'^2 \mu^2 > 0. \qquad (3.37)$$

However, in the two remaining cases ls and ss, no uniform statement can be made; depending on the specific choices of μ and μ', $(\mu' \cdot \mu)^2 - \mu'^2 \mu^2$ could have either sign.

The result is that if for the two elements A, A' in SU(1,1), at least one of the vectors μ, μ' is of type t, or if both are of type l, $\mathscr{C}(\mu')$ and $\mathscr{C}(\mu)$ definitely intersect on Σ; then Eq. (3.33) is adequate to give us a geometrical "addition" or composition law for screws, in a way similar to the use of Eq. (2.8b) for SU(2). But in the ls and ss cases, there is no guarantee that $\mathscr{C}(\mu')$ and $\mathscr{C}(\mu)$ will intersect; and if they do not, a choice of one common vector z as in Eq. (3.33) is not possible.

Fortunately the following decomposition theorem, which exhibits an interesting structural property of SU(1,1) comes to our aid, so that Eq. (3.33) can be used for composing screws in all situations:

Theorem: Any $A'' \in$ SU(1,1) can be expressed (in infinitely many ways) as a product $A'' = A'A$ where both factors A' and A are of type t.

Proof: We naturally construct the argument in the spirit of the theory of screws, and use the geometrical machinery in $\mathscr{M}$ already developed. Assuming $A'' \neq \pm \mathbb{1}$, we identify $\mu'' \neq 0$, construct $\mathscr{C}(\mu'') \subset \Sigma$, and pick $x, y \in \mathscr{C}(\mu'')$ such that

$$A'' = \lambda'' + i\mu'' \cdot \rho = A(x,y). \qquad (3.38)$$

Here x and y will be linearly independent, $x \neq \pm y$. Since each of them is spacelike, there are infinitely many timelike vectors n orthogonal to x, and similarly n' orthogonal to y. There are then infinitely many choices of n and n' obeying

$$\begin{aligned} &n^2, n'^2 < 0; \quad n \cdot x = n' \cdot y = 0; \\ &n \cdot y \neq 0, \quad n' \cdot x \neq 0. \end{aligned} \qquad (3.39)$$

These conditions are designed to ensure that n and n' are linearly independent and neither is proportional to $x\wedge y$. Since both n and n' are timelike, $n\wedge n'$ is spacelike and so when normalized will cut Σ:

$$\begin{aligned} &z=\alpha n\wedge n'\in\Sigma,\\ &\alpha=((n\cdot n')^2-n^2n'^2)^{-1/2}. \end{aligned}\tag{3.40}$$

Now we construct the SU(1,1) elements $A=A(x,z)$, $A'=A(z,y)$. Both of them are of type t since by Eq. (3.39)

$$\begin{aligned} &x\wedge z=\alpha x\wedge(n\wedge n')=-\alpha x\cdot n'n,\\ &z\wedge y=\alpha(n\wedge n')\wedge y=-\alpha y\cdot nn', \end{aligned}\tag{3.41}$$

are both nonvanishing timelike vectors. Finally,

$$A''=A(x,y)=A(z,y)A(x,z)=A'A,\tag{3.42}$$

which proves the result for $A''\neq\pm\mathbb{1}$. If $A''=\pm\mathbb{1}$, we can take A to be any element of type t, and $A'=\pm A^{-1}$.

To see by way of illustration how the choices of A and A' may be made, we can consider in turn A'' to be of type t,l,s. If A'' is already of type t, it lies on a t-type, one-parameter subgroup, so for instance A and A' could be chosen equal and essentially the square root of A''. If A'' is of type l or s, the situation is nontrivial. In these cases we can use the manifest covariance under SO(2,1), i.e., the conjugation relation (3.10), and assume without loss of generality that λ'' and μ'' are in some standard configuration. This makes possible choices of x,y,n,n',z easy to visualize. We record below the standard forms of λ'' and μ'', possibilities for n and n', and the resulting z, leaving it to the reader to check that all conditions (3.32), (3.38)–(3.40), and (3.42) are obeyed.

<u>A'' of type l:</u>

$$\begin{aligned} &\lambda''=\epsilon,\quad \mu''=\epsilon'(e_0+e_2),\quad \epsilon,\epsilon'=\pm1;\\ &x=e_1,\quad y=\epsilon e_1-\epsilon'(e_0+e_2);\\ &n=e_0,\quad n'=\epsilon'(2e_0+e_2)-\epsilon e_1;\\ &z=(\epsilon'e_1+\epsilon e_2)/\sqrt{2};\\ &\lambda=\epsilon'/\sqrt{2},\quad \mu=-\epsilon e_0/\sqrt{2};\\ &\lambda'=0,\quad \mu'=(2e_0-\epsilon\epsilon'e_1+e_2)/\sqrt{2}. \end{aligned}\tag{3.43}$$

<u>A'' of type s:</u>

$$\begin{aligned} &\lambda''=\epsilon\cosh\zeta,\quad \mu''=\epsilon\sinh\zeta e_2,\quad \epsilon=\pm1,\quad \zeta\neq0;\\ &x=e_1,\quad y=\epsilon(-\sinh\zeta e_0+\cosh\zeta e_1);\\ &n=e_0,\quad n'=\cosh\zeta e_0-\sinh\zeta e_1;\\ &z=e_2;\\ &\lambda=0,\quad \mu=-e_0;\\ &\lambda'=0,\quad \mu'=\epsilon(\cosh\zeta e_0-\sinh\zeta e_1). \end{aligned}\tag{3.44}$$

With the help of our theorem, then, the product AB of any two elements $A,B\in$SU(1,1) can be handled geometrically as an operation on screws, requiring at most two applications of Eq. (3.33). If A, B belong to one of the four cases tt, tl, ts, or ll, we "slide" the representative pairs of points on $\mathscr{C}(\mu_A)$, $\mathscr{C}(\mu_B)$ till the "head" (second element) of the B pair and the "tail" (first element) of the A pair become $z\in\mathscr{C}(\mu_A)\cap\mathscr{C}(\mu_B)$. Then a single use of Eq. (3.33) gives the screw for AB as determined by the pair (tail of B, head of A). If A,B belong to either the ls or ss cases, and $\mathscr{C}(\mu_A)\cap\mathscr{C}(\mu_B)=\phi$, we use the decomposition theorem to write $AB=A''A'B$, with both A'' and A' being of type t. The screws for A' and B can then be composed using (3.33) to give the screw for $A'B$; this can then be composed with the screw for A'', using (3.33) again, to give the final result.

We now go back to the device of the flag, introduced in Sec. III D so as to allow us to visualize *every* $A\in$SU(1,1) as a connected arc on Σ plus a flag, and show how it can be represented graphically. As noted in Eq. (3.29), the element $-\mathbb{1}$ in SU(1,1) corresponds to the (degenerate) screw determined by any pair $(x,-x)$:

$$x\in\Sigma:\quad A(x,-x)=-\mathbb{1}.\tag{3.45}$$

This is just like the great semicircle turn in the SU(2) case. Since x is spacelike, there are infinitely many timelike vectors μ orthogonal to it, which means there are infinitely many connected $\mathscr{C}(\mu)$'s of type t containing both x and $-x$. [In addition, there are infinitely many s-type $\mathscr{C}(\mu)$'s, and two l-type $\mathscr{C}(\mu)$'s, each made up of two branches, and each containing x and $-x$, but on separate branches; these however are not useful for the present purpose.] This degenerate screw, representable by a connected arc on any one of these t-type $\mathscr{C}(\mu)$'s, and in fact running halfway across it, is indeed the flag we have used above.

As an example, consider the element

$$A=\begin{pmatrix}-\cosh\theta & -\sinh\theta\\ -\sinh\theta & -\cosh\theta\end{pmatrix}\in\Phi\subset\mathrm{SU}(1,1),\tag{3.46}$$

for which

$$\lambda=-\cosh\theta,\quad \mu=e_1\sinh\theta.\tag{3.47}$$

Then $\mathscr{C}(\mu)$ consists of two branches of the hyperbola

$$(x^2)^2-(x^0)^2=1\tag{3.48}$$

in the 0-2 plane, as in Fig. 1. A choice of x,y so that $A=A(x,y)$ is

$$\begin{aligned} &x=(\sinh(\theta-\theta_0),\quad 0,\quad \cosh(\theta-\theta_0)),\\ &y=(\sinh\theta_0,\quad 0,\quad -\cosh\theta_0), \end{aligned}\tag{3.49}$$

where θ_0 may be chosen freely. Naturally they are on sepa-

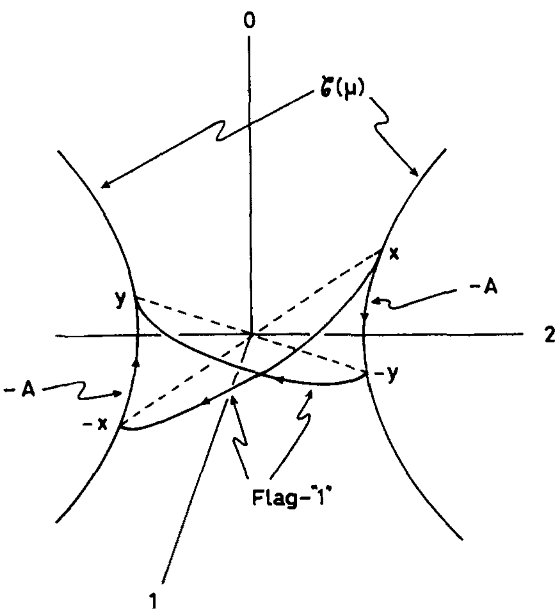

FIG. 1. Use of the flag $-$ "1."

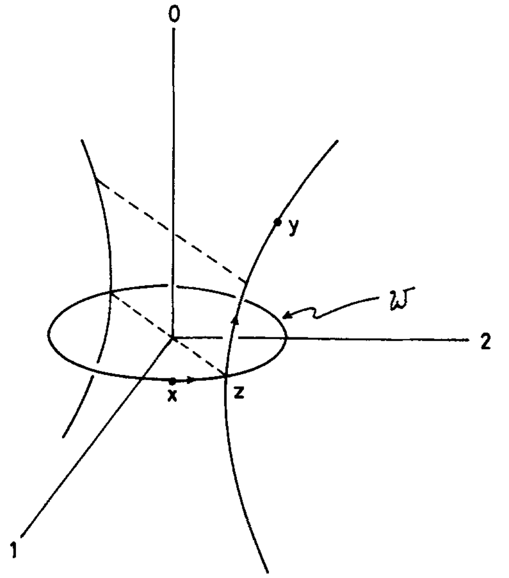

FIG. 2. Rotation boost decomposition of general element.

rate branches of $\mathscr{C}(\mu)$. Now, $-A = A(x, -y) = A(-x,y)$ is in the range of the exponential map, and can be represented by the connected arc $x \rightarrow -y$ or $-x \rightarrow y$ as we wish, along one branch of $\mathscr{C}(\mu)$. Coupled, respectively, with the flag screws $A(-y,y)$ or $A(x,-x)$ standing for -1, we get back $A(x,y)$ via Eq. (3.33):

$$A = A(x,y) = A(-y,y)A(x,-y) = A(-x,y)A(x,-x). \tag{3.50}$$

This is graphically seen in Fig. 1.

IV. CONCLUDING REMARKS

In this paper we have presented a generalization of Hamilton's method of turns for SU(2) to a theory of screws for SU(1,1), thus leading to a new and useful way of picturing the elements and structure of this noncompact group. The two distinguishing features of SU(1,1), as contrasted with SU(2), are that the range of the exponential map is a subset of SU(1,1); and that two planes passing through the origin in $\mathscr{M}$ may in some cases have a line of intersection that does not cut Σ. These features, which at first sight seem to pose problems for the development of a complete geometrical picture, can be taken care of by the use of the flag -1 when appropriate, and the use of the theorem of Sec. III E, so that all group elements and their products can be satisfactorily handled.

As an example of the usefulness of our geometrical picture for SU(1,1), we recall the result that any element of SU(1,1) is a "boost" in an appropriate direction followed by a rotation [element of the maximal compact subgroup U(1)], or a rotation followed by a boost. This fact is immediately and visually obvious in the screw representation, requiring no calculation at all! Elements of the U(1) subgroup have μ parallel to e_0, so for them $\mathscr{C}(\mu)$ is the unit circle in the 1-2 plane, or the "waist" $\mathscr{W}$ of Σ:

$$\mathscr{W}: \quad x^0 = 0, \quad (x^1)^2 + (x^2)^2 = 1.$$

This is like the equator on S^2 in the turns case. On the other hand, pure boosts are elements for which μ is a linear combination of e_1 and e_2. For such a μ, $\mathscr{C}(\mu)$ is the intersection of Σ with a "vertical" plane containing the e_0 axis; the two branches of the hyperbola making up $\mathscr{C}(\mu)$ are like "lines of longitude" on S^2. Now given a screw determined by the pair (x,y), we can assume without loss of generality that either x or y, whichever we wish, lies on $\mathscr{W}$. This is because *every* $\mathscr{C}(\mu)$ is guaranteed to intersect $\mathscr{W}$. Assuming for definiteness then that $x \in \mathscr{W}$, we can draw a hyperbola on Σ in the vertical plane containing y and the e_0 axis. If this cuts $\mathscr{W}$ at a point z, we can recover the pair (x,y) by composing the screws (x,z) and (z,y) in that order. [Note incidentally that both these are connected arcs, whereas no assumption was made about x and y being on a connected branch of $\mathscr{C}(x,y)$.] This is the proof by the present method that any SU(1,1) element is a pure rotation followed by some pure boost, as depicted in Fig. 2. If we had $y \in \mathscr{W}$ instead, the decomposition would have been in the opposite order.

We may note that the t,l,s classification of finite SU(1,1) elements is relevant to periodically focusing optical systems.[9] An example is a laser resonator. One sees that the system is stable if and only if the ray-transfer matrix for one period is a t-type element of SL(2,R). An interesting application of the theorem of subsection 3.5 is to squeezing that is an s-type element of SU(1,1). The theorem shows that squeezing can be realized by switching periodically between two t-type nonsqueezing transformations. A detailed analysis of these questions, as well as the development of a new representation for first-order Fourier optical systems, will be presented elsewhere.

ACKNOWLEDGMENT

This work was supported in part by the Department of Energy under Grant No. DE-FG 05-85 ER 40200.

[1]See, for instance, L. C. Biedenharn and H. van Dam, *Quantum Theory of Angular Momentum* (Academic, New York, 1965), and Ref. 4 below.
[2]R. Simon, N. Mukunda, and E. C. G. Sudarshan, "Hamilton's theory of turns and a new geometrical representation for polarization optics," preprint, Indian Institute of Science, Bangalore and Dept. of Physics, Univ. of Texas, Austin, Texas (1988).
[3]W. R. Hamilton, *Lectures on Quaternions* (Dublin, 1853).
[4]L. C. Biedenharn and J. D. Louck, "Angular momentum in quantum physics," *Encyclopedia of Mathematics and its Applications* (Addison–Wesley, Reading, MA, 1981), Vol. 8. This contains a detailed historical account of the subject, clarifying Hamilton's own contributions and those of later authors.
[5]A brief account of our main results has been presented in R. Simon, N. Mukunda, and E. C. G. Sudarshan, "Hamilton's theory of turns generalised to SU(1,1)," preprint, Indian Institute of Science, Bangalore, and Dept. of Physics, Univ. of Texas, Austin, Texas.
[6]H. Bacry and M. Cadilhac, Phys. Rev. A **23**, 2533 (1981); E. C. G. Sudarshan, N. Mukunda, and R. Simon, Opt. Acta **32**, 855 (1985); R. Simon, E. C. G. Sudarshan, and N. Mukunda, Phys. Rev. A **29**, 3273 (1984).
[7]D. F. Walls, Nature **306**, 141 (1983).
[8]S. Pancharatnam, Proc. Indian Acad. Sci. Sec. A **44**, 247 (1956); M. V. Berry, Proc. R. Soc. London **392**, 45 (1984); Y. Aharonov and J. Anandan, Phys. Rev. Lett. **58**, 1593 (1987).
[9]A. E. Siegman, *Lasers* (Oxford U.P., Oxford, 1986).

The Schwinger functions of a rational interaction: Local existence of the Borel transform

G. A. T. F. da Costa
Instituto de Física Teórica, Rua Pamploma, 145, CEP 01405, São Paulo SP, Brazil
M. Gomes
Instituto de Física da USP, Departmento de Física Matemática, C.P. 20516, São Paulo, S.P., Brazil

(Received 25 May 1988; accepted for publication 23 November 1988)

The local existence of the Borel transform of a two-dimensional field theoretical model characterized by the rational interaction $g^2x^6/(1+gx^2)$ is proven.

I. INTRODUCTION

In a recent paper,[1] we proved the Borel summability of the perturbative series for the energies of the quantum mechanical system characterized by the nonpolynomial interaction

$$g^2x^6/(1+gx^2)\,. \tag{1}$$

The proof of Borel summability was done by establishing that the perturbative spectrum is strongly asymptotic and then applying a modified version of Nevanlinna's theorem.[2] Here we want to extend these studies considering (1) as the self-interaction of a quantum field. To achieve renormalizability, due to the nonpolynomial character of the interaction (1), we shall restrict our analysis to two dimensions. We are then able to prove the local existence of the Borel transform of the Schwinger functions of the Euclidean theory. This is a first step towards a complete proof of Borel summability.

II. THEORY

The physical motivation for considering the interaction (1) comes from laser theory models where the reduction of the Fokker–Planck to the Schrödinger equation produces interactions similar to the above one (see, for example, Ref. 3 and the references mentioned therein). Besides that, the study of (1) has its own merits for, as it is a rational function, the perturbative series is singular both due to the bad behavior at x large and to the poles occurring in the denominator of the potential.

Basically, there are two reasons why, in general, the perturbative series in divergent. First, the number of diagrams can grow too fast (typically with $n!$) with the order n of perturbation and, second, some individual diagrams dominate giving too large contributions. The second reason is peculiar to renormalizable theories whereas the first phenomena occurs also in superrenormalizable models like ours. We will have to find therefore bounds in the number of diagrams that contribute at a given order. An important result concerning this is the following lemma.

Lemma 1: The number, $\gamma(n)$, of connected diagrams with V vertices contributing to order n to the E point Schwinger function of the two-dimensional model with Lagrangian

$$L=\frac{1}{2}\,\partial_\mu\varphi\,\partial^\mu\varphi+\frac{1}{2}\,m^2\varphi^2+\frac{g^2\varphi^6}{1+g\varphi^2} \tag{2}$$

satisfies the inequality

$$\gamma(n)<2^{7n}(n!)(V!)(E!)n^E\,. \tag{3}$$

To prove (3), it is convenient to collect some combinatoric relations valid for a generic connected graph G of our model.

Lemma 2: In a graph G contributing to the Schwinger functions of the model (2), the following relations are valid:

1) $L=I-V+1$,

2) $n=\frac{1}{2}\sum kV_k-V$,

3) $n=I+E/2-V$,

4) $V\leqslant n/2$,

5) $n=L-1+E/2$. (Thus $n=L$ if $E=2$.)

6) $I<\frac{3}{2}n-E/2+1<\frac{3}{2}n+1$,

7) $L<\frac{3}{2}n-E/2+2$,

8) $I<\frac{3}{2}n,\quad$ if $E=2$,

9) $\sum kV_k=E+2I$,

10) $6V\leqslant\sum kV_k\leqslant 3n$,

where $L=$ number of loops; $E=$ number of external lines; $I=$ number of internal lines; $V=$ number of vertices.

Relation (1) is just Euler's relation expressing the number of loops of a diagram in terms of the number of internal lines and vertices. To verify the other relations, we use

$$\frac{g^2\varphi^6}{1+g\varphi^2}=\sum_{\nu=2}(-1)^\nu g^\nu\varphi^{2\nu+2}\,. \tag{4}$$

A k vertex is a vertex in which k lines are met. From (4) a k vertex has a factor $g^{k/2-1}$. Thus if the graph has V_k k vertices then the diagram will be of order $n=\Sigma_k(k/2-1)V_k=\frac{1}{2}\Sigma_k kV_k-V$, which proves relation 2). Relation 3) is proven using relation 2) and the relation $\frac{1}{2}\Sigma kV_k=I+E/2$ obtained by counting the number of line endings in G. Now, since at a given vertex there are at least six lines, $n=\Sigma_k(k/2-1)V_k\leqslant 2\Sigma_k V_k=2V$, giving relation 4). The other relations are proven similarly.

The counting of the number of diagrams is greatly facili-

 0022-2488/89/051007-02$02.50

tated by a method developed in Ref. 4. Let G be a proper diagram. We can always draw it so that all its vertices are on a circumference. Therefore two lines of any given vertex of G are used to link it to two other vertices. There are $\binom{k}{2}$ ways of choosing two fields from the fields associated to a k vertex. Once these fields are chosen there is an additional factor, 2^{V-1}, coming from the different ways of contracting them. As $\binom{k}{2} \leqslant 2^k$, the number of ways of choosing two lines from each vertex is less than or equal to $2^{V-1} 2^{\Sigma\, k V_k} < 2^{4n}$.

To some vertices there will be also attached external lines. The number of ways of selecting the fields associated to a given k vertex so that j fields will be contacted with the external fields is $\binom{k-2}{j} < 2^{k-2}$. Hence, considering all the vertices of the diagram we get a number of contributions less or equal to $2^{\Sigma (k-2) V_k} \leqslant 2^{2n}$, using Lemma 2. The number of external lines j at a given k vertex can vary from zero to $k-2$. The consideration of these possibilities gives a factor less or equal to V^E. Moreover, having fixed the number of external lines at each vertex, there are $E!$ ways of attaching the external lines to them.

After having distributed the external lines there will be still a number l of fields to be contracted. These contractions will produce $(l-1)!!$ graphs. The number of diagrams with V vertices has also factor $V!$ coming from the permutations of the vertices. Putting all this together we arrive to the conclusion that the number of diagrams with V vertices is bounded by

$$\beta = 2^{6n} V! E! V^E (l-1)!! \tag{5}$$

The number l can be written in terms of n and E, $l = 2n - E$. Indeed, from its definition $l = 2I - 2V = 2n - E$, where (3) in the Lemma 2 has been used. Then

$$\beta = V! E! V^E 2^{7n - E/2} (n - E/2)! \leqslant V! E! V^E 2^{7n} n!. \tag{6}$$

This number bounds the number of graphs having a fixed, $\{V_k; k \geqslant 6\}$, configuration of vertices. The result (3) follows from relation (4) of Lemma 2.

The perturbative expansion for the Schwinger functions are obtained from (1) by expanding it in powers of g. The only divergences that are found in this process are associated to tadpole diagrams (graphs having just one loop and one internal line) that are removed by Wick ordering the interaction Lagrangian with respect to the mass m. In this situation it can be proven[5] that in order n individual amplitudes are bounded, i.e.,

$$A_n \leqslant K'^n, \tag{7}$$

where K' is a constant independent of the topology of the associated diagram. Now, the order n total amplitude, $G_n^{(E)}$, is defined as

$$G_n^{(E)} = (-1)^n \sum{}' (-1)^V \frac{A_{n,\{V_k\}}}{\Pi_k V_k!}, \tag{8}$$

where the sum Σ' is over all possible assignments $\{V_k\}; k \geqslant 6$ to the vertices such that $\Sigma (k/2 - 1) V_k = n$. Here, $A_{n,\{V_k\}}$ is the sum of all possible graphs corresponding to a given assignment. It can be easily seen that, in a given order n, there are not more than 2^n configurations satisfying $\Sigma (k/2 - 1) V_k = n$.

From (7), (8), and (3) we have therefore that the order n total amplitude is bounded by

$$2^{8n} n^E n! E! \sup \left(\frac{V!}{\Pi_k V_k!} \right) K'^n, \tag{9}$$

where the supremum is to be taken over all configurations satisfying $\Sigma (k/2 - 1) V_k = n$. This supremum is bounded by a positive constant to the power n. Indeed, we have

$$\frac{V!}{\Pi_k V_k!} = \binom{\Sigma_{k>6} V_k}{V_6} \binom{\Sigma_{k>8} V_k}{V_8} \cdots$$

$$\leqslant 2^{\Sigma_{k>6} V_k} 2^{\Sigma_{k>8} V_k} \cdots$$

$$= 2^{\Sigma_{k>6} (k/2 - 2) V_k} \leqslant 2^n. \tag{10}$$

We have therefore

$$|G_n^{(E)}| \geqslant 2^{9n} (n!) (E!) n^E K^n. \tag{11}$$

This equation implies that the perturbative series has a Borel transform free of singularities within a ball of radius $1/K$ with the center at the origin of the Borel plane.

III. CONCLUSION

We can now enunciate our main result.

Theorem: The expansion

$$\sum B_n b^n, \quad B_n = \frac{G_n^{(E)}}{n!} \tag{12}$$

for the Borel transform of the E point Schwinger functions of the theory (2) converges within a circle of radius $1/K$ where it defines an analytic function of b.

To go further, having proved the local existence of the Borel transform, we now need to extend its domain of analiticity to a neighborhood of the positive real axis in the complex Borel plane. Work in this direction is in progress.

ACKNOWLEDGMENT

The authors wish to thank V. Rivasseau for helpful discussions and a critical reading of the manuscript.

The work of one author (G.A.T.F.C.) was partially supported by FAPESP and the work of the other author (M.G.) was partially supported by CNPq.

[1] G. A. T. F. da Costa and M. Gomes, "A nonpolynomial interaction: Stability and Borel summability," preprint, Instituto de Fisica Teorica IFT/P-11/88, 1988.

[2] A. Sokal, J. Math. Phys. **21**, 2 (1980).

[3] H. Risken and H. D. Vollmer, Z. Phys. **201**, 323 (1967); H. Haken, "Laser Theory," in *Encyclopedia of Physics* (Van Nostrand, Princeton, NJ, 1970).

[4] A. Jaffe, Commun. Math. Phys. **1**, 127 (1965).

[5] J. Feldman, J. Magnen, V. Rivasseau, and R. Seneor, Commun. Math. Phys. **98**, 273 (1985).

A new derivation of the quadrupole formula for angular momentum

E. Nahmad-Achar

Department of Applied Mathematics and Astronomy, University College, Cardiff, United Kingdom and Instituto de Ciencias Nucleares,[a] *Universidad Nacional Autónoma de México, Apartado Postal 70-543, México D. F. 04510, Mexico*

(Received 9 August 1988; accepted for publication 30 November 1988)

A new derivation of the farfield quadrupole formula for radiated angular momentum is presented, based on the gravitational Noether operator. Those metrics for which the angular momentum flux is well defined are characterized, and it is shown that this is free of the supertranslation ambiguity in the Newtonian approximation.

I. INTRODUCTION

According to the theory of general relativity (as well as other theories of gravity), systems of moving masses and/or varying electromagnetic fields emit gravitational waves (just as a system of moving charges emits electromagnetic waves). These waves will, in general, carry away energy and angular momentum, so that the study of the fluxes of these quantities is a direct study of the evolution of the system.

Due to the complexity of working with exact solutions of the (nonlinear) field equations, in general, and perhaps also due to the fact that all such solutions available are necessarily oversimplified and thus correspond to unphysical situations, calculations of the energy and angular momentum fluxes of a system have been done employing certain approximation procedures like expanding in powers of the gravitational coupling constant (due to the weakness of the gravitational interaction), or in powers of the ratio of the velocities of the masses of the system to the velocity of light, or in inverse powers of the distance from the system, or even combinations of these. (See Refs. 1–4.) Working at large distances and for small velocities, in what is essentially the Newtonian approximation (see below) although not always explicitly saying so, these authors have derived the usual quadrupole formula

$$\int \frac{dL_i}{dt}\,dt = \frac{2}{5}\,\epsilon^{ijk}\int [\ddot{d}_{mj}\dddot{d}_{mk}]\,dt, \tag{1}$$

where $\mathbf{L}$ denotes here the angular momentum of the system, d_{ij} its quadrupole moment: $d_{ij} = \int T_{00}\bar{x}_i\bar{x}_j\, d\bar{x}$, and the integral is taken over the sources. The time averaging is introduced in (1) in order to make the integrals that define $d\,\mathbf{L}/dt$ converge. (A similar formula for energy has also been derived.)

We shall show in this paper that one can derive the quadrupole formula for all solutions of Einstein's field equations in the Newtonian approximation, without the need for averaging (cf. also Ref. 5). Our formalism will also allow us to characterize those metrics (no field equations assumed) for which the angular momentum flux is not divergent and well defined.

As a measure of energy and angular momentum density we use the Schutz and Sorkin[6] gravitational Noether operator (a pseudotensorial operator on vector fields that reduces to the familiar pseudotensors for particular choices of the fields), which is defined, for any vector field ξ^{μ}, by

$$8\pi t^{\mu}{}_{\nu}\xi^{\nu} := -(-g)^{1/2}G^{\mu}{}_{\nu}\xi^{\nu} + \tfrac{1}{2}\partial_{\alpha}[h^{\mu\alpha\nu\beta}{}_{,\beta}\xi_{\nu}(-g)^{-1/2}], \tag{2}$$

where

$$h^{\mu\alpha\nu\beta} = (-g)(g^{\mu\nu}g^{\alpha\beta} - g^{\alpha\nu}g^{\mu\beta}) \tag{3}$$

and G is the Einstein tensor. This operator has the advantages that it does not depend on second derivatives of the metric, it contains the Einstein[7] and the Landau–Lifshitz[1] complexes as particular cases, and when the field and matter equations are satisfied, the Schutz and Sorkin ξ-momentum, defined by

$$P[\xi,H] := \int_H (I^{\mu}{}_{\nu}\xi^{\nu} + t^{\mu}{}_{\nu}\xi^{\nu})\,d\sigma_{\mu} \tag{4a}$$

is conserved, and may be written as

$$P[\xi,H] = (16\pi)^{-1}\int_H \partial_{\alpha}(h^{\mu\alpha\nu\beta}{}_{,\beta}\xi_{\nu}(-g)^{1/2})\,d\sigma_{\mu} = (16\pi)^{-1}\oint_{\partial H} h^{\mu\alpha\nu\beta}{}_{,\beta}\xi_{\nu}(-g)^{-1/2}\,d\Sigma_{\mu\alpha}. \tag{4b}$$

Here, $I^{\mu}{}_{\nu}$ is the Noether operator for matter (a covariant generalization of the so-called canonical stress-energy tensor, see Schutz and Sorkin[6]), H is a spacelike or null hypersurface with boundary ∂H, and $d\sigma_{\mu}$ and $d\Sigma_{\mu\alpha}$ are the coordinate volume and surface elements. The quantity above is equivalent to other definitions when evaluated on solutions of Einstein's equations (Nahmad-Achar[8]). This, together with the fact that (4a) is a generalization of a quantity that would be the usual four-momentum or angular momentum for a classical field theory (according, respectively, to whether the vector field is chosen so as to be asymptotically equal to a timelike or a rotational Killing field in flat space), entitles us to call it by the same name. Furthermore, it has the added virtue of being rather insensitive to the asymptotic behavior of the metric, which allows one to develop very strong variational principles for solutions of Einstein's equations in a vacuum or in the presence of matter fields (cf. Nahmad-Achar and Schutz[9] and Nahmad-Achar[10] for details).

The reader should note that Eq. (4a) does not define the full four-momentum or the full angular momentum, but only one component of them. Different components may be

[a] Present address.

 0022-2488/89/051009-04$02.50

chosen by an appropriate choice of the vector field. Note also that this definition is an integral over a hypersurface, and not the (perhaps more usual) tensorial map of a Poincaré subgroup of the BMS group. There is no relation between these two definitions unless Einstein's field equations are assumed to hold. This is important because it allows us to evaluate the angular momentum (or the four-momentum) in metrics that are not necessarily solutions of Einstein's equations, a requirement for dealing with the variational principles alluded to above. It also means that we can define and evaluate quantities like "angular momentum flux" in a generic spacetime (Nahmad-Achar[10]).

II. ANGULAR MOMENTUM FLUX

Let us assume that the description of our physical system under consideration (matter + gravitational fields) is given completely by a set of field variables $Q_A(x)$, where x stands for the local coordinates, and that the field equations can be derived from the principle of stationary action

$$\delta S = 0, \tag{5a}$$

where

$$S = \int_{\Xi} L\,[x,Q(x)]\,d^4x, \tag{5b}$$

and $L[x,Q(x)]$ is the Lagrangian density, which is a functional of the field variables Q and their derivatives.

Let H be a hypersurface of constant time t (or constant retarded time u if it is null). Let H' be another hypersurface of constant time $t + \Delta t$. (Again, and in what follows, replace t by u if null surfaces are considered.) Let C be the "cylinder" at (spatial or null) infinity joining H and H'; note that C is of the form $C = \partial H \times \Delta t$. Call Ξ the region enclosed by H, H', and C. Under a pure dragging by a vector field $\boldsymbol{\xi}$, i.e., $\delta x^\alpha = \xi^\alpha$, we have

$$\delta_\xi S = \int_{\Xi} \frac{\delta L}{\delta Q}\,\delta_\xi Q\, d^4x + \oint_{\partial\Xi} \Im^{\mu}{}_{\nu}\cdot\xi^{\nu}\, d\sigma^{\mu}, \tag{6}$$

where $\Im^{\mu}{}_{\nu}\cdot\xi^{\nu} = I^{\mu}{}_{\nu}\cdot\xi^{\nu} + t^{\mu}{}_{\nu}\cdot\xi^{\nu}$ (cf. Schutz and Sorkin[6] and Nahmad-Achar[8]), so that if the action is invariant under the vector field $\boldsymbol{\xi}$, and the equations for the variational principle are satisfied, we obtain

$$\int_{H'} \Im^{\mu}{}_{\nu}\cdot\xi^{\nu}\, d\sigma_\mu - \int_{H} \Im^{\mu}{}_{\nu}\cdot\xi^{\nu}\, d\sigma_\mu = -\int_{C} \Im^{\mu}{}_{\nu}\cdot\xi^{\nu}\, d\sigma_\mu. \tag{7}$$

Using this expression we can write

$$\begin{aligned} \frac{\partial}{\partial t} P[\xi,H] &= \lim_{\Delta t \to 0} \frac{1}{\Delta t}\{P[\xi(t+\Delta t),H'] - P[\xi(t),H]\} \\ &= -\lim_{\Delta t\to 0}\frac{1}{\Delta t}\int_C \Im^{\mu}{}_{\nu}\cdot\xi^{\nu}\, d\sigma_\mu, \\ &= -2\oint_{\partial H} \Im^{\mu}{}_{\nu}\cdot\xi^{\nu} n_i \delta^i_\mu\, d^2S, \end{aligned} \tag{8}$$

where the factor 2 comes from the two boundaries of C, and $\mathbf{n}$ is the unit radial vector field. Equation (8) makes sense regardless of whether or not the field equations are satisfied, and so we may take it, in general, as a measure of the angular momentum flux. Since we are interested in angular momentum (the case for energy is totally analogous, and the calculations even simpler), we set $\xi^{\mu} = x\delta^{\mu}_2 - y\delta^{\mu}_1$ asymptotically, in an asymptotically Lorentzian coordinate system $[t,x,y,z]$ (by following the formalism introduced in Nahmad-Achar and Schutz,[11] one may choose to work with the Noether operator in any coordinate system, obtaining exactly the same result). Furthermore, we write the metric as

$$g^{\mu\nu} = \eta^{\mu\nu} - h^{\mu\nu}, \quad g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu} + h_{\mu\lambda}h^{\lambda}{}_{\nu} + O(1/r^3) \tag{9}$$

with, as usual, $h_{\mu\nu} = O(1/r)$. Then, after some algebra we have

$$\begin{aligned} \Im^{\mu}{}_{\nu}\cdot\xi^{\nu} n_i\delta^i_\mu = [(-g)^{1/2}/16\pi]\, n_i\epsilon^{3jk}\,[& -x_j h_{k\nu,0}h^{0\nu}{}_{,i} - x_j h_{k\nu,0}h^{0\nu}{}_{,0} - x_j h^{k\nu}{}_{,\nu}h_i{}^{\beta}{}_{,\beta} + x_j h_{k\lambda}h^{\lambda 0}{}_{,0}h_i{}^0{}_{,0} + h_{ji,k} - x_j h_{k\nu,0}h_i{}^{\nu}{}_{,0} \\ & - x_j h_{k\lambda}h^{\lambda}{}_{\nu,0}h_i{}^{\nu}{}_{,0} - x_j h_{k\lambda,0}h^{\lambda}{}_{\nu}h_i{}^{\nu}{}_{,0} + h_i{}^j{}_{,0}h^{0k} + x_j h_{k\nu,0}h_i{}^{\nu}{}_{,0}h^{00} + \delta_{ij}h^{k\beta}{}_{,\beta} + x_j h_{ki,\alpha}h^{\alpha\beta}{}_{,\beta} \\ & - x_j h_{k\lambda}h_i{}^{\lambda}{}_{,0}h_\alpha{}^{\alpha}{}_{,0} + \delta_{ij}h^{\rho}{}_{\rho,k} - x_j h_i{}^k{}_{,0}h^{\rho}{}_{\rho,0} + x_j h_{k0,0}h^{\rho}{}_{\rho,i} + x_j h_{k0,0}h_i{}^0 h^{\rho}{}_{\rho,0} \\ & + \delta_{ij}h^{k0}h^{\rho}{}_{\rho,0} + x_j h_{ki,0}h^{00}h^{\rho}{}_{\rho,0} - x_j h_{ki,0}h^{\sigma\rho}h_{\sigma\rho,0}] + O(1/r^3), \end{aligned} \tag{10}$$

where the index 0 refers to the coordinate t, and latin indices run in [1,3].

The angular momentum flux is calculated by substituting this last expression into Eq. (8) and performing the integral. However, some terms in Eq. (10) are $O(1/r)$ and will in principle diverge when integrated over ∂H and taken the limit to infinity. These are

$$\begin{aligned} [\text{Eq. } (10)]_{1/r} = (1/16\pi)\, n_i\epsilon^{3jk}\, [& x^j h_{ki,0}h_{aa,0} - x_j h_{k0,0}h_{i0,0} \\ & - x_j h_{k\nu,0}h^{i\nu}{}_{,0} + \delta_{ij}h^{k0}{}_{,0}]. \end{aligned} \tag{11}$$

In order to have a well-defined flux we would have to restrict ourselves to the class of metrics for which the rhs of Eq. (11) adds up to something which is $O(1/r^2)$ or vanishes under angular integration. If we choose our origin of coordinates to be at the center-of-mass of the source, as is often done in linearized theory, the linear momentum of the source vanishes and one has $h^{0k} = O(1/r^2)$, so that (11) reduces to

$$\begin{aligned} [\text{Eq. } (10)]_{1/r} = (1/16\pi)\, n_i\epsilon^{3jk}x_j\,[h^{ki}{}_{,0}h_{aa,0} - h_{ka,0}h^{ai}{}_{,0}] \\ [h^{0k} = O(1/r^2)]. \end{aligned} \tag{12}$$

III. NEWTONIAN APPROXIMATION

When dealing with a system whose gravitational field is weak, so that terms nonlinear in $h^{\mu\nu}$ and its derivatives can be neglected, and such that its characteristic velocity is small, $v \ll c$, we can relate the metric coefficients to the source in a simple way:

$$\bar{h}_{ij} = \frac{1}{r}\,(\text{quadrupole moment})^{\cdot\cdot} + \frac{1}{r}\,O\left(\frac{v}{c}\right) + O\left(\frac{1}{r^2}\right). \tag{13}$$

One usually neglects the second term on the rhs under the assumption $v \ll c$; but strictly speaking, one can only do so when evaluating the coefficients at a fixed value of r. At fixed velocity, however small, and in the limit as $r \to \infty$ (which is the case we are interested in here), this term, while formally small compared to the first one, will cause the angular momentum integral to diverge.

In the so-called *Newtonian limit*, as treated by Futamase and Schutz[4], the velocity v is assumed to be of the order of a small parameter ϵ that labels a sequence of relativistic solutions that approaches a Newtonian limit. The characteristic time scales of the system go as ϵ^{-1} since $v \sim \epsilon$, so that events at similar stages of evolution in the different space-times of the sequence are related by a map $(x^i,t) \to (x^i, \tau = \epsilon t)$. Since gravitational radiation emitted by the system will have a characteristic period given by a dynamical time interval $\Delta\tau$, and this scales as $1/\epsilon$ in t, any point at fixed x^i will find itself within one wavelength from the source for sufficiently small ϵ. Therefore, Futamase and Schutz define another coordinate $\eta^i = \epsilon x^i$ and study the radiation at fixed (η^i, τ), which essentially means at a fixed number of wavelengths away from the relativistic source. In this formalism, $v \sim \epsilon$ and $r \sim 1/\epsilon$, so that the second term in Eq. (13) is $O(\epsilon^2)$ just as the third one, while the first is $O(\epsilon)$. The second term (which can be seen to consist of a combination of the third time derivative of the mass octupole and the second time derivative of the current quadrupole of the source) can thus be neglected and the integral of Eq. (12) becomes

$$\oint [\text{Eq. (12)}]\,d^2S = -\frac{1}{4}\, r^3 \epsilon^{3ik} h_{ka,0} h_{ia,0} = 0. \tag{14}$$

That is, the $O(1/r)$ terms in Eq. (10) do not contribute to the flux integral in this approximation. This result is important since Futamase and Schutz have shown that the Newtonian-limiting sequence described above is an asymptotic approximation to a well-defined sequence of solutions of Einstein's field equations, and that both the near- and far-zone quadrupole formulae for the outgoing energy flux are asymptotic approximations to general relativity.

The usual argument found in the literature (Landau and Lifshitz,[1] Morgan and Peres,[2] Peters,[3] and Lightman *et al.*[12]) is that, by imposing an outgoing-radiation boundary condition, one gets

$$\frac{\partial}{\partial r}\bar{h}_{\mu\nu} + \frac{1}{c}\frac{\partial}{\partial t}\bar{h}_{\mu\nu} = 0, \tag{15}$$

so that

$$\bar{h}_{\mu\nu,i} = -\,n_i \bar{h}_{\mu\nu,0}, \tag{16}$$

and this allows us to convert all spatial derivatives into time derivatives multiplied by the appropriate direction vectors. Performing a time average, all the terms that are pure time derivatives are transformed into secular changes, and these are neglected under the assumption of periodic motion. This procedure gets rid of the leading-order terms that would otherwise lead to divergences. Equation (16), however, assumes that the $O(v/c)$ term remains negligible in the limit $r \to \infty$, so that the argument holds, again, only in the Newtonian approximation.

We shall now evaluate the integral (8) for the flux, using the expression given in (10) for the integrand, and relating the metric coefficients to the sources (in the limit of weak fields and small velocities). Using the zero linear momentum frame (ZLMF) mentioned above, $h_{0k} = O(1/r^2)$ and we are left with

$$\begin{aligned}
&\mathfrak{T}^{\mu}{}_{\nu}\cdot\xi^{\nu} n_i \delta^i_\mu \\
&\quad = (1/16\pi) n_i \epsilon^{3jk} \big[h_{ij,k} + x_j h_{ks} h_{sn,0} h^{ni}{}_{,0} \\
&\qquad + x_j h_{ks,0} h_{sn} h^{ni}{}_{,0} + x_j h^{00} h_{kn,0} h^{ni}{}_{,0} + \delta_{ij} h^{k\beta}{}_{,\beta} \\
&\qquad + \delta_{ij} h^{\rho}{}_{\rho,k} - x_j h_{ks} h^{si}{}_{,0} h^{aa}{}_{,0} + x_j h^{00} h_{ki,0} h^{\rho}{}_{\rho,0} \\
&\qquad - x_j h_{ki,0} h^{\sigma\rho} h_{\sigma\rho,0} + x_j h_{ki,0} h_{aa,0} - x_j h_{ka,0} h^{ai}{}_{,0} \big] \\
&\qquad + O(1/r^3).
\end{aligned} \tag{17}$$

Furthermore, using Eq. (16) for the last two terms in (17), we have finally

angular momentum flux

$$\begin{aligned}
&= -\frac{2}{16\pi}\epsilon^{3jk} \oint_{\partial H} \{ n_i [h_{ij,k} + \delta_{ij} h^{k\beta}{}_{,\beta} + \delta_{ij} h^{\rho}{}_{\rho,k}] \\
&\quad + x_j [-\bar{h}_{ki,i}\bar{h}_{00,0} - \tfrac{1}{2}\bar{h}^{\rho}{}_{\rho,k}\bar{h}_{00,0} + \bar{h}_{ai,i}\bar{h}_{ka,0} \\
&\quad + \tfrac{1}{2}\bar{h}^{\rho}{}_{\rho,a}\bar{h}_{ka,0}] \} d^2S.
\end{aligned} \tag{18}$$

Carrying out an expansion of the components $T^{\mu\nu}$ we obtain

$$\begin{aligned}
\bar{h}_{ij} &= -(2/r)\ddot{d}_{ij} + O(1/r^2), \\
\bar{h}_{00} &= -\frac{4M}{r} - \frac{2x^i x^j}{r^4}(3\dot{d}_{ij} + \eta_{ij}\dot{d}^k{}_k) \\
&\qquad - \frac{2x^i x^j}{r^3}\ddot{d}_{ij} + O\left(\frac{1}{r^3}\right),
\end{aligned} \tag{19}$$

which we use to evaluate each term in (18):

(a) first term vanishes;

(b + c) second and third terms vanish together;

(d) $\displaystyle \frac{1}{16\pi}\epsilon^{3jk}\oint [-x_j \bar{h}_{ki,i}\bar{h}_{00,0}]\,d^2S = \frac{2}{15}\epsilon^{3jk}\ddot{d}_{ks}\dddot{d}_{sj};$

(e) $\displaystyle \frac{1}{16\pi}\epsilon^{3jk}\oint \left[-\frac{1}{2} x_j \bar{h}^{\rho}{}_{\rho,k}\bar{h}_{00,0} \right] d^2S$

$\displaystyle = -\frac{2}{15}\epsilon^{3jk}\ddot{d}_{ks}\dddot{d}_{sj};$

(f) $\displaystyle \frac{1}{16\pi}\epsilon^{3jk}\oint x_j \bar{h}_{ai,i}\bar{h}_{ka,0}\, d^2S = \epsilon^{3jk}\ddot{d}_{ks}\dddot{d}_{sj};$

(g) $\displaystyle \frac{1}{16\pi}\epsilon^{3jk}\oint \frac{1}{2} x_j \bar{h}^{\rho}{}_{\rho,a}\bar{h}_{ka,0}\, d^2S$

$\displaystyle = \epsilon^{3jk}\left[-\ddot{d}_{ks}\dddot{d}_{sj} + \frac{3}{5}\ddot{d}_{ks}\dddot{d}_{sj} \right].$

Adding up all the terms, the angular momentum flux in the Newtonian approximation, evaluated in the ZLMF, is

$$\frac{\partial}{\partial t} P[\xi, H] = -\frac{2}{5}\epsilon^{3jk}\ddot{d}_{ks}\dddot{d}_{sj} \tag{20}$$

as expected.

IV. DEPENDENCE ON VECTOR FIELDS

We have noted that the integral defining the angular momentum is in general an $O(r)$ quantity; therefore, changes to order $1/r$ in the vector field may produce finite changes in the angular momentum itself (this seems to be the way in which supertranslations are reflected in this formalism). Thus, considering two null hypersurfaces $H(u)$ and $H(u+\Delta u)$, and changing the vector field ξ to order $1/r$ on $H(u+\Delta u)$ without altering it on $H(u)$, we can arbitrarily change the value of the angular momentum measured at a later retarded time while maintaining that at the present retarded time unaltered. This must of course be reflected in the measurement of the flux, and we shall here calculate this change.

The leading terms, and the only ones that would contribute to a change in the flux, are those given in Eq. (11):

$$[\Im^{\mu}{}_{\nu}\cdot\xi^{\nu}n_i\delta^i_{\mu}]_{1/r} = (1/16\pi)n_i\epsilon^{3jk}[x_j h_{ki,0}h_{aa,0} - x_j h_{k0,0}h_{i0,0} - x_j h_{k\nu,0}h^{i\nu}{}_{,0} + \delta_{ij}h^{k0}{}_{,0}]. \tag{21}$$

Suppose now that we change ξ to first order in $1/r$, i.e., consider

$$\xi^{\mu}\to\xi^{\mu}+\zeta^{\mu},\quad \zeta^{\mu}=\frac{\partial}{\partial\phi}\left[O\left(\frac{1}{r}\right)\right]. \tag{22}$$

The lhs of (21) will pick up new $O(1/r^2)$ terms under this transformation, which will contribute to the flux when integrated. Denoting these by Z, they can be shown to be

$$Z=(1/16\pi)[\zeta_{i,0}n_i h_{aa,0}-\zeta_{k,0}n_i h_{ik,0}], \tag{23}$$

whose surface integral need not vanish. When Einstein's equations hold, we can use them to relate the metric coefficients to the source. Then,

$$\text{change in flux} = \oint Z\, d^2S = -\frac{1}{2}r\dddot{d}_{kk}n_i\zeta_{i,0}+\frac{1}{8\pi}r\dddot{d}_{ik}\oint n_i\zeta_{k,0}\,d^2\Omega, \tag{24}$$

where $d^2\Omega=\sin\theta\, d\theta\, d\phi$.

Note that the change in angular momentum flux depends only on the *time* derivative of the *spatial* components of ζ_{ν}. In other words, if the spatial components of ζ_{ν} are constant in time, there will be no change in the flux. This is not surprising, for it is precisely those transformations with $\zeta_{k,0}\neq 0$ that take us away from a ZLMF, by leaving our origin of coordinates not at rest with respect to the center of mass of the source. What is interesting, is that even for these transformations with $\zeta_{k,0}\neq 0$ there will be no change in the flux when we work in the Newtonian limit, since they introduce an extra factor of v thereby increasing the order in $1/r$ (cf. see Sec. III). Thus, by relating the radiation terms to the sources we are fixing the null cones, in such a way that the angular momentum flux is free of the supertranslation ambiguity in the Newtonian approximation.

V. COMMENTS

We have shown that the Noether operator formalism reproduces the well-known results found in the literature for the angular momentum flux carried away by gravitational radiation, in linearized Einstein's theory. Furthermore, it allows us to calculate the flux for a wide class of metrics, for which it is shown to be well defined. We have calculated the dependence of the flux on the vector field, and shown that it is free of the supertranslation ambiguity in the Newtonian limit.

ACKNOWLEDGMENT

It is a pleasure to thank B. F. Schutz for useful comments.

[1]L. D. Landau and E. M. Lifshitz, *The Classical Theory of Fields* (Addison-Wesley, Reading, MA, 1951).

[2]T. A. Morgan and A. Peres, Phys. Rev. **131**, B494 (1963).

[3]P. C. Peters, Phys. Rev. **136**, B1225 (1964).

[4]T. Futamase and B. F. Schutz, Phys. Rev. D **32**, 2557 (1985).

[5]J. Winicour, Institute Report MPA-253, Max Planck Institut für Astrophysik, Garching bei München, 1986.

[6]B. F. Schutz and R. Sorkin, Ann. Phys. **107**, 1 (1977).

[7]A. Einstein, Ann. Phys. (Leipzig) **49**, 769 (1916).

[8]E. Nahmad-Achar, Ph.D. dissertation, Cambridge University, Cambridge, 1986.

[9]E. Nahmad-Achar and B. F. Schutz, Class. Quantum Gravit. **4**, 929 (1987).

[10]E. Nahmad-Achar, Class. Quantum Gravit. **4**, 943 (1987).

[11]E. Nahmad-Achar and B. F. Schutz, Gen. Relativ. Gravit. **19**, 655 (1987).

[12]A. P. Lightman, W. H. Press, R. H. Price, and S. A. Teukoplsky, *Problem Book in Relativity and Gravitation* (Princeton U. P., Princeton, NJ, 1979).

Painlevé property, auto-Bäcklund transformation, Lax pairs, and reduction to the standard form for the Korteweg–De Vries equation with nonuniformities

T. Brugarino
Dipartimento di Matematica ed Applicazioni, Università di Palermo, 90128 Palermo, Italy

(Received 26 February 1988; accepted for publication 11 January 1989)

It is demonstrated that the KdV equation with nonuniformities, $u_t + a(t)u + (b(x,t)u)_x + c(t)uu_x + d(t)u_{xxx} + e(x,t) = 0$, has the Painlevé property if the compatibility condition among the coefficients of it holds: $b_t + (a - Lc)b + bb_x + db_{xxx} = 2ah + hL(d/c^2) + (dh/dt) + ce + x[2a^2 + aL(d^3/c^4) + (da/dt) + L(d/c)L(d/c^2) + (d/dt)L(d/c)]$, where $L = (d/dt)\lg$ and $h(t)$ is an arbitrary function of t. The auto-Bäcklund transformation and Lax pairs for this equation are obtained by truncating the Laurent expansion. Furthermore, assuming the compatibility condition, then the KdV equation with nonuniformities is transformable, via suitable variable transformations, to the standard KdV.

I. INTRODUCTION

In the last decade there has been intense activity regarding the study of the complete integrability of nonlinear partial differential equations (PDE) and, moreover, their study has revealed unexpected connections between geometry, analysis, and physics. But the main domain of activity was generally restricted to equations without explicit (x,t) dependency.

Recently, it has been observed that, when all the ordinary differential equations obtained by exact similarity transformed from a given PDE have the Painlevé property (PP), then, perhaps after a change of variables, PDE is completely integrable.[1] The complete integrability is also defined in terms of the existence of the inverse scattering transform (IST) or the Bäcklund transformation (BT). The existence of an IST solution is assured by that of the Lax pairs.[2]

More recently, Weiss *et al.*[3] have proposed a generalized PP which is directly applicable to PDE. In the latter approach, a PDE has the PP if its solution can be expressed as a single-valued expansion about noncharacteristic singular manifold $f(x,t) = 0$. This approach has been applied to various field equations.[4–6]

II. PAINLEVÉ PROPERTY FOR THE KDV EQUATION WITH NONUNIFORMITIES

Here, we report the results of PP analysis of a KdV equation with nonuniformity terms:

$$u_t + a(t)u + (b(t,x)u)_x + c(t)uu_x + d(t)u_{xxx} + e(x,t) = 0. \tag{1}$$

Equation (1), for particular values of the coefficients, appears in many physical systems.[7–10]

We have found that Eq. (1) admits the PP if the following compatibility condition for the coefficients of the equation is satisfied:

$$\begin{aligned} b_t &+ (a - Lc)b + bb_x + db_{xxx} \\ &= 2ah + hL\frac{d}{c^2} + \frac{dh}{dt} + ce \\ &\quad + x\left(2a^2 + aL\frac{d^3}{c^4} + \frac{da}{dt}\right. \\ &\quad \left. + L\frac{d}{c}L\frac{d}{c^2} + \frac{d}{dt}L\frac{d}{c}\right), \end{aligned} \tag{2}$$

where L is the logarithmic differentiation operator, i.e., $L = (d/dt)\lg$, and $h(t)$ is an arbitrary and sufficiently smooth function of t.

Condition (2) is found using a property of Lax pairs obtained from the PP. The possible ABT is also developed, when the equation (1) is integrable.

We are now looking for a solution of (1) in the Laurent series expansion

$$u(x,t) = f^{\alpha}(x,t) \sum_{0}^{\infty}{}_j\, u_j(x,t) f^j(x,t), \tag{3}$$

where $u_j(x,t)$ are analytic functions in a neighborhood of the singular manifold $f(x,t) = 0$ and α is an integer to be determined.

Inserting the ansatz

$$u(x,t) \cong f^{\alpha} u_0$$

in (1) and comparing the exponents, we find that the leading-order analysis gives the value $\alpha = -2$. Inserting the ansatz (3) together with $\alpha = -2$ in (1), then we find the recursion relations for $u_j(x,t)$,

$$\begin{aligned} &u_{j-3,t} + (j-4)u_{j-2}f_t + (a + b_x)u_{j-3} + b\{u_{j-3,x} + (j-4)u_{j-2}f_x\} \\ &+ c\left\{\sum_{0}^{j}{}_m u_{j-m}[u_{m-1,x} + (m-2)u_m f_x]\right\} \\ &+ d\{u_{j-3,xxx} + 3(j-4)u_{j-2,xx}f_x + 3(j-3)(j-4)u_{j-1,x}f_x^2 \\ &+ 3(j-4)u_{j-2,x}f_{xx} + (j-2)(j-3)(j-4)u_j f_x^3 \\ &+ 3(j-3)(j-4)u_{j-1}f_x f_{xx} + (j-4)u_{j-2}f_{xxx}\} + e\delta_j^5 = 0. \end{aligned} \tag{4}$$

 0022-2488/89/051013-03$02.50

In collecting terms involving u_j, it is found that

$$df_x{}^3(j+1)(j-4)(j-6)u_j = F(x,t,u_{j-1},...,u_0,f_t,f_x,...) \quad \text{for } j=0,1,2,\ldots . \tag{5}$$

Equation (5) determines the coefficients u_j of the series expansion (3), provided that $j\neq -1, 4, 6$. These values of j are called the "resonance" of the recursion relations and allow the introduction of arbitrary functions u_4 and u_6. For $j=-1$, the series (3) is nondefined and therefore the resonance at $j=-1$ corresponds to "arbitrary" function f defining the singular manifold.

For KdV with nonuniformities (1), we find from (5),

$$j=0, \quad u_0 = -12(d/c)f_x^2, \tag{6}$$

$$j=1, \quad u_1 = 12(d/c)f_{xx}, \tag{7}$$

$$j=2, \quad f_x f_t + bf_x^2 + cu_2 f_x^2 + 4df_x f_{xxx} - 3df_{xx}^2 = 0, \tag{8}$$

$$j=3, \quad W \equiv L(d/c)f_x - cu_3 f_x^2 + df_{xxxx} + cu_2 f_{xx} + af_x + bf_{xx} + f_{xt} = 0, \tag{9}$$

and

$$j=4, \quad W_x = 0. \tag{10}$$

By (9) the compatibility condition (10) is satisfied identically. The compatibility condition at $j=6$ involves extensive calculations. (All the straightforward but long and tedious computations are performed in a P.C. by using the muMATH package.)

If we specialize $u_4 = u_6 = 0$ and, furthermore, $u_3 = 0$, then it is easily demonstrated that

$$u_j = 0 \quad \text{for } j \geq 3 \tag{11}$$

if

$$u_{2,t} + au_2 + (bu_2)_x + cu_2 u_{2,x} + du_{2,xxx} + e = 0. \tag{12}$$

From Eq. (3) and Eqs. (6)–(12) we get

$$u_0 = -12(d/c)f_x^2, \tag{13}$$

$$u_1 = 12(d/c)f_{xx}, \tag{14}$$

$$f_x f_t + bf_x{}^2 + cu_2 f_x^2 + 4df_x f_{xxx} - 3df_{xx}^2 = 0, \tag{15}$$

$$L(d/c)f_x + df_{xxxx} + cu_2 f_{xx} + f_{xt} + af_x + bf_{xx} = 0, \tag{16}$$

$$u_{2,t} + au_2 + (bu_2)_x + cu_2 u_{2,x} + du_{2,xxxx} + e = 0, \tag{17}$$

$$u = 12(d/c)(L_x f)_x + u_2, \tag{18}$$

and $u_j = 0$ for $j \geq 3$.

In transformation (18) u and u_2 satisfy the same equation (1) if Eqs. (15) and (16) are consistent. In this case (18) defines an ABT of the KdV equation with nonuniformities.

To prove this, we obtain

$$f_x = V^2, \tag{19}$$

and find by straightforward calculation that

$$-V_t = \tfrac{1}{2}b_x V + bV_x + cu_2 V_x + \tfrac{1}{2}cu_{2,x}V + 4dV_{xxx}, \tag{20}$$

$$-V_t = \tfrac{1}{2}[a + L(d/c)]V + 3dV_x V_{xx}/V + dV_{xxx} + cu_2 V_x + bV_x. \tag{21}$$

Eliminating V_t, we find

$$\tfrac{1}{2}[a - b_x + L(d/c)]V - \tfrac{1}{2}cu_{2,x}V + 3dV_x V_{xx}/V - 3dV_{xxx} = 0. \tag{22}$$

This equation can be written as

$$\tfrac{1}{2}[a - b_x + L(d/c)] - \tfrac{1}{2}cu_{2,x} - 3d(V_{xx}/V)_x = 0. \tag{23}$$

Integrating Eq. (23) with respect to x, we have

$$\tfrac{1}{2}[(a + L(d/c))x - b + h(t)] - \tfrac{1}{2}cu_2 - 3dV_{xx}/V = q(t), \tag{24}$$

where $h(t)$ and $q(t)$ are two arbitrary functions of the variable t.

Equation (24) can be rewritten as

$$s(t)\{\tfrac{1}{2}[(a + L(d/c))x - b + h(t)]V - \tfrac{1}{2}cu_2 V - 3dV_{xx}\} = \lambda V, \tag{25}$$

with $\lambda = q(t)s(t)$ constant.

We thus have found a candidate for a Lax formulation of

$$u_{2,t} + au_2 + (bu_2)_x + cu_2 u_{2,x} + du_{2,xxx} + e = 0$$

by defining the operators

$$\Lambda(u_2,t,x) = s(t)\{\tfrac{1}{2}[(a + L(d/c))x - b + h] - \tfrac{1}{2}cu_2 - 3dD^2\}, \tag{26}$$

$$B = \tfrac{1}{2}b_x + bD + cu_2 D + \tfrac{1}{2}cu_{2,x} + 4dD^3. \tag{27}$$

Equations (25) and (20) then read

$$\Lambda V = \lambda V, \tag{28}$$

$$V_t = -BV. \tag{29}$$

The eigenvalue problem (28), and (29) is compatible with the time evolution of eigenfunction V given by (29), if we have the operator identity

$$\Lambda_t = [\Lambda, B]. \tag{30}$$

From (28) and (29), then, (30) holds if

$$\frac{d}{dt}(cs) = cs\left(2a + L\frac{d}{c}\right), \tag{31}$$

$$\frac{d}{dt}(ds) = 2ds\left(a + L\frac{d}{c}\right), \tag{32}$$

$$\frac{d}{dt}\left\{s\left[\left(a + L\frac{d}{c}\right)x - b + h\right]\right\} + bs\left(a + L\frac{d}{c} - b_x\right) - sdb_{xxx} + sce = 0. \tag{33}$$

The equations (31) and (32) are simultaneously fulfilled by

$$s(t) = \frac{d(t)}{c^2(t)}\exp\left(2\int a(t)dt\right). \tag{34}$$

Substituting for $s(t)$ from Eq. (34) into Eq. (33), we obtain the constraint (2).

III. REDUCTION TO THE KDV STANDARD, EXAMPLES, AND FINAL REMARKS

The transformation

$$u(x,t) = \frac{1}{c(t)}(v(x,t) - b(x,t))$$

reduces Eq. (1), by using the condition (2), to

$$v_t + (a - Lc)v + vv_x + dv_{xxx} - x\Big(2a^2 + aL\frac{d^3}{c^4} + \frac{da}{dt} + L\frac{d}{c}L\frac{d}{c^2} + \frac{d}{dt}L\frac{d}{c}\Big) - 2ah - hL\frac{d}{c^2} - \frac{dh}{dt} = 0. \qquad (35)$$

Equation (35) can be transformed under

$$\begin{cases} v = w(\xi,t) + g(t) \\ \xi = x - \int g(t)dt \end{cases},$$

where

$$\frac{dg}{dt} + (a - Lc)g - \Big\{2a^2 - aL\frac{d^3}{c^4} + L\frac{d}{c}L\frac{d}{c^2} + \frac{d}{dt}L\frac{d}{c}\Big\}\int g\,dt - 2ah - hL\frac{d}{c^2} - \frac{dh}{dt} = 0,$$

to yield the following equation,

$$w_t + (a - Lc)w + ww_\xi + dw_{\xi\xi\xi} - \xi\Big(2a^2 + aL\frac{d^3}{c^4} + \frac{da}{dt} + L\frac{d}{c}L\frac{d}{c^2} + \frac{d}{dt}L\frac{d}{c}\Big) = 0. \qquad (36)$$

In turn, Eq. (36), via the transformation

$$w = (a + L(d/c))\xi + \Theta(\xi,t),$$

becomes

$$\Theta_t + (2a + L(d/c^2))\Theta + \xi(a + L(d/c))\Theta_\xi + \Theta\Theta_\xi + d\Theta_{\xi\xi\xi} = 0. \qquad (37)$$

Now Eq. (37) satisfies the condition[12] that guarantees its own reduction to the Korteweg–deVries equation with constant coefficients.

Several equations of physical interest satisfy the compatibility condition (2). These equations are:

The equation of KdV type for nonuniform media with relaxation effects,[10]

$$u_t + \gamma u + [(c_0 + \gamma x)u]_x + 6uu_x + u_{xxx} = 0,$$

with $h = c_0$,

the cylindrical KdV equation,[7]

$$u_t + (t/2t)u + 6uu_x + u_{xxx} = 0, \quad \text{with } h = 0,$$

the equation with nonuniform terms,[11]

$$u_t + u + 6uu_x + u_{xxx} = \gamma^2 x/3, \quad \text{with } h = 0, \text{ and}$$

the KdV equation for unidirectional waves in a channel of gradually varying width $b(t)$ and depth $d(t)$,[9]

$$2u_t + \frac{3}{cd}uu_x + \frac{1}{3}\frac{d^2}{c^3}u_{xxx} + uL(bc) = 0,$$

where $c^2 = gd$, and if $bd^{9/2}$ is a constant, then it satisfies the compatibility condition (2) with $h = 0$.

Concluding this final section, we remark that all the cases considered in the literature[11–15] are covered by condition (2) and that the slightly different equation

$$u_t + a(x,t)u + b(x,t)u_x + c(x,t)uu_x + d(x,t)u_{xxx} + e(x,t) = 0$$

has the PP if it is of the form (1) with (2) satisfied. Assuming this, then Eq. (1) is essentially the KdV standard.

ACKNOWLEDGMENTS

This work was supported by the Italian M.P.I.

[1] M. J. Ablowitz, A. Ramani, and H. Segur, J. Math. Phys. **21**, 715 (1980).
[2] F. Calogero and A. Degasperis, *Spectral Transform and Solitons* (North-Holland, Amsterdam, 1982), Vol. I.
[3] J. Weiss, M. Tabor, and G. Carnevale, J. Math. Phys. **24**, 522 (1983).
[4] W. H. Steeb, M. Kloke, B. M. Spieker, and W. Oevel, J. Phys. A. **16**, L447 (1983).
[5] W. Oevel and W. H. Steeb, Phys. Lett. A **103**, 239 (1984).
[6] S. Puri, Phys. Lett. A **107**, 359 (1985).
[7] S. Maxon and J. Viecelli, Phys. Fluids **17**, 1614 (1974).
[8] A. Greco, C. R. Acad. Sci. Paris **305**, 151 (1987).
[9] J. W. Miles, J. Fluid Mech. **91**, 81 (1979).
[10] R. Hirota and J. Satsuma, J. Phys. Soc. Jpn. Lett. **41**, 2141 (1976).
[11] R. Hirota, J. Phys. Soc. Jpn. Lett. **46**, 1681 (1979).
[12] T. Brugarino and P. Pantano, Phys. Lett. A **80**, 223 (1980).
[13] N. Nirmala, M. J. Vedan, and B. V. Baby, J. Math. Phys. **27**, 2640 (1986).
[14] L. Hlavaty, J. Phys. Soc. Jpn. **55**, 1405 (1986).
[15] N. Joshi, Phys. Lett. A **125**, 456 (1987).

A note on the triple sum series for the $9j$ coefficient

K. Srinivasa Rao
Institute of Mathematical Science, Madras 600-113, India and Center for Particle Theory, University of Texas, Austin, Texas 78712

V. Rajeswari
Institute of Mathematical Science, Madras 600-113, India

(Received 13 April 1988; accepted for publication 23 November 1988)

It is shown that the triple sum series of Jucys and Bandzaitis [*Angular Momentum Theory in Quantum Physics* (Vilnius, Mokslas, 1977)] for the $9j$ coefficient can be identified with a formal triple hypergeometric series due to Lauricella–Saran–Srivastava [G. Lauricella, Rend. Circ. Mat. Palermo 7, 111 (1893); L. Saran, Ganita 5, 77 (1954); H. M. Srivastava, Ganita 5, 97 (1964)].

I. THEORY

The conventional single sum expansion for the $9j$ coefficient,[1] derivable from the fundamental theorem of recoupling theory, is given by

$$\begin{Bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{Bmatrix} = \sum_k (-)^{2k}(2k+1) \times \begin{Bmatrix} a & d & g \\ h & i & k \end{Bmatrix} \begin{Bmatrix} b & e & h \\ d & k & f \end{Bmatrix} \begin{Bmatrix} c & f & i \\ k & a & b \end{Bmatrix}, \tag{1}$$

where $a,b,...,i$ can take integral or half-integral values and the summation index k takes the values:

$$\max(|a-i|,|d-h|,|b-f|) \leqslant k \leqslant \min(a+i,d+h,b+f). \tag{2}$$

In (1), the coefficients on the rhs are the $6j$ coefficients that have been defined[2] as sets of generalized hypergeometric functions of unit argument, ${}_4F_3(1)$s. The simplest known algebraic form, after a change of notation, for the $9j$ coefficient, due to Jucys and Bandzaitis[3] is the triple sum series:

$$\begin{Bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{Bmatrix} = (-1)^{x5} \frac{(d\,a\,g)(b\,e\,h)(i\,g\,h)}{(d\,e\,f)(b\,a\,c)(i\,c\,f)}$$
$$\cdot \frac{\Gamma(1+x1,1+x2,1+x3,1+y1,1+y2,1+z1,1+z2,1+p1)}{\Gamma(1+x4,1+x5,1+y3,1+y4,1+y5,1+z3,1+z4,1+z5,1+p2,1+p3)}$$
$$\times \sum_{x,y,z} \frac{1}{x!y!z!} \cdot \frac{(1+x2,x)(1+x3,x)(-x4,x)(-x5,x)}{(-x1,x)} \cdot \frac{(1+y1,y)(1+y2,y)(-y4,y)(-y5,y)}{(1+y3,y)}$$
$$\cdot \frac{(1+z2,z)(-z3,z)(-z4,z)(-z5,z)}{(-z1,z)} \cdot \frac{1}{(-p1,y+z)(1+p2,x+y)(1+p3,x+z)}, \tag{3}$$

where

$$\begin{aligned} &0 \leqslant x \leqslant \min(-d+e+f, c+f-i) = XF, \\ &0 \leqslant y \leqslant \min(g-h+i, b+e-h) = YF, \\ &0 \leqslant z \leqslant \min(a-b+c, a+d-g) = ZF, \end{aligned} \tag{4}$$

$$\begin{aligned} &x1 = 2f, \quad x2 = d+e-f, \quad x3 = c+i-f, \\ &x4 = e+f-d, \quad x5 = c+f-i, \quad y1 = -b+e+h, \\ &y2 = g+h-i, \quad y3 = 2h+1, \quad y4 = b+e-h, \\ &y5 = g-h+i, \quad z1 = 2a, \quad z2 = -a+b+c, \\ &z3 = a+d+g+1, \quad z4 = a+d-g, \quad z5 = a-b+c, \\ &p1 = a+d-h+i, \quad p2 = -b+d-f+h, \\ &p3 = -a+b-f+i, \end{aligned} \tag{5}$$

and

$$(abc) = [(a-b+c)!(a+b-c)!\,(a+b+c+1)!/(-a+b+c)!]^{1/2}, \tag{6}$$

$$\Gamma(x,y,...) = \Gamma(x)\Gamma(y)\cdots, \tag{7}$$

and the symbol (λ,k) is defined as

$$\begin{aligned} (\lambda,k) &= \Gamma(\lambda+k)/\Gamma(\lambda) \\ &= \lambda(\lambda+1)(\lambda+2)\cdots(\lambda+k-1), \quad k \geqslant 0, \\ (\lambda,-k) &= (-1)^k/(1-\lambda,k), \quad k<0. \end{aligned} \tag{8}$$

To identify the triple sum series in (3) with a triple hypergeometric series, consider the product of the following ${}_4F_3(1)$'s:

$${}_4F_3\begin{bmatrix} 1+x2 & 1+x3 & -x4 & -x5; & 1 \\ -x1 & 1+p2 & 1+p3 & & \end{bmatrix}$$
$$\cdot {}_4F_3\begin{bmatrix} 1+y1 & 1+y2 & -y4 & -y5; & 1 \\ 1+y3 & -p1 & 1+p2 & & \end{bmatrix}$$
$$\cdot {}_4F_3\begin{bmatrix} 1+z2 & -z3 & -z4 & -z5; & 1 \\ -z1 & -p1 & 1+p3 & & \end{bmatrix}$$

 0022-2488/89/051016-02$02.50

$$= \sum_{x,y,z} \frac{1}{x!y!z!} \frac{(1+x2,x)(1+x3,x)(-x4,x)(-x5,x)}{(-x1,x)(1+p2,x)(1+p3,x)}$$

$$\cdot \frac{(1+y1,y)(1+y2,y)(-y4,y)(-y5,y)}{(1+y3,y)(-p1,y)(1+p2,y)}$$

$$\cdot \frac{(1+z2,z)(-z3,z)(-z4,z)(-z5,z)}{(-z1,z)(-p1,z)(1+p3,z)}. \quad (9)$$

In (8), replace the pairs of products

$$(1+p2,x)(1+p2,y) \text{ by } (1+p2,x+y),$$
$$(1+p3,x)(1+p3,z) \text{ by } (1+p3,x+z), \quad (10)$$
$$(-p1,y)(-p1,z) \text{ by } (-p1,y+z),$$

to make the identification with the triple sum series that occurs in (3) possible. The product of three ${}_4F_3(1)$'s given by (9) with the replacements given in (10) leads us now to the new function:

$$F^{(3)}\left[\begin{matrix}(0)::(0);(0);(0):1+x2,1+x3,-x4,-x5;1+y1,1+y2,-y4,-y5;1+z2,-z3,-z4,-z5;1,1,1\\(0)::1+p2;-p1;1+p3:-x1;1+y3;-z1\end{matrix}\right], \quad (11)$$

which is a particular case of the function defined in three variables by Srivastava[4] and an elegant unification[5] of the triple hypergeometric functions of Lauricella,[6] Saran,[7] and Srivastava.[8] It is to be noted that the *new* generalized hypergeometric function in three variables $\Phi^3(\alpha_{kl};\beta_i,\gamma_m;w_k)$ defined by Wu,[9] is the same as $F^{(3)}$ given in (11), which is a particular case of an extremely general hypergeometric series defined in three variables by Srivastava[4] as

$$F^{(3)}\left[\begin{matrix}(a)::(b); & (b'); & (b''):(c); & (c'); & (c''); & x,y,z\\(e)::(f); & (f'); & (f''):(g); & (g'); & (g'') & \end{matrix}\right]$$
$$= \sum_{m,n,p} \frac{((a),m+n+p)((b),m+n)((b'),n+p)((b''),p+m)}{((e),m+n+p)((f),m+n)((f'),n+p)((f''),p+m)} \cdot \frac{((c),m)((c'),n)((c''),p)x^m y^n z^p}{((g),m)((g'),n)((g''),p)m!n!p!}, \quad (12)$$

where (a) denotes a sequence of parameters and we have used the notation of Srivastava.[4]

II. CONCLUSIONS

Finally, the identification of the triple sum series of Jucys and Bandzaitis[3] for the $9j$ coefficient as a special case of the triple hypergeometric series of Lauricella–Saran–Srivastava supersedes the claim of Wu[9] to his discovery of a "new" generalized hypergeometric function in three variables. Furthermore, this identification has two immediate consequences. They are (i) the possibility of studying,[10] for the first time, the polynomial zeros of $9j$ coefficients—a feature not revealed by the conventional single sum series (1); and (ii) the numerical computation[11] of the $9j$ coefficient adapting the Horner scheme[12] for the triple sum series, which has been found to be faster by a factor of 2 to 4 over the conventional method.

These aspects are beyond the scope of this note and are dealt with in detail elsewhere.[10,11]

ACKNOWLEDGMENTS

One of us (KSR) wishes to thank Professor L. C. Biedenharn for correspondence regarding the Jucys–Bandzaitis triple sum series, Professor E. C. G. Sudarshan and the Center for Particle Theory of the University of Texas at Austin for support, and Professor Charles B. Chiu for several interesting discussions. The other author (VR) thanks the Council of Scientific and Industrial Research of the Government of India, for the award of a Senior Research Fellowship.

[1] L. C. Biedenharn and J. D. Louck, "Angular momentum in quantum physics," in *Encyclopedia of Mathematics and its Applications* (Addison–Wesley, New York, 1981), Vol. 8. The expression (3.326), on p. 130 of this reference, contains two misprints. They are the factor $(a+d+h+1-z)!$ occurs twice instead of once and the factor $(b+f-a-j+x+z)!$ should be $(b-f-a+j+x+z)!$

[2] K. Srinivasa Rao, T. S. Santhanam, and K. Venkatesh, J. Math. Phys. **16**, 1528 (1975); K. Srinivasa Rao and K. Venkatesh, *Fifth International Colloquium on Group Theoretical Methods in Physics* (Academic, New York, 1977); K. Srinivasa Rao and V. Rajeswari, Int. J. Theor. Phys. **24**, 983 (1985).

[3] A. P. Jucys and A. A. Bandzaitis, *Angular Momentum Theory in Quantum Physics* (Vilnius, Mokslas, 1977).

[4] H. M. Srivastava, Proc. Cambridge Philos. Soc. **63**, 425 (1967).

[5] H. Exton, *Multiple Hypergeometric Functions and Applications* (Wiley, New York, 1976).

[6] G. Lauricella, Rend. Circ. Mat. Palermo **7**, 111 (1893).

[7] L. Saran, Ganita **5**, 77 (1954).

[8] H. M. Srivastava, Ganita **5**, 97 (1964).

[9] A. C. T. Wu, J. Math. Phys. **14**, 1222 (1973); also, S. J. Alisaukas and A. P. Jucys, *ibid.* **12**, 594 (1971).

[10] K. Srinivasa Rao and V. Rajeswari, J. Phys. A: Math. Gen. **21**, 4255 (1988).

[11] K. Srinivasa Rao, V. Rajeswari, and C. B. Chiu, to be published in Comput. Phys. Commun.

[12] J. A. N. Lee, *Numerical Analysis for Computers* (Reinhold, New York, 1966).

Hamilton–Jacobi formalism for geodesics and geodesic deviations

Stanisław L. Bażański
Institute of Theoretical Physics, University of Warsaw, ul. Hoża 69, 00-681 Warsaw, Poland

(Received 7 September 1988; accepted for publication 14 December 1988)

A formalism of integrating the equations of geodesics and of geodesic deviation is examined based upon the Hamilton–Jacobi equation for geodesics. The latter equation has been extended to the case of geodesic deviation and theorems analogous to Jacobi's theorem on the complete integral has been proved. As a result, a straightforward algorithm of integrating the geodesic deviation equations on Riemannian (or pseudo-Riemannian) manifolds is obtained.

I. INTRODUCTION

In analytical dynamics, the method based upon the partial differential equation due to Hamilton and Jacobi turns out to be the most powerful way of explicitly solving the equations of motion.[1] Also in both the special and general theory of relativity, in a number of physically interesting cases, an analogous method employing the relativistic Hamilton–Jacobi equation permits reducing the process of solving the equations of motion to quadratures. In particular, this method is a very effective one when integrating the geodesic equations on manifolds being the space-times of the general theory of relativity.[2,3]

The objective of this paper is to show that the method of finding geodesics by solving the Hamilton–Jacobi equation can be generalized in such a way that it also allows one to obtain simultaneously with a solution of the geodesic equations the general solution of the equations of the geodesic deviation on a given manifold. The physical significance of the notion of geodesic deviation follows from the interpretation of the general theory of relativity. There, in the sense of a thought experiment at least, it is the measurement of a geodesic deviation field defined along a geodesic world line that enables one to detect some of the components of the curvature of space-time, which are considered to be a measure of strength of the gravitational field. The sole knowledge of a single geodesic line, without any deviation field along it, cannot on the other hand supply us with information of such a type.[4]

In lectures on classical mechanics the Hamilton–Jacobi equation is usually derived as a result of searching for such canonical transformations that make the motion look simple. It is the generating function of these transformations that satisfies the Hamilton–Jacobi equation. There is known, however, also another approach to this equation. Its starting point is the action functional in Hamilton's variational principle. After one substitutes for the functional argument a true motion being a solution to the Lagrange equations, the action functional turns over into a function of coordinates and time, called the principal function of Hamilton, which must satisfy the Hamilton–Jacobi equation. A description of this approach to the Hamilton-Jacobi equation of classical dynamics is presented in Sec. II of the paper. The reason for including this reviewing section was not only that an analogous method would serve in the following sections to introduce the Hamilton–Jacobi formalism dealing with the relativistic problems considered here, but also that it would help emphasize certain differences between the Hamilton–Jacobi equations in the classical and in the relativistic mechanics, respectively.

Section III contains the derivation of the Hamilton–Jacobi equation for geodesics on manifolds, with the exclusion of the case of null geodesics, and a discussion of the differences just mentioned. The Hamilton–Jacobi equations obtained in Secs. II and III are only formal relations that must be satisfied by the two Hamilton's principal functions, respectively, and the meaning they have for the process of integration of the equations of motion does not at all follow from their derivation. The same could also be said about the canonical transformation method of deriving this equation. The significance of the Hamilton–Jacobi equation for dynamics results first from a theorem due to Jacobi on its complete integral.[5] In the case of classical Newtonian dynamics the proof of this theorem is given in nearly every textbook and it would be superfluous to repeat it here. It must also be noted that since the Hamilton–Jacobi equation for geodesics can always be solved with respect to the time derivative $\partial S/\partial t$, the classical Jacobi theorem on the complete integral is applicable to the relativistic case as well. Such a justification is, however, not manifestly relativistically covariant, and may leave one with a feeling of dissatisfaction when comparison is made with the beautiful geometric form of the equation itself; especially that there is no difficulty in formulating a theorem of this kind in a manifestly covariant way. The formulation and proof of such a theorem are presented in Sec. IV of the paper. The theorem, besides its obvious methodological advantage for the problem of geodesics considered for its own sake, is a starting point to a generalization enhancing both geodesics and geodesic deviations, which is the main goal of the paper.

The approach to the Hamilton–Jacobi equation presented in Secs. II and III is in Sec. V applied to a variational principle that was already formulated a few years ago[6] and which leads to both the geodesic and the geodesic deviation equations simultaneously. It turns out that in this case the principal function must fulfill not one but two relations that by analogy with the cases in former sections may be called the Hamilton–Jacobi equations of the dynamical system consisting of geodesic lines and deviation vector fields determined along these lines. Section VI contains the formulation and proofs of theorems that can be considered as an extension of the classical Jacobi theorem on the complete integral to the case of the equations of geodesic deviation.

 0022-2488/89/051018-12$02.50

The main result of the paper is an algorithm solving the geodesic deviation equations in cases when one knows a complete integral of the Hamilton–Jacobi equation for the problem of geodesics. A characteristic feature of this algorithm, of some practical importance, is that solving the geodesic deviation equations with the help of this method does not require any new integration nor any knowledge of the connection nor curvature coefficients on the given manifold. All that is needed for this procedure is computing derivatives and solving an algebraic system of linear equations.

The whole work makes use of the standard variational principle in which the action is the proper time measured along the world line, this being the functional argument. This variational principle determines a geodesic in an arbitrary parametrization. In the literature[7,8] one can also encounter another action principle, sometimes called the dynamic action principle, the variation of which leads directly to the geodesic equations in an affine parametrization. Since this action principle has some supporters, in Appendix A the formalism considered in the paper has been discussed from the point of view of this second principle as well. For the sake of sole computations the two variational principles are of course completely equivalent to each other. From an orthodox geometric or relativistic point of view, however, the action principle appearing in the main body of the paper must be considered as the only physically correct one, for the other principle requires introducing into the formalism an additional element that is completely superfluous at a rigorously relativistic formulation of the problem. A fuller presentation of such a standpoint is given in Appendix A.

The paper neglects completely the case of deviations between null geodesics, as it calls for an entirely different approach that is going to be a subject of another paper by the author.

II. HAMILTON-JACOBI EQUATION VERSUS THE ACTION PRINCIPLE

The subject of this section is a Newtonian dynamical theory based upon an action of the form

$$S[q] = \int_{t_0}^{t_1} L(q(t), \dot{q}(t), t)dt, \tag{2.1}$$

where $q(t) = (q^1(t), \ldots, q^n(t))$ is a set of functions of time t that describes in a configuration space Q_n the motion of a mechanical system characterized by a Lagrange function L, and where $\dot{q} = dq/dt$.

As is customary, the notion of the variation of the motion $q(t)$ is introduced as a one-parametric family of functions $q(t, \epsilon)$, where $t \in (t_0, t_1)$ and $\epsilon \in (\epsilon_0, \epsilon_1)$, that satisfy (i) certain regularity requirements,[9] (ii) the conditions $q(t, 0) = q(t)$, $q(t_i, \epsilon) = q(t_i)$ for $i = 0, 1$. The variation vector $\bar{\delta}$ (also called an infinitesimal variation, or just a variation of q) is then defined to be a vector whose components are

$$\bar{\delta} q^k = \frac{\partial q^k}{\partial \epsilon}(t, 0)\delta\epsilon, \tag{2.2}$$

where $\delta\epsilon$ is an arbitrary increment of the value of ϵ. Similarly, a variation of the time t is a one-parametric family of functions $t(\epsilon)$ satisfying conditions analogous to (i) and (ii), and the variation vector associated with this family is

$$\delta t = \frac{dt}{d\epsilon}(0)\delta\epsilon. \tag{2.3}$$

A complete infinitesimal variation $\delta q(t)$, a variation of the motion with the variation of time, is in turn constructed in terms of $\bar{\delta} q^k$ and δt as

$$\delta q^k = \bar{\delta} q^k + \dot{q}^k \, \delta t. \tag{2.4}$$

These definitions enable one to compute the complete variation δS of the action (2.1):

$$\begin{aligned} \delta S &= L\, \delta t \,|_{t_0}^{t_1} + \int_{t_0}^{t_1} \bar{\delta} L \, dt \\ &= \left[\frac{\partial L}{\partial \dot{q}^k} \delta q^k - \left(\frac{\partial L}{\partial \dot{q}^k} \dot{q}^k - L \right) \delta t \right]_{t_0}^{t_1} \\ &\quad + \int_{t_0}^{t_1} \left(\frac{\partial L}{\partial q^k} - \frac{d}{dt} \frac{\partial L}{\partial \dot{q}^k} \right) \bar{\delta} q^k \, dt. \end{aligned} \tag{2.5}$$

The expression above may serve various purposes.

First of all, as is well known, it can be applied, in the special case of $\delta t = 0$ and $\delta q = 0$ at the boundary, to the derivation of the equations of motion from the stationary action principle or, in the more general case, to the proof of the Noether identities.

Another, perhaps less known application of Eq. (2.5) is an old method, going back to Hamilton, of deriving the form of the Hamilton–Jacobi equation, which in the majority of lectures on mechanics is nowadays being derived by a different approach, based upon the canonical formalism. The old method, however, has turned out to be a very useful one in relativistic mechanics and some examples of such applications will be presented in the paper. That is why a general outline of deriving the Hamilton–Jacobi equation from an action functional of the form (2.1) is being reviewed in the present section. The purpose is not only to set up a general framework, but also to enable one to make a comparison between the classical and the relativistic cases that are discussed in the following sections.

The starting point of the Hamilton approach is constructing from the action functional (2.1) the so-called principal function $S(q, t)$ (cf. Ref. 5). If (q_0, t_0) is a point in the space of events (i.e., in the product of the configuration space and the time axis) and (q, t) is another point, belonging to a sufficiently close neighborhood of (q_0, t_0), then from the uniqueness theorem of solutions of differential equations it follows that through the two points (q_0, t_0) and (q, t) there is exactly one solution of the Lagrange equations, provided the Lagrangian L is sufficiently regular. Assuming that (q_0, t_0) is given, the values of the principal function at any point (q, t) from a sufficiently close neighborhood of (q_0, t_0) can thus be computed from (2.1) as a result of both substituting for $q(t)$ in the integrand the unique motion that passes through the points (q_0, t_0) and (q, t) and replacing at the same time the value t_1 in the upper limit of the integral in (2.1) by the value t. Owing to this procedure, the differential $dS(q, t)$ of the principal function is given by Eq. (2.5), in which one must take into account that $q(t)$ satisfies the Lagrange equations. Additionally, making in Eq. (2.5) use of the definitions of the conjugated momenta, $p_k = \partial L / \partial \dot{q}^k$, and of the Hamiltonian $H(q, p, t)$, one obtains

$$dS = p_k \, dq^k - H(q^k, p_l, t) dt. \tag{2.6}$$

On comparing now the coefficients in this expression to the appropriate derivatives of S, one derives the equations

$$\frac{\partial S}{\partial q^k}(q, t) = p_k(q, t), \quad \frac{\partial S}{\partial t}(q, t) = -H(q, p(q, t), t), \tag{2.7}$$

which result in the single relationship

$$\frac{\partial S(q, t)}{\partial t} = -H\left(q, \frac{\partial S(q, t)}{\partial q}, t\right) \tag{2.8}$$

being the Hamilton–Jacobi equation that must be satisfied by the principal function $S(q, t)$. Of course, the method of deriving Eq. (2.8) presented here results in a formal relationship for a very special function only and from a more general point of view is inferior to the method based on canonical transformations that reveal the beautiful dynamical meaning of every solution of Eq. (2.8). Nevertheless, in applications into which the canonical formalism would carry in complexities of its own, the principal function method can be used as a short cut when deriving the form of the Hamilton–Jacobi equation. The value of the so derived equation as a method of finding solutions to the equations of motion is, however, first established after one proves a theorem of the type of the Jacobi theorem about a complete integral of Eq. (2.8). In the case of a dynamical system defined by an action principle of the form (2.1), this well-known theorem is quoted here just for completeness and future reference.

Theorem 2.1 (Jacobi): If a function S: $Q_n \times \mathbb{R}^{n+1} \to \mathbb{R}$, defined in the form $S = S(q^k, a^k, t)$, where a^k (for $k = 1, \ldots, n$) are independent parameters, is a complete integral of Eq. (2.8), then the system of equations

$$\frac{\partial S}{\partial a^l}(q^k, a^m, t) = \alpha_l, \tag{2.9}$$

where α_l, $l = 1, \ldots, n$, are some new parameters, defines in an implicit way n functions $\xi_k : \mathbb{R}^{2n+1} \to \mathbb{R}$ such that (i) the relations

$$q^k = \xi^k(t, a^l, \alpha_m) \tag{2.10}$$

satisfy identically Eqs. (2.9); (ii) they enable one to determine n new functions $\eta_k : \mathbb{R}^{2n+1} \to \mathbb{R}$, given as

$$p_k = \eta_k(t, a^l, \alpha_m):$$
$$= \frac{\partial S}{\partial q^k}(\xi^l(t, a^i, \alpha_j), a^l, t), \tag{2.11}$$

and for any values of a^k, α_m the dynamical variables of q^k, p_l defined by Eqs. (2.10) and (2.11), respectively, are solutions to the Hamilton equations of motion of the dynamical problem defined by the action (2.1).

III. THE CASE OF GEODESICS

Let $(M, g_{\alpha\beta})$ be a four-dimensional pseudo-Riemannian space-time manifold of signature -2 in which γ is a timelike curve described in a coordinate map by four functions ξ^α of an arbitrary real parameter $\tau \in [\tau_0, \tau_1]$ in the form $x^\alpha = \xi^\alpha(\tau)$.

The action functional is

$$U[\gamma] = \int_{\tau_0}^{\tau_1} (g_{\alpha\beta} \dot{\xi}^\alpha \dot{\xi}^\beta)^{1/2} d\tau, \tag{3.1}$$

where $\dot{\xi}^\alpha = d\xi^\alpha / d\tau$. The variations of ξ^α and τ are defined in analogy to Eqs. (2.2) and (2.3) as

$$\bar{\delta}\xi^\alpha = \frac{\partial \xi^\alpha}{\partial \epsilon}(\tau, 0)\delta\epsilon, \quad \delta\tau = \frac{\partial \tau}{\partial \epsilon}(0)\delta\epsilon. \tag{3.2}$$

Similarly as in Sec. II, one defines the complete variation $\delta\xi^\alpha$ of ξ^α with the variation of the parameter τ by means of the relation

$$\delta\xi^\alpha = \bar{\delta}\xi^\alpha + \dot{\xi}^\alpha \delta\tau. \tag{3.3}$$

The corresponding complete variation of the action (3.1) can then be brought to the form

$$\delta U = \frac{g_{\alpha\beta}\dot{\xi}^\beta}{(\dot{\xi}_\sigma \dot{\xi}^\sigma)^{1/2}} \delta\xi^\alpha \Big|_{\tau_0}^{\tau_1} - \int_{\tau_0}^{\tau_1} d\tau \frac{D}{d\tau}\left(\frac{g_{\alpha\beta}\dot{\xi}^\beta}{(\dot{\xi}_\rho \dot{\xi}^\rho)^{1/2}}\right) \bar{\delta}\xi^\alpha. \tag{3.4}$$

Now, in a similar way as in Sec. II, the principal function $U(x^\alpha, \tau)$ can be constructed, and because of Eq. (3.4) its differential is

$$dU = \frac{g_{\alpha\beta}\dot{\xi}^\beta}{(\dot{\xi}_\rho \dot{\xi}^\rho)^{1/2}} dx^\alpha. \tag{3.5}$$

Hence

$$\frac{\partial U}{\partial x^\alpha} = \frac{g_{\alpha\beta}\dot{\xi}^\alpha}{(\dot{\xi}_\rho \dot{\xi}^\rho)^{1/2}}, \quad \frac{\partial U}{\partial \tau} = 0. \tag{3.6}$$

These relations correspond to Eqs. (2.7). Now, however, instead of the Hamilton–Jacobi equation (2.8) we have a trivial relation $\partial U / \partial \tau = 0$, which is, strictly speaking, the Hamilton–Jacobi equation for geodesics. Additionally, the derivatives $\partial U / \partial x^\alpha$ are, in accordance with Eqs. (3.6), components of a unit vector and thus the following constraint condition is satisfied:

$$g^{\alpha\beta} \frac{\partial U}{\partial x^\alpha} \frac{\partial U}{\partial x^\beta} = 1, \tag{3.7}$$

supplying us with an additional, but now nontrivial restriction on the derivatives of U.

The derivation of Eq. (3.7) presented here reveals only that this equation must be satisfied by the principal function obtained from the action (3.1), but gives one no hint about its dynamical significance. It can, however, always be solved with respect to $\partial U / \partial x^0$ and written in a form analogous to Eq. (2.8), i.e., in a form to which Theorem 2.1 is applicable. And this supplies one with an argument that justifies the widely known dynamical application of Eq. (3.7) as of the Hamilton–Jacobi equation in relativity. Solving (3.7) with respect to $\partial U / \partial x^0$ and making use of the nonrelativistic Theorem 2.1 spoil, unfortunately, the manifest relativistic covariance of Eq. (3.7). It is, however, possible to prove a theorem of a kind of Theorem 2.1 that is manifestly relativistic and geometric in its form. Such an theorem is presented in the next section. Its application to relativistic dynamics is in general simpler than that of Theorem 2.1. The new theorem also forms a basis of a method of the Hamilton–Jacobi type of solving the equations of geodesic deviation that is presented in this paper.

IV. THE JACOBI THEOREM FOR GEODESICS

Let Eq. (3.7) be considered now as a partial differential equation on a scalar function $U: M \to \mathbb{R}$, which is defined in a coordinate map of the space-time M by a relation of the form $U = U(x^\alpha)$.

Definition 4.1: A function $U: M \times \mathbb{R}^3 \to \mathbb{R}$, defined in a coordinate map of the manifold M by an equation of the form

$$U = U(x^\alpha, a^k), \tag{4.1}$$

where a^k $(k = 1,2,3)$ are three parameters taken from some open intervals $I_{(k)} \in \mathbb{R}$, is called a complete integral of the partial differential equation (3.7) iff (i) the function $U(x^\alpha, a^k)$ satisfies Eq. (3.7) identically for all admissible values of the parameters a^k, (ii) the 4×3 matrix $(\partial^2 U/\partial x^\mu\, \partial a^k)$ is of rank 3.

Remarks: (1) The knowledge of a complete integral of a partial differential equation permits one to reconstruct the equation. This can easily be seen by noting that as a result of differentiating Eq. (4.1) with respect to x^α one obtains four relations that involve the coordinates x^α, parameters a^k, and derivatives $\partial U/\partial x^\nu$. Then, due to the property (ii) of Definition 4.1, one can eliminate from these relations all three parameters a^k and one is left with a single relationship of the form $\Psi(x^\mu, \partial U/\partial x^\nu) = 0$. This relationship, in virtue of the property (i), must be equivalent to the differential equation (3.7).

(2) The parameters a^k are independent of the coordinates x^α. In the sequel it is thus assumed that under transformations of the coordinates x^α the parameters a^k are scalars. This property is, however, independent of the fact that one can always replace the parameters a^k in Eq. (4.1) by some new parameters being a result of a nonsingular transformation applied to the old set of a^k's

Theorem 4.1: Let the function (4.1) be a complete integral of the partial differential equation (3.7), then the system of algebraic equations

$$\frac{\partial U(x^\alpha, a^k)}{\partial a^k} = \alpha_k, \tag{4.2}$$

where α_k are some new parameters taken from certain real intervals, determines four functions $\xi^\alpha: \mathbb{R}^7 \to \mathbb{R}$ which are a coordinate description of a family of world lines Γ,

$$x^\alpha = \xi^\alpha(f(\tau), a^k, \alpha_l) \tag{4.3}$$

where f is an arbitrary monotonous function of a parameter τ, and the functions ξ^α are such that[10]

(i) $$\frac{\partial U}{\partial a^k}(\xi^\alpha(\tau, a^m, \alpha_n)) \equiv \alpha_k, \tag{4.4}$$

(ii) there exists a function $\lambda = \lambda(\tau)$ such that

$$\lambda(\tau)\frac{\partial U}{\partial x^\mu}(\xi^\alpha(\tau, a^k, \alpha_l), \alpha_m) = g_{\mu\nu}(\xi^\alpha(\tau, a^k, \alpha_l))\frac{d}{d\tau}\xi^\nu(\tau, a^k, \alpha_l), \tag{4.5}$$

(iii) every world line from the family (4.3) is a geodesic parametrized by an arbitrary parameter τ

$$\frac{D}{d\tau}\left[g_{\alpha\beta}(\xi^\mu(\tau, a^k, \alpha_l))\frac{d}{d\tau}(\xi^\alpha(\tau, a^k, \alpha_l))\right] = \frac{\dot\lambda(\tau)}{\lambda(\tau)} g_{\alpha\mu}(\xi^\alpha(\tau, a^k, \alpha_l))\frac{d}{d\tau}(\xi^\alpha(\tau, a^m, \alpha_n)), \tag{4.6}$$

(iv) the functions f in Eq. (4.3) and λ in Eq. (4.5) satisfy the relation

$$\frac{df(\tau)}{d\tau} = \pm\lambda(\tau) \tag{4.7}$$

iff the new parameters $s = f(\tau)$ is the proper time along the geodesic (4.3).

Proof: The set of equations (4.4) can be supplemented by an equation of the form

$$\varphi(x^\alpha, a^k) = \tau, \tag{4.8}$$

where τ is an additional parameter independent of x^α and of all the other parameters introduced so far, and the function φ is chosen in such a way that the extended, 4×4 matrix

$$\left(\frac{\partial^2 U}{\partial x^\mu\,\partial a^k} \quad \frac{\partial\varphi}{\partial x^m}\right)$$

is nonsingular. Due to Definition 4.1 (ii), such a function φ always exists. Moreover, the extended matrix will be nonsingular for any function $\Phi = f\circ\varphi$, where $f: \mathbb{R}\to\mathbb{R}$ is an arbitrary function[11] such that $f' \neq 0$. Thus there are four algebraic equations, i.e., Eqs. (4.2) and the equation

$$\Phi(x^\alpha, a^k) = f(\tau), \tag{4.9}$$

which can locally be solved with respect to x^α, giving as a result Eqs. (4.3). The values of the so derived functions ξ^α must, of course, satisfy the identities

$$\frac{\partial U}{\partial a^k}(\xi^\alpha(f(\tau), a^k, \alpha_l), a^m) \equiv \alpha_k. \tag{4.10}$$

Since the function (4.1) is a complete integral of Eq. (3.7), it fulfills identically the equation

$$g^{\alpha\beta}\frac{\partial U(x^\mu, a^k)}{\partial x^\alpha}\frac{\partial U(x^\nu, a^l)}{\partial x^\beta} \equiv 1, \tag{4.11}$$

which means that the covariant vector $\partial U/\partial x^\alpha$ is timelike. Differentiating Eq. (4.11) with respect to a^k, one obtains

$$g^{\alpha\beta}\frac{\partial U}{\partial x^\alpha}\frac{\partial^2 U}{\partial x^\beta\,\partial a^k} = 0. \tag{4.12}$$

The three $(k = 1,2,3)$ vectors $\partial^2 U/\partial x^\beta\,\partial a^k$ are, because of Definition 4.1 (ii), linearly independent, and therefore, due to Eq. (4.12), all three of them are spacelike. Thus also the four vectors

$$\frac{\partial U}{\partial x^\alpha}, \quad \frac{\partial^2 U}{\partial x^\beta\,\partial a^k} \tag{4.13}$$

are linearly independent.

On the other hand, after differentiating Eq. (4.10) with respect to τ, one obtains

$$\frac{\partial^2 U}{\partial x^\alpha\,\partial a^k}\frac{d\xi^\alpha}{d\tau} = 0. \tag{4.14}$$

Due to Eqs. (4.12), the linear independence of the vectors (4.13), and to the fact that the number of space-time dimensions is limited by four, the vectors $g_{\alpha\beta}\xi^{\beta}$ and $\partial U/\partial x^{\alpha}$ must be parallel to each other, although the corresponding proportionality factor λ may vary from one point to another. And this is just the statement which forms the contents of Eq. (4.5).

Substituting Eq. (4.5) into (4.11), one obtains the relation

$$\left(\frac{ds(\tau)}{d\tau}\right)^2 = \lambda^2, \tag{4.15}$$

which results in Eq. (4.7).

Now, in accordance with Eq. (4.5), one can compute the left-hand side of Eq. (4.6) as follows:

$$\begin{aligned}
&\frac{D}{d\tau}\left(g_{\alpha\beta}(\xi(\dots))\frac{d\xi^{\beta}}{d\tau}\right)\\
&= \dot\lambda\frac{\partial U}{\partial x^{\alpha}}(\xi(\dots),a) + \lambda\left(\frac{\partial U}{\partial x^{\alpha}}(\xi(\dots))\right)_{;\mu}\frac{d\xi^{\mu}}{d\tau}\\
&= \dot\lambda\frac{\partial U}{\partial x^{\alpha}}(\xi(\dots),a) + \lambda^2\frac{\partial}{\partial x^{\alpha}}\left(\frac{1}{2}g^{\mu\nu}\frac{\partial U}{\partial x^{\mu}}\frac{\partial U}{\partial x^{\nu}}\right)_{x=\xi(\dots)}\\
&= \frac{\dot\lambda}{\lambda}g_{\alpha\beta}\frac{d\xi^{\beta}}{d\tau},
\end{aligned}$$

where also Eq. (4.11) has been taken into account. This proves the validity of Eq. (4.5) and ends the proof of Theorem 4.1.

Remarks: (1) Equation (4.4) can of course also be regarded as the statement that the functions $\partial U/\partial a^k$ are constants of the motion of the geodesic equations (4.6).

(2) From the fact that the four vectors (4.13) are linearly independent one can conclude that the complete integral $U(x^{\alpha}, a^k)$ itself can be used instead of an arbitrary function Φ in Eq. (4.9). In such a case, due to Eq. (4.5) one obtains

$$\frac{df}{d\tau} = \frac{dU}{d\tau} = \frac{\partial U}{\partial x^{\alpha}}\frac{d\xi^{\alpha}}{d\tau} = \lambda(\tau)g^{\alpha\beta}\frac{\partial U}{\partial x^{\alpha}}\frac{\partial U}{\partial x^{\beta}} = \lambda(\tau), \tag{4.16}$$

which in accordance with (4.7) means that the family of geodesics

$$x^{\alpha} = \xi^{\alpha}(s - s_0, a^k, \alpha_l), \tag{4.17}$$

derived as the unique solution of the set of equations

$$\frac{\partial U(x^{\alpha}, a^k)}{\partial a^k} = \alpha_k,\ U(x^{\alpha}, a^k) = s - s_0, \tag{4.18}$$

is the general solution of the geodesic equations parametrized by the natural parameter s. This solution depends, as it should, on seven arbitrary constants s_0, a^k, and α_l, which is in agreement with the fact that a non-null geodesic is uniquely given by specifying four independent coordinates of its initial point and, due to the condition $g_{\alpha\beta}\, u^{\alpha}\, u^{\beta} = 1$, three independent components u^{α} of its initial tangent vector.

Thus, if one is interested in formulating a theorem analogous to Theorem 4.1, but for the case of natural parametrization, it can be easily done by replacing Eqs. (4.2) with Eqs. (4.18), and by substituting all over in Theorem 4.1 both $s - s_0$ for the function $f(\tau)$ and unity for $\lambda(\tau)$.

It should be noted that Theorem 4.1 as well as the last remark apply only to non-null geodesics. The null case requires a separate discussion that will be presented in another paper.

V. THE GEODESIC DEVIATION

As was already noted in Ref. 6, the equations that determine the law of evolution of geodesic deviation vector $r(\tau)$ along a geodesic γ: $x^{\alpha} = \xi^{\alpha}(\tau)$ parametrized by an arbitrary parameter τ can be derived together with the geodesic equations from a unified variational principle based upon the action functional

$$S[\gamma, r] = \int_{\tau_0}^{\tau_1} d\tau\, g_{\alpha\beta}(\xi)\frac{\dot\xi^{\alpha}}{(\dot\xi_{\rho}\dot\xi^{\rho})^{1/2}}\frac{Dr^{\alpha}}{d\tau}. \tag{5.1}$$

In this functional, the four functions $\xi^{\alpha}(\tau)$ determine, in a coordinate system $\{x^{\alpha}\}$, a curve which joins two fixed points p_0 and p_1 with the coordinates $\xi^{\alpha}(\tau_0)$ and $\xi^{\alpha}(\tau_1)$, respectively; and the functions $r^{\alpha}(\tau)$ are components of a vector $r(\tau)$ from the tangent space at the point $\xi^{\alpha}(\tau)$. The variation of (5.1) with respect to the variables $r^{\alpha}(\tau)$ leads to the equations of geodesics on $\xi^{\alpha}(\tau)$, whereas with respect to $\xi^{\alpha}(\tau)$, it leads to the equations of the geodesic deviation on $r^{\alpha}(\tau)$, both, respectively, in an arbitrary parametrization.

In this section, equations will be derived that are satisfied by the principal function determined by the action (5.1). Similarly as previously stated, one starts with the definitions of the ordinary variations $\bar\delta\xi^{\alpha}$, $\bar\delta r^{\alpha}$, and $\delta\tau$,

$$\bar\delta\xi^{\alpha} = \frac{\partial\xi^{\alpha}}{\partial\epsilon}(\tau,0),\quad \bar\delta r^{\alpha} = \frac{\partial r^{\alpha}}{\partial\epsilon}(\tau,0),\quad \delta\tau = \frac{\partial\tau}{\partial\epsilon}(\tau,0). \tag{5.2}$$

The complete variations $\delta\xi^{\alpha}$ and δr^{α} are then determined by the equations

$$\delta\xi^{\alpha} = \bar\delta\xi^{\alpha} + \dot\xi^{\alpha}\,\delta\tau,\quad \delta r^{\alpha} = \bar\delta^{\alpha} + \dot r^{\alpha}\,\delta\tau. \tag{5.3}$$

As a result of a computation, which can be facilitated by using the concept of covariant variation described, e.g., in Ref. 6, one obtains

$$\begin{aligned}
\delta S = {} & \frac{1}{(\dot\xi_{\rho}\dot\xi^{\rho})^{1/2}}\left(h_{\alpha\beta}\frac{Dr^{\beta}}{d\tau} + \Gamma^{\beta}{}_{\alpha\gamma}r^{\gamma}\dot\xi_{\beta}\right)\delta\xi^{\alpha}\Big|_{\tau_2}^{\tau_1}\\
& - \frac{1}{(\dot\xi_{\rho}\dot\xi^{\rho})^{1/2}}g_{\alpha\beta}\dot\xi^{\beta}\,\delta r^{\alpha}\Big|_{\tau_0}^{\tau_1}\\
& - \int_{\tau_0}^{\tau_1} d\tau\Bigg\{\Bigg[\frac{D}{d\tau}\left(\frac{h_{\alpha\beta}}{(\dot\xi_{\rho}\dot\xi^{\rho})^{1/2}}\frac{Dr^{\beta}}{d\tau}\right)\\
& + \frac{1}{(\dot\xi_{\rho}\dot\xi^{\rho})^{1/2}}R_{\alpha\beta\gamma\delta}\dot\xi^{\beta}r^{\gamma}\dot\xi^{\delta}\\
& + \Gamma^{\gamma}{}_{\alpha\beta}r^{\beta}\frac{D}{d\tau}\left(\frac{1}{(\dot\xi_{\rho}\dot\xi^{\rho})^{1/2}}\right)\Bigg]\bar\delta\xi^{\alpha}\\
& + \frac{D}{d\tau}\left(\frac{\dot\xi_{\alpha}}{(\dot\xi_{\rho}\dot\xi^{\rho})^{1/2}}\right)\bar\delta r^{\alpha}\Bigg\},
\end{aligned} \tag{5.4}$$

where

$$h_{\alpha\beta} = g_{\alpha\beta} - \dot\xi_{\alpha}\dot\xi_{\beta}/(\dot\xi_{\rho}\dot\xi^{\rho})^{1/2} \tag{5.5}$$

is the projection tensor on the hyperplane orthogonal to $\dot\xi^{\alpha}$.

In the special case, when the variation of the parameter vanishes, $\delta\tau = 0$, and when the variations $\delta\xi^{\alpha}$ and δr^{α} are independent and are vanishing for τ being equal to both τ_0 and τ_1, the requirement $\delta S = 0$ is equivalent to the system of equations that simultaneously describe geodesics and geodesic deviation in an arbitrary parametrization (cf. Ref. 6), that is to the equations

$$\frac{D}{d\tau}\left(\frac{\dot{\xi}^{\alpha}}{(\dot{\xi}_{\rho}\dot{\xi}^{\rho})^{1/2}}\right)=0 \tag{5.6}$$

and

$$\frac{D}{d\tau}\left(\frac{h_{\alpha\beta}}{(\dot{\xi}_{\rho}\dot{\xi}^{\rho})^{1/2}}\frac{Dr^{\beta}}{d\tau}\right)+\frac{1}{(\dot{\xi}_{\rho}\dot{\xi}^{\rho})^{1/2}}R_{\alpha\beta\gamma\delta}\dot{\xi}^{\beta}r^{\gamma}\dot{\xi}^{\delta}=0, \tag{5.7}$$

where the field quantities entering the coefficients in Eq. (5.7) are to be evaluated along the solutions of Eq. (5.6).

To start with determining the principal function corresponding to the action functional (5.1), consider a family of geodesics $x^{\alpha}=\xi^{\alpha}(\tau)$ all of which for $\tau=\tau_0$ pass through the same point $\xi^{\alpha}(\tau_0)$. Along every one of these geodesics consider further a family of geodesic deviation vector fields with the components $r^{\alpha}=r^{\alpha}(\tau)$ such that every one of its members assumes the same value $r^{\alpha}(\tau_0)$ at the point $\xi^{\alpha}(\tau_0)$, a value that is also the same for all the other geodesics passing through $\xi^{\alpha}(\tau_0)$. Every one of these geodesics together with a fixed geodesic deviation field along it defines a curve T_{γ} in the tangent bundle *TM* over *M*, i.e., a curve that passes through a point in *TM* with the local coordinates $(\xi^{\alpha}(\tau_0), r^{\alpha}(\tau_0))$. This family serves the construction of the principal function *S*. Its value $S(x^{\alpha}, r^{\beta}, \tau)$ is given by the integral (5.1) in that the upper limit is replaced by the value τ taken from the argument of *S*, and the integral is, furthermore, calculated along the curve T_{γ} in *TM* which joins the point $(\xi^{\alpha}(\tau_0), r^{\beta}(\tau_0))$ with the point (x^{α}, r^{α}) whose coordinates enter the argument of *S*. (It is tacitly assumed that the two points are close enough to each other to ensure the uniqueness of the construction, which is therefore in such a sense a local one.) Then, due to the assumption that T_{γ} satisfies Eqs. (5.6) and (5.7), the integrand in Eq. (5.4) vanishes and the differential of the principal function *S* is equal to

$$dS=\frac{1}{(\dot{\xi}_{\rho}\dot{\xi}^{\rho})^{1/2}}\left(h_{\alpha\beta}\frac{Dr^{\beta}}{d\tau}+g_{\mu\nu}\dot{\xi}^{\mu}\Gamma^{\nu}{}_{\sigma\alpha}r^{\sigma}\right)dx^{\alpha}+g_{\alpha\beta}\frac{\dot{\xi}^{\beta}}{(\dot{\xi}_{\rho}\dot{\xi}^{\rho})^{1/2}}dr^{\alpha}. \tag{5.8}$$

This implies that

$$\frac{\partial S}{\partial\tau}=0,\quad \frac{\partial S}{\partial r^{\alpha}}=g_{\alpha\beta}\frac{\dot{\xi}^{\beta}}{(\dot{\xi}_{\rho}\dot{\xi}^{\rho})^{1/2}},\qquad \frac{\partial S}{\partial x^{\alpha}}=\frac{1}{(\dot{\xi}^{\rho}\dot{\xi}_{\rho})^{1/2}}\left(h_{\alpha\beta}\frac{Dr^{\alpha\beta}}{d\tau}+g_{\mu\nu}\dot{\xi}^{\mu}\Gamma^{\nu}{}_{\sigma\alpha}r^{\sigma}\right). \tag{5.9}$$

Since the right-hand sides of Eqs. (5.9), as is easily seen, are not independent, the left-hand sides must satisfy certain relations. As a result, besides the trivial equation $\partial S/\partial\tau=0$, one obtains two other equations,

$$g^{\alpha\beta}\frac{\partial S}{\partial r^{\alpha}}\frac{\partial S}{\partial r^{\beta}}=1, \tag{5.10}$$

$$g^{\alpha\beta}\frac{\partial S}{\partial r^{\alpha}}\left(\frac{\partial S}{\partial x^{\beta}}-\Gamma^{\mu}{}_{\beta\nu}r^{\nu}\frac{\partial S}{\partial r^{\mu}}\right)=0, \tag{5.11}$$

that must be fulfilled by the principal function $S(x^{\alpha}, r^{\beta})$. By analogy with the derivation demonstrated in Secs. II and III, one can expect that Eqs. (5.10) and (5.11) will play the role of the Hamilton–Jacobi equation for the simultaneous problem of geodesics and geodesic deviations in space-time. This expectation, however, will be first confirmed by theorems formulated in the following section. The behavior of Eq. (5.11) under transformations of coordinates is briefly discussed in Appendix B.

VI. THEOREMS ON INTEGRATING THE GEODESIC DEVIATION EQUATIONS

Equations (5.10) and (5.11) will now be treated as a simultaneous set of partial differential equations of the first order on a single unknown function *S* defined in a region of the tangent bundle *TM* over the space-time manifold *M*. In a coordinate system $\{x^{\alpha}\}$ the values of the function *S* are equal to $S(x^{\alpha}, r^{\beta})$ and depend on eight variables x^{α}, r^{β}, where r^{β} are natural components of an arbitrary vector *r* from the tangent space T_pM at a point *p* with the coordinates x^{α}.

Definition 6.1: A function S: $M\times\mathbb{R}^6\rightarrow\mathbb{R}$, defined in a coordinate map of the manifold *M* by an equation of the form

$$S=S(x^{\alpha}, r^{\beta}, a^{k}, b^{l}), \tag{6.1}$$

where a^k and b^l $(k, l=1, 2, 3)$ are six parameters taken from some open real intervals, is called a complete integral of the simultaneous set of differential equations (5.10) and (5.11) iff (i) the function $S(x^{\alpha}, r^{\beta}, a^{k}, b^{l})$ is a simultaneous solution of Eqs. (5.10) and (5.11) for all admissible values of the parameters a^k, b^l; (ii) the 8×6 matrix

$$M_{8\times 6}=\begin{pmatrix}\dfrac{\partial^2 S}{\partial x^{\mu}\,\partial a^{k}} & \dfrac{\partial^2 S}{\partial x^{\mu}\,\partial b^{l}}\\ \dfrac{\partial^2 S}{\partial r^{\nu}\,\partial a^{k}} & \dfrac{\partial^2 S}{\partial r^{\nu}\,\partial b^{l}}\end{pmatrix} \tag{6.2}$$

is of rank 6.

From this definition it does not, of course, follow that such a first integral really exists. Its existence results, however, from the subsequent theorem.

Theorem 6.1: Every complete integral $U(x^{\alpha}, a^{k})$ of Eq. (3.7) generates a complete integral $S(x^{\alpha}, r^{\beta}, a^{k}, b^{l})$ of the simultaneous set of differential equations (5.10) and (5.11), which is of the form

$$S(x^{\alpha}, r^{\beta}, a^{k}, b^{l})=\frac{\partial U}{\partial x^{\sigma}}(x^{\alpha}, a^{k})r^{\sigma}+\frac{\partial U}{\partial a^{s}}(x^{\alpha}, a^{k})b^{s}. \tag{6.3}$$

Proof: From Eq. (6.3) it follows at once that

$$\frac{\partial S}{\partial r^{\alpha}}=\frac{\partial U}{\partial x^{\alpha}}, \tag{6.4}$$

which implies that Eq. (5.10) can simply be reduced to Eq. (3.7) and is thereby satisfied by the function (6.3).

Similarly, Eq. (6.3) gives

$$\frac{\partial S}{\partial x^{\alpha}}=\frac{\partial^2 U}{\partial x^{\alpha}\,\partial x^{\sigma}}r^{\sigma}+\frac{\partial^2 U}{\partial x^{\alpha}\,\partial a^{s}}b^{s}. \tag{6.5}$$

After substituting Eqs. (6.4) and (6.5) into Eq. (5.11), one has to take into account that the second covariant derivatives of a scalar function commute, $U_{;\sigma\beta}=U_{;\beta\sigma}$. Moreover, one must make use of Eq. (4.11) and of equations that are its consequence, that is, of Eq. (4.12) and the equation

$$g^{\alpha\beta}\frac{\partial U}{\partial x^\alpha}\left(\frac{\partial U}{\partial x^\beta}\right)_{;\sigma}=0. \tag{6.6}$$

As a result of all that, one obtains

$$\begin{aligned}g^{\alpha\beta}\frac{\partial S}{\partial r^\alpha}&\left(\frac{\partial S}{\partial x^\beta}-\Gamma^\mu{}_{\beta\sigma}r^\sigma\frac{\partial S}{\partial r^\mu}\right)\\&=g^{\alpha\beta}\frac{\partial U}{\partial x^\alpha}\left(\frac{\partial^2 U}{\partial x^\beta\,\partial x^\sigma}r^\sigma+\frac{\partial^2 U}{\partial x^\beta\,\partial a^s}b^s-\Gamma^\mu{}_{\beta\sigma}r^\sigma\frac{\partial U}{\partial x^\mu}\right)\\&=g^{\alpha\beta}\frac{\partial U}{\partial x^\alpha}\left(\frac{\partial U}{\partial x^\beta}\right)_{;\sigma}r^\sigma+g^{\alpha\beta}\frac{\partial U}{\partial x^\alpha}\frac{\partial^2 U}{\partial x^\beta\,\partial a^s}b^s\equiv 0,\end{aligned}$$

which shows that the function (6.3) solves Eq. (5.11) as well.

To verify that the function (6.3) is a complete integral of Eqs. (5.10) and (5.11), one must additionally evaluate the elements of the matrix (6.2) for S given by Eq. (6.3). An obvious computation leads to

$$M_{8\times 6}=\begin{pmatrix}\dfrac{\partial^2 S}{\partial x^\mu\,\partial a^k} & \dfrac{\partial^2 U}{\partial x^\mu\,\partial a^l}\\ \dfrac{\partial^2 U}{\partial x^\nu\,\partial a^k} & \mathbb{O}\end{pmatrix}, \tag{6.7}$$

where $\mathbb{O}$ stands for the zero 4×3 matrix. Since $\partial^2 U/\partial x^\alpha\,\partial a^k$ are three linearly independent four-dimensional vectors, the columns of the matrix (6.7) form six linearly independent eight-dimensional vectors and hence in the case considered the matrix (6.2) is of rank 6. This concludes the proof of Theorem 6.1.

Theorem 6.2: If the function $S(x^\alpha, r^\beta, a^k, b^l)$ is given by Eq. (6.3), where the function $U(x^\alpha, a^k)$ is a complete integral of Eq. (3.7), then the following conclusions are valid.

(i) The equations

$$\frac{\partial S}{\partial b^k}\equiv\frac{\partial U}{\partial a^k}(x^\alpha, a^l)=\alpha_k,\quad k=1,2,3, \tag{6.8}$$

determine in accordance with Theorem 4.1 a family of geodesics parametrized by a^k and α_l, $k, l=1, 2, 3$ which is described by the equations

$$x^\alpha=\xi^\alpha(f(\tau), a^\tau, \alpha_l). \tag{6.9}$$

(ii) The equations

$$\frac{\partial S}{\partial a^k}(\xi^\mu(\tau, a^k, \alpha_l), r^\nu, a^m, b^n)=\beta_k,\quad k=1,2,3,$$

$$\frac{\partial S}{\partial r^\nu}r^\nu=\mu(\tau), \tag{6.10}$$

where β_k are some additional parameters taken from certain real intervals, and $\mu(\tau)$ is an arbitrary (but considered as an *a priori* fixed) function, can locally be solved with respect to r^ν. As a result, they determine uniquely four functions ρ^α such that the equations

$$r^\alpha=\rho^\alpha(\tau, a^k, \alpha_l, b^m, \beta_n, \mu(\tau)) \tag{6.11}$$

define, for constant values of the parameters $a^k, \alpha_l, b^m, \beta_n$, a vector field r along the geodesic that is chosen from the family (6.9) by specifying the same as in (6.1), and in (6.11), values of the parameters a^k and α_l.

(iii) The vector fields ρ^α determined by Eq. (6.11) satisfy the following three conditions.

$$\text{(a)}\quad \frac{\partial S}{\partial r^\mu}\frac{D\rho^\mu}{d\tau}=\dot\mu(\tau), \tag{6.12}$$

where the derivatives $\partial S/\partial r^\mu$ are evaluated for x^α and r^α satisfying Eqs. (6.9) and (6.11), respectively. Moreover, in virture of (4.5) and (6.4), Eq. (6.12) is equivalent to

$$g_{\alpha\beta}\frac{d\xi^\alpha}{d\tau}\frac{D\rho^\beta}{d\tau}=\lambda\,\dot\mu, \tag{6.13}$$

$$\begin{aligned}\text{(b)}\quad \frac{D\rho^\alpha}{d\tau}&=\lambda(\tau)g^{\alpha\beta}\left(\frac{\partial S}{\partial x^\beta}-\Gamma^\mu{}_{\beta\sigma}r^\sigma\frac{\partial S}{\partial r^\mu}\right)\bigg|_{\substack{x=\xi\\ r=\rho}}\\&\quad+\dot\mu(\tau)g^{\alpha\beta}\frac{\partial S}{\partial r^\beta}\bigg|_{\substack{x=\xi\\ r=\rho}}\\&=\lambda(\tau)g^{\alpha\beta}\left[\left(\frac{\partial U}{\partial x^\beta}\right)_{;\sigma}r^\sigma+\frac{\partial^2 U}{\partial x^\beta\,\partial a^l}b^l\right]\bigg|_{\substack{x=\xi\\ r=\rho}}\\&\quad+\dot\mu(\tau)g^{\alpha\beta}\frac{\partial U}{\partial x^\beta}\bigg|_{x=\xi},\end{aligned} \tag{6.14}$$

where all the functions of x^α and r^α are evaluated in accordance with Eqs. (6.9) and (6.11), respectively,

$$\begin{aligned}\text{(c)}\quad &\frac{D^2\rho^\alpha}{d\tau^2}+R^\alpha{}_{\beta\gamma\delta}\frac{d\xi^\beta}{d\tau}\rho^\gamma\frac{d\xi^\delta}{d\tau}\\&=\frac{\dot\lambda}{\lambda}\frac{D\rho^\alpha}{d\tau}+\frac{d}{d\tau}\left(\frac{\dot\mu}{\mu}\right)\frac{d\xi^\alpha}{d\tau},\end{aligned} \tag{6.15}$$

where the curvature tensor is taken at points whose coordinates satisfy Eqs. (6.9).

Proof: From Eq. (6.3) it follows that Eq. (6.8) is equivalent to (4.2) and conclusion (i) of the theorem is therefore a consequence of Theorem 4.1.

Due to Eqs. (6.3) and (6.4), Eqs. (6.10) can be rewritten in the form

$$\frac{\partial^2 U}{\partial x^\sigma\,\partial a^k}r^\sigma+\frac{\partial^2 U}{\partial a^k\,\partial a^s}b^s=\beta_k,\quad \frac{\partial U}{\partial x^\sigma}r^\sigma=\mu(\tau), \tag{6.16}$$

in which none of the coefficients depends on r. Thus Eqs. (6.16) form a set of four algebraic linear equations on four unknowns r^σ. In virtue of the assumption of the theorem, the vectors (4.13) are linearly independent and the determinant of the set (6.16) is different from zero, which proves the existence and uniqueness of the solution (6.11).

To prove Eq. (6.13), take the absolute derivative with respect to τ of the relation

$$\lambda\frac{\partial U}{\partial x^\alpha}\rho^\alpha=\lambda\mu,$$

which is a consequence of the second of Eqs. (6.16) and of (6.11). Making use of Eqs. (4.5), (4.6), and again of (4.5), one obtains

$$\dot\lambda\frac{\partial U}{\partial x^\alpha}\rho^\alpha+\lambda\frac{\partial U}{\partial x^\alpha}\frac{D\rho^\alpha}{d\tau}=\dot\lambda\mu+\lambda\dot\mu,$$

and from here, with the help of the second of Eqs. (6.16), one derives the equation

$$\frac{\partial U}{\partial x^\alpha}\frac{D\rho^\alpha}{d\tau}=\dot\mu \tag{6.17}$$

that is equivalent to Eq. (6.13).

As a result of Eqs. (6.10), the $D/d\tau$ derivative of $\partial S/\partial a^k$

vanishes when evaluated for the solution (6.11). On the other hand, with the aid of (6.3), this derivative can be represented in the form

$$\frac{D}{d\tau}\left(\frac{\partial S}{\partial a^k}\right)=\frac{\partial^2 U}{\partial x^\sigma\,\partial a^k}\frac{D\rho^\sigma}{d\tau}+\frac{D}{d\tau}\left(\frac{\partial^2 U}{\partial x^\sigma\,\partial a^k}\right)\rho^\sigma+b^l\frac{\partial^3 U}{\partial x^\alpha\,\partial a^l\,\partial a^k}\frac{d\xi^\alpha}{d\tau}. \qquad (6.18)$$

Due to Eqs. (4.5) and a relation that follows from covariantly differentiating Eq. (4.12), the second term in the expression (6.18) is equal to

$$\frac{D}{d\tau}\left(\frac{\partial^2 U}{\partial x^\sigma\,\partial a^k}\right)=\left(\frac{\partial^2 U}{\partial x^\sigma\,\partial a^k}\right)_{;\alpha}\frac{d\xi^\alpha}{d\tau}=-\lambda g^{\alpha\beta}\frac{\partial^2 U}{\partial x^\alpha\,\partial a^k}\left(\frac{\partial U}{\partial x^\beta}\right)_{;\sigma}, \qquad (6.19)$$

while a replacement for the third term in Eq. (6.18) can be found in accordance with the identity

$$0=\frac{\partial}{\partial a^k}\left(g^{\alpha\beta}\frac{\partial^2 U}{\partial x^\alpha\,\partial a^l}\frac{\partial U}{\partial x^\beta}\right)=g^{\alpha\beta}\frac{\partial^3 U}{\partial x^\alpha\,\partial a^k\,\partial a^l}\frac{\partial U}{\partial x^\beta}+g^{\alpha\beta}\frac{\partial^2 U}{\partial x^\alpha\,\partial a^l}\frac{\partial^2 U}{\partial x^\beta\,\partial a^k},$$

which again is a simple consequence of Eq. (4.12). On summing up all that was just stated, one can bring Eq. (6.18) to the form

$$\frac{\partial^2 U}{\partial x^\alpha\,\partial a^k}\left\{\frac{D\rho^\alpha}{d\tau}-g^{\alpha\beta}\lambda(\tau)\left[\left(\frac{\partial U}{\partial x^\beta}\right)_{;\sigma}\rho^\sigma+\frac{\partial^2 U}{\partial x^\beta\,\partial a^l}b^l\right]\right\}=0.$$

Comparing now the equation above with Eq. (4.12) and taking into account that the vectors (4.13) are linearly independent, one infers that the expression in the curly brackets is proportional to the vector $g^{\alpha\beta}\,\partial U/\partial x^\beta$ or, in other words, there exists a function $\bar{\mu}(\tau)$ such that

$$\frac{D\rho^\alpha}{d\tau}=\lambda(\tau)g^{\alpha\beta}\left[\left(\frac{\partial U}{\partial x^\beta}\right)_{;\sigma}\rho^\sigma+\frac{\partial^2 U}{\partial x^\beta\,\partial a^l}b^l\right]+\bar{\mu}(\tau)g^{\alpha\beta}\frac{\partial U}{\partial x^\beta}.$$

Contracting this expression with $\partial U/\partial x^\alpha$ and making use of Eqs. (4.11), (4.12), (6.6), and (6.17), one easily obtains that $\bar{\mu}(\tau)=\dot{\mu}(\tau)$, thus proving the validity of Eq. (6.14).

To verify Eq. (6.15), one must compute the absolute derivative with respect to τ of both sides of Eq. (6.14). Taking into account Eqs. (4.5) and (4.6), one can obtain the relation

$$\frac{D^2\rho^\alpha}{d\tau^2}=\frac{\dot{\lambda}}{\lambda}\frac{D\rho^\alpha}{d\tau}+\frac{d}{d\tau}\left(\frac{\dot{\mu}}{\lambda}\right)\lambda g^{\alpha\beta}\frac{\partial U}{dx^\beta}\bigg|_{x=\xi}+\lambda\left\{\frac{D}{d\tau}\left[g^{\alpha\beta}\left(\frac{\partial U}{\partial x^\beta}\right)_{;\sigma}\rho^\sigma+g^{\alpha\beta}\frac{\partial^2 U}{\partial x^\beta\,\partial a^l}b^l\right]\right\}_{x=\xi}. \qquad (6.20)$$

The first term in the curly bracket here can be expressed in the form (after $U_{;\beta\sigma}=U_{;\sigma\beta}$ is taken into account)

$$\frac{D}{d\tau}\left[g^{\alpha\beta}\left(\frac{\partial U}{\partial x^\beta}\right)_{;\sigma}\rho^\sigma\right]=g^{\alpha\beta}\left(\frac{\partial U}{\partial x^\sigma}\right)_{;\beta\gamma}\rho^\sigma\frac{d\xi^\gamma}{d\tau}+g^{\alpha\beta}\left(\frac{\partial U}{\partial x^\sigma}\right)_{;\beta}\frac{D\rho^\sigma}{d\tau}, \qquad (6.21)$$

while the second term can be computed with the aid of a simple consequence of Eq. (6.6), rendering

$$\frac{D}{d\tau}\left(g^{\alpha\beta}\frac{\partial^2 U}{\partial x^\beta\,\partial a^l}b^l\right)=-\lambda g^{\alpha\beta}g^{\sigma\mu}\frac{\partial^2 U}{\partial x^\mu\,\partial a^l}\left(\frac{\partial U}{\partial x^\sigma}\right)_{;\beta}b^l. \qquad (6.22)$$

Now, after substituting for $D\rho^\sigma/d\tau$ in the last term of Eq. (6.21) the right-hand side of Eq. (6.14), and after making use of Eq. (6.6), the right-hand sides of Eqs. (6.21) and (6.22) can in turn be substituted into Eq. (6.20). The outcome is

$$\frac{D^2\rho^\alpha}{d\tau^2}=\frac{\dot{\lambda}}{\lambda}\frac{D\rho^\alpha}{d\tau}+\frac{d}{d\tau}\left(\frac{\dot{\mu}}{\lambda}\right)\frac{d\xi^\alpha}{d\tau}+\lambda g^{\alpha\beta}\left(\frac{\partial U}{\partial x^\sigma}\right)_{;\beta\gamma}\rho^\sigma\frac{d\xi^\gamma}{d\tau}+\lambda^2 g^{\alpha\beta}g^{\gamma\sigma}\left(\frac{\partial U}{\partial x^\sigma}\right)_{;\beta}\left(\frac{\partial U}{\partial x^\gamma}\right)_{;\rho}\rho^\rho. \qquad (6.23)$$

The last term here can be expressed in the form

$$g^{\alpha\beta}g^{\gamma\sigma}\left(\frac{\partial U}{\partial x^\sigma}\right)_{;\beta}\left(\frac{\partial U}{\partial x^\gamma}\right)_{;\rho}=g^{\alpha\beta}\left[\frac{\partial U}{\partial x^\sigma}\left(\frac{\partial U}{\partial x^\gamma}\right)_{;\rho}g^{\gamma\sigma}\right]_{;\beta}-g^{\alpha\beta}\left(\frac{\partial U}{\partial x^\gamma}\right)_{;\rho\beta}\frac{\partial U}{\partial x^\sigma}g^{\gamma\sigma}. \qquad (6.24)$$

The first term in Eq. (6.24) vanishes due to Eq. (6.6) and in the second term use can be made of Eq. (4.5). As a result, from Eqs. (6.23) and (6.24) one obtains

$$\frac{D^2\rho^\alpha}{d\tau^2}=\frac{\dot{\lambda}}{\lambda}\frac{D\rho^\alpha}{d\tau}+\frac{d}{d\tau}\left(\frac{\dot{\mu}}{\lambda}\right)\frac{d\xi^\alpha}{d\tau}+\lambda g^{\alpha\beta}\left[\left(\frac{\partial U}{\partial x^\sigma}\right)_{;\beta\gamma}-\left(\frac{\partial U}{\partial x^\sigma}\right)_{;\gamma\beta}\right]\rho^\sigma\frac{d\xi^\gamma}{d\tau}. \qquad (6.25)$$

Applying now the Ricci identity to the expression in the square bracket and using once again Eq. (4.5) permit one to bring Eq. (6.25) to the final form given by Eq. (6.15), thus completing the proof of Theorem 6.2.

Remark: In virtue of conclusion (iii c) of Theorem 2, the first three equations (6.10) mean of course that the functions $\partial S/\partial a^k$ are first integrals of the equations of the geodesic deviation between geodesics belonging to the family determined by the complete integral U, which enters the definition (6.3) of S.

VII. AN ALGORITHM OF SOLVING THE GEODESIC DEVIATION EQUATIONS

As was shown in Ref. 6, Eq. (6.15) is a geodesic deviation equation in the situation when the parameters along the two geodesics, the basic and neighboring ones, are arbitrary and independent of each other. The function λ in Eq.

(6.15) determines in accordance with Eq. (4.7) the parameter τ along the basic geodesic γ in terms of its proper time s (if $\lambda = \text{const}$, τ is an affine parameter). Due to the second of Eqs. (6.10), the function μ in (6.15) on the other hand determines the evolution along γ of the scalar product of the unit vector tangent to γ and the deviation vector r. Since the deviation vector is defined as "joining" the points with the same numeric values of the parameters at the two neighboring geodesics, respectively, the function μ could indirectly determine the second, independent parameter along the neighboring geodesic in terms of the proper time evaluated there. (In Ref. 6 it is shown that $\lambda = \text{const}$, $\dot{\mu} = \text{const}$ correspond to a situation in which the two geodesics are parametrized by two different affine parameters, while in the case of $\lambda = \text{const}$, $\mu = \text{const}$ the two affine parameters are exactly the same; i.e., along each of the two geodesics the corresponding $\Delta\tau$ is proportional to the respective Δs with exactly the same proportionality factor. Of course, $\tau = s$ iff $\lambda = \dot{\mu} = 0$.)

Assume now that one is given a single geodesic world line described by the equation

$$x^\alpha = \xi^\alpha(\tau) \tag{7.1}$$

and also that one knows a complete integral $U(x^\alpha, a^k)$ of Eq. (3.7). From Eq. (7.1) one can find the corresponding function λ in the form

$$\lambda(\tau) = (g_{\alpha\beta}(\xi^\mu(\tau))\dot{\xi}^\alpha\dot{\xi}^\beta)^{1/2} \tag{7.2}$$

where $\dot{\xi} = d\xi/d\tau$. Suppose further that one is interested in finding along the geodesic (7.1) a vector field r satisfying Eqs. (6.15) in which $\mu(\tau)$ is a given function. [In practice, the most frequent case is that of $\dot{\mu} \equiv 0$ in which r is a deviation vector between two geodesics parametrized by the same parameter τ. By the last statement one understands simply that the norms (7.2) of the tangent vectors along the two geodesics are given by the same function $\lambda(\tau)$.] The procedure of finding the solution of Eqs. (6.15) consists then of the following steps.

(i) Determining such a set of values of the parameters τ_0, a^k, and α_l for which the general solution (4.3) defined by the known complete integral $U(x^\alpha, a^k)$ will turn over into the world line (7.1).

(ii) Evaluating the coefficients in the system of equations (6.16) along the basic geodesic (7.1).

(iii) Solving the so derived algebraic linear equations for four unknowns r^α.

To perform step (i), one must compare the right-hand sides of Eqs. (7.1) and (4.3), and also the corresponding expressions for their first $d/d\tau$ derivatives. As a result, one can uniquely determine the values of the seven constants τ_0, a^k, and α_l for which Eqs. (4.3) turn over into (7.1). In step (ii) one computes the derivatives of U that enter Eqs. (6.16) and evaluates them along the world line (7.1) for the values of a^k found in step (i). Equations (6.16) then turn into

$$\begin{aligned} &\frac{\partial^2 U}{\partial x^\sigma \partial a^k}(\xi^\alpha(\tau), a^l)r^\sigma + \frac{\partial^2 U}{\partial a^k \partial a^s} b^s = \beta_k, \\ &\frac{\partial U}{\partial x^\sigma}(\xi^\alpha(\tau), a^l)r^\sigma = \mu(\tau). \end{aligned} \tag{7.3}$$

In these equations r^σ are four unknowns, b^s and β_k are arbitrary constants, and all the other quantities are known functions of τ. In virtue of Theorem 6.2, Eqs. (7.3) admit a solution of the form

$$r^\alpha = \rho^\alpha(\tau, b^k, \beta_l),$$

which satisfies Eqs. (6.15). The six arbitrary constants b^k, β_l can always be expressed in terms of the initial values of r^α and $Dr^\sigma/d\tau$ for $\tau = \tau_0$. Doing this, one must, however, take into account that the second of Eqs. (7.3) and its $D/d\tau$ derivative provide us with two constraint conditions on r^α and $Dr^\alpha/d\tau$.

Thus the algorithm of solving Eqs. (6.15) formulated now requires only computing derivatives and solving a system of linear algebraic equations, and does not assume any knowledge of the explicit form of the connection and curvature, in opposition to when one was solving Eqs. (6.15) directly.

In the special case when the two neighboring geodesics are parametrized by the proper time s, the basic geodesic (7.1) must be selected from the family (4.17) which is determined by solving the system of equations (4.18). The equations of geodesic deviation can then be obtained from Eqs. (6.15) by assuming that λ and μ in these equations accept constant values (cf. Ref. 6). The respective modification of the algorithm of finding the geodesic deviation reduces then in this special case to solving Eqs. (7.3) with μ being a constant (or being even zero if the geodesic deviation vector is to be orthogonal to the vector $d\xi^\alpha/ds$). Details of the algorithm for the proper time parametrization, although without proofs, were already reported by the author[12] some time ago.

An application of the procedure derived here will be published in a subsequent paper.[13]

ACKNOWLEDGMENT

This work has been partly supported by the Polish Research Project CPBP 01.03.

APPENDIX A: AN ALTERNATE VERSION OF THE FORMALISM

As is known, the equations of geodesics can be derived not only from a variational principle based upon the functional (3.1), but also from that starting with the action[7]

$$W[\gamma] = \frac{1}{2}\int_{\tau_0}^{\tau_1} g_{\alpha\beta}\dot{\xi}^\alpha\dot{\xi}^\beta\, d\tau. \tag{A1}$$

There is, however, an important difference between the two actions, (3.1) and (A1). The first is invariant under reparametrizations defined by arbitrary monotonous and differentiable functions: $\tau' = f(\tau), f' \neq 0$, whereas the second admits only adding a constant ϵ to the parameter $\tau' = \tau + \epsilon$. In other words, the group of invariance of (3.1) is formed by all the diffeomorphisms of the one-dimensional manifold represented by the geodesic line treated as a set of points, while that of (A1) is a one-parametric Abelian Lie group acting along a geodesic endowed with an affine parameter. Since in accordance with the theory of relativity it is the world line understood as a locus of points that is physically significant, and not a world line as a parametrized curve with a selected special parametrization along it, the action (3.1) is in my opinion physically preferable to that given by (A1). True enough, the action (A1) leads to both null and non-null geodesics, while the homogeneous action (3.1) leads to the non-null ones only. I do not, however, consider this as a significant property, for first, it is not too difficult to formulate a homogeneous action enhancing all geodesics,[14] and second, even with the action (A1) the null case still shows some singular features.

Nevertheless the action (A1), perhaps due to its similarity to the Newtonian action, is used quite commonly (cf. Ref. 8). It seems therefore to be useful to demonstrate how the algorithm formulated in the paper can be applied for this other case as well.

In the notation of Sec. III, the complete variation of the action (A1) is

$$\delta W = (g_{\mu\nu}\dot{\xi}^{\mu}\delta\xi^{\nu} - \tfrac{1}{2}g_{\mu\nu}\dot{\xi}^{\mu}\dot{\xi}^{\nu}\delta\tau)\big|_{\tau_0}^{\tau_1} - \int_{\tau_0}^{\tau_1}\frac{D}{d\tau}(g_{\alpha\rho}\dot{\xi}^{\rho})\bar{\delta}\xi^{\alpha}\,d\tau,$$

and the differential of the principal function $W(x^\alpha,\tau)$ determined by (A1) reads as

$$dW = g_{\mu\nu}\dot{\xi}^{\nu}\,dx^{\mu} - \tfrac{1}{2}g_{\mu\nu}\dot{\xi}^{\mu}\dot{\xi}^{\nu}\,d\tau.$$

Thus

$$\frac{\partial W}{\partial x^\alpha} = g_{\alpha\beta}\dot{\xi}^{\beta}, \quad \frac{\partial W}{\partial\tau} = -\frac{1}{2}g_{\mu\nu}\dot{\xi}^{\mu}\dot{\xi}^{\nu},$$

and the Hamilton–Jacobi equation has the form

$$\frac{\partial W}{\partial\tau} = -\frac{1}{2}g^{\alpha\beta}\frac{\partial W}{\partial x^\alpha}\frac{\partial W}{\partial x^\beta}. \tag{A2}$$

This is an equation of exactly the same form as Eq. (2.8) (with τ now playing the role of the absolute time) and can with the help of Theorem 1.1 determine a solution of the geodesic equations, provided it admits a complete integral $W(x^\alpha, a^\beta, \tau)$ that depends on four parameters a^β and satisfies the condition

$$\det\left(\frac{\partial^2 W}{\partial x^\alpha\,\partial a^\beta}\right) \neq 0. \tag{A3}$$

Since the functions $g^{\alpha\beta}$ do not depend explicitly on τ, one can expect the complete integral of (A2) to be of the form

$$W(x^\alpha, a^\beta, \tau) = -h\tau + W_0(x^\alpha, a^k, h), \tag{A4}$$

where $a^0 = h$ is a constant. The function W_0 must then satisfy the equation

$$g^{\alpha\beta}\frac{\partial W_0}{\partial x^\alpha}\frac{\partial W_0}{\partial x^\beta} = h, \tag{A5}$$

which for $h \neq 0$ and $W_0 = |h|^{1/2}U(x^\alpha, a^k)$ reduces itself to Eq. (3.7). Thus in the non-null case the complete integrals $W(x^\alpha, a^\beta, \tau)$ and $U(x^\alpha, a^k)$ of Eqs. (A2) and (3.7), respectively, are mutually related,

$$W(x^\alpha, a^\beta, \tau) = -a^0\tau + |a^0|^{1/2}U(x^\alpha, a^k). \tag{A6}$$

As a result, every complete integral of Eqs. (3.7) determines a complete integral of Eqs. (A2), and also the other way around. (In particular, for $a^0 = 1$, $\tau = s$.) Moreover, the equations

$$\frac{\partial W}{\partial a^\alpha} = \tilde{\alpha}_\alpha, \tag{A7}$$

which due to Theorem 1.1 determine the motion, reduce themselves to Eqs. (4.18) after the parameters $\tilde{\alpha}_k$, s, and s_0 are set to be

$$\tilde{\alpha}_k = \alpha_k|a^0|^{-1/2}, \quad s = 2|a^0|^{-1/2}\tau, \quad s_0 = -2\tilde{\alpha}_0|a^0|^{1/2}, \tag{A8}$$

and this in principle shows the equivalence of the two approaches of finding solutions to the equations of geodesics.

The separation of the "absolute time" τ done in (A6) produces a decomposition of the space $M\times\mathbb{R}$, being the domain of the function W, into the real line $\mathbb{R}$ (the space of the parameter τ) and the space-time M being the domain of U. In other words, in the present formalism a relativistic world line plays an analogous role to a Newtonian trajectory, with the only difference that now the "space of trajectories" is the four-dimensional space-time M.

The null case, excluded from the consideration, requires a separate discussion that will not be pursued here.

The action leading simultaneously to the geodesic and the geodesic deviation equations in the formalism considered in this Appendix can be derived, in accordance with the general rule,[15] from the action (A1). It is equal to

$$\mathcal{S}[\gamma, r] = \int_{\tau_0}^{\tau_1} g_{\alpha\beta}\dot{\xi}^{\alpha}\frac{Dr^\beta}{d\tau}\,d\tau, \tag{A9}$$

and replaces in the present approach the action (5.1) with the important difference that (5.1) was invariant both under arbitrary reparametrizations, $\tau' = f(\tau)$, $f' \neq 0$, and under changes of gauge of the form,[16] $r'^\alpha = r^\alpha + \mathcal{H}(\tau)\dot{\xi}^\alpha$, where f and $\mathcal{H}$ were arbitrary functions, while (A9) is invariant only under translations of the parameter $\tau' = \tau + \epsilon$, $\epsilon = \text{const}$, and under transformations of the form $r'^\alpha = r^\alpha + \lambda u^\alpha$, $\lambda = \text{const}$.

In the notation of Sec. V, the complete variation of the action (A9) is

$$\begin{aligned}\delta\mathcal{S} &= \left(g_{\alpha\beta}\frac{Dr^\beta}{d\tau} + \Gamma^{\beta}{}_{\alpha\gamma}r^\gamma\dot{\xi}_\beta\right)\delta\xi^\alpha\Big|_{\tau_0}^{\tau_1} + g_{\alpha\beta}\dot{\xi}^\beta\,\delta r^\alpha\Big|_{\tau_0}^{\tau_1} + g_{\alpha\beta}\dot{\xi}^\alpha\frac{Dr^\beta}{d\tau}\delta\tau\Big|_{\tau_0}^{\tau_1} - \int_{\tau_0}^{\tau_1}d\tau \\ &\times\left[\left(g_{\alpha\beta}\frac{D^2r^\beta}{d\tau^2} + R_{\alpha\beta\gamma\delta}\dot{\xi}^\beta r^\gamma\dot{\xi}^\delta + g_{\rho\sigma}\times\frac{Du^\rho}{d\tau}\Gamma^{\sigma}{}_{\mu\alpha}r^\mu\right)\bar{\delta}\xi^\alpha + g_{\alpha\beta}\frac{Du^\alpha}{d\tau}\bar{\delta}r^\beta\right].\end{aligned} \tag{A10}$$

Since due to (A10) the differential of the principal function $\mathcal{S}(x^\alpha, r^\beta, \tau)$ determined by (A9) is of the form

$$\begin{aligned}d\mathcal{S} &= \left(g_{\alpha\beta}\frac{Dr^\beta}{d\tau} + \Gamma^{\beta}{}_{\alpha\gamma}r^\gamma\dot{\xi}_\beta\right)d\xi^\alpha + g_{\alpha\beta}\dot{\xi}^\beta\,dr^\alpha - g_{\alpha\beta}\dot{\xi}^\alpha \\ &\times\frac{Dr^\beta}{d\tau}\,d\tau,\end{aligned} \tag{A11}$$

we have

$$\frac{\partial \mathscr{S}}{\partial x^\alpha} = g_{\alpha\beta}\frac{Dr^\beta}{d\tau} + \Gamma^\beta{}_{\alpha\gamma} r^\gamma \dot{\xi}_\beta,$$
$$\frac{\partial \mathscr{S}}{\partial r^\alpha} = g_{\alpha\beta}\dot{\xi}^\beta, \tag{A12}$$
$$\frac{\partial \mathscr{S}}{\partial \tau} = -g_{\alpha\beta}\frac{Dr^\alpha}{d\tau}\xi^\beta.$$

Eliminating from these equations the variables $\dot{\xi}^\alpha$ and $Dr^\alpha/d\tau$, we obtain the Hamilton–Jacobi equation for the dynamical system determined by the action (A9)

$$\frac{\partial \mathscr{S}}{\partial \tau} = -g^{\alpha\beta}\Big(\frac{\partial \mathscr{S}}{\partial x^\alpha} - \frac{\partial \mathscr{S}}{\partial x^\rho}\Gamma^\rho{}_{\mu\alpha} r^\mu\Big)\frac{\partial \mathscr{S}}{\partial x^\beta}. \tag{A13}$$

As usual, a function $\mathscr{S}(x^\alpha, r^\beta, \tau, a^\gamma, b^\delta)$ depending on eight additional parameters a^γ and b^δ is called a complete integral of the differential equation (A13) iff

(i) it is a solution of Eq. (A13) for any values of the parameters a and b from a suitably defined domain;

(ii) the 8×8 matrix

$$M_{8\times 8} = \begin{pmatrix} \dfrac{\partial^2 \mathscr{S}}{\partial x^\alpha \partial a^\gamma} & \dfrac{\partial^2 \mathscr{S}}{\partial x^\alpha \partial b^\delta} \\ \dfrac{\partial^2 \mathscr{S}}{\partial r^\beta \partial a^\gamma} & \dfrac{\partial^2 \mathscr{S}}{\partial r^\beta \partial b^\delta} \end{pmatrix} \tag{A14}$$

is of rank 8.

The equivalence of the formalisms based on Eqs. (5.10) and (5.11) on one side and on (A13) on the other follows from the next two theorems.

Theorem A1: For every complete integral $W(x^\alpha, a^\beta, \tau)$ of Eqs. (A2) the function

$$\mathscr{S}(x^\alpha, r^\beta, a^\gamma, b^\delta, \tau) = \frac{\partial W}{\partial x^\sigma}(x^\alpha, a^\beta, \tau) r^\sigma + \frac{\partial W}{\partial a^\sigma}(x^\alpha, a^\beta, \tau) b^\sigma \tag{A15}$$

is a complete integral of Eqs. (A13).

Proof: The proof is completely analogous to that of Theorem 6.1. First, one can directly show that for the function given by (A15) the matrix (A14) is nonsingular provided the function W in (A15) satisfies the condition (A3). Second, by a simple computation one can check that the function $\mathscr{S}$ determined by (A15) satisfies Eq. (A13) iff the function W is a solution of Eq. (A2).

Theorem A2: If the function (A15) is defined in terms of a function W given by the relation (A4), then the equations

$$\frac{\partial \mathscr{S}}{\partial b^\alpha} = \tilde{\alpha}_\alpha \tag{A16}$$

are equivalent to Eqs. (A7) [and by the same to (4.18)], and

$$\frac{\partial \mathscr{S}}{\partial a^\alpha} = \tilde{\beta}_\alpha \tag{A17}$$

to Eqs. (6.16) provided that the parameters β_k and the function $\mu(\tau)$ are set in (6.16) to be

$$\beta_k = \frac{1}{|a^0|^{1/2}}\Big(\tilde{\beta}_k - \frac{\alpha_k b^0}{2a^0}\Big), \tag{A18}$$

$$\mu(s) = \frac{1}{2a^0}(s - s_0)b^0 - \alpha_s b^0 + \tilde{\beta}_0. \tag{A19}$$

Proof: The equivalence of (A16) and (A7) follows from immediately differentiating (A15) with respect to b^α. To prove the remaining part of the theorem, one must first substitute (A4) into (A15) and then differentiate it with respect to a^k and a^0. The resulting equations are the same as the equations obtained from substituting (A18) and (A19) into Eqs. (6.16), which in principle proves the equivalence of the approaches based upon Eqs. (6.16) and (A17), respectively.

Remark: It should be noted that the substitution of (A18) into Eqs. (6.16) amounts simply to a redefinition of the integration constants when passing from one of the two approaches to the other, whereas the substitution of (A19) is a restriction on the function μ, which in Eqs. (6.16) was arbitrary. It is connected with the fact that the formalism in the main body of the paper deals with geodesics parametrized arbitrarily, while the formalism in this Appendix uses affine parametrization. Therefore, before one could even talk about the equivalence of the two approaches, a restriction to a linear function (A19) was needed. This is in agreement with the result (cf. Ref. 6) that $\dot{\mu} = \text{const}$ in Eqs. (6.16) describes two neighboring geodesics parametrized by two (possibly different) affine parameters. These affine parameters are in particular the same if the integration constant $b^0 = 0$ (i.e., $\dot{\mu} = 0$). Moreover, the basic line (and for $b^0 = 0$ also the neighboring one) is parametrized by the proper time s if additionally $a^0 = 1$.

APPENDIX B: THE GRADIENT OPERATOR ON THE TANGENT BUNDLE

The Hamilton–Jacobi equations (5.11) and (A13) contain terms depending explicity on the affine connection coefficients of the space-time manifold M. At first sight, this circumstance may give rise to some doubts concerning the covariance of these equations. The connection coefficients do appear here; however, mainly due to the fact that the scalar function S (or $\mathscr{S}$, respectively) in these Hamilton–Jacobi equations depends not only on points of the manifold M but in addition also on a vector field on M. And the transformation properties of the derivatives of such a function are more involved than those of a function depending on points on M only.

To examine this question in detail, let us assume that in local coordinates our function takes the values $S(x^\alpha, r^\beta)$ and that under nonsingular coordinate transformations of the form

$$x^{\alpha'} = X^{\alpha'}(x^\alpha), \tag{B1}$$

$$r^{\alpha'} = \frac{\partial X^{\alpha'}}{\partial x^\alpha} r^\alpha, \tag{B2}$$

we have

$$S(x^{\alpha'}, r^{\beta'}) = S(x^\alpha, r^\beta). \tag{B3}$$

By differentiating this equality with respect to $x^{\mu'}$, substituting the inverse transformations to (B1) and (B2) in place of the arguments at the right-hand side of (B3), and using the chain rule, one obtains

$$\frac{\partial S}{\partial x^{\mu'}} = \frac{\partial X^\mu}{\partial x^{\mu'}}\frac{\partial S}{\partial x^\mu} + \frac{\partial X^{\nu'}}{\partial x^\sigma}\frac{\partial^2 X^\rho}{\partial x^{\mu'}\partial x^{\nu'}}\frac{\partial S}{\partial r^\rho} r^\sigma, \tag{B4}$$

where $x^\alpha = X^\alpha(x^{\beta'})$ denotes the inverse transformation to (B1). Formula (B4) gives a transformation rule of the quantities $\partial S/\partial x^\alpha$ and demonstrates that they do not form, in the case considered now, any geometric object at all. Making, however, use of the transformation rule of the connection coefficients, one can instead easily show that the quantities

$$\frac{\delta S}{\delta x^\alpha} = \frac{\partial S}{\partial x^\alpha} - \Gamma^\rho{}_{\alpha\sigma} r^\sigma \frac{\partial S}{\partial r^\rho} \tag{B5}$$

are components of a covariant vector that can be considered as a generalization of the gradient of a function of the form $S = S(x^\alpha, r^\beta)$. On the other hand, it can be immediately shown that the derivatives $\partial S/\partial r^\alpha$ of such a function are still components of a covariant vector. This elucidates the transformation properties of all the quantities entering the Hamilton–Jacobi equations (5.11) and (A13).

The derivative (B5) can be generalized to the case of all tensor valued functions of the form $T^{\alpha\cdots}{}_{\beta\ldots} = T^{\alpha\cdots}{}_{\beta\ldots}(x^\mu, r^\nu)$, where $T^{\alpha\cdots}{}_{\beta\ldots}$ are components of a tensor of any valence. By a procedure analogous to that leading to (B5), it can easily be shown that it is not the covariant derivative $T^{\alpha\cdots}{}_{\beta\ldots;\mu}$ but the expression

$$\frac{\delta T^{\alpha\cdots}{}_{\beta\cdots}}{\delta x^\mu} = T^{\alpha\cdots}{}_{\beta\cdots;\mu} - \Gamma^\rho{}_{\mu\sigma} r^\sigma \frac{\partial T^{\alpha\cdots}{}_{\beta\cdots}}{\partial r^\rho} \tag{B6}$$

that is a tensor in the case considered now.

The generalization (B5) of the gradient of a scalar function finds a nice geometric interpretation in the formalism of fiber bundles. Let TM be the tangent bundle over M with the bundle projection π: $TM \to M$. As is known, a connection on M defines n-dimensional horizontal subspaces H_q of the tangent spaces $T_q(TM)$ at all points $q \in TM$ (cf., e.g., Ref. 9, pp. 53 and 54). A one-form on $T^*_q(TM)$ is called horizontal if it vanishes on all vertical vectors, i.e., vectors from $T_q(TM)/H_q$. Let S be a differentiable function on TM. In analogy to the definition of the differential of a function on the manifold M, I propose to call a one-form on TM a horizontal differential $d_h S$ of the function S: $TM \to \mathbb{R}$ iff (i) $d_h S$ is a horizontal one-form; (ii) for every vector $t \in H_q$

$$\langle d_h S, t\rangle = t[S], \tag{B7}$$

where the angular brackets on the left-hand side denote the dual pairing of forms from $T^*_q(TM)$ with vectors from $T_q(TM)$, and the quantity on the right is a functional value resulting from the action of a vector t on S. In terms of local coordinates $\{(x^\alpha, r^\beta)\}$ on TM, the horizontal space H_q is spanned by its basis vectors

$$h_\alpha = \frac{\partial}{\partial x^\alpha} - \Gamma^\mu{}_{\alpha\nu} r^\nu \frac{\partial}{\partial r^\mu}.$$

From here and from its definition, it follows that in the coordinate basis the horizontal differential is given by the expression

$$d_h S = \frac{\delta S}{\delta x^\alpha} dx^\alpha,$$

where the quantity $\delta S/\delta x^\alpha$ is just that defined by Eq. (B5).

[1] *C. G. J. Jacobi's Vorlesungen über Dynamik*, edited by A. Clebsch (Reimer, Berlin, 1884).

[2] L. D. Landau and E. M. Lifschitz, *The Classical Theory of Fields*, translated by M. Hamermesh, 3rd English ed. (Pergamon, Oxford, 1971).

[3] B. Carter, Phys. Rev. **174**, 1959 (1968).

[4] F. A. E. Pirani, Acta Phys. Pol. **15**, 389 (1956).

[5] E. T. Whittaker, *A Treatise on Analytic Dynamics of Particles and Rigid Bodies with an Introduction to the Problem of Three Bodies* (Cambridge U. P., Cambridge, 1952), Chap. 12, Secs. 141 and 142; J. L. Synge, "Classical dynamics," Sec. 72, in *Handbuch der Physik, Encyclopedia of Physics*, edited by F. Flügge (Springer, Berlin, 1960), Vol. III; F. Gantmacher, *Lectures in Analytical Mechanics* (Mir, Moscow, 1970), p. 106.

[6] S. L. Bażański, Ann. Inst. H. Poincaré A **27**, 145 (1977).

[7] M. v. Laue, *Die Relativitätstheorie* (Vieweg, Braunschweig, 1953), Vol. II, p. 32; W. Pauli, *Theory of Relativity* (Pergamon, London, 1958), Sec. 15; E. C. G. Stückelberg, Helv. Phys. Acta **15**, 23 (1942).

[8] C. W. Misner, K. S. Thorne, and J. A. Wheeler, *Gravitation* (Freeman, San Francisco, 1973).

[9] See, e.g., S. W. Hawking and G. F. R. Ellis, *The Large Scale Structure of Space-time* (Cambridge U. P., Cambridge, 1973), p. 106.

[10] In the notation adopted here, the dependence of the functions ξ^α on the function f will be very often suppressed.

[11] Strictly speaking the most general function Φ should be of the form $\Phi = f(\varphi(x^\alpha, a^k), a^l)$, where f is an arbitrary function such that $\partial f/\partial\varphi \neq 0$. This more general case would lead to an equation analogous to Eq. (4.3), in which $f(\tau)$ should be replaced by $f(\tau, a^k)$. For the sake of simplicity, however, the dependence of the functions f and λ on a^k is in the sequel suppressed.

[12] S. L. Bażański, in *Proceedings of the Fourth Marcel Grossman Meeting on General Relativity*, edited by R. Ruffini (Elsevier, Amsterdam, 1986), p. 1615.

[13] S. L. Bażański and P. Jaranowski, to be published in J. Math. Phys.

[14] S. L. Bażański, to be published in Acta Phys. Pol.

[15] S. L. Bażański, Acta Phys. Pol. B **7**, 305 (1976).

[16] Strictly speaking, this second transformation results in adding to the action an integral of a complete differential, which has no effect on the equations of motion.

All directing fields that are polynomial in the $(n-1)$ velocity are geodesic

Robert Alan Coleman and Herbert Korte
Interdisciplinary Studies in Mathematical Physics and Philosophy of Science, University of Regina, Regina, Saskatchewan S4S 0A2, Canada

(Received 16 August 1988; accepted for publication 30 November 1988)

It is proven that all directing fields that are polynomial in the $(n-1)$ velocities ξ_1^α are, in fact, cubic in these variables; so that they are geodesic directing fields (projective structures). This result, in conjunction with previously published work, entails that $(n-1)$ forces must be nonpolynomial in the $(n-1)$ velocities. A large class of acceleration fields that give rise to directing fields is briefly discussed in order to illustrate the source of this nonpolynomial behavior.

I. ACCELERATION AND DIRECTING FIELDS

Let M be an n-dimensional C^∞ manifold. A *curve* element of order k at $p\in M$ is an equivalence class $j_0^k\gamma$ of curves through p that have the same Taylor expansion with respect to some (and hence every) coordinate chart $(U,x)_p$ up to and including order k at $0\in\mathbb{R}$. A *path element* of order k at $p\in M$ is an equivalence class of paths $j_p^k\xi$ consisting of all paths corresponding to curves in $j_0^k\gamma$, where $\gamma\in\xi$.

A second-order curve element $j_0^2\gamma$ has local coordinates γ_1^i and γ_2^i, called n velocity and n acceleration, respectively, and given by

$$\gamma_1^i=\frac{d}{d\lambda}x^i\circ\gamma(0),\quad \gamma_2^i=\frac{d^2}{d\lambda^2}x^i\circ\gamma(0). \tag{1}$$

A second-order path element $j_p^2\xi$ has local coordinates ξ_1^α and ξ_2^α, called $(n-1)$ velocity and $(n-1)$ acceleration, respectively, and given by

$$\xi_1^\alpha=\frac{dx^\alpha\circ\gamma}{dx^n\circ\gamma}\bigg|_p,\quad \xi_2^\alpha=\frac{d^2x^\alpha\circ\gamma}{(dx^n\circ\gamma)^2}\bigg|_p. \tag{2}$$

Under a change of coordinate chart from $(U,x)_p$ to $(\overline{U},\overline{x})_p$, the coordinates of $j_0^2\gamma$ transform according to

$$\overline{\gamma}_1^i=\overline{X}_j^i\gamma_1^j,\quad \overline{\gamma}_2^i=\overline{X}_j^i\gamma_2^j+\overline{X}_{jk}^i\gamma_1^j\gamma_1^k, \tag{3}$$

and the coordinates of $j_p^2\xi$ transform according to

$$\overline{\xi}_1^\alpha=(\overline{X}_n^\alpha+\overline{X}_\beta^\alpha\xi_1^\beta)/(\overline{X}_n^n+\overline{X}_\gamma^n\xi_1^\gamma) \tag{4}$$

and

$$\begin{aligned}\overline{\xi}_2^\alpha=&(\overline{X}_\rho^\alpha\xi_2^\rho+\overline{X}_{\rho\sigma}^\alpha\xi_1^\rho\xi_1^\sigma+2\overline{X}_{n\rho}^\alpha\xi_1^\rho\\&+\overline{X}_{nn}^\alpha)/(\overline{X}_n^n+\overline{X}_\gamma^n\xi_1^\gamma)^2\\&-\overline{\xi}_1^\alpha(\overline{X}_\rho^n\xi_2^\rho+\overline{X}_{\rho\sigma}^n\xi_1^\rho\xi_1^\sigma+2\overline{X}_{n\rho}^n\xi_1^\rho\\&+\overline{X}_{nn}^n)/(\overline{X}_n^n+\overline{X}_\gamma^n\xi_1^\gamma)^2,\end{aligned} \tag{5}$$

where $\overline{X}=\overline{x}\circ x^{-1}$.

Denote by $\mathscr{L}_1^1(M)$ and $\mathscr{L}_1^2(M)$ the bundles of first- and second-order curve elements and by $\mathscr{D}^1(M)$ and $\mathscr{D}^2(M)$ the bundles of first- and second-order path elements. In each case, the bundle of second-order elements can be regarded as a bundle over the corresponding bundle of first-order elements.

An acceleration field is a cross section $A:L_1^1(M)\to L_1^2(M)$. Such a field is described in terms of local coordinates by functions $A_2^i(x^i,\gamma_1^i)$ that transform under a change of coordinates according to

$$\overline{A}_2^i(\overline{x}^i,\overline{\gamma}_1^i)=\overline{X}_j^i(x^i)A_2^j(x^i,\gamma_1^i)+\overline{X}_{jk}^i(x^i)\gamma_1^j\gamma_1^k. \tag{6}$$

An acceleration field is called geodesic iff for every $p\in M$, there is some coordinate chart $(\overline{U},\overline{x})_p$ such that the functions $\overline{A}_2^i(\overline{x}^i,\overline{\gamma}_1^i)$ vanish at p. A geodesic acceleration field is denoted by Γ and has the special functional form

$$\Gamma_2^i(x^i,\gamma_1^i)=-\Gamma_{jk}^i(x^i)\gamma_1^j\gamma_1^k. \tag{7}$$

A directing field is a cross section Ξ: $\mathbf{D}^1(M)\to\mathbf{D}^2(M)$. Such a field is described in terms of local coordinates by functions $\Xi_2^\alpha(x^i,\xi_1^\alpha)$ that transform under a change of coordinate chart according to

$$\begin{aligned}&\overline{\Xi}_2^\alpha(\overline{x}^i,\overline{\xi}_1^\alpha)\\&=[\overline{X}_\rho^\alpha(x^i)\Xi_2^\rho(x^i,\xi_1^\alpha)+\overline{X}_{\rho\sigma}^\alpha(x^i)\xi_1^\rho\xi_1^\sigma\\&\quad+2\overline{X}_{n\rho}^\alpha(x^i)\xi_1^\rho+\overline{X}_{nn}^\alpha(x^i)]/(\overline{X}_n^n(x^i)\\&\quad+\overline{X}_\gamma^n(x^i)\xi_1^\gamma)^2-\overline{\xi}_1^\alpha[\overline{X}_\rho^n(x^i)\Xi_2^\rho(x^i,\xi_1^\alpha)\\&\quad+\overline{X}_{\rho\sigma}^n(x^i)\xi_1^\rho\xi_1^\sigma+2\overline{X}_{n\rho}^n(x^i)\xi_1^\rho\\&\quad+\overline{X}_{nn}^n(x^i)]/(\overline{X}_n^n(x^i)+\overline{X}_\gamma^n(x^i)\xi_1^\gamma)^2.\end{aligned} \tag{8}$$

A directing field is called goedesic iff for every $p\in M$, there is some coordinate chart $(\overline{U},\overline{x})_p$ such that the functions $\overline{\Xi}_2^\alpha(\overline{x}^i,\overline{\xi}_1^\alpha)$ vanish at p. A geodesic acceleration field is denoted by Π and has the special functional form

$$\begin{aligned}&\Pi_2^\alpha(x^i,\xi_1^\alpha)\\&=\xi_1^\alpha(\Pi_{\rho\sigma}^n(x^i)\xi_1^\rho\xi_1^\sigma+2\Pi_{n\rho}^n(x^i)\xi_1^\rho+\Pi_{nn}^n(x^i))\\&\quad-(\Pi_{\rho\sigma}^\alpha(x^i)\xi_1^\rho\xi_1^\sigma+2\Pi_{n\rho}^\alpha(x^i)\xi_1^\rho+\Pi_{nn}^\alpha(x^i)),\end{aligned} \tag{9}$$

where the projective coefficients $\Pi_{jk}^i(x^i)$ are traceless so that $\Pi_{nn}^n(x^i)$ and $\Pi_{n\rho}^n(x^i)$ may be elminated from (9).

An acceleration field A determines a directing field Ξ iff A is of the form (Ref. 1, see Theorem 3.1)

$$A_2^i(x^i,\gamma_1^i)=B(x^i,\gamma_1^i)\gamma_1^i+C^i(x^i,\gamma_1^i), \tag{10}$$

where

$$C^i(x^i,\lambda\gamma_1^i)=\lambda^2C^i(x^i,\gamma_1^i). \tag{11}$$

 0022-2488/89/051030-04$02.50

II. GEODESICITY OF DIRECTING FIELDS THAT ARE POLYNOMIAL IN THE $(n-1)$ VELOCITIES

The theorem proved in this section was conjectured by Ehlers and Köhler (Ref. 2, see Theorem 1). It is also an extension of a theorem proved by us (Ref. 3, see Theorem 3), which states that a directing field $\Xi^{\alpha}_{2}(x^i,\xi^{\alpha}_{1})$ that is cubic in the $(n-1)$ velocities ξ^{α}_{1} is geodesic.

Theorem 1: If $n>1$, a directing field $\Xi^{\alpha}_{2}(x^i,\xi^{\alpha}_{1})$ that is polynominal in the $(n-1)$ velocities ξ^{α}_{1} with respect to all coordinate charts is, in fact, cubic in the $(n-1)$ velocities ξ^{α}_{1}, and hence is geodesic.

Proof: A directing field Ξ that is polynomial in the $(n-1)$ velocities ξ^{α}_{1} with respect to coordinate chart $(U,x)_p$ for a neighborhood of $p\in M$ has the form

$$\begin{aligned}\Xi^{\alpha}_{2}(x^i,\xi^{\alpha}_{1}) &= A^{\alpha}(x^i) + A^{\alpha}_{\beta}(x^i)\xi^{\beta}_{1} + A^{\alpha}_{\beta_1\beta_2}(x^i)\xi^{\beta_1}_{1}\xi^{\beta_2}_{1}\\ &\quad + A^{\alpha}_{\beta_1\beta_2\beta_3}(x^i)\xi^{\beta_1}_{1}\xi^{\beta_2}_{1}\xi^{\beta_3}_{1} + \cdots\\ &\quad + A^{\alpha}_{\beta_1\cdots\beta_r}(x^i)\xi^{\beta_1}_{1}\cdots\xi^{\beta_r}_{1}.\end{aligned} \tag{12}$$

Since the argument proceeds on a pointwise basis, the dependence on the variables x^i may be *suppressed* and the coefficients $(A^{\alpha},A^{\alpha}_{\beta_1},\ldots,A^{\alpha}_{\beta_1\cdots\beta_r})$ of (12) may be regarded as given *fixed* constants, which determine the directing field with respect to the coordinate chart $(U,x)_p$ at $p\in M$. To find the functions $\overline{\Xi}^{\alpha}_{2}(\bar\xi^{\alpha}_{1})$ that describe the directing field with respect to some other coordinate chart $(\overline{U},\bar x)_p$ substitute (12) into (8) and reexpress ξ^{α}_{1} in terms of $\bar\xi^{\alpha}_{1}$ using the inverse of (4), namely,

$$\xi^{\alpha}_{1} = (X^{\alpha}_{n} + X^{\alpha}_{\beta}\bar\xi^{\beta}_{1})/(X^{n}_{n} + X^{n}_{\gamma}\bar\xi^{\gamma}_{1}) \tag{13}$$

and

$$X^{n}_{n} + X^{n}_{\gamma}\bar\xi^{\gamma}_{1} = (\overline{X}^{n}_{n} + \overline{X}^{n}_{\gamma}\xi^{\gamma}_{1})^{-1}. \tag{14}$$

The factors $(\overline{X}^{n}_{n} + \overline{X}^{n}_{\gamma}\xi^{\gamma}_{1})^{-2}$ in (5) clearly yield a common factor $(X^{n}_{n} + X^{n}_{\gamma}\bar\xi^{\gamma}_{1})^{2}$ in the result; consequently, the only conceivable nonpolynomial terms in the result arise from the terms in the expression (12) for $\Xi^{\alpha}_{2}(\xi^{\alpha}_{1})$ that are of degree 3 or more in ξ^{α}_{1}. One obtains the result

$$\begin{aligned}\overline{\Xi}^{\alpha}_{2}(\bar\xi^{\alpha}_{1}) &= (\overline{X}^{\alpha}_{\rho_1\rho_2} - \bar\xi^{\alpha}_{1}\overline{X}^{n}_{\rho_1\rho_2})(X^{\rho_1}_{n} + X^{\rho_1}_{\sigma_1}\bar\xi^{\sigma_1}_{1})(X^{\rho_2}_{n} + X^{\rho_2}_{\sigma_2}\bar\xi^{\sigma_2}_{1})\\ &\quad + 2(\overline{X}^{\alpha}_{n\rho} - \bar\xi^{\alpha}_{1}\overline{X}^{n}_{n\rho})(X^{n}_{n} + X^{n}_{\sigma_1}\bar\xi^{\sigma_1}_{1})(X^{\rho}_{n} + X^{\rho}_{\sigma_2}\bar\xi^{\sigma_2}_{1})\\ &\quad + (\overline{X}^{\alpha}_{nn} - \bar\xi^{\alpha}_{1}\overline{X}^{n}_{nn})(X^{n}_{n} + X^{n}_{\sigma_1}\bar\xi^{\sigma_1}_{1})(X^{n}_{n} + X^{n}_{\sigma_2}\bar\xi^{\sigma_2}_{1})\\ &\quad + (\overline{X}^{\alpha}_{\beta} - \bar\xi^{\alpha}_{1}\overline{X}^{n}_{\beta})\Big[A^{\beta}(X^{n}_{n} + X^{n}_{\sigma_1}\bar\xi^{\sigma_1}_{1})(X^{n}_{n} + X^{n}_{\sigma_2}\bar\xi^{\sigma_2}_{1})\\ &\quad + A^{\beta}_{\rho}(X^{\rho}_{n} + X^{\rho}_{\sigma_1}\bar\xi^{\sigma_1}_{1})(X^{n}_{n} + X^{n}_{\sigma_2}\bar\xi^{\sigma_2}_{1})\\ &\quad + A^{\beta}_{\rho_1\rho_2}(X^{\rho_1}_{n} + X^{\rho_1}_{\sigma_1}\bar\xi^{\sigma_1}_{1})(X^{\rho_2}_{n} + X^{\rho_2}_{\sigma_2}\bar\xi^{\sigma_2}_{1})\\ &\quad + A^{\beta}_{\rho_1\rho_2\rho_3}\frac{(X^{\rho_1}_{n} + X^{\rho_1}_{\sigma_1}\bar\xi^{\sigma_1}_{1})(X^{\rho_2}_{n} + X^{\rho_2}_{\sigma_2}\bar\xi^{\sigma_2}_{1})(X^{\rho_3}_{n} + X^{\rho_3}_{\sigma_3}\bar\xi^{\sigma_3}_{1})}{(X^{n}_{n} + X^{n}_{\gamma}\bar\xi^{\gamma}_{1})}\\ &\quad + \cdots\\ &\quad + A^{\beta}_{\rho_1\cdots\rho_r}\frac{(X^{\rho_1}_{n} + X^{\rho_1}_{\sigma_1}\bar\xi^{\sigma_1}_{1})\cdots(X^{\rho_r}_{n} + X^{\rho_r}_{\sigma_r}\bar\xi^{\sigma_r}_{1})}{(X^{n}_{n} + X^{n}_{\gamma}\bar\xi^{\gamma}_{1})^{r-2}}\Big].\end{aligned} \tag{15}$$

By hypothesis, the field $\overline{\Xi}^{\alpha}_{2}(\bar\xi^{\alpha}_{1})$ must be a polynomial in the variables $\bar\xi^{\gamma}_{1}$ for all possible coordinate choices. For a given fixed point $p\in M$, the coefficients $(A^{\alpha},A^{\alpha}_{\beta_1},\ldots,A^{\alpha}_{\beta_1\cdots\beta_r})$ are constants, but the quantities $\bar\xi^{\gamma}_{1},X^{i}_{j}$, and X^{i}_{jk} are variables that may be freely chosen. By bringing the last $(r-2)$ terms to a common denominator, one obtains the constraint

$$\begin{aligned}&(X^{n}_{n} + X^{n}_{\gamma}\bar\xi^{\gamma}_{1})^{r-2}(B^{\alpha} + B^{\alpha}_{\rho}\bar\xi^{\rho}_{1} + B^{\alpha}_{\rho_1\rho_2}\bar\xi^{\rho_1}_{1}\bar\xi^{\rho_2}_{1} + B^{\alpha}_{\rho_1\rho_2\rho_3}\bar\xi^{\rho_1}_{1}\bar\xi^{\rho_2}_{1}\bar\xi^{\rho_3}_{1})\\ &\quad = (\overline{X}^{\alpha}_{\beta} - \bar\xi^{\alpha}_{1}\overline{X}^{n}_{\beta})\big[A^{\beta}_{\rho_1\cdots\rho_3}(X^{\rho_1}_{n} + X^{\rho_1}_{\sigma_1}\bar\xi^{\sigma_1}_{1})\cdots(X^{\rho_3}_{n} + X^{\rho_3}_{\sigma_3}\bar\xi^{\sigma_3}_{1})(X^{n}_{n} + X^{n}_{\gamma}\bar\xi^{\gamma}_{1})^{r-3}\\ &\qquad + A^{\beta}_{\rho_1\cdots\rho_4}(X^{\rho_1}_{n} + X^{\rho_1}_{\sigma_1}\bar\xi^{\sigma_1}_{1})\cdots(X^{\rho_4}_{n} + X^{\rho_4}_{\sigma_4}\bar\xi^{\sigma_4}_{1})(X^{n}_{n} + X^{n}_{\gamma}\bar\xi^{\gamma}_{1})^{r-4}\\ &\qquad + \cdots\\ &\qquad + A^{\beta}_{\rho_1\cdots\rho_r}(X^{\rho_1}_{n} + X^{\rho_1}_{\sigma_1}\bar\xi^{\sigma_1}_{1})\cdots(X^{\rho_r}_{n} + X^{\rho_r}_{\sigma_r}\bar\xi^{\sigma_r}_{1})\big].\end{aligned} \tag{16}$$

The terms of degree $(r+1)$ yield

$$\begin{aligned}&(X^{n}_{\gamma}\bar\xi^{\gamma}_{1})^{r-2}B^{\alpha}_{\rho_1\rho_2\rho_3}\bar\xi^{\rho_1}_{1}\bar\xi^{\rho_2}_{1}\bar\xi^{\rho_3}_{1}\\ &\quad = -\bar\xi^{\alpha}_{1}\overline{X}^{n}_{\beta}\big[A^{\beta}_{\rho_1\cdots\rho_3}X^{\rho_1}_{\sigma_1}\bar\xi^{\sigma_1}_{1}\cdots X^{\rho_3}_{\sigma_3}\bar\xi^{\sigma_3}_{1}(X^{n}_{\gamma}\bar\xi^{\gamma}_{1})^{r-3}\\ &\qquad + A^{\beta}_{\rho_1\cdots\rho_4}X^{\rho_1}_{\sigma_1}\bar\xi^{\sigma_1}_{1}\cdots X^{\rho_4}_{\sigma_4}\bar\xi^{\sigma_4}_{1}(X^{n}_{\gamma}\bar\xi^{\gamma}_{1})^{r-4}\\ &\qquad + \cdots\\ &\qquad + A^{\beta}_{\rho_1\cdots\rho_{r-1}}X^{\rho_1}_{\sigma_1}\bar\xi^{\sigma_1}_{1}\cdots X^{\rho_{r-1}}_{\sigma_{r-1}}\bar\xi^{\sigma_{r-1}}_{1}X^{n}_{\gamma}\bar\xi^{\gamma}_{1}\\ &\qquad + A^{\beta}_{\rho_1\cdots\rho_r}X^{\rho_1}_{\sigma_1}\bar\xi^{\sigma_1}_{1}\cdots X^{\rho_r}_{\sigma_r}\bar\xi^{\sigma_r}_{1}\big].\end{aligned} \tag{17}$$

The right-hand side of this equation is proportional to $\bar\xi^{\alpha}_{1}$. The left-hand side is proportional to $\bar\xi^{\alpha}_{1}$ iff

$$B^{\alpha}_{\rho_1\rho_2\rho_3} = \tfrac{1}{3}(\delta^{\alpha}_{\rho_1}C_{\rho_2\rho_3} + \delta^{\alpha}_{\rho_2}C_{\rho_3\rho_1} + \delta^{\alpha}_{\rho_3}C_{\rho_1\rho_2}). \tag{18}$$

Substitution of (18) into (17) yields

$$\begin{aligned}&(X^{n}_{\gamma}\bar\xi^{\gamma}_{1})^{r-2}C_{\rho_1\rho_2}\bar\xi^{\rho_1}_{1}\bar\xi^{\rho_2}_{1}\\ &\quad = -\overline{X}^{n}_{\beta}\big[A^{\beta}_{\rho_1\cdots\rho_3}X^{\rho_1}_{\sigma_1}\bar\xi^{\sigma_1}_{1}\cdots X^{\rho_3}_{\sigma_3}\bar\xi^{\sigma_3}_{1}(X^{n}_{\gamma}\bar\xi^{\gamma}_{1})^{r-3}\end{aligned}$$

$$+A^{\beta}_{\rho_1\cdots\rho_4}X^{\rho_1}_{\sigma_1}\bar\xi^{\sigma_1}_{1}\cdots X^{\rho_4}_{\sigma_4}\bar\xi^{\sigma_4}_{1}(X^{n}_{\gamma}\bar\xi^{\gamma}_{1})^{r-4}$$
$$+\cdots$$
$$+A^{\beta}_{\rho_1\cdots\rho_{r-1}}X^{\rho_1}_{\sigma_1}\bar\xi^{\sigma_1}_{1}\cdots X^{\rho_{r-1}}_{\sigma_{r-1}}\bar\xi^{\sigma_{r-1}}_{1}X^{n}_{\gamma}\bar\xi^{\gamma}_{1}$$
$$+A^{\beta}_{\rho_1\cdots\rho_r}X^{\rho_1}_{\sigma_1}\bar\xi^{\sigma_1}_{1}\cdots X^{\rho_r}_{\sigma_r}\bar\xi^{\sigma_r}_{1}]. \tag{19}$$

Each term in this equation except the last one explicitly contains the factor $X^{n}_{\gamma}\bar\xi^{\gamma}_{1}$. Without restricting the variables X^{i}_{j} in any way, choose $\bar\xi^{\alpha}_{1}$ so that $X^{n}_{\gamma}\bar\xi^{\gamma}_{1}=0$. Note that such a choice restricts the orientation but not the magnitude of the "vector" $\bar\xi^{\alpha}_{1}$. Since the X^{α}_{β} are arbitrary, the variables $\zeta^{\alpha}\equiv X^{\alpha}_{\beta}\bar\xi^{\beta}_{1}$ may still be freely chosen. With this choice, one obtains from (19) the constraint

$$\bar X^{n}_{\beta}A^{\beta}_{\rho_1\cdots\rho_r}X^{\rho_1}_{\sigma_1}\bar\xi^{\sigma_1}_{1}\cdots X^{\rho_r}_{\sigma_r}\bar\xi^{\sigma_r}_{1}=0. \tag{20}$$

If $A^{\beta}_{\rho_1\cdots\rho_r}$ has the form

$$A^{\beta}_{\rho_1\cdots\rho_r}=(1/r)(\delta^{\beta}_{\rho_1}E_{\rho_2\cdots\rho_{r-1}\rho_r}$$
$$+\delta^{\beta}_{\rho_r}E_{\rho_1\cdots\rho_{r-2}\rho_{r-1}}+\cdots+\delta^{\beta}_{\rho_2}E_{\rho_3\cdots\rho_r\rho_1}), \tag{21}$$

then the left side of (20) vanishes because it too contains the factor $X^{n}_{\gamma}\bar\xi^{\gamma}_{1}$ since

$$\bar X^{n}_{\beta}X^{\beta}_{\gamma}+\bar X^{n}_{n}X^{n}_{\gamma}=0. \tag{22}$$

If $A^{\beta}_{\rho_1\cdots\rho_r}$ does not have the form (21), then the factor $X^{n}_{\gamma}\bar\xi^{\gamma}_{1}$ is not present. Since the variables $\zeta^{\alpha}\equiv X^{\alpha}_{\beta}\bar\xi^{\beta}_{1}$ may be freely chosen, it follows from (20) that

$$\bar X^{n}_{\beta}A^{\beta}_{\rho_1\cdots\rho_r}=0. \tag{23}$$

But, the $\bar X^{n}_{\beta}$ may also be freely chosen. It follows that

$$A^{\beta}_{\rho_1\cdots\rho_r}=0. \tag{24}$$

If $A^{\beta}_{\rho_1\cdots\rho_r}$ has the form (21), then *every* term of (19) contains the factor $X^{n}_{\gamma}\bar\xi^{\gamma}_{1}$, which could, therefore, have been divided out at the start to yield the constraint

$$(X^{n}_{\gamma}\bar\xi^{\gamma}_{1})^{r-3}C_{\rho_1\rho_2}\bar\xi^{\rho_1}_{1}\bar\xi^{\rho_2}_{1}$$
$$=-\bar X^{n}_{\beta}[A^{\beta}_{\rho_1\cdots\rho_3}X^{\rho_1}_{\sigma_1}\bar\xi^{\sigma_1}_{1}\cdots X^{\rho_3}_{\sigma_3}\bar\xi^{\sigma_3}_{1}(X^{n}_{\gamma}\bar\xi^{\gamma}_{1})^{r-4}$$
$$+A^{\beta}_{\rho_1\cdots\rho_4}X^{\rho_1}_{\sigma_1}\bar\xi^{\sigma_1}_{1}\cdots X^{\rho_4}_{\sigma_4}\bar\xi^{\sigma_4}_{1}(X^{n}_{\gamma}\bar\xi^{\gamma}_{1})^{r-5}$$
$$+\cdots$$
$$+A^{\beta}_{\rho_1\cdots\rho_{r-1}}X^{\rho_1}_{\sigma_1}\bar\xi^{\sigma_1}_{1}\cdots X^{\rho_{r-1}}_{\sigma_{r-1}}\bar\xi^{\sigma_{r-1}}_{1}]$$
$$+\bar X^{n}_{n}E_{\rho_1\cdots\rho_{r-1}}X^{\rho_1}_{\sigma_1}\bar\xi^{\sigma_1}_{1}\cdots X^{\rho_{r-1}}_{\sigma_{r-1}}\bar\xi^{\sigma_{r-1}}_{1}=0. \tag{25}$$

Every term of this equation except the last two explicitly contains the factor $X^{n}_{\gamma}\bar\xi^{\gamma}_{1}$. If $\bar\xi^{\gamma}_{1}$ is chosen as before to make $X^{n}_{\gamma}\bar\xi^{\gamma}_{1}$ vanish, one obtains the constraint

$$-\bar X^{n}_{\beta}A^{\beta}_{\rho_1\cdots\rho_{r-1}}X^{\rho_1}_{\sigma_1}\bar\xi^{\sigma_1}_{1}\cdots X^{\rho_{r-1}}_{\sigma_{r-1}}\bar\xi^{\sigma_{r-1}}_{1}$$
$$+\bar X^{n}_{n}E_{\rho_1\cdots\rho_{r-1}}X^{\rho_1}_{\sigma_1}\bar\xi^{\sigma_1}_{1}\cdots X^{\rho_{r-1}}_{\sigma_{r-1}}\bar\xi^{\sigma_{r-1}}_{1}=0. \tag{26}$$

Again, there are two possibilities to consider. If $A^{\beta}_{\rho_1\cdots\rho_{r-1}}$ has the form

$$A^{\beta}_{\rho_1\cdots\rho_{r-1}}=[1/(r-1)](\delta^{\beta}_{\rho_1}E_{\rho_2\cdots\rho_{r-2}\rho_{r-1}}$$
$$+\delta^{\beta}_{\rho_{r-1}}E_{\rho_1\cdots\rho_{r-3}\rho_{r-2}}+\cdots+\delta^{\beta}_{\rho_2}E_{\rho_3\cdots\rho_{r-1}\rho_1}), \tag{27}$$

the first term of (26) will vanish because it contains a factor $X^{n}_{\gamma}\bar\xi^{\gamma}_{1}$ in which case one obtains

$$E_{\rho_1\cdots\rho_{r-1}}X^{\rho_1}_{\sigma_1}\bar\xi^{\sigma_1}_{1}\cdots X^{\rho_{r-1}}_{\sigma_{r-1}}\bar\xi^{\sigma_{r-1}}_{1}=0. \tag{28}$$

Since the variables $X^{\alpha}_{\beta}\bar\xi^{\beta}_{1}$ may be freely chosen, it follows that

$$E_{\rho_1\cdots\rho_{r-1}}=0. \tag{29}$$

If $A^{\beta}_{\rho_1\cdots\rho_{r-1}}$ does not have the form (27), then the first term of (26) will not contain a factor $X^{n}_{\gamma}\bar\xi^{\gamma}_{1}$ and since the variable $\zeta^{\alpha}\equiv X^{\alpha}_{\beta}\bar\xi^{\beta}_{1}$ may be freely chosen, one obtains

$$-\bar X^{n}_{\beta}A^{\beta}_{\rho_1\cdots\rho_{r-1}}+E_{\rho_1\cdots\rho_{r-1}}=0. \tag{30}$$

Since the variables $\bar X^{n}_{\beta}$ may also be freely chosen, it follows that $A^{\beta}_{\rho_1\cdots\rho_{r-1}}=0$ and hence that $E_{\rho_1\cdots\rho_{r-1}}=0$. In all cases, it follows that (24) holds, that is, the polynomial has degree 1 less than the degree r that was assumed.

The above argument can be repeated to show that all of the coefficients of (12) for degree greater than 3 must vanish and that the coefficients for degree 3 must have the form

$$A^{\alpha}_{\rho_1\rho_2\rho_3}=\tfrac{1}{3}(\delta^{\alpha}_{\rho_1}E_{\rho_2\rho_3}+\delta^{\alpha}_{\rho_2}E_{\rho_3\rho_1}+\delta^{\alpha}_{\rho_3}E_{\rho_1\rho_2}). \tag{31}$$

The directing field is then given by

$$\Xi^{\alpha}_{2}(\xi^{\alpha}_{1})=A^{\alpha}+A^{\alpha}_{\beta}\xi^{\beta}_{1}+A^{\alpha}_{\beta_1\beta_2}\xi^{\beta_1}_{1}\xi^{\beta_2}_{1}+\xi^{\alpha}_{1}E_{\beta_1\beta_2}\xi^{\beta_1}_{1}\xi^{\beta_2}_{1}. \tag{32}$$

As we have shown before (Ref. 3, see Theorem 3), a directing field of this form may be put into the form (9) as follows. Define $\tilde A^{\alpha}_{\rho}$ by

$$A^{\alpha}_{\rho}=2\tilde A^{\alpha}_{\rho}+\delta^{\alpha}_{\rho}A,\quad A=[1/(n+1)]A^{\alpha}_{\alpha}, \tag{33}$$

and $\tilde A^{\alpha}_{\rho_1\rho_2}$ by

$$A^{\alpha}_{\rho_1\rho_2}=2\tilde A^{\alpha}_{\rho_1\rho_2}+\delta^{\alpha}_{\rho_1}A_{\rho_2}+\delta^{\alpha}_{\rho_2}A_{\rho_1},$$
$$A_{\rho}=[1/(n+1)]A^{\alpha}_{\alpha\rho}. \tag{34}$$

One finds that

$$\tilde A^{\alpha}_{\alpha}=A,\quad\text{and}\quad\tilde A^{\alpha}_{\alpha\rho}=A_{\rho}. \tag{35}$$

The directing field (32) can be put into the form (9) by making the identifications

$$E_{\rho_1\rho_2}=\Pi^{n}_{\rho_1\rho_2},\quad\tilde A^{\alpha}_{\rho_1\rho_2}=-\Pi^{\alpha}_{\rho_1\rho_2},$$
$$\tilde A^{\alpha}_{\rho}=-\Pi^{\alpha}_{n\rho},\quad A^{\alpha}=-\Pi^{\alpha}_{nn}, \tag{36}$$

from which follow (recall that $\Pi^{i}_{ij}=0$)

$$A=-\Pi^{\alpha}_{\alpha n}=\Pi^{n}_{nn},\quad A_{\rho}=-\Pi^{\alpha}_{\alpha\rho}=\Pi^{n}_{n\rho}. \tag{37}$$

III. (n–1) FORCES ARE NONPOLYNOMIAL IN THE (n–1) VELOCITIES

In a previous paper,[4] we proved that, in the context of a conformal causal structure, (a) any acceleration field, such that its n force is orthogonal to the n velocity, uniquely decomposes into the sum of a symmetric affine structure that is compatible with the conformal structure and an n force, and (b) any directing field, such that the n force of the corresponding family of acceleration fields is due to tensor fields and is orthogonal to the n velocity, uniquely decomposes into the sum of a projective structure (geodesic directing field) that is compatible with the conformal structure and an $(n-1)$ force.

If the $(n-1)$ force were polynomial in the $(n-1)$ velocities ξ^{α}_{1} in all coordinate systems, then by the theorem proved in the previous section it would have to be cubic and have a form such that the "total" directing field is also geodesic. This total geodesic directing field must also be compatible with the conformal structure of space-time in the sense that no solution paths can "break the light barrier." The corresponding n force is then required to have the form given by Eq. (81) of Ref. 4; however, such an n force cannot satisfy the orthogonality condition (necessary for the absence of variable rest masses) unless it vanishes identically. It follows that the $(n-1)$ force must be nonpolynomial in the $(n-1)$ velocities ξ^{α}_{1}.

The source of this nonpolynomial behavior is illustrated by the large class of acceleration fields, each projective equivalence class of which is determined by the standard representative

$$\begin{aligned} A^{i}_{2}(\gamma^{i}_{1}) = {} & (g_{rs}\gamma^{r}_{1}\gamma^{s}_{1})^{1/2}T^{i}_{j_1}\gamma^{j_1}_{1} - \Pi^{i}_{j_1j_2}\gamma^{j_1}_{1}\gamma^{j_2}_{1} \\ & + (g_{rs}\gamma^{r}_{1}\gamma^{s}_{1})^{-1/2}T^{i}_{j_1j_2j_3}\gamma^{j_1}_{1}\gamma^{j_2}_{1}\gamma^{j_3}_{1} + \cdots \\ & + (g_{rs}\gamma^{r}_{1}\gamma^{s}_{1})^{-(k-2)/2}T^{i}_{j_1\cdots j_k}\gamma^{j_1}_{1}\cdots\gamma^{j_k}_{1}, \end{aligned} \tag{38}$$

where g_{jk} is the space-time metric tensor, the $\Pi^{i}_{j_1j_2}$ are the projective coefficients, and the

$$T_{ij_1\cdots j_r} = g_{ij}T^{j}_{j_1\cdots j_k} \tag{39}$$

are tensors that are antisymmetrized on the first two indices and then symmetrized on the last r indices so that

$$g_{ij}\gamma^{i}_{1}T^{j}_{j_1\cdots j_k} = 0 \tag{40}$$

and

$$A^{a}_{2,a} = 0. \tag{41}$$

For the directing field determined by (38), each of the $(n-1)$ force terms will contain a power, either negative or fractional, of $(g_{nn} + g_{n\rho}\xi^{\rho}_{1} + g_{\rho\sigma}\xi^{\rho}_{1}\xi^{\sigma}_{1})$ that is not polynomial in the $(n-1)$ velocities.

[1]R. A. Coleman and H. Korte, J. Math. Phys. **25**, 3513 (1984).
[2]J. Ehlers and E. Köhler, J. Math. Phys. **18**, 2014 (1977).
[3]R. A. Coleman and H. Korte, J. Math. Phys. **21**, 1340 (1980); **23**, 345 (E) (1982).
[4]R. A. Coleman and H. Korte, J. Math. Phys. **28**, 1492 (1987).

The metric of a space and quadratic algebras

G. P. Wene
Division of Mathematics, Computer Science, and Statistics, The University of Texas at San Antonio, San Antonio, Texas 78285

(Received 28 June 1988; accepted for publication 3 November 1988)

A collection of algebras with product that preserves the metric of the underlying vector space is investigated. It is shown that this collection includes the generalized quaternions, the composition algebras, and the algebra of color. All the algebras in this collection that are either commutative or alternative are determined; the ramifications of nondegenerate metrics are detailed. If the algebra is alternative and not of dimension 1, 2, 4, or 8 over the base field, the metric must be degenerate.

I. INTRODUCTION

There are three approaches used to analyze the structure of a vector space with metric; these approaches are based on the theories of either Lie algebras, Clifford algebras, or Cayley algebras. There are many interconnections between these theories, and any detailed analysis should utilize elements of all three theories. This paper will concentrate on the non-Lie theory, particularly on the role of quadratic algebras in these theories.

Quadratic algebras have enjoyed an almost continuous presence in physics literature since the discovery of the quaternion division ring by Sir William Rowan Hamilton on October 16, 1843.[1] Clifford, following Hamilton's success with the quaternions, generalized them to any dimension in the universal Clifford algebra.[2] While the Clifford algebras are all associative, they are not all quadratic algebras; however, all Clifford algebras are constructed from quadratic algebras. An extensive bibliography on research in Clifford algebras is available.[3] (See also Ref. 4.)

Nonassociative quadratic algebras appeared in the literature as early as 1845[5]; the eight-dimensional alternative division algebra (appearing in the literature as *octonions*, *Cayley numbers*, *Cayley–Graves numbers*, or *octaves*) is a quadratic division algebra, as are all algebras constructed via the Cayley–Dickson process. For a review of octonions in physics literature, see Wene,[6] and Sorgsepp and Lohmus.[7] Noting that the dimensions of the Clifford algebras and those constructed via the Cayley–Dickson process are powers of two, Wene[6] gives a construction relating both algebras.

Because of a growing interest in quadratic algebras, especially those of dimension *not* a power of two over the scalar field (Domokos and Kovesi-Domokos,[8,9] Plebanski and Przanowski,[10,11] and Wene[12–14]), a better understanding of the algebraic properties of quadratic algebras and their relation to the underlying vector space is desirable. We work in the setting of nonassociative quadratic algebras constructed from a vector space with a metric. Our main result concerns alternative algebras, and therefore completely describes the generalized quaterions in Refs. 10, 11, and 15.

Let V be an n-dimensional vector space over the field F of real numbers R or complex numbers C. Let $\mathscr{A}$ be an alternative algebra that is the vector space sum

$$\mathscr{A} = Fe \oplus V, \tag{1}$$

where e is the identity of $\mathscr{A}$ and $vv \in F$ for all v in V. Then this product induces a quadratic form on $\mathscr{A}$, $N(a)$, and an associated bilinear form,

$$B(a,b) = [N(a+b) - N(a) - N(b)]/2. \tag{2}$$

The algebra $\mathscr{A}$ is the vector space sum

$$\mathscr{A} = \mathscr{C} \oplus \mathscr{N}, \tag{3}$$

where $\mathscr{N}$ is a nilideal that is the *nilradical* of $\mathscr{A}$ and

$$\mathscr{N} = \{a \in \mathscr{A} \mid B(a,x) = 0 \quad \text{for all} \quad x \quad \text{in} \quad \mathscr{A}\}. \tag{4}$$

The bilinear form B restricted to $\mathscr{C}$ is nondegenerate, and $\mathscr{C}$ is one of the following: (i) F, (ii) $F \oplus F$, the algebra direct sum, (iii) $\mathscr{C}$, the complex numbers, (iv) $F(2)$, the ring of 2×2 matrices over F, (v) H, the quaternion division ring, (vi) $\mathscr{O}$, the octonions, or (vii) the split octonions.

Thus if the bilinear form is nondegenerate, the alternative algebra $\mathscr{A}$ must be one of these seven algebras. If $\mathscr{A}$ is associative, we have only five possible algebras.

If the algebra $\mathscr{A}$ is alternative, and the dimension of V over F is not 1, 3, or 7, then the bilinear form (2) is degenerate. One of the few discussions of algebras associated with degenerate bilinear forms is that of Ablamowicz.[16,17]

The construction in Sec. II assumes only that the vector space V has a metric defined on it. A construction that assumes a vector space with an anticommuting product defined on it is given by Czerwinski[18] and Osborn.[19] In Sec. III we derive general results for flexible quadratic algebras, and in Sec. IV we apply these results to the varieties of communicative algebras and alternative algebras.

All algebras will be finite dimensional over the field F, where F is either R or C.

II. THE CONSTRUCTION

Let $e_i, i = 1,2,\ldots,n$ be a basis for the vector space V over R, the field of real numbers, and let

$$X = x^i e_i, \tag{5a}$$

$$Y = y^j e_j, \tag{5b}$$

be two arbitrary vectors of that space. A symmetric bilinear form $G: V\times V \to R$,

$$G(X,Y) = g_{ij} x^i y^j, \tag{6}$$

is called a *metric* (or *metric form*) of the given basis and its

 0022-2488/89/051034-05$02.50

coefficients g_{ij} are called the metric coefficients of that basis. We will also use G to denote the matrix $(g)_{ij}$ of the form G.

A functional Q: $V\times V\to R$ is *quadratic* if the map B: $V\times V\to R$, defined by

$$B(X,Y)=[Q(X+Y)-Q(X)-Q(Y)]/2, \tag{7}$$

is a bilinear form on V. It therefore makes no difference whatsoever in principle whether one considers symmetric bilinear or quadratic forms, for any statement about quadratic forms can be reformulated as a statement about symmetric bilinear forms, and vice versa. We say that $Q(X)$ has property P if and only if $B(X,Y)$ has property P.

The metric G is called *positive definite* if

$$g_{ij}x^iy^j>0, \tag{8}$$

for all $(x^1,...,x^n)\neq(0,0,...,0)$. Here G is *positive* if

$$g_{ij}x^iy^j>0, \tag{9}$$

for all $(x^1,...,x^n)$. Similar definitions hold for *negative definite* and *negative* metrics. If there is some vector X in V and

$$G(X,Y)=0, \tag{10}$$

for all Y and V, then G is said to be *degenerate*.

Let G: $V\times V\to R$ be a metric and G its matrix relative to the basis $e_i, i=1,2,...,n$. Then G is nondegenerate if and only if the determinant $|G|\neq 0$. If G_i denotes the $i\times i$ submatrix of G in the upper left-hand corner of G, is positive definite if and only if for each $i=1,2,...,n$, then

$$|G_i|>0. \tag{11}$$

Equivalently, G is positive definite if and only if all of its eigenvalues are positive. There exists a change of basis for V over R, such that the matrix of G relative to this new basis will be a diagonal matrix D with diagonal entries -1, $+1$, or 0. If G is nondegenerate, then all diagonal entries of D will be nonzero.

A direct transfer of these concepts to the case of the ground field C is impossible, since the field C cannot be ordered by a positive class P. We proceed in a more intricate way.

For any complex number z, $\bar{z}$ will denote the conjugate of z. The symmetric bilinear functional G: $V\times V\to R$ is replaced by a *sesquilinear* functional S: $V\times V\to C$, i.e., a functional that is linear in the first argument and semilinear in the second argument,

$$S(\alpha X+\beta Y,Z)=\alpha S(X,Z)+\beta S(Y,Z), \tag{12}$$

$$S(X,Y+Z)=S(X,Y)+S(X,Z), \tag{13}$$

$$S(X,\alpha Y)=\bar{\alpha}S(X,Y), \tag{14}$$

for any vectors X, Y, Z in V and all α,β in C. A sesquilinear functional S is said to be *Hermitian* if

$$S(X,Y)=\overline{S(Y,X)}. \tag{15}$$

For a Hermitian functional, the number $S(X,X)$ is real for all vectors in V. Therefore the question of its sign is meaningful and we say that S is positive definite if

$$S(X,X)>0, \tag{16}$$

for any nonzero vector X of the space V. We note that the associated form B: $V\times V\to R$, where

$$B(X,Y)=[S(X,Y)+S(Y,X)]/2, \tag{17}$$

is real, bilinear, and symmetric.

Let V be an n-dimensional vector space over $R(X)$ and G: $V\times V\to R(C)$ a symmetric bilinear (sesquilinear) form on V. Choose a basis e_i, $i=1,2,...,n$ such that G is given by a diagnonal matrix with entries -1, $+1$, or $0(+1$ or $0)$. Let $\mathscr{A}$ be any $(n+1)$-dimensional algebra with identity e over $R(C)$ which is the vector space direct sum

$$\mathscr{A}=\mathrm{Re}\oplus V, \tag{18}$$

such that for each v in $V, vv=G(v,v)e$. The set $e_0=e, e_i$, $i=1,2,...,n$ will be a basis for $\mathscr{A}$ over $R(C)$. Call the elements of V *vectors*, and the elements of *Re scalars*.

For $x\in\mathscr{A}$, $x=\alpha e+v$, $\alpha\in R$, and $v\in V$, we have

$$(\alpha e+v)^2-[2\alpha(\alpha e)+v]+[\alpha^2-G(v,v)]e=0. \tag{19}$$

If we set $T(\alpha e)=\alpha$, $N(\alpha e)=\alpha^2$, and $N(v)=-G(v,v)$ for all in R and all v in V, then for each $x\in\mathscr{A}$, we have

$$x^2-2T(x)x+N(x)e=0, \tag{20}$$

where $T(x)=\alpha$ and $N(x)=\alpha^2-G(v,v)=\alpha^2+N(v)$ are scalars.

The qualities $T(x)$ and $N(x)$ are called the *trace* and *norm* of x, respectively. The trace is a linear functional on $\mathscr{A}$. The symmetric bilinear form $N(x,y)$ defined on $\mathscr{A}$ by

$$N(x,y)=[N(x+y)-N(x)-N(y)]/2, \tag{21}$$

satisfies

$$N(u,v)=-G(u,v)=-(uv+vu)/2, \tag{22}$$

for all u, v in V. Call $N(x)$ nondegenerate if $N(x,x)$ is. Clearly $N(x,y)$ is nondegenerate if and only if G is.

An algebra $\mathscr{A}$ with identity e satisfying Eq. (20) is called a *quadratic algebra*.

Theorem 1: Each algebra $\mathscr{A}$ constructed above is quadratic.

Example 1: Minkowski space with metric $g_{uv}=0$, $u\neq v$, $g_{uu}=(1,-1,-1,-1)$ can be constructed as an algebra from R^3 with the usual metric $h_{uv}=\delta_{uv}$ by setting

$$G(x,y)=[N(x+y)-N(x)-N(y)]/2. \tag{23}$$

If u and v are vectors (in R^3), then

$$G(u,v)=-H(u,v). \tag{24}$$

As we will see later, our construction gives us a simple algebra in all cases. If we insist that the algebra be alternative, we will have constructed the (associative)ring of real 2×2 matrices. The associative case with the related light cone is discussed by Ilamed and Salingaros.[20]

The next section reviews some facts about quadratic algebras.

III. QUADRATIC ALGEBRAS

A general theory of quadratic algebras has yet to be developed; the interested reader is referred to the appropriate sections of the books by Braun and Koecher,[21] Schafer,[22] and Zhevlakov, Slin'ko, Shestakov, and Shorshov,[23] as well as Refs. 18 and 19.

While a general theory is lacking, there is considerable literature concerning quadratic algebras that are alternative.

The alternative quadratic algebras include the composition algebras and all associative algebras. Quadratic algebras in which the map $x \to \bar{x} = 2T(x)e - x$ defines an involution are called Cayley algebras by Boubaki,[21] and are discussed in many places in the literature.

The class of quadratic is closed under the operations of forming subalgebras and homomorphic images; it is not closed under the formation of complete direct sums, as we will see in Theorem 2, and therefore it is not a variety of algebras.

Quantum mechanical operators are often identified with sets of pairwise orthogonal *idempotent* elements in an algebra. An element $f \neq 0$ is idempotent if $f^2 = f$, and idempotents f and g are orthogonal if $fg = 0 = gf$. We see that a quadratic algebra can have, at most, two pairwise orthogonal idempotents.

Theorem 2: Let $\mathscr{A}$ be a quadratic algebra such that $e = f_1 + f_2 + \cdots + f_n$ is a sum of pairwise orthogonal idempotents f_i. Then $n = 1$ or $n = 2$.

Proof: Consider the element $a = f_1 + 2f_2 + \cdots + nf_n$ of $\mathscr{A}$. Since it is quadratic, there exist scalars x and y such that

$$a^2 + xa + ye = 0. \tag{25}$$

Equivalently,

$$[1 + x + y]f_1 + [4 + 2x + y]f_2 + \dots + (n^2 + nx + y)f_n = 0. \tag{26}$$

But this system of equations has a solution if and only if $n = 1$ or $n = 2$.

Corollary 3: Let the quadratic algebra $\mathscr{A}$ be the algebra direct sum

$$\mathscr{A} = \mathscr{A}_1 \oplus \mathscr{A}_2 \oplus \cdots \oplus \mathscr{A}_n. \tag{27}$$

Then $n = 1$ or $n = 2$.

Proof: Since $\mathscr{A}$ has an identity, each $\mathscr{A}_i$ has an identity f_i, $i = 1,2,\dots,n$. Then the f_i's form a set of pairwise orthogonal idempotents.

Each quadratic algebra is power associative, in the sense that any subalgebra generated by a single element is associative. Any finite-dimensional power-associative algebra $\mathscr{B}$ has a unique maximal nilideal $\mathscr{N}$, and the quotient algebra $\mathscr{B}/\mathscr{N}$ has maximal nilideal 0. $\mathscr{N}$ is called the *nilradical* of $\mathscr{B}$, and $\mathscr{B}$ is called *semisimple* in case $\mathscr{N} = 0$.

The most general variety of quadratic algebras we will study are the (noncommutative) *Jordan algebras.* An algebra $\mathscr{J}$ is a *Jordan algebra* if it is *flexible*, that is,

$$(xy) - x(yx) = 0, \tag{28}$$

and it satisfies the *Jordan identity*

$$(xy)x^2 - x(yx^2) = 0, \tag{29}$$

for every pair x and y in $\mathscr{J}$. The variety of Jordan algebras includes both the associative and alternative algebras.

Lemma 4: A quadratic algebra $\mathscr{A}$ is Jordan if and only if it is flexible.

Proof: From

$$x^2 = 2T(x)x - N(x)e, \tag{30}$$

we see that

$$(xy)x^2 = 2T(x)(xy)x - N(x)xy, \tag{31}$$

$$x(yx^2) = 2T(x)x(yx) - N(x)xy, \tag{32}$$

will be Jordan if

$$T(x)(xy)x = T(x)x(yx). \tag{33}$$

If $\mathscr{A}$ is flexible, these will be equal.

A quadratic algebra $\mathscr{A}$ will be flexible if and only if the trace $T(x)$ is associative; that is, $T((xy)z) = T(x(yz))$ for all x, y, and z in $\mathscr{A}$. If $\mathscr{A}$ is flexible, then the mapping

$$x \to \bar{x} = 2T(x)e - x, \tag{34}$$

is an involution in $\mathscr{A}$, and satisfies

$$\overline{x + y} = \bar{x} + \bar{y}, \tag{35}$$

$$\bar{\bar{x}} = x, \tag{36}$$

$$\overline{xy} = \bar{y}\bar{x}, \tag{37}$$

for all x, y in $\mathscr{A}$. Equation (20) implies

$$x\bar{x} = \bar{x}x = N(x), \tag{38}$$

$$x + \bar{x} = 2T(x)e \tag{39}$$

(see Braun and Koecher,[21] p. 216). We know that the overbar is the unique involution in $\mathscr{A}$ satisfying (38) and (39). We will call the overbar *standard involution* in $\mathscr{A}$. The map (33) may define an involution, even if $\mathscr{A}$ is not flexible.

Define a bilinear form T: $\mathscr{A} \times \mathscr{A} \to F$ by

$$T(x,y) = T(xy). \tag{40}$$

Then T is an associative bilinear form on $\mathscr{A}$. The next result relates the bilinear form T and the metric G.

Lemma 5: If $\mathscr{A}$ is a flexible quadratic algebra, then

$$T(x,y) = T(xy) = N(\bar{x},y). \tag{41}$$

If $x,y \in V$, then $T(x,y) = -g(x,y)$. Here $T(x,y)$ is nondegenerate if and only if G is.

Proof: Let $x,y \in \mathscr{A}$. The equations

$$(x + y)^2 - 2T(x + y)[x + y] + N(x + y)e = 0, \tag{42a}$$

$$x^2 - 2T(x)x + N(x)e = 0, \tag{42b}$$

$$y^2 - 2T(y)y + N(y)e = 0, \tag{42c}$$

imply

$$xy + yx - 2[T(x)y + T(y)x] + 2N(x,y) = 0, \tag{43}$$

$$xy + yx - (x + \bar{x})y - (y + \bar{y})x + 2N(x,y) = 0, \tag{44}$$

$$\bar{x}y - \bar{y}x + 2N(x,y) = 0, \tag{45}$$

$$T(\bar{x}y) = N(xy). \tag{46}$$

If $u,v \in V$, then $T(uv) = -g(u,v) = g(\bar{u},v)$.

The even easier proof is omitted.

Lemma 6: Let $\mathscr{A}$ be a flexible quadratic algebra. Then

(a) The radical $\mathscr{N}$ of $\mathscr{A}$ is the set $\mathscr{N} = \{n \in \mathscr{A} \mid T(n,x) = 0 \text{ for all } x \text{ in } \mathscr{A}\}$.

(b) The quotient algebra $\bar{\mathscr{A}} = \mathscr{A}/\mathscr{N}$ is the algebra direct sum of (at most two) simple algebras.

(c) The map T: $\bar{\mathscr{A}} \times \bar{\mathscr{A}} \to F$ defined by

$$T(a + \mathscr{N}, b + \mathscr{N}) = T(a,b), \tag{47}$$

is an associative trace form on $\bar{\mathscr{A}}$ that is nondegenerate. This will be called the induced form on $\bar{\mathscr{A}}$.

The "algebra of color" discussed in Refs. 8, 9, 12, and 13 is a flexible quadratic algebra that is not commutative.

This is about as far as we can push the theory of flexible quadratic algebras without some additional assumptions. It would be nice to have a quadratic algebra analog of the Wedderburn principal theorem for associate algebras.

Theorem 7: Let $\mathscr{A}$ be a finite-dimensional associative F algebra. Then $\mathscr{A}$ has a vector space decomposition,

$$\mathscr{A} = \mathscr{C} + \mathscr{N}, \tag{48}$$

where C is a subalgebra of $\mathscr{A}$ and is isomorphic to the quotient algebra $\mathscr{A}/\mathscr{N}$, which is the nilideal of $\mathscr{A}$.

Proof: See p. 47 of Ref. 24.

There is no principal theorem for the variety of Jordan algebras. Such theorems do exist for the variety of alternative algebras and the variety of commutative Jordan algebras. The next two actions of this paper will examine these two cases. We end this section with a remark on the Lie structure of flexible quadratic algebras.

In general, the flexible quadratic algebras, under the so-called *Lie product* [,],

$$[a,b] = ab - ba, \tag{49}$$

do not form Lie algebras. To see this, note that the octonion algebra is a flexible quadratic algebra that does not form a Lie algebra under the product (49).[25]

IV. COMMUTATIVE QUADRATIC ALGEBRAS

Commutative quadratic algebras are commutative Jordan algebras. The finite-dimensional simple algebras in this variety are classified as type A, B, C, D, or E (Ref. 22, p. 101). Those of type D are associated with the Jordan algebra of a symmetric bilinear form f.

Example 2: Let V be a finite-dimensional vector space over the field F with symmetric bilinear form f. Construct the vector space

$$\mathscr{J} = Fe + V, \tag{50}$$

and define a product in $\mathscr{J}$ by

$$(\alpha e + x)(\beta e + y) = (\alpha\beta + f(x,y))e + \beta x + \alpha y, \tag{51}$$

for all α, β in F and x, y in V. Clearly $\mathscr{J}$ with this product is a commutative Jordan algebra.

The nilradical of $\mathscr{J}$ will be the set

$$\mathscr{N} = \{n \in \mathscr{J} \,|\, f(v,n) = 0 \text{ for all } v \text{ in } J\}. \tag{52}$$

The quotient algebra $\mathscr{J}/\mathscr{N}$ is isomorphic to the Jordan algebra of the form induced by f in $\bar{v} = v/\mathscr{N}$.

The algebra $\mathscr{J}$ will be simple, that is, $\mathscr{J}$ will have no ideal $\neq \mathscr{J}, 0$ if the dimension of V over F is greater than one and $f(x,y)$ is nondegenerate. Here $\mathscr{J}$ contains an idempotent element $f \neq 0$, e, if and only if there exists $v \in V$ such that $f(v,v) = 1$ (see Ref. 26, p. 14).

We return to the more general case and assume, only for discussion of properties of these algebras, that $\mathscr{J}$ is commutative. Then we have the following theorem.

Theorem 8: Let $\mathscr{A}$ be a commutative quadratic algebra. Then

$$\mathscr{A} = \mathscr{C} \oplus \mathscr{N}, \tag{53}$$

where $\mathscr{N}$ is the nilradical of $\mathscr{A}$, and $\mathscr{C}$ is one of the following: (a) F, (b) $F \oplus F$, (c) the Jordan algebra of the set of symmetric elements of the ring of 2×2 matrices with an involution over F or the quaternions, or (d) the Jordan algebra of nondegenerate symmetric bilinear form.

Proof: From Ref. 27, there is a Wedderburn principal theorem for commutative Jordan algebras that gives us a decomposition as in (53); however, we know only that the algebra is the direct sum of simples. Lemma 3 of Ref. 23 applies, and $\mathscr{C}$ is simple or isomorphic to the direct sum $F \oplus F$. The above classification scheme for simple Jordan algebras is based on the maximal number (degree) of pairwise orthogonal idempotents in the respective algebras, and in our case, the algebras must have at most two idempotents. The result now follows upon examining the degree two algebras in each case.

V. THE ALTERNATIVE CASE

In this section, $\mathscr{A}$ is an alternative algebra that has been constructed as in Sec. II.

Theorem 9: The alternative algebra $\mathscr{A}$ is a division algebra if and only if $g(v,v) < 0$ for all nonzero v in V, and the dimension of V over F is one, three, or seven.

Proof: We use the fact that if every nonzero element of an alternative algebra with identity has an inverse, then that algebra is a division algebra (Ref. 22, p. 38). Let α be in F and v in V. Then by (38),

$$\begin{aligned}(\alpha e + v)&[\alpha 1 + v + 2\alpha(\alpha e + v)]\\ &= [\alpha e + v + 2\alpha(\alpha e + v)](\alpha e + v)\\ &= -n(de + V)\\ &= [-\alpha^2 - (v,v)] = 0,\end{aligned}$$

if and only if $\alpha = 0$, $v = 0$.

The alternative division algebras over R are R, $\mathscr{C}$, the real quaternions, and the octonions.

Theorem 10: If the dimension of V over F is at least two, and $g(v,v) = 0$ only if $v = 0$, then $\mathscr{A}$ is simple.

Proof: Suppose $0 \in I \leqslant \mathscr{A}$ is an ideal. Let $\alpha e + v \in I$, $\alpha \in F$, $v \in V$, and $w \neq v$ an element of V. Then

$$\begin{aligned}&\alpha w + wv \in I,\\ &\alpha w + vw \in I,\\ &2\alpha w + wv + vw \in I, \quad wv + vw \in Ff,\\ &v + \alpha \in I,\end{aligned}$$

and thus

$$y = 2\alpha^2 w + (wv + vw)v \in I.$$

Thus $0 \neq x \in I \cup V$ and x^2.

We can do no better than this. The bound of the dimension of $\mathscr{A}$ over R must be at least 2, as the following example shows.

Example 3: Let $V = F$, $\mathscr{A} = Fe \oplus V$. Multiplication for $\mathscr{A}$ is given by $v^2 = 1$ for $v \in F$, and e the identity. Then $e + v$ generates an ideal in $\mathscr{A}$,

$$\begin{aligned}&N(e + v) = e - e = 0,\\ &(e + v)(\alpha e + \beta v) = (\alpha e + \beta)(e + v).\end{aligned}$$

Theorem 11: Let $\mathscr{A}$ be an alternative algebra. Then $\mathscr{A}$ has a vector space decomposition

$$\mathscr{A} = \mathscr{C} + \mathscr{N}, \tag{54}$$

where $\mathscr{N}$ is the nilradical of $\mathscr{A}$ and $\mathscr{C}$ is one of the following: (i) F, (ii) $F \oplus F$, (iii) C, (iv) F_2, the ring of 2×2 matrices over F, (v) the quaterion division ring, (vi) the octonions, or (vii) the split octonions.

Proof: Applying the Wedderburn principal theorem for alternative rings, we get a decomposition as in (51) that tells us that the algebra $\mathscr{C}$ is isomorphic to the quotient algebra, which is a direct sum of simple alternative algebras. The simple alternative algebras are either simple associative or an octonion algebra. Since the quotient algebra can have at most two pairwise orthogonal idempotents, it must be one of the possibilities listed in the theorem.

Corollary 12: If $\mathscr{A}$ is alternative and the bilinear form is nondegenerate, then the dimension of $\mathscr{A}$ over F is 1, 2, 4, or 8. In this case, it is possible to choose a basis for $\mathscr{A}$ in which the norm form $N(x)$ is given by one of the following:

(i) $N(x) = x^2$,

(ii) $N(x) = x_1^2 - \alpha x_2^2$,

(iii) $N(x) = x_1^2 - \alpha x_2^2 - \beta x_3^2 - \alpha\beta_4^2$,

(iv) $N(x) = x_1^2 - \alpha x_2^2 - \beta x_3^2 + \alpha\beta x_4^2 - \gamma x_5^2 + \alpha\gamma x_6^2 + \beta\gamma x_7^2 - \alpha\beta\gamma x_8$,

where $\alpha, \beta, \gamma \in F$, $\alpha\beta\gamma \neq v$.

Proof: See p. 33 of Ref. 22.

Corollary 13: If $\mathscr{A}$ is alternative, and the dimension of $\mathscr{A}$ over F is not 1, 2, 4, or 8, then the norm is degenerate.

VI. CONCLUSION

Thus we see that if we require any respectable behavior of the algebra of arbitrary dimension, we force the metric to be degenerate. In the alternative case, if the metric is not identically zero, the algebra must contain one of the algebras listed in Theorem 11, and a nilalgebra. The commutative case is equally limited as to possible algebras.

[1]E. T. Bell, *Men of Mathematics* (Simon and Schuster, New York, 1937), p. 360.

[2]W. K. Clifford, "On the classification of geometric algebras," Paper XLIII in *Mathematical Papers of W. K. Clifford*, edited by R. Tucker (MacMillan, London, 1882).

[3]N. A. Salingaros and G. P. Wene, "The Clifford algebra of differential forms," Acta Appl. Math. **4**, 271 (1985).

[4]D. Li, C. P. Poole, and H. A. Farach, "A general method of generating and classifying Clifford algebras," J. Math. Phys. **27**, 1173 (1986).

[5]A. Cayley, "On Jacobi's elliptic functions, in reply to the Rev. Brice Bronwin; and on quaternions," Philios. Mag. **26**, 210 (1845).

[6]G. P. Wene, "A construction relating Clifford algebras and Cayley–Dickson algebras," J. Math. Phys. **25**, 2351 (1984).

[7]L. Sorgsepp and J. Lohmus, "About nonassociativity in physics and Cayley–Graves octonions," Hadronic J. **2**, 1388 (1979).

[8]G. Domokos and S. Kovesi-Domokos, "The algebra of color," J. Math. Phys. **19**, 1477 (1978).

[9]G. Domokos and S. Kovesi-Domokos, "Towards an algebraic quantum chromodynamics," Phys. Rev. D **19**, 2984 (1979).

[10]J. F. Plebanski and M. Przanowski, "Generalizations of the quaternion algebra and Lie algebras," J. Math. Phys. **29**, 529 (1988).

[11]J. F. Plebanski and M. Przanowski, "Notes on a cross product of vectors," J. Math. Phys. **29**, 2334 (1988).

[12]G. P. Wene, "An example of a flexible, Jordan-admissible algebra of current use in hadron physics," Hadronic J. **1**, 944 (1978).

[13]G. P. Wene, "A generalization of the construction of Ilamed and Salingaros," J. Math. Phys. **24**, 221 (1983).

[14]G. P. Wene, "Algebras with anticommuting basal elements, space-time symmetries, and quantum theory," J. Math. Phys. **25**, 414 (1984).

[15]R. T. Jantzen, "Generalized quaterions and space-time symmetries," J. Math. Phys. **23**, 1741 (1982).

[16]R. Ablamowicz, "Structure of spin group associated with degenerate Clifford algebras," J. Math. Phys. **27**, 1 (1986).

[17]R. Ablamowicz, "Reformation and contraction in Clifford algebras," J. Math. Phys. **27**, 423 (1986).

[18]R. A. Czerwinski, "Bonded quadratic division algebras," Pacific J. Math. **64**, 341 (1976).

[19]J. M. Osborn, "Quadratic division algebras," Trans. Am. Math. Soc. **105**, 202 (1962).

[20]Y. Ilamed and N. Salingaros, "Algebras with three anticommuting elements. I. Spinors and quaternions," J. Math Phys. **22**, 2091 (1981).

[21]H. Braun and M. Koecher, *Jordan-Algebras* (Springer, New York, 1966).

[22]R. D. Schafer, *Introduction to Nonassociative Algebras* (Academic, New York, 1966).

[23]K. A. Zhevlakov, A. M. Slin'ko, I. P. Shestakov, and A. I. Shirshov, *Rings That Are Nearly Associative* (Academic, New York, 1982).

[24]A. A. Albert, *Structure of Algebras* (Am. Math. Soc., Providence, RI, 1961).

[25]A. A. Sagle, "Simple Malcev algebras over a field of characteristic zero," Pacific J. Math. **12**, 1057 (1962).

[26]N. Jacobson, *Structure and Representations of Jordan Algebras* (Am. Math. Soc., Providence, RI (1968).

[27]A. J. Penico, "The Wedderburn principal theorem for Jordan Algebras," Trans. Am. Math. Soc. **70**, 404 (1951).

A Newlander–Nirenberg theorem for supermanifolds

Andrew McHugh
Department of Mathematics, State University of New York at Stony Brook, Stony Brook, New York 11794

(Received 22 June 1988; accepted for publication 4 January 1989)

A supermanifold version of the Newlander–Nirenberg theorem is proven. The classical Newlander–Nirenberg theorem states the conditions under which an almost complex structure on a differentiable manifold gives rise to a complex structure. A definition is given here of an almost complex structure on a differentiable supermanifold along with the conditions under which this gives rise to a complex structure on the supermanifold.

I. INTRODUCTION

Let us recall the definitions of superalgebras and supermanifolds. A superalgebra or $\mathbf{Z}_2$-graded commutative algebra is an algebra in which every element can be written as a sum of an *even* element and an *odd* element. Even elements commute with all elements in the algebra and odd elements anticommute with odd elements. A differentiable supermanifold is a pair (X,A), where X is a differentiable manifold and A is a sheaf on X of $\mathbf{Z}_2$-graded commutative algebras over $\mathbf{R}$ which is locally isomorphic to the sheaf $\Lambda^*_{C^\infty}(C^\infty)^{\oplus m}$. Let N be the sheaf of nilpotents of A. We also require that globally $A/N \cong C^\infty$. Then N/N^2 is a locally free sheaf of C^∞ modules.

The coordinate neighborhoods for a differentiable supermanifold (X,A) are by definition open sets U that are coordinate neighborhoods of X and are such that $A|_U \cong \Lambda^*_{C^\infty}(C^\infty)^{\oplus m}|_U$. Let $s^1,s^2,\ldots,s^m$ be linearly independent sections of $(C^\infty(U))^{\oplus m}$. Then sections of $A(U)$ have the form

$$f=\sum_I f_I s^I,$$

where $f_I = f_I(x^1,x^2,\ldots,x_n) \in C^\infty(U)$ and $s^I = s^{i_1}s^{i_2}\cdots s^{i_p}$, $I=(i_1 \leqslant i_2 \leqslant \cdots \leqslant i_p)$. The $x^1,\ldots,x^n$ and $s^1,\ldots,s^m$ are referred to as even and odd coordinates, respectively. The $\mathbf{Z}_2$ grading on A is represented locally by the following: f is even if

$$f=\sum_{|I|\ \text{even}} f_I s^I$$

and f is odd if

$$f=\sum_{|I|\ \text{odd}} f_I s^I.$$

Note that a change of coordinates is required to preserve the $\mathbf{Z}_2$ grading.

A complex supermanifold is a pair (X,A), where X is a complex manifold and A is a sheaf of $\mathbf{Z}_2$-graded algebras over $\mathbf{C}$ that is locally isomorphic to $\Lambda^*_{\mathscr{O}}\mathscr{O}^{\oplus m}$. We also require that globally $A/N \cong \mathscr{O}$ and that N/N^2 is a locally free sheaf of $\mathscr{O}$ modules. Locally, sections of A on a coordinate neighborhood U will have the form

$$g=\sum_I g_I \eta^I,$$

where $g_I = g_I(z^1,z^2,\ldots,z^n) \in \mathscr{O}(U)$ and $\eta^1,\ldots,\eta^m$ are linearly independent sections of $\mathscr{O}^{\oplus m}$. The $z^1,\ldots,z^n$ and $\eta^1,\ldots,\eta^m$ are referred to, respectively, as the even and odd complex coordinates.

II. ALMOST COMPLEX STRUCTURES ON SUPERMANIFOLDS

We wish to prove an analog of the Newlander–Nirenberg theorem for supermanifolds. The classical Newlander–Nirenberg theorem[1] states the conditions under which an almost complex structure on a manifold gives rise to a complex structure. We give a definition of an almost complex structure on a differentiable supermanifold and conditions under which this gives rise to a complex structure. The proof applies the classical Newlander–Nirenberg theorem on the underlying manifold. We then build up the result onto the odd coordinates using a well-known lemma on the holomorphic structure of vector bundles and finally a finite iterative procedure.

Let (X,A) be a differentiable supermanifold of dimension $(2n,2m)$. Define an almost complex structure on (X,A) as an even automorphism

$$J:\mathrm{Der}(A,A) \to \mathrm{Der}(A,A)$$

of sheaves of A modules such that $J^2 = -id$. Here $\mathrm{Der}(A,A)$ is the sheaf of graded derivations of A into A. Let $x^1,\ldots,x^n,s^1,\ldots,s^m$ be local coordinates for (X,A) on $U \subset X$. Let

$$\Theta_\alpha = \frac{\partial}{\partial x^\alpha} - iJ\frac{\partial}{\partial x^\alpha}$$

and

$$\Phi_j = \frac{\partial}{\partial s^j} - iJ\frac{\partial}{\partial s^j}.$$

Then Θ_α, $\overline{\Theta}_\alpha$, Φ_j, $\overline{\Phi}_j$ is a local basis for

$$\mathrm{Der}(A(U),A(U)) \otimes \mathbf{C} = \mathrm{Der}(A_{\mathbf{C}}(U),A_{\mathbf{C}}(U)),$$

where $A_{\mathbf{C}}(U) = A(U) \otimes \mathbf{C}$. Consider

$$\Omega^1_{\mathbf{C}}(U,A) = \mathrm{Hom}_{A_{\mathbf{C}}(U)}(\mathrm{Der}(A_{\mathbf{C}}(U),A_{\mathbf{C}}(U)),A_{\mathbf{C}}(U))$$

and a basis θ^α, $\overline{\theta}^\alpha$, ϕ^j, $\overline{\phi}^j$ dual to Θ_α, $\overline{\Theta}_\alpha$, Φ_j $\overline{\Phi}_j$. Sections of $\Omega^1_{\mathbf{C}}(U,A)$ are referred to as super one-forms. One also has super p-forms $\Omega^p_{\mathbf{C}}(U,A)$ and covariant differentiation d: $\Omega^p_{\mathbf{C}}(U,A) \to \Omega^{p+1}_{\mathbf{C}}(U,A)$. We refer the reader to Ref. 2. Henceforth $\Omega^1_{\mathbf{C}}(U,A)$ is abbreviated as Ω^1.

An almost complex structure J on a differentiable supermanifold (X,A) is said to be integrable if one can find local complex superfunctions $z^\alpha = x^\alpha + iy^\alpha$ and $\eta^j = s^j + it^j$ such that x^α, y^α, $s^j + it^j$ such that x^α, y^α, s^j, t^j is a local coordinate system and such that

$$J\left(\frac{\partial}{\partial x^\alpha}\right) = \frac{\partial}{\partial y^\alpha}, \quad J\left(\frac{\partial}{\partial y^\alpha}\right) = -\frac{\partial}{\partial x^\alpha},$$

0022-2488/89/051039-04$02.50

$$J\left(\frac{\partial}{\partial s^j}\right)=\frac{\partial}{\partial t^j},\quad J\left(\frac{\partial}{\partial t^j}\right)=-\frac{\partial}{\partial s^j}.$$

III. ANALOG OF THE NEWLANDER–NIRENBERG THEOREM FOR SUPERMANIFOLDS

We now present the super Newlander–Nirenberg theorem. An almost complex structure J on a differentiable supermanifold (X,A) is integrable if and only if $d\theta^\alpha=0 \bmod \theta^\beta, \phi^k$ and $d\phi^j=0 \bmod \theta^\beta, \phi^k$.

As one direction is trivial, we need only show this to be a sufficient condition. Notice that J gives rise to a J^* on super one-forms and that θ^α, ϕ^j span the $+i$ eigenspace of J^*. We seek complex superfunctions z^α, η^j such that dz^α and $d\eta^j$ are in the span of θ^β and ϕ^k and such that $dz^\alpha, d\bar z^\alpha, d\eta^j, d\bar\eta^j$ span Ω^1.

Consider the one-forms $\mathrm{red}(\theta^\alpha)\in\Omega_{\mathrm{C}}(U)$ where red: $\Omega^1_{\mathrm{C}}(U,A)\to\Omega^1_{\mathrm{C}}(U)$ is induced by the quotient map red: $A_{\mathrm{C}}(U)\to A_{\mathrm{C}}(U)/N_{\mathrm{C}}\cong C^\infty(U)$. (red is for "reduced.") They satisfy the conditions of the classical Newlander–Nirenberg theorem. (See Ref. 3.) Thus we obtain local holomorphic coordinates z^α on U. One can find linear combinations of the θ^α, $\alpha=1,\ldots,\eta$ $\theta^{\alpha\prime}$ which form a new basis for the $+i$ eigenspace of J^* and are such that

$$\theta^{\alpha\prime}=dz^\alpha \bmod N\Omega^1,$$

where N is the nilpotent of $A(U)$. We also have

$$\phi^j=ds^j+i\sum_k ds^k f^j_k \bmod N\Omega^1,\quad f^j_k\in C^\infty\otimes\mathbf{C}.$$

Let

$$\eta^j=s^j+i\sum_k s^k f^j_k.$$

Then

$$\phi^j=d\eta^j \bmod N\Omega^1.$$

We proceed to refine our z^α and η^j, considering higher and higher terms of nilpotency and taking linear combinations in the span of θ^α and η^j when necessary. We first refine our choice of θ^α and ϕ^j with respect to the coefficients of dz^α and $d\eta^j$ by the following inductive procedure.

Fix a positive integer p. Suppose that

$$\theta^{\alpha\prime}=dz^\alpha+\sum_\beta dz^\beta A^\alpha_\beta+\sum_k d\eta^k B^\alpha_k+\sum_\beta d\bar z^\beta C^\alpha_\beta+\sum_k d\bar\eta^k D^\alpha_k, \tag{1}$$

$$\theta^{j\prime}=d\eta^j+\sum_\beta dz^\beta E^j_\beta+\sum_k d\eta^k F^j_k+\sum_\beta d\bar z^\beta G^j_\beta+\sum_k d\bar\eta^k H^j_k, \tag{2}$$

is a basis for the $+i$ eigenspace of J^* such that $A^\alpha_\beta, B^\alpha_k, E^j_\beta, F^j_k\in N^p$ and $C^\alpha_\beta, D^\alpha_k, G^j_\beta, H^j_k\in N$. Then there is a new basis for the $+i$ eigenspace of J^* given by

$$\begin{aligned}\theta^{\alpha\prime\prime}&=\theta^{\alpha\prime}-\sum_\beta\theta^{\beta\prime}A^\alpha_\beta-\sum_k\phi^{k\prime}B^\alpha_k\\&=dz^\alpha+\sum_\beta dz^\beta A^{\alpha\prime}_\beta+\sum_k d\eta^k B^{\alpha\prime}_k\\&\quad+\sum_\beta d\bar z^\beta C^{\alpha\prime}_\beta+\sum_k d\bar\eta^k D^{\alpha\prime}_k,\end{aligned} \tag{3}$$

$$\begin{aligned}\theta^{j\prime\prime}&=\theta^{j\prime}-\sum_\beta\theta^{\beta\prime}E^j_\beta-\sum_k\phi^k F^j_k\\&=dn^j+\sum_\beta dz^\beta E^{j\prime}_\beta+\sum_k d\eta^k F^{j\prime}_k\\&\quad+\sum_\beta d\bar z^\beta G^{j\prime}_\beta+\sum_k d\bar\eta^k H^{j\prime}_k,\end{aligned} \tag{4}$$

where $A^{\alpha\prime}_\beta, B^{\alpha\prime}_k, E^{j\prime}_\beta, F^{j\prime}_k\in N^{p+1}$ and $C^{\alpha\prime}_\beta, D^{\alpha\prime}_k, G^{j\prime}_\beta, H^{j\prime}_k\in N$. Since $N^{m+1}=0$ one obtains, after applying the procedure a finite number of times, a basis for the $+i$ eigenspace of J^* in the form

$$\theta^{\alpha\prime\prime}=dz^\alpha+\sum_\beta d\bar z^\beta C^{\alpha\prime\prime}_\beta+\sum_k d\bar\eta^k D^{\alpha\prime\prime}_k, \tag{5}$$

$$\theta^{j\prime\prime}=d\eta^j+\sum_\beta d\bar z^\beta G^{j\prime\prime}_\beta+\sum_k d\bar\eta^k H^{j\prime\prime}_k, \tag{6}$$

for some $C^{\alpha\prime\prime}_\beta, D^{\alpha\prime\prime}_k, G^{j\prime\prime}_\beta, H^{j\prime\prime}_k\in N$.

We now work with $\theta^{\alpha\prime\prime}$ and $\phi^{j\prime\prime}$ as a basis for our $+i$ eigenspace of J^* and hence drop the use of the double primes.

The real and imaginary parts of z^a and η^j form a local real coordinate system for U; in particular,

$$d=\sum_\alpha dz^\alpha\frac{\partial}{\partial z^\alpha}+\sum_\alpha d\bar z^\alpha\frac{\partial}{\partial\bar z^\alpha}+\sum_j d\eta^j+\sum_j d\bar\eta^j\frac{\partial}{\partial\bar\eta^j}.$$

Now if

$$\theta^\alpha=dz^\alpha+\sum_{j,l}d\bar\eta^j\,\eta^l\,b^\alpha_{jl}+\sum_{j,l}d\bar\eta^j\,\bar\eta^l c^\alpha_{jl} \bmod N^2\Omega^1$$

then the condition

$$d\theta^\alpha=0 \bmod \theta^\beta,\phi^k$$

requires that

$$d\bar n^j\,\bar\eta^l c^\alpha_{jl}=d(\bar\eta^j\bar\eta^l)c^\alpha_{jl}.$$

(Here we have used $d\bar z^\beta, d\bar\eta^k, \eta^\beta, \phi^k$ as generators for Ω^*.) If we set

$$z^{\alpha\prime}=z^\alpha+\sum_{j,l}\bar\eta^j\eta^l b^\alpha_{jl}+\sum_{j,l}\bar\eta^j\bar\eta^l c^\alpha_{jl},$$

then

$$\theta^\alpha=dz^{\alpha\prime}-\sum_{j,l}\bar\eta^j\phi^l b^\alpha_{jl} \bmod N^2\Omega^1.$$

Thus set

$$\theta^{\alpha\prime}=\theta^\alpha+\sum_{j,l}\bar\eta^j\phi^l b^\alpha_{jl}=dz^{\alpha\prime} \bmod N^2\Omega^1. \tag{7}$$

Now

$$\phi^j=d\eta^j+\sum_{\beta,k}d\bar z^\beta\,\eta^k f^j_{\beta k}+\sum_{\beta,k}d\bar z^\beta\,\bar\eta^k g^j_{\beta k} \bmod N^2\Omega^1.$$

The condition $d\phi^j=0 \bmod \theta^\beta,\phi^k$ requires $g^j_{\beta k}=0$ for all β, j, k and thus

$$\phi^j=d\eta^j+\sum_{\beta,k}d\bar z^\beta\,\eta^k f^j_{\beta k} \bmod N^2\Omega^1.$$

We are now ready to proceed with the next step in our proof. Namely, let us show that N/N^2 is a sheaf of sections of a holomorphic vector bundle. We first prove a lemma.

Lemma 1: The $f^j_{\beta k}$ above satisfy the equation

$$\frac{\partial f^j_{\alpha k}}{\partial \bar z^\beta} - \frac{\partial f^j_{\beta k}}{\partial \bar z^\alpha} - \sum_l (f^j_{\alpha l} f^l_{\beta k} - f^j_{\beta l} f^l_{\alpha k}) = 0. \qquad (8)$$

Proof: We use the condition $d\phi^j = 0 \bmod \theta^\alpha, \phi^k$. Write

$$\phi^j = d\eta^j + \sum_{\beta,k} d\bar z^\beta \eta^k f^j_{\beta k} + \sum_l d\bar\eta^l g^j_l \bmod N^3\Omega^1,$$

where $g^j_l \in N^2$. Remembering that $d\bar z^\beta$, $d\bar\eta^l$, θ^α, ϕ^k can be used as generators of Ω^*, consider the terms of $d\phi^j$ involving $d\bar z^\alpha \wedge d\bar z^\beta$. The sum of these terms must be zero, and in particular must be zero mod $N^2\Omega^1$. Now, in the expression

$$d\phi^j = \sum_{\beta,k} d\bar z^\beta \wedge d\eta^k f^j_{\beta,k} + \sum_{\beta,k} d\bar z^\beta \eta^k\, df^j_{\beta k} + \sum_l d\bar\eta^l g^j_l \bmod N^2\Omega^1,$$

only the first two summations shall have terms with $d\bar z^\alpha \wedge d\bar z^\beta$ when they are written in terms of the generators $d\bar z^\beta$, $d\bar\eta^l$, θ^β, ϕ^l. Recall that $dz^\beta = \theta^\beta \bmod N^1\Omega^1$ and

$$d\eta^k = \phi^k - \sum_{\alpha,l} d\bar z^\alpha \eta^l f^k_{\alpha l} \bmod N^2\Omega^1.$$

Making this replacement one obtains

$$d\phi^j = \sum_{\beta,k} d\bar z^\beta \wedge \Big(\phi^k - \sum_{\alpha,l} d\bar z^\alpha \eta^l f^k_{\alpha l}\Big) f^j_{\beta k} + \sum_{\beta,k} d\bar z^\beta \wedge \Big(\sum_\alpha \theta^\alpha \frac{\partial f^j_{\beta k}}{\partial z^\alpha} + d\bar z^\alpha \frac{\partial f^j_{\beta k}}{\partial \bar z^\alpha}\Big)\eta^k + \sum_l d\bar\eta^l \wedge dg^j_l \bmod N^2\Omega^1,$$

where we have used $\theta^\alpha \eta^k = dz^\alpha \eta^k \bmod N^2\Omega^1$. Since $d\phi^j = 0 \bmod \theta^\alpha \phi^k$ we have

$$\sum_{\beta,\alpha,k} d\bar z^\beta \wedge d\bar z^\alpha \eta^k \Big(\frac{\partial f^j_{\beta k}}{\partial \bar z^\alpha} - \sum_l f^j_{\beta l} f^l_{\alpha k}\Big) = 0.$$

Since $d\bar z^\beta \wedge d\bar z^\alpha = -d\bar z^\alpha \wedge d\bar z^\beta$ one obtains

$$\frac{\partial f^j_{\beta k}}{\partial \bar z^\alpha} - \frac{\partial f^j_{\alpha k}}{\partial \bar z^\beta} - \sum_l (f^j_{\beta l} f^l_{\alpha k} - f^j_{\alpha l} f^l_{\beta k}) = 0. \qquad \text{Q.E.D.}$$

Let us proceed to show that (X,A) has an underlying holomorphic vector bundle.

Lemma 2: Given $p \in X$, there is a neighborhood of p, U'', and functions h^k_j on U'' such that

$$\frac{\partial h^k_j}{\partial \bar z^\alpha} = \sum_l f^l_{\alpha j} h^k_l$$

for each j, k, α and such that the matrix h^k_j has full rank everywhere on U''. (See also Ref. 4.)

Proof: Let z^α, w^k be classical complex coordinates on $U \times \mathbf{C}^m$, where U is chosen as a coordinate neighborhood of p on which our $f^l_{\alpha j}$ are defined. Consider the vector fields

$$W^j = \frac{\partial}{\partial w^j}, \quad Z^\alpha = \frac{\partial}{\partial z^\alpha} - \sum_{l,k} f^k_{\alpha l}\, \bar w^l \frac{\partial}{\partial \bar w^k}.$$

Notice that $\mathrm{Re}(Z_\alpha)$, $\mathrm{Im}(Z_\alpha)$, $\mathrm{Re}(W_j)$, $\mathrm{Im}(W_j)$ form a basis for $T(U \times \mathbf{C}^m)$ at each point in U.

Define an almost complex structure $\mathbf{J}$ on $U \times \mathbf{C}^m$ by

$$\mathbf{J}(\mathrm{Re}\, Z_\alpha) = -\mathrm{Im}(Z_\alpha), \quad \mathbf{J}(\mathrm{Im}\, Z_\alpha) = \mathrm{Re}\, Z_\alpha,$$
$$\mathbf{J}(\mathrm{Re}\, W_j) = -\mathrm{Im}\, W_j, \quad \mathbf{J}(\mathrm{Im}\, W_j) = \mathrm{Re}\, W_j.$$

The condition that $\mathbf{J}$ is integrable is that $[W_j, W_k]$, $[W_j, Z_\alpha]$, $[Z_\alpha, Z_\beta]$ are in the span of W_k, Z_α. We have $[W_k, W_l] = [W_j, Z_\alpha] = 0$. When we calculate $[Z_\alpha, Z_\beta]$ making use of Eq. (8) and the fact that $f^j_{\alpha k}$ is independent of $\bar\omega^j$, we obtain $[Z_\alpha, Z_\beta] = 0$. By the classical Newlander–Nirenberg theorem, there are holomorphic coordinates v^q, $q = 1,\ldots,n+m$ on some neighborhood of $(p,0)$, $U \times V \subset U \times \mathbf{C}^m$. We can form a new holomorphic coordinate system out of $z^1,\ldots,z^n$ and a choice of m of the v^q's which we may take to be $v^1,\ldots,v^m$. We have $\bar W_l(v^q) = 0$, i.e., v^q can be written as a power series in w of the form

$$v^q = h^q(z,\bar z) + \sum_j h^q_j(z,\bar z) w^j + o((w)^2).$$

From $\bar Z_\alpha(v^q) = 0$ we have

$$\bar Z_\alpha(h^q) = \bar Z_\alpha\Big(\sum_j h^q_j \omega^j\Big) = 0,$$

which implies that

$$\frac{\partial h^k_j}{\partial \bar z^\alpha} = \sum_l f^l_{\alpha j} h^k_l.$$

In order for dz^α and dv^j to be linearly independent at $(p,0)$, the matrix h^q_j must be nonsingular in a neighborhood of p. Q.E.D.

Return to our supermanifold to see that

$$\sum_j h^k_j \phi^j = \sum_j h^k_j\, d\eta^j + \sum_{j,\alpha,l} h^k_j\, d\bar z^\alpha f^j_{\alpha l} \eta^l \bmod N^2\Omega^1$$
$$= \sum_j h^k_j\, d\eta^j + \sum_{\alpha,l} d\bar z^\alpha \frac{\partial h^k_l}{\partial \bar z^\alpha} \eta^l \bmod N^2\Omega^1$$
$$= \sum_j d(h^k_j \eta^j) - \sum_{\alpha,l} \theta^\alpha \frac{\partial h^k_l}{\partial z^\alpha} \eta^l \bmod N^2\Omega^1.$$

Thus set

$$\phi^{k\prime} = \sum_j h^k_h \phi^j + \sum_{\alpha,l} \theta^\alpha \frac{\partial h^k_l}{\partial z^\alpha} \eta^l$$

and

$$\eta^{k\prime} = \sum_j h^k_j \eta^j.$$

We then have

$$\phi^{k\prime} = d\eta^{k\prime} \bmod N^2\Omega^1.$$

Also rewrite our previous Eq. (7): $\theta^{\alpha\prime} = dz^{\alpha\prime} \bmod N^2\Omega^1$. We take $z^{\alpha\prime}, \eta^{k\prime}$ as local coordinates on our supermanifold (X,A) and hence write them without the primes.

We now show that N/N^2 is a sheaf of sections of a holomorphic vector bundle. Consider a change of odd coordinates on the intersection of two coordinate neighborhoods, $U \cap \hat U$,

$$\bar\eta^j = \sum_k b^j_k \eta^k + \sum_k c^j_k \bar\eta^k \bmod N^2.$$

Since $d\bar\eta^j = \hat\phi^j \bmod N^2\Omega^1$, where $\hat\phi^j$ is in the $+i$ eigenspace of J^* and thus in the span of ϕ^k and θ^α, we must have

$$\sum_\alpha d\bar z^\alpha \frac{\partial \hat n^k}{\partial \bar z^\alpha} + \sum_j d\bar\eta^j \frac{\partial \hat n^k}{\partial \bar\eta^j} = 0 \bmod N^2\Omega^1.$$

This produces $\partial b^j_k / \partial \bar z^\alpha = 0$, $\alpha = 1,\ldots,n$ and $c^j_k = 0$.

Thus $\hat{\eta}^j = \Sigma_k b^j_k \eta^k$, where b^j_k is a matrix of holomorphic functions, and we conclude that N/N^2 is a sheaf of sections of a holomorphic vector bundle.

Now consider terms of higher nilpotency.

Lemma 3: Fix $l \geqslant 2$. Assume there are one-forms θ^α, ϕ^j in the $+i$ eigenspace of J^*, and supercoordinate functions z^α, η^j such that

$$\theta^\alpha = dz^\alpha \bmod N^l\Omega^1,$$

$$\phi^j = d\eta^j \bmod N^l\Omega^1.$$

Then there are one-forms $\theta^{\alpha\prime}$, $\phi^{j\prime}$ in the $+i$ eigenspace of J^* and supercoordinate functions $z^{\alpha\prime}$, $\eta^{j\prime}$ such that

$$\theta^{\alpha\prime} = dz^{\alpha\prime} \bmod N^{l+1}\Omega^1,$$

$$\phi^{j\prime} = d\eta^{j\prime} \bmod N^{l+1}\Omega^1.$$

Proof: Consider the case in which l is even. The case in which l is odd is exactly similar with only minor changes. Make appropriate changes in θ^α, ϕ^j similar to those made above in Eqs. (1)–(6) so that

$$\theta^\alpha = dz^\alpha + \sum_\beta d\bar{z}^\beta E^\alpha_\beta + \sum_k d\bar{\eta}^k F^\alpha_k$$

and

$$\phi^j = d\eta^j + \sum_\beta d\bar{z}^\beta G^j_\beta + \sum_k d\bar{\eta}^k H^j_k,$$

where $E^\alpha_\beta, F^\alpha_k, G^j_\beta, H^j_k \in N^l$.

Expand θ^α, ϕ^j in local coordinates to the next order of nilpotency,

$$\theta^\alpha = dz^\alpha + \sum_{\beta,|I|=l} d\bar{z}^\beta \eta^I b^\alpha_{\beta,I} + \sum_{\beta,|I|=l} d\bar{z}^\beta \bar{\eta}^I c^\alpha_{\beta,I}$$

$$+ \sum_{\beta,|I|=l} d\bar{z}^\beta \eta^I \bar{\eta}^J d^\alpha_{\beta,I,J} \bmod N^{l+1}\Omega^1,$$

$$\phi^j = dn^j + \sum_{k,|I|=l} d\bar{\eta}^k \eta^I e^j_{k,I} \sum_{k,|I|=l} d\bar{\eta}^k \eta^I f^j_{k,I}$$

$$+ \sum_{k,|I+J|=l} d\bar{\eta}^k \eta^I \bar{\eta}^J g^j_{k,I,J} \bmod N^{l+1}\Omega^1.$$

Now $d\theta^\alpha = 0 \bmod \theta^\beta, \phi^k$ and $d\phi^j = 0 \bmod \theta^\beta, \phi^k$ requires $c^\alpha_{\beta,I} = d^\alpha_{\beta,I,J} = f^j_{k,I} = g^j_{k,I,J} = 0$ for all $\alpha, \beta, I, J, j, k$. Thus

$$\theta^\alpha = dz^\alpha + \sum_{\beta,|I|=l} d\bar{z}^\beta \eta^I b^\alpha_{\beta,I} \bmod N^{l+1}\Omega^1$$

and

$$\phi^j = d\eta^j + \sum_{k,|I|=l} d\bar{\eta}^k \eta^I e^j_{k,I}.$$

The lowest order terms in η of $d\theta^\alpha$ containing $d\bar{z}^\beta \wedge d\bar{z}^\nu$ are

$$\sum_{\beta,\nu} d\bar{z}^\beta \wedge d\bar{z}^\nu \eta^I \frac{\partial b^\alpha_{\beta,I}}{\partial \bar{z}^\nu}, \quad |I| = l.$$

This sum must be zero for each I and each α since $d\theta^\alpha = 0 \bmod \theta^\beta$, ϕ^j. This gives local $\bar{\partial}$ closed one-forms $\Sigma_\beta d\bar{z}^\beta b^\alpha_{\beta,I}$. By the Dolbeault lemma, there are complex functions h^α_I such that

$$\bar{\partial} h^\alpha_I = \sum \beta\, d\bar{z}^\beta b^\alpha_{\beta,I}.$$

Let $z^{\alpha\prime} = z^\alpha + \Sigma_{|I|=l} h^\alpha_I \eta^I$ and

$$\theta^{\alpha\prime} = \theta^\alpha + \sum_{\beta,|I|=l} \theta^\beta \frac{\partial h^\alpha_I}{\partial z^\beta} \eta^I$$

$$- \sum_{k,|I|=l-1} (-1)^{\epsilon_{I,k}} \phi^k h^\alpha_{I,k} \eta^I,$$

where $\epsilon_{I,k}$ is 0 or 1 depending only on I and k. Then

$$\theta^{\alpha\prime} = dz^{\alpha\prime} \bmod N^{l+1}\Omega^1.$$

Let $\eta^{j\prime} = \eta^j + \Sigma_{k,I} \bar{\eta}^k \eta^I e^j_{k,I}$ and

$$\phi^{j\prime} = \phi^j + \sum_{i,k,|I|=l-1} (-1)^{\epsilon_{I,i}} \phi^i \bar{\eta}^k \eta^I e^j_{k,I,i}.$$

Then

$$\phi^{j\prime} = d\eta^{j\prime} \bmod N^{l+1}\Omega^1.$$

Since $N^{2m+1} = 0$, a finite number of applications of this lemma produces supercoordinate functions z^α and η^j such that dz^α and $d\eta^j$ are a basis for the $+i$ eigenspace of the almost complex structure, J^*. Q.E.D.

The above proof may be modified to give a proof of the Frobenius theorem for supermanifolds. As in the classical case, the two theorems, Frobenius and Newlander–Nirenberg, are related. We refer the reader to Ref. 5 for the Frobenius theorem on supermanifolds.

Note added in proof: Dirmitri Leites has pointed out that a different proof of the above theorem has been given by A. Yu. Weintrob and is to appear in an English translation.

ACKNOWLEDGMENT

I would like to thank Professor Claude LeBrun for his encouragement and guidance during the writing of this paper.

[1]A. Newlander and L. Nirenberg, "Complex analytic coordinates in almost complex manifolds," Ann. Math. **65**, 391 (1957).

[2]B. Kostant, *Graded Manifolds, Graded Lie Algebras, and Prequantization, Lecture Notes in Mathematics,* Vol. 570 (Springer, Berlin, 1977), pp. 177-300; *Differential-Geometric Methods in Physics* (Springer, Heidelberg, 1988).

[3]S. S. Chern, *Complex Manifolds Without Potential Theory* (Springer, New York, 1979).

[4]M. F. Atiyah, *Geometry of Yang–Mills Fields* (Accademia Nazionale Dei Lincei, Scuola Normale Superiore, Lezioni Fermiane, Pisa, 1979), p. 46.

[5]Yu. I. Manin, *Gauge Fields and Complex Geometry,* English translation (Springer, Heidelberg, 1988).

Integrable dynamical systems with hierarchy. II. Solutions

Susumu Okubo
Department of Physics and Astronomy, University of Rochester, Rochester, New York 14627

(Received 15 November 1988; accepted for publication 11 January 1989)

A semisystematic method of finding solutions that satisfy the zero Nijenhuis tensor condition with dual symplectic structure is developed. In this manner, a few solutions, including that of the Toda lattice, have been found. It is also shown that the general solution for the linear problem is intimately connected with nontriviality of the second cohomology group of solvable Lie algebras. Moreover, the case of the quadratic problem requires studies of an algebra involving the triple-linear product system.

I. INTRODUCTION

In a previous paper[1] we have proved that any dual symplectic manifold with zero Nijenhuis tensor possesses many interesting properties. Especially, we can construct an infinite number of conserved quantities in involution which satisfy the hierarchy equation. Moreover, the dual symplectic manifold admits an infinite series of Poisson brackets and Lagrangians that nevertheless yield the same equation of motion. However, it is in practice a rather difficult task to find physically realistic solutions satisfying the conditions, although both Korteweg–deVries (KdV) and Toda lattices are known[2,3] to be such examples. In this paper, we restrict ourselves to consideration only of a finite system for the sake of definiteness and present a semisystematic method of finding possible solutions. In this way, we find a few solutions, including the case of the Toda lattice.

First, we will show in Sec. II that the simplest linear problem is reduced to a study of the second cohomology group of solvable Lie algebras. In contrast, the case of the general quadratic problem will lead to a complicated triple-product system, as we will explain in Sec. V. However, we will concentrate most of our effort in this paper to the simple practical study of finding physically reasonable integrable models, as shown in Secs. III and IV. Some of these solutions, including the Toda lattice, are found to possess an extra affine structure with zero Rieman curvature, but nonzero torsion tensors.

Before going into the details of these facts, we will briefly sketch some results of Ref. 1 in order to be as self-contained as possible and establish notations. Let x^μ $(\mu = 1,2...,D)$ with $D = 2N$ be a local coordinate system in the $2N$-dimensional symplectic manifold M, with two symplectic forms given by

$$f = \tfrac{1}{2} f_{\mu\nu}(x)\,dx^\mu \wedge dx^\nu, \tag{1.1}$$

$$F = \tfrac{1}{2} F_{\mu\nu}(x)\,dx^\mu \wedge dx^\nu, \tag{1.2}$$

so that we have

$$\partial_\lambda f_{\mu\nu} + \partial_\mu f_{\nu\lambda} + \partial_\nu f_{\lambda\mu} = 0\,, \tag{1.3}$$

$$\partial_\lambda F_{\mu\nu} + \partial_\mu F_{\nu\lambda} + \partial_\nu F_{\lambda\mu} = 0. \tag{1.4}$$

We write inverses of $f_{\mu\nu}$ and $F_{\mu\nu}$ as $f^{\mu\nu}$ and $F^{\mu\nu}$, respectively, so that

$$f^{\mu\lambda} f_{\lambda\nu} = \delta^\mu_\nu, \tag{1.5}$$

$$F^{\mu\lambda} F_{\lambda\nu} = \delta^\mu_\nu. \tag{1.6}$$

In this paper, the repeated lower case Greek indices are understood to imply automatical summations on the $2N$ values $1,2,...,2N$. We now introduce the mixed tensor $S_\mu^{\ \nu}$ by

$$S_\mu^{\ \nu} = F_{\mu\lambda} f^{\lambda\nu} \tag{1.7}$$

and define the Nijenhuis tensor by

$$N^{\lambda}_{\mu\nu} = -N^{\lambda}_{\nu\mu} = S_\mu^{\ \alpha}\,\partial_\alpha S_\nu^{\ \lambda} - S_\nu^{\ \alpha}\partial_\alpha S_\mu^{\ \lambda} - S_\alpha^{\ \lambda}(\partial_\mu S_\nu^{\ \alpha} - \partial_\nu S_\mu^{\ \alpha}). \tag{1.8}$$

Next, as in Ref. 1, we construct the nth power antisymmetric tensor $(F^n)_{\mu\nu}$ inductively by

$$(F^{n+1})_{\mu\nu} = F_{\mu\alpha} f^{\alpha\beta} (F^n)_{\beta\nu}, \tag{1.9}$$

with

$$(F^0)_{\mu\nu} = f_{\mu\nu} \tag{1.10}$$

for $n \geqslant 0$. For negative n, we set

$$(F^{-1})_{\mu\nu} = f_{\mu\alpha} F^{\alpha\beta} f_{\beta\nu} \tag{1.11}$$

and similarly define $(F^{-n})_{\mu\nu}$. Then, we find

$$(S^n)_\mu^{\ \nu} = (F^n)_{\mu\lambda} f^{\lambda\nu} = f_{\mu\lambda} (F^{-n})^{\lambda\nu}, \tag{1.12}$$

as in Ref. 1.

Now, we set

$$K_n = (1/2n)\operatorname{Tr} S^n = (1/2n)(S^n)_\mu^{\ \mu} \quad (n \neq 0), \tag{1.13a}$$

$$K_0 = \tfrac{1}{2}\log(\det S) \quad (n = 0). \tag{1.13b}$$

If the Nijenhuis tensor $N^\lambda_{\mu\nu}$ is identically zero, we have proved the following results in Ref. 1. First, K_n satisfies the hierarchy equation

$$S_\mu^{\ \lambda}\,\partial_\lambda K_n = \partial_\mu K_{n+1} \tag{1.14}$$

for any integer n. Second, the antisymmetric tensor $(F^n)_{\mu\nu}$ also satisfies the symplectic condition

$$\partial_\lambda (F^n)_{\mu\nu} + \partial_\mu (F^n)_{\nu\lambda} + \partial_\nu (F^n)_{\lambda\mu} = 0, \tag{1.15}$$

again for arbitrary integer values of n. As a result of Eq. (1.15), we can introduce an infinite series of Poisson brackets by

$$\{h,g\}_n = (F^n)^{\mu\nu}\,\partial_\mu h\,\partial_\nu g \tag{1.16}$$

for any two functions $h = h(x)$ and $g = g(x)$: Especially, the K_n are in involution with respect to any one of these brackets, i.e., we have

$$\{K_n, K_m\}_p = 0 \tag{1.17}$$

for any three integers n, m, and p. Moreover, all K_m can be

 0022-2488/89/051043-10$02.50

regarded as conserved quantities of the infinite series of Lagrangians

$$L_p^{(n)} = \theta_\mu^{(n)}(x)\frac{dx^\mu}{dt_p} - K_{n+p}(x) \quad (n = 0, \pm 1, \pm 2, \ldots) \tag{1.18}$$

for a fixed value of p with the respective time variable t_p, where $\theta_\mu^{(n)}(x)$ is defined by

$$(F^n)_{\mu\nu} = \partial_\mu \theta_\nu^{(n)} - \partial_\nu \theta_\mu^{(n)} \tag{1.19}$$

in view of Eq. (1.15). As a result of the hierarchy equation (1.14), all Lagrangians $L_p^{(n)}$ $(n = 0, \pm 1, \pm 2, \ldots)$ lead to the same single equation of motion

$$\frac{dx^\mu}{dt_p} = f^{\mu\nu}\partial_\nu K_p, \tag{1.20}$$

irrespective of the values of n. Therefore, the infinite series of Lagrangians $L_p^{(n)}$ $(n = 0, \pm 1, \pm 2, \ldots)$ are equivalent to each other. Finally, we have

$$\frac{d}{dt_p}K_n = \{K_n, K_p\}_0 = 0, \tag{1.21}$$

so that the K_n are indeed the conserved quantities of the Hamilton system. Therefore, if we can find N algebraically independent terms among the K_n, then the system is integrable in view of the Liouville theorem. As we noted elsewhere,[3] the Toda lattice satisfies all of these criteria.

The remaining problem that is the major concern of the present paper is a practical one: how we proceed to discover models satisfying the zero Nijenhuis tensor condition. In Ref. 1, we have also shown that $N_{\mu\nu}^\lambda = 0$ is equivalent to the validity of

$$\partial_\lambda (F\cdot F)_{\mu\nu} + \partial_\mu (F\cdot F)_{\nu\lambda} + \partial_\nu (F\cdot F)_{\lambda\mu} = 0, \tag{1.22}$$

where we have set

$$(F\cdot F)_{\mu\nu} = (F^2)_{\mu\nu} = F_{\mu\alpha} f^{\alpha\beta} F_{\beta\nu}, \tag{1.23}$$

provided that Eqs. (1.3) and (1.4) hold. Especially, the general validity of Eq. (1.15) is remarkably a consequence of its two special cases $n = 1$ and $n = 2$, provided that the two-form f is symplectic. Now, by the Darboux theorem,[4] there exists a local coordinate frame, called the canonical coordinate system, in which all $f_{\mu\nu}$ are constants. More specifically, we can set

$$q_j = x^j, \tag{1.24a}$$

$$p_j = x^{j+N} = x^{\bar{j}} \tag{1.24b}$$

for $j = 1, 2, \ldots, N$, with the canonical form of

$$f = \sum_{j=1}^N dp_j \wedge dq_j. \tag{1.25}$$

Especially, we have

$$f_{jk} = f_{\bar{j}\bar{k}} = 0, \tag{1.26a}$$

$$f_{j\bar{k}} = -f_{\bar{j}k} = -\delta_{jk}, \tag{1.26b}$$

as well as

$$f^{jk} = f^{\bar{j}\bar{k}} = 0, \tag{1.27a}$$

$$f^{j\bar{k}} = -f^{\bar{j}k} = \delta_{jk}, \tag{1.27b}$$

where for simplicity we have set

$$\bar{k} = N + k \tag{1.28}$$

for any lower case Latin index k which will assume the N values $1, 2, \ldots, N$. Hereafter in this paper, we will work only in the canonical coordinate system unless stated otherwise. Then, the problem is reduced to finding solutions only of Eqs. (1.4) and (1.22) since Eq. (1.3) is now trivially satisfied. For later convenience, we set

$$\Delta_{\lambda\mu\nu}(G) = \partial_\lambda G_{\mu\nu} + \partial_\mu G_{\nu\lambda} + \partial_\nu G_{\lambda\mu} \tag{1.29}$$

for any antisymmetric tensor $G_{\mu\nu} = -G_{\nu\mu}$. Then, $\Delta_{\lambda\mu\nu}(G)$ is totally antisymmetric in λ, μ, and ν. In the following sections, we will develop a semisystematic way of finding solutions of the following desired equations:

$$\Delta_{\lambda\mu\nu}(F) = \Delta_{\lambda\mu\nu}(F\cdot F) = 0; \tag{1.30}$$

in general, we need not worry about the existence of $F^{\mu\nu}$ for solving them. If $\det F_{\mu\nu} = 0$, then we may use $\widetilde{F}_{\mu\nu} = F_{\mu\nu} + Cf_{\mu\nu}$ for any constant C instead of $F_{\mu\nu}$ since $\widetilde{F}_{\mu\nu}$ evidently also satisfies Eq. (1.30). We may then take the limit $C \to 0$ at the end of the calculation if the limit exists.

II. LINEAR CASE

In this section, we will study the simplest solution for $F_{\mu\nu} = F_{\mu\nu}(x)$ as linear functions of x^λ in the canonical frame. However, for a while, we assume only that $f_{\mu\nu}$ are constants.

We set

$$F_{\mu\nu} = h_{\mu\nu} + C_{\mu\nu}^\alpha f_{\alpha\beta} x^\beta \tag{2.1}$$

for some constants $h_{\mu\nu}$ and $C_{\mu\nu}^\alpha$ satisfying

$$h_{\mu\nu} = -h_{\nu\mu}, \tag{2.2}$$

$$C_{\mu\nu}^\lambda = -C_{\nu\mu}^\lambda. \tag{2.3}$$

Then, the condition $\Delta_{\lambda\mu\nu}(F) = 0$ is clearly equivalent to the validity of

$$C_{\mu\nu}^\alpha f_{\alpha\lambda} + C_{\nu\lambda}^\alpha f_{\alpha\mu} + C_{\lambda\mu}^\alpha f_{\alpha\nu} = 0. \tag{2.4}$$

Next, we calculate

$$\begin{aligned}\Delta_{\lambda\mu\nu}(F\cdot F) = {} & \{(C_{\nu\alpha}^\theta f_{\theta\lambda} + C_{\alpha\lambda}^\theta f_{\theta\nu})h_{\mu\beta} + (C_{\lambda\alpha}^\theta f_{\theta\mu} + C_{\alpha\mu}^\theta f_{\theta\lambda})h_{\nu\beta} \\ & + (C_{\mu\alpha}^\theta f_{\theta\nu} + C_{\alpha\nu}^\theta f_{\theta\mu})h_{\lambda\beta}\}f^{\alpha\beta} + \{(C_{\mu\alpha}^\theta f_{\theta\lambda} + C_{\alpha\lambda}^\theta f_{\theta\mu})C_{\beta\nu}^\gamma \\ & + (C_{\nu\alpha}^\theta f_{\theta\mu} + C_{\alpha\mu}^\theta f_{\theta\nu})C_{\beta\lambda}^\gamma + (C_{\lambda\alpha}^\theta f_{\theta\nu} + C_{\alpha\nu}^\theta f_{\theta\lambda})C_{\beta\mu}^\gamma\}f^{\alpha\beta} f_{\gamma\tau} x^\tau.\end{aligned}$$

When we utilize Eq. (2.4), the above relation can be simplified to become

$$\Delta_{\lambda\mu\nu}(F\cdot F) = C^{\beta}_{\nu\lambda} h_{\mu\beta} + C^{\beta}_{\lambda\mu} h_{\nu\beta} + C^{\beta}_{\mu\nu} h_{\lambda\beta} - \{C^{\beta}_{\lambda\mu} C^{\gamma}_{\beta\nu} + C^{\beta}_{\mu\nu} C^{\gamma}_{\beta\lambda} + C^{\beta}_{\nu\lambda} C^{\gamma}_{\beta\mu}\} f_{\gamma\tau} x^{\tau}.$$

Therefore, the condition $\Delta_{\lambda\mu\nu}(F\cdot F) = 0$ requires

$$C^{\beta}_{\nu\lambda} h_{\mu\beta} + C^{\beta}_{\lambda\mu} h_{\nu\beta} + C^{\beta}_{\mu\nu} h_{\lambda\beta} = 0 \tag{2.5}$$

and

$$C^{\beta}_{\lambda\mu} C^{\gamma}_{\beta\nu} + C^{\beta}_{\mu\nu} C^{\gamma}_{\beta\lambda} + C^{\beta}_{\nu\lambda} C^{\gamma}_{\beta\mu} = 0. \tag{2.6}$$

Especially, we recognize Eq. (2.6) as the Jacobi identity of a $2N$-dimensional Lie algebra L. When we introduce $2N$ tangent vectors X_μ in M by

$$X_\mu = x^\alpha C^{\beta}_{\alpha\mu}\, \partial_\beta, \tag{2.7}$$

then Eq. (2.6) is equivalently rewritten as

$$[X_\mu, X_\nu] = C^{\lambda}_{\mu\nu} X_\lambda. \tag{2.8}$$

Equations (2.4) and (2.5) can now be analyzed as follows: Since both equations have exactly the same form, we will consider Eq. (2.4). We introduce a bilinear antisymmetric functional $\langle X, Y\rangle$ in the tangent space T_M by

$$\langle X_\mu, X_\nu\rangle = f_{\mu\nu}. \tag{2.9}$$

Since $f_{\mu\nu}$ is nondegenerate, $\langle X, Y\rangle$ is also nondegenerate. Then, Eq. (2.4) is rewritten to be

$$\langle [X_\mu, X_\nu], X_\lambda\rangle + \langle [X_\nu, X_\lambda], X_\mu\rangle + \langle [X_\lambda, X_\mu], X_\nu\rangle = 0. \tag{2.10}$$

Equation (2.10) is precisely the two-cocycle condition[5] of L in a trivial representation space V_0 of the Lie algebra L. Therefore, both $f_{\mu\nu}$ and $h_{\mu\nu}$ are components of two cocycles of L. First, we will prove that L cannot be semisimple. Suppose contrarily that L is semisimple. Then, the second cohomology group $H^2(L,V)$ is identically zero[5,6] for any finite-dimensional representation space V of L. Therefore, any two-cocycle is exact, so that there exists a one-cochain ϕ_1 satisfying

$$f_{\mu\nu} = \langle X_\mu, X_\nu\rangle = (\delta\phi_1)(X_\mu, X_\nu) = \phi_1([X_\mu, X_\nu]) = C^{\lambda}_{\mu\nu}\phi_1(X_\lambda). \tag{2.11}$$

Since L is assumed to be semisimple, the Killing form $g_{\mu\nu} = \mathrm{Tr}(\mathrm{ad}\, X_\mu \mathrm{ad}\, X_\nu)$ is nondegenerate with its inverse $g^{\mu\nu}$. Moreover,

$$f_{\mu\nu\lambda} = C^{\alpha}_{\mu\nu} g_{\alpha\lambda}$$

is totally antisymmetric in the three indices μ, ν, and λ. Hence, if we set

$$\xi^\lambda = g^{\lambda\alpha}\phi_1(X_\alpha),$$

then Eq. (2.11) is rewritten as

$$f_{\mu\nu} = f_{\mu\nu\lambda}\xi^\lambda,$$

so that we have

$$f_{\mu\nu}\xi^\nu = f_{\mu\nu\lambda}\xi^\nu\xi^\lambda = 0.$$

Multiplying $f^{\lambda\mu}$ to the above equation, we find $\xi^\lambda = 0$, which leads to $\phi_1(X_\lambda) = 0$. However, Eq. (2.11) then implies $f_{\mu\nu} = 0$, which is a contradiction. This proves that L cannot be semisimple. Also, if $H^2(L, V_0) = 0$, we can replace $f_{\mu\nu}$ by $h_{\mu\nu}$ in Eq. (2.11). Then, we can effectively set $h_{\mu\nu} = 0$ in Eq. (2.1) by making a coordinate translation

$$x^\mu \to x'^\mu = x^\mu + b^\mu$$

for some constants b^μ. Therefore, nontriviality of $h_{\mu\nu}$ necessitates the nonzero second cohomology group $H^2(L, V_0)$.

Moreover, let

$$L = L_S + L_R$$

be the Levi decomposition of L into its semisimple part L_S and radical L_R. Then, we know[6] that $H^2(L,V_0)$ is isomorphic to a subspace of $H^2(L_R, V_0)$ and that

$$\langle [X_1, Y], X_2\rangle = \langle X_1, [Y, X_2]\rangle$$

for $X_1, X_2 \in L_R$, and $Y \in L_S$.

In conclusion, the Lie algebra L is such that the symplectic functional $f_{\mu\nu}$ is its two-cocycle and $H^2(L,V_0)$ must be nonzero in order to have nontrivial $h_{\mu\nu}$. Especially, L cannot be semisimple. Moreover, since $h_{\mu\nu}$ is a two-cocycle of L, the central extension[5,6] of the Lie algebra L is also a Lie algebra $\hat{L}$ defined by

$$[Y_\mu, Y_\nu] = C^{\lambda}_{\mu\nu} Y_\lambda + h_{\mu\nu} 1, \tag{2.12a}$$

$$[Y_\mu, 1] = 0. \tag{2.12b}$$

Actually, $\hat{L}$ is isomorphic to a finite sub-Lie algebra of a general Poisson bracket algebra defined by [see Eq. (1.16) for $n = -1$]

$$\{h,g\}_{-1} = (F^{-1})^{\mu\nu}\partial_\mu h\, \partial_\nu g = -F_{\mu\nu}\, \partial^\mu h\, \partial^\nu g, \tag{2.13}$$

where we use

$$(F^{-1})^{\mu\nu} = f^{\mu\alpha} F_{\alpha\beta} f^{\beta\nu}$$

and set

$$\partial^\mu h = f^{\mu\lambda}\partial_\lambda h. \tag{2.14}$$

For the present solution $F_{\mu\nu}$ given by Eq. (2.1), we calculate especially

$$\{x_\mu, x_\nu\}_{-1} = -h_{\mu\nu} - C^{\lambda}_{\mu\nu} x_\lambda \tag{2.15}$$

when we set

$$x_\mu = f_{\mu\lambda} x^\lambda. \tag{2.16}$$

Clearly, the Lie algebra specified by Eq. (2.15) with the generators x_μ and 1 is isomorphic to $\hat{L}$ given by Eq. (2.12) since $\{x_\mu, 1\}_{-1} = 0$ as well. The possible relevance of Eq. (2.15) for the KdV equation and Virassoro algebra will be discussed elsewhere.

In this paper, we consider only cases of L as a solvable Lie algebra without any semisimple part. Moreover, on the physical ground, we assume that $F_{\mu\nu}$ depends upon the momentum variables p_j, but not upon the space coordinate q_j. Then, assuming hereafter the canonical coordinate frame, we can readily find a special class of solutions of the form

$$[X_j, X_k] = 0, \tag{2.17a}$$

$$[X_j, X_{\bar{k}}] = \delta_{jk}\xi_j X_j, \tag{2.17b}$$

$$[X_{\bar{j}}, X_{\bar{k}}] = \xi_{jk} X_j - \xi_{kj} X_k, \tag{2.17c}$$

or equivalently,

$$C^{l}_{j\bar{k}} = -C^{l}_{\bar{k}j} = \delta_{jk}\, \delta_{jl}\, \xi_l, \tag{2.18a}$$

$$C^{l}_{\bar{j}\bar{k}} = \delta_{lj}\, \xi_{jk} - \delta_{lk}\, \xi_{kj}, \tag{2.18b}$$

$$C^{l}_{jk} = C^{\bar{l}}_{jk} = C^{l}_{j\bar{k}} = C^{\bar{l}}_{j\bar{k}} = 0, \tag{2.18c}$$

where ξ_j and ξ_{jk} are arbitrary constants. We can easily verify the fact that Eqs. (2.4) and (2.6) are automatically satisfied by Eq. (2.18). The present Lie algebra L is solvable, but not nilpotent. Also, $X_1, X_2, \ldots, X_N$ generate an Abelian ideal of L. Now, Eq. (2.5) reduces to

$$\xi_j h_{jk} = 0, \tag{2.19a}$$

$$\xi_{jk} h_{jl} - \xi_{kj} h_{kl} = \delta_{kl}\, \xi_k h_{l\bar{j}} - \delta_{lj}\, \xi_j h_{l\bar{k}}, \tag{2.19b}$$

$$\xi_{jk} h_{j\bar{l}} + \xi_{kl} h_{k\bar{j}} + \xi_{lj} h_{l\bar{k}} = \xi_{kj} h_{k\bar{l}} + \xi_{lk} h_{l\bar{j}} + \xi_{jl} h_{j\bar{k}}. \tag{2.19c}$$

Here, we may remark that no summations are implied for repeated Latin indices, in contrast to the cases of repeated Greek indices. In this paper, we will be content with the following special solution of Eqs. (2.19):

$$h_{jk} = 0, \tag{2.20a}$$

$$h_{j\bar{k}} = -h_{\bar{k}j} = \eta_j\, \delta_{jk}, \tag{2.20b}$$

$$h_{\bar{j}\bar{k}} = -h_{\bar{k}\bar{j}}, \tag{2.20c}$$

where η_j, as well as $h_{\bar{j}\bar{k}}$, are arbitrary constants as long as $h_{\bar{j}\bar{k}}$ satisfies the antisymmetry condition (2.20c). Note that this allows nontrivial $H^2(L,V_0)$.

Combining Eqs. (2.1), (2.18), and (2.20), we have

$$F_{j\bar{k}} = -S^{k}_{j} = -S^{\bar{k}}_{\bar{j}} = \delta_{jk}\,(\eta_j - \xi_j p_j), \tag{2.21a}$$

$$F_{\bar{j}\bar{k}} = -S^{k}_{\bar{j}} = h_{\bar{j}\bar{k}} - \xi_{jk} p_j + \xi_{kj} p_k, \tag{2.21b}$$

$$F_{jk} = S^{\bar{k}}_{j} = 0. \tag{2.21c}$$

Therefore, we calculate

$$K_1 = \frac{1}{2}\,\mathrm{Tr}\, S = \sum_{j=1}^{N} (\xi_j p_j - \eta_j), \tag{2.22a}$$

$$K_2 = \frac{1}{4}\,\mathrm{Tr}\, S^2 = \frac{1}{2}\sum_{j=1}^{N} (\xi_j p_j - \eta_j)^2. \tag{2.22b}$$

We note that all K_n are functions of only the p_j and independent of ξ_{jk} and $h_{\bar{j}\bar{k}}$. Especially, if we choose

$$\eta_j = 0, \quad \xi_j = 1, \tag{2.23}$$

then K_1 and K_2 represent the total momentum and total energy of a free N-body system in the language of the canonical Hamiltonian formalism. This is the reason why we have chosen $F_{\mu\nu}$ to be independent of the space coordinates q_j. At any rate, we may say that the present solution corresponds to a free particle system.

As noted in Sec. I, once we find a solution of the problem, then $\widetilde{F}_{\mu\nu} = (F^n)_{\mu\nu}$ is also a solution for any integer value of n since in view of Eq. (1.15) it also satisfies

$$\Delta_{\lambda\mu\nu}(\widetilde{F}) = \Delta_{\lambda\mu\nu}(\widetilde{F}\cdot\widetilde{F}) = 0. \tag{2.24}$$

Especially, if we choose $n = 2$, Eq. (2.24) will give us a solution quadratic in the momentum. However, the solution is really not new since its conserved quantity $\widetilde{K}_1 = \frac{1}{2}\,\mathrm{Tr}\,\widetilde{S}$ is precisely the same as $2K_2 = \frac{1}{2}\,\mathrm{Tr}\, S^2$ of the old theory. We have also found a similar quadratic solution which is, however, not equivalent to the one discussed above: It is given by

$$F_{j\bar{k}} = [\eta_j - \xi_j (p_j)^2]\,\delta_{jk}, \tag{2.25a}$$

$$F_{\bar{j}\bar{k}} = h_{\bar{j}\bar{k}} - \xi_{jk} p_j + \xi_{kj} p_k, \tag{2.25b}$$

$$F_{jk} = 0, \tag{2.25c}$$

where η_j, ξ_j, ξ_{jk}, and $h_{\bar{k}\bar{j}}\,(= -h_{\bar{j}\bar{k}})$ are again arbitrary constants. This solution will play a certain role in Sec. IV. Also, a general study of the quadratic solution will be given in Sec. V.

Concluding this section, we remark the following. We noted that we can construct an infinite number of Poisson brackets once a solution $F_{\mu\nu}$ of our problem is given. For example, for the special solution (2.21) we can construct the ordinary bracket with $n = 0$,

$$\{h,g\}_0 = \sum_{j=1}^{N}\left(\frac{\partial h}{\partial q_j}\frac{\partial g}{\partial p_j} - \frac{\partial h}{\partial p_j}\frac{\partial g}{\partial q_j}\right), \tag{2.26}$$

as well as, for example, the following unconventional bracket with $n = -1$:

$$\{h,g\}_{-1} = \sum_{j=1}^{N} (\xi_j p_j - \eta_j)\left\{\frac{\partial h}{\partial q_j}\frac{\partial g}{\partial p_j} - \frac{\partial h}{\partial p_j}\frac{\partial g}{\partial q_j}\right\} + \sum_{j,k=1}^{N} (\xi_{jk} p_j - \xi_{kj} p_k - h_{\bar{j}\bar{k}})\frac{\partial h}{\partial q_j}\frac{\partial g}{\partial q_k} \tag{2.27}$$

from Eq. (2.12). We can directly verify the validity of the Jacobi identity,

$$\Big\{\{h,g\}_{-1},f\Big\}_{-1} + \Big\{\{g,f\}_{-1},h\Big\}_{-1} + \Big\{\{f,h\}_{-1},g\Big\}_{-1} = 0, \tag{2.28}$$

for three functions f, g, and h from Eq. (2.27); especially, we can dispense with the assumption of the existence of $F^{\mu\nu}$ for its validity. This fact can also be shown by using $\widetilde{F}_{\mu\nu} = F_{\mu\nu} + Cf_{\mu\nu}$ for a small constant C and letting $C \to 0$ if necessary. We remark here that Eq. (2.27) gives the following unconventional Poisson bracket relations:

$$\{p_j,p_k\}_{-1} = 0, \tag{2.29a}$$

$$\{p_j,q_k\}_{-1} = (\xi_j p_j - \eta_j)\delta_{jk}, \tag{2.29b}$$

$$\{q_j,q_k\}_{-1} = \xi_{jk} p_j - \xi_{kj} p_k - h_{\bar{j}\bar{k}}, \tag{2.29c}$$

among the canonical variables p_j and q_j. Together with the constant unit function 1, Eqs. (2.29) define a Lie algebra which is a central extension of the solvable Lie algebra L defined by Eq. (2.17), as we have already remarked. If we choose $\eta_j = -1$, but $\xi_j = \xi_{jk} = h_{\bar{j}\bar{k}} = 0$, then we note that Eq. (2.27) reduces to Eq. (2.26). Finally, to be complete, we calculate the inverse $F^{\mu\nu}$ for the solution (2.21) to be given by

$$F^{j\bar{k}} = -F^{\bar{k}j} = -\delta_{jk}/(\eta_j - \xi_j p_j), \tag{2.30a}$$

$$F^{jk} = (h_{\bar{j}\bar{k}} - \xi_{jk} p_j + \xi_{kj} p_k)/(\eta_j - \xi_j p_j)(\eta_k - \xi_k p_k), \tag{2.30b}$$

$$F^{\bar{j}\bar{k}} = 0. \tag{2.30c}$$

We can readily verify

$$F_{\mu\lambda} F^{\lambda\nu} = \delta^{\nu}_{\mu}.$$

Then, the Poisson bracket for $n = 1$ is given by

$$\begin{aligned}\{h,g\}_1 &= F^{\mu\nu}\,\partial_\mu h\,\partial_\nu g \\ &= \sum_{j,k=1}^{N} \frac{h_{\bar{j}\bar{k}} - \xi_{jk} p_j + \xi_{kj} p_k}{(\eta_j - \xi_j p_j)(\eta_k - \xi_k p_k)}\frac{\partial h}{\partial q_j}\frac{\partial g}{\partial q_k} \\ &\quad - \sum_{j=1}^{N} \frac{1}{\eta_j - \xi_j p_j}\left\{\frac{\partial h}{\partial q_j}\frac{\partial g}{\partial p_j} - \frac{\partial h}{\partial p_j}\frac{\partial g}{\partial q_j}\right\}.\end{aligned} \tag{2.31}$$

Also, as proved in Ref. 1, the arbitrary linear combination

$$\{h,g\} = C_0\{h,g\}_0 + C_1\{h,g\}_1 + C_{-1}\{h,g\}_{-1} \tag{2.32}$$

also defines a Lie algebra for the arbitrary constants C_0, C_1, and C_{-1}.

III. TODA LATTICE AND MOMENTUM-DEPENDENT TODA LATTICE

In Sec. II we have solved the problem for the linear case. In order to obtain physically more relevant solutions, we must add some interaction terms to the free energy solution. This is in general a rather difficult task, but we proceed as follows. Let us suppose that we have found two solutions $H_{\mu\nu}$ and $G_{\mu\nu}$ satisfying

$$\Delta_{\lambda\mu\nu}(H) = \Delta_{\lambda\mu\nu}(H\cdot H) = 0, \tag{3.1a}$$

$$\Delta_{\lambda\mu\nu}(G) = \Delta_{\lambda\mu\nu}(G\cdot G) = 0; \tag{3.1b}$$

then, their sum

$$F_{\mu\nu} = H_{\mu\nu} + G_{\mu\nu} \tag{3.2}$$

will also satisfy

$$\Delta_{\lambda\mu\nu}(F) = \Delta_{\lambda\mu\nu}(F\cdot F) = 0, \tag{3.3}$$

provided that we have the constraint

$$f^{\alpha\beta}\{H_{\mu\alpha}\,\partial_\beta G_{\nu\lambda} + H_{\nu\alpha}\,\partial_\beta G_{\lambda\mu} + H_{\lambda\alpha}\,\partial_\beta G_{\mu\nu} + G_{\mu\alpha}\,\partial_\beta H_{\nu\lambda} + G_{\nu\alpha}\,\partial_\beta H_{\lambda\mu} + G_{\lambda\alpha}\,\partial_\beta H_{\mu\nu}\} = 0 \tag{3.4}$$

when we utilize $\Delta_{\lambda\mu\nu}(H) = \Delta_{\lambda\mu\nu}(G) = 0$.

Initially, it would appear that this strategy would not offer any advantage at all. However, as we will see shortly, it is indeed the case. First, we have already found the linear solution in Sec. II which we identify with $H_{\mu\nu}$, so that

$$F_{\mu\nu} = h_{\mu\nu} + C^{\theta}_{\mu\nu} f_{\theta\tau} x^\tau + G_{\mu\nu}. \tag{3.5}$$

Then, Eq. (3.4) is also satisfied if we have

$$f^{\alpha\beta}\{C^{\theta}_{\mu\alpha}\,\partial_\beta G_{\nu\lambda} + C^{\theta}_{\nu\alpha}\,\partial_\beta G_{\lambda\mu} + C^{\theta}_{\lambda\alpha}\,\partial_\beta G_{\mu\nu}\} = 0 \tag{3.6}$$

and

$$C^{\beta}_{\lambda\mu} G_{\nu\beta} + C^{\beta}_{\mu\nu} G_{\lambda\beta} + C^{\beta}_{\nu\lambda} G_{\mu\beta} = f^{\alpha\beta}\{h_{\mu\alpha}\,\partial_\beta G_{\nu\lambda} + h_{\nu\alpha}\,\partial_\beta G_{\lambda\mu} + h_{\lambda\alpha}\,\partial_\beta G_{\mu\nu}\}, \tag{3.7}$$

both of which are now linear in $G_{\mu\nu}$. We note that Eq. (3.6) restricts the possible choices of $C^{\lambda}_{\mu\nu}$, while Eq. (3.7) will then determine the final form of $h_{\mu\nu}$.

The next crucial task is to find some solutions of $G_{\mu\nu}$ satisfying Eq. (3.1): This is really not difficult since we can choose some physically unrealistic solutions by themselves. We will explain this fact below. We consider first the solution corresponding to the Toda lattice.

A. Toda lattice solution

Let us suppose that we have

$$G_{\bar{k}j} = G_{j\bar{k}} = G_{\bar{j}\bar{k}} = 0. \tag{3.8}$$

Then, the condition $\Delta_{\lambda\mu\nu}(G\cdot G) = 0$ is trivially satisfied when we note

$$(G\cdot G)_{\mu\nu} = G_{\mu\alpha} f^{\alpha\beta} G_{\beta\nu} = \sum_{l=1}^{N} (G_{\mu l} G_{\bar{l}\nu} - G_{\mu\bar{l}} G_{l\nu}) = 0$$

identically. The remaining condition $\Delta_{\lambda\mu\nu}(G) = 0$ is also trivially obeyed except for the cases of $(\lambda,\mu,\nu) = (l,j,k)$ and $(\bar{l},j,k)$. However, the condition $\Delta_{ljk}(G) = 0$ can easily be seen to be satisfied also, provided that G_{jk} is a function only of the variables q_j and q_k, i.e.,

$$G_{jk} = -G_{kj} = G_{jk}(q_j,q_k). \tag{3.9}$$

Then, $\Delta_{\bar{l}jk}(G) = 0$ follows again trivially from this ansatz. In this way, we have found a solution for Eq. (3.1). However, $F_{\mu\nu} = G_{\mu\nu}$ alone offers a physically unsatisfactory solution since it will not contain any momentum dependence. The only remaining task is to study the validity of Eqs. (3.6) and (3.7), which are linear differential equations for $G_{\mu\nu}$. Equation (3.6) can be easily verified to hold if we choose $\xi_{jk} = 0$. The crucial equation (3.7) is satisfied, provided that we have

$$\sum_{m=1}^{N} h_{\bar{l}\bar{m}}\,\partial_m G_{jk} = \delta_{jl}\xi_j G_{kj} - \delta_{kl}\,\xi_k G_{jk},$$

where we have set $\eta_j = 0$ for simplicity. A study of this equation is not unlike the work on the Dynkin diagram of the Lie algebra A_{N-1}, whereby we assign a straight line between two points j and k when we have $G_{jk} \neq 0$. Considering only the indecomposable solution with $\xi_j = 1$, we find the following final solution:

$$F_{jk} = G_{jk} = f_j\,\delta_{j+1,k} e^{q_j - q_{j+1}} - f_k\,\delta_{k+1,j} e^{q_k - q_{k+1}}, \tag{3.10a}$$

$$F_{j\bar{k}} = -F_{\bar{k}j} = -p_j\,\delta_{jk}, \tag{3.10b}$$

$$F_{\bar{j}\bar{k}} = h_{\bar{j}\bar{k}} = -\epsilon(j-k), \tag{3.10c}$$

where $\epsilon(j-k)$ is the sign function

$$\epsilon(j-k) = \begin{cases} 1, & j>k, \\ 0, & j=k, \\ -1, & j<k, \end{cases} \tag{3.11}$$

and $f_j\,(j = 1,2,\ldots,N-1)$ are arbitrary coupling constants with

$$f_N = 0 \quad (j = N). \tag{3.12}$$

Note that Eq. (3.10) reproduces the result of Ref. 3, which has been obtained by an entirely different approach. We may remark that the identity

$$\epsilon(j-k) - \epsilon(j-k-1) = \delta_{j,k} + \delta_{j,k+1} \tag{3.13}$$

can be used[3] to verify the validity of $N^{\lambda}_{\mu\nu} = 0$. We note

$$K_1 = \frac{1}{2}\operatorname{Tr} S = \sum_{j=1}^{N} p_j,$$
$$K_2 = \frac{1}{4}\operatorname{Tr} S^2 = \frac{1}{2}\sum_{j=1}^{N} (p_j)^2 + \sum_{j=1}^{N-1} f_j e^{q_j - q_{j+1}}, \tag{3.14}$$

as in Ref. 3. Therefore, if we identify K_2 as the Hamiltonian of the system in Eq. (1.20) with $p = 2$, this will reproduce the standard Hamiltonian of the Toda lattice. We may interpret the first two terms $h_{\mu\nu} + C^{\alpha}_{\mu\nu} f_{\alpha\beta} x^\beta$ and the last term $G_{\mu\nu}$ in Eq. (3.5) to represent the free kinetic and interaction terms, respectively.

Our Toda lattice solution has been found to admit a rather unexpected extra affine structure. Define the affine connection coefficients $\Gamma^{\lambda}_{\mu\nu}$ by

$$\Gamma^k_{lj} = -\Gamma^{\bar j}_{l\bar k} = \delta_{jk}(-1)^{l+j}\epsilon(l-j), \tag{3.15}$$

while all other components of $\Gamma^\lambda_{\mu\nu}$ are taken to be identically zero. The covariant derivatives of both $G_{\mu\nu}$ and $f_{\mu\nu}$ are then found to vanish identically:

$$G_{\mu\nu;\lambda} = \partial_\lambda G_{\mu\nu} - \Gamma^\alpha_{\lambda\mu} G_{\alpha\nu} - \Gamma^\alpha_{\lambda\nu} G_{\mu\alpha} = 0, \tag{3.16}$$

$$f_{\mu\nu;\lambda} = \partial_\lambda f_{\mu\nu} - \Gamma^\alpha_{\lambda\mu} f_{\alpha\nu} - \Gamma^\alpha_{\lambda\nu} f_{\mu\alpha} = 0. \tag{3.17}$$

On the other hand, the Bianchi identity[7] requires the validity of

$$G_{\mu\nu;\lambda;\tau} - G_{\mu\nu;\tau;\lambda} = G_{\mu\alpha} R^{\ \alpha}_{\nu\lambda\tau} + G_{\alpha\nu} R^{\ \alpha}_{\mu\lambda\tau} + T^\alpha_{\lambda\tau} G_{\mu\nu;\alpha}, \tag{3.18}$$

where $R^{\ \nu}_{\mu\lambda\tau}$ and $T^\lambda_{\mu\nu}$ are Riemann curvature and torsion tensors, respectively, defined by

$$R^{\ \nu}_{\mu\lambda\tau} = \partial_\lambda \Gamma^\nu_{\tau\mu} - \partial_\tau \Gamma^\nu_{\lambda\mu} + \Gamma^\nu_{\lambda\alpha}\Gamma^\alpha_{\tau\mu} - \Gamma^\nu_{\tau\alpha}\Gamma^\alpha_{\lambda\mu}, \tag{3.19}$$

$$T^\lambda_{\mu\nu} = \Gamma^\lambda_{\mu\nu} - \Gamma^\lambda_{\nu\mu}. \tag{3.20}$$

For the present problem, we can verify $T^\lambda_{\mu\nu} \neq 0$, but

$$R^{\ \nu}_{\mu\lambda\tau} = 0 \tag{3.21}$$

identically, so that Eqs. (3.16) and (3.18) are compatible with each other. In spite of Eq. (3.16), we must have

$$F_{\mu\nu;\lambda} \neq 0 \tag{3.22}$$

because of the following reason. If we have the validity of both $F_{\mu\nu;\lambda} = f_{\mu\nu;\lambda} = 0$, it implies

$$S^{\ \nu}_{\mu;\lambda} = 0,$$

which has been shown in Ref. 1 to lead to $\partial_\lambda K_n = 0$ for any integer n. However, this is certainly not true for the present problem. Finally, we simply remark that we have $h_{\mu\nu;\lambda} \neq 0$ and $C^\lambda_{\mu\nu;\tau} \neq 0$. The probable reason for the validity of Eq. (3.16) can be traced to Eq. (3.7). Also, the affine connection introduced here should not be confused with the one used in Ref. 1 for the definition of the generalized Nijenhuis tensor.

B. Momentum-dependent Toda lattice

We can find a more general solution of $\Delta_{\lambda\mu\nu}(G) = \Delta_{\lambda\mu\nu}(G\cdot G) = 0$, as in the Appendix, if we allow linear dependence of $G_{\mu\nu}$ upon the momentum variable p_j. However, since the general case is rather complex, we consider here the simplest solution

$$G_{\bar j \bar k} = 0, \tag{3.23a}$$

$$G_{j\bar k} = -G_{\bar k j} = -A_{jk}(q), \tag{3.23b}$$

$$G_{jk} = \theta\{p_j A_{kj}(q) - p_k A_{jk}(q)\}, \tag{3.23c}$$

where $A_{jk}(q)$ is a function of q_j and q_k alone and θ is an arbitrary constant. If $A_{jk}(q)$ satisfies

$$\partial_j A_{kl} = \theta\{\delta_{jl} A_{kj} - \delta_{jk} A_{kl}\}, \tag{3.24}$$

we will then verify in the Appendix that $G_{\mu\nu}$ given by Eq. (3.23) indeed fulfills the condition $\Delta_{\lambda\mu\nu}(G) = \Delta_{\lambda\mu\nu}(G\cdot G) = 0$ after some calculations. With respect to the $h_{\mu\nu} + C^\alpha_{\mu\nu} f_{\alpha\beta} x^\beta$ term, we effectively adopt the same form as in the case of the Toda lattice and solve Eqs. (3.6) and (3.7). The final solutions are found to be

$$F_{\bar k j} = -F_{j\bar k} = S^k_j = S^{\bar j}_{\bar k} = p_j\delta_{jk} + A^{(\pm)}_{jk}(q), \tag{3.25a}$$

$$F_{\bar j \bar k} = -S^{\bar k}_j = -\epsilon(j-k), \tag{3.25b}$$

$$F_{jk} = S^{\bar k}_j = \pm(p_j A^{(\pm)}_{kj} - p_k A^{(\pm)}_{jk}), \tag{3.25c}$$

where we have chosen $\theta = \pm 1$ in Eq. (3.23) and set

$$A^{(+)}_{jk} = \delta_{j,k+1} f_k e^{q_k - q_{k+1}}, \tag{3.26a}$$

$$A^{(-)}_{jk} = \delta_{k,j+1} f_j e^{q_j - q_{j+1}}, \tag{3.26b}$$

with $f_N = 0$ as before. Especially, we calculate

$$K^{(\pm)}_1 = \frac{1}{2}\operatorname{Tr} S^{(\pm)} = \sum_{j=1}^N p_j, \tag{3.27a}$$

$$K^{(+)}_2 = \frac{1}{4}\operatorname{Tr}(S^{(+)})^2 = \frac{1}{2}\sum_{j=1}^N (p_j)^2 + \sum_{j=1}^{N-1} f_j p_j e^{q_j - q_{j+1}}, \tag{3.27b}$$

$$K^{(-)}_2 = \frac{1}{4}\operatorname{Tr}(S^{(-)})^2 = \frac{1}{2}\sum_{j=1}^N (p_j)^2 + \sum_{j=1}^{N-1} f_j p_{j+1} e^{q_j - q_{j+1}}, \tag{3.27c}$$

$$K^{(+)}_3 = \frac{1}{6}\operatorname{Tr}(S^{(+)})^3 = \frac{1}{3}\sum_{j=1}^N (p_j)^3 + \sum_{j=1}^{N-1} \{p_j(p_j + p_{j+1}) f_j e^{q_j - q_{j+1}} + p_j f_j f_{j+1} e^{q_j - q_{j+2}}\}. \tag{3.27d}$$

If we identify $K^{(+)}_2$ or $K^{(-)}_2$ as the Hamiltonian of the system, we then see that the interaction term depends linearly upon the momentum p_j. This is the reason we call the present case the momentum-dependent Toda lattice. Both the usual Toda lattice and the momentum-dependent Toda lattice are integrable since $K_1, K_2, \ldots, K_N$ are clearly[3] algebraically independent.

The present model also admits an extra affine structure with zero Riemann curvature tensor. Let us set

$$\Gamma^{\bar k}_{l\bar j} = \theta\delta_{jk}\,\delta_{lj}, \tag{3.28a}$$

$$\Gamma^j_{lk} = \Gamma^{\bar j}_{l\bar k} = \delta_{jk}(-1)^{l+k}\epsilon(l-k), \tag{3.28b}$$

while all other components of $\Gamma^\lambda_{\mu\nu}$ are set to be zero; we take $\theta = \pm 1$ in Eq. (3.28a) according to two solutions $A^{(\pm)}_{jk}$. Then, we can again find

$$G_{\mu\nu;\lambda} = 0, \tag{3.29a}$$

$$R^{\ \mu}_{\nu\lambda\tau} = 0. \tag{3.29b}$$

However, in contrast to the Toda lattice case, we now have

$$f_{\mu\nu;\lambda} \neq 0, \tag{3.30}$$

although nonzero components of $f_{\mu\nu;\lambda}$ are constants in the canonical frame. The main reason for Eq. (3.30) is the sign difference in $\Gamma^{\bar j}_{l\bar k}$ between Eqs. (3.15) and (3.28b).

IV. SOLUTIONS FOR THE CASE $N=2$

For the special case of $N=2$, we can find solutions of a more complicated type: this is because we have

$$\Delta_{ijk}(F) = \Delta_{\bar i\bar j\bar k}(F) = 0, \tag{4.1a}$$

$$\Delta_{ijk}(F\cdot F) = \Delta_{\bar i\bar j\bar k}(F\cdot F) = 0, \tag{4.1b}$$

identically for this case since the latin indices i, j, and k can assume only the two values 1 and 2. Therefore, we need only satisfy equations of the form

$$\Delta_{12\bar{k}}(F) = \Delta_{12\bar{k}}(F \cdot F) = 0, \tag{4.2a}$$

$$\Delta_{\bar{1}\bar{2}k}(F) = \Delta_{\bar{1}\bar{2}k}(F \cdot F) = 0 \tag{4.2b}$$

for $k = 1$ and 2, which simplifies the problem considerably. Also, for $N = 2$, it can be verified that the following special identity for any antisymmetric tensor $F_{\mu\nu}$ is valid:

$$(F \cdot F)_{\mu\nu} = (K_2 - \tfrac{1}{2}K_1^2) f_{\mu\nu} + K_1 F_{\mu\nu}. \tag{4.3}$$

Therefore, if we have $\Delta_{\lambda\mu\nu}(F) = 0$, then Eq. (4.3) gives

$$\Delta_{\lambda\mu\nu}(F \cdot F) = f_{\mu\nu}\, \partial_\lambda \widetilde{K}_2 + f_{\nu\lambda}\, \partial_\mu \widetilde{K}_2 + f_{\lambda\mu}\, \partial_\nu \widetilde{K}_2 + F_{\mu\nu}\, \partial_\lambda K_1 + F_{\nu\lambda}\, \partial_\mu K_1 + F_{\lambda\mu}\, \partial_\nu K_1, \tag{4.4}$$

where for simplicity, we have set,

$$\widetilde{K}_2 = K_2 - \tfrac{1}{2}K_1^2.$$

Equation (4.4) offers a simple check for the validity of $\Delta_{\lambda\mu\nu}(F \cdot F) = 0$. Moreover, Eq. (4.3) implies that all K_n $(n \geqslant 3)$ are polynomials of K_1 and K_2. For example, we will have

$$K_3 = \tfrac{1}{6}\,\mathrm{Tr}\, S^3 = K_1 K_2 - \tfrac{1}{6}(K_1)^3$$

by multiplying $f^{\nu\lambda} f_{\lambda\tau} f^{\tau\mu}$ to both sides of Eq. (4.3).

However, we have to consider the nontranslation invariant solution since the translation invariant solution for $N = 2$ is trivial: Because of this, we must now take the quadratic expression (2.25), rather than the linear expression (2.21), for its free parts since the latter always implies the conservation of the total momenta $p_1 + p_2$ and forces the theory to be trivial.

We make the following ansatz:

$$F_{j\bar{k}} = -F_{\bar{k}j} = -\delta_{jk}(p_j)^2 - A_{jk}(q), \tag{4.5a}$$

$$F_{\bar{j}\bar{k}} = -\xi_{jk} p_j + \xi_{kj} p_k, \tag{4.5b}$$

$$F_{jk} = p_j B_{kj}(q) - p_k B_{jk}(q), \tag{4.5c}$$

for some constant ξ_{jk} and some functions $A_{jk}(q)$ and $B_{jk}(q)$; note that this choice is a combination of Eq. (2.25) with $h_{\mu\nu} = 0$ and that of $G_{\mu\nu}$ studied in the Appendix. However, in this case, we can directly analyze Eq. (4.2) without many complications. We then find the following two types of solutions.

A. Solution (i)

We set

$$J = C_1 \exp\{[2/(\xi_{21} + \xi_{12})](q_1 + q_2)\}, \tag{4.6a}$$

$$G = C_2 \exp\{[2/(\xi_{21} - \xi_{12})](q_1 - q_2)\} \tag{4.6b}$$

for the arbitrary constants C_1 and C_2. If we have $\xi_{21} + \xi_{12} = 0$ or $\xi_{21} - \xi_{12} = 0$, then we set $J = 0$ or $G = 0$, accordingly. After some calculations, the solution is given by

$$A_{11} = \xi_{21}(J + G), \tag{4.7a}$$

$$A_{22} = \xi_{12}(J - G), \tag{4.7b}$$

$$A_{21} = -A_{12} = \tfrac{1}{2}(\xi_{21} - \xi_{12})J - \tfrac{1}{2}(\xi_{21} + \xi_{12})G, \tag{4.7c}$$

$$B_{12} = -J - G, \tag{4.7d}$$

$$B_{21} = -J + G. \tag{4.7e}$$

Then, we calculate

$$K_1 = \tfrac{1}{2}\,\mathrm{Tr}\, S = p_1^2 + p_2^2 + A_{11} + A_{22} = p_1^2 + p_2^2 + (\xi_{21} + \xi_{12}) C_1 \times \exp\{[2/(\xi_{21} + \xi_{12})](q_1 + q_2)\} + (\xi_{21} - \xi_{12}) C_2 \exp\{[2/(\xi_{21} - \xi_{12})](q_1 - q_2)\}, \tag{4.8a}$$

$$K_2 = \tfrac{1}{4}\,\mathrm{Tr}\, S^2 = \tfrac{1}{2}(K_1)^2 - \tfrac{1}{4}\{2p_1 p_2 + (\xi_{21} + \xi_{12})J - (\xi_{21} - \xi_{12})G\}^2. \tag{4.8b}$$

Since K_1 and K_2 are algebraically independent, this model is integrable. However, this case is dynamically trivial because of the following reason. We make the canonical transformation $(q_j, p_j) \to (Q_j, P_j)$ by

$$Q_1 = (1/\sqrt{2})(q_1 + q_2), \quad Q_2 = (1/\sqrt{2})(q_1 - q_2),$$

$$P_1 = (1/\sqrt{2})(p_1 + p_2), \quad p_2 = (1/\sqrt{2})(p_1 - p_2)$$

where we have

$$f = \sum_{j=1}^{2} dp_j \wedge dq_j = \sum_{j=1}^{2} dP_j \wedge dQ_j.$$

However, then,

$$K_1 = H_1(P_1, Q_1) + H_2(P_2, Q_2)$$

is a sum of two independent Hamiltonians H_1 and H_2, where H_1 and H_2 have the forms

$$H_1 = P_1^2 + C_1' \exp(\beta Q_1),$$

$$H_2 = P_2^2 + C_2' \exp(\gamma Q_2)$$

for some constants C_1', C_2', β, and γ. Nevertheless, the validity of the hierarchy equation is still nontrivial.

B. Solution (ii)

A more careful study of Eq. (4.2) with (4.5) reveals that we can find a completely new type of now dynamically nontrivial solution when one of ξ_{12} and ξ_{21} vanishes. Here, we assume $\xi_{21} = 0$, but $\xi_{12} \neq 0$. Setting

$$J = C_1 \exp[(2/\xi_{12})(q_1 + q_2)], \tag{4.9a}$$

$$G = C_2 \exp[(2/\xi_{12})(q_2 - q_1)], \tag{4.9b}$$

the new solution is found to be

$$F_{1\bar{1}} = -p_1^2 - \phi(y), \tag{4.10a}$$

$$F_{2\bar{2}} = -p_2^2 + \xi_{12}(J + G), \tag{4.10b}$$

$$F_{1\bar{2}} = \frac{1}{2}\xi_{12}(J - G) - \frac{d}{dy}\phi(y), \tag{4.10c}$$

$$F_{2\bar{1}} = -\frac{1}{2}\xi_{12}(J - G), \tag{4.10d}$$

$$F_{12} = p_1(J + G) - p_2(J - G), \tag{4.10e}$$

$$F_{\bar{1}\bar{2}} = -\xi_{12} p_1. \tag{4.10f}$$

In Eq. (4.10), we have defined

$$y = (2/\xi_{12}) q_1 \tag{4.11}$$

and $\phi(y)$ is any solution of the second-order differential equation

$$\frac{d^2}{dy^2}\phi(y) + 3\,\frac{C_1e^{y} + C_2e^{-y}}{C_1e^{y} - C_2e^{-y}}\frac{d}{dy}\phi(y) + 2\phi(y) = 0, \quad (4.12)$$

which can be integrated to give

$$\phi(y) = \{1/[C_1e^{y} - C_2e^{-y}]^2\}\{a_1 + a_2[C_1e^{y} + C_2e^{-y}]\} \quad (4.13)$$

for another set of constants a_1 and a_2. Then, we calculate

$$K_1 = p_1^2 + p_2^2 + \phi(y) - \xi_{12}(J+G), \quad (4.14a)$$

$$\widetilde{K}_2 = K_2 - \frac{1}{2}(K_1)^2 = -\left[p_1p_2 - \frac{1}{2}\xi_{12}(J-G)\right]^2 - (p_2)^2\phi(y) + \xi_{12}(J+G)\phi(y) + \frac{1}{2}\xi_{12}(J-G)\frac{d}{dy}\phi(y). \quad (4.14b)$$

We can again verify the validity of $\Delta_{\lambda\mu\nu}(F\cdot F) = 0$ from Eq. (4.4). Also, this case is clearly dynamically nontrivial, in contrast to solution (i).

We can find simple forms of $\phi(y)$ for some special choices of the constants C_1 and C_2, redefining a_1 and a_2 suitably, as follows.

For $C_1 = 0,\ C_2 \neq 0$,

$$\phi(y) = a_1e^{y} + a_2e^{2y}, \quad (4.15a)$$

for $C_2 = 0,\ C_1 \neq 0$,

$$\phi(y) = a_1e^{-y} + a_2e^{-2y}, \quad (4.15b)$$

for $C_1 = C_2 \neq 0$,

$$\phi(y) = [1/(\sinh y)^2][a_1 + a_2\cosh y], \quad (4.15c)$$

and for $C_1 = -C_2 \neq 0$,

$$\phi(y) = [1/(\cosh y)^2][a_1 + a_2\sinh y]. \quad (4.15d)$$

If we identify K_1 as the Hamiltonian of the system, and if, for simplicity, we set

$$\alpha = 2/\xi_{12}, \quad (4.16)$$

then the potential of the system is identified as

$$V(q_1,q_2) = \phi(\alpha q_1) - \xi_{12}\{C_1\exp[\alpha(q_1+q_2)] + C_2\exp[\alpha(q_2-q_1)]\}. \quad (4.17)$$

Especially, if we choose $C_1 = 0$ and set

$$C = \tfrac{1}{2}\xi_{12}C_2, \quad (4.18)$$

then the system possesses the two conserved quantities

$$K_1 = p_1^2 + p_2^2 + a_1\exp(\alpha q_1) + a_2\exp(2aq_1) - 2C\exp[\alpha(q_2-q_1)] \quad (4.19)$$

and

$$\widetilde{K}_2 = -\{p_1p_2 + C\exp[\alpha(q_2-q_1)]\}^2 - (p_2)^2 \times[a_1\exp(\alpha q_1) + a_2\exp(2\alpha q_1)] + a_1C\exp(\alpha q_2). \quad (4.20)$$

We would not have suspected the existence of the second conserved quantity $\widetilde{K}_2$ for such a relatively simple Hamiltonian K_1 unless we systematically proceeded, as in the present case.

The present case is also integrable since K_1 and K_2 (or $\widetilde{K}_2$) are algebraically independent. Also, the explicit form of the Hamiltonian K_1 given by Eq. (4.19) or (4.14a) suggests a close relationship between the present model and that of Olshansky and Perelomov,[8] who utilized an entirely different approach based upon the root system of the exceptional Lie algebra G_2. Note that K_1 in Eq. (4.14a) may be interpreted to represent a reduced Hamiltonian of a translation-invariant three-body system at a center-of-the-mass frame.

V. GENERAL QUADRATIC PROBLEM

In Sec. II, we analyzed the linear problem for $F_{\mu\nu}$. In this section, we will discuss the same for the quadratic case. For simplicity, we set

$$F_{\mu\nu} = H_{\mu\nu} + G_{\mu\nu}, \quad (5.1)$$

$$H_{\mu\nu} = h_{\mu\nu} + C^{\alpha}_{\mu\nu}f_{\alpha\beta}x^{\beta}, \quad (5.2)$$

$$G_{\mu\nu} = \tfrac{1}{2}\,\theta_{\mu\nu\alpha\beta}x^{\alpha}x^{\beta}, \quad (5.3)$$

where $\theta_{\mu\nu\alpha\beta}$ are constants satisfying

$$\theta_{\mu\nu\alpha\beta} = -\theta_{\nu\mu\alpha\beta}, \quad (5.4)$$

$$\theta_{\mu\nu\alpha\beta} = \theta_{\mu\nu\beta\alpha}. \quad (5.5)$$

Although we can solve the general problem

$$\Delta_{\lambda\mu\nu}(F) = \Delta_{\lambda\mu\nu}(F\cdot F) = 0, \quad (5.6)$$

we shall assume here that both $H_{\mu\nu}$ and $G_{\mu\nu}$ are separate solutions of the problem, as in Eqs. (3.1) and (3.2) with Eq. (3.4). First, the condition $\Delta_{\lambda\mu\nu}(G) = 0$ implies the validity of

$$\theta_{\mu\nu\lambda\alpha} + \theta_{\nu\lambda\mu\alpha} + \theta_{\lambda\mu\nu\alpha} = 0. \quad (5.7)$$

Then, with the help of Eq. (5.4), (5.5), and (5.7), the constraint $\Delta_{\lambda\mu\nu}(G\cdot G) = 0$ is found to lead to the following

$$T_{\mu\nu;\lambda\alpha\beta\gamma} + T_{\nu\lambda;\mu\alpha\beta\gamma} + T_{\lambda\mu;\nu\alpha\beta\gamma} = 0, \quad (5.8a)$$

where we have set

$$T_{\mu\nu;\lambda\alpha\beta\gamma} = \theta_{\tau\mu\alpha\lambda}f^{\tau\rho}\theta_{\rho\nu\beta\gamma} + \theta_{\tau\mu\beta\lambda}f^{\tau\rho}\theta_{\rho\nu\alpha\gamma} + \theta_{\tau\mu\gamma\lambda}f^{\tau\rho}\theta_{\rho\nu\beta\alpha} + \theta_{\tau\mu\alpha\beta}f^{\tau\rho}\theta_{\rho\nu\gamma\lambda} + \theta_{\tau\mu\gamma\beta}f^{\tau\rho}\theta_{\rho\nu\alpha\lambda} + \theta_{\tau\mu\alpha\gamma}f^{\tau\rho}\theta_{\rho\nu\beta\lambda}. \quad (5.8b)$$

We note that $T_{\mu\nu;\lambda\alpha\beta\gamma}$ is totally symmetric in the indices λ, α, β, and γ, but antisymmetric in μ and ν because of Eqs. (5.4) and (5.5). If we use Eq. (5.7), then Eqs. (5.8) can be slightly simplified to become

$$\theta_{\lambda\mu\alpha\tau}f^{\tau\rho}\theta_{\rho\nu\beta\gamma} + \theta_{\lambda\mu\beta\tau}f^{\tau\rho}\theta_{\rho\nu\gamma\alpha} + \theta_{\lambda\mu\gamma\tau}f^{\tau\rho}\theta_{\rho\nu\alpha\beta} + \theta_{\nu\lambda\alpha\tau}f^{\tau\rho}\theta_{\rho\mu\beta\gamma} + \theta_{\nu\lambda\beta\tau}f^{\tau\rho}\theta_{\rho\mu\gamma\alpha} + \theta_{\nu\lambda\gamma\tau}f^{\tau\rho}\theta_{\rho\mu\alpha\beta} + \theta_{\mu\nu\alpha\tau}f^{\tau\rho}\theta_{\rho\lambda\beta\gamma} + \theta_{\mu\nu\beta\tau}f^{\tau\rho}\theta_{\rho\lambda\gamma\alpha} + \theta_{\mu\nu\gamma\tau}f^{\tau\rho}\theta_{\rho\lambda\alpha\beta} = 0. \quad (5.9)$$

Evidently, the analysis of Eqs. (5.7) and (5.9) is still difficult. To simplify the notation, we introduce an abstract vector space V spanned by the $2N$ basis vector X_μ ($\mu = 1,2,\ldots,2N$) and define a triple-linear product $[X,Y,Z]\in V$ for any three elements X,Y, and $Z\in V$ by

$$[X_\mu,X_\nu,X_\lambda] = f^{\alpha\beta}\theta_{\beta\mu\nu\lambda}X_\alpha. \quad (5.10)$$

Then, first, Eq. (5.5) is rewritten as

$$[X,Y,Z] = [X,Z,Y]. \quad (5.11)$$

On the other hand, Eq. (5.9) can be shown to be equivalent to

$$[[Y,Z,X],X,X]-[[Z,Y,X],X,X]$$
$$=[Y,X,[Z,X,X]]-[Z,X,[Y,X,X]]. \quad (5.12)$$

This fact can be demonstrated as follows. Let ξ^μ, η^μ, and ζ^μ be three arbitrary numerical vectors and set

$$X=\xi^\mu X_\mu, \quad Y=\eta^\mu X_\mu, \quad Z=\zeta^\mu X_\mu. \quad (5.13)$$

When we insert Eq. (5.13) into Eq. (5.12) and use the arbitrariness of ξ^μ, η^ν, and ζ^λ, Eq. (5.12) gives a quadratic identity involving $\theta_{\mu\nu\lambda\tau}$ in view of Eq. (5.10), which can be shown to reduce to Eq. (5.9) when we see Eq. (5.7).

Next, we have to rewrite the remaining conditions (5.4) and (5.7) in the same basis-independent notation. Introducing the bilinear antisymmetric nondegenerate functional $\langle X,Y\rangle$ by Eq. (2.9), these are indeed recast in the forms of

$$\langle X,[Y,Z,W]\rangle=-\langle Y,[X,Z,W]\rangle, \quad (5.14)$$

$$\langle X,[Y,Z,W]\rangle+\langle Y,[Z,X,W]\rangle+\langle Z,[X,Y,W]\rangle=0. \quad (5.15)$$

In conclusion, the desired solution has been reduced to a study of a triple-linear system satisfying Eqs. (5.11), (5.12), (5.14), and (5.15) for the arbitrary vectors X, Y, Z, and $W\in V$. Note that we can rewrite $\theta_{\mu\nu\alpha\beta}$ as

$$\theta_{\mu\nu\alpha\beta}=\langle X_\mu,[X_\nu,X_\alpha,X_\beta]\rangle.$$

Finally, we must consider Eq. (3.4), which gives

$$f^{\alpha\beta}\{h_{\mu\alpha}\theta_{\nu\lambda\beta\tau}+h_{\nu\alpha}\theta_{\lambda\nu\beta\tau}+h_{\lambda\alpha}\theta_{\mu\nu\beta\tau}\}=0 \quad (5.16)$$

and

$$\begin{aligned}f^{\alpha\beta}\{&C^{\gamma}_{\mu\alpha}f_{\gamma\tau}\theta_{\nu\lambda\beta\rho}+C^{\gamma}_{\mu\alpha}f_{\gamma\rho}\theta_{\nu\lambda\beta\tau}+C^{\gamma}_{\nu\alpha}f_{\gamma\tau}\theta_{\lambda\mu\beta\rho}\\&+C^{\gamma}_{\nu\alpha}f_{\gamma\rho}\theta_{\lambda\mu\beta\tau}+C^{\gamma}_{\lambda\alpha}f_{\gamma\tau}\theta_{\mu\nu\beta\rho}+C^{\gamma}_{\lambda\alpha}f_{\gamma\rho}\theta_{\mu\nu\beta\tau}\\&+C^{\gamma}_{\nu\lambda}f_{\gamma\beta}\theta_{\mu\alpha\tau\rho}+C^{\gamma}_{\lambda\mu}f_{\gamma\beta}\theta_{\nu\alpha\tau\rho}+C^{\gamma}_{\mu\nu}f_{\gamma\beta}\theta_{\lambda\alpha\tau\rho}\}=0.\end{aligned} \quad (5.17)$$

When we introduce the commutor $[X,Y]$ in V by

$$[X_\mu,X_\nu]=C^{\lambda}_{\mu\nu}X_\lambda \quad (5.18)$$

then Eq. (5.17) is also further rewritten as

$$\begin{aligned}&\langle[X,[Z,Y,W]-[Y,Z,W]],W\rangle\\&+\langle[Y,[X,Z,W]-[Z,X,W]],W\rangle\\&+\langle[Z,[Y,X,W]-[X,Y,W]],W\rangle\\&+\langle[Y,Z],[X,W,W]\rangle+\langle[Z,X],[Y,W,W]\rangle\\&+\langle[X,Y],[Z,W,W]\rangle=0\end{aligned} \quad (5.19)$$

when we use Eq. (5.7). We can also rewrite Eq. (5.16) in a basis-independent way by introducing another antisymmetric bilinear functional $h(X,Y)$ by

$$h(X_\mu,X_\nu)=h_{\mu\nu}. \quad (5.20)$$

Then, we find

$$\begin{aligned}&h(X,[Y,Z,W]-[Z,Y,W])\\&+h(Y,[Z,X,W]-[X,Z,W])\\&+h(Z,[X,Y,W]-[Y,X,W])=0.\end{aligned} \quad (5.21)$$

Although many triple-linear systems satisfying similar relations have been studied by many authors[9–12] the present triple system is, unfortunately, quite different from those already investigated and its study will be postponed for the future. However, we would like to simply mention that our triple system possesses a nontrivial solution (2.25), as well as a once-iterated solution $F_{\mu\nu}=(H\cdot H)_{\mu\nu}$ for the linear form $H_{\mu\nu}$ when we consider only the dominant quadratic terms.

ACKNOWLEDGMENT

This work is supported in part by the U. S. Department of Energy under Contract No. DE-AC02-76ER13065.

APPENDIX: SOME SOLUTIONS FOR $\Delta_{\lambda\mu\nu}(G\cdot G)=0$

Here, we will study a generalization of the solution given by Eq. (3.23).

We seek a solution of form

$$G_{\bar j\bar k}=0, \quad (A1)$$

$$G_{j\bar k}=-A_{jk}, \quad (A2)$$

$$G_{jk}=p_jB_{kj}-p_kB_{jk}, \quad (A3)$$

where A_{jk} and B_{jk} are functions only of q_j and q_k. Then, we can easily verify that the condition $\Delta_{\lambda\mu\nu}(G)=0$ is satisfied if we have

$$\partial_jA_{kl}=\delta_{jl}B_{kj}-\delta_{jk}C_{kl} \quad (A4)$$

for some functions C_{kl}. Next, we set similarly

$$\partial_jB_{kl}=\delta_{jl}D_{kj}-\delta_{jk}E_{kl} \quad (A5)$$

for some functions D_{kj} and E_{kj} of q_j and q_k. Moreover, when we note

$$(G\cdot G)_{\bar j\bar k}=0, \quad (A6)$$

$$(G\cdot G)_{j\bar k}=-\sum_l A_{jl}A_{lk}, \quad (A7)$$

$$(G\cdot G)_{jk}=p_j\Big(\sum_l A_{kl}B_{lj}\Big)-p_k\Big(\sum_l A_{jl}B_{lk}\Big)+\sum_l p_l(A_{jl}B_{kl}-A_{kl}B_{jl}), \quad (A8)$$

then we can readily find that the second condition $\Delta_{\lambda\mu\nu}(G\cdot G)=0$ is also satisfied, provided that we have the following two equations: First, we must have

$$A_{jl}(B_{kl}-B_{kj})+A_{kl}(B_{jk}-B_{jl})+A_{kj}C_{jl}-A_{jk}C_{kl}=0 \quad (A9)$$

and second, we must have

$$p_jQ_{jkl}+p_kQ_{klj}+p_lQ_{ljk}=0, \quad (A10)$$

where we have set

$$Q_{jkl}=-Q_{jlk}=(B_{kl}B_{lj}-B_{lk}B_{kj})+(A_{lk}E_{kj}-A_{kl}E_{lj})+(A_{kj}D_{lj}-A_{lj}D_{kj}). \quad (A11)$$

Since the lhs of Eq. (A10) is totally antisymmetric in the indices j, k, and l, Eq. (A10) is trivially satisifed for $N\leqslant 2$. This is one reason why we found some complicated solutions for $N=2$, as in Sec. IV. For $N\geqslant 3$, Eq. (A10) can be also obeyed if we have

$$Q_{jkl}=\delta_{jk}P_{kl}-\delta_{jl}P_{lk} \quad (A12)$$

for some functions P_{kl}.

Next, let us set

$$H_{j\bar{k}} = -\delta_{jk} p_j , \tag{A13}$$

$$H_{\bar{j}\bar{k}} = h_{\bar{j}\bar{k}} , \tag{A14}$$

$$H_{jk} = 0 . \tag{A15}$$

Then, the sum

$$F_{\mu\nu} = G_{\mu\nu} + H_{\mu\nu} \tag{A16}$$

will satisfy Eq. (3.1). The remaining condition (3.4) can also be fulfilled, first if we have

$$h_{\bar{j}\bar{k}}(B_{lk} + B_{lj}) + h_{\bar{k}\bar{l}} C_{lj} - h_{\bar{j}\bar{l}} C_{lk} = \delta_{kl} A_{kj} - \delta_{lj} A_{jk} \tag{A17}$$

and second, if there exists a function Y_{jl} satisfying

$$h_{\bar{l}\bar{k}} D_{jk} - h_{\bar{l}\bar{j}} E_{jk} + (\delta_{kl} + \delta_{jl}) B_{jk} = \delta_{jk} Y_{jl} . \tag{A18}$$

These conditions (A17) and (A18) are still not easy to analyze. However, there exists a simple solution. Suppose that we have

$$C_{jk} = B_{jk} = \theta A_{jk} , \tag{A19}$$

$$D_{jk} = E_{jk} = \theta^2 A_{jk} \tag{A20}$$

for a constant θ. Then Eqs. (A9) and (A10) are identically satisfied with $Q_{jkl} = 0$, while Eqs. (A4) and (A5) reduce to

$$\partial_j A_{kl} = \theta(\delta_{jl} A_{kj} - \delta_{jk} A_{kl}), \tag{A21}$$

which is precisely Eq. (3.24). Finally, both Eqs. (A17) and (A18) become identical and are rewritten as

$$\theta(h_{\bar{l}\bar{k}} - h_{\bar{l}\bar{j}}) A_{jk} = -(\delta_{kl} + \delta_{jl}) A_{jk}, \tag{A22}$$

with a special choice of $Y_{jl} = 0$. Then, the final solution of Eqs. (A21) and (A22) yields the result of Eq. (3.25) for $\theta = \pm 1$.

[1]S. Okubo, J. Math. Phys. **30**, 834 (1989).
[2]F. Magri, J. Math. Phys. **19**, 1156 (1978).
[3]A. Das and S. Okubo, to be published in Ann. Phys. (NY).
[4]R. Abraham and J. E. Marsden, *Foundations of Mechanics* (Benjamin/Cummings, Reading, MA, 1978).
[5]See, e.g., P. J. Hilton and V. Stammbach, *A Course in Homological Algebra* (Springer, Berlin, 1971).
[6]A. A. Kirillov, *Elements of the Theory of Representations* (Springer, Berlin, 1976), translated from Russian by E. Hewitt; see p. 224, theorem 1.
[7]See, e.g., S. Kobayashi and K. Nomizu, *Foundations of Differential Geometry* (Wiley, Interscience, New York, 1963), Vol. I.
[8]M. A. Olshansky and A. M. Perelomov, Phys. Rep. **71**, 315 (1981); see, also, D. L. Olive and N. Turok, Nucl. Phys. B **220**, 491 (1983).
[9]I. L. Kantor; Sov. Math. Dokl. **14**, 254 (1973).
[10]B. N. Allison, Am. J. Math. **98**, 285 (1976).
[11]K. Yamaguchi and A. Ono, Bull. Fac. Sch. Edu. Hiroshima Univ. Part II **7**, 43 (1984).
[12]S. Okubo, Algebras Groups Geomet. **3**, 60 (1986).

On topological effects in quantum mechanics: The harmonic oscillator in the pointed plane

H. D. Doebner, H. J. Elmers, and W. F. Heidenreich
Institut für Theoretische Physik A, Technische Universität Clausthal, D-3392 Clausthal, Federal Republic of Germany

(Received 5 October 1988; accepted for publication 7 December 1988)

Quantum mechanics of a (nonrelativistic) system S localized on a topologically nontrivial manifold $\mathscr{M}$ as its configuration space is based on a quantization method, which, in general, reflects global properties of $\mathscr{M}$, i.e., some of the observables of S will "feel" the topology: There are topological effects and inequivalent quantizations on $\mathscr{M}$. Some straightforward examples are given for such effects, using Borel quantization (BQ), the pointed plane as manifold $\mathscr{M}$, and the energy operator with harmonic potential as observable. Two topological effects exist. There are unitarily inequivalent BQ on $\mathscr{M}$, which are equivalent to the usual quantization on the plane with a topological potential, which has the form of a Bohm–Aharonov potential. There are different self-adjoint extensions of the energy operator for a given BQ that in some cases are related to another kind of topological potential. These effects are discussed in detail, especially the self-adjoint extensions of the energy operator. An experimental setup to verify some of the results is suggested.

I. INTRODUCTION

Observables of a (nonrelativistic) physical system S that is constrained on a smooth manifold $\mathscr{M}$ as its configuration space can depend on global properties of $\mathscr{M}$, e.g., on the fundamental group $\pi_1(\mathscr{M})$, its character group $\pi_1^*(\mathscr{M})$, cohomology groups $H^p(\mathscr{M},\mathbf{Z})$, etc. For a quantized system this is the case for a large set of observables. Unitarily inequivalent quantizations exist reflecting the topology of $\mathscr{M}$. They predict different experimental results and show the relevance of *topological effects*. For some $\mathscr{M}$, inequivalent quantizations of certain observables can be parametrized through differential n-forms, behaving like potentials. They are called *topological potentials of the n+1st kind*.

The mechanisms that produce such effects depend on the quantization method and the choice of the observables.

We report on two mechanisms: the *Borel quantization*, i.e., quantization of the kinematic on $\mathscr{M}$ and the corresponding quantization of an energy observable $\mathbf{H}$. After a general outline of the method (Sec. II, see also Refs. 1 and 2) we apply the results to a one-particle system constrained on $\mathscr{M}_1 = \mathbf{R}^3 - \mathbf{R}^1$ moving under the influence of an external harmonic force (Sec. III); because of its symmetry the problem is equivalent to one that is constrained on $\mathscr{M}_2 = \mathbf{R}^2 - \{0\}$, i.e., on the pointed plane. There appears a topological potential of the second kind, which has the form of the Bohm–Aharonov[3] potential. The self-adjointness condition for the corresponding Hamilton operator creates in some cases an additional (singular or point) potential of the first kind. We present an interpretation of the results and suggest (Sec. IV) an experimental setup to verify some of them.

II. THE QUANTIZATION

A. Topology and kinematical observables, topological potentials of the second kind

We use a "conservative" quantization method. Classical (generalized) position and momentum observables are modeled via Borel sets B taken from a Borel field $\mathscr{B}(\mathscr{M})$ and vector fields X taken from the set $\mathscr{X}_c(\mathscr{M})$ of complete vector fields. The flow φ_t^X of X acts on $B \in \mathscr{B}(\mathscr{M})$ as

$$B_t = \{m' | m' = \varphi_t^X(m),\ m \in B\}. \tag{1}$$

The geometrical object $(\mathscr{B}(\mathscr{M}), \mathscr{X}_c(\mathscr{M}))$ is called *Borel kinematic* on $\mathscr{M}$.[1] For its quantization we need a map $\mathscr{Q}$ into the set of self-adjoint operators $\mathscr{S}\mathscr{A}(\mathscr{H})$ on some Hilbert space with a common, dense, invariant domain:

$$\mathscr{Q}:\ \mathscr{B}(\mathscr{M}) \ni B \mapsto \mathbf{E}(B) \in \mathscr{S}\mathscr{A}(\mathscr{H}),$$

$$\mathscr{Q}:\ \mathscr{X}_c(\mathscr{M}) \ni X \mapsto \mathbf{P}(X) \in \mathscr{S}\mathscr{A}(\mathscr{H}).$$

Here $\mathscr{Q}$ should respect measure-theoretic properties of $\mathscr{B}(\mathscr{M})$ and the Lie-algebraic structure connected with $\mathscr{X}_c$. Furthermore, some additional technical conditions should be fulfilled. We require that $\mathbf{P}(X)$ is *local* (see Ref. 1) and that for any X a one-dimensional unitary group $\mathbf{V}_t^X$ exists, such that—as quantum analog to (1)—a kind of imprimitivity condition holds

$$\mathbf{V}_{-t}^X \circ \mathbf{E}(B) \circ \mathbf{V}_t^X = \mathbf{E}(\varphi_{-t}^X(B)). \tag{2}$$

It turns out that this *Borel quantization*, i.e., the map $\mathscr{Q}$, is not unique. Up to unitary equivalence there is a 1:1 correspondence between different $\mathscr{Q}$ and the elements of $\pi_1^*(\mathscr{M}) \otimes \mathbf{R}$ [$\pi_1(\mathscr{M})$ is assumed to be finitely generated]; $\mathbf{R}$ is the real line. The explicit forms of $\mathbf{E}(B)$ and $\mathbf{P}(X)$ are available. The Hilbert space $\mathscr{H}$ can be realized as completion $L^2(\eta(\mathscr{M}); \langle\cdot|\cdot\rangle, \mu)$ of the linear space of sections in a complex line bundle $\eta(\mathscr{M})$ over $\mathscr{M}$ which are square integrable with respect to a Hermitian metric $\langle\cdot|\cdot\rangle$ and a finite Borel measure μ on $\mathscr{M}$. In general, several nonisomorphic η exist over $\mathscr{M}$. For trivial bundles the space $\mathscr{H}$ is spanned by square integrable functions, i.e., $\mathscr{H} = L^2(\mathscr{M}; \mu)$; this is denoted as the *standard* case. Here inequivalent quantizations are in 1:1 correspondence with the elements of $\pi_1^*(\mathscr{M}) / (\Gamma\pi_1)^*(\mathscr{M}) \otimes \mathbf{R}$ with $(\Gamma\pi_1)^*(\mathscr{M})$ as the subgroup of $\pi_1^*(\mathscr{M})$ generated by all elements of finite order.

 0022-2488/89/051053-07$02.50

We give now the results for the standard case. Here $\mathbf{E}(B)$ is a projection operator (as it should be); $\mathscr{Q}$ maps real functions f into multiplication operators

$$\mathscr{Q}: f \mapsto \mathbf{Q}(f): \quad \mathbf{Q}(f)\psi = f\cdot\psi; \\ f\in C^\infty(\mathscr{M},\mathbf{R}), \ \psi\in L^2(\mathscr{M},\mu). \tag{3}$$

Here $\mathbf{P}(X)$ is a differential operator of first order, essentially self-adjoint for all X on $C^\infty(\mathscr{M},\mathbf{C})$ (a similar formula holds for the general case.):

$$\mathscr{Q}: X \mapsto \mathbf{P}(X); \quad \mathbf{P}(X) = iX + \omega(X) + (i/2 + c) \\ \times \mathbf{Q}(\mathrm{div}_\mu X); \tag{4}$$

here ω is an arbirary element of the Abelian group $Z^1(\mathscr{M})$ of (real) closed differential one-forms and c some real number. The X acts as Lie derivative and the divergence of X is meant with respect to $\mu = \rho(x,y,z)dx\,dy\,dz, \rho > 0$, and smooth (locally). The physical interpretation of $\omega(X)$ as a potential, i.e., as topological potential of the second kind (vector potential), is obvious. (Topological effects can be traced also in other quantization methods, e.g., in the path integral approach.[4])

For different $\omega_1,\omega_2\in Z^1(\mathscr{M})$ and $c_1,c_2\in\mathbf{R}$ the map $\mathscr{Q}$ yields different results. These are unitarily equivalent if and only if

$$c_1 - c_2 \text{ is zero and } \omega_1 - \omega_2 \text{ is logarithmically exact.} \tag{5}$$

Logarithmically exact one-forms form an Abelian subgroup $L^1(\mathscr{M})$ in $Z^1(\mathscr{M})$. A closed one-form α is in $L^1(\mathscr{M})$, if there exists $h\in\mathbf{C}^\infty(\mathscr{M},S^1)$, with

$$\alpha(X) = ih^{-1}(Xh). \tag{6}$$

One can show[5] that $Z^1(\mathscr{M})/L^1(\mathscr{M}) \approx \pi_1^*(M)/(\Gamma\pi_1)^*(\mathscr{M})$; i.e., the elements of Z^1/L^1 give inequivalent quantizations. The dimension of Z^1/L^1 is finite and equal to the first Betti number b_1 of $\mathscr{M}$. The de Rham reconstruction theorem[6] tells us how to construct a basis $p_j, j = 1,...,b_1$ in the vector space Z^1/L^1 and linear combinations of p_j give those ω, or topological potentials, with inequivalent quantizations

$$\omega(X) = \sum_{j=1}^{b_1} \beta_j p_j(X), \quad 0 \leqslant \beta_j < 2\pi, \tag{7}$$

which are parametrized via $(\beta_1,...,\beta_{b_1},c)$. The explicit calculation of the p_j can be difficult, but is in principle possible. An example is treated in Sec. III.

The close connection between Borel kinematics and topology through the quantization method is related to the fact that the set $\mathscr{X}_c(\mathscr{M})$ contains infinite-dimensional Lie algebras comprising enough flows to feel the global geometry of $\mathscr{M}$.

B. Topology and energy observables, topological potentials of the first kind

We refer again to a conservative quantization of the energy, given through the Hamilton operator that is a function of the $\mathbf{P}(X)$ and $\mathbf{Q}(f)$, i.e., a partial differential operator. The ordering of the noncommutative factors must be prescribed, such that $\mathbf{H}$ is symmetric. (For a method to select certain ordering see Ref. 2. In general the measuring process of the energy observable should determine the ordering. This needs a theory of measurement, which is not yet available.) However, in general, this $\mathbf{H}$ is not essentially self-adjoint. Because this is needed for the quantum mechanical probability interpretation, we have to construct *self-adjoint extensions* $\mathbf{H}^\alpha$ of $\mathbf{H}$ (α is a label index). The extension $\mathbf{H}^\alpha$ should depend also on the topology of $\mathscr{M}$, as there are enough Borel sets on $\mathscr{M}$ and the inner product in $L^2(\mathscr{M},\mu)$ feels any $B\in\mathscr{B}(\mathscr{M})$. But unlike the situation in the previous section, a general extension theory for partial differential operators on a given $\mathscr{M}$ is not available that parametrize families of inequivalent extensions through geometrical objects living on $\mathscr{M}$. Different types of $\mathbf{H}$ have to be treated separately depending on $\mathscr{M}$. Powerful techniques are known. For the example in Sec. III we use a method of Rellich.[7] In special cases, extensions $\mathbf{H}^\alpha$ are equivalent to an additive term

$$\mathbf{H}^\alpha = \mathbf{H}^0 + V^\alpha, \quad V^\alpha: \mathscr{M} \mapsto \mathbf{R}, \quad V^0 = 0. \tag{8}$$

Here V^α is a 0-form and can be viewed as a topological potential of the first kind. For special Hamilton operators all inequivalent quantizations for some $\mathbf{H}$ are constructed in Sec. III and topological potentials are calculated.

III. THE HARMONIC OSCILLATOR ON THE POINTED R^2

A. The model, its quantization, and its Hamilton operator

Now we apply the approach to the manifold $\mathscr{M}_1 = \mathbf{R}^3 - \mathbf{R}^1$ and a particle (mass M) moving on $\mathscr{M}_1$ under an external harmonic potential $V(x,y,z) = (M\omega^2/2)(x^2+y^2)$ (x,y,z are local coordinates) and with the usual Hamilton function. The topology of $\mathscr{M}_1$ is not trivial: $\pi_1(\mathscr{M}) \cong \mathbf{Z}$, $\pi_1^*(\mathscr{M}_1) \cong \mathbf{R} \pmod{2\pi}$ and $b_1(\mathscr{M}_1) = 1$. For simplicity (see also Ref. 8) we choose $c = 0$ and furthermore $d\mu = dx\,dy\,dz$.

For a quantization of the kinematics a basis element p of $Z^1(\mathscr{M}_1)/L^1(\mathscr{M}_1)$ must be calculated[9]; one gets with the physically interesting vector fields $(X_s) = (\partial/\partial x, \partial/\partial y, \partial/\partial z)$ for the corresponding topological potential ω of the second kind (independent of the external potential)

$$p(X_s) = [1/(2\pi(x^2+y^2))](-y,x,0), \tag{9}$$

$$\omega(X_s) = \beta p(X_s), \quad 0 \leqslant \beta < 2\pi. \tag{10}$$

Here $\omega(X_s)$ has the form of a Bohm–Aharonov potential[3] $(e/\hbar)\mathbf{A}_{BA}(x,y,z)$, where $\hbar\beta/e$ is the magnetic flux of the (infinitely thin and long) $\mathbf{R}^1$ solenoid. The result is connected with the topology of the Bohm–Aharonov configuration of $\mathscr{M}_1$. We stress here that the appearance of $\omega(X_s)$ does not tell us anything on its physical interpretation. The quantization gives no hint how to realize a topological potential in practice. If there is some deeper connection between quantum mechanics on $\mathscr{M}_1$ and electromagnetism it is not known.

For the rest of the paper we interpret $\omega(X_s)$ as Bohm–Aharanov potential, in which a particle of charge e moves.

The quantization of the classical Hamiltonian for the system gives

$$\mathbf{H}^\alpha_{(1)} = \frac{1}{2M}\sum_s[-\hbar\mathbf{P}(X_s)]^2 + \mathbf{Q}(V)$$

or

$$\mathbf{H}^a_{(1)} = \frac{1}{2M}\left[-\hbar i\left(\frac{\partial}{\partial x},\frac{\partial}{\partial y},\frac{\partial}{\partial z}\right) - a\,\frac{1}{x^2+y^2}(-y,x,0)\right]^2 + \frac{M\omega^2}{2}(x^2+y^2), \tag{11}$$

where $a = \hbar\beta/(2\pi)$. Because of the axial symmetry the z part separates as a free system in $\mathbf{R}_z^{(1)}$ and we are left with a particle constrained on $\mathscr{M}_2 = \mathbf{R}^2 - \{0\} = \dot{\mathbf{R}}^2$ with a Hamilton operator

$$\mathbf{H}^a = \frac{1}{2M}\left[-\hbar i\left(\frac{\partial}{\partial x},\frac{\partial}{\partial y}\right) - a\frac{1}{x^2+y^2}(-y,x)\right]^2 + \frac{M\omega^2}{2}(x^2+y^2). \tag{12}$$

The $\mathbf{H}^a$ has to be essentially self-adjoint on a suitable domain $D\subset L^2(\dot{\mathbf{R}}^2, dx\,dy)$. These domains are not obvious [e.g., $\mathbf{H}^a$ is not even symmetric on the natural domain $D = \{\psi(x,y)\,|\,\psi\in L^2(\dot{\mathbf{R}}^2,dx,dy), C^\infty(\dot{\mathbf{R}}^2,C)\}$]. So we have to construct self-adjoint extensions $\mathbf{H}^{a,\alpha}$ of $\mathbf{H}^a$ as explained in Section II B. As a first step we will reduce $\mathbf{H}^a$ to differential operators on S^1 and $\mathbf{R}^+$ (Sec. III C 2). Then we will find all self-adjoint extensions $\mathbf{H}^{a,\alpha}$ (Sec. III C). We calculate the (discrete) spectra of $\mathbf{H}^{a,\alpha}$, discuss various cases, and show the appearance of (singular) topological potentials of the first kind (Sec. III D).

B. The radial Hamilton operator

In polar coordinates

$$x = r\cos\varphi,\quad y = r\sin\varphi,$$

the Hilbert space of square-integrable functions over $\dot{\mathbf{R}}^2$ can be decomposed as

$$\mathscr{L}^2(\dot{\mathbf{R}}^2,dx\,dy) = \mathscr{L}^2(\mathbf{R}^+\,r\,dr)\otimes\mathscr{L}^2(S^1,d\varphi) = \bigoplus_{m=-\infty}^{+\infty}\mathscr{L}^2(\mathbf{R}^+,r\,dr)\otimes[e^{im\varphi}], \tag{13}$$

where [ ] denotes span and $m\in\mathbf{Z}$. The differential operators (12) separate to

$$\mathbf{H}^a = -\frac{\hbar^2}{2M}\left(\partial_r^2 + \frac{1}{r}\partial_r - \frac{a^2}{\hbar^2 r^2}\right) + \frac{1}{2}M\omega^2 r^2 - \frac{\hbar^2}{2Mr^2}\left(\partial_\varphi^2 + \frac{2ia}{\hbar}\partial_\varphi\right). \tag{14}$$

The angular operator

$$\partial_\varphi^2 + (2ia/\hbar)\partial_\varphi \tag{15}$$

is essentially self-adjoint on $C^\infty(S^1)$,[10] with nondegenerate discrete spectrum $-m(m+2a/\hbar)$ and the eigenfunctions

$$e^{im\varphi},\quad m\in\mathbf{Z}. \tag{16}$$

Here $\mathbf{H}^a$ is essentially self-adjoint on a subset of $\mathscr{L}^2(\dot{\mathbf{R}}^2)$ iff all members of the family of radial Hamilton operators

$$\mathbf{H}^{a,m} \equiv \frac{\hbar^2}{2Mr}\left\{-\partial_r(r\,\partial_r) + \frac{1}{r}\left(m+\frac{a}{\hbar}\right)^2 + \frac{M^2\omega^2}{\hbar^2}r^3\right\} \tag{17}$$

on corresponding subsets $\vartheta_r^{a,m}$ of $\mathscr{L}^2(\mathbf{R}^+, r\,dr)$ are essentially self-adjoint.[10] Essentially self-adjoint extensions of $\mathbf{H}^a$ and of the family $\mathbf{H}^{a,m}$ have the same eigenvalues. The eigenfunctions differ by a factor $e^{im\varphi}$ only. Considered as differential operators on a suitable function space, $\mathbf{H}^{a,m}$ are singular Sturm–Liouville operators, for which Weyl's limit-point, limit-circle theory can be applied.[11] Note that $H^{a,m}$ depends on

$$\nu = |a/\hbar + m| \tag{18}$$

only.

For reasons of convenience we introduce the dimensionless quantity

$$k = E/(2\hbar\omega), \tag{19}$$

the new variable

$$s = (M\omega/\hbar)r^2, \tag{20}$$

and the new wave function

$$w(s) = ru(r). \tag{21}$$

Then the eigenvalue equation

$$\mathbf{H}^{a,m}u(\nu,E;r) = Eu(\nu,E;r)\quad\text{on } \vartheta_r^{a,m} \tag{22}$$

becomes on $\vartheta_s^{a,m}\subset\mathscr{L}^2(\mathbf{R}^+,s^{-1}\,ds)$

$$\mathbf{A}^\nu w(\nu,k;s) = -kw(\nu,k;s), \tag{23}$$

with

$$\mathbf{A}^\nu = s\left(\frac{d^2}{ds^2} - \frac{1}{4} + \frac{1-\nu^2}{4s^2}\right). \tag{24}$$

[We have a singular Sturm–Liouville eigenvalue problem $-(pu')' + qu = kKu$ in the interval $0<s<\infty$, with $p=1$, $K=1/s$, $q=\frac{1}{4}-(1-\nu^2)/(4s^2)$.]

C. Self-adjoint extensions of the radial Hamilton operators

Equation (23) is the Whittaker differential equation. So all solutions for given ν,k are linear combinations of the Whittaker functions[12]

$$w_1(\nu,k;s) = W_{k,\nu/2}(s), \tag{25}$$

$$w_2(\nu,k;s) = W_{-k,\nu/2}(-s). \tag{26}$$

For large s the asymptotic expansion holds,

$$W_{k,\nu/2}(s) = e^{-s/2}s^k\{1 + O(|s|^{-1})\}. \tag{27}$$

So w_1 falls off exponentially while w_2 grows exponentially. Hence the latter is not normalizable in $\mathscr{L}^2(\mathbf{R}^+,s^{-1}\,ds)$ and we have the limit-point case at $s=\infty$ for all ν.

For small s we recall the expansions

$$W_{k,\nu/2}(s) = e^{-s/2}s^{(\nu+1)/2}\begin{cases}\dfrac{\Gamma(\nu)}{\Gamma((1+\nu)/2-k)}s^{-\nu} + R_\nu & (\nu>0),\\[2mm] \dfrac{-1}{\Gamma(\frac{1}{2}-k)}F_k & (\nu=0,\ \frac{1}{2}-k\neq -n),\\[2mm] (-1)^n n! + O(s) & (\nu=0,\ \frac{1}{2}-k=-n),\end{cases} \tag{28}$$

where

$$F_k = [\ln s - 2\psi(1) + \psi(1/2-k) + O(|s \ln s|)],$$

$$R_\nu = \begin{cases} O(|s|^{\nu-1}), & \text{for } \nu>1, \\ O(\ln|s|), & \text{for } \nu=1, \\ O(1), & \text{for } 0<\nu<1, \end{cases}$$

ψ is the digamma function and $n = 0,1,\ldots$. To get the corresponding expansion for negative argument, replace $\ln(-s)$ by $i\pi + \ln s$.[12]

It can be easily seen that

$$\int_0^c |w_i|^2 \frac{ds}{s} < \infty, \quad \text{with } c>0, \quad i=1,2, \tag{29}$$

in the range $0 \leqslant \nu < 1$. But at least one of these solutions is not normalizable at $s=0$ for $\nu \geqslant 1$.

So at $s=0$ we have for $\nu \geqslant 1$ the limit-point case and for $0 \leqslant \nu < 1$ the limit circle case.

1. The case $\nu \geqslant 1$

We have the limit-point case at both ends of the interval $[0,\infty)$. Hence $\mathbf{A}^\nu$ is symmetric and even essentially self-adjoint[7] on

$$\vartheta_s^\nu = \{w | w \in \mathscr{L}^2(\mathbf{R}^+, s^{-1}ds), w'' \text{ absolutely continuous}\}. \tag{30}$$

To get the discrete spectrum we have to require that w_1 from Eq. (25) is normalizable at $s=0$, i.e., the leading term in the expansion (28) must vanish. Hence we get for the eigenvalues

$$k = n + [(\nu+1)/2], \quad n = 0,1,2,\ldots; \tag{31}$$

the eigenfunctions (25) can be expressed in terms of generalized Laguerre polynomials,

$$W_{k,\nu/2}(s) = (-1)^n n! e^{-s/2} s^{(\nu+1)/2} L_n^{(\nu)}(s). \tag{32}$$

2. The case $0 \leqslant \nu < 1$

Now we have the limit-point case at $s=\infty$ and the limit-circle case at $s=0$. Hence additional boundary conditions at $s=0$ are necessary. Given any normalizable function $v(s)$, $\mathbf{A}^\nu$ is symmetric on

$$\varphi_v = \{w | w \in \mathscr{L}^2(\mathbf{R}^+, s^{-1}\,ds), w'' \text{ absolutely continuous}, [w,v]_0 = 0\}, \tag{33}$$

with $[w,v]_s := v(s) w^{*\prime}(s) - v'(s) w^*(s)$. This can be seen easily by partial integration:

$$(w, \mathbf{A}^\nu v) - (\mathbf{A}^\nu u, v) = [w,v]_\infty - [w,v]_0. \tag{34}$$

For linear combinations w of functions w_1 from Eq. (25) with different k, we have $[w,v]_\infty = 0$. So $\mathbf{A}^\nu$ is symmetric on any domain, whose elements satisfy the boundary condition

$$[w,v]_0 = 0. \tag{35}$$

This is precisely the condition that makes $\mathbf{A}^\nu$ essentially self-adjoint.[7] Different functions $v(s)$ can label the same domain.

To make the boundary condition (35) more accessible, we follow Rellich[7] and use a special solution system of the Eq. (23). If linearly independent solutions $w_3(\nu,k;s)$, $w_4(\nu,k;s)$ of Eq. (23) satisfy

$$[w_i(\nu,k;s), w_i(\nu,k';s)]_0 = 0, \quad i=3,4,$$
$$[w_3(\nu,k;s), w_4(\nu,k';s)]_0 = -[w_4(\nu,k;s), w_3(\nu,k';s)]_0, \tag{36}$$

then the boundary condition (35) is valid for all linear combinations

$$w^\alpha(\nu,k;s) = \sin(\alpha) w_3(\nu,k;s) + \cos(\alpha) w_4(\nu,k;s), \tag{37}$$

with $0 \leqslant \alpha < \pi$ fixed. Using trigonometric functions here is just a convenient way to parametrize the coefficients. Different values of α specify different (inequivalent) essentially self-adjoint extensions $A^{\nu,\alpha}$.

For $\nu > 0$ a possible choice of ω_3 and ω_4 are the Whittaker functions[12]

$$w_3(\nu,k;s) = M_{k,-\nu/2}(s), \quad w_4(\nu,k;s) = M_{k,+\nu/2}(s), \tag{38}$$

as they behave at $s=0$ like

$$w_3(s) = s^{(1-\nu)/2}(1+O(s)),$$
$$w_4(s) = s^{(1+\nu)/2}(1+O(s)). \tag{39}$$

For $\nu=0$, these two solutions become identical. If w_4 has the same form as in Eqs. (39) and (38), and if w_3 has an expansion

$$w_3(s) = s^{1/2}(1+O(s)) \ln s, \tag{40}$$

then the conditions (36) are satisfied. The explicit form depends on k. We can use

$$w_3(0,k;s) = -\Gamma(\tfrac{1}{2}+k) W_{-k,0}(-s) + (2\psi(1) - \psi(\tfrac{1}{2}+k) - i\pi) M_{k,0}(s) \tag{41}$$

for $k \in \{-\frac{1}{2}, \frac{1}{2}, \frac{3}{2}, \ldots\}$, and

$$w_3(0,k;s) = -\Gamma(\tfrac{1}{2}-k) W_{k,0}(s) + (2\psi(1) - \psi(\tfrac{1}{2}-k)) M_{k,0}(s) \tag{42}$$

for all other $k \in \mathbf{R}$. Hence, for each value $0 \leqslant \nu < 1$ there exists an essentially self-adjoint extension $\mathbf{A}^{\nu,\alpha}$, with $0 \leqslant \alpha < \pi$, whose domain is

$$\vartheta_{\nu,\alpha} = \{w | w \in [w^\alpha], \ w \in \mathscr{L}^2(s^{-1}\,ds), w'' \text{ absolutely continuous}\}, \tag{43}$$

where $[w^\alpha]$ is the span of the ω^α in Eq. (37) with w_3, w_4 from Eqs. (38), (41), and (42).

Next we calculate the *spectra*. The normalizable eigenfunctions of $\mathbf{A}^{\nu,\alpha}$ are those functions w^α from Eq. (37) which are normalizable at $s=\infty$, i.e., which are multiples of w_1. For $0<\nu<1$ we use the expansion of w_1,[12]

$$w_1(\nu,k;s) = \frac{\Gamma(\nu)}{\Gamma((1+\nu)/2-k)} w_3(\nu,k;s) + \frac{\Gamma(-\nu)}{\Gamma((1-\nu)/2-k)} w_4(\nu,k;s). \tag{44}$$

Comparison with Eq. (37) gives the discrete spectrum of $\mathbf{A}^{\nu,\alpha}$ as solutions of (α and ν are fixed)

$$\tan\alpha = \frac{\Gamma(\nu)\Gamma((1-\nu)/2-k)}{\Gamma(-\nu)\Gamma((1+\nu)/2-k)}. \tag{45}$$

For $\alpha=0$ we have

$$k = [(1+\nu)/2] + n, \quad n = 0,1,2,3\ldots, \tag{46}$$

and for $\alpha = \pi/2$,

$$k = (1-\nu)/2 + n. \tag{47}$$

The eigenfunctions are given by Eq. (25).

For a numeric evaluation of the discrete spectrum (45) in the case of $\nu = \frac{1}{4}$ see Fig. 1. In the limit $\alpha \to \pi$, the eigenvalues tend towards those given for $\alpha = 0$, except for the lowest one, which goes to $-\infty$.

For $\nu = 0$, $k \in \{-\frac{1}{2}, \frac{1}{2}, \frac{3}{2}, \ldots\}$, we have to require that the coefficient of $W_{-k,0}(-s)$ vanishes ($w_1 \sim w_4$ in this case). This implies $\alpha = 0$. So for $\nu = 0$ and $\alpha = 0$ we get

$$k = \tfrac{1}{2} + n; \tag{48}$$

the normalized eigenfunctions (25) can be expressed using Laguerre polynomials

$$w_1(0,k;s) = n! e^{-s/2}\sqrt{s} L_n(s). \tag{49}$$

For all other k, comparison of Eqs. (28), (37), and (42) gives

$$\sin\alpha(2\psi(1) - \psi(\tfrac{1}{2} - k)) + \cos\alpha = 0$$

or

$$\tan\alpha = -(2\psi(1) - \psi(\tfrac{1}{2} - k))^{-1}. \tag{50}$$

Here $\alpha = 0$ yields $k = n + \frac{1}{2}$ (the case was discussed above), and $\alpha = \pi/2$ gives for k

$$\psi(\tfrac{1}{2} - k) = 2\psi(1). \tag{51}$$

For a numeric evaluation of Eq. (50) see Fig. 1. In the limit $\alpha \to \pi$, again the eigenvalues tend towards those given for $\alpha = 0$, except for the lowest one which goes to $-\infty$. The eigenfunctions are given by Eq. (25). All the eigenvalues calculated above are nondegenerate.

D. Self-adjoint extensions of the reduced Hamilton operator

The results in Secs. III B and III C give immediately the self-adjoint extensions of $\mathbf{H}^a$ together with their eigenvalues and eigenfunctions.

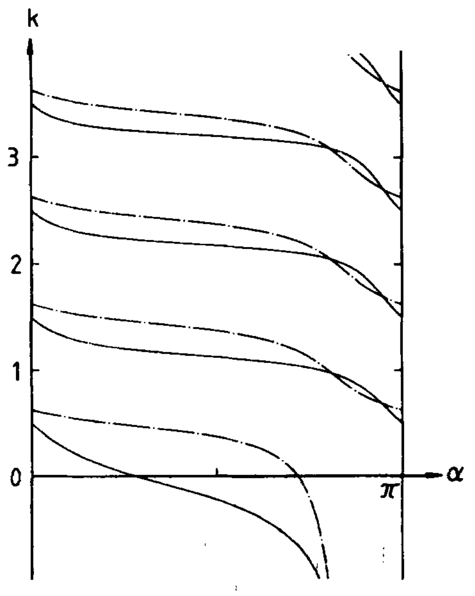

FIG. 1. Eigenvalues of the self-adjoint extensions $-A^{0,\alpha}$ (full lines) and $-A^{1/4,\alpha}$ (dashed lines).

For $0 \leqslant \nu = |a/\hbar + m| < 1$ the radial operators $\mathbf{A}^\nu$ have a one-dimensional family of self-adjoint extensions $\mathbf{A}^{\nu,\alpha}$ parametrized by $0 \leqslant \alpha < \pi$. The α are independent for different ν. For $\nu > 1$ the $\mathbf{A}^\nu$ are essentially self-adjoint.

For $\mathbf{H}^a$ with

$$\mathbf{H}^a \Psi(E,m;r,\varphi) = E\Psi(E,m;r,\varphi), \tag{52}$$

$$\Psi(E,m;r,\varphi) = W_{k,\nu/2}((M\omega/\hbar)r^2) r^{-1} e^{im\varphi}, \tag{53}$$

where $\Psi(E,m;r,\varphi)$ is a simultaneous eigenfunction of $\mathbf{H}^a$ and the angular operator (15), we get the following scenario.

For vanishing Bohm–Aharonov potential $\mathbf{H}^0$ has a one-dimensional family of self-adjoint extensions $\mathbf{H}^{0,\alpha_0}$, because $0 \leqslant \nu = |m| < 1$ implies $m = 0$.

For nonvanishing Bohm–Aharonov potential we only have to consider $0 < a < \hbar$; other values can always be obtained by shifting m by an integer value, as the spectrum depends only on the combination $\nu = |a/\hbar + m|$. This corresponds to the range of β in Eq. (10), and to a change of magnetic flux in the solenoid by $2\pi\hbar/e$. In the Bohm–Aharonov experiment this maps the interference pattern into itself. For $a \neq 0$ a two-dimensional family $\mathbf{H}^{a,\alpha_0,\alpha_1}$ appears, where α_1 parametrizes extensions for those $m = m_1 \neq 0$ with $0 \leqslant |a/\hbar + m| < 1$. The eigenfunctions are, as indicated in (53), up to a factor $r^{-1}e^{im\varphi}$ Whittaker functions.

We give now a detailed discussion.

1. The case $a = 0$

The system corresponds to an oscillator in $\dot{\mathbf{R}}^2$; $\alpha_0 = 0$ gives the kinetic-energy extension $\mathbf{H}^{0,0}$. Collecting the results for $\nu = 0$, $0 < \nu < 1$, and $1 \leqslant \nu$ we have the energy spectrum

$$E^{0,0}_{n,m} = \hbar\omega(1 + 2n + |m|), \tag{54}$$

which is the spectrum of the oscillator in $\mathbf{R}^2$. The spectrum is degenerate. It can be written as

$$E^{0,0}_{n'} = \hbar\omega(n' + 1), \quad n' = 0,1,\ldots \tag{55}$$

with a ($d_{n'} = n' + 1$)-fold degeneracy given by

$$m = -n', -n' + 2, \ldots, n'.$$

The normalized eigenfunctions are

$$\Psi(E^{0,0}_{n,m}, m; r,\varphi) = (-1)^n n! (M\omega/\hbar)^{(|m|+1)/2} r^{|m|} \times e^{-(M\omega/2\hbar)r^2} L_n^{|m|}\left(\frac{M\omega}{\hbar} r^2\right) e^{im\varphi}. \tag{56}$$

The kinetic-energy extension is not the only one. (As the introduction of polar coordinates restricts $\mathbf{R}^2$ to $\dot{\mathbf{R}}^2$, the treatment of any Hamiltonian on $\mathbf{R}^2$ with polar coordinates is delicate.) $\mathbf{H}^0$ has additional self-adjoint extensions $\mathbf{H}^{0,\alpha_0}$, which stem from those of $\mathbf{A}^0$. Hence the oscillator in $\dot{\mathbf{R}}^2$ can have energy spectra—depending on α_0—different from the oscillator spectrum in $\mathbf{R}^2$. As compared to $\mathbf{H}^{0,0}$, the degeneracy of the eigenvalues for even n' becomes reduced by 1 to $d_{n'} = n'$. There appear additional eigenvalues with $E^{0,0}_{n'-1} < E^{0,\alpha_0}_{n'} < E^{0,0}_{n'}$ for even n'. The ground state is shifted to an energy $E^{0,\alpha_0}_0 < E^{0,0}_0$. The situation is given in Fig. 2 (crosses) for $a = 0$, $\alpha_0 = \frac{1}{2}$. The eigenfunctions for $E^{0,\alpha_0}_{n'}$ are

$$\Psi^{\alpha_0}(E^{0,\alpha_0}_{n'}, 0; r,\varphi) = W_{k_n,0}((M\omega/\hbar)r^2) r^{-1}, \tag{57}$$

where k_n are solutions of Eq. (50) for $\alpha = \alpha_0$.

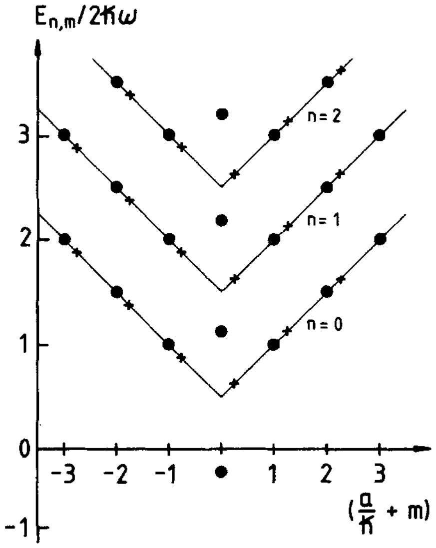

FIG. 2. Eigenvalues of the self-adjoint extensions $H^{0,1/2}$ (crosses) and $H^{\hbar/4,0,0}$ (dots) as a function of $a/\hbar + m$.

In connection with the discussion in Sec. II B it is interesting to check whether the nontrivial topological structure of $\dot{\mathbf{R}}^2$ which is responsible for the exotic oscillators or extensions, can be transferred into a potential that is, e.g., localized in $0 < r < r'$ with $r' < 1$. For this we have to analyze the behavior of $\Psi^{\alpha_0}(E,0;r,\varphi)$ for small r. The leading term is proportional to

$$\ln r. \tag{58}$$

The action of the differential operator $\mathbf{H}^0$ on $\Psi^{\alpha_0}(E,0;r,\varphi)$ gives an additional term that can be compensated via a point potential

$$V_p(\mathbf{x}) = -\frac{\hbar^2\pi}{M}\frac{\delta^2(\mathbf{x})}{\ln r} \tag{59}$$

in a distribution theoretic sense (see, e.g., Ref. 13). Obviously, V_p is a topological potential of the first kind.

2. The case $a \neq 0$

For $\alpha_1 = \alpha_0 = 0$ we have the kinetic-energy extension $\mathbf{H}^{a,0,0}$. As above the energy spectrum is (dots in Fig. 2)

$$E^{a,0,0}_{n,m} = \hbar\omega(1 + 2n + |a/\hbar + m|). \tag{60}$$

Comparing with the spectrum of $\mathbf{H}^{0,0}$ we see that the Bohm–Aharonov potential shifts all energies of the oscillator in $\mathbf{R}^2$. It does this by shifting the "orbital angular momentum" m to $m + a/\hbar$ via the centrifugal part of the potential energy $r^{-2}(m + a/\hbar)^2$ in $\mathbf{H}^a$. The eigenfunctions are

$$\Psi^{0,0}(E^{a,0,0}_{n,m},m;r,\varphi) = W_{k,\nu/2}(M\omega/\hbar)r^2 e^{im\varphi} r^{-1}, \tag{61}$$

with k from Eq. (19). For $\alpha_1 \neq 0$ and/or $\alpha_0 \neq 0$ we have, as in Sec. III D 1, exotic extensions. We find the same eigenvalues as for $\mathbf{H}^{a,0,0}$, but for any $m = 0$, and $m = m_1$ the corresponding eigenvalue $E^{a,\alpha_0,\alpha_1}_{n,m}$ is smaller than $E^{a,0,0}_{n,m}$ and decreases with increasing α_0, α_1 to the next eigenvalue of $\mathbf{H}^{a,0,0}$ respectively to $-\infty$. The eigenfunctions are as above, inserting the energies $E^{a,\alpha_0,\alpha_1}_{n,m}$. The case $a = 0$, $\alpha_0 > 0$ cannot be obtained continuously from $a > 0$, $\alpha_0 > 0$, as the two bases (35) and (34), which are used in the definition of the parameter α, are not connected continuously. We have no result as to whether this topological effect can be related to a (singular) topological potential.

IV. APPLICATION TO A CHARGE IN A HOMOGENEOUS MAGNETIC FIELD AND A BOHM–AHARONOV POTENTIAL

A physical setup to realize experimentally the boundary conditions necessary for the exotic self-adjoint extensions seems not to be known, but the effect of a Bohm–Aharonov potential on the energy eigenvalues of a two-dimensional oscillator (kinetic energy extension) could in principle be measurable.

To explain this we consider a particle moving in the superposition

$$\mathbf{A} = \mathbf{A}_{BA} + (B/2)(-y,x,0) \tag{62}$$

of a Bohm–Aharonov potential $\mathbf{A}_{BA}$ and a homogeneous magnetic field perpendicular to the xy plane. As in Sec. III C 1 the z part separates. The corresponding radial Hamilton operator is

$$\mathbf{H}^{B,a,m} = \mathbf{H}^{a,m} + e\hbar B/(2M)(m + a/\hbar), \quad m \in \mathbf{Z}, \tag{63}$$

with $\mathbf{H}^{a,m}$ from Eq. (17), where

$$\omega = eB/(2M), \tag{64}$$

that is, in our choice of gauge the homogeneous magnetic field gives the potential of a two-dimensional harmonic oscillator. The extensions of $H^{B,a,m}$ are obtained from those of $H^{a,m}$. The energies, i.e., the Landau levels in the presence of the Bohm–Aharonov potential, are for the kinetic energy extension:

$$E^{B,a}_{n,m} = (e\hbar B/2M)(2n + 1 + |m + a/\hbar| + m + a/\hbar). \tag{65}$$

They split in two level families: The usual Landau levels for $a = 0$ $(m \leqslant -1)$

$$E^{B}_{n'} = (e\hbar B/2M)(1 + 2n'), \quad n' = 0,1,2,..., \tag{66}$$

which are infinitely degenerate, and an additional family depending on a $(m \geqslant 0)$

$$E^{B,a}_{n'} = (e\hbar B/2M)(1 + 2n' + 2a/\hbar), \tag{67}$$

which are

$$(1 + n')\text{-fold} \tag{68}$$

degenerate; note that $0 \leqslant a < \hbar$.

The experimental problem is to measure the additional levels (67). Considerations on whether the magnitude of the effect is big enough to be measured with present day equipment are under way.

ACKNOWLEDGMENTS

We would like to thank B. Angermann and W. Lücke for discussions and J. Tolar for a critical reading of the manuscript. One of the authors (H.D.D.) is deeply indebted for help and encouragement to Professor H. Kreuzer, Dr. F. Schulze, and the staff of unit 3023 of the university hospital at Göttingen, where part of the paper was written.

[1]B. Angermann, H. D. Doebner, and J. Tolar, in "Non-linear partial differential equations and quantization procedures," edited by H. D. Doebner and S. I. Andersson, *Lecture Notes in Mathematics*, Vol. 1037 (Springer, Berlin, 1983), p. 171.
[2]S. T. Ali and H. D. Doebner, in "The physics of phase space," edited by Y. S. Kim and W. W. Zachary, *Lecture Notes in Physics*, Vol. 278 (Springer, Berlin, 1987), p. 330.
[3]A. Aharonov and D. Bohm, Phys. Rev. **115**, 485 (1959).
[4]L. S. Schulmann, J. Math. Phys. **12**, 304 (1971).
[5]B. Kostant, in "Lectures in modern analysis and applications III," edited by C. T. Taam, *Lecture Notes in Mathematics*, Vol. 170 (Springer, New York, 1970), p. 87.
[6]W. Greub, S. Halpern, and R. Vanstone, *Connections, Curvature, and Cohomology* (Academic, New York, 1972), Vol. 1, p. 233.
[7]K. Jörgens and F. Rellich, *Eigenwerttheorie gewöhnlicher Differentialgleichungen* (Springer, New York, 1976).
[8]G. A. Goldin, in *Infinite Dimensional Lie Algebras and Quantum Field Theory*, edited by H. D. Doebner, H. D. Hennig, and T. D. Palev (World Scientific, Singapore, 1988).
[9]H. D. Doebner, Lecture given at the Bulgarian Summer School at Primorsko (1980).
[10]M. Reed and B. Simon, *Methods of Modern Mathematical Physics* (Academic, New York, 1972), Vol. 1 and Vol. 2.
[11]H. Weyl, Math. Ann. **68**, 220 (1909).
[12]*Handbook of Mathematical Functions*, edited by M. Abramowitz and I. A. Stegun (Dover, New York, 1965).
[13]*Solvable Models in Quantum Mechanics*, edited by S. Albeverio, F. Gesztesy, R. Høegh-Krohn, and H. Holden (Springer, New York, 1988).

Technical methods for solving bound-state equations in momentum space

S. Boukraa
LPTHE,[a] Université P. et M. Curie, T16 E1, 4 place Jussieu, F-75230 Paris Cedex 05, France

J.-L. Basdevant
Division de Physique Thèorique,[a] Institut de Physique Nucleaire, F-91406 Orsay and LPTPE, Université P. et M. Curie, T16 E1, F-75230 Paris Cedex 05, France

(Received 17 June 1988; accepted for publication 30 November 1988)

Some technical methods used to solve two- and three-body bound-state equations in momentum space are described. These methods are, in particular, very efficient when applied to the calculation of spectra in semirelativistic confining potential Hamiltonians, which are used in hadron spectroscopy. They have a vast range of applicability.

I. INTRODUCTION

It is remarkable how well the structure of hadrons (mesons and baryons) is described by phenomenological potential models. The interest in these models arose after the discovery of the charmonium and bottomium families, which were calculated as nonrelativistic bound states of heavy quarks (c and b) in a confining potential.[1–4] Subsequently, it was shown that the masses of hadrons containing the light s quark could also be calculated with the same confining potential.[5] These successes have renewed interest in the study of the spectra and properties of the nonrelativistic Schrödinger equations with confining potentials and have led to a wide literature ranging from phenomenological models to more rigorous results.[6,7]

Recently, the desire to also include states composed of the lighter quarks (u and d) implied the inclusion of some relativistic effects. Many models have been proposed.[8–12] A simple way to incorporate these effects consists of replacing the nonrelativistic kinematics by the form $T=\Sigma_i(p_i^2+m_i^2)^{1/2}$. This leads naturally to work in momentum space with integral equations.[8,13–15] These integral equations appear to be singular for confining potentials and hence need a particular treatment.

On the other hand, the necessity of also considering baryons that are made of three quarks in the same framework led to the development of techniques for solving the three-body problem in momentum space, where the problem can be reduced to a set of coupled singular integral equations by using hyperspherical formalism.[16–19]

Our aim in this paper is to show these useful techniques, which are not discussed very much in the literature and which can also be applied to other problems in physics.

We begin by describing the two-body problem. In Secs. II A and B, after writing the integral equation for a two-particle system, we study the kernel of this equation for some power-law potentials that are the most frequently used in phenomenological applications. In Sec. II C we then describe the Vekua–Magnaradze method,[20] which can be applied for solving the Schrödinger equation with a linear potential. Then, in Sec. II D we show a very useful numerical method: the Multhopp technique[21,22] for finding the eigenstates of a Hamiltonian. In Sec. II E, we apply the Multhopp technique for calculating the spectra of the linear and logarithmic potentials.

We then consider the three-body problem. In Sec. III A, we work in the momentum space and then separate the coordinates into internal and external ones as in Omnès.[23] In Sec. III B, we define the hyperspherical harmonics and concentrate particularly on two different sets, as described and used by Fabre de la Rippelle, Simonov, and others.[24] In Sec. III C, using the internal and external coordinates, we show how one can separate the total angular momentum to end with a set of coupled integral equations in only the internal coordinates for each fixed value of the angular momentum. We also study the particular case of zero angular momentum and show that the problem reduces to a set of coupled integral equations in one variable, with kernels similar to those of the two-body problem. These equations are very suitable for numerical calculations. Finally, we give as an example the spectrum of hyper-radial potential in Sec. III D.

II. THE TWO-BODY PROBLEM

Consider a two-particle system described by the following general bound-state equation:

$$\left[\sum_{i=1}^{2} h(p_i)+V(|\mathbf{r}_1-\mathbf{r}_2|)-E\right]|\Psi\rangle=0, \tag{1}$$

where $|\Psi\rangle$ is the wavefunction of the system and E is its energy. The particle (i) is described by its position operator $\mathbf{r}_i$ and its momentum operator $\mathbf{p}_i$. We assume that the kinematics $\Sigma h(p_i)$ is a function of only p_i (the magnitude of $\mathbf{p}_i$) and that the potential is radial.

The interest in this general form of the kinetic part is to allow the treatment as special cases of a semirelativistic kinematics by setting $h(p)=(p^2+m^2)^{1/2}$ (m is the mass of the particle), of a nonrelativistic kinematics by setting $h(p)=m+p^2/2m$, or any intermediate form.

A. The two-body problem in momentum space

In the center of mass system, we obtain the l-wave projection of Eq. (1) in momentum space by extending the

[a] Laboratory associated with CNRS.

 0022-2488/89/051060-13$02.50

wavefunction $|\Psi\rangle$ on spherical harmonics. Equation (1) becomes an integral equation:

$$(2h(p) - E)u_l(p) + \int_0^\infty dp'\, u_l(p')V_l(p,p') = 0, \tag{2}$$

with

$$V_l(p,p') = (pp')^{1/2}\int_0^\infty r\,dr\, V(r)J_{l+1/2}(pr)J_{l+1/2}(p'r), \tag{3}$$

where p and r are the magnitudes of the relative momentum and coordinate of the two particles and $u_l(p)$ is the reduced wavefunction in momentum space. Here $V_l(p,p')$ is the kernel of the integral equation (2) expressed in terms of the usual Bessel functions $J_{l+1/2}(x)$.

In Sec. III, we will show that with kernels of the same form, similar coupled integral equations can be obtained for the three-body problem.

B. The kernels of the integral equations for power-law potentials

The analytic expression of the kernel $V_l(p,p')$ in Eq. (3) in terms of p and p' is not always easy to obtain for all potentials and for an arbitrary value of the angular momentum l.

The behavior of these kernels when $p \to p'$ is important. If the kernel is singular, Eq. (2) is a singular integral equation. The kernel can have a delta function (or derivative of delta functions) behavior, in which case Eq. (2) reduces to a differential equation. The kernel can also be regular and well behaved, leading to a nonsingular integral equation.

To see this explicitly, we will study these kernels for power-law potentials that contain as particular cases the confining potentials, which are of great interest in phenomenological applications.

For a potential $V(r) = r^\beta$, one can easily show the following general formula:

$$V_l(p,p') = \frac{2^{\beta/2-1}}{(2\pi)^{1/2}}\frac{\Gamma(l+3/2+\beta/2)}{\Gamma(l+1)\Gamma(-\beta/2)}\frac{1}{(pp')^{(1+\beta)/2}} \times \int_{-1}^{+1} du \frac{(1-u^2)^l}{[(p^2+p'^2)/2pp' - u]^{l+3/2+\beta/2}}. \tag{4}$$

Equation (4) is valid for all values of β where the gamma functions are defined. The values excluded are $\beta = 0,2,4,\ldots$ or $\beta = -3,-5,-7,\ldots$. We will discuss these two particular cases below: The integral on the rhs of Eq. (4) reduces to well-known functions.

1. Power-law potentials less singular than 1/r

(i) For the potentials $V(r) = r^{2n-1}$ with $n = 0,1,2,\ldots$, one has

$$V_l(p,p') = \frac{(2n)!}{2^n n!\pi}\frac{1}{(pp')^n} Q_l^{[n]}\left(\frac{p^2+p'^2}{2pp'}\right), \tag{5}$$

with $Q_l^{[n]}(x) = (d^n/dx^n)Q_l(x)$ and $Q_l(x)$ a Legendre function of the second kind (see, for example, Ref. 25).

(ii) For the potentials $V(r) = r^{2n}$ with $n = 0,1,2,\ldots$, $V_l(p,p')$ becomes a distribution expressed in terms of derivatives of delta functions as follows:

$$V_l(p,p') = \left[-\frac{d^2}{dp^2} + \frac{l(l+1)}{p^2}\right]^n \delta(p-p').$$

(iii) For the potentials $V(r) = r^\beta, \beta > -1$, where β is not an integer, the kernel can be expressed in terms of a generalized Legendre function of second order as in Eq. (5) by replacing n by $(\beta+1)/2$, which is not necessarily an integer.

Notice that for all these potentials the kernels $V_l(p,p')$ are singular when $p \to p'$. The singularity goes from logarithmic $\ln|p-p'|$ for the Coulomb potential $V(r) \to 1/r$ to more singular parts, for example, $1/(p-p')^2$ for the linear potential $V(r) = r$ or to delta function behavior, as in case (ii).

Knowing these singularities is important in the numerical calculations since one can integrate them out explicitly, leading to a rapid convergence of the numerical methods used: For this see the description of the Multhopp technique given in Sec. II D.

2. Power-law potentials more singular than 1/r

Attractive potentials more singular than $1/r$ cause problems in bound-state equations and the eigenvalues are not always defined. Let us recall that with nonrelativistc kinematics, $h = p^2$, the Hamiltonian is bounded from below if the potential is not more singular than $(-1/4r^2)$. We must also point out the remarkable analysis of Herbst,[26] who has shown that the relativistic Coulomb Hamiltonian $H = (p^2+m^2)^{1/2} - \lambda/r$ is bounded from below if and only if $\lambda < 2/\pi$.

However, combinations of these potentials can have a well-defined spectrum if a repulsive part dominates at small distances, at least in the nonrelativistic case. Exact solutions are known in this case,[27] and are also of particular interest in some physical problems (see, for example, a long discussion in Ref. 28). Therefore, it is also useful to know the kernels $V_l(p,p')$ for these singular potentials.

(i) From Eq. (4), the following formula can be easily written for the potentials $V(r) = 1/r^{2n+1}$ with $n \geqslant 0$ (not necessarily an integer):

$$V_l(p,p') = \frac{(-)^n n!(l-n)!}{\pi(2n)!(l+n)!}\frac{(p^2-p'^2)^{2n}}{(2pp')^n} Q_l^{[n]}\left(\frac{p^2+p'^2}{2pp'}\right).$$

We remark that the kernels $V_l(p,p')$ have no singularity when $p \to p'$.

(ii) However, the formula in (i) presents some problems. For example, when applied to the potential $1/r^3$, we see that the $l=0$ kernel is not defined. Similarly, for the potential $1/r^5$, the $l=0$ and $l=1$ kernels are not defined, etc. This problem comes from the definition of the gamma functions as mentioned above.

We can show as an example a regularization of the kernel $V_0(p,p')$ for $1/r^3$. For this, let us calculate $V_0(p,p')$ for the following potential:

$$V^{(\epsilon)}(r) = (1/2r_0^3)[(r_0/r)^{3+\epsilon} + (r_0/r)^{3-\epsilon}],$$

which is a regularization of the potential $1/r^3$ with a fixed value ϵ. The constant r_0 fixes a scale for this regularized potential and we can easily show that

$$V_0(p,p') = (1/2\pi)[(p-p')^2 \ln(r_0|p-p'|) - (p+p')^2 \times \ln(r_0|p+p'|)] + [(3-2\gamma)/\pi]pp' + o(\epsilon^2),$$

where γ is Euler's constant. We notice that the term of order $1/\epsilon$ that causes the problem of definition of the gamma functions has been canceled by adding the two contributions of $1/r^{3+\epsilon}$ and $1/r^{3-\epsilon}$. Thus the regular part of $V_0(p,p')$ is the kernel of the potential $\lim_{\epsilon\to 0} V^{(\epsilon)}(r)$. This behavior is always reproduced in any other different regularization of this singular potential. An application of this behavior is shown in Ref. 29.

We also notice that the kernel $V_1(p,p')$ for $l = 1$ has no problem and is well defined:

$$V_1(p,p') = \frac{(p^2-p'^2)^2}{8\pi pp'} \ln\left|\frac{p-p'}{p+p'}\right| + \frac{(p^2+p'^2)}{4\pi};$$

the same is also true for $V_l(p,p')$ with $l \geqslant 1$. However, notice that when regularized, $V_0(p,p')$ has lost its scaling property when $r \to \lambda r$, as expected.

3. The Fourier transforms for power-law potentials

It is also interesting to see that the kernel $V_0(p,p')$ can be easily related to the Fourier transform of the potential $V(r)$ by

$$-\pi \frac{d}{dp} V_0(p,p') = F(p-p') - F(p+p'),$$

with

$$\frac{2}{q} F(q) = \int dr\, V(r) e^{iqr}.$$

The above formulas allow us to calculate Fourier transforms of potentials once $V_0(p,p')$ is known. In Table I we give some Fourier transforms of the power-law and logarithmic potentials. These results generalize those of Ref. 30. Notice, also, that for any value of ν, one has

$$F^{\nu+2}(q) = -\frac{d^2}{dq^2} F^{\nu}(q),$$

where $F^{\nu}(q)$ is the Fourier transform of the potential r^{ν}.

TABLE I. Fourier transforms of power-law potentials. Here r_0 is the parameter chosen for the regularization of $1/r^3$ and $1/r^5$ and γ is the Euler constant.

Potential $V(r)$	Fourier transform $F(q)$
r^{ν}	$-(2^{\nu+1}/\pi^{1/2})\{\Gamma[(3+\nu)/2]$ $/\Gamma(-\nu/2)\}\lvert q\rvert^{-2-\nu}\,\mathrm{sgn}(q)$
$\nu \neq 0,2,4,\ldots$	
$\nu \neq -3,-5,\ldots$	
r^{2k}	$(-)^k \delta^{(2k+1)}(q)$
$k = 0,1,2,\ldots$	
$(1/r^3)^{\text{regularization}}$	$(q/\pi)[\ln\lvert r_0 q\rvert - 1 + \gamma]$
$(1/r^5)^{\text{regularization}}$	$-(q^3/6\pi)[\ln\lvert r_0 q\rvert - \frac{11}{6} + \gamma]$
$\vdots$	
$\ln r$	$\delta'(q)\ln(2/\gamma) + \mathrm{sgn}(q)/q^2$

C. Linear potential: The Vekua–Magnaradze method

When the most singular part of the potential at infinity is linear, the very clever method of Vekua and Magnaradze, described in Ref. 20 and used, e.g., in Ref. 31, gives particularly useful results.

1. The Vekua–Magnaradze method

We shall describe the Vekua–Magnaradze method on a simple example that can be easily generalized.

In what follows, we shall manipulate principal value integrals that we denote with the symbol $⨍$. We recall that

$$P\left(\frac{1}{x}\right) = \frac{1}{2}\left[\frac{1}{x-i\epsilon} + \frac{1}{x+i\epsilon}\right],$$

$$P\left(\frac{1}{x^2}\right) = \frac{1}{2}\left[\frac{1}{(x-i\epsilon)^2} + \frac{1}{(x+i\epsilon)^2}\right].$$

We also recall that if $\varphi(x)$ is regular on $x \in [a,b]$ and if $\varphi(a) = \varphi(b) = 0$, then

$$\frac{d}{dx} ⨍_a^b \frac{\varphi(y)dy}{y-x} = ⨍_a^b \frac{\varphi(y)dy}{(y-x)^2} = ⨍_a^b \frac{\varphi'(y)dy}{y-x}, \quad (6)$$

where $\varphi'(y) = d\varphi(y)/dy$.

Consider the one-dimensional eigenvalue problem

$$(K(p^2) + r)f(x) = Ef(x), \quad (7)$$

where $r = |x|$ and $K(p^2)$ is a kinetic energy operator, for example, $K = p^2/2m$, $K = (p^2+m^2)^{1/2}$, or any intermediate form depending on $p^2 = -d^2/dx^2$. The normalization conditions on f implies

$$\int_{-\infty}^{+\infty} f(x)dx = 1 \Rightarrow f(+\infty) = f(-\infty) = 0. \quad (8)$$

If we Fourier transform Eq. (7) and define

$$A(p^2) = K(p^2) - E,$$

we obtain the singular integral equation

$$A(p^2)f(p) = \frac{1}{\pi}\frac{d}{dp} ⨍_{-\infty}^{+\infty} \frac{f(q)dq}{q-p} = \frac{1}{\pi} ⨍_{-\infty}^{+\infty} \frac{f'(q)dq}{q-p}. \quad (9)$$

Equation (9) can be inverted by standard techniques,[20] giving

$$f'(p) = \frac{1}{\pi} ⨍_{-\infty}^{+\infty} \frac{A(q^2)f(q)}{q-p}\, dq. \quad (10)$$

The Vekua–Magnaradze trick consists in isolating the singular part of (10) as

$$f'(p) = -\frac{A(p^2)}{\pi} ⨍_{-\infty}^{+\infty} \frac{f(q)dq}{q-p} + g(p),$$

where the regular part $g(p)$ is

$$g(p) = \int_{-\infty}^{+\infty} R(p,q)f(q)dq, \quad R = \frac{1}{\pi}\frac{(A(q^2) - A(p^2))}{(q-p)}. \quad (11)$$

After some simple manipulations and making use of the initial equation (9), one obtains

$$\frac{1}{A}\frac{d}{dp}\left(\frac{1}{A}\frac{d}{dp} f(p)\right) + f(p) = G(p), \quad (12)$$

with

$$G(p) = \frac{1}{A}\frac{d}{dp}\left(\frac{1}{A} g(p)\right).$$

We introduce the variable $\tau(p)$ such that $d\tau = A(p^2)dp$, i.e.,

$$\tau(p) = \int_0^p A(p^2)dp = \int_0^p K(p^2)dp - Ep. \qquad (13)$$

Equation (12) is now

$$\frac{d^2}{d\tau^2}f + f = G,$$

which can be regarded as a second-order differential equation if G is considered given. After some simple manipulations, the general solution is expressed as

$$f(p) = a\cos\tau(p) + b\sin\tau(p) + \int_0^p g(p)\cos(\tau(p) - \tau(q))dq,$$

where a and b are arbitrary constants.

We note that $\tau(p)$ is odd in p and that if f is even g is odd and vice versa [see Eq. (11)]. If we impose the boundary conditions (8), we obtain an integral equation for f [remember g is expressed in terms of f through Eq. (11)]:

$$f(p) = \int_{-\infty}^p g(q)\cos(\tau(p) - \tau(q))dq \qquad (14)$$

and the quantization conditions for even and odd solutions:

$$\int_0^\infty g(q)\sin\tau(q)dq = 0 \quad \text{(even)}, \qquad (15a)$$

$$\int_0^\infty g(q)\cos\tau(q)dq = 0 \quad \text{(odd)}, \qquad (15b)$$

which are equivalent to $f'(0) = 0$ (even) and $f(0) = 0$ (odd). Referring to Eq. (13) we see that Eqs. (15) will determine the discrete set of energy eigenvalues $\{E_i\}$.

The following remarks are in order.

(i) In the case of nonrelativistic kinematics, $K = p^2$, Eq. (14) is no longer an integral equation, but gives an *analytic expression* for f. Indeed, in this case we have $R = -(p+q)/\pi$; therefore, $g(p) = \alpha p$ (even) or $g = \alpha$ (odd), where α is a normalization constant (determined as an integral over f).

In this case, we have $\tau(p) = p^3/3 - Ep$ and the quantization conditions (15) give as energy eigenvalues $E = -\epsilon_i$, where ϵ_i are the zeros of the Airy function or its derivative, as is well known.

(ii) More generally we have the following theorem.

Theorem. If the kinetic operator $K(p^2)$ is a *rational function* in p^2, then the eigenvalue problem (7) possesses *analytic solutions*.

The proof is straightforward since any rational function can be decomposed in a finite sum of simple elements. A polynomial will contribute a sum of separable terms to R (i.e., terms which factorize in p and q), as will any pole term: $1/(p^2 + a^2) \to (p+q)/[(p^2+a^2)(q^2+a^2)]$. Therefore, the function g entering Eq. (14) will be of the form

$$g(p) = \sum_{i=1}^N a_i(p)\alpha_i,$$

where the constants α_i are to be determined consistently in terms of f.

This has obvious consequences. The successive relativistic kinematic corrections can be calculated exactly to any given finite order, for example, by truncating the continued fraction expansion of $(p^2+m^2)^{1/2}$:

$$(p^2+m^2)^{1/2} = m + p^2/(2m + p^2/(2m + p^2/(2m + \cdots))),$$

which converges uniformly in the entire complex p^2 plane aside from the negative real axis $[-m^2, -\infty]$.

(iii) Equation (14) can be cast into the form of a Fredholm equation:

$$f(p) = \frac{1}{\pi}\int_{-\infty}^{+\infty} H(p,q)f(q)dq, \qquad (16)$$

with

$$H(p,q) = -\int_{-\infty}^p \cos(\tau(p) - \tau(k))\left(\frac{K(p^2) - K(q^2)}{q-k}\right)dk. \qquad (17)$$

Owing to the asymptotic behavior of $f(p)$ as $|p| \to +\infty$, a simple change of the function $f(p) = \phi(p)/(p^2+\lambda^2)^\gamma$ with a suitable choice of the exponent γ renders the kernel of the Hilbert–Schmitt type. This integral equation is very well suited for numerical calculations.

Note that the energy E appears as a parameter in the kernel H, Eq. (16). The spectrum $\{E_i\}$ corresponds to the discrete set of values of E for which the homogeneous equation (17) has nontrivial solutions ($f \neq 0$).

In the case of relativistic kinematics, one has

$$\frac{(K(p^2) - K(q^2))}{(q-p)} = \frac{(q+p)}{\sqrt{q^2+m^2} + \sqrt{p^2+m^2}}.$$

(iv) One can easily treat problems where the potential contains regular contributions besides the linear term r. In such cases, Eq. (9) takes the form

$$A(p^2)f(p) + v(p) = \frac{1}{\pi}\int_{-\infty}^{+\infty}\frac{f'(q)dq}{q-p},$$

where

$$v(p) = \int_{-\infty}^{+\infty} V(p,q)f(q)dq$$

represents these extra contributions.

If we define

$$h(p) = -\frac{1}{\pi}\int_{-\infty}^{+\infty}\frac{v(q)dq}{q-p},$$

then all the calculations proceed in the same way, with the function $g(p)$ replaced by $g(p) + h(p)$, e.g., (14) becomes

$$f(p) = \int_{-\infty}^p (g(q) + h(q))\cos(\tau(p) - \tau(q))dq$$

and a similar (but lengthier) Fredholm equation replaces Eqs. (16) and (17).

2. Three-dimensional problem

Consider now the three-dimensional problem

$$(\sqrt{-\Delta^2 + m^2} + (r - E))\Psi(\mathbf{r}) = 0 \qquad (18)$$

(any previous form of the kinetic energy operator could be written).

After partial-wave projection, the Laplacian operator is

$$\Delta^2 = \frac{1}{r}\frac{d^2}{dr^2}r - \frac{l(l+1)}{r^2}.$$

a. s wave $l=0$. In the s wave, we can introduce the operator

$$\mathcal{O} = \frac{1}{r}\left[-\frac{d^2}{dr^2} + m^2\right]^{1/2} r,$$

which satisfies

$$\mathcal{O}^2 = -\frac{1}{r}\frac{d^2}{dr^2} r + m^2.$$

If we introduce the reduced wavefunction $u(r) = r\psi(r)$, we end with a problem similar to the one-dimensional case:

$$\left[\left(-\frac{d^2}{dr^2} + m^2\right)^{1/2} + (r - E)\right]u(r) = 0,$$

except that we work on the positive real r axis:

$$u(0) = u(\infty) = 0, \quad r \in [0, +\infty], \quad u \in \mathcal{L}_0^2(R).$$

After Fourier transforming; setting $p^2 = s$; working on the reduced wavefunction $u(p)$, $u \in \mathcal{L}_0^2(R)$, $p \in [0, \infty]$; and writing $u(s)$ instead of $u(p)$, we end with

$$(\sqrt{s + m^2} - E)u(s) = \frac{2\sqrt{s}}{\pi} \fint_0^\infty \frac{u'(t)dt}{t - s},$$

which is basically the same equation as in the one-dimensional case and can be treated in the same way; therefore, we shall not repeat.

Since $u(0) = 0$, only the *odd* solutions of the one-dimensional problem are relevant, as usual.

b. Higher partial waves. Returning to Eq. (2), we now have, with the notations of Secs. II A and B,

$$(2h(p) - E)u_l(p) + \frac{1}{\pi}\int_0^\infty dp'\, u_l(p') \times \frac{1}{pp'} Q_l'\left(\frac{p^2 + p'^2}{2pp'}\right) = 0,$$

where for brevity, we have the following convention for the principal value integrals: If $f_1(p)$ has a real pole at p_0, i.e., $f_1(p) = g_1(p)/(p - p_0)$ or if $f_2(p)$ has a double pole at $p = p_0$, i.e., $f_2(p) = g_2(p)/(p - p_0)^2$, then

$$\fint_a^b f_1(p)dp = \int_a^b P\left(\frac{1}{p - p_0}\right)g_1(p)dp,$$

$$\fint_a^b f_2(p)dp = \int_a^b P\left(\frac{1}{(p - p_0)^2}\right)g_2(p)dp.$$

In our case, the function $Q_l'\{(p^2 + p'^2)/2pp'\}$ has a singularity at $p = p'$ of the form $1/(p^2 - p'^2)^2$, so that the latter is understood. However, the integrand vanishes at both endpoints ($p' = 0$ and $p' = +\infty$), so that we are within the conditions (6). Therefore, this case results in what was discussed in remark (iv) of Sec. II C 1.

However, for higher partial waves, the Vekua–Magnaradze method is more cumbersome to handle in practice than the Multhopp technique, which will be described below.

3. Analytic approximations

Explicit functions that constitute excellent approximations to the exact eigenfunctions of the relative Hamiltonian $H = (p^2 + m^2)^{1/2} + r$ (including the massless case $H = p + r$) may be obtained either in configuration or momentum space. We concentrate mainly on s waves. Once the s waves are obtained, WKB approximations can be used with the higher partial waves.[31]

a. Configuration space: Constant force. Consider the one-dimensional eigenvalue problem

$$(\sqrt{p^2 + m^2} + V(x))\psi = E\psi,$$

where

$$V(x) = x \quad \text{for} \quad x \geqslant 0, \quad V(x) = +\infty \quad \text{for} \quad x < 0.$$

The solutions are given by generalized Fresnel integrals defined as

$$G_n^m(x) = N\int_0^\infty dp\, \cos(\tilde{\tau}_m(p) + p(x - E_n)), \qquad (19)$$

where

$$\tilde{\tau}_m(p) = \int_0^p \sqrt{p^2 + m^2} = \frac{p}{2}\sqrt{p^2 + m^2} + \frac{m^2}{2}\,\text{argsh}\,\frac{p}{m}$$

for $x \geqslant 0$ and $G_n^m(x) = 0$ for $x < 0$. Here, m stands for the mass m and n stands for the radial quantum number ($n = 1,2,\ldots$). The eigenvalues E_n are determined by the condition $G_n^m(0) = 0$ and N is a normalization factor. This corresponds to a constant force for $x > 0$.

For $m = 0$ one recovers the Fresnel integral.

In the three-dimensional case, the functions

$$u_n(r) = G_n^m(r) \qquad (20)$$

constitute excellent approximations to the exact s-wave solutions of problem (18). In the limit $m \to \infty$ and under a suitable redefinition of the parameter E, one recovers the (exact) Airy functions.

Table II shows the comparison between the exact eigenvalues and those emerging from approximation (19). The mass scale is chosen in comparison with the eigenvalues of the massless case.

In Table II we notice that approximately (19) does not deviate by more than a few percent: The deviation decreases as the mass increases (in the limit $m \to \infty$ the approximation becomes exact) and as the radial quantum number increases.

TABLE II. Exact and approximate eigenvalues for the s-wave relativistic linear potential.

m (dimensionless)	n	Exact	Approximate
0	1	2.232	2.171
0	2	3.330	3.316
0	3	4.164	4.156
0	4	4.859	4.854
0	5	5.467	5.463
$\sqrt{2}$	1	2.929	2.912
$\sqrt{2}$	2	3.929	3.923
$2\sqrt{2}$	1	4.080	4.077
$4\sqrt{2}$	1	6.678	6.672
$6\sqrt{2}$	1	9.385	9.377
$10\sqrt{2}$	1	14.905	14.896

These facts can easily be understood if one calculates the expression

$$\delta(r) = (\sqrt{p^2+m^2} + r - E_n)u_n(r),$$

with $u_n(r)$ given by (20) and E_n the approximate eigenvalue. If u_n and E_n were exact, we would find $\delta(r)=0$. A straightforward calculation shows that one has

$$\delta(r) = \frac{2}{\pi}\int_0^\infty dp\cos(\tilde{\tau}(p) - pE_n) \times \int_0^\infty dk \frac{k\sin kr}{\sqrt{k^2+m^2}+\sqrt{p^2+m^2}}.$$

Obviously, the above equation goes to zero as $m\to\infty$; it also decreases with the radial quantum number owing to the oscillations of $\cos(\tilde{\tau}(p) - pE_n)$ and the fact that $\langle p\rangle$ increases with n.

b. Momentum space. An equivalent approximation can be obtained from the momentum space analysis performed in Sec. II C *1*. One can either perform a stationary phase approximation on Eqs. (16) and (17) or simply set $g = \text{const}$ in Eq. (14), which is what emerges in the nonrelativistic approximation. The reason for the latter approximation is that $f(p)$ is peaked around a value $p_0 = \langle p^2\rangle^{1/2}$; therefore, the integral defining $g(p)$ in Eq. (11) will vary slowly with p (remember we are only interested in the odd solutions of the one-dimensional problem).

One can readily verify that under this approximation,

$$f(p) \sim \lambda \int_0^p \cos(\tau(p) - \tau(q))dq,$$

with λ a normalization constant, and that the quantization condition (15b) reads as

$$\int_0^\infty \cos(\tilde{\tau}(p) - pE_n)dp = 0,$$

which is identical to the configuration space approximation.

D. The Multhopp technique for solving bound-state equations

One of the usual methods for finding eigenstates and eigenfunctions of bound-state equations is the discretization or colocation method. This means that the continuous variables are replaced by discrete ones, so that the differential or integral equations become simple matrix eigenvalue problems.

One of these very useful methods is the Multhopp technique. It has been used, for example, in the calculations of wing theory[21] and is very appropriate for singular integral equations. For a discussion of the notation we have adopted, see Ref. 22. Here we will discuss only integral equations, but the generalization to differential equations is straightforward.

1. Integral equations

Consider the following integral equation:

$$(H(p) - E)u(p) + \int_0^\infty dp'\, u(p')V(p,p') = 0$$

and make the change of variables $p = h\tan(\theta/2)$ and $p' = h\tan(\theta'/2)$, with h an arbitrary positive constant and $0\leqslant\theta\leqslant\pi$. The above equation then becomes

$$(H(\theta) - E)u(\theta) + \int_0^\pi \frac{h\,d\theta'}{2\cos^2(\theta'/2)} u(\theta')V(\theta,\theta') = 0, \tag{21}$$

where we have written, for simplicity, $u(h\tan(\theta/2))$ as $u(\theta)$, $H(h\tan(\theta/2))$ as $H(\theta)$, and $V(h\tan(\theta/2), h\tan(\theta'/2))$ as $V(\theta,\theta')$.

If we now expand $u(\theta)$ on an orthonormal basis in the interval $[0,\pi]$, namely, $\{(2/\pi)^{1/2}\sin i\theta, i\geqslant 1\}$ and truncate the series at a maximal value $i_{\max} = N$, we obtain, for the special values of θ: $\theta_k = k\pi/(N+1)$, $k = 1,\ldots,N$, the following expansion:

$$u(\theta_k) = \sum_{i=1}^N a_i \sin i\theta_k. \tag{22}$$

Replacing Eq. (22) in Eq. (21), we obtain

$$\sum_{i=1}^N a_i B(i,\theta_j) = Eu(\theta_j),$$

with

$$B(i,\theta_j) = H(\theta_j)\sin i\theta_j + \int_0^\pi \frac{h\,d\theta'}{2\cos^2(\theta'/2)} \sin i\theta' V(\theta_j,\theta').$$

Then from (22), we deduce that

$$a_j = \frac{2}{(N+1)} \sum_{i=1}^N \sin j\theta_k u(\theta_k),$$

where we have used the orthonormality condition

$$\frac{2}{(N+1)} \sum_{k=1}^N \sin i\theta_k \sin j\theta_k = \delta_{ij}.$$

Finally we end with the following matrix eigenvalue problem:

$$\sum_{k=1}^N B_{jk} u(\theta_k) = Eu(\theta_j),$$

with

$$B_{jk} = \frac{2}{(N+1)} \sum_{i=1}^N \sin i\theta_k B(i,\theta_j)$$

or, explicitly,

$$B_{jk} = H(\theta_j)\delta_{jk} + \frac{2}{(N+1)} \sum_{i=1}^N \sin i\theta_k \times \int_0^\pi \frac{h\,d\theta'}{2\cos^2(\theta'/2)} \sin i\theta' V(\theta_j,\theta').$$

Thus, the initial problem has been replaced by a simple matrix eigenvalue problem, which can be now solved numerically. The integrals in B_{jk} can be easily evaluated by standard numerical methods.

This method gives good results for the first low-lying eigenvalues and eigenfunctions. The accuracy increases with N. The method that shows that N of the order of 30 already gives very accurate results appeared in our calculations in Refs. 8 and 16. Notice, also, that the convergence of the method depends highly on the choice of the arbitrary constant h, which must be chosen such that the eigenfunctions are well centered around the points $\{p_i = h\tan(\theta_i/2), i = 1,...,N\}$. Thus h must be of the order of the mean value of p: Any variations of h around this value must have no effect on the first eigenstates.

2. Coupled integral equations

The Multhopp method can also be extended to coupled integral equations, which will appear later in the three-body problem. Thus we now consider the two coupled integral equations

$$(H_{11}(p) - E)u_1(p) + H_{12}(p)u_2(p) + \int_0^\infty dp'[V_{11}(p,p')u_1(p') + V_{12}(p,p')u_2(p')] = 0,$$

$$H_{21}(p)u_1(p) + (H_{22}(p) - E)u_2(p) + \int_0^\infty dp'[V_{21}(p,p')u_1(p') + V_{22}(p,p')u_2(p')] = 0.$$

By making the same calculations as before, but with $a_i^{(1)}$ and $a_i^{(2)}$ now defined by $u_1(\theta) = \Sigma_{i=1}^N a_i^{(1)} \sin i\theta$ and $u_2(\theta) = \Sigma_{i=1}^N a_i^{(2)} \sin i\theta$, we end with the following matrix eigenvalue problem:

$$\sum_{k=1}^N \begin{pmatrix} B_{jk}^{11} & B_{jk}^{12} \\ B_{jk}^{21} & B_{jk}^{22} \end{pmatrix} \begin{pmatrix} u_1(\theta_k) \\ u_2(\theta_k) \end{pmatrix} = E \begin{pmatrix} u_1(\theta_j) \\ u_2(\theta_j) \end{pmatrix},$$

with

$$B_{jk}^{\alpha\beta} = H_{\alpha\beta}(\theta_j)\delta_{jk} + \frac{2}{(N+1)} \sum_{i=1}^N \sin i\theta_k \times \int_0^\pi \frac{h\, d\theta'}{2\cos^2(\theta'/2)} \sin i\theta' V_{\alpha\beta}(\theta_j,\theta'), \quad \alpha,\beta = 1,2.$$

The Multhopp method can be generalized to any set of coupled equations. Compared to other methods—variational methods,[10,12,13] cutting the potential,[32] and variational Monte Carlo techniques[9]—the Multhopp technique is very efficient with confining potentials whose singularities in the kernels $V(p,p')$ can be integrated explicitly.

E. Examples

The Multhopp technique can be easily applied for the higher waves of the Hamiltonian $H = p^2 + r$. This can be done by first separating the singularity of the kernel $V_l(p,p')$, which is the same for all waves:

$$V_l(p,p') = \frac{1}{\pi pp'} Q_l'\left(\frac{p^2 + p'^2}{2pp'}\right) = -\frac{4}{(p^2 - p'^2)^2} + \text{regular part}.$$

The regular part has a divergence that is at most logarithmic and so can be easily evaluated numerically.

We give the eigenvalues for different values of the angular momentum l and the radial number n in Table III. For $l = 0$, the eigenvalues are just zeros of the Airy function $\mathrm{Ai}(-x)$, as seen in Sec. II C. We give the eigenvalues of the Hamiltonian $H = p + r$ in Table IV. The eigenvalues obtained using a WBK approximation, as given by Martin,[33] are in remarkable agreement with the exact values of Table IV. The WBK approximation gives

$$E(n,l) = [2l(l+1)]^{1/4} + 2^{1/2}(n + \tfrac{1}{2})/[l(l+1)]^{1/4}.$$

Notice, also, that when l is large, $E^2 \approx l$, which means that the Regge trajectories are linear. Thus this Hamiltonian can be applied as a toy model for bound states of light quarks (see Ref. 8).

Another interesting potential which can be easily treated by the Multhoop technique is the logarithmic one, $V(r) = \ln r$: Its kernel is given by

$$V_l(p,p') = (\ln 2 - \gamma)\delta(p - p') + \frac{1}{2}\left[\frac{1}{|p - p'|} - \frac{1}{p + p'}\right] \times \left[\frac{p^2 + p'^2 - |p^2 - p'^2|}{2pp'}\right]^l,$$

with γ Euler's constant.

The eigenvalues for the nonrelativistic and ultrarelativistic cases are given in Tables V and VI.

The Hamiltonian $H = p^2 + \ln r$ has many interesting features and can be applied for bound states made of heavy quarks (charmonium, bottomium, etc.). Notice that the splittings between the different eigenvalues remain constant if we rescale p to $p/m^{1/2}$ and r to $rm^{1/2}$, where m is a mass scale. This means that these splittings are independent of the quark masses. This is indeed what is observed approximately when the spectrum of the charmonium and bottomium are compared, i.e., when m goes from the mass of a c quark to that of a b quark.

TABLE III. Eigenvalues of $H = p^2 + r$.

n	$l=0$	$l=1$	$l=2$	$l=3$
1	2.3381	3.361	4.248	5.050
2	4.0879	4.884	5.629	6.331
3	5.5206	6.207	6.869	7.504
4	6.7867	7.405	8.009	8.596
5	7.944	8.514	9.076	9.626

TABLE IV. Eigenvalues of $H = p + r$.

n	$l=0$	$l=1$	$l=2$	$l=3$
1	2.2322	2.9872	3.5912	4.1084
2	3.3300	3.8586	4.3347	4.7676
3	4.1642	4.5835	4.9828	5.3592
4	4.8586	5.2144	5.5651	5.8956
5	5.4670	5.7803	6.091	6.392

TABLE V. Eigenvalues of $H = p^2 + \ln r$.

n	$l=0$	$l=1$	$l=2$	$l=3$
1	1.0443	1.6531	2.0132	2.286
2	1.8474	2.1504	2.3869	2.581
3	2.2897	2.4915	2.6623	2.811
4	2.5957	2.7456	2.880	2.999
5	2.8299	2.948	3.060	3.159

TABLE VI. Eigenvalues of $H = p + \ln r$.

n	$l=0$	$l=1$	$l=2$	$l=3$
1	1.0637	1.7116	2.1058	2.3886
2	1.7581	2.1262	2.4001	2.6163
3	2.1841	2.4294	2.6341	2.8064
4	2.4844	2.6655	2.8269	2.9688
5	2.7159	2.8581	2.9903	3.1100

A real phenomenological model, taking account of all the elementary particles, is in fact given by a Hamiltonian lying between $H = p + r$ and $H = p^2 + \ln r$. A phenomenological study of a Hamiltonian of the form $H = (p^2 + m^2)^{1/2} + V(r)$ (with m a mass scale) is given in Ref. 8. The potential $V(r)$ is logarithmic at short distances and linear at long distances.

III. THE THREE-BODY PROBLEM

The three-body problem can be treated in a similar manner. Consider the equation

$$\left[\sum_{i=1}^{3} h(p_i) + \sum_{i<j=1}^{3} V(|\mathbf{r}_i - \mathbf{r}_j|) - E\right]|\Psi\rangle = 0. \quad (23)$$

As Sec. I, we assume that the kinematics $\Sigma_i h(p_i)$ is a function of only the p_i's (magnitude of $\mathbf{p}_i$). Furthermore, the potential $\Sigma_{i<j=1}^{3} V(|\mathbf{r}_i - \mathbf{r}_j|)$ is supposed to be a pairwise interaction of the separations $r_{ij} = |\mathbf{r}_i - \mathbf{r}_j|$ between particles (i) and (j).

A. The three-body problem in momentum space

In the nonrelativistic case, $h(p) \sim p^2$, it is useful to introduce the Jacobi coordinates

$$\mathbf{Q} = \sum_{i=1}^{3} \frac{m_i \mathbf{r}_i}{M},$$
$$\boldsymbol{\eta} = (m_1 m_2 / m m_{12})^{1/2} (\mathbf{r}_1 - \mathbf{r}_2),$$
$$\boldsymbol{\xi} = (m_3 m_{12} / m M)^{1/2} [(m_1 \mathbf{r}_1 + m_2 \mathbf{r}_2)/m_{12} - \mathbf{r}_3]$$

and their conjugate momenta

$$\boldsymbol{\pi} = \sum_{i=1}^{3} \mathbf{p}_i,$$
$$\mathbf{q} = (m/(m_1 m_2 m_{12}))^{1/2} [m_2 \mathbf{p}_1 - m_1 \mathbf{p}_2],$$
$$\mathbf{p} = (m/(m_3 M m_{12}))^{1/2} [m_3(\mathbf{p}_1 + \mathbf{p}_2) - m_{12} \mathbf{p}_3],$$

with

$$M = \sum_{i=1}^{3} m_i, \quad m = \sum_{i<j=1}^{3} \frac{m_i m_j}{M}, \quad m_{ij} = m_i + m_j.$$

These new coordinates with their conjugate satisfy the canonical commutation relations and one can easily check that

$$\sum_{i=1}^{3} \frac{p_i^2}{m_i} = \frac{\pi^2}{M} + \frac{p^2 + q^2}{m},$$

which corresponds to the fact that in the nonrelativistic case, the center of mass motion can be eliminated since the pairwise potential $\Sigma_{i<j} V(r_{ij})$ is a function of only the two coordinates $\boldsymbol{\eta}$ and $\boldsymbol{\xi}$, as we will see below. Notice that this choice of coordinates is not unique and that for each permutation of the three particles, one can define another set of coordinates related to the others by unitary transformations.

For other kinematics, such as the semirelativistic $h(p) = (p^2 + m^2)^{1/2}$, the bound-state equation we consider is not covariant and one has to impose the center of mass condition explicitly by fixing $\mathbf{Q} = 0$ in the kinematics.

For our calculation, we suppose from now on that the three particles have equal masses ($m_1 = m_2 = m_3$). This will simplify the discussion, which, however, remains valid even when the masses are different. In this case, the Jacobi coordinates are defined simply by

$$\mathbf{Q} = \sum \frac{\mathbf{r}_i}{3}, \quad \boldsymbol{\eta} = \frac{(\mathbf{r}_1 - \mathbf{r}_2)}{\sqrt{2}},$$
$$\boldsymbol{\xi} = \sqrt{\frac{2}{3}} \left[\frac{(\mathbf{r}_1 + \mathbf{r}_2)}{2} - \mathbf{r}_3\right],$$

and

$$\boldsymbol{\pi} = \sum \mathbf{p}_i, \quad \mathbf{q} = \frac{(\mathbf{p}_1 - \mathbf{p}_2)}{\sqrt{2}},$$
$$\mathbf{p} = \sqrt{\frac{2}{3}} \left[\frac{(\mathbf{p}_1 + \mathbf{p}_2)}{2} - \mathbf{p}_3\right].$$

Other independent sets of coordinates can be constructed by cyclic permutations:

$$\begin{pmatrix} \mathbf{Q} \\ \boldsymbol{\eta}^{(k)} \\ \boldsymbol{\xi}^{(k)} \end{pmatrix} = T^{(k)} \begin{pmatrix} \mathbf{Q} \\ \boldsymbol{\eta} \\ \boldsymbol{\xi} \end{pmatrix}, \quad \begin{pmatrix} \boldsymbol{\pi} \\ \mathbf{q}^{(k)} \\ \mathbf{p}^{(k)} \end{pmatrix} = T^{(k)} \begin{pmatrix} \boldsymbol{\pi} \\ \mathbf{q} \\ \mathbf{p} \end{pmatrix},$$

where

$$T^{(1)} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad T^{(2)} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -\frac{1}{2} & \sqrt{3}/2 \\ 0 & -\sqrt{3}/2 & -\frac{1}{2} \end{pmatrix},$$
$$T^{(3)} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -\frac{1}{2} & -\sqrt{3}/2 \\ 0 & \sqrt{3}/2 & -\frac{1}{2} \end{pmatrix}$$

are the cyclic permutation matrices. We notice that $\mathbf{Q}$ and $\boldsymbol{\pi}$ are invariant, as expected.

Then, we can easily see for the potential and kinematics that

$$\sum_{i<j=1}^{3} V(r_{ij}) = \sum_{i<j=1}^{3} V(\sqrt{2}\,|T^{(i)}\boldsymbol{\eta}|)$$

and in the center of mass system

$$\sum_{i=1}^{3} h(p_i) = \sum_{i=1}^{3} h\left(\sqrt{\frac{2}{3}}\,|T^{(i)}\mathbf{p}|\right).$$

Concerning the potential one has

$$|T^{(k)}\boldsymbol{\eta}| = |\alpha\boldsymbol{\eta} + \beta\,\boldsymbol{\xi}|$$
$$= (\alpha^2\eta^2 + \beta^2\xi^2 + 2\alpha\beta\xi\eta\cos\gamma_R)^{1/2},$$

where α and β are numbers such that $\alpha^2 + \beta^2 = 1$ and γ_R is the angle between ξ and η. Thus $|T_{(k)}\boldsymbol{\eta}|$ and, consequently, the potential are functions of only ξ, η and the angle between $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$. These three coordinates will be denoted commonly by $\mathbf{R}$ and will be called internal coordinates since they specify the shape of the triangle formed by the three particles. The other three coordinates will be called external and indicate the orientation of this triangle: They will be chosen as the three Euler angles needed for going from a reference system Σ to a body-fixed system S.

More precisely, the body-fixed system S is defined by the conditions that $\boldsymbol{\eta}$ is along the z direction and the $\mathbf{x}$ and $\mathbf{y}$ axes are defined by imposing that $\boldsymbol{\xi}$ belongs to the (xz) plane. This prescription defines completely the three Euler angles $\mathscr{R}_R = (\phi_R,\theta_R,\beta_R)$. These coordinates will be used to separate the angular momentum.

In a similar manner, one can introduce internal and external coordinates for the conjugate coordinates $\mathbf{p}$ and $\mathbf{q}$. The kinematics will then be a function of only the internal conjugate coordinates denoted by $\mathbf{P}$.

This separation of coordinates was first introduced by Omnès in the nonrelativistic case for the same reasons[23] (see, also, Zickendraht[24]). Equation (23) can then be written in the momentum space as the following integral equation:

$$(h(\mathbf{P}) - E)\Psi(\mathbf{p},\mathbf{q}) + \int d\mathbf{p}'\, d\mathbf{q}'\, \Psi(\mathbf{p}',\mathbf{q}')\, U(\mathbf{p}-\mathbf{p}',\mathbf{q}-\mathbf{q}') = 0, \tag{24}$$

with

$$U(\mathbf{p},\mathbf{q}) = \frac{1}{(2\pi)^6}\int d\boldsymbol{\xi}\, d\boldsymbol{\eta}\, V(\mathbf{R}) e^{-i\mathbf{p}\boldsymbol{\xi} - i\mathbf{q}\boldsymbol{\eta}}.$$

As in the usual two-body problem, where the potential and kinematics were radial, here they appear also to be "radial" in the sense that they have no dependence on the external coordinates (i.e., the three Euler angles).

To separate the total angular momentum, one has to expand the wavefunction $\Psi(\mathbf{p},\mathbf{q})$ on the spherical functions $\mathscr{D}^J_{MM'}(\mathscr{R}_P)$ in the form

$$\Psi^{(JM)}(\mathbf{p},\mathbf{q}) = \sum_{M'=-J}^{+J} \Psi^{(JM)M'}(\mathbf{P})\mathscr{D}^J_{MM'}(\mathscr{R}_P) \tag{25}$$

and also expand the exponential $\exp(-i\mathbf{p}\boldsymbol{\xi} - i\mathbf{q}\boldsymbol{\eta})$ on the spherical functions $\mathscr{D}^J_{MM'}(\mathscr{R}_P)$ and $\mathscr{D}^J_{NN'}(\mathscr{R}_R)$. To do this, it is convenient to introduce the hyperspherical harmonics.

B. Hyperspherical coordinates; hyperspherical harmonics

Let us now describe briefly how to introduce the hyperspherical coordinates. The coordinate space generated by $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$ is a six-dimensional space in the three-body problem case. We then consider six equivalent radial coordinates replacing $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$. One of the new coordinates is the radius $R = (\xi^2 + \eta^2)^{1/2}$ (also called the hyper-radius) and the five others are angular coordinates. The different ways for defining the hyperspherical coordinates result from the different possible choices of these five angles.

1. Fabre de la Rippelle functions

A usual choice is the following. From $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$, we define the five angles ϕ, $\hat{\boldsymbol{\xi}} = (\theta_\xi,\phi_\xi)$ and $\hat{\boldsymbol{\eta}} = (\theta_\eta,\phi_\eta)$ by writing $\tan\phi = \xi/\eta$, which means that $\xi = R\cos\phi$ and $\eta = R\sin\phi$ and $\hat{\boldsymbol{\xi}}$ and $\hat{\boldsymbol{\eta}}$ define the two directions of $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$ in the usual three-dimensional space. Thus we have $0\leqslant\phi\leqslant\pi/2$, $0\leqslant\theta\leqslant\pi$, and $0\leqslant\phi\leqslant 2\pi$.

The volume element is then given by

$$d\boldsymbol{\xi}\, d\boldsymbol{\eta} = R^5\, dR\, d\Omega,$$

with

$$d\Omega = \cos^2\phi\,\sin^2\phi\, d\phi\,\sin\theta_\xi\, d\theta_\xi\, d\phi_\xi\,\sin\theta_\eta\, d\theta_\eta\, d\phi_\eta$$

and the Laplacian

$$\Delta = \Delta_R + \Delta(\Omega)/R^2,$$

where

$$\Delta_R = \frac{1}{R^5}\frac{\partial}{\partial R}R^5\frac{\partial}{\partial R}$$

and

$$\Delta(\Omega) = \frac{\partial^2}{\partial\phi^2} + 4\cot 2\phi\frac{\partial}{\partial\phi} + \frac{1}{\cos^2\phi}\Lambda_\xi + \frac{1}{\sin^2\phi}\Lambda_\eta .$$

Here Λ is the usual angular Laplacian in the three-dimensional space.

The eigenfunctions of the angular Laplacian $\Delta(\Omega)$ are called the hyperspherical harmonics (HH): They are a generalization of the usual spherical harmonics of the three-dimensional space, i.e., they are representations of the group SO(6). The HH with this choice of coordinates have been constructed explicitly by Fabre de la Rippelle.[24]

The HH denoted by $\boldsymbol{\phi}_K^{l_\xi m_\xi l_\eta m_\eta}(\Omega)$ obey the relation

$$\Delta(\Omega)\boldsymbol{\phi}_K^{l_\xi m_\xi l_\eta m_\eta}(\Omega) = -K(K+4)\boldsymbol{\phi}_K^{l_\xi m_\xi l_\eta m_\eta}(\Omega)$$

and are given by

$$\boldsymbol{\phi}_K^{l_\xi m_\xi l_\eta m_\eta}(\Omega) = N_K^{l_\xi l_\eta}(\cos\phi)^{l_\xi}(\sin\phi)^{l_\eta}P^{l_\eta+\frac12, l_\xi+\frac12}_{(K-l_\xi-l_\eta)/2} \times(\cos 2\phi)\, Y^{m_\xi}_{l_\xi}(\hat{\boldsymbol{\xi}})Y^{m_\eta}_{l_\eta}(\hat{\boldsymbol{\eta}}), \tag{26}$$

where K is the quantum number associated with the angle ϕ. Here (l_ξ,m_ξ) [resp. (l_η,m_η)] are associated with the direction of $\boldsymbol{\xi}$ (resp. $\boldsymbol{\eta}$). P is a Jacobi polynomial.

The HH are also eigenfunctions of the operators Λ_ξ, Λ_η, and the third projections of these two operators $\Lambda_\xi^{(3)}$ and $\Lambda_\eta^{(3)}$ with, respectively, the quantum numbers $l_\xi(l_\xi+1)$, $l_\eta(l_\eta+1)$, and m_ξ, m_η. Now $N_K^{l_\xi l_\eta}$ is a normalizing factor given by

$$|N_K^{l_\xi l_\eta}|^2 = \frac{2n!(n+l_\xi+l_\eta+1)!(2n+l_\xi+l_\eta+2)}{(n+l_\xi+\frac12)!(n+l_\eta+\frac12)!},$$

with

$$n = (K - l_\xi - l_\eta)/2,$$

where K is a positive integer and l_ξ and l_η are integers fixed by the condition $K - l_\xi - l_\eta \geqslant 0$. Here m_ξ and m_η are also integers such that $(-l_\xi\leqslant m_\xi\leqslant +l_\xi)$ and $(-l_\eta\leqslant m_\eta\leqslant +l_\eta)$.

The normalization condition for the HH is given by

$$\int \boldsymbol{\phi}_{K'}^{l'_\xi m'_\xi l'_\eta m'_\eta}(\Omega)\boldsymbol{\phi}_K^{l_\xi m_\xi l_\eta m_\eta}(\Omega)\, d\Omega = \delta^{KK'}\delta^{l_\xi l'_\xi}\delta^{l_\eta l'_\eta}\delta^{m_\xi m'_\xi}\delta^{m_\eta m'_\eta}.$$

The system of three particles can be coupled as usual to have a fixed total angular momentum J such that $\mathbf{J} = \mathbf{l}_1 + \mathbf{l}_2 + \mathbf{l}_3 = \mathbf{l}_\xi + \mathbf{l}_\eta$. A set of eigenfunctions having a conserved total angular momentum can be written after separating explicitly the dependence on the three Euler angles (external coordinates of Sec. III A).

2. Simonov functions

Another set of angles has been used by Simonov, Badalyan, and Simonov[24] for the case $J = 0$: It is given by the two coordinates A and λ related to $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$ by

$$A \cos \lambda = (\xi^2 - \eta^2)/(\xi^2 + \eta^2),$$
$$A \sin \lambda = 2\boldsymbol{\xi}\cdot\boldsymbol{\eta}/(\xi^2 + \eta^2), \tag{27}$$

with $0 \leqslant A \leqslant 1$ and $0 \leqslant \lambda \leqslant 2\pi$. The three other angles can be the Euler angles $\mathscr{R}$ defined previously. However, in the case $J = 0$, these angles do not appear in the HH since $\mathscr{D}^0_{00}(\mathscr{R}) = 1$.

The HH in this case are written as follows:

$$u_K^\nu(A,\lambda) = [(K+2)/2\pi^3]^{1/2} e^{-i\lambda\nu} A^{|\nu|} \times P^{|\nu|,0}_{\frac{1}{2}[K/2 - |\nu|]}(1 - 2A^2),$$

where K is an even integer and ν takes the values $-K/2$, $-K/2 + 2,\dots, +K/2$. P is a Jacobi polynomial.

The HH satisfy as before,

$$\Delta(\Omega) u_K^\nu(A,\lambda) = -K(K+4) u_K^\nu(A,\lambda)$$

and obey the following normalization:

$$\int u_{K'}^{\nu'}(A,\lambda) u_K^\nu(A,\lambda) \pi^2 A \, dA \, d\lambda = \delta^{KK'} \delta^{\nu\nu'}.$$

This particular set of Simonov coordinates has interesting properties since it exhibits naturally the symmetries under permutations, for example, a cyclic permutation of the three particles is equivalent to λ going to $\lambda + 2k\pi/3$, with $k = 1,2,3$. We will use this set of coordinates in Sec. III C 5 to reduce the three-body bound-state equations to a set of coupled equations in one variable: the hyper-radius R (or P, the corresponding hypermomentum) in the s-wave case ($J = 0$).

3. Example

A simple example of the application of these new coordinates is the expansion of the pairwise potential $\Sigma_{i<j} V(r_{ij})$ in terms of the HH.[24] One can easily show for three identical particles that

$$\sum_{i<j} V(r_{ij}) = \sum_{K\nu} u_K^\nu(A,\lambda) v_K^\nu(R),$$

with

$$v_K^\nu(R) = \frac{2}{\pi}\left(\frac{2\pi^3}{K+2}\right)^{1/2} (-)^{\frac{1}{2}[K/2 - |\nu|]} \left[1 + 2\cos\frac{2\pi\nu}{3}\right] \times \int_0^\pi d\phi \sin\phi \sin\left[\left(\frac{K}{2}+1\right)\phi\right] V\left(\sqrt{2}\, R \cos\frac{\phi}{2}\right). \tag{28}$$

The term $[1 + 2\cos(2\pi\nu/3)]$ means automatically that for identical particles, the pairwise potential can be expanded on only the HH $u_K^\nu(A,\lambda)$, which has ν as a multiple of 3, which means also that $\Sigma_{i<j} V(r_{ij})$ is invariant under permutation of the three particles. Notice, also, that for the same reason there is no term for $K = 2$ in this expansion since this would correspond to $\nu = 1$, which is not a multiple of 3. For example, the harmonic potential $V(r) = r^2$ gives only the term $v_0^0(R)$:

$$\sum_{i<j} r_{ij}^2 = R^2.$$

Now, if we consider the linear potential $V(r) = r$, we can show from (28) that

$$v_{2K}^\nu(R) = \left\{3\sqrt{2}(-)^{(K-\nu)/2} \frac{\pi^2}{K+1}\left[\frac{1}{2K+1} + \frac{1}{2K+3} - \frac{1}{2K-1} - \frac{1}{2K-5}\right]\right\} R,$$

with $K = 0,1,2,\dots$, $\nu = -K, -K+2,\dots, +K$, and where ν is a multiple of 3. This gives, for example, the following ratios between the different terms $v_{2K}^\nu(R)$:$|v_4^0/v_0^0| \sim 0.027$, $|v_6^3/v_0^0| \sim 0.009$, $|v_8^0/v_0^0| \sim 0.004$, etc., which means that for the linear potential, the dominant part of the expansion is given by the term $v_0^0(R)$. The linear potential is approximately hyper-radial.

Another interesting expansion is the one that can be done in momentum space when calculating the general kinematics $\Sigma_{i=1}^3 h(p_i)$. The expansion is now on the HH $u_K^\nu(A_P,\lambda_P)$, with the coordinates A_P and λ_P defined in momentum space. Thus we have

$$\sum_{i=1}^3 h(p_i) = \sum_{K\nu} h_K^\nu(P) u_K^\nu(A_P,\lambda_P).$$

The nonrelativistic kinematics is a particular case (as the harmonic potential above) where only the term $h_0^0(P)$ is different from zero: $\Sigma_{i=1}^3 p_i^2 = P^2$. This expansion is useful for kinematics that are not very different from the nonrelativistic one, as is the case for the relativistic kinematics $h(p) = (p^2 + m^2)^{1/2}$.

The linear potential and the semi-relativistic kinematics are both used for calculating the baryon masses in the potential models approach (see Ref. 16 and the references therein). Nonrelativistic calculations are given in Ref. 35.

C. Separation of angular momentum in momentum space

As stated previously, in order to calculate the spectrum and the eigenfunctions for a three-body problem, one has to first separate the internal and external coordinates and then integrate out the external coordinates. This is made possible by the fact that the potential and kinematics are functions of only the internal coordinates.

1. The HH

First, it is interesting to have the form of the HH in terms of internal and external coordinates. For this, one has to first couple the HH defined by Eq. (26) with the Clebsch–Gordan coefficients to have functions with a definite total angular momentum J and a fixed third component of the angular momentum M in the reference system:

$$\phi_K^{JMl_\xi l_\eta}(\Omega) = \sum_{m_\xi m_\eta} \langle l_\xi m_\xi l_\eta m_\eta | JM \rangle \phi_K^{l_\xi m_\xi l_\eta m_\eta}(\Omega).$$

Our conventions are those of Edmonds.[34] Using the explicit form of $\phi_K^{l_\xi m_\xi l_\eta m_\eta}(\Omega)$ and after applying the necessary rotations defined in Sec. III A, we can separate explicitly the external coordinates in the spherical functions $\mathscr{D}^J_{Mn}(\mathscr{R})$ as follows:

$$\phi_K^{JMl_\xi l_\eta}(\Omega) = \sum_{n=-J}^{+J} F_K^{n(l_\xi l_\eta J)}(\phi,\gamma)\mathscr{D}^J_{Mn}(\mathscr{R}).$$

The functions $F_K^{n(l_\xi l_\eta J)}(\phi,\gamma)$ define the HH for a fixed angular momentum and are related to the functions given by Zickendraht[24]: They are given explicitly by

$$F_K^{n(l_\xi l_\eta J)}(\phi,\gamma) = N_K^{l_\xi l_\eta} N^J_{l_\xi l_\eta n}(\cos\phi)^{l_\xi}(\sin\phi)^{l_\eta} P^{\,l_\eta+\frac12,\,l_\xi+\frac12}_{(K-l_\xi-l_\eta)/2}$$
$$\times(\cos 2\phi)P^n_{l_\xi}(\cos\gamma),$$

with $N_K^{l_\xi l_\eta}$ defined as before and

$$N^J_{l_\xi l_\eta n} = [(-)^n/4\pi]((l_\xi - n)!/(l_\xi + n)!)^{1/2}$$
$$\times(2l_\xi+1)(2l_\eta+1)^{1/2}\langle l_\xi n l_\eta 0|Jn\rangle.$$

As a special case, let us consider $J=0$. Since $\mathscr{D}^0_{00}(\mathscr{R})=1$, one has

$$\phi_K^{00ll}(\Omega) = F_K^{0(ll0)}(\phi,\gamma) = N_K^{ll}[(-)^l(2l+1)^{1/2}/4\pi 2^l]$$
$$\times\sin^l 2\phi P^{\,l+\frac12,\,l+\frac12}_{K/2-l}(\cos 2\phi)P_l(\cos\gamma),$$

with $K=0,2,4,\ldots$ and $l=0,1,\ldots,K/2$.

By simple unitary transformations, these functions can be related to the HH of Simonov,[24] $u_K^\nu(A,\lambda)$. However, notice that the symmetries of $F_K^{0(ll0)}(\phi,\gamma)$ under the permutation of the particles are not obvious, as opposed to the HH of Simonov.[24]

2. The exponential

As in the usual three-dimensional case, the exponential $\exp(-i\mathbf{p}\boldsymbol{\xi} - i\mathbf{q}\boldsymbol{\eta})$ can be expressed in terms of the HH as follows:

$$e^{-i\mathbf{p}\boldsymbol{\xi}-i\mathbf{q}\boldsymbol{\eta}} = (2\pi)^3\sum_{[L]} i^K \frac{J_{K+2}(PR)}{(PR)^2}\phi_{[L]}(\Omega_R)\phi^*_{[L]}(\Omega_P),$$

where $[L]$ stands for all the quantum numbers $(K,l_\xi,m_\xi,l_\eta,m_\eta)$ or (K,J,M,l_ξ,l_η) and (R,Ω_R) and (P,Ω_P) are the hyperspherical coordinates. For a discussion of this, see Ref. 24.

In terms of external and internal coordinates, we can write

$$e^{-i\mathbf{p}\boldsymbol{\xi}-i\mathbf{q}\boldsymbol{\eta}} = (2\pi)^3\sum_{Jnn'}\mathscr{F}^{Jnn'}(\mathbf{P},\mathbf{R})$$
$$\times\left\{\sum_M \mathscr{D}^J_{nM}(\mathscr{R}_R)\mathscr{D}^{*J}_{n'M}(\mathscr{R}_P)\right\}, \qquad (29)$$

where

$$\mathscr{F}^{Jnn'}(\mathbf{P},\mathbf{R})$$
$$= \sum_{Kl_\xi l_\eta} i^K \frac{J_{K+2}(PR)}{(PR)^2}\frac{2l_\eta+1}{4\pi}$$
$$\times\langle l_\xi n l_\eta 0|Jn\rangle\langle l_\xi n' l_\eta 0|Jn'\rangle$$
$$\times F_K^{n(l_\xi l_\eta J)}(\phi_R,\gamma_R)F_K^{n'(l_\xi l_\eta J)}(\phi_P,\gamma_P),$$

where $\mathscr{F}^{Jnn'}(\mathbf{P},\mathbf{R})$ is expressed only in terms of the internal coordinates $(P,\phi_R,\gamma_P)=\{\mathbf{P}\}$ and $(R,\phi_R,\gamma_R)=\{\mathbf{R}\}$ and generalizes in some sense the Bessel functions to the three-body problem.

3. The bound-state equations

The expansion of the exponential (29) can be used to integrate out all the external coordinates in the bound-state equation (24). Thus for a fixed angular momentum J and a space-fixed projection M_0, after expanding the wavefunction on the spherical functions as in Eq. (25), we end with the following coupled equations:

$$(H(\mathbf{P})-E)\Psi^{(JM_0)M}(\mathbf{P})$$
$$+\sum_{M'=-J}^{+J}\int d\mathbf{P}'\Psi^{(JM_0)M'}(\mathbf{P}')\mathscr{G}^{JMM'}(\mathbf{P},\mathbf{P}')=0, \qquad (30)$$

where the kernels $\mathscr{G}^{JMM'}(\mathbf{P},\mathbf{P}')$ can be calculated from the potential $V(\mathbf{R})$ and from the functions $\mathscr{F}^{JnM}(\mathbf{P},\mathbf{R})$ appearing in the expansion of the exponential in Eq. (29),

$$\mathscr{G}^{JMM'}(\mathbf{P},\mathbf{P}') = \left(\frac{8\pi^2}{2J+1}\right)^2\sum_{n=-J}^{+J}\int d\mathbf{R}\,V(\mathbf{R})$$
$$\times\mathscr{F}^{*JnM}(\mathbf{P},\mathbf{R})\mathscr{F}^{JnM'}(\mathbf{P}',\mathbf{R}).$$

With the coordinates $(R,\phi_R,\gamma_R)=\{\mathbf{R}\}$, the element of integration can be written as follows: $d\mathbf{R} = R^5\,dR\,d\cos\gamma_R\cos^2\phi_R\sin^2\phi_R d\phi_R$.

From Eq. (30), we see that the wavefunctions $\Psi^{(JM_0)M}(\mathbf{P})$ are coupled by the kernels $\mathscr{G}^{JMM'}(\mathbf{P},\mathbf{P}')$. By imposing some further symmetries between the component $\Psi^{(JM_0)M}(\mathbf{P})$ as parity, we can reduce the $(2J+1)$ coupled equations to only (J) or $(J+1)$ coupled equations.

4. The case J=0

The simplest example is $J=0$. In this case, the last equations reduce to only one equation since $J=M=0$; the kernel is given by

$$\mathscr{G}^{000}(\mathbf{P},\mathbf{P}')$$
$$=(8\pi^2)^2\int d\mathbf{R}\,V(\mathbf{R})\mathscr{F}^{*000}(\mathbf{P},\mathbf{R})\mathscr{F}^{000}(\mathbf{P}',\mathbf{R}),$$

where $\mathscr{F}^{000}(\mathbf{P},\mathbf{P}')$, the projection of the exponential on the s wave, can be expressed in terms of Simonov functions as

$$\mathscr{F}^{000}(\mathbf{P},\mathbf{R})$$
$$=\sum_K i^K\frac{J_{K+2}(PR)}{(PR)^2}\left\{\sum_\nu u_K^\nu(A_R,\lambda_R)u_K^{*\nu}(A_P,\lambda_P)\right\}.$$

This will allow us to obtain a set of coupled equations in one variable, the hyper-radius R (or P), very suitable for numerical calculations.

For this, consider the internal coordinates $\{\mathbf{R}\}=(R,A,\lambda)$ (and similarly for $\mathbf{P}$) and define the coefficients $v_K^\nu(R)$, $h_K^\nu(P)$, and $f_K^\nu(P)$ from the expansion of the potential $V(\mathbf{R})$, the kinematics $h(\mathbf{P})$, and the wavefunction for $J=0$, $\Psi(\mathbf{P})$ on the HH:

$$h(\mathbf{P}) = \sum_{K\nu} h_K^\nu(P)u_K^\nu(A_P,\lambda_P),$$

$$V(\mathbf{R}) = \sum_{K\nu} v_K^\nu(P)u_K^\nu(A_R,\lambda_R),$$

$$\Psi^{J=0}(\mathbf{P}) = \sum_{K\nu} \frac{f_K^\nu(P)}{P^5} u_K^\nu(A_P,\lambda_P).$$

One can then easily deduce from Eq. (30) the following set of coupled equations for fixed values of K_0 and ν_0:

$$\sum_{(K\nu)(K'\nu')} [K\nu K'\nu'|K_0\nu_0]\Big[h_K^\nu(P) f_{K'}^{\nu'}(P) + \int_0^\infty dP' f_{K'}^{\nu'}(P') V_{K_0K'}^{K\nu}(P,P')\Big] = Ef_{K_0}^{\nu_0}(P).$$

The kernels in the integral equations $V_{K_0K'}^{K\nu}(P,P')$ are now given from the coefficients of the potential $v_K^\nu(R)$ by

$$V_{K_0K'}^{K\nu}(P,P') = (-)^{(K_0+K')/2}(PP')^{1/2} \times \int_0^\infty R\,dR\, v_K^\nu(R) J_{K_0+2}(PR) J_{K'+2}(P'R).$$

These kernels are similar to those found in the integral equations of the two-body problem, with $K_0 = K'$ a half-integer related to the angular momentum l by $K = l + \frac{1}{2}$. The coefficients $[K\nu K'\nu'|K_0\nu_0]$ are the integrals of three HH's, similar to the Clebsch–Gordan coefficients for spherical harmonics:

$$[K\nu K'\nu'|K_0\nu_0] = (8\pi^2)\int d\Omega\, u_K^\nu(A,\lambda) u_{K'}^{\nu'}(A,\lambda) u_{K_0}^{*\nu_0}(A,\lambda).$$

We remark that the last coupled equations are valid for arbitrary masses. In the equal mass case, some reductions of the couplings appear as a result of the symmetries of the wavefunctions under permutations. For example, the $K=0$ equation is not coupled with the $K=2$ equation. This results in a fast convergence of the numerical methods used to solve these equations.

For the nonrelativistic case when $h_K^\nu(P) = (\pi)^{3/2} P^2 \delta_K^0\, \delta_\nu^0$, the coupled equations can be written after a Bessel transformation in the coordinate space as follows:

$$\left[-\frac{d^2}{dR^2} + \frac{(K_0+\frac{3}{2})(K_0+\frac{5}{2})}{R^2} - E\right] f_{K_0}^{\nu_0}(R) + \sum_{(K\nu)(K'\nu')} [K\nu K'\nu'|K_0\nu_0] V_K^\nu(R) f_{K'}^{\nu'}(R) = 0.$$

This set of coupled differential equations is given, for example, in Ref. 24 for the nonrelativistic problem.

Notice that if now $J \neq 0$, similar coupled equations can be written using the same methods, but their number increases rapidly.

Notice, also, that we can consider, instead of the set of internal coordinates $(R,A,\lambda) = \{\mathbf{R}\}$, another set $(\xi,\eta,\gamma) = \{\mathbf{R}\}$ given by Eqs. (27) and $\cos\gamma = \boldsymbol{\xi}\cdot\boldsymbol{\eta}/\xi\eta$. Then, a Legendre expansion in $\cos\gamma$ leads to other integral equations in the two variables ξ and η.

5. Kernels of the three-body problem

The three-body problem reduces to a set of infinite coupled integral equations which can be solved only approximately by truncating the set at a fixed number of equations. Since the quantum number K is not conserved, there appear diagonal kernels as $V_{K+3/2,K+3/2}^{(K\nu)}(P,P')$ and nondiagonal ones as $V_{K+3/2,K'+3/2}^{(K\nu)}(P,P')$, $K \neq K'$; all have to be known. For example, the linear potential $v_K^\nu(R) = R$ gives results similar to those of Sec. II B *1*. We have, for the diagonal kernels,

$$V_{K+3/2,K+3/2}(P,P') = \frac{1}{\pi PP'} Q'_{K+3/2}\left(\frac{P^2+P'^2}{2PP'}\right)$$

and for some nondiagonal kernels of particular interest,

$$V_{K+3/2,3/2}(P,P') = \frac{1}{\pi PP'} Q'_{3/2}\left(\frac{P^2+P'^2}{2PP'}\right) + \frac{I_K(P,P')}{\pi PP'},$$

with, for example,

$$I_0(P,P') = 0,$$

$$I_2(P,P') = \frac{6}{t(t^2-1)}\left[3tQ_{1/2}\left(\frac{P^2+P'^2}{2PP'}\right) + (t^2-4)Q_{3/2}\left(\frac{P^2+P'^2}{2PP'}\right)\right],$$

and

$$I_4(P,P') = \frac{32t^2}{(t^2-1)}\left[9\left(t^2-\frac{5}{6}\right)Q_{1/2}\left(\frac{P^2+P'^2}{2PP'}\right) - 12t\left(t^2-\frac{7}{8}\right)Q_{3/2}\left(\frac{P^2+P'^2}{2PP'}\right)\right],$$

where we have set $t = P/P'$.

These kernels possess extra singularities compared to the two-body problem and must also be taken into account in the Multhopp technique.

D. Special case: Hyper-radial potentials

We call the hyper-radial potential a potential depending only on R (the hyper-radius). The only hyper-radial potential evolving from a three-body pairwise potential is the harmonic one. In a similar manner, only the nonrelativistic kinematics gives a nonrelativistic hyper-radial kinematics in momentum space.

Let us now consider the spectrum of hyper-radial potential as an approximation for some kind of pairwise potential. Thus we impose that $V(\mathbf{R}) = V(R)$ and $h(\mathbf{P}) = P^2$.

This problem can be easily solved by using the HH of Fabre de la Rippelle.[24] We obtain the following decoupled equation with a fixed K, which is now a conserved quantum number:

$$-\frac{d^2}{dR^2} u_K(R) + [(K_0+\tfrac{3}{2})(K_0+\tfrac{5}{2})/R^2] u_K(R) + V(R) u_K(R) = E u_K(R).$$

The spectrum of the above equation is given in Fig. 1, together with the corresponding quantum numbers. In Fig. 2 we show some degeneracies that occur for the harmonic potential.

IV. CONCLUSION

We have shown that the inclusion of a semirelativistic kinematics in a Schrödinger equation can be easily done in a

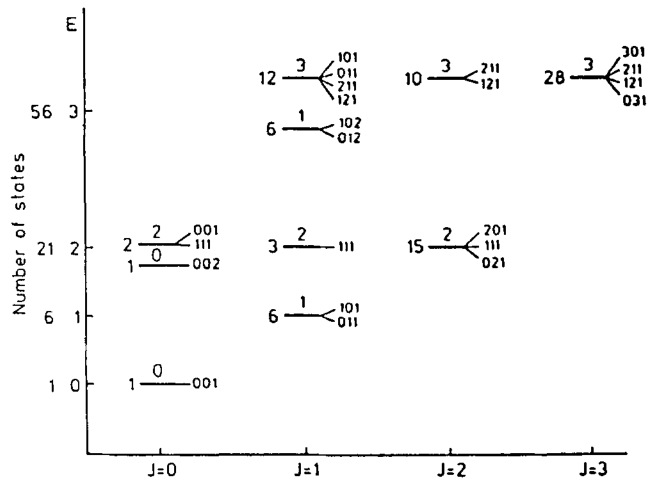

FIG. 1. Spectrum of a hyper-radial potential. The degeneracy is shown on the left of the line. The other quantum numbers are K and $l_\xi l_\eta n$ (n being the radial quantum number: 1,2,...).

mathematically well-defined manner: However, its resolution needs the introduction of new numerical methods.

We have described some of the technical methods we have used which seem to be the most appropriate for this kind of problem.

In the two-body problem case, relativistic Schrödinger equations with confining potentials are well defined as singular integral equations. We have given the kernels for the power-law potentials and described efficient numerical methods (the Vekua–Magnaradze and Multhopp techniques) for solving these equations.

Similarly, the three-body problem needs a special treatment. We have found that the hyperspherical formalism is very adequate. However, for a general pairwise potential, we end with a large number of coupled singular integral equations. The resolution can be done only approximately by truncating at a certain order; this gives accurate results with a few number of equations for some "almost hyper-radial" potentials such as the linear one.

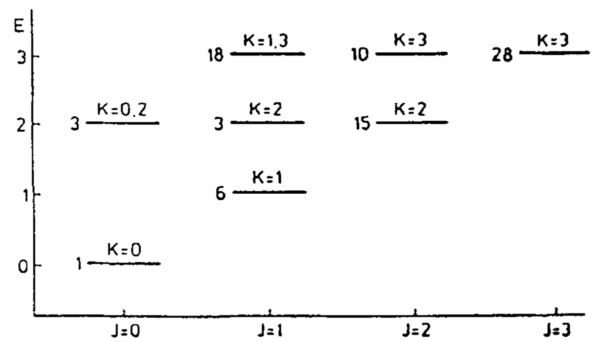

FIG. 2. Spectrum of a harmonic potential. The degeneracy is shown on the left of the solid line.

All these techniques have been applied in the potential models' approach to elementary particles.[8,16]

ACKNOWLEDGMENTS

We thank A. O. Barut, A. Martin, and J. M. Richard for their interest and enlightening discussions.

[1]A. de Rujula, H. Georgi, and S. Glashow, Phys. Rev. D **12**, 147 (1975).
[2]J. L. Richardson, Phys. Lett. B **82**, 272 (1979).
[3]W. Buchmüller and S.-H. H. Tye, Phys. Rev. D **24**, 132 (1981).
[4]E. Eichten, K. Gottfried, T. Kinoshita, J. Kogut, K. D. Lane, and T. M. Yan, Phys. Rev. Lett. **34**, 369 (1975); E. Eichten, K. Gottfried, T. Kinoshita, K. D. Lane, and T. M. Yan, Phys. Rev. D **17**, 3090 (1978); **21**, 203 (1980).
[5]A. Martin, Phys. Lett. B **100**, 511 (1981).
[6]H. Grosse and A. Martin, Phys. Rep. **60**, 341 (1980).
[7]C. Quigg and J. L. Rosner, Phys. Rep. **56**, 167 (1979).
[8]J. L. Basdevant and S. Boukraa, Z. Phys. C **28**, 413 (1985); Ann. Phys. **10**, 475 (1975).
[9]J. Carlson, J. Kogut, and V. R. Pandharipande, Phys. Rev. D **27**, 233 (1983); **28**, 2807 (1983).
[10]D. P. Stanley and D. Robson, Phys. Rev. D **21**, 3180 (1980).
[11]M. Bander, B. Klima, U. Maor, and M. Silvermann, Phys. Lett. B **134**, 258 (1984).
[12]S. Godfrey and N. Isgur, Phys. Rev. D **32**, 189 (1985).
[13]Y. R. Kwon and F. Tabakin, Phys. Rev. C **18**, 932 (1978); R. H. Landau, *ibid.* **27**, 2191 (1987).
[14]D. Eyre and J. P. Vary, Phys. Rev. D **34**, 3467 (1986); J. R. Spence and J. P. Vary, *ibid.* **35**, 2191 (1987).
[15]B. Durand, and L. Durand, Phys. Rev. D **28**, 396(1983); L. Durand, *ibid.* **32**, 1257 (1985); L. J. Nickisch, L. Durand, and B. Durand, *ibid.* **30**, 660 (1984).
[16]J. L. Basdevant and S. Boukraa, Z. Phys. C **30**, 103 (1986).
[17]J. M. Richard and P. Taxil, Phys. Lett. B **128**, 453 (1983).
[18]J. M. Richard, Phys. Lett. B **100**, 515 (1982).
[19]J. M. Richard and P. Taxil, Ann. Phys. **150**, 267 (1983).
[20]N. I. Muskhelishvili, *Singular Integral Equations* (Noordhoff, Groningen, Holland, 1953).
[21]A. Robinson and J. A. Laurmann, *Wing Theory* (Cambridge U.P., Cambridge, 1958); K. Karamchetti, *Principles of Ideal Fluid Dynamics* (New York, 1966).
[22]A. J. Hanson, R. D. Peccei, and M. K. Prasad, Nucl. Phys. B **121**, 477 (1977).
[23]R. L. Omnès, Phys. Rev. **134**, B 1358 (1964).
[24]L. M. Delves, Nucl. Phys. **9**, 391 (1959); Nucl. Phys. **20**, 275 (1960); W. Zickendraht, Ann. Phys. **35**, 18 (1965); Yu. A. Simonov. Sov. J. Nucl. Phys. **3**, 461 (1966); A. M. Badalyan and Yu. A. Simonov, *ibid.* **3**, 755 (1966); J. L. Ballot and M. Fabre de la Rippelle, Ann. Phys. **127**, 62 (1980); M. Fabre de la Rippelle, preprint WIS 80/11; R. I. Dzhibuti, N. B. Drupennikova, and V. Yu. Tomchinsky, Sov. J. Nucl. Phys. **23**, 285 (1976).
[25]M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1972).
[26]I. Herbst, Commun. Math. Phys. **53**, 285 (1977).
[27]A. O. Barut, Lett. Math. Phys. **10**, 195 (1985).
[28]W. M. Frank, D. J. Land, and R. M. Spector, Rev. Mod. Phys. **43**, 36 (1971) and references therein.
[29]A. O. Barut and S. Boukraa, Hadron. J. (to be published); A. O. Barut and A. Hacinliyan, Lett. Nuovo Cimento **123**, 1053 (1980).
[30]W. D. Heiss and G. M. Welke, J. Math. Phys. **27**, 936 (1986).
[31]J. L. Basdevant and G. Preparata, unpublished manuscript.
[32]T. W. Chiu, J. Phys. A **19**, 2537 (1986).
[33]A. Martin, Z. Phys. C **32**, 359 (1986).
[34]A. R. Edmonds, *Angular Momentum in Quantum Mechanics* (Princeton U.P., Princeton, NJ, 1960).
[35]A. M. Badalyan, Phys. Lett. B **199**, 267 (1987); M. Fabre de la Rippelle, *ibid.* **205**, 97 (1988).

On the universality of linear Lagrangians for gravitational field

Andrzej Jakubiec
Institute of Mathematical Physics, University of Warsaw, ul. Hoża 74, 00-682 Warsaw, Poland
Jerzy Kijowski
Institute for Theoretical Physics, Polish Academy of Sciences, Al. Lotnikow 32/46, 02-668 Warsaw, Poland

(Received 19 May 1988; accepted for publication 7 December 1988)

For a large class of purely metric, metric–affine, and purely affine theories of gravitation with nonlinear Lagrangians, it is proved that the theory is equivalent to the standard Einstein theory of gravitation interacting with additional matter fields.

I. INTRODUCTION

Recently, a lot of interest has been devoted to "generalizations of general relativity." Many authors consider theories with Lagrangians depending on the curvature in a nonlinear way.[1] The dynamical structure of such a theory is usually very complicated (e.g., fourth-order differential equations for the metric tensor). Especially obscure is the Cauchy problem, the description of the canonical structure, definition and positivity of energy, etc. In the present paper we prove that many of these generalized theories are equivalent to the standard Einstein theory of gravitation, interacting with additional matter fields. More precisely, for a theory derived from a Lagrangian depending on the Ricci tensor it is possible (under some regularity conditions) to define a new metric tensor and new matter fields in such a way that the field equations rewritten in terms of new quantities can be derived from the standard linear Einstein–Hilbert Lagrangian. Usually, such a transformation leads to considerable simplification of a theory (the number of independent fields or the differential degree of equations decreases). The mathematical analysis of the original theory based on our transformation is relatively easy because we have in general relativity a lot of standard techniques for studying the Cauchy problem, the problem of stability of the theory, etc. For example, the energy of the entire system is composed of the gravitational energy of the new metric (which—as we know—is positive) and the matter energy. To check the stability of the evolution it is, therefore, sufficient to check the algebraic properties (e.g., positivity) of the energy-momentum tensor of the new matter fields.

We use the following notation for the Ricci tensor of a connection Γ:

$$R_{\mu\nu}(j^1\Gamma) = \Gamma_\mu{}^\lambda{}_{\nu\lambda} - \Gamma_\mu{}^\lambda{}_{\lambda\nu} + \Gamma_\sigma{}^\lambda{}_\lambda\Gamma_\mu{}^\sigma{}_\nu - \Gamma_\sigma{}^\lambda{}_\nu\Gamma_\mu{}^\sigma{}_\lambda . \tag{1.1}$$

Here, for a field of a geometric object f, we denote by $j^1 f$ the first jet of f (the value of f and its derivatives). Whenever (f^B) is a coordinate representation of the field f then $(f^B, f^B{}_\sigma)$ is the coordinate representation of $j^1 f$, where we denote $f^B{}_\sigma := \partial_\sigma f^B$ (e.g., $\Gamma_\mu{}^\lambda{}_{\nu\sigma} := \partial_\sigma \Gamma_\mu{}^\lambda{}_\nu$). The covariant derivative with respect to Γ will be denoted by D.

The main result of the paper is contained in Sec. IV. The principal mathematical tool that we use is the appropriate analysis[2] of a general affine theory, contained in Sec. II.

II. AFFINE THEORY

We consider an affine theory based on a symmetric connection $\Gamma_\mu{}^\lambda{}_\nu$ and a matter field $\phi = (\phi^B)$. We assume that the Lagrangian $\mathbf{L}$ of the theory depends on the curvature *via* the Ricci tensor only:

$$\mathbf{L}(j^1\Gamma, j^1\phi) = \mathbf{L}_A(R_{\mu\nu}(j^1\Gamma), \Gamma, j^1\phi). \tag{2.1}$$

(The more general case[3] leads to weaker results.) From the point of view of field dynamics, nonsymmetric connections do not introduce new phenomena; indeed, the torsion can always be incorporated in ϕ as an additional matter field.[4] Introduce momenta canonically conjugate to both Γ and ϕ:

$$\pi^\mu{}_\lambda{}^{\nu\sigma} := \frac{\partial \mathbf{L}}{\partial \Gamma_\mu{}^\lambda{}_{\nu\sigma}}, \tag{2.2}$$

$$p_B{}^\sigma := \frac{\partial \mathbf{L}}{\partial \phi^B{}_\sigma}. \tag{2.3}$$

The derivative (2.2) is not uniquely defined unless we assume the symmetry $\pi^\mu{}_\lambda{}^{\nu\sigma} = \pi^\nu{}_\lambda{}^{\mu\sigma}$ corresponding to the symmetry of Γ. The momentum π is a tensor density. The character of the momentum p depends on the character of the matter field ϕ. For a tensorial matter field ϕ the momentum p is a tensor density, too. The Euler–Lagrange equations of the theory,

$$\frac{\delta \mathbf{L}}{\delta \Gamma_\mu{}^\lambda{}_\nu} = 0, \tag{2.4}$$

$$\frac{\delta \mathbf{L}}{\delta \phi^B} = 0, \tag{2.5}$$

can be rewritten[2,5] as

$$D_\sigma \pi^\mu{}_\lambda{}^{\nu\sigma} = \frac{\partial \mathbf{L}_A}{\partial \Gamma_\mu{}^\lambda{}_\nu}, \tag{2.6}$$

$$\partial_\sigma p_B{}^\sigma = \frac{\partial \mathbf{L}_A}{\partial \phi^B}, \tag{2.7}$$

where the momenta on the left-hand side are defined by (2.2) and (2.3). Introduce the symmetric and the antisymmetric part of the Ricci tensor:

$$K_{\mu\nu} = R_{(\mu\nu)} = \tfrac{1}{2}(R_{\mu\nu} + R_{\nu\mu}), \tag{2.8}$$

$$P_{\mu\nu} = R_{[\mu\nu]} = \tfrac{1}{2}(R_{\mu\nu} - R_{\nu\mu}). \tag{2.9}$$

Similarly, we split $\partial \mathbf{L}_A / \partial R_{\mu\nu}$:

 0022-2488/89/051073-04$02.50

$$\rho^{\mu\nu}=\rho^{(\mu\nu)}:=\frac{\partial \mathbf{L}_A}{\partial K_{\mu\nu}}, \tag{2.10}$$

$$\tau^{\mu\nu}=\tau^{[\mu\nu]}:=\frac{\partial \mathbf{L}_A}{\partial P_{\mu\nu}}, \tag{2.11}$$

where the convention for the derivatives of the Lagrangian is given by the formula:

$$d\mathbf{L}_A=\rho^{\mu\nu}\,dK_{\mu\nu}+\tau^{\mu\nu}\,dP_{\mu\nu}+\cdots.$$

Due to (2.1), (2.2), (2.10), and (2.11) we have

$$\pi^{\mu}{}_{\lambda}{}^{\nu\sigma}=\rho^{\alpha\beta}\frac{\partial K_{\alpha\beta}}{\partial \Gamma_{\mu}{}^{\lambda}{}_{\nu\sigma}}+\tau^{\alpha\beta}\frac{\partial P_{\alpha\beta}}{\partial \Gamma_{\mu}{}^{\lambda}{}_{\nu\sigma}}$$
$$=\delta_{\lambda}{}^{\sigma}\rho^{\mu\nu}-\delta_{\lambda}{}^{(\mu}\rho^{\nu)\sigma}-\delta_{\lambda}{}^{(\mu}\tau^{\nu)\sigma}. \tag{2.12}$$

The symmetric connection Γ splits into two independent geometric objects[6]:

$$\Gamma_{\mu}{}^{\lambda}{}_{\nu}=\Sigma_{\mu}{}^{\lambda}{}_{\nu}+\tfrac{1}{5}\delta_{\mu}{}^{\lambda}\alpha_{\nu}+\tfrac{1}{5}\delta_{\nu}{}^{\lambda}\alpha_{\mu}, \tag{2.13}$$

where $\Sigma_{\mu}{}^{\lambda}{}_{\nu}$ is a symmetric, projective connection ($\Sigma_{\lambda}{}^{\lambda}{}_{\mu}=0$), i.e., a connection in the projective bundle $\mathbf{P}(TM)$ and $\alpha_{\mu}:=\Gamma_{\lambda}{}^{\lambda}{}_{\mu}$ is a connection in the bundle of scalar densities. We have thus a theory of three independent fields; Σ, α, and ϕ. The curvature R can be expressed in terms of the first jets of Σ and α:

$$K_{\mu\nu}=\partial_{\sigma}\Sigma_{\mu}{}^{\sigma}{}_{\nu}-\tfrac{3}{5}\alpha_{(\mu\nu)}-\Sigma_{\mu}{}^{\lambda}{}_{\sigma}\Sigma_{\nu}{}^{\sigma}{}_{\lambda}$$
$$+\tfrac{3}{5}\Sigma_{\mu}{}^{\lambda}{}_{\nu}\alpha_{\lambda}+\tfrac{3}{25}\alpha_{\mu}\alpha_{\nu}, \tag{2.14}$$

$$P_{\mu\nu}=\tfrac{1}{2}(\alpha_{\nu\mu}-\alpha_{\mu\nu}). \tag{2.15}$$

Observe that $P_{\mu\nu}$ is the curvature of α. We define the momenta conjugate to Σ and α:

$$\eta^{\mu}{}_{\lambda}{}^{\nu\sigma}:=\frac{\partial \mathbf{L}}{\partial \Sigma_{\mu}{}^{\lambda}{}_{\nu\sigma}}, \tag{2.16}$$

$$\xi^{\mu\sigma}:=\frac{\partial \mathbf{L}}{\partial \alpha_{\mu\sigma}}, \tag{2.17}$$

equal to the traceless part ($\eta^{\lambda}{}_{\lambda}{}^{\nu\sigma}=0$) and the trace ($\xi^{\nu\sigma}=\tfrac{2}{5}\pi^{\lambda}{}_{\lambda}{}^{\nu\sigma}$) of π. Decomposition (2.13) of configurations implies the decomposition of momenta:

$$\pi^{\mu}{}_{\lambda}{}^{\nu\sigma}=\eta^{\mu}{}_{\lambda}{}^{\nu\sigma}+\delta_{\lambda}{}^{(\mu}\xi^{\nu)\sigma}. \tag{2.18}$$

The formulas (2.18) and (2.12) imply

$$\eta^{\mu}{}_{\lambda}{}^{\nu\sigma}=\delta_{\lambda}{}^{\sigma}\rho^{\mu\nu}-\tfrac{2}{5}\delta_{\lambda}{}^{(\mu}\rho^{\nu)\sigma}, \tag{2.19}$$

$$\xi^{\mu\nu}=-\tau^{\mu\nu}-\tfrac{3}{5}\rho^{\mu\nu}. \tag{2.20}$$

We see that the entire information about η is contained in ρ.

The Euler–Lagrange equation (2.4) splits into two equations:

$$\frac{\delta \mathbf{L}}{\delta \Sigma_{\mu}{}^{\lambda}{}_{\nu}}=0,\quad \frac{\delta \mathbf{L}}{\delta \alpha_{\mu}}=0 \tag{2.21}$$

[the decomposition of (2.4) into its traceless part and its trace]. Because of (2.18), (2.19), and (2.20), Eq. (2.6)—equivalent to (2.4)—can be rewritten as

$$D_{\lambda}\rho^{\mu\nu}-\tfrac{2}{5}\delta_{\lambda}{}^{(\mu}D_{\sigma}\rho^{\nu)\sigma}=\frac{\partial \mathbf{L}_A}{\partial \Sigma_{\mu}{}^{\lambda}{}_{\nu}}, \tag{2.22}$$

$$-D_{\sigma}\tau^{\mu\sigma}-\tfrac{3}{5}D_{\sigma}\rho^{\mu\sigma}=\frac{\partial \mathbf{L}_A}{\partial \alpha_{\mu}}. \tag{2.23}$$

Now following the method introduced in Refs. 2 and 7 we perform the complete Legendre transformation between Σ and η. We define the new Lagrangian $\mathfrak{U}$ by the formula:

$$\mathfrak{U}:=-\partial_{\sigma}(\Sigma_{\mu}{}^{\lambda}{}_{\nu}\eta^{\mu}{}_{\lambda}{}^{\nu\sigma})+\mathbf{L}_A=-\partial_{\sigma}(\Sigma_{\mu}{}^{\sigma}{}_{\nu}\rho^{\mu\nu})+\mathbf{L}_A. \tag{2.24}$$

To perform the Legendre transformation we have to express $\Sigma_{\mu}{}^{\lambda}{}_{\nu}$ and its derivatives in terms of $\eta^{\mu}{}_{\lambda}{}^{\nu\sigma}$ and its derivatives (i.e., in terms of $\rho^{\mu\nu}$ and its derivatives). Standard arguments show that Eqs. (2.5) and (2.21) are equivalent to

$$\frac{\delta \mathfrak{U}}{\delta \phi^{B}}=0,\quad \frac{\delta \mathfrak{U}}{\delta \rho^{\mu\nu}}=0,\quad \frac{\delta \mathfrak{U}}{\delta \alpha_{\mu}}=0. \tag{2.25}$$

Due to (2.24) and (2.14) we see that the only derivatives of Σ that are involved in the definition of $\mathfrak{U}$ are those equal to the "divergence" $\partial_{\sigma}\Sigma_{\mu}{}^{\sigma}{}_{\nu}$. It is therefore sufficient to express $\Sigma_{\mu}{}^{\lambda}{}_{\nu}$ and $K_{\mu\nu}$ in terms of $\alpha, P, j^1\phi$, and $j^1\rho$. For this purpose we solve $36+10$ algebraic equations (2.22) and (2.10) for $36+10$ variables Σ and K contained in $\mathbf{L}_A$. *We assume that these equations can be solved uniquely* (without this regularity condition we would obtain a theory with additional constraints[2,8]). However, formulas (2.22) and (2.24) imply that only the traceless part of $\rho^{\mu\nu}{}_{\lambda}=\partial_{\lambda}\rho^{\mu\nu}$ is involved. We denote by $\tilde{j}^1\rho$ the traceless part of $j^1\rho$. We conclude that our theory enables us to define functions $\mathfrak{S}_{\mu}{}^{\lambda}{}_{\nu}$ and $\mathfrak{K}_{\mu\nu}$ such that

$$\Sigma_{\mu}{}^{\lambda}{}_{\nu}=\mathfrak{S}_{\mu}{}^{\lambda}{}_{\nu}(P,\tilde{j}^1\rho,\alpha,j^1\phi), \tag{2.26}$$

$$K_{\mu\nu}=\mathfrak{K}_{\mu\nu}(P,\tilde{j}^1\rho,\alpha,j^1\phi) \tag{2.27}$$

are solutions of (2.22) and (2.10). Hence, the Lagrangian U can be expressed in terms of $\tilde{j}^1\rho$, $j^1\alpha$, and $j^1\phi$. Due to (2.24), (2.14), (2.26), and (2.27) we have

$$\mathfrak{U}=-\rho^{\mu\nu}(\mathfrak{K}_{\mu\nu}+\tfrac{3}{5}\alpha_{\mu\nu}+\mathfrak{S}_{\mu}{}^{\lambda}{}_{\sigma}\mathfrak{S}_{\nu}{}^{\sigma}{}_{\lambda}-\tfrac{3}{5}\mathfrak{S}_{\mu}{}^{\lambda}{}_{\nu}\alpha_{\lambda}$$
$$-\tfrac{3}{25}\alpha_{\mu}\alpha_{\nu})-\rho^{\mu\nu}{}_{\lambda}\mathfrak{S}_{\mu}{}^{\lambda}{}_{\nu}+\mathbf{L}_A(\mathfrak{K},P,\mathfrak{S},\alpha,j^1\phi). \tag{2.28}$$

The Euler–Lagrange equation,

$$\frac{\delta \mathfrak{U}}{\delta \rho^{\mu\nu}}=0, \tag{2.29}$$

can be replaced by the definition of the momentum

$$\Sigma_{\mu}{}^{\lambda}{}_{\nu}=-\frac{\partial \mathfrak{U}}{\partial \rho^{\mu\nu}{}_{\lambda}}=\mathfrak{S}_{\mu}{}^{\lambda}{}_{\nu}(P,\tilde{j}^1\rho,\alpha,j^1\phi) \tag{2.30}$$

canonically conjugate to $\rho^{\mu\nu}$ and the dynamical equation (2.14) with K and Σ being replaced by $\mathfrak{K}$ and $\mathfrak{S}$. The derivative (2.30) is traceless since $\mathfrak{U}$ does not depend on the trace of $\rho^{\mu\nu}{}_{\lambda}$. Equation (2.29) together with the definition (2.30) is equivalent to Eqs. (2.26) and (2.27), i.e., to (2.22) and (2.10).

The Lagrangian $\mathfrak{U}$ is coordinate dependent, as in the case of the first-order Einstein Lagrangian. We will show later [see Eq. (2.32)] that adding a complete divergence to (2.28) we can obtain an invariant Lagrangian of the second differential order in $\rho^{\mu\nu}$. For this purpose we introduce a new symmetric connection $\{_{\mu}{}^{\lambda}{}_{\nu}\}$, and the corresponding covariant derivative ∇, such that

$$\nabla_{\lambda}\rho^{\mu\nu}=0. \tag{2.31}$$

The above formula defines uniquely $\{_{\mu}{}^{\lambda}{}_{\nu}\}$ in terms of the first jet of ρ provided $\det(\rho^{\mu\nu})\neq 0$, *which we have to assume.* We define our invariant Lagrangian $\mathfrak{L}$ by the formula:

$$\mathfrak{L} = \mathfrak{U} + \partial_\sigma(\{_\mu{}^\lambda{}_\nu\}\eta^\mu{}_\lambda{}^{\nu\sigma}) + \tfrac{3}{5}\partial_\sigma(a_\mu\rho^{\mu\sigma}), \tag{2.32}$$

where by a_μ we denote the covector $a_\mu = \alpha_\mu - \{_\lambda{}^\lambda{}_\mu\}$. To produce an invariant quantity it is sufficient to use only the first divergence in the above formula. The last term is an invariant scalar density and has been added only for convenience (this way we cancel the term $\rho^{\mu\nu}\alpha_{\mu\nu}$ in $\mathfrak{U}$ that contains the symmetric part of $\alpha_{\mu\nu}$). Finally $\mathfrak{L}$ depends on the derivatives of α *via* P only. The price we pay for it is the dependence of $\mathfrak{L}$ on the entire set of derivatives of ρ ($j^1\rho$ instead of $\tilde{j}^1\rho$). The numerical value of $\mathfrak{L}$ can be calculated using (2.24):

$$\mathfrak{L} = \partial_\lambda[(\{_\mu{}^\lambda{}_\nu\} - \Gamma_\mu{}^\lambda{}_\nu)\rho^{\mu\nu}] + \partial_\sigma(a_\mu\rho^{\mu\sigma}) + L_A. \tag{2.33}$$

This proves the invariance of the Lagrangian $\mathfrak{L}$ depending on the second derivatives of $\rho^{\mu\nu}$. Easy calculations lead to the following result:

$$\mathfrak{L} = \rho^{\mu\nu}K_{\mu\nu}(j^2\rho) + \mathfrak{L}_{\rm mat}(P, j^1\rho, \alpha, j^1\phi), \tag{2.34}$$

where by $K_{\mu\nu}(j^2\rho)$ we denote the Ricci tensor (which is *a priori* symmetric) of the connection $\{_\mu{}^\lambda{}_\nu\}$. The "matter Lagrangian" $\mathfrak{L}_{\rm mat}$ is given by the formula:

$$\begin{aligned}\mathfrak{L}_{\rm mat} = &-\rho^{\mu\nu}[\mathfrak{R}_{\mu\nu} + \mathfrak{G}_\mu{}^\lambda{}_\sigma\mathfrak{G}_\nu{}^\sigma{}_\lambda - \mathfrak{G}_\mu{}^\lambda{}_\nu a_\lambda]\\ &+ L_A(\mathfrak{R}, P, \mathfrak{S}, \alpha, j^1\phi).\end{aligned} \tag{2.35}$$

We introduced the symmetric tensor valued function $\mathfrak{G}_\mu{}^\lambda{}_\nu = \mathfrak{G}_\mu{}^\lambda{}_\nu(P, j^1\rho, \alpha, j^1\phi)$:

$$\mathfrak{G}_\mu{}^\lambda{}_\nu := \mathfrak{S}_\mu{}^\lambda{}_\nu + \tfrac{2}{5}\delta_{(\mu}{}^\lambda\alpha_{\nu)} - \{_\mu{}^\lambda{}_\nu\}. \tag{2.36}$$

We observe that numerically $\mathfrak{G}_\mu{}^\lambda{}_\nu = \Gamma_\mu{}^\lambda{}_\nu - \{_\mu{}^\lambda{}_\nu\}$ is the nonmetricity of Γ. The assumption $\det(\rho^{\mu\nu}) \neq 0$ allows us to define the new "metric tensor" $h_{\mu\nu}$ and the inverse tensor $h^{\mu\nu}$ by the following formula[5]:

$$\rho^{\mu\nu} := -(1/2\varkappa)\sqrt{-\det h_{\alpha\beta}}\,h^{\mu\nu}, \tag{2.37}$$

where $\kappa = 8\pi G$ is the gravitational constant. We see that (2.31) is equivalent to $\nabla_\lambda h_{\mu\nu} = 0$, so that $\{_\mu{}^\lambda{}_\nu\}$ are the Christoffel symbols of h. The first term in (2.34) is the scalar curvature of the new metric. Observe that the expression (2.15) for $P_{\mu\nu}$ remains valid if we replace α_μ by a_μ since the curl of $\{_\lambda{}^\lambda{}_\mu\} = \partial_\mu(\ln\sqrt{-\det h_{\alpha\beta}})$ vanishes. Finally, we can choose "the metric" $h_{\mu\nu}$, the covector field a_μ, and the original matter field ϕ^B as independent variables. We see, that the Lagrangian (2.34) is the Einstein–Hilbert Lagrangian with its matter part depending on matter fields a_μ and ϕ^B, their first derivatives, the metric $h_{\mu\nu}$, and its Levi-Civita connection. Observe that a_μ is the nonlinear Proca field since the Lagrangian depends on its derivatives *via* $P_{\mu\nu}$ only. The Einstein equations of this theory ($\delta L/\delta h_{\mu\nu} = 0$) can be rewritten as:

$$\begin{aligned}K_{\mu\nu}(j^2h) = \mathfrak{R}_{\mu\nu} - \nabla_\sigma\mathfrak{G}_\mu{}^\sigma{}_\nu + \tfrac{1}{2}\nabla_{(\mu}a_{\nu)}\\ + \mathfrak{G}_\mu{}^\lambda{}_\sigma\mathfrak{G}_\nu{}^\sigma{}_\lambda - \mathfrak{G}_\mu{}^\lambda{}_\nu a_\lambda\end{aligned} \tag{2.38}$$

and are equivalent to (2.14). Variation of $\mathfrak{L}$ with respect to a_μ gives

$$-\partial_\sigma\tau^{\mu\sigma} + (3/10\varkappa)\sqrt{-\det h_{\alpha\beta}}\,h^{\lambda\sigma}\mathfrak{G}_\lambda{}^\mu{}_\sigma = \frac{\partial \mathbf{L}_A}{\partial a_\mu} \tag{2.39}$$

equivalent to (2.23). Variation with respect to ϕ reproduces Eq. (2.7).

III. METRIC–AFFINE THEORY

Now, we consider a metric-affine theory based on a symmetric connection $\Gamma_\mu{}^\lambda{}_\nu$, a metric tensor $g_{\mu\nu}$, and a Lagrangian L depending on the curvature *via* the Ricci tensor only:

$$\mathbf{L}(j^1\Gamma, j^1g, j^1\phi) = \mathbf{L}_{MA}(R_{\mu\nu}(j^1\Gamma), \Gamma, \gamma, g, j^1\phi), \tag{3.1}$$

where by γ we denote the Christoffel symbols of the metric g. From the purely mathematical point of view this theory is a special case of the one discussed in the previous section, with both g and ϕ treated as "matter fields." Due to the Legendre transformation that we described in the previous section we obtain the equivalence of the above theory with the Einstein theory for the new metric $h_{\mu\nu}$, interacting with three "matter fields": $g_{\mu\nu}$, a_μ, and ϕ^B. The matter fields enter into the Lagrangian *via* their first jets only.

We do not want to discuss the "philosophical" question of which of the fields, h or g, is the "true" metric and which one is merely an additional matter field. An important argument for h being the true metric is the role of light cones of h in the causal properties of the Einstein theory based on the Lagrangian (2.34). Generically, a solution of our field equations admits singularities, i.e., points where the signature of the metric becomes nonphysical. The transition between physical and nonphysical regions of space-time corresponds to extremal matter densities. Usually, regions which are nonphysical with respect to the metric g do not coincide with nonphysical regions for h. A deep analysis of these phenomena could probably help us to decide which metric is more physical. In most examples, however, regions corresponding to relatively weak density of matter are equally good for both g and h.

A special case is the Palatini formulation of general relativity. Here the Lagrangian (3.1) is linear with respect to the curvature:

$$\begin{aligned}\mathbf{L}(j^1\Gamma, j^1g, j^1\phi) = &-(1/2\kappa)\sqrt{-\det g}\,g^{\mu\nu}K_{\mu\nu}(j^1\Gamma)\\ &+ \mathbf{L}_{\rm mat}(\Gamma, \gamma, g, j^1\phi).\end{aligned} \tag{3.2}$$

In this case we have

$$\frac{\partial L_{MA}}{\partial K_{\mu\nu}} = -(1/2\kappa)\sqrt{-\det g}\,g^{\mu\nu}, \tag{3.3}$$

which together with (2.10) and (2.37) implies $h_{\mu\nu} = g_{\mu\nu}$. Moreover, the Lagrangian does not depend on derivatives of the field a_μ (the field a_μ can therefore be eliminated algebraically[7]).

Another example is the variational principle proposed recently by Moffat.[9] Using our techniques it is possible to prove that his theory of general nonsymmetric connection interacting with nonsymmetric metric is equivalent to the standard Einstein theory of the symmetric metric interacting with a second rank, antisymmetric matter field (see paper by the present authors in Ref. 4).

IV. METRIC THEORY

Consider now a purely metric case:

$$\mathbf{L}(j^2g, j^1\phi) = \mathbf{L}_M(R_{\mu\nu}(j^2g), \gamma, g, j^1\phi) \tag{4.1}$$

(obviously, $R_{\mu\nu} = K_{\mu\nu}$). This theory may be treated as an affine-metric theory with the Lagrangian,

$$L(j^1\Gamma, j^1 g, j^1\phi) = \mathbf{L}_M(K_{\mu\nu}(j^1\Gamma), \Gamma, g, j^1\phi) \qquad (4.2)$$

and with Lagrangian constraint,

$$\Gamma_\mu{}^\lambda{}_\nu = \gamma_\mu{}^\lambda{}_\nu. \qquad (4.3)$$

Following the method presented in previous sections, we perform the Legendre transformation between configuration $\Sigma_\mu{}^\lambda{}_\nu$ and the corresponding momentum $p^{\mu\nu}$ (the new metric tensor $h_{\mu\nu}$). Constraint (4.3) implies that the field α_μ is not an independent field $[\alpha_\mu = \partial_\mu(\ln\sqrt{-\det g})]$. We end up with the theory of the "new metric" h and two "matter fields": g and ϕ. Because of the formula (4.3) the solution of (2.22) is trivial: $\mathfrak{S}_\mu{}^\lambda{}_\nu$ is simply the traceless part of $\gamma_\mu{}^\lambda{}_\nu$. Therefore $\mathfrak{G}_\mu{}^\lambda{}_\nu(j^1 h, j^1 g) = \gamma_\mu{}^\lambda{}_\nu - \{_\mu{}^\lambda{}_\nu\}$. To express the new Lagrangian in terms of legal quantities it is therefore sufficient to solve only ten equations (2.10) in order to find functions $\mathfrak{R}_{\mu\nu}$. The invariant Lagrangian $\mathfrak{L}$ (2.34) now has the form:

$$\mathfrak{L} = -(1/2\kappa)\sqrt{-\det h}\, h^{\mu\nu} K_{\mu\nu}(j^2 h) + \mathfrak{L}_{\mathrm{mat}}(j^1 h, j^1 g, j^1\phi), \qquad (4.4)$$

where

$$\mathfrak{L}_{\mathrm{mat}} = (1/2\kappa)\sqrt{-\det h}\, h^{\mu\nu}(\mathfrak{R}_{\mu\nu} + \mathfrak{G}_\mu{}^\lambda{}_\sigma \mathfrak{G}_\nu{}^\sigma{}_\lambda - \mathfrak{G}_\mu{}^\lambda{}_\nu \mathfrak{G}_\lambda{}^\sigma{}_\sigma) + \mathbf{L}_M(\mathfrak{R}, \gamma, g, j^1\phi). \qquad (4.5)$$

Expressing $\mathfrak{G}$ in terms of $\nabla_\alpha g_{\mu\nu}$:

$$\mathfrak{G}_\mu{}^\lambda{}_\nu = \gamma_\mu{}^\lambda{}_\nu - \{_\mu{}^\lambda{}_\nu\} = \tfrac{1}{2} g^{\lambda\sigma}(\nabla_\mu g_{\nu\sigma} + \nabla_\nu g_{\mu\sigma} - \nabla_\sigma g_{\mu\nu}), \qquad (4.6)$$

we may rewrite (4.5) as

$$\mathfrak{L}_{\mathrm{mat}} = (1/2\kappa)\sqrt{-\det h}\, h^{\mu\nu}\{\mathfrak{R}_{\mu\nu} + \tfrac{1}{4} g^{\alpha\beta} g^{\lambda\sigma}[2\nabla_\lambda g_{\beta\nu}(\nabla_\alpha g_{\mu\sigma} - \nabla_\sigma g_{\mu\alpha}) + \nabla_\mu g_{\alpha\sigma}\nabla_\nu g_{\beta\lambda} + \nabla_\lambda g_{\alpha\beta}(\nabla_\sigma g_{\mu\nu} - 2\nabla_\nu g_{\mu\sigma})]\} + \mathbf{L}_M(\mathfrak{R}, \gamma, g, j^1\phi). \qquad (4.7)$$

Recently, a lot of interest has been devoted to quadratic Lagrangians[1]

$$\mathbf{L}(j^2 g, j^1\phi) = -(1/2\kappa)\sqrt{-\det g}\,(aR + bR^2 + c g^{\alpha\beta} g^{\mu\nu} R_{\alpha\mu} R_{\beta\nu}) + \mathbf{L}_{\mathrm{mat}}(j^1 g, j^1\phi), \qquad (4.8)$$

where $R = g^{\mu\nu} R_{\mu\nu}$. Equation (2.10) together with the definition (2.37) gives

$$\sqrt{-\det h}\, h^{\mu\nu} = \sqrt{-\det g}\,(a g^{\mu\nu} + 2b g^{\mu\nu} R + 2c g^{\alpha\mu} g^{\beta\nu} R_{\alpha\beta}). \qquad (4.9)$$

Solving the above equation with respect to $R_{\mu\nu}$ we obtain

$$\mathfrak{R}_{\mu\nu}(h,g) = -\frac{a}{2(4b+c)} g_{\mu\nu} + \frac{\sqrt{-\det h}}{2c\sqrt{-\det g}}\left[g_{\alpha\mu} g_{\beta\nu} h^{\alpha\beta} - \frac{b}{4b+c} g_{\mu\nu} g_{\alpha\beta} h^{\alpha\beta}\right]. \qquad (4.10)$$

Inserting (4.10) into (4.7) we obtain

$$\mathfrak{L}_{\mathrm{mat}} = \frac{1}{8\kappa}\sqrt{-\det h}\, h^{\mu\nu} g^{\alpha\beta} g^{\lambda\sigma}[2\nabla_\lambda g_{\beta\nu}(\nabla_\alpha g_{\mu\sigma} - \nabla_\sigma g_{\mu\alpha}) + \nabla_\mu g_{\alpha\sigma}\nabla_\nu g_{\beta\lambda} + \nabla_\lambda g_{\alpha\beta}(\nabla_\sigma g_{\mu\nu} - 2\nabla_\nu g_{\mu\sigma})] + \frac{1}{2\kappa}\left\{-\frac{a}{2(4b+c)}\sqrt{-\det h}\, h^{\mu\nu} g_{\mu\nu} + \frac{a^2}{4b+c}\sqrt{-\det g} + \frac{-\det h}{4c\sqrt{-\det g}}\left[h^{\mu\alpha} h^{\nu\beta} g_{\mu\nu} g_{\alpha\beta} - \frac{b}{4b+c}(h^{\mu\nu} g_{\mu\nu})^2\right]\right\} + \mathbf{L}_{\mathrm{mat}}(j^1 g, j^1\phi). \qquad (4.11)$$

The first ("kinetic") term of the above Lagrangian is universal and does not depend on the particular choice of the original Lagrangian $\mathbf{L}_M$. The second ("Higgs") term uniquely depends on the particular choice of field dynamics.

[1] A. D. Sakharov, Dokl. Akad. Nauk. SSSR **177**, 70 (1967); F. W. Hehl and G. D. Kerlick, Gen. Relativ. Gravit. **9**, 691 (1978); K. S. Stelle, *ibid.* **9**, 353 (1978); K. I. Macrae and R. J. Riegert, Phys. Rev. D **24**, 2555 (1981); Frenkel and K. Brecher, *ibid.* **26**, 368 (1982); P. Teyssandier and Ph. Tourrenc, J. Math. Phys. **24**, 2793 (1983); V. Müller and H.-J. Schmidt, Gen. Relativ. Gravit. **17**, 769, 971 (1985); H. J. Schmidt, Astron. Nachr. **308**, 183 (1987).

[2] J. Kijowski and W. M. Tulczyjew, *A Symplectic Framework for Fields Theories, Lecture Notes in Physics*, Vol. 107 (Springer, Berlin, 1979).

[3] A. Jakubiec and J. Kijowski, Gen. Relativ. Gravit. **19**, 719 (1987).

[4] E. Schrödinger, Proc. R. Irish Acad. Ser. A **51**, 163 (1947); M. Ferraris and J. Kijowski, Gen. Relativ. Gravit. **14**, 37 (1982); A. Jakubiec and J. Kijowski, J. Math. Phys. **30**, 1077 (1989).

[5] J. Kijowski, Gen. Relativ. Gravit. **9**, 857 (1978).

[6] L. P. Eisenhart, *Non-Riemannian Geometry* (Am. Math. Soc., Providence, RI, 1927); Ju. I. Manin, *Kalibrovocnyje polja i kompleksnaja geometrija* (Izdat. Nauka, Moscow, 1984).

[7] M. Ferraris and J. Kijowski, Gen. Relativ. Gravit. **14**, 165 (1982).

[8] A. Jakubiec and J. Kijowski, Phys. Rev. D **37**, 1406 (1988).

[9] J. W. Moffat, J. Math. Phys. **21**, 1798 (1980).

On theories of gravitation with nonsymmetric connection

Andrzej Jakubiec
Institute of Mathematical Physics, University of Warsaw, ul. Hoża 74;00-682 Warsaw, Poland

Jerzy Kijowski
Institute for Theoretical Physics, Polish Academy of Sciences, Al. Lotników 32/46; 02-668 Warsaw, Poland

(Received 19 May 1988; accepted for publication 7 December 1988)

For a large class of theories of gravitation with nonsymmetric connection, based on general nonlinear Lagrangians, it is proved that the theory is equivalent to the standard Einstein theory of gravitation interacting with additional matter fields.

I. INTRODUCTION

In the previous papers[1] we proved that theories of gravitation based on nonlinear Lagrangians depending on the Ricci tensor are essentially equivalent to the standard Einstein theory. All the effects due to nonlinear Lagrangians can be implemented by introducing additional matter fields. More precisely: to a collection of fields of a theory with a nonlinear Lagrangian $\mathbf{L}$ we are able to assign a new metric tensor and new matter fields in such a way that they satisfy equations of Einstein theory if and only if the original fields satisfy the Euler–Lagrange equations derived from the original Lagrangian $\mathbf{L}$. The above result was proved for metric, metric-affine, and purely affine theories with a symmetric connection. In the present paper we show that this result can be easily extended to the case of nonsymmetric connections. As a simple example we prove that Moffat's theory[2] with both nonsymmetric metric and connection is equivalent to the Einstein theory of a symmetric metric tensor interacting with two additional matter fields. This is a mathematical result and does not depend on the physical (or philosophical) interpretation of the variables (e.g., the problem of whether the metric tensor we introduce is the "true" metric or not). The relevance of such a result consists in the fact, that it enables us to analyze the dynamical content of the theory (Cauchy problem, energy positivity, stability, etc.) using standard tools (results concerning the Cauchy problem for non-Einsteinian theories are usually extremely difficult and give very weak results of the Cauchy–Kowalewska type[3].

II. LAGRANGIAN THEORY OF NONSYMMETRIC CONNECTION

We begin with a purely affine theory based on a general (nonsymmetric) connection $\underline{\Gamma}$. It is known[4] that the connection splits uniquely into the three independent geometric objects:

$$\underline{\Gamma}_\mu{}^\lambda{}_\nu = \Gamma_\mu{}^\lambda{}_\nu + \Delta_\mu{}^\lambda{}_\nu + \tfrac{1}{4}\delta_\mu{}^\lambda(A_\nu - \Gamma_\sigma{}^\sigma{}_\nu), \tag{1}$$

where Γ is a symmetric connection (in holonomic coordinates: $\Gamma_\mu{}^\lambda{}_\nu = \Gamma_\nu{}^\lambda{}_\mu$), Δ is a skew-symmetric, traceless tensor field ($\Delta_\mu{}^\lambda{}_\nu + \Delta_\nu{}^\lambda{}_\mu = 0$, $\Delta_\mu{}^\lambda{}_\lambda = 0$), and $A_\nu := \underline{\Gamma}_\lambda{}^\lambda{}_\nu$ is a connection in the bundle of scalar densities over space-time. We use the following notation for the curvature tensor $\mathbf{R}$ of the connection $\underline{\Gamma}$:

$$\mathbf{R}_\mu{}^\lambda{}_{\nu\sigma}(j^1\underline{\Gamma}) = \underline{\Gamma}_\mu{}^\lambda{}_{\sigma\nu} - \underline{\Gamma}_\mu{}^\lambda{}_{\nu\sigma} + \underline{\Gamma}_\alpha{}^\lambda{}_\nu\underline{\Gamma}_\mu{}^\alpha{}_\sigma - \underline{\Gamma}_\alpha{}^\lambda{}_\sigma\underline{\Gamma}_\mu{}^\alpha{}_\nu. \tag{2}$$

Here for a field of geometric objects f we denote by $j^1 f$ the first jet of f (the value of f and its derivatives). Whenever (f^B) is a coordinate representation of the field f then (f^B, f^B_σ) is the coordinate representation of $j^1 f$, where we denote $f^B{}_\sigma := \partial_\sigma f^B$ (e.g., $\underline{\Gamma}_\mu{}^\lambda{}_{\nu\sigma} := \partial_\sigma\underline{\Gamma}_\mu{}^\lambda{}_\nu$).

We consider the following three independent traces of the curvature:

$$\mathbf{K}_{\mu\nu} = \mathbf{R}_{(\mu\nu)} = \tfrac{1}{2}(\mathbf{R}_\mu{}^\lambda{}_{\lambda\nu} + \mathbf{R}_\nu{}^\lambda{}_{\lambda\mu}), \tag{3}$$

$$\mathbf{F}_{\mu\nu} = \mathbf{R}_\lambda{}^\lambda{}_{\mu\nu}, \tag{4}$$

$$\mathbf{P}_{\mu\nu} = 2\mathbf{R}_{[\mu\nu]} - \tfrac{1}{2}\mathbf{F}_{\mu\nu} = \mathbf{R}_\mu{}^\lambda{}_{\lambda\nu} - \mathbf{R}_\nu{}^\lambda{}_{\lambda\mu} - \tfrac{1}{2}\mathbf{F}_{\mu\nu}. \tag{5}$$

We consider an affine theory based on an affine connection $\underline{\Gamma}_\mu{}^\lambda{}_\nu$ and a matter field $\phi = (\phi^A)$. Moreover, we assume that the theory is based on a Lagrangian $\mathbf{L}$ depending on derivatives of the connection *via* the above traces of the curvature only:

$$\mathbf{L}(j^1\underline{\Gamma}, j^1\phi) = \mathbf{L}_A(\mathbf{K}_{\mu\nu}(j^1\underline{\Gamma}), \mathbf{P}_{\mu\nu}(j^1\underline{\Gamma}), \mathbf{F}_{\mu\nu}(j^1\underline{\Gamma}), \underline{\Gamma}, j^1\phi). \tag{6}$$

It is easy to calculate that

$$\mathbf{K}_{\mu\nu}(j^1\underline{\Gamma}) = K_{\mu\nu}(j^1\Gamma) + \Delta_\mu{}^\lambda{}_\sigma\Delta_\nu{}^\sigma{}_\lambda, \tag{7}$$

$$\mathbf{F}_{\mu\nu}(j^1\underline{\Gamma}) = A_{\nu\mu} - A_{\mu\nu}, \tag{8}$$

$$\mathbf{P}_{\mu\nu}(j^1\underline{\Gamma}) = P_{\mu\nu}(j^1\Gamma) + 2D_\lambda\Delta_\mu{}^\lambda{}_\nu, \tag{9}$$

where by $K_{\mu\nu}$ ($P_{\mu\nu}$) we denote the symmetric (skew-symmetric) part of the Ricci tensor of the symmetric connection Γ:

$$K_{\mu\nu} = R_{(\mu\nu)} = \tfrac{1}{2}(R_\mu{}^\lambda{}_{\lambda\nu} + R_\nu{}^\lambda{}_{\lambda\mu}), \tag{10}$$

$$P_{\mu\nu} = R_{[\mu\nu]} = \tfrac{1}{2}(R_\mu{}^\lambda{}_{\lambda\nu} - R_\nu{}^\lambda{}_{\lambda\mu}) = \tfrac{1}{2}(\Gamma_\lambda{}^\lambda{}_{\nu\mu} - \Gamma_\lambda{}^\lambda{}_{\mu\nu}), \tag{11}$$

and by D we denote the covariant derivative with respect to the symmetric connection Γ.

From the point of view of field dynamics, nonsymmetric connections do not introduce new phenomena; indeed, the torsion can always be incorporated into ϕ as an additional matter field. This way our theory can be treated as an affine theory based on the symmetric connection $\Gamma_\mu{}^\lambda{}_\nu$ interacting with the following "matter fields": $\Delta_\mu{}^\lambda{}_\nu$, A_μ, and ϕ^A. Fol-

 0022-2488/89/051077-04$02.50

lowing our method[1] we use the following notation for canonical momenta:

$$\rho^{\mu\nu}=\rho^{(\mu\nu)}:=\frac{\partial \mathbf{L}_A}{\partial \mathbf{K}_{\mu\nu}}=\frac{\partial \mathbf{L}_A}{\partial K_{\mu\nu}}, \tag{12}$$

$$\tau^{\mu\nu}=\tau^{[\mu\nu]}:=\frac{\partial \mathbf{L}_A}{\partial \mathbf{P}_{\mu\nu}}=\frac{\partial \mathbf{L}_A}{\partial P_{\mu\nu}}, \tag{13}$$

$$\omega^{\mu\nu}=\omega^{[\mu\nu]}:=\frac{\partial \mathbf{L}_A}{\partial \mathbf{F}_{\mu\nu}}, \tag{14}$$

$$p_B{}^{\sigma}:=\frac{\partial \mathbf{L}}{\partial \phi^B{}_\sigma}, \tag{15}$$

and for currents of the theory:

$$j^\mu{}_\lambda{}^\nu=\frac{\partial \mathbf{L}}{\partial \Gamma_\mu{}^\lambda{}_\nu}\quad d^\mu{}_\lambda{}^\nu=\frac{\partial \mathbf{L}}{\partial \Delta_\mu{}^\lambda{}_\nu},$$
$$r^\mu=\frac{\partial \mathbf{L}}{\partial A_\mu},\quad p_B=\frac{\partial \mathbf{L}}{\partial \phi^B}. \tag{16}$$

The above formulas do not define uniquely our currents unless we impose the symmetries corresponding to the symmetries of Γ and Δ. The currents j, d, and r are tensor densities. The character of the momentum p depends on the character of the matter field ϕ. For a tensorial matter field ϕ, the momentum p is a tensor density, too.

The Euler–Lagrange equation $\delta\mathbf{L}/\delta\Gamma=0$ has the form

$$D_\lambda\rho^{\mu\nu}+\tfrac{2}{3}\delta_\lambda{}^{(\mu}\partial_\sigma\tau^{\nu)\sigma}=4\Delta_\sigma{}^{(\mu}{}_\lambda\tau^{\nu)\sigma}+j^\mu{}_\lambda{}^\nu-\tfrac{2}{3}j^\sigma{}_\sigma{}^{(\mu}\delta_\lambda{}^{\nu)}. \tag{17}$$

The Euler–Lagrange equations for "matter" fields Δ, A, and ϕ can be written as

$$D_\lambda\tau^{\mu\nu}+\tfrac{2}{3}\delta_\lambda{}^{[\mu}\partial_\sigma\tau^{\nu]\sigma}=-\Delta_\sigma{}^{[\mu}{}_\lambda\rho^{\nu]\sigma}+\tfrac{1}{2}d^\mu{}_\lambda{}^\nu, \tag{18}$$

which is equivalent to $\delta\mathbf{L}/\delta\Delta=0$ and

$$\partial_\sigma\omega^{\mu\sigma}=-\tfrac{1}{2}r^\mu,\quad \partial_\sigma p_B{}^\sigma=p_B, \tag{19}$$

which is equivalent to $\delta\mathbf{L}/\delta A=0$ and $\delta\mathbf{L}/\delta\phi=0$, respectively.

III. METRIC THEORY AS A RESULT OF LEGENDRE TRANSFORMATION

Now we perform the complete Legendre transformation[1] between ρ and the traceless part Σ of Γ ($\Sigma_\mu{}^\lambda{}_\nu:=\Gamma_\mu{}^\lambda{}_\nu-\tfrac{1}{5}\delta_\mu{}^\lambda\Gamma_\sigma{}^\sigma{}_\nu-\tfrac{1}{5}\delta_\nu{}^\lambda\Gamma_\sigma{}^\sigma{}_\mu$). Numerically, the transformation consists in subtracting the term $\partial_\sigma(\Sigma_\mu{}^\sigma{}_\nu\rho^{\mu\nu})$ from the Lagrangian. Then both $\Sigma_\mu{}^\lambda{}_\nu$ and its derivatives have to be calculated in terms of $\rho^{\mu\nu}$, its derivatives and the fields that were not involved into the transformation. As a result of such an operation we obtain a noninvariant Lagrangian that can be improved by adding a complete divergence.[1,5] Such an improved, second-order (in $\rho^{\mu\nu}$) Lagrangian is already an invariant scalar density. Following the ideas of the purely affine theories of gravitation[6] we interprete ρ as a dynamically defined metric tensor. More precisely, we define a new metric tensor $h_{\mu\nu}$ and the inverse (contravariant) tensor $h^{\mu\nu}$ by the formula

$$\rho^{\mu\nu}=:-(1/2\kappa)\sqrt{-\det h_{\alpha\beta}}\,h^{\mu\nu}, \tag{20}$$

where $\kappa=8\pi G$ is the gravitational constant. *The definition is meaningful if we impose an additional assumption that* $\det(-\rho^{\mu\nu})\neq 0$. We denote by $\{_\mu{}^\lambda{}_\nu\}$ the Christoffel symbols of the metric h and by ∇ the corresponding covariant derivative. After the Legendre transformation, our theory may be treated as a metric theory based on the metric tensor h coupled to the following matter fields: Δ, A, ϕ, and a, where the covector field $a_\mu:=\Gamma_\lambda{}^\lambda{}_\mu-\{_\lambda{}^\lambda{}_\mu\}$ describes the nonmetricity of the trace of Γ. It was proved[1] that the new Lagrangian of the theory is equal numerically to the following expression:

$$\mathfrak{L}=\partial_\lambda[(\{_\mu{}^\lambda{}_\nu\}-\Gamma_\mu{}^\lambda{}_\nu)\rho^{\mu\nu}]+\partial_\lambda(a_\mu\rho^{\mu\lambda})+\mathbf{L}_A. \tag{21}$$

Due to (11) we have

$$P_{\mu\nu}=\tfrac{1}{2}(a_{\nu\mu}-a_{\mu\nu}). \tag{22}$$

In order to eliminate the first jet of Σ from the right-hand side of (21) we have to solve Eqs. (12) and the traceless part of (17) with respect to $\mathbf{K}_{\mu\nu}$ and $\Sigma_\mu{}^\lambda{}_\nu$. Let us denote this solutions by $\mathfrak{K}_{\mu\nu}$ and $\mathfrak{S}_\mu{}^\lambda{}_\nu$, respectively. The formula (7) implies

$$K_{\mu\nu}=\mathfrak{K}_{\mu\nu}(j^1h,j^1a,j^1\Delta,j^1A,j^1\phi)-\Delta_\mu{}^\alpha{}_\beta\Delta_\nu{}^\beta{}_\alpha. \tag{23}$$

Due to the definitions of $\Sigma_\mu{}^\lambda{}_\nu$ and a_μ we have

$$\Gamma_\mu{}^\lambda{}_\nu=\mathfrak{S}_\mu{}^\lambda{}_\nu(j^1h,j^1a,j^1\Delta,j^1A,j^1\phi)+\tfrac{2}{5}\delta_{(\mu}{}^\lambda a_{\nu)}$$
$$+\tfrac{1}{5}\delta_\mu{}^\lambda\{_\sigma{}^\sigma{}_\nu\}+\tfrac{1}{5}\delta_\nu{}^\lambda\{_\sigma{}^\sigma{}_\mu\}. \tag{24}$$

We introduce the following tensor-valued function:

$$\mathfrak{G}_\mu{}^\lambda{}_\nu=\mathfrak{G}_\mu{}^\lambda{}_\nu(j^1h,j^1a,j^1\Delta,j^1A,j^1\phi)$$
$$:=\mathfrak{S}_\mu{}^\lambda{}_\nu(j^1h,j^1a,j^1\Delta,j^1A,j^1\phi)$$
$$+\tfrac{2}{5}\delta_{(\mu}{}^\lambda a_{\nu)}+\tfrac{1}{5}\delta_\mu{}^\lambda\{_\sigma{}^\sigma{}_\nu\}$$
$$+\tfrac{1}{5}\delta_\nu{}^\lambda\{_\sigma{}^\sigma{}_\mu\}-\{_\mu{}^\lambda{}_\nu\}. \tag{25}$$

Formula (24) implies that numerically $\mathfrak{G}_\mu{}^\lambda{}_\nu=\Gamma_\mu{}^\lambda{}_\nu-\{_\mu{}^\lambda{}_\nu\}$ is the nonmetricity of Γ.

Observe that derivatives of a_μ and $\Delta_\mu{}^\lambda{}_\nu$ enter into the original Lagrangian *via* $\mathbf{P}_{\mu\nu}$ only. Rewriting formula (9) in terms of the new covariant derivative ∇ and using (22) and (25) we have

$$\mathbf{P}_{\mu\nu}=\tfrac{1}{2}(a_{\nu\mu}-a_{\mu\nu})+2D_\lambda\Delta_\mu{}^\lambda{}_\nu$$
$$=\tfrac{1}{2}(a_{\nu\mu}-a_{\mu\nu})+2(\nabla_\lambda\Delta_\mu{}^\lambda{}_\nu-\mathfrak{G}_\mu{}^\sigma{}_\lambda\Delta_\sigma{}^\lambda{}_\nu$$
$$-\mathfrak{G}_\nu{}^\sigma{}_\lambda\Delta_\mu{}^\lambda{}_\sigma+a_\lambda\Delta_\mu{}^\lambda{}_\nu). \tag{26}$$

It is convenient to code information carried by the tensor $\Delta_\mu{}^\lambda{}_\nu$ and the covector a_μ into a single tensor field:

$$B_\mu{}^\lambda{}_\nu:=2\Delta_\mu{}^\lambda{}_\nu+\tfrac{1}{2}(\delta_\mu{}^\lambda a_\nu-\delta_\nu{}^\lambda a_\mu). \tag{27}$$

We see that derivatives of B enter into the theory *via* the following skew-symmetric tensor field only:

$$\mathbf{B}_{\mu\nu}:=\nabla_\lambda B_\mu{}^\lambda{}_\nu. \tag{28}$$

Now formula (26) can be rewritten as

$$\mathbf{P}_{\mu\nu}=\mathbf{B}_{\mu\nu}-\mathfrak{G}_\mu{}^\sigma{}_\lambda B_\sigma{}^\lambda{}_\nu-\mathfrak{G}_\nu{}^\sigma{}_\lambda B_\mu{}^\lambda{}_\sigma+\tfrac{2}{3}B_\sigma{}^\sigma{}_\lambda B_\mu{}^\lambda{}_\nu. \tag{29}$$

We have therefore $\mathfrak{K}_{\mu\nu}=\mathfrak{K}_{\mu\nu}(j^1h,\mathbf{B},\mathbf{F},B,A,j^1\phi)$ and $\mathfrak{G}_\mu{}^\lambda{}_\nu=\mathfrak{G}_\mu{}^\lambda{}_\nu(j^1h,\mathbf{B},\mathbf{F},B,A,j^1\phi)$. It was proved[1] that the new invariant Lagrangian $\mathfrak{L}$ given by (21) is the Einstein–Hilbert Lagrangian of the metric h with the matter part depending on matter fields A, B, and ϕ:

$$\mathfrak{L} = -(1/2\kappa)\sqrt{-\det h_{\alpha\beta}}\, h^{\mu\nu} K_{\mu\nu}(j^2 h) + \mathfrak{L}_{\rm mat}(j^1 h, \mathbf{B}, \mathbf{F}, B, A, j^1\phi), \tag{30}$$

where

$$\mathfrak{L}_{\rm mat} = (1/2\kappa)\sqrt{-\det h_{\alpha\beta}}\, h^{\mu\nu}[\mathfrak{R}_{\mu\nu} - \Delta_\mu{}^\lambda{}_\sigma \Delta_\nu{}^\sigma{}_\lambda + \mathfrak{G}_\mu{}^\lambda{}_\sigma \mathfrak{G}_\nu{}^\sigma{}_\lambda - \mathfrak{G}_\mu{}^\lambda{}_\nu a_\lambda] + L_A(\mathfrak{R}, \mathbf{P}, \mathbf{F}, \mathfrak{G}, B, A, j^1\phi). \tag{31}$$

Here a and Δ denote the trace and the traceless part of B. The quantities $\mathbf{P}$ and $\mathbf{F}$ are given by (29) and (8), respectively. The function $\mathfrak{G}$ is defined by (25). We stress that the construction of the quantity B was possible only after the first Legendre transformation. To the traceless part Δ of the torsion we add a part of its trace represented by the covector a_μ. However, the definition of a_μ depends on the metric tensor $h_{\mu\nu}$. The transformation from the initial variables Γ and ϕ to the new variables h, A, B, ϕ is not a "point transformation" but is a canonical transformation. This way we have shown that the theory with the affine Lagrangian (6) is equivalent to the Einstein theory based on the Einstein–Hilbert Lagrangian (30).

However, it is possible to simplify further the theory reducing the number of independent matter fields. For this purpose we perform the Legendre transformation between the configuration $B_\mu{}^\lambda{}_\nu$ and the momentum $\tau^{\mu\nu}$. This way we replace a 24-component matter field B by a 6-component field τ. The transformation consists in subtracting the term $\nabla_\lambda(\tau^{\mu\nu}B_\mu{}^\lambda{}_\nu) = \partial_\lambda(\tau^{\mu\nu}B_\mu{}^\lambda{}_\nu)$ from the Lagrangian and calculating $(B_\mu{}^\lambda{}_\nu, \nabla_\lambda B_\mu{}^\lambda{}_\nu)$ in terms of $(\nabla_\lambda \tau^{\mu\nu}, \tau^{\mu\nu})$. We use for this purpose the equation

$$\frac{\partial \mathfrak{L}}{\partial \mathbf{B}_{\mu\nu}} = \tau^{\mu\nu}, \tag{32}$$

equivalent to (13) and the Euler–Lagrange equation

$$\frac{\partial \mathfrak{L}}{\partial B_\mu{}^\lambda{}_\nu} = \nabla_\lambda \tau^{\mu\nu}, \tag{33}$$

equivalent to (18) and the trace of (17). Finally we obtain Einstein theory of the metric h interacting with the following matter fields: τ, A, ϕ. The field equations of the theory can be derived from the Einstein–Hilbert Lagrangian with the following matter–Lagrangian part:

$$\mathscr{L}_{\rm mat} = (1/2\kappa)\sqrt{-\det h_{\alpha\beta}}\, h^{\mu\nu}[\mathfrak{R}_{\mu\nu} - \tfrac{1}{4}B_\mu{}^\lambda{}_\sigma B_\nu{}^\sigma{}_\lambda + \tfrac{1}{12}B_\lambda{}^\lambda{}_\mu B_\sigma{}^\sigma{}_\nu + \mathfrak{G}_\mu{}^\lambda{}_\sigma \mathfrak{G}_\nu{}^\sigma{}_\lambda - \tfrac{2}{3}\mathfrak{G}_\mu{}^\lambda{}_\nu B_\sigma{}^\sigma{}_\lambda] + \mathbf{L}_A(\mathfrak{R}, \mathbf{P}, \mathbf{F}, \mathfrak{G}, B, A, j^1\phi) - \nabla_\lambda(\tau^{\mu\nu}B_\mu{}^\lambda{}_\nu). \tag{34}$$

We remember that the above numerical value of $\mathscr{L}_{\rm mat}$ has to be expressed in terms of τ, A, ϕ, and other derivatives.

For calculational convenience the last Legendre transformation between B and τ can also be performed in two steps. First we perform the Legendre transformation between a_μ and $\tau^{\mu\nu}$. Then we subtract the complete divergence $\partial_\lambda(2\tau^{\mu\nu}\Delta_\mu{}^\lambda{}_\nu)$ from the Lagrangian. Finally we observe that the Lagrangian depends algebraically on Δ. Equations $\delta L/\delta\Delta = 0$ are therefore algebraic with respect to Δ. In a generic situation these equations can be treated as a definition of Δ in terms of $j^1 h$ and $j^1\tau$.

IV. SPECIAL CASE: LAGRANGIAN WITHOUT POTENTIALS

As an example of the above procedure we consider a Lagrangian (6) which does not depend on $\Gamma_\mu{}^\lambda{}_\nu$. We have therefore $j^\mu{}_\lambda{}^\nu = 0$ and $d^\mu{}_\lambda{}^\nu = 0$. One can check that the nonmetricity Eq. (17) for Γ can be solved algebraically with respect to Σ. This way we obtain an explicit form of the function $\mathfrak{S}$. Putting this solution into formula (25) we get the following expression:

$$\begin{aligned}\mathfrak{G}_\mu{}^\lambda{}_\nu = &-(2\kappa/\sqrt{-\det h_{\alpha\beta}})[h_{\alpha\nu}(\tau^{\lambda\beta}\Delta_\beta{}^\alpha{}_\mu + \tau^{\alpha\beta}\Delta_\beta{}^\lambda{}_\mu) \\ &+ h_{\alpha\mu}(\tau^{\lambda\beta}\Delta_\beta{}^\alpha{}_\nu + \tau^{\alpha\beta}\Delta_\beta{}^\lambda{}_\nu) \\ &+ h_{\alpha\mu}h_{\beta\nu}h^{\lambda\sigma}(\tau^{\alpha\xi}\Delta_\xi{}^\beta{}_\sigma + \tau^{\beta\xi}\Delta_\xi{}^\alpha{}_\sigma) \\ &- 2h_{\mu\nu}h_{\alpha\beta}h^{\lambda\sigma}\tau^{\alpha\xi}\Delta_\xi{}^\beta{}_\sigma] \\ &+ \tfrac{1}{2}(\delta_\mu{}^\lambda a_\nu + \delta_\nu{}^\lambda a_\mu - 3h_{\mu\nu}h^{\lambda\sigma}a_\sigma),\end{aligned} \tag{35}$$

where $\tau = \tau(\mathfrak{R}, \mathbf{P}, \mathbf{F}, B, A, j^1\phi)$ is defined by (13). We can therefore calculate $\mathfrak{L}_{\rm mat}$ (31) as follows:

$$\begin{aligned}\mathfrak{L}_{\rm mat}&(j^1 h, \mathbf{B}, \mathbf{F}, B, A, j^1\phi) \\ &= (1/2\kappa)\sqrt{-\det h_{\alpha\beta}}(h^{\mu\nu}\mathfrak{R}_{\mu\nu} + \tfrac{3}{2}h^{\mu\nu}a_\mu a_\nu) \\ &\quad + U(h,\Delta) + W_1(h,\Delta,\tau) + L_A(\mathfrak{R}, \mathbf{P}, \mathbf{F}, B, A, j^1\phi),\end{aligned} \tag{36}$$

where by $U(h,\Delta)$ and $W_1(h,\Delta,\tau)$ we denote the algebraic functions of the metric h,Δ (the traceless part of B) and of $\tau = \tau(\mathfrak{R}, \mathbf{P}, \mathbf{F}, B, A, j^1\phi)$:

$$U(h,\Delta) = -(1/2\kappa)\sqrt{-\det h_{\alpha\beta}}\, h^{\mu\nu}\Delta_\mu{}^\lambda{}_\sigma \Delta_\nu{}^\sigma{}_\lambda, \tag{37}$$

$$\begin{aligned}W_1(h,\Delta,\tau) = &(4\kappa/\sqrt{-\det h_{\alpha\beta}})\tau^{\alpha\beta}\tau^{\lambda\sigma}\{h_{\xi\eta}\Delta_\alpha{}^\xi{}_\lambda \Delta_\beta{}^\eta{}_\sigma \\ &+ 2h_{\sigma\eta}\Delta_\alpha{}^\xi{}_\lambda \Delta_\beta{}^\eta{}_\xi \\ &+ h_{\alpha\lambda}\Delta_\eta{}^\xi{}_\beta \Delta_\xi{}^\eta{}_\sigma - h^{\mu\nu}[h_{\alpha\xi}h_{\lambda\eta}\Delta_\sigma{}^\xi{}_\mu \Delta_\beta{}^\eta{}_\nu \\ &+ (h_{\alpha\lambda}h_{\xi\eta} + 2h_{\alpha\xi}h_{\lambda\eta})\Delta_\beta{}^\xi{}_\mu \Delta_\sigma{}^\eta{}_\nu]\}.\end{aligned} \tag{38}$$

Now we perform the Legendre transformation between the configuration B and the momentum τ. The new matter Lagrangian $\mathscr{L}_{\rm mat}$ equals

$$\begin{aligned}\mathscr{L}_{\rm mat} &= \mathscr{L}_{\rm mat}(j^1 h, j^1\tau, \mathbf{F}, A, j^1\phi) \\ &= \mathfrak{L}_{\rm mat} - \partial_\lambda(\tau^{\mu\nu}B_\mu{}^\lambda{}_\nu) \\ &= \mathfrak{L}_{\rm mat} - \tau^{\mu\nu}P_{\mu\nu} + (\partial_\lambda \tau^{\mu\lambda})a_\mu - 2\nabla_\lambda(\tau^{\mu\nu}\Delta_\mu{}^\lambda{}_\nu).\end{aligned} \tag{39}$$

In order to express the Lagrangian in terms of legal variables we have to solve Eqs. (13) with respect to $\mathbf{B}$. Moreover, we have to solve Eqs. (18) and the trace of (17) with respect to B (i.e., Δ and a). Let us denote the algebraic solution of (13) with respect to $\mathbf{P}$ by $\mathfrak{P}$. Due to (29) we have

$$\mathbf{B}_{\mu\nu} = \mathfrak{P}_{\mu\nu}(j^1 h, j^1\tau, j^1 A, j^1\phi) + \mathfrak{G}_\mu{}^\sigma{}_\lambda B_\sigma{}^\lambda{}_\nu + \mathfrak{G}_\nu{}^\sigma{}_\lambda B_\mu{}^\lambda{}_\sigma - \tfrac{2}{3}B_\sigma{}^\sigma{}_\lambda B_\mu{}^\lambda{}_\nu. \tag{40}$$

The trace of Eq. (17) can be rewritten as

$$a_\mu = (4\kappa/\sqrt{-\det h_{\alpha\beta}})(h_{\lambda\sigma}\tau^{\nu\lambda}\Delta_\mu{}^\sigma{}_\nu - \tfrac{1}{6}h_{\mu\nu}\partial_\lambda \tau^{\nu\lambda}). \tag{41}$$

Equation (18) [with $\Gamma_\mu{}^\lambda{}_\nu$ defined by (24)] is linear with respect to $\Delta_\mu{}^\lambda{}_\nu$. Let $\Delta_\mu{}^\lambda{}_\nu = \Delta_\mu{}^\lambda{}_\nu(j^1 h, j^1\tau)$ be the solution of this equation. Easy calculations lead to the following result:

$$\mathscr{L}_{\text{mat}} = (1/2\kappa)\sqrt{-\det h_{\alpha\beta}}\, h^{\mu\nu}\mathfrak{R}_{\mu\nu} - \tau^{\mu\nu}\mathfrak{P}_{\mu\nu} + L_A(\mathfrak{R},\mathfrak{P},\mathbf{F},\mathfrak{G},\Delta,A,j^1\phi) + (4\kappa/\sqrt{-\det h_{\alpha\beta}})(h_{\lambda\sigma}\tau^{\sigma\mu}\Delta_\mu{}^\lambda{}_\nu\partial_\alpha\tau^{\nu\alpha} - \tfrac{1}{12}h_{\mu\nu}\partial_\lambda\tau^{\mu\lambda}\partial_\sigma\tau^{\nu\sigma}) - 2\Delta_\mu{}^\lambda{}_\nu\nabla_\lambda\tau^{\mu\nu} + U(h,\Delta) - W_2(h,\Delta,\tau) \tag{42}$$

where by $W_2(h,\Delta,\tau)$ we denote the algebraic function of the fields h, τ and of $\Delta = \Delta(j^1h, j^1\tau)$:

$$\begin{aligned} W_2(h,\Delta,\tau) &= W_1(h,\Delta\tau) + (12\kappa/\sqrt{-\det h_{\alpha\beta}})h^{\mu\nu}h_{\alpha\xi}h_{\lambda\eta} \times\tau^{\alpha\beta}\tau^{\lambda\sigma}\Delta_\beta{}^\xi{}_\mu\Delta_\sigma{}^\eta{}_\nu \\ &= (4\kappa/\sqrt{-\det h_{\alpha\beta}})\tau^{\alpha\beta}\tau^{\lambda\sigma}\{h_{\xi\eta}\Delta_\alpha{}^\xi{}_\lambda\Delta_\beta{}^\eta{}_\sigma + 2h_{\sigma\eta}\Delta_\alpha{}^\xi{}_\lambda\Delta_\beta{}^\eta{}_\xi \\ &\quad + h_{\alpha\lambda}\Delta_\eta{}^\xi{}_\beta\Delta_\xi{}^\eta{}_\sigma - h^{\mu\nu}[h_{\alpha\xi}h_{\lambda\eta}\Delta_\sigma{}^\xi{}_\mu\Delta_\beta{}^\eta{}_\nu + (h_{\alpha\lambda}h_{\xi\eta} - h_{\alpha\xi}h_{\lambda\eta})\Delta_\beta{}^\xi{}_\mu\Delta_\sigma{}^\eta{}_\nu]\}. \end{aligned} \tag{43}$$

V. MOFFAT'S THEORY

The above considerations can be applied to Moffat's theory[2] (the gravitation theory with a nonsymmetric "metric"). The theory is based on the following Lagrangian $\mathbf{L}_{\text{mof}}$:

$$\mathbf{L}_{\text{Mof}} = \mathfrak{g}^{\mu\nu}\mathfrak{M}_{\mu\nu}(j^1\underline{\Gamma}) + \mathbf{L}_{\text{mat}}(\mathfrak{g}^{\mu\nu}, W_\lambda, j^1\phi), \tag{44}$$

where $\mathfrak{g}^{\mu\nu}$ is a (nonsymmetric) tensor density,

$$\mathfrak{M}_{\mu\nu}(j^1\underline{\Gamma}) = R_{\mu\nu}(j^1\underline{\Gamma}) + \tfrac{1}{2}(\underline{\Gamma}_\mu{}^\lambda{}_{\lambda\nu} - \underline{\Gamma}_\nu{}^\lambda{}_{\lambda\mu}) \tag{45}$$

and

$$W_\mu = \tfrac{3}{8}(\Gamma_\lambda{}^\lambda{}_\mu - \underline{\Gamma}_\lambda{}^\lambda{}_\mu) = \tfrac{3}{8}(\Gamma_\lambda{}^\lambda{}_\mu - A_\mu). \tag{46}$$

The variation is meant with respect to $\underline{\Gamma}$ and $\mathfrak{g}$ independently. Here $\mathbf{L}_{\text{mat}}$ is a Lagrangian density for the "phenomenological matter sources."[2]

Due to the decomposition of $\mathfrak{g}^{\mu\nu}$ into symmetric and skew-symmetric parts:

$$\rho^{\mu\nu} = \mathfrak{g}^{(\mu\nu)}, \tag{47}$$

$$\tau^{\mu\nu} = -\tfrac{1}{4}\mathfrak{g}^{[\mu\nu]}, \tag{48}$$

we can rewrite formula (44):

$$\mathbf{L}_{\text{Mof}} = \rho^{\mu\nu}\mathbf{K}_{\mu\nu}(j^1\underline{\Gamma}) + \tau^{\mu\nu}[\mathbf{P}_{\mu\nu}(j^1\underline{\Gamma}) - \tfrac{1}{2}\mathbf{F}_{\mu\nu}(j^1\underline{\Gamma}) - 6D_\lambda\Delta_\mu{}^\lambda{}_\nu] + \mathbf{L}_{\text{mat}}. \tag{49}$$

This Lagrangian differs from (6) by the term containing the derivatives of the torsion. However, the derivatives enter *via* the same combinations as the ones contained already in $\mathbf{P}_{\mu\nu}$. Therefore, our techniques can be applied also in this case.

The variation with respect to $\mathfrak{g}$ can be replaced by the independent variations with respect to ρ and τ. The Euler–Lagrange equations $\delta\mathbf{L}/\delta A = 0$, $\delta\mathbf{L}/\delta\Gamma = 0$, and $\delta\mathbf{L}/\delta\Delta = 0$ have the form:

$$\partial_\sigma\tau^{\mu\sigma} = r^\mu = -\tfrac{8}{3}\frac{\partial\mathbf{L}}{\partial W_\mu} = -\tfrac{8}{3}\frac{\partial\mathbf{L}_{\text{mat}}}{\partial W_\mu}, \tag{50}$$

$$D_\lambda\rho^{\mu\nu} = -8\Delta_\sigma{}^{(\mu}{}_\lambda\tau^{\nu)\sigma}, \tag{51}$$

$$D_\lambda\tau^{\mu\nu} + \tfrac{2}{3}\delta_\lambda{}^{[\mu}\partial_\sigma\tau^{\nu]\sigma} = \tfrac{1}{2}\Delta_\sigma{}^{[\mu}{}_\lambda\rho^{\nu]\sigma}. \tag{52}$$

Equation (51) implies

$$a_\mu = -(8\kappa/\sqrt{-\det h_{\alpha\beta}})h_{\lambda\sigma}\tau^{\nu\lambda}\Delta_\mu{}^\sigma{}_\nu. \tag{53}$$

Comparing the above result with formula (41) we see that a_μ is already expressed algebraically. Consequently, after the first Legendre transformation between $\Sigma_\mu{}^\lambda{}_\nu$ and $\rho^{\mu\nu}$, reexpressing $D_\lambda\Delta_\mu{}^\lambda{}_\nu$ in terms of j^1h, τ, $j^1\Delta$ and using (53) we obtain the following numerical value of the matter Lagrangian:

$$\mathfrak{L}_{\text{mat}} = \tau^{\mu\nu}[P_{\mu\nu}(j^1\Gamma) - \tfrac{1}{2}\mathbf{F}_{\mu\nu}(j^1\underline{\Gamma}) - 4\nabla_\lambda\Delta_\mu{}^\lambda{}_\nu] + U(h,\Delta) - 4W_2(h,\Delta,\tau) + \mathbf{L}_{\text{mat}}. \tag{54}$$

Moreover, it is convenient to use the field W_μ instead of a couple of fields A_μ and $\Gamma_\lambda{}^\lambda{}_\mu$. Finally, adding the complete divergence $4\nabla_\lambda(\tau^{\mu\nu}\Delta_\mu{}^\lambda{}_\nu) = 4\partial_\lambda(\tau^{\mu\nu}\Delta_\mu{}^\lambda{}_\nu)$ we obtain the matter Lagrangian for the Moffat's theory:

$$\begin{aligned} \mathscr{L}_{\text{mat}} &= \mathscr{L}_{\text{mat}}(j^1h, j^1\tau, j^1W, \Delta, j^1\phi) \\ &= -\tfrac{8}{3}\tau^{\mu\nu}W_{\mu\nu} + 4\Delta_\mu{}^\lambda{}_\nu\nabla_\lambda\tau^{\mu\nu} + U(h,\Delta) - 4W_2(h,\Delta,\tau) + \mathbf{L}_{\text{mat}}(h,\tau,W_\lambda,j^1\phi). \end{aligned} \tag{55}$$

We notice that Δ enters only algebraically into the above Lagrangian. Equations $\delta\mathbf{L}/\delta\Delta = 0$ are therefore algebraic equations for Δ. In a generic situation these equations can be treated as a definition of Δ in terms of j^1h and $j^1\tau$. This way we can eliminate Δ as an independent variable of the theory. Analytically, the function $\Delta(j^1h, j^1\tau)$ is very complicated. Finally, we obtain the theory with two independent matter fields: $\tau^{\mu\nu}$ and W_μ coupled to the (symmetric) metric tensor $h_{\mu\nu}$. The field equations can be derived from the following Einstein–Hilbert Lagrangian:

$$\mathfrak{L}(j^2h, j^1W) = -(1/2\kappa)\sqrt{-\det h_{\alpha\beta}}\, h^{\mu\nu}K_{\mu\nu}(j^2h) + \mathscr{L}_{\text{mat}}(j^1h, j^1\tau, j^1W, \Delta(j^1h, j^1\tau)j^1\phi). \tag{56}$$

Observe, that both metric tensor h and matter fields (τ and W) are original Moffat's fields. We therefore proved, that subtracting the complete divergence from the Lagrangian, and reexpressing $\underline{\Gamma}$ in terms of (h,τ,W) Moffat's theory becomes Einstein theory.

The situation here is similar as in classical mechanics, where the variational formula based on the Lagrangian $L = -q\dot{p} - H(p,q)$ can be used with p and q independent. Instead, it is better to add the complete time derivative $(d/dt)(pq)$ and express p in terms of q, using the part of equations of motion. This way we prove that p are not independent degrees of freedom of the theory.

[1] A. Jakubiec and J. Kijowski, Phys. Rev. D **37**, 1406 (1988); J. Math. Phys. **30**, 1073 (1989).
[2] J. W. Moffat, J. Math. Phys. **21**, 1798 (1980).
[3] P. Teyssandier and Ph. Tourrenc, J. Math. Phys. **24**, 2793 (1983).
[4] E. Schrödinger, Proc. R. Irish Acad. Ser. A **51**, 163 (1947); M. Ferraris and J. Kijowski, Gen. Relativ. Gravit. **14**, 37 (1982).
[5] M. Ferraris and J. Kijowski, Gen. Relativ. Gravit. **14**, 165 (1982).,
[6] J. Kijowski, Gen. Relativ. Gravit. **9**, 857 (1978).

Symmetries of the self-dual Einstein equations. I. The infinite-dimensional symmetry group and its low-dimensional subgroups

C. P. Boyer
Department of Mathematics and Statistics, University of New Mexico, Albuquerque, New Mexico 87131

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

(Received 26 April 1988; accepted for publication 19 October 1988)

This is the first of two papers in which the authors give a complete classification of symmetry reduced solutions of Plebanski's potential equation for self-dual Einstein spaces. In this first part the infinite pseudogroup of symmetries of Plebanski's equation is described, and the conjugacy classes of all local subgroups of dimensions one, two, and three over both the real and complex numbers are classified. Then in the second paper, this classification is used to obtain all symmetry-reduced solutions.

I. INTRODUCTION

The purpose of this investigation is twofold. On the one hand, we plan to apply, in a systematic manner, the method of symmetry reduction, to obtain group invariant solutions of the Euclidean signature self-dual Einstein equations. On the other hand, a first step in this program is to obtain the relevant local symmetry group of local point transformations and then give a complete classification of its low-dimensional local subgroups. Since this treatment is entirely local, we work infinitesimally with the corresponding Lie algebras.

In the present paper (Part I of a series of two) we show that the infinite-dimensional symmetry group is essentially one of Cartan's infinite-dimensional primitive groups[1] and provide a classification of its complex and real subalgebras of dimension less than four. It is precisely the real subalgebras that correspond to groups having orbits of codimension k, with $1 \leqslant k \leqslant 3$ in the underlying space of independent variables $\mathbb{R}^4$, and that hence provide a reduction of the considered equations to lower-dimensional ones.

Actually we do not deal with the self-dual Einstein equations[2,3] per se but with the potential equation obtained by Plebanski,[4]

$$\Omega_{x\bar{x}}\Omega_{y\bar{y}} - \Omega_{x\bar{y}}\Omega_{\bar{x}y} = 1, \qquad (*)$$

where $x, y \in \mathbb{C}$, Ω is a real function, and subindices indicate partial derivatives. This Monge–Ampere type of equation is of independent interest, having a rich history. Calabi[5–7] has studied the n-dimensional analog that he calls the Levi invariant equation. The function Ω appears as a local Kähler potential on some domain of $\mathbb{C}^2$, and gives rise to what has become known in the literature as Calabi–Yau spaces. These spaces are not only Kähler but hyperkähler,[7] that is, there is a two sphere's worth of complex structures on $\mathbb{R}^4$ that are compatible with the metric and the metric is Kähler with respect to all of these complex structures. In all that follows here we shall make a choice of complex structure and hence, a choice of Kähler potential.

It has been known for some time that Eq. $(*)$ has an infinite-dimensional symmetry "group" although it does not seem to have appeared in the literature. (The symmetry group of a related equation was determined in Ref. 8.) It is at least partially the purpose of the present paper to not only rectify this but also to show that the "symmetry group" of $(*)$ is an Abelian extension of one of the infinite primitive pseudogroups of Cartan, namely the pseudogroup of all biholomorphic maps from domains in $\mathbb{C}^2$ to itself with constant Jacobian determinant.

In Part II of this series we shall use the results obtained here to systematically apply the method of symmetry reduction[9,10] to Eq. $(*)$ to find all solutions that can be obtained by this method. From these solutions we can then write down the corresponding self-dual Einstein metrics that should be of interest from the point of view of gravitational instantons.[11–13] In the process we show how Eq. $(*)$ is related to many other interesting partial (and ordinary) differential equations.

The outline of the present paper is as follows: In this section we give a brief discussion of self-dual Einstein spaces and their relationship with Eq. $(*)$. In Sec. II we determine the Lie algebra of infinitesimal symmetries of Eq. $(*)$. Section III consists of relevant comments about the classification of subalgebras. Sections IV and V make up the heart of the present article. They give the classification of conjugacy classes of complex and real subalgebras, respectively, of dimension less than or equal to three under the relevant pseudogroup of transformations. Finally Sec. VI gives a brief conclusion of our results as well as a preview of things to come.

Let us recall briefly how the Ω equation $(*)$ arises from the self-dual Einstein equations. This was first derived in this context for the complex self-dual Einstein equations by Plebanski.[4] Let (M,g) be a four-dimensional Riemannian manifold. The Levi-Civita connection Γ has values in the Lie algebra so(4). But there is a well-known isomorphism $\mathrm{so}(4) \simeq \mathrm{su}(2)_+ \oplus \mathrm{su}(2)_-$, where we have labeled the two copies of su(2) by $\pm$ for convenience. Thus the connection splits $\Gamma = \Gamma_+ + \Gamma_-$ and we let $R_\pm = \mathrm{curv}\,\Gamma_\pm$. The Riemannian manifold (M,g) is called self-dual if $R_- = 0$. It is not difficult to see by writing things out explicitly that $R_- = 0$ or $R_+ = 0$ implies the vacuum Einstein equations, in fact, the Ricci tensor vanishes identically. Thus $R_- = 0$ is

 0022-2488/89/051081-14$02.50

called the self-dual Einstein equation. Now $R_- = 0$ means that Γ_- is a flat connection. Here $\Gamma_\pm$ are connections on the (locally defined) spin bundles $V_\pm M$. Although $V_\pm M$ may not be defined globally as bundles on M, their second-order symmetric tensor product bundles $S^2_\pm M$ are globally defined bundles on M. Furthermore, $\Gamma_\pm$ induce connections, also denoted $\Gamma_\pm$, on $S^2_\pm M$. There is another important splitting related to the Lie algebra splitting described above, namely, the splitting induced by Hodge's $*$ operator on two-forms. In four-dimensional Riemannian geometry $*$ is an involution on two-forms $\Lambda^2 M$ split into the plus and minus eigenspaces of $*$, viz., $\Lambda^2 M = \Lambda^2_+ M \oplus \Lambda^2_- M$. Furthermore, this splitting is compatible with the Lie algebra splitting and this gives rise to bundle isomorphisms $S^2_\pm M \simeq \Lambda^2_\pm M$. Thus $\Gamma_\pm$ can be considered as connections in $\Lambda^2_\pm M$, the bundles of self-dual (+) and anti-self-dual (−) two-forms.

We now describe the consequences of having a flat connection Γ_- on $\Lambda^2_- M \simeq S^2_- M$. From now on our considerations will be entirely local. We use the standard dotted and undotted spinor indices that are to be raised and lowered by the symplectic form ϵ_{AB} according to the convention $\psi_A = \epsilon_{AB}\psi^B$, where $A,B = 1,2$ and the Einstein summation convention is used on repeated indices. In spinor notation an orthonormal moving coframe is written as $\theta^{A\dot A}$. We define an almost complex structure on T^*M by requiring $\bar\theta^{A\dot A} = \theta_{A\dot A}$. Then a basis (locally) for $\Lambda^2_- M$ is given by

$$S^{\dot A\dot B} = \tfrac{1}{2}\epsilon_{AB}\theta^{A\dot A}\wedge\theta^{B\dot B}, \tag{1.1}$$

where $S^{\dot B\dot A} = S^{\dot A\dot B}$ and $\bar S^{\dot A\dot B} = S_{\dot A\dot B}$. Now as mentioned previously $R_- = 0$ if and only if Γ_- is flat, that is, there is a choice of orthonormal frame such that the connection coefficients $\Gamma_{\dot A\dot B}$ with respect to this frame vanish. It follows that the two-forms $S^{\dot A\dot B}$ are all closed, i.e.,

$$dS^{\dot A\dot B} = 0. \tag{1.2}$$

Considering first $dS^{\dot 1\dot 1} = 0$, by Darboux's theorem there are local coordinates $\{q^A; A = 1,2\}$ such that

$$S^{\dot 1\dot 1} = dq^1\wedge dq^2 = \tfrac{1}{2}dq^A\wedge dq_A. \tag{1.3}$$

Moreover, $S^{\dot 2\dot 2} = S_{\dot 1\dot 1} = \bar S^{\dot 1\dot 1} = d\bar q'\wedge d\bar q^2$. Now let us consider $S^{\dot 1\dot 2}$. First, $S^{\dot 1\dot 2} = S^{\dot 2\dot 1} = -S_{\dot 1\dot 2} = -S_{\dot 2\dot 1}$. Second, this form is nondegenerate since an explicit computation shows

$$2S^{\dot 1\dot 2}\wedge S^{\dot 1\dot 2} = S^{\dot 1\dot 1}\wedge S^{\dot 2\dot 2} = S^{\dot 1\dot 1}\wedge\bar S^{\dot 1\dot 1} \tag{1.4}$$

and the latter is proportional to the volume element on M. Since $S^{\dot 1\dot 2}$ is closed and nondegenerate g is a Kähler metric on M and $S^{\dot 1\dot 2}$ is the Kähler form. So locally there is a smooth function Ω such that

$$S^{\dot 1\dot 2} = \Omega_{q^A\bar q^B}\, dq^A\wedge d\bar q^B, \tag{1.5}$$

where $\Omega_{q^A\bar q^B} = \partial\Omega/\partial q^A\,\partial\bar q^B$. Now since the Levi-Civita connection $\Gamma = \Gamma_+$ has values in su(2), the local holomony group must be SU(2) or a subgroup thereof. It follows that under a change of SU(2) frame, the volume element $S^{\dot 1\dot 1}\wedge S^{\dot 2\dot 2}$ is preserved, thus about every point of M there are coordinates such that (1.3)–(1.5) hold. Plugging (1.5) into (1.4) and using (1.3) gives the Ω equation $(*)$.

II. THE SYMMETRY GROUP OF THE Ω EQUATION

In order to apply the method of symmetry reduction to $(*)$ we need to compute its symmetry group. We shall show that this group is an Abelian extension of one of the six infinite primitive pseudogroups of Cartan (Ref. 1, Theorem IX). More precisely, we compute the Lie algebra of infinitesimal symmetries; but rather than work with the differential equation itself, we write it as an equivalent Pfaffian system. First, notice that $(*)$ can be written as a conservation law, namely,

$$\partial_{\bar q_B}(\partial_{q_A}\Omega_{q^A\bar q^B} - \bar q_B) = 0 \tag{2.1a}$$

or equivalently, the complex conjugate equation

$$\partial_{q_A}(\Omega_{\bar q_B}\Omega_{q^A\bar q^B} - q_A) = 0. \tag{2.1b}$$

Equation (2.1a) is the integrability condition of the existence of a local complex-valued smooth function Σ satisfying

$$\Sigma_{\bar q^B} = \Omega_{q_A}\Omega_{q^A\bar q^B} - \bar q_B. \tag{2.2}$$

Similarly, we get the complex conjugate equation arising from (2.1b). In order to write the corresponding Pfaffian system we first construct the contact form θ_0 for Ω and then add the one-forms describing (2.2) and its complex conjugate. We arrive at

$$\begin{aligned}\theta_0 &= d\Omega - p_A\, dq^A - \bar p_A\, d\bar q^A,\\ \theta_1 &= d\Sigma - \tfrac{1}{2}p_A\, dp^A - s_A\, dq^A + \tfrac{1}{2}\bar q_A\, d\bar q^A,\\ \bar\theta_1 &= d\bar\Sigma - \tfrac{1}{2}\bar p_A\, d\bar p^A - \bar s_A\, d\bar q^A + \tfrac{1}{2}q_A\, dq^A,\end{aligned} \tag{2.3}$$

which is a Pfaffian system $\mathscr{I}$ on the two-jet $J^2(\mathbb{C}^2,\mathbb{R}\times\mathbb{C})$ (actually on a certain submanifold $\mathbb{R}^{15}$ of the two-jet) where q^A, $A = 1,2$ are complex coordinates on $\mathbb{C}^2$ and $(\Omega,\Sigma)\in\mathbb{R}\times\mathbb{C}$. The integral submanifolds $N\hookrightarrow J^2(\mathbb{C},\mathbb{R}\times\mathbb{C})$ which annihilate $\mathscr{I}$ and satisfy the independence condition $dq^1_A\wedge dq^2_A\wedge d\bar q^1_A\wedge d\bar q^2 \neq 0$ are precisely the solutions of $(*)$. Thus any infinitesimal contact transformation symmetry of $(*)$ must map $\mathscr{I}$ to $\mathscr{I}$. So we consider the vector fields on $J^2(\mathbb{C}^2,\mathbb{R}\times\mathbb{C})$ which satisfy

$$L_X\theta_i = \sum\lambda_{ij}\theta_j,\quad i,j = 0,\pm 1, \tag{2.4}$$

where $\theta_{-1} = \bar\theta_1$ and λ^i_j are smooth functions. Of course, if $L_X\theta_1\in\mathscr{I}$ then $L_X\bar\theta_1\in\mathscr{I}$. The method for solving (2.4) is by now quite standard so we omit the details, just presenting the infinitesimal symmetries and vector fields on $\mathbb{R}^{15}$,

$$\begin{aligned}X^{q^A} &= \tfrac{1}{2}aq_A + \alpha_{q^A},\\ X^{p^A} &= \tfrac{1}{2}\bar a p_A + \alpha_{q^A q^B}p^B + \beta q^A,\\ X^{\Omega} &= \tfrac{1}{2}(a+\bar a)\Omega + \beta + \bar\beta,\\ X^{s^A} &= (-\tfrac{1}{2} + \bar a)s_A + \alpha_{q^A q^B}s^B\\ &\quad + \alpha_{q^A q^B q^C}p^B p^C + \beta_{q^A q^B}p^B + \gamma_{q^A},\\ X^{\Sigma} &= \bar a\Sigma + \tfrac{1}{2}\alpha_{q^A q^B}p^A p^B + \tfrac{1}{2}\beta_{q^A}p^A + \tfrac{1}{2}\bar\alpha_{\bar q^A}\bar q^A + \bar\alpha + \gamma,\end{aligned} \tag{2.5}$$

and their complex conjugates. Here, α, β, and γ are holomorphic functions on open sets of the origin in $\mathbb{C}^2$.

Let us now discuss the structure of the symmetry algebra $\tilde T$ of the Pfaffian system (2.3). There are three arbitrary holomorphic functions of the complex variables $\{q^1,q^2\}$, namely, α, β, and γ, and one complex parameter a. Furthermore, the vector fields

$$\widetilde{X}_1 = [\beta(a) + \bar{\beta}(\bar{q})]\partial_\Omega,$$
$$\widetilde{X}_2 = \mathrm{Re}[\gamma(q)\partial_\Sigma], \qquad (2.6)$$
$$\widetilde{X}_3 = \mathrm{Im}[\gamma(q)\partial_\Sigma],$$

generate an infinite-dimensional Abelian ideal $\widetilde{A}$ in $\widetilde{T}$. Moreover, $\widetilde{X}_2$ and $\widetilde{X}_3$ generate q-dependent translations in Σ and $\bar{\Sigma}$, the first integrals of Ω. Therefore, they do not appear as infinitesimal symmetries of the Ω equation $(*)$. Thus the full infinitesimal symmetry algebra $\widehat{T}$ of the Ω equation is generated by two holomorphic functions α and β and one complex parameter a. In order to discuss further this symmetry algebra we introduce an affine bundle (not a vector bundle) $A \to \mathbb{C}^2$ where $\{q^A: A = 1,2\}$ are complex coordinates on $\mathbb{C}^2$ and a local section of A is given by the graph $(q^1,q^2) \mapsto (q^1,q^2,\Omega(q,\bar{q}))$, so the fibers of A are real affine lines, i.e., there is no fixed origin in the fibers. In fact, the transformation generated by $\widetilde{X}$ sends the section Ω to $\Omega + \beta(q) + \bar{\beta}(\bar{q})$, which is a base point dependent translation in the fibers of A. With this in mind the representation of $\widehat{T}$ given by the vector fields (1.2b) is the prolongation to the one-jet $J^1(A,\mathbb{R})$ of fiber preserving local transformations of $A \to \mathbb{C}^2$. However, we shall see later that from the point of view of symmetry reduction the Abelian ideal A generated by $\widetilde{X}_1$ plays no role whatsoever. It is thus only the factor algebra $T = \widehat{T}/A$ which is of interest to us, and T is easily identified with one of the transitive primitive algebras of Carãn,[1] namely the Lie algebra of holomorphic vector fields on $\mathbb{C}^2$ with constant divergence. It should also be mentioned that the transformations of T fix the zero section of A. Thus $A \simeq \mathbb{C}^2 \times \mathbb{R}$ can now be viewed as a trivial vector bundle over $\mathbb{C}^2$, i.e., as a product.

Let us summarize our results as the following theorem.

Theorem 2.1: The infinitesimal symmetry algebra $\widehat{T}$ of infinitesimal contact transformations of the Ω equation $(*)$ is the prolongation to $J^2(A,\mathbb{R})$ of infinitesimal transformations $\psi: A \to A$ preserving the affine structure of A. Furthermore, $\widehat{T}$ is isomorphic to an Abelian extension of the Lie algebra T of holomorphic vector fields on $\mathbb{C}^2$ with constant divergence by the infinite-dimensional Abelian ideal generated by $\widetilde{X}_1$. Moreover, the infinitesimal transformations of T preserve the product structure $A \simeq \mathbb{C}^2 \times \mathbb{R}$.

As mentioned previously, it is the factor algebra T that is of interest for the purposes of symmetry reduction, and an important consequence of Theorem 2.1 is that as a complex Lie algebra, T consists of holomorphic vector fields. We shall be interested mainly in real subalgebras L of T and we shall associate to such subalgebras a complex invariant, namely, the complex divergence of the complexification X^C of vector fields X in L. To this end we consider the simple subalgebra $T_0 \subset T$ consisting of all holomorphic vector fields on $\mathbb{C}^2$ with zero divergence. Then we have the following lemma.

Lemma 2.2: Let L be a real subalgebra of T and L^1 its derived algebra, then $L^1 \subset T_0$.

Proof: Let E denote the one-dimensional complex Lie algebra generated by the Euler vector field on $\mathbb{C}^2$. Then T_0 is an ideal in T and we have an exact sequence of Lie algebras

$$0 \to T_0 \to T \to E \to 0.$$

It follows that $T^1 \subset T_0$. Moreover, as a real Lie algebra E is a two-dimensional Abelian Lie algebra, so for any real subalgebra $L \subset T$, $L^1 \subset T^1$ and this proves the lemma.

Now any infinite-dimensional Lie algebra T of vector fields on $\mathbb{C}^n$ generates a pseudogroup of transformations on $\mathbb{C}^2$ as follows[14]: Let $\{X_\alpha\}$ by any collection of (locally defined) vector fields in T. We can integrate these vector fields locally to get a collection of local one-parameter groups $\{\phi_{t_\alpha}\}$. The set of all such local diffeomorphisms generates a pseudogroup $P(T)$ (the pseudogroup generated by T). In our case the Lie algebra T generates the pseudogroup of local biholomorphic diffeomorphisms of $\mathbb{C}^2$ with constant Jacobian determinant P, whereas T_0 generates the subpseudogroup with unit Jacobian determinant P_0. Now given any subalgebra $L \subset T$ we will make use of the subpseudogroup $N(L) \subset P$ that normalizes the subalgebra L.

Finally to end this section we mention the connection between the infinitesimal symmetries of $(*)$ and Killing vector fields on a given solution of $(*)$. In Ref. 15 the homothetic Killing equations were integrated and a standard form for any homothetic Killing vector field was given. It is easy to see that any $X \in T$ is a homothetic Killing vector field. However, the pseudogroup S of allowed transformations for the homothetic Killing equations is larger. In general, S does not preserve the complex structure, but $P \subset S$. The action of S on the coordinates is induced by the action of SO(3) on $\Lambda^2\_ M \simeq S^2 V\_M$, which leaves (2.2) invariant. The unit sphere bundle in $\Lambda^2\_ M$ can be identified with the bundle of complex structures on M. So any homothetic Killing vector field will fix a complex structure and thus is equivalent to a vector field in T. However, there is an action of SO(3) which permutes all complex structures and so this is not equivalent to any subgroup of P. This gives a class of Bianchi IX solutions[16] of $(*)$ that cannot be obtained by symmetry reduction of $(*)$.

III. GENERAL COMMENTS ON THE CLASSIFICATION OF SUBALGEBRAS

In order to perform a symmetry reduction for the Ω equation we need to know all low-dimensional subgroups of the symmetry pseudogroup P of this equation. More precisely, we need a classification of all local subgroups that will have generic orbits of dimension $d = 1, 2, or\ 3$ in the underlying four-dimensional Euclidean space-time. Our procedure will be an algebraic one: we shall classify subalgebras of the symmetry algebra, namely the algebra of holomorphic vector fields in two complex variables, having constant divergence. The classification will be under the action of the pseudogroup P of holomorphic transformations with constant Jacobian determinant.

Lie in his classical lecture notes[16] (see also Cartan[16]) has solved a related problem, namely that of classifying all continuous groups of point transformations in two complex variables. He obtained an exhaustive list of representatives of Lie algebras that can be realized in terms of holomorphic vector fields in two complex variables. The vector fields do not necessarily have constant divergence. Furthermore, Lie's classifying group is correspondingly larger: the transformations do not necessarily have constant Jacobian determinant. Lie's results cannot be directly adapted to our case, mainly because the value of the divergence of a vector field is coordinate dependent and is thus not invariant under arbitrary transformations.

We shall be interested in low-dimensional Lie algebras, realized in terms of either complex or real vector fields. However, before discussing the general procedure we shall make some simplifications. First, we show that for symmetry reductions it suffices to consider the factor algebra $T \simeq \hat{T}/A$. Moreover, according to Theorem 2.1, we may restrict ourselves to vector fields that represent infinitesimal symmetry transformations of the affine bundle A. Such complex vector fields may be written as

$$X = (aq^A + \alpha_{q_A})\partial_{q^A} + a\Omega\partial_\Omega + \beta\partial_\Omega. \tag{3.1}$$

The corresponding real vector fields are obtained by simply taking the real part of (3.1). If a and α_{q^A} do not both vanish, we can make the q-dependent translation $\Omega \mapsto \Omega + \gamma$ and remove β by choosing γ to be a solution of the first-order inhomogeneous partial differential equation

$$(aq^A + \alpha_{q_A})\gamma_{q^A} - a\gamma + \beta = 0. \tag{3.2}$$

On the other hand, if both a and α_{q^A} vanish, the invariant is an arbitrary function $F(q^1,q^2,\bar{q}^1,\bar{q}^2)$ which is independent of Ω and thus does not give rise to symmetry reduced solutions of the Ω equation (*). Thus we may restrict our considerations to the Lie algebra T. Furthermore, notice that the projection π: $A \to \mathbb{R}^4$ induces a Lie algebra $\pi_* T$ that is isomorphic to T. So in order to simplify notations in Secs. IV and V and especially in the tables, we will drop the terms involving ∂_Ω in the vector fields. The notation X and X^R below will be used for the projections of the vector fields onto the base manifold, i.e., elements of $\pi_* T$. The full vector fields are always recovered by adding $\frac{1}{2}(\operatorname{div} X)\Omega\partial_\Omega$, or $\frac{1}{2}(\operatorname{div} X + \operatorname{div} \bar{X})\Omega\partial_\Omega$, to the corresponding complex or real vector fields.

The complex vector fields under consideration have the form

$$X = f(x,y)\partial_x + g(x,y)\partial_y \tag{3.3a}$$

satisfying

$$\operatorname{div} X \equiv f_x + g_y = \alpha = \text{const}. \tag{3.3b}$$

Real vector fields that we are dealing with have the form

$$X^R = f(x,y)\partial_x + g(x,y)\partial_y + \bar{f}(\bar{x},\bar{y})\partial_{\bar{x}} + \bar{g}(\bar{x},\bar{y})\partial_{\bar{y}}. \tag{3.4a}$$

We have

$$\operatorname{div} X^R = f_x + g_y + \bar{f}_{\bar{x}} + \bar{g}_{\bar{y}} = \alpha + \bar{\alpha}. \tag{3.4b}$$

Thus we may have $\operatorname{div} X^R = 0$, but $\operatorname{div} X = \alpha = -\bar{\alpha} \neq 0$.

Vector fields and algebras of vector fields will be classified under coordinate transformations of the pseudogroup P which we write explicitly as

$$\xi = F(x,y), \quad \eta = G(x,y), \tag{3.5a}$$

with constant Jacobian determinant,

$$\det J \equiv J_0 = F_x G_y - F_y G_x. \tag{3.5b}$$

Let us first present some general results that will be used below in the subalgebra classification.

Lemma 3.1: Let X be a vector field of the type (3.1) in coordinates (x,y) and $\tilde{X}$ the same vector field in coordinates (ξ,η) of (3.5a). The divergences of X and $\tilde{X}$ are related by

$$\operatorname{div} \tilde{X} = \operatorname{div} X + X \ln(\det J). \tag{3.6}$$

In particular, if $\det J = J_0 = \text{const}$, then $\operatorname{div} X$ is invariant under (3.5).

Proof: The transformed vector field is

$$\tilde{X} = (fF_x + gF_y)\partial_\xi + (fG_x + gG_y)\partial_\eta. \tag{3.7}$$

A simple calculation yields (3.6). Q.E.D.

Lemma 3.2: If A, B, and C are three vector fields, satisfying $A = [B,C]$ and $\operatorname{div} B = \beta$, $\operatorname{div} C = \gamma$, where β and γ are constants, then $\operatorname{div} A = 0$.

Proof: A simple calculation yields

$$\operatorname{div} A = B \operatorname{div} C - C \operatorname{div} B. \tag{3.8}$$

Hence

$$\operatorname{div} A = B\gamma - C\beta = 0. \qquad \text{Q.E.D.}$$

Remark: Lemma 2.2 is a simple consequence of Lemma 3.2. We shall make use of the known classification of two- and three-dimensional Lie algebras into isomorphism classes.[18,19] They can be summed up in two lemmas.

Lemma 3.3: Any two-dimensional Lie algebra over either $\mathbb{C}$ or $\mathbb{R}$ is isomorphic to one of the following ones: (1) Abelian: $2A_1$,

$$[X_1,X_2] = 0. \tag{3.9a}$$

(2) Solvable, non-Abelian: $A_{2,1}$,

$$[X_1,X_2] = X_1. \tag{3.9b}$$

Lemma 3.4: Any three-dimensional Lie algebra over either $\mathbb{C}$ or $\mathbb{R}$ is isomorphic to one of the following ones: (1) Abelian: $3A_1$,

$$[X_1,X_2] = [X_2,X_3] = [X_3,X_1] = 0. \tag{3.10a}$$

(2) Decomposable, non-Abelian: $A_{2,1} \oplus A_1$,

$$[X_1,X_3] = X_1, \quad [X_2,X_3] = 0, \quad [X_1,X_2] = 0. \tag{3.10b}$$

(3) Indecomposable, nilpotent: $A_{3,1}$,

$$[X_2,X_3] = X_1, \quad [X_1,X_3] = 0, \quad [X_1,X_2] = 0. \tag{3.10c}$$

(4) Indecomposable, solvable, non-nilpotent, with a two-dimensional Abelian ideal $\{X_1,X_2\}$,

$$\begin{pmatrix} [X_1,X_3] \\ [X_2,X_3] \end{pmatrix} = M \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}, \quad [X_1,X_2] = 0, \quad M = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}. \tag{3.10d}$$

Over $\mathbb{C}$ the matrix M has one of the following forms:

$$A^{\alpha}_{3,2}: \; M^C_1 = \begin{pmatrix} 1 & 0 \\ 0 & \alpha \end{pmatrix},$$

$$0 \leqslant |\alpha| \leqslant 1, \quad \alpha \in \mathbb{C}, \quad \text{if } |\alpha| = 1,$$

$$\text{then } 0 \leqslant \arg \alpha \leqslant \pi, \tag{3.10e}$$

$$A_{3,3}: \; M^C_2 = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}.$$

Over $\mathbb{R}$ the matrix has one of the following forms:

$$A^a_{3,2}:\ M_1^R=\begin{pmatrix}1 & 0\\ 0 & a\end{pmatrix},\quad a\in\mathbb{R},\quad 0<|a|\leqslant 1,$$

$$A_{3,3}:\ M_2^R=\begin{pmatrix}1 & 0\\ 1 & 1\end{pmatrix},\tag{3.10f}$$

$$A^a_{3,4}:\ M_3^R=\begin{pmatrix}a & -1\\ 1 & a\end{pmatrix},\quad a\in\mathbb{R},\quad 0\leqslant a.$$

(5) Simple.

Over $\mathbb{C}$ the only possibility is $\mathrm{sl}(2,\mathbb{C})$,

$$A_{3,5}:\ [X_1,X_2]=X_1,\quad [X_2,X_3]=X_3,\quad [X_1,X_3]=-2X_2.\tag{3.10g}$$

Over $\mathbb{R}$ there are two cases: (a) $\mathrm{sl}(2,\mathbb{R})\sim\mathrm{su}(1,1)\sim\mathrm{o}(2,1)$ with commutation relations (3.10g), and (b) $\mathrm{su}(2)\sim\mathrm{o}(3)$ with commutation relations

$$A_{3,6}:\ [X_i,X_k]=\epsilon_{ikl}X_l,\quad \{i,k,l\}=\{1,2,3\}.\tag{3.10h}$$

Two useful lemmas that can be proven by direct calculations are the following.

Lemma 3.5: The most general transformation of the type (3.5) leaving the vector space $\{X_1=\partial_x\}$ invariant is

$$\xi=\lambda x+H(y),\quad \eta=\beta y+\nu,\quad \lambda\mu\neq 0,\tag{3.11}$$

where $\lambda,\mu,\nu\in\mathbb{C}$ are constants.

Lemma 3.6: The most general transformation of the type (3.5) leaving the vector space $\{X_1=x\partial_x\}$ invariant is

$$\xi=\lambda x/\dot G(y),\quad \eta=G(y),\quad \lambda\neq 0,\quad \dot G(y)\neq 0.\tag{3.12}$$

We restrict our classification to subalgebras of dimension d with $1\leqslant d\leqslant 3$. The justification for this is that the $d=3$ subalgebras already provide reductions of the Ω equation to ordinary differential equations, or to algebraic ones. Higher-dimensional algebras will always contain at least one three-dimensional subalgebra, so the algebras of dimension $d\geqslant 4$ will not provide any new reductions. Indeed, every simple Lie algebra except $\mathrm{su}(1,1)$ contains at least one class of $\mathrm{su}(2)$ subalgebras. Every solvable Lie algebra of dimension n has subalgebras (and ideals) of all dimensions $1\leqslant d\leqslant n$. Algebras that are neither simple nor solvable have a Levi decomposition in which the semisimple part is either $\mathrm{su}(2)$ or $\mathrm{su}(1,1)$, or contains at least one of them as a subalgebra.

The problem of classifying subalgebras of the Lie algebra of holomorphic vector fields with constant divergence into conjugacy classes and the action of the pseudogroup P is conceptually similar to that of classifying the subalgebras of a finite-dimensional Lie algebra under the action of the group of inner automorphisms. The classification methods have been developed and applied in a series of earlier papers.[20–22]

IV. CLASSIFICATION OF LOW-DIMENSIONAL COMPLEX SUBALGEBRAS

A. One-dimensional subalgebras

We are given a vector field (3.1) satisfying (3.2) and perform a change of variables (3.5). The vector field transforms into (3.7). We set

$$fG_x+gG_y=0,\tag{4.1a}$$

$$fF_x+gF_y=-(g/G_x)J_0.\tag{4.1b}$$

Relation (4.1a) amounts to a choice of the function $G(x,y)$ that can always be made; (4.1b) is then a consequence of (4.1a) and (3.5b). The vector field X in the coordinates (ξ,η) is now

$$X=-(g/G_x)J_0\partial_\xi.\tag{4.2}$$

By assumption, J_0 is constant; if we require that $-g/G_x$ be constant then (4.1a) implies the compatibility condition $f_x+g_y=0$. If this is satisfied, i.e., if we have $\operatorname{div}X=0$, then X is conjugate to $\widetilde X=\partial_\xi$. If on the other hand, we have $\operatorname{div}X=\alpha\neq 0$, we set $-g(J_0/G_x)=\alpha F$, and obtain equations for G and F that are always compatible.

We arrive at the following theorem.

Theorem 4.1: An arbitrary one-dimensional subalgebra $\{X\}$ of the symmetry algebra is conjugate to one of the two following Lie algebras:

$$L_{1,1}(\mathbb{C})=\{\partial_x\},\tag{4.3a}$$

$$L_{1,2}(\mathbb{C})=\{x\partial_x\}.\tag{4.3b}$$

They are distinguished by the fact that $\operatorname{div}X=0$ for $X\in L^{\mathbb{C}}_{1,1}$ and $\operatorname{div}X\neq 0$ for $X\in L^{\mathbb{C}}_{1,2}$. □

B. Two-dimensional subalgebras

Consider a Lie algebra $\{X_1,X_2\}$ with a commutation relation as in (3.9). We shall assume that X_1 has already been transformed to standard form as in (4.3); X_2, on the other hand, is left general, as in (3.1). The procedure is first to implement the commutation relation, then simplify X_2, using the normalizer subpseudogroup, N or $P\{X_1\}$ of X_1 in P.

1. Abelian algebras

(a) Take X_1 as in (4.3a) and require $[X_1,X_2]=0$, $\operatorname{div}X_2=\alpha=\text{const}$. We obtain

$$X_1=\partial_x,\quad X_2=f(y)\partial_x+(\alpha y+\beta)\partial_y.\tag{4.4a}$$

The normalizer pseudogroup of X_1 is given in (3.11) and it transforms the vector fields into

$$X_1=\lambda\partial_\xi,$$

$$X_2=[\lambda f(y)+(\alpha y+\beta)\dot H(y)]\partial_\xi+(\alpha y+\beta)\mu\partial_\eta.\tag{4.4b}$$

If $(\alpha,\beta)\neq(0,0)$ we choose $H(y)$ to satisfy $(\alpha y+\beta)\dot H+\lambda f(y)=0$. If $\alpha\neq 0$ we put $\alpha\nu=\beta\mu$. If $\alpha=0$, $\beta\neq 0$ we put $\beta\mu=1$. Finally, if $(\alpha,\beta)=(0,0)$ no simplification occurs.

(b) Take X_1 as in (4.3b). Imposing $[X_1,X_2]=0$ and replacing X_2 by $X_2-\alpha X_1$ we obtain

$$X_1=x\partial_x,\quad X_2=-\dot g(y)x\partial_x+g(y)\partial_y.\tag{4.5}$$

Performing a transformation (3.12) with $G(y)=\alpha/g(y)$ we reduce the algebra (4.5) to $X_1=\xi\partial_\xi$, $X_2=\partial_\eta$.

2. Non-Abelian algebras

The commutation relation is given by (3.9b). According to Lemma 3.2 we have $\operatorname{div}X_1=0$. With no loss of generality we can take $X_1=\partial_x$. From (3.9b) and (3.2) we obtain

$$X_1 = \partial_x,\quad X_2 = [x + h(y)]\partial_x + (\alpha y + \beta)\partial_y. \quad (4.6)$$

Performing a transformation of the form (3.11) with $H(y)$ satisfying $h(y)\lambda + (\alpha y + \beta)\dot{H} = H$ we reduce (4.6) to

$$X_1 = \partial_\xi,\quad X_2 = \xi\partial_\xi + (\alpha\eta - \alpha\nu + \beta\mu)\partial_\eta. \quad (4.7)$$

If $\alpha \neq 0$ we choose ν and μ so that $\beta\mu - \alpha\nu = 0$. If $\alpha = 0$, $\beta \neq 0$ we choose μ so that $\beta\mu = 1$.

The results of this subsection are summarized in Table I. We have denoted V the vector space spanned by the vectors $\{X_1,X_2\}$ at any generic point (x, y). The three Abelian algebras $L_{2,1}(\mathbb{C})$, $L_{2,2}(\mathbb{C})$, and $L^f_{2,3}(\mathbb{C})$ are distinguished from each other by the invariants in columns 5 and 6, namely $\operatorname{div} X$, the divergence of the general element $X = \mu_1 X_1 + \mu_2 X_2$, and the dimension of V. For the $A_{2,1}$ type algebras these invariants coincide for $L_{2,4}(\mathbb{C})$ and $L^\alpha_{2,5}(\mathbb{C})$ if $\alpha \neq 0$, $\alpha \neq -1$. In this case X_1 is uniquely defined as the vector field spanning the derived algebra. The value of α itself is an invariant under the action of the isotropy group of X_1, i.e., the transformation (3.11).

The algebras $L^f_{2,3}(\mathbb{C})$ are somewhat exceptional. They depend on one arbitrary function $f(y)$.

In order to avoid redundancy in Table I and in other subalgebra lists we establish the following equivalence relation. The two sets of linearly independent functions

$$\{f_1(y),\ldots,f_n(y)\} \quad\text{and}\quad \{g_1(y),\ldots,g_n(y)\},$$

are equivalent, if constants $\lambda \neq 0$, μ, and a matrix $\rho \in \mathrm{GL}(n,\mathbb{C})$ exist, such that

$$g_i(y) = \sum_{k=1}^{n} \rho_{ik} f_k(\lambda y + \mu). \quad (4.8)$$

Using this equivalence concept we arrive at the following theorem.

Theorem 4.2: Every complex two-dimensional subalgebra of the algebra g of holomorphic vector fields with constant divergence is conjugate under the pseudogroup P to an algebra in Table I. Two algebras in Table I are mutually conjugate if and only if they are in the classes $L^f_{2,3}(\mathbb{C})$ and $L^g_{2,3}(\mathbb{C})$, and the pairs $(1, f(y))$ and $(1, g(y))$ are equivalent under relation (4.8).

C. Three-dimensional subalgebras

It follows from Lemma 3.4 that all complex three-dimensional Lie algebras except $\mathrm{sl}(2,\mathbb{C})$ have a two-dimensional Abelian ideal. We choose it to be $\{X_1,X_2\}$. For solvable non-nilpotent Lie algebras this ideal is unique. Our procedure will be to assume that $\{X_1,X_2\}$ is in one of the standard forms $L_{2,1}(\mathbb{C})$, $L_{2,2}(\mathbb{C})$, or $L^f_{2,3}(\mathbb{C})$, whereas X_3 has the general form (3.1) satisfying (3.2). We then impose the commutation relations

$$\begin{pmatrix}[X_1,X_3]\\ [X_2,X_3]\end{pmatrix} = \begin{pmatrix}a & b\\ c & d\end{pmatrix}\begin{pmatrix}X_1\\ X_2\end{pmatrix},\quad M = \begin{pmatrix}a & b\\ c & d\end{pmatrix}, \quad (4.9)$$

for each standard form of M. Finally, we standardize M, using the normalizer pseudogroup of the ideal $\{X_1,X_2\}$. For Abelian and nilpotent algebras we must afterwards weed out possible redundancies, due to the uniqueness of the Abelian ideal.

Simple subalgebras, i.e., $\mathrm{sl}(2,\mathbb{C})$ subalgebras, will be treated separately. In this case use will be made of the fact that $\mathrm{sl}(2,\mathbb{C})$ contains a subalgebra of the type $A_{2,1}$, that is, however, not an ideal.

1. Solvable subalgebras

(a) Ideal $L_{2,1}(\mathbb{C}) = \{\partial_x,\partial_y\}$. We have

$$\begin{pmatrix}[X_1,X_3]\\ [X_2,X_3]\end{pmatrix} = \begin{pmatrix}f_x\partial_x + g_x\partial_y\\ f_y\partial_x + g_y\partial_y\end{pmatrix}. \quad (4.10)$$

Combining (4.9) and (4.10), we obtain

$$f = ax + cy + p,\quad g = bx + dy + q. \quad (4.11)$$

Replacing X_3 by $X_3 - pX_1 - qX_2$ we effectively set $p = q = 0$. Performing linear transformation of variables ($\xi = \mu x + \nu y, \eta = \rho x + \sigma y, \mu\sigma - \rho\nu \neq 0$) we can transform the matrix M to its standard form.

The Abelian case (3.10a) is excluded, since

$$\begin{pmatrix}a & b\\ c & d\end{pmatrix} = \begin{pmatrix}0 & 0\\ 0 & 0\end{pmatrix}$$

implies $X_3 = 0$. The decomposable case occurs if $a = 1$, $b = c = d = 0$ in M. This yields the decomposable algebra $\{\partial_x,\partial_y,x\partial_x\}$. The nilpotent case (3.10c) occurs for $c = 1$, $a = b = d = 0$ and leads to $\{\partial_x,\partial_y, y\partial_x\}$. The diagonalizable case (3.10e) corresponds to $a = 1$, $d = \alpha (0 \leqslant \alpha \leqslant 1)$ and yields $\{\partial_x,\partial_y,x\partial_x + \alpha y\partial_y\}$. Finally, the Jordan case (3.10e) corresponds to $a = d = 1$, $c = 1$, $b = 0$ and the algebra $\{\partial_x,\partial_y, (x + y)\partial_x + y\partial_y\}$.

(b) Ideal $L_{2,2} = \{\partial_x, y\partial_y\}$. According to Lemma 3.2, $X_2 = y\partial_y$ cannot figure in the derived algebra. Hence $b = d = 0$ in (4.9). Moreover, we have $f_x = a$, $yf_y = c$, $g_x = 0$, $yg_y - g = 0$. Taking an appropriate linear combination of X_3 with X_1, and X_2, we obtain $X_3 = (ax + c\ln y)\partial_x$. The transformation in P that leaves $L_{2,2}$ invariant is

$$\xi = \alpha x + \gamma\ln y + \mu,\quad \eta = \nu y,\quad \alpha\nu \neq 0. \quad (4.12)$$

The vector fields transform into

$$X_1 = \alpha\partial_\xi,\quad X_2 = \gamma\partial_\xi + \eta\partial_\eta,$$
$$X_3 = [a\xi + (-a\gamma + \alpha c)\ln\eta + (a\gamma - \alpha c)\ln\nu - a\mu]\partial_\xi. \quad (4.13)$$

If $a \neq 0$ we choose $\gamma = \alpha c/a$, $\mu = 0$ and obtain a decomposable algebra $\{\partial_x, y\partial_y, x\partial_x\}$. If $a = 0$ we obtain a nilpotent Lie algebra $\{\partial_x, y\partial_y, \ln y\partial_x\}$.

TABLE I. Two-dimensional complex subalgebras.

N_0	Type	Basis X_1	Basis X_2	$\operatorname{div} X$, $X = \mu_1X_1 + \mu_2X_2$	$\dim V$
$L_{2,1}(\mathbb{C})$	$2A_1$	∂_x	∂_y	0	2
$L_{2,2}(\mathbb{C})$	$2A_1$	∂_x	$y\partial_y$	μ_2	2
$L^f_{2,3}(\mathbb{C})$	$2A_1$	∂_x	$f(y)\partial_x$, $f(y) \neq 0$	0	1
$L_{2,4}(\mathbb{C})$	$A_{2,1}$	∂_x	$x\partial_x + \partial_y$	μ_2	2
$L^\alpha_{2,5}(\mathbb{C})$	$A_{2,1}$	∂_x	$x\partial_x + \alpha y\partial_y$	$\mu_2(1+\alpha)$, 0 if $\alpha = -1$	2 if $\alpha \neq 0$, 1 if $\alpha = 0$

(c) Ideal $L_{2,3}^{\phi}(\mathbb{C}) = \{\partial_x, \phi(y)\partial_x; \dot{\phi}(y) \neq 0\}$. The commutation relations in this case are

$$\begin{pmatrix}[X_1,X_3]\\ [X_2,X_3]\end{pmatrix} = \begin{pmatrix}f_x\partial_x + g_x\partial_y\\ (\phi f_x - g\dot{\phi})\partial_x + \phi g_x\partial_y\end{pmatrix} = \begin{pmatrix}[a+b\phi(y)]\partial_x\\ [c+d\phi(y)]\partial_x\end{pmatrix}. \tag{4.14}$$

Taking linear combinations of X_1 and X_2 and performing an appropriate change of variables we can simultaneously assume that $\{X_1,X_2\}$ remain in their standardized form $X_1 = \partial_x$, $X_2 = \phi(y)\partial_x$ and that the matrix M is in its standard form (as in Lemma 3.4).

From (4.14) we obtain

$$g = g(y), \quad f = [a + b\phi(y)]x + \psi(y), \tag{4.15}$$

$$g\dot{\phi} = b\phi^2 + (a-d)\phi - c. \tag{4.16}$$

Further, div $X_3 = \lambda$ implies

$$a + b\phi(y) + g(y) = \lambda. \tag{4.17}$$

The transformation (3.5) that leaves the ideal $\{X_1,X_2\}$ invariant is

$$\xi = x + H(y), \quad \eta = py + q, \quad p = 0. \tag{4.18}$$

It transforms the algebra to

$$X_1 = \partial_\xi, \quad X_2 = \phi\partial_\xi,$$
$$X_3 = [(a+b\phi)\xi + \psi + g\dot{H} - (a+b\phi)H]\partial_\xi + pg\partial_\eta. \tag{4.19}$$

Let us now run through all standard forms of M.

(i) Abelian algebras: $a = b = c = d = 0$. Since $\dot{\phi}(y) \neq 0$, (4.16) implies $g = 0$ and hence $\lambda = 0$. We obtain the Lie algebra $\{\partial_x, \phi(y)\partial_y, \psi(y)\partial_y\}$ where 1, ϕ, and ψ are linearly independent.

(ii) Decomposable non-Abelian algebras: We take $a = b = c = 0$, $d = 1$. Equation (4.17) implies $g = \lambda y + \nu$. If $\lambda = \nu = 0$ we reobtain the Abelian algebra considered above. Hence we have $g(y) \neq 0$ and we can put $\dot{H} = -\psi g^{-1}$. Two cases arise, namely, the Lie algebra $\{\partial_x\, y^p\partial_x, -(1/p)y\partial_y\}$ if $\lambda \neq 0$ (we have put $p = -\lambda^{-1}$) and $\{\partial_x, e^y\partial_x, -\partial_y\}$ if $\lambda = 0$ (we put $p = -\nu^{-1}$).

(iii) Nilpotent algebras: We have $a = b = d = 0$ and $c = 1$. Equation (4.17) implies $g = \lambda y + \mu$ and (4.16) tells us that $\lambda = \mu = 0$ is excluded. Integrating (4.16) for $\lambda \neq 0$ and for $\lambda = 0$, $\mu \neq 0$ we get two nilpotent algebras that are not new: they have appeared above in cases (a) and (b) respectively.

(iv) Algebras of type $A_{3,2}^{\alpha}$: We have $a = 1$, $b = c = 0$, $d \equiv \alpha$, $0 < |\alpha| \leqslant 1$. From (4.16) and (4.17) we find

$$g = (\lambda - 1)y + \mu, \quad [(\lambda-1)y + \mu]\dot{\phi} = (1-\alpha)\phi.$$

For $\lambda = 1$, $\mu = 0$, $\alpha = 1$ we obtain the algebra $\{\partial_x, \phi(y)\partial_x, x\partial_x\}$; $\lambda \neq 1$, $\alpha = 1$ is not allowed; $\lambda \neq 1$, $\alpha \neq 1$ yields $\{\partial_x, py\partial_x, x\partial_x + [(1-\alpha)/p]y\partial_y\}$ with $p \neq 0$. It is possible to transform p into $-p(p+1)^{-1}$, hence we restrict to $-2 \leqslant p \leqslant 0$. Finally $\lambda = 1$, $\mu \neq 0$ leads to $\{\partial_x, e^y\partial_x, x\partial_x + (1-\alpha)y\partial_y\}$.

(v) Algebras of type $A_{3,3}$: In this case $a = c = d = 1$, $b = 0$. Solving (4.16) we again have $g = (\lambda - 1)y + \mu$. For $\lambda \neq 1$ we obtain $\{\partial_x, [1/(1-a)]\ln y\partial_x, x\partial_x + (a-1)y\partial_y, a \neq 1\}$, for $\lambda = 1$, $\mu \neq 0$ we find another algebra, namely, $\{\partial_x, y\partial_x, x\partial_x - \partial_y\}$.

This completes the enumeration of all conjugacy classes of three-dimensional solvable Lie subalgebras over $\mathbb{C}$. To complete this section let us now construct the simple three-dimensional subalgebras of the considered algebra of holomorphic vector fields with constant divergence. According to Lemma 3.4 all such algebras over $\mathbb{C}$ must be isomorphic to sl(2,$\mathbb{C}$).

2. The sl(2,C) subalgebras

We choose a basis $\{X_1,X_2,X_3\}$ of sl(2,$\mathbb{C}$) such that the commutation relations are

$$[X_1,X_2] = X_1, \quad [X_2,X_3] = X_3, \quad [X_3,X_1] = 2X_2. \tag{4.20}$$

We can assume that the subalgebra $\{X_1,X_2\}$ is already in standard form, namely $L_{2,4}(\mathbb{C})$ or $L_{2,5}^{\alpha}(\mathbb{C})$ of Table I. Since sl(2,$\mathbb{C}$) is simple its derived algebra equals the original sl(2,$\mathbb{C}$) algebra. It follows from Lemma 3.2 that we must have div $X_i = 0$, $i = 1,2,3$. Hence the only allowed $A_{2,1}$ subalgebra as a candidate for $\{X_1,X_2\}$ is $L_{2,5}^{\alpha=-1}(\mathbb{C})$. We have

$$X_1 = \partial_x, \quad X_2 = x\partial_x - y\partial_y,$$
$$X_3 = f(x,y)\partial_x + g(x,y)\partial_y, \quad f_x + g_y = 0. \tag{4.21}$$

Imposing the commutation relations (4.20) we find

$$f = -x^2 + \alpha/y^2, \quad g = 2xy + \beta, \tag{4.22}$$

where α and β are constants. Transformations with constant Jacobian determinant leaving the subalgebra $\{X_1,X_2\}$ invariant are

$$\xi = rx + s/y - p, \quad \eta = qy,$$
$$rq \neq 0, \quad r,s,p,q = \text{const}. \tag{4.23}$$

Using (4.23) to simplify (4.21) and (4.22) we find that two cases must be distinguished. For $\alpha = \beta^2/4$ in (4.22) we can transform X_3 into

$$X_3 = -x^2\partial_x + 2xy\partial_y, \tag{4.24}$$

where $\alpha \neq \beta^2/4$ and we can transform X_3 into

$$X_3 = (-x^2 + 1/y^2)\partial_x + 2xy\partial_y. \tag{4.25}$$

An alternative form of these two sl(2,$\mathbb{C}$) algebras is obtained by a coordinate transformation

$$x = v/u, \quad y = u^2. \tag{4.26}$$

The resulting representatives of conjugacy classes of sl(2,$\mathbb{C}$) algebras are

$$X_1 = u\partial_v, \quad X_2 = \tfrac{1}{2}(-u\partial_u + v\partial_v), \quad X_3 = -v\partial_u, \tag{4.27}$$

and

$$X_1 = u\partial_v, \quad X_2 = \tfrac{1}{2}(-u\partial_u + v\partial_v),$$
$$X_3 = (1/u^3)\partial_v + v\partial_u. \tag{4.28}$$

The classification of three-dimensional subalgebras over $\mathbb{C}$ is summarized in Table II. In the first column we list the isomorphism class of each subalgebra, following Lemma 3.4. For solvable algebras, namely, $L_{3,1},\ldots,L_{3,4}$, $\{X_1,X_2\}$ is an Abelian ideal. This ideal is uniquely defined in all isomorphism classes except $3A_1$ and $A_{3,1}$ (Abelian and nilpotent, respectively). For solvable algebras we also give the matrix

TABLE II. Three-dimensional complex subalgebras.

Type	N_0	Basis X_1	X_2	X_3	M	div X	div X_I	dim V	dim V_I
$3A_1$	$L_{3,1}^{f_1,f_2}(\mathbb{C})$	∂_x	$f_1(y)\partial_x$	$f_2(y)\partial_x$	$\begin{pmatrix}0&0\\0&0\end{pmatrix}$	0	0	1	1
	$(1, f_1, f_2$ linearly independent)								
$A_1 \oplus A_2$	$L_{3,2}(\mathbb{C})$	∂_x	∂_y	$y\partial_y$	$\begin{pmatrix}0&0\\0&1\end{pmatrix}$	μ_3	0	2	2
$A_2 = \{X_2, X_3\}$									
	$L_{3,3}(\mathbb{C})$	$y\partial_y$	∂_x	$x\partial_x$	$\begin{pmatrix}0&0\\0&1\end{pmatrix}$	$\mu_1 + \mu_3$	μ_1	2	2
	$L_{3,4}^{\alpha}(\mathbb{C})$ $(\alpha \neq 0)$	∂_x	$y^{\alpha}\partial_x$	$-(1/\alpha)y\partial_y$	$\begin{pmatrix}0&0\\0&1\end{pmatrix}$	$-(1/\alpha)\mu_3$	0	2	1
	$L_{3,5}(\mathbb{C})$	∂_x	$e^{-y}\partial_x$	∂_y	$\begin{pmatrix}0&0\\0&1\end{pmatrix}$	0	0	2	1
$A_{3,1}$	$L_{3,6}(\mathbb{C})$	∂_x	∂_y	$y\partial_x$	$\begin{pmatrix}0&0\\1&0\end{pmatrix}$	0	0	2	2
	$L_{3,7}(\mathbb{C})$	∂_x	$y\partial_y$	$\ln y\partial_x$	$\begin{pmatrix}0&0\\1&0\end{pmatrix}$	μ_2	μ_2	2	2
$A_{3,2}^{\alpha}$	$L_{3,8}^{\alpha}(\mathbb{C})$	∂_x	∂_y	$x\partial_x + \alpha y\partial_y$	$\begin{pmatrix}1&0\\0&\alpha\end{pmatrix}$	$\mu_3(1+\alpha)$	0	2	2
$0 < \lvert\alpha\rvert \leqslant 1$									
	$L_{3,9}^{\alpha,p}(\mathbb{C})$	∂_x	$y^{p}\partial_x$	$x\partial_x + [(1-\alpha)/p]y\partial_y$	$\begin{pmatrix}1&0\\0&\alpha\end{pmatrix}$	$\mu_3(1 + (1 - \alpha/p))$	0	2	1
	$-2 \leqslant p < 0, \quad \alpha \neq 1$								
	$L_{3,10}^{\alpha}(\mathbb{C})$ $(\alpha \neq 1)$	∂_x	$e^{y}\partial_x$	$x\partial_x + (1-\alpha)\partial_y$	$\begin{pmatrix}1&0\\0&\alpha\end{pmatrix}$	μ_3	0	2	1
	$L_{3,11}^{f}(\mathbb{C})$	∂_x	$f(y)\partial_x$	$x\partial_x$	$\begin{pmatrix}1&0\\0&1\end{pmatrix}$	μ_3	0	1	1
	$\dot{f}(y) \neq 0$								
$A_{3,3}$	$L_{3,12}(\mathbb{C})$	∂_x	∂_y	$(x+y)\partial_x + y\partial_y$	$\begin{pmatrix}1&0\\1&1\end{pmatrix}$	$2\mu_3$	0	2	2
	$L_{3,13}^{\alpha}(\mathbb{C})$ $(\alpha \neq 0)$	∂_x	$\alpha \ln y\partial_x$	$x\partial_x - (1/\alpha)y\partial_y$	$\begin{pmatrix}1&0\\1&1\end{pmatrix}$	$-(1/\alpha)\mu_3$	0	2	1
	$L_{3,13}(\mathbb{C})$	∂_x	$y\partial_x$	$x\partial_x - \partial_y$	$\begin{pmatrix}1&0\\1&1\end{pmatrix}$	μ_3	0	2	1
$A_{3,5}$, $(\mathrm{sl}(2,\mathbb{C}))$	$L_{3,15}(\mathbb{C})$	$x\partial_y$	$\frac{1}{2}(-x\partial_x + y\partial_y)$	$-y\partial_x$		0	0	2	
	$L_{3,16}(\mathbb{C})$	$x\partial_y$	$\frac{1}{2}(-x\partial_x + y\partial_x)$	$x^{-3}\partial_y + y\partial_x$		0	0	2	

M of Lemma 3.4 in the fourth column. In the last columns $X = \Sigma_{i=1}^{3}\mu_i X_i$ is a general element of the Lie algebra, $X_I = \Sigma_{\mu=1}^{2}\mu_i X_i$ a general element of the ideal, V is the vector space spanned by elements of the Lie algebra at a generic point (x, y), V_I the vector space spanned by elements of the ideal at a generic point.

We have arrived at the following theorem.

Theorem 4.3: Every complex three-dimensional subalgebra of the algebra T is conjugate under the pseudogroup P to an algebra in Table III. Two algebras in Table III are mutually conjugate precisely in one of the two following cases:

TABLE III. One-dimensional real subalgebras.

N_0	Basis element X^R	Complex form	div X^R	div X $(p\in\mathbf{R}, \lambda, \eta, H(y)\in\mathbb{C})$	Normalizer in P
$L_{1,1}(\mathbf{R})$	$i(\partial_x - \partial_{\bar{x}})$	$L_{1,1}^{\mathbb{C}}$	0	0	$\xi = px + H(y)$, $\eta = \lambda y + \mu$, $\lambda p \neq 0$
$L_{1,2}(\mathbf{R})$	$i(x\partial_x - \bar{x}\partial_{\bar{x}})$	$L_{1,2}^{\mathbb{C}}$	0	$\neq 0$	$\xi = \lambda \dot{H}^{-1}(y)x, \eta = H(y), \dot{H}(y) \neq 0$
$L_{1,3}^{a}(\mathbb{C})$	$x\partial_x + \bar{x}\partial_{\bar{x}} + ia(x\partial_x - \bar{x}\partial_{\bar{x}})$	$L_{1,2}(\mathbb{C})$	$\neq 0$	$\neq 0$	$\xi = \lambda \dot{H}^{-1}(y)x, \eta = H(y), \dot{H}(y) \neq 0$

$$L_{3,1}^{f_1 f_2}(\mathbb{C}) \sim L_{3,1}^{g_1 g_2}(\mathbb{C}),$$

if $\{1, f_1, f_2\}$ and $\{1, g_1, g_2\}$ are equivalent under relation (4.8),

$$L_{3,11}^{f}(\mathbb{C}) \sim L_{3,11}^{g}(\mathbb{C}),$$

if $\{1, f\}$ and $\{1, g\}$ are equivalent under (4.8).

V. CLASSIFICATION OF LOW-DIMENSIONAL REAL SUBALGEBRAS

We are interested in algebras of vector fields of the form (3.3) satisfying $\operatorname{div} X = \text{const}$, i.e., the divergence of the complexification of X^R is constant. We proceed as in the complex case, remembering that the ground field now is $\mathbb{R}$ rather than $\mathbb{C}$. Whenever possible we make use of the results of Sec. IV.

A. One-dimensional real subalgebras

Starting with a vector field X^R in the form (3.4), we perform the general transformation (3.5), taking X into (3.7). The choice (4.1) takes X into (4.2) and hence X^R into

$$X^R = -(J_0 g/G_x)\partial_\xi - (\bar{J}_0 \bar{g}/\bar{G}_x)\partial_{\bar{\xi}}. \tag{5.1}$$

Requiring that the divergence be constant we obtain

$$\operatorname{div} X = -(J_0 g/G_x)_\xi - (\bar{J}_0 \bar{g}/\bar{G}_x)_{\bar{\xi}} = \alpha + \bar{\alpha}. \tag{5.2}$$

Differentiating with respect to ξ we obtain $(-J_0 g/G_x)_{\xi\xi} = 0$ and hence

$$-J_0 g/G_x = \alpha\xi + \beta. \tag{5.3}$$

The vector field X^R in the coordinates (ξ,η) now is

$$X^R = (\alpha\xi + \beta)\partial_\xi + (\bar{\alpha}\bar{\xi} + \bar{\beta})\partial_{\bar{\xi}}. \tag{5.4}$$

We can multiply X^R by an arbitrary nonzero real constant and also translate ξ. We find three different classes, corresponding to $\alpha = 0$, α pure imaginary and $\operatorname{Re}\alpha \neq 0$, respectively. We thus arrive at the following result.

Theorem 5.1: An arbitrary one-dimensional real subalgebra $\{X^R\}$ of the Lie algebra T of constant divergence holomorphic vector fields is conjugate under the pseudogroup P to precisely one of the algebras in Table III. □

The value of a in $L_{1,3}^{a}(\mathbb{R})$ is invariant under P, since we have $\operatorname{div} X = 1 + ia$ and Theorem 3.1 tells us that the divergence of a vector field X is not changed by holomorphic transformations with constant Jacobian determinant.

We shall need the normalizers of the one-dimensional subalgebras in P, i.e., the transformations (3.5) leaving $L_{1,i}$ $(i = 1,2,3)$ invariant. They are easy to calculate and are given in the sixth column of Table III. In the third column we give the complex Lie algebra, generated by X (rather than X^R).

B. Two-dimensional real subalgebras

Similarly as in the complex case, we shall assume that one element, X_1^R, is already in its standard form, namely one given in Table III. The other basis element X_2^R is in the general form (3.3). We first impose the commutation relation, then simplify X_2^R, using the normalizer of X_1^R in P (given in the sixth column of Table III).

1. Abelian subalgebras

(A) $X_1^R = i(\partial_x - \partial_{\bar{x}})$. Requiring $[X_1^R, X_2^R] = 0$ and using $\operatorname{Nor}_p L_{1,1}(\mathbb{R})$, we obtain

$$X_2^R = \{f(y)p + \dot{H}(y)[(\alpha/\lambda)(y-\mu) + \beta]\}\partial_x + [\alpha y - \mu\alpha + \beta\lambda]\partial_y + \text{c.c.}, \tag{5.5}$$

where $f(y)$, α, and β are given and p, λ, μ, and $H(y)$ are our choices. Depending on the original values of α, β, and $f(y)$, the following possibilities occur:

(A_1) $X_2^R = y\partial_y + \bar{y}\partial_{\bar{y}} + ia(y\partial_y - \bar{y}\partial_{\bar{y}})$,

(A_2) $X_2^R = i(y\partial_y - \bar{y}\partial_{\bar{y}})$,

(X_3) $X_2^R = i(\partial_y - \partial_{\bar{y}})$,

(X_4) $X_2^R = \partial_x + \partial_{\bar{x}}$,

(X_5) $X_2^R = f(y)\partial_x + \bar{f}(\bar{y})\partial_{\bar{x}}$, $\dot{f}(y) \neq 0$.

(B) $X_1^R = i(x\partial_x - \bar{x}\partial_{\bar{x}})$. Requiring $[X_1^R, X_2^R] = 0$ using $\operatorname{Nor}_p(L_{1,2}(\mathbb{R}))$ we obtain

$$X_2^R = [a - \dot{g}(y) - [\ddot{G}(y)/\dot{G}(y)]g(y)]x\partial_x + g\dot{G}(y)\partial_y + \text{c.c.}, \tag{5.6}$$

where a and $g(y)$ are given and $G(y)$ is our choice. Two possibilities occur, namely, $g(y) \neq 0$ and $g(y) = 0$,

(B_6) $X_2^R = y\partial_y + \bar{y}\partial_{\bar{y}}$,

(B_7) $X_2^R = x\partial_x + \bar{x}\partial_{\bar{x}}$.

(C) $X_1^R = x\partial_x + \bar{x}\partial_{\bar{x}} + ia(x\partial_x - \bar{x}\partial_{\bar{x}})$. Proceeding as above we obtain

$$X_2^R = [ib - \dot{g}(y) - (g/\dot{G})\ddot{G}]x\partial_x + g(y)\dot{G}\partial_y + \text{c.c.} \tag{5.7}$$

For $g \neq 0$, $b \neq 0$ we obtain an algebra conjugate to B_6. For $q \neq 0$, $b = 0$ we obtain an algebra conjugate to A_1. For $g = 0$ we reobtain B_7.

2. Non-Abelian subalgebras

Since the vector fields X_1 and X_2 satisfy $[X_1, X_2] = X_1$ it follows from Lemma 3.2 that $\operatorname{div} X = 0$. Hence we can always put

$$X_1^R = i(\partial_x - \partial_{\bar{x}}).$$

Using the commutation relation and the normalizer $\operatorname{Nor}_p L_{1,1}$ we obtain

$$X_2^R = [x - H(y) + ph + \{[(\alpha y - \mu)/\lambda] + \beta\}\dot{H}]\partial_x + (\alpha y_\alpha \mu + \beta\lambda)\partial_y + \text{c.c.} \tag{5.8}$$

Here $h(y)$, α, and β are given, p, λ, μ, and $H(y)$ are our choices. We distinguish three cases, namely, $\alpha = 0$, $\beta \neq 0$; $\alpha = 0$, $\beta = 0$; and $\alpha \neq 0$. We obtain

(D_8) $X_2^R = x\partial_x + \bar{x}\partial_{\bar{x}} + i(\partial_y - \partial_{\bar{y}})$,

(D_9) $X_2^R = x\partial_x + \bar{x}\partial_{\bar{x}}$,

(D_{10}) $X_2^R = x\partial_x + \bar{x}\partial_{\bar{x}} + a(y\partial_y + \bar{y}\partial_{\bar{y}}) + ib(y\partial_y - \bar{y}\partial_y)$, $(a,b) \neq (0,0)$.

For each algebra $L_{2,k}$ $(k = 1,\ldots,10)$, we calculate its normalizer $\operatorname{Nor}_p L_{2,k}(\mathbb{R})$ in the pseudogroup P. All results are summarized in Table IV. In the fourth column we give

TABLE IV. Two-dimensional real subalgebras.

Type	N_0	Basis X_1^R	X_2^R	Complexification	div X $X=p_1X_1+p_2X_2$	$\mathrm{Nor}_P L_{2,k}(\mathbb{R})$
$2A_1$	$L_{2,1}(\mathbb{R})=A_4$	$i(\partial_x-\partial_{\bar{x}})$	$\partial_x+\partial_{\bar{x}}$	$L_{1,1}(\mathbb{C})$	0	$\xi=\lambda x+H(y),\ \eta=\mu y+\nu,\ \lambda\mu\neq 0$
	$L_{2,2}(\mathbb{R})=B_7$	$i(x\partial_x-\bar{x}\partial_{\bar{x}})$	$x\partial_x+\bar{x}\partial_{\bar{x}}$	$L_{1,2}(\mathbb{C})$	ip_1+p_2	$\xi=[\lambda/H(y)]x,\ \eta=H(y),\ \lambda\dot{H}(y)\neq 0$
	$L_{2,3}(\mathbb{R})=A_3$	$i(\partial_x-\partial_{\bar{x}})$	$i(\partial_y-\partial_{\bar{y}})$	$L_{2,1}(\mathbb{C})$	0	$\xi=px+qy+\lambda,\ \eta=rx+sy+\nu,$ $ps-qr\neq 0$
	$L_{2,4}(\mathbb{R})=A_2$	$i(\partial_x-\partial_{\bar{x}})$	$i(y\partial_y-\bar{y}\partial_{\bar{y}})$	$L_{2,2}(\mathbb{C})$	ip_2	$\xi=px+q\ln y+\mu,\ \eta=\lambda y,\ p\lambda\neq 0$
	$L_{2,5}^{a}(\mathbb{R})=A_1$	$i(\partial_x-\partial_{\bar{x}})$	$y\partial_y+\bar{y}\partial_{\bar{y}}+ia(y\partial_y-\bar{y}\partial_{\bar{y}})$	$L_{2,2}(\mathbb{C})$	$(1+ia)p_2$	$\xi=px+[iq/(1+ia)]\ln y+\mu,\ \eta=\lambda y,\ p\lambda\neq 0$
	$L_{2,6}(\mathbb{R})=B_6$	$i(x\partial_x-\bar{x}\partial_{\bar{x}})$	$y\partial_y+\bar{y}\partial_{\bar{y}}$	$L_{2,2}(\mathbb{C})$	ip_1+p_2	$\xi=\alpha x,\ \eta=\beta y,\ \alpha\beta\neq 0$
	$L_{2,7}^{f}(\mathbb{R})=A_5$ $\dot{f}(y)\neq 0$	$i(\partial_x-\partial_{\bar{x}})$	$f(y)\partial_x+f(\bar{y})\partial_{\bar{x}}$	$L_{2,3}^{f}(\mathbb{C})$	0	$\xi=[\alpha/\dot{H}(y)]x+K(y),\ \eta=H(y),\ \alpha\dot{H}(y)\neq 0$ $i\beta f(y)f[H(y)]+(-af(y)+\delta f[H(y)])$ $+ic=0,\ a\delta-\beta c=0$
$A_{2,1}$	$L_{2,8}(\mathbb{R})=D_8$	$i(\partial_x-\partial_{\bar{x}})$	$x\partial_x+\bar{x}\partial_{\bar{x}}+i(\partial_y-\partial_{\bar{y}})$	$L_{2,4}(\mathbb{C})$	p_2	$\xi=px+\lambda e^{-iqy}+ir,\ \eta=y+\mu,\ p\neq 0$
	$L_{2,9}(\mathbb{R})=D_9$	$i(\partial_x-\partial_{\bar{x}})$	$x\partial_x+\bar{x}\partial_{\bar{x}}$	$L_{2,5}^{0}(\mathbb{C})$	p_2	$\xi=px+q,\ \eta=\lambda x+\mu,\ p\lambda\neq 0$
	$L_{2,10}^{a,b}(\mathbb{R})=D_{10}$ $(a,b)\neq(0,0)$	$i(\partial_x-\partial_{\bar{x}})$	$x\partial_x+\bar{x}\partial_{\bar{x}}+a(y\partial_y+\bar{y}\partial_{\bar{y}})$ $+ib(y\partial_y-\bar{y}\partial_{\bar{y}})$	$L_{2,5}^{a+ib}(\mathbb{C})$	$p_2(1+a+ib)$	$\xi=x+\mu y[p/(a+ib)]-iq,\ \eta=\lambda y,\ \lambda\neq 0$

the corresponding complex algebra of Table I. The normalizers are in the sixth column.

We arrive at the following statement.

Theorem 5.2: Every two-dimensional real subalgebra of the Lie algebra T of constant divergence holomorphic vector fields is conjugate under the pseudogroup P to an algebra in Table IV. Two algebras in Table IV are mutually conjugate if and only if they are of the type $L_{2,7}^{f}(\mathbb{R})$ and $L_{2,7}^{g}(\mathbb{R})$, where $\{1,f(y)\}$ and $\{1,g(y)\}$ are equivalent under the relation (4.8).

Notice that a complication occurs for the algebras $L_{2,7}^{f}(\mathbb{R})$: The function $H(y)$ in the normalizer (see Table IV) satisfies a functional relation

$$i\beta f(y)f(H(y))+\{-af(y)+\delta f[H(y)]\}+ic=0. \tag{5.9}$$

Thus four numbers a, $c\in\mathbb{R}$, β, $\delta\in\mathbb{C}$, satisfying $a\delta-\beta c\neq 0$ must exist, such that $H(y)$ satisfies (5.9).

C. Three-dimensional real subalgebras

As in the case of complex three-dimensional subalgebras, we start with the solvable ones and assume that their two-dimensional Abelian ideal $\{X_1^R,X_2^R\}$ is in one of the standard forms $L_{2,i}(\mathbb{R})$ $(i=1,\dots,7)$ of Table IV. We take X_3^R in the general form (3.4). We first impose the commutation relations

$$\begin{pmatrix}[X_1^R,X_3^R]\\ [X_2^R,X_3^R]\end{pmatrix}\times\begin{pmatrix}a & b\\ c & d\end{pmatrix}\begin{pmatrix}X_1^R\\ X_2^R\end{pmatrix},\quad a,b,c,d\in\mathbb{R}, \tag{5.10}$$

and then simplify X_3^R using the appropriate normalizer $\mathrm{Nor}_P\, L_{2,i}(\mathbb{R})$, listed in the sixth column of Table IV.

1. Solvable subalgebras

(A) Ideal $L_{2,1}(\mathbb{R})$. Implementing the commutation relations (5.10) in this case we find

$$X_3^R=[(a+ib)x+\phi(y)]\partial_x+(\alpha y+\beta)\partial_y+\text{c.c.} \tag{5.11}$$

and moreover

$$c=-b,\quad d=a, \tag{5.12}$$

in (5.9). Using the normalizer $\mathrm{Nor}_P(\mathbb{R})$ we transform X_3^R to

$$X_3^R=\{(a+ib)\xi+[(\alpha\eta-\alpha\nu+\beta\mu)(\dot{H}/\mu)-(a-ib)H+\lambda\phi]\}\partial_\xi+(\alpha\eta-\alpha\nu+\beta\mu)\partial_\eta+\text{c.c.}, \tag{5.13}$$

where λ, μ, ν, and $H(y)$ can be chosen at will $(\lambda\mu\neq 0)$.

The following possibilities occur. (i) Abelian algebras: $a=b=0$.

(A_1) $X_3^R=y\partial_y+\bar{y}\partial_{\bar{y}}+ip(y\partial_y-\bar{y}\partial_{\bar{y}})$,

(A_2) $X_3^R=i(y\partial_y-\bar{y}\partial_{\bar{y}})$,

(A_3) $X_3^R=i(\partial_y-\partial_{\bar{y}})$,

(A_4) $\phi(y)\partial_x+\bar{\phi}(\bar{y})\partial_{\bar{x}}$, $\dot{\phi}(y)\neq 0$.

(ii) Algebras $A_{3,2}^{1}$: $a=1$, $b=0$.

(A_5) $X_3^R=x\partial_x+\bar{x}\partial_{\bar{x}}+i(\partial_y-\partial_{\bar{y}})$,

(A_6) $X_3^R=x\partial_x+\bar{x}\partial_{\bar{x}}+p(y\partial_y+\bar{y}\partial_{\bar{y}})+iq(y\partial_y-\bar{y}\partial_{\bar{y}})$.

(iii) Algebras $A_{3,4}^{a}$: $b=1$.

(A_7) $X_3^R=a(x\partial_x+\bar{x}\partial_{\bar{x}})+i(x\partial_x-\bar{x}\partial_{\bar{x}})+\partial_y+\partial_{\bar{y}}$.

(A_8) $X_3^R=a(x\partial_x+\bar{x}\partial_{\bar{x}})+i(x\partial_x-\bar{x}\partial_{\bar{x}})+p(y\partial_y+\bar{y}\partial_{\bar{y}})+iq(y\partial_y-\bar{y}\partial_{\bar{y}})$.

(B) Ideal $L_{2,2}(\mathbb{R})$. In view of Lemma 3.2 neither of the operators $\{X_1^R,X_2^R\}\in L_{2,2}(\mathbb{R})$ can be in the derived algebra, hence any algebra containing $L_{2,2}$ as an ideal must be Abelian. We find

$$X_3^R=-x\dot{g}(y)\partial_x+g(y)\partial_y+\text{c.c.} \tag{5.14}$$

A transformation in $\mathrm{Nor}_P\, L_{2,2}(\mathbb{R})$ can be found, taking (5.14) into

$$(B_9)\quad X_3^R=i(\partial_y-\partial_{\bar{y}}). \tag{5.15}$$

(C) Ideal $L_{2,3}(\mathbb{R})$. For $L_{2,3}(\mathbb{R})$ to be an ideal X_3^R must have coefficients that are linear in x and y. We find

$$X_3^R = (ax + cy + r)\partial_x + (bx + dy + s)\partial_y + \text{c.c.} \qquad (5.16)$$

The normalizer $\text{Nor}_P\, L_{2,3}(\mathbb{R})$ is then used to transform the matrix in (5.10) into its standard form. The following cases occur: (i) Abelian algebras: $a = b = c = d = 0$.

$$X_3^R = r(\partial_x + \partial_{\bar{x}}) + s(\partial_y + \partial_{\bar{y}}).$$

The corresponding algebra is conjugate under P to the algebra A_3. (ii) Decomposable algebras: $a = 1$, $b = c = d = 0$. Depending on whether $s = 0$ or $s \neq 0$, we obtain,

$$(C_{10})\ \ X_3^R = x\partial_x + \bar{x}\partial_{\bar{x}},$$

$$(C_{11})\ \ X_3^R = x\partial_x + \bar{x}\partial_{\bar{x}} + \partial_y + \partial_{\bar{y}}.$$

(iii) Nilpotent algebras: $a = b = d = 0, c = 1$. Two cases occur:

$$(C_{12})\ \ X_3^R = y\partial_x + \bar{y}\partial_{\bar{x}},$$

$$(C_{13})\ \ X_3^R = y\partial_x + \bar{y}\partial_{\bar{x}} + \partial_y + \partial_{\bar{y}}.$$

(iv) $A_{3,2}^P$ algebras: $a = 1$, $d = \alpha$, $b = c = 0$.

$$(C_{14})\ \ X_3^R = x\partial_x + \bar{x}\partial_{\bar{x}} + p(y\partial_y + \bar{y}\partial_{\bar{y}}),$$

$$-1 \leqslant p \leqslant 1, \quad p \neq 0.$$

(v) $A_{3,3}$ algebras: $a = c = d = 1$, $b = 0$.

$$(C_{15})\ \ X_3^R = (x + y)\partial_x + (\bar{x} + \bar{y})\partial_{\bar{x}} + y\partial_y + \bar{y}\partial_{\bar{y}}.$$

(vi) $A_{3,4}^p$ algebras: $a = d = p$, $b = -c = 1$.

$$(C_{16})\ \ X_3^R = p(x\partial_x + \bar{x}\partial_{\bar{x}} + y\partial_y + \bar{y}\partial_{\bar{y}}) - y\partial_x + x\partial_y - \bar{y}\partial_{\bar{x}} + \bar{x}\partial_{\bar{y}}.$$

(D) Ideal $L_{2,4}(\mathbb{R})$. Imposing the usual commutation relations and performing a transformation by the normalizer of $L_{2,3}(\mathbb{R})$, we find

$$X_3 = [ax + (-a\beta + c\alpha)\ln(y/\delta) - a\gamma + \beta q + p\alpha]\partial_x + qy\partial_y + \text{c.c.} \qquad (5.17)$$

where $a, c, p, q \in \mathbb{R}$ are given and $\alpha, \beta, \gamma, \delta$ are our choice. Notice that we have $b = d = 0$ in (5.10), since $\operatorname{div} X_2 \neq 0$ (Lemma 3.2).

The following possibilities occur: (i) Abelian algebra: $a = c = 0$. We must have $q \neq 0$; we choose $\beta = -p\alpha/q$ and reobtain the algebra B_9 corresponding to $X_3^{R(9)}$ and the ideal $L_{2,2}(\mathbb{R})$. (ii) Decomposable algebra: $a = 1$. Choosing β and γ appropriately we obtain

$$(D_{17})\ \ X_3^{R(17)} = x\partial_x + \bar{x}\partial_{\bar{x}} + p(y\partial_y + \bar{y}\partial_{\bar{y}}).$$

(iii) Nilpotent algebras: $a = 0$, $c = 1$. We obtain, for $q \neq 0$ and $q = 0$, respectively,

$$(D_{18})\ \ X_3^{R(18)} = \ln y\partial_x + \ln \bar{y}\partial_{\bar{x}} + (y\partial_y + \bar{y}\partial_{\bar{y}}).$$

$$D_{19}\text{:}\ \ X_3^{R(19)} = \ln y\partial_x + \ln \bar{y}\partial_{\bar{x}}.$$

(E) Ideal $L_{2,5}^p(\mathbb{R})$. In this case $\operatorname{div} X_2 = (1 + ip) \neq 0$, hence X_2 cannot be in the derived algebra and we have $b = d = 0$ in the matrix M of (5.10). From (5.10) we obtain

$$X_3^R = (ax + [ic/(1 + ip)]\ln y + r)\partial_x + isy\partial_y + \text{c.c.}, \quad a,c,r,s,p \in \mathbb{R}. \qquad (5.18)$$

Using the normalizer $\text{Nor}_P\, L_{2,5}^p(\mathbb{R})$ we transform X_3^R into

$$X_3^R = \{a\xi + [i/(1 + ip)](-ab_2 + cb_1)\ln(\eta/\beta) - a\alpha + rb_1 - b_2 s/(1 + ip)\}\partial_\xi + is\eta\partial_\eta + \text{c.c.},$$

where $b_1, b_2 \in \mathbb{R}$ and $\alpha, \beta \in \mathbb{C}$ are at our disposal ($b_1\beta \neq 0$). The following cases occur: (i) Abelian algebras: $a = c = 0$. If $s \neq 0$ we obtain an algebra conjugate to B_9, if $s = 0$ we reobtain A_1. (ii) Decomposable algebras: $a = 1, c = 0$. We obtain one new algebra

$$E_{20}\text{:}\ \ X_3^R = x\partial_x + \bar{x}\partial_{\bar{x}} + is(y\partial_y - \bar{y}\partial_{\bar{y}}).$$

(iii) Nilpotent algebras: $a = 0$, $c = 1$. For $s = 0$ we obtain a new algebra

$$E_{21}\text{:}\ \ X_3^R = [i/(1 + ip)]\ln y\partial_x - [i/(1 - ip)]\ln \bar{y}\partial_{\bar{x}}.$$

For $s \neq 0$ we obtain the algebra

$$X_1^R = i(\partial_x - \partial_{\bar{x}}), \quad X_2^R = (1 + ia)y\partial_y + \text{c.c.},$$

$$X_3^R = iy\partial_y + [i/(1 + ip)]\ln y\partial_x + \text{c.c.} \qquad (5.19)$$

Putting

$$X_1^R = i(\partial_x - \partial_{\bar{x}}),$$

$$\tilde{X}_2^R = X_3^R = iy\partial_y + [i/(1 + ip)]\ln y\partial_x + \text{c.c.},$$

$$\tilde{X}_3^R = X_2^R - pX_3^R = y\partial_y - [ip/(1 + ip)]\ln y\partial_x + \text{c.c.},$$

and transforming to new variables

$$\xi = x + H(y), \quad \eta = y, \quad \text{with } y\dot{H} = -[1/(1 + ip)]\ln y,$$

we show that (5.19) is conjugate to D_{18}.

(F) Ideal $L_{2,6}(\mathbb{R})$. Since $\operatorname{div} X_1 \neq 0$, $\operatorname{div} X_2 \neq 0$ any algebra with $L_{2,6}(\mathbb{R})$ as an ideal must be Abelian. Hence $a = b = c = d = 0$ in (5.10). Imposing the commutation relations, we obtain a simple algebra, conjugate to B_9.

(G) Ideal $L_{2,7}^\phi(\mathbb{R})$. The ideal in this case depends on an arbitrary function that we shall here denote $\phi(y)$; both X_1 and X_2 can be present in the derived algebra. With no loss of generality we can assume that the matrix M in (5.10) is already in one of its standard forms. Imposing (5.10) we find

$$X_1^R = i(\partial_x - \partial_{\bar{x}}), \quad X_2^R = \phi(y)\partial_x + \bar{\phi}(\bar{y})\partial_{\bar{x}},$$

$$X_3^R = \{[a - ib\phi(y)]x + \psi(y)\}\partial_x + g(y)\partial_y + \text{c.c.}, \qquad (5.20)$$

with

$$g\dot{\phi} = -ib\phi^2 + (a - d)\phi - ic, \quad -ib\dot{\phi}(y) + \ddot{g}(y) = 0. \qquad (5.21)$$

Rather than use the normalizer of $L_{2,7}^\phi(\mathbb{R})$ in P, we use a simpler transformation

$$\xi = x + H(y), \quad \eta = \lambda y + \mu, \qquad (5.22)$$

that takes (5.20) into

$$X_1^R = i(\partial_\xi - \partial_{\bar{\xi}}), \quad X_2^R = \phi\partial_\xi + \text{c.c.},$$

$$X_3^R = [(a - ib\phi)\xi + \psi + g\dot{H} - (a - ib\phi)H]\partial_\xi + \lambda g\partial_\eta. \qquad (5.23)$$

In (5.23) we have actually put $\phi(y) = \phi((\eta - \mu)/\lambda) = \tilde{\phi}(\eta)$ and then dropped the $\sim$ sign; the same holds for $\psi(y)$, $g(y)$ and the auxiliary function $H(y)$ (all are now considered as functions of η). We now run through all possible types of algebras. In each case we must solve Eqs. (5.21) and then use $H(y)$, λ, and μ to simplify the result. (i) Abelian algebras: $a=b=c=d=0$. We obtain one type of algebra, namely

$$G_{22}\colon X_1^R=i(\partial_x-\partial_{\bar x}),$$
$$X_2^R=\phi_1(y)\partial_x+\bar\phi_1(\bar y)\partial_{\bar x},$$
$$X_3^R=\phi_2(y)\partial_x+\bar\phi_2(\bar y)\partial_{\bar x},$$

where 1, $\phi_1(y)$, and $\phi_2(y)$ are linearly independent. (ii) Decomposable algebras: $a=b=c=0, d=1$. From (5.21) we find $g=\alpha\eta+\beta$. For $\alpha\neq 0$ and $\alpha=0, \beta\neq 0$, we obtain, respectively,

$$G_{23}\colon X_1^R=i(\partial_x-\partial_{\bar x}),\quad X_2^R=y^\alpha\partial_x+\bar y^{\bar\alpha}\partial_{\bar x},$$
$$X_3^R=-(1/\alpha)y\partial_y-(1/\bar\alpha)\bar y\partial_{\bar y}.$$
$$G_{24}\colon X_1^R=i(\partial_x-\partial_{\bar x}),$$
$$X_2^R=e^{-y}\partial_x+e^{-\bar y}\partial_{\bar x},\quad X_3^R=\partial_y+\partial_{\bar y}.$$

(iii) Nilpotent algebra: $a=b=d=0, c=1$. From (5.21) we again have $g=\alpha\eta+\beta$ and again two cases occur: $\alpha\neq 0$ or $\alpha=0, \beta\neq 0$,

$$X_1^R=i(\partial_x-\partial_{\bar x}),\quad X_2=-i\ln y\partial_x+i\ln\bar y\partial_{\bar x},$$
$$X_3=y\partial_y+\bar y\partial_{\bar y}$$

and

$$X_1^R=i(\partial_x-\partial_{\bar x}),\quad X_2=y\partial_x+\bar y\partial_{\bar x},$$
$$X_3=-i(\partial_y-\partial_{\bar y}).$$

Neither of these are new; the first coincides with a special case of E_{22}, the second with C_{12}. (iv) Algebras $A_{3,2}^p\colon a=1, b=c=0, d=p, -1\leqslant p<1, p\neq 0$. Standard calculations lead to three new types of algebras, namely,

$$G_{25}^{p,\alpha}\colon X_1=i(\partial_x-\partial_{\bar x}),\quad X_2=y^\alpha\partial_x+\bar y^{\bar\alpha}\partial_{\bar x},$$
$$X_3=x\partial_x+\bar x\partial_{\bar x}+[(1-p)/\alpha]y\partial_y$$
$$+[(1-a)/\bar\alpha]\bar y\partial_{\bar y},\quad \alpha\in\mathbb{C},\quad \alpha\neq 0,$$
$$G_{26}^p\colon X_1=i(\partial_x-\partial_{\bar x}),\quad X_2=e^y\partial_x+e^{\bar y}\partial_{\bar x},$$
$$X_3=x\partial_x+\bar x\partial_{\bar x}+(1-p)\partial_y+(1-p)\partial_{\bar y},$$
$$G_{27}^\phi\colon X_1=i(\partial_x-\partial_{\bar x}),\quad X_2=\phi(y)\partial_x+\bar\phi(\bar y)\partial_{\bar y},$$
$$X_3=x\partial_x+\bar x\partial_{\bar x},\quad p=1,\quad \phi'(y)\neq 0.$$

(v) Algebras $A_{3,3}$: $a=c=d=1, b=0$. Since $b=0$ we have $g(y)=\lambda y+\mu$ from (5.21). Depending on whether $\lambda=0$ or $\lambda\neq 0$, we obtain one of the following algebras:

$$G_{28}^\lambda\colon X_1=i(\partial_x-\partial_{\bar x}),$$
$$X_2=(-(i/\lambda)\ln y)\partial_x+((i/\lambda)\ln\bar y)\partial_{\bar x},$$
$$X_3=x\partial_x+\bar x\partial_x+\lambda y\partial_y+\bar\lambda y\partial_{\bar\lambda},\quad \lambda\neq 0,$$
$$G_{29}\colon X_1=i(\partial_x-\partial_{\bar x}),\quad X_2=i(y\partial_x-\bar y\partial_{\bar x}),$$
$$X_3=x\partial_x+\bar x\partial_{\bar x}-\partial_y-\partial_{\bar y}.$$

(vi) Algebras $A_{3,4}^a\colon d=a, b=-c=1$. In this case the first of Eqs. (5.21) is nonlinear and difficult to solve. To avoid solving it we first diagonalize the matrix $M=\begin{pmatrix} a & 1\\ -1 & a\end{pmatrix}$ over the field of complex numbers, solve (5.21) for that case, then transform back. This leads to two algebras, namely,

$$G_{30}^{a,\lambda}\colon X_1^R=(i+y^{2i/\lambda})\partial_x+(-i+\bar y^{-2i/\lambda})\partial_{\bar x},$$
$$X_2^R=(1+iy^{2i/\lambda})\partial_x+(1-i\bar y^{-2i/\lambda})\partial_{\bar x},$$
$$X_3^R=a(x\partial_x+\bar x\partial_{\bar x})+i(x\partial_x-\bar x\partial_{\bar x})$$
$$+\lambda y\partial_y+\bar\lambda\bar y\partial_{\bar y},$$

and

$$G_{31}\colon X_1^R=i(\partial_x-\partial_{\bar x}),$$
$$X_2=\left(\frac{1+e^{-2iy}}{1-e^{-2iy}}\right)\partial_x+\left(\frac{1+e^{2iy}}{1-e^{2iy}}\right)\partial_{\bar x},$$
$$X_3^R=\left(a+\frac{1+e^{-2iy}}{1-e^{-2iy}}\right)x\partial_x$$
$$+\left(a+\frac{1+e^{2iy}}{1-e^{2iy}}\right)\bar x\partial_{\bar x}+i(\partial_y-\partial_{\bar y}).$$

The ideal $\{X_1^R,X_2^R\}$ in the case G_{30} is not in its standard form $L_{2,7}^\phi(\mathbb{R})$. A transformation taking it into its standard form exists, but we have not constructed it explicitly.

The results obtained so far are summarized in Table V where the subalgebras are listed by isomorphism classes. The algebras $L_{3,2}(\mathbb{R})$, $L_{3,3}^G(\mathbb{R})$, and $L_{3,4}(\mathbb{R})$ all have the same complexification, namely the two-dimensional algebra $L_{2,2}(\mathbb{C})$. However, they are not conjugate to each other. Indeed, putting $X=\Sigma_{i=1}^3 p_iX_i$ we find $\operatorname{div}X=ip_3$ in $L_{3,2}(\mathbb{R})$, $\operatorname{div}X=(1+ia)p_3$ in $L_{3,3}(\mathbb{R})$ and $\operatorname{div}X=p_1+ip_2$ in $L_{3,4}(\mathbb{R})$. For the algebras $A_{3,1}$ (nilpotent) we have weeded out all redundancies, due to the fact that in this case the Abelian ideal is not unique. In all other cases the Abelian ideal is unique. Once it is fixed, the only allowed transformations are in the normalizer of the ideal and these were used to the maximal possible degree in the text.

2. Simple subalgebras

Let us now turn to the simple three-dimensional real subalgebras of the algebra of holomorphic vector fields with constant divergence. Upon complexification such an algebra will turn into sl(2,$\mathbb{C}$), i.e., into either $L_{3,15}(\mathbb{C})$ or $L_{3,16}(\mathbb{C})$ of Table II. We shall rewrite these two algebras as the real algebras o(3,1) and then pick out the corresponding o(3) and o(2,1) subalgebras [unique up to conjugacy under the corresponding o(3,1) group].

We start with the "linear" sl(2,$\mathbb{C}$) algebra $L_{3,15}(\mathbb{C})$. Its o(3,1) realization is represented by

$$L_1=(i/2)[+x\partial_x-y\partial_y-\bar x\partial_{\bar x}+\bar y\partial_{\bar y}],$$
$$L_2=(i/2)[-x\partial_y-y\partial_x+\bar x\partial_{\bar y}+\bar y\partial_{\bar x}],$$
$$L_3=\tfrac12[x\partial_y-y\partial_x+\bar x\partial_{\bar y}-\bar y\partial_{\bar x}],$$
$$K_1=\tfrac12[-x\partial_x+y\partial_y-\bar x\partial_{\bar x}+\bar y\partial_{\bar y}],$$
$$K_2=\tfrac12[x\partial_y+y\partial_x+\bar x\partial_{\bar y}+\bar y\partial_{\bar x}],$$
$$K_3=(i/2)[x\partial_y-y\partial_x-\bar x\partial_{\bar y}+\bar y\partial_{\bar x}].\tag{5.24}$$

The commutation relations are the standard ones, namely,

$$[L_i,L_k]=\epsilon_{ikl}L_l,$$
$$[L_i,K_k]=\epsilon_{ikl}K_l,$$
$$[K_i,K_k]=-\epsilon_{ikl}L_l.\tag{5.25}$$

TABLE V. Three-dimensional real subalgebras.

Type	N_0	Basis X_1^R	X_2^R	X_3^R	Complexification
$3A_1$	$L_{3,1}(\mathbb{R}) = A_3$	$i(\partial_x - \partial_{\bar x})$	$\partial_x + \partial_{\bar x}$	$i(\partial_y - \partial_{\bar y})$	$L_{2,1}(\mathbb{C})$
	$L_{3,2}(\mathbb{R}) = A_2$	$i(\partial_x - \partial_{\bar x})$	$\partial_x + \partial_{\bar x}$	$i(y\partial_y - \bar y\partial_{\bar y})$	$L_{2,2}(\mathbb{C})$
	$L_{3,3}(\mathbb{R}) = A_1$	$i(\partial_x - \partial_{\bar x})$	$\partial_x + \partial_{\bar x}$	$y\partial_y + \bar y\partial_{\bar y} + ia(y\partial_y - \bar y\partial_{\bar y})$	$L_{2,2}(\mathbb{C})$
	$L_{3,4}(\mathbb{R}) = B_9$	$x\partial_x + \bar x\partial_{\bar x}$	$i(x\partial_x - \bar x\partial_{\bar x})$	$i(\partial_y - \partial_{\bar y})$	$L_{2,2}(\mathbb{C})$
	$L^f_{3,5}(\mathbb{R}) = A_4, (\dot f \neq 0)$	$i(\partial_x - \partial_{\bar x})$	$\partial_x + \partial_{\bar x}$	$f(y)\partial_x + \bar f(\bar y)\partial_{\bar x}$	$L^f_{2,3}(\mathbb{C})$
	$L^{f_1 f_2}_{3,6}(\mathbb{R}) = G_{22}$ $(1, f_1, f_2$ linearly independent$)$	$i(\partial_x - \partial_{\bar x})$	$f_1(y)\partial_x + \bar f_1(\bar y)\partial_{\bar x}$	$f_2(y)\partial_x + f_2(\bar y)\partial_{\bar x}$	$L^{f_1 f_2}_{3,1}(\mathbb{C})$
$A_1 \oplus A_2$	$L_{3,7}(\mathbb{R}) = C_{10}$	$i(\partial_x - \partial_{\bar x})$	$i(\partial_y - \partial_{\bar y})$	$y\partial_y + \bar y\partial_{\bar y}$	$L_{3,2}(\mathbb{C})$
$A_2 = \{X_2^R, X_3^R\}$	$L_{3,8}(\mathbb{R}) = C_{11}$	$i(\partial_x - \partial_{\bar x})$	$i(\partial_y - \partial_{\bar y})$	$y\partial_y + \bar y\partial_{\bar y} + \partial_x + \partial_{\bar x}$	$L_{3,2}(\mathbb{C})$
	$L^a_{3,9}(\mathbb{C}) = D_{17}$	$i(y\partial_y - y\partial_{\bar y})$	$i(\partial_x - \partial_{\bar x})$	$x\partial_x + \bar x\partial_{\bar x} + a(y\partial_y + \bar y\partial_{\bar y})$	$L_{3,3}(\mathbb{C})$
	$L^{ab}_{3,10}(\mathbb{R}) = E_{20}$	$y\partial_y + \bar y\partial_{\bar y} + ia(y\partial_y - \bar y\partial_{\bar y})$	$i(\partial_x - \partial_{\bar x})$	$x\partial_x + \bar x\partial_{\bar x} + ib(y\partial_y - \bar y\partial_{\bar y})$	$L_{3,3}(\mathbb{C})$
	$L^\alpha_{3,11}(\mathbb{R}) = G_{23}, (\alpha \neq 0)$	$i(\partial_x - \partial_{\bar x})$	$y^\alpha\partial_x + \bar y^{\bar\alpha}\partial_{\bar x}$	$-(1/\alpha)y\partial_y - (1/\bar\alpha)\bar y\partial_{\bar y}$	$L^\alpha_{3,4}(\mathbb{C})$
	$L_{3,12}(\mathbb{R}) = G_{24}$	$i(\partial_x - \partial_{\bar x})$	$e^{-y}\partial_x + e^{-\bar y}\partial_{\bar x}$	$\partial_y + \partial_{\bar y}$	$L_{3,5}(\mathbb{C})$
$A_{3,1}$	$L_{3,13}(\mathbb{R}) = C_{12}$	$i(\partial_x - \partial_{\bar x})$	$i(\partial_y - \partial_{\bar y})$	$y\partial_x + \bar y\partial_{\bar x}$	$L_{3,6}(\mathbb{C})$
	$L_{3,14}(\mathbb{R}) = C_{13}$	$i(\partial_x - \partial_{\bar x})$	$i(\partial_y - \partial_{\bar y})$	$y\partial_x + \bar y\partial_{\bar x} + \partial_y + \partial_{\bar y}$	$L_{3,6}(\mathbb{C})$
	$L_{3,15}(\mathbb{R}) = D_{19}$	$i(\partial_x - \partial_{\bar x})$	$i(y\partial_y - \bar y\partial_{\bar y})$	$\ln y\partial_x + \ln \bar y\partial_{\bar x}$	$L_{3,7}(\mathbb{C})$
	$L_{3,16}(\mathbb{R}) = D_{18}$	$i(\partial_x - \partial_{\bar x})$	$i(y\partial_y - \bar y\partial_{\bar y})$	$\ln y\partial_x + \ln \bar y\partial_{\bar x} + y\partial_y + \bar y\partial_{\bar y}$	$L_{3,7}(\mathbb{C})$
	$L^a_{3,17}(\mathbb{R}) = E_{21}$	$i(\partial_x - \partial_{\bar x})$	$y\partial_y + \bar y\partial_{\bar y} + ia(y\partial_y - \bar y\partial_{\bar y})$	$[i/(1+ia)]\ln y\partial_x - [i/(1-ia)]\ln \bar y\partial_{\bar x}$	$L_{3,7}(\mathbb{C})$
$A^a_{3,2}$: $-1<a<1$ $a\neq 0$	$L_{3,18}(\mathbb{R}) = A_5, (a=1)$	$i(\partial_x - \partial_{\bar x})$	$\partial_x + \partial_{\bar x}$	$x\partial_x + x\partial_{\bar x} + i(\partial_y - \partial_{\bar y})$	$L_{2,4}(\mathbb{C})$
	$L_{3,19}(\mathbb{R}) = A_6, (a=1)$	$i(\partial_x - \partial_{\bar x})$	$\partial_x + \partial_{\bar x}$	$x\partial_x + \bar x\partial_{\bar x} + p(y\partial_y + \bar y\partial_{\bar y}) + iq(y\partial_y - \bar y\partial_{\bar y})$	$L^a_{2,5}(\mathbb{C})$
	$L^a_{3,20}(\mathbb{R}) = C_{14}$	$i(\partial_x - \partial_{\bar x}$	$i(\partial_y - \partial_{\bar y})$	$x\partial_x + \bar x\partial_{\bar x} + a(y\partial_y + \bar y\partial_{\bar y})$	$L^a_{2,8}(\mathbb{C})$
	$L^{a\alpha}_{3,21}(\mathbb{R}) = G_{25}, (\alpha\neq 0)$	$i(\partial_x - \partial_{\bar x})$	$y^\alpha\partial_x + \bar y^{\bar\alpha}\partial_{\bar x}$	$x\partial_x + \bar x\partial_{\bar x} + [(1-a)/\alpha]\partial_y + [(1-a)/\bar\alpha]\partial_{\bar y}$	$L^{a\alpha}_{3,9}(\mathbb{C})$
	$L^a_{3,22}(\mathbb{R}) = G_{26}$	$i(\partial_x - \partial_{\bar x})$	$e^y\partial_x + e^{\bar y}\partial_{\bar x}$	$x\partial_x + \bar x\partial_{\bar x} + (1-a)\partial_y + (1-a)\partial_{\bar y}$	$L^a_{3,10}(\mathbb{C})$
	$L^f_{3,23}(\mathbb{R}) = G^f_{27}$	$i(\partial_x - \partial_{\bar x})$	$f(y)\partial_x + \bar f(\bar y)\partial_{\bar y}$	$x\partial_x + \bar x\partial_{\bar x}$	$L^f_{3,11}(\mathbb{C})$
$A_{3,3}$	$L_{3,24}(\mathbb{R}) = C_{15}$	$i(\partial_x - \partial_{\bar x})$	$i(\partial_y - \partial_{\bar y})$	$(x+y)\partial_x + (\bar x + \bar y)\partial_{\bar x} + y\partial_y + \bar y\partial_{\bar y}$	$L_{3,12}(\mathbb{C})$
	$L^\alpha_{3,25}(\mathbb{R}) = G_{28}$	$i(\partial_x - \partial_{\bar x})$	$\big(-(i/\alpha)\ln y\big)\partial_x + \big((i/\bar\alpha)\ln \bar y\big)\partial_{\bar x}$	$x\partial_x + \bar x\partial_{\bar x} + \alpha y\partial_y + \bar\alpha\bar y\partial_{\bar y}$	$L^\alpha_{3,13}(\mathbb{C})$
	$L^a_{3,26}(\mathbb{R}) = G_{29}$	$i(\partial_x - \partial_{\bar x})$	$i(y\partial_x - \bar y\partial_{\bar x})$	$x\partial_x + \bar x\partial_{\bar x} - \partial_y - \partial_{\bar y}$	$L_{3,14}(\mathbb{C})$
$A^a_{3,4}$ $0<a<\infty$	$L_{3,27}(\mathbb{R}) = A^a_7$	$i(\partial_x - \partial_{\bar x})$	$\partial_x + \partial_{\bar x}$	$(a+i)x\partial_x + (a-i)\bar x\partial_{\bar x} + \partial_y + \partial_{\bar y}$	$L_{2,4}(\mathbb{C})$
	$L^{a,b,c}_{3,28}(\mathbb{R}) = A^a_8$	$i(\partial_x - \partial_{\bar x})$	$\partial_x + \partial_{\bar x}$	$(a+i)x\partial_x + (b+ic)y\partial_y + \text{c.c.}$	$L^a_{2,5}(\mathbb{C})$
	$L^a_{3,29}(\mathbb{R}) = C^a_{16}$	$i(\partial_x - \partial_{\bar x})$	$i(\partial_y - \partial_{\bar y})$	$a(x\partial_x + y\partial_y) - y\partial_x + x\partial_y + \text{c.c.}$	$L^a_{3,8}(\mathbb{C})$
	$L^{a\lambda}_{3,30}(\mathbb{R}) = G^{a\lambda}_{30}, (\lambda\neq 0)$	$(i + y^{2i/\lambda})\partial_x + \text{c.c.}$	$(1 + iy^{2i/\lambda})\partial_x + \text{c.c.}$	$(a+i)x\partial_x + \lambda y\partial_y + \text{c.c.}$	$L^a_{3,9}(\mathbb{C})$
	$L^a_{3,31}(\mathbb{R}) = G^{a\lambda}_{31}$	$i(\partial_x - \partial_{\bar x})$	$[(1+e^{-2iy})/(1-e^{-2iy})]\partial_x + \text{c.c.}$	$\big(a + (1+e^{-2iy})/(1-e^{-2iy})\big)x\partial_x + \text{c.c.}$	$L^a_{3,10}(\mathbb{C})$
$A_{3,5}$ $(\mathrm{sl}(2,\mathbb{R}))$	$L_{3,32}(\mathbb{R})$	$\frac{1}{2}(-x\partial_x + y\partial_y + \text{c.c.})$	$\frac{1}{2}(x\partial_y + y\partial_x + \text{c.c.})$	$\frac{1}{2}(x\partial_y - y\partial_x + \text{c.c.})$	$L_{15}(\mathbb{C})$
	$L_{3,33}(\mathbb{R})$	$\frac{1}{2}(x\partial_x - y\partial_y + \text{c.c.})$	$\frac{1}{2}[(x + 1/x^3)\partial_y + y\partial_x + \text{c.c.}]$	$\frac{1}{2}[(-x + 1/x^3)\partial_y + y\partial_x + \text{c.c.}]$	$L_{16}(\mathbb{C})$
$A_{3,6}$ $(\mathrm{su}(2))$	$L_{3,34}(\mathbb{R})$	$(i/2)(-x\partial_x + y\partial_y) + \text{c.c.}$	$(i/2)(x\partial_y + y\partial_x) + \text{c.c.}$	$\frac{1}{2}(x\partial_y - y\partial_x + \text{c.c.})$	$L_{15}(\mathbb{C})$
	$L_{3,35}(\mathbb{R})$	$(i/2)(x\partial_x - y\partial_y) + \text{c.c.}$	$(i/2)[(x + 1/x^2)\partial_y + y\partial_x] + \text{c.c.}$	$\frac{1}{2}[(1 - x + 1/x^3)\partial_y + y\partial_x + \text{c.c.}]$	$L_{16}(\mathbb{C})$

The o(3,1) realization of $L_{3,16}(\mathbb{C})$ is

$$\begin{aligned}
L_1 &= \tfrac{1}{2}[(-x + 1/x^3)\partial_y + y\partial_x \\
&\quad + (-\bar x - 1/\bar x^3)\partial_{\bar y} + \bar y\partial_{\bar x}], \\
L_2 &= (i/2)[-x\partial_x + y\partial_y + \bar x\partial_{\bar x} - \bar y\partial_{\bar y}], \\
L_3 &= \tfrac{1}{2}[-(x + 1/x^3)\partial_y - y\partial_x \\
&\quad + (\bar x + 1/\bar x^3)\partial_{\bar y} + \bar y\partial_{\bar x}], \\
K_1 &= \tfrac{1}{2}[(-x + 1/x^3)\partial_y \\
&\quad + y\partial_x - (-\bar x + 1/\bar x^3)\partial_{\bar y} - \bar y\partial_{\bar x}], \\
K_2 &= \tfrac{1}{2}[x\partial_x - y\partial_y + \bar x\partial_{\bar x} - \bar y\partial_{\bar y}], \\
K_3 &= \tfrac{1}{2}[(x + 1/x^3)\partial_y + y\partial_x \\
&\quad + (-\bar x - 1/\bar x^3)\partial_{\bar y} + \bar y\partial_{\bar x}],
\end{aligned} \tag{5.26}$$

and the commutation relations are again (5.25). We thus obtain two different o(3) [or su(2)] algebras, namely L_1, L_2, and L_3 in both cases. Convenient choices of the o(2,1) [i.e., sl(2,$\mathbb{R}$)] subalgebras are $\{K_1, K_2, L_3\}$ in the first case and $\{K_2, K_3, L_1\}$ in the second. Four further real subalgebras are thus obtained and they are included in Table V.

VI. CONCLUSIONS AND PREVIEW OF FUTURE ATTRACTIONS

The main results of this paper are summed up in Tables III–V providing representatives of the conjugacy classes of one-, two-, and three-dimensional *real* subalgebras of the algebra of holomorphic vector fields in two complex variables, having constant divergence. The classification is performed under the pseudogroup P of biholomorphic transformations with constant Jacobian determinant.

The stage is now set for performing the actual symmetry reduction of the Ω equation and obtaining solutions and metrics. To give an example of the type of application we have in mind, consider the subalgebra $L_{3,4}(\mathbb{R})$ of Table V. Calculating its invariants in a standard manner[20–23] we find the expression

$$\Omega(x,\bar{x},y,\bar{y}) = \sqrt{x\bar{x}}F(\xi), \quad \xi = y + \bar{y}. \tag{6.1}$$

Substituting into the Ω equation $(*)$ we obtain an equation for $F(\xi)$,

$$F\ddot{F} - \dot{F}^2 = 4. \tag{6.2}$$

This nonlinear ordinary differential equation is invariant under translations and dilations, and can hence easily be solved. Substituting the solution back into (6.1), we obtain

$$\Omega = (1/K\sqrt{x\bar{x}})\cosh 2K(y + \bar{y} - c), \tag{6.3}$$

where K and c are integration constants. From this expression for Ω we obtain the metric tensor

$$ds^2 = \cosh(y+\bar{y})((1/4\sqrt{x\bar{x}})dx\,d\bar{x} + \sqrt{x\bar{x}}\,dy\,d\bar{y}) + \sinh(y+\bar{y})(\tfrac{1}{2}\sqrt{(\bar{x}/x)}\,dx\,d\bar{y} + \tfrac{1}{2}\sqrt{(x/\bar{x})}\,d\bar{x}\,dy). \tag{6.4}$$

A straightforward curvature computation shows that this is the flat metric on $\mathbb{R}^4$. In Part II we shall use all obtained subalgebras in a similar manner and describe the metrics obtained. For example the su(2) algebra $L_{3,34}(\mathbb{R})$ gives the well-known Eguchi–Hanson metric.

ACKNOWLEDGMENTS

One of the authors (C.P.B.) was supported in part by NSF Grants DMS-8508950 and 8875581. The other author (P.W.) was supported in part by research grants from NSERC of Canada and the "Fonds FCAR du Gouvernment du Québec."

[1]E. Cartan, Ann. Ec. Normale **26**, 93 (1909) [Ouevres Complète, Vol. II, 857, Editions CNRS, Paris, 1984].
[2]R. Penrose, Gen. Relativ. Gravit. **7**, 31 (1976).
[3]E. T. Newman, Gen. Relativ. Gravit. **7**, 107 (1976).
[4]J. F. Plebanski, J. Math. Phys. **16**, 2396 (1975).
[5]E. Calabi, Proc. Symp. Pure Math. **XV**, 223 (1970).
[6]E. Calabi, Proc. Symp. Pure Math. **XXVII**, 17 (1975).
[7]E. Calabi, Ann. Sci. Ec. Norm. Sup. **12**, 269 (1979).
[8]C. P. Boyer and J. F. Plebanski, J. Math. Phys. **18**, 1022 (1977).
[9]P. J. Olver, *Applications of Lie Groups to Differential Equations* (Springer, New York, 1986).
[10]P. Winternitz, *Lecture Notes in Physics*, Vol. 189 (Springer, Berlin, 1983), p. 263.
[11]S. W. Hawking, Phys. Lett. A **60**, 81 (1977).
[12]G. W. Gibbons and S. W. Hawking, Phys. Lett. B **78**, 430 (1978).
[13]M. J. Perry, in *Seminar on Differential Geometry* (Princeton U.P., Princeton, 1982), p. 603.
[14]I. M. Singer and S. Sternberg, J. Analyse Math. **XV**, 1 (1965).
[15]C. P. Boyer and J. D. Finley III, J. Math. Phys. **23**, 1126 (1982).
[16]S. Lie and F. Engel, *Theorie der Transformationgruppen* (Teubner, Leipzig, 1893) Vol. 3, 71–73; E. Cartan, Ann. Ec. Norm. **25**, 57 (1908) [Ouevres Complète, Vol. II, p. 719, Editions CNRS, Paris, 1984].
[17]G. W. Gibbons and C. N. Pope, Commun. Math. Phys. **66**, 267 (1979).
[18]N. Jacobson, *Lie Algebras* (Interscience, New York, 1962).
[19]G. M. Mubarakzyanov, Izv. Vys. Uiebn. Zaved. Mat. **1**, 114, **3**, 99, **4**, 104 (1963).
[20]J. Patera, R. T. Sharp, P. Winternitz, and H. Zassenhaus, J. Math. Phys. **17**, 986 (1976).
[21]J. Patera, P. Winternitz, and H. Zassenhaus, J. Math. Phys. **16**, 1597, 1615 (1975); **17**, 717 (1976).
[22]J. Patera, R. T. Sharp, P. Winternitz, and H. Zassenhaus, J. Math. Phys. **18**, 2259 (1977).
[23]A. M. Grundland, J. Harnad, and P. Winternitz, J. Math. Phys. **25**, 791 (1984).

The gravitational field of a spinning pencil of light

Nikolai Vsevolodovich Mitskievic
Department of Theoretical Physics, Patrice Lumumba People's Friendship University, Moscow, Union of Soviet Socialist Republics

Krishnadeva K. Kumaradtya
227-11 108th Avenue, Queens Village, New York 11429

(Received 7 March 1988; accepted for publication 4 January 1989)

An exact solution of Einstein's equations in a vacuum (outside of singularities), belonging to Kundt's class and Petrov type N, is interpreted as the metric of a spinning pencil of light (a linear source infinitely extended in one direction and moving with the speed of light). It is shown that the gravitational fields of two parallel pencils of light do not interact with each other, i.e., the superposition of the metrics of two parallel pencils of light is an exact solution of Einstein's equations in a vacuum.

I. INTRODUCTION

An approximate (weak field) metric for a pencil of light was first found by Tolman *et al.*[1] in 1934. They discovered a remarkable property of this metric, namely that the field of two parallel (not antiparallel) pencils of light when added together give a solution of Einstein's equations in the same approximation, i.e., they do not interact. One of us noticed recently that a special case of the Peres gravitational wave should be interpreted as an exact solution for the gravitational field of a pencil of light.[2] The fields of two parallel pencils of light have the property of exact superposition. But since these are vacuum solutions outside of singularities, how could one infer that the source represents a pencil of light?

We define a pencil of light as a linear flow of energy propagating along itself with the speed of light, thus not necessarily connected with a real beam of light, especially if one is talking about an infinitesimally thin pencil of light. The most characteristic property of a pencil of light is that its linear energy density is equal to the absolute value of its linear momentum density. We show this explicitly for the case of the Peres wave using Einstein's equations. Another way to come to a physical interpretation of the gravitational field under question is to consider the action of this field on test particles and if possible, the interaction of two or more identical sources. This action can be interpreted in some cases as a dragging phenomenon since it is due to the fact that the source's motion cannot be transformed away because it corresponds to the fundamental velocity (or in other, more conventional cases, to a rotational motion of the source, e.g., in the Kerr field).

In this paper we consider the possibility of the combination of both the luminal and rotational motions. An extended thin rectilinear source of this kind is a spinning pencil of light (SPL). A strange property of this source emerges: no SPL can exist if its angular momentum (spin) does not depend on the retarded time (i.e., on time and the coordinate along the symmetry axis z). The property of superposition continues to exist and an interplay of dragging effects occurs. The general metric (solution I) of a SPL is of Petrov type N and it belongs to Kundt's class (see Ref. 3), but we prefer to derive it here in some detail, especially since the identification of the metrics is not easy. As a by-product we come to solution II (of type II or D if two arbitrary functions vanish). In Sec. II we review the results of Ref. 2 concerning the stationary Peres wave as a pencil of light field. We give the derivation of solutions I and II in Sec. III. We study the effects due to dragging in Sec. IV, the superposition of metrics in Sec. V, and in Sec. VI we summarize the reasons for interpreting solution I as the gravitational field of a SPL.

II. PERES WAVE AS A FIELD OF A PENCIL OF LIGHT (REF. 2)

The metric of a Peres wave[4] is

$$ds^2 = \eta_{\mu\nu}\, dx^\mu\, dx^\nu - 2H\,[dt - dz]^2, \tag{2.1}$$

where the Minkowski metric can be written as

$$\eta_{\mu\nu}\, dx^\mu\, dx^\nu = dt^2 - dx^2 - dy^2 - dz^2$$
$$= dt^2 - d\rho^2 - \rho^2\, d\varphi^2 - dz^2$$

and

$$H = H(t - z;\rho;\varphi)$$

is an arbitrary function of the retarded time $t - z$ satisfying the two-dimensional flat space Laplace equation

$$\Delta_2 H = \frac{\partial^2 H}{\partial \rho^2} + \frac{1}{\rho}\frac{\partial H}{\partial \rho} + \frac{1}{\rho^2}\frac{\partial^2 H}{\partial \varphi^2} = 0$$

(except at singular points or lines). We shall consider the simplest case of the metric (2.1) when

$$H = -(k/2)\ln \sigma\rho. \tag{2.2}$$

In order to determine the physical meaning of the constant k we use one of Einstein's equations, namely,

$$R_{(0)(0)} - \tfrac{1}{2} g_{(0)(0)} R = -\varkappa T_{(0)(0)},$$

where $T_{(0)(0)} = \epsilon$ (the energy density) in an orthonormalized frame. We take for the latter

$$\theta^{(\alpha)} = dx^\alpha - H l^\alpha (dt - dz),$$
$$l^\alpha \partial_\alpha = \partial_t + \partial_z, \quad l_\alpha\, dx^\alpha = dt - dz,$$
$$l_\alpha l^\alpha = 0.$$

Then it can be found[5] that in this frame

$$R_{(\alpha)(\beta)} = (H l_\alpha l^\sigma)_{,\sigma,\beta} + (H l_\beta l^\sigma)_{,\sigma,\alpha} - H(l_\alpha l_\beta)_{,\sigma,\tau}\eta^{\sigma\tau},$$

 0022-2488/89/051095-05$02.50

so that if $H_{,t} = H_{,z} = 0$, in fact when $\partial H/\partial(t+z) = 0$, then we have $R = 0$,

$$R_{(0)(0)} = -H_{,\sigma,\tau}\eta^{\sigma\tau} = \left(\frac{\partial}{\partial x^2} + \frac{\partial}{\partial y^2}\right) H = \Delta_2 H.$$

Hence

$$\epsilon = (1/\varkappa)\Delta_2 H = (k/2\varkappa)\Delta_2 \ln \sigma\rho = (\pi k/\varkappa)\delta(x)\delta(y).$$

Taking

$$\epsilon = \epsilon_0 \delta(x)\delta(y),$$

where ϵ is the linear density on the z axis, we find finally

$$k = \epsilon_0 \varkappa/\pi = 8\gamma\epsilon_0, \tag{2.3}$$

with γ being the Newtonian gravitational constant. Similar consideration gives the same value for the momentum density in the z direction, thus meaning that the source of the Peres wave field is moving in the positive z direction with the speed of light. This is the reason why the source should be understood as an example of the pencil of light.

The Peres wave is indeed a stationary object, not a gravitational wave, but a field of a stationary pencil of light. It has no horizon, but the whole space-time has a property of the ergosphere: the field is only locally stationary (as it is for the case of the Kerr field in the ergosphere) since the Killing vector ∂_t is timelike only if $H < \frac{1}{2}$. At the same time another Killing vector ∂_z is spacelike only if $H > -\frac{1}{2}$, and we have the conventional definition of cylindrical symmetry only in the band $-\frac{1}{2} < H < \frac{1}{2}$. However, we can combine ∂_t and ∂_z and get Killing vectors that are good in shifted bands of the values of H. This corresponds to a coordinate transformation

$$t = (1-L)t' + Lz',$$

$$z = -Lt' + (1+L)z',$$

which leads to

$$H' = -4\gamma\epsilon_0 \ln \sigma'\rho,$$

with σ' corresponding to

$$L = -4\gamma\,\epsilon_0 \ln(\sigma'/\sigma),$$

as it should be for a locally stationary space-time. This transformation of coordinates is unnecessary if one takes retarded and advanced time coordinates lying always on the light cone. In this case the field of a pencil of light (or of the Peres wave which is no wave whatsoever) becomes

$$ds^2 = 2\,dv(du - 2H) - d\rho^2 - \rho^2\,d\varphi^2, \tag{2.4}$$

where $v = (1/\sqrt{2})\ (t-z)$ and $u = (1/\sqrt{2})\ (t+z)$.

Another approach to the interpretation of the Peres wave as the field of a pencil of light is connected with the dragging phenomenon. Dragging is a general property of all stationary gravitational fields, and it reflects their nature similar to that of magnetic fields. Therefore we call them quasimagnetic fields. Dragging occurs when the source of the gravitational field performs such a motion that it cannot be globally (or even locally, as it is in our case) transformed away. No observer can indeed move with the speed of light. The dragging manifests itself as a tendency of all test particles to accelerate in the direction of the pencil of light's motion (in addition to the free fall acceleration onto the z axis). We mean here under the acceleration of course not a nonzero absolute (covariant) derivative, but only an ordinary one, d^2s/dt^2, where z and t coordinates are determined by the Killing congruences inside the above-mentioned band. The proof of this statement is a particular case of a more general consideration given below (see Sec. IV). It is worth mentioning here that only a photon moving along the pencil of light in the same direction in which the energy of the pencil propagates does not experience any change in its motion (it does not even fall onto the singularity)—no interaction exists in this extreme case. This is most directly seen from the fact that the one-form dv is both null and exact; thus it is geodesic, being exactly the four-velocity form of such a photon. Similarly, since the field equation for H is linear, not only test photons but parallel (not antiparallel!) pencils of light do not interact in an exact sense[6]; a generalization of this observation is also discussed below (Sec. V). All these conclusions are exact results corresponding to the approximate ones found already in 1931 by Tolman, Ehrenfest, and Podolsky.[1,7]

III. EINSTEIN'S EQUATIONS AND TWO FAMILIES OF METRICS

In this section we give the derivation of the metric for the gravitational field of a SPL. Simultaneously we obtain another quite interesting metric. Our choice for the metric of a gravitational field of a SPL is the following:

$$ds^2 = 2e^{\alpha}\,dv(du + F\,dv + G\,d\varphi) - e^{2\beta}\,d\rho^2 - \rho^2\,d\varphi^2. \tag{3.1}$$

Here α, β, F, and G are functions of the retarded time v and the cylindrical radial coordinate ρ, the corresponding derivatives being denoted by the dot and the prime. This metric is a natural generalization of the metric (2.1) for a nonspinning pencil of light, with a new term $G\,d\varphi$ which leads to additional rotation of the covector field $(du + F\,dv + G\,d\varphi)$. When $G = 0$, the rotation was a manifestation of the source's motion along the z axis (i.e., $\rho = 0$) with the speed of light; now $G \neq 0$, a rotation about the same axis can be expected. As we shall see below, this is exactly the case.

A natural choice of the tetrad

$$\theta^{(0)} = e^{\alpha}\,dv,\quad \theta^{(1)} = du + F\,dv + G\,d\varphi,$$
$$\theta^{(2)} = e^{\beta}\,d\rho,\quad \theta^{(3)} = \rho\,d\varphi, \tag{3.2}$$

brings the metric (3.1) into the form

$$ds^2 = 2\theta^{(0)}\theta^{(1)} - \theta^{(2)2} - \theta^{(3)2}. \tag{3.3}$$

In this basis the independent nonzero components of the Riemann curvature tensor are

$$R_{(0)(1)(0)(1)} = -(\alpha'^2/4)e^{-2\beta},$$

$$R_{(0)(1)(0)(2)} = (\dot\alpha'/2)e^{-\alpha-\beta} - (\alpha'/2)\dot\beta e^{-\alpha-\beta},$$

$$R_{(0)(1)(0)(3)} = -(G'\alpha'/4\rho)e^{-2\beta},$$

$$R_{(0)(2)(1)(2)} = -(\alpha'' + \alpha'^2/2 - \alpha'\beta')(e^{-2\beta}/2),$$

$$R_{(0)(2)(0)(2)} = -(F'' - F'\beta' + F'\alpha')e^{-2\beta-\alpha} - (G'^2/4\rho^2)e^{-2\beta} + (\ddot\beta + \dot\beta(\dot\beta - \dot\alpha))e^{-2\alpha},$$
$$R_{(0)(2)(0)(3)} = (\alpha'\dot G + \dot G - G'\dot\beta - 2G'/\rho)(e^{-\alpha-\beta}/2\rho),$$
$$R_{(0)(3)(0)(3)} = -(F'/\rho)e^{-\alpha-\beta} - (G'^2/4\rho^2)e^{-2\beta},$$
$$R_{(0)(3)(1)(3)} = -\alpha' e^{-2\beta}/2\rho,$$
$$R_{(2)(3)(0)(2)} = (G'' + \tfrac{3}{2}G'\alpha' - G'\beta' - G'/\rho)(e^{-2\beta}/2\rho),$$
$$R_{(2)(3)(0)(3)} = -(\dot\beta/\rho)e^{-2\beta},$$
$$R_{(2)(3)(2)(3)} = -(\beta'/\rho)e^{-2\beta}.$$

Consequently we have the nontrivial components of the Ricci tensor being

$$R_{(0)(0)} = -(F'' - F'\beta' + F'\alpha' + F'/\rho)e^{-2\beta-\alpha} + (\ddot\beta + \dot\beta^2 - \dot\alpha\dot\beta)(e^{-2\alpha}/2) - (G'^2/2\rho^2)e^{-2\beta},$$
$$R_{(0)(1)} = -(\alpha'' + \alpha'^2/2 - \alpha'\beta' + \alpha'/\rho)(e^{-2\beta}/2),$$
$$R_{(0)(2)} = (\dot\alpha'/2 - \alpha'\dot\beta)(e^{-\alpha-\beta}/2) - (\dot\beta/\rho)e^{-2\beta},$$
$$R_{(0)(3)} = -(G'\alpha' + G''/2 - (G'/2)\beta' - G'/2\rho)(e^{-2\beta}/\rho),$$
$$R_{(2)(2)} = (\alpha'' - \alpha'\beta' - \beta'/\rho)e^{-2\beta},$$
$$R_{(3)(3)} = (\alpha' - \beta')(e^{-2\beta}/\rho).$$

However, since we are searching for a solution of Einstein's equations in a vacuum, all the components of the Ricci tensor must be equal to zero. Solving this system of differential equations we can find the functions α, β, F, and G. It is easy to show that there exist two possible families of exact solutions to these equations, forming together a general vacuum solution for the metric (3.1):

Solution I:
$$ds^2 = 2\,dv(du + (k\ln\sigma\rho - \tfrac{1}{2}f^2\rho^2)dv + (g + f\rho^2)d\varphi) - \rho^2\,d\varphi^2 - d\rho^2. \tag{3.4}$$

Solution II:
$$ds^2 = \frac{2\,dv}{\rho^4}(du + (r\ln\sigma\rho - \tfrac{1}{2}d^2\rho^6)dv + (d\rho^6 + h)d\varphi) - \rho^2 d\varphi^2 - \frac{d\rho^2}{\rho^8}. \tag{3.5}$$

In both of the above solutions the functions k, g, f, r, d, and h are arbitrary functions of the variable v. Solutions I and II can be simplified by transforming away the functions $f(v)$ and $d(v)$, respectively. Taking $d\tilde\varphi = d\varphi - f(v)dv$ in (3.4) we can write solution I as

$$ds^2 = 2\,dv(du + k\ln\sigma\rho\,dv + g(v)d\varphi) - \rho^2\,d\varphi^2 - d\rho^2.$$

Similarly using the transformation $d\tilde\varphi = d\varphi - d(v)dv$, solution II can be written as

$$ds^2 = \frac{2\,dv}{\rho^4}(du + k\ln\sigma\rho\,dv + h\,d\varphi) - \rho^2\,d\varphi^2 - \frac{d\rho^2}{\rho^8}.$$

As to the algebraic classification of these solutions the multiple principal null direction of the Weyl tensor is $\theta^{(0)}$ for both solutions. This vector field forms a geodesic nontwisting congruence without both expansion and shear. Solution I is of Petrov type N and solution II is of Petrov type II; however when $r = d = 0$, it degenerates into Petrov type D. All nontwisting and nonexpanding metrics of Petrov types D and N are known (see, e.g., Ref. 3), while this is true for only some metrics of type II. However, it is not easy at all to put these metrics explicitly into the standard form of the Kundt metrics to which they (including probably the solution II with r and $d \neq 0$) belong.

IV. DRAGGING IN THE FIELD OF A PENCIL OF LIGHT

We shall now study the motion of test particles in the field of solution I,

$$ds^2 = 2\,dv(du + k\ln\sigma\rho\,dv + g\,d\varphi) - d\rho^2 - \rho^2\,d\varphi^2.$$

The motion of test particles possessing no interior degrees of freedom is described by the geodesic equation, which can be conveniently written as

$$\frac{d}{d\lambda}\left(g_{\mu\nu}\frac{d\chi^\mu}{d\lambda}\right) = \frac{1}{2}g_{\alpha\beta,\nu}\frac{d\chi^\alpha}{d\lambda}\frac{d\chi^\beta}{d\lambda}. \tag{4.1}$$

Here λ is an affine parameter coinciding with the proper time on its world line for a massive particle. Since the metric coefficients do not depend on u and φ, the corresponding components of the equations (4.1) readily yield two first integrals of motion,

$$g_{vu}\frac{dv}{ds} = A = \text{const} > 0$$

and

$$g_{\varphi v}\frac{dv}{ds} + g_{\varphi\varphi}\frac{d\varphi}{ds} = -B = \text{const}.$$

The further integration of Eqs. (4.1) cannot be always performed exactly, and since we are not interested here in any perturbation expansions, another way of dealing with the dragging effect should be chosen. Let us consider only the tendency of the test particle's motion, i.e., the first nonzero higher derivative of its spatial coordinates when the initial state of motion is given. The initial state of motion is most naturally chosen as a state of rest (if our test particle is not massless). Then

$$\left(\frac{d\rho}{ds}\right)_0 = \left(\frac{d\varphi}{ds}\right)_0 = \left(\frac{dz}{ds}\right)_0 = 0, \tag{4.2}$$

but the problem is how to express z through u and v. In analogy with the Peres wave, we put here

$$z = (1/\sqrt{2})(u - v)$$

then

$$\left(\frac{dv}{ds}\right)_0 = \left(\frac{du}{ds}\right)_0.$$

Then from the two remaining components of Eqs. (4.1) and the two first integrals of motion, we have

$$\left(\frac{d^2\rho}{ds^2}\right)_0 = \frac{-kA^2}{\rho}, \tag{4.3}$$
$$\left(\frac{d^2z}{ds^2}\right)_0 = \frac{-g\dot gA^2}{\sqrt{2}\rho^2} - \frac{A^2\dot k\ln\sigma\rho}{\sqrt{2}}, \tag{4.4}$$
$$\left(\frac{d^2\varphi}{ds^2}\right)_0 = \frac{\dot gA^2}{\rho^2}, \tag{4.5}$$

where the right-hand side quantities are taken at the initial point of the test particle's world line. In Sec. II we have seen that $k = 8\gamma\epsilon_0 > 0$ (the linear energy density of the source must be positive), so that (4.2) describes an attraction of the test particle to the pencil of light. The noncovariant "acceleration" in the directions of z and φ, i.e., (4.4) and (4.5), can be either positive or negative depending on the signs of g, $\dot{g}$, and $\dot{k}$ and the relationship between the two right-hand terms in (4.4) (if we do not choose σ so that ρ_0 would correspond to the middle of the $-1 < k < 1$ band, cf. Sec. II). The dragging in the φ direction has the sign coinciding with that of $\dot{g}$, and this means that not the function g, but its first derivative is directly connected with the angular momentum of the SPL which is responsible for dragging in the φ direction. Moreover, if g was constant, it could be transformed away by merely introducing $\tilde{u} = u + g\varphi$, so that only $\dot{g}$ can have a physical significance.

We consider now the motion of a lightlike particle. The initial state of motion should be chosen now in a different way since conditions (4.2) are incompatible with $ds^2 = 0$. We shall take only the first two conditions of (4.2),

$$\left(\frac{d\rho}{d\lambda}\right)_0 = \left(\frac{d\varphi}{d\lambda}\right)_0 = 0. \tag{4.6}$$

Putting them into ds^2 we have

$$\left(\frac{dv}{d\lambda}\right)_0 \left(\frac{du}{d\lambda} + k \ln \sigma\rho \frac{dv}{d\lambda}\right)_0 = 0,$$

which leads to two possibilities

$$\left(\frac{dv}{d\lambda}\right)_0 = 0 \tag{4.7}$$

and

$$\left(\frac{du}{d\lambda} + k \ln \sigma\rho \frac{dv}{d\lambda}\right)_0 = 0. \tag{4.8}$$

The first one is that which we have already mentioned in Sec. II. This geodesic corresponds to the first integral of motion $A = 0$. The alternative case (4.8) does not admit $(dv/d\lambda)_0 = 0$ since such a case would correspond to a world point and not a world line. Combining the condition (4.8) with the geodesic equation (including the first integrals of motion), we come to the relations

$$\left(\frac{d^2\rho}{d\lambda^2}\right)_0 = -\frac{kA^2}{\rho},$$

$$\left(\frac{d^2\varphi}{d\lambda^2}\right)_0 = \frac{\dot{g}A^2}{\rho^2},$$

$$\left(\frac{d^2u}{d\lambda^2}\right)_0 = -\frac{g\dot{g}A^2}{\rho^2} - kA^2 \ln \sigma\rho.$$

These accelerations are in fact the same as in the case of a massive test particle, though we have here u instead of z. We see that a photon moving parallel to a SPL does not interact with it (in particular it does not feel dragging in the φ direction), while a photon moving antiparallel to it both falls onto the pencil of light and starts to rotate in the φ direction.

V. SUPERPOSITION OF METRICS

In the foregoing section we saw that a lightlike particle, moving parallel to a pencil of light, does not interact with it. The question that naturally arises is what happens when another pencil of light (a linear flow of the lightlike matter) is moving parallel to the first pencil of light. As we shall show below, a most remarkable property of parallel pencils of light is their additivity, i.e., parallel pencils of light do not interact with each other.

The metric of a SPL moving along the z axis at the locus $\rho = 0$ can be written as

$$ds^2 = 2\,dv(du + k \ln \sigma\rho\, dv + g\, d\varphi) - d\rho^2 - \rho^2\, d\varphi^2. \tag{5.1}$$

The metric of another SPL moving parallel to the z axis at a distance a in the x direction takes the form

$$ds^2 = 2\,dv\Big[du + \frac{k_2}{2} \ln \sigma_2(\rho^2 + 2a\rho\cos\varphi + a^2)dv + \frac{g_2 a \sin\varphi\, d\rho}{(\rho^2 + 2a\rho\cos\varphi + a^2)} + \frac{g_2(\rho^2 + \rho a\cos\varphi)}{(\rho^2 + 2a\rho\cos\rho + a^2)}\, d\varphi\Big] - d\rho^2 - \rho^2\, d\varphi^2. \tag{5.2}$$

Then it is natural to think that the superposition of two parallel pencils of light can be written as

$$ds^2 = 2\,dv\Big[du + \Big(k_1 \ln \sigma_1\rho + \frac{k_2}{2} \ln \sigma_2(\rho^2 + 2a\rho \times\cos\varphi + a^2)dv + \frac{g_2 a\sin\varphi\, d\rho}{(\rho^2 + 2a\rho\cos\varphi + a^2)} + \Big(g_1 + \frac{g_2(\rho^2 + \rho a\cos\varphi)}{(\rho^2 + 2a\rho\cos\varphi + a^2)}\Big)\, d\varphi\Big] - d\rho^2 - \rho^2\, d\varphi^2. \tag{5.3}$$

Using the following notation,

$$F = k_1 \text{ in } \sigma_1\rho + (k/2) \ln \sigma_2(\rho^2 + 2a\rho\cos\varphi + a^2),$$

$$G = g_1 + \frac{g_2(\rho^2 + \rho a\cos\varphi)}{(\rho^2 + 2a\rho\cos\varphi + a^2)},$$

$$H = \frac{g_2 a\sin\varphi}{(\rho^2 + 2a\rho\cos\varphi + a^2)}, \tag{5.4}$$

we can rewrite the metric (5.3) as

$$ds^2 = 2\,dv(du + F\,du + G\,d\varphi + H\,d\rho) - d\rho^2 - \rho^2\,d\varphi^2.$$

The most simple tetrad is

$$\theta^{(0)} = dv, \quad \theta^{(1)} = du + F\,dv + G\,d\varphi + H\,d\rho,$$
$$\theta^{(2)} = d\rho, \quad \theta^{(3)} = \rho\, d\varphi.$$

Such a choice of the tetrad gives the following nontrivial components of the Ricci tensor:

$$R_{(0)(0)} = -\left(F_{,\rho,\rho} + \frac{F_{,\rho}}{\rho} + \frac{F_{,\varphi,\varphi}}{\rho^2}\right) + \left(\frac{G_{,v,\rho}}{\rho^2} + \frac{H_{,v}}{\rho} - H_{,v,\varphi}\right) - \frac{2}{\rho^2}\left(\frac{G_{,\rho}}{2} - \frac{H_{,\varphi}}{2}\right)^2_1,$$

$$R_{(0)(2)} = (G_{,\rho} - H_{,\varphi})_{,\varphi}/2\rho^2$$

$$R_{(0)(3)} = -(G_{,\rho} - H_{,\varphi})_{,\rho}/2\rho + (G_{,\rho} - H_{,\varphi})/2\rho^2.$$

However, the substitution of (5.4) shows that

$$R_{(0)(0)} = R_{(0)(2)} = R_{(0)(3)} = 0.$$

Consequently the metric (5.3) is an exact solution of Einstein's equations for a vacuum. It can be shown that such a conclusion is valid not only for two parallel pencils of light, but for any number of parallel pencils of light.

VI. PHYSICAL INTERPRETATION OF SOLUTION I

We now summarize some results of this paper beginning with solution I. Our main aim was to establish the physical meaning of the cylindrically symmetric vacuum solutions of Einstein's equations that are traditionally interpreted as a kind of gravitational wave (in particular, the Peres wave). We consider the evidence presented here to be very convincing that these "waves" are in fact exterior gravitational fields of pencils of light, i.e., of sources extended only in one direction and propagating in this direction with the speed of light. Several reasons support this conclusion. First, both a lightlike particle and a pencil of light do not interact with a pencil of light if they are moving in the same direction as the latter. Here the "direction of motion" is given unambiguously by the anisotropy of the mixed $dt\,dz$ term in ds^2, while such a direction for a photon is obvious. Second, in the same direction the dragging effect acts on all massive test particles, and this dragging cannot be transformed away by coordinate transformations, thus proving the velocity of the source to be the fundamental one. Third, an integration of the left-hand side of Einstein's equations shows that their right-hand side contains for T_{tt} and T_{tz} (in the chosen tetrad basis) two-dimensional delta-function terms along the z axis equal to each other, so that the linear energy density of the source is equal to the absolute value of its linear momentum density. Our opinion is that the Petrov type N metrics (which is the case here) admit either pure wave solutions (which is the conventional interpretation of these metrics) or fields of lightlike sources (which is the case under consideration). And, finally, as early as 1931 Tolman *et al.*[1,7] found an approximate solution for the field of a pencil of light fully compatible with the simplest special case of the Peres wave, a fact first previously noticed by one of us.[2] Along these lines we also established that the parameter k in the metric (this parameter may also depend arbitrarily on the retarded time) represents the linear energy density of the source (the pencil of light).

We studied here in the realm of solution I more general pencil of light metrics with spinning sources, thus leading to an extra rotational dragging around the z axis. The two dragging effects (the z and φ the ones) do mutually interact: it can be shown that in the coordinates t, z, ρ, and φ (our calculations were given in u, v, ρ, and φ coordinates) no vacuum solutions of the symmetry under consideration and the mixed $dt\,d\varphi$ and $dt\,dz$ terms (but without $dz\,d\varphi$) exist at all. Another interesting fact is that the function g in ds^2, which is due to the source's rotation, is not allowed to degenerate to a constant without making the rotation transformable away. This means that a pencil of light cannot spin with a constant linear density of the angular momentum, which should be evidently understood as a too strong influence of the gravitational field of an infinitely extended linear homogeneous source on the global geometry of space-time. One could fancy here that a linearly homogeneous rotation of such a source should lead, in accordance with Mach's principle, to a rotation of the universe itself, but in contrast to, e.g., the Kerr metric where the rotation falls off quite quickly, in a cylindrically symmetric case it behaves very differently and even degenerates to a purely coordinate rotation (thus becoming transformable away). The noninteraction property of two or more pencils of light (which move in one and the same direction) is closely connected with the additivity of their gravitational fields which we established here as (5.3). It is worth mentioning that the resulting metric retains only one Killing vector ∂_u and becomes explicitly dependent on φ. These results hold when the pencil(s) of light is (are) spinning. From here it is natural to infer that also a spinning test photon should not interact with an arbitrarily spinning pencil of light when their directions of propagation coincide. The Papapetrou–Mathisson equation of motion for such a photon[8] includes the curvature tensor, thus not reducing to the geodesic equation in general, but in this special case such a reduction is to be expected.

As to solution II we did not consider it here closely. However, it is important to note that when $r = h = 0$ in solution II, this metric coincides with that of Levi-Civita ($m = 2$, see Eq. 20.8 in Ref. 3), though in nonstandard coordinates. Thus the full metric solution II is a generalization of the Levi-Civita metric (for $m = 2$).

ACKNOWLEDGMENTS

The authors express their gratitude to Dr. G. A. Alekseev, Dr. A. A. Garcia Diaz, and the members of the research group of Professor E. Schmutzer for their stimulating interest in this work and helpful suggestions.

[1] R. C. Tolman, P. Ehrenfest, and B. Podolsky, Phys. Rev. **37**, 602 (1931).

[2] N. V. Mitskievic, "Gravitational field of a pencil of light," Experimentelle Technik der Physik, Report 29, No. 3, 1981, pp. 213–215.

[3] D. Kramer, H. Stephani, M. MacCallum, and E. Herlt, *Exact Solutions of Einstein's Field Equations* (VEB Deutscher, Berlin, 1980).

[4] A. Peres, Phys. Rev. Lett. **3**, 571 (1959).

[5] N. V. Mitskievic, *Rotation in General Relativity, Problems of Gravitation and Elementary Particles* (Atomizdat, Moscow, 1976), Vol. 7, pp. 15–34.

[6] The fact of linearity makes it possible to derive new potentials H from (2.2), by preforming a complex transformation of the (x,y) coordinates leaving the Laplacian unchanged, with a subsequent use of the real and imaginary parts of the resulting complex function H as potentials, the physical meaning of which should however be then studied *ab initio*. For example, we thus come to

$$H = k \ln [\sigma(\rho^4 + 2a\rho^2 \cos 2\varphi + a^2)],$$

$$H = k \tan^{-1}[2a\rho \cos \varphi / (\rho^2 - a^2)],$$

$$H = (k/\rho) \cos \varphi.$$

[7] R. C. Tolman, *Relativity, Thermodynamics and Cosmology* (Clarendon, Oxford, 1934), p. 274.

[8] E. N. Epikhin and N. V. Mitskievic, Acta Phys. Pol. **87**, 543 (1976).

Explicit description of ansatz E_n for the Ernst equation in general relativity

Kiyokazu Nagatomo
Department of Mathematics, Osaka University, Toyonaka, Osaka 560, Japan

(Received 29 July 1988; accepted for publication 4 January 1989)

For the Ernst equation, a hierachy of ansatz that generates determinantal solutions of the Ernst equation is proposed. The ansatz is described explicitly in the inverse scattering formalism and it is shown that the corresponding exact solutions are determinantal solutions that have been constructed by Kyriakopoulos [Phys. Rev. D **30**, 1158 (1984)] and Vein [Class. Quantum Gravit. **2**, 899 (1985)].

I. INTRODUCTION

In this paper we consider determinantal solutions of the Ernst equation, which are expressed in terms of cofactors of particular matrices. Exact solutions discussed here have been constructed by various methods.[1-3] However, an ansatz that produces these solutions has not been given. In the inverse scattering formalism, we present an explicit form of ansatz E_n for the Ernst equation and construct the corresponding exact solutions by using Riemann–Hilbert transform.

The Ernst equation[4] is a system of nonlinear differential equations for two unknown functions $f=f(z,\rho)$ and $e=e(z,\rho)$:

$$\begin{aligned} &f\nabla^2 f-(\partial_z f)^2-(\partial_\rho f)^2+(\partial_z e)^2+(\partial_\rho e)^2=0,\\ &f\nabla^2 e-2(\partial_z f\cdot\partial_z e+\partial_\rho f\cdot\partial_\rho e)=0, \end{aligned} \tag{1}$$

in $\mathbf{R}_z\times\mathbf{R}_\rho^+$ where $\nabla^2=\partial_\rho^2+(1/\rho)\partial_\rho+\partial_z^2$ and $\mathbf{R}_\rho^+=\{\rho\in\mathbf{R}:\rho>0\}$. This equation is closely related with the SDYM equation. In the static axially symmetric case, the Yang's equation—the SDYM equation in the R gauge of Yang—coincides with the Ernst equation.[5] This fact can be used to generate solutions of the Ernst equation. Fortunately a family of determinantal solutions of the Yang's equation have been constructed by ansatz due to Atiyah and Ward.[6,7] By using an explicit formula giving these solutions, Kyriakopoulos[1] has constructed a family of determinantal solutions. The question arises as to how ansatz for the Ernst equation is expressed in contrast to that of Atiyah–Ward for Yang's equation. This is the motivation for the present work. Our ansatz E_n has an expression similar to that of Atiyah–Ward. It is known that the Atiyah–Ward ansatz solutions are constructed by applying a particular Riemann–Hilbert transform to a trivial solution.[7] In a similar way, E_n ansatz solutions are obtained; however, the seed solution is nontrivial: $f=\rho^{-2n+1}$, $e=0$.

In Sec. II we review the inverse scattering method to the Ernst equation. The explicit description of an ansatz is stated in Sec. III and here the action of RHT is defined by solving a Riemann–Hilbert problem. Determinantal solutions corresponding to ansatz E_n are derived in Sec. IV.

The method of Vein,[2] which uses Bäcklund transform of Nakamura,[8] seems to be closely related to our method via Bäcklund transform for the SDYM equation. The relation between them is now under consideration. Finally we note that recently Candler and Freeman[3] have obtained two families of determinantal solutions by using bilinear representation of the Ernst equation.

II. INVERSE SCATTERING METHOD

Throughout this paper we assume $f\neq 0$, for the function f represents physically a nonvanishing coefficient of a Lorentz metric. Let us define a matrix $\tau(z,\rho)$ by

$$\tau=\begin{pmatrix} -\dfrac{f^2+e^2}{f} & \dfrac{e}{f}\\ \dfrac{e}{f} & -\dfrac{1}{f}\end{pmatrix}, \tag{2}$$

then Eq. (1) is represented in the matrix form

$$\partial_\rho(\rho\partial_\rho\tau\cdot\tau^{-1})+\partial_z(\rho\partial_z\tau\cdot\tau^{-1})=0, \tag{3}$$

$$\det\tau=1,\quad {}^t\tau=\tau. \tag{4}$$

Following Belinsky–Zakharov,[9] we rewrite Eq. (3) in terms of the matrices $U=\rho\partial_\rho\tau\cdot\tau^{-1}$ and $V=\rho\partial_z\tau\cdot\tau^{-1}$:

$$\partial_\rho U+\partial_z V=0,\quad \rho(\partial_z U-\partial_\rho V)+V+[U,V]=0. \tag{5}$$

Then we can see that a system of linear differential equations for $W=W(z,\rho,\lambda)$, $\lambda\in\mathbf{C}$,

$$D_1W=U\cdot W,\ D_2W=V\cdot W, \tag{6}$$

is involutive if and only if U and V satisfy Eq. (5), where $D_1=\lambda\partial_z+\rho\partial_\rho+2\lambda\partial_\lambda$, $D_2=\rho\partial_z-\lambda\partial_\rho$. A solution of Eq. (6) is said to be a wave function associated with $\tau(z,\rho)$. There are two important wave functions $W_+(z,\rho,\lambda)$ and $W_-(z,\rho,\lambda)$; the former is holomorphic in a neighborhood of $\lambda=0$ as a function of λ and the latter is holomorphic in a neighborhood of $\lambda=\infty$ and satisfies $W_-(z,\rho,\lambda)|_{\lambda=\infty}=1_2$. Equation (6) with $\lambda=0$ yields that $\rho\partial_\rho W_+(z,\rho,0)=U\cdot W_+(z,\rho,0)$ and $\rho\partial_z W_+(z,\rho,0)=V\cdot W_+(z,\rho,0)$. Thus $W_+(z,\rho,0)$ satisfies the same equation as τ and hence we can assume $W_+(z,\rho,0)=\tau(z,\rho)$. It is clear that the wave function $W_+(z,\rho,\lambda)$ has the following properties: (i) $W_+(z,\rho,\lambda)$ is invertible and holomorphic in a neighborhood of $\lambda=0$; (ii) $W_+(z,\rho,0)=\tau(z,\rho)$; (iii) $D_1W_+\cdot W_+^{-1}$ and $D_2W_+\cdot W_+^{-1}$ are independent of variable λ. The above conditions (i) and (iii) completely characterize wave function $W_+(z,\rho,\lambda)$. If a square matrix function $W_+(z,\rho,\lambda)$ satisfies the conditions (i) and (iii), then it follows from the property (iii) that $\tau(z,\rho)=W_+(z,\rho,\lambda)|_{\lambda=0}$ is a solution of Eq. (3). We note that another wave function $W_-(z,\rho,\lambda)$ will play an important role when we attempt to find W_+ with desired properties by solving a Riemann–Hilbert problem.

We give the simplest example, which is used in the next section.

 0022-2488/89/051100-03$02.50

Example: Let $\tau_0^{(s)}(z,\rho)$ be a solution of Eq. (3) defined by $\tau_0^{(s)}(z,\rho) = \mathrm{diag}(\rho^{-2s},\rho^{2s})$, $s\in\mathbf{R}$. The corresponding wave functions $\mathsf{H}_\pm^{(s)}(z,\rho,\lambda)$ are given by

$$\mathsf{H}_+^{(s)}(z,\rho,\lambda) = \mathrm{diag}(p^{-s},p^s), \quad \mathsf{H}_-^{(s)}(z,\rho,\lambda) = \mathrm{diag}(q^s,q^{-s}),$$

where $p = \rho^2 + 2z\lambda - \lambda^2$ and $q = 1 - 2z/\lambda - \rho^2/\lambda^2$.

III. DESCRIPTION OF ANSATZ

We first review the method of Riemann–Hilbert transform (RHT).[4,9] Let $W_+^{(0)}$ be the wave function of a seed solution of Eq. (3). The method of RHT is a way of finding a new solution of Eq. (3) whose wave function is expressed as $W_+ = X_+\cdot W_+^{(0)}$, where X_+ is a holomorphic function of λ in a neighborhood of $\lambda = 0$. Let $u(t)$ be a square matrix function that is invertible and holomorphic on $\mathbf{C}$. Solve the following Riemann–Hilbert problem for X_- and X_+ on a closed curve C in $\mathbf{C}_\lambda$ surrounding the origin:

$$X_+ = X_-\cdot W_+^{(0)}\cdot u(\lambda - 2z - \rho^2/\lambda)\cdot[W_+^{(0)}]^{-1} \quad \text{on } C,$$
$$X_-(z,\rho,\infty) = 1_2,$$

where X_+ and X_- are holomorphic inside and outside of C, respectively. Then $W_+ = X_+\cdot W_+^{(0)}$ has the desired properties (i) and (iii) of wave functions and hence a new solution of Eq. (3) is obtained by $\tau(z,\rho) = X_+(z,\rho,0)\cdot W_+^{(0)}(z,\rho,0)$.

Let us consider a class of solutions of Eq. (3) whose wave functions $W_\pm$ have the form

$$W_+ = X_+\cdot\mathsf{H}_+^{(n/2)}, \quad W_- = X_-\cdot\mathsf{H}_-^{(n/2)}, \tag{7}$$

where X_+ and X_- are holomorphic and invertible on $\mathbf{C}_\lambda$ and $\mathbf{C}_\lambda\cup\{\infty\}\setminus\{0\}$, respectively. Then the function $w = [W_-]^{-1}\cdot W_+$ is holomorphic except $\lambda\in[\lambda_-,\lambda_+]\cup\{\infty\}$, where $\lambda_\pm = z\pm\sqrt{z^2+\rho^2}$. Since by Eq. (6) the function w satisfies $D_1 w = D_2 w = 0$, there exists a holomorphic function $w(t)$ on $\mathbf{C}\setminus\mathbf{R}^+$ such that $w = w(\lambda - 2z - \rho^2/\lambda)$. Hence we obtain the important relation

$$W_+ = W_-\cdot w(\lambda - 2z - \rho^2/\lambda), \tag{8}$$

for $\lambda\in\mathbf{C}\setminus[\lambda_-,\lambda_+]$.

We now consider the case when a function w has the following form.

Ansatz E_n: The function $w(t)$ has the form

$$w(t) = \begin{pmatrix} 0 & \omega_n \\ -\omega_n^{-1} & t^n v(t)\end{pmatrix},$$

where $\omega_n = \exp(\sqrt{-1}n\pi/2)$, $n\in\mathbf{N}$ and $v(t)$ is an entire function on $\mathbf{C}$ such that $v(0)\neq 0$.

By means of the method of RHT, our ansatz E_n is expressed as follows. Consider the RHT with

$$W_+^{(0)} = \mathsf{H}_+^{(n/2)}, \quad u(t) = \begin{pmatrix} 0 & (-t)^n \\ -(-t)^n & v(t)\end{pmatrix}.$$

Then we can easily verify that the equation which defines this RHT coincides with Eq. (8).

We now look for $W_\pm$ corresponding to $w(t)$ of ansatz E_n. In terms of $X_\pm$, Eq. (8) is equivalent to

$$X_+(z,\rho,\lambda) = X_-(z,\rho,\lambda)\cdot\begin{pmatrix} 0 & \lambda^n \\ -\lambda^n & v(\lambda - 2z - \rho^2/\lambda)\end{pmatrix},$$
$$\text{for any } \lambda\in\mathbf{C}_\lambda\setminus\{0\}. \tag{9}$$

Since $w(\lambda - 2z - \rho^2/\lambda)$ is holomorphic and invertible on $\mathbf{C}_\lambda\setminus\{0\}$, there exist[10] uniquely $X_\pm$ which satisfy (9) and $X_-(z,\rho,\infty) = 1_2$.

Lemma: Let $X_\pm$ be solutions of the Riemann–Hilbert problem (9) such that $X_-(z,\rho,\infty) = 1_2$. Then (i) $\det X_\pm = 1$; (ii) the matrix $\tau(z,\rho) = X_+(z,\rho,0)\cdot\mathsf{H}_+^{(n/2)}(z,\rho,0)$ is a solution of Eq. (3) such that $\det\tau = 1$.

Proof: From Eq. (9) there follows $\det X_+ = \det X_-$. Since the left- and right-hand sides of this equality are holomorphic, respectively, on $\mathbf{C}_\lambda$ and $\mathbf{C}_\lambda\cup\{\infty\}\setminus\{0\}$, it follows that both sides are analytically continued to $\mathbf{C}_\lambda\cup\{\infty\}$, so they must be constant on $\mathbf{C}_\lambda$ by virtue of Liouville's theorem. Noting $X_-(z,\rho,\infty) = 1_2$, we have $\det X_\pm = 1$.

The second statement (ii) is proved in a standard manner as follows. From the definition (7) we have

$$D_1 W_+\cdot[W_+]^{-1} = D_1 X_+\cdot[X_+]^{-1} + X_+\begin{pmatrix} -n & 0 \\ 0 & n\end{pmatrix}\cdot[X_+]^{-1},$$
$$D_1 W_-\cdot[W_-]^{-1} = D_1 X_-\cdot[X_-]^{-1} + X_-\begin{pmatrix} n & 0 \\ 0 & -n\end{pmatrix}\cdot[X_-]^{-1}.$$

Therefore $D_1 W_+\cdot[W_+]^{-1}$ and $D_1 W_-\cdot[W_-]^{-1}$ are holomorphic, respectively, on $\mathbf{C}_\lambda$ and $\mathbf{C}_\lambda\cup\{\infty\}\setminus\{0\}$. On the other hand, operating D_1 on both sides of Eq. (8), we see that $D_1 W_+\cdot[W_+]^{-1} = D_1 W_-\cdot[W_-]^{-1}$ for $\lambda\in\mathbf{C}\setminus[\lambda_-,\lambda_+]$. Hence, again, Liouville's theorem shows that $D_1 W_+\cdot[W_+]^{-1}$ is dependent of the variable λ. Similarly we can prove that $D_2 W_+\cdot[W_+]^{-1}$ is also independent of λ. Consequently $\tau(z,\rho) = W_+(z,\rho,0)$ is a solution of Eq. (3). Finally, as a corollary of the first statement, we obtain $\det\tau = 1$. Thus the lemma is proved.

In the next section we shall give a formula that represents $\tau(z,\rho)$ by means of the function $v(t)$. If n is an odd number, then the matrix τ shall turn out to be automatically symmetric. Hence a solution of the Ernst equation is obtained by (2).

IV. EXPLICIT FORMULA

We now present an explicit formula that represents the solution $\tau(z,\rho)$ of Eq. (3) corresponding to ansatz E_n.

Theorem: Let $K = (k_{rs})_{r,s=0}^n$ be the matrix defined by

$$k_{rs} = \frac{1}{2\pi\sqrt{-1}}\oint_C \frac{d\lambda}{\lambda}\, v\left(\lambda - 2z - \frac{\rho^2}{\lambda}\right)\lambda^{r+s-n},$$

where the integration contour C encircles the origin. Let $\tilde{k}_{rs}$'s be the cofactors of K. Then:

(i) There exists a neighborhood of $(z,\rho) = (0,0)$ such that $\det K\neq 0$ and $\tilde{k}_{0n}\neq 0$.

(ii) The solution $\tau(z,\rho)$ of Eq. (3) corresponding to ansatz E_n is given by

$$\tau(z,\rho) = \begin{pmatrix} a_0 & b_0 \\ c_0 & d_0\end{pmatrix}\cdot\begin{pmatrix}\rho^{-n} & 0 \\ 0 & \rho^n\end{pmatrix},$$

where

$$b_0=\frac{\tilde{k}_{nn}}{\tilde{k}_{0n}},\quad d_0=\frac{\det K}{\tilde{k}_{0n}},$$

$$a_0=-\frac{\tilde{k}_{00}}{\det K}b_0+\frac{\tilde{k}_{n0}}{\det K},\quad c_0=-\frac{\tilde{k}_{00}}{\det K}d_0.$$

(iii) In the case where n is odd, τ is symmetric, whereas it is skew symmetric in the case where n is even.

Proof: (i) From the definition of $K=(k_{rs})$, we have

$$k_{rs}|_{z=\rho=0}=\frac{1}{2\pi\sqrt{-1}}\oint_C\frac{d\lambda}{\lambda}v(\lambda)\lambda^{r+s-n}.$$

Since the function $v(\lambda)$ is entire, it follows that

$$k_{rs}|_{z=\rho=0}=\begin{cases}0, & \text{for } r+s>n+1,\\ v(0), & \text{for } r+s=n.\end{cases}$$

This shows that $\det K|_{z=\rho=0}=\epsilon v(0)^{n+1}$ and $\tilde{k}_{0n}|_{z=\rho=0}=\epsilon v(0)^n$, where $|\epsilon|=1$. Hence there exists a neighborhood of $(z,\rho)=(0,0)$ such that $\det K\neq 0$ and $\tilde{k}_{0n}\neq 0$. (ii) Let us write the matrices occurring in (9) as

$$X_+=\begin{pmatrix}a & b\\ c & d\end{pmatrix},\ X_-=\begin{pmatrix}\alpha & \beta\\ \gamma & \delta\end{pmatrix},$$

where $ad-bc=\alpha\delta-\beta\gamma=1$ and $\alpha,\beta,\gamma,\delta$ are holomorphic as functions of λ on $\mathbf{C}\cup\{\infty\}\setminus\{0\}$ while a, b, c, d are holomorphic on $\mathbf{C}$. Then Eq. (9) is equivalent to

$$a=-\lambda^n\beta,\quad c=-\lambda^n\delta, \tag{10}$$

$$b=\lambda^{-n}\alpha+\beta v,\quad d=\lambda^{-n}\gamma+\delta v. \tag{11}$$

It follows from Eq. (10) that a and c are polynomials of λ of degree at most n:

$$a=\sum_{r=0}^n a_r\lambda^r,\quad c=\sum_{r=0}^n c_r\lambda^r.$$

The coefficient of λ^{-r} in the Laurent series of $b-\lambda^{-n}\alpha$, $d-\lambda^{-n}\gamma$ vanishes for $0<r<n$, so that

$$\oint_C\frac{d\lambda}{\lambda}(\beta v)\lambda^r=\oint_C\frac{d\lambda}{\lambda}(\delta v)\lambda^r=0,\quad 0<r<n.$$

Substituting $\beta=-a\lambda^{-n}$, $\delta=-c\lambda^{-n}$ into these equations, we obtain

$$\sum_{s=0}^n v_{r+s}a_s=\sum_{s=0}^n v_{r+s}c_s=0,\quad 0<r<n.$$

Further, $\alpha(z,\rho,\infty)=1$ and $\gamma(z,\rho,\infty)=0$, a fact that leads to

$$\sum_{s=0}^n v_{n+s}a_s=1,\quad \sum_{s=0}^n v_{n+s}c_s=0.$$

Let us denote the values of b and d at $\lambda=0$ by b_0 and d_0, respectively. Then we get

$$\sum_{s=0}^n v_s a_s=-b_0,\quad \sum_{s=0}^n v_s c_s=-d_0.$$

Hence we have the following linear equations for column vectors $(a_r)_{r=0}^n$ and $(c_r)_{r=0}^n$:

$$K\cdot(a_r)={}^t(-b_0,0,...,0,1),\quad K\cdot(c_r)={}^t(-d_0,0,...,0,0), \tag{12}$$

where $K=(k_{rs})_{r,s=0}^n$, $k_{rs}=v_{r+s}$. Since the matrix K is invertible,

$$a_r=-\frac{\tilde{k}_{0r}}{\det K}b_0+\frac{\tilde{k}_{nr}}{\det K},\quad c_r=-\frac{\tilde{k}_{0r}}{\det K}d_0,\quad 0\leqslant r\leqslant n. \tag{13}$$

Further we obtain $a_n=0$ and $c_n=-1$ from Eq. (10) by using the conditions $\beta(z,\rho,\infty)=0$ and $\delta(z,\rho,\infty)=1$. Consequently Eq. (13) with $r=n$ implies

$$b_0=\frac{\tilde{k}_{nn}}{\tilde{k}_{0n}},\quad d_0=\frac{\det K}{\tilde{k}_{0n}}.$$

(iii) Since $b_0=\tilde{k}_{nn}/\tilde{k}_{0n}$ and $c_0=-\tilde{k}_{00}/\tilde{k}_{0n}$, we have to show

$$\rho^n\tilde{k}_{nn}=(-1)^n\rho^{-n}\tilde{k}_{00}. \tag{14}$$

To this end we introduce a matrix $\Delta=(\Delta_{rs})_{r,s=0}^n$ defined by

$$\Delta_{rs}=\Delta_{-r-s+n},$$

$$\Delta_k=\frac{1}{2\pi\sqrt{-1}}\oint_C\frac{d\lambda}{\lambda}v\left(\frac{\rho}{\lambda}+2z-\rho\lambda\right)\lambda^k.$$

By using this matrix, we can express the matrix K as $k_{rs}=\rho^{r+s-n}\Delta_{r,s}$, obtaining

$$\tilde{k}_{00}=\rho^n\tilde{\Delta}_{00},\quad \tilde{k}_{nn}=\rho^{-n}\tilde{\Delta}_{nn}.$$

Further, we can easily verify that $\Delta_{-r}=(-1)^r\Delta_r$ so that $\tilde{\Delta}_{00}=(-1)^n\tilde{\Delta}_{nn}$, which yields Eq. (14). Thus the theorem is proved.

It follows from this theorem that the matrix $\tau(z,\rho)$ with $n=2m-1$, $m\in\mathbf{N}$ gives a solution of the Ernst equation. The corresponding exact solution is $f=\rho^{-2m+1}\Delta_{2m-1,0}/\det K$ and $e=\tilde{k}_{2n-1,2n-1}/\det K$. In terms of the matrix Δ we get

$$f=\rho^{-2n+1}\frac{\tilde{\Delta}_{2m-1,0}}{\det\Delta},\quad e=\rho^{-2n+1}\frac{\tilde{\Delta}_{2m-1,2m-1}}{\det\Delta},$$

an expression that was originally derived by Kyriakopoulos.[1]

[1] E. Kyriakopoulos, Phys. Rev. D **30**, 1158 (1984).
[2] P. V. Vein, Class. Quantum Gravit. **2**, 899 (1985).
[3] S. Candler and N. C. Freeman, University of Newcastle upon Tyne preprint.
[4] I. Hauser and F. J. Ernst, J. Math. Phys. **21**, 1126 (1979).
[5] L. Witten, Phys. Rev. D **19**, 718 (1978).
[6] M. F. Atiyah and R. S. Ward, Commun. Math. Phys. **55**, 117 (1977).
[7] E. F. Corrigan, D. B. Fairly, R. G. Yates, and P. Goddard, Commun. Math. Phys. **58**, 223 (1978).
[8] Y. Nakamura, J. Math. Phys. **24**, 606 (1983).
[9] V. A. Belinsky and V. E. Zakharov, Sov. Phys. JETP **48**, 985 (1978).
[10] N. I. Muskhelishivili, *Singular Integral Equations* (Noordhoff, Holland, 1963).

Massless scalar fields at null and spatial infinity in the Schwarzschild space-time

Chris X. Habisohn
Department of Physics and the Enrico Fermi Institute, University of Chicago, S. Ellis Avenue, Chicago, Illinois 60637

(Received 14 July 1988; accepted for publication 28 December 1988)

It is known that massless scalar, Maxwell, and linearized metric fields (in an appropriate gauge) having data of compact support will evolve to be asymptotically flat on any asymptotically flat background space-time. However, little is known about the evolution of data that is reasonably well behaved but has nontrivial falloff at spatial infinity. Is the set of such data that evolves to be asymptotically flat at null infinity in a curved asymptotically flat space-time of the same size as, and does it consist of elements with falloff rates similar to the set of such data in Minkowski space-time? Stewart and Schmidt analyzed massless scalar fields on both the Minkowski and Schwarzschild space-times. Their calculations indicated that the set of Schwarzschild data in question was much smaller than the Minkowski set. In this paper, this problem is reexamined and it is determined, contrary to the indications of Stewart and Schmidt, that the Schwarzschild set is of the same size, and its elements have falloff rates similar to the corresponding Minkowski set. This result supports the ability of the definition of asymptotic flatness to admit a large class of space-times.

I. INTRODUCTION

The introduction of the definition of an asymptotically flat space-time by Penrose[1] and the subsequent development of the theory of asymptotics has proved fruitful in our understanding of gravitational radiation. This modern approach to studying the far-field limit of general relativity has, in particular, yielded precise definitions of important quantities such as the radiated Bondi energy flux and the Bondi mass. However, very little is known concerning the existence of exact solutions to Einstein's equation that contain gravitational radiation and have a complete future null infinity.

The question of whether the presence of gravitational radiation will inhibit a smooth differential structure at null infinity has been studied in the literature. Winicour,[2] Christodoulou, Klainerman,[3] and other authors have obtained results indicating that it may be necessary to reduce the required differentiability class of null infinity to some finite order to guarantee that the definition of an asymptotically flat space-time will admit a large enough collection of radiating space-times to make it useful. On the other hand, Friedrich[4] proved that given data on a hyperboloidal initial data surface—that is, a surface having a cross section of future null infinity for a boundary—such that these data are "close" to that of the Minkowsi space-time, then these data have a maximal future evolution that can be smoothly conformally extended to include that part of future null infinity lying to the future of this surface. However, Friedrich did not prove that the maximal evolution to the past could be smoothly conformally extended to spatial infinity or past null infinity, or even to that part of future null infinity lying to the past of the boundary cross section of the initial surface.

To shed light on the compatibility of the definition of an asymptotically flat space-time and the presence of radiation, some authors have studied linear test fields on a curved background. Geroch and Xanthopolous,[5] in the framework of linearized gravity, demonstrated that in an appropriate gauge any solution to the linearized Einstein equation with data of compact support will smoothly conformally extend to null infinity. Their result, however, still leaves open the possibility that linear test fields having data with nontrivial falloff at spatial infinity may somehow fail to evolve to fields that are smoothly conformally extendible to null infinity when curvature is present.

Let us now focus on this issue for test massless scalar fields on a Schwarzschild background. In the Schwarzschild space-time, Stewart and Schmidt[6] analyzed the asymptotic behavior of massless test scalar fields having nontrivial data in a neighborhood of spatial infinity. They interpreted their results as indicating that, besides the known static solutions, almost no such massless scalar fields exist that are smoothly conformally extendible to both past and future null infinity. This work was later extended to higher spin fields by Porill and Stewart.[7]

In this paper, we reexamine the issue of whether the curvature of the Schwarzschild space-time tends to inhibit massless scalar test fields with nontrivial data at spatial infinity from evolving to fields smoothly conformally extendible to null infinity. We determine, contrary to the indications of the work of Stewart and Schmidt, that it does not.

This paper contains two results. The first result is a demonstration that the class of Schwarzschild test data having asymptotically flat evolution is similar in size to the corresponding Minkowski class. This demonstration, given at the end of Sec. II, basically consists of constructing a one-to-one map from Minkowski data to Schwarzschild data that preserves asymptotically flat evolution. The drawback of this first result, however, is that it gives us no clue as to how the falloff behavior of Schwarzschild data that evolves to be asymptotically flat compares with that of the Minkowski data. It could be that such Schwarzschild data would tend to behave badly in a neighborhood of spatial infinity.

Our second result clears up this drawback. It is a demonstration that the Schwarzschild data that evolves to be

 0022-2488/89/051103-12$02.50

asymptotically flat does indeed have falloff behavior similar to the corresponding Minkowski data. In both the Minkowski and Schwarzschild space-times, we will specify initial data on a constant time surface, i.e., a surface orthogonal to the timelike killing field. On such a surface, initial data for a scalar field is the pair $(\Phi_{(0)}, \dot{\Phi}_{(0)})$, where $\Phi_{(0)}$ and $\dot{\Phi}_{(0)}$ are the restriction of Φ and its time derivative to the surface. In constructing this second result, we consider the time symmetric and time antisymmetric pieces of the data separately and will focus upon the case of time symmetric data, $(\Phi_{(0)}, 0)$. (An extension of this result to time antisymmetric data will then be described briefly in Sec. VII.) We shall construct maps $\beta_{n,\ell}$ that yield an injective correspondence from time symmetric Minkowski data that falls off like $r^{-(n+1)}$, is proportional to $Y_{\ell,m}$, and evolves to be asymptotically flat at $\mathscr{I}$ to time symmetric Schwarzschild data that falls off like $\tilde{r}^{-(n+1)}$, is proportional to $Y_{\ell,m}$, and evolves to be asymptotically flat at $\mathscr{I}$. (Here r denotes the Minkowski radial coordinate and $\tilde{r}$ denotes the Schwarzschild radial coordinate.) This shows that there are "as many Schwarzschild data sets falling off like $\tilde{r}^{-(n+1)}$ as Minkowski data sets falling off like $r^{-(n+1)}$ that evolve to be asymptotically flat."

The paper proceeds as follows. We begin in Sec. II by restricting attention to radial and time-dependent fields proportional to a fixed $Y_{\ell,m}$, thus effectively reducing the problem to that of studying fields in two dimensions. We then identify both the Minkowski and Schwarzschild equations as evolution equations on a single two-dimensional flat space-time that we denote as $\mathscr{L}$ space. Our first result, the construction of a map Π_ℓ between Minkowski and Schwarzschild data is then obtained by associating a Minkowski data set with a Schwarzschild data set when they induce the same null data on a certain pair of intersecting null lines. An appropriate application of the null initial value formulation will guarantee that a field will smoothly evolve to null infinity if its null data smoothly extends there. The preservation of asymptotically flat evolution then follows because, by construction, the map preserves the null data.

In Sec. III, we determine the general form of time symmetric Minkowski data proportional to $Y_{\ell,m}$ that evolves to be asymptotically flat. We find that, aside from a linear combination of ℓ particular solutions, the general form is $(r^{-(\ell+1)} f(1/r) Y_{\ell,m}, 0)$ where f is an arbitrary smooth function of its one variable. Depending on the order of the zero f has at zero, the term $f(1/r) r^{-(\ell+1)}$ may fall off like $r^{-(n+1)}$ or faster for any integer $n \geqslant \ell$.

In Sec. IV, we will introduce the notion of an evolution equation of type ℓ and construct differential operators that will take time symmetric data from a type-ℓ equation to time symmetric data for a type-$(\ell+1)$ equation while preserving asymptotically flat evolution. The type-ℓ equation is a certain generalization of the ℓth Schwarzschild evolution equation. The main result of this section will be to associate with the ℓth Schwarzschild evolution equation a type-0 equation and a sequence of differential operators that will map data for the type-0 equation to data for the ℓth Schwarzschild evolution equation while preserving asymptotically flat evolution. Hence the study of the ℓth Schwarzschild evolution equation is essentially reduced to studying the type-0 equation.

In Sec. V, we introduce the notion of an asymptotically regular function. Basically, an asymptotically regular function of r (or $\tilde{r}$) is a generalization of a function that is smooth in the variable $1/r$ (or $1/\tilde{r}$) in the sense that the kth derivative of an asymptotically regular function will fall off roughly k powers of r (or $\tilde{r}$) faster than the function itself. Appendix A contains a statement and proof of a theorem that makes precise the falloff properties of the derivatives of an asymptotically regular function. We will need the notion of an asymptotically regular function to control the falloff rate of certain data sets under the action of the differential operators introduced in Sec. IV. We then proceed to prove a theorem which demonstrates that any time symmetric data set for a type-0 equation that evolves to be asymptotically flat will be asymptotically regular in $\tilde{r}$.

In Sec. VI, we will then construct for a fixed type-0 equation a sequence of injective maps W_n. Each map will take spherically symmetric Minkowski data that fall off like r^{-n} and which have asymptotically flat evolution to data for the type-0 equation that fall off like r^{-n} and which have asymptotically flat evolution. Appendix B contains a theorem that is used to demonstrate that our W maps so constructed do indeed preserve the falloff rate. The $\beta_{n,0}$ maps will be taken to be the W_n maps when the type-0 equation considered is the Schwarzschild $\ell = 0$ evolution equation. We then construct the $\beta_{n,\ell}$ maps for the cases $\ell \geqslant 1$. Roughly, this will be accomplished by composing the W maps with the appropriate ℓ-raising operators introduced in Sec. IV.

In Sec. VII we extend the action of the $\beta_{n,\ell}$ maps to time antisymmetric data as well.

II. REDUCTION TO TWO DIMENSIONS

In this section we reduce the study of massless scalar fields in the Minkowski and Schwarzschild space-times to that of studying fields satisfying wave equations on a single two-dimensional flat Lorentz space which we denote as $\mathscr{L}$ space. We end the section by constructing our one-to-one map Π_ℓ from Minkowski data to Schwarzschild data that preserves asymptotically flat evolution.

We start our reduction by separating out the radial and time dependence of scalar fields in the Minkowski space-time by restricting our attention to fields proportional to a fixed $Y_{\ell,m}$. Let the Minkowski metric in standard spherical coordinates be

$$ds^2 = dr^2 - dt^2 + r^2 \sin^2(\theta) d\theta^2 + r^2 d\varphi^2, \qquad (2.1)$$

and let Ψ be a massless scalar field of the form

$$\Psi = (\phi/r) Y_{\ell,m}, \qquad (2.2)$$

where ϕ is a function of r and t. That Ψ obeys the massless wave equation in the Minkowski space-time implies that ϕ obeys the two-dimensional wave equation

$$\left(\frac{\partial^2}{\partial r^2} - \frac{\partial^2}{\partial t^2}\right)\phi = \frac{\ell(\ell+1)}{r^2}\phi. \qquad (2.3)$$

By definition, Ψ is an asymptotically flat field if and only if the conformally related field $\Omega^{-1}\Psi$ (where Ω is a conformal factor that compactifies the Minkowski space-time) has a smooth extension to both future and past null infinity. This definition is independent of the particular conformal factor used. Since the conformal factor $1/r$ is known to compactify the Minkowski space-time, Ψ will be asymptotically flat if and only if $\phi Y_{\ell,m}$ ($= r\Psi$), and therefore ϕ itself, is smoothly extendible to null infinity.

A similar situation exists for scalar fields in Schwarzschild. Let the Schwarzschild metric in the spherical coordinates $(\tilde{t},\tilde{r},\theta,\varphi)$ be

$$d\tilde{s}^2 = -\left(1-\frac{2m}{\tilde{r}}\right)d\tilde{t}^{\,2} + \left(1-\frac{2m}{\tilde{r}}\right)^{-1} d\tilde{r}^2 + \tilde{r}^2 \sin^2(\theta)d\theta^2 + \tilde{r}^2\, d\varphi^2. \tag{2.4}$$

Let $\tilde{\Psi}$ be a massless scalar field in the Schwarzschild space-time of the form

$$\tilde{\Psi} = (\tilde{\phi}/\tilde{r})\, Y_{\ell,m}, \tag{2.5}$$

where $\tilde{\phi}$ is a function of $\tilde{r}$ and $\tilde{t}$. The fact that $\tilde{\Psi}$ obeys the massless wave equation implies $\tilde{\phi}$ obeys

$$\left(\frac{\partial^2}{\partial\tilde{r}^{*2}} - \frac{\partial^2}{\partial\tilde{t}^{\,2}}\right)\tilde{\phi} = \tilde{V}_\ell\left(\frac{1}{\tilde{r}}\right)\tilde{\phi}, \tag{2.6}$$

where $\tilde{V}_\ell(1/\tilde{r})$ is the static potential given by

$$\tilde{V}_\ell(1/\tilde{r}) = \left(1-\frac{2m}{\tilde{r}}\right)\left(\frac{\ell(\ell+1)}{\tilde{r}^2} + \frac{2m}{\tilde{r}^3}\right) \tag{2.7}$$

and

$$\tilde{r}^* = P(\tilde{r}) = \tilde{r} - 2m + 2m\ln(\tilde{r}-2m). \tag{2.8}$$

Since the conformal factor $1/\tilde{r}$ is known to compactify the Schwarzschild space-time, $\tilde{\Psi}$ will be asymptotically flat if and only if $\tilde{\phi} Y_{\ell,m}$ ($= \tilde{r}\tilde{\Psi}$), and therefore $\tilde{\phi}$ itself, is smoothly extendible to null infinity.

Let us now conformally compacify the wedge $r > |t|$ of the Minkowski space-time so that the conformal metric is both flat and smoothly extendible to i^0. On this wedge, choose the conformal factor Ω as

$$\Omega = 1/(r^2 - t^2). \tag{2.9}$$

Then, in terms of the new spherical coordinates $(\hat{t},\hat{r},\theta,\varphi)$ defined by

$$\hat{t} = \Omega t, \quad \hat{r} = \Omega r, \tag{2.10}$$

the conformal metric $d\hat{s}^2$ defined by $d\hat{s}^2 = \Omega^2\, ds^2$ is

$$d\hat{s}^2 = d\hat{r}^2 - d\hat{t}^{\,2} + \hat{r}^2 \sin^2(\theta)d\theta^2 + \hat{r}^2\, d\varphi^2. \tag{2.11}$$

Let us interpret (2.11) as defining an auxiliary flat space-time with global spherical coordinates $(\hat{t},\hat{r},\theta,\varphi)$ whose origin is the point i^0. In this picture, then, the wedge $r > |t|$ in the Minkowski space-time is conformally embedded onto the $\hat{r} > |\hat{t}|$ wedge of this auxiliary space-time. The future null cone of i^0 is to be identified with that part of future null infinity on the boundary of the wedge $r > |t|$. Likewise, the past null cone of i^0 is to be identified with that part of past null infinity on the boundary of the same wedge.

Now, define $\mathscr{L}$ space as a flat two-dimensional Lorentz space with global inertial coordinates $\hat{x}$ and $\hat{t}$, metric $-d\hat{t}^{\,2} + d\hat{x}^2$, and identify $\hat{x}$ with $\hat{r}$ for $\hat{x} > 0$. Through this identification we may view the field ϕ as existing on the region $\hat{x} > |\hat{t}|$ of $\mathscr{L}$ space. Define the future and past null boundaries of this region as σ^+ and σ^-, respectively. Let us further denote the $\hat{t} = 0$ and $\hat{x} > 0$ surface in this two-dimensional space-time as Σ. The surface Σ is the one-dimensional version of the $t = 0$ initial data surface in the Minkowski space-time and the line segments σ^+ and σ^- represent the one-dimensional versions of those pieces of future and past null infinity on the boundary of the $r > |t|$ wedge in the Minkowski space-time. These structures are depicted in Fig. 1(a). We will also use the convention of denoting the origin of $\mathscr{L}$ space as i^0 since it represents the two-dimensional version of spatial infinity. Viewed as a field on $\mathscr{L}$ space, ϕ will represent an asymptotically flat field if it has a smooth extension to σ^+ and σ^-. Writing the evolution equation (2.3) for ϕ in terms of $\hat{x}$ and $\hat{t}$, we get

$$\left(\frac{\partial^2}{\hat{x}^2} - \frac{\partial^2}{\partial\hat{t}^{\,2}}\right)\phi = \frac{\ell(\ell+1)}{\hat{x}^2}\,\phi. \tag{2.12}$$

We note that this evolution equation is regular on σ^+ and σ^-.

Remarkably, we can also identify the coordinates $\hat{x}$ and $\hat{t}$ on the $\hat{x} > |\hat{t}|$ region of $\mathscr{L}$ space with the radial and time coordinates of the Schwarzschild space-time in such a way that σ^+ and σ^- will represent a one-dimensional version of its future and past null infinity. This is a rather surprising result since it is known that the Schwarzschild space-time can not be conformally extended to smoothly admit a point at spatial infinity. However, in two dimensions, all metrics are conformally flat. Hence it is possible to compactify the Schwarzschild space-time so that the radial and time components of the conformal metric will be smoothly extendible to spatial infinity and appear flat. It would then be a simple matter to identify these radial and time components with the flat metric of $\mathscr{L}$ space. We will omit the steps needed to compactify the Schwarzschild space-time in this way and proceed directly to writing down the coordinate identification.

Define the null coordinates $(\tilde{u},\tilde{v})$ in the Schwarzschild space-time by

$$\tilde{u} = \tilde{r}^* - \tilde{t}, \tag{2.13}$$

$$\tilde{v} = \tilde{r}^* + \tilde{t}, \tag{2.14}$$

and the null coordinates $(\hat{u},\hat{v})$ in $\mathscr{L}$ space by

$$\hat{u} = \hat{x} - \hat{t}, \quad \hat{v} = \hat{x} + \hat{t},$$

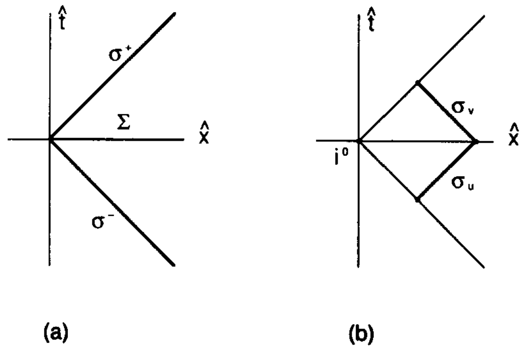

FIG. 1. Some structures on $\mathscr{L}$ space are shown. In (a) the two null surfaces σ^+ and σ^- are depicted along with the initial data surface Σ. In (b) the origin i^0 representing the point at spatial infinity is depicted along with the null surfaces σ_u and σ_v.

and identify these coordinates by

$$\hat{v} = 1/(2P^{-1}(\tilde{u}/2) - 4m), \tag{2.15}$$

$$\hat{u} = 1/(2P^{-1}(\tilde{v}/2) - 4m). \tag{2.16}$$

Unlike the Minkowski identification where only the wedge $r > |t|$ was identified, this identification will map the entire Schwarzschild space-time external to the event horizon onto the $\hat{x} > |\hat{t}|$ region of $\mathscr{L}$ space. The rather complicated looking form of these identifications stems mostly from the requirement that the differential structure between null infinity and σ^+ and σ^- should be preserved. Preservation of the differential structure is most easily shown by demonstrating that $\hat{v}$, a smooth coordinate on σ^- is also, via the coordinate identifications, a smooth coordinate on past null infinity in the Schwarzschild space-time. Along the line $\tilde{v} = 0$ we have $\tilde{u} = 2\tilde{r}^*$. By substituting $2\tilde{r}^*$ for $\tilde{u}$ into (2.15) we have $\hat{v} = 1/(2\tilde{r} - 4m)$. Now, $1/\tilde{r}$, being a conformal factor that compactifies Schwarzschild, must be a good coordinate on null infinity. Therefore $\hat{v}$, which is a smooth function of $1/\tilde{r}$ along the line $\tilde{v} = 0$, must be smoothly extendible to past null infinity along this line. Since $\hat{v}$ is a null coordinate, being smoothly extendible to past null infinity along this one line is enough to guarantee that it is a good coordinate along all of past null infinity. Hence the differential structure is preserved. To regularize the evolution equation (2.6) for $\tilde{\phi}$ on σ^+ and σ^- simply multiply both sides by $(4m\hat{v} + 1)(4m\hat{u} + 1)\Omega^{-2}$ and express it in terms of the "caretted" coordinates as

$$\left(\frac{\partial^2}{\partial \hat{x}^2} - \frac{\partial^2}{\partial \hat{t}^2}\right)\tilde{\phi} = (4m\hat{v} + 1)(4m\hat{u} + 1)\Omega^{-2}\tilde{V}\left(\frac{1}{\tilde{r}}\right)\tilde{\phi}. \tag{2.17}$$

The potential on the right is regular on σ^+ and σ^- because $\tilde{V}(1/\tilde{r})$ is a smooth function of $1/\tilde{r}$ that falls off like $(1/\tilde{r})^2$ or faster, so when it is multiplied by Ω^{-2} it can be viewed as a smooth function of $1/\tilde{r}$ multiplied by $(1/\tilde{r})^2/\Omega^2$—the ratio of two conformal factors squared—which itself must be smoothly extendible to σ^+ and σ^-. Thus we may view the original separated scalar wave equations on the Minkowski and Schwarzschild space-time as the regularized wave equations (2.12) and (2.17) on $\mathscr{L}$ space.

We now construct a simple one-to-one map Π_ℓ that takes scalar data to scalar data and preserves asymptotically flat evolution in the sense that if a data set in the domain is evolved using the Minkowski ℓth evolution equation (2.3) and the resulting field is smoothly extendible to σ^+ and σ^-, then the image data set evolved using the ℓth Schwarzschild evolution equation (2.6) will result in a field smoothly extendible to σ^+ and σ^-. In the $\hat{x} \geqslant |\hat{t}|$ region of $\mathscr{L}$ space place two intersecting null line segments, σ_u and σ_v so that σ_u has an end point on σ, σ_V has an end point on σ^+, and both intersect the point $\hat{x} = \hat{x}_0$ and $\hat{t} = 0$. These line segments are depicted in Fig. 1(b). Let $(\phi_{(0)}, \dot{\phi}_{(0)})$ be data on Σ. Evolve these data into both the future and past domains of dependence of Σ using (2.3), the Minkowski ℓth evolution equation. Restrict the resulting field ϕ to the two null segments σ_u and σ_v and, using the null initial value formulation, evolve this null data using (2.6), the Schwarzschild ℓth evolution equation, into the entire domain of dependence of Σ. (Recall that wave equations may also be evolved in spacelike directions in two dimensions.) The resulting field $\tilde{\phi}$ now induces data, say $(\tilde{\phi}_{(0)}, \dot{\tilde{\phi}}_{(0)})$, onto Σ. The action of Π_ℓ on $(\phi_{(0)}, \dot{\phi}_{(0)})$ is defined to be $(\tilde{\phi}_{(0)}, \dot{\tilde{\phi}}_{(0)})$.

Now, assume ϕ is smoothly extendible to both σ^+ and σ^-. Then the null data it induces on σ_u and σ_v will be smoothly extendible to their end points on σ^+ and σ^-. A smooth extension of $\tilde{\phi}$ to all of σ^+ and σ^- can be produced by evolving this extended data with the regularized equation (2.17). Hence Π_ℓ will preserve asymptotically flat evolution.

We summarize the above discussion about Π_ℓ in the following theorem.

Theorem 1: The bijection Π_ℓ from data on Σ to data on Σ preserves asymptotically flat evolution in the sense that if a data set evolved with the ℓth Minkowski evolution equation (2.3) has a smooth extension to all of σ^+ and σ^- then the evolution under the ℓth Schwarzschild evolution equation (2.6) of its image under Π_ℓ will also have a smooth extension to all of σ^+ and σ^-.

III. THE MINKOWSKI ANALYSIS

In this and the next three sections, we shall focus attention upon the case of time symmetric data. Let us denote as $M(\ell)$ the set of functions f on Σ such that $(f,0)$ evolved with the ℓth Minkowski equation results in a field smoothly extendible to σ^+ and σ^-. In this section we will prove the following theorem about the general form of the functions in $M(\ell)$.

Theorem 2: A function f defined on the initial surface Σ is in $M(0)$ if and only if there exists a smooth function of one variable, say g, such that $f = g(1/r)$. Furthermore, f is in $M(\ell)$ for $\ell > 0$ if and only if there exists a smooth function of one variable g and a collection of ℓ constants, say $C_0, C_1, \ldots, C_{\ell-1}$, such that $f = g(1/r)r^{-\ell} + C_0 r^{2-\ell} + C_1 r^{4-\ell} + \cdots + C_{\ell-1} r^{\ell}$.

Proof: On the wedge $\hat{x} > |\hat{t}|$ of $\mathscr{L}$ space, let ϕ_0 satisfy the $\ell = 0$ Minkowski evolution equation (2.12). This is just the standard wave equation on $\mathscr{L}$ space. The general time symmetric solution is

$$\phi_0 = [f(\hat{u}) + f(\hat{v})]/2. \tag{3.1}$$

The initial data of such a solution on Σ is $((f(\hat{x}), 0)$. The field ϕ_0 will be smoothly extendible to σ^+ and σ^- if and only if f is smooth at zero. Hence

$$M(0) = \{f(\hat{x}) \mid f \text{ is a smooth function at zero}\}.$$

Now, for each non-negative integer j define the operator Λ_j on fields in $\mathscr{L}$ space by

$$\Lambda_j[\phi] = \hat{x}^{-j}\frac{\partial}{\partial \hat{x}}(\hat{x}^j \phi) \tag{3.2}$$

and the operator $\overline{\Lambda}_j$ by

$$\overline{\Lambda}_j[\phi] = \hat{x}^{j+1}\frac{\partial}{\partial \hat{x}}\left(\frac{\phi}{\hat{x}^{j+1}}\right). \tag{3.3}$$

Here and in the future, any partial derivatives taken with respect to r, $\hat{x}$, and $\tilde{r}^*$ will be taken holding t, $\hat{t}$, and $\tilde{t}$ fixed, respectively. In terms of these operators, the ℓth Minkowski equation can be written as either $(\partial_{\hat{t}}\partial_{\hat{t}} - \overline{\Lambda}_{\ell-1}\Lambda_\ell)\phi = 0$ or $(\partial_{\hat{t}}\partial_{\hat{t}} - \Lambda_{\ell+1}\overline{\Lambda}_\ell)\phi = 0$. It follows that if ϕ satisfies the ℓth equation then $\overline{\Lambda}_\ell[\phi]$ will satisfy the $(\ell+1)$st Minkowski evolution equation and $\Lambda_\ell[\phi]$ satisfies the $(\ell-1)$th equation. Indeed, $(\partial_{\hat t}\partial_{\hat t}-\Lambda_{\ell+1}\overline{\Lambda}_\ell)\phi=0$ implies $\overline{\Lambda}_\ell(\partial_{\hat t}\partial_{\hat t}-\Lambda_{\ell+1}\Lambda_\ell)\phi=0$, which implies $(\partial_{\hat t}\partial_{\hat t}-\overline{\Lambda}_\ell\Lambda_{\ell+1})\overline{\Lambda}_\ell[\phi]=0$, which shows that $\overline{\Lambda}_\ell[\phi]$ satisfies the $(\ell+1)$st equation. Moreover, $\overline{\Lambda}_\ell$ as an operator is regular on σ^+ and σ^-. Hence $\overline{\Lambda}_\ell[\phi]$ smoothly extends to these boundaries if ϕ does. The converse is also true: if $\phi_{\ell+1}$ satisfies the ℓth + first Minkowski evolution equation, is time symmetric, and smoothly extendible to σ^+ and σ^-, then there exists a time symmetric ϕ_ℓ satisfying the ℓth equation, which is smoothly extendible to σ^+ and σ^- and satisfies

$$\phi_{\ell+1}=\overline{\Lambda}_\ell[\phi_\ell]. \tag{3.4}$$

To see this, let $(f(\hat x)_{\ell+1},0)$ be the initial data for $\phi_{\ell+1}$, where $f_{\ell+1}$ is smooth except, possibly, at $\hat x=0$, and define the function f_ℓ by

$$f_\ell(\hat x)=\hat x^{\ell+1}\int\frac{f(\hat x_{\ell+1})}{\hat x^{\ell+1}}\,d\hat x. \tag{3.5}$$

Define ϕ_ℓ as the field satisfying the ℓth Minkowski evolution equation with $(f(\hat x)_\ell,0)$ as initial data. It follows that $\phi_{\ell+1}=\overline{\Lambda}_\ell[\phi_\ell]$ since both sides satisfy the $(\ell+1)$st evolution equation and both have the same initial data. In order to show that ϕ_ℓ is smoothly extendible to σ^+ and σ^-, view (3.4) as a differential equation for ϕ_ℓ with source $\phi_{\ell+1}$ and integrate this equation along the lines of constant $\hat t$ to get the appropriate extension.

From the above remarks about the operators $\overline{\Lambda}_j$, it is clear that each solution to the ℓth evolution equation that smoothly extends to σ^+ and σ^- can be generated by successively applying the appropriate sequence of the $\overline{\Lambda}_j$ operators for $0\leqslant j\leqslant\ell$ to solutions of the $\ell=0$ evolution equation that have a smooth extension to σ^+ and σ^-. In a similar manner, because the operators $\overline{\Lambda}_j$ act naturally on initial data, each element of $M(\ell)$ can be generated from an element of $M(0)$ by successively applying the appropriate sequence of $\overline{\Lambda}_j$ operators. For example, in the $\ell=1$ case, the time symmetric data that evolve to fields smoothly extendible to σ^+ and σ^- are precisely of the form $(\overline{\Lambda}_0[h(\hat x)],0)$ where h is a smooth function. To determine the behavior of this data, write the function $\overline{\Lambda}_0[h(\hat x)]$ as

$$\overline{\Lambda}_0[h(\hat x)]=\hat x\frac{\partial}{\partial\hat x}\left[\frac{h(\hat x)-h(0)}{\hat x}+\frac{h(0)}{\hat x}\right]. \tag{3.6}$$

The first term in brackets on the right side is, as a function of one variable, equal to a smooth function of $\hat x$ multiplied by $\hat x$. The second term is equal to $-h(0)\hat x^{-1}$, which is singular at $\hat x=0$. Therefore,

$$M(1)=\{f(\hat x)\,|\,f(\hat x)=\hat x g(\hat x)+C_0\hat x^{-1},\ \text{where } g \text{ is smooth and } C_0 \text{ is an arbitrary constant}\}.$$

In a similar manner, we can apply the operator $\overline{\Lambda}_1$ to this $\ell=1$ initial data to generate the $\ell=2$ initial data that evolve to fields smoothly extendible to σ^+ and σ^-. The result of this calculation is

$$M(2)=\{f(\hat x)\,|\,f(\hat x)=\hat x^2g(\hat x)+C_0+C_1\hat x^{-2},\ \text{where } g \text{ is smooth and } C_0 \text{ and } C_1 \text{ are arbitrary constants}\}.$$

By iterating this procedure we can construct the form of the time symmetric data that evolve to fields smoothly extendible to σ^+ and σ^- for any ℓ. For fixed $\ell>0$, we get

$$M(\ell)=\{f(\hat x)\,|\,f(\hat x)=g(\hat x)\hat x^\ell+C_0\hat x^{\ell-2}+C_1\hat x^{\ell-4}+\cdots+C_{\ell-1}\hat x^{-\ell},\ \text{where } g \text{ is smooth and } C_0, C_1,\ldots,C_{\ell-1} \text{ are arbitrary}\}.$$

The desired result of the theorem now follows because $1/r=\hat x$ on Σ.

IV. OPERATORS FOR THE TYPE-ℓ EQUATION

In the previous section we found that the differential operator $\overline{\Lambda}_\ell$ mapped solutions for the ℓth Minkowski equation that evolved smoothly to σ^+ and σ^- to solutions for the $\ell+$ first Minkowski equation that evolved smoothly to σ^+ and σ^-. Note that Eqs. (2.3) and (2.12), two different versions of the same Minkowski evolution equation, have the same form (simply replace $\hat x$ with r and $\hat t$ with t). Therefore one can construct a set of ℓ-raising operators, say $\overline{\Theta}_j$, different from $\overline{\Lambda}_i$ by simply replacing $\hat x$ with r in the definition of the $\overline{\Lambda}_i$ operators. The action of $\overline{\Theta}_j$ on a field ϕ is then

$$\overline{\Theta}_j(\phi)=r^{j+1}\frac{\partial}{\partial r}\left(\frac{\phi}{r^{j+1}}\right). \tag{4.1}$$

The ℓth such operator, being smooth on σ^+ and σ^-, will map time symmetric solutions of the ℓth equation that are smoothly extendible to σ^+ and σ^- to time symmetric solutions of the $(\ell+1)$st equation that are smoothly extendible to σ^+ and σ^-. Unlike the $\overline{\Lambda}_i$ operators, however, the $\overline{\Theta}_j$ do not work backwards. That is, if $\phi_{\ell+1}$ is a time symmetric solution of the $(\ell+1)$st equation, which is smoothly exendible to null infinity, then there may not exist a time symmetric solution ϕ_ℓ of the ℓth equation which is smoothly extendible to null infinity and satisfies

$$\phi_{\ell+1}=\overline{\Theta}_\ell[\phi_\ell]. \tag{4.2}$$

This means that one will not be able to construct all of $M(\ell)$ by applying the appropriate sequence such $\overline{\Theta}_j$ operators to elements in $M(0)$. One can ask, however, what subset of $M(\ell)$ is generated by applying the appropriate sequence. The answer is precisely those elements of the form $f(1/r)r^{-\ell}$, where f is smooth.

In this section, we will construct operators for the Schwarzschild space-time analogous to the $\overline{\Theta}_j$ operators. [As we shall soon see, the construction of these operators depends heavily on the fact that the associated equation is static. For this reason there is no analog of $\overline{\Lambda}_i$ for Schwarzschild because, unlike the Minkowski equation (2.12), the ℓth Schwarzschild evolution equation written in terms of the caretted coordinates of $\mathscr{L}$ space is not static.]

To begin our constructions we need to introduce a certain generalization of the ℓth Schwarzschild equation termed a type-ℓ equation.

Definition 1: A type-ℓ equation is an equation of the form

$$\left(\frac{\partial^2}{\partial\bar r^{*2}}-\frac{\partial^2}{\partial\bar t^2}\right)\Psi=V_\ell\left(\frac{1}{\bar r}\right)\Psi, \tag{4.3}$$

where the potential V_ℓ is a smooth function of one variable, the function $\tilde{r}^2 V_\ell(1/\tilde{r})$ has a limit of $\ell(\ell+1)$ as $\tilde{r}\to\infty$, and such that a static solution $S_\ell(1/\tilde{r})$ exists where S_ℓ is a smooth function of one variable that falls off like $\tilde{r}^{-l}$.

The ℓth Schwarzschild equation is type ℓ. Indeed, its potential $\tilde{V}_\ell(1/\tilde{r}) = (1-2m/\tilde{r})(\ell(\ell+1)/\tilde{r}^2 + 2m/\tilde{r}^3)$ satisfies the potential fall off requirement and the static solution $\tilde{r}Q_\ell(\tilde{r}/2m-1)$, where Q_ℓ is the ℓth Legendre function of the second kind, satisfies the static solution requirement.

We also wish to introduce the notion of the lowering operation on a type-ℓ equation.

Definition 2: The lowering operation of the type-ℓ equation with static solution $S_\ell(1/\tilde{r})$ consists of the replacement of $V_\ell(1/\tilde{r})$ by

$$V_{\ell-1}\left(\frac{1}{\tilde{r}}\right) = -V_\ell\left(\frac{1}{\tilde{r}}\right) + 2\left(\frac{\partial \ln(S_\ell(1/\tilde{r}))}{\partial \tilde{r}^*}\right)^2. \quad (4.4)$$

The resulting equation, which is of type $(\ell-1)$, is said to be L related to the original equation.

To show that a type-$(\ell-1)$ equation is indeed generated by this prescription we need to demonstrate that $V_{\ell-1}$ $\times(1/\tilde{r})$ has the appropriate leading falloff term and a static solution exists with the appropriate falloff rate. To show that $V_{\ell-1}(1/\tilde{r})$ has the appropriate leading term we note that $S_\ell(1/\tilde{r})$ falls off like $\tilde{r}^{-\ell}$ so the leading falloff term in the derivative of the logarithm of $S_\ell(1/\tilde{r})$ is $\ell/\tilde{r}$. By squaring this result, multiplying by 2, and subtracting $\ell(\ell+1)/\tilde{r}^2$ we get that the leading falloff term in $V_{\ell-1}(1/\tilde{r})$ is $\ell(\ell-1)/\tilde{r}^2$. To show that the L-related equation has a static solution $S_{\ell-1}(1/\tilde{r})$ with the appropriate falloff rate we first discuss some further connections between it and the original type-ℓ equation. It is possible to construct differential operators Γ_ℓ and $\overline{\Gamma}_{\ell-1}$ for a type-ℓ equation so that Γ_ℓ maps solutions of the type-ℓ equation to solutions of its L-related equation and $\overline{\Gamma}_{\ell-1}$ maps solutions of its L-related equation to solutions of itself. Indeed, let Γ_ℓ be defined by

$$\Gamma_\ell\psi = S_\ell\left(\frac{1}{\tilde{r}}\right)\frac{\partial}{\partial\tilde{r}^*}\left(\frac{\psi}{S_\ell(1/\tilde{r})}\right) \quad (4.5)$$

and $\overline{\Gamma}_{\ell-1}$ by

$$\overline{\Gamma}_{\ell-1}\psi = \overline{S}_{\ell-1}\left(\frac{1}{\tilde{r}}\right)\frac{\partial}{\partial\tilde{r}^*}\left(\frac{\psi}{\overline{S}_{\ell-1}(1/\tilde{r})}\right), \quad (4.6)$$

where $\overline{S}_{\ell-1}(1/\tilde{r}) = 1/S_\ell(1/\tilde{r})$. In terms of these operators, the type-ℓ equation can be written as $(\overline{\Gamma}_{\ell-1}\Gamma_\ell - \partial_{\tilde{t}}\,\partial_{\tilde{t}})$ $\psi=0$ and the L-related equation as $(\Gamma_\ell\overline{\Gamma}_{\ell-1} - \partial_{\tilde{\tau}}\,\partial_{\tilde{t}}))$ $\times\psi=0$. If ψ satisfies the original equation then Γ_ℓ $(\overline{\Gamma}_{\ell-1}\Gamma^{(\ell)} - \partial_{\tilde{t}}\,\partial_{\tilde{t}})\psi=0$ implying $(\Gamma_\ell\overline{\Gamma}_{\ell-1} - \partial_\tau\,\partial_\tau)$ $\times\Gamma_\ell\psi=0$, which just states that $\Gamma_\ell\psi$ satisfies the L-related equation. By a similar argument, $\overline{\Gamma}_{\ell-1}\psi$ will satisfy the type-ℓ equation if ψ satisfies its L-related equation.

It follows from the form of $\overline{\Gamma}_{\ell-1}$ that $\overline{S}_{\ell-1}(1/\tilde{r})$ is a static solution of the L-related equation. The static solution $S_{\ell-1}(1/\tilde{r})$ of the L-related equation that falls off like $\tilde{r}^{1-\ell}$ can now be obtained as follows. The L-related equation acting on $S_{\ell-1}(1/\tilde{r})$ is $\Gamma_\ell[\overline{\Gamma}_{\ell-1}S_{\ell-1}(1/\tilde{r})] = 0$. Therefore $\overline{\Gamma}_{\ell-1}S_{\ell-1}(1/\tilde{r}) = S_\ell(1/\tilde{r})$. Inverting the $\overline{\Gamma}_{\ell-1}$ operator by multiplying by $S_\ell(1/\tilde{r})$, integrating with respect to $\tilde{r}^*$, and then dividing by $S_\ell(1/\tilde{r})$ we get

$$S_{\ell-1}\left(\frac{1}{\tilde{r}}\right) = \frac{1}{S_\ell(1/\tilde{r})}\int S_\ell\left(\frac{1}{\tilde{r}}\right)^2 d\tilde{r}^*. \quad (4.7)$$

The smoothness of $S^{\ell-1}(1/\tilde{r})$ in $1/\tilde{r}$ and its falloff rate of $\tilde{r}^{1-\ell}$ now follows from this integral expression.

We can now construct the equation L related to the L-related equation to get a type-$(\ell-2)$ equation along with its associated operators $\Gamma_{\ell-1}$ and $\overline{\Gamma}_{\ell-2}$. By iterating this procedure we will construct a series of type-j equations for $0\leqslant j\leqslant\ell$, ending with a type-0 equation. The composite operator $\overline{\Gamma}_{\ell-1}\overline{\Gamma}_{\ell-2}\cdots\overline{\Gamma}_0$ will map solutions of this type-0 equation to solutions of the original type-ℓ equation. Each $\overline{\Gamma}_j$ operator is smoothly extendible to σ^+ and σ^- because the static solutions $S_j(1/\tilde{r})$, being smooth in $1/\tilde{r}$, all are. Hence the composite $\overline{\Gamma}_{\ell-1}\overline{\Gamma}_{\ell-2}\cdots\overline{\Gamma}_0$ operator will also preserve smooth evolution to σ^+ and σ^-.

The operators $\overline{\Gamma}_j$ are analogs of the $\overline{\Theta}_j$ operators in the sense that each increases the ℓ value of the field each is applied to, each has a natural action on initial data, and each preserves smooth evolution to σ^+ and σ^-. There is, however, an important distinction that should be kept in mind. The $\overline{\Theta}_j$ operators for the Minkowski equation are closed in the sense that $\overline{\Theta}_\ell$ will map solutions from ℓth Minkowski evolution equation to the $(\ell+1)$st. However, the $\overline{\Gamma}_j$ operators for the Schwarzschild equations are not closed in this sense. One is only guaranteed that $\overline{\Gamma}_\ell$ will map solutions to solutions of some other type-$(\ell+1)$ equation.

In closing this section we wish to comment on the special case of the equation L related to a type-0 equation. As calculated above, the leading falloff term in the potential of the equation L related to a type-ℓ equation is $\ell(\ell-1)/\tilde{r}^2$. For the case $\ell=0$, this implies the $\tilde{r}^{-2}$ term in the potential of the L-related equation will vanish. It follows that the equation L related to a type-0 equation is itself a type-0 equation. And, in fact, the equation L related to the equation L related to a type-0 equation is the original type-0 equation. To accommodate this particular instance we need to alter our notation slightly. Let us denote Γ_0 as the operator defined by (4.5) for $\ell=0$ that maps solutions from the original type-0 equation to its L-related equation. Let us further denote Γ_0' as the operator defined by (4.6) for $\ell=0$, which maps solutions from the L-related equation to the original equation. These two operators will be used in Sec. VI in our analysis of the falloff rates of initial data that, when evolved with a type-0 equation, has an evolution smoothly extendible to σ^+ and σ^-.

V. ASYMPTOTIC REGULARITY AND THE TYPE-0 EQUATION

In Sec. VI, we will use the differential operators constructed in the previous section in constructing our $\beta_{n,\ell}$ maps. However, since the $\beta_{n,\ell}$ maps are intended to preserve the falloff rates of the data, we need to understand how the falloff rates of certain data will be affected by these differential operators. In Sec. III, where we determined the falloff behavior of the Minkowski data that evolved to be smoothly extendible to σ^+ and σ^- we were able to control the falloff rates of data under differentiation since we were dealing with a function smooth in the variable $1/r$. The derivative of a function smooth in $1/r$ and falling off like r^{-n} for $n \geqslant 1$ will fall off like $r^{-(n+1)}$. However, the data for type-ℓ equations that evolve to be smoothly extendible to σ^+ and σ^- need not be smooth in any readily identifiable variable. What we need is a generalization, that will apply to such data, of the notion of a function that is smooth in the variable $1/\tilde{r}$ or $1/\tilde{r}^*$ in the sense that the derivative with respect to $\tilde{r}^*$ of such a function will fall off, roughly, one power of $\tilde{r}$ (or equivalently $\tilde{r}^*$) faster than the function itself. The following definition of asymptotic regularity provides such a notion. Our definition differs from the notion of regularity put forth by Ashtekar and Hansen in Ref. 8 in that ours deals with a function of one variable while theirs deals with four-dimensional scalar fields.

Definition 3: We say that function f: $\mathbb{R} \to \mathbb{R}$ is an asymptotically regular function if f is smooth, the limit of $f(x)$ exists as $x \to \infty$, and $\lim x^n f^{(n)} = 0$ as $x \to \infty$ for all positive integers n, where $f^{(n)}$ denotes the nth derivative of f.

For functions on the initial surface Σ, being asymptotically regular in the coordinate r is equivalent to being asymptotically regular in the coordinate $\tilde{r}$ and $\tilde{r}^*$. This fact follows from relations (A1) and (A2) in Appendix A.

The specific falloff behavior of asymptotically regular functions under differentiation is captured by the following theorem proved in Appendix A.

Theorem 3: Let f be an asymptotically regular function in x and let $c \geqslant 0$ be a constant such that $x^m f$ is bounded as $x \to \infty$ for all $m < c$. Then, for any positive integer k, $x^q f^{(k)}$ is bounded as $x \to \infty$ for all $q < c + k$.

It turns out that a type-0 equation has enough structure to allow us to demonstrate that its time symmetric data that evolves to be smoothly extendible to σ^+ and σ^- are asymptotically regular in r. Indeed the following theorem holds.

Theorem 4: Let ψ be a time symmetric solution of a type-0 equation that smoothly extends to σ^+ and σ^- and has data $(f,0)$. Then ψ is continuous at i^0 and f is an asymptotically regular function of r (equivalently $\tilde{r}$ or $\tilde{r}^*$).

Proof: We begin by expressing the type-0 evolution equation in terms of $\hat{u}$ and $\hat{v}$. This can be accomplished by multiplying (4.3) by $(4m\hat{v}+1)(4m\hat{u}+1)(\hat{v}\hat{u})^{-2}$. We get

$$\partial_{\hat{v}}\,\partial_{\hat{u}}\Psi = (4m\hat{v}+1)(4m\hat{u}+1)(\hat{v}\hat{u})^{-2}V(1/\tilde{r})\psi\,. \tag{5.1}$$

We note that the potential on the right side of (5.1), $(4m\hat{v}+1)(4m\hat{u}+1)(\hat{v}\hat{u})^{-2}V(1/\tilde{r})$, blows up at the origin i^0 when approached along the surface Σ. Indeed, $(\hat{v}\hat{u})^{-2} \sim \hat{x}^{-4}$ while $V(1/\tilde{r}) \sim \hat{x}^3$.

Now, introduce the finer coordinate system $(\bar{u}, \bar{v})$ on the $\hat{u} \geqslant 0$, $\hat{v} \geqslant 0$ wedge of $\mathscr{L}$ space by the expressions

$$\hat{u} = \bar{u}^n\,, \tag{5.2}$$

$$\hat{v} = \bar{v}^n\,, \tag{5.3}$$

where n is a positive even integer. In terms of these new "barred" coordinates, the evolution equation (5.1) can be written as

$$\partial_{\bar{v}}\,\partial_{\bar{u}}\psi = (4m\bar{v}^n+1)(4m\bar{u}^n+1)n^2(\bar{u}\bar{v})^{-(n+1)}V(1/\tilde{r})\Psi\,. \tag{5.4}$$

Here we note that the potential term arising in (5.4) is equal to the potential term in (5.1) multiplied by $n^2(\bar{u}\bar{v})^{(n-1)}$. With $n \geqslant 4$, this extra factor, $n^2(\bar{u}\bar{v})^{(n-1)}$, falls off fast enough at i^0 to make the potential vanish there. Indeed, along Σ we have $(\bar{u}\bar{v})^{-(n+1)} \sim \hat{x}^{-(2+2/n)}$ while $V(1/\tilde{r}) \sim \hat{x}^3$. In fact, with respect to the barred coordinates the potential in (5.4) will be at least $C^{(n-3)}$ at i^0. To see this, let us first express—using (5.2), (5.3), and (2.13)–(2.16)—the function $\zeta = 1/\tilde{r}^*$ in terms of these new coordinates as

$$\zeta = \frac{1}{\tilde{r}^*} = \frac{4\bar{u}^n\bar{v}^n}{2\bar{u}^n + 2\bar{v}^n - 8mn^2\bar{u}^n\bar{v}^n\ln(4\bar{u}\bar{v})}\,. \tag{5.5}$$

This function ζ, along with the combination $\zeta\ln(\zeta)$, are both $C^{(n-1)}$ extendible to i^0. Let us further define the function γ by $\gamma = (\tilde{r} - 2m)/\tilde{r}^*$. Using (2.8) we find that γ satisfies the following equation:

$$\gamma = 1 + 2m\zeta\ln(\zeta) - 2m\zeta\ln(\gamma)\,. \tag{5.6}$$

For large $\tilde{r}$, the function γ has a limit of unity. Hence as a function on $\mathscr{L}$ space γ can be extended continuously to i^0 with this same limiting value. One can now apply the implicit function theorem to Eq. (5.6) to solve for γ as a smooth function of ζ and $\zeta\ln(\zeta)$ in a neighborhood of i^0. Since both ζ and $\zeta\ln(\zeta)$ are $C^{(n-1)}$ at i^0 it follows that γ is too. Now, let us substitute the quantity $\zeta^3\gamma^{-3}T(\zeta\gamma^{-1})$ into (5.4) for $V(1/\tilde{r})$, where T is the appropriate smooth function of one variable. We get

$$\partial_{\bar{v}}\partial_{\bar{u}}\psi = (4m\bar{v}^n+1)(4m\bar{u}^n+1)n^2(\bar{u}\bar{v})^{-(n+1)} \times \zeta^3\gamma^{-3}T(\zeta\gamma^{-1})\psi\,. \tag{5.7}$$

Excluding the $(\bar{u}\bar{v})^{-(n+1)}\zeta^3$ term, each factor in the potential of (5.7) is $C^{(n-1)}$ extendible to i^0. By using (5.5), one can explicitly compute that $(\bar{u}\bar{v})^{-(n+1)}\zeta^3$ is $C^{(n-3)}$ extendible to i^0. Therefore the entire potential term in (5.7) is $C^{(n-3)}$ extendible to i^0.

The next step in the proof is to show that ψ, being the solution of (5.7) that smoothly extends to σ^+ and σ^-, will have a $C^{(n-5)}$ extension to i^0 with respect to the barred coordinates. From the general initial value formulation theorem 7.4.7 in Hawking and Ellis[9] it follows that smooth data for a two-dimensional wave equation with a flat wave operator and with a potential in the jth Sobolev space can be evolved uniquely into the domain of dependence of the initial data surface and the resulting evolution will also be in the jth Sobolev space. The collection of C^j functions is a subset of the jth Sobolev space and, in two dimensions, is itself a subset of the collection of $C^{(j-2)}$ functions. Thus evolving smooth data with a two-dimensional wave equation having a flat wave operator and a $C^{(n-3)}$ potential will result in a $C^{(n-5)}$ evolution. (Actually, by integrating the two-dimensional wave equation it is possible to demonstrate that the resulting evolution will be $C^{(n-2)}$. However, the $C^{(n-5)}$ result is sufficient for our purpose.) Now, recall that in two dimensions a wave equation can be evolved in spacelike directions. In particular, consider the intial data for (5.7) induced by ψ on some timelike initial data surface that intersects both σ^+ and σ^-. This data will be smooth in the barred coordinates since ψ, being a smooth function of $\hat{u}$ and $\hat{v}$ in a neighborhood of the initial data surface, is trivially smooth in the finer coordinates $\bar{u}$ and $\bar{v}$. Furthermore, since the potential in (5.7) is

$C^{(n-3)}$ and i^0 is in the domain of dependence of the timelike initial data surface, we are guaranteed that this smooth data can be evolved to a field that is $C^{(n-5)}$ and includes i^0. This field, then, constitutes a $C^{(n-5)}$ extension of ψ to i^0. The first conclusion of the theorem, that ψ has a continuous extension to i^0, follows trivially from the existence of this $C^{(n-5)}$ extension.

Now define the rectangular barred coordinates $(\bar{x},\bar{t})$ on $\mathscr{L}$ space by

$$\bar{x} = (\bar{u} + \bar{v})/2\,, \tag{5.8}$$

$$\bar{t} = (\bar{u} - \bar{v})/2\,. \tag{5.9}$$

Since ψ is $C^{(n-5)}$ at i^0, it follows that, for k a positive integer, the field f_k defined by

$$f_k = \left(\bar{x}\frac{\partial}{\partial \bar{x}}\right)^k \psi \tag{5.10}$$

is continuous and vanishes at i^0 for $1 \leqslant k \leqslant (n-5)$. On Σ, the coordinates $\bar{x}$ and r are related by

$$1/r = \bar{x}^n \tag{5.11}$$

from which it follows that f_k, on Σ, is given by

$$f_k = \left(-nr\frac{\partial}{\partial r}\right)^k \psi\,. \tag{5.12}$$

The integer n was arbitrary. Therefore on Σ, the limit of f_k as $r \to \infty$ is zero for any $k \geqslant 1$. This implies that f, the restriction of ψ to the surface Σ, is an asymptotically regular function of r. Q.E.D.

VI. CONSTRUCTION OF THE MAPS $\beta_{n\ell}$

Fix a type-0 equation. In this section we will construct for each integer $n \geqslant 0$ an injective map W_n that takes functions smooth in $1/r$ [i.e., functions in $M(0)$] that fall off like r^{-n} to other functions on Σ that fall off like r^{-n} (equivalently $\tilde{r}^{-n}$) and which when viewed as time symmetric data for the type-0 equation will have an evolution smoothly extendible to σ^+ and σ^-. Hence we can conclude that time symmetric data for a type-0 equation that has an evolution smoothly extendible to σ^+ and σ^- has falloff behavior similar to the Minkowski $\ell = 0$ data having asymptotically flat evolution. Our desired Schwarzschild maps for the spherically symmetric case $\beta_{n,0}$ will be taken to be the W_n maps when the type-0 equation under consideration is the $\ell = 0$ Schwarzschild evolution equation. We will then construct the $\beta_{n,\ell}$ maps for the cases $\ell > 0$. Roughly, this will be accomplished by composing the W maps with the appropriate ℓ-raising operators introduced in Sec. IV.

We start by constructing an injective map α which will take functions in $M(0)$ to other functions on Σ which, when viewed as time symmetric data for a type-0 equation, will have an evolution smoothly extendible to σ^+ and σ^-. The maps W_0 and W_1 will then be obtained by appropriately restricting the domain of α. Basically, the construction of this injective map will be similar to that of the bijection Π_ℓ given in Sec. II except that we add to the image a multiple of a data set which produces unit data on the two intersecting null line segments σ_u and σ_v. The multiple is chosen so that the resulting map will preserve the limiting value of the data at i^0. Let $f \in M(0)$ and let ϕ be the field evolved from $(f,0)$ using the $\ell = 0$ Minkowski equation. Then restrict ϕ to σ_u and σ_v and add to this restriction the constant c_1. View the resulting restricted function as null data for the type-0 equation and evolve these data into the entire domain of dependence of Σ to get ψ. This field ψ is smoothly extendible to σ^+ and σ^- since its null data is. From Theorem 4, ψ will be continuous at i^0. Similarly, the solution to the type-0 equation that has unit data on the null surfaces σ_u and σ_v will also be continuous at i^0. By placing the null surfaces σ_u and σ_v close enough to i^0 one can guarantee that the value of the solution with unit null data will not vanish at i^0. Therefore it is possible to tune the constant c_1 so that the limit of ψ at i^0 is $f(i^0)$. Let $(g,0)$ be the data for ψ on Σ for this choice of c_1. The action of α on f is defined to be g.

We will now use α to construct the maps W_0 and W_1. The domain of W_0 consists of those elements of $M(0)$ that have a nonvanishing limit at i^0. We define W_0 on such elements to have the same action as the map α. That the map W_0 preserves the limiting value of data at i^0 is, of course, a direct consequence of the construction of α. The domain of W_1 consists of those elements of $M(0)$ which fall off like $1/r$. We define W_1 on such elements to have, again, the same action as the map α. That this choice preserves the $1/r$ falloff rates follows from the following proposition, the proof of which is given in Appendix B.

Proposition 1: Let $f \in M(0)$ and $\tilde{f} = \alpha(f)$ and denote the difference $\tilde{f} - f$ as Δ. Then the limit as $r \to \infty$ of the function $r^\epsilon \Delta$ vanishes for all $\epsilon < 1$. Furthermore if f vanishes at i^0 then this limit vanishes for all $\epsilon < 2$.

We now proceed to construct the W maps for $n \geqslant 2$. First, we need to show that the kth derivative with respect to $\tilde{r}^*$ of a function in the range of W_1 will fall off like $\tilde{r}^{-(k+1)}$.

Lemma 1: The kth derivative with respect to $\tilde{r}^*$ of a function in the range of W_1 will fall off like $\tilde{r}^{-(k+1)}$.

Proof: Let $\tilde{f} = W_1(f)$ and $\Delta = \tilde{f} - f$. Using (B1), the derivative of $\tilde{f}$ with respect to $\tilde{r}^*$ is

$$\frac{d}{d\tilde{r}^*}\tilde{f} = \left(1 - \frac{4m}{r+2m}\right)\frac{d}{dr}(f + \Delta), \tag{6.1}$$

which can be written as

$$\frac{d}{d\tilde{r}^*}\tilde{f} = (1+g)\frac{d}{dr}(f + \Delta), \tag{6.2}$$

where g is an asymptotically regular function that falls off like $1/r$. By expanding the right-hand side and collecting terms we have

$$\frac{d}{d\tilde{r}^*}\tilde{f} = \frac{d}{dr}(f) + \Delta^{(1)}, \tag{6.3}$$

where

$$\Delta^{(1)} = \left(g\frac{d}{dr}(f) + (1+g)\,\frac{d}{dr}(\Delta)\right). \tag{6.4}$$

The first term on the right-hand side of (6.3)—the derivative of f with respect to r—will fall off like $1/r^2$ since f is smooth in $1/r$ and falls off like $1/r$. From the above theorem, we know that Δ will fall off faster than $r^{-\epsilon}$ for all $\epsilon < 2$. Applying the falloff theorem in Sec. IV to Δ we conclude that its derivative with respect to r, and hence the remainder term $\Delta^{(1)}$, must fall off at least as fast as $r^{-\epsilon}$ for any $\epsilon < 3$. Therefore, the derivative of $\tilde{f}$ with respect to $\tilde{r}^*$ falls off like

$1/r^2$. By iterating this procedure, one can show that the k th derivative of $\tilde{f}$ with respect to $\tilde{r}^*$ can be expressed as

$$\frac{d^k}{d\tilde{r}^{*k}}\tilde{f}=\frac{d^k}{dr^k}(f)+\Delta^{(k)}, \tag{6.5}$$

where $\Delta^{(k)}$ falls off at least as fast as $r^{-\epsilon}$ for any $\epsilon<k+2$. The k th derivative of f with respect to r will fall off like $1/r^{k+1}$ since f is smooth in $1/r$ and falls off like $1/r$. Hence the k th derivative of $\tilde{f}$ with respect to $\tilde{r}$ will also fall off like $1/r^{k+1}$.

We are now ready to construct W_2. Let f be an element of $M(0)$ that falls off like $1/r^2$. Let $\overline{W}_1$ be constructed for the equation L related to the type-0 equation by the same procedure as W_1 was for the original type-0 equation. Finally, define the action of W_2 on f to be $\Gamma_0'(\overline{W}_1(rf))$. Here Γ_0' is the operator, as discussed at the end of Sec. IV, that maps data for the L-related equation to the original type-0 equation while preserving asymptotically flat evolution to σ^+ and σ^-. Functions in the range of W_2 are guaranteed to fall off like $1/r^2$ by Lemma 1.

To construct W_{n+1} for $n\geqslant 2$ we use induction. Let f be an element of $M(0)$ that falls off like $r^{-(n+1)}$. Let $\overline{W}_n$ be the analog of W_n for the L-related equation and define the action of W_{n+1} on f to be $\Gamma_0'(\overline{W}_n(rf))$. Again Lemma 1 will guarantee that elements in the range of W_{n+1} will fall off like $1/r^{n+1}$.

We summarize the properites of W_n in the following theorem.

Theorem 5: The map W_n constructed above is injective, takes functions smooth in $1/r$ [i.e., functions in $M(0)$] that fall off like r^{-n} to other functions on Σ that fall off like r^{-n} (equivalently $\tilde{r}^{-n}$) and which when viewed as time symmetric data for the type-0 equation will have an evolution smoothly extendible to σ^+ and σ^-.

We are now ready to construct the $\beta_{n,\ell}$ maps for the Schwarzschild evolution equation. For $\ell=0$, we take $\beta_{n,0}=W_n$. For $\ell\geqslant 1$ let $f\in M(\ell)$ fall off like r^{-n} for $n\geqslant\ell$ and define the action of $\beta_{n,\ell}$ on f to be $\overline{\Gamma}_{\ell-1}\overline{\Gamma}_{\ell-2}\cdots\overline{\Gamma}_0 W'_{n-\ell}(r^\ell f)$. Here the operators $\overline{\Gamma}_j$ are generated by the iteration procedure discussed in Sec. IV starting with the ℓth Schwarzschild evolution equation and $W'_{n-\ell}$ is the map, analogous to $W_{n-\ell}$, but associated with the type-0 equation resulting at the end of this iteration procedure. The idea behind the construction of the $\beta_{n,\ell}$ for $\ell\geqslant 1$ is as follows. We first raise the falloff rate of f to $r^{\ell-n}$ by multiplying by r^ℓ. We then use $W'_{n-\ell}$ to map the result over to data that falls off like $r^{\ell-n}$ for the type-0 equation. Finally, this result is mapped to data for the ℓth Schwarzschild evolution equation by applying $\overline{\Gamma}_{\ell-1}\overline{\Gamma}_{\ell-2}\cdots\overline{\Gamma}_0$. The operator $\overline{\Gamma}_{\ell-1}\overline{\Gamma}_{\ell-2}\cdots\overline{\Gamma}_0$ which can be roughly thought of as acting like $\partial^\ell/\partial\tilde{r}^{*\ell}$ will increase the falloff rate back to r^{-n} due to Lemma 1 for the cases when $n>\ell$. However, for $n=\ell$ the map $\beta_{n,\ell}$ is $\overline{\Gamma}_{\ell-1}\overline{\Gamma}_{\ell-2}\cdots\overline{\Gamma}_0 W_0'$ and since W_0' was not defined by iterative procedure from W_1' as were all the other W' maps, we can not invoke Lemma 1 to demonstrate that falloff rates are preserved in this case. However, one can use Proposition 1 to prove a lemma for the map W_0' similar to Lemma 1 which would insure that fall off rates are preserved in this case.

We summarize the properties of $\beta_{n,\ell}$ in the following theorem.

Theorem 6: The map $\beta_{n,\ell}$ constructed above is injective, takes functions in $M(\ell)$ that fall off like r^{-n} for $n\geqslant\ell$ to other functions on Σ that fall off like r^{-n} (equivalently $\tilde{r}^{-n}$) and which when viewed as time symmetric data for the ℓth Schwarzschild evolution equation will have an evolution smoothly extendible to σ^+ and σ^-.

VII. THE TIME ANTISYMMETRIC CASE

We now want to comment on how the $\beta_{n,\ell}$ map can be adapted to handle time antisymmetric data. Using a method similar to that presented in Sec. III, one can show that aside from ℓ particular data sets, time antisymmetric data of the form $(0,g)$ for the ℓth Minkowski equation will smoothly evolve to σ^+ and σ^- if g is of the form $r^{-(\ell+3)}f(1/r)$ where f is smooth. Given that g falls off like r^{-n} for an $n\geqslant\ell+3$, we propose to extend the action of $\beta_{n,\ell}$ to time antisymmetric data by mapping g to $\beta_{n,\ell}(g)$. We will now argue that $(0,\beta_{n,\ell}(g))$ evolved with the ℓth Schwarzschild equation will smoothly extend to σ^+ and σ^-.

Time differentiation in both the Minkowski and Schwarzschild space-times preserves solutions of the scalar wave equations, maps time symmetric fields to time antisymmetric fields, and because the timelike killing fields are smoothly extendible to null infinity, time differentiation preserves asymptotically flat evolution. The action of time differentiation on initial data can naturally be expressed in terms of the static parts of the evolution equations. For example, if ϕ is a solution of the ℓth Minkowski equation with data $(f,0)$ then its time derivative has data $(0,(\partial_r\partial_r-\ell(\ell+1)/r^2)f)$. Therefore, to show that $(0,\beta_{n,\ell}(g))$ evolves to be smoothly extendible to σ^+ and σ^- we need only show that $\beta_{n,\ell}(g)=\Delta_s\beta_{n-2,\ell}(f)$ for some f, where Δ_s is the static part of the ℓth Schwarzschild evolution equation. We claim that the unique choice for f is $f=r^2g$. To demonstrate this, recall that by definition, $\beta_{n,\ell}(g)$ is given by

$$\beta_{n,\ell}(g)=\overline{\Gamma}_{\ell-1}\overline{\Gamma}_{\ell-2}\cdots\overline{\Gamma}_0 W_0' W_{n-\ell}(r^\ell g). \tag{7.1}$$

Now the operator $W_{n-\ell}$, built by iteration, is equal to $\Gamma_0'\Gamma_0 W_{n-(\ell+2)}$. (To see this recall that $\overline{W}_{n-(\ell+1)}=\Gamma_0 W_{n-(\ell+2)}$ and $W_{n-\ell}=\Gamma_0'\overline{W}_{n-(\ell+1)}$.) The operator $\Gamma_0'\Gamma_0$ is just the static part of the type-0 equation the W map is based on. This same static part can be written as $\Gamma_1\overline{\Gamma}_0$. By substituting $\Gamma_1\overline{\Gamma}_0 W_{n-(\ell+2)}$ into (7.1) we get

$$\beta_{n,\ell}(g)=\overline{\Gamma}_{\ell-1}\overline{\Gamma}_{\ell-2}\cdots\overline{\Gamma}_0\Gamma_1\overline{\Gamma}_0 W_{n-(\ell+2)}(r^{\ell+2}g). \tag{7.2}$$

Now, the static part of the type-j equation can be expressed as $\Gamma_{j+1}\overline{\Gamma}_j$ or $\overline{\Gamma}_{j-1}\Gamma_j$. Therefore these two expressions are equivalent. We can use this equivalency to rewrite (7.2) as

$$\beta_{n,\ell}(g)=\Gamma_{\ell-1}\overline{\Gamma}_\ell\overline{\Gamma}_{\ell-1}\overline{\Gamma}_{\ell-2}\cdots\overline{\Gamma}_0 W_{n-(\ell+2)}(r^{\ell+2}g). \tag{7.3}$$

But $\Gamma_{\ell-1}\overline{\Gamma}_\ell$ is just the static part of the ℓth Schwarzschild evolution equation.

By substituting Δ_s for this term we get

$$\beta_{n,\ell}(g) = \Delta_s \bar{\Gamma}_{\ell-1} \bar{\Gamma}_{\ell-2} \cdots \bar{\Gamma}_0 W_{n-(\ell+2)} (r^{\ell+2} g). \tag{7.4}$$

By substituting $\Delta_s \beta_{n-2,\ell}(g)$ on the right we get

$$\beta_{n,\ell}(g) = \Delta_s \beta_{n-2,\ell}(r^2 g). \tag{7.5}$$

Therefore $(0,\beta_{n,\ell}(g))$ is the time derivative of the time symmetric data set $(\beta_{n-2,\ell}(r^2 g),0)$ from which it follows that it will evolve, using the ℓth Schwarzschild evolution equation, to be smoothly extendible to σ^+ and σ^-. Thus we have our final theorem.

Theorem 7: Let g be a function smooth in the variable $1/r$ and let g fall off like r^{-n} for an $n \geqslant \ell + 3$. Then the time antisymmetric data set $(0\beta_{n,\ell}(g))$ evolved with the ℓth Schwarzschild equation will smoothly extend to σ^+ and σ^-.

ACKNOWLEDGMENT

I wish to thank Bob Wald for his persistent encouragement and helpful discussions.

This work was supported in part by NSF Grant PHY 84-16691 to the University of Chicago.

APPENDIX A: ASYMPTOTICALLY REGULAR FUNCTIONS

As defined in Sec. V, $f(x)$ is an asymptotically regular function if f is smooth, the limit of $f(x)$ exists as $x \to \infty$, and $\lim x^n f^{(n)} = 0$ as $x \to \infty$ for all positive integers n, where $f^{(n)}$ is the nth derivative of f. An asymptotically regular function is a generalization of the notion of a function that is smooth in the variable $1/x$ in the sense that the derivative of such a function will fall off, roughly, one power of x faster than the function itself.

For functions on the initial surface Σ, being asymptotically regular in the coordinate r is equivalent to being asymptotically regular in the coordinate $\tilde{r}$ and $\tilde{r}^*$. This fact follows from the relations that hold on Σ:

$$\frac{d}{d\tilde{r}^*} = \left(1 - \frac{4m}{r+2m}\right)\frac{d}{dr}, \tag{A1}$$

$$\frac{d}{d\tilde{r}^*} = \left(1 - \frac{2m}{\tilde{r}}\right)\frac{d}{d\tilde{r}}. \tag{A2}$$

Relation (A1) can be used to show that asymptotic regularity in r is equivalent to asymptotic regularity in $\tilde{r}^*$ while (A2) can be used to show that asymptotic regularity in $\tilde{r}$ is equivalent to asymptotic regularity in $\tilde{r}^*$.

The specific falloff behavior of the derivatives of an asymptotically regular function is captured by the following theorem.

Theorem 4: Let f be an asymptotically regular function and let $c \geqslant 0$ be a constant such that $x^m f$ is bounded as $x \to \infty$ for all $m < c$. Then, for any positive integer k, $x^q f^{(k)}$ is bounded as $x \to \infty$ for all $q < c + k$. Here $f^{(k)}$ is the k th derivative of f with respect to x.

To prove the above theorem we need the following lemma that gives an estimate of the second derivative of a function in terms of the first derivative when the function is bounded.

Lemma 2: Let H be a smooth function satisfying $|H| < b$ for some constant A. Then for every $x_0 \in \mathbb{R}$ there exists an $s \in [x_0 - 4b/|H'(x_0)|, x_0]$ satisfying

$$|H''(s)| \geqslant H'(x)^2/8b. \tag{A3}$$

Proof: Fix x_0 where $H'(x_0) \neq 0$. The function $H'(x)$ must drop to the value $(1/2)H'(x_0)$ somewhere in the domain $[x_0 - 4b/|H'(x_0)|, x_0]$ or the bound, $|H| < b$, would be violated. Let $t \in [x_0 - 4b/|H'(x_0)|, x_0]$ be a point where this occurs. Applying the mean value theorem, we find there exists an $s \in [t, x_0]$ satisfying

$$H''(s) = \frac{H'(x_0) - H'(t)}{x_0 - t} = \frac{(1/2)H'(x_0)}{x_0 - t}. \tag{A4}$$

By using $4b/|H'(x_0)| \geqslant x_0 - t$ we arrive at (A1). Q.E.D.

We now prove the falloff theorem.

Proof: We prove the case $k = 1$ by contradiction. The other cases follow by induction. Assume f is asymptotically regular and that there exists a pair q,c with $q \in (c, c+1)$ such that $x^m f$ is bounded for all $m < c$ and $x^q f'$ is unbounded. Fix m so that $q - m < 1$. Define $g(x) = f(\exp(x))$. The asymptotic regularity of f implies g and all derivatives of g are bounded. That $x^m f$ is bounded and $x^q f'$ is not implies that $h(x) = \exp(mx)g(x)$ is bounded and that $\exp(-\sigma x)h'(x)$ is not, where $\sigma = m + 1 - q$ is positive.

Now, $\exp(-\sigma x)h'(x)$ being unbounded means $h'(x)$ must increase like $\exp(\sigma x)$ or faster. It follows that the higher derivatives $h^{(n)}(x)$ must also increase like $\exp(\sigma x)$ or faster, therefore $\exp(-\sigma x)h^{(n)}(x)$ is unbounded. From the definition $h(x) = \exp(mx)g(x)$ and the fact that $g(x)$ and all its derivatives are bounded, we known that $\exp(-mx)h^{(n)}(x)$ must be bounded. Therefore the supremum α_n of all the α's such that $\exp(-\alpha x)h^{(n)}(x)$ is unbounded must exist and satisfy $\sigma \leqslant \alpha_n \leqslant m$.

We now use the lemma to estimate the α_2. Let $H(x) = h(x)$ and apply the lemma to get $|h''(s)| \geqslant b_1 |h'(x_0)|^2$ for some constant b_1. Multiply the left-hand side by $\exp(-2\kappa s)$ and the right-hand side by $\exp(-2\kappa x_0)$ to get $|\exp(-2\kappa s)h''(s)| \geqslant b_1 \times \exp(-\kappa x_0)h'(x_0)|^2$. If $\kappa < \alpha_1$ then the resulting right-hand side will be unbounded as $x_0 \to \infty$, thus implying the left-hand side $\exp(-2\kappa s)h''(s)$ is unbounded as $s \to \infty$. Therefore α_2 satisfies $2\kappa < \alpha_2$. Since κ was arbitrary, we have $\alpha_2 \geqslant 2\alpha_1$.

To estimate α_3 let $H(x) = \exp(-\lambda x)h'(x)$ where λ is between α_1 and α_2. Then $H(x)$ will be bounded while $H'(x)$ will not be. By applying the lemma we get $|H''(s)| \geqslant b_2 |H'(x_0)|^2$ for some constant b_2. Multiply the left-hand side by $\exp(-2\kappa s)$ and the right-hand side by $\exp(-2\kappa x_0)$ to get $|\exp(-2\kappa s)H''(s)| \geqslant b_2| \times \exp(-\kappa x_0)H'(x_0)|^2$. If $\lambda + \kappa < \alpha_2$ then the right-hand side will remain unbounded as $x_0 \to \infty$, thus implying the left side is unbounded as $s \to \infty$. For values of λ sufficiently close to α_1 and values of κ sufficiently close to $\alpha_2 - \lambda$, all terms on the left-hand side will be bounded except $\exp[-(2\kappa + \lambda)s]h^{(3)}(s)$ which then must be unbounded. Therefore $\alpha_3 \geqslant 2\kappa + \lambda$. Since λ and κ are arbitrary up to the restrictions $\lambda + \kappa < \alpha_2$ and $\alpha_1 < \lambda < \alpha_2$, the supremum α_3 must obey the estimate $\alpha_3 \geqslant 2\alpha_2 - \alpha_1$.

By induction on n, one attains the inequality

$\alpha_{n+1} \geqslant 2\alpha_n - \alpha_{n-1}$. This together with $\alpha_1 \geqslant \sigma$ implies $\alpha_n \geqslant n\sigma$ for all n. But this is a contradiction since $\alpha_n \leqslant m$. Q.E.D.

APPENDIX B: A FALLOFF THEOREM

In this Appendix we prove Proposition 1 that was used in Sec. VI.

Proposition 1: Let $f \in M(0)$ and $\tilde{f} = \alpha(f)$ and denote the difference $\tilde{f} - f$ as Δ. Then the limit as $r \to \infty$ of the function $r^\epsilon \Delta$ vanishes for all $\epsilon < 1$. Furthermore if f vanishes at i^0 then this limit vanishes for all $\epsilon < 2$.

Proof: Let ϕ be the field on $\mathscr{L}$ space resulting from evolving the initial data $(f,0)$ with the $\ell = 0$ Minkowski evolution equation (2.3). Likewise, denote $\tilde{\phi}$ as the field resulting from evolving the initial data $(\tilde{f},0)$ with the type-0 equation having potential $V(1/\tilde{r})$. From our discussion in Sec. V we know that $\tilde{\phi}$ is continuous at i^0 and therefore bounded in a neighborhood. Observing that $V(1/\tilde{r})$ falls off like $1/\tilde{r}^3$, choose c_1 so that in this neighborhood $|\tilde{\phi} V(1/\tilde{r})|$ obeys the bound

$$|\tilde{\phi} V(1/\tilde{r})| \leqslant c_1/\tilde{r}^{*3}. \tag{B1}$$

Let $\Phi_{(1)}$ be the unique solution that vanishes on σ^+ and σ^- of the sourced equation

$$\left(\frac{\partial^2}{\partial \tilde{r}^{*2}} - \frac{\partial^2}{\partial \tilde{t}^2}\right)\Phi = V\left(\frac{1}{\tilde{r}}\right)\tilde{\phi}. \tag{B2}$$

In terms of the coordinates $\hat{u}$ and $\hat{v}$, Eq. (B2) becomes

$$\frac{\partial}{\partial \hat{u}\, \partial \hat{v}}\Phi = \frac{\tilde{\phi} V(1/\tilde{r})}{(1-4m\hat{u})(1-4m\hat{v})\hat{u}^2\hat{v}^2}. \tag{B3}$$

By integrating with respect to $d\hat{u}$ and $d\hat{v}$, we can explicitly write $\Phi_{(1)}(p)$ as

$$\Phi_{(1)}(p) = \int_{H(p)}\int \frac{V(1/\tilde{r})\tilde{\phi}}{(1-4m\hat{u})(1-4m\hat{v})\hat{u}^2\hat{v}^2} d\hat{u}\, d\hat{v}, \tag{B4}$$

where the region $H(p)$ is given in Fig. 2. [We note here that the integrand in (B4) blows up at i^0 and thus threatens the existence $\Phi_{(1)}$. However, we could express the integrand in terms of the finer barred coordinates of Sec. V thus making the integrand continuous at i^0 and hence insuring the existence of $\Phi_{(1)}$. This technique can also be used to insure the existence of the following integrals in this proof whose integrands appear to become singular at i^0.] Next we bound $|\Phi_{(1)}(p)|$ in the neighborhood as follows:

$$|\Phi_{(1)}(p)| \leqslant \int_{H(p)}\int \frac{|V(1/\tilde{r})\tilde{\phi}|}{(1-4m\hat{u})(1-4m\hat{v})\hat{u}^2\hat{v}^2} d\hat{u}\, d\hat{v}, \tag{B5}$$

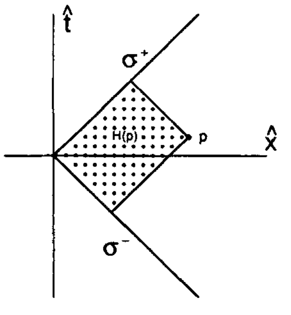

FIG. 2. The region of integration $H(p)$ is shaded.

$$\leqslant \int_{H(p)}\int \frac{c_1/\tilde{r}^{*3}}{(1-4m\hat{u})(1-4m\hat{v})\hat{u}^2\hat{v}^2} d\hat{u}\, d\hat{v}. \tag{B6}$$

The integral (B6) can be viewed as the unique solution that vanishes on σ^+ and σ^- of equation (B2) with source $V(1/\tilde{r})\tilde{\phi}$ replaced by $c_1/\tilde{r}^{*3}$. It simple to check that $c_1(2\tilde{r}^*)$ also satisfies this equation and vanishes on σ^+ and σ^-, therefore it must be equal to the integral in (B6). By substituting $c_1/(2\tilde{r}^*)$ into (B6) we get

$$|\Phi_{(1)}| \leqslant = c_1/2\tilde{r}^*. \tag{B7}$$

Define $\Phi_{(2)}$, by

$$\Phi_{(2)} = \tilde{\phi} - \phi - \Phi_{(1)}. \tag{B8}$$

Since $\Phi_{(2)}$ is a homogeneous solution of (B2) (because $\tilde{\phi}_0$ and $\Phi_{(1)}$ are both particular solutions and ϕ is a homogeneous solution), we can express it as

$$\Phi_{(2)} = \tfrac{1}{2}(g(\tilde{u}) + g(\tilde{v})). \tag{B9}$$

The construction of the map $\alpha[\tilde{V}]$ requires $\tilde{\phi}$ and ϕ differ on σ_v and σ_u by a constant. Therefore $\Phi_{(2)}$ and $-\Phi_{(1)}$ must differ on σ_v and σ_u by the same constant, implying $g(\tilde{u})/2$ restricted to σ_v must equal $-\Phi_{(1)} + c_2$ restricted to σ_v for some c_2. Because $\Phi_{(2)}$ vanishes at i^0 (because $\tilde{\phi}, \phi, \Phi_{(1)}$ each do), we know that $\lim g(\tilde{u})/2 = 0$ as $\tilde{u} \to \infty$, implying that $\lim(-\Phi_{(1)} + c_2) = 0$ as $\tilde{u} \to \infty$ on σ_v. Since $\Phi_{(1)}$ vanishes on σ^- we conclude that $c_2 = 0$. The bound on $\Phi_{(1)}$, (B7), restricted to σ_v, implies the following bound on $g(\tilde{u})$

$$|g(\tilde{u})| \leqslant c_1/(\tilde{u} + \tilde{v}_0), \tag{B10}$$

where $\tilde{v}_0$ is the value of $\tilde{v}$ on σ_v. This, in turn, implies the following bound on $\Phi_{(2)}$:

$$|\Phi_{(2)}| \leqslant \frac{c_1}{2}\left(\frac{1}{\tilde{u} + \tilde{v}_0} + \frac{1}{\tilde{u}_0 + \tilde{v}}\right) \tag{B11}$$

$$\leqslant \frac{c_1}{2}\left(\frac{1}{\tilde{u}} + \frac{1}{\tilde{v}}\right) \tag{B12}$$

$$= \frac{c_1 \tilde{r}^*}{(\tilde{r}^{*2} - \tilde{t}^2)}. \tag{B13}$$

Therefore $(\tilde{r}^{*2} - \tilde{t}^2)\Phi_{(2)}/\tilde{r}^*$ is bounded. Note that ϕ obeys a similar bound and from (B7), so does $\Phi_{(1)}$. Hence $(\tilde{r}^{*2} - \tilde{t}^2)\tilde{\phi}/\tilde{r}^*$ is bounded.

Therefore $(\tilde{r}^{*2} - \tilde{t}^2)\Phi_{(2)}/\tilde{r}^*$ is bounded. From (B7) we can conclude that $\Phi_{(1)}$ obeys the same bound. Hence $(\tilde{r}^{*2} - \tilde{t}^2)(\tilde{\phi} - \phi)/\tilde{r}^*$ is bounded from which it follows that $\tilde{r}^*\Delta$ is bounded, forcing the first conclusion of the theorem.

Now suppose that f vanishes at i^0. Then $(\tilde{r}^{*2} - \tilde{t}^2)\phi/\tilde{r}^*$ will be bounded since ϕ is now a homogeneous solution that vanishes at i^0. Therefore $(\tilde{r}^{*2} - \tilde{t}^2)\tilde{\phi}/\tilde{r}^*$ is bounded implying there exists a positive constant c_2 such that

$$|\tilde{\phi} V(1/\tilde{r})| \leqslant c_2/\tilde{r}^{*2}(\tilde{r}^{*2} - \tilde{t}^2). \tag{B14}$$

The corresponding bound on $\Phi_{(1)}$ is

$$|\Phi_{(1)}(p)| \leqslant \int_{H(p)}\int \frac{|V(1/\tilde{r})\tilde{\phi}|}{(1-4m\hat{u})(1-4m\hat{v})\hat{u}^2\hat{v}^2} d\hat{u}\, d\hat{v} \tag{B15}$$

$$\leqslant \int_{H(p)} \int \left(\frac{1}{(1-4m\hat{u})(1-4m\hat{v})\hat{u}^2\hat{v}^2} \right) \times \frac{c_2}{\tilde{r}^{*2}(\tilde{r}^{*2}-\tilde{t}^{2})} d\hat{u}\, d\hat{v} \tag{B16}$$

$$= 2c_2\left(\left(\frac{1}{\tilde{u}^2} + \frac{1}{\tilde{v}^2} \right) \ln \left(\frac{1}{\tilde{u}} + \frac{1}{\tilde{v}} \right) - \frac{1}{\tilde{u}\tilde{v}} + \frac{1}{\tilde{u}^2} \ln(\tilde{u}) + \frac{1}{\tilde{v}^2} \ln(\tilde{v}) \right) \Bigg|_p . \tag{B17}$$

The last integral, (B16), was evaluated to be (B17) because it represents the unique solution, that vanishes on σ^+ and σ, to Eq. (B2) with the source $V(1/\tilde{r})\tilde{\phi}$ replaced by terms on the right side of (B14). Let $\Phi_{0(1)}$ and $\Phi_{0(2)}$ denote the restriction of $\Phi_{(1)}$ and $\Phi_{(2)}$ to the Σ, respectively. The bound (B17) on $\Phi_{(1)}$ implies that $\Phi_{0(1)}$ satisfies

$$\lim_{r\to\infty} (r^{\epsilon}\Phi_{0(1)}) = 0 \tag{B18}$$

for $\epsilon < 2$. This bound on $\Phi_{(1)}$ also leads to the bound on $g(\tilde{u})$ of

$$|g(\tilde{u})| \leqslant 4c_2\left(\left(\frac{1}{\tilde{u}^2} + \frac{1}{\tilde{v}_0^2} \right) \ln \left(\frac{1}{\tilde{u}} + \frac{1}{\tilde{v}_0} \right) - \frac{1}{\tilde{u}\tilde{v}_0} + \frac{1}{\tilde{u}^2} \ln(\tilde{u}) + \frac{1}{\tilde{v}_0^2} \ln(\tilde{v}_0) \right), \tag{B19}$$

implying that $\Phi_{0(2)}$ satisfies

$$\lim_{r\to\infty} (r^{\epsilon}\Phi_{0(2)}) = 0 \tag{B20}$$

for $\epsilon < 2$. But $\Delta = \Phi_{0(2)} + \Phi_{0(1)}$, thus proving the theorem. Q.E.D.

[1]R. Penrose, Phys. Rev. Lett. **10**, 66 (1963).

[2]J. Winicour, Found. Phys. **15**, 605 (1985).

[3]S. Klainerman in "Mathematics and general relativity," in *Contemporary Mathematics*, edited by J. Isenberg, (Am. Math. Soc., Providence, RI, 1988), Vol. 71.

[4]H. Friedrich, Commun. Math. Phys. **107**, 587 (1986).

[5]R. P. Geroch and A. C. Xanthopoulos, J. Math. Phys. **19**, 714 (1978).

[6]A. G. Schmidt and J. M. Stewart, Proc. R. Soc. London Ser. A **367**, 503 (1979).

[7]J. Porill and J. M. Stewart, Proc. R. Soc. London Ser. A **376**, 451 (1981).

[8]A. Ashtekar and R. Hansen, J. Math. Phys. **19**, 1542 (1978).

[9]S. Hawking and G. Ellis, *The Large Scale Structure of Space-Time* (Cambridge U. P., Cambridge, 1973).

Second-order equations from a second-order formalism

Richard T. Hammond
Code 5052, Naval Air Development Center, Warminster, Pennsylvania 18974

(Received 11 March 1988; accepted for publication 28 December 1988)

It is often assumed that Lagrangians of gravitation that are quadratic in the curvature tensor produce field equations of fourth differential order in the metric tensor from a Hilbert variational principle. It is shown here, for the Lagrangian given by $R + R_{\mu\nu}R^{\mu\nu}$, that independent variations of the metric tensor and the torsion tensor produce gravitational field equations of second, not fourth, differential order.

I. INTRODUCTION

It is generally argued that if terms quadratic in the curvature tensor are present in the Lagrangian of a metric theory of gravity, then, excluding trivial cases that reduce to general relativity, the gravitational field equations will be of fourth differential order in the metric tensor if the Hilbert variational principle is used. This argument, which is indeed true without torsion present, is generally used either to exclude quadratic terms from the Lagrangian, or to cause one to adopt the so-called first-order formalism, in which independent variations of the metric tensor and affine connection are performed. In fact, it has been proven[1] that the only Lagrangian in v_4 that yields second-order equations is given by R, the curvature scalar, plus Gauss–Bonnet type collections of quadratic terms in the curvature scalar. Thus for any general collection of quadratic terms in the curvature tensor, i.e., terms with arbitrary coupling constants, this theorem proves that the equations will be of fourth differential order.

It is the purpose of this paper to show that in a u_4 spacetime with independent variations of the metric tensor and the torsion tensor, this theorem fails. This is shown for the particular Lagrangian used in Eq. (1). In particular, calling the equations resulting from varying the metric tensor the gravitational field equations (GFE's), and the term torsional field equations (TFE's) meaning those resulting from the torsional variations, it is shown that the TFE's can be used in the GFE's to reduce the order from four to two. The TFE's will be of second differential order in the torsion, but will contain third derivative terms of the metric tensor. However, a set of four independent TFE's will be derived that contain no derivative higher than the second.

II. THE FIELD EQUATIONS

There have been numerous uses of quadratic Lagrangians over the years.[2] The particular Lagrangian used here contains, in addition to the usual scalar of general relativity, a quadratic term in the Ricci tensor. This is the simplest Lagrangian quadratic in the curvature tensor which will produce propagating torsion. Details concerning how to perform the following variations may be found in the literature.[3] Thus the variation principle takes the form

$$\delta \int \sqrt{-g}\,(R + nR_{\mu\nu}R^{\mu\nu})\,d^4x = 0, \tag{1}$$

where $R = g^{\mu\nu}R_{\mu\nu}$ and $R_{\mu\nu} = R_{\sigma\mu\nu}{}^{\sigma}$. Also, the torsion is defined as $S_{\alpha\beta}{}^{\gamma} = \Gamma_{[\alpha\beta\,]}{}^{\gamma}$, where brackets (parentheses) imply taking the antisymmetric (symmetric) part, and the definitions here and in the following are those of Schouten.[4] Considering variations of $g_{\mu\nu}$ while holding $S_{\alpha\beta}{}^{\gamma}$ fixed yields

$$-G^{\mu\nu} + \overset{*}{\nabla}_{\sigma}T^{\mu\nu\sigma} + nH^{\mu\nu} + \underset{\mu\nu}{\mathrm{SYM}}\Big[\overset{*}{\nabla}_{\sigma}(T^{\nu\mu\sigma} + T^{\sigma\mu\nu} + T^{\sigma\nu\mu}) + n\overset{*}{\nabla}_{\sigma}(C^{\nu\sigma\mu} + C^{\sigma\nu\mu} - C^{\mu\nu\sigma})\Big] = 0, \tag{2}$$

where

$$\overset{*}{\nabla}_{\sigma} A^{\mu} = \nabla_{\sigma}A^{\mu} + 2S_{\sigma}A^{\mu}, \tag{3}$$

$$H^{\mu\nu} = \tfrac{1}{2}g^{\mu\nu}R^{\alpha\beta}R_{\alpha\beta} - R^{\mu\sigma}R^{\nu}{}_{\sigma} - R^{\sigma\mu}R_{\sigma}{}^{\nu}, \tag{4}$$

$$T^{\alpha\beta\gamma} = S^{\alpha\beta\gamma} + S^{\beta}g^{\alpha\gamma} - S^{\alpha}g^{\beta\gamma}, \tag{5}$$

$$C^{\alpha\beta\gamma} = -\overset{*}{\nabla}{}^{\gamma}R^{\alpha\beta} + g^{\alpha\gamma}\overset{*}{\nabla}_{\phi}R^{\phi\beta} + 2R_{\phi}{}^{\beta}S^{\gamma\phi\alpha}, \tag{6}$$

and where $G^{\mu\nu}$ is the Einstein tensor. It has been assumed that $\nabla_{\sigma}g_{\mu\nu} = 0$ so that

$$\Gamma_{\alpha\beta}{}^{\sigma} = \{{}^{\sigma}_{\alpha\beta}\} + S_{\alpha\beta}{}^{\sigma} + 2S^{\sigma}{}_{(\alpha\beta)}, \tag{7}$$

where $\{{}^{\sigma}_{\alpha\beta}\}$ represents the Christoffel symbol.

As is often stated, (2), with (6), shows that the GFE's are of fourth differential order in the metric tensor. With no torsion, and assuming that $\nabla_{\sigma}g_{\mu\nu} = 0$, this conclusion is inescapable. With torsion present, however, the TFE's impose the constraints necessary to cause the higher derivatives to drop out of the GFE's.

To show this, consider the TFE's which are obtained by varying $S_{\alpha\beta}{}^{\gamma}$ while holding $g_{\mu\nu}$ fixed. This yields, from (1),

$$\underset{\alpha\beta}{\mathrm{ANT}}[T^{\alpha\beta\gamma} - T^{\beta\gamma\alpha} + T^{\alpha\gamma\beta} + n(C^{\alpha\beta\gamma} + C^{\beta\gamma\alpha} + C^{\gamma\beta\alpha})] = 0, \tag{8}$$

where $\mathrm{ANT}_{\alpha\beta}$ implies antisymmetrization in $\alpha\beta$.

From this, it is straightforward to show that (8) may be put in the form

$$T^{\gamma\beta\alpha} = -nC^{\alpha[\beta\gamma]}. \tag{9}$$

Also, using the definition

$$D^{\alpha\beta\gamma} \equiv T^{\alpha\beta\gamma} + nC^{\gamma\beta\alpha}, \tag{10}$$

one may show that

$$P\{D^{[\alpha\beta]\gamma}\} = 0, \tag{11}$$

where P stands for permutation over all the indices. This form will be used shortly.

Now turn back to (2), which with (9) and (10) can be written as

 0022-2488/89/051115-02$02.50

$$-G^{\mu\nu}+\overset{*}{\nabla}_\sigma T^{\mu\nu\sigma}+nH^{\mu\nu}-\overset{*}{\nabla}_\sigma D^{\mu\nu\sigma}=0. \tag{12}$$

To boil this down, first note the identity,[5] for any tensor $A^{\sigma\lambda}$,

$$\nabla_{[\nu}\nabla_{\mu]}A^{\sigma\lambda}=\tfrac{1}{2}R_{\nu\mu\rho}{}^{\sigma}A^{\rho\lambda}+\tfrac{1}{2}R_{\nu\mu\rho}{}^{\lambda}A^{\sigma\rho}-S_{\nu\mu}{}^{\rho}\nabla_\rho A^{\sigma\lambda}. \tag{13}$$

When this is used in (12) with (6), one obtains

$$\begin{aligned}\overset{*}{\nabla}_\sigma D^{\mu\nu\sigma}/n = {} & -R^{\mu}{}_{\rho}R^{\rho\nu}-R^{\sigma\mu\rho\nu}R_{\sigma\rho} \\ & +2S_\sigma{}^{\mu\rho}\nabla_\rho R^{\sigma\nu}+2T^{\mu\xi\sigma}\nabla_\sigma R_\xi{}^{\nu} \\ & +2R_\xi{}^{\nu}\nabla_\sigma T^{\mu\xi\sigma}+\nabla_\sigma T^{\mu\nu\sigma}/n+2S_\sigma(T^{\mu\nu\sigma}/n \\ & -\nabla^{\mu}R^{\sigma\nu}+2R_\xi{}^{\nu}T^{\mu\xi\sigma})+2S^{\mu}\nabla_\xi R^{\xi\nu}.\end{aligned} \tag{14}$$

This simplifies with (5) and the GFE's (12) become, using (14),

$$\begin{aligned}&-G^{\mu\nu}+n(H^{\mu\nu}+R^{\mu}{}_{\sigma}R^{\sigma\nu} \\ &\quad +R_{\sigma\rho}R^{\sigma\mu\rho\nu}-2R_\xi{}^{\nu}\overset{*}{\nabla}_\sigma T^{\mu\xi\sigma})=0,\end{aligned} \tag{15}$$

or,

$$-G^{\mu\nu}+nN^{\mu\nu}=0, \tag{16}$$

$$N^{\mu\nu}=\tfrac{1}{2}g^{\mu\nu}R^{\alpha\beta}R_{\alpha\beta}-R^{\mu\sigma}R^{\nu}{}_{\sigma}-R^{\mu}{}_{\alpha\beta}{}^{\nu}R^{\alpha\beta}. \tag{17}$$

Thus it is evident that the GFE's are of second, not fourth, differential order. Moreover, (16) and (17) show that the field equations are much simpler than their original form (2) due to the use of the TFE's. For weak fields, the quadratic terms will be small compared to the linear terms (in $R_{\sigma\mu\nu}{}^{\gamma}$) and the field equations reduce to the same form as those of general relativity, but, of course, $G^{\mu\nu}$ contains the torsional terms.

Finally, consider the TFE's (9). It is seen that they are of second differential order in the torsion tensor, but they do contain third derivatives of the metric tensor. However, using (5) and (6) with (11), and performing the permutation, one obtains

$$P\{S^{\alpha\beta\gamma}+n\nabla^{\alpha}(\overset{*}{\nabla}_\phi T^{\beta\gamma\phi})-n(R_\phi{}^{\alpha}T^{\beta\phi\gamma}-R_\phi{}^{\alpha}T^{\gamma\phi\beta})\}=0, \tag{18}$$

which contains no derivatives of higher than second order. However, (18) represents only four independent equations, and therefore, the total set of TFE's still contains third derivatives of the metric tensor.

The main conclusion of this paper is contained in (16), which shows that the GFE's are of second differential order. This is particularly useful when considering coupling to matter, since, e.g., Havas[6] has shown that fourth-order GFE's (resulting from a quadratic Lagrangian) lead to an incorrect equation of motion for an extended source. The work presented in the present paper does not constitute a complete theory at this stage because the coupling to matter has not been considered. This area of investigation is currently under way.

[1]D. Lovelock and H. Rund, *Tensors, Differential Forms, and Variational Principles* (Wiley, New York, 1975), Sec. 8.4.

[2]See, e.g., V. Szczrba, Phys. Rev. D **36**, 351 (1987), and the many references contained therein.

[3]R. Hammond, Gen. Relativ. Gravit. **20**, 813 (1988).

[4]J. Schouten, *Ricci Calculus* (Springer, Berlin, 1954).

[5]See Ref. 4, p. 140.

[6]P. Havas, Gen. Relativ. Gravit. **8**, 631 (1977).

The case against intermediate statistics

B. H. Lavenda
Università di Camerino, Dipartimento di Scienze Chimiche, via Sant'Agostino 1, Camerino 62032 (MC), Italy

J. Dunning-Davies
Department of Applied Mathematics, University of Hull, Cottingham Road, Hull, HU6 7RX, United Kingdom

(Received 27 July 1988; accepted for publication 14 December 1988)

It is shown that intermediate statistics do not correspond to any physical process. The stationary probability distributions of intermediate statistics are not compatible with any mechanism which allows a variation between Fermi–Dirac and Bose–Einstein statistics. The binomial and negative binomial distributions, characterizing Fermi–Dirac and Bose–Einstein statistics, respectively, transform into the Poisson distribution, descriptive of classical statistics, as the number of energy cells increases without limit. These distributions are shown to be the laws of error leading to the average value as the most probable value.

I. INTRODUCTION

Over the years speculations have been made on an intermediate statistics lying between Bose–Einstein (BE) and Fermi–Dirac (FD) statistics. Instead of the occupation number of the energy levels being infinite or one, as in the case of BE or FD statistics, respectively, the occupation number would have some finite value. Such an intermediate form of statistics was first proposed by Gentile[1] who compared its thermodynamic properties with the two known statistics. Basing his argument on quantum mechanical considerations, Sommerfeld[2] concluded that the only application that intermediate statistics may have is in the case where the number of particles n equals the number of energy levels or "cells" d. Müller[3] treated intermediate statistics from a purely academic point of view. Both Wergeland[4] and Schubert[5] contended that it did not matter whether the case $d = n$ or $d = \infty$ was considered since both reduce to BE statistics. ter Haar[6] argued that by putting the occupancy number equal to the number of particles there should be no difference between intermediate statistics and BE statistics simply because there are not enough particles to make the differences apparent. Finally, Guénault and MacDonald[7] showed how intermediate statistics could be exploited to make a gradual transition from FD statistics to BE statistics. The thermodynamics of intermediate statistics was also commented upon by Landsberg.[8]

The argument of ter Haar,[6] and also of Sommerfeld,[2] Wergeland,[4] and Schubert,[5] appears to go against the grain of certain limit theorems in probability theory which assert that as the number of indistinguishable particles increases without limit, while the number of energy cells remains finite, the probability distribution should converge to a normal one.[9] According to the above mentioned authors, it is in this case where the differences between the intermediate and the known forms of quantum statistics should be manifested. Rather, according to the limit theorems of probability theory, the statistics should become classical in the limit.

In this paper, we show that BE statistics arises *only* in the limit as $d = \infty$. All values of d, between one and infinity, give rise to intermediate statistics which, however, cannot be described by stationary probability distributions that tend to the binomial and negative binomial distributions as $d \rightarrow 1$ and $d \rightarrow \infty$, respectively.[10] In other words, there is no stationary probability distribution which is compatible with the physical processes that generate FD and BE statistics. Realizing that there is no purely *thermodynamic* argument that can be given in favor or against the existence of intermediate statistics, we employ a *statistical* argument by showing that the stationary probability distribution can only satisfy the recursion relation obtained from the time independent master equation when it coincides with the binomial, negative binomial, or Poisson distribution corresponding, respectively, to FD, BE, or classical statistics. The Poisson case was overlooked by all the previous studies on intermediate statistics. Classical statistics will be shown to emerge as the number of energy cells increases without limit independently of the magnitude of the occupation number. It can never depend on the size of the particle number since the conclusions drawn from the generating function and its binomial expansion, in which the particle number appears as a mere index, must lead to the same conclusions. This lays to rest the possibility of there being still yet other types of particles with spins different from semi-integral or integral values.

II. DERIVATION OF THE DISTRIBUTION FUNCTION

Consider the generating function

$$\Xi(s) = \left(p \sum_{k=0}^{d} (qs)^k\right)^m = \left(\frac{p[1-(qs)^{d+1}]}{(1-qs)}\right)^m, \quad (1)$$

without necessarily requiring that $p + q = 1$. Two limiting cases are known to have direct physical meanings: In the case $d = 1$ and $q = 1/p - 1$ the generating function reduces to

$$\Xi_{\mathrm{FD}}(s) = \{p(1+qs)\}^m, \quad (2)$$

which is the generating function for the binomial distribution; while for $d = \infty$ and $p + q = 1$ it becomes

$$\Xi_{\mathrm{BE}}(s) = (p/(1-qs))^m, \quad (3)$$

the generating function for the negative binomial distribu-

 0022-2488/89/051117-05$02.50

tion. Using the binomial expansion in (1), the numerator and denominator can be expanded as

$$[1-(qs)^a]^m=\sum_{k=0}^{m}(-1)^k\binom{m}{k}(qs)^{ak} \tag{4}$$

and

$$[1-qs]^{-m}=\sum_{j=0}^{\infty}\binom{-m}{j}(-qs)^j, \tag{5}$$

where $\binom{-m}{j}=(-1)^j\binom{m+j-1}{j}$ and $a\equiv d+1$. The product of the two series (4) and (5) can be written as the convolution

$$\Xi(s)=p^m\sum_{n=0}^{dm}\vartheta(n;m,d)(qs)^n \tag{6}$$

since only $dm+1$ terms in the sum are different from zero. The coefficient

$$\vartheta(n;m,d)=\sum_{k=0}^{[n/a]}(-1)^k\binom{m+n-ak-1}{m-1}\binom{m}{k}, \tag{7}$$

where $[n/a]$ is the largest integer $\leqslant n/a$. The probability of n

$$f^s(n;m,d)=\vartheta(n;m,d)q^np^m \tag{8}$$

is simply the coefficient of s^n in (6).

Apart from the missing $(-1)^k$, this coefficient was found by ter Haar.[6] He observed that for $d=1$, $\vartheta(n;m,1)=\binom{m}{n}$ while for $d=\infty$, $\vartheta(n;m,\infty)=\binom{m+n-1}{n}$ since $d=\infty$ means that we must take only the $k=0$ term in (7). ter Haar, like Wergeland[4] and Schubert,[5] remarked that it did not make any difference whether one took $d=n$ or $d=\infty$ and went on to conclude that "any effects arising from intermediate statistics should be spurious." However, to set $d=n$ has no meaning since n, the realization of a random variable representing the number of particles, can take on all values from 0 to dm. Rather, the occupancy number d is fixed in advance so that the only way for the coefficient $\vartheta(n;m,d)$ to reduce to the binomial coefficient and

$$p^m\sum_{n=0}^{dm}\binom{m+n-1}{n}(qs)^n$$

to coincide with the generating function is to set $d=\infty$. In fact, it is apparent from the generating function (1) that any finite value of d will lead to a finite number of terms in the binomial series expansion and the resulting probability distribution will be different from the negative binomial distribution. In other words, BE statistics arises *only* in the strict limit $d=\infty$. It is important to emphasize that this conclusion could have only been reached by working with *bona fide* probability distributions, or generating functions, rather than with the usual binomial coefficients.[10]

In the case $d=1$, application of the binomial theorem gives

$$\sum_{n=0}^{m}\binom{m}{n}q^n=(1+q)^m=p^{-m} \tag{9}$$

or $p=1/(1+q)$. In the other extreme limit where $d=\infty$, the binomial theorem gives

$$\sum_{n=0}^{m}\binom{-m}{n}(-q)^n=(1-q)^{-m}=p^{-m}. \tag{10}$$

In the limit where $m\to\infty$ and $q\to 0$ such that $qm\to\lambda$ we have

$$\sum_{n=0}^{\infty}\frac{\lambda^n}{n!}=e^\lambda=p^{-m}. \tag{11}$$

The fact that the generating function (1) is the mth power of the ratio $p[1-(qs)^a]/(1-qs)$ means that $f^s(n;m,d)$ is the distribution of a sum $\mathbf{S}_m=\mathbf{X}_1+\cdots+\mathbf{X}_m$ of m mutually independent variables with a common generating function. In the case $d=1$, each variable $\mathbf{X}_i$ assumes the value 0 with probability p and the value 1 with probability pq such that $p(1+q)=1$. For $d=\infty$, the random variable $\mathbf{X}_i$ can be thought of as the number of failures following the $(i-1)$st and preceding the ith success. The sum $\mathbf{S}_m$ is then the total number of failures preceding the mth success. The probability that there are exactly n failures preceding the mth success is $\Pr\{\mathbf{S}_m=n\}=f^s(n;m,\infty)$, where $\mathbf{S}_m+m$ is the number of trials up to and including the mth success.

For the historical record, we note that the derivation of the probability distribution (8) can be traced back to De Moivre for the probability of obtaining a score $m+n$ in a throw of m dice.[11] Suppose that each of the random variables $\mathbf{X}_i$ can assume only a finite number of values, say $0, 1, 2, \ldots, d$ and that each occurs with probability $1/a$ where $a=d+1$. As far as intermediate statistics is concerned, $\mathbf{X}_i$ can be thought of as level i which can accommodate d particles. In the case $d=1$, the probability of there being a particle is simply $\frac{1}{2}$. The probability that $\mathbf{S}_m=n$ is

$$\Pr\{\mathbf{S}_m=n\}=\frac{1}{a^m}\sum_{k=0}^{[n/a]}(-1)^k\binom{m}{k}\binom{m+n-ak-1}{m-1} \tag{12}$$

and the probability that the levels will contain no more than n particles is

$$\Pr\{\mathbf{S}_m\leqslant n\}=\frac{1}{a^m}\sum_{k=0}^{[n/a]}(-1)^k\binom{m}{k}\binom{m+n-ak}{m}. \tag{13}$$

In the limit as $n\to\infty$ and $a\to\infty$ such that their ratio $n/a\to x$,

$$\Pr\{\mathbf{S}_m\leqslant n\}\to\frac{1}{m!}\sum_{k=0}^{[x]}(-1)^k\binom{m}{k}(x-k)^m, \tag{14}$$

where the sum is over all k for which $0\leqslant k<x$. This expression was first derived by Lagrange[11] and states that in the limit, the sum of m independent random variables is uniformly distributed in the interval 0,1 with a constant density of one.[12] In the case $d=1$, $\Pr\{\mathbf{S}_m\leqslant n\}=(1/2^m)\binom{m-1}{n}$ while in the case $d=\infty$ with $n\to\infty$ such that their ratio tends to a constant, $x<1$, $\Pr\{\mathbf{S}_m\leqslant\infty\}=(1/m!)(n/a)^m$.

We will now show that distributions of the form (12) are not compatible with the physical processes that generate the known forms of physical statistics. Said differently, there is no stationary probability distribution that can transform from one statistics to the other with the variation of a characteristic parameter.

III. CRITERION FOR THE STATIONARY PROBABILITY DISTRIBUTION

The stationary probability distributions can be derived from a dynamical equilibrium between the rates of adsorption, $\beta(m-\eta n)$, and desorption, αn, on a lattice of m sites, in the case $\eta = 1$ or, in the case $\eta = -1$, the two rates can be thought of as those for absorption and emission (spontaneous + stimulated) of radiation, as in the Einstein mechanism.[13] In the intermediate case, $\eta \in [-1,1]$, where an intermediate form of statistics should result lying between FD and BE statistics. However, the intermediate statistics is itself spurious insofar as the stationary probability distribution (8) provides for only three constant values of η. All other values of η depend on the number of particles n and consequently do not correspond to the physical processes under consideration. Figuratively speaking, one may consider the values of η to be "quantized" which correspond to stationary probability distributions belonging to exponential families of distributions.

The master equation governing the elementary transitions is

$$\dot{f}(n,t) = \alpha(n+1) f(n+1,t) + \beta[m - \eta(n-1)] f(n-1,t) - [\alpha n + \beta(m - \eta n)] f(n,t). \tag{15}$$

If (8) is to be the stationary solution to this equation then $q = \beta/\alpha$ and setting equal to zero the coefficients of the different powers of β/α give

$$\eta = \frac{m}{n} - \left(\frac{n+1}{n}\right) \frac{\vartheta(n+1;m,d)}{\vartheta(n;m,d)} \tag{16}$$

and

$$\eta = \frac{m}{n-1} - \frac{n}{n-1} \frac{\vartheta(n;m,d)}{\vartheta(n-1;m,d)}. \tag{17}$$

Together, (16) and (17) imply that ϑ satisfies the recursion relation

$$m + (n^2 - 1)\frac{\vartheta(n+1;m,d)}{\vartheta(n;m,d)} = n^2 \frac{\vartheta(n;m,d)}{\vartheta(n-1;m,d)}, \tag{18}$$

which has been obtained by eliminating η between the two equations.

Not any coefficient $\vartheta(n;m,d)$ will do, for it was supposed that η is a parameter and independent of n. This means that any acceptable value of the ratio of the ϑ's must produce a fixed value of η. There are only two cases, plus one limiting form, for which this will be true. Hence there is no stationary probability distribution, describing an intermediate form of statistics, which can be derived from the stationary condition of the master equation (15). The following three forms of statistics exhaust the possibilities and there is no additional form that could describe another kind of particle.

For $\eta = 1$, expression (16) reduces to

$$\frac{m-n}{n+1} = \frac{\varphi(n+1;m,1)}{\vartheta(n;m,1)} = \frac{\binom{m}{n+1}}{\binom{m}{n}}, \tag{19}$$

corresponding to FD statistics, while for $\eta = -1$ there results

$$\frac{m+n}{n+1} = \frac{\vartheta(n+1;m,\infty)}{\vartheta(n;m,\infty)} = \frac{\binom{m+n}{n+1}}{\binom{m+n-1}{n}}, \tag{20}$$

corresponding to BE statistics. In addition, there is a third case, $\eta = 0$, which has gone unnoticed by all the authors cited above. For $\eta = 0$ (16) reduces to

$$\frac{m}{n+1} = \frac{\vartheta(n+1;m,d)}{\vartheta(n;m,d)} = \frac{m^{n+1}/(n+1)!}{m^n/n!}, \tag{21}$$

for which the stationary probability distribution (8) transforms into the Poisson distribution

$$f^s(n) = \frac{(qm)^n}{n!} e^{-qm}. \tag{22}$$

It will now be appreciated that the Poisson distribution (22) is a limiting probability distribution and yet, remarkably enough, it satisfies the recursion formula (16) exactly.

Unlike the binomial and negative binomial distributions, corresponding to $d = 1$ and $d = \infty$, respectively, the Poisson distribution depends only on the fact that m is sufficiently large or, more precisely, as $m \to \infty$ and $q \to 0$ such that $qm = \lambda$. The generating function (1) can then be written as

$$\Xi(s) = \left(\frac{(1-\lambda/m)[1-(\lambda s/m)^a]}{1-\lambda s/m}\right)^m. \tag{23}$$

Passing to logarithms it can easily be seen that the right-hand side of (23) tends to $e^{-\lambda + \lambda s}$, independently of the magnitude of the occupation number. We also emphasize that there can never be a condition on the number of particles which appears only as an index in the convolution (6) of binomial series expansions (4) and (5) and not in the expression for the generating function (1) itself. In the limit as $m \to \infty$, we have

$$\sum_{k=0}^{[n/a]} (-1)^k \binom{m}{k} \binom{m+n-ak-1}{m-1} \to \frac{m^n}{n!}, \tag{24}$$

where terms of the order m^{-d} have been neglected. This shows that BE statistics transforms into classical statistics at a much earlier stage than FD statistics with the increase of m. It also shows that the distinguishability or lack of distinguishability of the particles depends on the fact that the number of energy cells increases without limit independently of d.

The average value $\bar{n}^s$ can be calculated by differentiating the logarithm of the generating function (1)

$$\ln \Xi(s) = m \ln p + m \ln[1 - (qs)^a] - m \ln(1 - qs) \tag{25}$$

with respect to s and evaluating the expression at $s = 1$. We then obtain

$$\frac{\bar{n}^s}{m} = \frac{q}{1-q} - a\frac{q^a}{1-q^a}, \tag{26}$$

which was first derived by Gentile.[1] For $a = 2$ ($d = 1$), expression (26) reduces to

$$\bar{n}^s = mq/(1+q), \tag{27}$$

while for $a = \infty$ it becomes

$$\bar{n}^s = mq/(1-q). \tag{28}$$

A general expression for $\bar{n}^s$ can be found with the aid of expression (16). Multiplying (16) by q^n and summing over n gives $q = \bar{n}^s/(m - \eta\bar{n}^s)$ or

$$\bar{n}^s = m/(\eta + q^{-1}). \tag{29}$$

We now turn to a thermodynamic analysis that will allow us to derive an expression for q. This will enable us to identify (27) and (28) with FD and BE statistics, respectively, while (29) will be identified with intermediate statistics for $\eta \in (-1,1)$.

IV. GAUSS'S PRINCIPLE AND THE SECOND LAW

There is nothing thermodynamically prohibitive to restricting the occupancy of the energy levels. Yet while thermodynamics makes no pronouncement on the existence or nonexistence of intermediate statistics, it can clarify some of the results obtained above. Furthermore, we will show that the average number of particles coincides with the most probable value only for the three known statistics.

With the aid of Gauss's principle, the stationary probability distribution (8) can be cast as a law of error for which the average number of particles is the most probable value. The law of error has the form[10]

$$\ln f^s(n) = -(n-\bar{n}^s)\frac{d\mathscr{S}}{d\bar{n}^s} - \mathscr{S}(\bar{n}^s) + \Sigma(n), \quad (30)$$

where $\mathscr{S}(\bar{n}^s)$ is the entropy and $\Sigma(n)$ is the stochastic entropy. According to the second law,

$$\frac{d\mathscr{S}}{d\bar{n}^s} = \frac{\partial\mathscr{S}}{\partial\bar{n}^s} + \frac{\partial\mathscr{S}}{\partial\mathscr{E}}\frac{d\mathscr{E}}{d\bar{n}^s} = \frac{-\mu+\epsilon}{T}, \quad (31)$$

where the average energy $\mathscr{E} = \bar{n}^s\epsilon$, μ is the chemical potential, and T is the absolute temperature measured in energy units. Introducing the probability distribution (8) into the left-hand side of (30) and comparing terms that are independent and dependent on n give

$$\frac{d\mathscr{S}}{d\bar{n}^s} = \ln\left(\frac{m-\eta\bar{n}^s}{\bar{n}^s}\right) = -\ln q, \quad (32)$$

$$\bar{n}^s\frac{d\mathscr{S}}{d\bar{n}^s} - \mathscr{S} = \frac{m}{\eta}\ln\left(\frac{m-\eta\bar{n}^s}{m}\right) = m\ln p, \quad (33)$$

and

$$\Sigma(n) = \ln\vartheta(n;m,d). \quad (34)$$

Equating (31) and (32) results in

$$\ln q = (\mu-\epsilon)/T. \quad (35)$$

For $\eta=-1$, we have $q=\bar{n}^s/(m+\bar{n}^s)$ and $p=m/(m+\bar{n}^s)$ while for $\eta=1$, $q=\bar{n}^s/(m-\bar{n}^s)$ and $p=(m-\bar{n}^s)/m$ with $q=1/p-1$. In terms of the probabilities of "success" and "failure," with $\tilde{p}+\tilde{q}=1$, we get $\tilde{p}=q/(1+q)$ and $\tilde{q}=1/(1+q)$. Finally, in the case $\eta=0$, we have $p=e^{-q}$ where $q=\bar{n}^s/m$ or, in terms of the normalized probabilities, $\tilde{p}=p$ and $\tilde{q}=1-e^{-q}$.

For values of $\eta\neq 0$, Eq. (32) can be integrated to obtain

$$\mathscr{S}(\bar{n}^s) = -\bar{n}^s\ln\bar{n}^s - (1/\eta)(m-\eta\bar{n}^s)\ln(m-\eta\bar{n}^s) + \text{const.} \quad (36)$$

Equating the derivative of (36) with (31) yields

$$\bar{n}^s = m/(e^{(\epsilon-\mu)/T}+\eta). \quad (37)$$

Introducing (37) into (36) results in

$$\mathscr{S}(\bar{n}^s) = (\mathscr{E}-\mu\bar{n}^s)/T + (m/\eta)\ln(1+\eta e^{-(\epsilon-\mu)/T}), \quad (38)$$

where we have set the constant in expression (36) equal to $(m/\eta)\ln m$. Furthermore, from thermodynamics, we have $T\mathscr{S}-\mathscr{E}+\mu\bar{n}^s = PV$ so that the last term in (38) is the product of the logarithm of grand partition function, Ω_η and the absolute temperature,

$$PV = T\ln\Omega_\eta = T(m/\eta)\ln(1+\eta e^{-(\epsilon-\mu)/T}). \quad (39)$$

For $\eta=1$ and $\eta=-1$, Ω_η is the grand partition function for FD and BE statistics, respectively. Expression (33) can now be written as

$$m\ln p = -PV/T. \quad (40)$$

To derive an expression for the stochastic entropy (34), we take the logarithm of (16) written as

$$m-\eta n = (n+1)\frac{\vartheta(n+1;m,d)}{\vartheta(n;m,d)}.$$

Then approximating the finite difference by the differential

$$\frac{d\ln\vartheta}{dn} = \ln\left(\frac{m-\eta n}{n}\right) \quad (41)$$

for $n \gg 1$, we get

$$\ln\vartheta(n;m,d) = -n\ln n - (1/\eta)(m-\eta n)\ln(m-\eta n) + \text{const} = \Sigma(n), \quad (42)$$

provided $\eta\neq 0$, upon integration. For $n=\bar{n}^s$, the stochastic entropy (42) coincides with the entropy (36). But since the recursion relation (16) is valid for $\eta=\pm 1$, in addition to $\eta=0$, the stochastic entropy will coincide with the entropy at $n=\bar{n}^s$ for the negative binomial and binomial distributions, corresponding to BE and FD statistics. Only for these distributions, and the limiting Poisson distribution, will the stationary distribution $f^s(n)$ be a maximum at $n=\bar{n}^s$. In other words, these distributions are the laws of error leading to the average value as the most probable value. Let us now consider the limiting form in greater detail.

The classical limit is achieved by examining the entropy expression (36), or equivalently,

$$\mathscr{S}(\bar{n}^s) = (m/\eta)\ln(1+q\eta) - [mq/(1+q\eta)]\ln q \quad (43)$$

in the limit as $q\to 0$ and $m\to\infty$ such that $qm\to\bar{n}^s$. The entropy expression then reduces to

$$\mathscr{S}(\bar{n}^s)\to\bar{n}^s - \bar{n}^s\ln(\bar{n}^s/m), \quad (44)$$

which is seen to be independent of the parameter η. In the same limit, the stochastic entropy (42) tends to

$$\Sigma(n)\to n - n\ln(n/m). \quad (45)$$

Comparing expressions (44) and (45), in conjunction with the form of the probability distribution (30), shows that the average value $\bar{n}^s$ coincides with the most probable value of n. The stochastic entropy (45) contains the logarithm of the term m^n which it would correspond to in classical statistics as the number of ways one can distribute n particles over m cells. The term $n\ln n - n \approx \ln n!$, according to Stirling's approximation for sufficiently large n, takes into account the fact that the number of arrangements has been over counted exactly $n!$ times.

ACKNOWLEDGMENT

This work was supported in part by contributions from the Italian Ministry of Public Education (M.P.I.) and the Italian National Research Council (C.N.R.).

[1]G. Gentile, Nuovo Cimento **19**, 106, 493 (1940).
[2]A. Sommerfield, Ber. Dtsch. Chem. Ges. **75**, 1988 (1942).
[3]H. Müller, Ann. Phys. (Leipzig) 7, 420 (1950).
[4]H. Wergeland, Kgl. Norske Vid. Selsk. Forh. **17**, 51 (1944).
[5]G. Schubert, Z. Naturforsch. **1**, 113 (1946).
[6]D. ter Haar, Physica **43**, 199 (1952).
[7]A. M. Guénault and D. K. C. MacDonald, Mol. Phys. **5**, 525 (1962).
[8]P. T. Landsberg, Mol. Phys. **6**, 341 (1963).
[9]See, for example, R. von Mises, *Mathematical Theory of Probability and Statistics* (Academic, New York, 1964), p. 202.
[10]B. H. Lavenda, Int. J. Theor. Phys. **27**, 1371 (1988).
[11]W. Feller, *An Introduction to Probability Theory and its Applications* (Wiley, New York, (1950), Vol. I, pp. 284 and 285.
[12]W. Feller, *An Introduction to Probability Theory and its Applications* (Wiley, New York, 1971), p. 27.
[13]A. Einstein, Phys. Z. **18**, 121 (1917), reprinted in *Sources in Quantum Mechanics*, edited by B. van der Waerden (North-Holland, Amsterdam, 1967), pp. 63–77; B. H. Lavenda, "On Einstein's quantum theory of radiation," submitted to Int. J. Theor. Phys.

A dynamical formalism of singular Lagrangian system with higher derivatives

Yoshihiko Saito and Reiji Sugano
Department of Physics, Osaka City University, Osaka 558, Japan
Tadayuki Ohta
Department of Physics, Miyagi University of Education, Sendai 980, Japan
Toshiei Kimura
Research Institute of Theoretical Physics, Hiroshima University, Takehara, Hiroshima 725, Japan

(Received 11 March 1988; accepted for publication 11 January 1989)

The singular Lagrangian system with higher derivatives is analyzed with the aid of the Ostrogradski transformation and the Dirac formalism. The formulation of canonical theory is developed so that the equivalence between the Lagrange formalism and the Hamilton one is maintained. As a practical example, the acceleration-dependent potentials appearing in the Lagrangian of two-point particles interacting gravitationally are dealt with and the equivalence between the two Hamiltonians that follow from the two Lagrangians which are related by the coordinate transformations is shown. It is also shown, when the constraints are all first class, that a consistent generator of gauge transformation is constructed. Typical examples are given.

I. INTRODUCTION

Dynamical systems described in terms of higher derivative variables have been investigated for a long time in connection with nonlocal field theory[1] and relativistic dynamics of particles moving in a field, and so on.[2] Recently, dynamical models of a Lagrangian with the higher derivatives have been acquiring importances motivated by gravity theory with a Lagrangian containing the quadratic term in the curvature tensor[3] and supersymmetry theory in higher dimensions, in spite of the possibility of unitarity violation associated with ghost states.

The canonical formalism for the system with higher derivatives was presented by Ostrogradski.[4] Analysis of a singular Lagrangian with higher derivatives involves more than that of an ordinary singular Lagrangian, that is, it is not a trivial extension of the ordinary theory.

In this paper, we shall develop the canonical formalism of the singular Lagrangian with the aid of the Ostrogradski transformation and the Dirac formalism[5] for singular Lagrangians. Our considerations are mainly concerned with the relation between constraints and gauge transformations of the system and the equivalence of the Lagrange formalism to the Hamilton one. In Sec. II it is shown that all secondary constraints derived from the conditions of primary constraints, being stationary, are classified into two groups; one group being contained in equations defining momenta in the Ostrogradski transformations and other group being on the outside of them. The latter constraints correspond to the constraints in the ordinary singular system. If the order of the highest derivative in L is reduced by adding a total time derivative term, the number of constraints are also reduced. The relation between the number of the constraints and the order of the highest derivative of the additive term is given. The addition of the total derivative term to the Lagrangian is confirmed to have no physical effect, even in the theory with higher derivatives.

In the case of the particle systems interacting with the gravitational field, acceleration terms appear, in general, in the interaction Lagrangian when calculated by means of the perturbation with respect to $1/c$, c being the velocity of light. If we eliminate the acceleration $\boldsymbol{\alpha}$ in the Lagrangian by the use of equations of motion of the lower-order Langrangian, correct equations of motion cannot be obtained.[6] In Sec. III, we treat gravitational interaction. In the post Newtonian approximation ($1/c^2$ order), there occurs a gauge-dependent interaction potential of the form[7]

$$-y\frac{G}{2c^2}\sum_a\sum_{b\neq a}m_am_b\mathbf{n}_{ab}\cdot\left(\boldsymbol{\alpha}_a+G\sum_{c\neq a}\frac{m_c}{r_{ac}^2}\mathbf{n}_{ac}\right),$$

where y is a gauge parameter and G is Newton's gravitational constant (detailed definition of notations are given in Sec. III). The acceleration terms in this potential can be dropped by adding a total time derivative term,

$$\frac{d}{dt}\left\{y\frac{G}{2c^2}\sum_a\sum_{b\neq a}m_am_b(\mathbf{n}_{ab}\cdot\mathbf{v}_a)\right\}.$$

However, even when $y=0$, a post-post-Newtonian interaction potential contains acceleration terms[8]

$$\frac{G}{4c^4}\sum_a\sum_{c\neq a}m_am_c\left(\boldsymbol{\alpha}_a+G\sum_{b\neq a}\frac{m_b\mathbf{n}_{ab}}{r_{ab}^2}\right)\cdot\mathbf{n}_{ac}\left\{\mathbf{v}_c^2-G\sum_{d\neq c}\frac{m_d}{r_{cd}}\right\},$$

which occurs in a physically acceptable coordinate system.[7,9] This potential cannot be transformed to an acceleration-independent one by adding total time derivative terms. So we shall investigate the problem of the acceleration-dependent gravitational potential from the standpoint of Hamiltonian dynamics based on the canonical formalism. We show that the canonical equations of motion in terms of Dirac brackets coincides with the Euler–Lagrange equations

 0022-2488/89/051122-11$02.50

derived from the Lagrangian with the acceleration terms. A relation between Lagrangian systems with acceleration terms and without them is obtained. Two Hamiltonians derived from the Lagrangians with and without acceleration terms are related to each other by a canonical transformation and they are physically equivalent.

On the other hand, as shown by Schäfer,[10] the elimination of the acceleration in the Lagrangian by the use of the equations of motion is permitted in the gravitational case, because it is equivalent to a coordinate transformation. Then the Dirac bracket formalism allows one to exhibit the equivalence between the two Hamiltonians that follow from the two Lagrangians that are related by the coordinate transformation.

In Sec. IV, for a gauge invariant system having only first-class constraints, we shall give the procedure to construct the generator of a gauge transformation. Though the procedure is almost the same as that in the ordinary case,[11] we must show the mutual consistency of the transformations among variables with different orders of derivatives. The consistency condition will be examined. In Sec. V, typical examples will be presented for illustration of Sec. IV.

II. CANONICAL FORMALISM FOR SINGULAR LAGRANGIAN WITH HIGHER DERIVATIVES

We consider a system with n degrees of freedom. Let us assume a Lagrangian of the system with higher derivatives to be given by

$$L\left(q^i,\dot q^i,\ldots,\overset{(N)}{q}{}^i\right), \tag{2.1}$$

where $i = 1 \sim n$ and

$$\overset{(s)}{q}{}^i \equiv \frac{d^s}{dt^s} q^i \quad (s = 1 \sim N) .$$

We suppose the extended Hessian matrix

$$A_{ij} \equiv \frac{\partial^2 L}{\partial \overset{(N)}{q}{}^i \, \partial \overset{(N)}{q}{}^j}, \tag{2.2}$$

to be singular and its rank to be $n - r$.

Now, we apply the Ostrogradski transformation and the Dirac formalism[5] for constrained systems to this singular L. Then, it is shown that the Lagrangian formalism is transformed to the Hamiltonian one by using a total Hamiltonian given by Dirac.

If the order of the highest derivatives in L is changed by adding a total time derivative term, the number of constraints is also changed. It, therefore, will be worthwhile to reveal the relation of the number of constraints to the order of derivative of the total time derivative term.

The Ostrogradski transformation introduces canonical momenta $p_i^{(s)}$ conjugate to $q^i_{(s)} \equiv \overset{(s)}{q}{}^i$

$$p_i^{(N-1)} \equiv \frac{\partial L}{\partial \overset{(N)}{q}{}^i}, \tag{2.3}$$

$$p_i^{(s-1)} \equiv \frac{\partial L}{\partial \overset{(s)}{q}{}^i} - \dot p_i^{(s)} \quad (s = 1 \sim N-1) , \tag{2.4}$$

and gives the transformation from the "velocity" phase space (VPS) to the phase space (PS);

$$\left(q^i,\dot q^i,\ddot q^i,\ldots,\overset{(2N-1)}{q}{}^i\right)$$
$$\to (q^i_{(0)},q^i_{(1)},q^i_{(2)},\ldots,q^i_{(N-1)};p_i^{(0)},p_i^{(1)},p_i^{(2)},\ldots,p_i^{(N-1)}) ,$$

where $q^i_{(0)} \equiv q^i$, $q^i_{(s)} \equiv \overset{(s)}{q}{}^i$, $(s = 1 \sim N)$. The Hamiltonian is defined by

$$H \equiv \sum_{s=0}^{N-1} p_i^{(s)} q^i_{(s+1)} - L\left(q,\ldots,\overset{(N)}{q}\right). \tag{2.5}$$

The summation convention is assumed for dummy upper and lower indices. Since A_{ij} is singular, (2.3) is not solvable for all $\overset{(N)}{q}{}^i$. But, we have r constraints

$$\phi^0_\alpha (q_{(0)},q_{(1)},\ldots,q_{(N-1)},p^{(N-1)}) = 0 \quad (\alpha = 1 \sim r) , \tag{2.6}$$

which are derived from (2.3) and called the primary constraints. Using these constraints we obtain a canonical Hamiltonian H_0 and a total Hamiltonian

$$H_T = H_0 + v^\alpha \phi^0_\alpha , \tag{2.7}$$

where v^α are undetermined multipliers. It is noticed that the primary constraints $\phi^0_\alpha \simeq 0$ depend only on $p_i^{(N-1)}$, not on $p_i^{(s)} (s < N-1)$, as shown later.

With the aid of H_T, we can formulate in PS the Hamiltonian formalism equivalent to the Lagrangian formalism. In order to show it, let us find the explicit form of H_0. Since the rank of matrix A_{ij} is $n - r$, without loss of generality, we can assume

$$p_a^{(N-1)} = \frac{\partial L}{\partial \overset{(N)}{q}{}^a} \quad (a = 1 \sim n - r) , \tag{2.8}$$

in (2.3), to be solvable with respect to $\overset{(N)}{q}{}^a = \dot q^a_{(N-1)}$ and obtain

$$\dot q^a_{(N-1)} = \overset{(N)}{q}{}^a = f^a(q^i_{(0)},q^i_{(1)},\ldots,q^i_{(N-1)},p_b^{(N-1)},\dot q^\alpha_{(N-1)}) , \tag{2.9}$$

where $b = 1 \sim n - r$ and $\alpha = n - r + 1,\ldots,n$. The substitution of (2.9) into (2.5) yields

$$H = p_a^{(N-1)} f^a + p_\alpha^{(N-1)} \dot q^\alpha_{(N-1)} + \sum_{s=0}^{N-2} p_i^{(s)} q^i_{(s+1)} - L\,\Big|_{\dot q^a_{(N-1)} = f^a} . \tag{2.10}$$

This H is linearly dependent on $\dot q^\alpha_{(N-1)}$, since

$$\frac{\partial H}{\partial \dot q^\alpha_{(N-1)}} = p_\alpha^{(N-1)} - \left(\frac{\partial L}{\partial \dot q^\alpha_{(N-1)}}\right)\Bigg|_{\dot q^a_{(N-1)} = f^a}$$

is independent of $\dot q^\alpha_{(N-1)}$ (see Appendix). Then H can be expressed as

$$H = H_0(q^i_{(0)},\dots,q^i_{(N-1)},p^{(0)}_i,\dots,p^{(N-2)}_i,p^{(N-1)}_a)$$
$$+ \dot{q}^\alpha_{(N-1)}\phi^0_\alpha(q_{(0)},\dots,q_{(N-1)},p^{(N-1)}) , \qquad (2.11)$$

with

$$H_0 = H|_{\dot{q}^\alpha_{(N-1)}=0} , \qquad (2.12)$$

$$\phi^0_\alpha \equiv p^{(N-1)}_\alpha - \left.\frac{\partial L}{\partial \dot{q}^\alpha_{(N-1)}}\right|_{\dot{q}^a_{(N-1)}=f^a} . \qquad (2.13)$$

As proved in the Appendix, $(\partial L/\partial \dot{q}^\alpha_{(N-1)})|_{\dot{q}^a_{(N-1)}=f_a}$ and also ϕ^0_α are independent of $\dot{q}^\alpha_{(N-1)}$. The constraints

$$\phi^0_\alpha = 0 \qquad (2.14)$$

are the primary constraints (2.6). Thus H can be expressed in terms of variables $q^i_{(0)},\dots,q^i_{(N-1)},p^{(0)}_i,\dots,p^{(N-1)}_i$, if the primary constraints are used. The total Hamiltonian H_T is given by

$$H_T = H|_{\dot{q}^\alpha_{(N-1)}=v^\alpha} . \qquad (2.15)$$

By resorting to the Dirac formalism[5] on the constraints, the Hamilton equations of motion are represented as

$$\dot{q}^i_{(s)} = \frac{\partial H_T}{\partial p^{(s)}_i} \simeq q^i_{(s+1)} \quad (s = 0\sim N-2) , \qquad (2.16a)$$

$$\dot{q}^a_{(N-1)} = \frac{\partial H_T}{\partial p^{(N-1)}_a} \simeq f^a(q_{(0)},\dots,q_{(N-1)},p^{(N-1)}_b,v) , \qquad (2.16b)$$

$$\dot{q}^\alpha_{(N-1)} = \frac{\partial H_T}{\partial p^{(N-1)}_\alpha} \simeq v^\alpha , \qquad (2.16c)$$

$$\dot{p}^{(0)}_i = -\frac{\partial H_T}{\partial q^i_{(0)}} \simeq \frac{\partial L}{\partial q^i_{(0)}} , \qquad (2.16d)$$

$$\dot{p}^{(s)}_i = -\frac{\partial H_T}{\partial q^i_{(s)}} \simeq -p^{(s-1)}_i + \frac{\partial L}{\partial q^i_{(s)}}$$
$$(s = 1\sim N-1) , \qquad (2.16e)$$

where the weak equality $\simeq$ means to impose $\phi^0_\alpha = 0$. On the right-hand sides of these equations and in the following equations up to Eq. (2.22), the derivative of L with respect to $q^i_{(s)}$ means

$$\frac{\partial L}{\partial q^i_{(s)}} \equiv \left.\left\{\frac{\partial}{\partial q^i_{(s)}} L(q_{(0)},\dots,q_{(N-1)},q_{(N)})\right\}\right|_{\dot{q}^a_{(N-1)}=f^a}$$
$$(s = 0\sim N) .$$

In these equations, only (2.16d) is a dynamical equation, but all other equations together with $\phi^0_\alpha = 0$ are equivalent to the Ostrogradski transformation (2.4). Equation (2.16c) indicates for H_T to be equivalent to H of (2.10) or (2.11).

In the Ostrogradski transformation, if some equations of (2.4) are independent of $\overset{(N+k)}{q}$ $(k\geqslant 0)$, they are regarded as constraints. When the Ostrogradski transformation (2.4) gives such constraints other than the primary ones, all these constraints can be derived from the condition of the primary ones, $\phi^0_\alpha = 0$ being stationary, as shown below. In (2.4), the relation for $s = N-2$,

$$p^{(N-2)}_\alpha = \frac{\partial L}{\partial q^\alpha_{(N-1)}} - \frac{d}{dt}\frac{\partial L}{\partial \dot{q}^\alpha_{(N-1)}} ,$$

contains at least $\overset{(N+1)a}{q}$ and hence does not yield constraints, since $\det(A_{ab}) \neq 0$ by assumption. For a particular α, suppose

$$p^{(N-2)}_\alpha = \frac{\partial L}{\partial q^\alpha_{(N-1)}} - \frac{d}{dt}\frac{\partial L}{\partial \dot{q}^\alpha_{(N-1)}} \qquad (2.17)$$

to be the constraint. Then, from (2.10), (2.13), (2.16a)–(2.16e) it follows that the condition of ϕ^0_α being stationary turns out to be

$$\{\phi^0_\alpha,H\} \simeq -p^{(N-2)}_\alpha + \frac{\partial L}{\partial q^\alpha_{(N-1)}} - \frac{d}{dt}\frac{\partial L}{\partial \dot{q}^\alpha_{(N-1)}} = 0 . \qquad (2.18)$$

The Poission bracket $\{\ ,\ \}$ is defined by

$$\{f,g\} = \sum_{s=0}^{N-1}\left(\frac{\partial f}{\partial q^i_{(s)}}\frac{\partial g}{\partial p^{(s)}_i} - \frac{\partial f}{\partial p^{(s)}_i}\frac{\partial g}{\partial q^i_{(s)}}\right),$$

and the fundamental Poisson brackets are

$$\{q^i_{(s)},p^{(s')}_j\} = \delta^i_j\delta^{s'}_s, \quad \{q^i_{(s)},q^j_{(s')}\} = \{p^{(s)}_i,p^{(s')}_j\} = 0$$
$$(i,j = 1\sim n, s,s' = 0\sim N-1) .$$

On the other hand, we have

$$\{\phi^0_\alpha,H\} \simeq \{\phi^0_\alpha,H_0\} + \dot{q}^\beta_{(N-1)}\{\phi^0_\alpha,\phi^0_\beta\} . \qquad (2.19)$$

Since ϕ^0_α and H_0 are independent of $\dot{q}^\alpha_{(N-1)}$ and (2.17) is also assumed to be so, we obtain from (2.17), (2.18), (2.19), and the form of ϕ^0_α of (2.13)

$$\{\phi^0_\alpha,\phi^0_\beta\} = 0 . \qquad (2.20)$$

Hence, the condition of $\phi^0_\alpha \simeq 0$ being stationary,

$$\phi^1_\alpha \equiv \{\phi^0_\alpha,H_0\} \simeq 0 \qquad (2.21)$$

is equivalent to (2.17) which is a "secondary" constraint. Furthermore, if

$$p^{(N-3)}_\alpha = \frac{\partial L}{\partial q^\alpha_{(N-2)}} - \dot{p}^{(N-2)}_\alpha \qquad (2.22)$$

does not contain any $\overset{(N+k)}{q}$ $(k\geqslant 0)$, we can prove that (2.22) also is equivalent to

$$\phi^2_\alpha \equiv \{\phi^1_\alpha,H_0\} \simeq 0 , \quad \text{with } \{\phi^0_\alpha,\phi^1_\beta\} \simeq 0 . \qquad (2.23)$$

It should be noticed that if (2.17) is not a constraint, (2.22) necessarily contains $\overset{(N+k)}{q}$ $(k>0)$ and then is not the constraint. Thus it has been seen that all constraints contained in the Ostrogradski transformations are successively derived from the conditions of the primary constraints being stationary. If the relations for $s\geqslant M$ in (2.4) are constraints $\phi^{N-s}_\alpha = 0$ but the ones for $s<M$ are not, there exists ϕ^0_β such as $\{\phi^0_\beta,\phi^{N-M}_\alpha\} \neq 0$. In this case the condition of ϕ^{N-M}_α being stationary no longer yields new constraints, as is well known in the Dirac algorithm using H_T, and all the constraints ϕ^s_α following ϕ^0_α are exhausted with the constraints contained in the Ostrogradski transformations.

From the Hamilton equations (2.16) and (2.13) we can reproduce the Euler-Lagrange equations derived from L of (2.1)

$$\sum_{s=0}^{N}(-1)^s\frac{d^s}{dt^s}\left(\frac{\partial L}{\partial \overset{(s)}{q}{}^{i}}\right)=0 . \tag{2.24}$$

Thus it has been verified that even for the singular Lagrangian with higher derivatives, the Hamiltonian formalism based on the Ostrogradski transformation is equivalent to the Lagrangian one.

Next we investigate the relation between the number of the constraints and the number of the canonical variables for a singular Lagrangian obtained by adding a total time derivative term dF/dt. By adding $\dot F$ to L, the forms of $p_i^{(s)}$ and H change. The Euler–Lagrange equations (2.24), however, are not affected by it, provided that the boundary conditions $\delta \overset{(s)}{q}{}^{i}(t_1)=\delta \overset{(s)}{q}{}^{i}(t_2)=0$ for all s are imposed in the variation principle. Further, it is shown that the addition of $\dot F$ gives no effect to H, in spite of the change of its functional form. The changes of momenta $\Delta p_i^{(s)}$ due to the $\dot F$ term are given by

$$\Delta p_i^{(s)}=\frac{\partial F}{\partial \overset{(s)}{q}{}^{i}},$$

and the change of Hamiltonian ΔH vanishes;

$$\Delta H=\sum_s \Delta p_i^{(s)}\overset{(s+1)}{q}{}^{i}-\dot F=0 .$$

So, we may conclude that the addition of $\dot F$ brings no physical effect.

Now let the Lagrangian be

$$L=L_0\Big(q,\dot q,\ldots,\overset{(M)}{q}\Big)+\frac{d}{dt}F\Big(q,\dot q,\ldots,\overset{(N-1)}{q}\Big)\quad (M\leqslant N-1) . \tag{2.25}$$

Here we assume, for simplicity, that the Hessian matrix for L_0,

$$A^0_{ij}\equiv\frac{\partial^2 L_0}{\partial \overset{(M)}{q}{}^{i}\,\partial \overset{(M)}{q}{}^{j}}, \tag{2.26}$$

is regular. By applying (2.3) and (2.4) to this L, the momenta conjugate to $q^i_{(s)}$ turn out to be

$$p_i^{(s)}=\frac{\partial F}{\partial \overset{(s)}{q}{}^{i}}\quad (M\leqslant s\leqslant N-1) , \tag{2.27a}$$

$$p_i^{(M-r)}=\frac{\partial F}{\partial \overset{(M-r)}{q}{}^{i}}+\sum_{a=1}^{r}(-1)^{a-1}\frac{d^{a-1}}{dt^{a-1}}\frac{\partial L_0}{\partial \overset{(M-r+a)}{q}{}^{i}}$$

$$(1\leqslant r\leqslant M) . \tag{2.27b}$$

Depending on $N>2M$ or $N\leqslant 2M$, there are two cases concerning features of the constraints.

(i) The case of $N\leqslant 2M$. The relations (2.27) give the $2n(N-M)$ constraints

$$\phi_i^u\equiv p_i^{(N-u-1)}-\frac{\partial F}{\partial q^i_{(N-u-1)}}-\sum_{a=1}^{M-N+u+1}(-1)^{a-1}\frac{d^{a-1}}{dt^{a-1}}\left(\frac{\partial L_0}{\partial \overset{(N-u+a-1)}{q}{}^{i}}\right)$$

$$=0\quad (0\leqslant u\leqslant 2N-2M-1) . \tag{2.28}$$

As mentioned before, these constraints $\phi_i^u(u\geqslant 1)$ are also derived from the stationarity conditions of the primary ones ϕ_i^0.

In general, the constraints ϕ's are classified into two classes. A ϕ_a is defined to be first class if $\{\phi_a,\phi_b\}\equiv 0 \bmod(\phi_c)$ for all ϕ's and $\{\phi_a,H\}\equiv 0 \bmod(\phi_b)$, and the other is second class. In order to treat the second-class constraints $\theta_s=0$ consistently, we should use the following Dirac bracket in place of the Poisson bracket:

$$\{f,g\}^*=\{f,g\}-\{f,\theta_s\}(C^{-1})^{st}\{\theta_t,g\} ,$$

where $(C^{-1})^{st}$ satisfy the relationship $\{\theta_s,\theta_t\}(C^{-1})^{tu}=\delta_s^u$. If we work with Dirac bracket, we can put $\theta_s=0$ strongly, since $\{\theta_s,f\}^*=0$ for any dynamical variable f.

In the present case, all ϕ_i^u of (2.28) are second class. Indeed, owing to (2.26), we find

$$\det\{\phi^u,\phi^v\}=\det\left|\begin{array}{cccc:cccc} & & & & A^0 & & & * \\ & 0 & & & & A^0 & & \\ & & & & & & \ddots & \\ & & & & & 0 & & A^0 \\ \hdashline -A^0 & & & & & & & \\ & -A^0 & & 0 & & & & \\ & & \ddots & & & * & & \\ * & & & -A^0 & & & & \end{array}\right|$$

$$=(-1)^{N-M}(\det A^0)^{2(N-M)}\neq 0 . \tag{2.29}$$

Hence, due to the $2n(N-M)$ second-class constraints, the number of independent variables reduces to $2nM$, which is equal to the number of the canonical variables of L_0.

A simple example of this case is

$$L=\sum_{s=0}^{N} g_{ij}^{s} q^{i} \overset{(2s)_j}{q} = \sum_{s=0}^{N} (-1)^{s} g_{ij}^{s} \overset{(s)_i}{q} \overset{(s)_j}{q} + \dot{F}, \qquad (2.30)$$

with

$$F=\sum_{s=1}^{N} g_{ij}^{s} \Big(q^{i} \overset{(2s-1)_j}{q} - \dot{q}^{i} \overset{(2s-2)_j}{q} + \cdots + (-1)^{s-1} \overset{(s-1)_i}{q} \overset{(s)_j}{q}\Big),$$

where g_{ij}^{s} are constants. There appear $2nN$ constraints and L is equivalent to

$$L_0 \equiv \sum_{s=0}^{N} (-1)^{s} g_{ij}^{s} \overset{(s)_i}{q} \overset{(s)_j}{q}. \qquad (2.31)$$

(ii) The case of $N>2M$. Though the constraints are presented by (2.28), the series of the constraints are extended to $u=N-1$, due to $N>2M$. Among them, $\{\phi^0,\phi^1,\ldots,\phi^{N-2M-1}\}$ are first class and $\{\phi^{N-2M},\phi^{N-2M+1},\ldots,\phi^{N-1}\}$ are second class. There exist $n(N-2M)$ trivial degrees of freedom that are associated with the $n(N-2M)$ first-class constraints. Combining the $2nM$ second-class constraints with them, we have $2nM$ independent physical variables in this system, which again coincide with the number of canonical variables of L_0.

In both cases, since the Hamilton equations of motion with the primary constraints reduce to the Euler–Lagrange equations (2.24), we obtain the Hamiltonian formalism equivalent to the Lagrangian one, by resorting to the Dirac bracket. Consequently, if we employ a Lagrangian whose order of time derivative is reduced by adding a total time derivative term to L, redundant constraints can be removed from the system.

If A_{ij}^{0} of (2.26) is singular, the stationarity conditions of ϕ^u yield more constraints that are out of the Ostrogradski transformation. So, a further investigation is needed for such a system.

III. EXAMPLES IN GRAVITATIONAL INTERACTIONS

In the approximation of second order of $1/c$, the Lagrangian of two-point particles interacting gravitationally is given by[7]

$$\widetilde{L}^{*}_{PN} = L_{PN} + y\Big[\frac{G}{4c^2}\sum_a \sum_{b\neq a} \frac{m_a m_b}{r_{ab}} \times\{(\mathbf{v}_a-\mathbf{v}_b)^2-(\mathbf{n}_{ab}\cdot\mathbf{v}_a-\mathbf{v}_b)^2\} - \frac{G^2}{2c^2}\sum_a\sum_{b\neq a}\sum_{c\neq a}\frac{m_a m_b m_c}{r_{ab}^2}(\mathbf{n}_{ab}\cdot\mathbf{n}_{ac})\Big], \qquad (3.1)$$

with

$$L_{PN} = \frac{1}{2}\sum_a m_a \mathbf{v}_a^2 + \frac{1}{8c^2}\sum_a m_a(\mathbf{v}_a^2)^2 + \frac{G}{2}\sum_a\sum_{b\neq a}\frac{m_a m_b}{r_{ab}} + \frac{G}{4c^2}\sum_a\sum_{b\neq a}\frac{m_a m_b}{r_{ab}} \times\{6\mathbf{v}_a^2 - 7(\mathbf{v}_a\cdot\mathbf{v}_b) - (\mathbf{n}_{ab}\cdot\mathbf{v}_a)(\mathbf{n}_{ab}\cdot\mathbf{v}_b)\} - \frac{G^2}{2c^2}\sum_a\sum_{b\neq a}\sum_{c\neq a}\frac{m_a m_b m_c}{r_{ab} r_{ac}}, \qquad (3.2)$$

where m_a and $\mathbf{v}_a$ are the rest mass and the velocity of ath particle ($a=1,2$), respectively, $r_{ab}=|\mathbf{z}_a-\mathbf{z}_b|$ ($\mathbf{z}_a$ being the coordinate of ath particle), $\mathbf{n}_{ab}=(\mathbf{z}_a-\mathbf{z}_b)/r_{ab}$, and y is a gauge parameter.

In order to illustrate the essence of our procedure, we use here a system of center of inertia for simplicity's sake. The $\widetilde{L}^{*}_{PN}$ is rewritten as

$$\widetilde{L}^{*}_{PN} = \frac{1}{2}\mu\dot{\mathbf{r}}^2 + \frac{1}{8}\frac{\mu(M-3\mu)}{Mc^2}(\dot{\mathbf{r}}^2)^2 + \frac{G\mu M}{r} - \frac{G^2\mu M^2}{2c^2r^2} + \frac{G\mu}{2c^2r}\Big\{(3M+\mu)(\dot{\mathbf{r}})^2+\mu\frac{(\mathbf{r}\cdot\dot{\mathbf{r}})^2}{r^2}\Big\} + \frac{y}{2}\frac{G\mu M}{c^2r}\Big\{\dot{\mathbf{r}}^2-\frac{(\mathbf{r}\cdot\dot{\mathbf{r}})^2}{r^2}-\frac{G}{r}M\Big\}, \qquad (3.3)$$

where $M=m_1+m_2$, $\mu=m_1m_2/(m_1+m_2)$, $\mathbf{r}=\mathbf{z}_1-\mathbf{z}_2$, and $r=(\mathbf{r}^2)^{1/2}$.

If we add a total time derivative term

$$L_{\text{add}} = -\frac{y}{2c^2}G\mu M\frac{d^2}{dt^2}r \qquad (3.4)$$

to (3.3), we have the acceleration-dependent Lagrangian

$$L^{*}_{PN} = \frac{1}{2}\mu\dot{\mathbf{r}}^2 + \frac{1}{8}\frac{\mu(M-3\mu)}{c^2M}(\dot{\mathbf{r}}^2)^2 + \frac{G\mu M}{r} - \frac{G^2\mu M^2}{2c^2r^2} + \frac{G\mu}{2c^2r}\Big\{(3M+\mu)\dot{\mathbf{r}}^2+\mu\frac{(\mathbf{r}\cdot\dot{\mathbf{r}})^2}{r^2}\Big\} - \frac{y}{2}\frac{G\mu M}{c^2r}\mathbf{r}\cdot\Big(\ddot{\mathbf{r}}+\frac{GM}{r^3}\mathbf{r}\Big). \qquad (3.5)$$

The y-dependent term in (3.5) would reduce to the order of $1/c^4$, if we use the equation of motion that is derived from L_{PN}. However, it is evident that one cannot adopt the equations of motion in the Lagrangian. We shall show that the Lagrangian L^{*}_{PN} is physically equivalent to the L_{PN} irrespective of the value of y.

The canonical momenta conjugate to $\mathbf{r}_{(1)}\equiv\dot{\mathbf{r}}$ and $\mathbf{r}$ are, respectively ($\mathbf{r}_{(0)}=\mathbf{r}$, $\mathbf{p}^{(0)}=\mathbf{p}$),

$$\mathbf{p}^{(1)} = -y(G\mu M/2c^2r)\mathbf{r}, \qquad (3.6)$$

$$\mathbf{p} = \mu\dot{\mathbf{r}} + \frac{\mu(M-3\mu)}{2c^2M}(\dot{\mathbf{r}}^2)\dot{\mathbf{r}} + \frac{G\mu}{c^2r}\Big\{(3M+\mu)\dot{\mathbf{r}}+\frac{\mu}{r^2}(\mathbf{r}\cdot\dot{\mathbf{r}})\mathbf{r}\Big\} + y\frac{G\mu M}{2c^2r}\Big\{\dot{\mathbf{r}}-\frac{(\mathbf{r}\cdot\dot{\mathbf{r}})\mathbf{r}}{r^2}\Big\}. \qquad (3.7)$$

The Hamiltonian is

$$H^{*}_{PN} = (\mathbf{p}^{(1)}\cdot\ddot{\mathbf{r}}) + (\mathbf{p}\cdot\dot{\mathbf{r}}) - L^{*}_{PN}. \qquad (3.8)$$

When $\dot{\mathbf{r}}$ is replaced by $\mathbf{r}_{(1)}$ in (3.7), Eqs. (3.6) and (3.7) become the second-class constraints:

$$\boldsymbol{\phi}^0 \equiv \mathbf{p}^{(1)} + y(G\mu M/2c^2r)\mathbf{r} = 0, \qquad (3.9)$$

$$\phi^1 \equiv \mathbf{p} - \mu \mathbf{r}_{(1)} - \frac{\mu(M-3\mu)}{2c^2M}(\mathbf{r}_{(1)}^2)\mathbf{r}_{(1)} - \frac{G\mu}{c^2 r}\left\{(3M+\mu)\mathbf{r}_{(1)} + \frac{\mu}{r^2}(\mathbf{r}\cdot\mathbf{r}_{(1)})\mathbf{r}\right\} - y\frac{G\mu M}{2c^2 r}\left\{\mathbf{r}_{(1)} - \frac{(\mathbf{r}\cdot\mathbf{r}_{(1)})\mathbf{r}}{r^2}\right\} = 0 . \tag{3.10}$$

The Poisson brackets

$$\begin{aligned} &\{r^i,p_j\} = \delta^i_j, \qquad \{r^i,r^j\} = \{p_i,p_j\} = 0, \\ &\{r^i_{(1)},p^{(1)}_j\} = \delta^i_j, \quad \{r^i_{(1)},r^j_{(1)}\} = \{p^{(1)}_i,p^{(1)}_j\} = 0 , \end{aligned} \tag{3.11}$$

should be replaced with the Dirac bracket defined by, up to $1/c^2$,

$$\{f,g\}^* = \{f,g\} - \{f,\phi^0_i\}(C^{-1})^{ij}\{\phi^1_j,g\} + \{f,\phi^1_i\}(C^{-1})^{ij}\{\phi^0_j,g\} ,$$

in which $\{\phi^0_i,\phi^1_j\}(C^{-1})^{jk} = \delta^k_i$. Hence, we have (up to $1/c^2$)

$$\begin{aligned} &\{r^i,p_j\}^* = \delta^i_j - y(GM/2c^2r)(\delta^i_j - r^ir^j/r^2) , \\ &\{r^i,r^j\}^* = \{p_i,p_j\}^* = 0 . \end{aligned} \tag{3.12}$$

If we introduce $\underline{\mathbf{r}}$ and $\underline{\mathbf{p}}$ through

$$\underline{\mathbf{r}} = \mathbf{r} + y(GM/2c^2)(\mathbf{r}/r), \quad \underline{\mathbf{p}} = \mathbf{p} , \tag{3.13}$$

they satisfy the ordinary canonical form

$$\{\underline{r}^i,\underline{p}_j\}^* = \delta^i_j, \quad \{\underline{r}^i,\underline{r}^j\}^* = \{\underline{p}_i,\underline{p}_j\}^* = 0 . \tag{3.14}$$

Since we work with the Dirac brackets, the constraints $\phi^0 = 0$ and $\phi^1 = 0$ hold strongly. The Hamiltonian is, up to the order of $1/c^2$, reduced to

$$H^*_{PN} = \frac{1}{2\mu}\underline{\mathbf{p}}^2 - \frac{M-3\mu}{8c^2\mu^3 M}(\underline{\mathbf{p}}^2)^2 - \frac{G\mu M}{\underline{r}} - \frac{G}{2c^2\mu \underline{r}}\left\{(3M+\mu)\underline{\mathbf{p}}^2 + \frac{\mu(\underline{\mathbf{r}}\cdot\underline{\mathbf{p}})^2}{\underline{r}^2}\right\} + \frac{G^2\mu M^2}{2c^2\underline{r}^2} + O\left(\frac{1}{c^4}\right), \tag{3.15}$$

which has the same form as the one obtained from the Lagrangian (3.5) with $y = 0$.

As mentioned in Sec. II, the total time derivative term does not affect the physical effect. We shall investigate, in the Hamilton formalism, the relation between the two systems described by the Lagrangian L^*_{PN} (3.5) and the $\tilde{L}^*_{PN}$ (3.3), respectively [cf. case (i) in Sec. II].

The Lagrangian $\tilde{L}^*_{PN}$ does not contain acceleration terms. The canonical momenta and the Hamiltonian are given by

$$\mathbf{p} = \mu\dot{\mathbf{r}} + \frac{\mu(M-3\mu)}{2c^2M}(\dot{\mathbf{r}}^2)\dot{\mathbf{r}} + \frac{G\mu}{c^2r}\left\{(3M+\mu)\dot{\mathbf{r}} + \frac{\mu(\mathbf{r}\cdot\dot{\mathbf{r}})\mathbf{r}}{r^2}\right\} + y\frac{G\mu M}{c^2r}\left\{\dot{\mathbf{r}} - \frac{(\mathbf{r}\cdot\dot{\mathbf{r}})\mathbf{r}}{r^2}\right\}, \tag{3.16}$$

$$\tilde{H}^*_{PN} = \frac{1}{2\mu}\mathbf{p}^2 - \frac{M-3\mu}{8c^2\mu^3M}(\mathbf{p}^2)^2 - \frac{G\mu M}{r} - \frac{G}{2c^2\mu r}\left\{(3M+\mu)\mathbf{p}^2 + \mu\frac{(\mathbf{r}\cdot\mathbf{p})^2}{r^2}\right\} + \frac{G^2\mu M^2}{2c^2r^2} - y\frac{GM}{2c^2\mu r}\left\{\mathbf{p}^2 - \frac{(\mathbf{r}\cdot\mathbf{p})^2}{r^2}\right\} + y\frac{G^2\mu M^2}{2c^2r^2} . \tag{3.17}$$

We apply a canonical transformation defined by

$$\underset{\sim}{q} = \frac{\partial}{\partial \underset{\sim}{p}}S(q,\underset{\sim}{p},t), \quad p = \frac{\partial}{\partial q}S(q,\underset{\sim}{p},t), \quad H = \underset{\sim}{H} - \frac{\partial}{\partial t}S(q,\underset{\sim}{p},t)$$

to (3.17), where $S(z,\underset{\sim}{p},t)$ is its generating function. If we take

$$S = (\mathbf{r}\cdot\underset{\sim}{\mathbf{p}}) + y(GM/2c^2r)(\mathbf{r}\cdot\underset{\sim}{\mathbf{p}}) , \tag{3.18}$$

we get

$$\underset{\sim}{\mathbf{r}} = \mathbf{r} + y\frac{GM}{2c^2r}\mathbf{r}, \quad \mathbf{p} = \underset{\sim}{\mathbf{p}} + y\frac{GM}{2c^2r}\left\{\underset{\sim}{\mathbf{p}} - \frac{(\mathbf{r}\cdot\underset{\sim}{\mathbf{p}})\mathbf{r}}{r^2}\right\}, \tag{3.19}$$

and

$$\underset{\sim}{\tilde{H}}{}^*_{PN} = \frac{1}{2\mu}\underset{\sim}{\mathbf{p}}^2 - \frac{M-3\mu}{8c^2\mu^3M}(\underset{\sim}{\mathbf{p}}^2)^2 - \frac{G\mu M}{\underset{\sim}{r}} - \frac{G}{2c^2\mu\underset{\sim}{r}}\left\{(3M+\mu)\underset{\sim}{\mathbf{p}}^2 + \mu\frac{(\underset{\sim}{\mathbf{r}}\cdot\underset{\sim}{\mathbf{p}})^2}{\underset{\sim}{r}^2}\right\} + \frac{G^2\mu M^2}{2c^2\underset{\sim}{r}^2} + O\left(\frac{1}{c^4}\right), \tag{3.20}$$

which has the same form as H^*_{PN} of (3.15).

Thus it has been shown that the Hamiltonian with $y \neq 0$ is obtained from the Hamiltonian with $y = 0$ by a canonical transformation and then both Hamiltonians are physically equivalent to each other up to order $1/c^2$. Furthermore, L_{add} given by (3.4) has no physical effect as mentioned in Sec. II. Consequently, L^*_{PN} given by (3.5) is physically equivalent with L_{PN} given by (3.2). It should be stressed that the Dirac bracket plays a vital role in proving the equivalence. There exist the relations $\underline{\mathbf{r}} = \underset{\sim}{\mathbf{r}}$, $\underline{\mathbf{p}} = \underset{\sim}{\mathbf{p}}$ that are seen from (3.6), (3.7), (3.13), (3.16), and (3.19) by expressing $\underline{\mathbf{p}}$ and $\underset{\sim}{\mathbf{p}}$ in terms of $\mathbf{r}$ and $\dot{\mathbf{r}}$. This guarantees that the Hamilton equations described by (3.15) and (3.20) give rise to the same Euler–Lagrange equations up to the order $1/c^2$.

Next, we consider the post-post-Newtonian approximated Lagrangian in a physically acceptable coordinate system [in which $(g_{\mu\nu} - \eta_{\mu\nu}) \underset{r\to+\infty}{\to} O(1/r)$ under approximation processes].[7,9] As mentioned in the Introduction, there exists an acceleration-dependent Lagrangian that cannot be transformed to an acceleration-independent one by adding total time derivative terms. In the case of the two-body system, the Lagrangian is given by[8]

$$L_{PPN} = L_0 + L_\alpha , \tag{3.21}$$

with

$$L_0 = \frac{1}{2}(m_1\mathbf{v}_1^2 + m_2\mathbf{v}_2^2) + \frac{1}{8c^2}\{m_1(\mathbf{v}_1^2)^2 + m_2(\mathbf{v}_2^2)^2\} + \frac{1}{16c^4}\{m_1(\mathbf{v}_1^2)^3 + m_2(\mathbf{v}_2^2)^3\} + \frac{Gm_1m_2}{r}$$
$$- \frac{G^2m_1m_2(m_1+m_2)}{2c^2r^2} + \frac{G^3m_1m_2}{4c^4r^3}(m_1^2 + m_2^2 + 5m_1m_2) + \frac{Gm_1m_2}{2c^2r}\{3(\mathbf{v}_1^2 + \mathbf{v}_2^2) - 7(\mathbf{v}_1\cdot\mathbf{v}_2) - (\mathbf{n}\cdot\mathbf{v}_1)(\mathbf{n}\cdot\mathbf{v}_2)\}$$
$$+ \frac{G^2m_1m_2^2}{8c^4r^2}\{18\mathbf{v}_1^2 + 13\mathbf{v}_2^2 - 34(\mathbf{v}_1\cdot\mathbf{v}_2) + 15(\mathbf{n}\cdot\mathbf{v}_2)^2\} + \frac{G^2m_1^2m_2}{8c^4r^2}\{13\mathbf{v}_1^2 + 18\mathbf{v}_2^2 - 34(\mathbf{v}_1\cdot\mathbf{v}_2) + 15(\mathbf{n}\cdot\mathbf{v}_1)^2\}$$
$$- \frac{G^2m_1m_2(m_1+m_2)}{4c^4r^2}\{\mathbf{v}_1^2 + \mathbf{v}_2^2 - 2(\mathbf{v}_1\cdot\mathbf{v}_2)\} + \frac{Gm_1m_2}{8c^4r}\{7[(\mathbf{v}_1^2)^2 + (\mathbf{v}_2^2)^2]$$
$$- 14(\mathbf{v}_1\cdot\mathbf{v}_2)(\mathbf{v}_1^2 + \mathbf{v}_2^2) - 2(\mathbf{n}\cdot\mathbf{v}_1)(\mathbf{n}\cdot\mathbf{v}_2)(\mathbf{v}_1^2 + \mathbf{v}_2^2) + 11\mathbf{v}_1^2\mathbf{v}_2^2 + 2(\mathbf{v}_1\cdot\mathbf{v}_2)^2 - 5[(\mathbf{v}_1^2(\mathbf{n}\cdot\mathbf{v}_2)^2 + \mathbf{v}_2^2(\mathbf{n}\cdot\mathbf{v}_1)^2]$$
$$+ 12(\mathbf{v}_1\cdot\mathbf{v}_2)(\mathbf{n}\cdot\mathbf{v}_1)(\mathbf{n}\cdot\mathbf{v}_2) + 3(\mathbf{n}\cdot\mathbf{v}_1)^2(\mathbf{n}\cdot\mathbf{v}_2)^2\}, \tag{3.22}$$
$$L_\alpha = \frac{G}{4c^4}\sum_{a=1}^{2}\mathbf{f}_a\cdot\left(m_a\boldsymbol{\alpha}_a + \epsilon_a G\frac{m_1m_2}{r^2}\mathbf{n}\right), \tag{3.23}$$

where

$$\mathbf{f}_a = \epsilon_a\mathbf{n}(G(m_1m_2/r) - m_a\mathbf{v}_{a'}^2) \tag{3.24}$$

in which $\mathbf{v}_a = \dot{\mathbf{z}}_a$, $\boldsymbol{\alpha}_a = \ddot{\mathbf{z}}_a$ $(a=1,2)$, $\mathbf{n} = \mathbf{n}_{12}$, $\epsilon_1 = -\epsilon_2 = 1$, and a' means "not a."

The canonical momenta conjugate to $\mathbf{z}_{(1)a} = \mathbf{v}_a$ and $\mathbf{z}_a$ are

$$\mathbf{p}_a^{(1)} = (G/4c^4)m_a\mathbf{f}_a, \tag{3.25a}$$

$$\mathbf{p}_a = \frac{\partial L_0}{\partial \mathbf{v}_a} + \frac{\partial L_\alpha}{\partial \mathbf{v}_a} - \frac{d}{dt}\frac{\partial L_\alpha}{\partial \boldsymbol{\alpha}_a}$$
$$= m_a\mathbf{z}_{(1)a} + (1/c^2)\mathbf{W}_a, \tag{3.25b}$$

with

$$\frac{1}{c^2}\mathbf{W}_a = \frac{\partial}{\partial \mathbf{v}_a}\left\{L_0 - \frac{1}{2}\sum_b m_b\mathbf{v}_b^2\right\} - \frac{Gm_1m_2}{4c^4}\left\{\epsilon_a\left[(\mathbf{v}_1 - \mathbf{v}_2)\cdot\frac{\partial}{\partial \mathbf{z}_a}\right]\frac{\partial}{\partial \mathbf{z}_a}(Gm_a\ln r - \mathbf{v}_{a'}^2 r) - 2Gm_a\frac{(\mathbf{n}\cdot\mathbf{v}_{a'})}{r^2}\mathbf{n}\right\}. \tag{3.25c}$$

In deriving (3.25b) with (3.25c), use has been made of the equation of motion. According to the procedure given in Sec. II, we have the following second-class constraints with 12 components,

$$\boldsymbol{\phi}_a^0 \equiv \mathbf{p}_a^{(1)} - (G/4c^4)m_a\mathbf{f}_a = 0, \tag{3.26a}$$
$$\boldsymbol{\phi}_a^1 \equiv \mathbf{p}_a - m_a\mathbf{z}_{(1)a} - (1/c^2)\mathbf{W}_a(z_b, z_{(1)b}) = 0. \tag{3.26b}$$

The i,j indices labeling the degrees of freedom in Sec. II are here written (ai), (bj), $a,b = 1,2$ (particle's number), $i,j = 1,2,3$ (spatial indices). To obtain the Dirac brackets, let us take the 12×12 matrix that is relevant to our post-post-Newtonian approximation

$$\begin{pmatrix}\{\phi_{ai}^0,\phi_{bj}^0\}, & \{\phi_{ai}^0,\phi_{bj}^1\}\\ \{\phi_{ai}^1,\phi_{bj}^0\}, & \{\phi_{ai}^1,\phi_{bj}^1\}\end{pmatrix} = \begin{pmatrix} -\frac{G}{4c^4}\left(m_a\frac{\partial f_{ai}}{\partial v_b^j} - m_b\frac{\partial f_{bj}}{\partial v_a^i}\right), & m_a\delta_{ab}\delta_{ij} + O(1/c^2)\\ -m_a\delta_{ab}\delta_{ij} + O(1/c^2), & O(1/c^2)\end{pmatrix} \tag{3.27}$$

whose inverse is

$$\{\phi,\phi\}^{-1} = \begin{pmatrix} O(1/c^2) & -\frac{1}{m_a}\delta_{ab}\delta_{ij} + O(1/c^2)\\ \frac{1}{m_a}\delta_{ab}\delta_{ij} + O(1/c^2) & -\frac{G}{4c^4}\left(\frac{1}{m_b}\frac{\partial f_{ai}}{\partial v_b^j} - \frac{1}{m_a}\frac{\partial f_{bj}}{\partial v_a^i}\right)\end{pmatrix}. \tag{3.28}$$

Using (3.28), we get

$$\{z_a^i,z_b^j\}^* = \sum_{c,d}\sum_{k,l}\{z_a^i,\phi_{ck}^1\}\frac{G}{4c^4}\left(\frac{1}{m_d}\frac{\partial f_{ck}}{\partial v_d^l} - \frac{1}{m_c}\frac{\partial f_{dl}}{\partial v_c^k}\right)\{\phi_{dl}^1,z_b^j\} + O\left(\frac{1}{c^6}\right)$$
$$= -\frac{G}{4c^4}\left(\frac{1}{m_b}\frac{\partial f_{ai}}{\partial v_b^j} - \frac{1}{m_a}\frac{\partial f_{bj}}{\partial v_a^i}\right) + O\left(\frac{1}{c^6}\right) \tag{3.29}$$
$$= \delta_{ab'}\frac{G}{2c^4}\varepsilon_a\{v_a^i n^j + v_b^j n^i\} + O\left(\frac{1}{c^6}\right), \tag{3.29'}$$

$$\{z_a^i, p_{bj}\}^* = \delta_{ab}\delta_{ij} - \sum_{c,d}\sum_{k,l}\{z_a^i, \phi_{ck}^1\}\frac{1}{m_c}\delta_{cd}\delta_{kl}\{\phi_{dl}^0, p_{bj}\} + O\left(\frac{1}{c^6}\right)$$

$$= \delta_{ab}\delta_{ij} + \frac{G}{4c^4}\frac{\partial}{\partial z_b^j} f_{ai} + O\left(\frac{1}{c^6}\right) \tag{3.30}$$

$$= \delta_{ab}\delta_{ij} + \varepsilon_a\varepsilon_b\frac{Gm_1m_2}{4c^4 r}\left\{\frac{G}{r}(\delta_{ij} - 2n_in_j) - \frac{\mathbf{v}_{a'}^2}{m_a}(\delta_{ij} - n_in_j)\right\} + O(1/c^6), \tag{3.30'}$$

$$\{p_{ai}, p_{bj}\}^* = O(1/c^6). \tag{3.31}$$

Under the framework of Dirac brackets, we have the Hamiltonian

$$H = H_0(\mathbf{p}_a, \mathbf{z}_a) + \frac{G^2}{4c^4}\left\{-\frac{2Gm_1^2m_2^2}{r^3} + \frac{1}{r^2}(m_1\mathbf{p}_2^2 + m_2\mathbf{p}_1^2)\right\}, \tag{3.32}$$

in which[12]

$$\begin{aligned}
H_0(\mathbf{p}_a, \mathbf{z}_a) = {} & \frac{\mathbf{p}_1^2}{2m_1} + \frac{\mathbf{p}_2^2}{2m_2} - \frac{Gm_1m_2}{r} - \frac{1}{8c^2}\left\{m_1\left(\frac{\mathbf{p}_1^2}{m_1^2}\right)^2 + m_2\left(\frac{\mathbf{p}_2^2}{m_2^2}\right)^2\right\} \\
& + \frac{G^2}{2c^2}\frac{m_1m_2(m_1+m_2)}{r^2} + \frac{G}{8c^2}\frac{m_1m_2}{r}\left\{-12\left(\frac{\mathbf{p}_1^2}{m_1^2} + \frac{\mathbf{p}_2^2}{m_2^2}\right) + 28\frac{(\mathbf{p}_1\cdot\mathbf{p}_2)}{m_1m_2} + 4\frac{(\mathbf{n}\cdot\mathbf{p}_1)(\mathbf{n}\cdot\mathbf{p}_2)}{m_1m_2}\right\} \\
& + \frac{1}{16c^4}\left\{m_1\left(\frac{\mathbf{p}_1^2}{m_1^2}\right)^3 + m_2\left(\frac{\mathbf{p}_2^2}{m_2^2}\right)^3\right\} - \frac{G^3}{4c^4}\frac{m_1m_2(m_1^2 + m_2^2 + 5m_1m_2)}{r^3} \\
& + \frac{G^2}{4c^4}\frac{m_1m_2^2}{r^2}\left(10\frac{\mathbf{p}_1^2}{m_1^2} + 19\frac{\mathbf{p}_2^2}{m_2^2}\right) + \frac{G^2}{4c^4}\frac{m_1^2m_2}{r^2}\left(19\frac{\mathbf{p}_1^2}{m_1^2} + 10\frac{\mathbf{p}_2^2}{m_2^2}\right) \\
& - \frac{G^2}{4c^4}\frac{m_1m_2(m_1+m_2)}{r^2}\left\{27\frac{(\mathbf{p}_1\cdot\mathbf{p}_2)}{m_1m_2} + 6\frac{(\mathbf{n}\cdot\mathbf{p}_1)(\mathbf{n}\cdot\mathbf{p}_2)}{m_1m_2}\right\} \\
& + \frac{G}{8c^4}\frac{m_1m_2}{r}\left\{5\left[\left(\frac{\mathbf{p}_1^2}{m_1^2}\right)^2 + \left(\frac{\mathbf{p}_2^2}{m_2^2}\right)^2\right] - 11\frac{\mathbf{p}_1^2\mathbf{p}_2^2}{m_1^2m_2^2} - 2\frac{(\mathbf{p}_1\cdot\mathbf{p}_2)^2}{m_1^2m_2^2} + 5\frac{\mathbf{p}_1^2(\mathbf{n}\cdot\mathbf{p}_2)^2 + (\mathbf{n}\cdot\mathbf{p}_1)^2\mathbf{p}_2^2}{m_1^2m_2^2}\right. \\
& \left. - 12\frac{(\mathbf{p}_1\cdot\mathbf{p}_2)(\mathbf{n}\cdot\mathbf{p}_1)(\mathbf{n}\cdot\mathbf{p}_2)}{m_1^2m_2^2} - 3\frac{(\mathbf{n}\cdot\mathbf{p}_1)^2(\mathbf{n}\cdot\mathbf{p}_2)^2}{m_1^2m_2^2}\right\}.
\end{aligned} \tag{3.33}$$

We can directly confirm that the canonical equation of motion in terms of Dirac brackets coincides with the reduced equation of motion, which is obtained from the Euler–Lagrange equation derived from (3.22) with (3.23) by the iteration method using the equations of motion in the lowest order. A relation between the Lagrangian system with acceleration terms and the Lagrangian system without acceleration terms is obtained by introducing new coordinate variables as done in (3.13).

Instead of (3.29), (3.30), and (3.31), we can get the ordinary canonical forms

$$\begin{aligned}
&\{\underline{z}_a^i, \underline{z}_b^j\}^* = \{\underline{p}_{ai}, \underline{p}_{bj}\}^* = 0, \\
&\{\underline{z}_a^i, \underline{p}_{bj}\}^* = \delta_{ab}\delta_{ij},
\end{aligned} \tag{3.34}$$

up to the approximation of order $1/c^4$, in which

$$\underline{\mathbf{z}}_a = \mathbf{z}_a - (G/4c^4)\mathbf{f}_a, \quad \underline{\mathbf{p}}_a = \mathbf{p}_a. \tag{3.35}$$

If we take account of the relationship between $\underline{r} = \{(\underline{\mathbf{z}}_1 - \underline{\mathbf{z}}_2)^2\}^{1/2}$ and $r = \{(\mathbf{z}_1 - \mathbf{z}_2)^2\}^{1/2}$,

$$\underline{r} = r\left\{1 - \frac{G}{4c^4 r}\left(2G\frac{m_1m_2}{r} - m_1\mathbf{v}_1^2 - m_2\mathbf{v}_2^2\right)\right\} + O\left(\frac{1}{c^6}\right), \tag{3.36}$$

we get

$$H = H_0(\underline{\mathbf{p}}_a, \underline{\mathbf{z}}_a), \tag{3.37}$$

where the second term on the right-hand side of (3.32) is absorbed in $-Gm_1m_2/\underline{r}$ of $H_0(\underline{\mathbf{p}}_a, \underline{\mathbf{z}}_a)$.

A similar procedure has been applied by Jaén *et al.*[2] to Wheeler–Feynman electrodynamics for two charged point particles up to order $1/c^4$. Jaén *et al.* stand on the assumption that the Hessian matrix $\partial^2 L/\partial \overset{(N)}{q}_i \partial \overset{(N)}{q}_j$ is regular, while we start with the singular Hessian matrix.

IV. GAUGE SYMMETRY AND ITS GENERATOR

For the Lagrangian with higher derivatives, if all the constraints are first class, the system is gauge invariant. In this case we can construct the generator of the gauge transformation by following the procedure similar to that of the ordinary Lagrangian[11] and prove the consistency between the gauge transformations in VPS and in PS, by using the equations of motion. In this section we will show them.

A. The construction of the gauge generator

Here all constraints are assumed to be first class. Using a Dirac algorithm, we define successively the secondary constraints from the primary ones ϕ_α^0;

$$\phi_\alpha^k \equiv \{\phi_\alpha^{k-1}, H_0\}. \tag{4.1}$$

Though H_T should be used in this algorithm, H_0 can be substituted instead of H_T, owing to the assumption that all constraints are first class.[13] This algorithm is continued until ϕ_α^m satisfies[14]

$$\phi_\alpha^{m+1} = \{\phi_\alpha^m, H_0\} = c_{\alpha k}^\beta \, \phi_\beta^k \quad (k \leqslant m). \tag{4.2}$$

The generator G of the gauge transformation is expressed as

$$G \equiv \varepsilon_k^\alpha(t)\phi_\alpha^k \quad (k = 0 \sim m), \tag{4.3}$$

where ε_k^α are determined by the condition

$$\frac{\partial G}{\partial t} + \{G, H_0\} \equiv 0 \mod(\phi^0), \tag{4.4}$$

which is nothing but the conservation law of G. From (4.3) and (4.4) it follows that

$$\frac{\partial \varepsilon_k^\alpha}{\partial t} + \{\varepsilon_k^\alpha, H^0\} + \varepsilon_{k-1}^\alpha + \varepsilon_m^\beta \, c_{\beta k}^\alpha = 0. \tag{4.5}$$

By using (4.5), $\varepsilon_k^\alpha (k<m)$ is successively determined and expressed in terms of $\varepsilon_m^\alpha(t)$ which is an arbitrary function of t.

If G satisfies the following conditions for all α;

$$\{G, \phi_\alpha^0\} \equiv 0 \mod (\phi^0), \tag{4.6}$$

G is the generator of the gauge transformation leaving the action invariant

$$\delta q_{(0)}^i \equiv \{q_{(0)}^i, G\} = \frac{\partial G}{\partial p_i^{(0)}}, \tag{4.7a}$$

$$\delta q_{(s)}^i \equiv \frac{d^s}{dt^s} \delta q_{(0)}^i. \tag{4.7b}$$

Proof:

$$\delta L = \sum \frac{\partial L}{\partial q_{(s)}^i} \delta q_{(s)}^i$$

$$= \left(\frac{\partial L}{\partial q_{(0)}^i} - \dot{p}_i^{(0)}\right) \delta q_{(0)}^i + \frac{d}{dt}\left(\sum_{s=0}^{N-1} p_i^{(s)} \, \delta q_{(s)}^i\right). \tag{4.8}$$

Owing to (4.6), (4.4) is equivalent to

$$\frac{\partial G}{\partial t} + \{G, H\} \equiv 0 \mod(\phi^0), \tag{4.9}$$

with H of (2.10) or (2.11). With the help of (2.16) [except for the equation of motion (2.16_4)], (4.9) in VPS turns out to be

$$\frac{\partial G}{\partial t} + \sum_{s=0}^{N-1}\left[\frac{\partial G}{\partial q_{(s)}^i} q^{(s+1)_i} + \frac{\partial G}{\partial p_i^{(s)}}\left(-p_i^{(s-1)} + \frac{\partial L}{\partial q_{(s)}^i}\right)\right] = 0, \tag{4.10}$$

where $p^{(-1)} = 0$. Using (4.10) we obtain[15]

$$\delta L = \left(\frac{\partial L}{\partial q_{(0)}^i} - \dot{p}_i^{(0)}\right) \frac{\partial G}{\partial p_i^{(0)}} + \frac{d}{dt}\left(\sum_{s=0}^{N-1} p_i^{(s)} \, \delta q_{(s)}^i\right)$$

$$= \frac{d}{dt}\left(-G + \sum_{s=0}^{N-1} p_i^{(s)} \, \delta q_{(s)}^i\right). \tag{4.11}$$

Thus the assertion has been proved.

Even when second-class constraints appear, if the series of first-class constraints (4.1) and (4.2) derived from primary constraints are completely separated from the series of second-class ones, this formulation on the gauge symmetry is valid for such a system.

B. The consistency of gauge transformations in VPS and PS

We define the gauge transformation in PS as

$$\begin{aligned} \bar{\delta} q_{(s)}^i &\equiv \{q_{(s)}^i, G\}, \\ \bar{\delta} p_i^{(s)} &\equiv \{p_i^{(s)}, G\}. \end{aligned} \tag{4.12}$$

The gauge transformation (4.7) in VPS is equivalent to the one of (4.12) under the modulo of the equations of motion;

$$\begin{aligned} \delta q_{(s)}^i &= \frac{d^s}{dt^s} \delta q_{(0)}^i \simeq \bar{\delta} q_{(s)}^i, \\ \delta p_i^{(s)}(q_{(0)}, \ldots, q_{(N)}) &\simeq \bar{\delta} p_i^{(s)}. \end{aligned} \tag{4.13}$$

We emphasize that these consistency conditions are much more serious than the ordinary case. For the ordinary Lagrangian, the relation $\delta \dot{q}^i = \bar{\delta} \dot{q}^i$ does not hold without using equations of motion if G contains terms higher than quadratic in p_i, as shown in Ref. 11.

Proof: Equation (4.9) can be written as

$$\frac{\partial G}{\partial t} + \sum_{s=0}^{N-1}\left(\frac{\partial G}{\partial q_{(s)}^i} \frac{\partial H}{\partial p_i^{(s)}} - \frac{\partial G}{\partial p_i^{(s)}} \frac{\partial H}{\partial q_{(s)}^i}\right) = \varepsilon^\alpha \phi_\alpha^0. \tag{4.14}$$

Differentiating (4.14) with respect to $p_j^{(r)}$ $(r \leqslant N-2)$ we obtain

$$\frac{\partial^2 G}{\partial t \partial p_j^{(r)}} + \sum_{s=0}^{N-1}\left(\frac{\partial^2 G}{\partial p_j^{(r)} \partial q_{(s)}^i} \frac{\partial H}{\partial p_i^{(s)}} - \frac{\partial^2 G}{\partial p_i^{(s)} \partial p_j^{(r)}} \frac{\partial H}{\partial q_{(s)}^i}\right)$$
$$+ \sum_{s=0}^{N-1}\left(\frac{\partial G}{\partial q_{(s)}^i} \frac{\partial^2 H}{\partial p_j^{(r)} \partial p_i^{(s)}} - \frac{\partial G}{\partial p_i^{(s)}} \frac{\partial^2 H}{\partial p_j^{(r)} \partial q_{(s)}^i}\right)$$
$$= \frac{\partial \varepsilon^\alpha}{\partial p_j^{(r)}} \phi_\alpha^0. \tag{4.15}$$

In the above, use has been made of the fact that ϕ_α^0 is independent of momenta other than $p_i^{(N-1)}$. From the expression (2.10) for H it follows that

$$\frac{\partial^2 H}{\partial p_j^{(r)} \partial p_i^{(s)}} = \frac{\partial}{\partial p_i^{(s)}} q_{(r+1)}^j = 0 \quad (\text{for } r \leqslant N-2),$$

$$\frac{\partial^2 H}{\partial p_j^{(r)} \partial q_{(s)}^i} = \frac{\partial}{\partial q_{(s)}^i} q_{(r+1)}^j = \delta_i^j \delta_{s,r+1}.$$

The substitution of these into (4.15) leads us to

$$\frac{\partial^2 G}{\partial t \, \partial p_j^{(r)}} + \sum_{s=0}^{N-1}\left(\frac{\partial^2 G}{\partial p_j^{(r)} \partial q_{(s)}^i} \frac{\partial H}{\partial p_i^{(s)}} - \frac{\partial^2 G}{\partial p_j^{(r)} \partial p_i^{(s)}} \frac{\partial H}{\partial q_{(s)}^i}\right)$$
$$- \frac{\partial G}{\partial p_j^{(r+1)}} = \frac{\partial \varepsilon^\alpha}{\partial p_j^{(r)}} \phi_\alpha^0. \tag{4.16}$$

With the aid of (2.16a)–(2.16e), (4.16) turns out to be

$$\frac{\partial^2 G}{\partial t\partial p_j^{(r)}} + \sum_{s=0}^{N-1} \frac{\partial^2 G}{\partial p_j^{(r)}\,\partial q^i_{(s)}}\, q^i_{(s+1)} + \sum_{s=1}^{N-1} \frac{\partial^2 G}{\partial p_j^{(r)}\,\partial p_i^{(s)}}\left(-p_i^{(s-1)} + \frac{\partial L}{\partial q^i_{(s)}}\right) + \frac{\partial^2 G}{\partial p_i^{(0)}\,\partial p_j^{(r)}}\,\frac{\partial L}{\partial q^i_{(0)}} - \frac{\partial G}{\partial p_j^{(r+1)}} \equiv 0 \mod(\phi^0). \quad (4.17)$$

From the Ostrogradski transformation and (4.17) it follows that

$$\frac{\partial^2 G}{\partial t\partial p_j^{(r)}} + \sum_{s=0}^{N-1} \frac{\partial^2 G}{\partial p_j^{(r)}\,\partial q^i_{(s)}}\, q^i_{(s+1)} + \sum_{s=1}^{N-1} \frac{\partial^2 G}{\partial p_j^{(r)}\,\partial p_i^{(s)}}\,\dot p_i^{(s)}. + \frac{\partial^2 G}{\partial p_i^{(0)}\,\partial p_j^{(r)}}\,\frac{\partial L}{\partial q^i_{(0)}} - \frac{\partial G}{\partial p_j^{(r+1)}} \equiv 0 \mod(\phi^0). \quad (4.18)$$

This is transformed as

$$\frac{d}{dt}\{q^i_{(s)},G\} \equiv \{q^i_{(s+1)},G\} - \frac{\partial^2 G}{\partial p_i^{(0)}\,\partial p_j^{(s)}}\left(\frac{\partial L}{\partial q^j_{(0)}} - \dot p_j^{(0)}\right) \mod(\phi^0). \quad (4.19)$$

It should be noticed that in the derivation of (4.19) we have not used the dynamical equation (2.16d), but only kinematical relations.

Since the primary constraints ϕ^0_α identically vanish in VPS, by using the equations of motion (4.19) gives the first relation of (4.13) in VPS.

Next to prove the second relation of (4.13) we differentiate (4.10) with respect to $q^j_{(r+1)}$ $(r\leqslant N-2)$ and use (4.18) or (4.19). Then we obtain

$$\left(\frac{d}{dt}\,\bar\delta q^i_{(N-1)}\right)\frac{\partial^2 L}{\partial q^i_{(N)}\,\partial q^j_{(r+1)}} - \left(\frac{d}{dt}\,\bar\delta p_j^{(r+1)}\right) - \bar\delta p_j^{(r)} + \sum_{s=0}^{N-1} \frac{\partial^2 L}{\partial q^j_{(r+1)}\,\partial q^i_{(s)}}\,\bar\delta q^i_{(s)} + \left(\frac{\partial L}{\partial q^i_{(0)}} - \dot p_i^{(0)}\right) \times\left(\sum_{s=0}^{N-1} \frac{\partial^2 G}{\partial p_i^{(0)}\,\partial p_k^{(s)}}\,\frac{\partial p_k^{(s)}}{\partial q^j_{(r+1)}} + \frac{\partial^2 G}{\partial p_i^{(0)}\partial q^j_{(r+1)}}\right) = 0, \quad (4.20)$$

with $p_i^{(N)} \equiv 0$, $q^i_{(N)} \equiv \dot q^i_{(N-1)}$. Using (4.20) we are led to

$$\begin{aligned}
\delta p_i^{(s)} &= \delta\left(\frac{\partial L}{\partial q^i_{(s+1)}} - \dot p_i^{(s+1)}\right)\\
&= \sum_{r=0}^{N} \frac{\partial^2 L}{\partial q^j_{(r)}\partial q^i_{(s+1)}}\,\delta q^j_{(r)} - \frac{d}{dt}\,\delta p_i^{(s+1)}\\
&= \bar\delta p_i^{(s)} + \sum_{r=0}^{N-1} \frac{\partial^2 L}{\partial q^i_{(s)}\,\partial q^j_{(r)}}\left(\delta q^j_{(r)} - \bar\delta q^j_{(r)}\right)\\
&\quad + \frac{\partial^2 L}{\partial q^i_{(s)}\,\partial q^j_{(N)}}\left(\delta q^j_{(N)} - \frac{d}{dt}\,\bar\delta q^j_{(N-1)}\right)\\
&\quad - \frac{d}{dt}\left(\delta p_i^{(s+1)} - \bar\delta p_i^{(s+1)}\right) + \left(\dot p_j^{(0)} - \frac{\partial L}{\partial q^j_{(0)}}\right)\\
&\quad \times\left(\sum_{r=0}^{N-1} \frac{\partial^2 G}{\partial p_j^{(0)}\,\partial p_k^{(r)}}\,\frac{\partial p_k^{(r)}}{\partial q^i_{(s+1)}} + \frac{\partial^2 G}{\partial p_j^{(0)}\,\partial q^i_{(s+1)}}\right).
\end{aligned} \quad (4.21)$$

By taking into account the equation of motion and $\delta q^i_{(s)} \simeq \bar\delta q^i_{(s)}$, (4.21) reduces to

$$\delta p_i^{(s)} \simeq \bar\delta p_i^{(s)} - \frac{d}{dt}\left(\delta p_i^{(s+1)} - \bar\delta p_i^{(s+1)}\right). \quad (4.22)$$

Starting from $s = N-1$ in (4.22), we obtain successively

$$\delta p_i^{(s)} \simeq \bar\delta p_i^{(s)}.$$

Thus it is verified that for the transformation $\bar\delta q^i_{(s)}$ and $\bar\delta p_i^{(s)}$, the action which is invariant for $\delta q^i_{(s)}$ and $\delta p_i^{(s)}$, is also invariant under the modulo of the equations of motion (2.16). Then the Hamilton equations of (2.16) are invariant under the transformation given by (4.12).

V. EXAMPLES OF GAUGE TRANSFORMATIONS

To illustrate the results of Sec. IV, we present a few examples. (1) A model with the Lagrangian

$$L = \dddot q^1 \dddot q^2 + \ddot q^1(\ddot q^2 - \ddot q^3) + \dot q^1(\dot q^2 - \dot q^3) - q^1 q^3. \quad (5.1)$$

The momenta conjugate to $\overset{(s)i}{q}$ are

$$\begin{aligned}
&p_1^{(2)} = \dddot q^2, \quad p_2^{(2)} = \dddot q^1, \quad p_3^{(2)} = 0,\\
&p_1^{(1)} = \ddot q^2 - \ddot q^3 - \dot p_1^{(2)} = \ddot q^2 - \ddot q^3 - \overset{(4)}{q}{}^2,\\
&p_2^{(1)} = \ddot q^1 - \dot p_2^{(2)}, \quad p_3^{(1)} = -\ddot q^1,\\
&p_1^{(0)} = \dot q^2 - \dot q^3 - \dot p_1^{(1)}, \quad p_2^{(0)} = \dot q^1 - \dot p_2^{(1)},\\
&p_3^{(0)} = -\dot q^1 - \dot p_3^{(3)}.
\end{aligned} \quad (5.2)$$

The primary constraint is only

$$\phi^0 = p_3^{(2)}. \quad (5.3)$$

The Hamiltonian is given by

$$\begin{aligned}
H_0 &= p_1^{(2)} p_2^{(2)} + p_1^{(1)} q^1_{(2)} + p_2^{(1)} q^2_{(2)} + p_3^{(1)} q^3_{(2)} + p_1^{(0)} q^1_{(1)}\\
&\quad + p_2^{(0)} q^2_{(1)} + p_3^{(0)} q^3_{(1)} + q^1_{(2)}\left(q^3_{(2)} - q^2_{(2)}\right)\\
&\quad + q^1_{(1)}\left(q^3_{(1)} - q^2_{(1)}\right) + q^1_{(0)} q^3_{(0)},
\end{aligned} \quad (5.4)$$

and

$$H_T = H_0 + v p_3^{(2)}. \quad (5.5)$$

The stationarity condition of ϕ^0 yields the following secondary constraints:

$$\begin{aligned}
&\phi^1 = \{\phi^0,H_0\} = -p_3^{(1)} - q^1_{(2)},\\
&\phi^2 = \{\phi^1,H_0\} = p_3^{(0)} + q^1_{(1)} - p_2^{(2)},\\
&\phi^3 = -q^1_{(0)} + p_2^{(1)}, \quad \phi^4 = -p_2^{(0)}.
\end{aligned} \quad (5.6)$$

All constraints in (5.3) and (5.6) are first class. Hence the system is gauge invariant and we can construct the generator of the gauge transformation:

$$\begin{aligned}
G &= -\epsilon p_2^{(0)} + \dot\epsilon\left(q^1_{(0)} - p_2^{(1)}\right) + \ddot\epsilon\left(p_3^{(0)} + q^1_{(1)} - p_2^{(2)}\right)\\
&\quad + \dddot\epsilon\left(p_3^{(1)} + q^1_{(2)}\right) + \overset{(4)}{\epsilon}\, p_3^{(2)}.
\end{aligned} \quad (5.7)$$

This G produces the following transformations:

$$\begin{aligned}
&\bar\delta q^1_{(s)} = \{q^1_{(s)},G\} = 0, \quad \bar\delta q^2_{(s)} = -\overset{(s)}{\epsilon}, \quad \bar\delta q^3_{(s)} = \overset{(s+1)}{\epsilon},\\
&\bar\delta p_1^{(s)} = \{p_1^{(s)},G\} = -\overset{(s+1)}{\epsilon}, \quad \bar\delta p_2^{(s)} = \bar\delta p_3^{(s)} = 0.
\end{aligned} \quad (5.8)$$

These $\bar{\delta}q^i_{(s)}$ satisfy $\delta q^i_{(s)} = (d^s/dt^s)\delta q^i_{(0)} = \bar{\delta}q^i_{(s)}$, and $\bar{\delta}p_i^{(s)}$ are consistent with the definition of the momenta in (5.2), that is $\delta p_i^{(s)} = \bar{\delta}p_i^{(s)}$. Under this gauge transformation, we observe the invariance of the action

$$\delta L = -\frac{d}{dt}(\ddot{q}^1\ddot{\epsilon} + \dot{q}^1\ddot{\epsilon} + q^1\dot{\epsilon}). \tag{5.9}$$

(2) Yang–Mills theory with higher derivatives

$$L = -\tfrac{1}{4}F_{\mu\nu}F^{\mu\nu} - \kappa D_\rho F_{\mu\nu}D^\rho F^{\mu\nu}, \tag{5.10}$$

where

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu + [A_\mu, A_\nu],$$

$$D_\mu X = \partial_\mu X + [A_\mu, X],$$

and

$$A_\mu = T_a A^a_\mu.$$

The generator for this system was already given by Galvao *et al.*[16] We take this model in order to compare with our formulation.

By denoting momenta conjugate to A^μ and $\dot{A}^\mu \equiv A^\mu_{(1)}$ by $p_\mu \equiv p_\mu^{(0)}$ and $p_\mu^{(1)}$, respectively, we obtain

$$p_0^{(1)} = 0, \quad p_i^{(1)} = 4\kappa D_0 F_{0i},$$

$$p_0 = 4\kappa D^i D_0 F_{0i}, \tag{5.11}$$

$$p_i = 4\kappa(D^j D_j F_{0i} + D^j D_0 F_{ji}) - D_0 p_i^{(1)} + F_{0i},$$

and

$$H_0 = (1/8\kappa)(p_i^{(1)})^2 + p_i^{(1)}(D_i A^0_{(1)} + [A^0, A^i_{(1)}] + [F_{0i}, A^0]) + \kappa D_0 F_{ij} D^0 F^{ij} + 2\kappa D_i F_{0j} D^i F^{0j} + \kappa D_i F_{jk} D^i F^{jk} + \tfrac{1}{4}F_{\mu\nu}F^{\mu\nu} + p_\mu A^\mu_{(1)}. \tag{5.12}$$

The constraints are as follows:

$$\begin{aligned} \phi^0 &= p_0^{(1)}, \\ \phi^1 &= -p_0 - D^i p_i^{(1)}, \\ \phi^2 &= -D^i p_i + D^i[A^0, p_i^{(1)}] + [F^{0i}, p_i], \\ \phi^3 &= [\phi^2, A^0], \end{aligned} \tag{5.13}$$

which are first class. The generator of the gauge transformation turns out to be

$$G = \int d^3x (p_\mu D^\mu \epsilon + p_\mu^{(1)} \partial_0 D^\mu \epsilon), \tag{5.14}$$

after a partial integration with respect to x^i. This G satisfies the conditions (4.4) and (4.6). The transformations induced by G are

$$\begin{aligned} &\bar{\delta}A^\mu = D^\mu \epsilon, \quad \bar{\delta}A^\mu_{(1)} = \partial_0 D^\mu \epsilon, \\ &\bar{\delta}p_\mu = -[\epsilon, p_\mu] - [\dot{\epsilon}, p_\mu^{(1)}], \\ &\bar{\delta}p_\mu^{(1)} = -[\epsilon, p_\mu^{(1)}], \end{aligned} \tag{5.15}$$

which also satisfy the consistency condition without using equations of motion. Under the transformation (5.15) we obtain

$$\delta L = 0. \tag{5.16}$$

APPENDIX: $\overset{(N)\alpha}{q}$ INDEPENDENCE OF $\partial L/\partial \overset{(N)\alpha}{q}$

We shall show that $\partial L/\partial \overset{(N)\alpha}{q}\big|_{\overset{(N)\alpha}{q} = f^\alpha}$ is independent of $\overset{(N)\alpha}{q}$.

From (2.8) and (2.9), we obtain the identity

$$p_a^{(N-1)} = \frac{\partial L}{\partial \overset{(N)a}{q}}\bigg|_{\overset{(N)a}{q} = f_a} \qquad (a = 1 \sim n - r). \tag{A1}$$

Differentiating (A1) with respect to $\overset{(N)\alpha}{q}$, we get

$$A_{ab}\frac{\partial f^b}{\partial \overset{(N)\alpha}{q}} + A_{a\alpha} = 0. \tag{A2}$$

On the other hand, there are C^a_α such that

$$A_{\alpha i} = C^a_\alpha A_{ai} \quad (\alpha = n - r + 1, \ldots, n), \tag{A3}$$

owing to the assumption (2.8),

$$\det(A_{ab}) \neq 0 \tag{A4}$$

and

$$\operatorname{rank}(A_{ij}) = n - r. \tag{A5}$$

Using (A2) and (A3), we obtain

$$\begin{aligned} \frac{d}{d\overset{(N)\beta}{q}}\left[\frac{\partial L}{\partial \overset{(N)\alpha}{q}}\bigg|_{\overset{(N)\alpha}{q} = f^\alpha}\right] &= A_{\alpha a}\frac{\partial f^a}{\partial \overset{(N)\beta}{q}} + A_{\alpha\beta} \\ &= C^a_\alpha\left(A_{ab}\frac{\partial f^b}{\partial \overset{(N)\beta}{q}} + A_{a\beta}\right) = 0. \end{aligned} \tag{A6}$$

[1] A. Pais and G. E. Uhlenbeck, Phys. Rev. **79**, 145 (1950); Y. Katayama, Prog. Theor. Phys. **9**, 31 (1953).

[2] X. Jaén, J. Llosa, and A. Molina, Phys. Rev. D **34**, 2302 (1986).

[3] R. Utiyama and B. S. DeWitt, J. Math. Phys. **3**, 608 (1962); K. S. Stelle, Phys. Rev. D **16**, 953 (1977); E. S. Fradkin and A. A. Tseytlin, Nucl. Phys. B **201**, 469 (1982).

[4] M. Ostrogradski, see E. T. Whittaker, *Analytical Dynamics* (Cambridge U.P., London, 1937), 4th ed., p. 265.

[5] P. A. M. Dirac, *Lecture on Quantum Mechanics* (Belfer Graduate School, Yeshiva University, New York, 1964).

[6] B. M. Barker and R. F. O'Connell, Phys. Lett. A **78**, 231 (1980).

[7] T. Ohta, H. Okamura, T. Kimura, and K. Hiida, Prog. Theor. Phys. **50**, 492 (1973).

[8] T. Ohta and T. Kimura, Prog. Theor. Phys. **79**, 819 (1988).

[9] T. Ohta, H. Okamura, T. Kimura, and K. Hiida, Prog. Theor. Phys. **51**, 1220 (1974).

[10] G. Schäfer, Phys. Lett. A **100**, 128 (1984).

[11] R. Sugano, Y. Saito, and T. Kimura, Prog. Theor. Phys. **73**, 283 (1986).

[12] T. Ohta and T. Kimura, Prog. Theor. Phys. **76**, 329 (1986).

[13] It is noticed that all Ostrogradski transformations (2.3) and (2.4) for $i = \alpha$ turn out to be constraints due to this assumption, since $\{\phi^0_\alpha, \phi^k_\beta\} \simeq 0 (k \geqslant N - 1)$ (see Sec. II).

[14] In general m depends on α. For simplicity we use the same m. Its generalization is easy.

[15] For a detail of the derivation of (4.11), see Ref. 11.

[16] C. A. P. Galvao and J. B. T. Boechat, preprint CBPF-NF-006/89 (1987).

Expressions for the zeta-function regularized Casimir energy

E. Elizalde and A. Romeo
Department of Structure and Constituents of Matter, Faculty of Physics, University of Barcelona, Diagonal 647, 08028 Barcelona, Spain

(Received 18 October 1988; accepted for publication 11 January 1989)

The expansion of the Casimir energy for a scalar field with mass m, in a space where one dimension has been compactified into a circle of length a, leads to a double-infinite series that can be regularized by analytic continuation in the space dimension. The dimensionally regularized sum is then expressed as a power series in am by means of zeta-function expansions. The two possibilities of odd and even space dimensions are distinguished. In the odd space dimension we give a power expansion for small am, in addition to the asymptotic behavior. For the even space dimension, an expansion valid for any value of am is obtained. The contribution of higher-order terms is studied and, for the three-dimensional space, results for different values of the compactification length are shown.

I. INTRODUCTION

In 1948 Casimir showed that neutral, perfectly conducting, parallel plates in an electromagnetic field attract each other. The attractive force can be thought of as caused by the change in the zero-point energy of the field when the plates are brought into position. In fact, for any quantum field, the zero-point modes are affected by the presence of any sort of boundaries or external constraints, so that the zero-point energy is modified. The evaluation of these vacuum energies is sometimes ambiguous and the outcome is usually a divergent sum. In canonical quantization, this can be regarded as a consequence of the fact that this scheme does not fix the ordering of noncommuting operators, making additional prescriptions necessary for removing ambiguities.

One of the most commonly used procedures for obtaining the vacuum energy is direct evaluation of infinite sums over eigenvalues of zero-point field modes. These sums, which happen to be formally divergent, may be regularized by a variety of techniques, e.g., momentum cutoff[1] or dimensional regularization.[2]

Here we will consider a massive noninteracting scalar field in a region bounded by two $(d-1)$-dimensional hyperplanes contained in a d-dimensional space. By subjecting the field to periodic boundary conditions one obtains the theory that corresponds to a space where one dimension has been made compact.

Basic quantum field theory (QFT) tells us that for a scalar field with mass m, the Hamiltonian takes the form

$$H=\frac{1}{2}\sum_k \omega_k (a_k^\dagger a_k + a_k a_k^\dagger)$$
$$=\frac{1}{2}\sum_k \omega_k\left(n_k+\frac{1}{2}\right), \tag{1.1}$$

where $\omega_k^2 = k^2+m^2$ are the eigenvalues of the Klein–Gordon operator; $a_k^\dagger$ and a_k satisfy

$$[a_k,a_{k'}^\dagger]=\delta_{kk'},\quad [a_k,a_{k'}]=[a_k^\dagger,a_{k'}^\dagger]=0; \tag{1.2}$$

and $n_k \equiv a_k a_k^\dagger$ is the number operator whose eigenvalues are non-negative integers.

Since the vacuum state $|0\rangle$ is defined by $a_k|0\rangle=0$, when computing the vacuum expectation value $E_0 \equiv \langle 0|H|0\rangle$ one obtains

$$E_0=\frac{1}{2}\sum_k \omega_k, \tag{1.3}$$

which is a divergent quantity. The reason why this is so is that we have introduced no additional prescription, i.e., normal ordering for the Fock space operators in the Hamiltonian. Had we defined H to be normal ordered, we would have obtained

$$:H: = \sum_k \omega_k a_k^\dagger a_k \tag{1.4}$$

and thus

$$E_0' \equiv \langle 0|:H:|0\rangle = 0. \tag{1.5}$$

Nevertheless, introducing normal ordering amounts to making an additional postulate in order to render the vacuum state energy finite. The arbitrariness of this prescription points out the existence of an ambiguity in the zero-point energy.

On the other hand, it is known that when a physical field is forced to satisfy certain boundary conditions, the presence of the boundaries induces a change in the energy spectrum that modifies the zero-point energy. Thus in order to remove the above-mentioned ambiguity, it is reasonable to define the physical vacuum energy as a difference in zero-point energy. Let $\partial\Gamma$ be an arbitrary boundary for the field in question, $E_0[\partial\Gamma]$ the zero-point energy in the presence of that boundary, and $E_0[0]$ the zero-point energy without boundary. Then, the Casimir energy is formally defined as

$$E_c[\partial\Gamma]=E_0[\partial\Gamma]-E_0[0]. \tag{1.6}$$

The definition (1.6) gives $E_c[0]=0$, so that the intuitive idea of a noninteracting vacuum is preserved.

II. EVALUATION OF THE CASIMIR ENERGY FOR A MASSIVE NONINTERACTING BOSONIC FIELD

Given a scalar field ϕ in $\mathbf{R}^d$ with mass m, the constraint of having two neutral parallel plates, orthogonal to some certain direction of the space—say the x_1 axis—can be implemented by the periodic boundary condition

$$\phi(x_1=0)=\phi(x_1=a), \tag{2.1}$$

which means compactification in a circle of length a, giving rise to the space $\mathbf{S}^1\times\mathbf{R}^{d-1}$.

 0022-2488/89/051133-07$02.50

By solving the Klein–Gordon equation with condition (2.1), the field modes are found to be

$$\phi(x_1,\mathbf{x}_T,t) \sim e^{2\pi i n x_1/a} e^{i\mathbf{k}_T\cdot\mathbf{x}_T} e^{-i\omega_{n,\mathbf{k}_T} t}, \tag{2.2}$$

with integer n, and

$$\omega_{n,\mathbf{k}_T} = \sqrt{(2n\pi/a)^2 + \mathbf{k}_T^2 + m^2}, \tag{2.3}$$

where $\mathbf{x}_T$ corresponds to the transverse coordinates.

Now, assuming that the plates have sides of length L, with $L \gg a$, the zero-point energy of the field inside the resulting cavity (whose volume is $L^{d-1}a$) will be given according to (1.3) by

$$\begin{aligned} E(d,a,m) &= \left(\frac{L}{2\pi}\right)^{d-1} \int d^{d-1}\mathbf{k}_T \frac{1}{2} \sum_n \omega_{n,\mathbf{k}_T} \\ &= \left(\frac{L}{2\pi}\right)^{d-1} \int d^{d-1}\mathbf{k}_T \sum_{n=1}^{\infty} \\ &\quad \times \sqrt{\left(\frac{2n\pi}{a}\right)^2 + \mathbf{k}_T^2 + m^2}. \end{aligned} \tag{2.4}$$

The mode for which $n = 0$ has not been included because, since it carries zero longitudinal momentum, it is irrelevant. As it stands (2.4) is infinite. It is quite clear that what makes this sum divergent is the contribution of high-frequency modes. Nevertheless, this contribution ought to be independent of a because these modes are not affected by the presence of the plates; thus they should cancel in calculations of forces or in comparisons of the energy density in the presence and absence of plates.

The quantity (2.4) will be regularized by analytic continuation in d. Making use of

$$d^D\mathbf{k}_T = \frac{2\pi^{D/2}}{\Gamma(D/2)} \|\mathbf{k}_T\|^{D-1} d\,\|\mathbf{k}_T\| \tag{2.5}$$

and

$$\int_0^\infty dt\, t^{a-1}(1+t)^{-a-b} = B(a,b), \tag{2.6}$$

where B is Euler's beta function, one is led to

$$\begin{aligned} E(d,a,m) &= \left(\frac{L}{2\pi}\right)^{d-1} \pi^{(d-1)/2} \frac{\Gamma(-d/2)}{\Gamma(-1/2)} \left(\frac{2\pi}{a}\right)^d \\ &\quad \times \sum_{n=1}^{\infty} \left(n^2 + \left(\frac{am}{2\pi}\right)^2\right)^{d/2}. \end{aligned} \tag{2.7}$$

After evaluating the n summation (see Appendix A), (2.7) takes the form

$$\begin{aligned} &E(d,a,m) \\ &= -\frac{1}{2}\frac{L^{d-1}}{a^d} \pi^{-(d+1)/2} \left\{ -\sqrt{\pi}\,\Gamma\left(-\frac{d}{2}\right)\left(\frac{am}{2}\right)^d \right. \\ &\quad + \left(\frac{am}{2}\right)^{d+1} \left[\Gamma\left(-\frac{d+1}{2}\right) \right. \\ &\quad \left.\left. + 4\sum_{n=1}^{\infty} \frac{K_{(d+1)/2}(amn)}{(amn/2)(d+1)/2}\right]\right\}. \end{aligned} \tag{2.8}$$

Terms independent of a have no physical significance and may differ from one regularization scheme to another: Looking at (2,8) we notice at once that the first term in the curly braces is one of them, so it is to be dropped. The first term in the square brackets can be thought of as an energy density which would occur even in the absence of plates: For consistency with (1.6) we have to substract it as well. Then, what we obtain is the physically relevant energy, say ϵ:

$$\begin{aligned} \epsilon(d,a,m) &= -2\frac{L^{d-1}}{a^d} \pi^{-(d+1)/2} \left(\frac{am}{2}\right)^{d+1} \\ &\quad \times \sum_{n=1}^{\infty} \frac{K_{(d+1)/2}(amn)}{(amn/2)(d+1)/2} \\ &\equiv -2\frac{L^{d-1}}{a^d} \pi^{-(d+1)/2} S(d,a,m), \end{aligned} \tag{2.9}$$

where the definition of the sum $S(d,a,m)$ has been chosen for convenience. For small am, it is possible to use the expansion

$$(x/2)^\nu K_\nu(x) = \tfrac{1}{2}\Gamma(\nu) + O(x^2), \quad (\nu > 0), \quad x \ll 1 \tag{2.10}$$

to obtain

$$\begin{aligned} &\epsilon(d,a,m) \\ &\quad \simeq -(L^{d-1}/a^d)\pi^{-(d+1)/2}[\Gamma(d+1/2)\zeta(d+1) \\ &\qquad + O((am)^2)], \quad am \ll 1, \end{aligned} \tag{2.11}$$

where ζ is the Riemann zeta function. Equation (2.11) can be viewed as the result for $m = 0$ plus small-m corrections. From here on it is clear that in the massless case, the result is finite for all positive d and always negative. Thus this energy gives rise to a force which tends to contract the system. In general, lower modes can be considered to be responsible for the dependence of the Casimir energy on a. On the other hand, if we take $ma \gg 1$, then the leading term in the n sum will be the one with $n = 1$. When we are in these conditions, the asymptotic behavior

$$K_\nu(x) \simeq \sqrt{\pi/2x}\, e^{-x}, \quad x \gg 1 \tag{2.12}$$

gives

$$\epsilon(d,a,m) \simeq (L^{d-1}/a^d)(am/8\pi)^{d/2} e^{-am}, \quad am \gg 1. \tag{2.13}$$

In this case, the energy coming from the lower modes is dominated by m and does not depend on a so strongly, as in the previous case (small-am limit).

III. ZETA-FUNCTION REGULARIZATION OF THE CASIMIR ENERGY

The method used here, which rests upon the properties of the Riemann zeta function, is actually different from the procedure of direct zeta-function regularization. By this we mean that now one starts from a sum mode that has already been regularized (in our case, dimensionally). Nevertheless, in analogy with the observation made by Actor[3] concerning thermodynamic potentials, for nonzero mass we have a rather complicated function whose power series in a, or rather in am, would be, as a general rule, difficult to obtain.

As a matter of fact, our problem involves a Bessel function series of the same type as the one that appears when evaluating the partition function for a relativistic Bose gas with nonzero mass[3] and in the absence of any chemical potential.

From now on, we will focus on the sum $S(d,a,m)$ defined in (2.9), which can also be expressed as

$$S(d,a,m) = \sum_{n=1}^{\infty} \frac{1}{n^{d+1}}\left(\frac{amn}{2}\right)^{(d+1)/2} K_{(d+1)/2}(amn), \tag{3.1}$$

and we shall consider two possibilities.

A. Odd space dimension

Now we have $d = 2p - 1$, $p = 1,2,3,...,$ so that we are posed with

$$S(2p-1,a,m) = \sum_{n=1}^{\infty} \frac{1}{n^{2p}}\left(\frac{amn}{2}\right)^{p} K_p(amn). \tag{3.2}$$

It can be found that there exists an ascending power series for $K_p(z)$ for the non-negative integer p, namely,

$$\left(\frac{z}{2}\right)^p K_p(z) = \frac{1}{2}\sum_{k=0}^{p-1} \frac{(p-k-1)!}{k!}\left(-\frac{z^2}{4}\right)^k + (-1)^p \sum_{k=0}^{\infty}\left(\frac{z^2}{4}\right)^{p+2k} \frac{1}{k!(p+k)!} \times\left(-\ln\left(\frac{z}{2}\right) + C(p,k)\right), \tag{3.3}$$

where

$$C(p,k) \equiv \psi(k+1) + \psi(k+p+1). \tag{3.4}$$

The occurrence of the logarithm in (3.3) is quite important as it shows that $K_p(z)$ necessarily has a branch point at $z = 0$ and a cut to infinity which is placed along the negative z axis.

By putting (3.3) into (3.1) it turns out that

$$S(2p-1,a,m) = S_1 + S_2, \tag{3.5}$$

with

$$S_1 \equiv \sum_{n=1}^{\infty} \frac{1}{n^{2p}} \sum_{k=0}^{p-1} \frac{1}{2}\frac{(p-k-1)!}{k!}(-1)^k\left(\frac{amn}{2}\right)^{2k} \tag{3.6}$$

and

$$S_2 \equiv (-1)^p \sum_{n=1}^{\infty} \frac{1}{n^{2p}} \sum_{k=0}^{\infty}\left(\frac{amn}{2}\right)^{2k+2p} \frac{1}{k!(p+k)!} \times\left(-\ln\left(\frac{amn}{2}\right) + C(p,k)\right). \tag{3.7}$$

As for S_1, the convergence of the series for $k < p$ makes it possible to interchange the summatories, obtaining

$$S_1 = \sum_{k=0}^{\infty} \frac{1}{2}\left(\frac{am}{2}\right)^{2k} \frac{(p-k-1)!}{k!}(-1)^k \zeta(2p-2k). \tag{3.8}$$

In fact, S_2 is a bit more tricky: By separating $\ln(amn/2)$ into $\ln n$ and $\ln(am/2)$, it can also be written as

$$S_2 = (-1)^p (am/2)^{2p}(-S_{21} - \ln(am/2)S_{22} + S_{23}), \tag{3.9}$$

where

$$S_{21} \equiv \sum_{n=1}^{\infty} \ln n \sum_{k=0}^{\infty} n^{2k} \frac{(am/2)^{2k}}{k!(p+k)!},$$
$$S_{22} \equiv \sum_{n=1}^{\infty}\sum_{k=0}^{\infty} n^{2k} \frac{(am/2)^{2k}}{k!(p+k)!}, \tag{3.10}$$
$$S_{23} \equiv \sum_{n=1}^{\infty}\sum_{k=0}^{\infty} n^{2k} \frac{(am/2)^{2k}}{k!(p+k)!} C(p,k).$$

The results for the double sums (3.10) can be obtained from Appendix B and give rise to

$$S_2 = (-1)^p\left(\frac{am}{2}\right)^{2p} \times\left[-\frac{1}{2p!}\ln 2\pi - \sum_{k=1}^{\infty} \frac{(-1)^k(2k)!}{2(2\pi)^{2k}k!(k+p)!} \times\zeta(2k+1)\left(\frac{am}{2}\right)^{2k} + \frac{\pi^{3/2}}{4(p-\frac{1}{2})!} + \frac{1}{2p!}\ln\left(\frac{am}{2}\right) - \frac{1}{2p!}C(p,0)\right]. \tag{3.11}$$

By putting the results (3.10) and (3.11) together, we can see that the regularized, physically relevant vacuum energy becomes

$$\epsilon(2p-1,a,m) = -2\frac{L^{2p-2}}{a^{2p-1}}\pi^{-p} S(2p-1,a,m)$$
$$= -2\frac{L^{2p-2}}{a^{2p-1}}\pi^{-p}\left\{\sum_{k=0}^{p-1}(-1)^k \frac{(p-k-1)!}{2k!}\zeta(2p-2k)\left(\frac{am}{2}\right)^{2k} + (-1)^{p+1}\frac{\pi^{3/2}}{4(p-\frac{1}{2})!}\left(\frac{am}{2}\right)^{2p-1} + \frac{(-1)^p}{2p!}\left(\ln\left(\frac{am}{4\pi}\right) - C(p,0)\right)\left(\frac{am}{2}\right)^{2p} + (-1)^p\sum_{k=1}^{\infty}\frac{(-1)^{k+1}(2k)!}{2(2\pi)^{2k}k!(k+p)!}\zeta(2k+1)\left(\frac{am}{2}\right)^{2p+2k}\right\}. \tag{3.12}$$

For large am we obtain, from (2.13),

$$\epsilon(2p-1,a,m) \simeq (L^{2p-2}/a^{2p-1})(am/8\pi)^{p-1/2}e^{-am}. \tag{3.13}$$

B. Even space dimension

We will now assume that $d = 2p$, $p = 1,2,3,...;$ it is necessary to find a power expansion for

$$S(2p,a,m) = \sum_{n=1}^{\infty} \frac{1}{n^{2p+1}}\left(\frac{amn}{2}\right)^{p+1/2} K_{p+1/2}(amn). \tag{3.14}$$

For the modified Bessel function there is an asymptotic expansion for the large argument and real parameter, i.e.,

$$K_\nu(z) = \sqrt{\frac{\pi}{2z}}\,e^{-z}\sum_{k=0}^{\infty}\frac{\Gamma(\nu+k+\frac{1}{2})}{\Gamma(\nu-k+\frac{1}{2})}\frac{1}{k!}\frac{1}{(2z)^k}. \tag{3.15}$$

Actually, (2.13) comes from (3.15). Notice that when ν is half an odd integer, owing to the singularities of the gamma function, the series does in fact turn into a polynomial. Taking this into account, one arrives at

$$S(2p,a,m) = \sum_{n=1}^{\infty} \frac{1}{2^{p+1/2}n^{2p+1}} \sqrt{\frac{\pi}{2}}\, e^{-amn} \times \sum_{k=0}^{p} \frac{(p+k)!}{2^k k!(p-k)!} (amn)^{p-k}. \quad (3.16)$$

The finiteness of the second sum in (3.16) makes it possible to naively interchange the summatories and obtain

$$S(2p,a,m) = \frac{\sqrt{\pi}}{2^{p+1}} \sum_{k=0}^{p} \frac{(p+k)!}{2^k k!(p-k)!} \times (am)^{p-k} \mathrm{Li}_{p+k+1}(e^{-am}), \quad (3.17)$$

where $\mathrm{Li}_N(x)$ is the polylogarithm function

$$\mathrm{Li}_{N(x)} = \sum_{n=1}^{\infty} \frac{1}{n^N} x^n. \quad (3.18)$$

Zeta regularization of (3.18) (see Appendix B) allows us to write an am expansion for the physically relevant vacuum energy:

$$\begin{aligned}\epsilon(2p,a,m) &= -2\frac{L^{2p-1}}{a^{2p}} \pi^{-(p+1/2)} S(2p,a,m) \\ &= -\frac{L^{2p-1}}{2^p a^{2p}} \pi^{-p} \sum_{k=0}^{p} \frac{(p+k)!}{2^k k!(p-k)!} \\ &\quad \times \Bigg\{ \sum_{\substack{n=0 \\ n\neq p+k}}^{\infty} \frac{(-1)^n}{n!} \zeta(p+k+1-n)(am)^{p-k+n} \\ &\quad + \frac{(-1)^{p+k+1}}{(p+k)!} \times [\ln am - (\psi(p+k)+\gamma)](am)^{2p} \Bigg\}. \quad (3.19)\end{aligned}$$

However, it is important to notice that expansion (3.19) is also valid for any value of am, as in fact should be expected of a negative exponential times a polynomial. There is a simple way of checking this property, i.e., to find $\epsilon(2p,a,0)$ from expression (3.19). When setting $m = 0$, only the terms with $k = p$ and $n = 0$ survive and the remaining value is

$$\epsilon(2p,a,0) = -\frac{L^{2p-1}}{a^{2p}} \pi^{-(2p+1)/2} \Gamma(p+\tfrac{1}{2}) \zeta(2p+1), \quad (3.20)$$

in agreement with (2.11).

C. Numerical results

Application of (3.12) and (3.19) for the space dimensions 1, 2, and 3 gives

$$\begin{aligned}\epsilon(1,a,m) &= -\frac{2}{a\pi} \Bigg\{ \frac{\pi^2}{12} - \frac{\pi}{16} am - \frac{1}{8}(am)^2\left(1 + \ln\left(\frac{am}{4\pi}\right)\right) \\ &\quad - \frac{\zeta(3)}{128\pi^2}(am)^4 + \frac{\zeta(5)}{1024\pi^4}(am^6) + O((am)^8) \Bigg\}, \\ \epsilon(2,a,m) &= -(L/2\pi a^2)\{\zeta(3) + \tfrac{1}{2}(am)^2(1 + \ln am) \\ &\quad - \tfrac{1}{6}(am)^3 + \tfrac{1}{96}(am)^4 - (1/17280)(am)^6 \\ &\quad + O((am)^8)\}, \\ \epsilon(3,a,m) &= -\frac{2L^2}{\pi^2 a^3} \Bigg\{ \frac{\pi^4}{180} - \frac{\pi^2}{48}(am)^2 + \frac{\pi}{32}(am)^3 \\ &\quad + \frac{1}{64}(am)^4\left(\frac{3}{2} + \ln\left(\frac{am}{4\pi}\right)\right) \\ &\quad + \frac{\zeta(3)}{1536\pi^2}(am)^6 + O((am)^8) \Bigg\}, \quad (3.21)\end{aligned}$$

where the first neglected coefficients in the curly braces, i.e., those for the terms of eighth order in am, have been evaluated, with the results -1.60×10^{-7}, 6.89×10^{-7}, and -6.50×10^{-7}, respectively. The rates between the eighth-order terms and those of order six turn out to be, in absolute values, 0.015, 0.012, and 0.008. Thus up to these orders of

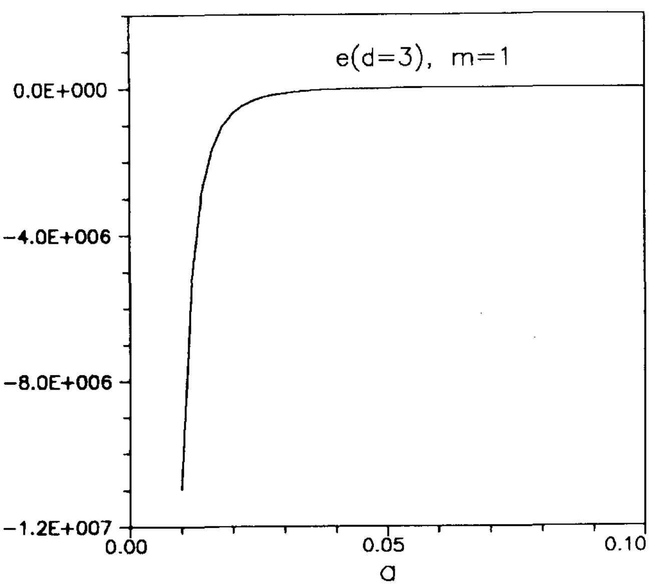

FIG. 1. The Casimir energy density (e) for a unit mass scalar field in $\mathbf{S}^1 \times \mathbf{R}^2$ as a function of the length a.

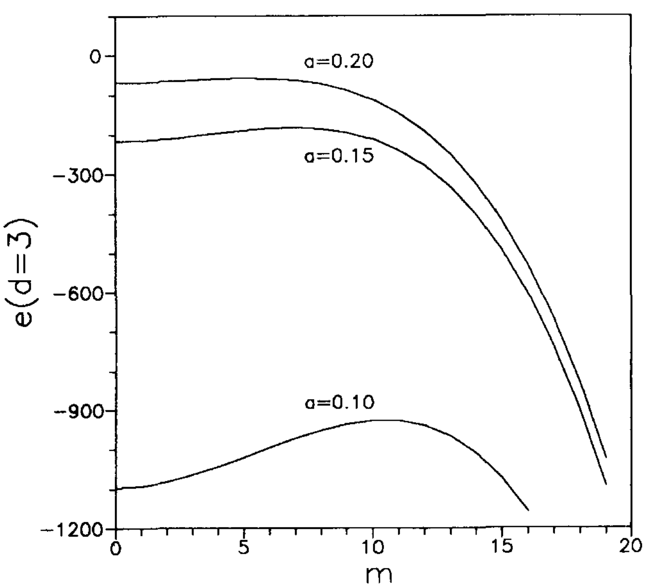

FIG. 2. The Casimir energy per volume unit for a scalar field compactified in a circle of fixed length a versus the field mass. The curves correspond to three different values of a.

accuracy, the truncated series (3.21) provide a reliable approximation for small values of am. In particular, the third equation in (3.21) supplies a way of computing the Casimir energy per volume unit $e(3,a,m) \equiv \epsilon(3,a,m)/L^2 a$ in a three-dimensional space ($\mathbf{S}^1 \times \mathbf{R}^2$ in this case) for a relatively wide range of values of a and m. Some of these results are shown in Figs. 1 and 2.

ACKNOWLEDGMENTS

We have greatly enjoyed discussing this subject with Professor A. Actor and feel very thankful to him for all his encouragement.

This work has been partially supported by Comisión Asesora de Investigación Científica y Técnica (CAICYT) Contract No. AE 87-0016-3. One of us, AR, has also obtained financial help through an FPPI grant from Ministerio de Educación y Ciencia (MEC, Spain).

APPENDIX A: MODE-SUM REGULARIZATION

The Casimir energy of a massive scalar field is very closely related to the formal series

$$\sum_{n=1}^{\infty} \left(n^2 + \left(\frac{am}{2\pi}\right)^2\right)^{d/2}. \tag{A1}$$

Equation (A1) can be regarded as a particular value of the "Epstein–Hurwitz" zeta function:

$$Z(s;a) \equiv \sum_{n=1}^{\infty} (n^2 + a^2)^{-s}, \quad \operatorname{Re} s > \frac{1}{2}, \tag{A2}$$

which admits, however, an analytic continuation to $\operatorname{Re} s < \frac{1}{2}$. In the genuine Hurwitz function, we would have n instead of n^2, while for a true Epstein one, no n-independent term should appear.

The analytic continuation can be done in the following way. One writes

$$\begin{aligned} Z(s;a) &= \sum_{n=1}^{\infty} \frac{1}{\Gamma(s)} \int_0^\infty dt\, t^{s-1} e^{-t(n^2+a^2)} \\ &= \frac{1}{\Gamma(s)} \int_0^\infty dt\, t^{s-1} e^{-ta^2} S_2(t), \end{aligned} \tag{A3}$$

where S_2 is the analytic function

$$S_2(t) \equiv \sum_{n=1}^{\infty} e^{-n^2 t} \tag{A4}$$

having the useful property[4]

$$S_2(t) = -\tfrac{1}{2} + \tfrac{1}{2}\sqrt{\pi/t} + \sqrt{\pi/t}\, S_2(\pi^2/t). \tag{A5}$$

By putting identity (A5) into the integral in (A3) and after doing two simple Gaussian integrations, we obtain

$$\begin{aligned} Z(s;a) = {}& -\frac{1}{2}(a^2)^{-s} + \frac{\sqrt{\pi}}{2\Gamma(s)} \Gamma\left(s - \frac{1}{2}\right)(a^2)^{-s-1/2} \\ & + \frac{\sqrt{\pi}}{\Gamma(s)} \sum_{n=1}^{\infty} \int_0^\infty dt\, t^{s-3/2} e^{-\pi^2 n^2/t - ta^2}. \end{aligned} \tag{A6}$$

Using

$$\int_0^\infty dx\, x^{\nu-1} e^{-a/x - bx} = 2\left(\frac{a}{b}\right)^{\nu/2} K_\nu(2\sqrt{ab}), \quad a,b>0, \tag{A7}$$

where K_ν is the modified Bessel function, $Z(s;a)$ can be written as

$$\begin{aligned} Z(s;a) &= -\frac{1}{2} a^{-2s} + \frac{\sqrt{\pi}}{2a^{2s-1}\Gamma(s)} \\ &\quad \times \left[\Gamma\left(s - \frac{1}{2}\right) + 4 \sum_{n=1}^{\infty} (\pi n a)^{s-1/2} K_{s-1/2}(2\pi n a)\right], \end{aligned} \tag{A8}$$

which does in fact provide the analytic continuation of (A2). For $s = -d/2$, $a = am/2\pi$, formula (A8) allows us to obtain (2.8) from (2.7).

APPENDIX B: DOUBLE-SUM CALCULATIONS

Here we shall evaluate some cases of double series. Using a notation similar to that in Ref. 5, let us consider the general types

$$\begin{aligned} S_B^{(\alpha)}[f,s] &\equiv \sum_{m=1}^{\infty} \frac{1}{m^{s+1}} \sum_{a=0}^{\infty} m^{\alpha a} f(a), \\ S_{BL}^{(\alpha)}[f,s] &\equiv \sum_{m=1}^{\infty} \frac{\ln m}{m^{s+1}} \sum_{a=0}^{\infty} m^{\alpha a} f(a), \end{aligned} \tag{B1}$$

where f is supposed to satisfy (i) $f(a) \geqslant 0$ for $a \in \mathbf{N}$ and regular for $\operatorname{Re} a \geqslant 0$ and (ii) $a m^a f(a) e^{-\pi|\operatorname{Im} a|} \to 0$ as $|a| \to \infty$ for $\operatorname{Re} a \geqslant 0$ and fixed m. By applying a method analogous to Weldon's,[5] but taking into account possible contour corrections,[6] we arrive at

$$S_B^{(\alpha)}[f,s] = \begin{cases} \displaystyle\sum_{k=0}^{\infty} \zeta(s+1-\alpha a) f(a) - f\left(\frac{s}{\alpha}\right) \frac{\pi}{\alpha} \cot \frac{\pi s}{\alpha} - \Delta_B^{(\alpha)}[f,s], & \dfrac{s}{\alpha} \notin \mathbf{N}, \\ \displaystyle\sum_{\substack{k=0 \\ a \neq s/\alpha}}^{\infty} \zeta(s+1-\alpha a) f(a) + \gamma f\left(\frac{s}{\alpha}\right) - \frac{1}{\alpha} f'\left(\frac{s}{\alpha}\right) - \Delta_B^{(\alpha)}[f,s], & \dfrac{s}{\alpha} \in \mathbf{N}. \end{cases} \tag{B2}$$

The second terms on the rhs of (B2) are referred to as *Weldon's* and $\Delta_B^{(\alpha)}[f,s]$ is the correction term coming from integration over the infinite semicircle on the right side of the complex plane (see Ref. 6), which is given by

$$\Delta_B^{(\alpha)}[f,s] \equiv \frac{1}{2\pi i} \int_{\supset} da\, \zeta(s+1-\alpha a) f(a) \pi \cot \pi a. \tag{B3}$$

We will be concerned with $\alpha = 2$ and functions of the sort

$$f(a) = b^a / a!(a+p)!, \tag{B4}$$

where p, b have fixed values. In this case, the asymptotic behavior of the integrand for $|a| \gg 1$ is, for α and s real,

$$\zeta(s+1-\alpha a)\frac{b^a}{a!(a+p)!}\pi\cot\pi a$$
$$\sim 2(2\pi)^{s-1/2}\alpha^{-s-1/2}\exp\left\{\left(2-\alpha+\ln b\left(\frac{\alpha}{2\pi}\right)^{\alpha}\right)a+\left((\alpha-2)a-2p-s-\frac{3}{2}\right)\ln a\right\}$$
$$\times\begin{cases}-\frac{\pi i}{2}\operatorname{sgn}(\operatorname{Im}a)\exp\left\{\frac{\pi}{2}|\operatorname{Im}(2a-s)|+i\frac{\pi}{2}\operatorname{Re}(2a-s)\right\}, & \text{for } \operatorname{Im}a\neq 0,\\ \pi\cot\pi a\cos\frac{\pi}{2}(2a-s), & \text{for } \operatorname{Im}a=0.\end{cases} \quad \text{(B5)}$$

If $\operatorname{Im} a = 0$, $\cot \pi a$ may oscillate very quickly between $-\infty$ and $+\infty$; however, it happens that the curved contour contains only one real point, which has zero measure, so its contribution can be neglected. The type of asymptotic behavior of this integrand depends on α, with a critical value precisely at $\alpha = 2$. Thus for $\alpha = 2$ and given a fixed range for the parameter s, we would like to know under which conditions it is possible to ensure the vanishing of the correction term. Setting $\alpha = 2$ for $\operatorname{Im} a \neq 0$ and for $|a| \to \infty$, $\operatorname{Re} a \geqslant 0$, the two dominating terms in the exponential are

$$[\ln(b/\pi^2)]a + \pi|\operatorname{Im} a|. \quad \text{(B6)}$$

It is not difficult to see that a sufficient condition for the asymptotic vanishing of (B5) is that obtained by imposing

$$\ln(b/\pi^2) + \pi < 0, \quad \text{(B7)}$$

i.e.,

$$b < b_c, \quad b_c = \pi^2 e^{-\pi} = 0.4265. \quad \text{(B8)}$$

Therefore, b_c is a bound below which there is no correction term: Above this value, there may be a small correction term; it will certainly exist for much larger values of b. For the sake of simplicity, we will assume that our b satisfies (B8): This will make it possible that

$$\Delta_B^{(\alpha)}\left[f(a)=\frac{b^a}{a!(a+p)!},s\right]=0. \quad \text{(B9)}$$

It is quite interesting to observe that for the series appearing in the expansion of the Casimir energy, b is none other than $am/2$, where a is the compactification length and m is the mass of the field [see (3.7) and the resulting double series (3.10)]. The existence of b_c, or any more accurate upper bound for the absence of correction terms, should not seriously worry us because it was already known that expansion (3.7) in powers of $am/2$, which gave rise to the double series (3.10), was valid only for small values of am [and so were the series in (3.10)]. Arbitrarily large values of am would not make sense anyway. Thus from now on we will assume that the value of b is small enough to forget about correction terms.

Since we are interested in the value $s = -1$, we have $s/\alpha = -\frac{1}{2} \notin \mathbb{N}$ and therefore, (B2) gives

$$S_B^{(2)}\left[f(a)=\frac{b^a}{a!(a+p)!},s=1\right]$$
$$=\sum_{a=0}^{\infty}\zeta(2a)\frac{b^a}{a!(a+p)!}$$
$$-\frac{b^{-1/2}}{(-1/2)!(p-1/2)!}\frac{\pi}{2}\cot\left(-\frac{\pi}{2}\right). \quad \text{(B10)}$$

As pointed out in Ref. 5 (Appendix A), the vanishing of the second term caused by the occurrence of $\cot(-\pi/2)$ would make a naive summation interchange valid for particular cases such as this one, although that cannot be acceptable in general. Furthermore, since $\zeta(-2k) = 0$ for k positive integer and $\zeta(0) = -\frac{1}{2}$, the result (B10) reduces to

$$S_B^{(2)}\left[f(a)=\frac{b^a}{a!(a+p)!},s=-1\right]=-\frac{1}{2p!}, \quad \text{(B11)}$$

which gives the value for the second sum in (3.10).

Another case of interest is the one given by

$$f(a) = [b^a/a!(a+p)!]C(p,a),$$
$$C(p,a) \equiv \psi(a+1) + \psi(a+p+1), \quad \text{(B12)}$$
$$\psi(z) \equiv \frac{1}{\Gamma(z)}\frac{d}{dz}\Gamma(z),$$

with $\alpha = 2$ and $s = -1$, as before. Actually, because of the decrease of $C(p,a)$ as $|a| \to \infty$, there is no correction from the semicircumference either. Retracing these steps, one can quickly arrive at

$$S_B^{(2)}\left[f(a)=\frac{b^a}{a!(a+p)!}C(p,a),s=-1\right]$$
$$=-\frac{1}{2p!}C(p,0), \quad \text{(B13)}$$

which is the result for the third series in (3.10).

Now let us turn to $S_{BL}^{(\alpha)}[f,s]$: This double series can be easily dealt with by noticing that since $\ln m/m \to 0$ when $m \to \infty$, as far as the m summation is concerned, its behavior is of the same sort as that of $S_{BL}^{(\alpha)}[f,s]$, provided that s can arbitrarily vary before making the analytic continuation to $s = -1$. In fact, it would be enough to think of s as being one unit larger than in the previous cases. Taking this into consideration, it is correct to put

$$S_{BL}^{(\alpha)}[f,s] = -\frac{d}{ds}S_B^{(\alpha)}[f,s]. \quad \text{(B14)}$$

Let us again take the case where $f(a)$ is given by (B4) and $\alpha = 2$. Since b is supposed to be small enough, the same reasoning concerning the vanishing of the correction term applies. Bearing in mind that for $s = -1$, $\alpha = 2$, and $s/\alpha \notin \mathbb{N}$, one can evaluate the derivative of (B2) with respect to s for the f in (B4) and set $s = -1$. The outcome is

$$S_{BL}^{(2)}\left[f(a)=\frac{b^a}{a!(a+p)!},s=-1\right]$$
$$=-\frac{1}{2p!}+\sum_{a=1}^{\infty}(-1)^{a+1}\frac{(2a!)}{2(2\pi)^{2a}a!(a+p)!}$$
$$\times\zeta(2a+1)b^a-\frac{\pi^2}{4}\frac{1}{\sqrt{\pi}(p-\frac{1}{2})!}b^{-1/2}. \quad \text{(B15)}$$

Equation (B15) is the result for the first series in (3.10). To obtain (B15), use has been made of the identities

$$\zeta'(0) = -\tfrac{1}{2}\ln 2\pi,$$
$$\zeta'(-2k) = \tfrac{1}{2}(-1)^k (2\pi)^{-2k}\Gamma(2k+1)\zeta(2k+1),$$
$$k = 1,2,3,\ldots, \tag{B16}$$

where the second can be easily found through derivation of the zeta-function reflection formula.

Finally, we are also concerned with sums of the sort

$$S_{AB}^{(\alpha=1)}[f,s] = \sum_{m=1}^{\infty} \frac{1}{m^{s+1}} \sum_{a=0}^{\infty} (-1)^a m^a f(a). \tag{B17}$$

When $\alpha = 1$, Weldon's complex integration method gives a totally correct result[6]:

$$S_{AB}^{(\alpha=1)}[f,s] = \begin{cases} \displaystyle\sum_{a=0}^{\infty} \zeta(s+1-a)(-1)^a f(a) - f(s)\pi\csc(\pi s), & \text{for } s\notin\mathbf{N}, \\ \displaystyle\sum_{\substack{a=0\\ a\neq s}}^{\infty} \zeta(s+1-a)(-1)^a f(a) + (-1)^s(\gamma f(s) - f'(s)), & \text{for } s\in\mathbf{N}. \end{cases} \tag{B18}$$

The particular case we will be looking at is the one corresponding to

$$\mathrm{Li}_N(e^{-x}) = \sum_{n=1}^{\infty} \frac{1}{n^N} e^{-nx}, \quad N \text{ positive integer.} \tag{B19}$$

By expanding the exponential, it is clear that $f(a) = a^k/k!$, which has the appropriate behavior. Since we have $s = N-1\in\mathbf{N}$, we must take the second possibility in (B18), which gives

$$\mathrm{Li}_N(e^{-x}) = \sum_{\substack{a=0\\ a\neq N-1}}^{\infty} \zeta(N-a)(-1)^a \frac{x^a}{a!}$$
$$+ (-1)^{N-1}\frac{x^{N-1}}{\Gamma(N)}(\gamma - \ln x + \psi(N)). \tag{B20}$$

This allows us to obtain (3.19).

[1]G. Plunien, B. Müller, and W. Greiner, Phys. Rep. **134**, 87 (1986).
[2]J. Ambjørn and S. Wolfram, Ann. Phys. (NY) **147**, 1 (1983).
[3]A. Actor, Nucl. Phys. B **265**, (**FS 16**), 689 (1986).
[4]A. Erdélyi, W. Magnus, F. Oberhettinger, and F. G. Tricomi, *Higher Transcendental Functions* (McGraw-Hill, New York, 1953).
[5]H. A. Weldon, Nucl. Phys. B **270**, (**FS 16**), 79 (1986).
[6]E. Elizalde and A. Romeo, *Rigorous Extension of the Proof of Zeta Function Regularization*, Univ. of Barcelona preprint, UB-ECM-FP-14/88.

R structures, Yang–Baxter equations, and related involution theorems

Walter Oevel
Universitaet Paderborn, D4790 Paderborn, Federal Republic of Germany

(Received 22 September 1987; accepted for publication 21 December 1988)

There are several ways to construct functions in involution on a Lie bi-algebra, a Lie algebra equipped with a second Lie bracket. For the solvable systems associated to the Casimir functions a second Hamiltonian formulation can be constructed and a class of bi-Hamiltonian Korteweg–de Vries-like evolution equations with explicit space dependence is derived. Translating the Casimir functions with the flow of a special vector field yields another set of functions in involution. Lenard relations are found for the corresponding Hamiltonian systems. Finally, solutions of the classical Yang–Baxter equation lead to an analog of compatible Hamiltonian pairs. The invariants of the resulting hereditary operators are in involution.

I. INTRODUCTION

Many of the integrable systems discovered in the last two decades have a Lie algebra background giving direct access to most of the striking features found for these systems. One of the most essential tools for solving these equations is given by their Lax representation. One may think of the Lax equation as the abstract dynamical system from which the "physical" dynamical systems are obtained by introducing suitable charts. Hence the phase space for most of these equations can be regarded as given by the set of Lax operators taking values in a Lie algebra or the Hilbert space of currents over a Lie algebra. In this sense in many cases (e.g., the Toda lattice) the phase space itself is a Lie algebra; other examples (e.g., the Calogero–Moser system) are obtained by reduction techniques applied to certain Lie algebras. A review of the relevant involution theorems is found in Ref. 1. Also, infinite-dimensional equations such as the celebrated Korteweg–de Vries (KdV) equation can be regarded as systems on an infinite-dimensional Lie algebra of pseudodifferential operators.[2]

It has turned out that the factorization method and the group of dressing transformations for these integrable systems can be understood in terms of Poisson–Drinfeld groups acting on the phase space.[3] In this setup, dressing transformations are Poisson maps relative to the Poisson structure on the product of the group with this phase space. The Poisson structure on the group corresponds to a Lie bracket on the dual of the associated Lie algebra or a second Lie bracket on the Lie algebra, respectively. It is engendered by a linear map called the "*R* matrix," which leads to a direct construction of integrable systems using the Casimir functions on the dual algebra. In special situations the *R* matrix also gives a direct construction of the second Hamiltonian formulation for these equations.[4]

In this paper we will investigate how functions in involution can be constructed from a given *R* matrix. After giving the definitions in Sec. II we will review the Adler–Kostant–Symes scheme[2,1] in terms of *R* structures in Sec. III. It is shown how bi-Hamiltonian systems involving two arbitrary functions can be constructed systematically by applying these considerations to the Lie algebra of pseudodifferential operators as in Ref. 2 and a class of KdV-like equations with explicit space dependence is obtained. In Sec. IV we will give a generalization of an involution theorem by Mishenko and Fomenko.[5] Special *R* matrices provide examples for this theorem: Translations of Casimir functions by the flow of a special vector field related to the *R* structure yield functions in involution. In Sec. V we show that a Lie Poisson structure and *R* matrices given by solutions of the classical Yang–Baxter equation[6–8] form an analog of compatible Hamiltonian pairs in the sense of Refs. 9–11. The spectrum of the resulting hereditary operator is in involution relative to a hierarchy of Poisson brackets.

The basic reference for the first sections is the article by Semenov-Tian-Shansky[4]; further references to *R* matrices and the structure of Poisson–Drinfeld groups are found in Ref. 12.

II. DEFINITIONS AND BASIC PROPERTIES

Definition 1: An *R* structure is a Lie algebra g equipped with a linear map $R:g\to g$ (called the *R* matrix) such that the bracket $[a,b]_R := [Ra,b] + [a,Rb]$ is a second Lie product on g.

When the Lie algebra can be identified with its dual via a nondegenerate, symmetric invariant metric

$$\langle L_1,L_2\rangle = \langle L_2,L_1\rangle,\quad \langle L_1,[L_2,L_3]\rangle = \langle L_2,[L_3,L_1]\rangle, \tag{2.1}$$

we sometimes will assume the "unitarity" condition

$$\langle RL_1,L_2\rangle + \langle L_1,RL_2\rangle = 0 \tag{2.2}$$

for the *R* matrix, i.e., *R* is assumed to be skew symmetric relative to the pairing (2.1). One easily checks that the Jacobi identity for $[\ ,\ ]_R$ can be rewritten as the Jacobi identity of the expression $[[Ra,Rb] - R[a,b]_R,c]$. Hence claiming the first entry to be just a scalar multiple of the original Lie bracket $[\ ,\ ]$ is a sufficient condition for R to be an *R* matrix, i.e.,

$$[Ra,Rb] - R\,[a,b]_R = -\alpha[a,b], \tag{2.3}$$

where α is a scalar parameter. Following Ref. 4 we will refer to (2.3) as the Yang–Baxter equation, or YB(α) for short. The correspondence of this terminology to the classical tensor notation is explained in Ref. 4. In reference to the depen-

 0022-2488/89/051140-10$02.50

dence of the parameter α, obviously only the two cases $\alpha = 0$ and $\alpha = 1$ have to be considered since for any $\alpha \neq 0$ the dilation $R \to (1/\sqrt{\alpha})R$ maps the solution of YB(α) to solutions of YB(1). The case $\alpha = 1$ is also called the modified Yang–Baxter equation.

Observation 1: Any solution R of YB(α) for the Lie bracket $[\ ,\ ]$ also solves YB(α) for the Lie bracket $[\ ,\ ]_R$.

As a result of this observation one can iterate the construction of a "modified" Lie bracket by starting with $[\ ,\ ]_R$ instead of $[\ ,\ ]$, thus obtaining a third Lie bracket $[a,b]_{RR} = [Ra,b]_R + [a,Rb]_R$, etc. Hence an R structure equips a Lie algebra with a hierarchy of Lie brackets, where the nth iterated bracket is given by

$$[a,b]_{RR\cdots R} = \frac{d^n}{dt^n}\bigg|_{t=0} [e^{tR}a, e^{tR}b] = \sum_{k=0}^{n} \binom{n}{k} [R^k a, R^{n-k} b]. \tag{2.4}$$

Each bracket induces a Lie Poisson structure on the dual g^* given by

$$\{f_1,f_2\}(L) = \langle L,[df_1,df_2]\rangle. \tag{2.5}$$

Here the f_i are scalar-valued functions on g^* and their differentials are interpreted as elements of g (rather than of the bidual g^{**}). Here L is chosen for a point g^* since in the applications L will be the Lax operator for the considered integrable systems.

There is a special class of solutions to YB(1) that arises in a very simple manner: Assume that the Lie algebra can be split into a direct sum of subalgebras, i.e., $g = g_+ \oplus g_-$. Denoting the projections onto the subalgebras by P_+ and P_-, it is easily verified that

$$R = P_+ - P_- \tag{2.6}$$

solves YB(1) and hence defines an R structure on g. In this case the hierarchy of Lie brackets generated by R reduces to just three different brackets:

$$[a,b],[a,b]_R = 2[a_+,b_+] - 2[a_-,b_-],$$
$$[a,b]_{RR} = 4[a_+,b_+] + 4[a_-,b_-],$$
$$[a,b]_{RRR} = 4[a,b]_R, \ldots,$$

where the subscripts denote the projections to the corresponding subalgebras.

For YB(0) there does not seem to be such a natural class of solutions. Assuming invertibility, though, YB(0) can be rewritten as

$$\begin{aligned} 0 &= [Ra,Rb] - R[a,b]_R \\ &= R(R^{-1}[\tilde a,\tilde b] - [R^{-1}\tilde a,\tilde b] \\ &\quad - [\tilde a,R^{-1}\tilde b]), \quad \tilde a = Ra, \tilde b = Rb; \end{aligned} \tag{2.7}$$

hence solutions of YB(0) model the algebraic properties of the inverse of a derivation on g. Hence in the presence of an invariant metric, a unitary solution of YB(0) models the inverse of a symplectic two-cocycle

$$B(L_1,L_2) = \langle R^{-1}L_1,L_2\rangle, \quad B(L_1,[L_2,L_3]) + \text{cycl} = 0. \tag{2.8}$$

For both cases ($\alpha = 0,1$) a systematic scheme of solving the YB equations is given in Ref. 6, even for the more general case when R depends on some additional parameter in a specific way.

III. INVOLUTION OF RESTRICTED CASIMIRS

The following results hold for arbitrary R structures, i.e., not only for those arising from solutions of YB equations. Let R be an R matrix on a Lie algebra $(g,[\ ,\ ])$. By g^*, g^*_R we denote the dual of g endowed with the Lie Poisson structures arising from $[\ ,\ ]$ and $[\ ,\ ]_R$, respectively.

Theorem 1 (Ref. 4): (i) Casimirs of g^* are in involution on g^*_R. (ii) Let C be a Casimir function of g^*. The associated g^*_R- Hamiltonian vector field has orbits in the symplectic leaves of g^*. The form of this Hamiltonian system is

$$\frac{d}{dt} L = \mathrm{ad}^*_{R\,dC} L, \quad L \in g^*, \tag{3.1}$$

where ad* is the coadjoint representation of g (w.r.t $[\ ,\ ]$). If g and g^* can be identified by a $[\ ,\ ]$-invariant metric, then (3.1) becomes a Lax equation:

$$\frac{d}{dt} L = [L,R\,dC], \quad L \in g. \tag{3.2}$$

Proof: It is most convenient to think about the Poisson bracket in terms of the Poisson tensor P defined by $\{f_1,f_2\} = \langle df_2,P\,df_1\rangle$, i.e., P can be regarded as a skew-symmetric linear map from the covector fields to the vector fields. In the Lie Poisson case this tensor is given by $P(L): g \to g^*$, $P(L)\gamma = \mathrm{ad}^*_\gamma L$, $\gamma \in g$. The Poisson tensors arising from the Lie brackets $[\ ,\ ]$, and $[\ ,\ ]_R$ are related by $P_R = R^*P + PR$, where R here has to be understood as the pointwise lift of the map R on g to the vector fields over g, where R^* is the transpose of this map. Casimirs on a Poisson manifold are those functions Poisson commuting with all functions on the manifold; their differentials lie in the kernel of the Poisson tensor.[13] Thus the Hamiltonian vector field $X_C = P_R\,dC$ of such a function has the form $X_C = PR\,dC$ and hence takes values in the image of the Poisson tensor P which spans the tangent spaces of the symplectic leaves in g^*.

We want to look at the special case of an R structure (2.6) given by a splitting into two subalgebras. With $g = g_+ \oplus g_-$, $R = P_+ - P_-$, $L \in g^*$, $f_1,f_2 \in C^\infty(g^*)$ one calculates the Lie Poisson bracket arising from $[\ ,\ ]_R$

$$\begin{aligned} \{f_1,f_2\}_R(L) &= 2\langle L,[(df_1)_+,(df_2)_+]\rangle \\ &\quad - 2\langle L,[(df_1)_-,(df_2)_-]\rangle. \end{aligned} \tag{3.3}$$

The projections $P_\pm\,df$ of a differential coincide with the differential of the restrictions $f|_{g^*_\pm}$. Hence restricting L to $g^*_\pm$ (identified with the null spaces $g^0_\mp = \{L \in g^*, \langle L,\tilde L_\mp\rangle = 0 \text{ for all } \tilde L_\mp \in g_\mp\}$) one recovers the usual Lie Poisson structures $\{\ ,\ \}_{g^*_\pm}$ on the duals of the subalgebras:

$$\{f_1,f_2\}_R(L) = \pm 2\langle L,[(df_1)_\pm,(df_2)_\pm]\rangle$$
$$= \pm 2\{f_1|_{g^*_\pm},f_2|_{g^*_\pm}\}_{g^*_\pm}(L), \quad L\in g^*_\pm. \tag{3.4}$$

Hence functions in involution on g^* can be restricted to functions in involution on $g^*_\pm$. For Casimir functions C on g^* (characterized by $[L,dC]=0$) the explicit form of the Hamiltonian systems $(d/dt)L = X_C = [L,R\,dC]$ is given by

$$\frac{d}{dt}L_+ = 2[L_+,dC_+] + 2[L_-,dC_+]_+$$
$$= -2[L_+,dC_-]_+,$$
$$\frac{d}{dt}L_- = -2[L_+,dC_-]_- - 2[L_-,dC_-]$$
$$= 2[L_-,dC_+]_-, \tag{3.5}$$

where we now have assumed an invariant metric on g. These dynamical systems obviously leave the subspaces $g_\pm$ invariant and coincide with the Lie Poisson Hamiltonian systems of the restricted Casimirs on these subspaces.

Thus it turns out that the Hamiltonian systems described by Theorem 1 are (up to irrelevant constants) those described by the Adler–Kostant–Symes scheme.[1,2] Thinking about these systems in terms of R structures at this stage gives the additional information that these equations on g_+ (or g_-) are restrictions of larger Hamiltonian systems on g.

In special situations the R structure gives a direct approach to the second Hamiltonian formulation encountered for many of the integrable equations as follows.

Theorem 2 (Ref. 4): Assume g to be an algebra identified with g^* via a nondegenerate, symmetric "trace form" $\mathrm{Tr}: g\to\mathbb{R}$, i.e., $\langle L,\tilde L\rangle = \mathrm{Tr}(L\tilde L) = \mathrm{Tr}(L\tilde L), L,\tilde L\in g$. Let R be an antisymmetric solution of YB(α) relative to the Lie bracket $[L,\tilde L] = L\tilde L - \tilde L L$. Then for $f_1,f_2\in C^\infty(g)$ we define

$$\{f_1,f_2\}_2(L) = \langle L\,df_1,RL\,df_2\rangle - \langle df_1\,L,R\,df_2\,L\rangle. \tag{3.6}$$

(i) This bracket defines a Poisson structure.

(ii) This bracket is compatible with the Lie Poisson structure

$$\{f_1,f_2\}_R(L) = \langle L,[df_1,df_2]_R\rangle \tag{3.7}$$

in the sense of Refs. 9–11, i.e., the sum of these brackets is again a Poisson structure.

(iii) Casimir functions of $(g,[\ ,\ ])$ are in involution w.r.t. (3.6).

Proof: Statement (i) is shown by a lengthy computation. Then (ii) is obtained easily by substituting $L\to L+\epsilon 1$ into (3.6). As the differentials of $[\ ,\ ]$ Casimirs commute with L, (iii) follows trivially.

The situation described by Theorem 2 certainly seems rather special: Nevertheless, in many applications the underlying Lie algebra g is indeed a matrix or operator algebra such that this theorem can be applied to that situation. The most restrictive point is claiming R to be antisymmetric w.r.t. the trace form. For the splitting case $g = g_+\oplus g_-, R = P_+ - P_-$ this is granted if $g_- = g^*_+$, i.e.,

$$\mathrm{Tr}(L\tilde L) = \mathrm{Tr}(L_+\tilde L_-) + \mathrm{Tr}(L_-\tilde L_+). \tag{3.8}$$

A natural example of such a structure is given by

$$g = g^0\otimes\mathbb{C}[\lambda,\lambda^{-1}] = \Big\{\sum_{k\in\mathbb{Z}} l_k\lambda^k; l_k\in g^0\Big\},$$
$$g_- = \Big\{\sum_{k<0} l_k\lambda^k\Big\}, \quad g_+ = \Big\{\sum_{k\geq 0} l_k\lambda^k\Big\}, \tag{3.9}$$

where g^0 is some algebra with a trace form tr that is lifted to

$$\mathrm{Tr}(L) = \mathrm{Res}_{\lambda=0}\,\mathrm{tr}\Big(\sum_{k\in\mathbb{Z}} l_k\lambda^k\Big) = \mathrm{tr}(l_{-1}). \tag{3.10}$$

Note that in this situation the quadratic Poisson bracket (3.6) can also be projected to the subalgebras. One finds

$$\{f_1,f_2\}_2(L) = \langle L(df_1)_\mp,RL(df_2)_\mp\rangle$$
$$- \langle (df_1)_\mp L,R(df_2)_\mp L\rangle, \quad L\in g_\pm. \tag{3.11}$$

Obviously, restrictions of functions in involution on g yield functions in involution on the subalgebras.

Although the Casimir functions on g are in involution w.r.t. both brackets (3.6) and (3.7), the associated Hamiltonian systems (3.2) are not necessarily bi-Hamiltonian w.r.t. both Poisson structures in the sense that these equations also leave the bracket (3.6) invariant. However, in most applications a natural set of Casimirs is given by the traces of some powers of the Lax operator, i.e., $C_q := \mathrm{tr}(L^q)$, $dC_q = qL^{q-1}$ and hence

$$\frac{d}{dt}L = [L,R\,dC_q] = \frac{q}{q-1}[L,RL\,dC_{q-1}]. \tag{3.12}$$

Note that for Casimir functions the Hamiltonian system associated with (3.6) assumes the form

$$\frac{d}{dt}L = [L,RL\,dC], \tag{3.13}$$

so the system (3.12) is not only Hamiltonian relative to (3.7) (with the Hamiltonian function C_q), but also relative to (3.6) (with the Hamiltonian function $[q/(q-1)]C_{q-1}$). In the following application we will exploit this to derive the bi-Hamiltonian structure (and hence the recursion operator) for a class of KdV-like equations with explicit space dependence.

Example 1: Following Ref. 2 we will consider the ring of formal pseudodifferential operators as the formal Laurent series in the variable ξ, with coefficients in a ring S of functions of one variable x, i.e.,

$$g = \bigcup_m \Big\{\sum_{-\infty}^{m} d_k(x)\xi^k \,|\, d_k\in \mathrm{S}\Big\}$$

endowed with the multiplication

$$L\tilde L = \sum_{\alpha\geq 0}\frac{1}{\alpha!}\Big[\Big(\frac{\partial}{\partial\xi}\Big)^\alpha L\Big]\Big[\Big(\frac{\partial}{\partial x}\Big)^\alpha \tilde L\Big],$$
$$L = \sum_{-\infty}^{m} d_k(x)\xi^k, \quad \tilde L = \sum_{-\infty}^{\tilde m}\tilde d_k(x)\xi^k\in g. \tag{3.14}$$

Indeed, with $D = \partial/\partial x, I = (\partial/\partial x)^{-1}, f_x = \partial f/\partial x$ the multiplication (3.14) models the properties of the algebra of pseudodifferential operators:

$$g=\bigcup_m\left\{\sum_{k=0}^{m}a_k(x)D^k\right\}\oplus\left\{\sum_{k=1}^{\infty}c_k(x)I^k\right\}=g_+\oplus g_-, \tag{3.15}$$

where the formal integration operator I is subject to the (purely algebraic) rule

$$\begin{aligned}If&=fI-f_xI^2+f_{xx}I^3-\cdots,\\ fI&=If+I^2f_x+I^3f_{xx}+\cdots.\end{aligned}\tag{3.16}$$

Using (3.16) it is more convenient to rewrite the elements of g_- as

$$g_-=\left\{\sum_{k=0}^{\infty}I^{k+1}b_k(x)\right\}. \tag{3.17}$$

Introducing a "trace formula"

$$\mathrm{Tr}(L)=\mathrm{Tr}\left(\sum_{k=0}^{\infty}I^{k+1}b_k+\sum_{k=0}^{m}a_kD^k\right):=\int_{-\infty}^{\infty}b_0\,dx, \tag{3.18}$$

one observes

$$\langle L_+,L_-\rangle:=\mathrm{Tr}(L_+L_-)=\int_{-\infty}^{\infty}\left(\sum_{k=0}^{m}a_kb_k\right)dx, \tag{3.19}$$

with

$$L_-=\sum_{k=0}^{\infty}I^{k+1}b_k\in g_-,\ L_+=\sum_{k=0}^{m}a_kD^k\in g_+.$$

Assuming the function ring S to be equipped with the usual L_2 scalar product given by integration along the real axis we impose g_- to carry the topology of $l_2(\mathrm{S})$. Thus it can be seen from (3.19) that g_+ is dense in the dual of g_-. Hence up to closure (which we will omit in our notation since it is of no relevance for the following) the above trace engenders a nondegenerate duality between g_+ and g_-, i.e., $g_+=g_-^*$. In Ref. 2 it is shown that the I coefficient of a commutator $[L,\tilde L]$, L, $\tilde L\in g$ is always in the image of the differential operator D. Assuming that the coefficients a_k, b_k of the operators vanish rapidly as $x\to\pm\infty$, e.g., thinking of S as the Schwartz space of smooth functions vanishing rapidly at infinity, one finds

$$\langle L,\tilde L\rangle:=\mathrm{Tr}(L\tilde L)=\langle\tilde L,L\rangle, \tag{3.20}$$

thus giving a symmetric, invariant, nondegenerate trace form on g such that Theorems 1 and 2 can be applied to this setup.

We will calculate explicitly the bi-Hamiltonian formulation (3.12) of the integrable equations obtained from Theorems 1 and 2. To this end one has to calculate the Poisson structures (3.6) and (3.7) at a given point $L=\cdots+I^2b_1+Ib_0+a_0+\cdots+a_mD^m$, where the functions $\ldots b_1,b_0,a_0,\ldots,a_m$ will be considered as the "physical" coordinates of the abstract equations (3.12). If one thinks of a scalar valued function f on g as given by a functional

$$f(L)=\int_{-\infty}^{\infty}F(\ldots b_1,b_0,a_0,\ldots,a_m)dx \tag{3.21}$$

of the physical coordinates (and their derivatives), then its differential, interpreted as an element of g, is given by

$$df(L)=\sum_{k\geq 0}I^{k+1}\frac{\delta f}{\delta a_k}+\sum_{k\geq 0}\frac{\delta f}{\delta b_k}D^k, \tag{3.22}$$

where

$$\frac{\delta f}{\delta a}=\frac{\partial F}{\partial a}-\left(\frac{\partial F}{\partial a_x}\right)_x+\left(\frac{\partial F}{\partial a_{xx}}\right)_{xx}-\cdots \tag{3.23}$$

is the Euler operator. Indeed, with $\tilde L=\Sigma I^{k+1}\tilde b_k+\Sigma\tilde a_kD^k$ one obtains

$$\langle df(L),\tilde L\rangle=\int_{-\infty}^{\infty}\left(\sum\tilde a_k\frac{\delta f}{\delta a_k}+\sum\tilde b_k\frac{\delta f}{\delta b_k}\right)\times dx=\frac{\partial}{\partial\epsilon}\Big|_{\epsilon=0}f(L+\epsilon\tilde L). \tag{3.24}$$

Inserting (3.22) into (3.6) and (3.7) gives the explicit form of the Poisson brackets in terms of the physical coordinates. Of course, since the elements of g_- consist of infinite sums, the resulting brackets involve infinitely many of the variables b_i and hence are difficult to use. However, as noted previously, both Poisson structures are projectible to g_+, where only finitely many variables are involved when calculating the tensor at a given point $L=\Sigma_{k=0}^m a_kD^k\in g_+$. Inserting

$$(df_i)_-=\sum_{k\geq 0}I^{k+1}\frac{\delta f_i}{\delta a_k},\quad i=1,2 \tag{3.25}$$

into the projected bracket (3.4), i.e.,

$$\{f_1,f_2\}_R(L)=-2\,\mathrm{tr}\left(\sum_{k=0}^{m}a_kD^k\times\left[\sum_{p\geq 0}I^{p+1}\frac{\delta f_1}{\delta a_p},\sum_{q\geq 0}I^{q+1}\frac{\delta f_2}{\delta a_q}\right]\right) \tag{3.26}$$

shows that only $\Sigma_{k=0}^{m-2}I^{k+1}(\delta f_i/\delta a_k)$ contributes, i.e., the Poisson bracket (3.7) is actually projected to the submanifold given by fixing the two highest coefficients a_m and a_{m-1} of the operator L. Indeed, it is easily seen that the Hamiltonian systems

$$\begin{aligned}\dot L&=\dot a_mD^m+\dot a_{m-1}D^{m-1}+\cdots\\ &=-2[L,df_-]_+=-2\left[a_mD^m+a_{m-1}D^{m-1}+\cdots,I\frac{\delta f}{\delta a_0}+I^2\frac{\delta f}{\delta a_1}+\cdots\right]_+\\ &=-2\left((m-1)a_m\left(\frac{\delta f}{\delta a_0}\right)_x+\left(a_m\frac{\delta f}{\delta a_0}\right)_x\right)D^{m-2}+\cdots D^{m-3}+\cdots D^{m-4}+\cdots\end{aligned} \tag{3.27}$$

leave these two highest coefficients constant. We rewrite (3.26) in terms of the Poisson tensor P_R as

$$\{f_1,f_2\}_R(L) = \langle df_2, P_R(L) df_1\rangle = \int_{-\infty}^{\infty} \sum_{ij\geqslant 0} \frac{\delta f_2}{\delta a_i} P_{Rij}(L) \frac{\delta f_1}{\delta a_j}\, dx \tag{3.28}$$

and obtain

$$P_R = \begin{pmatrix} -2a_2D - 2Da_2 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \tag{3.29}$$

when evaluating for $m = 2$, i.e., at the point $L = a_0 + a_1D + a_2D^2$.

Similarly, using the g_+ projection (3.11) of the Poisson bracket (3.6), inserting (3.25), and again evaluating for $m = 2$, the Poisson tensor

$$\{f_1,f_2\}_2(L) = \langle df_2, P_2(L) df_1\rangle = \int_{-\infty}^{\infty} \sum_{ij\geqslant 0} \frac{\delta f_2}{\delta a_i} P_{2ij}(L) \frac{\delta f_1}{\delta a_j}\, dx \tag{3.30}$$

is found to be

$$P_2 = \begin{pmatrix} 2(-a_2D^3a_2 + a_2D^2a_1 - a_1D^2a_2 + a_1Da_1 - a_0a_2D - Da_0a_2) & 2(a_1 + a_2D)Da_2 & 0 \\ 2a_2D(a_1 - Da_2) & 4a_2Da_2 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \tag{3.31}$$

Checking the explicit form of (3.11), it is easily seen that at any point $L = \Sigma_{k=0}^{m} a_k D^k \in g_+$ the components $P_{2im}, P_{2mi}, i = 0,...,m$ vanish, i.e., the second Poisson structure is projectible to the submanifold given by fixing the highest coefficient a_m. Hence all Hamiltonian systems relative to P_2 leave the coefficient a_m invariant. To obtain the bi-Hamiltonian vector fields (3.12) relative to P_R and P_2 we will now consider explicitly the Casimir functions $C_q(L) = \mathrm{Tr}(L^q)$ for $m = 2$, i.e., at the point $L = a_0 + a_1D + a_2D^2$.

Obviously, the functions $C_q, q\in\mathbb{N}$ vanish identically when restricting them to g_+. However, for certain $q\in\mathbb{Q}$ the fractional powers L^q can be calculated easily on a purely algebraic level, yielding nontrivial functions in involution on g_+. For $L = \Sigma_{k=0}^{m} a_k D^k$ one sets up the ansatz

$$L^{1/m} = a_m^{1/m} D + A_0 + IA_1 + I^2A_2 + \cdots \tag{3.32}$$

and calculates the mth power of this ansatz and compares it with L, thus obtaining a straightforward recursion scheme for the coefficients $A_0, A_1, \ldots$, e.g.,

$$\begin{aligned}(a_0 + a_1D + a_2D^2)^{1/2} &= \sqrt{a_2}D + \frac{1}{2}\left(\frac{a_1}{\sqrt{a_2}} - (\sqrt{a_2})_x\right) \\ &\quad + I\left(\frac{a_0}{2\sqrt{a_2}} - \frac{3a_{2x}^2}{32\sqrt{a_2^3}} + \frac{a_{2xx}}{8\sqrt{a_2}} - \frac{a_{1x}}{4\sqrt{a_2}} + \frac{a_1a_{2x}}{4\sqrt{a_2^3}} - \frac{a_1^2}{8\sqrt{a_2^3}}\right) \\ &\quad + I^2(\cdots) + I^3(\cdots) + \cdots .\end{aligned} \tag{3.33}$$

For $m = 2$ an interesting set of Casimir functions $C_{n/2}(L) = \mathrm{Tr}(L^{n/2})$, $n\in\mathbb{N}$ thus can be calculated explicitly by multiplying the above root of L with integer powers of L. From (3.5) the associated Hamiltonian systems are calculated as

$$\frac{d}{dt} L = 2[L,(dC_{n/2}(L))_+] = n[L,(L^{n/2-1})_+]. \tag{3.34}$$

In the physical coordinates a_0, a_1, and a_2 the first nontrivial of these equations ($n = 3$) has the form

$$\begin{aligned}\frac{d}{dt} a_0 = \frac{3}{16\sqrt{a_2^3}}(&-16a_{0x}a_2^2 \\ &- 3a_{2x}^3 - 4a_2^2a_{2xxx} + 6a_2a_{2x}a_{2xx} - 8a_{1x}a_2a_{2x} \\ &+ 8a_{1xx}a_2^2 + 8a_1a_{2x}^2 - 8a_1a_2a_{2xx} \\ &+ 8a_1a_{1x}a_2 - 4a_1^2a_{2x}),\end{aligned} \tag{3.35}$$

$$\frac{d}{dt} a_1 = 0, \quad \frac{d}{dt} a_2 = 0.$$

The higher equations are very complicated since for $n = 5$ the time evolution of a_0 is a lengthy expression of 50 terms.

All equations (3.34) admit the bi-Hamiltonian formulation (3.12) relative to the Poisson operators (3.29) and (3.31). As discussed previously, all equations (3.34) imply $(d/dt)a_1 = (d/dt)a_2 = 0$ and hence can be interpreted as an evolution equation for a_0 in which a_1 and a_2 turn up as arbitrary and fixed (i.e., time-independent) parameters. Obviously, the first Poisson operator (3.29) can be projected to the submanifold spanned by the a_0 coordinate, i.e.,

$$\Theta_R = -2a_2D - 2Da_2 \tag{3.36}$$

is a Hamiltonian operator for the a_0 component of Eqs. (3.34). Here a_2 is now regarded as a parameter rather than as a variable. For the second Poisson structure (3.31) one observes that resulting from (3.12) the a_1 component vanishes, i.e.,

$$P_{2,10}\frac{\delta C}{\delta a_0}+P_{2,11}\frac{\delta C}{\delta a_1}=0 \tag{3.37}$$

for the relevant functions C. Solving (3.37) for $\delta C/\delta a_1$ yields

$$\begin{aligned}\frac{d}{dt}a_0&=P_{2,00}\frac{\delta C}{\delta a_0}+P_{2,01}\frac{\delta C}{\delta a_1}\\&=(P_{2,00}-P_{2,01}(P_{2,11})^{-1}P_{2,10})\frac{\delta C}{\delta a_0}.\end{aligned} \tag{3.38}$$

Indeed, the operator

$$\begin{aligned}\Theta_2&=P_{2,00}-P_{2,01}(P_{2,11})^{-1}P_{2,10}\\&=-a_2D^3a_2+a_2D^2a_1-a_1D^2a_2\\&\quad+a_1Da_1-2a_0a_2D-2Da_0a_2\end{aligned} \tag{3.39}$$

turns out to be a Poisson operator (relative to the a_0 coordinate) and compatible with (3.36). Hence (3.39) and (3.36) form a compatible pair of Hamiltonian operators for the a_0 components of (3.34) and a hereditary recursion operator is given by

$$\Phi=\Theta_2\Theta_R^{-1},$$
$$\Theta_R^{-1}=-\frac{1}{4\sqrt{a_2}}D^{-1}\frac{1}{\sqrt{a_2}},\quad D^{-1}=\int_{-\infty}^{x}\cdots d\xi. \tag{3.40}$$

The hierarchy of commuting integrable equations obtained by applying (3.40) to the first of Eqs. (3.35) gives (up to constants) the a_0 component of Eqs. (3.34). Here a_1 and a_2 are now arbitrary functions.

By construction all these equations admit $L=a_0+a_1D+a_2D^2$ as the Lax operator and the Lax formulation is given by (3.34). Obviously, for the special choice $a_1(x)=0$, $a_2(x)=1$ one encounters the Schrödinger operator and all structures constructed here reduce to the well-known structures of the KdV hierarchy.

For the special choice $a_1(x)=\epsilon/x$, $a_2(x)=1$ (3.35) reduces to

$$\frac{d}{dt}a_0=-3a_{0x}+\frac{3\epsilon(2-\epsilon)}{2x^3} \tag{3.41}$$

and the next equation of the hierarchy ($n=5$) reduces to the following explicitly space-dependent KdV:

$$\begin{aligned}\frac{d}{dt}a_0=&-\frac{5}{4}(a_{0xxx}+6a_0a_{0x})+\frac{15\epsilon(\epsilon-2)}{8x^2}a_{0x}\\&-\frac{15\epsilon(\epsilon-2)}{4x^3}a_0+\frac{15(\epsilon+2)\epsilon(\epsilon-2)(\epsilon-4)}{16x^5}.\end{aligned} \tag{3.42}$$

Example 2: The analysis of Example 1 can be carried out at any point $L\in g_+$. The hierarchy of bi-Hamiltonian equations corresponding to the Casimir functions $C_q=\mathrm{tr}(L^q)$, $q\in\mathbb{Q}$ always leaves the two highest coefficients of L invariant, so that the coefficients enter the hierarchies as two arbitrary functions of the space variable x. Hence each evaluation of the structures discussed previously at a given point $L\in g_+$ provides an example of an infinite-dimensional integrable system, with the explicit space dependence given by two arbitrary functions.

Having done the explicit calculations at the point $L=a_0+a_1D+a_2D^2$ in Example 1 the next simplest example is given by the point $L=a_0+a_1D+a_2D^2+a_3D^3$. In terms of these coordinates one obtains a hierarchy of bi-Hamiltonian systems in the field variables a_0 and a_1, with the two arbitrary functions a_2 and a_3. Already, it turns out that the first equations, as well as the two Hamiltonian formulations, are very lengthy for arbitrary a_2 and a_3; hence we only give the results for the special restriction $a_2=0$ and $a_3=1$. For $L=a_0+a_1D+D^3$ one finds the following first elements of the hierarchy (3.12):

$$\begin{aligned}\frac{d}{dt}L&=[L,(L^{1/3})_+]\Leftrightarrow\frac{d}{dt}\begin{pmatrix}a_0\\a_1\end{pmatrix}=\begin{pmatrix}-a_{0x}\\-a_{1x}\end{pmatrix},\\\frac{d}{dt}L&=[L,(L^{2/3})_+]\Leftrightarrow\frac{d}{dt}\begin{pmatrix}a_0\\a_1\end{pmatrix}\\&=\begin{pmatrix}-a_{0xx}+\frac{2}{3}a_{1xxx}+\frac{2}{3}a_1a_{1x}\\a_{1xx}-2a_{0x}\end{pmatrix}.\end{aligned} \tag{3.43}$$

From (3.4) and (3.11) the Hamiltonian pair for the hierarchy is (3.12) is found to be

$$\begin{aligned}\Theta_R&=\begin{pmatrix}0&-6D\\-6D&0\end{pmatrix},\\\Theta_2&=\begin{pmatrix}\frac{4}{3}(D^5+D^3a_1+a_1D^3+a_1Da_1)+2a_0D^2-2D^2a_0&-2D^4-2a_1D^2-2Da_0-4a_0D\\2D^4+2D^2a_1-2a_0D-4Da_0&-4D^3-2a_1D-2Da_1\end{pmatrix}.\end{aligned} \tag{3.44}$$

As in Example 1, the tensors (3.44) are found by calculating the Poisson tensors P_R and P_2 from (3.7) and (3.6) at the point $L=a_0+a_1+a_2D^2+a_3D^3$ and then projecting these operators to the submanifold given by $a_2=0,a_3=1$. An argument similar to (3.37) and (3.38) is applied, i.e., Θ_2 is given by

$$\Theta_2=\begin{pmatrix}P_{2,00}-P_{2,02}(P_{2,22})^{-1}P_{2,20}&P_{2,01}-P_{2,02}(P_{2,22})^{-1}P_{2,21}\\P_{2,10}-P_{2,12}(P_{2,22})^{-1}P_{2,20}&P_{2,11}-P_{2,12}(P_{2,22})^{-1}P_{2,21}\end{pmatrix}. \tag{3.45}$$

The first conservation laws associated to the hierarchy (3.12) are given by

$$\mathrm{tr}(L^{1/3}) = \int_{-\infty}^{\infty} \frac{1}{3} a_1\, dx,\quad \mathrm{tr}(L^{2/3}) = \int_{-\infty}^{\infty} \frac{2}{3} a_0\, dx,\quad \mathrm{tr}(L^{4/3}) = \int_{-\infty}^{\infty} \frac{4}{9} a_0 a_1\, dx. \tag{3.46}$$

IV. INVOLUTION OF TRANSLATED CASIMIRS

Throughout this section we will assume that R solves the modified Yang–Baxter equation YB(1) and $R^2 = 1$. We remark that YB(1) is now just the hereditary property, as introduced in Ref. 10. Note that the typical example $R = P_+ - P_-$ of the splitting case does indeed satisfy $R^2 = 1$; hence any splitting of an algebra into subalgebras provides an example for the structures to be discussed in this section.

In this case R equips g with three additional Lie brackets:

$$\begin{aligned} [a,b]_R &= [Ra,b] + [a,Rb],\\ [a,b]_S &= R\,[a,b]_R,\\ [a,b]_T &= [a,b]_R - R\,[a,b]. \end{aligned} \tag{4.1}$$

The bracket $[\ ,\ ]_S$ coincides with $\frac{1}{2}[\ ,\ ]_{RR}$ of the general iterative scheme of Sec. II and $[\ ,\ ]_T$ is an additional bracket special for the case of a hereditary operator.

As a result of observation 1 the R matrix again solves YB(1), starting with $[\ ,\ ]_R$ or $[\ ,\ ]_S$ instead of the original bracket; it is easily checked that the same is true for $[\ ,\ ]_T$. Iterating the construction of modified brackets, i.e., calculating the R, S, or T modifications of these new brackets, one just recovers the original bracket and its modifications (4.1), e.g., $[\ ,\ ]_{TT} = [\ ,\ ]$, $[\ ,\ ]_{RT} = [\ ,\ ]_S$, etc.

Via P, P_R, P_S, and P_T we denote the Lie Poisson tensors associated to the four brackets on g. The following simple observation immediately gives further information about the involutivity of Casimir functions.

Observation 2: Consider the vector field $X(L) = R^*L$ on g^*. Then

$$\begin{aligned} &L_X P = -P_T,\ L_X P_T = -P,\\ &L_X P_R = -P_S,\quad L_X P_S = -P_R, \end{aligned} \tag{4.2}$$

where L_X is the Lie derivative into the direction of X.

In addition to the results of Theorem 1 we obtain that Casimir functions on g (w.r.t. P, i.e., $[\ ,\ ]$) are not only in involution w.r.t. P_R, but also w.r.t. P_T, as

$$\begin{aligned} 0 = L_X\{C_1,C_2\} &= \langle L_X\, dC_2, P\, dC_1\rangle\\ &\quad + \langle dC_2,(L_X P)dC_1\rangle + \langle dC_2, PL_x\, dC_1\rangle\\ &= \langle dC_2,(L_X P)dC_1\rangle = -\{C_1,C_2\}_T. \end{aligned} \tag{4.3}$$

With the same argument Casimirs of the $[\ ,\ ]_T$ bracket are in involution w.r.t. P; by Theorem 1 they are also in involution w.r.t. P_S (since $[\ ,\ ]_S$ turns out to be the R modification of $[\ ,\ ]_T$). Note that the Casimirs of $[\ ,\ ]$ and $[\ ,\ ]_T$ (as well as those of $[\ ,\ ]_R$ and $[\ ,\ ]_S$) are in 1:1 correspondence, as can be seen easily from the identities

$$\begin{aligned} \{f\circ R^*, g\circ R^*\} &= \{f,g\}_T \circ R^*,\\ \{f\circ R^*, g\circ R^*\}_R &= \{f,g\}_S \circ R^*. \end{aligned} \tag{4.4}$$

However, there is still another way of obtaining functions in involution from Observation 2. With $L_X L_X P = P$ (and the same for P_R, P_S, and P_T) one finds an application for the following lemma which holds for arbitrary Poisson manifolds.

Lemma 1: Let P be a Poisson tensor on some manifold and let X be a vector field with flow F_t. If $L_X L_X P = \alpha P$ (where α is a scalar factor), then

$$\{C_1\circ F_{t_1}, C_2\circ F_{t_2}\}_P = \{C_1\circ F_{t_1}, C_2\circ F_{t_2}\}_{L_X P} = 0 \tag{4.5}$$

for all Casimir functions C_1, C_2 and all times $t_1, t_2 \in \mathbb{R}$.

The proof is given in the Appendix. The above statement does not depend on the Poisson property of the tensor P; it holds for any skew-symmetric (or symmetric) tensor field of the type (0,2). In fact, the Lie derivative $L_X P$ of a Poisson tensor does not necessarily yield a Poisson tensor, i.e., $\{f,g\}_{L_X P} := \langle dg, (L_X P) df\rangle$ might not satisfy the Jacobi identity. If $L_X P$ turns out to be Poisson, then it is automatically compatible with P, i.e., the sum of these two tensors is again Poisson. In this case Lemma 1 leads to the construction of a family of functions in involution w.r.t. a Hamiltonian pair.

An example for the case $\alpha = 0$ is given as follows: Let P be the Lie Poisson tensor on the dual g^* of a Lie algebra. Choose $X(L) = L_0$ (a constant element of g^*). Then $L_X P$ yields a coboundary $\{f,g\}_{L_X P} = \langle L_0,\ [df,dg]\rangle$ and $L_X L_X P = 0$. The flow of X is given by F_t: $L \to L + tL_0$. Hence Lemma 1 yields the result: The set of one-parameter families of functions $C_t(L) = C(L + tL_0)$, where C is a Casimir function, is in involution relative to the Hamiltonian pair P and $L_X P$. This was used in Ref. 5 to show the integrability of the geodesic flow on semisimple Lie algebras.

Another example for the case $\alpha = 0$ is given by the Poisson–Drinfeld structure on Poisson Lie groups since its second Lie derivative into the direction of any bi-invariant vector field on this group vanishes (Ref. 14).

Examples for the case $\alpha \neq 0$ are given by the Lie algebra splittings (2.6) and the resulting Lie Poisson tensors P, P_R, P_S, and P_T. Again, the flow of $X(L) = R^*L$ is calculated easily in this case:

$$F_t = \exp(tR^*)\colon L \to \cosh(t)L + \sinh(t)R^*L. \tag{4.6}$$

Remark: In this construction it is quite easy to find vector fields satisfying Lenard relations, i.e., admitting Hamiltonian formulations relative to different Poisson structures: As explained in the Appendix the translations of Lemma 1 are obtained by applying the Lie series $\exp(tL_X)$ to the Casimir functions. Applying this series to the identity

$$P\,dC = 0 \tag{4.7}$$

(identifying C as a Casimir function), one obtains

$$\begin{aligned} 0 &= (\exp(tL_X)P)(\exp(tL_X)dC) \\ &= a(t)(P + \epsilon(L_X P))d(C\circ F_t), \end{aligned}$$

i.e.,

$$P\,d(C\circ F_t) = -\epsilon(L_X P)d(C\circ F_t),\ \epsilon = b(t)/a(t) \tag{4.8}$$

where a and b are the functions given in the Appendix. Expanding (4.8) into powers of t obviously yields Lenard relations, e.g., if the Casimir functions are given as $C_q(L) = \mathrm{tr}(L^q)$ and the R matrix arises from a Lie algebra splitting, then according to (4.6) $(C_q \circ F_t)(L) = \mathrm{tr}((\exp(tR^*)L)^q)$ is proportional to $\mathrm{tr}((L + \epsilon R^* L)^q)$ and one obtains an expansion

$$\mathrm{tr}((L + \epsilon R^* L)^q) = \sum_{k=0}^{q} C_q^k(L)\epsilon^k. \tag{4.9}$$

According to Lemma 1 the set of functions C_q^k is again in involution relative to P and $L_X P = -P_T$ and Eq. (4.9) yields

$$P\,dC_q^k = P_T C_q^{k-1},\ k = 1\cdots q. \tag{4.10}$$

V. AN INVOLUTION THEOREM FOR UNITARY *R* STRUCTURES

Throughout this section we assume that R is unitary in the sense of (2.2). We will consider the presymplectic two-form

$$\omega_0(L,\widetilde{L}) = \langle RL,\widetilde{L}\rangle,\quad L,\widetilde{L}\in g \tag{5.1}$$

on g. Note that in the splitting case the R structure (2.6) is unitary only if $g_- = g_+^*$. In this situation (5.1) is just the canonical symplectic form on the symplectic space $g = g_+ \oplus g_+^*$.

It turns out that an analog of a Hamiltonian pair can be constructed with (5.1) and the Lie Poisson structure on g. One has the more general lemma which follows.

Lemma 2 (Refs. 15, 16)*:* Let $\omega_0(X,Y)(u) = \langle B_0(u)X,Y\rangle; u\in M,\ X,Y\in T_u M$ be a closed two-form on a Poisson manifold M with the Poisson tensor P_1. Define $\Phi := P_1 B_0$, $P_{n+1} := \Phi^n P_1$, and $\omega_n(X,Y) := \langle B_0\Phi^n X,Y\rangle$. Then one finds the implications (i) $\Leftrightarrow$ (ii) $\Rightarrow$ (iii) $\Rightarrow$ (iv) $\Leftrightarrow$ (v) for the following statements: (i) all ω_m's are closed; (ii) ω_1 is closed; (iii) ϕ has vanishing Nijenhuis Torsion, i.e., $L_{\phi x}\phi = \phi L_x \phi$ for all vector fields x; (iv) P_2 is a Poisson tensor; and (v) all P_n's are Poisson tensors.

For injective P_1 one also has (iii) $\Rightarrow$ (ii) and for surjective P_1 one has (iv) $\Rightarrow$ (iii). If B_0 is invertible, then (i) and (ii) are equivalent to $P_1 + B_0^{-1}$ being a Poisson tensor, i.e., to the notion of a compatible Hamiltonian pair.

According to this lemma it is sufficient to check (ii) to find such a compatibility between $B_0 = R$ [i.e., ω_0 given by (5.1)] and the Lie Poisson tensor $P(=P_1)$ on g in the following.

Lemma 3: The two-form $\omega_1(X,Y)(L) = \langle RP(L)RX,Y\rangle = \langle L,[RY,RX]\rangle$ is closed if R solves $YB(0)$. This is shown easily since the invariance of the metric implies

$$\begin{aligned} d\omega_1(X,Y,Z) &= \langle Z,[RY,RX]\rangle + \mathrm{cycl} \\ &= \langle Z,[RY,RX] - R[Y,X]_R\rangle. \end{aligned} \tag{5.2}$$

For invertible R another way of looking at this structure is to regard (2.8) as a Poisson bracket

$$\{f_1,f_2\}_0 = B(df_2,df_1) = \langle R^{-1} df_2,df_1\rangle, \tag{5.3}$$

which—as a two-cocycle—forms a compatible Hamiltonian pair with the Lie Poisson structure. Hence Lemma 3 follows by the compatibility criteria of Lemma 2. We remark that

$$\{f_1,f_2\}_{P_2}(L) = \langle [L,df_1],R[L,df_2]\rangle \tag{5.4}$$

and the square of the Nijenhuis tensor

$$\Phi(L) = P(L)R: X \to [RX,L] \tag{5.5}$$

constitute the quadratic parts of the bi-Hamiltonian scheme set up by Magri[12] to describe the hierarchies of KdV and the Kadomtsev–Petviashvili equation in a unified way.

The vanishing Nijenhuis torsion of the operator Φ implies that the underlying manifold, i.e., g can be foliated by the eigenspaces of this tensor field. Up to certain additional technical assumptions (essentially, diagonalizability of Φ) it can be shown[17] that the invariants of Φ are in involution relative to all Poisson structures $P_1, P_2, \ldots$ of Lemma 2.

As the Lie Poisson tensor P constitutes a factor of Φ [5.5], one of the eigenvalues vanishes identically. For invertible R the zero eigenspaces of Φ are spanned by the vector fields $R^{-1}\,dC$, where C is a Casimir function of the Lie Poisson structure. We remark that these vector fields leave ω_0, $P_1 = P$ and hence all closed forms and all Poisson structures of Lemma 2 invariant.

For solutions of YB(0) we refer to Ref. 8. As remarked in Sec. II, the inverse of a derivation on g solves YB(0) and hence leads to a hereditary operator (5.5). On finite-dimensional Lie algebras the construction of invertible derivations is not straightforward or does not even exist, but for the infinite-dimensional example of currents $C^\infty(\mathbb{R}, g^0)$ over a finite-dimensional Lie algebra g^0 the differential operator D provides a derivation that, subject to boundary conditions, can be inverted.

As an example let us consider $g = C^\infty(\mathbb{R},\mathrm{gl}(2,\mathbb{R}))$, with the charts

$$a = \begin{pmatrix} a_1(x) & a_3(x) \\ a_2(x) & a_4(x) \end{pmatrix} \in g,\quad a^* = \begin{pmatrix} a_1^*(x) & a_2^*(x) \\ a_3^*(x) & a_4^*(x) \end{pmatrix} \in g^*, \tag{5.6}$$

endowed with the duality

$$\langle a,a^*\rangle = \int_{-\infty}^{\infty} \mathrm{tr}(a^* a)dx = \int_{-\infty}^{\infty} \sum_{i=1}^{4} a_i^* a_i\,dx. \tag{5.7}$$

An antisymmetric derivation on g is given by $R^{-1} = D + ad_{b_0}$, where D is differentiation and b_0 is a fixed element of g. Relative to the charts (5.6) R^{-1} (regarded as the map from g to g^*) and its inverse are found to be

$$R^{-1}=\begin{pmatrix} D & b & 0 & 0\\ -b & 0 & D & b\\ 0 & D & 0 & 0\\ 0 & -b & 0 & D\end{pmatrix},$$

$$R=\begin{pmatrix} I & 0 & -IbI & 0\\ 0 & 0 & I & 0\\ IbI & I & -2IbIbI & -IbI\\ 0 & 0 & IbI & I\end{pmatrix}, \tag{5.8}$$

with $I=D^{-1}=\int_{-\infty}^{x}$, assuming the elements of g vanish rapidly at $|x|=\infty$. Here b_0 has been chosen as

$$b_0=\begin{pmatrix} 0 & b(x)\\ 0 & 0\end{pmatrix}. \tag{5.9}$$

The Lie Poisson tensor, here regarded as the map $P(a):g^*\to g$, has the form

$$P=\begin{pmatrix} 0 & -a_2 & a_3 & 0\\ a_2 & 0 & a_4-a_1 & -a_2\\ -a_3 & a_1-a_4 & 0 & a_3\\ 0 & a_2 & -a_3 & 0\end{pmatrix}. \tag{5.10}$$

The symplectic form given by R and the Poisson structure (5.10) form a compatible pair in the sense of Lemma 2 and the operator $\Phi=PR$ is hereditary.

Rather than looking at the spectrum of Φ it is more convenient to look for a bi-Hamiltonian system generated by R^{-1} and P. One finds a bi-Hamiltonian vector field

$$\begin{pmatrix} a_3\\ a_4-a_1\\ 0\\ -a_3\end{pmatrix}=P\,df_0=R^{-1}\,df_1, \tag{5.11}$$

with

$$f_0=\int_{-\infty}^{\infty} a_3\,dx,\quad f_1=\int_{-\infty}^{\infty}(a_1Ia_3+a_3Ia_4-b(Ia_3)^2)dx. \tag{5.12}$$

According to the bi-Hamiltonian scheme (see, e.g., Refs. 9–11) all the covector fields $\gamma_n:=(\Phi^*)^n\,df_0$ are closed, i.e., locally $\gamma_n:=df_n$. The functions f_n are in involution relative to R^{-1} (regarded as the Poisson tensor), $\Phi R^{-1}=P$, $\Phi^2R^{-1}=PRP$,... . These functions can also be obtained via "master symmetries."[15,16] Note that the vector field $\tau_0(L)=L$ is a conformal symmetry for R, as well as for the Lie Poisson structure P, i.e.,

$$L_{\tau_0}R^{-1}=2R^{-1},\quad L_{\tau_0}P=-P. \tag{5.13}$$

Checking $L_{\tau_0}f_0=f_0$ one concludes that

$$L_{\tau_n}\gamma_m=(1+m+n)\gamma_{n+m} \tag{5.14}$$

for $\tau_n:=\Phi^n\tau_0$. Hence

$$f_{n+m}=[1/(1+m+n)]L_{\tau_n}f_m \tag{5.15}$$

provides a scheme to obtain the higher functions using the master symmetries $\tau_n, n\geqslant 1$.

ACKNOWLEDGMENTS

The author would like to thank A. Weinstein for his kind hospitality during the author's visit to the University of California, Berkeley.

APPENDIX: PROOF OF LEMMA 1

For any smooth function f the composition with the flow F_t of a vector field X can be represented via the Lie series

$$f\circ F_t=\exp(tL_X)f=\sum\frac{t^k}{k!}(L_X)^kf.$$

Since the Lie series provides a homomorphism w.r.t. the composition of arbitrary tensor fields T_1, T_2 on the manifold, i.e.,

$$\exp(tL_X)(T_1\circ T_2)=(\exp(tL_X)T_1)\circ(\exp(tL_X)T_2),$$

one obtains

$$\begin{aligned}&\{f\circ F_{-t},g\circ F_{-t}\}_P\circ F_t\\ &\quad=\exp(tL_X)\langle d(g\circ F_{-t}),P\,d(f\circ F_{-t})\rangle\\ &\quad=\langle\exp(tL_X)d(g\circ F_{-t}),(\exp(tL_X)P)\\ &\qquad\times\exp(tL_X)d(f\circ F_{-t})\rangle\\ &\quad=\langle dg,(\exp(tL_X)P)df\rangle.\end{aligned}$$

From $L_XL_XP=\alpha P$ the remaining Lie series can be calculated explicitly:

$$\begin{aligned}\exp(tL_X)P&=a(t)P+b(t)L_XP\\ &=\begin{cases}P+tL_XP, & \text{for } \alpha=0,\\ \cosh(\sqrt{\alpha t})P\\ +[\sinh(\sqrt{\alpha t})/\sqrt{\alpha}]L_XP, & \text{for } \alpha\neq 0.\end{cases}\end{aligned}$$

Inserting $f=C_1\circ F_{t_1}$, $g=C_2\circ F_{t_2}$, one obtains

$$\begin{aligned}&\{C_1\circ F_{t_1-t},C_2\circ F_{t_2-t}\}_P\circ F_t\\ &\quad=a(t)\{C_1\circ F_{t_1},C_2\circ F_{t_2}\}_P+b(t)\{C_1\circ F_{t_1},C_2\circ F_{t_2}\}_{L_XP}.\end{aligned}$$

For $t=t_1$ and $t=t_2$ the lhs vanishes since C_1, C_2 are to be Casimir functions. The resulting two homogeneous equations show that the brackets on the rhs vanish for all times $t_1\neq t_2$. By continuity this result extends to $t_1=t_2$.

[1] T. Ratiu, "Involution theorems," Lect. Notes Math. **775**, 219 (1980).
[2] M. Adler, "On a trace functional for formal pseudo-differential operators and the symplectic structure of the Korteweg-de Vries equations," Inventiones Math. **50**, 219 (1979).
[3] M. A. Semenov-Tian-Shansky, "Dressing transformations and Poisson group actions," Publ. RIMS, Kyoto Univ. **21**, 1237 (1985).
[4] M. A. Semenov-Tian-Shansky, "What is a classical R-matrix?," Funct. Anal. Appl. **17**, 259 (1983).
[5] A. S. Mishenko and A. T. Fomenko, "Euler equations on finite dimensional Lie groups," Math. USSR Izv. **12**, 371 (1978).
[6] I. M. Gel'fand and I. Y. Dorfman, "Hamiltonian operators and the classical Yang-Baxter equation," Funct. Anal. Appl. **17**, 241 (1982).
[7] V. G. Drinfel'd, "Hamiltonian structures on Lie groups, Lie bi-algebras and the geometric meaning of the classical Yang–Baxter equations," Sov.

Math. Dokl. **27**, 68 (1983).
[8]A. A. Belavin and V. G. Drinfel'd, "Triangle equations and simple Lie algebras," Math. Phys. Rev. Moscow **4**, 94 (1987).
[9]F. Magri, "A simple model of the integrable Hamiltonian equation," J. Math. Phys. **19**, 1156 (1978).
[10]B. Fuchssteiner and A. S. Fokas, "Symplectic structures, their Baecklund transformations, and hereditary symmetries," Physica D **4**, 47 (1981).
[11]I. M. Gel'fand and I. Y. Dorfman, "Hamiltonian operators and algebraic structures related to them," Funct. Anal. Eqs. Priloz. **13**, 248 (1979); "The Schouten bracket and Hamiltonian operators," Funct. Anal. Eqs. Priloz. **14**, 223 (1980).
[12]M. Ablowitz, B. Fuchssteiner, and M. Kruskal, *Topics in Soliton Theory and Exactly Solvable Nonlinear Equations* (World Scientific, Singapore, 1986).
[13]A. Weinstein, "The local structure of Poisson manifolds," J. Diff. Geom. **18**, 523 (1983).
[14]A. Weinstein, "Some remarks on dressing transformations," preprint UTYO-MATH 87-16, Tokyo (1987).
[15]H. M. M. ten Eikelder, "Symmetries for dynamical and Hamiltonian systems," CWI tract **17**, CWI, Amsterdam (1985).
[16]W. Oevel, "Master symmetries: weak action/angle structure for Hamiltonian and non-Hamiltonian systems," preprint, Paderborn (1986).
[17]H. M. M. ten Eikelder, "On the local structure of recursion operators for symmetries," Indag. Math. **48**, 389 (1986).

Partial waves for the linearized Yang–Mills equation about a monopole background

Shahn Majid
Lyman Laboratory of Physics, Harvard University, Cambridge, Massachusetts 02138

(Received 1 July 1988; accepted for publication 12 October 1988)

The properties of the monopole-vector spherical harmonics that are needed to transform covariant differential equations on charged vector fields in the presence of a monopole into purely radial ones are computed. This further extends the work of Wu and Yang [Nucl. Phys. B **107**, 365 (1976)]. As an application, the complete set of partial waves of SU(2) Yang–Mills fluctuations about the U(1) monopole configuration are explicitly computed. They are generated from the scalar solutions to the covariant Laplacian by the covariant operators $\mathbf{D}$ and $\mathbf{r}\times\mathbf{D}$, in analogy with the uncharged case.

I. INTRODUCTION

The fundamental paper of Wu and Yang[1] computed the properties of the monopole-spherical harmonics $\{Y_{qjk}\}$. These can be used to solve charged scalar field equations in the presence of a monopole of charge q. The monopole-vector spherical harmonics $\{\mathbf{Y}_q^{jkl}\}$ are also known. For example, they were used to study general properties of gauge field fluctuations in Ref. 2. We continue this line of development by computing certain useful properties of the $\{\mathbf{Y}_q^{jkl}\}$.

Thus, when $q=0$, the spherical harmonics reduce to the ordinary scalar and vector spherical harmonics. These are very useful in electromagnetism[3] and other situations, e.g., Ref. 4, because an expansion in terms of them reduces invariant differential equations on $\mathbb{R}^3$ to purely radial ones. The work of Wu and Yang makes this equally easy for $q\neq 0$ in the scalar case, but the vector case has been severely limited by the complexity of the $\{\mathbf{Y}_q^{jkl}\}$. This is addressed by Lemma 3.1, in Sec. III below, the main result of the paper. It provides the necessary data to reduce covariant vector differential equations in the presence of the monopole to radial ones. The proof of Lemma 3.1 occupies Sec. III and the Appendix.

As an elementary application, the linearized Yang–Mills equation for SU(2) gauge field fluctuations about the monopole configuration is completely solved. This is done by applying our lemma to the equations formulated in Refs. 2 and 5. These are set up in Sec. II along the lines of Ref. 5. Previously, Brandt and Neri[2] had studied these equations and concluded the existence of certain unstable modes. These will be evident among our explicit results for the complete set of partial waves. The results also explicitly confirm expectations about magnetic antiscreening, cf. Ref. 6. An independent check of these modes is provided by geometrical methods in Sec. IV.

The remainder of the introduction sets up the explicit application and states the resulting solutions. From a mathematical point of view, we compute the tangent space to the space of SU(2) Yang–Mills connections on $\mathbb{R}^3-0$ or S^3-0 at a reducible connection, that of a U(1) monopole. On S^4 reducible connections are known to have the structure of conical singularities in $\mathscr{A}/\mathscr{G}$. In our case, the structure will also be interesting, exhibiting an irrational power law for the radial dependence (as known for the scalar field case) and a mixing of the orbital angular momentum modes.

Note that gauge field fluctuations in the monopole U(1) direction are not charged—they do not see the monopole background (i.e., they can be solved with $q=0$). Hence for these modes we just have to solve Laplace's equation for vector fields on $\mathbb{R}^3-0$. As stated above, this is easily done explicitly in terms of vector spherical harmonics $\mathbf{Y}^{jkl}$ where $j\in 0,1,\ldots$, $k\in -j,\ldots,j$ and $l\in j-1,j,j+1$. The answer for solutions of the form $\nabla\cdot\boldsymbol{\phi}=0$ is

$$\boldsymbol{\phi}\in\{(1/r^{j+1})\mathbf{Y}^{jkl=j},\quad j=1,2,\ldots\}$$
$$\oplus\{(1/r^{j+2})\mathbf{Y}^{jkl=j+1},\quad j=0,1,\ldots\}\tag{1.1}$$

as a vector space. The first terms are all tangential to the sphere at radius r while the second terms are all pure gradient. Here we intend standard polar coordinates on $\mathbb{R}^3-0$. This can also be obtained from solving the Laplacian on $S^2=\mathrm{SO}(3)/\mathrm{SO}(2)$ by group theory[7] and then the radial equation. For the non-Abelian fluctuations we shall use the known form of the corresponding $\{\mathbf{Y}_q^{jkl}\}$ to carry through the corresponding more complicated computation. Group theory is not directly applicable here, but our methods will suggest some elegant underlying structure.

More precisely, let $P\rightarrow(\mathbb{R}^3-0)$ be the trivial SU(2) bundle. Among the connections on this let

$$\mathscr{M}=\text{smooth solutions of the Yang–Mills equations on } P.\tag{1.2}$$

In this space lies the reducible connection of the $\mathrm{U}(1)\subset\mathrm{SU}(2)$ monopole. Explicitly let U_+,U_- be patches covering the $z>0$ and $z<0$ halves of $\mathbb{R}^3-0$, then

$$g_{+-}(\phi)=e^{2iQ\phi},\quad \mathbf{A}_q=-iQ\,\mathbf{A}_{\mathrm{Dirac}},$$
$$\mathbf{A}_{\mathrm{Dirac}\pm}=\pm(\hat{z}\times\hat{r})/r(1\pm\hat{\mathbf{r}}\cdot\hat{\mathbf{z}}),\tag{1.3}$$
$$Q=q\sigma_3/2\in\underline{\mathrm{SU}}(2),\quad q\in\mathbb{Z}/2,$$

in an explicit gauge where $\partial\cdot\mathbf{A}_q=0$. Here it should be pointed out that any SU(2) bundle over $\mathbb{R}^3-0$ is bundle equivalent to the trivial cross-product bundle and ∂ denotes the flat connection in these cross-product coordinates on $\mathrm{SU}(2)\times(\mathbb{R}^3-0)$. [However, the bundle reduced to an SO(3) or (as above) a U(1) subgroup need not be trivial as $\pi_1(\mathrm{SO}(3))=\mathbb{Z}_2$ and $\pi_1(\mathrm{U}(1))=\mathbb{Z}$. Thus the transition functions in (1.3) cannot be transformed to the identity by SU(2) gauge transformations that map the U(1) onto itself.] It is also convenient to take the flat metric on $\mathbb{R}^3-0$ in these coordinates.

 0022-2488/89/051150-08$02.50

In this paper then we compute

$$T_{\mathbf{A}_q}\mathscr{M}\cap\Sigma_{\mathbf{A}_q}, \tag{1.4}$$

where $\Sigma_{\mathbf{A}_q}$ is the gauge-fixing slice defined by $D_{\mathbf{A}_q}\cdot\delta\mathbf{A}=0$,

$$\Sigma_{\mathbf{A}_q}=\{\delta\mathbf{A}\in\Gamma(\mathrm{ad}\,P\otimes T^*(\mathbb{R}^3-0))\quad \text{s.t. } D_{\mathbf{A}_q}\cdot\delta\mathbf{A}=0\},$$

where $D_{\mathbf{A}_q}$ is the covariant derivative. Since $\mathbf{A}_q$ is a reducible connection, the usual theorem[8] that this is locally a good slice, i.e., that there exists an open $\mathscr{O}_{[\mathbf{A}_q]}\ni[\mathbf{A}_q]$ in $\mathscr{A}/\mathscr{G}$ such that

$$\pi:\ \Sigma_{\mathbf{A}_q}\cap\pi^{-1}(\mathscr{O}_{[\mathbf{A}_q]})\to\mathscr{O}_{[\mathbf{A}_q]}$$

is a bijection, need not be true. Indeed, we shall find in Sec. III that half of our solutions are "Gribov copies" (we call them type II) that lie in $\Sigma_{\mathbf{A}_q}$ but are gauge transforms of $\mathbf{A}_q$.

The holonomy group of the connection $\mathbf{A}_q$ is "broken down to U(1)." (In a physical setting such an $\mathbf{A}_q$ can arise as the approximate connection outside of a non-Abelian core, here modeled by $\{0\}$, due to the presence of a Higgs field Φ obeying $D_{\mathbf{A}}\Phi=0$.[9] The $\mathbf{A}_q$ themselves do not have finite action due to the singularity at 0.) Now, because $\mathbf{A}_q$ is reducible, the adjoint bundle ad P also splits so that $\delta\mathbf{A}$ may be decomposed as fluctuations $\boldsymbol{\phi}$ in the U(1) Abelian direction in SU(2) and non-Abelian fluctuations $\boldsymbol{\Psi}$ in the orthogonal noncommuting directions. Explicitly in the coordinates of (1.3),

$$\delta\mathbf{A}=-\imath\boldsymbol{\phi}\,\frac{\sigma_3}{2}-\imath\boldsymbol{\psi}\,\frac{\sigma_+}{2}-\imath\boldsymbol{\psi}^*\,\frac{\sigma_-}{2}. \tag{1.5}$$

With this decomposition in mind, we find the (physical) type I solutions

$$T_{\mathbf{A}_q}\mathscr{M}\cap\Sigma_{\mathbf{A}_q}|_{\boldsymbol{\psi}}$$
$$=\left\{\frac{\Sigma_{l=j-1}^{j+1}\mathbf{Y}_q^{jkl}b_l}{r^{1/2+[j(j+1)-q^2+1/4]^{1/2}}};\right.$$
$$b=\left(\frac{q}{\sqrt{j(2j+1)}},\frac{\sqrt{j^2-q^2}}{\sqrt{(j+1)j}},\right.$$
$$\left.\left.\frac{q\sqrt{j^2-q^2}}{\sqrt{((j+1)^2-q^2)(j+1)(2j+1)}}\right),\ j=q,q+1,\ldots\right\}, \tag{1.6}$$

which are purely tangential to the sphere at radius r, and similar pure gauge modes for type II. The commuting directions $\boldsymbol{\phi}$ are also described by this by setting $q=0$ in (1.6), to recover (1.1).

The monopole vector spherical harmonics $\{\mathbf{Y}_q^{jkl}\}$ are given in detail in Ref. 2 where they were used in another context (to show that for $q>\frac{1}{2}$ there is an unstable mode—a solution regular near 0 but unbounded as $r\mapsto\infty$). [For each of the solutions of decay r^α, the equations for $\alpha(\alpha+1)$ that arise have another solution $r^{-\alpha-1}$ that therefore blows up at ∞ if $\alpha<-1$.]

As a check on these results, note that for $q=\frac{1}{2}$ the slowest possible decay in this Coulomb gauge for $\delta\mathbf{A}$ in the non-Abelian direction in SU(2) is for $j=\frac{3}{2}$. This is because the $j=\frac{1}{2}$ mode, in fact, vanishes identically. Hence we find

$$\boldsymbol{\psi}\sim r^{-(1/2)(1+\sqrt{15})}, \tag{1.7}$$

so that at large r the U(1) $\mathbf{A}_q$ and $\boldsymbol{\phi}$ necessarily dominate. This is the phenomenon known as magnetic SU(2)-color antiscreening[6] whereby the classical Yang–Mills equations on $\mathbb{R}^3-0$ keep non-Abelian modes "classically confined" to near the origin. In all explicit exact (as opposed to linearized fluctuation) solutions of Yang–Mills, the non-Abelian fields, in fact, decay exponentially, e.g., see Ref. 9.

The present work was motivated by the following elementary consideration. Consider an Abelian configuration such as (1.3) viewed as a time-independent connection on $(\mathbb{R}^3-0)\times\mathbb{R}$. Suppose at $t=0^+$ that the connection $\mathbf{A}_q$ is suddenly perturbed by $\delta\mathbf{A}=Q'\boldsymbol{\psi}$ in some direction $Q'\in\underline{\mathrm{SU}}(2)$, where $\boldsymbol{\psi}$ is real with compact support in the patch U_+, say, in the coordinates of (1.3). Then the energy $\mathscr{E}=\frac{1}{2}\int_{\mathbb{R}^3-B_\epsilon}d^3x|\partial\times\mathbf{A}|^2$ jumps,

$$\mathscr{E}(0^+)=\mathscr{E}(0)+\frac{1}{2}\,\mathrm{tr}\,Q'^2\int_{\mathbb{R}^3-B_\epsilon}d^3x|\partial\times\boldsymbol{\psi}|^2$$
$$+\frac{1}{8}\,\mathrm{tr}[Q,Q']^2\int_{\mathbb{R}^3-B_\epsilon}d^3x|\mathbf{A}_{\mathrm{Dirac}}\times\boldsymbol{\psi}|^2,$$

where $\mathscr{E}(0)$ and $\mathscr{E}(0^+)$ are not finite as $\epsilon\mapsto 0$ in the present case (1.3) but their difference is. The second term is the energy of the U(1) curvature of $\boldsymbol{\psi}$ and the third term is the interaction. From this one may conclude from the principle of virtual work (or by looking at the Yang–Mills equation explicitly) that as time evolves, non-Abelian fluctuations about an Abelian background evolving according to the Yang–Mills equations, tend to align themselves by rotating in SU(2). Equation (1.6) was obtained from a desire to see this explicitly—we see that only asymptotically Abelian configurations are possible in the steady state.

II. FURTHER PRELIMINARIES

(a) For slightly greater generality in Sec. IV, consider $\mathbf{A}_q$ in (1.3) a connection on the $\mathbb{R}^4$-time axis, with metric $(+\,-\,-\,-)$ and $(\mathbf{A}_q)_0=0$. Time-dependent fluctuations $\delta\mathbf{A}$ are then allowed. For $F(\mathbf{A}+\delta\mathbf{A})$ to obey the Yang–Mills equation when $\mathbf{A}$ does, we need, with $\mu,\nu=0,1,2,3$ (summation convention) and $\mathbf{D}\cdot\delta\mathbf{A}=0$,

$$\Box_{\mathbf{A}}\delta\mathbf{A}_\mu+2[F_\mu{}^\nu,\delta\mathbf{A}_\nu]=\tfrac{1}{2}[\delta\mathbf{A}_\mu,\mathbf{D}_{\mathbf{A}+\delta\mathbf{A}}\,\delta\mathbf{A}^\mu]$$
$$=O(\delta\mathbf{A}^2),\quad \Box_{\mathbf{A}}=\mathbf{D}^2,\quad \mathbf{D}\equiv\mathbf{D}_{\mathbf{A}}. \tag{2.1}$$

For us $F_0{}^i=0$, $F_i{}^j=(-q\imath\sigma_3/2)(-F_{ij}^{\mathrm{Dirac}})$. Making the decomposition analogous to (1.5), we have, to linear order,

$$\Box_{\mathbf{A}}\psi_0=0,\quad \Box\phi_0=0,\quad \Box\boldsymbol{\phi}=0,$$
$$\Box_{\mathbf{A}_q}\boldsymbol{\psi}-2\imath q[(\mathbf{r}\times\boldsymbol{\psi})/r^3]=0, \tag{2.2}$$
$$\partial^0\phi_0-\nabla\cdot\boldsymbol{\phi}=0,\quad \partial^0\psi_0-\mathbf{D}\cdot\boldsymbol{\psi}=0, \tag{2.3}$$

where $\cdot$ and $\times$ denote the usual inner and cross product on $\mathbb{R}^3$ given by $\delta_{ij}=(+\,+\,+)$ and ϵ_{ijk}. For us $\Box_{\mathbf{A}}=(\partial^0)^2-\mathbf{D}\cdot\mathbf{D}$ and $\mathbf{D}=\nabla-\imath q\mathbf{A}_{\mathrm{Dirac}}$ is the derivative on fields $\boldsymbol{\psi}$, ψ_0 as sections of a time-independent line bundle. Here, $\psi_0,\phi_0,\boldsymbol{\phi}$ decouple and can be treated separately. Since we are interested in time-independent solutions we shall henceforth set $\psi_0\equiv\phi_0\equiv 0$. The solution for $\boldsymbol{\phi}$ is standard, (1.1). It remains to solve for $\boldsymbol{\psi}$.

(b) Now under rotations R_θ we have an SO(3) action

$$R_\theta \psi = e^{\imath\theta\cdot S} \exp\left(\int_{R_{\imath\theta}^{-1}\mathbf{x}}^{\mathbf{x}} \mathbf{A}_q\right) \psi(R_\theta^{-1}\mathbf{x})$$
$$\equiv \exp[\imath\theta\cdot(\mathbf{S}+\mathbf{L})]\psi,$$

where the orbital angular momentum generator is

$$\mathbf{L} = -\imath\mathbf{r}\times\mathbf{D} - q\mathbf{r}/r,$$

which obeys $[L_i,L_j] = \imath\epsilon_{ijk}L_k$ as desired. Here, $-q\mathbf{r}/r$ can be interpreted[5] as the orbital angular momentum of the monopole. The integral is taken over the arc described by $R_{t\theta}^{-1}\mathbf{x}$, $t\in[0,1]$.

Let

$$S_i a_j = -\imath\epsilon_{ijk}a_k$$

act pointwise on all vector fields a_k. Then $[S_i,S_j] = -\imath\epsilon_{ijk}S_k$ as required. Let $\mathbf{J}=\mathbf{L}+\mathbf{S}$ and $[S_i,L_j] = -\imath\epsilon_{ijk}L_k$ (i.e., in the present formalism **S** acts on **D** and **r** as vectors corresponding to the Poincaré group as a semidirect product). Hence

$$[J_i,J_j] = \imath\epsilon_{ijk}(\mathbf{L}-2\mathbf{L}-\mathbf{S})_k = -\imath\epsilon_{ijk}J_k.$$

This defines the spin and angular momentum operators. Note that

$$[J_i,L_j]=0,\quad [J_i,r_j]=0,$$
$$[J_i,D_j]=0,\quad [J^2,J_z]=0. \tag{2.4}$$

Let $J^2=j(j+1)$, $L^2=l(l+1)$, and $J_z\equiv J_3=k$ on their simultaneous eigenfunctions.

Lemma 2.1 (cf. Ref. 5)*:* Let ψ be an eigenstate of $J^2=j(j+1)$. Then for $\mathbf{D}\cdot\psi=0$, (2.2) becomes

$$(\partial^0)^2\psi\equiv H\psi = -H_J\psi + 2\mathbf{D}\left(\frac{\psi\cdot\mathbf{r}}{r^2}\right) + 4\frac{\mathbf{r}}{r^2}\frac{\psi\cdot\mathbf{r}}{r^2}, \tag{2.5}$$

where

$$H_J = -\frac{\partial^2}{\partial r^2} - \frac{2}{r}\frac{\partial}{\partial r} + \frac{j(j+1)-q^2}{r^2}.$$

Proof: Note that $\mathbf{D}^2 = -H_L$ [cf. H_J but $l(l+1)$ in place of $j(j+1)$] and $\mathbf{J}^2 = \mathbf{S}^2+\mathbf{L}^2+2\mathbf{L}\cdot\mathbf{S}$. Now

$$\mathbf{L}\cdot\mathbf{S}\psi_k = \imath L_i\psi_j\epsilon_{ijk} = \imath(-\imath\epsilon_{ilm}r_lD_m - qr_i/r)\psi_j\epsilon_{ijk}$$
$$=\{-q[(\mathbf{S}\cdot\mathbf{r})/r]\psi - \mathbf{r}\mathbf{D}\cdot\psi - \psi + \mathbf{D}\psi\cdot\mathbf{r}\}_k.$$

With this in mind, (2.2) for ψ becomes

$$(\partial^0)^2\psi \equiv \mathbf{D}^2\psi + 2q\frac{\mathbf{S}\cdot\mathbf{r}}{r^3}\psi$$
$$= -H_J\psi + \left(\frac{\mathbf{S}^2+2\mathbf{L}\cdot\mathbf{S}}{r^2} + 2q\frac{\mathbf{S}\cdot\mathbf{r}}{r^3}\right)\psi$$
$$= -H_J\psi - 2(\mathbf{r}/r^2)\mathbf{D}\cdot\psi + (2/r^2)\mathbf{D}(\psi\cdot\mathbf{r}).$$

We wish to solve (2.5) in the form $H\psi=0$, $\psi=O(1/r)$, $\mathbf{D}\cdot\psi=0$ since we are interested in time-independent perturbations ψ that are small compared to $\mathbf{A}_q$ at sufficiently large r (i.e., outside the monopole core, here modeled by $\{0\}$). In fact, it will be just as easy to solve for

$$H\psi = \mathbf{D}\chi,\quad \psi=O(1/r),\quad \mathbf{D}\cdot\psi=0, \tag{2.6}$$

for some χ. The directly relevant modes are $\chi=0$. Note that $[J_i,H]=0$.

(c) In standard polar coordinates, let

$$u=\cos\theta,\quad v=\sin\theta=\sqrt{1-u^2},\quad 0\leqslant\theta\leqslant\pi.$$

Then the monopole spherical harmonics that we need[1] are

$$Y_{qlm_\pm} = 2^m e^{\imath(m\pm q)\phi}\left[\frac{2l+1}{4\pi}\,\frac{(l-m)!(l+m)!}{(l-q)!(l+q)!}\right]^{1/2}$$
$$\times(1-u)^{(m+q)/2}(1+u)^{(m-q)/2}$$
$$\times\frac{d^{l+m}}{du^{l+m}}(1-u)^{l-q}(1+u)^{l+q},$$

which can be combined with Clebsch–Gordan coefficients for the tensor product with spin 1 to obtain a complete orthonormal set of monopole vector spherical harmonics, $\mathbf{Y}_q^{jkl}$ labeled by the eigenvalues of $\mathbf{J}^2=j(j+1)$, $\mathbf{L}^2=l(l+1)$, and $J_z=k$ given simply in a standard Cartesian basis in Ref. 2,

$$\mathbf{Y}_q^{jkl} = \sum_{m=-l}^{m=l}\sum_{m'=-1}^{m'=1}\langle j,k\,|l,m;1,m'\rangle Y_{qlm}\mathbf{e}_{m'},$$
$$k\in -j,\ldots,j,\quad l\in j-1,j,j+1,\quad l\geqslant 0. \tag{2.7}$$

Here $\mathbf{e}_{\pm1}=(\hat{\mathbf{x}}\pm\imath\hat{\mathbf{y}})/\sqrt{2}$ and $\mathbf{e}_0=\hat{\mathbf{z}}$ is the spin operator eigenbasis in terms of standard Cartesian coordinates on $\mathbb{R}^3$. As a first step we change to a polar coordinate basis for the tangent space,

$$\hat{\mathbf{r}} = \begin{pmatrix}\sin\theta\cos\phi\\ \sin\theta\sin\phi\\ \cos\theta\end{pmatrix},\quad \hat{\boldsymbol{\theta}} = \begin{pmatrix}\cos\theta\cos\phi\\ \cos\theta\sin\phi\\ -\sin\theta\end{pmatrix},\quad \hat{\boldsymbol{\phi}} = \begin{pmatrix}-\sin\phi\\ \cos\phi\\ 0\end{pmatrix}.$$

In this basis (which we adopt) we obtain, after come computation,

$$\mathbf{Y}_q^{jjl=j-1} = \frac{1}{\sqrt{2}}\begin{pmatrix}-vY_{qj-1j-1}\\ -uY_{qj-1j-1}\\ -\imath Y_{qj-1j-1}\end{pmatrix},$$

$$\mathbf{Y}_q^{jjl=j} = \frac{1}{\sqrt{2(j+1)}}\begin{pmatrix}vY_{qjj-1}e^{\imath\phi}+u\sqrt{2j}Y_{qjj}\\ uY_{qjj-1}e^{\imath\phi}-v\sqrt{2j}Y_{qjj}\\ \imath Y_{qjj-1}e^{\imath\phi}\end{pmatrix}, \tag{2.8}$$

$$\mathbf{Y}_q^{jjl=j+1} = \frac{1}{\sqrt{2(2j+3)(j+1)}}\begin{pmatrix}v(-Y_{qj+1j-1}e^{\imath\phi}+e^{-\imath\phi}\sqrt{(2j+1)(j+1)}Y_{qj+1j+1})-\sqrt{2(2j+1)}uY_{qj+1j}\\ u(-Y_{qj+1j-1}e^{\imath\phi}+e^{-\imath\phi}\sqrt{(2j+1)(j+1)}Y_{qj+1j+1})+v\sqrt{2(2j+1)}uY_{qj+1j}\\ -\imath Y_{qj+1j-1}e^{\imath\phi}-\imath e^{-\imath\phi}\sqrt{(2j+1)(j+1)}Y_{qj+1j+1}\end{pmatrix}.$$

III. ORBITAL TRANSFORM AND SOLUTIONS

The key step in solving (2.6) will be the "orbital transform" Lemma 3.1. Thus let ψ denote a section with $\mathbf{J}^2 = j(j+1)$ and $J_z = k$ and write

$$\psi = \sum_{l=j-1}^{l=j+1} \mathbf{Y}_q^{jkl}(\theta,\phi) f_l(r,t) \equiv \mathbf{Y}_q^{jk} \cdot f, \tag{3.1}$$

where we think of f as a vector in the three-dimensional real vector space, "orbital space" with orthonormal basis $\{\mathbf{Y}_q^{jkl},\ l \in j-1, j, j+1\}$ (jk fixed). It will suffice to work with the highest weight $k=j$ for the general case is identical with $qjj \mapsto qjk$.

Lemma 3.1: For a spherically symmetric U(1) connection $\mathbf{D}$, (i) $\exists$ vectors a,d such that

$$\hat{\mathbf{r}} \cdot \mathbf{Y}_q^{jj} = Y_{qjj} a, \quad \mathbf{D}^{\text{angular}} \cdot \mathbf{Y}_q^{jj} = Y_{qjj} d,$$

where

$$\mathbf{D} \cdot \psi = \frac{1}{r^2} \frac{\partial}{\partial r} r^2 \hat{\mathbf{r}} \cdot \psi + \frac{1}{r} \mathbf{D}^{\text{angular}}(\theta,\phi) \cdot \psi$$

in spherical polar coordinates; (ii) explicitly for connection $\mathbf{A}_q$,

$$a = \left(\sqrt{\frac{j^2 - q^2}{j(2j+1)}}, -\frac{q}{\sqrt{j(j+1)}}, -\sqrt{\frac{(j+1)^2 - q^2}{(j+1)(j+2)}} \right),$$

$$d = \left(-\frac{(j+1)\sqrt{j^2 - q^2}}{\sqrt{j(2j+1)}}, \frac{q}{\sqrt{(j+1)j}}, -\frac{j\sqrt{(j+1)^2 - q^2}}{\sqrt{(j+1)(2j+1)}} \right)$$

and

$$a \times d = \left(\frac{q}{\sqrt{j(2j+1)}}, \frac{\sqrt{j^2 - q^2}}{\sqrt{(j+1)j}}, -\frac{q\sqrt{j^2 - q^2}}{\sqrt{((j+1)^2 - q^2)(j+1)(2j+1)}} \right) \times \sqrt{(j+1)^2 - q^2}$$

completes an orthogonal basis for the orbital space; and (iii) with χ an arbitrary function of r,

$$\hat{\mathbf{r}} Y_{qjj} = \mathbf{Y}_q^{jj} \cdot a, \quad \mathbf{D}^{\text{angular}} Y_{qjj} = \mathbf{Y}_q^{jj} \cdot (-d),$$

$$\mathbf{D}(Y_{qjj}\chi) = \mathbf{Y}_q^{jj} \cdot \left(\frac{d\chi}{dr} a - \frac{\chi}{r} d \right).$$

Proof: (i) There must exist such orbital vectors a,d because $\mathbf{D}$ and $\hat{\mathbf{r}}$ are spherically symmetric and therefore commute with $\mathbf{J}$ as explained in Sec. II(b). Therefore the left-hand sides, which are SO(3) scalars, must have each component proportional to Y_{qjj}.

(ii) The coefficients a may be found by manipulating the top entries in Eqs. (2.8) using the identities among the $\{Y_{qjk}\}$ from their definition in Sec. II (b). The identities needed are spelled out in the Appendix. Next, let $a = (a_1, a_2, a_3)$ be as in Lemma 3.1. Note $a^2 = 1$. A similar and rather tedious direct computation (see the Appendix) gives

$$-\imath v \hat{\boldsymbol{\phi}} \cdot \mathbf{Y}_q^{jj} = (a_{\perp\perp} + u a_\perp) Y_{qjj}, \quad v \hat{\boldsymbol{\theta}} \cdot \mathbf{Y}_q^{jj} = (u a_{\perp\perp} + a_\perp) Y_{qjj},$$

where

$$a_{\perp\perp} = \left(a_1, a_2, a_3 - \frac{1}{a_3} \right), \quad a_\perp = \frac{q}{j+1} \left(0, \frac{1}{a_2}, -\frac{1}{a_3} \right),$$

$$a_\perp{}^2 = a_{\perp\perp}{}^2 = \frac{j(j+1) + q^2}{(j+1)^2 - q^2},$$

$$a_\perp \cdot a_{\perp\perp} = \frac{(2j+1)q}{(j+1)^2 - q^2}, \quad a_\perp \cdot a = 0 = a_{\perp\perp} \cdot a.$$

This then yields, after further direct computation (see the Appendix),

$$d = q a_\perp - (j+1) a_{\perp\perp} = (-(j+1)a_1, -a_2, j a_3),$$

$$d^2 = j(j+1) - q^2, \quad a \cdot d = 0,$$

$$a_\perp \cdot d = -q, \quad a_{\perp\perp} \cdot d = -j,$$

so that $(a, d, q a_{\perp\perp} - j a_\perp)$ constitute an orthogonal basis. One may compute that

$$\frac{q a_{\perp\perp} - j a_\perp}{\sqrt{j^2 - q^2}} = \frac{a \times d}{\sqrt{(j+1)^2 - q^2}}.$$

(iii) From Ref. 1, acting on Y_{qjk} one has

$$D_\phi = \mp \imath q (1 \mp u) + \frac{\partial}{\partial \phi} = \imath (j \pm q) \mp \imath q + \imath q = \imath (j + q),$$

$$D_\theta = \partial_\theta = -\sqrt{1 - u^2} \frac{\partial}{\partial u}$$

$$= e^{-\imath\phi}(L_x + \imath L_y) - \frac{\imath u}{\sqrt{1 - u^2}} \frac{\partial}{\partial \phi} + q\sqrt{\frac{1 \mp u}{1 \pm u}}$$

$$= e^{-\imath\phi} L_+ + (1/v)(q + ju),$$

so that from part (ii) we find

$$\mathbf{D}^{\text{angular}} Y_{qjj} = \frac{1}{v} \begin{pmatrix} 0 \\ q + ju \\ \imath(j + qu) \end{pmatrix} Y_{qjj} = -\mathbf{Y}_q^{jj\,\text{angular}} \cdot d,$$

while

$$\hat{\mathbf{r}} Y_{qjj} \equiv \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} Y_{qjj}$$

has zero angular part. Therefore its orbital coefficients are perpendicular to $a_\perp$ and to $a_{\perp\perp}$ and thus are proportional to a. Since $a \cdot a = 1$ we see from part (i) that they are just a. Putting these together we have

$$\mathbf{D}(Y_{qjj}\chi) = \begin{pmatrix} \left(\frac{d\chi}{dr}\right) \\ \frac{\chi}{r} \frac{1}{v} \begin{pmatrix} q + ju \\ \imath(j + qu) \end{pmatrix} \end{pmatrix} Y_{qjj}$$

$$= \mathbf{Y}_q^{jj} \cdot \left(a \frac{d}{dr} \chi - \frac{\chi}{r} d \right).$$

This completes the sketch of the computations leading to Lemma 3.1.

Using this lemma the equation to be solved, (2.6), for f

according to (3.1) and $\psi_0\equiv 0$, is

$$\Big(-(\partial^0)^2+\frac{\partial^2}{\partial r^2}+\frac{2}{r}\frac{\partial}{\partial r}-\frac{\tilde J}{r^2}\Big)f+4\frac{a}{r^2}a\cdot f$$
$$=\Big(a\frac{d}{dr}-\frac{d}{r}\Big)\Big(\chi-2\frac{f\cdot a}{r}\Big) \tag{3.2}$$

and the divergence constraint is

$$\frac{1}{r^2}\frac{d}{dr}r^2 f\cdot a+\frac{1}{r}d\cdot f=0.$$

We employ the notation

$$j(j+1)-q^2\equiv \tilde J\equiv j'(j'+1),$$

say, and we solve the equation $-(\partial^0)^2\psi=\mathbf{D}\chi(t,r,\theta,\phi)$ for some $\chi(t,r,\theta,\phi)$ that must be of the form $\chi(t,r,\theta,\phi)=\chi(t,r)Y_{qjj}(\theta,\phi)$ since H in (2.6) commutes with J_i. In the time-independent case that we consider first, the equation is homogeneous in r so it is convenient to write

$$f(t,r)=f(t)r^\alpha,\quad \chi(t,r)=\chi(t)r^{\alpha-1},\quad \tilde\alpha=\alpha(\alpha+1).$$

Then

$$(\tilde\alpha-\tilde J)f+4f\cdot aa=((\alpha-1)a-d)(\chi-2f\cdot a), \tag{3.3}$$
$$(\alpha+2)f\cdot a+f\cdot d=0, \tag{3.4}$$

which can now be solved by elementary linear algebra. Thus, contracting (3.3) with a and with d and then cross-multiplying we have, respectively,

$$(\tilde\alpha-\tilde J+4)f\cdot a=(\alpha-1)(\chi-2f\cdot a)$$
$$\Leftrightarrow(\alpha-1)\chi=(\alpha\mp 1-\tilde J)f\cdot a, \tag{3.5}$$
$$(\tilde\alpha-\tilde J)f\cdot d=(-\tilde J)(\chi-2f\cdot a), \tag{3.6}$$
$$(\chi-2f\cdot a)((\tilde\alpha-\tilde J+4)(-\tilde J)+(\tilde\alpha-2)(\tilde\alpha-\tilde J))=0, \tag{3.7}$$

so that we classify the solutions as follows.

For type I $[(\chi-2f\cdot a)=0]$, if $f\cdot a\neq 0$ (3.5) implies $\tilde\alpha-\tilde J+4=0$ so (3.6) implies $f\cdot d=0$. In this case (3.4) implies $\alpha=-2$. The solution is therefore

$$f=\lambda a,\quad \alpha=-2,\quad \chi=2\lambda,\quad \tilde J=6, \tag{3.8}$$

which will be included among type III below.

If $f\cdot a=0$ then (3.4) implies $f\cdot d=0$, and (3.3) implies $\tilde\alpha-\tilde J=0$ and from Lemma 3.1 (ii) we have the solution of type I:

$$f=\lambda(qa_{\perp\perp}-ja_\perp)/\sqrt{j^2-q^2}=\lambda b,$$
$$\chi=0,\quad \alpha=-\tfrac12-\sqrt{j(j+1)-q^2+\tfrac14}=-j'-1, \tag{3.9}$$
$$j=q,q+1,\ldots,$$

where only the $-ve$ roots of $\tilde\alpha=\tilde J$ are given in accordance with our decaying boundary conditions. Note that the $j=q$ mode exists only for $j>\frac12$ as $\mathbf{Y}_q^{jkl}$ is defined to be zero for $l<0$.

For types II and III,

$$(\tilde\alpha-\tilde J+4)(-\tilde J)+(\tilde\alpha-2)(\tilde\alpha-\tilde J)$$
$$=0\Leftrightarrow\tilde\alpha=\tilde J+1\pm\sqrt{4\tilde J+1}$$
$$=j'(j'+1)\pm(2j'+1). \tag{3.10}$$

Here $f\cdot a\neq 0$ since $f\cdot a=0$ implies $f\cdot d=0$ by (3.4); hence $f=0$ by (3.3) as $\tilde\alpha-\tilde J\neq 0$ for (3.10). Therefore (3.5) implies $\alpha\neq 1$ and thus by writing $\chi-2f\cdot a=\lambda(\tilde\alpha-\tilde J)$, (3.4) gives a solution proportional to (3.8) again if $\alpha=2$ or else

$$f=-\lambda d+\lambda\frac{\tilde J}{\alpha+2}a,$$
$$\chi=\lambda\frac{(\alpha\mp 1-\tilde J)\tilde J}{(\alpha-1)(\alpha+2)},\quad j=q,q+1,\ldots,$$

with α as in (3.10). The choice of the minus sign in (3.10) gives $\alpha\mp 1=\tilde J$, which we call solutions of type II:

$$f=-\lambda d+\lambda(\alpha+1)a,\quad \chi=0,$$
$$\alpha=-j'-2,\quad j=q,q+1,\ldots, \tag{3.11}$$

where only the $-ve$ roots of $\alpha\mp 1=\tilde J$ are taken in accordance with our chosen decaying boundary conditions.

For type III, these are the remaining solutions corresponding to the $+ve$ root in (3.10), which, in fact, include (3.8) by rescaling λ:

$$f=-\lambda(\alpha+2)d+\lambda\tilde J a,$$
$$\chi=\lambda[(\alpha\mp 1-\tilde J)\tilde J/(\alpha-1)], \tag{3.12}$$
$$\alpha=-\tfrac12-\sqrt{(j'+1)(j'+2)-\tfrac34},\quad j=q,q+1,\ldots.$$

The physical significance of these three types of solution to (2.6) is as follows. Type I, being orthogonal to a and d, is, by Lemma 3.1, tangential to the sphere at radius r and not of the form $\mathbf{D}\chi$. These type I are the physical non-Abelian fluctuations about $\mathbf{A}_q$. Type II are precisely the solutions of the form $\mathbf{D}\chi$ since, according to Lemma 3.1 (iii),

$$\mathbf{D}r^{\alpha+1}Y_{qjj}=\mathbf{Y}_q^{jj}\cdot(-d+(\alpha+1)a),$$

i.e., tangent to the action of time-independent gauge transformations of the form $-\imath\chi\sigma_+/2-\imath\chi^*\sigma_-/2$ corresponding to (1.5). They correspond to the infinitesimal "Gribov ambiguity" expected at the reducible connection $\mathbf{A}_q$. Here only $\chi\propto r^{\alpha+1}Y_{qjj}$ needed to be considered as $[J_i,D_j]=0$, see (2.4). Explicitly,

$$-d+(\alpha+1)a=\Big(\frac{\sqrt{j^2-q^2}(j-j')}{\sqrt{j(2j+1)}},\frac{-j'q}{\sqrt{j(j+1)}},\frac{(j+j'+1)\sqrt{(j+1)^2-q^2}}{\sqrt{(j+1)(2j+1)}}\Big).$$

Finally, the type III solutions are the non-Abelian modes that preserve the Yang–Mills equation only up to a gauge transform,

$$\mathbf{D}_{\mathbf{A}_q+\delta\mathbf{A}}\cdot F(\mathbf{A}_q+\delta\mathbf{A})=\mathbf{D}_{\mathbf{A}_q}\chi.$$

They do not correspond to anything physical but rather to points that are critical points of the Yang–Mills action restricted to $\Sigma_{\mathbf{A}_q}$, since any fluctuations, $\delta\mathbf{A}'$, that lie on $\Sigma_{\mathbf{A}_q}$, are restricted to obey $\mathbf{D}\cdot\delta\mathbf{A}'=0$. Hence they give zero change in the action at configurations $\mathbf{A}_q+\delta\mathbf{A}$ that only obey the Yang–Mills equations up to a gauge transform (assuming integration by parts.) Rather, these modes correspond to an ambiguity in a coupled source $\mathbf{J}$ when, in the background field method,[10] a functional Legendre transform exchanges $\mathbf{J}$ for $\mathbf{A}$ as the fundamental variable. They are therefore somewhat dual to the infinitesimal "Gribov ambiguity" modes of type II, and it is interesting that they arise precisely as the opposite root of the indicial equation (3.10) to type II.

The $q=0$ limit of the type I solution is

$$\psi = \mathbf{Y}_q^{jj}\cdot(0, j\sqrt{j/(j+1)}, 0)r^{-j-1}, \quad j>0, \tag{3.13}$$

and of type II,

$$\psi = \mathbf{Y}_q^{jj}\cdot(0,0,\sqrt{(j+1)(2j+1)})r^{-j-2}, \quad j\geqslant 0. \tag{3.14}$$

Here the $j=0$ mode is the solution $\hat{\mathbf{r}}/r^2$ familiar in a different context.

Since $[J_i, H]=0$ from Sec. II(b), we have without computation that all the lower weight modes are obtained from the highest weights given above by applying $J_- = J_x - \imath J_y$ and hence are obtained from the above simply by replacing Y_{qjj} and $\mathbf{Y}_q^{jj}$ by Y_{qjk} and $\mathbf{Y}_q^{jk}$, respectively, where $k\in -j,...,j$. Lemma 3.1 also holds in this greater generality (with the same vectors a,b,d) in view of (2.4).

IV. ALTERNATIVE SOLUTION OF LINEARIZED YANG–MILLS FLUCTUATIONS: CONCLUSIONS

As an independent check of the computations of Sec. II and the Appendix, we shall give an alternative less explicit solution of (2.6) for $\chi=0$ and indicate that it agrees with the above. It is modeled on the standard treatment of the $q=0$ case (Ref. 3, Sec. 16) made possible by the following lemma.

Lemma 4.1: (i) Let $\mathbf{A}$ be a connection on a principle bundle P and $\mathbf{D}$ its associated covariant derivative and $\Box_{\mathbf{A}} = \mathbf{D}\cdot\mathbf{D}$. Suppose that $\mathbf{A}$ obeys the Yang–Mills equations. For all smooth sections χ of ad P we have

$$(\Box_{\mathbf{A}} + 2F)\mathbf{D}\chi = \mathbf{D}(\Box_{\mathbf{A}}\chi), \tag{4.1}$$

where F denotes the curvature acting by matrix multiplication with commutator in the Lie-algebra values (the adjoint representation).

(ii) For the $\mathbf{A}_q$ of (1.3) or more generally in $n=3$ dimensions if $\mathbf{D}\cdot F=0$ and $x^i\epsilon_{ij[k}F_{l]}{}^j=0$ (related to rotational invariance) we have (with $\Box_{\mathbf{A}}$ denoting $\mathbf{D}\cdot\mathbf{D}$),

$$(\Box_{\mathbf{A}} + 2F)\cdot(\mathbf{r}\times\mathbf{D})\chi = \mathbf{r}\times\mathbf{D}(\Box_{\mathbf{A}}\chi), \quad \forall\chi. \tag{4.2}$$

(iii) Because of parts (i) and (ii), if χ is an eigenvector of $\Box_{\mathbf{A}}$ then $\mathbf{D}\chi$ and $\mathbf{r}\times\mathbf{D}\chi$ are eigenvectors of $\Box_{\mathbf{A}} + 2F$ [the operator in (2.2)]. The zero modes obey

$$\mathbf{D}\cdot(\mathbf{D}\chi) = \Box_{\mathbf{A}}\chi = 0, \quad \mathbf{D}\cdot(\mathbf{r}\times\mathbf{D}\chi) = -\mathbf{r}\cdot\mathbf{B}\chi = 0, \tag{4.3}$$

provided $\mathbf{r}\cdot\mathbf{B}\equiv x^i\tfrac{1}{2}\epsilon_{ijk}F^{jk}=0$ as is the case for $\mathbf{A}_q$.

Proof: We have

$$\begin{aligned}
(\Box_{\mathbf{A}}\delta_\mu{}^\nu &+ 2F_\mu{}^\nu)D_\nu\chi \\
&\equiv D^\nu D_\nu D_\mu\chi + [2F_\mu{}^\nu, D_\nu\chi] \\
&= D^\nu[F_{\nu\mu},\chi] + D^\nu D_\mu D_\nu\chi + 2[F_\mu{}^\nu, D_\nu\chi] \\
&= D_\mu\Box_{\mathbf{A}}\chi + [F^\nu{}_\mu, D_\nu\chi] + D^\nu[F_{\nu\mu},\chi] + 2[F_\mu{}^\nu, D_\nu\chi] \\
&= D_\mu\Box_{\mathbf{A}}\chi + [D\cdot F_\mu,\chi].
\end{aligned}$$

Similarly for part (ii), and not explicitly writing the commutators of the adjoint representation acting on χ,

$$\begin{aligned}
(\Box_{\mathbf{A}}\delta_i{}^j &+ F_i{}^j)\epsilon_{jkl}x^k D^l \\
&= D\cdot D\epsilon_{ikl}x^k D^l + 2F_i{}^m\epsilon_{mkl}x^k D^l \\
&= D^n(\epsilon_{inl}D^l + \epsilon_{ikl}x^k F_n{}^l) \\
&\quad + D^n\epsilon_{ikl}x^k D^l D_n + 2F_i{}^m\epsilon_{mkl}x^k D^l \\
&= (\mathbf{r}\times\mathbf{D})_i\Box_{\mathbf{A}} + \epsilon_{ikl}x^k(\mathbf{D}\cdot F)^l + 2x^k\epsilon_{kl[j}F_{n]}{}^l D^n.
\end{aligned}$$

Note that $\Box_{\mathbf{A}} + 2F$ does *not* commute in this sense with the orbital angular momentum $\mathbf{L}$ defined in Sec. II(b), except for $q=0$.

Now the spectrum of $\Box_{\mathbf{A}}$ on scalar sections χ was found long ago. The results stated in Ref. 1 for the zero modes on $\mathbb{R}^3-0$ are

$$\chi = r^{-j'-1}Y_{qjk},$$

where

$$j'(j'+1) = j(j+1) - q^2, \quad j'\geqslant 0.$$

Hence, according to Lemma 4.1, the general solution to (2.6) with $\chi=0$ is

$$\psi\in\{\mathbf{r}\times\mathbf{D}r^{-j'-1}Y_{qjk}\}\oplus\{\mathbf{D}r^{-j'-1}Y_{qjk}\}.$$

Comparing this with the results of Sec. III we see that we can conclude as an addendum to Lemma 3.1 that

$$\mathbf{r}\times\mathbf{D}Y_{qjj}\propto\mathbf{Y}_q^{jj}\cdot(a\times d).$$

One may perform some explicit computations to fix the constants of proportionality and check the inverse $(\mathbf{r}\times\mathbf{D})\cdot\mathbf{Y}_q^{jj}\propto Y_{qjj}(a\times d)$, but this is to be expected in view of parts (i) and (ii) of Lemma 3.1.

The time-dependent solutions, i.e., writing $\psi = e^{\imath\omega t}\psi(\mathbf{r})$, the positive eigenvalue modes of $\Box_{\mathbf{A}} + 2F$, are obtained simply by replacing $r^{-j'-1}$ by suitable combinations of spherical Bessel and Neumann functions. For example, $j_{j'}(\omega r)/\sqrt{\omega t}$ in the type I solutions. This and an analysis of scattering can proceed almost in analogy with the $q=0$ case (Ref. 3, Sec. 16), but shifted by an irrational amount by using j' in place of j. This is a topic of further work.

V. CONCLUSIONS

It is concluded that the solution of the SU(2) Yang–Mills equation on $\mathbb{R}^3-0$, and other differential equations on sections associated to the monopole bundle, can be developed along the lines of the familiar $q=0$ case, with some essential differences due to the effective angular momentum j' being irrationally shifted from the total angular momentum j (as in the scalar case) and due to the associated orbital angular momentum mixing. In addition to scalar and vector monopole spherical harmonics, spin-$\frac{1}{2}$ sections have also been of interest, e.g., Ref. 11, and further work might include explicit partial waves for this situation also.

The explicit Lemmas 3.1 and 4.1 are intended as a useful explicit complement to the alternative highly abstract approach to monopole dynamics. The important point is that for charged fields in background (external) fields, there is, in general, no Poincaré group. But when the background field has symmetry under a subgroup, then that subgroup can usefully be applied in analogy with the uncharged case.

ACKNOWLEDGMENT

This work was supported in part by the Herchel Smith Scholarship of Emmanuel College, Cambridge, England.

APPENDIX: ADDITIONAL COMPUTATIONS FOR SEC. III

(a) Since we know from (2.4) that $\hat{\mathbf{r}}\cdot\mathbf{Y}_q^{jj}$, $v\hat{\boldsymbol{\theta}}\cdot\mathbf{Y}_q^{jj}$, and $v\hat{\boldsymbol{\phi}}\cdot\mathbf{Y}_q^{jj}$ are of the form $Y_{q\,jj}\,a$, etc., we have to use identities among the $\{Y_{q\,jk}\}$ to put the scalar coordinate expressions (3.1) explicitly into this form. A general expression for the needed identities is given in the Appendix of Ref. 2. Rather than develop an abstract theory, it is perhaps simplest for our limited purpose just to list the needed identities for the convenience of reference while checking the answer. Thus

$$uY_{q\,j-1j-1}=-\frac{q}{j}Y_{q\,j-1j-1}+\frac{\sqrt{j^2-q^2}}{j\sqrt{2j+1}}Y_{q\,jj-1},$$

$$-ve^{i\phi}Y_{q\,j-1j-1}=\frac{\sqrt{j^2-q^2}}{j\sqrt{2j+1}}Y_{q\,jj},$$

$$-ve^{i\phi}Y_{q\,jj-1}=q\frac{\sqrt{2j}}{j(j+1)}Y_{q\,jj}+\frac{\sqrt{2j}\sqrt{(j+1)^2-q^2}}{\sqrt{2j+3}(j+1)}Y_{q\,j+1j},$$

$$-ve^{i\phi}Y_{q\,jj}=\frac{\sqrt{2j+2}\sqrt{(j+1)^2-q^2}}{\sqrt{2j+3}(j+1)}Y_{q\,j+1j+1},$$

$$uY_{q\,jj}=-\frac{q}{j+1}Y_{q\,jj}+\frac{\sqrt{(j+1)^2-q^2}}{\sqrt{2j+3}(j+1)}Y_{q\,j+1j},$$

$$-ve^{i\phi}Y_{q\,j+1j-1}=q\frac{\sqrt{2}\sqrt{2j+1}}{(j+1)(j+2)}Y_{q\,j+1j}+\frac{\sqrt{2j+1}\sqrt{2j+2}\sqrt{(j+2)-q^2}}{\sqrt{2j+3}\sqrt{2j+5}(j+2)}Y_{q\,j+2j}-\frac{\sqrt{2}\sqrt{(j+1)^2-q^2}}{\sqrt{2j+1}\sqrt{2j+3}(j+1)}Y_{q\,jj},$$

$$-ve^{-i\phi}Y_{q\,j+1j+1}=q\frac{\sqrt{2}}{\sqrt{j+1}(j+2)}Y_{q\,j+1j}-\frac{\sqrt{2}\sqrt{(j+2)-q^2}}{\sqrt{2j+3}\sqrt{2j+5}(j+2)}Y_{q\,j+2j}+\frac{\sqrt{2}\sqrt{(j+1)^2-q^2}}{\sqrt{2j+3}\sqrt{j+1}}Y_{q\,jj},$$

$$uY_{q\,j+1j}=\frac{-qj}{(j+1)(j+2)}Y_{q\,j+1j}+\frac{\sqrt{2j+2}\sqrt{2}\sqrt{(j+2)-q^2}}{\sqrt{2j+3}\sqrt{2j+5}(j+2)}Y_{q\,j+2j}+\frac{\sqrt{(j+1)^2-q^2}}{\sqrt{2j+3}(j+1)}Y_{q\,jj},$$

$$-ve^{-i\phi}Y_{q\,j+1j}=q\frac{\sqrt{2j+1}\sqrt{2}}{(j+2)(j+1)}Y_{q\,j+1j-1}-\frac{\sqrt{2}\sqrt{3}\sqrt{(j+2)-q^2}}{\sqrt{2j+3}\sqrt{2j+5}(j+2)}Y_{q\,j+2j-1}.$$

Using these various identities to find $\hat{\mathbf{r}}\cdot\mathbf{Y}_q^{jj}$ from (2.8) is easy. For the other two spatial components of $\mathbf{Y}_q^{jj}$, the computation is also straightforward so I shall just pause at key steps as an aid in checking:

$$\begin{aligned}v\hat{\boldsymbol{\theta}}\cdot\mathbf{Y}_q^{jj}=\Bigg(&\frac{u\sqrt{2j}\sqrt{j^2-q^2}}{\sqrt{2}\sqrt{2j+1}j}Y_{q\,jj},\frac{-uq\sqrt{2j}}{\sqrt{2j+2}j(j+1)}Y_{q\,jj}-\frac{\sqrt{2j}}{\sqrt{2}\sqrt{j+1}}Y_{q\,jj}+\frac{u\sqrt{2j}(-q)j}{\sqrt{2}\sqrt{j+1}j(j+1)}Y_{q\,jj},\\&-\frac{u\sqrt{(j+1)^2-q^2}}{\sqrt{2j+1}\sqrt{j+1}(2j+3)(j+1)}Y_{q\,jj}-\frac{u\sqrt{2j+1}\sqrt{(j+1)^2-q^2}}{\sqrt{2j+3}\sqrt{2j+5}\sqrt{j+1}}Y_{q\,jj}\\&-\frac{u\sqrt{2j+1}\sqrt{(j+1)^2-q^2}}{(2j+3)\sqrt{j+1}(j+1)}Y_{q\,jj}+\frac{\sqrt{2j+1}}{\sqrt{2j+3}\sqrt{j+1}}Y_{q\,j+1j}\Bigg),\end{aligned}$$

where eight similar intercanceling terms have been omitted. Defining a as in the statement of Lemma 3.1, we have

$$\begin{aligned}v\hat{\boldsymbol{\theta}}\cdot\mathbf{Y}_q^{jj}&=\Bigg(ua_1Y_{q\,jj},ua_2Y_{q\,jj}-\frac{\sqrt{j}}{\sqrt{j+1}}Y_{q\,jj},\frac{\sqrt{2j+1}}{\sqrt{2j+3}\sqrt{j+1}}Y_{q\,j+1j}\\&\quad-\frac{u\sqrt{(j+1)^2-q^2}\sqrt{2j+1}}{(2j+3)\sqrt{j+1}}Y_{q\,jj}\left(\frac{1}{j+1}+1+\frac{1}{(j+1)(j+2)}\right)\Bigg)\\&=\left('','',\left(u+\frac{q}{j+1}\right)Y_{q\,jj}\frac{\sqrt{j+1}\sqrt{2j+1}}{\sqrt{(j+1)^2-q^2}}-\frac{u\sqrt{(j+1)^2-q^2}2}{\sqrt{2j+1}\sqrt{j+1}}Y_{q\,jj}\right)\\&=\left(ua_1,ua_2-\frac{\sqrt{j}}{\sqrt{j+1}},\left(u+\frac{q}{j+1}\right)\left(-\frac{1}{a_3}\right)+ua_3\right)Y_{q\,jj}=(ua_{\perp\perp}+a_\perp)Y_{q\,jj},\end{aligned}$$

as required in the proof of Lemma 3.1(ii). Similarly

$$
\begin{aligned}
-\imath v\hat{\boldsymbol{\phi}}\cdot\mathbf{Y}^{jj}_q &= \Bigg(\frac{\sqrt{j}\sqrt{j^2-q^2}}{\sqrt{2j+1}j}\,Y_{qjj},\ -\frac{q\sqrt{2j}}{\sqrt{2}\sqrt{j+1}j(j+1)}\,Y_{qjj}-\frac{\sqrt{2j}\sqrt{(j+1)^2-q^2}}{\sqrt{2}\sqrt{j+1}\sqrt{2j+3}(j+1)}\,Y_{qj+1j},\\
&\qquad \frac{1}{\sqrt{2}\sqrt{2j+3}\sqrt{j+1}}\Bigg[\frac{q\sqrt{2j+1}\sqrt{2}}{(j+1)(j+2)}\,Y_{qj+1j}-\frac{\sqrt{2}\sqrt{(j+1)^2-q^2}}{\sqrt{2j+1}\sqrt{2j+3}(j+1)}\,Y_{qjj}\\
&\qquad +\frac{\sqrt{2j+1}\sqrt{2}q}{(j+2)}\,Y_{qj+1j}+\frac{\sqrt{2j+1}\sqrt{2}\sqrt{(j+1)^2-q^2}}{\sqrt{2j+3}}\,Y_{qjj}\Bigg]\Bigg)\\
&= \Bigg(a_1Y_{qjj},\ \frac{-\sqrt{j}\sqrt{(j+1)^2-q^2}}{\sqrt{j+1}\sqrt{2j+3}(j+1)}\,Y_{qj+1j}+\frac{a_2}{j+2}\,Y_{qjj},\\
&\qquad \frac{q\sqrt{2j+1}}{(j+1)\sqrt{j+1}\sqrt{2j+3}}\,Y_{qj+1j}+\frac{\sqrt{(j+1)^2-q^2j}}{\sqrt{2j+1}\sqrt{j+1}(j+1)}\,Y_{qjj}\Bigg)\\
&= \left(a_1,\ \frac{a_2}{j+1}-\frac{\sqrt{j}}{\sqrt{j+1}}\left(x+\frac{q}{j+1}\right),\ -\frac{j}{j+1}a_3-\frac{q}{j+1}\,\frac{(u+q/(j+1))}{a_3}\right)Y_{qjj}\\
&= \left(a_1, a_2-\frac{\sqrt{j}}{j+1}u,\ -\frac{q}{j+1}\frac{u}{a_3}+a_3-\frac{1}{a_3}\right)Y_{qjj}=(ua_\perp+a_{\perp\perp})Y_{qjj}.
\end{aligned}
$$

(b) Using the results of part (a) and Ref. 1,

$$
-v\frac{\partial}{\partial u}=e^{\imath\phi}L_+ +\frac{(q+ju)}{v}=\frac{(q+ju)}{v}
$$

acting in Y_{qjj}, we find in spherical polars

$$
\begin{aligned}
\mathbf{D}^{\text{angular}}\cdot\mathbf{Y}^{jj}_q &= -\frac{\partial}{\partial u}v\hat{\boldsymbol{\theta}}\cdot\mathbf{Y}^{jj}_q+\frac{\imath(j+qu)}{v}\hat{\boldsymbol{\phi}}\cdot\mathbf{Y}^{jj}_q\\
&= -\frac{\partial}{\partial u}(ua_{\perp\perp}+a_\perp)Y_{qjj}-(1-u^2)(j+qu)(a_{\perp\perp}+ua_\perp)Y_{qjj}\\
&= \left(-a_{\perp\perp}+\frac{(q+ju)}{1-u^2}(ua_{\perp\perp}+a_\perp)-(1-u^2)(j+qu)(a_{\perp\perp}+ua_\perp)\right)Y_{qjj}\\
&= ((-a_{\perp\perp}+qa_\perp-ja_{\perp\perp})+O(u))Y_{qjj},
\end{aligned}
$$

where the terms labeled $O(\mathrm{u})$ must cancel among themselves in view of (2.4).

[1]T. T. Wu and C. N. Yang, Nucl. Phys. B **107**, 365 (1976).

[2]R. A. Brandt and F. Neri, "Stability analysis for singular non-Abelian monopoles," Nucl. Phys. B **161**, 253 (1979).

[3]J. D. Jackson, *Classical Electrodynamics* (Wiley, New York, 1975), 2nd ed.

[4]D. R. Stump, "Instantaneous Coulomb interaction for a long-range background field in quantum chromodynamics," Phys. Rev. D **20**, 1965 (1979).

[5]S. Coleman, "Magnetic Monopoles 50 years later," in *Lectures at the International School of Physics* (Ettore Majorana, Erice, 1975).

[6]See, for example, L. Soffer, "Long-range scattering in non-Abelian gauge theories," Phys. Rev. D **29**, 1866 (1984).

[7]S. Helgason, *Differential Geometry, Lie Groups and Symmetric Spaces* (Academic, New York, 1978).

[8]D. Freed and K. Uhlenbeck, *Instantons and Four-Manifolds* (Springer, New York, 1984).

[9]R. S. Ward, "Magnetic monopoles with gauge group SU(3) broken down to U(2)," Phys. Lett. B **107**, 281 (1981).

[10]L. F. Abbott, "The background-field method beyond one loop," Nucl. Phys. B **185**, 189 (1981); G. M. Shore, "Symmetry restoration and the background field method in gauge theories," Ann. Phys. (NY) **137**, 262 (1981).

[11]A. N. Skellekens, "Fermions and spherically symmetric monopoles," Nucl. Phys. B **246**, 494 (1984).

Double-solution family of self-dual SU(2) gauge fields

Zai-Zhe Zhong
Department of Physics, Liaoning Normal University, Dalian, People's Republic of China

(Received 17 August 1988; accepted for publication 23 November 1988)

The method of Witten [Phys. Rev. D **19**, 718 (1979)] is extended to a double-complex form. Many new solutions of the self-dual SU(2) gauge fields are generated by the double-inverse scattering method, and the soliton solutions can form an infinite network.

I. INTRODUCTION

In this paper, the double-complex function and the double-inverse scattering method[1,2] are used to generate the new solutions of the self-dual SU(2) gauge field (SDSGF).[3] According to Witten,[4] some static axisymmetric SDSGF equation can be changed into the Ernst equation.[5] In addition, Letelier[6] had used the inverse scattering method of Belinsky and Zakharov[7] to find the soliton solutions of the SDSGF. However, since in these methods only the ordinary complex numbers (with the imaginary unit $i, i^2 = -1$) are used, the double-complex duality symmetry is hidden. If we use this symmetry, many new and more complex solutions of the SDSGF are found.

According to Ref. 4, finding a solution of some static $(\partial/\partial x_4 \equiv 0)$ axisymmetric (about x_3) SDSGF on Euclidean space (x_1,x_2,x_3,x_4) and in the R gauge is changed into finding an ordinary complex Ernst potential $\mathscr{E}_c = \phi + i\sigma$, where both ϕ and σ are functions of x_1,x_2 only. Once $\mathscr{E}_c$ has been obtained, then the potential $\mathbf{b}_\mu$ of SDSGF is determined by the equations

$$\begin{aligned}
&\phi \mathbf{b}_{x_1} = (\sigma_{x_2}, \sigma_{x_1}, -\phi_{x_1}),\\
&\phi \mathbf{b}_{x_2} = (-\sigma_{x_1}, \sigma_{x_2}, \phi_{x_1}),\\
&\phi \mathbf{b}_{x_3} = (0, \sigma_{x_3}, 0),\\
&\phi \mathbf{b}_{x_4} = (\sigma_{x_3}, 0, -\phi_{x_3}), \quad \sigma_{x_i} \equiv \frac{\partial \sigma}{\partial x_i} \quad (i = 1,2,3).
\end{aligned} \tag{1}$$

In Sec. II, the concept of double-complex duality symmetry is introduced. A number of applications are illustrated in Sec. III.

II. DOUBLE-COMPLEX DUALITY SYMMETRY

Let J denote the double imaginary unit, i.e., $J = i(i^2 = -1)$ or $J =$ the hyperbolic imaginary unit ϵ $(\epsilon^2 = +1, \epsilon \neq \pm 1)$. Here $a(J) = \Sigma_{n=0}^{\infty} a_n J^{2n}$ is called a double-real number, if the real series Σa_n is absolutely convergent. Let $\mathscr{E}(J) = F(J) + J\cdot\Omega(J)$ be a double-complex Ernst potential, where $F(J) = F(r,z;J)$ and $\Omega(J) = \Omega(r,z,;J)$ are double-real functions of r and z, and (r,z,θ) are the cylindrical coordinates

$$\begin{aligned}
&x_1 = r\cos\theta, \quad x_2 = r\sin\theta,\\
&r = (x_1^2 + x_2^2)^{1/2}, \quad \tan\theta = x_2/x_1.
\end{aligned} \tag{2}$$

The double-complex Ernst equation is[1,2]

$$\mathrm{Re}(\mathscr{E}(J))\nabla^2\mathscr{E}(J) = \nabla\mathscr{E}(J)\cdot\nabla\mathscr{E}(J), \tag{3}$$

where the symbol Re denotes the real part of a double-complex function. This equation and its solutions are discussed in Refs. 1 and 2.

If a double solution $\mathscr{E}(J)$ of Eq. (3) is given, then $\mathscr{E}_c = \mathscr{E}(J = i)$ is an ordinary complex Ernst potential. According to Witten,[4] a solution (ϕ,σ) of SDSGF is determined by $\mathscr{E}_c$, i.e.,

$$(\phi,\sigma) = (F_c,\Omega_c) \equiv (F(J=i), \Omega(J=i)). \tag{4}$$

Now, we prove that there is yet another solution $(\hat\phi,\hat\sigma)$ of the SDSGF that is determined by

$$\mathscr{E}_H = F_H + \epsilon\cdot\Omega_H \equiv F(J=\epsilon) + \epsilon\cdot\Omega(J=\epsilon)$$

as follows. Let T and W_ϕ be the Kramer–Neugebauer transformations

$$\begin{aligned}
&T\colon\ \phi \to T(\phi) = r\phi^{-1},\\
&W_\phi\colon\ \sigma \to \omega = W_\phi(\sigma) = \int r^{-1}\phi^2(\sigma_x\, dr - \sigma_r\, dz),\\
&\text{i.e.,}\ \partial_r\omega = r^{-1}\phi^2\,\partial_z\sigma, \quad \partial_z\omega = -r^{-1}\phi^2\,\partial_r\sigma.
\end{aligned} \tag{5}$$

Therefore $\mathscr{D}_c = T(F_H) + i\cdot W_{F_H}(\Omega_H)$ is[1] also an ordinary complex Ernst solution of Eq. (3) with $J = i$. This means that we obtain a new solution of SDSGF, i.e.,

$$(\hat\phi,\hat\sigma) = (T(F_H), W_{F_H}(\Omega_H)). \tag{6}$$

The relations among the solutions of the SDSGF, the double-complex Ernst equation, and the stationary axisymmetric gravitation field solutions (f,ω), which correspond to the line element

$$ds^2 = f(dt - \omega\, d\theta)^2 - f^{-1}[e^\gamma(dr^2 + dz^2) + r^2\, d\theta^2], \tag{7}$$

can be expressed as in the following configuration:

$$\begin{tikzcd}
(f,\omega) \arrow[d, dash] & (F_c,\Omega_c) \arrow[l, "{(1,V)}"'] \arrow[r, "{(1,1)}"] \arrow[d, dash] & (\phi,\sigma) \arrow[d, dash] \\
(\hat f,\hat\omega) & (F_H,\Omega_H) \arrow[l, "{(T,1)}"'] \arrow[r, "{(T,W)}"] & (\hat\phi,\hat\sigma),
\end{tikzcd} \tag{8}$$

where $V = W^{-1}$, 1 is the identity transformation, and the symbol (,) denotes a transformation pair, e.g.,

$$(1,V)\colon\ (f,\omega) \to (1(f), V(\omega)) = (F_c,\Omega_c),$$

etc. Evidently, the result of Witten only corresponds to the upper line in Eq. (8). Since $TT^{-1} = VW = 1$, the transformation from $(\hat f,\hat\omega)$ to $(\hat\phi,\hat\sigma)$ is essentially the same as from (f,ω) to (ϕ,σ). In general relativity, we generally select the asymptotically flat solutions; however, this restriction is not

 0022-2488/89/051158-03$02.50

necessary for the SDSGF. Thus, in SDSGF theory, the above method is more effective than those in general relativity.

III. APPLICATIONS

In the following, we give some examples of how to find new solutions of SDSGF.

(A) Let $E(J) = (1 + \mathscr{E}(J))/(1 - \mathscr{E}(J))$, thus Eq. (3) is changed into

$$(EE^* - 1)\nabla^2 E = 2E^*\nabla E\cdot\nabla E \quad (* = \text{conjugation}) . \tag{9}$$

When $J = i$, this equation has the Tomimatsu–Sato[8] solution series $E_c(\delta)$

$$E_c(\delta) = \frac{\alpha(\xi,\eta,p,q;\delta)}{\beta(\xi,\eta,p,q;\delta)} \quad (\delta = 1,2,3,...) , \tag{10}$$

where α and β are certain polynomials, p and q are real parameters such that $p^2 + q^2 = 1$, and (ξ,η) are prolate spheroidal coordinates, related to (r,z) by $r = c\sqrt{(\xi^2 - 1)(1 - \eta^2)}$, $z = c\xi\eta$. This solution series has been extended by us[1] to a double solution series $E(J,\delta)$,

$$E(\xi,\eta;J,\delta) = \frac{A(\xi,\eta,C[J\lambda],S[J\lambda];\delta)}{B(\xi,\eta,C[J\lambda],S[J\lambda];\delta)} , \tag{11}$$

where double-cosine

$$C[J\lambda] = \sum_{n=0}^{\infty} (1/2n!)(J^2\lambda)^n ,$$

double-sine

$$S[J\lambda] = \sum_{n=0}^{\infty} [1/(2n+1)!](J^2\lambda)^n\lambda ,$$

$$p = C[i\lambda] = \cos\lambda ,$$

and

$$q = S[i\lambda] = \sin\lambda .$$

The polynomials A and B are generated from α and β, respectively, by a simple algebraic substitution

$$[p,q^{2n},iq^{2n-1}] \rightarrow [p,(-1)^n(Jq)^{2n},(-1)^{n-1}(Jq)^{2n-1}] . \tag{12}$$

From $E_c(\delta)$ the results of Witten are obtained, however, $(\hat{\phi},\hat{\sigma})$ corresponding to $E_H(\delta)$ are new solutions. For $\delta = 1$,

$$\begin{aligned} &E(J) = C[J\lambda]\xi + J\cdot S[J\lambda]\eta , \\ &E_H = \hat{p}\xi + \epsilon\hat{q}\eta \equiv \xi\cosh\lambda + \eta\epsilon\sinh\lambda , \\ &F_H = \frac{\hat{p}^2\xi^2 - 1 - \hat{q}^2\eta^2}{(\hat{p}\xi + 1)^2 - \hat{q}^2\eta^2} , \\ &\Omega_H = \frac{2\hat{q}\eta}{(\hat{p}\xi + 1)^2 - \hat{q}^2\eta^2} . \end{aligned} \tag{13}$$

If $\lambda = 0$, then $p = 1$ and $q = 0$, and $\mathbf{b}_\mu$ had been calculated in Ref. 4. In this case, $\hat{p} = 1$, $\hat{q} = 0$, and the dual gauge potential $\hat{\mathbf{b}}_\mu$ is defined by

$$\begin{aligned} \hat{\mathbf{b}}_{x_1} &= \left[0,0,\frac{x_2}{c(\xi^2 - 1)}\left(\frac{1}{r_1} + \frac{1}{r_2}\right) - \cos\theta\right] , \\ \hat{\mathbf{b}}_{x_2} &= \left[0,0,\frac{-x_1}{c(\xi^2 - 1)}\left(\frac{1}{r_1} + \frac{1}{r_2}\right) + \sin\theta\right] , \\ \hat{\mathbf{b}}_{x_3} &= [0,0,0] , \\ \hat{\mathbf{b}}_{x_4} &= \left[0,0,\frac{1}{c(\xi^2 - 1)}\left(\frac{x_3 + c}{r_2} + \frac{x_3 - c}{r_1}\right)\right] , \end{aligned} \tag{14}$$

where $r_1^2 = (z - c)^2 + r^2$, $r_2^2 = (z + c)^2 + r^2$. It is interesting that the difference between $\mathbf{b}_\mu$ and $\hat{\mathbf{b}}_\mu$ is only a periodic component.

Next, Eq. (9) has a double solution

$$E(J) = e^{-J\alpha}\coth\psi, \quad \nabla^2\psi = 0 , \tag{15}$$

where α is a real number,

$$e^{J\lambda} = \sum_{n=0}^{\infty} \frac{(J\lambda)^n}{n!} = C[J\lambda] + J\cdot S[J\lambda] .$$

Thus E_c is the ordinary Ernst solution,[4,5] and from E_H we can obtain a new solution

$$(\hat{\phi},\hat{\sigma}) = [T(\mathrm{Re}(\mathscr{E}_H)), W_{\mathrm{Re}(\mathscr{E}_H)}(\mathrm{Im}(\mathscr{E}_H))] .$$

(B) The Eq. (3) is invariant under the double transformation[1]

$$K(J):\mathscr{E}(J) \rightarrow \mathscr{E}'(J) = \frac{a(J)\mathscr{E}(J) + J\cdot b(J)}{J\cdot c(J)\mathscr{E}(J) + d(J)} , \tag{16}$$

where a, b, c, and d are double-real parameters and

$$a(J)d(J) - J^2 b(J)c(J) = 1.$$

Here, $K(J = i)$ is just the Ehlers transformation. By $K(J = \epsilon)$, a new solution $(\hat{\phi}',\hat{\sigma}')$ of SDSGF can be generated from a known $(\hat{\phi},\hat{\sigma})$.

(C) In Ref. 2 we have given a double-inverse scattering method; it can also be used to generate new soliton solutions of SDSGF. In fact, if $(\phi_0(J),\sigma_0(J))$ is a double seed solution of SDSGF corresponding to a background double Ernst potential $\mathscr{E}_0(J)$, and we let

$$M_{00}(J) = \frac{1}{\phi_0(J)}\begin{pmatrix} 1 & \sigma_0(J) \\ \sigma_0(J) & \sigma_0^2(J) - J^2\phi_0^2(J) \end{pmatrix} , \tag{17}$$

then according to Ref. 2, we obtain an infinite soliton solution network as follows:

$$\begin{tikzcd}
\cdots \arrow[r] \arrow[dr, phantom, "\mathscr{T}_\beta"] & M_{0-1}(J) \arrow[r] \arrow[d, "BZ"] \arrow[dr, phantom, "\mathscr{T}"] & M_{00}(J) \arrow[r] \arrow[d, "BZ"] \arrow[dr, phantom, "\mathscr{T}_\beta"] & M_{01}(J) \arrow[r] \arrow[d, "BZ"] \arrow[dr, phantom, "\mathscr{T}"] & \cdots \\
\cdots \arrow[r] \arrow[dr, phantom, "\mathscr{T}_\beta"] & M_{1-1}(J) \arrow[r] \arrow[d, "BZ"] \arrow[dr, phantom, "\mathscr{T}"] & M_{10}(J) \arrow[r] \arrow[d, "BZ"] \arrow[dr, phantom, "\mathscr{T}_\beta"] & M_{11}(J) \arrow[r] \arrow[d, "BZ"] \arrow[dr, phantom, "\mathscr{T}"] & \cdots \\
\cdots \arrow[r] \arrow[dr, phantom, "\mathscr{T}_\beta"] & M_{2-1}(J) \arrow[r] \arrow[d] \arrow[dr, phantom, "\mathscr{T}"] & M_{20}(J) \arrow[r] \arrow[d] \arrow[dr, phantom, "\mathscr{T}_\beta"] & M_{21}(J) \arrow[r] \arrow[d] \arrow[dr, phantom, "\mathscr{T}"] & \cdots \\
 & \vdots & \vdots & \vdots &
\end{tikzcd} , \tag{18}$$

where BZ denotes the inverse scattering transformation[2] from $M_{ij}(J)$ to $M_{i+1j}(J)$ by an operation of adding one simple real pole μ_{n+1}. Here, $\mathscr{T}$ and $\mathscr{T}_\beta$ are transformations from $M_{ij}(J)$ to $M_{ij+1}(J)$ defined, respectively, as follows. If $M(J)$ corresponds to $\mathscr{E}(J)$ by Eq. (17), then

$$\mathscr{T}(M) = M', \quad \mathscr{E}' = -J^2 \frac{F - J\cdot\Omega}{F^2 - J^2\Omega^2} = -J^2 \frac{\mathscr{E}^*}{\mathscr{E}\mathscr{E}^*} \tag{19a}$$

and

$$\begin{aligned} &\mathscr{T}_\beta(M) = M'', \\ &\mathscr{E}''_H = T(F_c) + \epsilon\cdot W_{F_c}(\Omega_c), \\ &\mathscr{E}''_c = T(F_H) + i\cdot V_{F_H}(\Omega_H). \end{aligned} \tag{19b}$$

Now, for each M_{nm} ($|m|$, $n = 1,2,...$) we obtain a soliton solution pair $[(\phi_{nm},\sigma_{nm}),(\hat{\phi}_{nm},\hat{\sigma}_{nm})]$ of SDSGF,

$$\phi_{nm} = \frac{1}{[\widetilde{M}_{nm}(J=i)]_{11}}, \quad \sigma_{nm} = \frac{[\widetilde{M}_{nm}(J=i)]_{12}}{[\widetilde{M}_{nm}(J=i)]_{11}},$$

$$\hat{\phi}_{nm} = T([\widetilde{M}_{nm}(J=\epsilon)]_{11}),$$

$$\hat{\sigma}_{nm} = W_{\hat{\phi}_{nm}}\left(\frac{[\widetilde{M}_{nm}(J=\epsilon)]_{12}}{[\widetilde{M}_{nm}(J=\epsilon)]_{11}}\right), \tag{20}$$

where $[M]_{ij}$ is an element of the matrix M, and $\widetilde{M}_{nm}(J) = M_{nm}(J)$ when n is an even number, but $\widetilde{M}_{nm}(J=i) = M_{nm}(J=\epsilon)$, $\widetilde{M}_{nm}(J=\epsilon) = M_{nm}(J=i)$ when n is an odd number. Notice that, from a fixed seed solution and by the method of Letelier,[6] we can only obtain a subset, which, in fact, corresponds to $M_{00}(J=i) \to M_{20}(J=i) \to M_{40}(J=i) \to \cdots$.

[1] Z. Z. Zhong, J. Math. Phys. **26**, 2589 (1985).
[2] Z. Z. Zhong, Sci. Sin. A **31**, 436 (1988).
[3] C. N. Yang, Phys. Rev. Lett. **38**, 1377 (1977).
[4] L. Witten, Phys. Rev. D **19**, 718 (1979).
[5] F. J. Ernst, Phys. Rev. **167**, 1175 (1968).
[6] P. S. Letelier, J. Math. Phys. **23**, 1175 (1982).
[7] V. A. Belinsky and V. E. Zakharov, Zh. Eksp. Teor. Fiz. **77**, 3 (1979).
[8] A. Tomimatsu and H. Sato, Prog. Theor. Phys. **50**, 95 (1973).

Ordered exponentials and differential equations

H. M. Fried[a)]
Observatoire de Nice, B. P. 139, 06003 Nice Cedex, France

(Received 6 October 1988; accepted for publication 21 December 1988)

An independent and elementary method of guessing the leading form of certain SU(2) and SU(3) ordered exponentials in the strong-coupling limit of rapidly fluctuating input is presented. For SU(N), with $N \geqslant 3$, the method is not unambiguous, but it can be used to check previous estimates.

I. INTRODUCTION

Ordered exponentials (OE's) satisfying first-order differential equations (DE's) and describing the evolution of a multicomponent, linear system have a rich structure of solutions that may be estimated in certain adiabatic and rapidly fluctuating input (RFI) limits of strong coupling.[1-3] One may ask if any part of this structure may be inferred from an examination of higher-order DE's written for individual components of the system, as relevant to the particular input information of the problem. The purpose of this paper is to answer this question in the affirmative and to display, in two simple SU(2) and SU(3) contexts, an independent method of finding results obtained previously in the RFI limit. In the course of this analysis, one learns—it was a surprise to the author—that certain, relatively simple, but important problems can be solved exactly; these exact solutions can be easily generalized to produce the desired RFI limits.

II. THEORY

It will be useful to state the essential problem in a convenient SU(N) context of finding an explicit solution to the first-order DE for the $N \times N$ unitary matrix satisfying

$$\frac{dU(t)}{dt} = i[\lambda \times E(t)]\, U(t), \tag{1}$$

with $U(0) = 1$. Here the input $E_a(t)$ denote real components of an n-dimensional vector, with $n = N^2 - 1$, and the λ_a represent the defining (Gell-Mann) matrices of SU(N). The solution of (1) is the unitary OE of interest, written in the standard (physicists') form as

$$U(t) = \left[\exp\left(i\int_0^t dt'\lambda \times E(t')\right)\right]_+, \tag{2}$$

where the ordering symbol $(\ )_+$ signifies that in the expansion of (2) in powers of E_a, those terms carrying the largest t' values are to be placed on the left, as in

$$([\lambda \times E]^n)_+ = \lambda \times E(t_1)\lambda \times E(t_2) \cdots \lambda \times E(t_n), \quad t_1 \geqslant t_2 \geqslant \cdots t_n.$$

We first consider the simplest case of SU(2). Here, the λ_a are the Pauli matrices σ_a given by

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

and we restrict E_a to nonzero components in directions orthogonal to that direction whose associated σ_a is diagonal [in general SU(N) to those directions whose λ_a are diagonal]. That is, for SU(2), $E_{1,2} \neq 0$; while using the Gell-Mann matrices in SU(3), we exclude E_3 and E_8. The reason for this simplification is that such "diagonal" dependence may always be removed by an appropriate unitary transformation on U, leaving the result dependent upon only the "orthogonal" components E_a.

We therefore consider

$$\frac{dU}{dt} = i[\sigma_1 E_1(t) + \sigma_2 E_2(t)]\, U(t), \quad U(0) = 1 \tag{3}$$

and next calculate the equivalent DE that would be satisfied by the components of the column vector $v(t) = \binom{x(t)}{y(t)}$, where $v(t) = U(t) \times v(0)$. The components $x(t)$ and $y(t)$ are complex functions which for real E_a satisfy the unitarity relation

$$|x(t)|^2 + |y(t)|^2 + |z(t)|^2 = \text{const.} \tag{4}$$

Denoting $E_\pm = E_1 - E_2$, these components satisfy the pair of first-order equations

$$\dot{x} = iyE_-, \quad \dot{y} = ixE_+, \tag{5}$$

with $E_+E_- = E_1^2 + E_2^2 = E^2$. From (5), one may construct the second-order DE

$$\ddot{x} + E^2 x - \dot{x}\frac{d}{dt}\ln\left(\frac{dE_-}{dt}\right) = 0, \tag{6}$$

which for general $E_{1,2}$ is far too complicated to be solved exactly by any method.

A general form for $E_{1,2}$ may be written as

$$E_1(t) = E(t)\cos\left(\int_0^t dt'\,\omega(t')\right),$$
$$E_2(t) = E(t)\sin\left(\int_0^t dt\,\omega(t')\right), \tag{7}$$

where $E(t)$ and $\omega(t)$ represent an alternate pair of arbitrary functions. The coefficient of $-dx/dt$ in (6) would then become $d(\ln[dE/dt - i\omega E])/dt - i\omega(t)$. If one imagines passing to the limit of large $\rho = \omega/E$, with variations of ρ (or, more precisely, with the appropriate dimensionless variations of ω and/or E) negligible compared to ρ, then for sufficiently long time scales ($\gg \omega^{-1}$), one might expect the behavior of (6) to be given by solutions of

$$\ddot{x} + E^2 x - i\omega\dot{x} = 0, \tag{8}$$

where E and ω are slowly varying; on the time scale $\gg \omega^{-1}$ they may be taken to be constants. This is analogous to the special case of constant E and ω used in Refs. 1 and 2, which

[a)] Permanent address: Physics Department, Brown University, Providence RI 02912.

 0022-2488/89/051161-03$02.50

when generalized to slowly varying functions—but always with $\rho \gg 1$—give the leading RFI approximations for U.

With this as motivation, one may now consider (8) with the constant coefficients E and ω. Of course, it is immediately soluble in the form

$$x(t) = A_t e^{i\Omega_+ t} + A_- e^{-i\Omega_- t}, \quad (9)$$

with

$$\Omega^\pm = E[\pm\sqrt{1+\rho^2/4} - \rho/2]$$

and A_+, A_- chosen as appropriate constants. Note that Ω_+/E is precisely the form previously found for $\xi(\rho)$.[2]

From the pair (5), one can see that a corresponding solution for $y(t)$,

$$y(t) = B_+ e^{-\Omega_+ t} + B_- e^{i\Omega_- t}, \quad (10)$$

will require $B_+ = A_+\Omega_+$, $B_- = A_-\Omega_-$. Of course, these constants may be reexpressed in terms of the integration constants $x_0 = x(0)$, $y_0 = y(0)$ of the original first-order DE's:

$$A_+ = \frac{x_0 - y_0/\Omega_-}{1_-\Omega_+/\Omega_-}, \quad A_- = \frac{x_0 - y_0/\Omega_+}{1 - \Omega_-/\Omega_+}. \quad (11)$$

We now extract the appropriate $F_{0,a}$ of the exact $U = F_0 + i\boldsymbol{\sigma}\times\mathbf{F}$, where the $F_{0,a}$ are real functions. Multiplying out the matrix product $U(t)\times v(0)$ yields the column vector

$$\begin{pmatrix} [F_0 + iF_3]x_0 + [F_2 + iF_1]y_0 \\ [F_0 - iF_3]y_0 - [F_2 - iF_1]x_0 \end{pmatrix}, \quad (12)$$

which must be the same as $\binom{x(t)}{y(t)}$. Using (9)–(11), one obtains a pair of complex equations by matching coefficients of the *independent* x_0, y_0; then one can take real and imaginary parts of these equations to obtain the four desired relations for $F_{0,a}$:

$$F_0 = [\Omega_+ \cos(\Omega_- t) - \Omega_- \cos(\Omega_+ t)]/(\Omega_+ - \Omega_-),$$
$$F_3 = [\Omega_+ \sin(\Omega_- t) - \Omega_- \sin(\Omega_+ t)]/(\Omega_+ - \Omega_-),$$
$$F_2 = [\cos(\Omega_- t) - \cos(\Omega_+ t)]/(\Omega_+ - \Omega_-),$$
$$F_1 = [\sin(\Omega_+ t) - \sin(\Omega_- t)]/(\Omega_+ - \Omega_-). \quad (13)$$

Equations (13) are exact and represent the solution for the special case of constant E and ω. In particular, the solution can be evaluated in the large ρ limit using $\Omega_- = -(\omega + \Omega_+)$; one finds that it reproduces exactly the leading forms of the $F_{0,a}$ given previously,[1,2]

$$\overline{F} = \cos G, \quad \overline{F}_3 = \sin G, \quad \overline{F}_1 = \xi[\sin(G+L) + \sin G],$$
$$\overline{F}_2 = \xi[\cos G - \cos(G+L)], \quad (14)$$

where $\xi \simeq 1/\rho$ (just the leading term in the expansion of Ω_+/E) and

$$G = \int_0^t dt' \frac{E(t')}{\rho(t')}, \quad L = \int_0^t dt'\,\omega(t'). \quad (15)$$

In writing (14) and (15), we have supposed that the input dependence on time scales large compared to ω^{-1} is obtained by replacing the $\xi\, Et$ and ωt of this special solution by the integral statements of (15).

For sufficiently large $\rho(t)$ this seems to be an "experimentally" valid prescription; one can now see its justification at a glance from (6) and (8): The replacements $\Omega_\pm t \to \int_0^t dt'\,\Omega_\pm(t')$ will reproduce Eq. (8), in leading $1/\rho$ order, with slowly varying functions E and ω. Here, then, is another, independent method for estimating $U(t)$ in the RFI limit: Solve for $x(t)$, $y(t)$ exactly, in the constant E, ω case; then, with the aid of the independent constants x_0 and y_0, solve for the $F_{0,a}$; and, finally, take the limit $\rho \gg 1$ and allow for the possibility of slow variations in E and ω by writing (15).

Turning to SU(3), one can ask if these same techniques are applicable. In general, the answer is no, although much of the SU(3) structure is reproduced. We consider the simplest nontrivial model of Ref. 3, where but two E_a components were nonzero:

$$E_a = \delta_{a2} E\cos(\omega t) + \delta_{as} E\sin(\omega t)$$

and first consider E and ω as constants. Writing

$$v(t) = \begin{pmatrix} x(t) \\ y(t) \\ z(t) \end{pmatrix} = U(t) \times v(0),$$

where $U = M_0 + \lambda_a M_a$ is the object of interest, the component equations corresponding to those of (5) are found to be

$$\dot{x} = zE_5 + yE_2, \quad \dot{y} = -xE_2, \quad \dot{z} = -xE_5. \quad (16)$$

From (16) it immediately follows that $R^2 = x^2 + y^2 + z^2$ is a constant of the motion; if $R(t=0)$ is real so is $R(t)$, leading to the inference that $x(t)$, $y(t)$, and $z(t)$ will all be real if $x(0)$, $y(0)$, and $z(0)$ are real.

To solve this triplet of equations, it is useful to define the new linear combinations u and w according to $Eu(t) = E_5 z + E_2 y$ and $Ew(t) = E_2 z - E_5 y$. The equations corresponding to (16) for the triplet x, u, and w are then

$$\dot{w} = -\omega u, \quad \dot{x} = Eu, \quad \dot{u} = \omega w - Ex, \quad (17)$$

and it is a trivial matter to eliminate any two of these variables to obtain an equation for one of them along, e.g.,

$$\ddot{u} + \Omega^2 u = 0, \quad \Omega^2 = E^2 + \omega^2. \quad (18)$$

Because (18) is a second-order DE, it has but two constants of integration. Once $u(t)$ is known, a trivial integration [the first of Eqs. (17)] yields $w(t)$, with one new integration constant; but the remaining pair of Eqs. (17) relate that constant to those two of $u(t)$. In other words, after one has made the transition to $x(t)$, $y(t)$, and $z(t)$, one finds an explicit solution to (16), but with one relation between the three constants of integration, which here takes the form of $z(0) = -(\omega/E)x(0)$ and independent $x(0)$, $y(0)$. For $(\omega/E) \gg 1$, one finds

$$x(t) \simeq x_0 \cos(\Omega t) + y_0(E/\Omega)\sin(\Omega t),$$
$$y(t) \simeq -x_0(\omega/E)\sin G + y_0 \cos G, \quad (19)$$
$$z(t) \simeq -x_0(\omega/E)\cos G - y_0 \sin G,$$

with $G = (\Omega - \omega)t \simeq Et/\rho$ and $\rho = 2\omega/E$.

Were the three integration constants independent, it would then be possible in principle to repeat the SU(2) procedure for SU(3), achieving for the nine complex, independent parameters $M_{0,a}$ of U a total of nine independent relations; and, upon combining the nine relations with the nine independent statements of unitarity, to determine unambiguously the real and imaginary parts of the nine coefficient functions $M_{0,a}$. More simply, if the solutions x, y, and z are real, then $M_{0,1,3,4,6,8}$ are real, while $M_{2,5,7} = iJ_{2,5,7}$ are imaginary, so that nine relations would suffice to determine the nine components.

However, in the present case, with but two independent constants of integration, we can construct but six independent relations, which are insufficient to specify the $M_{0,a}$. In the large ω/E limit, one will be forced to make certain assumptions; we do this in a way that resembles a part of the output of Ref. 3: choosing $M_{1,6,4}$ and $J_{2,5}$ as negligible in the $(\omega/E) \gg 1$ limit, one finds that the remaining four quantities

$$M_0 \simeq \tfrac{1}{3}(1 + 2\cos G), \quad M_3 \simeq \tfrac{1}{2}(1 - \cos G),$$
$$J_7 \simeq \sin G, \quad M_8 \simeq (1/2\sqrt{3})(1 - \cos G) \tag{20}$$

are exactly those obtained previously.[3] Although one cannot view this as an independent construction of the leading RFI behavior in this SU(3) example, one may infer from (20) that the method used in Ref. 3 is correct, although not unambiguous.

III. SUMMARY

In summary, one sees that by considering second-order DE's for vector components, instead of the first-order DE for the OE of interest, the leading behavior of the RFI can be reproduced. More important, this illustrates the close connection between OE approximations at large ρ and "infrared" approximations of second-order DE's, where the essential part of the approximate solution comes from the large, low-frequency input dependence, here the $-i(\omega/E)$ of $E^{-1}d[\ln(dE_-/dt)]dt$. Calculating the solution to the second-order DE written for constant E and ω and then subsequently modifying it for slow variations of these quantities—as long as $\rho \gg 1$—is a particularly simple way of defining a "strong coupling by infrared extraction" approximation, such as those which have been used in a variety of "eikonal" problems elsewhere.[4] One sees that the relatively low-frequency components of the input E and ω parameters, treated in a nonperturbative way, define the leading, relatively low-frequency behavior of the desired solution for the OE.

[1] M.-E. Brachet and H. M. Fried, Phys. Lett. A **103**, 309 (1984).
[2] M.-E. Brachet and H. M. Fried, J. Math. Phys. **28**, 15 (1987).
[3] H. M. Fried, J. Math. Phys. **28**, 1275 (1987).
[4] In a quantum field theory context, see F. Guèrin and H. M. Fried, Phys. Ref. D **33**, 3039 (1986); H. M. Fried and T. Grandou, *ibid.* **33**, 1151 (1986); T. Grandou, H.-T. Cho, and H. M. Fried, *ibid.* **37**, 946, 960 (1988).

Existence of infinitely many zero-energy states in a model of supersymmetric quantum mechanics

Asao Arai
Department of Mathematics, Hokkaido University, Sapporo 060, Japan

(Received 9 September 1988; accepted for publication 21 December 1988)

The general framework of the $N=2$ Wess–Zumino holomorphic supersymmetric quantum mechanics with polynomial superpotentials is extended to the case of nonpolynomial superpotentials $V(z)$ ($z\in\mathbb{C}$) in a mathematically rigorous way. It is also proved that there exist no fermionic zero-energy states. Under some conditions for V, the operator domain of the supercharges and the supersymmetric Hamiltonian are identified. As an example, the model with $V(z)=\lambda e^{\alpha z}$ ($\lambda\in\mathbb{C}\setminus\{0\}$, $\alpha>0$) is analyzed in view of index theory. The following remarkable result is proved: There exist infinitely many bosonic zero-energy states which are localized in the momentum space dual to the Im z direction. The results are applied to two models in atomic and nuclear physics.

I. INTRODUCTION

In Ref. 1, Jaffe *et al.* considered two models of supersymmetric quantum mechanics (SSQM), which are the quantum mechanics versions of the two-dimensional, $N=1$ and $N=2$ Wess–Zumino quantum field models; they also computed the Witten index ($\equiv$the number of bosonic zero-energy states minus the number of fermionic zero-energy states) in each model. In particular, it was proved that in the $N=2$ Wess–Zumino SSQM with an arbitrary polynomial superpotential $V(z)$ ($z\in\mathbb{C}$), there exist no fermionic zero-energy states (the "vanishing theorem") and the Witten index I_W is equal to the number of bosonic zero-energy states, with

$$I_W=\deg V-1\,. \tag{1.1}$$

In this paper, we consider the $N=2$ Wess–Zumino SSQM with *nonpolynomial* holomorphic superpotentials and try to extend the results for the case of the polynomial potentials considered in Ref. 1. This is at least mathematically interesting: Formula (1.1) shows that the Witten index is determined by the order of singularity of the superpotential at $z=\infty$ and suggests formally that the Witten index is infinite in the case of nonpolynomial holomorphic superpotentials, for they have the essential singularity at $z=\infty$. Note, also, that a nonpolynomial entire function is the limit of a sequence $\{V_n(z)\}_{n=1}^{\infty}$ of polynomials with deg $V_n=n$.

In Sec. II, we describe in a mathematically rigorous way a fundamental framework for the Wess–Zumino holomorphic SSQM with not necessarily polynomial superpotentials. We shall show that some results in Ref. 1 can be extended. For example, the "vanishing theorem" holds also in the present case (Proposition 2.5). In Sec. III, we consider the case with $V(z)=\lambda e^{\alpha z}$ ($\lambda\in\mathbb{C}\setminus\{0\}$, $\alpha>0$) and prove by identifying the space of the bosonic zero-energy states exactly that there exist infinitely many bosonic zero-energy states. Thus as far as this special model is concerned, the result justifies the above formal argument. It is noted that every bosonic zero-energy state is localized in the momentum space dual to the Im z direction. In Sec. IV, we apply the result in Sec. III to two models in atomic and nuclear physics: One is a model of a nonrelativistic spin-$\frac{1}{2}$ particle in an external SU(2) gauge field and the other is a model of a nonrelativistic nucleon in a pion field. These models were discussed in Ref. 1 in order to give physical interpretation to the Wess–Zumino holomorphic SSQM. In each model, the potential is two-dimensional and periodic in one direction (e.g., the y direction). It is shown that each model has infinitely many zero-energy states which are localized in the momentum space dual to the y direction.

II. WESS–ZUMINO SSQM WITH GENERAL HOLOMORPHIC SUPERPOTENTIALS

In this section we recapitulate the definition of the $N=2$ Wess–Zumino holomorphic SSQM[1] (cf., also, Refs. 2 and 3) and extend some mathematical results obtained in the case of polynomial potentials[1] to the case of general holomorphic potentials.

The Hilbert space $\mathscr{H}$ of state vectors for the model is given by

$$\mathscr{H}=L^2(\mathbb{R}^2;\mathbb{C}^4)\,. \tag{2.1}$$

In order to define the supercharges, we introduce 4×4 matrices ψ_1 and ψ_2 by

$$\psi_1=\frac{1}{2}\begin{pmatrix}0 & I+\sigma_3\\ I-\sigma_3 & 0\end{pmatrix},\qquad \psi_2=\frac{1}{2}\begin{pmatrix}0 & i\sigma_1+\sigma_2\\ -i\sigma_1-\sigma_2 & 0\end{pmatrix}, \tag{2.2}$$

where $\sigma_j, j=1,2,3$ are the Pauli matrices

$$\sigma_1=\begin{pmatrix}0&1\\1&0\end{pmatrix},\quad \sigma_2=\begin{pmatrix}0&-i\\i&0\end{pmatrix},\quad \sigma_3=\begin{pmatrix}1&0\\0&-1\end{pmatrix} \tag{2.3}$$

and I is the 2×2 identity matrix. The matrices ψ_1 and ψ_2 satisfy the anticommutation relations

$$\{\psi_j,\psi_k^*\}=\delta_{jk}\,, \tag{2.4}$$

$$\{\psi_j,\psi_k\}=0,\quad j,k=1,2\,, \tag{2.5}$$

where $\{A,B\}\equiv AB+BA$.

Let $V(z)$ be a holomorphic function on $\mathbb{C}$ (not necessarily polynomial) and consider the operators

$$Q_1=i(\psi_2\bar\partial+\psi_2^*\partial)+i\{\psi_1(\partial V)-\psi_1^*(\partial V)^*\}\,, \tag{2.6}$$

$$Q_2=\psi_2\bar\partial-\psi_2^*\partial+\psi_1(\partial V)+\psi_1^*(\partial V)^*\,, \tag{2.7}$$

where $\partial=\partial/\partial z$ and $\bar\partial=\partial/\partial z^*$ [the operators Q_j given by

 0022-2488/89/051164-07$02.50

(2.6) and (2.7) are different from those in Ref. 1]. We shall use the usual identification of $\mathbb{C}$ with $\mathbb{R}^2$ through the correspondence $z = x + iy\in\mathbb{C}\leftrightarrow(x,y)\in\mathbb{R}^2$. Then Q_1 and Q_2 can be considered as operators acting in $\mathscr{H}$ given by (2.1). We put

$$D = C_0^\infty(\mathbb{R}^2;\mathbb{C}^4) . \tag{2.8}$$

Proposition 2.1: The operators Q_1 and Q_2 are essentially self-adjoint on D. Further, every power of Q_1 (respectively, Q_2) is essentially self-adjoint on D.

Remark: In the case of polynomial potentials V, Proposition 2.1 was proved for Q_1 and Q_1^2 (Ref. 1).

Proof: It is obvious that Q_j is symmetric on D. We write it as

$$Q_1 = -iL$$

on D, where

$$L = A_1\frac{\partial}{\partial x} + A_2\frac{\partial}{\partial y} + B(x,y) ,$$

with

$$A_1 = -\tfrac{1}{2}(\psi_2 + \psi_2^*),\quad A_2 = (i/2)(\psi_2^* - \psi_2) ,$$
$$B(x,y) = \psi_1^*(\partial V(z))^* - \psi_1\,\partial V(z) .$$

The operator L is of the form of the first-order differential operators considered in Ref. 4 because $A_j; j = 1,2$; and B are C^∞ 4×4 matrix-valued functions. Note that A_1 and A_2 are constant matrices. Hence the "velocity of propagation" (Ref. 4) associated to L is constant. Then a direct application of Ref. 4, Theorem 2.2 gives the desired result on Q_1. The proof for the case of Q_2 is quite similar. □

Remark: Note that $\{A_j,A_k\} = \delta_{jk}/2$, $j,k = 1,2$. Hence Q_1 is a Dirac-type operator. The same holds for Q_2.

We shall denote the closure of $Q_j \restriction D$ by $\overline{Q}_j$. Then we have the following lemma.

Lemma 2.2:

$$\overline{Q}_1^2 = \overline{Q}_2^2 . \tag{2.9}$$

Proof: Direct computations give

$$Q_1^2\Psi = Q_2^2\Psi$$

for all Ψ in D. Then, Proposition 2.1 implies (2.9). □

We define the non-negative self-adjoint operator H by

$$H\equiv\overline{Q}_1^2 = \overline{Q}_2^2 . \tag{2.10}$$

Then we have

$$H^{1/2} = |\overline{Q}_1| = |\overline{Q}_2| \tag{2.11}$$

and hence, in particular,

$$D(\overline{Q}_1) = D(\overline{Q}_2) = D(H^{1/2}) , \tag{2.12}$$

where $D(A)$ denotes the operator domain of operator A.

Lemma 2.3: Each Q_j maps D into itself and the anticommutation relation

$$\{Q_1,Q_2\}\Psi = 0,\quad \Psi\in D \tag{2.13}$$

holds. Further, we have

$$(\overline{Q}_1\Psi,\overline{Q}_2\Phi) + (\overline{Q}_2\Psi,\overline{Q}_1\Phi) = 0,\quad \Phi,\Psi\in D(H^{1/2}) . \tag{2.14}$$

Proof: Equation (2.13) follows from direct computations. By a limiting argument using Proposition 2.1, one can extend (2.13) in the form of (2.14). □

Let

$$N_{\rm F} = \begin{pmatrix} I & 0 \\ 0 & -I \end{pmatrix} . \tag{2.15}$$

Lemma 2.4: For each $j = 1,2$, $N_{\rm F}$ maps $D(\overline{Q}_j)$ into itself and the anticommutation relations

$$\{N_{\rm F},\overline{Q}_j\}\Psi = 0,\quad \Psi\in D(\overline{Q}_j),\quad j = 1,2 \tag{2.16}$$

hold.

Proof: We first prove (2.16) for Ψ in D. Then, a limiting argument using Proposition 2.1 gives the desired result. □

The Hilbert space $\mathscr{H}$ given by (2.1) has the following orthogonal decomposition:

$$\mathscr{H} = \mathscr{H}_+ \oplus \mathscr{H}_- , \tag{2.17}$$

with

$$\mathscr{H}_+ = \left\{\begin{pmatrix} f_1 \\ f_2 \\ 0 \\ 0 \end{pmatrix} \middle| f_1,f_2\in L^2(\mathbb{R}^2)\right\} , \tag{2.18}$$

$$\mathscr{H}_- = \left\{\begin{pmatrix} 0 \\ 0 \\ f_1 \\ f_2 \end{pmatrix} \middle| f_1,f_2\in L^2(\mathbb{R}^2)\right\} . \tag{2.19}$$

Obviously, we have

$$N_{\rm F}\Psi_\pm = \pm\Psi_\pm ,\quad \Psi_\pm\in\mathscr{H}_\pm . \tag{2.20}$$

In summary, we have proved that *the quadruple* $\{\mathscr{H},\{\overline{Q}_1,\overline{Q}_2\},H,N_{\rm F}\}$ *is a SSQT with N=2 supersymmetry in the sense of Ref. 5* (cf., also, Ref. 6); the operators $\overline{Q}_1$ and $\overline{Q}_2$ are the self-adjoint supercharges, H is the supersymmetric Hamiltonian, and $N_{\rm F}$ is the fermion number operator. The closed subspace $\mathscr{H}_+$ (respectively, $\mathscr{H}_-$) is the Hilbert space consisting of bosonic (respectively, fermionic) states.

On the domain D, H is explicitly given as

$$H = -\partial\bar\partial - \psi_1^*\psi_2(\partial^2V)^* - \psi_2^*\psi_1\,\partial^2V + |\partial V|^2 . \tag{2.21}$$

By a general fact of a SSQT, H is reduced by $\mathscr{H}_\pm$; we shall denote the reduced part of H to $\mathscr{H}_\pm$ by $H_\pm$. We have

$$H_+ = H_- + \begin{pmatrix} 0 & -i(\partial^2V) \\ i(\partial^2V)^* & 0 \end{pmatrix} \tag{2.22}$$

and

$$H_- = (-\partial\bar\partial + |\partial V|^2)\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \tag{2.23}$$

on D, where we identify $\mathscr{H}_\pm$ with $L^2(\mathbb{R}^2;\mathbb{C}^2)$.

We now proceed to the index problem. The Witten index $I_{\rm W}$ is defined by the number of bosonic zero-energy states minus the number of fermionic zero-energy states[3]:

$$I_{\rm W} = n_+ - n_- , \tag{2.24}$$

with

$$n_\pm = \dim\operatorname{Ker} H_\pm . \tag{2.25}$$

By Lemma 2.4 and (2.20), for each $j = 1,2$ there exists a unique closed linear operator $Q_{j+}:\mathscr{H}_+\to\mathscr{H}_-$ such that

$$\overline{Q}_j = \begin{pmatrix} 0 & Q_{j+}^* \\ Q_{j+} & 0 \end{pmatrix} . \tag{2.26}$$

It follows from (2.10) that

$$H_+ = Q_{j+}^* Q_{j+}, \quad H_- = Q_{j+} Q_{j+}^*, \quad j = 1,2, \tag{2.27}$$

which imply that

$$n_+ = \dim \operatorname{Ker} Q_{j+}, \quad n_- = \dim \operatorname{Ker} Q_{j+}^*, \quad j = 1,2. \tag{2.28}$$

Hence we obtain

$$\begin{aligned} I_W &= \dim \operatorname{Ker} Q_{j+} - \dim \operatorname{Ker} Q_{j+}^* \\ &\equiv \operatorname{index} Q_{j+}, \quad j = 1,2. \end{aligned} \tag{2.29}$$

Remark: The arguments leading to (2.26)–(2.29) apply to every SSQT and the results are well known.

Proposition 2.5 (Vanishing theorem)*:* There exist no fermionic zero-energy states:

$$n_- = 0. \tag{2.30}$$

Remark: This result of (2.30) has been established in the case where $V(z)$ is a polynomial (Ref. 1, Proposition 6). In the case where V is not necessarily a polynomial, it may happen that H_- is not closed on $D(\partial\bar\partial) \cap D(|\partial V|^2)$ and hence that $H_-\Omega = 0$ is not equivalent to $-\partial\bar\partial\Omega = 0$ and $|\partial V|^2\Omega = 0$, as in the case of polynomial potentials V.

Proof: The operator $h \equiv -\partial\bar\partial + |\partial V|^2$ is a two-dimensional Schrödinger operator with a non-negative potential. By Proposition 2.1, $C_0^2(\mathbb{R}^2)$ is a core for h. Thus we can apply Lemma A1 in the Appendix to obtain the desired result. □

We next consider conditions for a vector to be in Ker $H_+ = \operatorname{Ker} Q_{j+}$ [see (2.25), (2.27), and (2.28)].

Formulas (2.28) and (2.29) show that as far as the index problem is concerned, it is sufficient to consider one of $\bar Q_j, j = 1,2$. Henceforth we write

$$Q = \bar Q_1, \quad Q_+ = Q_{1+}. \tag{2.31}$$

Lemma 2.6: Suppose that $D(Q_+) = D(\partial) \cap D(|\partial V|)$. Then the following hold.

(i) Every vector (f,g) in Ker Q_+ satisfies

$$(-\partial\bar\partial + |\partial V|^2) f + (\partial^2 V)(\partial V)^{-1} \bar\partial f = 0, \tag{2.32}$$

$$(-\partial\bar\partial + |\partial V|^2) g + (\partial^2 V)^* (\partial V)^{*-1} \partial g = 0 \tag{2.33}$$

in the generalized sense.

(ii) Let $f \in L^2(\mathbb{R}^2)$ be a vector satisfying (2.32) in the generalized sense. Then (f,g) is in Ker Q_+ if and only if f is in $D(\partial) \cap D(|\partial V|)$ and $(\partial V)^{-1}\bar\partial f$ is in $L^2(\mathbb{R}^2)$, with

$$g = i(\partial V)^{-1} \bar\partial f. \tag{2.34}$$

(iii) Let $g \in L^2(\mathbb{R}^2)$ be a vector satisfying (2.33) in the generalized sense. Then (f,g) is in Ker Q_+ if and only if g is in $D(\partial) \cap D(|\partial V|)$ and $(\partial V)^{*-1}\partial g$ is in $L^2(\mathbb{R}^2)$, with

$$f = -i(\partial V)^{*-1} \partial g. \tag{2.35}$$

Remark: Lemma 2.6 is an elaborate and extended version of Ref. 1, Lemma 8.

Proof: (i) By (2.2), (2.3), (2.6), and the assumption $D(Q_+) = D(\partial) \cap D(|\partial V|)$, we have

$$Q_+ = \begin{pmatrix} -i(\partial V)^* & \partial \\ \bar\partial & i\,\partial V \end{pmatrix} \tag{2.36}$$

on $D(Q_+)$. Hence every vector (f,g) is in Ker Q_+ if and only if (f,g) is in $D(Q_+)$ and satisfies

$$\partial g - i(\partial V)^* f = 0 \tag{2.37}$$

and

$$\bar\partial f + i(\partial V) g = 0. \tag{2.38}$$

Equation (2.37) and the condition $f \in D(\bar\partial)$ imply that $\bar\partial\,\partial g$ exists as an L^1_{loc} function with

$$\bar\partial\,\partial g - i(\partial^2 V)^* f - i(\partial V)^* \bar\partial f = 0. \tag{2.39}$$

It follows from (2.37) and (2.38) that

$$f = -i(\partial V)^{*-1} \partial g, \quad \bar\partial f = -i(\partial V) g.$$

Substituting these relations into (2.39), we obtain (2.33). Similarly, using (2.38), we can show that (2.32) holds.

(ii) The part "only if" is obvious by (2.38). To prove the part "if," we first note that (2.34) gives (2.38). Hence it follows that $(\partial V)g$ is in $L^2(\mathbb{R}^2)$ [i.e., $g \in D(\partial V)$] and

$$\partial\bar\partial f + i(\partial^2 V) g + i(\partial V)\partial g = 0. \tag{2.40}$$

Using (2.32) and (2.38) to rewrite (2.40), we obtain (2.37). In particular, we have $g \in D(\partial)$. Thus we have proved that (f,g) is in $D(\partial) \cap D(|\partial V|)$ and satisfies (2.37) and (2.38). Therefore (f,g) is in Ker Q_+.

(iii) Similar to the proof of (ii). □

Finally, we consider conditions under which the assumption $D(Q_+) = D(\partial) \cap D(\partial V)$ in Lemma 2.6 holds.

Lemma 2.7: Suppose that there exists a constant $r > 0$ such that for all $z \in \mathbb{C}$ satisfying $|\partial^2 V(z)| \geqslant r$, the estimate

$$|\partial^2 V(z)|^2 \leqslant a |\partial V(z)|^4 + b \tag{2.41}$$

holds with the constants $0 < a < 1$ and $b \geqslant 0$. Then $D(H_-) = D(\partial\bar\partial) \cap D(|\partial V|^2)$ and (2.23) holds as an operator equality.

Proof: Let f be in $C_0^\infty(\mathbb{R}^2; \mathbb{C}^2)$. Then we have

$$\begin{aligned} \|H_- f\|^2 &= \|\partial\bar\partial f\|^2 + \|(\partial V)^2 f\|^2 \\ &\quad - 2\,\mathrm{Re}(\partial\bar\partial f, |\partial V|^2 f). \end{aligned}$$

Via integration by parts, one can see that

$$\begin{aligned} 2\,\mathrm{Re}(\partial\bar\partial f, |\partial V|^2 f) &= \|(\partial^2 V) f\|^2 \\ &\quad - \|(\partial f)(\partial V)\|^2 - \|(\bar\partial f)(\partial V)\|^2. \end{aligned}$$

Hence we obtain

$$\|H_- f\|^2 \geqslant \|\partial\bar\partial f\|^2 + \|(\partial V)^2 f\|^2 - \|(\partial^2 V) f\|^2.$$

Using (2.41), we can show that

$$\|(\partial^2 V) f\|^2 \leqslant a \|(\partial V)^2 f\|^2 + (r^2 + b)\|f\|^2.$$

Therefore, we obtain the estimate

$$\|\partial\bar\partial f\|^2 + (1-a)\|(\partial V)^2 f\|^2 \leqslant \|H_- f\|^2 + (r^2 + b)\|f\|^2. \tag{2.42}$$

Since $C_0^\infty(\mathbb{R}^2; \mathbb{C}^2)$ is a core for H_- (Proposition 2.1), (2.42) extends to all f in $D(H_-)$, showing at the same time that $D(H_-) \subset D(\partial\bar\partial) \cap D(|\partial V|^2)$. Since H_- is self-adjoint, we conclude that $D(H_-) = D(\partial\bar\partial) \cap D(|\partial V|^2)$ and $H_- = -\partial\bar\partial + |\partial V|^2$. □

Remark: In the case of polynomial potentials V, it is easy to see that (2.41) holds, where a can be made arbitrarily small if r is taken sufficiently large.

Lemma 2.8: Suppose that there exists $r > 0$ such that (2.41) holds with $0 < a < \frac{1}{2}$ and $b \geqslant 0$. Then $D(H_+) = D(\partial\bar\partial) \cap D(|\partial V|^2)$ and (2.22) holds as an operator equality.

Proof: Let

$$H_I = \begin{pmatrix} 0 & -i(\partial^2 V) \\ i(\partial^2 V)^* & 0 \end{pmatrix}$$

and f be in $D(H_-)$. Then by Lemma 2.7 and the above assumption, we have $D(H_-) \subset D(H_I)$ and

$$\|H_I f\|^2 \leqslant a\|(\partial V)^2 f\|^2 + (r^2+b)\|f\|^2 .$$

Using (2.42) extended to $f \in D(H_-)$, we obtain

$$\|H_I f\|^2 \leqslant \frac{a}{1-a}\|H_- f\|^2 + \frac{r^2+b}{1-a}\|f\|^2 .$$

Since $0<a<\frac{1}{2}$, we have $0<a/(1-a)<1$. Therefore, by a standard theorem (the Kato–Rellich theorem) (see, e.g., Ref. 7, Chap. V, Theorem 4.3 and Ref. 8, Theorem X.12), $H_+ = H_- + H_I$ [see (2.22)] is self-adjoint on $D(H_-)$. □

Remark: Polynomial potentials V satisfy the assumption of Lemma 2.8.

Lemma 2.9: Under the same assumption as in Lemma 2.8, we have $D(Q_+) = D(\partial) \cap D(\partial V)$.

Proof: By Lemma 2.8 and its proof, the operators $\partial\bar{\partial}$ and $|\partial V|^2$ are relatively bounded with respect to H_+ and hence, by a standard theorem (see e.g., Ref. 7, Chap. VI, Theorem I.38 and Ref. 8, Theorem X.18), relatively form bounded with respect to H_+. Therefore, in particular, it follows that $D(\partial)\cap D(\partial V) \supset D(H_+^{1/2}) = D(Q_+)$. Define the operators

$$L_+ = \begin{pmatrix} -i(\partial V)^* & \partial \\ \bar{\partial} & i\partial V \end{pmatrix}$$

and

$$L_- = \begin{pmatrix} i\partial V & -\partial \\ -\bar{\partial} & -i(\partial V)^* \end{pmatrix}$$

on $D(\partial)\cap D(\partial V)$. Then the above result shows that

$$Q_+ \subset L_+ . \tag{2.43}$$

Obviously, the operator

$$L \equiv \begin{pmatrix} 0 & L_- \\ L_+ & 0 \end{pmatrix}$$

is a symmetric extension of $Q \restriction D$. Hence we obtain $Q = \bar{L}$ since $Q = \overline{Q \restriction D}$ is self-adjoint. On the other hand, we have

$$\bar{L} = \begin{pmatrix} 0 & \bar{L}_- \\ \bar{L}_+ & 0 \end{pmatrix}.$$

By the uniqueness of Q_+, we obtain $Q_+ = \bar{L}_+$ and $Q_+^* = \bar{L}_-$. Hence, by (2.43), we have $Q_+ = L_+ = \bar{L}_+$. Therefore, L_+ is closed on $D(\partial)\cap D(\partial V)$ and equal to Q_+. □

III. THE MODEL WITH A SUPERPOTENTIAL OF THE EXPONENTIAL TYPE

In this section, we consider the model with the potential

$$V(z) = \lambda e^{\alpha z}, \quad z \in \mathbb{C}, \tag{3.1}$$

where $\lambda \in \mathbb{C}\setminus\{0\}$ and $\alpha > 0$. Then H_+ takes the form

$$H_+ = -\partial\bar{\partial} + |\lambda|^2\alpha^2 e^{2\alpha \operatorname{Re} z} + i\alpha^2 \begin{pmatrix} 0 & -\lambda e^{\alpha z} \\ \lambda^* e^{\alpha z^*} & 0 \end{pmatrix} \tag{3.2}$$

on D [see (2.22)], for we have

$$\partial V(z) = \lambda\alpha e^{\alpha z}, \quad \partial^2 V(z) = \lambda \alpha^2 e^{\alpha z}. \tag{3.3}$$

For a measurable function u on $[0,2\alpha]$, we define the functions f_u and $\tilde{f}_u$ on $\mathbb{R}^2$ by

$$f_u(x,y) = \int_0^{2\alpha} u(p) e^{\alpha x/2} W_{0,|\alpha - p|/\alpha}(4|\lambda| e^{\alpha x}) e^{ipy}\, dp , \tag{3.4}$$

$$\tilde{f}_u(x,y) = \frac{i}{2\lambda\alpha} e^{-\alpha(x+iy)}\left(\frac{\partial}{\partial x} + i\frac{\partial}{\partial y}\right) f_u(x,y) , \tag{3.5}$$

provided that the right-hand sides are meaningful, where $W_{k,m}(\cdot)$ is the Whittaker function (see e.g., Ref. 9, Chap. XVI). Let

$$D_\alpha = \{u \in L^2(\mathbb{R}) \mid \operatorname{supp} u \subset (0,\alpha/2)\cup(\alpha/2,\alpha)\} . \tag{3.6}$$

We shall prove the following proposition.

Proposition 3.1: Let H_+ be given by (3.2). Then every vector in $\operatorname{Ker} H_+$ is of the form $(f_u, \tilde{f}_u)$ or $((\tilde{f}_u)^*, f_u^*)$, with some u satisfying the condition $\operatorname{supp} u \subset [0,\alpha]$ and

$$\operatorname{Ker} H_+ \supset \{(f_u,\tilde{f}_u) \mid u \in D_\alpha\} \cup \{((\tilde{f}_u)^*, f_u^*) \mid u \in D_\alpha\} . \tag{3.7}$$

In particular, H_+ has infinitely many zero-energy states:

$$n_+ = \infty . \tag{3.8}$$

Remarks: (i) The above result shows that in the present model, supersymmetry is unbroken with infinitely degenerate vacua.

(ii) It should be noticed that the Fourier transform $\hat{f}_u(x,p)$ [respectively, $\hat{\tilde{f}}_u(x,p)$] of $f_u(x,y)$ [respectively, $\tilde{f}_u(x,y)$] with respect to y has compact support in $p \in \mathbb{R}$. This means physically that every zero-energy state of H_+ is strictly localized in the momentum space dual to the y direction.

(iii) Let

$$V_N(z) = \lambda \sum_{n=0}^{N} \frac{(\alpha z)^n}{n!}$$

and let $Q^{(N)}$ and $H_+^{(N)}$ be Q and H_+ with $V = V_N$, respectively. Then by Ref. 1, Proposition 9 [see (1.1)], we have $n_+(N) \equiv \dim \operatorname{Ker} Q_+^{(N)} = \dim \operatorname{Ker} H_+^{(N)} = N-1$. On the other hand, it is easy to see that $Q^{(N)}$ and $H_+^{(N)}$ converge Q and H_+ in the strong resolvent sense as $N \to \infty$. Formula (3.8) may be regarded as $n_+ = \infty = \lim_{N\to\infty} n_+(N)$.

(iv) The corresponding model in the $N=1$ Wess–Zumino SSQM (the Witten model[2,3]) is given by the Hamiltonian

$$H = \begin{pmatrix} H_+ & 0 \\ 0 & H_- \end{pmatrix}, \quad H_\pm = -\frac{d^2}{dx^2} + \lambda^2\alpha^2 e^{2\alpha x} \pm \lambda\alpha^2 e^{\alpha x},$$

where $\alpha, \lambda \in \mathbb{R}\setminus\{0\}$. This model was discussed in Ref. 10 and it can be shown that $\operatorname{Ker} H_\pm = \{0\}$.[11] This result also shows an essential difference between the $N=1$ and $N=2$ Wess–Zumino SSQM's.

Lemma 3.2: For all $r>0$ and all $z \in \mathbb{C}$ with $|\partial^2 V(z)| \geqslant r$, the estimate

$$|\partial^2 V(z)| \leqslant (\alpha^2/r)|\partial V(z)|^2 \tag{3.9}$$

holds.

Proof: An elementary exercise. □

Lemma 3.3: The operator H_+ is self-adjoint on $D(\partial\bar{\partial}) \cap D(|\partial V|^2)$ and $D(Q_+) = D(\partial)\cap D(\partial V)$.

Proof: By taking r sufficiently large in (3.9), the assumption of Lemma 2.8 is satisfied. Thus Lemmas 2.8 and 2.9 give the desired results. □

The following lemma is derived from Lemma 2.6 (i) and (3.3).

Lemma 3.4: Every vector (f,g) in Ker Q_+ satisfies

$$\Big(-\frac{\partial^2}{\partial x^2}-\frac{\partial^2}{\partial y^2}+2\alpha\frac{\partial}{\partial x}+2i\alpha\frac{\partial}{\partial y}+4|\lambda|^2\alpha^2e^{2\alpha x}\Big)f(x,y)=0, \qquad (3.10)$$

$$\Big(-\frac{\partial^2}{\partial x^2}-\frac{\partial^2}{\partial y^2}+2\alpha\frac{\partial}{\partial x}-2i\alpha\frac{\partial}{\partial y}+4|\lambda|^2\alpha^2e^{2\alpha x}\Big)g(x,y)=0, \qquad (3.11)$$

where we put $z=x+iy$.

Lemma 3.5: Let $a\in\mathbb{R}$, $b>0$, $c>0$, and $E\in\mathbb{C}$ be constants.

(i) If Re $E\leqslant 0$, then

$$\Big(-\frac{d^2}{dx^2}+2a\frac{d}{dx}+ce^{2bx}\Big)f(x)=Ef(x) \qquad (3.12)$$

has no solutions $f\neq 0$ with $f\in D(d^2/dx^2)\cap D(e^{2bx})\subset L^2(\mathbb{R})$.

(ii) Suppose that

$$0\leqslant|\mathrm{Re}\sqrt{a^2-E}|<a,\quad \mathrm{Re}\,E>0$$

and $2\sqrt{a^2-E}/b$ is not an integer.

Then every nonzero solution $f\in L^2(\mathbb{R})$ to Eq. (3.12) is given by

$$f(x)=Ke^{(a-b/2)x}W_{0,\sqrt{a^2-E}/b}(2\sqrt{c}e^{bx}/b), \qquad (3.13)$$

with a constant $K\in\mathbb{C}\setminus\{0\}$, where $W_{k,m}(z)$ is the Whittaker function (see e.g., Ref. 9, Chap. XVI).

Proof: (i) Via the elliptic regularity, every solution f to Eq. (3.12) is C^∞. Let $f\in D(d^2/dx^2)\cap D(e^{2bx})$ be a nonzero solution to Eq. (3.12). Taking the inner product of f with (3.12) in $L^2(\mathbb{R})$, we obtain

$$\|f'\|^2+2a(f',f)+c\|e^{bx}f\|^2=E\|f\|^2.$$

Since (f',f) is pure imaginary, it follows that Re $E>0$. Thus the desired result follows.

(ii) Let $f\in L^2(\mathbb{R})$ be a nonzero solution to Eq. (3.12) and define

$$v(t)=t^{(b-2a)/2b}f(1/2b\log(b^2t^2/4c)),\quad t>0. \qquad (3.14)$$

Then we have

$$\|f\|_2^2=\frac{1}{b}\int_0^\infty t^{2(a-b)/b}|v(t)|^2\,dt<\infty \qquad (3.15)$$

and v satisfies the Whittaker equation

$$\frac{d^2}{dt^2}v(t)+\Big\{-\frac{1}{4}+\frac{\frac14-(a^2-E)/b^2}{t^2}\Big\}v(t)=0. \qquad (3.16)$$

Since $2\sqrt{a^2-E}/b$ is not an integer by assumption, the confluent hypergeometric functions $M_{0,\sqrt{a^2-E}/b}(t)$ and $M_{0,-\sqrt{a^2-E}/b}(t)$ (see, e.g., Ref. 9, Chap. XVI) form a fundamental system of solutions to (3.16); every solution to (3.16) is given by a linear combination of these functions. By taking the asymptotic property of $M_{0,\mu}(t)$ as $t\to 0$ and $t\to\infty$ into account, we see that possible solutions to (3.16) with condition (3.15) are scalar multiples of the Whittaker function $W_{0,\sqrt{a^2-E}/b}(t)$, with

$$0\leqslant|\mathrm{Re}\sqrt{a^2-E}|<a. \qquad (3.17)$$

Condition (3.17) comes from the integrability condition of $t^{2(a-b)/b}|v(t)|^2$ near $t=0$. On the contrary, if we define $f(x)$ by relation (3.14) with $v(t)=W_{0,\sqrt{a^2-E}/b}(t)$ under condition (3.15), then f is in $L^2(\mathbb{R})$ $(f\neq 0)$ and satisfies Eq. (3.12). □

Lemma 3.6: (i) Every solution $f\in D(\partial\bar\partial)\cap D(e^{2\alpha x})$ to Eq. (3.10) has the form (3.4).

(ii) For all $u\in D_\alpha$, the function f_u given by (3.4) is a solution to Eq. (3.10) with $f_u\in D(\partial\bar\partial)\cap D(e^{2\alpha x})$.

Remark: The sets of the solutions g of Eq. (3.11) consist of the complex conjugates of the solutions f to (3.10).

Proof: (i) If f is in $D(\partial\bar\partial)\cap D(e^{2\alpha x})$, then Eq. (3.10) is equivalent to

$$\Big(-\frac{\partial^2}{\partial x^2}+2\alpha\frac{\partial}{\partial x}+4|\lambda|^2\alpha^2e^{2\alpha x}\Big)\hat f(x,p)=E(p)\hat f(x,p), \qquad (3.18)$$

with

$$E(p)=p(2\alpha-p),\quad p\in\mathbb{R}, \qquad (3.19)$$

where $\hat f(x,p)$ is the Fourier transform of $f(x,y)$ with respect to y:

$$\hat f(x,p)=\frac{1}{\sqrt{2\pi}}\int e^{-ipy}f(x,y)dy.$$

Via Lemma 3.5 (i), Eq. (3.18) has no nonzero solutions $\hat f(\cdot,p)\in D(d^2/dx^2)\cap D(e^{2\alpha x})$ if $E(p)\leqslant 0$. Hence $\hat f(\cdot,p)=0$ for all $p\notin(0,2\alpha)$.

Let $E(p)>0$, i.e., $0<p<2\alpha$. Then by Lemma 3.5 (ii), every solution $\hat f(\cdot,p)\in L^2(\mathbb{R})$ to Eq. (3.18) is given by

$$\hat f(x,p)=u(p)e^{\alpha x/2}W_{0,|\alpha-p|/\alpha}(4|\lambda|e^{\alpha x})=\hat f_u(x,p),$$

where u is a function on the set $S\equiv\{p\in\mathbb{R}|0<p<2\alpha, p\neq\alpha/2,\alpha,\frac32\alpha\}$. Thus the desired result follows.

(ii) Let $u\in D_\alpha$. We need only show that f_u is in $D(\partial\bar\partial)\cap D(e^{2\alpha x})$: We write it as $f_u=f$. By the asymptotics of the Whittaker function $W_{0,\mu}(t)$ at $t=0$ and $t=+\infty$ (see, e.g., Ref. 9, Chap. XVI), we see that $\hat f$ and $e^{2\alpha x}\hat f(x,p)$ are in $L^2(\mathbb{R}^2)$. By using the recursion relation

$$zW'_{0,m}(z)=(z/2)W_{0,m}(z)-W_{1,m}(z),$$

we can show that

$$\frac{\partial}{\partial x}\hat f(x,p)=\frac{\alpha}{2}\hat f(x,p)+2|\lambda|\alpha e^{\alpha x}\hat f(x,p)-\alpha u(p)e^{\alpha x/2}W_{1,|\alpha-p|/\alpha}(4|\lambda|e^{\alpha x}). \qquad (3.20)$$

Each term on the rhs of (3.20) is in $L^2(\mathbb{R}^2)$ and hence $\partial\hat f(x,p)/\partial x$ is in $L^2(\mathbb{R}^2)$. By virtue of (3.18), this implies that $\partial^2\hat f(x,p)/\partial x^2$ is in $L^2(\mathbb{R}^2)$. We can also see that $p^2\hat f(x,p)$ is in $L^2(\mathbb{R}^2)$. Thus the function f is in $D(\partial\bar\partial)\cap D(e^{2\alpha x})$. □

Lemma 3.7: Let f_u be given by (3.4) with $u\in D_\alpha$. Then f_u is in $D(\partial)\cap D(e^{\alpha x})$ and $e^{-\alpha x}\partial f_u$ is in $L^2(\mathbb{R}^2)$.

Proof: By the proof of Lemma 3.6 we have $f_u\in D(\partial\bar\partial)\cap D(e^{2\alpha x})$. Since $D(\partial\bar\partial)\cap D(e^{2\alpha x})\subset D(\partial)\cap D(e^{\alpha x})$, we obtain $f_u\in D(\partial)\cap D(e^{\alpha x})$.

Let

$$h(x,y)=e^{-\alpha x}(\bar\partial f_u)(x,y).$$

Then we have

$$\hat h(x,p)=\frac12 e^{-\alpha x}\Big(\frac{\partial}{\partial x}\hat f_u(x,p)-p\hat f_u(x,p)\Big).$$

By using (3.20), we obtain

$$2\hat{h}(x,p) = R(x,p) + 2|\lambda|\alpha \hat{f}_u(x,p),$$

where

$$R(x,p) = u(p)e^{-\alpha x/2}\{(\alpha/2 - p)W_{0,|\alpha-p|/\alpha}(4|\lambda|e^{\alpha x}) - \alpha W_{1,|\alpha-p|/\alpha}(4|\lambda|e^{\alpha x})\}. \qquad (3.21)$$

Since $\hat{f}_u$ is in $L^2(\mathbb{R}^2)$, we need only show that the function R is in $L^2(\mathbb{R}^2)$. It is easy to see that for every $c\in\mathbb{R}$,

$$\int_{\mathbb{R}} dp \int_c^\infty dx |R(x,p)|^2 < \infty .$$

By the asymptotics of $W_{k,m}(z)$ at $z=0$ and by virtue of the condition $\operatorname{supp} u \subset (0,\alpha/2)\cup(\alpha/2,\alpha)$, we see that

$$e^{\alpha x/2}R(x,p) \underset{x\to-\infty}{\sim} u(p)w_1(p) \times (4|\lambda|e^{\alpha x})^{-|\alpha-p|/\alpha+1/2}\cdot(4|\lambda|e^{\alpha x})$$

and hence

$$R(x,p) \underset{x\to-\infty}{\sim} u(p)w_2(p)e^{(\alpha-|\alpha-p|)x},$$

with the continuous functions w_1 and w_2 on the support of u. Therefore, we have

$$\int_{\mathbb{R}} dp \int_{-\infty}^c dx |R(x,p)|^2 < \infty .$$

Thus we obtain $R\in L^2(\mathbb{R}^2)$. □

Proof of Proposition 3.1: Via Lemma 3.6, the set of solutions to Eq. (3.10) in $D(\partial\bar{\partial})\cap D(e^{2\alpha x})$ consist of just functions of the form f_u given by (3.4). If f_u and $\tilde{f}_u$ given by (3.4) and (3.5), respectively, are in $L^2(\mathbb{R}^2)$, then $R(\cdot,p)$ defined by (3.21) must be in $L^2(\mathbb{R})$ for a.e. $p\in[0,2\alpha]$. We can show that if $\alpha\leqslant p\leqslant 2\alpha$, then $R(\cdot,p)$ is not in $L^2(\mathbb{R})$ (cf. the proof of Lemma 3.7). Hence we have $\operatorname{supp} u\subset[0,\alpha]$. Then the first assertion follows from Lemma 2.6(ii) and (iii) and the remark after Lemma 3.6.

Via Lemma 3.7, every f_u with $u\in D_\alpha$ satisfies the assumption of Lemma 2.6(ii). Therefore, for every $u\in D_\alpha$, the pair $(f_u,\tilde{f}_u)$, with $\tilde{f}_u$ given by (3.5), is in $\operatorname{Ker} Q_+ = \operatorname{Ker} H_+$. Via Lemma 2.6(iii) and the remark after Lemma 3.6, we also have $((\tilde{f}_u)^*, f_u^*)\in\operatorname{Ker} H_+$ for every $u\in D_\alpha$. Thus we obtain (3.7). We have $\dim D_\alpha = \infty$ and, if the vectors $u_1,\dots,u_n\in D_\alpha$ are linearly independent, then so are the vectors $(f_{u_1},\tilde{f}_{u_1}),\dots,(f_{u_n},\tilde{f}_{u_n})$. Thus (3.8) follows. □

IV. APPLICATION

In this section, we apply the result in Sec. III to models in atomic and nuclear physics, which were discussed in Ref. 1.

A. Nonrelativistic spin-$\frac{1}{2}$ particle in an external SU(2) gauge field

Let τ_a, $a=1,2,3$ be the generators of the SU(2) group. An external SU(2) gauge field $A(x) = (A_1(x), A_2(x), A_3(x))$, $x=(x_1,x_2,x_3)\in\mathbb{R}^3$ is given as $A_j(x) = \Sigma_{a=1}^3 A_j^a(x)\tau_a$, $j=1,2,3$, where we take A_j^a to be real valued. Then the Hamiltonian of a nonrelativistic spin-$\frac{1}{2}$ particle with mass 1 coupled minimally to the gauge field is given by

$$H_A = \tfrac{1}{2}(-i\nabla - gA)^2 - \tfrac{1}{2}\sigma\cdot B, \qquad (4.1)$$

where $g\in\mathbb{R}\setminus\{0\}$ is a coupling constant, $\nabla = (\partial/\partial x_1, \partial/\partial x_2, \partial/\partial x_3)$, $\sigma=(\sigma_1,\sigma_2,\sigma_3)$, and $B_j = g(\operatorname{curl} A)_j + \frac{1}{2}g^2\Sigma_{k,m=1}^3 \epsilon_{jkm}[A_k,A_m]$ (where ϵ_{jkm} is the Kronecker antisymmetric symbol).

We consider the following situation.

(a): (i) $A_1=A_2=0$, (ii) $A_3^3=0$, (iii) $A_3^1(x)$ and $A_3^2(x)$ depend only on x_1 and x_2, and (iv) $\nabla_3=0$, where condition (iv) means that we consider only state vectors independent of x_3 and identify them with elements in $L^2(\mathbb{R}^2;\mathbb{C}^4)$. Let V be a holomorphic function on $\mathbb{C}$ and H be given by (2.21). Then it was shown in Ref. 1 that under condition (a), with $A_3^1 - iA_3^2 = 2\,\partial V/g$, H_A can be written as

$$H_A = 2H \qquad (4.2)$$

with a rearrangement of components.

Proposition 4.1: Suppose that condition (a) is satisfied. Let

$$A_3^1(x_1,x_2) = (2\alpha\lambda/g)e^{\alpha x_1}\cos\alpha x_2, \qquad (4.3)$$

$$A_3^2(x_1,x_2) = -(2\alpha\lambda/g)e^{\alpha x_1}\sin\alpha x_2, \qquad (4.4)$$

with $\alpha>0$ and $\lambda\in\mathbb{R}\setminus\{0\}$. Then H_A has infinitely many zero-energy states Ψ in $L^2(\mathbb{R}^2;\mathbb{C}^4)$: $H_A\Psi=0$. Further, the Fourier transform $\hat{\Psi}(x_1,p)$ of every $\Psi(x_1,x_2)$ with respect to x_2 has compact support in $p\in\mathbb{R}$.

Proof: By (4.3) and (4.4), we have

$$A_3^1 - iA_3^2 = 2\frac{\partial V}{g},$$

with $V(z)=\lambda e^{\alpha z}$ $(z=x_1+ix_2)$. Thus by (4.2) and Proposition 3.1, we obtain the desired result. □

Remark: The gauge field given by (4.3) and (4.4) is periodic in the x_2 direction. This may be an origin of the existence of infinitely many zero-energy states and the localization of the states in the momentum space dual to the x_2 direction.

B. Nonrelativistic nucleon in an external pion field

The Hamiltonian of a nonrelativistic nucleon in an external pion field $\phi(x) = (\phi_1(x),\phi_2(x),\phi_3(x))$, $x\in\mathbb{R}^3$ is given by

$$H_\phi = -\tfrac{1}{2}\nabla^2 + \tfrac{1}{2}g\sigma\cdot\nabla(\tau\cdot\phi(x)) + \tfrac{1}{2}g^2\phi(x)^2. \qquad (4.5)$$

(See Ref. 1.) Suppose that the following condition is satisfied.

Condition (ϕ): (i) $\phi_3=0$, (ii) ϕ_1 and ϕ_2 depend only on x_1 and x_2, and (iii) $\nabla_3=0$.

Then it was shown in Ref. 1 that

$$H_\phi = 2H, \qquad (4.6)$$

with $\phi_1 - i\phi_2 = -2i\,\partial V/g$.

Proposition 4.2: Suppose that condition (ϕ) is satisfied. Let

$$\phi_1(x_1,x_2) = (2\lambda\alpha/g)e^{\alpha x_1}\sin\alpha x_2,$$

$$\phi_2(x_1,x_2) = (2\lambda\alpha/g)e^{\alpha x_1}\cos\alpha x_2,$$

with $\alpha>0$ and $\lambda\in\mathbb{R}\setminus\{0\}$. Then, H_ϕ has infinitely many zero-energy states Φ in $L^2(\mathbb{R}^2;\mathbb{C}^4)$: $H_\phi\Phi=0$. Further, the Fourier transform $\hat{\Phi}(x_1,p)$ of every $\Phi(x_1,x_2)$ with respect to x_2 has compact support in $p\in\mathbb{R}$.

Proof: Similar to the proof of Proposition 4.1. □

ACKNOWLEDGMENTS

This research was supported in part by Grant-In-Aid Nos. 63740063 and 62460001 for science research from the Ministry of Education, Japan.

APPENDIX: KERNEL OF A SCHRÖDINGER OPERATOR

Let d be an arbitrarily fixed positive integer. Let U be a real-valued measurable function on $\mathbb{R}^d$ and Δ be the d-dimensional Laplacian: $\Delta = \Sigma_{j=1}^{d} \partial^2/\partial x_j^2$, $x = (x_1,...,x_d) \in \mathbb{R}^d$. We consider the Schrödinger operator

$$H_S = -\Delta + U \tag{A1}$$

acting in $L^2(\mathbb{R}^d)$.

Lemma A1: Suppose that $C_0^2(\mathbb{R}^d) \subset D(U)$ and $U \geqslant 0$. Let $\overline{H}_S$ be the closure of $H_S \restriction C_0^2(\mathbb{R}^2)$. Then we have

$$\operatorname{Ker} \overline{H}_S = \{0\}. \tag{A2}$$

Remark: (i) It is obvious that under the condition $C_0^2(\mathbb{R}^d) \subset D(U)$, $H_S \restriction C_0^2(\mathbb{R}^d)$ is closable and symmetric.

(ii) The domain $D(\overline{H}_S)$ is not necessarily equal to $D(\Delta) \cap D(U)$. This is the reason why we need a limiting argument to prove (A2) (see below).

Proof: Let $f \in \operatorname{Ker} \overline{H}_S : \overline{H}_S f = 0$. Then we can take a sequence $\{f_n\} \subset C_0^2(\mathbb{R}^d)$ such that $f_n \xrightarrow{s} f$ and $H_S f_n \xrightarrow{s} 0 (n \to \infty)$: It follows from the latter that

$$\sum_{j=1}^{d} \|D_j f_n\|^2 + \|U^{1/2} f_n\|^2 \to 0,$$

where $D_j = \partial/\partial x_j$. Hence we have $D_j f_n \xrightarrow{s} 0$, $j = 1,...,d$, and $U^{1/2} f_n \xrightarrow{s} 0$. Since D_j, $j = 1,...,d$, are closed, it follows that $f \in D(D_j)$, $j = 1,...,d$, and

$$D_j f = 0, \quad j = 1,...,d,$$

which, together with $f \in L^2(\mathbb{R}^d)$, imply that $f = 0$. □

[1]A. Jaffe, A. Lesniewski, and M. Lewenstein, "Ground state structure in supersymmetric quantum mechanics," Ann. Phys. **178**, 313 (1987).

[2]E. Witten, "Dynamical breaking of supersymmetry," Nucl. Phys. B **185**, 513 (1981).

[3]E. Witten, "Constraints on supersymmetry breaking," Nucl. Phys. B **202**, 253 (1982).

[4]P. Chernoff, "Essential self-adjointness of powers of generators of hyperbolic equations," J. Funct. Anal. **12**, 401 (1973).

[5]A. Arai, "Supersymmetry and singular perturbations," J. Funct. Anal. **60**, 378 (1985).

[6]H. Grosse and L. Pittner, "Supersymmetric quantum mechanics defined as sesquilinear forms," J. Phys. A Math. Gen. **20**, 4265 (1987).

[7]T. Kato, *Perturbation Theory for Linear Operators* (Springer, Berlin, 1976), 2nd ed.

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

[9]E. T. Whittaker and G. N. Watson, *A Course of Modern Analysis* (Cambridge U. P., London, 1969), 4th ed.

[10]R. Akhoury and A. Comtet, "Anomalous behavior of the Witten index—Exactly soluble models," Nucl. Phys. B **246**, 253 (1984).

[11]In Ref. 10, the Witten index I_W is computed formally and the two different results $I_W = -\frac{1}{2}$ and $I_W = 0$ are derived according to the two different methods of calculation. We can prove rigorously that $\operatorname{Ker} H_\pm = \{0\}$ and hence $I_W = 0$. The method of the proof is similar to that in the present paper.

On infinite-dimensional variational principles with constraints

Rudolf Schmid
Department of Mathematics and Computer Science, Emory University, Atlanta, Georgia 30322

Alberto Simoni
Dipartimento di Scienze Fisiche dell' Universita, Napoli, Italy and Istituto Nazionale di Fisica Nucleare, Sezione di Napoli, Napoli, Italy

(Received 7 July 1988; accepted for publication 14 December 1988)

The Lagrange multiplier theorem is generalized for constrained functions on dual pairs of Banach spaces. Then a variational principle for dual pairs of Banach spaces is proven for the case when the constraint set is given by a symmetry and it is generalized to Banach manifolds.

I. INTRODUCTION

The classical variational principle states that for a differentiable function $F: \mathbf{R}^n \to \mathbf{R}$ a necessary condition for a point $x_0 \in \mathbf{R}^n$ to be a local extremum of F is that $\nabla F(x_0) = 0$. This is easily generalized to Banach spaces. Let X be a Banach space and $F: X \to \mathbf{R}$ a differentiable function. A necessary condition for F to have an extremum at a point $x_0 \in X$ is that the Frechet derivative of F at x_0 vanishes, i.e., $DF(x_0) = 0$. By definition $DF(x_0)$ is a continuous linear map from X to $\mathbf{R}$ hence an element of the dual space X^*; $DF(x_0) \in X^*$.

In physical applications the Banach spaces in question are often function spaces such as $X = C^1[a,b]$ or $X = C^\infty[a,b]$, etc., and it can be difficult to determine the corresponding dual space X^*. However, in order to just compute $DF(x_0)$ we do not need to know what the dual space X^* is because the Frechet derivative can be computed in using the directional derivative (Gateaux derivative), i.e., $DF(x_0)h = (d/dt|_{t=0})F(x+th)$ for any $h \in X$. The knowledge of the dual space X^* becomes more important when one considers the problem of finding an extremum of a function F subject to a *constraint*. As in classical theory, one formulates this problem as a Lagrange multiplier problem (Kolmogorov,[1] Weinstock,[2] Whittacker,[3] Goldstein[4]), but now the Lagrange multiplier is no longer a real number as in the finite dimensional case, but an element of the dual space X^*, i.e., a continuous linear functional on X. To avoid the problem of determining the dual space X^* we consider a more general concept of duality of Banach spaces. The idea is that one chooses a convenient "dual" space with respect to a pairing (see Schmid[5] for details). In Sec. II we introduce the concept of dual pairs of Banach spaces and in Sec. III we formulate the variational principle in this context and prove a Lagrange multiplier theorem with respect to a constraint surface. We study several examples; in particular, we derive Maxwell's equations from this variational principle by varying the field and not the potential. The existence of the vector potential follows from the Lagrange multiplier theorem, i.e., the potential arises as a Lagrange multiplier associated to the inhomogeneous equation. In Sec. IV we study the variational principle in the case where the functional is invariant under some symmetry group and the constraint surface is then given by the fixed point set of the action. The question now is whether critical symmetric points are also symmetric critical points. We give a necessary and sufficient condition for this "Principle" to hold.

II. DUAL PAIRS OF BANACH SPACES

A pair of Banach spaces (X,X') are called *dual* to each other with respect to a pairing if there exists a continuous bilinear map $\langle\ ,\ \rangle: X' \times X \to \mathbf{R}$ that is *weakly nondegenerate*. This means that if $\langle x',x\rangle = 0$ for all $x \in X$, then $x' = 0$, and if $\langle x',x\rangle = 0$ for all $x' \in X'$ then $x = 0$. This is equivalent to the condition that the induced linear maps from X' to X^* and from X to $(X')^*$ defined by $x' \to \langle x',\cdot\rangle$, $x' \in X'$, and $x \to \langle\cdot,x\rangle$, $x \in X$, are one-to-one. If these maps are isomorphisms then the pairing $\langle\ ,\ \rangle$ is called *nondegenerate*.

A weakly nondegenerate pairing $\langle\ ,\ \rangle$ thus represents certain linear functionals on X in terms of elements in X', i.e., each element $x' \in X'$ defines an $x^* \in X^*$ by $x^*(x) = \langle x',x\rangle$, $x \in X$.

Examples: (1) Let $\Omega \subset \mathbf{R}^n$ be open and consider the Banach space $X = C^k(\Omega)$ of real-valued functions of class C^k on Ω. Take $X' = C^k(\Omega) = X$. Then a pairing (called the L^2 *pairing*) between X and X' is given by

$$\langle\ ,\ \rangle: C^k(\Omega) \times C^k(\Omega) \to \mathbf{R}; \quad \langle f,g\rangle = \int_\Omega f(x)g(x)dx.$$

The dual space of $C^k(\Omega)$ is the space of distributions on Ω and $\langle\ ,\ \rangle$ defines a one-to-one linear map from X to $(X')^*$.

(2) Let $X = H^s(\mathbf{R}^n)$ be the space of H^s vector fields on $\mathbf{R}^n$ and take $X' = H^s(\Lambda^1(\mathbf{R}^n))$ the space of H^s one-forms on $\mathbf{R}^n$. Then a weakly nondegenerate pairing

$$\langle\ ,\ \rangle: H^s(\mathbf{R}^n) \times H^s(\Lambda^1(\mathbf{R}^n)) \to \mathbf{R}$$

called the L^2 pairing is given by

$$\langle X,\alpha\rangle = \int_{\mathbf{R}^n} X(x)\cdot\alpha(x)dx, \quad X \in H^s(\mathbf{R}^n), \quad \alpha \in H^s(\Lambda^1(\mathbf{R}^n)).$$

(3) If X is a Banach space with an inner product $\langle\ ,\ \rangle$: $X \times X \to \mathbf{R}$; then take $X' = X$ and $\langle\ ,\ \rangle$ defines a weakly nondegenerate pairing since $\langle\ ,\ \rangle$ is positive definite. If X is a Hilbert space then the inner product $\langle\ ,\ \rangle$ defines a nondegenerate pairing by the Riesz representation theorem.

(4) Let X by a Banach space and take $X' = X^*$. Then a pairing $\langle\ ,\ \rangle$: $X^* \times X \to \mathbf{R}$ is naturally given by $\langle x^*, x\rangle = x^*(x)$ and the map $X \to X^*$ is the identity. Thus if X is

 0022-2488/89/051171-06$02.50

reflexive then the pairing $\langle\ ,\ \rangle$ is nondegenerate by the Hahn–Banach theorem.

Let (Y,Y') be another pair of Banach spaces with a weakly nondegenerate pairing $\langle\ ,\ \rangle_Y$: $Y'\times Y\to\mathbf{R}$ and let A: $X\to Y$ be a linear map. The *adjoint* of A, if it exists, is the linear map A^*: $Y'\to X'$ defined by $\langle A^*y',x\rangle_X = \langle y', Ax\rangle_Y$, $x\in X$, $y'\in Y'$, where $\langle\ ,\ \rangle_X$ denotes the pairing between X and X'. If A is closed and linear then $A^{**}=A$.

Example 5: Let M be an n-dimensional manifold and $\Lambda^k(M)$ the space of differentiable k-forms on M [completed in some suitable H^s-Sobolev topology so $\Lambda^k(M)$ becomes a Banach space], and let $*:\Lambda^k(M)\to\Lambda^{n-k}(M)$ be the Hodge-star operator. Then for $X=X'=\Lambda^k(M)$ the L^2 pairing $\langle\ ,\ \rangle$: $\Lambda^k(M)\times\Lambda^k(M)\to\mathbf{R}$ is given by $\langle\alpha,\beta\rangle=\int_M\alpha\wedge *\beta$, $\alpha,\beta\in\Lambda^k(M)$. If d: $\Lambda^k(M)\to\Lambda^{k+1}(M)$ denotes the exterior derivative then $\delta\equiv *d*$: $\Lambda^k(M)\to\Lambda^{k-1}(M)$ is the adjoint operator, $\langle d\alpha,\beta\rangle=\langle\alpha,\delta\beta\rangle$, i.e., $\int_M d\alpha\wedge *\beta=\int_M\alpha\wedge\delta *\beta$.

Let L be a linear subspace of X. The *orthogonal* complement $L^\perp$ of L in X' is the linear subspace of X' defined by $L^\perp=\{x'\in X'|\langle x',L\rangle=0\}$. If L is a closed split linear subspace of X, i.e., if there exists a topological complement W of L in X, $L\oplus W=X$ then $L=(L^\perp)^\perp$.

Lemma 1: Let A: $X\to Y$ be linear such that ker A splits. Then $(\operatorname{Ker} A)^\perp=\operatorname{Im} A^*$.

Proof: Let $y'\in(\operatorname{Im} A)^\perp$. Then $\langle y', Ax\rangle_Y=0$ for all $x\in X$, hence $\langle A^*y',x\rangle_X=0$ for all $x\in X$, which implies $A^*y'=0$ and so $y'\in\ker A^*$. Reversing these steps shows that $(\operatorname{Im} A)^\perp=\operatorname{Ker} A^*$. Therefore $(\operatorname{Im} A^*)^\perp=\ker A^{**}=\ker A$ and $(\operatorname{Im} A^*)^{\perp\perp}=\operatorname{Im} A^*=(\operatorname{Ker} A)^\perp$. ■

III. LAGRANGE MULTIPLIERS IN DUAL PAIRS OF BANACH SPACES

With this notion of duality we introduce the *functional derivative*. Let F: $X\to\mathbf{R}$ be differentiable at $x_0\in X$. The *functional derivative* (*or variational derivative*) $\delta F/\delta x_0$ of F with respect to x_0, is the unique element $\delta F/\delta x_0\in X'$, if it exists, such that

$$DF(x_0)h=\left\langle\frac{\delta F}{\delta x_0},h\right\rangle,\quad\text{for all } h\in X.$$

Remark 1: For a functional of several variables, e.g., F: $X_1\times X_2\to\mathbf{R}$ (with weak duals X_1',X_2') the *partial* functional derivatives $\delta F/\delta x_1\in X_1'$, $\delta F/\delta x_2\in X_2'$, $x_1\in X_1'$, $x_2\in X_2'$, are defined in a similar way (for details see, e.g., Schmid[5]).

Proposition 1: Let F: $X\to\mathbf{R}$ be C^1. If $x_0\in X$ is an extremum of F then $\delta F/\delta x_0=0$.

Proof: For each $h\in X$ let $f(t)=F(x_0+th)$. If x_0 is an extremum for F, then $t=0$ is an extremum for f: $\mathbf{R}\to\mathbf{R}$; hence $f'(0)=0$. So for any $h\in X$

$$0=f'(0)=\frac{d}{dt}\bigg|_{t=0}F(x_0+th)=DF(x_0)h=\left\langle\frac{\delta F}{\delta x_0},h\right\rangle.$$

Hence $\langle\delta F/\delta x_0,h\rangle=0$ for all $h\in X$, therefore $\delta F/\delta x_0=0$. ■

Let (X,X') and (Y,Y') be two pairs of Banach spaces with the corresponding weakly nondegenerate pairings $\langle\ ,\ \rangle_X$: $X'\times X\to\mathbf{R}$, $\langle\ ,\ \rangle_Y$: $Y'\times Y\to\mathbf{R}$, and let Φ: $X\to Y$ be a differentiable map. Let F: $X\to\mathbf{R}$ be a differentiable function. We want to find a criterion for an extremum of the function $F(x)$ subject to the constraint condition $\Phi(x)=0$; i.e., we consider Φ as constraint map and look for extrema of F on the constraint surface $\Sigma=\Phi^{-1}(0)\subset X$.

Theorem 1: Let F: $X\to\mathbf{R}$ be C^1 and a constraint map Φ: $X\to Y$ be C^1. Assume $0\in Y$ is a regular value of Φ and let $x_0\in\Sigma\equiv\Phi^{-1}(0)$. If x_0 is an extremum of F/Σ then there exists a $y_0'\in Y'$ (*Lagrange multiplier*) such that

$$\frac{\delta F}{\delta x_0}=D\Phi(x_0)^*y_0'.$$

Lemma 2: In the situation above, if $h\in\operatorname{Ker} D\Phi(x_0)$, then $\langle\delta F/\delta x_0,h\rangle_X=0$.

Proof: If $0\in Y$ is a regular value of Φ then $\Sigma=\Phi^{-1}(0)$ is a submanifold of X and the tangent space to the surface Σ at $x_0\in\Sigma$ is given by $T_{x_0}\Sigma=\operatorname{Ker} D\Phi(x_0)$, which splits in $T_{x_0}X$, the tangent space of X at x_0 ($T_{x_0}X\cong X$). For $h\in\operatorname{Ker} D\Phi(x_0)$ choose a curve γ: $\mathbf{R}\to\Sigma$ such that $\gamma(0)=x_0$ and $\dot\gamma(0)=h\in T_{x_0}\Sigma$. If x_0 is an extremum of F/Σ then $t=0$ is an extremum for $f(t)=(F\circ\gamma)(t)$, hence $f'(0)=0$. Therefore

$$0=\frac{d}{dt}\bigg|_{t=0}(F\circ\gamma)(t)=DF(\gamma(0))\cdot\dot\gamma(0)$$

$$=DF(x_0)\cdot h=\left\langle\frac{\delta F}{\delta x_0},h\right\rangle.\qquad\blacksquare$$

Proof of Theorem 1: From Lemma 2 we get that if x_0 is an extremum of F/Σ then $\langle\delta F/\delta x_0,h\rangle=0$ for all $h\in\operatorname{Ker} D\Phi(x_0)$, hence $\delta F/\delta x_0\in[\operatorname{Ker} D\Phi(x_0)]^\perp\subset X'$. Lemma 1 implies that $[\operatorname{Ker} D\Phi(x_0)]^\perp=\operatorname{Im} D\Phi(x_0)^*\subset X'$ and therefore $\delta F/\delta x_0\in\operatorname{Im} D\Phi(x_0)^*$, i.e., there exists a $y_0'\in Y'$ such that $\delta F/\delta x_0=D\Phi(x_0)^*y_0'$. ■

Corollary 1: Theorem 1 is true under the following weaker assumptions: The constraint map Φ: $X\to Y$ need not be differentiable on the whole space X; it is enough that $\Sigma=\Phi^{-1}(0)$ is a submanifold of X and Φ is differentiable in a neighborhood of x_0 in X. Furthermore we assume that $D\Phi(x_0)(X)$ is closed in Y and that the tangent space at x_0 to Σ is given by $T_{x_0}\Sigma=\operatorname{Ker} D\Phi(x_0)$. Then if x_0 is an extremum of F/Σ then there exist $\lambda_0\in\mathbf{R}$ and $y_0'\in Y'$ (Lagrange multiplier) such that

$$\lambda_0\frac{\delta F}{\delta x_0}+D\Phi(x_0)^*y_0'=0.$$

Proof: Under these conditions we have to consider in addition the case when $D\Phi(x_0)(X)$ is not the whole space Y. Denote by $L\equiv D\Phi(x_0)(X)\subset Y$. By the Hahn–Banach theorem there exists a $y_0'\in Y'$, $y_0'\neq 0$ such that $\langle y_0',L\rangle=0$. So for any $x\in X$ we have $\langle D\Phi(x_0)^*y_0',x\rangle_X=\langle y_0',D\Phi(x_0)x\rangle_Y=0$, since $D\Phi(x_0)x\in L$. Hence $D\Phi(x_0)^*y_0'=0$ and we can choose $\lambda_0=0$.

For the case where $D\Phi(x_0)=Y$ the previous proof of Theorem 1 works with the choice of $\lambda_0=-1$. ■

Remark 2: If $X'=X^*$ the dual space of X and the pairing is the natural one $\langle\ ,\ \rangle$: $X^*\times X\to\mathbf{R}$, $\langle\xi,x\rangle=\xi(x)$ then Corollary 1 becomes the usual Lagrange multiplier theorem in Banach spaces, e.g., Kolmogorov and Fomin.[1]

Assume we are in the situation of Theorem 1 and let us define a function H: $X\times Y'\to\mathbf{R}$ by

$$H(x,y')=F(x)+\langle y',\Phi(x)\rangle_Y,\quad x\in X,\quad y'\in Y'.$$

Theorem 2: If $x_0 \in \Sigma = \Phi^{-1}(0)$ is an extremum of F/Σ then:

(1) $\dfrac{\delta H}{\delta x_0} = 0,$

(2) $\dfrac{\delta H}{\delta y'} = 0.$

Proof: (1) We compute $\delta H/\delta x$ for $x \in X$: $\delta H/\delta x = \delta F/\delta x + \delta/\delta x \langle y', \Phi(x)\rangle_Y$. Let us denote $h(y',x) = \langle y', \Phi(x)\rangle$ and compute $\delta h/\delta x$, which is defined by $D_2 h(x)\cdot v = \langle \delta h/\delta x, v\rangle_X$, $v \in X$, where $D_2 h$ denotes the partial derivative of h with respect to the second variable, so

$$\begin{aligned} D_2 h(y',x)\cdot v &= \frac{d}{dt}\Big|_{t=0} h(y', x+tv) \\ &= \left\langle y', \frac{d}{dt}\Big|_{t=0} \Phi(x+tv)\right\rangle_Y \\ &= \langle y', D\Phi(x)\cdot v\rangle_Y = \langle D\Phi(x)^* y', v\rangle_X. \end{aligned}$$

Hence $\delta h/\delta x = D\Phi(x)^* y'$, and we get $\delta H/\delta x = \delta F/\delta x + D\Phi(X)^* y'$. If $x_0 \in \Sigma$ is an extremum of F/Σ we get from Theorem 1 that $0 = \delta F/\delta x_0 + D\Phi(x_0)^* y' = \delta H/\delta x_0$.

(2) We compute $\delta H/\delta y'$ for $y' \in Y'$: $\delta H/\delta y' = (\delta/\delta y')\langle y', \Phi(x)\rangle_Y = \delta h/\delta y'$, which is defined by $D_1 h(y')\cdot w = \langle w, \delta h/\delta y'\rangle_Y$, where $D_1 h$ denotes the partial derivative of h with respect to the first variable, i.e.,

$$\begin{aligned} D_1 h(y')\cdot w &= \frac{d}{dt}\Big|_{t=0} h(y'+tw, \Phi(x)) \\ &= \frac{d}{dt}\Big|_{t=0} \langle y'+tw, \Phi(x)\rangle_Y \\ &= \langle w, \Phi(x)\rangle_Y; \end{aligned}$$

hence $\delta h/\delta y' = \Phi(x)$. If $x_0 \in \Sigma$ we get $\delta H/\delta y' = \Phi(x_0) = 0$, which recovers the constraint equation. ∎

In many examples the constraint map is a real valued function $\Phi: X \to \mathbf{R}$. In this special case, i.e., where $Y = \mathbf{R}$, our theorems take simpler forms. Let $\langle\ ,\ \rangle: X' \times X \to \mathbf{R}$ be a weakly nondegenerate pairing between the Banach spaces X' and X and let $\Phi: X \to \mathbf{R}$ be a C^1 constraint map and $0 \in \mathbf{R}$ be a regular value of Φ. Then $\Sigma = \Phi^{-1}(0)$ is a submanifold of X and at each point $x \in \Sigma$ the tangent space to Σ is $T_x\Sigma \cong \operatorname{Ker} D\Phi(x)$, which splits in $T_x X \cong X$. A vector $x' \in X'$ is called *perpendicular* to Σ at the point $x \in \Sigma$ if $\langle x', h\rangle = 0$ for all $h \in T_x\Sigma$. With this terminology we get the analog result as in finite dimensions.

Lemma 3: For any $x \in \Sigma$, $\delta\Phi/\delta x$ is perpendicular to Σ at x.

Proof: Let $h \in T_x\Sigma = \operatorname{Ker} D\Phi(x)$, then $0 = D\Phi(x)\cdot h = \langle \delta\Phi/\delta x, h\rangle$. ∎

Lemma 4: Let $F: X \to \mathbf{R}$ be C^1. If $x_0 \in \Sigma$ is an extremum of F/Σ then $\delta F/\delta x_0$ is perpendicular to Σ at x_0.

Proof: Let $h \in T_{x_0}\Sigma = \operatorname{Ker} D\Phi(x_0)$ and choose a curve $\gamma: \mathbf{R} \to \Sigma$ such that $\gamma(0) = x_0$ and $\dot\gamma(0) = h$. Then if x_0 is an extremum of F/Σ the function $f(t) = (F\circ\gamma)(t): \mathbf{R} \to \mathbf{R}$ has an extremum at $t = 0$, hence $f'(0) = 0$. So

$$\begin{aligned} 0 &= \frac{d}{dt}\Big|_{t=0} (F\circ\gamma)(t) = DF(\gamma(0))\cdot\dot\gamma(0) \\ &= DF(x_0)\cdot h = \left\langle \frac{\delta F}{\delta x_0}, h\right\rangle. \end{aligned}$$ ∎

Corollary 2: If $x_0 \in \Sigma$ is an extremum of F/Σ then $\langle \delta F/\delta x_0, h\rangle = 0 = \langle \delta\Phi/\delta x_0, h\rangle$ for all $h \in T_{x_0}\Sigma$, i.e., there exists a number $\lambda \in \mathbf{R}$ such that

$$\frac{\delta F}{\delta x_0} + \lambda\frac{\delta\Phi}{\delta x_0} = 0.$$

This result resembles the classical finite-dimensional Lagrange multiplier theorem, which states that at an extremum $x_0 \in \mathbf{R}^n$ of a function $F: \mathbf{R}^n \to \mathbf{R}$ with respect to a constraint $\Phi: \mathbf{R}^n \to \mathbf{R}$ we have

$$\nabla F(x_0) + \lambda\,\nabla\Phi(x_0) = 0$$

for some $\lambda \in \mathbf{R}$.

If we define as in the general case the function $H: X \times \mathbf{R} \to \mathbf{R}$ by $H(x,\lambda) = F(x) + \lambda\Phi(x)$, we obtain from Theorem 2:

Corollary 3: If x_0 is an extremum of F/Σ then

(1) $\dfrac{\delta H}{\delta x_0} = \dfrac{\delta F}{\delta x_0} + \lambda\dfrac{\delta\Phi}{\delta x_0} = 0$

(2) $\dfrac{\delta H}{\delta\lambda} = \Phi(x_0) = 0.$

Example 6: This example of a constraint variational problem is a generalization of the Euler–Lagrange equation where the constraint is given by the time derivative of the position equal to the velocity. Let (X,X') and (Y,Y') be dual pairs of Banach spaces with $X = U \times V$ and $Y = V$. Let $F: U \times V \to \mathbf{R}$ be C^1 and $\Phi: U \times V \to V$ be linear of the form $\Phi(u,v) = Au + Bv$ where $A: U \to V$ and $B: V \to V$ are linear operators. Then we get from Theorem 2:

$$(*)\begin{cases} \dfrac{\delta F}{\delta u_0} = -A^* y', \\ \dfrac{\delta F}{\delta v_0} = -B^* y', \\ Au_0 + Bv_0 = 0. \end{cases}$$

Suppose B^* is invertible, then

$$\begin{aligned} &y' = -B^{*-1}\frac{\delta F}{\delta v_0}, \\ &\frac{\delta F}{\delta u_0} - A^* B^{*-1}\frac{\delta F}{\delta v_0} = 0, \\ &Au_0 + Bv_0 = 0. \end{aligned}$$

For a concrete application we consider the following: Let $U = \{u \in C^1([0,1],\mathbf{R}) \mid u(0) = u(1) = 0\}$ with the usual C^1-norm; i.e.,

$$\|u\|_1 = \sup_{t\in[0,1]} |u(t)| + \sup_{t\in[0,1]} |u'(t)|,$$

and let $V = \{v \in C^0([0,1],\mathbf{R}) \mid v(0) = v(1) = 0\}$ with the usual C^0-norm; i.e.,

$$\|v\|_0 = \sup_{t\in[0,1]} |v(t)|.$$

Let $X = U \times V$ and $F: X \to \mathbf{R}$ be given by $F(u,v)$

$= \int_0^1 L(u(t),v(t))dt$, for some Lagrangian $L(u,v)$. Let $Y = V$ and Φ: $U \times V \to V$ be given by $\Phi(u,v) = (d/dt)u - v$. Then $A = (d/dt)$: $U \to V$ is bounded; indeed

$$\|Au\|_0 = \|\frac{d}{dt} u\|_0 = \sup_{t\in[0,1]} |u'(t)|$$

$$\leqslant \sup_{t\in[0,1]} |u'(t)| + \sup_{t\in[0,1]} |u(t)| = \|u\|_1.$$

Moreover $B = -Id$: $V \to V$, hence Φ is bounded.

We return for a moment to the general case: The Lagrange multiplier is defined only on the stationary points. If a differentiable function can be found that extends the Lagrange multiplier to the entire space y': $X \to Y$ then one can pose a related variational problem as follows. Defining $H(x) = F(x) + \langle y'(x),\Phi(x)\rangle$ one finds stationary points of H by

$$\frac{\delta F}{\delta x_0} + D\Phi(x_0)^* y'(x_0) + Dy'(x_0)^*\Phi(x_0) = 0.$$

This problem has the same solution as the original constraint problem on the constraining surface but may have other stationary points that are not on the surface. If not, we call the functional F *regular* with respect to Φ, or Φ *regular*.

We apply the preceding remark to our example. An extension of the Lagrange multiplier is given by $y' = -B^{*-1}(\delta F/\delta y)$ and a computation shows that F being regular is equivalent to the Hessian being nonsingular. We get the following two equations:

$$\frac{\delta F}{\delta u_0} - A^* B^{*-1} \frac{\delta F}{\delta v_0} - \left[D_1 B^{*-1} \frac{\delta F}{\delta v_0}\right]^* \Phi(u_0,v_0) = 0,$$

$$\left[D_2 B^{*-1} \frac{\delta F}{\delta v_0}\right]^* \Phi(u_0,v_0) = 0.$$

If $\mathrm{Ker}[D_2 B^{*-1}(\delta F/\delta v_0)]^* = 0$ then $\Phi = 0$ is the only solution of the second equation and therefore F is regular.

Example 7: Let $X = \mathbf{H}$ be a real Hilbert space and $Y = \mathbf{R}, X' = X^* \cong \mathbf{H}, Y' = Y^* \cong \mathbf{R}$. Let $F(x) = (x,Ax)$ and $\Phi(x) = (x,x) - 1$, where A is a bounded self-adjoint operator on $\mathbf{H}$. All the hypotheses of Theorem 1 are satisfied and one computes easily: $D\Phi(x)\cdot h = 2(h,x)$, $\delta F/\delta x = 2Ax$, $D\Phi(x)^*$: $\mathbf{R} \to \mathbf{H}$; $D\Phi(x)^*\lambda = \lambda x$ and from Theorem 1 we get $Ax = -\lambda x$.

Example 8: In this example we derive Maxwell's equations from a variational principle with constraints. The electromagnetic potential will appear as the Lagrange multiplier and the homogeneous Maxwell equation as the condition of extremum of the variational principle. We put this example in a geometric setting. Let M be a four-dimensional Riemannian manifold and consider the Banach spaces $X = X' = \Lambda^2(M)$ and $Y = Y' = \Lambda^1(M)$ with pairings as described in Example 5, $\langle \alpha,\beta\rangle = \int_M \alpha \wedge *\beta$ and $\delta = *d*$ the adjoint of d. Consider the action F: $X \to \mathbf{R}$, $F(\Omega) = \frac{1}{2}\langle\Omega,\Omega\rangle = \frac{1}{2}\int_M \Omega \wedge *\Omega$ and think of $\Omega \in \Lambda^2(M)$ as representing the electromagnetic field. Let $J \in \Lambda^1(M)$ be a fixed one-form representing the external current. Then the field Ω must satisfy $\delta\Omega = J$ and we get the constraint equation with Φ: $X = \Lambda^2(M) \to Y = \Lambda^1(M)$ as $\Phi(\Omega) = \delta\Omega - J = 0$. Since Φ is linear we have $D\Phi = \delta\Omega$ and hence for $(D\Phi)^*$: $Y' = \Lambda^1(M) \to X' = \Lambda^2(M)$ we get for any $A \in \Lambda^1(M)$, $\Omega \in \Lambda^2(M)$,

$$\langle (D\Phi)^* A,\Omega\rangle_X = \langle A,(D\Phi)\Omega\rangle_Y$$

$$= \langle A,\delta\Omega\rangle_Y = \langle dA,\Omega\rangle_X$$

or equivalently

$$\int_M (D\Phi)^* \wedge *\Omega = \int_M A \wedge D\Phi(*\Omega) = \int_M dA \wedge *\Omega,$$

hence $(D\Phi)^* A = dA$ for any $A \in \Lambda^1(M)$. Moreover $\delta F/\delta\Omega = \Omega$ and if $\Phi(\Omega) = 0$ and Ω is an extremum of F then by Theorem 1 there exists an $A \in \Lambda^1(M)$ such that

$$\Omega = \frac{\delta F}{\delta \Omega} = (D\Phi)^* A = dA,$$

i.e., $\Omega = dA$. Therefore the Lagrange multiplier is identified as the vector potential. Furthermore, from the constraint equation $\delta\Omega = J$ we get immediately $\delta J = 0$, which is the continuity equation for the current. So we have derived Maxwell's equations by making a variation of the field Ω instead of the potential A as in the usual treatment, by imposing the constraint condition on Ω in terms of the current J.

IV. VARIATIONAL PRINCIPLES AND SYMMETRY GROUPS

Next we study the variational principle in the case when it is invariant under some symmetry group G. Then the question is the following: For a fixed point of the action of G to be an extremum for a G-invariant function F, does it suffice to check the vanishing of the first variation of F with respect to variations that are *symmetric*? In other words, the question is whether a critical symmetric point is also a symmetric critical point. In Ref. 6 Palais gave a necessary and sufficient condition for this "Principle" to be valid for any Banach space. This is again formulated in terms of the dual space X^*. We generalize this result to the case where one is allowed to choose any generalized dual space X' of X with respect to any weakly nondegenerate pairing $\langle\ ,\ \rangle$: $X' \times X \to \mathbf{R}$.

Let G be a Lie group acting on the Banach space X by linear transformations ϕ_g: $X \to X$, $g \in G$, $\phi_g(x) = g\cdot x$ and assume that the dual action $\phi_g{}^*$: $X' \to X'$ exists. Denote by $\Sigma = \{x \in X \mid \phi_g(x) = x$, for all $g \in G\}$ the set of fixed points (symmetric points) of X under the action of G. Since the action is linear, Σ is a linear subspace of X. Similarly let $\Sigma_* = \{x' \in X' \mid \phi_g{}^*(x') = x'$, for all $g \in G\}$ denote the set of fixed points of the dual action. Σ_* is a linear subspace of X'.

Lemma 5: (Chain rule for the functional derivative): Let ϕ $X \to X$ be a differentiable map and F: $X \to \mathbf{R}$ be differentiable at $x_0 \in X$, then:

$$\frac{\delta(F\circ\phi)}{\delta x_0} = D\phi(x_0)^* \frac{\delta F}{\delta(\phi(x_0))}.$$

Proof: From the chain rule of the Frechet derivative $D(F\circ\phi)(x_0)\cdot h = DF(\phi(x_0))\cdot(D\phi(x_0)\cdot h)$ for any $h \in X$ we get, assuming that all functional derivatives and adjoints exist,

$$D(F\circ\phi)(x_0)\cdot h = \left\langle \frac{\delta(F\circ\phi)}{\delta x_0},h\right\rangle = \left\langle \frac{\delta F}{\delta(\phi(x_0))},D\phi(x_0)\cdot h\right\rangle$$

$$= \left\langle D\phi(x_0)^* \frac{\delta F}{\delta(\phi(x_0))},h\right\rangle$$

for all $h\in X$. ■

Lemma 6: Let F: $X\to\mathbf{R}$ be a smooth G-invariant function, i.e., $F\circ\phi_g = F$ for all $g\in G$. If $p\in\Sigma$ then $\delta F/\delta p\in\Sigma_*$.

Proof: From Lemma 1 and G invariance we get for all $g\in G$:

$$\frac{\delta F}{\delta p}=\frac{\delta(F\circ\phi_g)}{\delta p}=D\phi_g(p)^*\frac{\delta F}{\delta\phi_g(p)}.$$

Since ϕ_g is linear and $p\in\Sigma$, this equals $\phi_g{}^*(\delta F/\delta p)$. ■

This lemma says that the variation of an invariant function with respect to a symmetric point is symmetric with respect to the dual action. Next we show that the variation of any functional subject to a constraint, with respect to a critical point, is orthogonal to the constraint set.

Lemma 7: Let $W\subset X$ be a subspace and F: $X\to\mathbf{R}$ be differentiable. If $p\in W$ is a critical point of F/W then $\delta F/\delta p\in W^\perp$.

Proof: Let $h\in W$ and choose a curve γ: $\mathbf{R}\to W$ such that $\gamma(0)=p$ and $\dot\gamma(0)=h$. If p is a critical point of F/W then the function $f(t)=(F\circ\gamma)(t)$: $\mathbf{R}\to\mathbf{R}$ has a critical point at $t=0$, hence $f'(0)=0$. Therefore

$$0=\frac{d}{dt}\bigg|_{t=0}(F\circ\gamma(t))=DF(\gamma(0))\cdot\dot\gamma(0)$$

$$=DF(p)\cdot h=\left\langle\frac{\delta F}{\delta p},h\right\rangle,$$

hence $\langle\delta F/\delta p,h\rangle=0$ for all $h\in W$, i.e., $\delta F/\delta p\in W^\perp$.

If $W=X$ we get as a corollary that at any critical point p of F we have $\delta F/\delta p=0$. Indeed, since the pairing $\langle\ ,\ \rangle$ is weakly nondegenerate $\langle\delta F/\delta p,x\rangle=0$ for all $x\in X$ implies $\delta F/\delta p=0$.

Now we choose for the constraint subspace W the set Σ of fixed points of the G action and we obtain the following:

Theorem 3: Let G act linearly on the Banach space X and let F: $X\to\mathbf{R}$ be a smooth G invariant function. Let $p\in\Sigma$ be a (symmetric) critical point of F/Σ. Then p is a critical point of F if and only if $\Sigma_*\cap\Sigma^\perp=0$.

Proof: (1) Let $p\in\Sigma$ be a critical point of F/Σ. It follows from Lemma 6 that $\delta F/\delta p\in\Sigma_*$ and from Lemma 7 that $\delta F/\delta p\in\Sigma^\perp$. Now if p is also a critical point of F then $\delta F/\delta p=0$ therefore $\Sigma_*\cap\Sigma^\perp=0$. (2) Assume that $\Sigma_*\cap\Sigma^\perp=0$ and let $p\in\Sigma$ be a critical point of F/Σ. From Lemma 7 it follows that $\delta F/\delta p\in\Sigma^\perp$. Since F is G invariant we get from Lemma 6 that $\delta F/\delta p\in\Sigma_*$, hence $\delta F/\delta p=0$. ■

This theorem is a very simple check of whether a critical symmetric point is also a symmetric critical point in the linear case, i.e., for a linear action of a Lie group on a dual pair of Banach spaces. In order to generalize this theorem to G actions on a Banach manifold M, one can require that at each point p of M the action of G on M being linearizable about p. That means that there exists a local coordinate system about p in which the action is linear, and that in this linear coordinate system the condition of the (linear) Theorem 3 is satisfied, i.e., $\Sigma_*\cap\Sigma^\perp=0$. It is shown in Palais[6] that these hypotheses are satisfied for semisimple Lie groups acting real analytically on a finite-dimensional real analytic manifold. Therefore in these cases the theorem generalizes to manifolds.

We now prove a more general form of the "Principle," which is valid for a general nonlinear action on a Hilbert manifold. This is usually the framework of physical applications. In Remark 3 we will consider the case of an action on a Banach manifold.

Theorem 4: Let ϕ_g be a differentiable action of a group G on a Hilbert manifold M; let Σ be the set of fixed point of ϕ_g. Consider $(D\phi_g)^*(p)$ (* denotes the Hermitian conjugate) at $p\in\Sigma$ and the vector subspace Σ^* of T_pM given by $\Sigma^*=\{x\in T_pM\,|\,(D\phi_g)^*(p)x=x\}$. Then for any invariant functional F the condition $dF(h)=0$ for all $h\in\Sigma^*$ implies $dF\equiv0$.

Proof: This is an immediate consequence of Lemma 7 applied to the case where $W=\Sigma^*$, and of Theorem 3.

Theorem 4 is like the proposed generalized "Principle" in Ref. 6, the difference being that here we use the adjoint of the linearized action instead of the derivative itself. We test Theorem 4 on the two examples in Ref. 6 where the "Principle" in its original form fails.

Example 9: Let $G=\mathbf{R}$, $M=\mathbf{R}^2$ with the usual scalar product and

$$\phi_t\begin{pmatrix}x\\y\end{pmatrix}=\begin{pmatrix}1&t\\0&1\end{pmatrix}\begin{pmatrix}x\\y\end{pmatrix},\ \text{then}\ \Sigma\equiv\begin{pmatrix}x\\0\end{pmatrix},$$

$$D\phi_t{}^*=\begin{pmatrix}1&0\\t&1\end{pmatrix},\ \text{and}\ \Sigma^*\equiv\begin{pmatrix}0\\y\end{pmatrix}.$$

Theorem 4 states that for any invariant function F on $\mathbf{R}^2$, $\partial F/\partial y=0$ implies $DF=0$ on Σ. This is the case because any invariant function of this action is of the form $F(x,y)=f(y)$.

Example 10: Let $G=\mathrm{SO}(2,1)$ and $M=L_2(S^1)$ with the usual scalar product. The action is generated by the following operators:

$$A_1=-\sin(2\alpha)+\frac{1}{2}\cos(2\alpha)\frac{d}{d\alpha},$$

$$A_2=\frac{1}{2}\frac{d}{d\alpha},$$

$$A_3=\cos(2\alpha)+\frac{1}{2}\sin(2\alpha)\frac{d}{d\alpha}.$$

It follows immediately that $\Sigma=0\subset L_2(S^1)$ and that $D\phi_g^*=\phi_g^*$ is generated by the negative Hermitian conjugate operators:

$$-A_1^*=\frac{1}{2}\cos(2\alpha)\frac{d}{d\alpha},$$

$$-A_2^*=\frac{1}{2}\frac{d}{d\alpha},$$

$$-A_3^*=\frac{1}{2}\sin(2\alpha)\frac{d}{d\alpha}.$$

From these we get $\Sigma^*\equiv\{f=\text{const}\}$. Theorem 4 states that at the origin any invariant functional has an extremum if and only if it has vanishing derivative in the direction of this constant function.

Theorem 4 reduces to the "Principle" in the linear or linearizable case for compact group actions or for semisimple groups acting on finite-dimensional Hilbert manifolds. In these cases Σ is in fact a submanifold of M and $D\phi_g$ is completely reducible and therefore Σ^*

$= \{x | D\phi_g^*(p)x = x\} = \{x | D\phi_g(p)x = x\} \equiv T_p\Sigma, p\in\Sigma$.

Remark 3: For an action on a Banach manifold, Σ^* cannot be identified with a subspace of $T\Sigma$. One can recover an analog of Theorem 4 by considering a topological complement W of Σ^*, i.e., $W \oplus \Sigma^* = X'$ and replacing Σ^* with $W^\perp \subset X$.

Remark 4: When the action is not linearizable, Theorem 4 (while still true) may reduce to a trivial statement as in this example: Let $G = \mathbf{R}$ and $M = \mathbf{R}$ with the usual scalar product. Let

$$\phi_t \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 & yt \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}.$$

Then

$$\Sigma \equiv \begin{pmatrix} x \\ 0 \end{pmatrix}$$

and on Σ we have

$$D\phi_t^* = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \text{ and } \Sigma^* \equiv \begin{pmatrix} x \\ y \end{pmatrix} \equiv \mathbf{R}^2.$$

Theorem 4 states that if both partial derivatives are 0 then $DF = 0$.

Remark 5: This remark is based on an argument about finding symmetric solutions by Coleman[7] in his Erice lectures. Let G act on the Banach manifold M and let $F: M \to \mathbf{R}$ be an invariant functional with Σ its set of fixed points. From G invariance of F, $F \circ g = F$, we get $TF \circ Tg = TF$, where T denotes the tangent map. Computed at any point $x\in\Sigma$ we get locally

$$F'(x)Dg(x) = F'(x). \qquad (*)$$

This holds for any action of G, linearizable or not. Here $Dg(x)$ is a linear representation of the group G. If the representation $Dg(x)$ is completely reducible then let $[Dg(x)]_\alpha$ denote the irreducible pieces, $Dg(x) = \Sigma\, [Dg(x)]_\alpha$ and $(*)$ becomes

$$F'(x)[Dg(x)]_\alpha = F'(x). \qquad (**)$$

Equation $(**)$ states that $F'(x)$ intertwines the irreducible representations $[Dg(x)]_\alpha$ of G with the trivial representation on $\mathbf{R}$. Now Schur's lemma implies that $F'(x) \neq 0$ only on a subspace E of $T_x\Sigma$ on which $Dg(x)$ acts trivially, i.e., $E = \{v\in T_xM \,|\, Dg(x)v = v\}$. If P denotes the projection onto E, $P: T_xM \to E$, we get $F'(x)h = F'(x)Ph$. Taylor's formula for F then becomes

$$F(x+h) = F(x) + F'(x)Ph + \tfrac{1}{2}F''(x)(h,h) + \epsilon.$$

This shows that x is a (nonrestricted) extremum if $F'(x)h = 0$ for all $h\in E$, i.e., one defines x to be a critical point of F/Σ to mean that $F'(x)Ph = 0$. Therefore the "Principle" holds. This is true whether Σ is a manifold or not (Palais[7]). If Σ is a smooth manifold and the action is linearizable at all points $x\in\Sigma$ then $E = T_x\Sigma$. In this situation the "Principle" holds in its usual interpretation because $F'(x)Ph = 0$ implies x is an extremum of F/Σ.

ACKNOWLEDGMENTS

We are grateful to G. Marmo and E. J. Saletan for many useful discussions. This work was done while R. S. was visiting the Dipartimento di Scienze Fisiche of the University of Naples and he wishes to express thanks for its hospitality.

This research was supported in part by the Emory Research Grant #2-50027 and the Emory Summer Fellowship, the Dipartimento di Scienze Fisiche of the University of Naples, the INFN, and the GNFM del CNR (Italy).

[1]A. N. Kolmogorov and S. V. Fomin, *Elementi di Teoria delle Funzioni e di Analisi Funzionale* (Mir, Moscow, 1980).

[2]R. Weinstock, *Calculus of Variations* (Dover, New York, 1974).

[3]E. T. Whittaker, *A Treatise on the Analytical Dynamics of Particles and Rigid Bodies* (Cambridge U.P., Cambridge, 1959).

[4]H. Goldstein, *Classical Mechanics* (Addison-Wesley, Reading, MA, 1980).

[5]R. Schmid, *Infinite Dimensional Hamiltonian Systems* (Biblionopolis, Naples, 1987).

[6]R. S. Palais, Commun. Math. Phys. **69**, 19 (1979).

[7]S. Coleman, *Classical Lumps and their Quantum Descendants*, Lectures International School of Subnuclear Physics (Ettore Majorana, Erice, 1975).

Conditions for runaway phenomena in the kinetic theory of particle swarms

Giovanni Frosali
Istituto di Matematica, Università di Ancona, Via Brecce Bianche, I-60100 Ancona, Italy
Cornelis V. M. van der Mee
Department of Mathematical Sciences, University of Delaware, Newark, Delaware 19716
Stefano L. Paveri-Fontana
Dipartimento di Metodi e Modelli Matematici, Università di Roma "La Sapienza," Via Scarpa 10, I-00161 Roma, Italy

(Received 16 October 1987; accepted for publication 21 December 1988)

The velocity distribution of a spatially uniform diluted guest population of charged particles moving within a host medium under the influence of a D. C. electric field is studied. A simplified one-dimensional Boltzmann model of the Kač type is adopted. Necessary conditions and sufficient conditions are established for the existence, uniqueness, and attractivity of a stationary non-negative distribution corresponding to a specified concentration level. Conditions for the onset of the runaway process are established.

I. INTRODUCTION AND STATEMENT OF THE PROBLEM

In this paper we are concerned with some mathematical aspects of the behavior of a population of charged particles under the influence of a spatially uniform D. C. (i.e., time independent) electric field. Problems of this type appear in a number of distinct scientific areas, e.g., in the theory of swarms of charged particles in a neutral background gas,[1,2] in the study of "runaway" electrons in fully ionized plasmas,[3–5] in the calculation of D. C. conductivity in biological membranes,[6] and in semiconductor theory. In many of these cases the charged particles of interest are electrons; however, in some instances ions or positive vacancies are considered as well.

Let us consider a spatially uniform dilute population of charged particles that are initially at thermal equilibrium with a neutral environment. Suppose that for times $t \geqslant 0$ a uniform D. C. electric field is applied to the system. The charged particles are accelerated by the electric field but return some of the acquired kinetic energy to the host medium via some interaction process (collisions). The heating of the host medium is assumed to be negligible enough for the temperature of the background host medium to remain approximately time independent. Further, we assume the existence of a balance between ionization of host particles and recombination of charged particles, so that the total number of charged particles appears to be invariant.

Two main physical situations may occur: (i) the collision process is sufficiently effective to force the velocity distribution of the charged particles towards a steady state nonzero profile, which is usually a heavily distorted Maxwellian at a temperature exceeding the reference temperature of the background gas, or (ii) the collision process is not effective in removing kinetic energy from the population of charged guest particles, so that no relaxation of the distribution function towards a nonzero steady state distribution occurs. On the contrary, a "travelling wave in velocity space" is generated (the so-called runaway case). In case (i) the contribution to D. C. conductivity of the guest particles is the ratio of the magnitude of the current due to their motion (in steady state conditions) to the intensity of the D. C. field; in the runaway case (ii) one does not have a (finite) D. C. conductivity, since the speed of the charged particles increases indefinitely. In the Appendix we present two simple model problems, based on the BGK approximation, to illustrate the two kinds of behavior. A more sophisticated model problem has been presented by Corngold and Rollins.[7]

The physical aspects of the picture sketched above have been well understood for a number of years.[2,3] For instance, it is recognized that the key ingredient in determining whether a given process will involve "relaxation" [case (i)] or "runaway" [case (ii)] is the dependence of the collision frequency $\nu(v)$ upon the speed v of the charged particles for large values of v. Indeed, if $\nu(v)$ drops towards zero too rapidly as $v \to \infty$, the collision process can be shown[2] to be unable to slow down the most energetic charged particles. As a consequence, these particles "runaway."

In spite of this body of existing knowledge, we feel that the mathematical aspects of the runaway process—as opposed to the strictly phenomenological physical ones—still require some study. For one thing, the approximations adopted in the literature are often so drastic[4] as to make one wonder about the reliability of the results (beyond, maybe, an order of magnitude level of precision). On the other hand, at a more fundamental level, even the physical outline given above is open to some criticism. In fact, one could consider intermediate cases [besides the cases (i) and (ii) given above]. For instance, one could construct an ad hoc model according to which the charged particle distribution function relaxes towards an asymptotic profile whose first (or second) velocity moment is unbounded; then, the drift velocity (or the temperature) would diverge even under case (i) conditions. Conversely, under case (ii) conditions one could envisage, as an alternative to the travelling wave in velocity space, a distribution function which relaxes (uniformly with respect to velocity) towards zero as time grows; under such conditions the velocity moments may or may not converge to finite values as $t \to \infty$. Thus there are cases in which the distinction between the runaway and the "absence of runaway" situation becomes blurred.

 0022-2488/89/051177-10$02.50

Other instances of confusion can be encountered. For instance, one author has erroneously presented estimates of the D. C. conductivity even in cases when the steady state distribution fails to exist (see the cases $p < -1$ in Ref. 6).

These questions have motivated the present introductory study on some mathematical aspects of the behavior of a collection of charged particles under the influence of an electric field. At this point we would like to present some additional remarks. First of all, we recall that—as observed by Corngold and Rollins[7]—much of the literature in the field deals with the problem of a steady state population of charged particles generated by a time independent source of cold particles. It should be noted that the two cases mentioned above for the sourceless problem—namely, case (i) of no runaways and case (ii) where runaways are present—correspond to the impossibility or the existence of a steady state population, respectively, when the source is present. The sourceless point of view taken in this paper has been described above. Another question concerns the choice of the mathematical model to employ in the description of the collective dynamics of the population of charged particles. In this preliminary study we assume, somewhat artificially, that charged particles move on a straight line parallel to the electric field (cf. the celebrated Kač model[8] of the Boltzmann equation); moreover, we usually assume that the collision process is described by a collision integral; the differential counterpart has been studied by Corngold and Rollins.[7] Finally we would like to mention that one of the problems we face is that of deciding upon the mathematical environment to adopt. On the one hand, we may introduce an L_1 space of distribution functions with at all times a finite total number of charged particles. On the other hand, we may adopt an L_1 space of distribution functions with at all times a finite number of collisions between the charged particles and the host medium. In part for reasons of mathematical convenience, we have made the former choice, especially as the general solution of the time dependent problem will turn out to have both a finite total number of charged particles and a finite number of guest-host interactions at all times. For the steady state problem we will be in the same rather fortunate situation, provided we assume the charged particle cross section $\nu(v)\geqslant 0$ to be nonintegrable with respect to velocity [in the sense that $\int_{-\infty}^{+\infty}\nu(v)dv = +\infty$]. On the other hand, if the cross section is integrable with respect to velocity [i.e., if $\int_{-\infty}^{+\infty}\nu(v)dv < +\infty$], we will have the rather unphysical situation of a "steady state" with finite total number of collisions but a nonzero particle density for infinite speed. We will make our assumptions more precise below.

Thus let us consider the simplified linear Boltzmann equation

$$\frac{\partial f}{\partial t}(v,t) + a\frac{\partial f}{\partial v}(v,t) + \nu(v)f(v,t) = \int_{-\infty}^{+\infty} k(v,v')\nu(v')f(v',t)dv'. \tag{1.1}$$

This equation describes the electron distribution $f(v,t)$ in a weakly ionized host medium as a function of the velocity $v\in(-\infty,+\infty)$ and time $t\geqslant 0$. The electrostatic acceleration a is assumed constant and positive. Recombination and ionization effects are assumed to balance each other. The expressions $\nu(v)$ and $k(v,v')$ denote the collision frequency (between an electron and the host medium) and the corresponding scattering kernel; accordingly, $k(v,v')dv$ is the probability that an electron entering the collision with velocity v' will come out of the collision with its velocity in the interval $(v,v+dv)$. We have

$$k(v,v')\geqslant 0, \tag{1.2}$$

$$\int_{-\infty}^{+\infty} k(v,v')dv\equiv 1. \tag{1.3}$$

The electron distribution $f(v,t)$ and the collision frequency $\nu(v)$ must, of course, be non-negative. By reciprocity symmetry, we also have

$$\nu(-v) = \nu(v), \tag{1.4}$$

$$k(-v,-v') = k(v,v'). \tag{1.5}$$

In connection with Eq. (1.1), we will study two mathematical problems. In the first place we will prove the unique solvability of the time-dependent evolution equation (1.1) under the initial condition

$$f(v,0) = f_0(v) \tag{1.6}$$

in a suitable functional setting, as well as the non-negativity of the solution for a non-negative initial condition, and establish the appropriate semigroup properties and bounds on the solution. In the second place we will establish necessary and sufficient conditions for the existence of a (unique) non-negative solution of the corresponding stationary equation

$$a\frac{\partial f}{\partial v}(v) + \nu(v)f(v) = \int_{-\infty}^{+\infty} k(v,v')\nu(v')f(v')dv'. \tag{1.7a}$$

We add the plausible requirements of a finite electron concentration and a finite collision rate (per unit volume); namely, we require

$$\int_{-\infty}^{+\infty} f(v)dv < +\infty, \tag{1.7b}$$

$$\int_{-\infty}^{+\infty} \nu(v)f(v)dv < +\infty. \tag{1.7c}$$

An additional plausible requirement is that in velocity space there should be no electrons entering or leaking out from the system. Since the acceleration a takes the role of "velocity" in velocity space, we require

$$\lim_{v\to-\infty} af(v) = \lim_{v\to+\infty} af(v) = 0. \tag{1.7d}$$

Hence

$$f(-\infty) = f(+\infty) = 0. \tag{1.7e}$$

Along with it we will establish under which conditions the stationary solution can be obtained from the solution of the time-dependent problem at $t\to\infty$.

In this paper we will investigate both the stationary and the time-dependent problem, as well as the decay to equilibrium of the solution in the time-dependent case. The time-dependent problem was already solved in Sec. XIII.4 of Ref. 9 as an application of the theory of time-dependent kinetic equations of Beals and Protopopescu (see Ref. 10; also Ref. 9, Chap. XI and Sec. XII.1–2). Here we shall give a direct proof based on semigroup considerations, which does not rely on this theory. In part we shall recover a well-known result. Note that, if the collision frequency is unbounded, the initial-value problem cannot be treated directly within the framework of Refs. 9 and 10; however, our proof will extend to this case. In fact, we will develop one of the few theories of kinetic equations where the usual cutoff leads to an unbounded collision frequency and an unbounded gain part of the collision operator. Different theories of this type were recently developed, for linearized Maxwell–Boltzmann equations by Arlotti[11] and for Fokker–Planck type equations by Cosner *et al.*[12]

Prior to developing the proper functional formulation of the problem, we make a number of assumptions on a, $\nu(v)$, and $k(v,v')$. Concerning a and ν we have:

Assumption (i): the acceleration a is a fixed positive constant;

Assumption (ii): the collision frequency $\nu(v)$ is a Lebesgue measurable, non-negative, and even function of v on $(-\infty,+\infty)$, which is almost everywhere nonzero and Lebesgue integrable on every bounded Lebesgue measurable set.

It is more complicated to formulate proper conditions on $k(v,v')$. On the one hand, we shall consider measurable functions $k(v,v')$ on $\Re^2$ satisfying (1.2), (1.3), and (1.5); on the other hand, we would like to include $k(v,v') = \delta(v-\alpha v')$ in our description. For this reason we shall consider the Banach spaces $L_1(\Re,dv)$ and $L_1(\Re,\nu\, dv)$ with the norms

$$\|f\|_1 = \int_{-\infty}^{+\infty} |f(v)|dv,$$

$$\|f\|_\nu = \int_{-\infty}^{+\infty} \nu(v)|f(v)|dv,$$

and postulate the following assumption.

Assumption (iii): The operator K which is formally represented as

$$(Kf)(v) = \int_{-\infty}^{+\infty} k(v,v')\nu(v')f(v')dv'$$

is a positive linear operator $K: L_1(\Re,\nu\, dv) \to L_1(\Re,dv)$ satisfying

$$\|Kf\|_1 = \|f\|_\nu, \text{ if } f\in L_1(\Re,\nu\, dv) \text{ and } f\geqslant 0, \tag{1.8a}$$

as well as the reciprocity principle

$$(Kf)(v) = (Kg)(-v), \quad \text{if } f(v) = g(-v) \text{ and } f\in L_1(\Re,\nu\, dv). \tag{1.8b}$$

If we define the (distributional) derivative f' of a function f in $L_1(\Re,dv)$ by

$$\int_{-\infty}^{+\infty} f'(v)g(v)dv = -\int_{-\infty}^{+\infty} f(v)g'(v)dv,$$

for every $g\in C_c^1(\Re)$, where $C_c^1(\Re)$ is the set of continuously differentiable complex functions on $\Re$ of compact support, *by a solution of the stationary equation* (1.7a) *we mean a function* φ *satisfying*

$$\varphi'(v) = -(1/a)\nu(v)\varphi(v) + (1/a)(K\varphi)(v), \quad v\in\Re, \tag{1.9a}$$

$$\varphi\in L_1(\Re,\nu dv). \tag{1.9b}$$

Since such a solution obviously has its derivative in $L_1(\Re,dv)$, each solution of the stationary problem will be absolutely continuous on $[-b,b]$ for all $b>0$. We have seen above that a physically acceptable solution ought to be non-negative and to obey requirements (1.7b), (1.7c), (1.7d), and (1.7e). Accordingly, among the possible solutions of problem (1.9) we shall be mostly interested in those non-negative solutions φ which also belong to $L_1(\Re,dv)$ and satisfy $\varphi(-\infty) = \varphi(+\infty) = 0$. The following theorem gives a necessary condition for the existence of a non-negative solution of (1.9) in $L_1(\Re,dv)$.

Theorem 1: Let a, ν, and K satisfy the assumptions (i), (ii), and (iii) stated above. Then a necessary condition for problem (1.9) to admit a nontrivial non-negative solution $\varphi\in L_1(\Re,dv)$ is that

$$\int_{-\infty}^{+\infty} \nu(v)dv = +\infty.$$

Proof: Let φ be a nontrivial non-negative solution of problem (1.9). Then φ is continuous on $\Re$ and there exists $v_0\in\Re$ such that $\varphi(v_0)>0$. However, since K is positive, Eq. (1.9a) yields

$$\begin{aligned}\varphi(v) &= \exp\left\{-\frac{1}{a}\int_{v_0}^{v}\nu(v'')dv''\right\}\varphi(v_0) \\ &\quad + \frac{1}{a}\int_{v_0}^{v}\exp\left\{-\frac{1}{a}\int_{v'}^{v}\nu(v'')dv''\right\}(K\varphi)(v')dv' \\ &\geqslant \exp\left\{-\frac{1}{a}\int_{v_0}^{v}\nu(v'')dv''\right\}\varphi(v_0),\end{aligned}$$

so that

$$\liminf_{v\to+\infty}\varphi(v) \geqslant \exp\left\{-\frac{1}{a}\int_{v_0}^{\infty}\nu(v'')dv''\right\}\varphi(v_0).$$

Then $\varphi\in L_1(\Re,dv)$ only if $\int_{v_0}^{\infty}\nu(v'')dv'' = +\infty$, i.e., only if $\int_{-\infty}^{+\infty}\nu(v'')dv'' = +\infty$. □

II. THE STATIONARY PROBLEM

In this section we shall discuss the stationary problem (1.9). Throughout this section, with the exception of the final part, we shall also make the additional assumption

$$\int_{-\infty}^{+\infty} \nu(v)dv = +\infty, \tag{2.1}$$

whose motivation is given by Theorem 1 above. Note that, as a consequence of assumption (ii), Eq. (2.1) characterizes the frequency behavior of $\nu(v)$ as $v\to\infty$. As observed above, (2.1) is equivalent to

$$\int_{\alpha}^{+\infty} \nu(v)dv = +\infty, \quad \text{for some } \alpha\in\Re.$$

Our first step is to convert the integrodifferential equation (1.9) into an (equivalent) integral equation. To this purpose, we define the following operator:

$$L: L_1(\Re,dv) \to L_1(\Re,\nu\, dv),$$

$$(Lf)(v) = \int_{-\infty}^{v}\frac{1}{a}\exp\left\{-\frac{1}{a}\int_{v'}^{v}\nu(v'')dv''\right\}f(v')dv'.$$

On integrating Lf with respect to the measure $\nu(v)dv$ and changing the order of integration we obtain

$$\|Lf\|_\nu = \int_{-\infty}^{+\infty}\left(1 - \lim_{v\to+\infty} \exp\left\{-\frac{1}{a}\int_{v'}^{v} \nu(v'')dv''\right\}\right)f(v')dv', \quad f\geqslant 0, \tag{2.2}$$

which implies that L is a positive contraction from $L_1(\mathfrak{R},dv)$ to $L_1(\mathfrak{R},\nu\, dv)$ and hence LK is a positive contraction on $L_1(\mathfrak{R},\nu\, dv)$. Under our assumptions, if $\nu(v)$ obeys (2.1), then $\|Lf\|_\nu = \|f\|_1$ for all non-negative $f\in L_1(\mathfrak{R},dv)$.

Theorem 2: If (2.1) holds, then every solution of the integrodifferential equation (1.9a) in $L_1(\mathfrak{R},\nu\, dv)$ is a solution of the linear stationary problem

$$\varphi = LK\varphi, \quad \varphi\in L_1(\mathfrak{R},\nu\, dv), \tag{2.3}$$

and conversely. Moreover, for every solution φ of the two equivalent problems we have $\varphi(-\infty) = \varphi(+\infty) = 0$.

Proof: Let φ be a solution of problem (1.9). Setting

$$H(v) = \exp\left\{\frac{1}{a}\int_{v_0}^{v} \nu(v')dv'\right\},$$

where v_0 is some real number, we obtain from (1.9),

$$(H\varphi)'(v) = (1/a)H(v)(K\varphi)(v), \quad v\in\mathfrak{R},$$

which in turn implies

$$H(v)\varphi(v) = \varphi(v_0) + \frac{1}{a}\int_{v_0}^{v} H(v')(K\varphi)(v')dv'.$$

Here we observe that the integral on the right-hand side is finite, because $H(v)$ is bounded on every interval of the type $(-\infty,A]$ with $A< +\infty$. As a result we find

$$\varphi(v) = \exp\left\{-\frac{1}{a}\int_{v_0}^{v} \nu(v')dv'\right\}\varphi(v_0) + \frac{1}{a}\int_{v_0}^{v}\exp\left\{-\frac{1}{a}\int_{v'}^{v}\nu(v'')dv''\right\}(K\varphi)(v')dv', \tag{2.4}$$

where $v\in\mathfrak{R}$. We now note that φ is of bounded variation on $\mathfrak{R}$, due to the fact that $\varphi'\in L_1(\mathfrak{R},dv)$. This obviously implies the boundedness of φ on $\mathfrak{R}$. Letting v_0 tend to $-\infty$ and taking account of (2.1) in combination with the boundedness of φ, we must have $\varphi(-\infty) = 0$ by dominated convergence. Similarly, if $v_0\to+\infty$, we get $\varphi(+\infty) = 0$. Thus any solution φ of problem (1.9) obeys $\varphi = LK\varphi$, with $\varphi(\pm\infty) = 0$.

Conversely, directly from the explicit form of $LK\varphi$, every solution of Eq. (2.3) in $L_1(\mathfrak{R},\nu\, dv)$ is absolutely continuous on $[-b,b]$ for all $b>0$ and of bounded variation on $(-\infty,+\infty)$. Moreover,

$$\frac{d}{dv}(LK\varphi)(v) = \frac{1}{a}(K\varphi)(v) - \frac{1}{a}\nu(v)(LK\varphi)(v)$$

and the right-hand side belongs to $L_1(\mathfrak{R},dv)$; hence the solution φ of Eq. (2.3) satisfies Eq. (1.9a). □

Theorem 3: If condition (2.1) is satisfied, then the set of all φ satisfying problem (2.3) is at most one dimensional and, when nontrivial, contains a nontrivial non-negative function.

Proof: Let us suppose that Eq. (2.3) admits solutions. Then every such solution is non-negative, apart from a constant factor. Indeed, if $\varphi = LK\varphi$ for some $\varphi\in L_1(\mathfrak{R},\nu\, dv)$, one first chooses φ real. We then have

$$|\varphi| = |LK\varphi|\leqslant LK|\varphi|,$$

while

$$\begin{aligned}\int_{-\infty}^{+\infty}\nu(v)\{(LK|\varphi|)(v) - |\varphi(v)|\}dv &= \|LK|\varphi|\|_\nu - \|\varphi\|_\nu\\ &= \|K|\varphi|\|_1 - \|\varphi\|_\nu\\ &= \|\varphi\|_\nu - \|\varphi\|_\nu = 0.\end{aligned}$$

Hence $|\varphi| = LK|\varphi|$ and the existence of a nontrivial solution of Eq. (2.3) in $L_1(\mathfrak{R},\nu\, dv)$ implies the existence of a nontrivial non-negative solution of Eqs. (1.9) in $L_1(\mathfrak{R},\nu\, dv)$. In fact, if the solution φ is real, then $\|LK\{|\varphi|\pm\varphi\}\|_\nu = 0$ and $LK\{|\varphi|\pm\varphi\} = \{|\varphi|\pm\varphi\}$ imply that φ does not change sign.

Finally, if φ is a nontrivial non-negative solution of Eq. (2.3) in $L_1(\mathfrak{R},\nu\, dv)$ and $\varphi(v) = 0$ for some $v\in\mathfrak{R}$, then $(LK\varphi)(v) = 0$ yields $(K\varphi)(v')\equiv 0$ for $v'<v$ and hence $\varphi(v'') = (LK\varphi)(v'') = 0$ for all $v''\leqslant v$. Putting $v_0 = \sup\{v\in\mathfrak{R}: \varphi(v'') = 0 \text{ for all } v''\leqslant v\}$ we find $\varphi(v)>0$ for all $v>v_0$, since otherwise $\varphi(v)$ would vanish for some $v>v_0$. Thus if φ_1 and φ_2 are two different non-negative solutions of Eq. (2.3) of unit norm in $L_1(\mathfrak{R},\nu\, dv)$, then $(\varphi_1-\varphi_2)$ will be a nontrivial real solution of Eq. (2.3), which must have constant sign. Since both φ_1 and φ_2 have unit norm in $L_1(\mathfrak{R},\nu\, dv)$, we obtain $\|\varphi_1-\varphi_2\|_\nu = |\|\varphi_1\|_\nu - \|\varphi_2\|_\nu|$, which is a contradiction. Hence the solution space of Eq. (2.3) is at most one dimensional. □

Let φ be a nontrivial non-negative stationary solution in $L_1(\mathfrak{R},dv)$. Then, apparently, either $\varphi(v)>0$ for all $v\in\mathfrak{R}$ or $\varphi(v) = 0$ for $v\leqslant v_0$ and $\varphi(v)>0$ for $v>v_0$ where v_0 is some real constant. In the latter case we have $(K\varphi)(v)\equiv 0$ for all $v<v_0$, as a result of the equation $\varphi = LK\varphi$. Since, by assumption, $\nu(v)$ does not vanish on a set of positive measure, we must then have $k(v,v')\equiv 0$ for all $v<v_0$ and $v'>v_0$.

Theorem 4: If condition (2.1) is satisfied and if, in addition, LK is a weakly compact operator on $L_1(\mathfrak{R},\nu\, dv)$, then the stationary problem has a unique non-negative solution in $L_1(\mathfrak{R},\nu\, dv)$ of unit norm.

Proof: If condition (2.1) is satisfied and f is non-negative, we have $\|LKf\|_\nu = \|f\|_\nu$. Consequently, the spectral radius of LK, $\mathrm{spr}(LK)$, is one. Moreover, $(LK)^2$ is compact as an operator on $L_1(\mathfrak{R},\nu\, dv)$, because the square of a weakly compact operator in L_1 is compact. Then the compactness of $(LK)^2$ in combination with $\mathrm{spr}(LK) = 1$ implies the existence of at least one non-negative $\varphi\in L_1(\mathfrak{R},\nu\, dv)$ of unit norm such that Eq. (2.3) holds true (see Ref. 13, Chap. 2). By the previous theorem this solution is unique. □

Corollary 5: Let condition (2.1) be satisfied. If the operator

$$(Bf)(v) = \int_{-\infty}^{+\infty} k(v,v')f(v')dv'$$

is weakly compact on $L_1(\mathfrak{R},dv)$ then the problem (2.3) has a unique non-negative solution of unit norm.

Proof: If the above operator B is weakly compact on $L_1(\mathfrak{R},dv)$, then $LK = LB\nu$ is weakly compact on $L_1(\mathfrak{R},\nu\, dv)$. Here we observe that ν is a bounded operator from $L_1(\mathfrak{R},\nu\,dv)$ into $L_1(\mathfrak{R},dv)$. The result follows directly from the previous theorem. □

We now consider some simple models for the collision term. They satisfy the assumptions we formulated to ensure the existence of stationary solutions.

Example 1 [*The Bhatnagar–Gross–Krook (BGK) model*]*:* The idea behind this model is the assumption that the average effect of collisions is to provide a "source" which is proportional to the deviation of the distribution function $f(v)$ from a Maxwellian $f_m(v)$. Thus the collision term is assumed to take the following form:

$$\int_{-\infty}^{+\infty} k(v,v')\nu(v')f(v')dv' = \nu(v)F_m(v)\int_{-\infty}^{+\infty} \nu(v')f(v')dv', \tag{2.5}$$

where

$$F_m(v) = \frac{f_m(v)}{\int_{-\infty}^{+\infty}\nu(v')f_m(v')dv'}.$$

In this case, the operator K defined by (2.5) is a compact operator from $L_1(\mathfrak{R},\nu\,dv)$ to $L_1(\mathfrak{R},dv)$. Hence, if condition (2.1) is satisfied, the results of Theorem 4 hold true. In the Appendix we will give a more elaborate account of the BGK model. For a discussion of the reliability of the BGK model in the transport theory of charged particles see, for instance, the paper of Corngold and Rollins.[14] □

Example 2: Consider a particular class of integral kernels $k(v,v')$, which is a finite linear combination of functions separated in the variables v and v'. A kernel of this type is said to be degenerate and can be written in the form

$$k(v,v') = \nu(v)f_m(v)f_m(v')\sum_{i=1}^{\infty}\alpha_i\psi_i(v)\varphi_i(v'), \tag{2.6}$$

where ψ_i and φ_i are given functions and the α_i are suitable positive accommodation coefficients. In the literature of the kinetic theory of gases this model is known as the generalized BGK model and is obtained by generalizing the linearized BGK model. If we suppose the functions φ_i and ψ_i to be essentially bounded, the operator K defined by (2.6) and (1.8) is a compact operator from $L_1(\mathfrak{R},\nu\,dv)$ into $L_1(\mathfrak{R},dv)$. □

Example 3: Consider the integral kernel defined by

$$k(v,v') = \begin{cases} 1/2r, & v\in[-r,r], \quad v'\in\mathfrak{R}, \\ 0, & \text{otherwise.}\end{cases}$$

Then the integral operator B defined by the above kernel has the property that for every $\epsilon>0$ there exists $\delta>0$ such that

$$\int_{-\infty}^{+\infty}|(Bf)(v+h)-(Bf)(v)|dv<\epsilon$$

for every f belonging to a bounded subset of $L_1(\mathfrak{R},dv)$ and every h with $|h|<\delta$. Moreover, there exists a subset $[-r,r]\subset G\subset\mathfrak{R}$ such that

$$\int_{\mathfrak{R}\setminus\overline{G}}|(Kf)(v)|dv<\epsilon,$$

is trivially satisfied for every $\epsilon>0$. From Theorem 2.1 of Ref. 15 it follows that B is compact on $L_1(\mathfrak{R},dv)$. By virtue of Corollary 5, we have a stationary solution if $\int_{-\infty}^{+\infty}\nu(v')dv' = +\infty$. More generally, we may replace $k(v,v')$ with a bounded continuous non-negative function with support on $[-r,r]\times\mathfrak{R}$. A sufficient condition for compactness will then be the existence, for every $\epsilon>0$, of a number $\delta>0$ such that

$$\int_{-r}^{+r}|k(v+h,v')-k(v,v')|dv<\epsilon$$

for $|h|<\delta$, uniformly in v' on $\mathfrak{R}$. □

In the remaining part of this section, we consider the case when the behavior of the collision frequency $\nu(v)$ at infinity is such that condition (2.1) is not satisfied. In other words, from now on in this section, we replace (2.1) by the alternative assumption

$$\int_{-\infty}^{+\infty}\nu(v)dv<+\infty. \tag{2.7}$$

If condition (2.7) is satisfied, then $\|Lf\|_\nu \leqslant(1-\delta)\|f\|_1$ for all non-negative $f\in L_1(\mathfrak{R},dv)$ where

$$\delta=\exp\left\{-\frac{1}{a}\int_{-\infty}^{+\infty}\nu(v')dv'\right\}>0.$$

In this case dominated convergence applied to Eq. (2.4) yields the existence of the continuous limits $\varphi(\pm\infty)$, whence the integrodifferential equation (1.9) can be put in the equivalent form

$$\varphi(v)-(LK\varphi)(v)=\exp\left\{-\frac{1}{a}\int_{-\infty}^{v}\nu(v')dv'\right\}\varphi(-\infty). \tag{2.8}$$

As a result we find that $\varphi(\pm\infty)$ are finite, while an easy integration of Eq. (1.9) over $\mathfrak{R}$ yields $\varphi(-\infty)=\varphi(+\infty)$. Now the integral equation to be investigated is Eq. (2.8). Equation (2.8) is uniquely solvable in $L_1(\mathfrak{R},\nu\,dv)$, which is easily seen from the norm estimate

$$\|LK\varphi\|_\nu\leqslant(1-\delta)\|K\varphi\|_1=(1-\delta)\|\varphi\|_\nu,$$

where $\delta\in(0,1)$. We summarize the result as follows.

Theorem 6: If condition (2.7) is satisfied, then the stationary problem (2.8) has a unique non-negative solution φ in $L_1(\mathfrak{R},v\,dv)$ of unit norm with $\varphi(-\infty)=\varphi(+\infty)>0$.

Remark: Note that the solution of (2.8) under assumption (2.7), which is referred to in Theorem 6, is physically irrelevant, since it corresponds to an infinite population level.

It is possible to give necessary and sufficient conditions for the existence of a stationary solution in $L_1(\mathfrak{R},\nu\,dv)$ in terms of the spectral properties of LK. The stationary solution will be unique apart from a normalization factor. If condition (2.1) holds true, the necessary and sufficient condition is that 1 is an eigenvalue of LK. The corresponding eigenfunction will then be non-negative. In particular, if $(LK)^n$ is compact for some $n\in\mathbb{N}$, there will be a stationary solution. On the other hand, if condition (2.7) is satisfied, there always exists a unique non-negative stationary (unphysical) solution in $L_1(\mathfrak{R},\nu\,dv)$ of unit norm, because $\mathrm{spr}(LK)<1$; its values at $\pm\infty$ are equal and positive.

Instead of Eqs. (2.3) and (2.8) on $L_1(\mathfrak{R},\nu\,dv)$, we may also study the equivalent equations

$$\psi-KL\psi=0, \tag{2.9}$$

$$\psi - KL\psi = \varphi(-\infty)K\omega, \qquad (2.10)$$

on $L_1(\Re,dv)$, where

$$\omega(v) = \exp\left\{ -\frac{1}{a}\int_{-\infty}^{v} \nu(v'')dv'' \right\}.$$

In fact, if φ is a (non-negative) solution of Eq. (2.3) in $L_1(\Re,\nu\, dv)$, then $K\varphi$ is a (non-negative) solution of Eq. (2.9) in $L_1(\Re,dv)$; conversely, if ψ is a (non-negative) solution of Eq. (2.9) in $L_1(\Re,dv)$, then $L\psi$ is a (non-negative) solution of Eq. (2.3) in $L_1(\Re,\nu\, dv)$. Moreover, in this manner nontrivial solutions of Eq. (2.3) in $L_1(\Re,\nu\, dv)$ correspond to nontrivial solutions of Eq. (2.9) in $L_1(\Re,dv)$. A similar connection exists between solutions of Eq. (2.8) in $L_1(\Re,\nu\, dv)$ and solutions of Eq. (2.10) in $L_1(\Re,dv)$, but now $K\varphi$ is a solution of Eq. (2.10) if φ is a solution of Eq. (2.8), while $\varphi(-\infty)\omega + L\psi$ is a solution of Eq. (2.8) if ψ is a solution of Eq. (2.10). However, since in general the operator K does not map absolutely continuous functions of $L_1(\Re,\nu\, dv)$ into continuous functions, the solutions of Eqs. (2.9) and (2.10) need not be continuous. On the other hand, if K (or the above operator B) has finite rank, it is much easier to solve Eqs. (2.9) and (2.10) than to solve Eqs. (2.3) and (2.8). Finally, it should be observed that the nonzero spectra and eigenvalue spectra of LK and KL coincide.

III. THE TIME-DEPENDENT PROBLEM

In order to study the time-dependent problem, we shall analyze the operator

$$(Tf)(v) = -a\frac{\partial f}{\partial v} - \nu(v)f(v) + (Kf)(v)$$

on the intersection $\mathscr{M}$ of $L_1(\Re,dv)$, $L_1(\Re,\nu\, dv)$ and the set of functions which are absolutely continuous on $[-b,b]$ for all $b>0$, are of bounded variation and vanish at $-\infty$. We shall prove an extension of T to be the generator of a strongly continuous semigroup on $L_1(\Re,dv)$ using the Hille–Phillips theorem. For this purpose we solve the equation

$$(\lambda - T)f = g \qquad (3.1)$$

for $f\in\mathscr{M}$, where g is an arbitrary function in $L_1(\Re,dv)$ and $\lambda>0$. We obtain

$$f = L_\lambda Kf + L_\lambda g, \qquad (3.2)$$

where

$$(L_\lambda f)(v) = \frac{1}{a}\int_{-\infty}^{v} \exp\left\{ -\frac{1}{a}\int_{v'}^{v} [\nu(v'') + \lambda]dv'' \right\} f(v')dv'.$$

The derivation of Eq. (3.2) is the same as the derivation of (2.3) with $\nu(v)$ replaced by $\nu(v)+\lambda$, since for $\lambda>0$ the integral $\int_{-\infty}^{+\infty}\{\nu(v)+\lambda\}dv$ is infinite. As a result we obtain

$$\|L_\lambda f\|_\nu + \lambda\,\|L_\lambda f\|_1 = \|f\|_1, \quad f\geqslant 0. \qquad (3.3)$$

Here we have replaced $\nu(v)$ by $\nu(v)+\lambda$ in the identity $\|Lf\|_\nu = \|f\|_1$ for $f\geqslant 0$. This is allowed, since $L_\lambda f$ coincides with Lf on replacing $\nu(v)+\lambda$. Consequently,

$$\|L_\lambda Kf\|_\nu + \lambda\,\|L_\lambda Kf\|_1 = \|f\|_\nu, \quad f\geqslant 0, \qquad (3.4)$$

whence L_λ maps $L_1(\Re,dv)$ and $L_\lambda K$ maps $L_1(\Re,\nu\, dv)$ into the intersection of $L_1(\Re,dv)$ and $L_1(\Re,\nu\, dv)$. Hence for every $\lambda>0$ and $g\in L_1(\Re,dv)$ the solutions f of Eq. (3.2) belong to this intersection.

Theorem 7: For every $\lambda>0$ and $g\in L_1(\Re,dv)$ there exists a unique solution $T_\lambda g$ of Eq. (3.2), which belongs to $L_1(\Re,dv)$. Then T_λ is the resolvent of a strongly continuous positive contraction semigroup $\{S(t)\}_{t>0}$ on $L_1(\Re,dv)$ whose generator G is a closed extension of T. Moreover, the semigroup $\{S(t)\}_{t>0}$ satisfies

$$\|S(t)f\|_1 = \|f\|_1, \quad f\geqslant 0, \qquad (3.5)$$

if and only if G is the (minimal) closure of T.

Proof: Put

$$T_\lambda g = \sum_{n=0}^{\infty}(L_\lambda K)^n L_\lambda g, \quad g\in L_1(\Re,dv).$$

Then for $g\geqslant 0$ in $L_1(\Re,dv)$ Eq. (3.4) implies

$$\lambda\,\|(L_\lambda K)^n L_\lambda g\|_1 = \|(L_\lambda K)^{n-1}L_\lambda g\|_\nu - \|(L_\lambda K)^n L_\lambda g\|_\nu, \quad n=1,2,3,\ldots,$$

and therefore for $g\geqslant 0$

$$\begin{aligned}\|T_\lambda g\|_1 &= \sum_{n=0}^{\infty}\|(L_\lambda K)^n L_\lambda g\|_1 = \|L_\lambda g\|_1 + \frac{1}{\lambda}\|L_\lambda g\|_\nu - \frac{1}{\lambda}\beta_\lambda(L_\lambda g)\\ &= \frac{1}{\lambda}\|g\|_1 - \frac{1}{\lambda}\beta_\lambda(L_\lambda g) \leqslant \frac{1}{\lambda}\|g\|_1.\end{aligned}$$

Here

$$\beta_\lambda(f) = \lim_{n\to\infty}\|(L_\lambda K)^n f\|_\nu, \quad 0\leqslant f\in L_1(\Re,\nu\, dv),$$

extends uniquely to a positive linear functional of $L_1(\Re,\nu\, dv)$. Thus there exists a non-negative function $\varphi_\lambda\in L_\infty(\Re,\nu\, dv)$ such that

$$\beta_\lambda(f) = \int_{-\infty}^{+\infty}\nu(v)f(v)\varphi_\lambda(v)dv, \quad f\in L_1(\Re,\nu\, dv).$$

Since obviously $\beta_\lambda(L_\lambda Kf) = \beta_\lambda(f)$ for all $f\in L_1(\Re,\nu\, dv)$, we have $(L_\lambda K)^*\varphi_\lambda = \varphi_\lambda$, where the adjoint is defined on $L_\infty(\Re,\nu\, dv)$.

To prove that 1 is not an eigenvalue of $L_\lambda K$, suppose $c(\lambda)$ is an eigenvalue of $L_\lambda K$. If φ is a corresponding eigenfunction, then

$$\begin{aligned}|c(\lambda)|\{\|\varphi\|_\nu + \lambda\,\|\varphi\|_1\} &= \|L_\lambda K\varphi\|_\nu + \lambda\,\|L_\lambda K\varphi\|_1\\ &\leqslant \|L_\lambda K|\varphi|\|_\nu + \lambda\,\|L_\lambda K|\varphi|\|_1\\ &= \||\varphi|\|_\nu = \|\varphi\|_\nu,\end{aligned}$$

which implies that $|c(\lambda)|\leqslant 1$. Moreover, $|c(\lambda)|\neq 1$, since otherwise $\|\varphi\|_1 = 0$ and thus $\varphi = 0$. Thus $f = T_\lambda g$ is the unique solution of Eq. (3.2) in $L_1(\Re,dv)$.

Clearly, T_λ is the resolvent of a bounded strongly continuous positive contraction semigroup on $L_1(\Re,dv)$, i.e., $T_\lambda = (\lambda - G)^{-1}$, where G is the generator and $D(G) = \operatorname{Ran} T_\lambda$. Moreover,

$$T_\lambda g = \int_0^\infty e^{-\lambda t}S(t)g\,dt, \quad g\geqslant 0 \text{ in } L_1(\Re,dv),$$

while

$$\|T_\lambda g\|_1 + (1/\lambda)\beta_\lambda(L_\lambda g) = (1/\lambda)\|g\|_1, \quad g \geqslant 0 \text{ in } L_1(\Re, dv).$$

Thus if $1 \notin \sigma_r(L_\lambda K)$, the residual spectrum of $L_\lambda K$, and hence $1 \notin \sigma_p((L_\lambda K)^*)$, we have $\beta_\lambda(L_\lambda g) = 0$ and therefore

$$\|T_\lambda g\|_1 = (1/\lambda)\|g\|_1, \quad g \geqslant 0 \text{ in } L_1(\Re, dv).$$

Hence

$$\|S(t)g\|_1 = \|g\|_1, \quad g \geqslant 0 \text{ in } L_1(\Re, dv).$$

Conversely, if the last two equations are true, $\beta_\lambda(L_\lambda g) = 0$ for all $g \geqslant 0$ in $L_1(\Re, dv)$. Since $\{L_\lambda g: g \in L_1(\Re, dv)\}$ is dense in $L_1(\Re, \nu\, dv)$, we get $\varphi_\lambda = 0$ and hence $1 \notin \sigma_r(L_\lambda K)$.

If Eq. (3.5) is true and $1 \notin \sigma_r(L_\lambda K)$, then $\mathbb{1} - L_\lambda K$ maps $L_1(\Re, \nu\, dv)$ into a dense subspace of $L_1(\Re, \nu\, dv)$ and Eq. (3.2) can be solved in $\mathscr{M}$ for all $g \in \mathscr{D}$, where $\mathscr{D}$ is a suitable dense subset of $L_1(\Re, dv)$. For every $g \in L_1(\Re, dv)$ and $\epsilon > 0$ we then choose $g_0 \in \mathscr{D}$ such that $\|g - g_0\|_1 < \epsilon\lambda / (\lambda + 1)$ so that

$$\|T_\lambda[g - g_0]\|_1 < \epsilon/(\lambda + 1).$$

Therefore $f = T_\lambda g \in D(\lambda - G)$ can be approximated by some $f_0 = T_\lambda g_0 \in \mathscr{M}$ such that

$$\|f - f_0\|_1 + \|(\lambda - G)f - (\lambda - T)f_0\|_1$$
$$= \|f - f_0\|_1 + \|g - g_0\|_1 < \epsilon.$$

Consequently, $(\lambda - G)$ is the closure of $(\lambda - T)$. On the other hand, if $(\lambda - G)$ is not the closure of $(\lambda - T)$, there exist $g \in L_1(\Re, dv)$ and $\epsilon > 0$ such that

$$\|T_\lambda g - f\|_1 + \|g - (\lambda - T)f\|_1 \geqslant \epsilon, \quad f \in \mathscr{M}.$$

Hence if $f = \Sigma_{n=0}^{N}(L_\lambda K)^n L_\lambda g$, we have

$$\|T_\lambda(KL_\lambda)^{N+1}g\|_1 + \|(L_\lambda K)^{N+1}g\|_1 \geqslant \epsilon.$$

By choosing N large enough, the first term can be made arbitrarily small, because the series defining $T_\lambda g$ converges absolutely in $L_1(\Re, dv)$. But then

$$\beta_\lambda(g) = \lim_{N \to \infty} \|(L_\lambda K)^{N+1}g\|_1 \geqslant \epsilon,$$

whence $1 \in \sigma_r(L_\lambda K)$. □

Remarks: (1) If there exists a nonzero stationary solution φ, then $\varphi \geqslant 0$ apart from a constant factor, $S(t)\varphi \equiv \varphi$ and $\lambda T_\lambda \varphi \equiv \varphi$. But then $\beta_\lambda(L_\lambda \varphi) = 0$ and hence $\beta_\lambda = 0$, which implies that $1 \notin \sigma_r(L_\lambda K)$ for all $\lambda > 0$. Consequently, the existence of a nonzero stationary solution implies (3.5).

(2) Another case when (3.5) is true occurs if $\nu(v)$ is essentially bounded. In that case $\|g\|_\nu \leqslant \|\nu\|_\infty \|g\|_1$ for all $g \in L_1(\Re, dv) \subseteq L_1(\Re, \nu\, dv)$ Thus if $\{\alpha_m\}_{m=1}^\infty$ is an increasing sequence in (0,1) with limit 1, then $f_m = (\mathbb{1} - \alpha_m L_\lambda K)^{-1} L_\lambda g \in D(T)$, increases with m if $g \geqslant 0$ and satisfies

$$(\lambda - T)f_m = g - (1 - \alpha_m)Kf_m.$$

As a result, $\|f - f_m\|_1 \to 0$ as $m \to \infty$. On the other hand,

$$\|g - (\lambda - T)f_m\|_1$$
$$= (1 - \alpha_m)\|Kf_m\|_1 \leqslant (1 - \alpha_m)\|f_m\|_\nu$$
$$\leqslant (1 - \alpha_m)\|\nu\|_\infty \|f_m\|_1,$$

which vanishes as $m \to \infty$. Consequently, $f \in D(G)$ and $(\lambda - G)f = g$. Moreover, $G = \overline{T}$.

(3) If $\nu(v)$ is integrable, then $\|L_\lambda K\| \leqslant \|LK\| < 1$ on $L_1(\Re, \nu\, dv)$. Then $1 \in \sigma_r(L_\lambda K)$ and Eq. (3.5) must be satisfied. This is also the case if KL_λ is weakly compact on $L_1(\Re, dv)$ [or $L_\lambda K$ is weakly compact on $L_1(\Re, \nu\, dv)$]. The reason is that power compact operators do not have a residual spectrum.

(4) In general, T is not a closed operator and hence G is a proper extension of T. This is, for instance, the case if $k(v, v') = \delta(v - v')$ and $\nu(v)$ is not essentially bounded. In this case $(Tf)(v) = -a(\partial f/\partial v)$ defined on $\mathscr{M}$ while $\mathscr{M}$ does not coincide with the (generally) larger domain of the generator of the semigroup $\{S_{0,\nu=0}(t)\}_{t \geqslant 0}$ on $L_1(\Re, dv)$ defined by $(S_{0,\nu=0}(t)g)(v) = g(v - at)$. However, if $\operatorname{spr}(L_\lambda K) < 1$ [which occurs, for instance, if KL_λ is weakly compact on $L_1(\Re, dv)$ or if $\nu(v)$ is integrable], we can easily prove that T_λ maps $L_1(\Re, dv)$ into $\mathscr{M}$ and therefore that $G = T$.

Suppose there is a nontrivial non-negative stationary solution φ in $L_1(\Re, dv)$ Then, as known, either $\varphi(v) > 0$ for all $v \in \Re$ or $\varphi(v) = 0$ for $v \leqslant v_0$ and $\varphi(v) > 0$ for $v > v_0$. In order to derive some properties of the semigroup $\{S(t)\}_{t \geqslant 0}$ in the latter case, we consider the free streaming semigroup $\{S_0(t)\}_{t \geqslant 0}$ on $L_1(\Re, dv)$ generated by the operator $T_0 = -(a(d/dv) + \nu(v))$ on the domain $D(T_0) = \mathscr{M}$. Then $\{S_0(t)\}_{t \geqslant 0}$ is a contraction semigroup whose generator satisfies

$$(\lambda - T_0)^{-1}g = L_\lambda g = \int_0^\infty e^{-\lambda t} S_0(t) g\, dt, \quad \operatorname{Re} \lambda > 0. \tag{3.6}$$

It is possible to write down $S_0(t)$ in closed form. In fact,

$$(S_0(t)g)(v) = M(t, v)g(v - at), \tag{3.7}$$

where

$$M(t, v) = \exp\left\{-\frac{1}{a}\int_{v - at}^{v} \nu(v')dv'\right\}. \tag{3.8}$$

Hence $\|S_0(t)\| = \operatorname{ess\,sup}\{M(t, v): v \in \Re\}$, so that the type $\omega_0(S_0)$ of the semigroup $\{S_0(t)\}_{t \geqslant 0}$ is given by

$$\omega_0(S_0) = \lim_{t \to \infty}(1/t)\log \operatorname*{ess\,sup}_{v \in \Re} M(t, v). \tag{3.9}$$

Since in an L_1 space the type and the spectral bound of a positive semigroup coincide (see Ref. 16), we may extend (3.6) to all $\operatorname{Re} \lambda > \omega_0(S_0)$. Writing $(R_\alpha g)(v) = \exp\{i\alpha v/a\}g(v)$ we have for $\lambda = \sigma + i\tau$ with $\sigma, \tau \in \Re$

$$L_\lambda = R_\tau^{-1} L_\sigma R_\tau,$$

while $\|L_\sigma g\|_1$ increases monotonically as σ decreases from $+\infty$ to $-\infty$ for all non-negative $g \in L_1(\Re, dv)$. Hence

$$\sigma(T_0) = \{\lambda \in \mathbb{C}: \operatorname{Re} \lambda \leqslant \omega_0(S_0)\}$$

whenever $\omega_0(S_0) > -\infty$, while $\sigma(T_0) = \varnothing$ whenever $\omega_0(S_0) = -\infty$. We now observe that

$$[(\lambda - G)^{-1} - (\lambda - T_0)^{-1}]g$$
$$= [T_\lambda - L_\lambda]g = \sum_{n=1}^{\infty}(L_\lambda K)^n L_\lambda g$$

implies that, for all $\operatorname{Re} \lambda > 0$, $[T_\lambda - L_\lambda]g = 0$ for all g of support within $[v_0, \infty)$. Since

$$[T_\lambda - L_\lambda]g = \int_0^\infty e^{-\lambda t}[S(t) - S_0(t)]g\,dt, \quad \mathrm{Re}\,\lambda > 0,$$

we find that $[S(t) - S_0(t)]g = 0$ for all g of support within $[v_0, \infty)$. For later use we also mention that

$$\lim_{t\to\infty} \|S_0(t)g\|_1 = 0, \quad g\in L_1(\mathfrak{R}, dv)$$

as a consequence of (3.7), (3.8) and the nonintegrability of $\nu(v)$.

We call $\{S(t)\}_{t\geq 0}$ *mean ergodic*[17,18] if for every $g\in L_1(\mathfrak{R}, dv)$ there exists a vector $Pg\in L_1(\mathfrak{R}, dv)$ such that

$$\lim_{t\to\infty} \left\| \frac{1}{t}\int_0^t S(t')g\,dt' - Pg \right\|_1 = 0. \tag{3.10}$$

It then follows that P is the (bounded) projection of $L_1(\mathfrak{R}, dv)$ onto the fixed space

$$\mathscr{F} = \{g\in L_1(\mathfrak{R}, dv): S(t)g = g \text{ for all } t\geq 0\}$$

of the semigroup $\{S(t)\}_{t\geq 0}$ along the space

$$\mathfrak{g} = \overline{\mathrm{span}\{[\mathbb{1} - S(t)]g: t>0, g\in L_1(\mathfrak{R}, dv)\}}.$$

Theorem 8: Suppose there is a nontrivial stationary solution φ in $L_1(\mathfrak{R}, dv)$. Then the semigroup $\{S(t)\}_{t\geq 0}$ is mean ergodic and the limit Pg is a one-dimensional projection of the form

$$(Pg)(v) = \alpha(g)\varphi(v), \quad v\in\mathfrak{R}, \tag{3.11}$$

where

$$\alpha(g) = \int_{-\infty}^{\infty} \psi(v')g(v')dv'$$

for some non-negative function $\psi\in L_\infty(\mathfrak{R}, dv)$ with $\|\psi\|_\infty < +\infty$ and $\int_{-\infty}^{+\infty}\psi(v')\varphi(v')dv' = 1$.

Proof: First, if G is the generator of $\{S(t)\}_{t\geq 0}$, then $\mathscr{F} = \{\varphi\in L_1(\mathfrak{R}, \nu\,dv): G\varphi = 0\}$. Thus, if $G = T$, then $\mathscr{F}$ coincides with the set of stationary solutions in $L_1(\mathfrak{R}, dv)$. Now recall that φ is continuous and suppose that φ does not have (finite) zeros. Then $0\leq S(t)\varphi = \varphi$ for all $t\geq 0$, and the mean ergodicity of $\{S(t)\}_{t\geq 0}$ is immediate from Ref. 18 (Corollary 1 of Theorem V 8.4).

Next, suppose φ has a finite zero. Then there exists $v_0\in\mathfrak{R}$ such that $\varphi(v)\equiv 0$ on $(-\infty, v_0]$ and $\varphi(v) > 0$ on (v_0, ∞). Then $(K\varphi)(v)\equiv 0$ on $(-\infty, v_0]$, and therefore $(Ku)(v)\equiv 0$ on $(-\infty, v_0]$ and for all characteristic functions u of compact support within (v_0, ∞). Since every nonnegative function in $L_1(\mathfrak{R}, \nu\,dv)$ of support within $[v_0, \infty)$ is the monotone limit of a sequence of finite linear combinations of characteristic functions of compact support within (v_0, ∞), we have $(Ku)(v)\equiv 0$ on $(-\infty, v_0)$ for all $u\in L_1(\mathfrak{R}, \nu\,dv)$ of support within $[v_0, \infty)$. Thus K leaves invariant the closed invariant ideal in $L_1(\mathfrak{R}, \nu\,dv)$ of functions with support in $[v_0, \infty)$. Then, by the second paragraph following the proof of Theorem 7, this must also be the case for $S(t)$. We may now restrict $S(t)$ to $L_1([v_0, \infty), dv)$ and apply the same corollary in Ref. 18 to get the ergodicity of the reduced semigroup. From the ergodicity of the reduced semigroup and the special form of φ we immediately have the ergodicity of the full semigroup $\{S(t)\}_{t\geq 0}$.

Finally, as the stationary problem has at most one linearly independent solution in $L_1(\mathfrak{R}, \nu\,dv)$, we easily obtain the specific form (3.11) of the projection P. □

IV. DECAY TO EQUILIBRIUM

In this section we shall prove that under certain quite natural conditions the solution of the time-dependent problem converges in the norm of $L_1(\mathfrak{R}, dv)$ to a solution of the stationary problem. Obviously, one of these conditions is that there exists a nontrivial stationary solution in $L_1(\mathfrak{R}, dv)$. The other condition is that the generator G of the time evolution semigroup $\{S(t)\}_{t\geq 0}$ of Eq. (1.1) does not have purely imaginary eigenvalues. Of course, the second condition is suggested by the fact that if $i\alpha$ is a purely imaginary eigenvalue of the generator G and g is a corresponding eigenfunction, then the solution of Eq. (1.1) with initial condition g has the form

$$S(t)g = e^{i\alpha t}g$$

and therefore does not converge at $t\to\infty$.

More specifically, if the generator G of the semigroup $\{S(t)\}_{t\geq 0}$ has purely imaginary eigenvalues, then under certain conditions one may prove that every solution $S(t)g$ of the time-dependent problem converges in the strong topology of $L_1(\mathfrak{R}, dv)$ to a periodic function (cf. Ref. 19, Theorem C IV 2.14). Indeed, suppose that $\lambda = 0$ is an isolated eigenvalue of G and that $i\alpha\in\sigma_p(T)$ for some nonzero real α. Let us also suppose that the distribution kernel $k(v,v')$ does not vanish on a set of positive measure, so that the semigroup $L_1(\mathfrak{R}, dv)$ is irreducible (i.e., does not have nontrivial closed invariant ideals). Then the spectrum of G on the imaginary line consists of a sequence $\{in\alpha\}_{n=-\infty}^{\infty}$ of simple eigenvalues. On denoting by Q a suitable projection of $L_1(\mathfrak{R}, dv)$ onto the closed linear span of the corresponding eigenfunctions, we obtain

$$\lim_{t\to\infty} \|S(t)g - e^{i\gamma t}Qg\|_1 = 0$$

for some period $\gamma > 0$.

Assume now that there are no purely imaginary eigenvalues and that a nontrivial stationary solution exists.

In order to establish the decay to equilibrium we shall apply the 0-2 law for positive semigroups in L_1 spaces (see Ref. 19, Theorem C IV 2.6 plus corollary), which may be formulated as follows. Let $\{S(t)\}_{t\geq 0}$ be a positive semigroup on the Banach space $L_1(E,\Sigma,\mu)$ and let $e(\mu)$ be a non-negative function in the kernel of its generator which does not vanish on a set of positive μ measure. Then for every $\tau > 0$ there exists a partition of E into two μ-measurable subsets $E_{0\tau}$ and $E_{2\tau}$ with the following properties:

(1) For every $t > 0$ the closed ideals of all functions in $L_1(E,\Sigma,\mu)$ having their support on $E_{0\tau}$ and $E_{2\tau}$, respectively, are invariant under $S(t)$.

(2) $|S(t) - S(t+\tau)|e_{0\tau}\downarrow 0$ as $t\to\infty$.

(3) $|S(t) - S(t+\tau)|e_{2\tau} = 2e_{2\tau}$ for all $t\geq 0$.

Here $e_{0\tau} = e\chi_{0\tau}$ and $e_{2\tau} = e\chi_{2\tau}$, where $\chi_{0\tau}$ and $\chi_{2\tau}$ denote the characteristic functions of $E_{0\tau}$ and $E_{2\tau}$, respectively.

Moreover, if the point spectrum $\sigma_p(G)$ of the generator G of $\{S(t)\}_{t\geq 0}$ satisfies $\sigma_p(G)\cap\{\mathrm{Re}\,\lambda = 0\} = \{0\}$, then

$S(t)g$ converges in $L_1(E,\Sigma,\mu)$ strongly as $t\to\infty$ for all $g\in L_1(E,\Sigma,\mu)$ that vanish on $E_{2\tau}$.

Theorem 9: Suppose there is a nontrivial stationary solution φ in $L_1(\mathfrak{R},dv)$, while G does not have purely imaginary eigenvalues. Then

$$\lim_{t\to\infty}\|S(t)g-Pg\|_1=0,\quad g\in L_1(\mathfrak{R},dv), \tag{4.1}$$

where P is the projection given by (3.11).

Proof: As a result of the previous section we may write

$$T_\lambda g=\sum_{n=0}^{\infty}(L_\lambda K)^n L_\lambda g,\quad \text{Re }\lambda>0.$$

From this equality we easily derive that the closed invariant ideals of T_λ in $L_1(\mathfrak{R},dv)$ are ideals of all functions in $L_1(\mathfrak{R},dv)$ that have their support in $[v_0,\infty)$ for some $v_0\in\mathfrak{R}$. This in turn implies that for all $\tau\geqslant 0$ one of the sets $E_{0\tau}$ and $E_{2\tau}$ in the 0-2 law has zero measure.

First suppose $E_{2\tau}=\mathfrak{R}$. Then $e_{0\tau}=0$ and $e_{2\tau}=\varphi$ and hence

$$|S(t)-S(t+\tau)|\varphi=2\varphi=\{S(t)+S(t+\tau)\}\varphi,\quad t\geqslant 0,$$

which is impossible. Indeed, there is a sequence of functions $g_n\in L_1(\mathfrak{R},dv)$ with $|g_n|\leqslant\varphi$ such that for every $\epsilon>0$ there exists $n_0\in\mathbb{N}$ such that for $n\geqslant n_0$ and for $t\geqslant 0$

$$(\{S(t)+S(t+\tau)\}\varphi)(v)$$
$$\leqslant(\{S(t)-S(t+\tau)\}g_n)(v)+\epsilon,\quad v\in\mathfrak{R}.$$

Writing $g_n^\pm=\sup(\pm g_n,0)$ we have

$$(S(t)[\varphi-g_n^+])(v)+(S(t)g_n^-)(v)$$
$$+(S(t+\tau)[\varphi-g_n^-])(v)+(S(t+\tau)g_n^+)(v)\leqslant\epsilon.$$

If $g_n^\pm=0$, then $|g_n|=g_n^\mp\leqslant\varphi$ yields $\varphi(v)\leqslant\epsilon$ for all $v\in\mathfrak{R}$, which is a contradiction for sufficiently small ϵ. Consequently, $E_{0\tau}=\mathfrak{R}$ for all $\tau>0$, which implies that for every $g\in L_1(\mathfrak{R},dv)$

$$\lim_{t\to\infty}\|S(t)g-Qg\|_1=0$$

for some vector Qg (cf. Ref. 19 Corollary to Theorem C IV 2.6). But then the inequality

$$\|S(\tau)-Qg\|_1\leqslant\|S(\tau)-\mathbf{1}\|\;\|S(t)g-Qg\|_1$$
$$+\|[S(t+\tau)-S(t)]g\|_1$$

implies $Q=P$, which completes the proof. □

Remarks: (1) If the generator $G\neq\overline{T}$, then there are no nontrivial stationary solutions (see Remark 1, after Theorem 7). On the other hand, if $G=T$, then T_λ is bounded as an operator from $L_1(\mathfrak{R},dv)$ into $L_1(\mathfrak{R},\nu\,dv)$. Then

$$T_\lambda=L_\lambda+L_\lambda KT_\lambda,\quad \lambda>0, \tag{4.2}$$

implies the Duhamel formula

$$S(t)=S_0(t)+\int_0^t S_0(t-\tau)KS(\tau)d\tau. \tag{4.3}$$

Now, if φ is an eigenfunction to the imaginary eigenvalue $i\lambda$ of T and hence $S(t)\varphi=e^{i\lambda t}\varphi$ and $S(t)|\varphi|=|\varphi|$ (see Nagel,[19] Corollary 2.3 on p. 297), we find, following an argument by Arlotti,[20]

$$e^{i\lambda t}\varphi=S_0(t)\varphi+\int_0^t e^{i\lambda\tau}S_0(t-\tau)K\varphi\,d\tau$$

and

$$|\varphi|=S_0(t)|\varphi|+\int_0^t S_0(t-\tau)K|\varphi|d\tau.$$

A simple comparison of the L_1 norms yields $\varphi\geqslant 0$ and $\lambda=0$. Hence if $G=T$ [which occurs if $\nu(v)$ is integrable or if $L_\lambda K$ is weakly compact on $L_1(\mathfrak{R},\nu\,dv)$], G does not have purely imaginary eigenvalues. Finally, if ν is essentially bounded, then $\|Kg\|_1\leqslant\|\nu\|_\infty\|g\|_1$ implies Eqs. (4.2) and (4.3). We may then repeat the above reasoning and conclude that $G=\overline{T}$ does not have purely imaginary eigenvalues.

(2) If we consider the case $k(v,v')=\delta(v-v')$ where $[S(t)g](v)=g(v-at)$ and hence $G=\overline{T}$, we see that $\varphi(v)=\exp\{-i\lambda v/a\}$ seemingly is an eigenfunction of $S(t)$ corresponding to the eigenvalue $e^{i\lambda t}$. However, $\varphi\notin L_1(\mathfrak{R},dv)$, though $\varphi\in L_1(\mathfrak{R},\nu\,dv)$ if ν is integrable.

ACKNOWLEDGMENTS

The authors are greatly indebted to Professor Luisa Arlotti for a number of illuminating discussions on the functional analysis of the problem.

This work was performed under the auspices of C.N.R.–G.N.F.M. and of the M.P.I. project "Equations of Evolution". C.v.d.M. is grateful to C.N.R.–G.N.F.M. for a visiting position and to the Department of Mathematics of the University of Pavia (Italy) for its hospitality. Part of the work of G.F. and S.L.P.F. on this project was performed during a visit to the Center for Transport Theory and Mathematical Physics, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061.

APPENDIX: AN ILLUSTRATIVE EXAMPLE

The object of this Appendix is to illustrate—with the help of two simple model problems—the typical patterns of behavior that one may expect from a population of charged particles moving through a host medium under the influence of a D.C. electric field. We shall consider two distinct versions of a simplified one-dimensional BGK model. A parallel and more sophisticated treatment has been proposed by Corngold and Rollins[7] who have adapted a one-dimensional Fokker–Planck model.

One simplified model is represented by the kinetic equation

$$\frac{\partial f}{\partial t}(v,t)+a\frac{\partial f}{\partial v}(v,t)=\nu(v)\{c(t)f_m(v)-f(v,t)\},$$
$$v\in\mathfrak{R},\quad t\geqslant 0, \tag{A1}$$

where

$$c(t)=\frac{\int_{-\infty}^{\infty}\nu(v)f(v,t)dv}{\int_{-\infty}^{\infty}\nu(v)f_m(v)dv}$$

is a normalization parameter, and $f_m(v)=\sqrt{\beta/\pi}\times\exp(-\beta v^2)$ is the normalized Maxwellian with

$$\langle v^2\rangle=\int_{-\infty}^{\infty}v^2 f_m(v)dv=(2\beta)^{-1}.$$

We consider the two following cases:

(i) $\nu(v) \equiv \nu_0 > 0$,

$$\text{(ii)}\ \ \nu(v) = \begin{cases} \nu_0, & -w \leqslant v \leqslant w, \\ 0, & |v| > w, \end{cases} \tag{A2}$$

where w and ν_0 are positive constants. Note that

$$\int_{-\infty}^{\infty} \nu(v)\,dv = +\infty$$

in case (i) whereas

$$0 < \int_{-\infty}^{\infty} \nu(v)\,dv = 2\nu_0 w < +\infty$$

in case (ii). Moreover, note that in case (ii) assumption (ii) is violated. We shall study the typical problem of the time evolution of a swarm of guest particles following the switching-on of the acceleration field at time $t = 0$, with $f(v,0) = f_m(v)$.

It is easy to establish the following results.

Case (i): Here $f(v,t)$ relaxes towards a steady profile. In fact, one can show that

$$f(v,t) = f_\infty(v) + \frac{a}{\nu_0}\exp(-r_0 t)\frac{d}{dv} f_\infty(v - at),$$

where

$$f_\infty(v) = (\nu_0/2a)\exp\{-\beta(v^2 - \lambda^2)\}\operatorname{erfc}(\lambda\sqrt{\beta}),$$
$$\text{with } \lambda = -v + (\nu_0/2a\beta)$$

and $f(v,t) \to f_\infty(v)$, as $t \to +\infty$.

Now we introduce the normalized velocity moments $\mu_k(t)$ which are defined by

$$\mu_k(t) = \int_{-\infty}^{\infty} v^k f(v,t)\,dv \left(\int_{-\infty}^{\infty} f(v,t)\,dv\right)^{-1}, \quad k = 0,1,2,\ldots .$$

Then it is easy to show that

$$\begin{aligned} \langle v\rangle(t) &= \mu_1(t) \\ &= (a/\nu_0)\{1 - \exp(-\nu_0 t)\} \sim a/\nu_0, \quad \text{as } t \to +\infty. \end{aligned}$$

Further

$$\begin{aligned} \tfrac{1}{2}\langle v^2\rangle(t) &= \mu_2(t) \\ &= (1/4\beta) + (a/\nu_0)^2\{1 - (1 + \nu_0 t) \\ &\qquad \times \exp(-\nu_0 t)\} \sim (1/4\beta) + (a/\nu_0)^2, \quad \text{as } t \to +\infty. \end{aligned}$$

Accordingly, for thermal agitation (relative to average velocity) we have

$$\tfrac{1}{2}(\langle v^2\rangle(t) - \langle v\rangle^2(t)) \sim 1/4\beta + \tfrac{1}{2}(a/\nu_0)^2, \quad \text{as } t \to +\infty.$$

In this case there is no runaway process.

Case (ii): For problem (A1) subject to (A2) and to the initial condition $f(v,0) = f_m(v)$, it is obvious that, within the quadrant $v \geqslant w$, $t \geqslant 0$, the solution $f(v,t)$ remains constant along the characteristics $v = \bar{v} + as$, $t = s$ $(s \geqslant 0)$. Accordingly, we can write

$$f(\bar{v} + at,t) = g_t(\bar{v}) = \begin{cases} f_m(\bar{v}), & \bar{v} \geqslant w \\ f(w,(w - \bar{v})/a), & w - at < \bar{v} < w, \end{cases}$$

where $t \geqslant 0$. Therefore,

$$f(\bar{v} + at,t) \sim g_\infty(\bar{v}), \quad \text{as } t \to +\infty, \quad \bar{v} \in \Re,$$

i.e., there is convergence towards a travelling wave. Note that this is not an explicit expression.

We can summarize the results as follows. Under case (i) conditions [for which $\int_{-\infty}^{\infty} \nu(v)\,dv = +\infty$] there are no runaways and the distribution function relaxes towards an asymptotic profile $f_\infty \in L_1(\Re,dv) \cap L_1(\Re,\nu\,dv)$ whose velocity moments are finite. Note that, in this case, if cold charged particles were fed continuously into the system, then the distribution function would not relax towards a steady state value. However, in both situations the velocity moments would relax towards finite values.

On the contrary, in case (ii) [for which $0 < \int_{-\infty}^{\infty} \nu(v)\,dv < +\infty$] $f(v,t)$ converges towards a "travelling wave" and all velocity moments diverge as $t \to +\infty$. Under steady feeding, the velocity moments would diverge as $t \to +\infty$, whereas $f(v,t)$ would converge to a steady profile belonging to $L_1(\Re,\nu\,dv)$; however, this profile would not belong to $L_1(\Re,dv)$.

[1]K. Kumar, H. R. Skullerud, and R. E. Robson, Aust. J. Phys. **33**, 343 (1980); K. Kumar, Phys. Rep. **112**, 319 (1984).

[2]G. Cavalleri and S. L. Paveri-Fontana, Phys. Rev. A **6**, 327 (1972).

[3]H. Dreicer, Phys. Rev. **115**, 238 (1959); **117**, 329 (1960).

[4]E. M. Lifshitz and L. P. Pitaesvkii, *Physical Kinetics*, Volume 10 of the Landau and Lifshitz "Course of Theoretical Physics" (Pergamon, New York, 1981).

[5]V. V. Parail and O. P. Pogutse, "Runaway electrons in a plasma," *Reviews of Plasma Physics*, edited by M. A. Leontovich (Consultants Bureau, New York, 1986), Vol. 11.

[6]M. C. Mackey, Biophys. J. **11**, 75 (1971).

[7]N. Corngold and D. Rollins, Phys. Fluids **29**, 1042 (1986); **30**, 393 (1987).

[8]M. Kač, "Foundations of kinetic theory," in *Proceedings of the third Berkeley Symposium on Mathematical Statistics and Probability*, edited by J. Neyman (University of California Press, Berkeley, 1956), Vol. III, pp. 171–197.

[9]W. Greenberg, C. V. M. van der Mee, and V. Protopopescu, *Boundary Value Problems in Abstract Kinetic Theory* (Basel, Birkhäuser, 1987), Vol. 23.

[10]R. Beals and V. Protopopescu, J. Math. Anal. Appl. **121**, 370 (1987).

[11]L. Arlotti, J. Diff. Eqs. **69**, 166 (1987).

[12]C. Cosner, S. M. Lenhart, and V. Protopopescu, SIAM J. Math. Anal. **19**, 797 (1988).

[13]M. A. Krasnoselskii, *Positive Solutions of Operator Equations* (Groningen, Noordhoff, 1964) [Fizmatgiz, Moscow, 1962 (Russian)].

[14]R. E. Robson, Aust. J. Phys. **28**, 523 (1975); **29**, 171 (1976).

[15]R. A. Adams, *Sobolev Spaces* (Academic, New York, 1975).

[16]R. Derndinger, Math. Z. **172**, 281 (1980).

[17]E. B. Davies, *One-parameter Semigroups* (Academic, New York, 1980).

[18]H. H. Schaefer, *Banach Lattices and Positive Operators*, Grundlehren Math. Wiss. (Springer, Berlin, 1974), Vol. 215.

[19]*One-parameter Semigroups of Positive Operators, Lecture Notes in Mathematics*, Vol. 1184, edited by R. Nagel (Springer, Berlin, 1986).

[20]L. Arlotti, Proceedings of the Conference on Transport Theory, Invariant Imbedding, and Integral Equations in honor of G. M. Wing's 65th birthday, Santa Fe, 1988, to appear.